Conditional Mean and Maximum Likelihood Approaches to Multiharmonic Frequency Estimation

Ben James, Brian D. O. Anderson, *Fellow, IEEE*, and Robert C. Williamson

***Abstract*— The performance of an extended Kalman filter (EKF) applied to the problem of estimating the (assumed constant) parameters (fundamental frequency, harmonic phases, and amplitudes) of a complex multiharmonic signal measured in noise is shown to be asymptotically (i.e., as the number of measurements tends to infinity) efficient. The Cramer–Rao (CR) bounds associated with the estimation problem are derived for the case where the measurements commence at an arbitrary time distinct from zero.**

I. Introduction

This paper is concerned with the discrete-time estimation of the parameters of a signal comprising a sum of harmonically related sinusoidal tones contaminated by additive noise, given noisy observations of the tones over a finite interval. Such a harmonic series is parametrized by its fundamental frequency and its harmonic phases (and amplitudes). The amplitudes, frequency, and phases are assumed not to vary over the interval of observation.

The work of [1] and [2] analyzed the maximum likelihood approach applied to a closely related parameter estimation problem, with consideration in [1] of the single-tone case and in [2] the general multiple-tone case. The latter did not consider the special situation where the tones are harmonically related, but rather one where no special relationship between the tones exists. The work of [3] contained, in part, a derivation of the Cramer–Rao (CR) bounds for the case of a real multiharmonic signal, measurements of which are assumed to commence at time $t = 0$. (The CR bounds are lower bounds on the estimation error variance for any unbiased estimator.) One of the results of this paper is the generalization of the derivation of [3] to the case of a complex multiharmonic signal, with noisy measurements assumed to start at an arbitrary time distinct from zero. Reasons for preferring a complex signal formulation are discussed in Section II. The details of our slightly generalized derivation are summarized in Theorem 4.1. (We remark that this theorem is essentially no different from that which could be condensed from the treatment of [3] in which no explicit theorem statement is made.)

The statement of Theorem 4.1 sets the scene for the major result of this paper, which is primarily concerned with a particular approximate conditional mean estimator, the extended Kalman filter (EKF). The estimation problem is cast in state-space form, and two parameters are defined which reflect the knowledge of the frequency and phases prior to estimation. The so-called *information* formulation of the EKF equations is applied to the resulting state-space signal model, and, after derivation of approximate expressions governing the performance of the EKF, it is argued (through appeal to Theorem 4.1) that the EKF is asymptotically efficient for sufficiently high signal-to-noise-ratio (SNR). Finally, conclusions are drawn and directions of future research are discussed.

II. Problem Formulation

The estimation problem with which we are concerned is to determine the "best" estimate (in some sense) of a constant, but unknown, parameter vector, given a finite set of noisy observations of some function of the parameter vector.

For the multiharmonic (MH) estimation problem with m harmonics, the parameters of interest are the amplitudes $b_1, \cdots, b_m$ of each of the harmonics, their relative phases $\theta_1, \cdots, \theta_m$, and the fundamental frequency ω_0. The parameter vector is then defined to be

$$\alpha_0 \triangleq [b_1 \cdots b_m \, \omega_0 \, \theta_1 \cdots \theta_m]^T \tag{2.1}$$

and an arbitrary estimate $\hat{\alpha}$ of α_0 is defined by

$$\hat{\alpha} \triangleq [\hat{b}_1 \cdots \hat{b}_m \, \hat{\omega} \, \hat{\theta}_1 \cdots \hat{\theta}_m]^T. \tag{2.2}$$

The underlying real signal comprising m harmonics is a nonlinear function of the parameter vector α_0 and is defined by

$$s_{\alpha_0}(t) \triangleq \sum_{k=1}^{m} b_k \cos(k\omega_o t + \theta_k) \tag{2.3}$$

along with its *in-quadrature* counterpart (perhaps obtained via a Hilbert transform: see Appendix A and the remark below):

$$\check{s}_{\alpha_0}(t) \triangleq \sum_{k=1}^{m} b_k \sin(k\omega_o t + \theta_k). \tag{2.4}$$

The noisy measurements are complex and are defined by

$$Z_n \triangleq X_n + jY_n, \qquad 0 \le n \le N-1 \tag{2.5}$$

Manuscript received March 13, 1991; revised April 14, 1993. This work was supported by the Defence Science and Technology Organization, the Australian Telecommunications and Electronics Research Board, the ANU Centre for Information Science Research, and the Cooperative Research Centre for Robust and Adaptive Systems. The associate editor coordinating the review of this paper and approving it for publication was Dr. David Rossi.

B. James is with the Industrial Systems Group, Department of Electrical and Electronic Engineering, Imperial College of Science, Technology and Medicine, London, SW7 2BT, England.

B. D. O. Anderson and R. C. Williamson are with the Department of Systems Engineering, Research School of Physical Sciences and Engineering, Australian National University, Canberra, A.C.T., Australia.

IEEE Log Number 9400385.

where

$$X_n \triangleq s_{\alpha_0}(t_0 + nT) + w(t_0 + nT) \quad (2.6a)$$

$$Y_n \triangleq \check{s}_{\alpha_0}(t_0 + nT) + \check{w}(t_0 + nT). \quad (2.6b)$$

The time at which measurements commence is denoted by t_0. The measurement noise processes w and $\check{w}$ are assumed to be independent, zero mean, white, and Gaussian with variance σ^2.

Remark: There are two basic reasons for preferring a formulation of the estimation problem in terms of the complex multiharmonic signal defined by (2.5)–(2.4). (Note that Z_n as defined in (2.5) is sometimes referred to as the *analytic* signal.) The first is that it provides a degree of analytical simplification that facilitates subsequent analysis not afforded by the corresponding real signal formulation. (In the real signal formulation, the imaginary part of the RHS of (2.5) evanesces.) The second is simply a desire to remain consistent with existing treatments of the problem, particularly those of [1], [2], and [4]. It should be stressed that there is essentially no loss of generality associated with our appeal to the complex model; it is first and foremost a matter of analytical convenience. Having said this, it is perhaps nevertheless desirable that some connection be made between the complex and real formulations since, in practice, it is only the real signal X_n that is available for measurement. This we do in Appendix A, where a particular means of constructing Y_n from the received real signal is described. This may be summarized by saying that the underlying continuous-time received signal is sampled at twice the rate associated with X_n (i.e., with sampling period $T/2$) and passed through an ideal discrete-time Hilbert transformer, the output of which is downsampled to give Y_n.

III. Maximum Likelihood Estimation

The estimation problem generally is that of determining an estimate $\hat{\alpha}$ of α_0 given the N measurements $Z_0, \cdots, Z_{N-1}$ that is "best" in some sense. The ML approach is just one way of assigning meaning to the word "best." More specifically, the maximum likelihood (ML) estimate $\hat{\alpha}_{ml}$ is that parameter vector which most likely would result in the observed measurement sequence. Mathematically, this is expressed by

$$\hat{\alpha}_{ml} \triangleq \arg\max f(Z_0, \cdots, Z_{N-1}|\hat{\alpha}) \quad (3.1)$$

where the conditional probability density function f is given by

$$\begin{aligned} f(Z_0, \cdots, Z_{N-1}|\alpha) &= \left(\frac{1}{2\pi\sigma^2}\right)^{\hat{N}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}\right. \\ &\cdot [(X_n - s_{\hat{\alpha}}((n+n_0)T))^2 \\ &\left. + (Y_n - \check{s}_{\hat{\alpha}}((n+n_0)T))^2]\right\}. \end{aligned} \quad (3.2)$$

Formulas for the ML estimate of the amplitude, phase, and frequency are given for the single-tone case ($m = 1$) in [1]. These vary depending upon whether the phase is known or unknown, and similarly for the frequency and amplitude. This paper is concerned with equivalent bounds on the MLE performance for the multiharmonic case that are known to be tight in the linear or above-threshold region.

IV. Cramer–Rao Bounds

In this section, the Cramer–Rao bounds for the complex multiharmonic parameter estimation problem formulated earlier are derived. These bounds represent the best performance achievable by any unbiased estimator applied to the measurements $Z_0, \cdots, Z_{N-1}$. They also enable us to define the above-threshold or linear region as those values of the signal-to-noise ratio (SNR) for which an estimator's performance meets the CR bounds, and the below-threshold or nonlinear region as those values for which it does not.

As already remarked in the Introduction, the CR bounds for a *real* multiharmonic signal have been calculated in [3]. This derivation assumed that measurements commenced at $t_0 = 0$ and that the received signal is real, rather than complex. For the sake of completeness (many of the calculations here appear closely related to those for the EKF performed later in the paper) and consistency with [1] (where the CR bounds for a single complex tone are presented), we repeat the calculations for the more general case of a complex multiharmonic signal with measurements assumed to commence at an arbitrary time $t_0 = n_0 T$.

To simplify matters, the harmonic amplitudes are intially assumed known. It proves straightforward to subsequently generalize this treatment to the case where the harmonic amplitudes are unknown.

Let $Z \triangleq \{Z_0, \cdots, Z_{N-1}\}$. The Fisher Information matrix J is defined (see [5]) by

$$J(\alpha) \triangleq E\left\{\left[\frac{\partial \ln f(Z|\alpha)}{\partial \alpha}\right]\left[\frac{\partial \ln f(Z|\alpha)}{\partial \alpha}\right]^T\right\} \quad (4.1)$$

and the Cramer–Rao bounds are the diagonal elements of $J^{-1}(\alpha_0)$. As in [1], the elements of J are given by

$$\begin{aligned} J_{ij}(\alpha) = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}&\left[\frac{\partial s_\alpha(t_0+nT)}{\partial \alpha_i}\frac{\partial s_\alpha(t_0+nT)}{\partial \alpha_j}\right. \\ &\left.+ \frac{\partial \check{s}_\alpha(t_0+nT)}{\partial \alpha_i}\frac{\partial \check{s}_\alpha(t_0+nT)}{\partial \alpha_j}\right]. \end{aligned} \quad (4.2)$$

After some calculation (Appendix B), there holds

$$J(\alpha_0) = \overline{J} + \tilde{J} \quad (4.3)$$

where

$$\overline{J} = \begin{pmatrix} A & B \\ B^T & C \end{pmatrix} \quad (4.4)$$

$$A = \frac{T^2(Q + 2n_0 P + n_0^2 N)}{\sigma^2}\sum_{k=1}^{m} k^2 b_k^2 \quad (4.5a)$$

$$B = \frac{T(n_0 N + P)}{\sigma^2}[b_1^2 \; 2b_2^2 \cdots m b_m^2] \quad (4.5b)$$

$$C = \frac{N}{\sigma^2}\,\mathrm{diag}\,\{b_1^2, b_2^2, \cdots, b_m^2\} \quad (4.5c)$$

and $\tilde{J}$ is a perturbation term, the magnitude of which is given by Theorem 4.1 below. The quantities P and Q are defined as in [1]:

$$P \triangleq \sum_{n=0}^{N-1} n = \frac{N(N-1)}{2} \tag{4.6}$$

$$Q \triangleq \sum_{n=0}^{N-1} n^2 = \frac{N(N-1)(2N-1)}{6}. \tag{4.7}$$

Let $O(x)$ denote a matrix whose components are asymptotically linear in x, i.e., $\lim_{x\to\infty} |O_{ij}(x)|/x = c$, where $0 < c < \infty$. The following theorem (which describes in a compact form the derivation of the CR bounds in [3]) permits approximation of the CR bounds valid for large N.

Theorem 4.1 (Nehorai and Porat [3]): Let $J(\alpha_0), \overline{J}$, and $\tilde{J}$ be as previously defined. Then

1)

$$\tilde{J} = J(\alpha_0) - \overline{J} = \begin{pmatrix} O(N^2) & O(N) \\ O(N) & O(1) \end{pmatrix} \tag{4.8}$$

[where $\tilde{J}$ is partitioned conformally with the RHS of (4.4)].

2. With $\tilde{J}$ as given in (4.8), $\|J^{-1}(\alpha_0) - \overline{J}^{-1}\| \to 0$ as $N \to \infty$.

Proof: See Appendix B.

Hence, for sufficiently large N, the inverse of the Fisher information matrix $J(\alpha_0)$ is closely approximated by $\overline{J}^{-1}$. (Note that the theorem holds for the general case, where both the frequency and the harmonic phases are assumed unknown.)

The diagonal nature of C permits ready calculation of the diagonal elements of $\overline{J}^{-1}$ when *all* the components of α_0 are unknown (the most general case) by a straightforward application of the Block Matrix Inversion Lemma (see Appendix B). The task is simplified further if the frequency is unknown and all the harmonic phases are assumed known, in which case $\overline{J}$ reverts to a scalar, namely, A. On the other hand, if the frequency is assumed known and the harmonic phases unknown, then $\overline{J}$ becomes an $m \times m$ diagonal matrix, namely C. Summarized below are the bounds for each of the possible combinations.

Performance Bounds for Unbiased Frequency Estimators:

Phases known: $\mathrm{var}(\hat{\omega})$

$$\geq \frac{\sigma^2}{T^2(Q + 2n_0P + n_0^2N)\left(\sum_{k=1}^{m} k^2 b_k^2\right)} \tag{4.9}$$

Phases unknown: $\mathrm{var}(\hat{\omega})$

$$\geq \frac{12\sigma^2}{T^2N(N^2-1)\left(\sum_{k=1}^{m} k^2 b_k^2\right)}. \tag{4.10}$$

Performance Bounds for Unbiased Phase Estimators:

Frequency known: $\mathrm{var}(\hat{\theta}_k)$

$$\geq \frac{\sigma^2}{Nb_k^2}, \qquad 1 \leq k \leq m \tag{4.11}$$

Frequency Unknown: $\mathrm{var}(\hat{\theta}_k)$

$$\geq \frac{\sigma^2}{Nb_k^2} + \frac{3k^2(N-1)\sigma^2}{N(N+1)\left(\sum_{k=1}^{m} k^2 b_k^2\right)} \qquad 1 \leq k \leq m. \tag{4.12}$$

A. Effect of Amplitude Uncertainty

The foregoing assumed that the harmonic amplitudes $b_1, \cdots, b_m$ are known. How are the bounds affected if this assumption is relaxed? In this case, the parameter vector is $\alpha_0 \triangleq [b_1 \cdots b_m\, \omega\, \theta_1 \cdots \theta_m]^T$. The approximation to the Fisher information matrix $\overline{J}$ becomes the $(2m+1) \times (2m+1)$ matrix below:

$$\overline{J} = \begin{pmatrix} \frac{N}{\sigma^2} I_m & 0 & 0 \\ 0 & A & B \\ 0 & B^T & C \end{pmatrix}. \tag{4.13}$$

From the block-diagonal nature of $\overline{J}$, it is easily seen that the bounds for phase and frequency variation are unaffected by uncertainty in the harmonic amplitudes. Similarly, the bounds for amplitude estimation are unaffected by certainty or uncertainty in either the phases or frequency. The amplitude estimation bounds are given as follows.

Bounds for Amplitude Estimation:

$$\mathrm{var}(\hat{b}_k) \geq \frac{\sigma^2}{N}, \qquad 1 \leq k \leq m. \tag{4.14}$$

Remarks:

- The multiharmonic frequency estimation error variance bounds are identical to the single-tone bounds given in [1] when one makes the replacement $b_0^2 \to \Sigma_{k=1}^{m} k^2 b_k^2$. (Reference [1] defined b_0 to be the amplitude of the single tone.)
- It is apparent that the frequency estimation bounds for a multiharmonic signal are lower than the corresponding bounds for a single sinusoidal tone with the same signal power (i.e., with $b_0^2 = \Sigma_{k=1}^{m} b_k^2$). This appears to confirm the intuitive expectation that the presence of harmonics should improve the accuracy of the (fundamental) frequency estimate over that for a single tone.

V. Performance of Extended Kalman Filter

In this section, an equivalent state-space formulation of the discrete-time estimation problem is given. The state-space description of the complex multiharmonic signal is, of course, nonlinear. Standard linear, optimal filtering techniques are not therefore applicable. A technique based upon, and for sufficiently high SNR virtually equivalent to, the standard Kalman filter is the so-called extended Kalman filter (EKF). In essence, this object is simply a Kalman filter applied to a linearized version of the original signal model, the linearization being about the current state estimate. Provided the error incurred by the linearization is not too great, the EKF generates an estimate of the state close to the optimal conditional mean estimate.

The work of [6]–[8] dealt with the application of the EKF to the multiharmonic *tracking* problem (i.e., where α_0 is time varying). Some of the analytical techniques used in the tracking problem prove fruitful for the estimation problem, and conversely.

The state-space formulation of the multiharmonic estimation problem proceeds as follows. First, the measured multiharmonic signal is rewritten in the equivalent vector form below. (We remark that this formulation is unconventional, although nonetheless valid, in the sense that state-space signal models most commonly deal with real, scalar signals.)

$$Z((n+n_0)T) = \sum_{k=1}^{m} \begin{bmatrix} b_k \cos \Theta_k((n+n_0)T) \\ b_k \sin \Theta_k((n+n_0)T) \end{bmatrix} + \begin{bmatrix} w((n+n_0)T) \\ \tilde{w}((n+n_0)T) \end{bmatrix}, \qquad 0 \le n \le N-1. \tag{5.1}$$

The kth total harmonic phase $\Theta_k(nT)$ is defined (for the case of the multiharmonic signal with constant parameter vector α_0) by

$$\Theta_k(nT) \triangleq k\omega_0 nT + \theta_k. \tag{5.2}$$

Also, as before, n_0 is the point in time at which measurements commence. The sequence $W(nT) \triangleq [w(nT)\ \tilde{w}(nT)]^T$ defines a vector random process (the measurement noise) that is white, zero mean, and Gaussian, with covariance matrix

$$R \triangleq \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}. \tag{5.3}$$

The state vector $x(nT)$ is defined as follows:

$$x(nT) \triangleq [\omega(nT)\ \Theta_1(nT) \cdots \Theta_m(nT)]^T. \tag{5.4}$$

Again, the harmonic amplitudes are assumed known. (This assumption will be later relaxed.) The complex multiharmoinc signal model is then given by

$$x((n+1)T) = Fx(nT) \tag{5.5a}$$
$$Z(nT) = h[x(nT)] + W(nT) \tag{5.5b}$$

where

$$F \triangleq \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ T & 1 & 0 & & 0 \\ 2T & 0 & 1 & & 0 \\ \vdots & & & \ddots & 0 \\ mT & & & & 1 \end{pmatrix} \tag{5.6}$$

and the definition of $h(\cdot)$ is apparent from (5.1).

It is evident from the definition of the system matrix F that the frequency $\omega(nT)$ (the first component of the state vector) is constant in time[1] (and equal to the fundamental frequency ω_0, assuming correct initialization) due to the absence of a driving term from (5.5a). [We remark that a state-space formulation of the *tracking* problem is easily achieved by the inclusion of a stochastic driving term in (5.5a).]

[1] State variables are, in general, time varying and the usual notation reflects this. Our use of the notation is in compliance with convention only, and is not intended to imply that the fundamental frequency is time varying when, in our case, it patently is not.

The difference equation (5.5a) is initialized at time $n = n_0$ in such a way that $x(0) = \alpha_0$. (Recall that α_0 is the quantity to be estimated.) In the absence of a stochastic driving term (normally referred to as the *input* or *process* noise) from (5.5a), and given that F is nonsingular, this is achieved by setting

$$x(n_0T) = \begin{cases} F^{n_0}\alpha_0, & n_0 > 0 \\ (F^{-1})^{-n_0}\alpha_0, & n_0 < 0 \end{cases}. \tag{5.7}$$

The estimation problem may now be stated in terms of the state-space formulation of (5.5a)–(5.5b):

Determine the "best" estimate of $x(0) = \alpha_0$ in some sense, given the sequence of measurements $Z(n_0T), \cdots, Z((n_0+N-1)T)$.

The "best" estimate, in the sense of minimum mean-square error, is that provided approximately by the EKF, with small approximation error for sufficiently high SNR. (Of course, there are other, more complex nonlinear estimators that will give close to optimal performance at lower values of SNR.)

Some clarification is required here. In general, the "best" estimate of $x(0)$ given the N measurements commencing at $n = n_0$ is a *smoothed* estimate in the sense that, for arbitrary n_0, the time of interest $n = 0$ does not necessarily coincide with the time of conclusion of measurements $n = n_0 + N - 1$. The EKF, however, provides a *filtered* state estimate: that is, an estimate at time k based on measurements up to and including time k. This poses no difficulty for the problem at hand since the absence of input noise and the nonsingularity of F means that the filtered estimate of $x((n_0+N-1)T)$ is identical to the smoothed estimate of $x(0)$ to within an invertible linear transformation [of the same nature as that described in (5.7)].

For the signal model given by (5.5a)–(5.5b), the equations defining the EKF (we have omitted T for the sake of convenience) are given by [see [9]]

$$\hat{x}(k|k) = \hat{x}(k|k-1) + L(k)\{Z(k) - h[\hat{x}(k|k-1)]\} \tag{5.8a}$$
$$\hat{x}(k+l|k) = F^l\hat{x}(k|k), \qquad l \ge 0 \tag{5.8b}$$

where

$$L(k) \triangleq \Sigma(k|k-1)H(k)\Omega^{-1}(k) \tag{5.9a}$$
$$\Omega(k) \triangleq H^T(k)\Sigma(k|k-1)H(k) + R \tag{5.9b}$$

and

$$\Sigma(k|k) = \Sigma(k|k-1) - \Sigma(k|k-1)H(k)\Omega^{-1}(k)H^T(k) \cdot \Sigma(k|k-1) \tag{5.10a}$$
$$\Sigma(k+l|k) = F^l\,\Sigma(k|k)(F^T)^l, \qquad l \ge 0 \tag{5.10b}$$

along with

$$H^T(k) \triangleq \left.\frac{\partial h(x)}{\partial x}\right|_{x=\hat{x}(k|k-1)} \tag{5.11}$$

The EKF so defined is an *exact* Kalman filter for a particular linearized version of the nonlinear model. It is, of course, no longer optimal when applied to signals generated by the original signal model. The quantities $\Sigma(k|k-1)$ and $\Sigma(k|k)$ therefore denote *approximations* to the actual prior and posterior conditional error covariance matrices; the region where these approximations are very close to the true covariances is termed the linear region.

This paper is not concerned with the actual estimates *per se*, but rather with their associated error covariances. The question to be answered is: Does the performance of the EKF in its linear region meet the limits imposed by the CR bounds? In other words, do the EKF and the MLE perform equally well (in terms of the accuracy of their estimates) in their respective linear regions?

We will attempt to answer this question by calculating the approximate covariance as defined in (5.10a) and (5.10b). To ensure a fair comparison, assumptions concerning prior knowledge of the parameter vector underlying the ML approach must be reflected in the EKF formulation. To this end, define the initial error covariance

$$\Sigma_s(0) \triangleq E\{[\hat{\alpha}-x(0)][\hat{\alpha}-x(0)]^T\} = \begin{pmatrix} a & 0 \\ 0 & \epsilon I_m \end{pmatrix} \qquad (5.12)$$

where $\hat{\alpha}$ is an initial estimate of $x(0)$ in the absence of measurements, i.e., an estimate based on prior knowledge of $x(0)$. [Either $\hat{\alpha}$ or a propagated version thereof would be used to initialize the EKF estimate update equation of (5.8a).]

The initial error covariance $\Sigma_s(0)$ reflects knowledge of the parameters prior to estimation. Thus, for example, the problem of estimating a completely uncertain ω_0 when each of the harmonic phases is perfectly known corresponds to letting $a \to \infty$ and $\epsilon \to 0$. The appropriate manipulation of the two parameters a and ϵ in this fashion enables a fair comparison for each of the situations in (4.9)–(4.12). Our goal is then to show that the estimation error variances associated with the estimate $\hat{x}(0)$ meet the bounds for each of the limiting situations there specified.

VI. Calculation of EKF Estimation Error Covariance

Define

$$\hat{x}(0) = E[x(0)|Z(n_0),\cdots,Z(n_0+N-1)]. \qquad (6.1)$$

The quantity of interest is the *smoothed* conditional error covariance

$$E\{[\hat{x}(0)-x(0)][\hat{x}(0)-x(0)]^T |Z(n_0),\cdots,Z(n_0+N-1)\}, \qquad (6.2)$$

an approximation to which is given by a suitable transformation of the approximate *filtered* error covarinace obtained by solution of the difference equation (5.10a) [along with (5.10b)]. (See earlier comments regarding filtered versus smoothed estimates.)

The definition of the EKF given in (5.8a)–(5.11) is one expressed in terms of covariance matrices. A dual formulation is that of the so-called information filter, which re-expresses the defining equations of the EKF in terms of the information matrices, which are inverses of the covariance matrices. This formulation will be preferred for three main reasons.

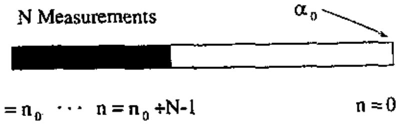

Fig. 1. Relationship of measurements to point of interest.

- The form of the so-called *measurement update step* [see (5.10a)] is generally simpler for the information filter than in the dual covariance formulation.
- The absence of input or process from the signal model ensures that the *time update step* [see (5.10b)] has the same simple form in both formulations.
- For the case most commonly encountered in practice, where the initial error covariance $\Sigma_s(0)$ reflects complete uncertainty in the frequency and phases (i.e., $a \to \infty, \epsilon \to \infty$), the equations of the covariance formulation would be initialized by a matrix containing infinite entries. This is avoided in the information formulation.

For the sake of simplicity and to avoid clouding the main issues, the assumption is made that $n_0 \leq 1-N$; that is, the measurements commence and conclude before the time of interest, $n = 0$ (see Fig. 1). Under this assumption, the approximate information matrices associated wiht a filter initialized at time $n = n_0$ and running forwards in time to $n = n_0+N-1$ are calculated. We claim that this assumption does not entail any loss of generality, although this is not proved here. (The claim is based on the reversible, symmetric nature of the signal model from which input noise is absent.)

VII. Information Filter

Define the measurement-update and time-update information matrices as follows:

$$\Gamma(k|k) \triangleq \Sigma^{-1}(k|k) \qquad (7.1a)$$

$$\Gamma(k|k-1) \triangleq \Sigma^{-1}(k|k-1). \qquad (7.1b)$$

The respective update equations are [cf. (5.10a) and (5.10b)]

$$\Gamma(k|k) = \Gamma(k|k-1) + H(k)R^{-1}H^T(k) \qquad (7.2a)$$

$$\Gamma(k+l|k) = (F^{-T})^l\Gamma(k|k)(F^{-1})^l, \qquad l \geq 0. \qquad (7.2b)$$

From (7.2a) and (7.2b), it is evident that the measurement-update (or *posterior*) information matrix satisfies the Lyapunov difference equation

$$\Gamma(k+1|k+1) = F^{-T}\Gamma(k|k)F^{-1} + H(k)R^{-1}H^T(k). \qquad (7.32)$$

The difference equation is initialized as follows. The initial error covariance at time $n = 0, \Sigma_s(0)$ is, in the absence of measurements, simply the propagation of the initial error covariance $\Sigma(n_0|n_0-1)$ to time $n = 0$. From (5.10b), this is given by

$$\Sigma_s(0) = F^{-n_0}\Sigma(n_0|n_0-1)(F^T)^{-n_0}. \qquad (7.4)$$

Therefore,

$$\Gamma(n_0|n_0-1) = (F^T)^{-n_0}\Sigma_s^{-1}(0)F^{-n_0} \qquad (7.5)$$

with the result that, from (7.2a),

$$\Gamma(n_0|n_0) = (F^{-T})^{n_0}\Sigma_s^{-1}(0)F^{-n_0} + H(n_0)R^{-1}H^T(n_0). \qquad (7.6)$$

In terms of the information formulation definitions so far, the quantity defined in (6.2) is given by

$$\Sigma(0|n_0+N-1) = (F)^{1-N-n_0}\Gamma^{-1}(n_0+N-1|n_0+N-1)(F^T)^{1-N-n_0}. \qquad (7.7)$$

VIII. Solution of Lyapunov Equation

From the definition of $H(\cdot)$ in (5.11) and $h(\cdot)$ in (5.5b) and (5.1), a straightforward calculation gives

$$H(k) = \begin{bmatrix} 0 & 0 \\ -b_1\sin\hat{\Theta}_1 & b_1\cos\hat{\Theta}_1 \\ \vdots & \vdots \\ -b_m\sin\hat{\Theta}_m & b_m\cos\hat{\Theta}_m \end{bmatrix} \qquad (8.1)$$

where $\hat{\Theta}_l \triangleq \hat{\Theta}_l(k|k-1), 1 \le l \le m$.

From (8.1), it is easy to see that the $(i+1),(j+1)$th component of $H(\cdot)R^{-1}H^T(\cdot)$ is given by $(1/\sigma^2)b_ib_j\cos(\hat{\Theta}_i - \hat{\Theta}_j)$. The diagonal elements of $H(\cdot)R^{-1}H^T(\cdot)$ are therefore known and constant [they are equal to the diagonal elements of the matrix S defined below in (8.3)]. As it stands, the time-varying matrix $H(\cdot)R^{-1}H^T(\cdot)$ has a complicated dependence upon the measurements via the estimates $\hat{\Theta}_k, 1 \le k \le m$, which renders the solution of (7.3) an intractable problem. We note, however, that the off-diagonal elements of $H(\cdot)R^{-1}H^T(\cdot)$ are oscillatory (thus averaging over time roughly to zero), which suggests that we approximate $H(\cdot)R^{-1}H^T(\cdot)$ by its average value, given below by

$$\frac{1}{M}\sum_{n=0}^{M-1} H(n)R^{-1}H^T(n), \qquad (8.2)$$

prior to solving (7.3). Provided M corresponds to a sufficiently large number of periods of the slowest oscillation in $H(\cdot)R^{-1}H^T(\cdot)$, then this quantity is given roughly by

$$S \triangleq \frac{1}{\sigma^2}\operatorname{diag}\{0, b_1^2, \cdots, b_m^2\}. \qquad (8.3)$$

We remark that the use of S, rather than $H(\cdot)R^{-1}H^T(\cdot)$, leads to an approximate expression for Γ identical to that for the approximate Fisher matrix $\overline{J}$ defined in (4.4). An approximate expression for $\Gamma, \overline{\Gamma}$, is derived as follows. First, observe that

$$\Gamma(k|k) = \overline{\Gamma}(k|k) + \tilde{\Gamma}(k|k) \qquad (8.4)$$

where

$$\overline{\Gamma}(k|k) = F^{-T}\overline{\Gamma}(k-1|k-1)F^{-1} + S \qquad (8.5a)$$

$$\overline{\Gamma}(n_0|n_0) = (F^{-T})^{n_0}\Sigma_s^{-1}(0)F^{-n_0} + S \qquad (8.5b)$$

and

$$\tilde{\Gamma}(k|k) = F^{-T}\tilde{\Gamma}(k-1|k-1)F^{-1} + H(k)R^{-1}H^T(k) - S \qquad (8.6a)$$

$$\tilde{\Gamma}(n_0|n_0) = H(n_0)R^{-1}H^T(n_0) - S. \qquad (8.6b)$$

(Here, $\overline{\Gamma}(\cdot|\cdot)$ is analogous to $\overline{J}$ and $\tilde{\Gamma}(\cdot|\cdot)$ to $\tilde{J}$.)

The quantity of interest is the information matrix at time $n = 0$ condtioned on all the measurements, viz.

$$\overline{\Gamma}(0|n_0+N-1) = (F^{-T})^{1-N-n_0}\overline{\Gamma}(n_0+N-1|n_0+N-1)(F^{-1})^{1-N-n_0}. \qquad (8.7)$$

[The LHS of (8.7) will be denoted $\Gamma(0)$ to keep the notation simple, similarly for $\tilde{\Gamma}(0)$ and $\overline{\Gamma}(0)$.] From (8.6a) and (8.6b), it is evident, after some calculation, that

$$\tilde{\Gamma}(0) = \sum_{n=0}^{N-1}(F^T)^{n+n_0}\tilde{S}(n+n_0)F^{n+n_0} \qquad (8.8)$$

where

$$\tilde{S}(k) \triangleq H(k)R^{-1}H^T(k) - S. \qquad (8.9)$$

Similarly, from (8.5a) and (8.5b), there follows

$$\overline{\Gamma}(0) = \Sigma_s^{-1}(0) + \sum_{n=0}^{N-1}(F^T)^{n+n_0}SF^{n+n_0}. \qquad (8.10)$$

Observing that

$$F^{n+n_0} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ (n+n_0)T & 1 & & & \\ 2(n+n_0)T & 0 & 1 & & \\ \vdots & & & \ddots & \\ m(n+n_0)T & & & & 1 \end{bmatrix} \qquad (8.11)$$

some further calculation shows that

$$\tilde{\Gamma}_{ij}(0) = \begin{cases} \dfrac{2T^2}{\sigma^2}\displaystyle\sum_{n=0}^{N-1}(n+n_0)^2\sum_{l\neq p} lpb_lb_p\cos\Delta\hat{\Theta}_{lp}, \\ \quad i=j=1 \\ \dfrac{T}{\sigma^2}\displaystyle\sum_{n=0}^{N-1}(n+n_0)\sum_{k\neq j-1} kb_kb_{j-1}\cos\Delta\hat{\Theta}_{k,j-1}, \\ \quad i=1, j>1 \\ \dfrac{1}{\sigma^2}\displaystyle\sum_{n=0}^{N-1} b_{i-1}b_{j-1}\cos\Delta\hat{\Theta}_{i-1,j-1}, \\ \quad i>1, j>1 \end{cases} \qquad (8.12)$$

(where $\Delta\hat{\Theta}_{jk} \triangleq \hat{\Theta}_j - \hat{\Theta}_k$) and

$$\overline{\Gamma}(0) = \begin{pmatrix} A + a^{-1} & B \\ B^T & C + \epsilon^{-1}I_m \end{pmatrix} \qquad (8.13)$$

where A, B, and C are defined in (4.5a)–(4.5c). (Apart from the perturbing terms on the diagonal, $\overline{\Gamma}(0)$ is identical to the Fisher information matrix $\overline{J}$ earlier derived.)

Inspection of (8.12) reveals that $\tilde{\Gamma}(0)$ depends indirectly upon the measurements via the state estimates $\hat{\Theta}_k$. Also observe that $\tilde{\Gamma}(0)$ is identical to $\tilde{J}$ (see the proof of Theorem 4.1 in Appendix B), except that the total phase values are replaced by their *estimates*. It is therefore reasonable to expect that Theorem 4.1 might be used to give credibility to the approximation of $\Gamma(0)$ by $\overline{\Gamma}(0)$ since, in the absence of process noise, we would expect the EKF state estimates to converge *almost surely* to the true state values (given sufficiently accurate initial state esimtates). Therefore, for sufficiently large

$N, \tilde{\Gamma}(0)$ would closely approximate $\tilde{J}$, enabling "application" of Theorem 4.1.[2]

Inversion of $\overline{\Gamma}(0)$ yields approximately the quantity of interest, $\Sigma(0|n_0 + N - 1)$, the diagonal components of which are the desired parameter estimation error variances. These are given as follows for the various limiting cases.

Frequency Estimation Error Variance $(a \to \infty)$:
Phases known $(\epsilon \to 0)$:

$$\mathrm{var}(\hat{\omega}) = \frac{\sigma^2}{T^2(Q + 2n_0P + n_0^2N)\left(\sum_{k=1}^{m} k^2 b_k^2\right)} \tag{8.14}$$

Phases unknown: $(\epsilon \to \infty)$:

$$\mathrm{var}(\hat{\omega}) = \frac{12\sigma^2}{T^2N(N^2-1)\left(\sum_{k=1}^{m} k^2 b_k^2\right)}. \tag{8.15}$$

Phase Estimation Error Variance $(\epsilon \to \infty)$:
Frequency known: $(a \to 0)$:

$$\mathrm{var}(\hat{\theta}_k) = \frac{\sigma^2}{Nb_k^2}, \qquad 1 \le k \le m \tag{8.16}$$

Frequency Unknown: $(a \to \infty)$:

$$\mathrm{var}(\hat{\theta}_k) = \frac{\sigma^2}{Nb_k^2} + \frac{3k^2(N-1)\sigma^2}{N(N+1)\left(\sum_{k=1}^{m} k^2 b_k^2\right)}, \qquad 1 \le k \le m. \tag{8.17}$$

These are identical to the CR bounds of (4.9)–(4.12).

IX. Amplitudes Unknown

In this section, the previous assumption concerning complete knowledge of the harmonic amplitudes is relaxed. We give a heuristic argument in favor of the conjecture that the amplitude uncertainty has, for large N, no effect on the phase and frequency estimation performance.

Define the augmented state vector

$$x_a(n) \triangleq [b_1(n) \cdots b_m(n)\, x^T(n)]^T \tag{9.1}$$

[where $x(n)$ was earlier defined in (5.4)]. The amplitudes $b_1, \cdots, b_m$ are assumed constant but unknown. Therefore, define the augmented system matrix

$$F_a \triangleq \begin{bmatrix} I_m & 0 \\ 0 & F \end{bmatrix} \tag{9.2}$$

where F is as defined in (5.6). By arguments similar to the case of known amplitudes, the augmented information matrix $\Gamma_a(0)$ is given by the expression

$$\Gamma_a(0) = \Sigma_a^{-1}(0) + \sum_{n=0}^{N-1} (F_a^T)^{n+n_0} S_a(n) F_a^{n+n_0} \tag{9.3}$$

where

$$S_a(n) \triangleq \begin{bmatrix} I_m & 0 \\ 0 & S^*(n) \end{bmatrix} \tag{9.4}$$

and

$$\Sigma_a(0) \triangleq \begin{bmatrix} \delta I_m & 0 \\ 0 & \Sigma_s(0) \end{bmatrix}. \tag{9.5}$$

Note that $S^*(n)$ is a time-varying version of S in (8.3) and is given by

$$S^*(n) \triangleq \frac{1}{\sigma^2} \mathrm{diag}\,\{0, \hat{b}_1^2(n|n-1), \cdots, \hat{b}_m^2(n|n-1)\}. \tag{9.6}$$

Due to the block diagonal nature of $F_a, \Gamma_a(0)$ is also block diagonal, viz.

$$\Gamma_a(0) = \begin{bmatrix} \Gamma_1(0) & 0 \\ 0 & \Gamma_2(0) \end{bmatrix}. \tag{9.7}$$

Here, $\Gamma_1(0)$ is an $m \times m$ matrix, In fact, it is easy to see that

$$\Gamma_1(0) = \left(\frac{1}{\delta} + \frac{N}{\sigma^2}\right) I_m. \tag{9.8}$$

If the harmonic ampitudes are assumed completely unknown (i.e., $\delta \to \infty$), then $\Gamma_1(0) = (N/\sigma^2) I_m$. Thus, the estimation error variances for the harmonic amplitudes are given by

$$\mathrm{var}\,(\hat{b}_k(0)) = \frac{\sigma^2}{N}, \qquad 1 \le k \le m. \tag{9.9}$$

For large N, we argue, as in the last section, that the amplitude estimates have essentially converged to the true amplitude values so that $S^*(n)$ can be replaced in (9.4) by S. This then yields the same expression for $\Gamma_2(0)$ as given by (8.10) for $\Gamma(0)$. Hence, it appears reasonable to suppose that the phase and frequency estimation performance is unaffected by any requirement to estimate the harmonic amplitudes for large N. This is consistent with the result for the CR bounds.

X. Conclusions

This paper has considered the estimation of the parameters (frequency, amplitudes, phases) of a complex multiharmonic signal. It has presented a slightly generalized derivation of the Cramer–Rao bounds on unbiased estimation performance for such a signal, and has shown how the EKF meets these bounds approximately in its linear (or high-SNR) region for a sufficiently large number of measurements. It was seen that the MLE and EKF approaches are equivalent (at least in the linear region) in the sense that the respective estimation performances are approximately the same. However, an important unresolved question relates to the relative performance of the two approaches as the SNR is lowered and thresholding

[2] A rigorous proof of this claim (complicated as it is by the dependence of $\tilde{\Gamma}$ via the phase estimates on the measurement noise) lies outside the scope of this paper.

behavior becomes apparent. Even though their performance is equivalent for high SNR's, it is unclear that the two approaches would have the same threshold point (i.e., value of SNR where performance suddenly collapses). The relative location of the threshold points is obviously an important point of comparison between the two approaches. Of importance also is the relative performance below threshold. Monte Carlo simulation should give guidance here in relation to both these issues.

The EKF approach has the advantage that estimates are computed recursively and that it can cope easily with time variation in the signal parameters, whereas the MLE approach depends on the parameters remaining constant in time. Of future interest in this regard is an investigation of the possibly deleterious effects on ML estimation performance if slow time variations occur in the signal parameters.

APPENDIX A
THE HILBERT TRANSFORM AND WHITE NOISE

For consistency with the approach of [1] and [2], the socalaled analytic signal has been considered in this paper. This Appendix derives properties of the associated analytic noise sequence and aims to clarify statements elsewhere in the literature concerning, e.g., [1], [2], and [4].

We first consider a *real* measurement sequence of length $2N$, with each element of the sequence separated in time by $(T/2)$. Assume a received signal of the form

$$\begin{aligned} X(n) \triangleq & \sum_{k=1}^{m} b_k \cos\left[k\omega_0(n+2n_0)\frac{T}{2} + \theta_k\right] \\ & + w\left[(n+2n_0)\frac{T}{2}\right], \\ & n = 0, 1, 2, \cdots, 2N-1. \end{aligned} \tag{11.1}$$

(This is an upsampled version of (2.3) with sampling beginning at the same point in time, n_0T.) The noise component $w(\cdot)$ is, as before, zero mean, white, and Gaussian with variance σ^2. We then suppose that $2N$ samples of $X(\cdot)$ are taken (beginning at n_0T) and passed through an ideal discrete Hilbert transformer to yield $Y(\cdot)$ [given by an appropriately modified version of (2.4)]. The *analytical signal* is then constructed as $Z(\cdot) = X(\cdot) + jY(\cdot)$. The transformed noise component for the kth sample is defined (dropping $(T/2)$) by (see [10])

$$w^H(k) \triangleq \sum_{\substack{-\infty \\ m\,\text{odd}}}^{\infty} \frac{2w(k-m)}{\pi m}. \tag{11.2}$$

Given that the variance of the original white noise sequence is σ^2 and in conjunction with (11.2), it is straightforward to show that

$$E[w(k)w^H(l)] = \begin{cases} 0, & (k-l) \text{ even} \\ \dfrac{2\sigma^2}{\pi(l-k)}, & (k-l) \text{ odd} \end{cases} \tag{11.3}$$

and

$$E[w^H(k)w^H(l)] = \sigma^2\delta_{kl}. \tag{11.4}$$

From this, we conclude that the variance of the *analytical noise* $\overline{w}(k) \triangleq w(k) + jw^H(k)$ is given by

$$E[\overline{w}(k)\overline{w}(l)^*] = \begin{cases} 2\sigma^2\delta_{kl}, & (k-l) \text{ even} \\ \dfrac{4j\sigma^2}{\pi(k-l)}, & (k-l) \text{ odd.} \end{cases} \tag{11.5}$$

The analytical noise sequence is therefore not white. Consider, however, a sequence composed of the even (or odd) samples of the original analytic noise sequence. This sequence has N samples and sample period T. Furthermore, from (11.5), it is *white* with variance $2\sigma^2$. Thus, the analytic signal Z_n defined in Section II may be obtained by taking the even (or odd) samples of a length $2N$ sequence of samples of the "unsampled" analytic signal as defined above. The signal Z_n so obtained therefore contains a noise component that is *white*, a property that is vital to the subsequent analysis of Section II.

(*Note:* Consider the vector noise sequence $\tilde{w} \triangleq [w\, w^H]^T$ composed of the even (or odd) samples of the original white noise sequence and the Hilbert-transformed noise sequence, respectively. Then it is easy to see that $E[\tilde{w}(k)\tilde{w}(l)^T] = \sigma^2 I_2$. This fact is used in Section III.)

APPENDIX B
PROOF OF THEOREM 4.1

This proof, which is slightly different from that in [3], is given for the sake of completeness. Also, we remark that Theorem 4.1 is slightly more general than the result given in [3] since it deals with a complex signal with an arbitrary time of commencement of measurements n_0. We also reiterate that the result proved below applies to the Fisher matrix for the problem where the frequency *and* the phases are unknown. For the case where the frequency is unknown and the phases are known (or *vice versa*), the proof is not significantly different and will not be given here.

The following standard result, which we state without proof (see [3]) will be of use:

$$\sum_{n=0}^{N-1} n^i \cos n\phi = O(N^i) \tag{12.1}$$

where $i \geq 0$, ϕ is not an integral multiple of π, and $O(x)$ is defined immediately prior to Theorem 4.1.

Recalling (2.1), (2.3), (2.4), and (4.2), and defining

$$\phi_k \triangleq k\omega_0(n+n_0)T + \theta_k, \qquad k = 1, \cdots m \tag{12.2}$$

the Fisher information matrix J is straightforwardly calculated to be (12.3), which appears at the top of the next page. From the definition of $\overline{J}$ in (4.4), $\tilde{J}$ is given by (12.4), which also

$$J = \begin{cases} \dfrac{T^2(Q+2n_0P+n_0^2N)}{\sigma^2}\sum_{k=1}^{m} k^2 b_k^2 + \dfrac{2T^2}{\sigma^2}\sum_{n=0}^{N-1}(n+n_0)^2 \sum_{l\neq p} l_p b_l b_p \cos(\phi_l - \phi_p), \\ \quad i=1, j=1 \\ \dfrac{(j-1)b_{j-1}^2 T(n_0 N + P)}{\sigma^2} + \dfrac{T}{\sigma^2}\sum_{n=0}^{N-1}(n+n_0)\sum_{k\neq j-1} k b_k b_{j-1}\cos(\phi_k - \phi_{j-1}), \\ \quad i=1, j\neq i \\ \dfrac{1}{\sigma^2}\sum_{n=0}^{N-1} b_{i-1}b_{j-1}\cos(\phi_{i-1}-\phi_{j-1}), \\ \quad i>1, j>1. \end{cases} \tag{12.3}$$

$$\tilde{J} = \begin{cases} \dfrac{2T^2}{\sigma^2}\sum_{n=0}^{N-1}(n+n_0)^2 \sum_{l\neq p} lpb_l b_p \cos(\phi_l - \phi_p), & i=1, j=1 \\ \dfrac{T}{\sigma^2}\sum_{n=0}^{N-1}(n+n_0)\sum_{k\neq j-1} k b_k b_{j-1}\cos(\phi_k - \phi_{j-1}), & i=1, j\neq i \\ \dfrac{1}{\sigma^2}\sum_{n=0}^{N-1} b_{i-1}b_{j-1}\cos(\phi_{i-1}-\phi_{j-1}), & i>1, j>1, i\neq j \\ 0, & i=j>1. \end{cases} \tag{12.4}$$

$$\begin{pmatrix} A & B \\ B^T & C \end{pmatrix}^{-1} = \begin{pmatrix} (A-BC^{-1}B^T)^{-1} & -(A-BC^{-1}B^T)^{-1}BC^{-1} \\ -C^{-1}B^T(A-BC^{-1}B^T)^{-1} & (C-B^TA^{-1}B)^{-1} \end{pmatrix} \tag{12.6}$$

appears at the top of the page. Partitioning $\tilde{J}$ conformally with the RHS of (4.4), it is easy to see from (12.1) and (12.4) that

$$\tilde{J} = \begin{pmatrix} O(N^2) & O(N) \\ O(N) & O(1) \end{pmatrix}, \tag{12.5}$$

which proves the first part of Theorem 4.1.

Straightforward application of the Block Matrix Inversion Lemma (see (12.6) at the top of the page) reveals that for large N,

$$\overline{J}^{-1} = \begin{pmatrix} \alpha N^{-3} & \beta N^{-2} \\ \beta^T N^{-2} & \gamma N^{-1} \end{pmatrix} \tag{12.7}$$

where α is a constant scalar, β is a constant row vector, and γ is a constant $m\times m$ matrix. [Note that the diagonal components of $\overline{J}$ are given by (4.10) and (4.12).]

It is easy to see from (12.5) and (12.7) that

$$\overline{J}^{-1}(\tilde{J}\overline{J}^{-1})^k = \begin{pmatrix} O(1)N^{-(4+k)} & O(1)N^{-(3+k)} \\ O(1)N^{-(3+k)} & O(1)N^{-(2+k)} \end{pmatrix}. \tag{12.8}$$

Some work shows that J^{-1} is given by an infinite sum of the terms defined in (12.8) as follows:

$$J^{-1} = \sum_{k=0}^{\infty}(-1)^k \overline{J}^{-1}(\tilde{J}\overline{J}^{-1})^k. \tag{12.9}$$

From (12.8), it is clear that this power series is convergent provided N is sufficiently large. We can rewrite (12.9) as follows:

$$J^{-1} - \overline{J}^{-1} = \sum_{k=1}^{\infty}(-1)^k \overline{J}^{-1}(\tilde{J}\overline{J}^{-1})^k. \tag{12.10}$$

It is therefore also clear from (12.8) that $\|J^{-1} - \overline{J}^{-1}\| \to 0$ as $N\to\infty$ for any suitable matrix norm $\|\cdot\|$. ∇∇

References

[1] D. C. Rife and R. R. Boorstyn, "Signal tone parameter estimation from discrete time observations," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 591-598, Sept. 1974.

[2] ——, "Multiple tone parameter estimation from discrete time observations," *Bell Syst. Tech. J.*, vol. 55, pp. 1389-1410, Nov. 1976

[3] A. Nehorai and B. Porat, "Adaptive comb filtering for harmonic signal enhancement," *IEEE Trans. Acoust. Speech Signal Processing*, vol. ASSP-34, pp. 1124-1138, Oct. 1986.

[4] R. F. Barrett and D. R. A. McMahon, "ML estimation of the fundamental frequency of a harmonic series," in *Proc. Int. Symp. Signal Processing Applications*, 1967, pp. 333-336.
[5] F. L. Lewis, *Optimal Estimation*. New York: Wiley, 1986.
[6] P. J. Parker and B. D. O. Anderson, "Frequency tracking of nonsinusoidal periodic signals in noise, *IEEE Trans. Signal Processing*, vol. 20, pp. 127-152, 1990.
[7] B. James and B. D. O. Anderson, "The amplitude, phase and frequency estimation of multiharmonic signals in noise—An investigation of the general phase-frequency estimator," in *Proc. ISSPA – 90*. Gold Coast, Australia, Aug. 1990, pp. 141-145.
[8] B. James, B. D. O. Anderson and R. C. Williamson, "Multiharmonic frequency estimation in noise," submitted to *IEEE Trans. Signal Processing*.
[9] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall, 1979.
[10] B. C. Lovell and R. C. Williamson, "The statistical performance of some instantaneous frequency estimators," *IEEE Trans. Signal Processing*, July 1992.

Ben James was born in Brisbane, Australia, in 1965. He received the B.Sc. degree from the Australian National University in 1987, the B.E. degree (in electrical engineering) from the James Cook University of North Queensland in 1989, and the Ph.D. degree in systems engineering from the Australian National University in 1992.

In 1992 he was a Lecturer in the Department of Engineering at the Australian National University, and is currently a Research Associate in the Industrial Systems Group, Imperial College, London. His interests include signal processing and multivariable frequency domain control.

Brian D. O. Anderson (S'62–M'66–SM'74–F'75) was born in Sydney, Australia. He received his undergraduate education at the University of Sydney, with majors in pure mathematics and electrical engineering. He subsequently received the Ph.D. degree in electrical engineering from Stanford University.

Following completion of his education, he worked in industry in Silicon Valley and served as an Assistant Professor in the Department of Electrical Engineering at Stanford. He was Foundation Professor of Electrical Engineering at the University of Newcastle, Australia, from 1967 to 1981, and is now Professor of Systems Engineering at the Australian National University. His interests are in control and signal processing.

Dr. Anderson is a Fellow of the Royal Society, the Australian Academy of Science, the Australian Academy of Technological Sciences and Engineering, and an Honorary Fellow of the Institution of Engineers, Australia. He holds a doctorate (honoris causa) from the Université Catholique de Louvain, Belgium. He served a term as President of the International Federation of Automatic Control from 1990 to 1993.

Robert C. Williamson was born in Brisbane, Australia, in 1962. He received the B.E. degree from QIT in 1984, and the M.Eng.Sc. and Ph.D. degrees from the University of Queensland in 1986 and 1990, respectively, all in electrical engineering.

In 1989 he was a Lecturer at the University of Queensland. Since 1990 he has been at the Australian National University in the Department of Systems Engineering and the Department of Engineering where he is a Senior Lecturer. His research interests include signal processing and neural networks.