Maximum Likelihood Estimation of the Parameters of Multiple Sinusoids from Noisy Measurements

PETRE STOICA, RANDOLPH L. MOSES, MEMBER, IEEE, BENJAMIN FRIEDLANDER, FELLOW, IEEE, AND TORSTEN SÖDERSTRÖM, SENIOR MEMBER, IEEE

Abstract—**The problem of estimating the frequencies, phases, and amplitudes of sinusoidal signals is considered. A simplified maximum-likelihood Gauss-Newton algorithm which provides asymptotically efficient estimates of these parameters is proposed. Initial estimates for this algorithm are obtained by a variation of the overdetermined Yule-Walker method and a periodogram-based procedure. Use of the maximum-likelihood Gauss-Newton algorithm is not, however, limited to this particular initialization method. Some other possibilities to get suitable initial estimates are briefly discussed. An analytical and numerical study of the shape of the likelihood function associated with the sinusoids-in-noise process reveals its multimodal structure and clearly sets the importance of the initialization procedure. Some numerical examples are presented to illustrate the performance of the proposed estimation procedure. Comparison to the performance corresponding to the Cramer-Rao lower bound is also presented, using a simple expression for the asymptotic Cramer-Rao bound covariance matrix derived in the paper.**

I. Introduction

THE problem of estimating the parameters of sinusoidal signals from noisy data has received considerable attention recently [1]-[9], [18]-[21]. The sinusoid parameters can be estimated using correlation-based techniques. These include Prony's method, Pisarenko's harmonic decomposition procedure, and the Yule-Walker method in one of its many versions. Prony's method (see [2] for a recent survey) is known to give inconsistent estimates. It cannot be used in cases with a low signal-to-noise ratio since the resulting estimates may be highly biased. In Pisarenko's procedure [2] this problem is eliminated. This method gives consistent estimates, but in some cases it has poor accuracy.

The basic Yule-Walker method [1], [2] does not eliminate this deficiency of Pisarenko's method. It gives consistent estimates, but its accuracy may be poor. Since the Yule-Walker method is attractive from the computational standpoint, much effort has been spent in recent years to improve its accuracy properties.

The overdetermined or high-order Yule-Walker method is a modification of the basic Yule-Walker procedure, which was reported to lead to a considerable increase in estimation accuracy [3]-[6]. This method was proposed heuristically, and the properties of the corresponding estimates were analyzed by Monte Carlo simulations only. The reasons for the increase of the parameter estimation accuracy when the number of Yule-Walker equations and the model order are increased were not too well understood. In [11] and [12] we have tried to fill this gap. Very briefly, the conclusions of [11] and [12] are that the asymptotic accuracy of the Yule-Walker estimates will increase with the number of Yule-Walker equations used and with the model order, although not necessarily monotonically. However, even when the number of Yule-Walker equations and the model order are increased without bound, the limiting accuracy may still be worse than that corresponding to the Cramer-Rao lower bound (CRLB). Thus, in general, it is possible to improve the accuracy of the Yule-Walker-based estimates.

To improve the performance of correlation-based techniques, a number of researchers have studied the use of the maximum likelihood (ML) method for estimation of sinusoid parameters; see [7], [8], and [18]-[21]. (Some of these studies, such as [20] and [21], treat more general problems that include the sinusoid parameter estimation as a special case.) The basic idea to all of the estimation schemes in the above references consists of first obtaining suboptimal initial parameter estimates and then refining them through an iterative maximization of the likelihood function. The suboptimal estimation method and the iterative maximization procedure used vary from one reference to another.

In this paper we consider a similar "two-stage" ML estimation procedure. We use the overdetermined Yule-Walker (OYW) method to get initial estimates and a periodogram-based procedure to get improved initial estimates. We also discuss briefly other possibilities for obtaining initial estimates, such as the overdetermined and high-order Yule-Walker method [3], [5], [6] and Tufts and Kumaresan eigenanalysis procedure [4], [8].

An analysis of the likelihood function associated with the sinusoid estimation problem reveals that it has a complicated multimodal shape with a narrow trough corre-

Manuscript received May 3, 1986; revised May 12, 1988. The work of B. Friedlander was supported by the Army Research Office under Contract DAAG29-83-C-0027.

P. Stoica is with the Facultatea de Automatica, Institutul Politehnic Bucuresti, Splaiul Independentei 313, Sector 6, R-77 206 Bucharest, Romania.

R. L. Moses is with the Department of Electrical Engineering, The Ohio State University, Columbus, OH 43210.

B. Friedlander is with Signal Processing Technology, Ltd., 703 Coastland Drive, Palo Alto, CA 94303.

T. Söderström is with the Department of Automatic Control and Systems Analysis, Institute of Technology, Uppsala University, P.O. Box 534, S-751 21 Uppsala, Sweden.

IEEE Log Number 8825662.

sponding to the global maximum. Thus, there is a strong possibility that an initial estimate will not be sufficiently accurate to fall in the global maximum trough. To overcome this problem, we include a second step in the initial estimation to search locally for estimates that lie in the trough. Again, any of several search procedures could be used; we have used a series of one-dimensional searches for the maximum of the periodogram in a certain frequency range. This method is relatively simple computationally, and produces estimates that have a high probability of lying in the narrow trough of the likelihood function.

The two-step initialization can be thought of as a compromise between pure Yule–Walker-based estimates of [8] and [19] and pure periodogram-based estimates of [7] and [18]. The OYW method produces estimates that may not be in the desired trough of the likelihood function. However, these estimates are close enough so that the periodogram step can operate on a small frequency region. Other initialization procedures are given in [20] and [21], in conjunction with an iterative algorithm of a special type.

The refining iterative step of our ML procedure consists of a simplified Gauss–Newton maximization algorithm. This algorithm has the computational simplicity of a pure gradient technique and the convergence rate of a Newton method. It is tailored to the special structure of the problem under discussion, which makes it more attractive than the general Newton-based or other algorithms used in the works referenced above for the iterative maximization of the likelihood function. The development of the simplified Gauss–Newton algorithm is a main contribution of this paper. Its derivation is based on an expression for the asymptotic Cramer–Rao lower bound (CRLB) on the covariance matrix of any consistent estimator of sinusoid parameters, which appears interesting in its own right. Note that the finite-sample CRLB is known (see, e.g., [7]) but the expression for the asymptotic bound presented here appears to be novel (its usefulness is discussed later in the paper).

We prove that the two-step ML estimation procedure outlined above achieves the asymptotic accuracy corresponding to the CRLB. This result does not follow from the "standard" properties of the ML estimator. The finite-sample properties of the procedure are studied by Monte Carlo simulations.

Finally, let us make some general remarks on two-stage (initialization followed by iterative maximization) ML procedures. The whole procedure cannot, in general, have a better threshold SNR (signal-to-noise ratio) than the initialization step. The reason, of course, is that the iterative algorithm of the second step cannot converge to the global maximum of the likelihood function if the initial estimates have poor accuracy (at some low SNR). However, for SNR's above the threshold, the accuracy properties of the two-step ML method are expected to be better than those of the initial step (by the statistical efficiency of the ML estimator). Furthermore, the second ML step may improve the "resolution" properties of the initial step, too.

II. Statement of the Problem

Consider the following sinusoidal signal:

$$x(t) = \sum_{i=1}^{m} \alpha_i \sin(\omega_i t + \phi_i), \qquad t = 1, 2, \cdots, \tag{2.1a}$$

where

$$\alpha_i, \phi_i \in R, \qquad \omega_i \in (0, \pi), \quad \text{and} \quad \omega_i \neq \omega_j \quad \text{for } i \neq j. \tag{2.1b}$$

The assumption $\omega_i \neq 0$ means that a possible nonzero constant level of $x(t)$ has been removed. The condition $\omega_i < \pi$ is a consequence of Shannon's sampling theorem.

Let $y(t)$ denote the noise-corrupted measurements of $x(t)$

$$y(t) = x(t) + \epsilon(t), \tag{2.2}$$

where $\{\epsilon(t)\}$ is a sequence of independent and identically distributed Gaussian random variables with zero mean and variance λ^2. We assume that $x(t)$ and $\epsilon(s)$ are independent for any t and s.

The assumption that $\epsilon(t)$ is Gaussian may appear somewhat restrictive. Under the Gaussian hypothesis, it is easy to write the likelihood function of the data and to obtain an explicit expression for the CRLB. If, in some application, the Gaussian hypothesis fails to be true, the algorithm of this paper is still applicable, but it will no longer provide the ML estimates. Nevertheless, the estimates obtained by using the algorithm will still give the minimum variance in a fairly large class of estimators whose covariance matrices depend only on the second-order statistics of the data. This is explained further in Section IV. The commonly used assumption that $\epsilon(t)$ is a sequence of i.i.d. random variables is restrictive. However, in Section VI it will be shown that the proposed estimation method may provide accurate estimates even if $\epsilon(t)$ is a correlated sequence (see also [23]). For colored $\epsilon(t)$, the estimates provided by the algorithms are no longer pure ML estimates; in such a case, they are "minimum output error estimates" (using a term from the system identification literature).

Next we denote by r_n the covariance of $y(t)$ at lag n ($n = 0, 1, 2, \cdots$)

$$r_n = E\{y(t)\, y(t+n)\}. \tag{2.3}$$

The operator $E\{\cdot\}$ denotes statistical expectation. The sample covariances corresponding to (2.3) shall be denoted by $\hat{r}_n$. We will use the following definition of $\hat{r}_n$:

$$\hat{r}_n = \frac{1}{N-n} \sum_{t=1}^{N-n} y(t)\, y(t+n), \qquad n = 0, 1, 2, \cdots$$

$$\hat{r}_{-n} = \hat{r}_n, \tag{2.4}$$

where N denotes the length of the data sample.

Collecting the amplitudes $\{\alpha_i\}$, phases $\{\phi_i\}$, and frequencies $\{\omega_i\}$ in a single parameter vector, we define

$$\theta = [\alpha_1, \cdots, \alpha_m, \phi_1, \cdots, \phi_m, \omega_1, \cdots, \omega_m]^T. \tag{2.5}$$

The problem considered in this paper is the estimation of θ from N samples of noisy measurements $\{y(1), \cdots, y(N)\}$.

III. The Initial Overdetermined Yule-Walker Estimates

As is well known $x(t)$, (2.1), obeys a homogeneous difference equation of order $2m$,

$$x(t) + a_1 x(t-1) + \cdots + a_n x(t-n) = 0, \qquad n \triangleq 2m, \tag{3.1}$$

where $\{a_i\} \in R$ are such that the polynomial

$$A(z) = 1 + a_1 z + \cdots + a_n z^n \tag{3.2a}$$

has all its zeros located on the unit circle at $e^{\pm i\omega_k}$, i.e.,

$$A(e^{\pm i\omega_k}) = 0, \qquad k = 1, \cdots, m \tag{3.2b}$$

(see [2], [4], [5], and [14]). Since we have

$$r_n = E\{x(t)\, x(t+n)\} + \lambda^2 \delta_{n,0}, \tag{3.3a}$$

where $\delta_{i,j}$ is the Dirac delta

$$\delta_{i,j} = \begin{cases} 1 & i = j \\ 0 & i \neq j, \end{cases} \tag{3.3b}$$

it follows from (3.1) that the coefficients $\{a_i\}$ obey the so-called (modified) Yule-Walker equations

$$\sum_{i=0}^{n} a_i r_{n+k-i} = 0, \qquad k \geq 1 \qquad (a_0 = 1). \tag{3.4}$$

A commonly used technique for estimating the frequencies $\{\omega_i\}$ is based on (3.4). Consistent estimates $\{\hat{a}_i\}$ can be obtained by solving the following linear system of equations:

$$\begin{bmatrix} \hat{r}_n & \cdots & \hat{r}_1 \\ \hat{r}_{n+1} & \cdots & \hat{r}_2 \\ \vdots & & \vdots \\ \hat{r}_{L-1} & \cdots & \hat{r}_{L-n} \end{bmatrix} \begin{bmatrix} \hat{a}_1 \\ \vdots \\ \hat{a}_n \end{bmatrix} = - \begin{bmatrix} \hat{r}_{n+1} \\ \hat{r}_{n+2} \\ \vdots \\ \hat{r}_L \end{bmatrix}, \qquad L \geq 2n, \tag{3.5}$$

where $\{\hat{r}_i\}$ are the sample covariances (see [2], [3], and [22]). The matrix appearing in (3.5) has full rank, at least for large N [14]. Note that for $L > 2n$, the system (3.5) is overdetermined and needs to be solved in a least-squares sense. Intuitively we can expect that the larger L, the more accurate will be the estimates $\{\hat{a}_i\}$, since the covariances for large lags contain "useful information" about the covariance structure of the data. While it is not always true that increasing L increases estimation accuracy [12], it was shown by simulations [3], [6] that increasing L is often useful. A theoretical explanation of this empirically noticed fact was recently presented in [12]. It was shown there that while the asymptotic (for $N \to \infty$) accuracy of $\{\hat{a}_i\}$ does not increase monotonically with L, it improves considerably in the limit as $L \to \infty$. For $L < \infty$, the estimation errors $(\hat{a}_i - a_i)$ are of order $1/\sqrt{N}$, and for $L \to \infty$, they are of order $1/L\sqrt{N}$. The estimation technique based on (3.5) with $L > 2n$ is the so-called overdetermined Yule-Walker (OYW) method [3]-[6].

The frequencies $\{\omega_i\}$ can now be estimated by determining the roots of

$$\hat{A}(z) = 1 + \hat{a}_1 z + \cdots + \hat{a}_n z^n = 0. \tag{3.6}$$

Note that determining the estimates $\{\hat{\omega}_i\}$ from (3.6) implies, in general, some approximations since $\hat{A}(z)$ is not guaranteed to have all of its zeros on the unit circle. For example, one may look at the peaks of $1/|\hat{A}(e^{j\omega})|^2$, or at the angles of the roots of $\hat{A}(z)$.

The problem of determining estimates of $\{\alpha_i\}$ and $\{\phi_i\}$, once estimates $\{\hat{\omega}_i\}$ of the frequencies are given, can be reduced to a least-squares fit. Rewrite (2.1) and (2.2) as

$$y(t) = \sum_{k=1}^{m} (\beta_k \sin \omega_k t + b_k \cos \omega_k t) + \epsilon(t), \tag{3.7a}$$

where

$$\beta_k = \alpha_k \cos \phi_k, \qquad b_k = \alpha_k \sin \phi_k. \tag{3.7b}$$

Replacing $\{\omega_i\}$ in (3.7) by their estimates $\{\hat{\omega}_i\}$, the problem of estimating $\{\beta_k, b_k\}$ can be formulated as the following minimization problem:

$$\min_{\{\beta_k, b_k\}} \sum_{t=1}^{M} \left\{ y(t) - \sum_{k=1}^{m} (\beta_k \sin \hat{\omega}_k t + b_k \cos \hat{\omega}_k t) \right\}^2, \qquad M < N. \tag{3.8}$$

The solution to this probem is given by

$$\hat{\psi} \triangleq \left[\hat{\beta}_1 \cdots \frac{\hat{\beta}_m}{\hat{b}_1} \cdots \hat{b}_m \right] = \left\{ \frac{1}{M} \sum_{t=1}^{M} V(t)\, V(t)^T \right\}^{-1} \cdot \left\{ \frac{1}{M} \sum_{t=1}^{M} V(t)\, y(t) \right\}, \tag{3.9a}$$

where

$$V(t) \triangleq [\sin \hat{\omega}_1 t, \cdots, \sin \hat{\omega}_m t, \cos \hat{\omega}_1 t, \cdots, \cos \hat{\omega}_m t]^T. \tag{3.9b}$$

The reason for using only the first M data points in (3.8) and (3.9) will be explained later. It will be shown that if

M in (3.8) is too large (e.g., $M = N$), then the estimation accuracy may deteriorate considerably. Note that for $M < N$, the computational burden is smaller.

Using $\{\hat{\beta}_i\}$ and $\{\hat{b}_i\}$ in (3.7b), we readily obtain estimates of $\{\alpha_i\}$ and $\{\phi_i\}$ as given by

$$\begin{aligned} \hat{\phi}_j &= \arctan\{\hat{b}_j/\hat{\beta}_j\} \quad (\text{mod } 2\pi) \\ \hat{\alpha}_j &= \hat{\beta}_j/\cos\hat{\phi}_j \end{aligned} \qquad j = 1, \cdots, m. \tag{3.10}$$

Next, we discuss some implementation issues related to (3.9). Straightforward programming of (3.9) would lead to a large computational burden. The main reason is that calculation of trigonometric functions on a computer is time consuming. Note, however, that the solution $c_i(t)$ of the following second-order difference equation:

$$c_i(t) - (2\cos\omega_i)c_i(t-1) + c_i(t-2) = 0, \quad t = 3, 4, \cdots, \tag{3.11}$$

with initial conditions

$$c_i(1) = \cos\omega_i, \qquad c_i(2) = \cos 2\omega_i, \tag{3.12}$$

is given by

$$c_i(t) = \cos\omega_i t, \qquad t = 1, 2, \cdots. \tag{3.13}$$

A different set of initial conditions ($c_i(1) = \sin\omega_i$, $c_i(2) = \sin 2\omega_i$) will produce $c_i(t) = \sin\omega_i t$. Thus, the sequences $\{\sin\omega_i t, \cos\omega_i t;\ t = 1, \cdots, M;\ i = 1, \cdots, m\}$ can be generated using (3.11) at a cost of approximately $2mM$ multiplications, and the vector $\Sigma\, V(t)\, y(t)$ in (3.9) will require a total of $4mM$ multiplications.

Next we note that the matrix $\Sigma\, V(t)\, V(t)^T$ in (3.9) can be efficiently computed using Lemma A.1. For large M, we can further simplify the computation of this matrix by using some approximations. Thus, from Lemma A.1 it follows that

$$\frac{1}{M}\sum_{t=1}^{M} V(t)\, V(t)^T = \frac{1}{2} I_{2m} + 0(1/M). \tag{3.14}$$

We conclude from (3.14) that for large M, the following simple estimate:

$$\tilde{\psi} = \frac{2}{M}\sum_{t=1}^{M} V(t)\, y(t) \tag{3.15}$$

is an approximation of order $1/M$ of $\hat{\psi}$ (3.9). Note, however, that the smaller $\inf_{i\neq j}|\hat{\omega}_i - \hat{\omega}_j|$, the larger the value of M needed for the approximation in (3.14) to be valid (see the discussion in Appendix A.) If M is not large enough, then $\tilde{\psi}$ may not be a good approximation of $\hat{\psi}$. Furthermore, the calculation of $\hat{\psi}$ may be problematic in such a case since the matrix in (3.9) will be ill conditioned.

We conclude this section with a discussion of the asymptotic properties of the estimates introduced above. The frequency estimates $\{\hat{\omega}_i\}$ obtained by the OYW method are consistent [15]. The asymptotic (as $N, L \to \infty$) standard deviations of $\{\hat{\omega}_i - \omega_i\}$ are of order $1/L\sqrt{N}$, provided that L increases not faster than N^γ, with $\gamma < 1/2$ [12]. The condition $\gamma < 1/2$ is sufficient but probably not necessary. A necessary and sufficient condition on γ is not known. Since the CRLB on the standard deviation of $\{\hat{\omega}_i\}$ is $0(1/N^{3/2})$, as is shown in Appendix A, it seems possible to improve significantly the accuracy of the OYW estimates.

An analysis of the asymptotic behavior of $\{\hat{\alpha}_i, \hat{\phi}_j\}$ (3.9), (3.10) does not seem to be available in literature. Due to the use in (3.9) of $\{\hat{\omega}_i\}$ instead of $\{\omega_i\}$, such an analysis is not easy. Since $\{\hat{\alpha}_i\}$ and $\{\hat{\phi}_i\}$ are used as initial estimates, their accuracy is not too important, and will not be discussed in detail. What is, however, quite important is the choice of M i . (3.9). To simplify notation, we will consider the case of a single sinuosid ($m = 1$). It should be emphasized, however, that the same conclusions apply also for $m > 1$.

For $m = 1$ and large M, we have from (3.9) and (3.14)

$$\begin{aligned} \hat{\psi} - \psi &= \frac{2}{M}\sum_{t=1}^{M}\begin{bmatrix}\sin\hat{\omega}t\\ \cos\hat{\omega}t\end{bmatrix}\{y(t) - [\sin\hat{\omega}t\ \cos\hat{\omega}t]\psi\} \\ &\quad + 0(1/M) \\ &= \frac{2}{M}\sum_{t=1}^{M}\begin{bmatrix}\sin\omega t\\ \cos\omega t\end{bmatrix}\epsilon(t) \\ &\quad + \frac{2}{M}\sum_{t=1}^{M}\left\{\begin{bmatrix} t\cos\omega t\\ -t\sin\omega t\end{bmatrix}\epsilon(t)\right. \\ &\quad \left. - \begin{bmatrix}\sin\omega t\\ \cos\omega t\end{bmatrix}[t\cos\omega t, -t\sin\omega t]\psi\right\}(\hat{\omega}-\omega) \\ &\quad + \frac{1}{M}\sum_{t=1}^{M}\left\{\begin{bmatrix}-t^2\sin\omega t\\ -t^2\cos\omega t\end{bmatrix}\epsilon(t)\right. \\ &\quad - 2\begin{bmatrix} t\cos\omega t\\ -t\sin\omega t\end{bmatrix}[t\cos\omega t, -t\sin\omega t]\psi \\ &\quad \left. + \begin{bmatrix}\sin\omega t\\ \cos\omega t\end{bmatrix}[t^2\sin\omega t, t^2\cos\omega t]\psi\right\}(\hat{\omega}-\omega)^2 \\ &\quad + \cdots + 0(1/M), \end{aligned} \tag{3.16}$$

where $\psi = [\beta, b]^T$ is the vector of the true parameters. It is not difficult to see that the first term in (3.16) is $0(1/\sqrt{M})$. Since $\hat{\omega} - \omega = 0(1/L\sqrt{N})$ (see the discussion above), it can be shown that the second term is $0(M/L\sqrt{N})$, the third is $0(M^2/(L\sqrt{N})^2)$, etc. Thus, if M increases faster than $L\sqrt{N}$ (for example, if we set $L = N^{1/2-\delta}$ for some $\delta > 0$, and $M = N$), then difficulties may occur. Indeed, in such a case, the estimate $\hat{\psi}$ may not even be consistent. If the condition $M << L\sqrt{N}$ is imposed, then the first and second terms in (3.16) are asymptotically the dominant ones. Note that the magnitude of the first term decreases with M while that of the second increases with M. To get good asymptotic properties for $\hat{\psi}$ (i.e., small estimation error $\hat{\psi} - \psi$), M should

be chosen such that these two terms have the same magnitude. Thus, the "optimal" rate of increase of M is given by

$$M = (L\sqrt{N})^{2/3}. \tag{3.17}$$

The estimation error $(\hat{\psi} - \psi)$ corresponding to this choice of M is of the order $1/\sqrt{M}$.

Note from the analysis above that we can use all N data points in (3.9) only if the estimation error $(\hat{\omega} - \omega)$ is $O(1/N^{3/2})$, i.e., $\hat{\omega}$ achieves the CRLB (see Appendix A). If the estimation error of $\hat{\omega}$ is larger than $O(1/N^{3/2})$, only a fraction of the data samples should be used in (3.9). The poorer the frequency estimates, the smaller this fraction has to be. This important property, which albeit is rather intuitive, is often overlooked in the literature.

IV. A Maximum Likelihood Algorithm

The ML estimate of θ is obtained as the minimum point of the following loss function (see [7], [8], [19]–[21] and also Appendix A):

$$LF = \sum_{t=1}^{N} \epsilon^2(t, \theta), \tag{4.1a}$$

where

$$\epsilon(t, \theta) = y(t) - \sum_{i=1}^{m} \alpha_i \sin(\omega_i t + \phi_i). \tag{4.1b}$$

We use the Gauss–Newton algorithm to minimize (4.1). Let $\hat{\theta}^k$ denote the parameter estimate at iteration k. The updated estimate $\hat{\theta}^{k+1}$ is computed by

$$\hat{\theta}^{k+1} = \hat{\theta}^k - \left[\sum_{t=1}^{N} \epsilon_\theta(t, \hat{\theta}^k)\, \epsilon_\theta^T(t, \hat{\theta}^k)\right]^{-1} \cdot \left[\sum_{t=1}^{N} \epsilon_\theta(t, \hat{\theta}^k)\, \epsilon(t, \hat{\theta}^k)\right] \tag{4.2a}$$

where

$$\epsilon_\theta(t, \theta) \triangleq \frac{\partial \epsilon(t, \theta)}{\partial \theta}, \tag{4.2b}$$

and where $\hat{\theta}^0$, the initial estimate, should be given. The elements of the gradient vector $\epsilon_\theta(t, \theta)$ are given by

$$\left.\begin{aligned} \frac{\partial \epsilon(t, \theta)}{\partial \alpha_i} &= -\sin(\omega_i t + \phi_i) \\ \frac{\partial \epsilon(t, \theta)}{\partial \phi_i} &= -\alpha_i \cos(\omega_i t + \phi_i) \\ \frac{\partial \epsilon(t, \theta)}{\partial \omega_i} &= -t\alpha_i \cos(\omega_i t + \phi_i) \end{aligned}\right\} \quad i = 1, \cdots, m. \tag{4.3}$$

The matrix to be inverted in (4.2) contains entries of very different magnitudes. The elements of its left-top $m \times m$ block are of order N, while those of the right-bottom $m \times m$ block are of the order N^3. Thus, it is desirable from the numerical standpoint to "balance" the elements of the matrix. This will also be convenient for some subsequent theoretical considerations.

Let us introduce the notation

$$K_N = \begin{bmatrix} N^{1/2} I_{2m} & 0 \\ 0 & N^{3/2} I_m \end{bmatrix} \tag{4.4}$$

where I_K denotes the $K \times K$ identify matrix. The following recursion is equivalent to, but numerically more reliable than, (4.2a):

$$K_N \hat{\theta}^{k+1} = K_N \hat{\theta}^k - H_N^{-1}(\hat{\theta}^k) \cdot \left[K_N^{-1} \sum_{t=1}^{N} \epsilon_\theta(t, \hat{\theta}^k)\, \epsilon(t, \hat{\theta}^k)\right], \tag{4.5a}$$

where

$$H_N(\theta) = K_N^{-1}\left[\sum_{t=1}^{N} \epsilon_\theta(t, \theta)\, \epsilon_\theta^T(t, \theta)\right] K_N^{-1}. \tag{4.5b}$$

Evaluation of the vector $\Sigma_{t=1}^{N} \epsilon_\theta(t, \hat{\theta}^k)\, \epsilon(t, \hat{\theta}^k)$ is straightforward. Its elements contain trigonometric functions which could be computed efficiently by the technique discussed in the previous section. Evaluation of the matrix $H_N(\hat{\theta}^k)$ can be done similarly but it appears quite costly. To overcome this difficulty, we propose an approximate version of the iterative algorithm (4.2a).

As is shown in Appendix A,

$$H_N^{-1}(\theta) = G(\theta) + 0(1/N), \tag{4.6}$$

where

$$G(\theta) = 2\left[\begin{array}{c|ccc|ccc} I_m & \multicolumn{3}{c|}{0} & \multicolumn{3}{c}{0} \\ \hline & \frac{4}{\alpha_1^2} & & 0 & -\frac{6}{\alpha_1^2} & & 0 \\ 0 & & \ddots & & & \ddots & \\ & 0 & & \frac{4}{\alpha_m^2} & 0 & & -\frac{6}{\alpha_m^2} \\ \hline & -\frac{6}{\alpha_1^2} & & 0 & \frac{12}{\alpha_1^2} & & 0 \\ 0 & & \ddots & & & \ddots & \\ & 0 & & -\frac{6}{\alpha_m^2} & 0 & & \frac{12}{\alpha_m^2} \end{array}\right]. \tag{4.7}$$

Replacing $H_N^{-1}(\hat{\theta}^k)$ in (4.5) by its large sample approximation $G(\hat{\theta}^k)$, we get

$$K_N \hat{\theta}^{k+1} = K_N \hat{\theta}^k - \mu_k G(\hat{\theta}^k) \cdot \left[K_N^{-1} \sum_{t=1}^{N} \epsilon_\theta(t, \hat{\theta}^k)\, \epsilon(t, \hat{\theta}^k)\right] \tag{4.8}$$

where $\{\mu_k\}$ is a sequence of positive scalars which can be used for controlling the step size (μ_k can be determined, for example, by using a line search algorithm).

The algorithm (4.8) is much simpler than (4.2a). The two algorithms have clearly the same convergence point. Furthermore, for large N, they will also have similar convergence rates.

We conclude this section with a discussion of the asymptotic accuracy of the limiting (as $k \to \infty$) estimate obtained by (4.8). Let this estimate be denoted by $\tilde{\theta}$:

$$\tilde{\theta} = \lim_{k \to \infty} \hat{\theta}^k. \tag{4.9}$$

Since we will initialize the recursion (4.8) with a consistent estimate, it is expected to converge in a few iterations. In fact, paralleling the calculations in the proof of Theorem 4.1 below, it is possible to show that (4.8) will asymptotically (as $N \to \infty$) converge in *one iteration* provided that the initial estimate $\hat{\theta}^0$ is good enough.

Under the Gaussian hypothesis, $\tilde{\theta}$ is the ML estimate. We expect, therefore, that its asymptotic covariance matrix equals the CRLB $P_{CR}^{\theta} = \lambda^2 G(\theta)$ (see Appendix A for the derivation of P_{CR}^{θ}). However, this does not follow immediately since some of the standard assumptions of ML theory [10] fail to hold in our case (e.g., $\epsilon_\theta(t, \theta)$ is a nonstationary process).

If we relax the Gaussian hypothesis, then $\tilde{\theta}$ is the prediction error (PE) estimate [16]. Again, the standard PE theory does not apply to our problem. If it were applicable, it would follow from [16] that the asymptotic covariance matrix of $\tilde{\theta}$ is still given by the matrix P_{CR}^{θ} defined below.

The asymptotic covariance matrix of the normalized estimation errors $K_N(\tilde{\theta} - \theta)$ is derived next. We show that this matrix equals P_{CR}^{θ} for any distribution function of the data.

Theorem 4.1: Consider the process $y(t)$ generated by (2.1) and (2.2) under the assumptions stated except that $\epsilon(t)$ is allowed to be non-Gaussian. Let $\tilde{\theta}$ be the estimate given by (4.9). Then

$$\lim_{N \to \infty} E\left[K_N(\tilde{\theta} - \theta)(\tilde{\theta} - \theta)^T K_N\right] = P_{CR}^{\theta}, \tag{4.10}$$

where K_N is defined in (4.4) and $P_{CR}^{\theta} = \lambda^2 G(\theta)$.

Proof: See Appendix C.

It follows from the above theorem that in the case of Gaussian data, the estimate $\tilde{\theta}$ of θ is asymptotically efficient. For non-Gaussian data, $\tilde{\theta}$ will asymptotically be the minimum-variance estimate in a fairly large class of estimators whose covariance matrices depend only on the second-order statistics of the data.

V. The Problem of Local Minima

A major concern in any iterative minimization algorithm is the presence of local minima in the function to be minimized. Below we analyze the shape of the Loss Function (LF). For an arbitrary parameter vector $\hat{\theta}$ we can express $\mathrm{LF}(\hat{\theta})$ as

$$\mathrm{LF}(\hat{\theta}) = N \cdot \left\{\mathrm{LF}_s(\hat{\theta}) + \mathrm{LF}_n(\hat{\theta}) + R\right\} \tag{5.1a}$$

where

$$\mathrm{LF}_s(\hat{\theta}) = \frac{1}{N} \sum_{t=1}^{N} \left[x(t) - \hat{x}(t)\right]^2 \tag{5.1b}$$

$$\mathrm{LF}_n(\hat{\theta}) = \frac{2}{N} \sum_{t=1}^{N} \epsilon(t)\left[x(t) - \hat{x}(t)\right] \tag{5.1c}$$

$$R = \frac{1}{N} \sum_{t=1}^{N} \epsilon^2(t) \tag{5.1d}$$

and where $\hat{x}(t)$ is defined as in (2.1) but with elements of $\hat{\theta}$ replacing elements of θ there. Comparing (5.1c) and (C.7) in Appendix C, we see that $\mathrm{LF}_n(\hat{\theta})$ is $0(1/\sqrt{N})$. Also, writing out (5.1b) and using Lemma A.1 in Appendix A, it is easy to show that

$$\begin{aligned}\mathrm{LF}_s(\hat{\theta}) &= \sum_{i=1}^{m} \left(\frac{1}{N} \sum_{t=1}^{N} \left[\alpha_i \sin(\omega_i t + \phi_i) \right.\right. \\ &\qquad \left.\left. - \hat{\alpha}_i \sin(\hat{\omega}_i t + \hat{\phi}_i)\right]^2 \right) + 0(1/N) \\ &= \sum_{i=1}^{m} F_i(\hat{\alpha}_i, \hat{\omega}_i, \hat{\phi}_i) + 0(1/N)\end{aligned} \tag{5.2a}$$

where

$$\begin{aligned}F_i(\hat{\alpha}_i, \hat{\omega}_i, \hat{\phi}_i) = \frac{1}{N} \sum_{t=1}^{N} &\left[\alpha_i \sin(\omega_i t + \phi_i) \right. \\ &\left. - \hat{\alpha}_i \sin(\hat{\omega}_i t + \hat{\phi}_i)\right]^2.\end{aligned} \tag{5.2b}$$

Thus, to within $0(1/N)$, $\mathrm{LF}_s(\hat{\theta})$ is the sum of m decoupled functions F_i; moreover, all of the F_i's have the same form. Understanding the shape of the LF in (5.2a) asymptotically reduces to understanding the shape of the function

$$F(\hat{\alpha}, \hat{\omega}, \hat{\phi}) = \frac{1}{N} \sum_{t=1}^{N} \left[\alpha \sin(\omega t + \phi) - \hat{\alpha} \sin(\hat{\omega} t + \hat{\phi})\right]^2. \tag{5.3}$$

It is easy to check that F is quadratic in $\hat{\alpha}$ and sinusoidal in $\hat{\phi}$; thus, the local minima of F with respect to these two variables are the global minima. However, F is not so well behaved as a function of $\hat{\omega}$. A plot of $F(\hat{\omega})$ for $N = 40$, $\omega = 0.4\pi$, $\phi = \hat{\phi} = 0$, and $\alpha = \hat{\alpha} = 1$ is shown in Fig. 1. From this figure, it is apparent that the initial estimate of $\hat{\omega}$ must be within the deep valley if we expect the Gauss–Newton algorithm to find the global minimum.

In Appendix B we show that the width of the valley is in the range $\Delta\omega \in [2\pi/N, 4\pi/N]$. Thus, the initial estimate must have a standard deviation that is on the order of $1/N$. However, the standard deviation of $\hat{\omega}$ estimates obtained from (3.5) are $0(1/L\sqrt{N})$ (when $L < N^{1/2}$), which asymptotically is too large for use with the Gauss–Newton method. Thus, we need to improve the initial frequency estimates before starting the minimization.

It is known [7] that, for large N, the ML estimates of

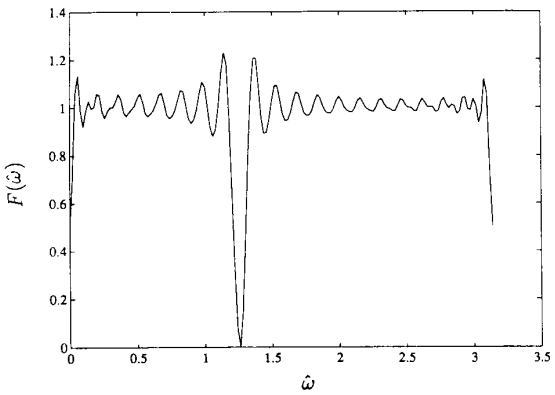


Fig. 1. A plot of $F(\hat{\omega})$ for $N = 40$, $\omega = 0.4\pi$, $\phi = \hat{\phi} = 0$, and $\alpha = \hat{\alpha} = 1$.

$\{\omega_i\}$ are approximately given by the maxima of the periodogram. Therefore, one method for improving initial frequency estimates is to search in some small interval about each $\hat{\omega}_i$ say $[\hat{\omega}_i - \epsilon, \hat{\omega}_i + \epsilon]$, for the maximum of the periodogram. Note that we need only perform m one-dimensional searches (one for each frequency) instead of a more computationally intensive search on an m-dimensional region; this simplification is possible because the LF in (5.2a) asymptotically decouples to the sum of m independent functions. Thus, the following search procedure can be used.

Choose appropriate values for $\Delta\omega$ and $l_{\max}$.

For each $i = 1, 2, \cdots, m$:

1) compute the periodogram α_{il} of the data at frequencies

$$\tilde{\omega}_{il} = \hat{\omega}_i \pm l\Delta\omega \qquad l = 0, 1, \cdots, l_{\max} \tag{5.4}$$

using

$$\alpha_{il} = (\beta_{il}^2 + b_{il}^2)^{1/2},$$

where β_{il} and b_{il} are computed using (3.15) but with $M = N$;

2) choose as the new initial frequency estimate the $\hat{\omega}_{il}$ whose corresponding α_{il} is largest; compute the new initial amplitude and phase estimates using (3.15).

From the above discussion, $\Delta\omega$ should be chosen less than $2\pi/N$ to ensure that one of the $\hat{\omega}_{il}$ is in the deep trough of LF; in our simulations, we used $\Delta\omega\epsilon[2\pi/3N, \pi/N]$. In addition, $l_{\max}$ should be chosen so that Pr $\{\omega_i\epsilon[\hat{\omega}_i - l_{\max} \cdot \Delta\omega, \hat{\omega}_i + l_{\max} \cdot \Delta\omega]\}$ is sufficiently large. For example, since the OYW method was used to obtain $\hat{\omega}_i$, $l_{\max}$ could be chosen based on the asymptotic probability distribution of the $\hat{\omega}_i$ given in [12]. Finally, one must ensure that the search interval for two adjacent frequencies do not overlap.

Because the improvement step is periodogram based, it is unable to resolve sinusoids whose frequencies are within $2\pi/N$ of each other. To avoid adverse effects of this resolution limit, we imposed the following restriction to the improvement step: no search interval could include a "dead zone" of width $4\pi/N$ between any two frequencies. Specifically, for two adjacent OYW frequency estimates $\hat{\omega}_i$ and $\hat{\omega}_{i+1}$, no improvement search could include the interval $[\bar{\omega} - (2\pi/N), \bar{\omega} + (2\pi/N)]$ where $\bar{\omega} = (\hat{\omega}_i + \hat{\omega}_{i+1})/2$. This restriction, in essence, permits the improvement step to operate only in cases where the finite resolution of the periodogram will not adversely affect the estimation procedure.

A consequence of the restriction is that, in most cases, if the OYW frequency estimates are within $4\pi/N$ of each other, no improvement takes place. Thus, in order for the Gauss–Newton algorithm to converge, the OYW frequency estimates must lie in the narrow trough of the loss function. For large N, the OYW frequency estimates have a standard deviation of $c/\sqrt{N}$, where c is some constant that depends on (among other things) the SNR and the frequency separation. If the frequency separation and/or SNR are too small, the initial OYW estimates will not be sufficiently accurate and the Gauss–Newton algorithm will converge to a false local minimum; it is this effect that results in a threshold SNR for good performance of the composite algorithm for closely spaced sinuosids.

In order to reduce the threshold SNR, other algorithms for obtaining initial estimates for the ML Gauss–Newton algorithm can be used. The overdetermined and high-order Yule–Walker (HOYW) method of Cadzow and Chan [3], [5], [6] and the eigenstructure analysis (EA) method of Tufts and Kumaresan [4], [8] can be used to obtain accurate frequency estimates. It can be shown (see [11]) that the asymptotic estimation errors $\{\hat{\omega}_i - \omega_i\}$ corresponding to these two methods are of the order $1/L^{3/2}N^{1/2}$, where L denotes the number of YW equations and the model order used. Thus, for large L, the parameter estimates obtained by HOYW or EA methods are likely to lie in the deep valley of the likelihood loss function and can, therefore, be used for properly initializing the ML Gauss–Newton algorithm (a periodogram-based "refining" step being no longer needed). Note that the HOYW and EA methods are expected to be more accurate than the OYW/periodogram-based procedure and to have lower SNR thresholds, but they are also more complex computationally (a main computational task for both HOYW and EA lies in finding the roots of a large degree polynomial). In general, one should choose an initialization estimation algorithm which is the most computationally simple, and which is effective at resolving the sinusoidal frequencies for the data length and SNR of interest.

In summary, the initial estimates provided to the iterative ML Gauss–Newton algorithm are obtained by the following procedure.

i) Determine rough initial estimates of $\{\omega_i, \alpha_i, \text{and } \theta_i\}$ by the OYW method (3.5), (3.6), (3.15).

ii) Obtain improved initial estimates of $\{\omega_i, \alpha_i, \text{and } \theta_i\}$ by the periodogram-based procedure (5.4).

We used the initialization method above in the numerical experiments, which we report in this paper.

VI. Numerical Examples

We present some numerical experiments that indicate the performance of the proposed algorithm. We first consider the problem of estimating θ from a signal of the form (2.1), where $m = 2$,

$$\alpha_1 = 1.0 \quad \omega_1 = 0.25\pi \quad \phi_1 = 1.0$$

$$\alpha_2 = 1.0 \quad \omega_2 = 0.2\pi \quad \phi_2 = 0.0.$$

In all examples, $L = \sqrt{N}$, and M is chosen as in (3.17). The white noise variance λ^2 was varied so that the SNR ranges between 10–40 dB in 5 dB increments. (Here, SNR is defined for each signal, e.g., SNR $= \alpha_1^2/2\lambda^2$.) For each SNR, J independent data sets were generated, and average sum-squared errors (SSE) of the resulting estimates $K_N\hat{\theta}$ were computed. For each element of θ, the normalized SSE is computed by

$$\frac{1}{J}\sum_{j=1}^{J}\left[(K_N)(\hat{\theta}_j - \theta)\right]_i^2$$

($J = 50$ in these simulations) where $[\ ]_i$ denotes the ith element of the vector, and where $\hat{\theta}_j$ is the jth estimate vector.

The SSE of the estimated parameters for $N = 500$ are shown in Figs. 2–4. In these figures, initial estimates are those obtained using the method of Section III. Equation (3.15) was used for estimates in these plots; however, the SSE for estimates obtained using (3.9) are not significantly different (and, in particular, no better on the average). From these initial estimates, improved estimates were obtained as outlined in the previous section, then the Gauss–Newton algorithm [equation (4.8)] was used. In (4.8), μ_k was at each iteration set to 1; if LF increased, μ_k was decreased by a factor of 4 until the resulting step was such that LF decreased.

From these figures, it is apparent that the SSE's of the ML estimates are very close to the Cramer–Rao bound for SNR's at and above 20 dB. (In a few instances, the estimated SSE is below the CR lower bound; this is a result of using a relatively small number of Monte–Carlo experiments.) High SSE's for SNR's below this threshold are due to poor initial estimates which result in convergence to a false local minimum of the loss function. For a given initial estimate algorithm, the threshold depends most strongly on the relative frequency separation $\Delta\omega \cdot N$; as this ratio decreases, the threshold SNR increases.

The effect of the improvement step in the estimation procedure can be seen in Table I. Listed in this table are the number of Monte Carlo experiments (out of the 50) which gave "good" initial estimates from the HOYW equations alone, and from the HOYW equations followed by the periodogram-based improvement step. Good initial estimates are those from which the ML algorithm converges to the global minimum of the loss function. For SNR's in the 5–20 dB range, the improvement step is effective in increasing the likelihood that the ML algorithm will converge to the global minimum. For SNR's below or above this range, the improvement step is of little use. Note that the improvement step lowers the threshold SNR by 5 dB in this case. Note also that even though only one of the 50 Monte Carlo experiments for 15 dB had a poor initial guess, the SSE's for this case in Figs. 2–4 are much

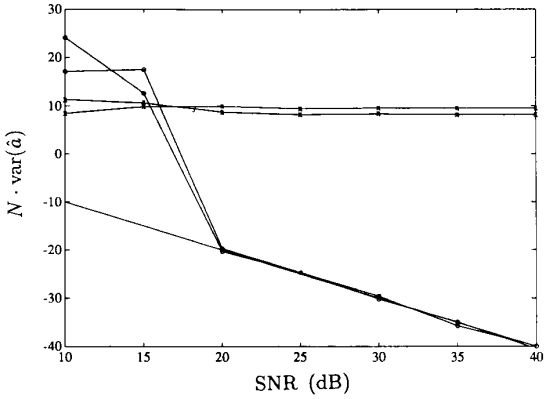


Fig. 2. SSE's of amplitude estimates for $N = 500$ and white noise. Curves with "x" are initial estimates; curves with "o" are ML estimates, and the solid line is the Cramer–Rao bound.

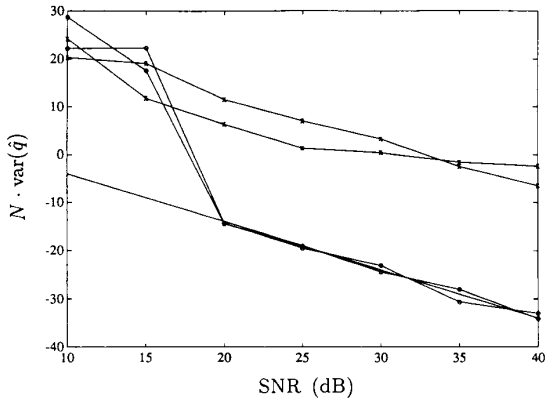


Fig. 3. SSE's of phase estimates for $N = 500$ and white noise. Curve labels are as in Fig. 2.

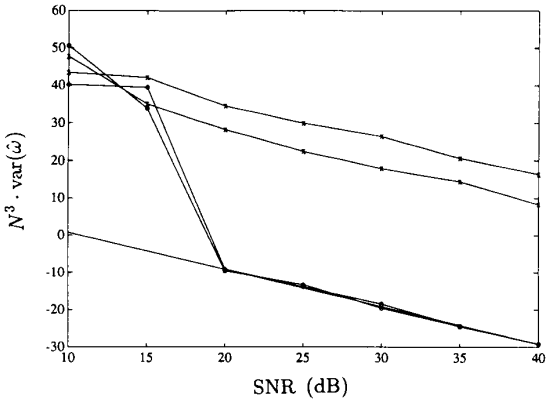


Fig. 4. SSE's of frequency estimates for $N = 500$ and white noise. Curve labels are as in Fig. 2.

TABLE I
NUMBER OF CORRECT ML ESTIMATES OBTAINED FROM 50 MONTE CARLO SIMULATIONS WITH AND WITHOUT THE INITIAL FREQUENCY ESTIMATE IMPROVEMENT STEP

SNR (dB)	Without Improvement	With Improvement
0	0	1
5	4	30
10	28	46
15	41	49
20	42	50
25	50	50

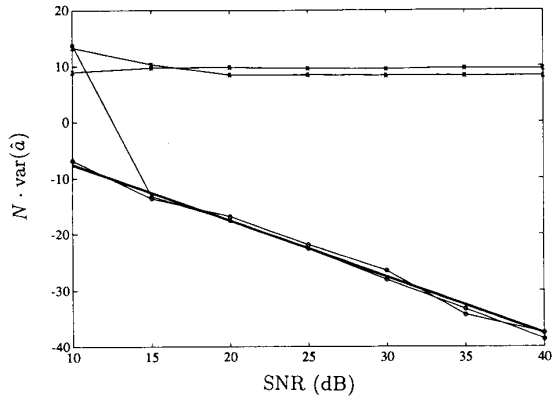


Fig. 5. SSE's of amplitude estimates for $N = 500$ and colored noise. Curve labels are as in Fig. 2.

higher than the CR lower bound; thus, the SSE for one bad estimate dominates the total SSE in these figures.

Figs. 5–7 show average error results for $N = 500$ data points when the additive noise is colored. The noise used is MA(1):

$$n(t) = [\epsilon(t) + 0.9\epsilon(t-1)]/\sqrt{1.81}.$$

Note that $n(t)$ has the same total power as $\epsilon(t)$ does. It can be seen that the ML method provides significant improvement over Yule–Walker estimates for colored noise. The Yule–Walker method does not give consistent estimates in this case because the first row of (3.5) should not be used (the data can be modeled as a limiting ARMA (4, 5) process, in which case (3.5) holds only for $k \geq 2$). However, for large L, the effect of the first equation is small, and "reasonable" estimates could still result (as seen in the figures). We note also that for colored noise, the proposed method is not a maximum likelihood estimate, but it is still an output error method.

The minimization procedure converged in a few iterations in most cases. For the above example, the average number of iterations in the 50 Monte Carlo experiments ranged from 9 to 4.5 (as the SNR ranged from 10–40 dB) to achieve a tolerance of 10^{-4} (where no element of $K_N\hat{\theta}$ changed more than "tolerance" in one iteration). This rapid convergence rate was seen for $(\Delta\omega/2\pi)N$ greater than about 4; for smaller frequency differences,

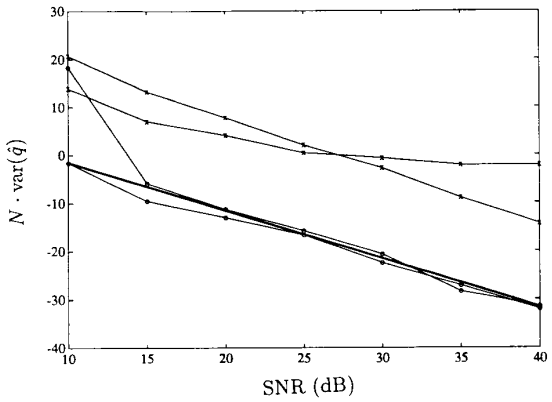


Fig. 6. SSE's of phase estimates for $N = 500$ and colored noise. Curve labels are as in Fig. 2.

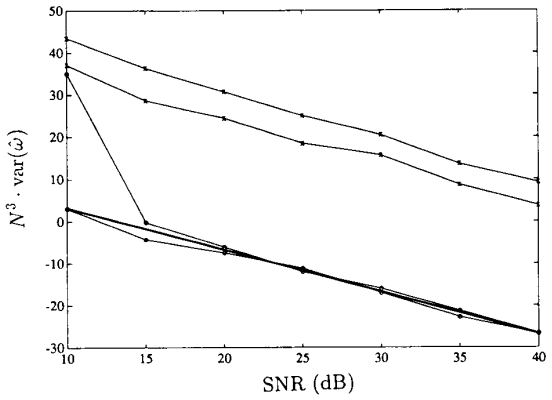


Fig. 7. SSE's of frequency estimates for $N = 500$ and colored noise. Curve labels are as in Fig. 2.

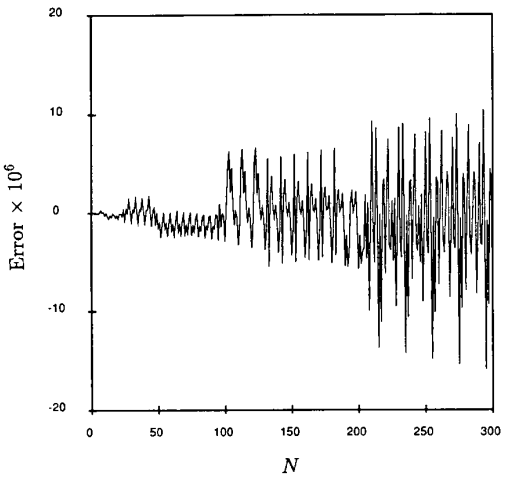


Fig. 8. Error between $\sin(\omega t)$ computed directly and computed by difference equation for $\omega = 0.4\pi$.

the convergence was slower. Slower convergence is not unexpected in this case because the algorithm was derived under the assumption that $\Delta\omega >> 2\pi/N$.

As a final note, the recursive computation of sin ωt or cos ωt using (3.11) required only about 1/6 the CPU time of direct computation. The error between the recursively and directly computed values remained below 10^{-3} for $N < 1000$ (using single-precision arithmetic); a typical plot of the error is shown in Fig. 8.

VII. Conclusions

We derived a simplified maximum-likelihood Gauss–Newton algorithm for estimating the parameters of sinusoidal signals in noise. The algorithm has the computational simplicity of a pure gradient technique and the convergence rate of a Newton method. It can be initialized by a set of preliminary estimates obtained via the overdetermined Yule–Walker method and a periodogram-based procedure.

The asymptotic properties of the proposed techniques are discussed, and it is shown that the parameter estimates are consistent and asymptotically efficient for the Gaussian case. In the non-Gaussian case, the estimator provides a minimum-variance solution within a large class of estimators based on second-order statistics.

The finite-sample performance of the proposed technique was studied by Monte Carlo simulations. It was shown that the maximum-likelihood Gauss–Newton procedure can improve the accuracy of the initial estimates significantly. Comparisons to the performance corresponding to the CRLB were also presented, using a simple expression for the asymptotic CRLB covariance matrix derived in this paper.

Appendix A
Cramer–Rao Lower Bounds

The estimation problem formulated in Section II falls into the class of nonlinear regression problems. The CRLB, say P_{CR}^N, for any unbiased estimator of θ and λ^2 can be easily derived [7]. In this appendix, we will be interested in the asymptotic CRLB: P_{CR}^∞. The reason for this interest is threefold.

i) P_{CR}^∞ has a much simpler expression than P_{CR}^N and is, therefore, much easier to compute. Yet P_{CR}^∞ is a good approximation of P_{CR}^N whenever

$$\inf_{i \neq j} |\omega_i - \omega_j| > \frac{2\pi}{N}. \tag{A.1}$$

This will become apparent in the following, where it will be shown that the smaller the minimum separation in frequency $\inf |\omega_i - \omega_j|$, the slower the convergence of P_{CR}^N to P_{CR}^∞. It is worth noting that a main conclusion of the study of P_{CR}^N in [7] was that P_{CR}^N increases rapidly as the minimum frequency separation goes below the critical value $2\pi/N$. In such a case, P_{CR}^N is much larger than P_{CR}^∞.

ii) P_{CR}^N can be attained only under certain restrictive conditions [10] which apparently are not satisfied for the problem under study. On the other hand, P_{CR}^∞ is attained in the limit (as $N \to \infty$) by the covariance matrix of the ML estimate; see Theorem 4.1. Furthermore, for other estimation methods (such as the OYW method) only asymptotic results are available. Thus, it is P_{CR}^∞ which is of interest in any analytical study comparing the performance of the ML method to that of other estimation methods.

iii) The expression of P_{CR}^∞ is useful in the derivation of the simplified ML Gauss–Newton algorithm in Section IV.

Note than an expression for P_{CR}^∞ does not seem to be available in the literature, except for the special case of $m = 1$; see [9] and its references.

For the estimation problem under discussion, the log-likelihood function is given by

$$L(\theta, \lambda^2) = -\frac{N}{2} \ln (2\pi) - \frac{N}{2} \ln \lambda^2 - \frac{1}{2\lambda^2} \sum_{t=1}^{N} \epsilon^2(t), \tag{A.2}$$

where

$$\epsilon(t) = y(t) - \sum_{i=1}^{m} \alpha_i \sin (\omega_i t + \phi_i). \tag{A.3}$$

The CRLB,

$$P_{CR}^N = \left[E\left\{ \begin{bmatrix} \dfrac{\partial L(\theta, \lambda^2)}{\partial \theta} \\ \dfrac{\partial L(\theta, \lambda^2)}{\partial \lambda^2} \end{bmatrix} \cdot \left[\left(\frac{\partial L(\theta, \lambda^2)}{\partial \theta} \right)^T \frac{\partial L(\theta, \lambda^2)}{\partial \lambda^2} \right] \right\} \right]^{-1}, \tag{A.4}$$

can be evaluated by straightforward calculations:

$$\frac{\partial L}{\partial \theta} = -\frac{1}{\lambda^2} \sum_{t=1}^{N} \epsilon(t)\, \epsilon_\theta(t)$$

$$\frac{\partial L}{\partial \lambda^2} = -\frac{N}{2\lambda^2} + \frac{1}{2\lambda^4} \sum_{t=1}^{N} \epsilon^2(t) \tag{A.5}$$

$$E\left[\frac{\partial L}{\partial \theta} \cdot \left(\frac{\partial L}{\partial \theta} \right)^T \right] = \frac{1}{\lambda^2} \sum_{t=1}^{N} \epsilon_\theta(t)\, \epsilon_\theta^T(t)$$

$$E\left[\frac{\partial L}{\partial \theta} \cdot \frac{\partial L}{\partial \lambda^2} \right] = \frac{N}{2\lambda^4} \sum_{t=1}^{N} \epsilon_\theta(t)\, E\{\epsilon(t)\} - \frac{1}{2\lambda^6} \sum_{t=1}^{N} \sum_{s=1}^{N} \cdot \epsilon_\theta(t)\, E\{\epsilon(t)\, \epsilon^2(s)\} = 0$$

$$E\left[\frac{\partial L}{\partial \lambda^2} \right]^2 = \frac{N^2}{4\lambda^4} - \frac{N^2}{2\lambda^4} + \frac{1}{4\lambda^8} \sum_{t=1}^{N} \sum_{s=1}^{N} E\{\epsilon^2(t)\, \epsilon^2(s)\}$$

$$= -\frac{N^2}{4\lambda^4} + \left[\frac{3N}{4\lambda^4} + \frac{N(N-1)}{4\lambda^4} \right]$$

$$= \frac{N}{2\lambda^4}. \tag{A.6}$$

In (A.5) and (A.6), we used the assumption that $\epsilon(t)$ is white Gaussian noise. It follows that

$$P_{CR}^{N} = \begin{bmatrix} P_{CR}^{\theta,N} & 0 \\ 0 & \dfrac{2\lambda^4}{N} \end{bmatrix}, \tag{A.7}$$

where

$$P_{CR}^{\theta,N} = \lambda^2 \left[\sum_{t=1}^{N} \epsilon_\theta(t)\, \epsilon_\theta^T(t) \right]^{-1}, \tag{A.8}$$

and where the derivatives of $\epsilon(t)$ with respect to the parameters $\{\alpha_i, \phi_i, \omega_i\}$ are given by (4.3). The expression (A.8) for $P_{CR}^{\theta,N}$ appears, for example, in [7]. However, the calculations necessary to show that P_{CR}^{N} has the block-diagonal form of (A.7), which in turn implies that $P_{CR}^{\theta,N}$ is given by (4.8), were not included there.

In the following, we will study the limit of $P_{CR}^{\theta,N}$ as $N \to \infty$. The following results will be useful for this study.

Lemma A.1: For $\omega \in [0, 2\pi)$

$$\frac{1}{N}\sum_{t=1}^{N} \cos(\omega t + \phi) = \begin{cases} \cos\phi & \text{for } \omega = 0 \\ \dfrac{1}{N}\dfrac{\sin(N\omega/2)\cos[(N+1)\omega/2 + \phi]}{\sin(\omega/2)} & \text{for } \omega \neq 0. \end{cases} \tag{A.9}$$

Proof: [17].

Corollary: For $\omega \in [0, 2\pi)$

$$\lim_{N\to\infty} \frac{1}{N^{k+1}} \sum_{t=1}^{N} t^k \cos(\omega t + \phi) = \begin{cases} \dfrac{1}{k+1}\cos\phi & \omega = 0 \\ 0 & \omega \neq 0 \end{cases}, \quad k \geq 0. \tag{A.10}$$

Proof: For $k = 0$, the limit follows immediately from (A.9). For $k > 0$, the limits follow from relations similar to (A.9) obtained by differentiation of (A.9) with respect to ω.

Let us denote

$$\bar{P}_{CR}^{\theta,N} = K_N P_{CR}^{\theta,N} K_N, \tag{A.11}$$

where K_N is given by (4.4). Clearly, $\bar{P}_{CR}^{\theta,N}$ is the CRLB on the covariance matrix of the following normalized estimation error vector:

$$\begin{bmatrix} \sqrt{N}(\hat{\alpha} - \alpha) \\ \sqrt{N}(\hat{\phi} - \phi) \\ N\sqrt{N}(\hat{\omega} - \omega) \end{bmatrix}, \tag{A.12}$$

where $\alpha = [\alpha_1, \cdots, \alpha_m]^T$, and $\hat{\alpha}$ is any unbiased estimator of α. ϕ, $\hat{\phi}$, and ω, $\hat{\omega}$ are similarly defined. In the following we will show that

$$P_{CR}^{\theta} \triangleq \lim_{N\to\infty} \bar{P}_{CR}^{\theta,N} \tag{A.13}$$

exists and has a simple expression.

By making repeated use of Lemma A.1 and its corollary, we can write

$$\lim_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N} \frac{\partial\epsilon(t)}{\partial\alpha_i} \cdot \frac{\partial\epsilon(t)}{\partial\alpha_j} = \frac{1}{2}\lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}\{\cos[(\omega_i - \omega_j)t + \phi_i - \phi_j] - \cos[(\omega_i + \omega_j)t + \phi_i + \phi_j]\} = \frac{1}{2}\delta_{i,j}$$

$$\lim_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N} \frac{\partial\epsilon(t)}{\partial\alpha_i} \cdot \frac{\partial\epsilon(t)}{\partial\phi_j} = \frac{\alpha_j}{2}\lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}\{\sin[(\omega_i + \omega_j)t + \phi_i + \phi_j] + \sin[(\omega_i - \omega_j)t + \phi_i - \phi_j]\} = 0$$

$$\lim_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N} \frac{\partial\epsilon(t)}{\partial\phi_i} \cdot \frac{\partial\epsilon(t)}{\partial\phi_j} = \frac{\alpha_i\alpha_j}{2}\lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}\{\cos[(\omega_i - \omega_j)t + \phi_i - \phi_j] + \cos[(\omega_i + \omega_j)t + \phi_i + \phi_j]\} = \frac{\alpha_i\alpha_j}{2}\delta_{i,j}$$

$$\lim_{N\to\infty} \frac{1}{N^2}\sum_{t=1}^{N} \frac{\partial\epsilon(t)}{\partial\alpha_i} \cdot \frac{\partial\epsilon(t)}{\partial\omega_j} = \frac{\alpha_j}{2}\lim_{N\to\infty}\frac{1}{N^2}\sum_{t=1}^{N} t\{\sin[(\omega_i + \omega_j)t + \phi_i + \phi_j] + \sin[(\omega_i - \omega_j)t + \phi_i - \phi_j]\} = 0$$

$$\lim_{N\to\infty} \frac{1}{N^2}\sum_{t=1}^{N} \frac{\partial\epsilon(t)}{\partial\phi_i} \cdot \frac{\partial\epsilon(t)}{\partial\omega_j} = \frac{\alpha_i\alpha_j}{2}\lim_{N\to\infty}\frac{1}{N^2}\sum_{t=1}^{N} t\{\cos[(\omega_i - \omega_j)t + \phi_i - \phi_j] + \cos[(\omega_i + \omega_j)t + \phi_i + \phi_j]\} = \frac{\alpha_i\alpha_j}{4}\delta_{i,j}$$

$$\lim_{N\to\infty} \frac{1}{N^3}\sum_{t=1}^{N} \frac{\partial\epsilon(t)}{\partial\omega_i} \cdot \frac{\partial\epsilon(t)}{\partial\omega_j}$$

$$= \frac{\alpha_i\alpha_j}{2} \lim_{N\to\infty} \frac{1}{N^3} \sum_{t=1}^{N} t^2 \left\{ \cos\left[(\omega_i - \omega_j)t + \phi_i - \phi_j\right] + \cos\left[(\omega_i + \omega_j)t + \phi_i + \phi_j\right] \right\}$$

$$= \frac{\alpha_i\alpha_j}{6} \delta_{i,j} \tag{A.14}$$

where $\delta_{i,j}$ denotes the Dirac delta function (3.3b). Therefore,

$$P_{CR}^{\theta} = \lambda^2 \left[\begin{array}{c|ccc|ccc} \frac{1}{2} I_m & 0 & & & 0 & & \\ \hline & \frac{\alpha_1^2}{2} & & 0 & \frac{\alpha_1^2}{4} & & 0 \\ 0 & & \ddots & & & \ddots & \\ & 0 & & \frac{\alpha_m^2}{2} & 0 & & \frac{\alpha_m^2}{4} \\ \hline & \frac{\alpha_1^2}{4} & & 0 & \frac{\alpha_1^2}{6} & & 0 \\ 0 & & \ddots & & & \ddots & \\ & 0 & & \frac{\alpha_m^2}{4} & 0 & & \frac{\alpha_m^2}{6} \end{array}\right]^{-1}, \tag{A.15}$$

which, after some straightforward calculations, gives

$$P_{CR}^{\theta} = 2\lambda^2 \left[\begin{array}{c|ccc|ccc} I_m & 0 & & & 0 & & \\ \hline & \frac{4}{\alpha_1^2} & & 0 & -\frac{6}{\alpha_1^2} & & 0 \\ 0 & & \ddots & & & \ddots & \\ & 0 & & \frac{4}{\alpha_m^2} & 0 & & -\frac{6}{\alpha_m^2} \\ \hline & -\frac{6}{\alpha_1^2} & & 0 & \frac{12}{\alpha_1^2} & & 0 \\ 0 & & \ddots & & & \ddots & \\ & 0 & & -\frac{6}{\alpha_m^2} & 0 & & \frac{12}{\alpha_m^2} \end{array}\right]. \tag{A.16}$$

Note that the bounds for phases and frequencies are proportional to the noise-to-signal ratios corresponding to the frequency in question. However, somewhat contrary to intuition, the bound for the amplitudes of the sinusoids is independent of these amplitudes. Note also the almost diagonal structure of P_{CR}^{θ}. The estimation errors of the phase and frequency of the same sinusoid are asymptotically cross correlated. All the other estimation errors are asymptotically uncorrelated.

It is also interesting to note that the bounds for $(\hat{\omega}_i - \omega_i)$ are of order $N^{-3/2}$ (see also [9] and its references). This order of the CRLB is rather unusual for a stationary estimation problem for which the corresponding bounds are, in general, of the order $N^{-1/2}$. However, the problem of estimating the parameters of a sinusoidal signal is not a strictly stationary estimation problem: the derivative of $\epsilon(t)$ with respect to ω_i is clearly a nonstationary signal.

It follows from Lemma A.1 that the smaller the minimum frequency separation $\inf_{i\neq j} |\omega_i - \omega_j|$, the slower the convergence in (A.13). Consider, for example, (A.9) for ω small but nonzero. Then the left-hand side of (A.9) will generally be small provided that $N\omega$, rather than N, is large enough [see the right-hand side of (A.9)].

Appendix B
Width of the Loss Function Trough

We restrict attention to the case $\omega \approx \hat{\omega}$, $\alpha \approx \hat{\alpha}$, and $\phi \approx \hat{\phi}$. From (5.3), the derivative of F with respect to $\hat{\omega}$ is

$$\frac{\partial F}{\partial \hat{\omega}} = \frac{1}{N} \sum_{t=1}^{N} t\left[\alpha \sin(\omega t + \phi) - \hat{\alpha} \sin(\hat{\omega} t + \hat{\phi})\right] \cdot \left[-2\hat{\alpha} \cos(\hat{\omega} t + \hat{\phi})\right] \tag{B.1a}$$

$$= -\frac{\alpha\hat{\alpha}}{N} \sum_{t=1}^{N} t \sin\left[(\omega + \hat{\omega})t + (\phi + \hat{\phi})\right] + \frac{\hat{\alpha}^2}{N} \sum_{t=1}^{N} t \sin(2\hat{\omega} t + 2\hat{\phi}) - \frac{\alpha\hat{\alpha}}{N} \sum_{t=1}^{N} t \sin\left[(\omega - \hat{\omega})t + (\phi - \hat{\phi})\right]. \tag{B.1b}$$

We claim that the zeros of $\partial F/\partial\hat{\omega}$ in the region of interest are nearly equal to those of the third term of (B.1b). To support this, a plot of

$$\frac{1}{N} \sum_{t=1}^{N} t \sin \omega t \tag{B.2}$$

versus ω is shown for $N = 100$ in Fig. 9. It can be seen that for ω not near zero, this function is near zero. Also, near $\omega = 0$, the zero crossings have large slopes and are therefore insensitive to small additive disturbances.

Defining $\tilde{\omega} = \omega - \hat{\omega}$ and $\tilde{\phi} = \phi - \hat{\phi}$, the third term in (B.1b) can be expressed as

$$-\frac{\alpha\hat{\alpha}}{N} \sum_{t=1}^{N} t \sin(\tilde{\omega} t + \tilde{\phi}) = -\alpha\hat{\alpha}\left[\cos\tilde{\phi}\left(\frac{1}{N}\sum_{t=1}^{N} t \sin\tilde{\omega} t\right) + \sin\tilde{\phi}\left(\frac{1}{N}\sum_{t=1}^{N} t\cos\tilde{\omega} t\right)\right]. \tag{B.3}$$

Since $\tilde{\phi}$ is small, the second term in (B.3) can be neglected. Thus, for $\omega \approx \hat{\omega}$, $\alpha \approx \hat{\alpha}$, and $\phi \approx \hat{\phi}$, the zeros of $\partial F/\partial\hat{\omega}$ are nearly those of the function

$$\frac{1}{N} \sum_{t=1}^{N} t \sin \tilde{\omega} t. \tag{B.4}$$

It is not difficult to see that (B.4) is zero for $\tilde{\omega} = 0$. More-

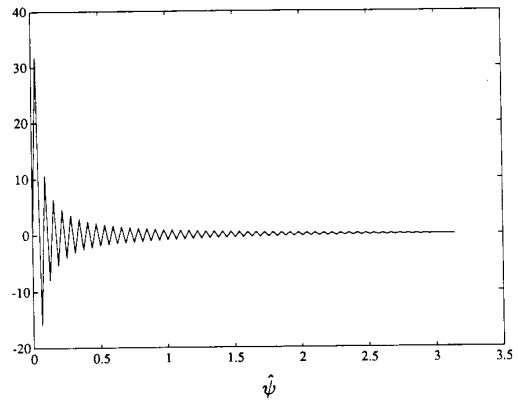


Fig. 9. A plot of $(1/N)\sum_{t=1}^{N} t \sin(\hat{\psi}t)$ versus $\hat{\psi}$ for $N = 100$.

over, for $0 < \tilde{\omega} < \pi/N$, (B.4) is positive (since each element of the sum is positive). For $\tilde{\omega} = 2\pi/N$, (B.4) is negative for large N; thus, the first zero of (B.4) occurs for $\tilde{\omega} \in [\pi/N, 2\pi/N]$. Since (B.4) is an odd function of $\tilde{\omega}$, and since the zeros of (B.4) are approximately equal to the zeros of $\partial F/\partial\hat{\omega}$, we conclude that the width of the main valley of $F(\hat{\omega})$ is in the range

$$(\omega - \hat{\omega}) \in \left[\frac{\pi}{N} - \left(-\frac{\pi}{N}\right), \frac{2\pi}{N} - \left(\frac{-2\pi}{N}\right)\right] = \left[\frac{2\pi}{N}, \frac{4\pi}{N}\right].$$

Appendix C
Proof of Theorem 4.1

Note that

$$K_N^{-1} \sum_{t=1}^{N} \epsilon_\theta(t, \tilde{\theta})\, \epsilon(t, \tilde{\theta}) = 0. \tag{C.1}$$

Thus, for large N, we can write

$$0 = K_N^{-1} \sum_{t=1}^{N} \epsilon_\theta(t, \theta)\, \epsilon(t) + F(\theta)\, K_N(\tilde{\theta} - \theta) + \frac{1}{2} \sum_{i=1}^{3m} (\tilde{\theta}_i - \theta_i) \frac{\partial F(\theta)}{\partial \theta_i} K_N(\tilde{\theta} - \theta) + \cdots, \tag{C.2}$$

where θ_i is the ith component of θ and

$$F(\theta) = K_N^{-1} \sum_{t=1}^{N} \left\{\epsilon_\theta(t, \theta)\, \epsilon_\theta^T(t, \theta) + \epsilon_{\theta\theta}(t, \theta)\, \epsilon(t, \theta)\right\} K_N^{-1}. \tag{C.3}$$

The first term in (C.2) is asymptotically independent of N. To see this, note that its asymptotic covariance matrix, say P, is given by [see (4.5b) and (4.6)]

$$\begin{aligned} P &\triangleq \lim_{N\to\infty} E\left\{K_N^{-1}\left[\sum_{t=1}^{N}\sum_{s=1}^{N} \epsilon_\theta(t, \theta) \cdot \epsilon_\theta^T(s, \theta)\, \epsilon(t)\, \epsilon(s)\right] K_N^{-1}\right\} \\ &= \lambda^2 \lim_{N\to\infty} K_N^{-1}\left[\sum_{t=1}^{N} \epsilon_\theta(t, \theta)\, \epsilon_\theta^T(t, \theta)\right] K_N^{-1} \\ &= \lambda^4 (P_{CR}^{\theta})^{-1}, \end{aligned} \tag{C.4}$$

where P_{CR}^{θ} is defined in Appendix A. The last equality in (C.4) is also proven in Appendix A.

Next we show that for large N,

$$K_N^{-1}\left[\sum_{t=1}^{N} \epsilon_{\theta\theta}(t, \theta)\, \epsilon(t)\right] K_N^{-1} = 0(1/\sqrt{N}). \tag{C.5}$$

The matrix $\epsilon_{\theta\theta}(\cdot, \cdot)$ of second-order derivatives is given by

$$\epsilon_{\theta\theta}(t, \theta) = \begin{bmatrix} 0 & -\text{diag}\left[\cos(\omega_i t + \phi_i)\right] & -\text{diag}\left[t\cos(\omega_i t + \phi_i)\right] \\ -\text{diag}\left[\cos(\omega_i t + \phi_i)\right] & \text{diag}\left[\alpha_i \sin(\omega_i t + \phi_i)\right] & \text{diag}\left[t\alpha_i \sin(\omega_i t + \phi_i)\right] \\ -\text{diag}\left[t\cos(\omega_i t + \phi_i)\right] & \text{diag}\left[t\alpha_i \sin(\omega_i t + \phi_i)\right] & \text{diag}\left[t^2\alpha_i \sin(\omega_i t + \phi_i)\right] \end{bmatrix}, \tag{C.6}$$

where each block of the matrix has size $m \times m$. The generic element of the matrix in (C.5) can therefore be written as

$$V_N \triangleq \frac{\alpha}{N^{\beta+1}} \sum_{t=1}^{N} t^\beta \sin(\omega t + \phi)\, \epsilon(t), \tag{C.7}$$

where $\beta = \{0, 1, \text{or } 2\}$, $\alpha = \{\pm\alpha_i \text{ or } \pm 1\}$, $\omega = \omega_i$, and $\phi = \{\phi_i \text{ or } \phi_i + (\pi/2)\}$. The variance of V_N is readily evaluated:

$$\begin{aligned} E\{V_N^2\} &= \frac{\alpha^2}{N^{2\beta+2}} E\left\{\sum_{t=1}^{N}\sum_{s=1}^{N} t^\beta s^\beta \sin(\omega t + \phi) \cdot \sin(\omega s + \phi)\, \epsilon(t)\, \epsilon(s)\right\} \\ &= \frac{\alpha^2\lambda^2}{N^{2\beta+2}} \sum_{t=1}^{N} t^{2\beta} \sin^2(\omega t + \phi). \end{aligned}$$

Thus,

$$E\{V_N^2\} \le \text{const.} \frac{1}{N^{2\beta+2}} \sum_{t=1}^{N} t^{2\beta} = 0(1/N),$$

which proves (C.5).

It follows from the calculations above and from Appendix A that

$$F(\theta) = \lambda^2 (P_{CR}^{\theta})^{-1} + 0(1/\sqrt{N}).$$

Next we show that the higher order terms in (C.2) can be neglected asymptotically. Note that

$\dfrac{\partial F(\theta)}{\partial \theta_i}$ is of the same order of magnitude as $F(\theta)$,

$$\text{for } i = 1, \cdots, 2m,$$

$\dfrac{\partial F(\theta)}{\partial \theta_i}$ is of the order of magnitude of $F(\theta) \cdot N$,

$$\text{for } i = 2m + 1, \cdots, 3m.$$

Since

$$(\tilde{\theta}_i - \theta_i) = \begin{cases} 0(1/\sqrt{N}), & i = 1, \cdots, 2m, \\ 0(1/N\sqrt{N}), & i = 2m + 1, \cdots, 3m, \end{cases}$$

and since $F(\theta)$ is asymptotically independent of N as shown above, we conclude that the higher order terms in (C.2) are $0(1/\sqrt{N})$. Thus, for large N,

$$K_N(\tilde{\theta} - \theta) = -\frac{1}{\lambda^2} P_{CR}^{\theta} \cdot K_N^{-1} \sum_{t=1}^{N} \epsilon_{\theta}(t, \theta)\, \epsilon(t)$$

which implies that [cf. (C.4)]

$$\lim_{N \to \infty} E\left\{K_N(\tilde{\theta} - \theta)(\tilde{\theta} - \theta)^T K_N\right\}$$

$$= \frac{1}{\lambda^2} P_{CR}^{\theta} \cdot \lambda^4 (P_{CR}^{\theta})^{-1} \cdot \frac{1}{\lambda^2} P_{CR}^{\theta} = P_{CR}^{\theta}.$$

References

[1] *Proc. IEEE*, Special Issue on Spectral Estimation, vol. 70, Sept. 1982.

[2] S. Kay and S. L. Marple, "Spectrum analysis—A modern perspective," *Proc. IEEE*, vol. 69, pp. 1380–1419, Nov. 1981.

[3] J. A. Cadzow, "Spectral estimation: An overdetermined rational model equation approach," *Proc. IEEE*, vol. 70, pp. 907–939, Sept. 1982.

[4] D. W. Tufts and R. Kumaresan, "Estimation of frequencies of multiple sinusoids: Making linear prediction perform like maximum likelihood," *Proc. IEEE*, vol. 70, pp. 975–989, Sept. 1982.

[5] Y. T. Chan and R. P. Langford, "Spectral estimation via the high-order Yule–Walker equations," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-30, pp. 689–698, Oct. 1982.

[6] J. A. Cadzow, "High performance estimation—A new ARMA method," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, pp. 524–529, Oct. 1980.

[7] D. C. Rife and R. R. Boorstyn, "Multiple tone parameter estimation from discrete-time observations," *Bell Syst. Tech. J.*, pp. 1389–1410, Nov. 1976.

[8] D. W. Tufts and R. Kumaresan, "Improved spectral resolution II," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, pp. 392–397, Apr. 1980.

[9] S. W. Lang and J. H. McClellan, "Frequency estimation with maximum entropy spectral estimators," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, pp. 716–723, Dec. 1980.

[10] M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics*, vol. II. London: Griffin, 1966.

[11] P. Stoica, B. Friedlander, and T. Söderström, "On the accuracy of high-order Yule-Walker estimates of frequencies of multiple sinusoids," in *Proc. 1986 Int. Conf. Acoust., Speech, Signal Processing*, Tokyo, Japan, Apr. 1986, pp. 609–612.

[12] P. Stoica, T. Söderström, F. Ti, and B. Friedlander, "Overdetermined Yule-Walker estimation of frequencies of multiple sinusoids: Some accuracy aspects," in *Proc. 8th IFAC Symp. Syst. Identification Parameter Estimation*, Beijing, China, Aug. 1988.

[13] T. Söderström and P. Stoica, *Instrumental Variable Methods for System Identification*. Berlin: Springer-Verlag, 1983.

[14] P. Stoica and T. Söderström, "Optimization with respect to a covariance sequence: Algorithmic aspects and some applications," Rep. UPTEC 83112R, Inst. Technol., Uppsala Univ., Sweden, Nov. 1983. A short version appeared as P. Stoica and T. Söderström, "Optimization with respect to covariance sequence parameters," *Automatica*, vol. 21, pp. 671–675, 1985.

[15] P. Stoica, T. Söderström, and B. Friedlander, "Optimal instrumental variable estimates of the AR parameters of an ARMA process," *IEEE Trans. Automat. Contr.*, vol. AC-30, pp. 1066–1074, Nov. 1985.

[16] L. Ljung and P. E. Caines, "Asymptotic normality of prediction error estimates for approximate system models," *Stochastics*, vol. 3, pp. 29–46, 1979.

[17] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*. New York: Academic, 1980.

[18] T. Abatzoglou, "A fast maximum-likelihood algorithm for frequency estimation of a sinusoid based on Newton's method," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 77–89, Feb. 1985.

[19] S. Parthasarathy and D. W. Tufts, "Maximum-likelihood estimation of parameters of exponentially damped sinusoids," *Proc. IEEE*, vol. 73, pp. 1528–1530, Oct. 1985.

[20] Y. Bresler and A. Macovski, "Exact maximum likelihood parameter estimation of superimposed exponential signals in noise," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 1081–1089, Oct. 1986.

[21] R. Kumaresan, L. L. Scharf, and A. K. Shaw, "An algorithm for pole-zero modeling and spectral analysis," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 637–640, June 1986.

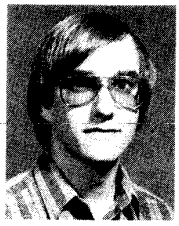
[22] R. K. Mehra, "On-line identification of linear dynamic systems with applications to Kalman filtering," *IEEE Trans. Automat. Contr.*, vol. AC-16, pp. 12–21, Feb. 1971.

[23] P. Stoica and A. Nehorai, "Statistical analysis of two nonlinear least-squares estimators of sine wave parameters in the colored noise case," in *Proc. 1988 Int. Conf. Acoust., Speech, Signal Processing*, New York, Apr. 11–14, 1988.

Petre Stoica received the M.Sc. and Ph.D. degrees, both in automatic control, in 1972 and 1979, respectively.

Since 1972 he has been with the Department of Automatic Control, the Polytechnic Institute of Bucharest, Romania. His research interests include various aspects of system identification, time series analysis, and signal processing. For papers on these topics he received three national prizes. He is co-author of five books, the most recent being *System Identification* (Englewood Cliffs, NJ: Prentice-Hall, 1988).

Dr. Stoica is a member of the Board of Directors of the Time Series Analysis and Forecasting (TSA&F) Society, and an Associate Editor for the *Journal of Forecasting*. He was given the Member of TSA&F (MTSA&F) honors award.



Randolph L. Moses (S'78–M'85) received the B.S., M.S., and Ph.D. degrees in electrical engineering from Virginia Polytechnic Institute and State University in 1979, 1980, and 1984, respectively.

During the summer of 1983 he was an SCEEE Summer Faculty Research Fellow at Rome Air Development Center, Rome, NY. From 1984 to 1985 he was with the Eindhoven University of Technology, Eindhoven, The Netherlands, as a NATO Postdoctoral Fellow. Since 1985 he has been an Assistant Professor in the Department of Electrical Engineering, The Ohio State University. His research interests are in digital signal processing, and include parametric time series analysis, system identification, and model reduction.

Dr. Moses is a member of Eta Kappa Nu, Tau Beta Pi, Phi Kappa Phi, and Sigma Xi.

Benjamin Friedlander (S'74-M'76-SM'82-F'87) received the B.Sc. and M.Sc. degrees in electrical engineering from the Technion-Israel Institute of Technology in 1968 and 1972, respectively, and the Ph.D. degree in electrical engineering and the M.Sc. degree in statistics from Stanford University, Stanford, CA, in 1976.

From 1968 to 1972 he served in the Israel Defence Forces as an Electronic Engineer. From 1976 to 1985 he was at Systems Control Technology, Inc., Palo Alto, CA, as Manager of the Advanced Technology Division. He was responsible for a number of research and development projects in the areas of signal processing and control. During this period he was also a Lecturer at Stanford University, teaching graduate courses on linear systems, nonlinear detection and estimation, and system identification. In November 1985 he joined Saxpy Computer Corporation, Sunnyvale, CA, as the Director of Advanced Technology. He was responsible for the research and development activities of the company and for the development of signal processing applications for the MATRIX 1—a 1000 Mflop supercomputer. He now heads Signal Processing Technology, Ltd., in Palo Alto, CA. His current interests include: parallel and systolic processing architectures and their VLSI implementations, advanced processing techniques for underwater surveillance, high resolution spectral analysis and array processing, adaptive filtering, radar and communication processing, detection tracking and localization of multiple targets, image processing, and knowledge based signal processing.

Dr. Friedlander has over 210 publications in the areas of signal processing and estimation. He was an Associate Editor of the IEEE TRANSACTIONS ON AUTOMATIC CONTROL in 1984, and is a member of the Administrative Committee of the Acoustics, Speech, and Signal Processing (ASSP) Society. He serves as a member of the Technical Committee on Spectrum Estimation of the ASSP Society, and is the Vice Chairman of the Bay Area Chapter of that society. He is also a member of the IEEE Admission and Advancement Committee. He was the recipient of the 1983 ASSP Senior Award for the paper "Recursive lattice forms for spectral estimation," and the 1985 Award for the Best Paper of the Year from the European Association for Signal Processing (EURASIP) for the paper "On the computation of an asymptotic bound for estimating autoregressive signals in white noise." He is a member of Sigma Xi.

Torsten Söderström (M'76-SM'83) was born in Malmö, Sweden, in 1945. He received the M.S. degree ("civilingenjör") in engineering physics in 1969 and the Ph.D. degree in automatic control in 1973, both from the Lund Institute of Technology, Lund, Sweden. In 1976 he was awarded the title of Docent in automatic control.

In the period 1967-1974 he held various teaching positions at the Lund Institute of Technology. Since 1974 he has been working at the Institute of Technology, Uppsala University, Uppsala, Sweden, where he also is the Head of the Automatic Control and Systems Analysis Group since 1975. He has been employed as Lecturer and Docent and is currently Professor of Automatic Control. He is author or coauthor of many technical papers. His main research interests are in the fields of system identification, signal processing, process control, and adaptive systems. In 1981 he was, with coauthors, given an Automatica Prize Paper Award. He is a coauthor of three books: *Theory and Practice of Recursive Identification* (Cambridge, MA: M.I.T. Press, 1983), with L. Ljung; *The Instrumental Variable Methods for System Identification* (New York: Springer, 1983), with P. Stoica; and *System Identification* (Englewood Cliffs, NJ: Prentice-Hall, 1988), with P. Stoica.