Optimal High-Order Yule–Walker Estimation of Sinusoidal Frequencies

Petre Stoica, Randolph L. Moses, *Senior Member, IEEE*, Torsten Söderström, *Senior Member, IEEE*, and Jian Li, *Student Member, IEEE*

Abstract—**The asymptotic properties of sinusoidal frequency estimators based on the high-order Yule–Walker (HOYW) equations were analyzed recently. The results of that analysis are used here to propose two classes of frequency estimators; one class uses singular value decomposition, and the other uses a sparse matrix solution. Both classes entail two estimation steps: the first step generates initial estimates which are used to obtain an optimal weighting matrix, and the second step generates an optimally weighted estimate. Each two-step method produces asymptotically minimum variance estimates over all estimators of their class. The implementation of the proposed estimators is described in detail and numerical examples are presented to evaluate their performance.**

I. Introduction

THE problem of estimating the frequencies of sinusoidal signals from noise-corrupted measurements has attracted considerable interest [1]–[19]. A number of solutions to this problem have been proposed and analyzed. The Yule–Walker equation approach to sinusoidal frequency estimation has been discussed in a large number of papers [1], [2], [5], [15], [16]. It leads to numerically simple estimators which often exhibit satisfactory statistical properties. The high resolution of the overdetermined and high-order Yule–Walker (HOYW) estimators was proven in many numerical simulations. However, a theoretical analysis of these Yule–Walker methods was lacking.

Recently we analyzed the accuracy properties of the HOYW estimator [9], [10], [12]. Among other things, this analysis provided theoretical explanation of the empirically observed fact that estimation accuracy increases considerably with increasing model order and number of Yule–Walker equations.

This paper can be viewed as a continuation of the work reported in [9], [10], [12]. The insight into the problem of sinusoidal frequency estimation gained by the theoretical analyses in these references is used here to propose two classes of two-step optimal HOYW estimators. The implementation of these optimal estimators is discussed in detail and their performance is studied and compared.

Manuscript received May 25, 1989; revised June 25, 1990.

P. Stoica is with the Department of Automatic Control, Polytechnic Institute of Bucharest, R 77 206 Bucharest, Romania.

R. L. Moses and J. Li are with the Department of Electrical Engineering, Ohio State University, Columbus, OH 43210.

T. Söderström is with the Automatic Control and Systems Analysis Group, Department of Technology, Uppsala University, S-751 21 Uppsala, Sweden.

IEEE Log Number 9132802.

II. High-Order ARMA Representation of the Sinusoids in Noise Process

Consider the following sinusoidal signal:

$$x(t) = \sum_{k=1}^{m} \gamma_k \sin(\omega_k t + \psi_k) \tag{2.1}$$

where $\gamma_k > 0$, $\omega_k \in (0, \pi)$, $\omega_i \neq \omega_j$ for $i \neq j$, and $\{\psi_k\}$ are mutually independent uniform random variables distributed on $[0, 2\pi]$. Let $y(t)$ denote the noise corrupted measurement of $x(t)$

$$y(t) = x(t) + e(t), \qquad t = 1, 2, \cdots \tag{2.2}$$

where $e(t)$ is a sequence of independent and identically distributed random variables with zero mean and variance λ^2. It is assumed that $x(t)$ and $e(s)$ are uncorrelated for all t and s. For later use, define

$$r_k = E\{y(t)y(t+k)\}. \tag{2.3}$$

The problem is to estimate the angular frequencies $\{\omega_k\}_{k=1}^m$ from N samples of noisy measurements $\{y(1), \cdots, y(N)\}$.

It is well known that $x(t)$ obeys the following homogeneous autoregressive (AR) equation [5]:

$$A(q^{-1})x(t) = 0 \tag{2.4}$$

where q^{-1} denotes the unit delay operator $[q^{-1}x(t) = x(t-1)]$, and $A(q^{-1})$ is given by

$$\begin{aligned} A(q^{-1}) &\triangleq 1 + a_1 q^{-1} + \cdots + a_n q^{-n} \\ &= \prod_{k=1}^{m} (1 - 2\cos\omega_k q^{-1} + q^{-2}), \qquad n \triangleq 2m. \end{aligned} \tag{2.5}$$

From (2.2) and (2.4) it follows that $y(t)$ obeys the following autoregressive moving-average (ARMA) equation:

$$A(q^{-1})y(t) = A(q^{-1})e(t). \tag{2.6}$$

Note from (2.5) that the poles (and zeros) of (2.6) are located on the unit circle at $e^{\pm i\omega_k}$, $k = 1, \cdots, m$. It follows from (2.6) that $y(t)$ also obeys the following high-

1053-587X/91/0600-1360$01.00 © 1991 IEEE

order ARMA equation:

$$C(q^{-1})y(t) = C(q^{-1})e(t) \tag{2.7a}$$

where

$$C(q^{-1}) = c_0 + c_1 q^{-1} + \cdots + c_L q^{-L} \triangleq B(q^{-1})A(q^{-1}) \tag{2.7b}$$

and where $B(q^{-1})$ is an arbitrary polynomial of degree $L - n$:

$$B(q^{-1}) = b_0 + b_1 q^{-1} + \cdots + b_{L-n} q^{-(L-n)}. \tag{2.7c}$$

The reason for considering a high-order ARMA model as in (2.7a) is that frequency estimates which use the high-order model often have lower variances than estimates obtained for the minimal order model (i.e., when $C(q^{-1}) = A(q^{-1})$). This has been shown both experimentally and theoretically by a number of authors; see, e.g., [1], [2], [9], [10], and the next sections.

III. HOYW Equations and Their Solutions

It is well known that if a time series satisfies the difference equation (2.7a), the corresponding c_i coefficients satisfy the following (high order) Yule–Walker equations:

$$\begin{bmatrix} r_{L+1} & r_L & \cdots & r_1 \\ r_{L+2} & r_{L+1} & \cdots & r_2 \\ \vdots & & \ddots & \vdots \\ r_{L+M} & \cdots & \cdots & r_M \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_L \end{bmatrix} = 0, \qquad L, M \geq n. \tag{3.1}$$

Both the model order L and the number M of equations in (3.1) can be chosen so as to improve the accuracy of the Yule–Walker frequency estimates (see the next sections).

The solutions of the linear system of equations (3.1) are given by (2.7b), where the $\{a_i\}$ coefficients are given by (2.5) and the $\{b_i\}$ coefficients are arbitrary. A particular algorithm for solving (3.1) corresponds to a particular choice of a set of $\{b_i\}$ coefficients.

The standard HOYW-based frequency estimation problem consists of first determining a particular solution to equation (3.1). Then one forms $C(z)$ and finds its zeros. Finally, the zeros of $C(z)$ are separated into the "signal zeros" of $A(z)$ and the "spurious zeros" of $B(z)$, cf. (2.7b). Most often, these zeros are separated by their relationship to the unit circle; for example, zeros whose magnitudes are closest to one are chosen as signal zeros.[1] Therefore, to avoid incorrectly choosing a spurious zero as a signal zero, it is desirable that $B(z)$ have the property that its zeros are all bounded away from the unit circle.

To apply standard numerical algorithms to (3.1) we need to constrain $\{c_k\}$ in some way. The most convenient constraint is to set $c_k = 1$, for some $k \in [0, L]$. In the following, for conciseness of the discussion, we assume that

$$c_0 = 1, \qquad \text{(or equivalently } b_0 = 1\text{)}. \tag{3.2}$$

Other constraints of the form $c_k = 1$ for $k \in [1, L]$ will lead to similar algorithms.

Next we introduce the following notation:

$$R = \begin{bmatrix} r_L & \cdots & r_1 \\ \vdots & & \vdots \\ r_{L+M-1} & \cdots & r_M \end{bmatrix} \tag{3.3a}$$

$$r = [r_{L+1} \cdots r_{L+M}]^T \tag{3.3b}$$

$$\theta = [c_1 \cdots c_L]^T. \tag{3.3c}$$

With this notation, the YW equations (3.1) along with (3.2) can be written compactly as

$$R\theta = -r. \tag{3.4}$$

For any nonsingular $M \times M$ matrix Q the system of equations (3.4) is equivalent to

$$QR\theta = -Qr. \tag{3.5}$$

The reason for introducing Q in (3.5) can be explained as follows. In the known covariance case the solutions of (3.5) do not depend on Q. However, in the unknown covariance case, when the system of equations (3.5) no longer holds exactly, its "solution" will depend on Q, and hence Q can be selected so as to improve estimation accuracy. The choice of Q is addressed in Section V.

Let $(QR)^-$ denote any pseudoinverse of QR. (A^- is called a pseudoinverse of a matrix A iff $AA^-A = A$). Then

$$\theta = -(QR)^-(Qr) \tag{3.6}$$

is a solution of (3.5). In the following we discuss two possible choices of the pseudoinverse in (3.6), leading to two particularly useful solutions θ.

A. Minimum-Norm Solution

Let

$$(QR)^- = (QR)^+ \tag{3.7}$$

where $(\cdot)^+$ denotes the Moore–Penrose pseudoinverse. Then (3.6) becomes

$$\theta_1 \triangleq -(QR)^+(Qr) \tag{3.8}$$

which is known to be the minimum Euclidean norm solution of (3.5) [14], [20]. (We will use the notation $(\cdot)_1$ to denote quantities corresponding to the minimum norm solution above.) Numerically well-behaved algorithms for computing the Moore–Penrose pseudoinverse are readily available (for example, the singular value decomposition (SVD) procedure). Furthermore, the polynomial

$$C_1(z) \triangleq B_1(z)A(z) \tag{3.9a}$$

[1] Theoretically, the zeros of $A(z)$ have modulus exactly equal to one. In practice, numerical roundoff and inaccuracies in the $\{r_k\}$ coefficients result in signal zeros which do not lie exactly on the unit circle.

corresponding to θ_1 is such that [3], [9]

$$B_1(z) = 0 \Rightarrow |z| > 1 \tag{3.9b}$$

for any finite L. This property makes it possible to extract $A(z)$ from $C_1(z)$. Note, however, that zeros of $B_1(z)$ tend to the unit circle as L increases [9], [10]. Thus, for too large a value of L it may be difficult to separate the zeros of $A(z)$ from those of $B_1(z)$ once $C_1(z)$ has been found.

B. Sparse Solution

Let Γ be an $M \times n$ matrix made up of any n columns of R. We can express Γ compactly by

$$\Gamma = RJ \begin{bmatrix} 0 \\ I_n \end{bmatrix} \tag{3.10}$$

where J is an $L \times L$ (orthogonal) permutation matrix formed by reordering the rows of the identity matrix I_L. The last n columns of RJ gives Γ. Thus,

$$R = [X \quad \Gamma]J^T \tag{3.11}$$

where the explicit expression of X has no importance for the following discussion. Let the matrix Γ be chosen to have full rank n (see [9], [15])

$$\text{rank } \{\Gamma\} = n. \tag{3.12}$$

Introduce

$$(QR)^- = J \begin{bmatrix} 0 \\ (\Gamma^T Q^T Q \Gamma)^{-1} \Gamma^T Q^T \end{bmatrix}. \tag{3.13}$$

The matrix on the right-hand side of (3.13) is a pseudoinverse of (QR). This can be shown as follows. Since rank $\{R\}$ = rank $\{\Gamma\} = n$, the columns of X must be linearly dependent on the columns of Γ. Thus

$$X = \Gamma Z \tag{3.14}$$

for some $n \times (L - n)$ matrix Z. Using this observation, we can write

$$\begin{aligned}(QR)(QR)^-(QR) &= [QX \quad Q\Gamma]J^T J \begin{bmatrix} 0 \\ (\Gamma^T Q^T Q \Gamma)^{-1} \Gamma^T Q^T \end{bmatrix} [QX \quad Q\Gamma]J^T \\ &= Q\Gamma(\Gamma^T Q^T Q \Gamma)^{-1} \Gamma^T Q^T Q \Gamma [Z \quad I_n]J^T \\ &= Q\Gamma[Z \quad I_n]J^T = QR\end{aligned}$$

which shows that (3.13) is indeed a pseudoinverse of QR. It follows from (3.6) and (3.13) that

$$\theta_2 = J \begin{bmatrix} 0 \\ \alpha \end{bmatrix} \tag{3.15a}$$

is a solution of (3.5), where

$$\alpha = -(\Gamma^T W \Gamma)^{-1} (\Gamma^T W r) \tag{3.15b}$$

and where

$$W = Q^T Q. \tag{3.15c}$$

Since for $L >> n$ the vector θ_2 contains many zero elements, we call it a "sparse solution." We denote quantities corresponding to the sparse solution as $(\cdot)_2$.

There are numerically well-behaved algorithms for computing α. Note that α in (3.15b) is the least squares solution of the following system of equations:

$$Q\Gamma\alpha = -Qr. \tag{3.16}$$

The so-called QR algorithm [20] can be used to determine α given by (3.16). It is worth stressing that solving the linear system (3.16) is a much simpler task than evaluating θ_1 in (3.8), unless $L = n$ (for $L = n$ (3.8) and (3.16) are identical). The price paid for the numerical simplicity of (3.16) is that the polynomial $C_2(z) = B_2(z)A(z)$ corresponding to θ_2 is no longer guaranteed to be such that the zeros of $B_2(z)$ are bounded away from the unit circle. Extraction of $A(z)$ and $C_2(z)$ may therefore be difficult in some cases; however, this difficulty has rarely surfaced in the numerical experiments we performed (see Section VI).

We end this section with some comments on possible extensions and modifications of the two solutions above.

Remark 3.1: In Section III-B we pointed out that different sparse solutions of (3.5) can be obtained by selecting different sets of n linearly independent columns of R to form Γ. For example, the n "most linearly independent" columns of R could be chosen. This choice of columns is defined as HOYWE-3, and discussed in more detail in Section VI.

Remark 3.2: Use of other pseudoinverses in (3.6) may lead to frequency estimators which, depending on the case under study, may have better properties than the estimators corresponding to (3.8) or (3.15). This aspect is discussed in [19].

IV. HOYW Estimators

In this section we treat the unknown covariance case. Let $\hat{r}_k$ denote a consistent estimate of r_k. For example,

$$\hat{r}_k = \frac{1}{N-k} \sum_{t=1}^{N-k} y(t)y(t+k) \tag{4.1}$$

where N is the number of data points. Replace $\{r_k\}$ by $\{\hat{r}_k\}$ in R, Γ, and r and denote the results by $\hat{R}$, $\hat{\Gamma}$, and $\hat{r}$. In the following, we introduce two estimators of θ based on the expressions (3.8) and (3.15) developed in the previous section. These two estimators are called HOYWE-1 and HOYWE-2, respectively. Note that once an estimate $\hat{\theta}$ is available, the frequencies $\{\hat{\omega}_k\}$ can be determined from the zeros of the polynomial $\hat{C}(z)$ corresponding to $\hat{\theta}$. More specifically, let $\{\hat{\rho}_k e^{\pm i\hat{\omega}_k}\}_{k=1}^m$ denote the m complex-conjugate zeros of $\hat{C}(z)$ which are nearest the unit circle. The angular positions $\{\hat{\omega}_k\}$ of these zeros are taken as the estimates of $\{\omega_k\}$. The number m of sinusoids is assumed to be known, so $n = 2m$ is also known. Procedures for determining m are described in [13], [17].

A. HOYWE-1

First we note that replacement of R and r in (3.8) by $\hat{R}$ and $\hat{r}$ produces an estimator of questionable utility since

the matrix $Q\hat{R}$ is very likely to be ill conditioned. It is well known (see, e.g., [1], [2], [9]) that a much better estimator is the following one, which we call HOYWE-1:

$$\hat{\theta}_1 = -(\overline{Q\hat{R}})^+ (Q\hat{r}) \tag{4.2}$$

where $\overline{Q\hat{R}}$ is the best rank n approximation of $Q\hat{R}$ in the Frobenius norm sense [20]. The Moore–Penrose pseudoinverse of $Q\hat{R}$ is given by ([14], [20])

$$(\overline{Q\hat{R}})^+ = V\Sigma^{-1}U^T \tag{4.3}$$

where $\Sigma = \text{diag}\,(\sigma_1, \cdots, \sigma_n)$, $\{\sigma_i\}$ are the n principal singular values of $Q\hat{R}$, and V and U^T are the corresponding matrices of right and left singular vectors.

B. HOYWE-2

HOYWE-2 is obtained by replacing the true autocovariances in Γ and r in (3.15) by estimates obtained from equation (4.1). The resulting estimator is given by

$$\hat{\theta}_2 = J\begin{bmatrix} 0 \\ \hat{\alpha} \end{bmatrix}, \qquad \hat{\alpha} = (\hat{\Gamma}^T W \hat{\Gamma})^{-1} \hat{\Gamma}^T W \hat{r}. \tag{4.4}$$

For $L = n$, (4.4) reduces to the overdetermined weighted Yule–Walker estimator. For $L > n$ the estimator (4.4) appears to be new. Note that $\hat{\alpha}$ should be determined as the least squares solution of $Q\hat{\Gamma}\alpha = -Q\hat{r}$ using, for example, the QR algorithm (see (3.16)). The evaluation of $\hat{\alpha}$ using (4.4) is not recommended for numerical computations because it is not as well numerically conditioned as a QR algorithm based solution.

V. Optimal HOYW Estimators and Their Implementation

The (asymptotic) accuracy properties of the HOYW estimators have been derived in [9], [19], [21]. In particular, the choice of Q, L, and M to achieve the maximum possible accuracy of the frequency estimates has been studied in these papers. It was shown that the following choice of Q maximizes the estimation accuracy for given L and M:

$$Q = S^{-1/2} \tag{5.1}$$

where

$$S = \frac{1}{\lambda^2} E\left\{ C(q^{-1}) \begin{bmatrix} e(t-1) \\ \vdots \\ e(t-M) \end{bmatrix} \cdot C(q^{-1}) \, [e(t-1) \cdots e(t-M)] \right\}. \tag{5.2}$$

It was also shown that the optimal accuracy corresponding to the choice (5.1) of Q increases significantly with increasing M or L (the optimal accuracy increases monotonically as M increases, but not necessarily monotonically as L increases).

The above results are general in the sense that they apply to the HOYW estimators obtained from any solution of the Yule–Walker equations (3.5). In the following, the S matrices corresponding to the specific solutions (3.8) and (3.15) considered in this paper will be denoted by S_1 and S_2, respectively. Thus, each S_i is given by (5.2) with $C(q^{-1}) = C_i(q^{-1})$ where $C_i(q^{-1})$ corresponds to θ_i, $i = 1, 2$.

To summarize, for improved estimation accuracy we should use the minimum norm and sparse HOYW estimators with: a) $Q = S_i^{-1/2}$, $i = 1, 2$; and b) large L and M. Two issues which must still be addressed are those of how to handle the problem that Q in (5.1) is not known, and how large to choose L and M. These are discussed below.

A. The Optimal HOYWE-1

The matrix S_1 in (5.2) can be written as

$$S_1 = \phi_1^T \phi_1 \tag{5.3}$$

where

$$\phi_1^T = \begin{bmatrix} 1 & c_{11} & \cdots & c_{1L} & & 0 \\ & \ddots & \ddots & & \ddots & \\ 0 & & 1 & c_{11} & \cdots & c_{1L} \end{bmatrix}, \qquad (M \times (M+L)) \tag{5.4}$$

and $\{c_{1i}\}$ are the elements of θ_1. Let $\hat{\phi}_1$ denote a consistent estimates of ϕ_1 obtained by replacing $\{c_{1i}\}$ in (5.4) by some consistent estimates $\{\hat{c}_{1i}\}$. It can be shown that use of $\hat{\phi}_1^T \hat{\phi}_1$ instead of S_1 in the optimal HOYWE-1 does not affect the asymptotic accuracy [9]. Thus, the following two-step procedure can be used to implement the optimal HOYWE-1.

Step 1: Use (4.2) with $Q = I$ and some large values of M and L to calculate initial estimates of the $\{c_{1i}\}$ parameters.

Step 2: Using the estimates from step 1 to form $\hat{\phi}_1$, determine optimal HOYW estimates of $\{c_{1i}\}$ with (4.2) where Q is set to $Q = (\hat{\phi}_1^T \hat{\phi}_1)^{-(1/2)}$ (since the matrix $(\hat{\phi}_1^T \hat{\phi}_1)$ is Toeplitz, its inverse square root can be efficiently computed). Finally, estimate the frequencies $\{\omega_k\}$ from the zeros of the estimated polynomial $\hat{C}_1(z)$.

Next we turn to the problem of choosing L and M. The values of L and M should be only a fraction of the number N of data points to guarantee that the high-lag sample covariances are computed with reasonable statistical accuracy. The value of L is limited also by another consideration. If L is too large, the zeros of $\hat{B}_1(z)$ may be located too close to the unit circle to be reliably separated from the zeros of $\hat{A}_1(z)$, and this may produce spurious frequency estimates as already explained. Therefore some upper bounds on L and M should be imposed. The values of these bounds, however, are problem dependent and there is no general rule for choosing them.

Some guidelines for choosing L and M in specific cases can be found in [1], [2], [9], [10] and in the next section. Here we present some further general considerations that should be taken into account when choosing L and M.

The value of M should not be "much larger" than L. This is so since for fixed L the minimum eigenvalue of S_1 goes to zero when M increases [10]. Thus, for $M >> L$ the calculations in step 2 of the two-step procedure above are expected to be ill conditioned. Conversely, if we let L be much larger than M then it can be shown that ([9])

$$S_1^{-1} = [I + O(1/L)]^{-1} = I + O(1/L) \qquad (5.5)$$

which implies that in such a case it is approximately optimal to set $Q = I$. However, M should not be too small compared to L since the accuracy increases considerably with increasing M. Moreover, the two-step procedure requires solving an algebraic equation of degree L. The task of solving a high order equation should be avoided. Recall also that for large values of L spurious frequency estimates may occur.

To summarize the above discussion, we recommend to use the two-step procedure with M larger than L, but not much larger, to guarantee that the calculations in step 2 are well conditioned.

B. The Optimal HOYWE-2

Similarly to the discussion in Section V-A, we can use the following two-step procedure to implement the optimal HOYWE-2 (corresponding to the choice (5.1), (5.2) of Q).

Step 1: Use (4.4) with $Q = I$ and large values of L and M, to obtain initial estimates of the $\{c_{2i}\}$ parameters.

Step 2: Use the estimates from step 1 to form $\hat{\phi}_2$ (an estimate of ϕ_2 defined similarly to ϕ_1 (5.4)). Next, compute optimal estimates of $\{c_{2i}\}$ with (4.4), where Q is set to $Q = (\hat{\phi}_2^T \hat{\phi}_2)^{-(1/2)}$. Finally, estimate the frequencies $\{\omega_k\}$ from the zeros of the estimated polynomial $\hat{C}_2(z)$.

Concerning the choice of L and M, the principal conclusions pertaining to this choice which were presented in Section V-A still apply. Thus we recommend using the optimal HOYWE-2 procedure with M larger than L, but not "much larger."

VI. Numerical Examples

In this section we present some numerical examples which illustrate the performance of the above HOYW algorithms. The examples consider the case of two closely spaced sinusoids in white noise. The frequencies, amplitudes, and phases of the sinusoids are given by: $\omega_1 = 1.00$, $\omega_2 = 1.05$; $\gamma_1 = \gamma_2 = 1$; and $\psi_1 = \psi_2 = 0$. The noise variance (λ^2) is chosen to give a signal-to-noise ratio of 10 dB (where SNR = 10 log ($\gamma_1^2/2\lambda^2$)). The numerical results were obtained using Matlab, and operations such as SVD and polynomial root finding were performed using standard Matlab functions.

The tables and figures in this section give normalized sum-squared errors (SSE's) $N \cdot \mathrm{SSE}(\hat{\omega}_{i1})$ for $i = 1$ or 2 (the subscript i indicates the HOYWE variant in question, i.e., HOYWE-1 or 2). The theoretical SSE's were computed using the formulas derived in [9], [10], whereas the empirical SSE's were obtained by averaging the results from 50 Monte Carlo experiments. Only results for one of the two frequency estimates, $\hat{\omega}_{i1}$, are shown; the results for $\hat{\omega}_{i2}$ are similar. All SSE values are given in decibels. For comparison purposes, the asymptotic Cramér–Rao lower bound (CRLB) for the variance of each frequency estimates is given by $(24\lambda^2)/(\gamma_i^2 N^3)$; for this example, $N \cdot \mathrm{CRLB} = -42.7$ dB when $N = 150$ and -62.7 dB when $N = 1500$ ($N \cdot$ CRLB is given to facilitate comparison with the $N \cdot$ SSE values in the figures).

TABLE I
THEORETICAL NORMALIZED VARIANCES OF $\hat{\omega}_{11}$ (IN DECIBELS) FOR THE INITIAL, MINIMUM NORM ESTIMATOR ($N = 150$)

L	$M = 5$	$M = 9$	$M = 13$	$M = 17$	$M = 21$
5	23.3	5.99	1.37	−1.93	−7.78
9	7.20	0.34	−6.02	−12.1	−16.0
13	2.01	−5.08	−11.6	−16.7	−19.6
17	1.63	−9.14	−13.9	−17.3	−21.9
21	−4.48	−13.7	−17.5	−21.2	−25.2

TABLE II
THEORETICAL NORMALIZED VARIANCES OF $\hat{\omega}_{11}$ (IN DECIBELS) FOR THE TWO-STEP OPTIMAL, MINIMUM NORM ESTIMATOR ($N = 150$)

L	$M = 5$	$M = 9$	$M = 13$	$M = 17$	$M = 21$
5	23.3	5.42	−1.55	−7.26	−13.6
9	7.19	0.32	−6.11	−12.6	−17.5
13	2.01	−5.10	−11.8	−16.7	−19.9
17	1.63	−9.15	−13.9	−17.3	−22.0
21	−4.48	−13.8	−17.5	−21.3	−25.3

Table I shows the theoretical SSE's of $\hat{\omega}_{11}$ for several values of L and M, corresponding to the initial minimum norm estimate ($\hat{\theta}_1$ with $Q = I$). Note that the error decreases by nearly 50 dB as L and M increase. Table II shows the variances for the two-step optimum minimum norm estimate; here, similar decrease in error is seen for increasing L and M. Note that the optimal variances are smaller than the initial variances for large M and small L; in other ranges the initial and optimum variances are almost equal. The latter observation is in agreement with the theoretical discussion immediately following (5.5). The former implies that the optimal two-step procedure with $L = L_1$ can achieve the same statistical accuracy as the initial procedure with $L = L_2$ greater than L_1. In particular, the optimum two-step procedure with $L = n$ may provide satisfactory statistical performance without the risk of spurious frequency estimates that often occur when L is chosen greater than n. This is an important advantage which may constitute enough motivation for the second (optimal) step of the two-step procedure.

Fig. 1 shows theoretical and simulation SSE's of $\hat{\omega}_{11}$ for the minimum norm estimator HOYWE-1 with $M = 20$, for two values of the number of data points: $N = 150$ and $N = 1500$. It can be seen that the optimal SSE is significantly smaller than the initial SSE only for small L.

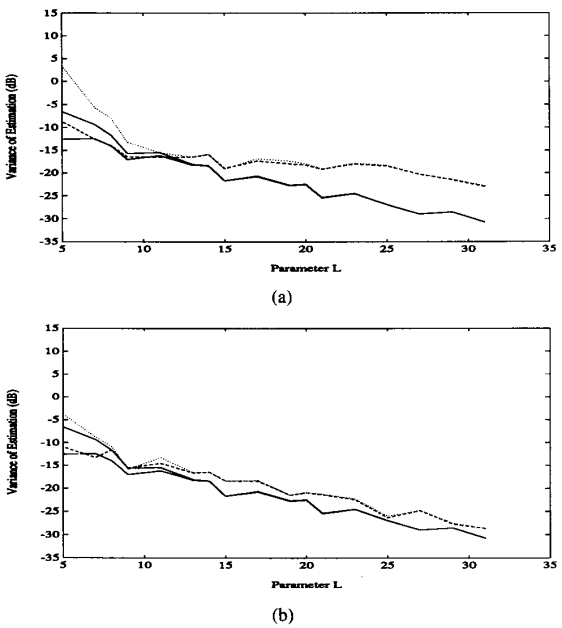


Fig. 1. Normalized variances of $\hat{\omega}_{11}$ using the minimum-norm estimator ($M = 20$). Solid lines are theoretical variances (initial and optimal). The dotted line is the estimated initial variance with $Q = I$, and the dashed line is the estimated two-step optimal variance. The estimated variances are obtained by averaging over 50 independent Monte Carlo runs. (a) $N = 150$. (b) $N = 1500$.

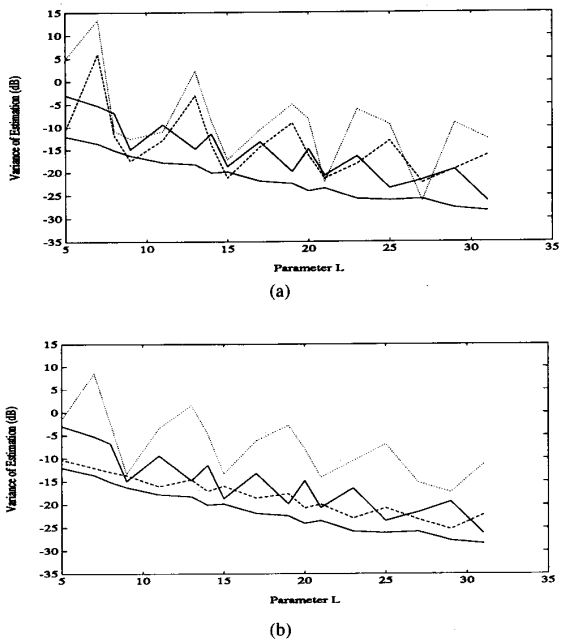


Fig. 2. Normalized variances of $\hat{\omega}_{21}$ using the sparse matrix estimator with $J = I$ ($M = 20$). Lines are as in Fig. 1. (a) $N = 150$. (b) $N = 1500$.

The empirical SSE's follow the theoretical results fairly closely when $N = 150$ and more closely for $N = 1500$ data points; because the theortical variance expressions are asymptotic results (as $N \to \infty$), closer agreement between theory and experiment is expected for larger N.

Fig. 2 shows similar SSE curves for the sparse solution estimate HOYWE-2 with $M = 20$ and $J = I$, again for $N = 150$ and $N = 1500$. First note from the solid lines in Figs. 1 and 2 that while the theoretical optimal estimator variance for $\hat{\omega}_{11}$ and $\hat{\omega}_{21}$ are similar, the theoretical initial estimator variance is larger for the HOYWE-2 estimate. The initial HOYWE-2 variances do not approach the optimal variances as L increases, so here, the use of the second step of the optimal two-step procedure appears motivated for the whole range of values of L. Moreover, the initial HOYWE-2 variances are sensitive to the particular choice of L.

The estimated SSE's in Fig. 2 closely agree with the theoretical results in some cases, but deviate significantly in other cases; in particular, the optimal HOYWE-2 estimate at $N = 1500$ agree closely with theory for all values of L, but the other estimated SSE's agree only for certain values of L. Also, the estimated SSE curves fluctuate much more as a function of L than the theoretical curves do, especially for the initial HOYWE-2 estimator.

Figs. 3 and 4 show pole plots which correspond to the $L = 9$ and $L = 13$ cases in Fig. 2. The $L = 9$ estimate corresponds to a point where experimental SSE's and the theoretical variances match closely, and the $L = 13$ estimate corresponds to a point where they do not. It can be seen that for $L = 9$ the two frequencies are estimated about equally well. The pole plots for $L = 13$ show a much larger variation in the $\hat{\omega}_{21}$ pole estimate. In fact, for $N = 150$ the two sinusoids are not resolved in the initial estimates. Even for $N = 1500$, the initial $\hat{\omega}_{21}$ poles are not tightly clustered, and the centroid of this cluster is biased in both magnitude and frequency. In all cases the optimal method significantly improves the pole estimates.

The pole plots in Figs. 3 and 4 also give an indication of occurrence of extraneous frequency estimates. From these pole plots it can be seen that for $N = 150$, there are some cases in which extraneous zeros are near the unit circle. For $L = 9$, there are several zero estimates near $\omega = 0.9\pi$ which are outside the unit circle; for $L = 13$, there are only two cases. For $N = 1500$, extraneous zeros did not appear in our examples. In order to eliminate these spurious frequency estimates, more elaborate methods than just choosing the poles nearest the unit circle must be employed; for example, one could select frequencies based on the corresponding amplitude of the estimated sinusoid.

The higher simulation errors of $\hat{\omega}_{21}$ and the large swings in the performance as a function of L are primarily caused by ill conditioning of the $Q\hat{\Gamma}$ matrix for some of the Monte Carlo experiments.[2] To reduce the ill-conditioning effect, an alternate sparse matrix estimator was tested. In this estimate (denoted $\hat{\omega}_{31}$), the n most linearly independent columns of $\hat{R}$ were used to form $\hat{\Gamma}$ at each Monte Carlo

[2]Note that the columns of R are nearly linearly dependent for closely spaced sinusoids.

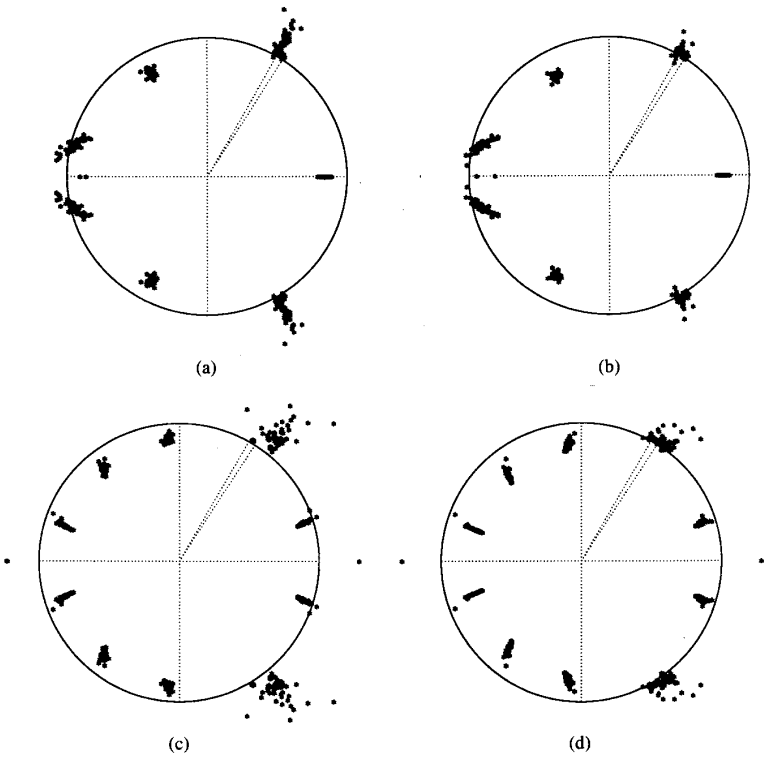


Fig. 3. Pole plots for HOYWE-2 for $N = 150$ and $M = 20$. (a) Initial estimate, $L = 9$. (b) Optimal estimate, $L = 9$. (c) Initial estimate, $L = 13$. (d) Optimal estimate, $L = 13$.

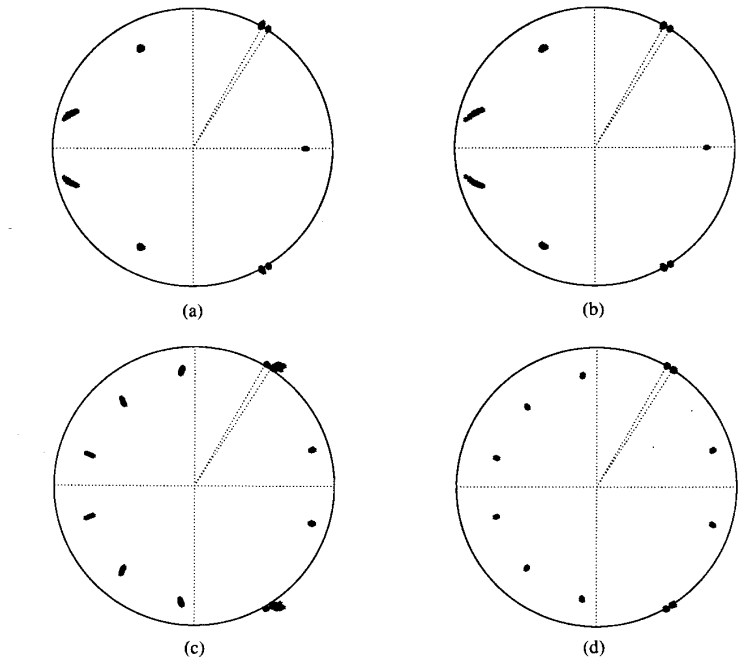


Fig. 4. Pole plots for HOYWE-2 for $N = 1500$ and $M = 20$. (a) Initial estimate, $L = 9$. (b) Optimal estimate, $L = 9$. (c) Initial estimate, $L = 13$. (d) Optimal estimate, $L = 13$.

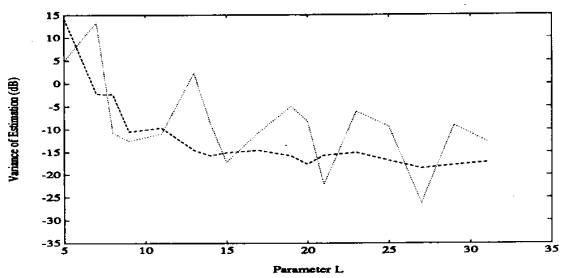


Fig. 5. Normalized variances of $\hat{\omega}_{21}$ and $\hat{\omega}_{31}$ for two sparse matrix estimators, initial solution ($M = 20$, $N = 150$). Dotted line is the $\hat{\omega}_{21}$ estimated variance. Dashed line is the $\hat{\omega}_{31}$ estimated variance.

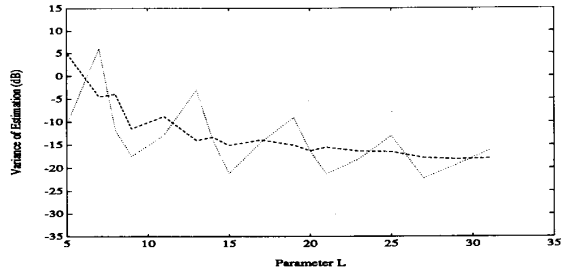


Fig. 6. Normalized variances of $\hat{\omega}_{21}$ and $\hat{\omega}_{31}$ for two sparse matrix estimators, two-step procedure ($M = 20$, $N = 150$). Lines are as in Fig. 5.

experiment. Specifically, the n columns of $\hat{R}$ are chosen recursively and each is chosen as the one which gives the minimum residual error of (3.16). Note that this method only requires n additional pivoting operations in a QR algorithm used to solve (3.16), and as such has very little computational overhead compared with the sparse solution using $J = I$ in (3.10). When N is very large $\hat{R} \approx R$, so the same n columns are chosen for each simulation; however, for small N, this is not the case. Thus, for small N, the estimated SSE for this sparse solution-based HOYWE are expected to differ significantly from the asymptotic SSE (the former correspond to a data-dependent J matrix and the latter to a fixed J matrix).

Figs. 5 and 6 compare the sparse matrix estimator $\hat{\omega}_{21}$ as in Fig. 2 with $\hat{\omega}_{31}$. Fig. 5 shows initial estimates, and Fig. 6 shows two-step optimum estimates; in all cases, $N = 150$ and $M = 20$. It can be seen that the $\hat{\omega}_{31}$ estimate exhibits much less sensitivity to the parameter L. However, the $\hat{\omega}_{31}$ SSE's are not always lower than the $\hat{\omega}_{21}$ SSE's. We note that using the most linearly independent columns of $\hat{R}$ does not necessarily yield the lowest asymptotic error estimates, so we do not necessarily expect the $\hat{\omega}_{31}$ SSE's to be lower than the $\hat{\omega}_{21}$ SSE's. However, the simulation curves suggest that choosing the most linearly independent columns of $\hat{R}$ may improve performance for small N, especially when a "good" choice for the L parameter is not known *a priori*. Since the computational overhead in choosing these n columns is small, the procedure provides a simple but effective means of reducing the SSE sensitivity to L for small data lengths.

Finally, we remark that even though the second step in the two-step procedures can be repeated, no significant improvement was seen in the estimates for any simulation performed.

VII. Conclusions

In this paper, two forms of HOYW estimators for estimating the frequencies of sinusoidal signals have been investigated. One estimator is based on a singular value decomposition approach, and the other on a sparse least squares solution to the HOYW equations. For each method, an initial solution, in which a fixed weighting matrix Q was used, and a two-step optimum solution, in which a data-dependent weight Q was employed, were considered.

The SVD-based solutions provided lower error estimates than the sparse matrix solutions in our experiments. Sparse solutions, however, require much less computation to realize the estimates. Estimates obtained with the two-step procedures were generally better than the initial estimates obtained with $Q = I$. Repeating the second step of the two-step algorithms provided no additional decrease of estimate error.

A second sparse matrix estimator, wherein the most linearly independent columns of the estimated autocorrelation matrix were chosen, provided better estimates than did the fixed-column sparse solution for short data lengths. However, for large N, this column selection technique may not be as useful. Indeed, for large values of N it may often happen that the n most linearly independent columns of $\hat{R}$ are not the best n columns to choose in order to minimize the sum-squared error of the frequency estimates. The question of how to choose these n columns to minimize the error of the estimate remains an open problem.

References

[1] 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, no. 5, pp. 689-698, Oct. 1982.

[2] J. A. Cadzow, "Spectrum estimation: An overdetermined rational model equation approach," *Proc. IEEE*, vol. 70, no. 9, pp. 907-939, Sept. 1982.

[3] D. W. Tufts and R. Kumaresan, "Estimation of frequencies of multiple sinusoids: Making linear prediction perform like maximum likelihood," *Proc. IEEE*, vol. 70, no. 9, pp. 975-989, Sept. 1982.

[4] Special Issue on Spectral Estimation, *Proc. IEEE*, vol. 72, no. 9, Sept. 1982.

[5] S. Kay and S. L. Marple, Spectrum analysis—a modern perspective," *Proc. IEEE*, vol. 69, no. 11, pp. 1380-1419, Nov. 1981.

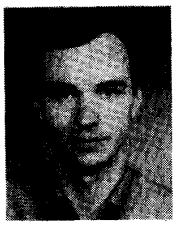
[6] S. W. Lang and J. M. McClellan, "Frequency estimation with maximum entropy spectral estimators," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, no. 6, pp. 716-724, Dec. 1980.

[7] R. Kumaresan, "Accurate frequency estimation using an all-pole filter with mostly zero coefficients," *Proc. IEEE*, vol. 70, no. 8, pp. 873-875, 1982.

[8] P. Stoica, R. L. Moses, B. Friedlander, and T. Söderström, "Maximum likelihood estimation of the parameters of multiple sinusoids from noisy measurements," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-37, no. 3, pp. 378-392, Mar. 1989.

[9] P. Stoica, T. Söderström, and F. Ti, "Asymptotic properties of the high-order Yule-Walker estimates of sinusoidal frequencies," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-37, no. 11, pp. 1721-1734, Nov. 1989.

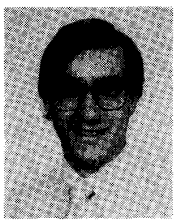
[10] P. Stoica, T. Söderström, and F. Ti, "Overdetermined Yule-Walker estimation of the frequencies of multiple sinusoids: accuracy aspects," *Signal Processing*, vol. 16, pp. 155-174, Feb. 1989.
[11] J. Li, "On the estimation of damped and undamped sinusoids with application to speech analysis," M.S. thesis, Ohio State University, Columbus, OH, 1987.
[12] R. Moses, J. Li, and P. Stoica, "Accuracy properties of high order Yule-Walker equation estimators of sinusoidal frequencies," in *Proc. 8th IFAC Symp. Identification Syst. Parameter Estimation*, Aug. 27-31, 1988.
[13] T. Söderström and P. Stoica, *Instrumental Variable Methods for System Identification*. Berlin: Springer, 1983.
[14] T. Söderström and P. Stoica, *System Identification*. London: Prentice-Hall, 1989.
[15] Y. T. Chan, Y. M. M. Lavoie, and J. B. Plant, "A parameter estimation approach to estimation of frequency of sinusoids," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, no. 2, pp. 214-219, Apr. 1981.
[16] P. Stoica, B. Friedlander, and T. Söderström, "On instrumental variable estimation of sinusoid frequencies and the parsimony principle," *IEEE Trans. Automat. Contr.*, vol. AC-31, no. 8, pp. 793-795, Aug. 1986.
[17] J. Fuchs, "Estimating the number of sinusoids in additive white noise," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-36, no. 12, pp. 1846-1854, Dec. 1988.
[18] P. Stoica, B. Friedlander, and T. Söderström, "Asymptotic bias of the high order autoregressive estimates of sinusoidal frequencies," *J. Circuits, Syst., Signal Processing*, vol. 6, pp. 287-298, 1987.
[19] P. Stoica and T. Söderström, "High-order Yule-Walker systems of equations for estimating the sinusoidal frequencies: The complete set of solutions," *Signal Processing*, July 1990.
[20] G. W. Stewart, *Introduction to Matrix Computations*. New York: Academic, 1973.
[21] P. Stoica, R. Moses, and T. Söderström, "Optimal high-order Yule-Walker estimation of sinusoidal frequencies," Tech. Rep. SAMPL-89-02, Signal Analysis Machine Perception Lab., Ohio State University, Columbus, OH, Apr. 1989.



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 coauthor of five books, the most recent being *System Identification* (Englewood Cliffs, NJ: Prentice-Hall, 1989).

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 has been given the Member of TSA&F (MTSA&F) honors award. He was corecipient with A. Nehorai of the IEEE Signal Processing Society's Senior Award for the paper, "MUSIC, Maximum Likelihood, and the Cramér-Rao Bound," published in 1989.



Randolph L. Moses (S'78-M'83-SM'90) 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 a 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, Ohio State University. His research interests are in digital signal processing, and include parametric time series analysis, radar signal processing, system identification, and model reduction.

Dr. Moses is a member of Eta Kappa Nu, Tau Beta Pi, Phi Kappa Phi, and 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, 1989), with P. Stoica.

Jian Li (S'87) was born on April 17, 1965. She received the M.Sc. degree in electrical engineering from Ohio State University, Columbus, in 1987. She is presently completing the Ph.D. degree in electrical engineering at OSU.

She is a Presidential Fellow at the Ohio State University. Her current research interests include sensor array signal processing, spectral estimation, and communications.

Ms. Li is a member of Sigma Xi and Phi Kappa Phi.