Reprinted from Journal of the Optical Society of America, Vol. 73, page 1799, December 1983

State-space and singular-value decomposition-based approximation methods for the harmonic retrieval problem

S. Y. Kung, K. S. Arun, and D. V. Bhaskar Rao

Department of Electrical Engineering—Systems, University of Southern California University Park, Los Angeles, California 90089-0272

Received April 5, 1983

We present new high-resolution methods for the problem of retrieving sinusoidal processes from noisy measurements. The approach taken is by use of the so-called principal-components method, which is a singular-value-decomposition-based approximate modeling method. The low-rank property and the algebraic structure of both the data matrix and the covariance matrix (under noise-free conditions) form the basis of exact modeling methods. In a noisy environment, however, the rank property is often perturbed, and singular-value decomposition is used to obtain a low-rank approximant in factored form. The underlying algebraic structure of these factors leads naturally to least-squares estimates of the state-space parameters of the sinusoidal process. This forms the basis of the Toeplitz approximation method, which offers a robust Pisarenko-like spectral estimate from the covariance sequence. Furthermore, the principle of Pisarenko's method is extended to harmonic retrieval directly from time-series data, which leads to a direct-data approximation method. Our simulation results indicate that favorable resolution capability (compared with existing methods) can be achieved by the above methods. The application of these principles to two-dimensional signals is also discussed.

1. INTRODUCTION

Spectral analysis forms the basis of a major part of signal processing, typically for distinguishing and tracking signals of interest and for extracting relevant information from the data. For a majority of modern signal-processing applications, such as in radar, sonar, and passive arrays, the spectral analysis problem often involves estimating the locations of spectral lines or spectral peaks, which usually represent physical quantities such as speed and bearing. In the case of two-dimensional signals, for instance, in the processing of spatiotemporal data for direction finding in passive sonar[1] or in star-bearing estimation in astronomy, the locations of two-dimensional spectral lines are of interest. In these contexts, a key measure of performance is frequency resolution, i.e., the ability to distinguish and identify spectral lines that are closely spaced in frequency.[2–4] Furthermore, in some modern signal-processing applications, the spectral estimate has to be based on short data records and yet low-bias, low-variance, high-resolution estimates are desired. In this paper we propose some singular-value-decomposition (SVD) based modeling methods for the above problem.

A. Spectral Estimation and Autoregressive Moving Average Models

The power spectrum of a discrete-time stochastic process represents the distribution of power over frequencies and is usually defined in terms of its autocovariance sequence.[5] Suppose that $y(k)$ is a zero-mean, wide-sense-stationary, discrete-time stochastic process; then its autocovariance sequence is defined as

$$r(m) = E[y(n)y(m+n)^*],$$

where E denotes the expectation operator.

The power spectrum $P(\omega)$ is related to the infinite autocovariance sequence $\{r(n)\}$ of the process by the Fourier transform[6]:

$$P(\omega) = \sum_{-\infty}^{\infty} r(m)\exp(-j\omega m), \qquad \omega \in (-\pi, \pi)$$

It can be shown to be equivalent to

$$P(\omega) = \lim_{N\to\infty} E\left[\frac{1}{N}\left|\sum_{n=0}^{N-1} X(n)e^{-jn\omega}\right|^2\right].$$

Conventional methods of spectral estimation[2,5] use one of the above two formulas for $P(\omega)$ and assume that the data outside the observation interval are zero. It is well known that this limits the frequency resolution to the reciprocal of the observation interval length. The modern approach to overcoming this fundamental limit and achieving high resolution from a short data segment is by extrapolating the data, based on certain *a priori* knowledge. In certain applications, the physical environment generating the signal can be modeled well by a linear rational system of low order. By using such *a priori* information, current methods model the process as the output of a linear rational system driven by white noise. The problem of spectral estimation is then reduced to that of estimating the model parameters. Enhanced performance can be achieved by an appropriate choice of model.

A.1 *Transfer-Function Representation*

The parameterization of a linear rational system can be done in terms of either its transfer function or its state-space parameters. Transfer-function parameterization has been the popular approach for model-based spectral estimation methods. The input–output relationship for the general autoregressive moving average (ARMA) model is given by the following difference equation:

$$y(k) = \sum_{i=1}^{p} a_i y(k-i) + \sum_{i=1}^{q} b_i v(k-i) + b_0 v(k), \qquad (1)$$

where the input is $\{v(k)\}$ and the output is $\{y(k)\}$. The transfer function of this system is

0030-3941/83/121799-13$01.00

$$H(z) = \frac{B(z)}{A(z)},$$

where

$$B(z) = b_0 + \sum_{i=1}^{q} b_i z^{-i}$$

and

$$A(z) = 1 - \sum_{i=1}^{p} a_i z^{-i}$$

The roots of $A(z)$ are the poles of the system, whereas the roots of $B(z)$ determine the zeros. For the spectral estimation problem, the input $v(k)$ to the model is a white-noise process of variance ρ. Then the covariance of the output $y(k)$ satisfies the recurrence relations

$$r(m) = \sum_{i=1}^{p} a_i r(m-i) \qquad \text{for all } m > q, \tag{2}$$

and the power spectrum of the output is simply $\rho|H(e^{j\omega})|^2$, which can be computed in terms of the transfer-function parameters $\{a_i, b_j\}$, which in turn can be estimated from the given information. In general, model-based spectral estimation consists of model identification (parameter estimation) from the given information, followed by spectrum computation from the model parameters. Since the ARMA power-spectrum estimate corresponds to an infinite covariance sequence, the ARMA modeling approach (in effect) extends the covariance sequence beyond the finite observation interval [cf. Eq. (2)] and results in higher resolution.

An interesting and popular special case of the ARMA model [Eq. (1)] is the autoregressive (AR) model, where $b_i = 0$ for all $i \neq 0$, so that

$$y(k) = \sum_{i=1}^{p} a_i y(k-i) + b_0 v(k),$$

$$r(m) - \sum_{i=1}^{p} a_i r(m-i) = \begin{cases} b_0^2, & m = 0 \\ 0, & m > 0 \end{cases},$$

and

$$P(\omega) = \rho b_0^2 / |A(e^{j\omega})|^2 \tag{3}$$

The covariance recurrence relations for $m = 0, 1, \ldots, p$ form a Toeplitz system, usually called the Yule–Walker equations or the normal equations. Given exact covariances $r(0), r(1), \ldots, r(p)$, the AR parameters may be obtained by solving this Toeplitz system. The AR model has been used often in spectral estimation, mainly because

(1) There exists a computationally efficient algorithm, the Levinson algorithm, to solve the Toeplitz system of normal equations for the AR parameters.[7]

(2) The covariance extension provided by the AR model maximizes the entropy of the process among all possible semipositive extensions.[8] Hence the AR modeling approach is also known as the maximum entropy method (MEM).

For an ARMA model, the recurrence relations are satisfied only for $m > q$. Hence the denominator parameters $\{a_i\}$ are obtained from exact covariances[9] by solving Eq. (2) for $m = q+1, q+2, \ldots, q+p$. This $p \times p$ system of equations is often called the higher-order Yule–Walker equations.[3]

A.2 State-Space Representation

An alternative representation of a linear rational system that has been popular in systems theory and control literature is the state-space representation, in which the input–output description of the ARMA model, instead of Eq. (1), is given by the state-space equations:

$$x(k+1) = Fx(k) + Tv(k),$$
$$y(k) = hx(k) + v(k) \tag{4}$$

Here $x(k)$, the state vector, is a $p \times 1$ vector process and F, T, and h are constant matrices of sizes $p \times p$, $p \times 1$, and $1 \times p$, respectively. The transfer function of this system is $H(z) = h(zI - F)^{-1}T + 1$. The ARMA model for a discrete-time process $y(k)$ in state-space notation is related to the transfer-function representation in that the poles of the model [roots of $A(z)$] are the eigenvalues of F, whereas the zeros [roots of $B(z)$] are the eigenvalues of $(F - Th)$. For a given transfer function, the triplet (F, T, h) of a minimal realization is unique modulo a similarity (coordinate) transformation. An interesting choice of coordinates leads to a canonical form realization, with

$$F = \begin{bmatrix} a_1 & 1 & 0 & \cdots & 0 \\ a_2 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ a_p & 0 & 0 & \cdots & 0 \end{bmatrix}. \tag{5}$$

This is a form that directly relates the state-space model to the transfer-function parameters.

B. State-Space Identification

B.1 Deterministic Case

It can be shown that the relationship between the impulse response of the model and the state-space parameters is

$$i(k) = hF^{k-1}T, \qquad k > 0 \tag{6}$$

This indicates that the infinite Hankel matrix formed from the impulse response sequence can be factorized as

$$\mathcal{H} = \begin{bmatrix} i(1) & i(2) & i(3) & \cdots \\ i(2) & i(3) & i(4) & \cdots \\ i(3) & i(4) & i(5) & \cdots \\ \vdots & \vdots & \vdots & \end{bmatrix}$$
$$= \begin{bmatrix} h \\ hF \\ hF^2 \\ \vdots \end{bmatrix} [T, FT, F^2T, \cdots] = \mathcal{O} \cdot \mathcal{C} \tag{7}$$

The matrices $\mathcal{O}$ and $\mathcal{C}$ are known as the observability and controllability matrices in linear systems theory.[10] Observe that the observability matrix $\mathcal{O}$ (or the controllability matrix $\mathcal{C}$) has only p columns (or rows) and that consequently $\mathcal{H}$ has finite rank ($\leq p$). The controllability matrix $\mathcal{C}$ is an ∞-to-p mapping that maps the infinite-dimensional past input

$$V^- = [v(-1), v(-2), v(-3), \cdots]'$$

into the current $p \times 1$ state vector $x(0)$. The p-dimensional state x summarizes all the relevant information in the past input history that is needed for the future outputs. As a matter of fact, *if there is no future input*, the future output is simply

$$Y^+ = [y(0), y(1), y(2), \cdots]' = \mathcal{O} \cdot x(0)$$

Hence the observability matrix $\mathcal{O}$ maps the state vector into the future output. The Hankel matrix $\mathcal{H} = \mathcal{O} \times \mathcal{C}$ is thus an operator from the past input to the future output, and it necessarily has to be of rank p because of the component operators $\mathcal{O}$ and $\mathcal{C}$. This rank property can be traced back to Kronecker,[11] who noted that an impulse-response sequence admits of a finite-dimensional realization of order p, if and only if the infinite Hankel matrix formed from the sequence has rank equal to p. Also note that $\mathcal{O}$ satisfies $\mathcal{O}F = \mathcal{O}\uparrow$. Thus the state transition matrix F can be computed from $\mathcal{O}$ as

$$F = \mathcal{O}^{\dagger}\mathcal{O}\uparrow, \tag{8}$$

where

$$\mathcal{O}\uparrow = \begin{bmatrix} hF \\ hF^2 \\ hF^3 \\ \vdots \end{bmatrix}, \qquad \mathcal{O}^{\dagger} = (\mathcal{O}'\mathcal{O})^{-1} \cdot \mathcal{O}'.$$

The aim of deterministic identification is to obtain the state-space model from the impulse-response sequence. Classical realization methods, such as the Ho–Kalman algorithm,[12,13] factorize the Hankel matrix as in Eq. (7) and use Eq. (8) to arrive at the state-space parameters. It might be noted that in the Ho–Kalman algorithm only a $(p+1) \times (p+1)$ Hankel matrix, constructed from the first $(2p+1)$ exact impulse-response coefficients, is sufficient to obtain the $\mathcal{O}$ factor and the state-space parameters.

B.2 *Stochastic Case*

The covariance of the output of an ARMA model driven by white noise of variance ρ is given by

$$r(m) = \begin{cases} hPh' + \rho, & m = 0 \\ hF^{m-1}g, & m > 0 \end{cases}, \tag{9}$$

where $g = FPh' + \rho T$ and where P is the $p \times p$ state covariance matrix that satisfies the Lyapunov equation, $P = FPF' + \rho TT'$.

In general, the infinite Toeplitz matrix

$$R = \begin{bmatrix} r(0) & r(-1) & r(-2) & \cdots \\ r(1) & r(0) & r(-1) & \cdots \\ r(2) & r(1) & r(0) & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix} \tag{10}$$

does not have finite rank. However, it can be shown that, for the special case of sinusoidal processes, R turns out to have finite rank. This fact is exploited in a new algorithm developed in Section 2.

At this point it is worth noting that the Hankel matrix formed from the covariance sequence is factorizable as

$$H = \begin{bmatrix} r(1) & r(2) & r(3) & \cdots \\ r(2) & r(3) & r(4) & \cdots \\ r(3) & r(4) & r(5) & \cdots \\ \vdots & \vdots & \vdots & \end{bmatrix}$$

$$= \begin{bmatrix} h \\ hF \\ hF^2 \\ \vdots \end{bmatrix} [g, Fg, F^2g, \quad] = \mathcal{O} \cdot \mathcal{G} \tag{11}$$

The aim of stochastic identification is to obtain the state-space model from the covariance sequence. Deterministic identification schemes can be extended to the stochastic case in a straightforward fashion, and the above factorization of the covariance Hankel H, along with Eq. (8), can be used to estimate the state-space parameters from the exact covariances.[14] Note that, just as in the Ho–Kalman algorithm, only the first $(2p+1)$ covariance lags are needed to obtain the state-space parameters. In fact, solving the so-called higher-order Yule–Walker Eqs. (2) [for the denominator parameters $\{a_p{}^i\}$ of the transfer-function model Eq. (1)] is equivalent to finding the null vector of the same $(p+1) \times (p+1)$ principal submatrix of the covariance Hankel matrix H.

C. Identification By Use of Approximate Modeling

In practice, the given information is often inexact, with perturbations that may be attributed to the following possible sources:

(1) The given measurements [of the time series $y(k)$, the impulse response $i(k)$, or the covariance sequence $r(m)$] are usually corrupted by noise, white or colored.

(2) There will be inevitable covariance estimation errors because of finite data record length, word length, or bad choice of estimators.

In such a situation, the Hankel matrices formed from either the noisy impulse-response sequence or the perturbed covariance sequence will not have low rank. We then have to resort to approximate modeling of the data (measurements) to smooth out the perturbations. Fortunately, we normally have more than $(2p+1)$ covariance lags or impulse-response measurements. Hence, instead of exact modeling based on only $(2p+1)$ measurements, a model that approximately fits all the available information is more desirable.

The objective of the approximation procedure should be to determine best the exact underlying model for the (pure) signal part based on the given information (e.g., the time-series measurements y). In a typical (conceptual) formulation,

(1) A vector y in a normed (measurement) vector space is given.

(2) The aim of the approximation is to find a point m in the model space $\mathcal{M}$ so that the model output $\hat{y}(m)$ is as close to y as possible.

(3) The performance is measured in terms of the norm of the approximation error, $y - \hat{y}(m)$.

The actual choice of the model space $\mathcal{M}$ (which is determined by the model type and the upper bound on the model order) depends on an often complicated trade-off between model order and error norm.

C.1 *Principal-Components Approach*

In this paper we propose approximate modeling algorithms using SVD of a nearly low-rank matrix to determine its rank and to obtain a low-rank approximant from the principal components. In the presence of perturbations, normally low-rank matrices, such as the Hankel matrices $\mathcal{H}$ and H, tend to have full rank. However, the singular values that ideally should have been zero will be much smaller than the other (principal) singular values. So a possible scheme for model-

order estimation involves looking for a break in the singular values.

At this point, it is not clear how one can construct an optimal approximant to the given perturbed information in any particular norm. Instead, we adopt a two-step (approximate) modeling procedure, based on the principal components obtained from the SVD. The procedure is best illustrated by its application to the deterministic identification problem of state-space model estimation from noisy measurements of the impulse response.

Given measurements of the impulse response $i(k)$, then the Hankel operator constructed from $i(k)$ is a mapping from the past input vector V^- to the future output Y^+, i.e.,

$$Y^+ = \mathcal{H} \cdot V^-.$$

Assume that we are seeking a pth-order approximation; the first step of the principal-components (PC) procedure provides a mechanism for obtaining an $\infty \to p$ controllability-type map: $x = \mathcal{C}V^-$, from the (past) input space to the state space, and a $p \to \infty$ observability-type map: $Y^+ = \mathcal{O}x$, from the state space to the (future) output space. This basically ensures that the state space is p dimensional. The aim of the first approximation step is to find $p \times \infty$ and $\infty \times p$ maps $\mathcal{C}$ and $\mathcal{O}$, respectively, such that

$$\overline{Y}^+(\mathcal{O}, \mathcal{C}) = \theta \mathcal{C} V^-$$

is closest to Y^+ in the *minimax sense*, i.e., we wish to minimize

$$\sup_{\|V^-\|_2 = 1} \|Y^+ - \overline{Y}^+(\mathcal{O}, \mathcal{C})\|_2$$

This is equivalent to finding the best $p \times \infty$ and $\infty \times$ p maps $\mathcal{C}$ and $\mathcal{O}$, respectively, that make $\bar{\mathcal{H}} = \mathcal{O} \cdot \mathcal{C}$ closest to the given Hankel matrix $\mathcal{H}$ in the spectral norm sense. (The spectral norm of A is defined as $\sup\|Ax\|_2$ subject to $\|x\|_2 = 1$.) This can be achieved by a SVD of $\mathcal{H}$, as described in the following procedure.

Procedure PC 1

Perform an SVD of $\mathcal{H}$ and arrange the singular values $\{\sigma_i^2\}$ of $\mathcal{H}$ in decreasing order

$$= (U_1\ U_2) \begin{pmatrix} \Sigma_1^2 & 0 \\ 0 & \Sigma_2^2 \end{pmatrix} \begin{pmatrix} V_1 \\ V_2 \end{pmatrix},$$

where the $p \times p$ matrix Σ_1^2 contains the dominant singular values and Σ_2^2 contains the smaller singular values. It is well known that a p-rank matrix $\bar{\mathcal{H}}$ that best approximates $\mathcal{H}$ in the spectral norm sense is obtained by retaining only the principal components:

$$\bar{\mathcal{H}} = U_1\ \Sigma_1^2\ V_1$$

The (optimal) approximation error in the spectral norm $\|\mathcal{H} - \bar{\mathcal{H}}\|_s$ is given by σ_{p+1}. Therefore, if $\sigma_p \gg \sigma_{p+1}$, then $\bar{\mathcal{H}}$ is a good approximant to $\mathcal{H}$.

Moreover, $\bar{\mathcal{H}}$ is obtained in factored form:

$$\bar{\mathcal{H}} = U_1 \cdot \Sigma_1^2 \cdot V_1 = \mathcal{O} \cdot \mathcal{C}$$

so that $\mathcal{O}$ and $\mathcal{C}$ can be immediately identified. A realization in balanced coordinates[17] is obtained by choosing the controllability-type and observability-type maps as

$$\mathcal{O} = U_1 \cdot \Sigma_1,$$

$$\mathcal{C} = \Sigma_1 \cdot V_1$$

The second step involves determining the model parameters. Ideally, $\mathcal{O}$ would have the exact observability matrix structure, and a $p \times p$ solution F to the matrix equation

$$\theta F = \theta\uparrow$$

would exist. However, because of the approximation, no exact solution exists and we have to resort to a least-squares solution that minimimizes $\|\mathcal{O}F - \mathcal{O}\uparrow\|_E$, where subscript E denotes the Euclidean norm. [The Euclidean norm of A is defined as the square root of trace (A′A).] This leads to least-squares estimates of the state-space parameters, as described in the following procedure.

Procedure PC 2

The state-space parameters can be derived from the matrices $\mathcal{O}$ and $\mathcal{C}$, as follows:

h is the first row of $\mathcal{O}$,

T is the first column of $\mathcal{C}$,

$F = \mathcal{O} \dagger \mathcal{O}\uparrow$.

The optimal error $\|\mathcal{O}F - \mathcal{O}\uparrow\|_E$ can be shown to be $O(\sigma_{p+1})$.[16] Hence state-space model identification seems to be a natural end product of the SVD approximation approach.

Numerical Properties of the Principal Component Approximation

The PC method (consisting of procedures PC1 and PC2) enjoys many desirable numerical properties. First, SVD has been widely recognized as a numerically reliable tool for displaying closeness to low rank. Second, the entire process leading to the state-space solution requires no matrix inversion, as the pseudoinverse of $\mathcal{O}$ is in fact the transpose of $\mathcal{O}$. Moreover, the error bound of $O(\sigma_{p+1})$ ensures a good, although suboptimal, approximation. Also, the PC method yields a balanced realization,[17] which is less sensitive to finite-word-length effects.[18] Morever, from Eq. (8), we can show that $\|F\|_s \leq 1$, which implies that the eigenvalues of F are bounded by unity and ensures stability of the model. Consequently, PC methods have consistently performed well.

C.2 Applicability of the Principal Components Method

The PC method is applicable to any problem in which either the θ or the $\mathcal{C}$ map can be extracted (by use of SVD) from the given information. To ensure a p-dimensional state vector, only the p principal components from the SVD are retained to construct $\mathcal{O}$ and $\mathcal{C}$. Then the next step extracts (by use of least-squares approximation) the state-space parameters from the matrix $\mathcal{O}$ or $\mathcal{C}$.

Apart from the deterministic identification problem, the PC method has found applications in stochastic identification and harmonic retrieval. In the stochastic identification

problem, the given information could be in the form of a noise corrupted data sequence (i.e., time series y) or noisy measurements of its covariance sequence $r(m)$, from which the state-space model is to be estimated.[19–21] Again, the p-dimensional map θ can be extracted from y. One possibility is to use the covariance Hankel matrix H:

$$H = E(Y^+Y^-).$$

Since the input is white noise, Y^- is uncorrelated with the future inputs and is correlated only with that part of Y^+ that depends on the past input V^-. Therefore

$$H = \theta E(xY^-).$$

This indicates why H has rank p and how θ can be obtained from the principal components of H. Since the covariance Hankel matrix admits of a factorization [cf. Eq. (11)] similar to the factorization of the impulse-response Hankel matrix [cf. Eq. (7)], a PC algorithm was recently proposed[13] for the approximate covariance modeling problem. There are several schemes that use the PC approach on other related matrices. For examples, see Refs. 20 and 21.

In the harmonic retrieval problem, which is the problem addressed in this paper, it will be shown that $\mathcal{O}$ can be extracted from the data $y(k)$ or the covariance $r(m)$ in a similar fashion. It is shown in Section 2 that for sinusoidal data

$$R = \mathcal{O}E(xx')\mathcal{O}',$$

and that $\mathcal{O}$ can be obtained by a PC approximation of R. This is done by the use of a special sinusoidal model developed for the high-resolution spectral line estimation problem. The model is used to establish the finite rank and factorization property [similar to Eqs. (7) and (11)] of the Hankel data and Toeplitz covariance matrices. This allows the PC approach to be applied and leads to the so-called Toeplitz approximation method (TAM). The TAM provides an improvement over exact matching schemes, such as Pisarenko's method, in estimating the sinusoid parameters. In Section 3 it is shown how the PC approach can be applied directly on the given data without forming a covariance estimate. In Section 4, the PC method is applied to the two-dimensional harmonic retrieval problem.

2. HARMONIC RETRIEVAL PROBLEM

The problem of retrieving sinusoids (with frequencies close to one another) from perturbed covariance information is of special interest in a vast range of signal-processing applications. Often the covariance sequence may have to be estimated from time-series data, as in Doppler processing in radar. However, it is not uncommon to encounter applications in which the (time-series) data are not measurable, whereas the covariance information is directly available. Such situations arise in astronomical star bearing estimation, interferometry, and passive sonar applications.

When the covariance information is exact, Fourier-transform methods, such as the Blackman–Tukey estimator,[5] or AR modeling methods, such as the MEM,[8] may be used. Although the MEM provides better resolution than conventional Fourier-transform methods, both methods perform poorly when the covariance information is inexact, as is often the case in practice.[2] By incorporating the extra information that the signal is sinusoidal into the model, the special structure inherent in the harmonic retrieval problem can be exploited to get better resolution. Such a method was first proposed by Pisarenko[22] in 1973, and since then many variants have surfaced.

A. Special Model for Harmonic Processes

The key to achieving high resolution is to use a model for harmonic processes that incorporates all the *a priori* information regarding the special structure of sinusoidal signals. For line enhancement and for tracking of sinusoidal signals with slowly varying frequencies, it has been seen that a special ARMA model can lead to high performance.[23] The ARMA model of the previous section is, however, too general for our purpose. Hence we develop a special model for harmonic processes and establish the low-rank property of the Toeplitz covariance matrix and the data Hankel matrix in a straightforward fashion.

A signal composed of, say, $p/2$ sinusoids can be mathematically represented by

$$y(k) = \sum_{i=1}^{p} c_i \exp[j(\omega_i k + \phi_i)],$$

where c_i, ω_i, and ϕ_i are the amplitudes, frequencies, and phases of the ith complex exponential. The phases are assumed to be independent random variables distributed between 0 and 2π. Such a signal can be considered to be the output of a special ARMA model with poles on the unit circle and input noise power $\rho = 0$. (Poles on the unit circle make the system self-generating.) Hence the difference equation representation for sinusoidal signals is a special case of Eqs. (1) and (3) with $b_i = 0$ for all i:

$$y(k) = \sum_{i=1}^{p} a_i y(n-i), \tag{12}$$

where the roots of the polynomial $A(z)$ are on the unit circle at $e^{j\omega_1}, e^{j\omega_2}, \ldots, e^{j\omega_p}$. This indicates that the Hankel matrix

$$Y = \begin{bmatrix} y(1) & y(2) & y(3) & \cdots \\ y(2) & y(3) & y(4) & \cdots \\ y(3) & y(4) & y(5) & \cdots \\ \vdots & \vdots & \vdots & \end{bmatrix},$$

built from the sinusoidal data has rank p because the vector $\mathbf{a} = (1, -a_1, -a_2, \ldots, -a_p)'$ is a null vector of *any* $(p+1)$ consecutive columns of Y. The above model indicates that $y(k)$ is *exactly* linearly predictable, and so linear prediction methods such as Prony's method[24] and MEM have been used to estimate the $\{a_i\}$ parameters and the sinusoid frequencies.

It can be shown that the covariance also satisfies a similar recurrence relation, i.e.,

$$r(m) = \sum_{i=1}^{p} a_i r(m-i). \tag{13}$$

This indicates that the Toeplitz covariance matrix R [cf. Eq. (10)] also has rank p. Popular harmonic retrieval methods, such as Pisarenko's method and the Tufts–Kumaresan method, use this property to estimate $\mathbf{a}$ and the sinusoid parameters.

B. Pisarenko's Method

Given exact sinusoidal data, Prony's method and related linear prediction methods obtain $\mathbf{a}$ as the null vector of Y and obtain the sinusoidal frequencies from the roots of $A(z)$. In the presence of additive noise, a least-squares (AR) linear prediction fit leads to the solution of the normal Eqs. (2) and the MEM spectral estimate. Although MEM is capable of high resolution, it requires a substantially larger model order to account for the additive noise in low signal-to-noise ratio (SNR) situations. Moreover, MEM is not optimum for the spectral line estimation problem, since it does not exploit the sinusoidal structure of the signal. When it is known *a priori* that the additive noise in the time-series data is white, the sinusoidal structure of the signal (i.e., finite rank properties of Y and R) can be exploited to reduce the effect of the noise.

In the presence of noise, the observed covariance matrix $\tilde{R}$ is the sum of the signal covariance matrix R and the noise covariance matrix $\sigma^2 I$, i.e., $\tilde{R} = R + \sigma^2 I$. (It is assumed here that the noise is white with variance σ^2.) It is desired to reduce or to remove the noise contribution before carrying out the estimation. For this purpose we need the following observations:

(1) For any $N > p$, the $N \times N$ leading submatrix R_N of the infinite Toeplitz covariance matrix R has rank p; therefore σ^2 has to be an eigenvalue of the matrix $\tilde{R}_N$ with multiplicity $(N - p)$.

(2) Moreover, σ^2 can only be the minimum eigenvalue, so that $(\tilde{R}_N - \sigma^2 I)$ $(= R_N)$ remains semipositive definite.

(3) Finally, the vector $\mathbf{a}$ [for the model, Eq. (12)] is the null vector of R_{p+1}, and consequently $\mathbf{a}$ is the eigenvector associated with the minimum eigenvalue σ^2 of the $(p + 1) \times (p + 1)$ leading principal minor of $\tilde{R}_{p+1}$.

When these results are combined they lead to Pisarenko's spectral estimation method, which is summarized below.

(1) Diagonalize a sufficiently large matrix $\tilde{R}$ and determine the number of sinusoids by examining the multiplicity of the minimum eigenvalue. (From now on we drop the subscript N for the covariance matrix R and its estimates, but it should be understood that the matrix size is $N \times N$, i.e., R stands for R_N.)

(2) Compute the eigenvector $\mathbf{a}$ of the $(p + 1) \times (p + 1)$ leading principal minor of $\tilde{R}$ corresponding to the minimum eigenvalue.

(3) Solve for the roots of $a(z) = 0$ to obtain the location of the spectral lines.

(4) The amplitude (power) of each of the sinusoidal components can be derived by solving a linear system of equations.[22]

The above method is directly applicable when the covariance information is available. When the information is in the form of time-series data, then the covariance has to be estimated first. There are a number of ways to estimate the covariance, and the choice of the covariance estimator can be critical. An improper estimator may adversely affect the underlying algebraic properties, such as positivity and low rank. In Section 2.C we discuss some covariance estimation schemes particularly suited for our problem.

C. Covariance Estimation from Time-Series Data

Numerous covariance matrix estimators exist in the literature, and for the sinusoid retrieval problem the following three unbiased estimators have often been used.

Type 1. Given L consecutive values of a time series, one popular choice is the unbiased Toeplitz estimator:

$$\hat{R} = [\hat{r}(i - j)], \qquad i = 1, \ldots, N; \qquad j = 1, \ldots, N, \tag{14a}$$

where

$$\hat{r}(m) = \frac{1}{L - |m|} \sum_{n=1}^{L - |m|} y(n) \cdot y(n + m)^*, \qquad m = 0, 1, \ldots, N - 1. \tag{14b}$$

If the time series is composed of $p/2$ sinusoids, the covariance matrix R should be a rank p Toeplitz matrix. However, the estimate $\hat{R}$, although it is Toeplitz, will not have rank p even when there is no noise in the time series.

Type 2. In sharp contrast, the second estimate

$$\bar{R} = (Y_N \cdot Y_N')/(L - N + 1) \tag{15}$$

[where Y_N is the leading $N \times (L - N + 1)$ submatrix of the Hankel data matrix Y] is not Toeplitz. But it has rank p when there is no additive noise because of the finite rank of Y. It is immediately obvious that when there is no additive noise in $y(k)$ one can obtain $\mathbf{a}$ exactly as the null vector of any $(p + 1)$ consecutive rows of $\bar{R}$, and the sinusoidal parameters can be exactly found. However, in the presence of additive noise in $y(k)$, $\bar{R}$ may become an inferior covariance estimate (especially when L is small) compared to $\hat{R}$, because $\hat{R}$ will smooth out the noise better than $\bar{R}$. Note that the entire data record of length L is used in estimating the covariance lags of $\hat{R}$, whereas only $L - N + 1$ terms are used in the estimates of $\bar{R}$.

Type 3. To improve the estimate $\hat{R}$, Ulrych and Clayton[25] suggested a new estimator that utilizes all the available data more effectively:

$$\check{R} = (DD')/2(L - N + 1), \tag{16a}$$

where

$$D = \begin{bmatrix} y(1) & y(2) & \cdots & y(L - N + 1) & y(L) & y(L - 1) & \cdots & y(N) \\ y(2) & y(3) & \cdots & y(L - N + 2) & y(L - 1) & y(L - 2) & \cdots & y(N - 1) \\ y(3) & y(4) & \cdots & y(L - N + 3) & y(L - 2) & y(L - 3) & \cdots & y(N - 2) \\ \vdots & & & & & & & \\ y(N) & y(N + 1) & \cdots & y(L) & y(L - N + 1) & y(L - N) & \cdots & y(1) \end{bmatrix}. \tag{16b}$$

It turns out that, for noise-free sinusoidal data, D also has rank p for the following reasons:

(1) The second half of D is the Hankel matrix (say, Y^R)

constructed from the time-reversed series $y(L), y(L-1), \ldots, y(1)$

(2) Since the poles of $A(z)$ are on the unit circle, $\mathbf{a}$ is symmetric, and therefore $\mathbf{a}$ is a null vector of any $(p+1)$ consecutive rows of Y^R as well

(3) Hence the vector $\mathbf{a}$ is also the null vector of any $(p+1)$ consecutive rows of D

This indicates that, under noise-free conditions, the sinusoidal parameters can be exactly identified from $\check{R}$.

Comparison among the Estimators: We first note that $\overline{R}$, $\hat{R}$, and $\check{R}$ are all unbiased estimates of R, since

$$E(\overline{R}) = E(\hat{R}) = E(\check{R}) = R$$

But when the estimates are based on short data records, other factors, such as estimation error variance, finite rank, and structure, have to be taken into consideration. In the presence of additive noise, $\check{R}$ has better noise-smoothing capability than $\hat{R}$ since it uses more data. However, the choice between $\overline{R}$ and $\check{R}$ depends on the SNR and the harmonic retrieval method that is adopted. For instance, if the data record is short, $\overline{R}$ will be Toeplitz but not low rank (when noise free), whereas $\check{R}$ will be ideally low rank but not Toeplitz. Simulations (cf. Section 2 F) indicate that $\overline{R}$ has better noise-rejection properties than $\check{R}$ and that R performs better at a low SNR. But for the harmonic retrieval problem, the low-rank property of the estimate under noise-free conditions is rather crucial. Consequently, $\check{R}$ seems to perform better than $\overline{R}$ at a high SNR.

In linear prediction (AR modeling) methods, for example, all the aforementioned estimators have been used with different levels of success. However, as mentioned before, MEM is not optimum for the spectral line estimation problem, since it does not exploit the sinusoidal structure. On the other hand, Pisarenko's method uses this structure but is sensitive to perturbations in the covariance matrix and performs rather poorly when any of the three covariance estimates is used.[25] Hence more robust methods are called for. An effective improvement was made by Tufts and Kumaresan,[4] who developed approximate linear prediction methods that also exploit the sinusoidal nature of the signal. Their method involves a SVD-based low-rank approximation of the covariance matrix, followed by estimation of the linear prediction vector $\mathbf{a}$ from the approximant. (The covariance estimate used is $\check{R}$.) From a state-space perspective, linear prediction methods such as the Tufts–Kumaresan method can be viewed as procedures for computing the first column of a state transition matrix F for the model, assuming $\theta = R$. Under *exact* conditions, because of the Toeplitz structure of R, this estimate

$$F = R^{\downarrow} R^{\uparrow}$$

will be in the canonical form of Eq. (5), and the first column will be the prediction polynomial. However, after the approximation of R by its principal components, the Toeplitz structure is not preserved, and the estimate of F will no longer be in canonical form. Then a method that computes the entire matrix F may be more desirable. Moreover, as was indicated earlier (cf. Section I.C.1), the state-space-based PC approach has many desirable numerical properties, and so we now seek a state-space formulation of the problem. The state-space formulation will show that Y and R are not only low rank but also highly structured and factorizable and that an observability matrix can be extracted, which means that the PC method can be applied.

D. State-Space Formulation

The state-space representation of the special model [Eq. (12)] for sinusoidal signals is a special case of Eq. (3) (with $T = 0$ and no input):

$$\begin{aligned} x(k+1) &= Fx(k) \\ y(k) &= hx(k), \end{aligned} \tag{17}$$

where the eigenvalues of F are of unit magnitude and equal the roots of $A(z)$. Since the triplet $(F, x(0), h)$ is unique only up to a (coordinate) similarity transformation, one interesting realization for the model is when

$$F = \mathrm{diag}(e^{j\omega_1}, e^{j\omega_2}, \ldots, e^{j\omega_p})$$

and

$$h^{(i)} \times (0)^{(i)} = c_i e^{j\phi_i} \tag{18}$$

Here the ith element of a vector is denoted by a superscript (i). Note that this is a (diagonal) canonical form that directly relates the state-space parameters to the sinusoid frequencies. Therefore, when the state transition matrix F is diagonalized, its diagonal elements (which equal the eigenvalues of F) will give the frequencies of the sinusoids [cf. Eq. (18)]. Moreover, the transformed vectors h and x_0 will give us their amplitudes and phases.

From the sinusoidal state-space equations [Eqs. (17)], we can show that

$$y(k) = hF^k x(0), \qquad k \geq 0$$

and

$$r(m) = hPF'^m h' = hF^m P h', \qquad m \geq 0,$$

where the state variance P satisfies $P = FPF'$. When these equations are compared with Eqs. (6) and (9) for the general ARMA model, we see that the output process behaves like the impulse response. Hence the Hankel matrix formed from the sinusoidal data is itself factorizable:

$$Y = \begin{bmatrix} h \\ hF \\ hF^2 \\ \vdots \end{bmatrix} [x(1), Fx(1), F^2x(1), \ldots] = \theta\, \mathcal{R}. \tag{19}$$

Moreover, the Toeplitz covariance matrix R is also factorizable, as shown below:

$$\begin{aligned} R &= \begin{bmatrix} h \\ hF \\ hF^2 \\ \vdots \end{bmatrix} [Ph', F^{-1}Ph', F^{-2}Ph', \ldots] \\ &= \theta P \theta' \qquad (\text{using } F^{-1}P = PF') \end{aligned} \tag{20}$$

Equations (19) and (20) also indicate that Y and R have rank p. The above factorizations are not valid for the general ARMA model and result from incorporating the *a priori* information about sinusoidal signals in the special model. [For the diagonal realization (9b), the factorizations of Y and R

take on a special form: [see Appendix A.] Based on the factorization of R, we present an approximate realization method for our special model that works on the covariance matrix or its estimate (either $\bar{R}$ or $\hat{R}$). It turns out that, for the diagonal form of Eq. (18), when the elements are real the double Hankel matrix D is also factorizable as

$$D = \begin{bmatrix} h \\ hF \\ hF^2 \\ \vdots \\ hF^{N-1} \end{bmatrix} [x(1), Fx(1), F^2x(1), \ldots, F^{L-N}x(1)] \; x^*(L), Fx^*(L), F^2x^*(L), \ldots, F^{L-N}x^*(L)] \quad (21)$$

By using similarity transformations it can be shown that D is factorizable for any realization, with the left-hand factor having the observability structure. The proof uses the fact that the poles of the model are on the unit circle and is not valid for the double Hankel matrix built from the impulse response of a general ARMA model. The factorization of D proves to be useful for the spectral line estimation problem, since $\hat{R}$ will be factorizable (under noise-free conditions) even for short data records, unlike $\bar{R}$.

E. Toeplitz Approximation Method

For the purpose of robust harmonic retrieval from covariance information, we make note of the following two properties of the sinusoidal covariance matrix R.

(1) It has finite rank under no-noise conditions.
(2) It admits of a factorization [as indicated by Eq. (20)] from which an observability-type matrix can be obtained.

Hence we can conclude that our problem is amenable to the PC approach. On applying the PC approach on $\hat{R}$, we have

$$\hat{R} = U \Sigma^2 V = [U_1 \; U_2] \begin{bmatrix} \Sigma_1^2 & 0 \\ 0 & \Sigma_2^2 \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix},$$

where Σ_1 is $p \times p$ and Σ_2 is $(N-p) \times (N-p)$. The observability matrix is obtained from the principal singular vectors U_1 and the principal singular values Σ_1^2. In the presence of white noise, although the singular vectors are unchanged, the singular values are affected. In fact, all the singular values are increased by the noise variance, and so the smallest singular value has to be substracted to compensate for this effect, i.e.,

$$\Sigma_1^2 = \Sigma_1^2 - \sigma_N^2 I,$$

where σ_N^2 is the smallest singular value of R. Hence the observability-type matrix is given by

$$\mathcal{O} = U_1 \Sigma_1,$$

where

$$\mathcal{C} = \Sigma_1 V_1.$$

Then the state-space parameters can be estimated as follows:

$$E = \theta^{\dagger} \theta^{\dagger},$$

h is the first row of $\mathcal{O}$,

Ph' is the first column of $\mathcal{C}$.

The eigenvalues of F give the frequencies of the sinusoids, i.e.,

$$\tilde{F} = \text{diag}\,[e^{j\omega_i}] = Q F Q^{-1},$$

Here (F, Ph', h) is similar to

$$(Q^{-1}FQ, Q^{-1}Ph', hQ) = (\tilde{F}, \tilde{P}\tilde{h}', \tilde{h}) \quad (22)$$

Since $\tilde{F}$ is diagonal, $\tilde{P}$ must also be diagonal and contain part of the amplitude information. So

$$c_i = (\tilde{P}\tilde{h}')^{(i)} = \tilde{h}^{(i)} \quad (23)$$

Note that the computation of F reduces, by using the matrix-inversion lemma and the orthonormality of the singularity vectors, to a simple matrix multiplication.[16] Also, the discussion on state-space parameter estimation using the PC approach indicates that the TAM will exhibit good numerical properties. This claim is further supported by simulation results discussed below.

F. Simulations and Discussions

This section provides some simulation results that should help to illustrate the theoretical discussion so far. The problem considered is the retrieval of a single sinusoid (effectively, the problem of resolving two closely spaced complex exponentials) in additive white noise from 25 data samples. More precisely, assuming that the sampling frequency is 2 Hz, the sinusoid frequency is chosen close to 1 Hz, say, 0.98 Hz [cf. Fig. 1(a)], which gives rise to the problem of resolving complex exponentials of 0.98 and 1.02 Hz, which are 0.04 Hz apart. (The simulations were performed in double precision on a 36-bit PDP-10 computer.) First the popular MEM method is applied. Five representative simulation results are shown in Fig. 2, which shows a high percentage of failure to resolve the lines. Often a higher-order MEM will improve the resolution, but this is generally accompanied by the problem of spurious peaks. This problem is significantly alleviated in the Tufts–Kumaresan[4] method (TKM) and in the TAM as shown in our simulation study.

In this study, the TKM with polynomial order 13 and the TAM applied on 11×11 $\bar{R}$ and 13×13 $\hat{R}$ were each tested on the harmonic process corrupted by 200 different (white) pseudonoise sequences. The results in terms of the mean, the standard deviation, the root-mean-square (rms) error of the frequency estimates, and the failure rate (failure to resolve the spectral lines) are shown in Table 1. (The mean and the standard deviation of the estimates are computed only for the resolved cases, whereas the rms error takes the unresolved samples into consideration also.) A trial is considered a failure if the method identifies the frequencies to be (1.0 Hz, 1.0 Hz), i.e., both z-plane roots are on the negative real axis [cf. Fig. 1(b)]. Other simulation parameters are the rank of

the approximant (model order) and the SNR. Although the rank can be determined by examining the singular values, for convenience the rank of the approximant is predetermined to be 2 in our simulation study. The SNR is defined as the ratio of the power in each exponential to the variance of the noise.

The results indicate that the TAM on $\bar{R}$ [cf. Eq (14)] is suitable for low-SNR situations, whereas the TAM on the covariance estimate $\hat{R}$ [cf. Eq (16)] performs well for high-resolution problems. As an example, we note that, for the low-SNR (0-dB) case, the TAM on $\bar{R}$ performs better than the TKM, with 7 versus 48 failures in 200 trials. For the high-SNR (20-dB) case, the TAM on $\hat{R}$ compares favorably with the TKM, with 5 versus 43 failures in the 200 trials.

Cramer–Rao Bound: A comparison of the variance can give an idea of the relative performance. However, for an absolute performance evaluation it is useful to compare with the Cramer–Rao lower bounds. For example, for the simulation performed on data consisting of a sinusoid of frequency 0.97 Hz in white noise with 0-dB SNR, the Cramer–Rao bound (on the standard deviation of the frequency estimate) can be computed to be 0.014092. The results indicate that the standard deviations of 0.0201 and 0.0236 in the TAM estimates are fairly close to the lower bound. Note that the deviations are also comparable with the standard deviation of 0.0226 in the Tufts–Kumaresan estimate.

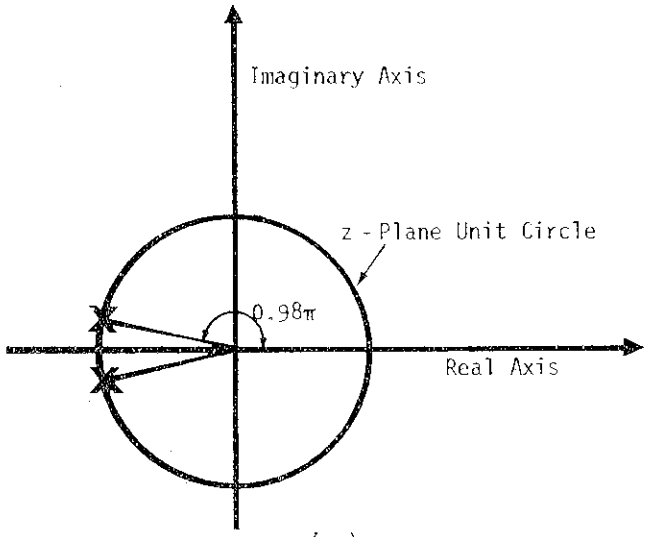

Fig. 1. (a) Actual pole positions (indicated by crosses) for 0.98-Hz sinusoid sampled at 2.0 Hz. (b) Pole positions found in a failed trial.

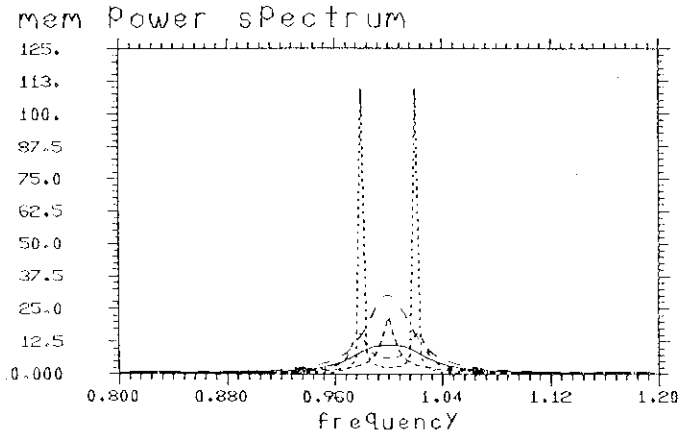

Fig. 2. Simulation results for MEM (five trials).

Table 1. Simulation Results

Method	Frequencies To Be Resolved (0.97, 1.031) (SNR: 0 dB)	(0.98, 1.02) (SNR: 10 dB	(0.99, 1.01) SNR: 20 dB
TAM (on $\bar{R}$)			
Miss ratio	7/200	2/200	100/200
Rms error	0.020154	0.003738	0.008007
Mean	0.952043	0.979278	0.995135
Stand. dev.	0.008113	0.003090	0.001359
TAM (on $\hat{R}$)			
Miss ratio	72/200	9/200	5/200
Rms error	0.023608	0.007115	0.003163
Mean	0.959822	0.980397	0.990353
Staud. dev.	0.016156	0.005831	0.002752
TKM			
Miss ratio	48/200	15/200	43/200
Rms error	0.022631	0.007489	0.005431
Mean	0.956483	0.979600	0.991312
Stand. dev.	0.014386	0.005295	0.002910

Numerical Aspects. The prediction vector computed by the TKM method has dimension (equal to 11) much higher than the true order, and the method calls for the rooting of an oversized polynomial, whereas the 2×2 matrix F obtained in the TAM is of a much smaller size. Computationally, the eigendecomposition of a 2×2 matrix is simpler and less sensitive to errors than the rooting of a thirteenth-order polynomial. As a general remark, since the TAM is based on the PC method, all its numerical-stability properties are inherited.

3. HARMONIC RETRIEVAL DIRECTLY FROM DATA

We saw in the last section that an improper covariance estimator may adversely disturb the original underlying algebraic properties, such as positivity and low rank. So methods that work directly on data, avoiding the covariance-estimation step

altogether, become attractive. For instance, in maximum entropy spectral estimation, Burg proposed a direct data method[27] that has gained considerable popularity.

For direct-data estimation of the sinusoid parameters, an examination of the Hankel data matrix is useful. As was indicated earlier, the infinite Hankel matrix formed from the time series ideally has low rank equal to twice the number of sinusoids. To exploit the low-rank property, when the data consist of sinusoids corrupted by additive white noise, we need to be able to remove the contribution of noise (just as in Pisarenko's method for the Toeplitz covariance). At first sight, the problem appears to be more difficult, because here all the elements of the Hankel matrix, not just the diagonal elements as in the covariance case, are corrupted by noise. But the following claim establishes a curious fact that the addition of white noise on the data will not change the singular vectors of the infinite Hankel matrix.

Claim. If Y is the Hankel matrix built from the uncorrupted data and $\tilde{Y}$ is the Hankel matrix corresponding to data plus white noise, then asymptotically Y and $\tilde{Y}$ have the same singular vectors. Similarly if D is constructed [according to Eq. (16b)] from the uncorrupted data and, correspondingly, $\tilde{D}$ constructed from the noise corrupted data, then asymptotically (as the record length $L \to \infty$), D and $\tilde{D}$ have the same singular vectors.

Briefly, this claim can be verified by noting that, for ergodic processes, the left singular vectors of Y (and asymptotically those of D) are the eigenvectors of R, whereas the left singular vectors of $\tilde{Y}$ and $\tilde{D}$ (asymptotically) are the eigenvectors of $\tilde{R} = R + \sigma^2 I$, which are the same as the eigenvectors of R.

This claim leads to Pisarenko's spectral estimation directly from time-series data. For a sufficiently large double Hankel matrix $\tilde{D}$ formed from the corrupt data, we expect (ideally) the smallest singular value to have multiplicity $(N - p)$. Then the prediction vector $\mathbf{a}$ can be obtained as the left singular vector corresponding to the smallest singular value of a new matrix D_{p+1}, which is defined as in Eq. (16b) with N replaced by $p + 1$. The direct-data Pisarenko algorithm is theoretically equivalent to Pisarenko's method applied on $D_{p+1}D_{p+1}'$ and, again, it is expected that the method will be sensitive (cf. Section 2.D).

Since the matrices Y and D are themselves factorizable and have finite rank [cf. Eqs. (19) and (21)] by using the above claim, it seems feasible to apply the PC approach on the data matrices directly to obtain robust estimates. A robust improvement of the direct-data Pisarenko scheme is the direct-data version of the TKM, using SVD-based low-rank approximation of D, followed by computation of the linear prediction vector $\mathbf{a}$ from this approximant. But, as before, we adopt a state-space approach in the following section.

Direct-Data Approximation

An obvious possibility is to treat the time series as the noisy impulse response of our special model and to use an approximate deterministic realization algorithm, the PC method (cf. Section I.C.1), on the Hankel data matrix Y to compute (F, x_0, h). But the performance of the method will be limited by the quality of $\hat{R}$ as a covariance estimator. Since $\hat{R}$ is a better covariance estimate and D also enjoys the factorization property [Eq. (21)], the PC approach can be used on D as well. Application of the PC method on D will result in a direct-data version of the Toeplitz approximation of DD'. Here, SVD is performed on D instead of on R, resulting in

$$D = U\Sigma V'$$

The effect of noise is suppressed by substracting the smallest singular value squared:

$$\Sigma_1 = (\Sigma^2 - \sigma_N^2 I)^{1/2},$$

where σ_N is the smallest singular value. Then the observability matrix $\mathcal{O}$ is formed from the principal singular vectors and singular values

$$\mathcal{O} = U_1 \Sigma_1^{1/2}$$

Similarly,

$$\mathcal{D} = \Sigma_1^{1/2} V_1'$$

The state-space parameters are computed as

$$F = \mathcal{O}^{\dagger}\mathcal{O}\uparrow,$$

$x(1)$ is the first column of $\mathcal{D}$,

h is the first row of $\mathcal{O}$.

From the state-space parameters, one can retrieve the sinusoidal information (frequencies and amplitudes) after diagonalizing F, just as in the TAM [cf. Eqs. (22) and (23)]. In addition, we can also obtain the phase information from the transformed coordinates by using Eq. (18).

Recall that the left singular vectors of D are the same as the eigenvectors of DD'. Hence the approximation of D is theoretically equivalent to the Toeplitz approximation of DD'. However, working on D avoids the numerical problems associated with the increased condition number of DD' and, moreover, there exists a stable and efficient SVD algorithm that can work on D directly.[28] Therefore a direct-data approximation is preferable for numerical reasons.

4. TWO-DIMENSIONAL HARMONIC RETRIEVAL

Two-dimensional (2-D) spectral line estimation has a number of important applications. For example, in the star-location problem in astronomical signal processing, the stars are the point sources in space whose bearings are to be estimated. The light from the stars is focused by a space telescope lens and recorded by 2-D optical sensors placed at its focal plane. Therefore we have a 2-D spatial spectral line retrieval problem from finite 2-D data or covariance information.

In this section we develop a special model for 2-D sinusoids, establish the low-rank property of both the data and covariance matrices, and exploit this fact to derive 2-D PC methods. The notion of matrix factorization will be again applied; however, the matrices used are not direct generalizations of the corresponding one-dimensional (1-D) matrices. A signal that is the sum of 2-D sinusoidal signals,

$$y(n, m) = \sum_{i=1}^{p} c_i \exp[j(\omega_{1i} n + \omega_{2i} m + \phi_i)],$$

can be represented by a 2-D state-space model. Here we make a simplifying but generic assumption that the

frequencies are distinct in both directions, i.e.,

$$\omega_{1i} \neq \omega_{1k}, \quad \omega_{2i} \neq \omega_{2k}, \quad \forall i, k, i \neq k$$

A special separable model[29] can be used for this signal as shown below:

$$\begin{aligned} x(n+1, m) &= F_1 x(n, m), \\ x(n, m+1) &= F_2 x(n, m), \\ y(n, m) &= h x(n, m), \quad n = 1, \quad , N \qquad m = 1, \quad , N \end{aligned} \tag{24}$$

Here $x(\ ,\ )$ represents the state vector. The eigenvalues of F_1 give the frequencies in one direction, whereas the eigenvalues of F_2 give the frequencies in the other direction. The model can be better understood by examining the following special choice of $F_1, F_2, h, x(1, 1)$:

$$\begin{aligned} F_1 &= \text{diag}[e^{j\omega_{1i}}] \\ F_2 &= \text{diag}[e^{j\omega_{2i}}] \\ h &= [c_1, \quad , c_p], \\ x(1, 1) &= [e^{j\phi_1}, \quad , e^{j\phi_p}]^T \end{aligned} \tag{25}$$

All other realizations can be obtained through similarity transformations of the special case. Just like the Hankel data matrix in the 1-D case, the data matrix Y, where $y_{ij} = y(i, j)$, can be factored as

$$\begin{aligned} Y &= \begin{bmatrix} y(1,1), \quad , y(1,N) \\ \vdots \qquad\qquad \vdots \\ y(N,1), \quad , y(N,N) \end{bmatrix} \\ &= \begin{bmatrix} h \\ hF_1 \\ \vdots \\ hF_1^{N-1} \end{bmatrix} [x(1,1), F_2 x(1,1), \quad , F_2^{N-1} x(1,1)] \end{aligned} \tag{26}$$

In the presence of white-noise perturbations on the data, a result similar to the claim in the 1-D case is possible. Also, the covariance matrix shown below is factorizable.

$$R = \begin{bmatrix} r(0,1) & r(0,2) & \quad & r(0,N) \\ r(1,1) & r(1,2) & \quad & r(1,N) \\ r(2,1) & r(2,2) & \quad & r(2,N) \\ \vdots & \vdots & & \vdots \\ r(N,1) & r(N,2) & \quad & r(N,N) \end{bmatrix},$$

where

$$\begin{aligned} r(i, j) &= E\,[y(n+i, m+j) y^*(n, m)] \\ & \quad i = 0, 1, \quad N, \qquad j = 1, 2, \quad N \\ &= hF_1^i F_2^j P h' \end{aligned}$$

and P is the state variance.

Note that R can be factored as

$$\begin{aligned} R &= \begin{bmatrix} h \\ hF_1 \\ \vdots \\ hF_1^N \end{bmatrix} [F_2 P h', F_2^2 P h', \quad F_2^N P h'] \\ &= \mathcal{O}_1 \cdot \mathcal{O}_2' \end{aligned} \tag{27}$$

Table 2. Comparison of Variance

		First Frequency		Second Frequency	
		X	Y	X	Y
Actual values		0.25	0.30	0.35	0.45
Data based method:	mean	0.24972	0.30115	0.35075	0.45005
standard deviation		0.00189	0.00253	0.00370	0.00362
Covariance method:	mean	0.24983	0.30051	0.35195	0.45010
standard deviation		0.00188	0.00188	0.00265	0.00177

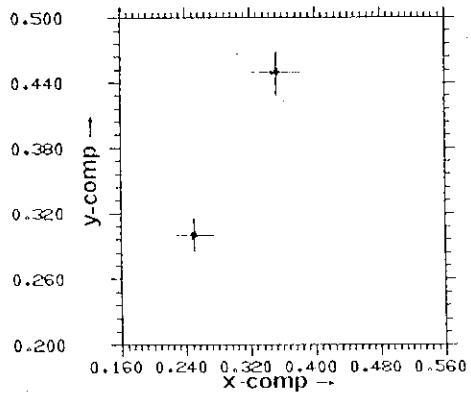

Fig. 3. Simulation results for two-dimensional covariance method (10 estimates of point-source locations).

This indicates that Y and R have finite rank and the PC method can be applied. More precisely, the factorization [cf. Eq. (25)] and the structure, i.e., $\mathcal{O}_1 F_1 = \mathcal{O}_1\uparrow$ and $\mathcal{O}_2' F_2' = \mathcal{O}_2'\uparrow$, can be exploited to obtain F_1 and F_2 as in Eq. (8). Again, the frequencies can be computed by diagonalizing F_1 and F_2. The pairing of the frequency components from the two axes deserves more caution, but it can be accomplished by reducing it to the diagonal form given by Eq. (25) and examining the amplitudes (see Appendix A).

A similar procedure can be adopted for the direct-data case. More precisely, the PC approach can be applied on the data matrix Y, and the factorization [Eq. (26)] can be used to obtain the state-space parameters. In fact, the data approach is equivalent to the 1-D Toeplitz approximation of two type-2 covariance estimates $(YY')/N$ and $(Y'Y)/N$, one for each direction.

To demonstrate the performance, we present some simulation examples below. The data used for the 2-D simulations consist of two real sinusoids of amplitude 2.0 each, in additive white noise of variance 1.0. The data matrix used is of size 25×25, and 10 independent trials were conducted. The covariance estimate used for the covariance method is a type-1 estimate. The simulation results are summarized in Table 2.

The data-based method utilizes the data less effectively than the type-1 covariance estimates. (The use of type-3 covariance estimates for the 2-D problem is more difficult than in the 1-D problem.) However, the data approach may be useful for high-resolution, high-SNR problems (cf. Section 2.D). For low-SNR situations, the type-1 covariance estimate appears to be more suitable. The results of the covariance method are displayed in Fig. 3.

5. CONCLUSIONS

In this paper, we have indicated how state-space realization theory can be used for harmonic retrieval. We have shown

that there is a rich algebraic structure embedded in the harmonic retrieval problem that can be exploited by using the (state-space) PC method. Moreover, SVD possesses desirable numerical properties that are inherited by the PC method. The PC approach, when applied to the covariance matrix, leads to the derivation of an approximate modeling method: the TAM, which offers a robust Pisarenko spectral estimate from estimated or observed covariance. The PC approach is also directly applicable to the data matrix, leading to a direct-data approximation approach that has better numerical properties.

As a historical note, the potential of Kronecker's theorem (circa 1881) and realization theory[1,20,21] has been largely ignored by the signal-processing community, which has instead leaned heavily toward Prony's linear prediction approach. The PC approach proposed in this paper extends realization theory to a general approximate modeling strategy that is applicable to a large class of identification problems. For example, it is a strong candidate for the problem of retrieving exponentially damped sinusoids from noisy data. (A damped sinusoid can be treated as the impulse response of a system with poles inside the unit circle.) In conclusion, the algebraic and numerical significance of using SVD and the state-space approach for approximation appears to be promising, and it will be an important area for future in-depth exploration.

APPENDIX A

For the special diagonal form given by Eq. (18), the factorization of Y [cf. Eq. (19)] reduces to

$$Y=\begin{bmatrix}1 & 1 & 1\\ e^{j\omega_1} & e^{j\omega_2} & e^{j\omega_p}\\ \vdots & \vdots & \vdots\\ e^{jn\omega_1} & e^{jn\omega_2} & e^{jn\omega_p}\end{bmatrix}\begin{bmatrix}c_1e^{j\phi_1} & & 0\\ & c_2e^{j\phi_2} & \\ & & \\ 0 & & c_pe^{j\phi_p}\end{bmatrix}\begin{bmatrix}1 & e^{j\omega_1} & e^{jn\omega_1}\\ 1 & e^{j\omega_2} & e^{jn\omega_2}\\ \vdots & \vdots & \vdots\\ 1 & e^{j\omega_p} & e^{jn\omega_p}\end{bmatrix},$$

and that of R [cf. Eq. (20)] reduces to

$$R=\begin{bmatrix}1 & 1 & 1\\ e^{j\omega_1} & e^{j\omega_2} & e^{j\omega_p}\\ \vdots & \vdots & \vdots\\ e^{jn\omega_1} & e^{jn\omega_2} & e^{jn\omega_p}\end{bmatrix}\begin{bmatrix}c_1{}^2 & & 0\\ & c_2{}^2 & \\ & & \\ 0 & & c_p{}^2\end{bmatrix}\begin{bmatrix}1 & e^{-j\omega_1} & & e^{-jn\omega_1}\\ 1 & e^{-j\omega_2} & \cdots & e^{-jn\omega_2}\\ \vdots & \vdots & & \vdots\\ 1 & e^{-j\omega_p} & & e^{-jn\omega_p}\end{bmatrix}.$$

Similarly, for the 2-D signal case, the special diagonal representation given by Eq. (25) reduces to the factorization of R [cf. Eq. (27)] to

$$R=\begin{bmatrix}1 & 1 & 1\\ e^{j\omega_{11}} & e^{j\omega_{12}} & e^{j\omega_{1p}}\\ \vdots & \vdots & \vdots\\ e^{jn\omega_{11}} & e^{jn\omega_{12}} & e^{jn\omega_{1p}}\end{bmatrix}\begin{bmatrix}c_1{}^2 & & 0\\ & & \\ & & \\ 0 & & c_p{}^2\end{bmatrix}\begin{bmatrix}e^{j\omega_{21}} & & e^{jn\omega_{21}}\\ e^{j\omega_{22}} & \cdots & e^{jn\omega_{22}}\\ \vdots & & \vdots\\ e^{j\omega_{2p}} & & e^{jn\omega_{2p}}\end{bmatrix}.$$

This is useful in determining the amplitudes and the pairing of the frequencies.

ACKNOWLEDGMENTS

This research was supported in part by the U.S. Office of Naval Research under contract N00014-81-K-0191, by the National Science Foundation under grant ENG-7908673, and by the U.S. Army Research Office under grant DAAG29-79-C-0054.

REFERENCES

1. O. S. Halpeny and D. G. Childers, "Composite wavefront decomposition via multidimensional digital filtering of array data," IEEE Trans. Circuits Syst. **CAS-22,** 552–562 (1975).
2. S. M. Kay and S. L. Marple, Jr., "Spectrum analysis—a modern perspective," Proc. IEEE **69,** 1380–1418 (1981).
3. J. A. Cadzow, "Spectral estimation: an overdetermined rational model equation approach," Proc. IEEE **70,** 907–939 (1982).
4. D. W. Tufts and R. Kumaresen, "Estimation of frequencies of multiple sinusoids: making linear prediction perform like maximum likelihood," Proc. IEEE **70,** 975–989 (1982).
5. G. M. Jenkins and D. G. Watts, *Spectral Analysis and its Applications* (Holden-Day, San Francisco, Calif., 1966).
6. A. Papoulis, *Probability, Random Variables, and Stochastic Processes* (McGraw-Hill, New York, 1965).
7. N. Levinson, "The Wiener rms (root mean square) error criterion in filter design and prediction," J. Math. Phys. **25,** 261–278 (1947).
8. J. P. Burg, "Maximum entropy spectral analysis," Ph.D. Thesis (Stanford University, Stanford, Calif., 1975).
9. W. Gersh, "Estimation of the autoregressive parameters of mixed autoregressive moving-average time series," IEEE Trans. Autom. Control **AC-15,** 583–588 (1970).
10. T. Kailath, *Linear Systems* (Prentice-Hall, Englewood Cliffs, N.J., 1980).
11. L. Kronecker, "Zur Theorie der Elimination einer Variabeln aus zwei Algebraischen Gleichungen," Trans. Royal Prussian Academy of Science (see collected works, Vol. 2), 1881.
12. B. L. Ho and R. E. Kalman, "Effective construction of linear state variable models from input/output data," in *Proceedings of the 3rd Allerton Conference on Circuits and System Theory* (U. Illinois Press, Urbana, Ill., 1965), pp. 449–459.
13. J. Rissanen, "Recursive identification of linear systems," J. SIAM Control **9,** 420–430 (1971).
14. P. Faurre, "Stochastic realization algorithms" in *System Identification: Advances and Case Studies*, R. K. Mehra and D. G. Lainiotis, eds. (Academic, New York, 1976).
15. V. C. Klemma and A. J. Laub, "The singular value decomposition: its computation and some applications," IEEE Trans. Autom. Control **AC-25,** 164–176 (1980).
16. S. Y. Kung, "A new identification and model reduction algorithm via singular value decomposition", in *Proceedings of the 12th Asilomar Conference on Circuits, Systems and Computers*, (Institute of Electrical and Electronics Engineers, New York, 1978), pp. 705–714.
17. B. C. Moore, "Principal component analysis in linear systems:

controllability, observability, and model reduction," IEEE Trans. Autom. Control **AC-26,** 17–32 (1981).
18. C. T. Mullis and R. A. Roberts, "Synthesis of minimum round-off noise fixed point digital filters," IEEE Trans. Circuits Syst. **CAS-23,** 551–562 (1976).
19. S. Y. Kung and K. S. Arun, "A novel Hankel approximation method for ARMA pole-zero estimation from noisy covariance data," in *Digest of the Topical Meeting on Signal Recovery and Synthesis with Incomplete Information and Partial Constraints* (Optical Society of America, Washington, D.C., 1983), pp. WA19-1–WA19-5.
20. U. B. Desai and D. Pal, "A realization approach to stochastic model reduction and balanced stochastic realizations," in *Proceedings of the Twenty-First IEEE Conference on Decision and Control* (Institute of Electronics and Electrical Engineers, New York, 1982), pp. 1105–1112.
21. S. Y. Kung and K. S. Arun, "Approximate realization methods for ARMA spectral estimation," in *Proceedings of the IEEE International Symposium on Circuits and Systems* (Institute of Electrical and Electronics Engineers, New York, 1983).
22. V. F. Pisarenko, "The retrieval of harmonics from a covariance function," Geophys. J. R. Astron. Soc. Can. **33,** 347–366 (1973).
23. D. V. Bhaskar Rao, "Adaptive notch filtering for the retrieval of harmonics," Ph.D. Thesis (University of Southern California, Los Angeles, Calif., 1983).
24. G. R. B. Prony, "Essai experimental et analytique, etc.," J. Ec. Polytech. **1,** 24–76 (1795).
25. T. J. Ulyrch and R. W. Clayton, "Time series modeling and maximum entropy," Phys. Earth Planet. Inter. **12,** 188–200 (1976).
26. S. Y. Kung, "A Toeplitz approximation method and some applications," in *Proceedings of the International Symposium on Mathematical Theory of Networks and Systems* (Western, North Hollywood, Calif., 1981), pp. 262–266.
27. J. P. Burg, "A new analysis technique for time series data," presented at NATO Advanced Study Institute on Signal Processing with Emphasis on Underwater Acoustics, August 12–23, 1968.
28. G. H. Golub and C. Reinsch, "Singular value decomposition and least squares solutions," Numer. Math. **14,** 403–420 (1970).
29. S. Attasi, "Modelling and recursive estimation for double indexed sequences," in *System Identification: Advances and Case Studies*, R. K. Mehra and D. G. Lainiotis, eds. (Academic, New York, 1976).