Journal of Sound and Vibration (2001) **242**(1), 85–102
doi:10.1006/jsvi.2000.3361, available online at http://www.idealibrary.com on IDEAL®

A NEW PRONY-BASED CIRCULAR-HYPERBOLIC DECOMPOSITION

S. O'F. Fahey, B. M. Suleiman, A. H. Nayfeh and M. R. Hajj

Department of Engineering Science and Mechanics, MC 0219, *Virginia Polytechnic Institute and State University, Blacksburg, VA* 24061

(*Received* 13 *December* 1999, *and in final form* 13 *September* 2000)

We introduce a new closed-form decomposition technique for estimating the model parameters of an evenly sampled signal known to be composed of circular and hyperbolic sine and cosine functions in the presence of Gaussian white noise. The techniqe is closely related to Prony's method and hereditary algorithms that fit complex exponential functions to evenly sampled data. The circular and hyperbolic sine and cosine functions are obtained by adding constraints that limit the form of the characteristic polynomial coefficients. It avoids the leakage effects associated with the discrete fourier transform (DFT) for circular sine and cosine functions. When the signal contains frequency components that are not rational multiples of each other, the proposed decomposition yields amplitude and phase parameters that are more accurate than those obtained with the DFT in moderate levels of noise. First, we review Prony's method and one hereditary algorithm (the complex exponential algorithm). Then, we detail three implementation procedures of the new technique. The first is a two-stage least-squares approach. The second utilises a novel concept of noise reduction which is attributed to Pisarenko. The last provides additional means of noise reduction through a covariance formulation that avoids zero-lag terms. Experimental and numerical examples of the application of the circular-hyperbolic decomposition (CHD) are given.

1. INTRODUCTION

In many applications, there is a need to fit a discrete time-limited signal to one or more complex exponentials, such as the Markov parameters, the impulse-response function, and the ring down of a linear structural dynamic system. The Prony family of algorithms has been shown to be very useful in achieving such an objective [2]. A real signal composed of complex exponentials can be expressed as

$$v(t_\ell) = \sum_{i=1}^{m} A_i \exp(\lambda_i t_\ell) \tag{1}$$

and in the presence of Gaussian white noise as

$$y(t_\ell) = v(t_\ell) + \varepsilon(t_\ell), \tag{2}$$

where m, λ_i, A_i, $v(t_\ell)$, $y(t_\ell)$, $\varepsilon(t_\ell)$, and t_ℓ are the number of complex exponentials, poles (frequencies and decay rates), residues (amplitudes and phases), signal, signal with noise, noise, and evenly sampled control variable (often taken as time or distance along a path) with the associated index ℓ respectively.

0022-460X/01/160085 + 18 $35.00/0

Since $v(t_\ell)$ is real, the complex pole λ_i must have a complex conjugate $\lambda_k = \lambda_i^*$ and its corresponding complex conjugate residue $A_k = A_i^*$. On the other hand, the poles of hyperbolic functions are real and occur in reciprocal pairs (e.g., $\lambda_2 = 1/\lambda_1$). In practice, we often make the restriction that m is an even number. Yet, there can be unpaired real poles in which Case m could be odd. Here, we restrict m to be an even number $2n$ for notational purposes. We note that, although not all data can be expressed in terms of either equation (1) or (2), these equations encompass a very large class of signals [1–6]. Details of Prony's method and the complex exponential algorithm for uncovering the parameters in equation (1) are reviewed in sections 2.1 and 2.2.

Besides the need to fit data with complex exponentials, there is a need to fit data with circular and hyperbolic sine and cosine functions, which are referred to as *circular–hyperbolic* functions. Signals that can be described in terms of these functions include stationary motions of rotating machines; deformation of cables, beams, and shells; trajectories of celestial bodies. Prony's hereditary algorithms can be applied to these signals because circular–hyperbolic functions represent a special case of complex exponentials. However, unless the signal is noiseless, one must contend with non-zero decay rates for circular sine and cosine functions and unpaired exponentials for hyperbolic sine and cosine functions, which can cause a dilemma in the interpretation of parameter estimates. Thus, we consider the special case of circular–hyperbolic functions, expressed as

$$\begin{aligned} v(t_\ell) = & \sum_{i=1}^{n_\omega} p_i \cos(\omega_i t_\ell) + q_i \sin(\omega_i t_\ell) \\ & + \sum_{i=1}^{n_\kappa} a_i \cosh(\kappa_i t_\ell) + b_i \sinh(\kappa_i t_\ell) + (-1)^{\ell-1} \sum_{i=1}^{n_\eta} c_i \cosh(\eta_i t_\ell) + d_i \sinh(\eta_i t_\ell) \\ & + \sum_{i=1}^{n_\lambda} \{e_i \cos[\mathrm{Im}(\lambda_i) t_\ell] \cosh[\mathrm{Re}(\lambda_i) t_\ell] + f_i \sin[\mathrm{Im}(\lambda_i) t_\ell] \cosh[\mathrm{Re}(\lambda_i) t_\ell] \\ & + g_i \cos[\mathrm{Im}(\lambda_i) t_\ell] \sinh[\mathrm{Re}(\lambda_i) t_\ell] + h_i \sin[\mathrm{Im}(\lambda_i) t_\ell] \sinh[\mathrm{Re}(\lambda_i) t_\ell]\}, \end{aligned} \tag{3}$$

$$y(t_\ell) = v(t_\ell) + \varepsilon(t_\ell), \tag{4}$$

where n_ω, n_κ, n_η, and n_λ are the number of frequencies ω_i, κ_i, η_i, and λ_i, respectively, and p_i, q_i, a_i, b_i, c_i, d_i, e_i, f_i, g_i, and h_i are the amplitudes. The parameters ω_i, κ_i, p_i, q_i, a_i, b_i, c_i, d_i, e_i, f_i, g_i, and h_i are real-valued for real signals. The λ_i are complex-valued. The transition from equation (1) to (3) provides $n(= n_\omega + n_\kappa + n_\eta + 2n_\lambda)$ constraints that can be integrated into the estimation procedure. On the right-hand side of equation (3), the terms 1 and 2 represent circular functions, the terms 3 and 4 represent hyperbolic functions, the terms 5 and 6 represent hyperbolic functions multiplied by a circular component at the Nyquist frequency, and the terms 7–10 represent the product of hyperbolic functions (3) are given in sections 3.1–3.3. Experimental and numerical examples of the proposed decomposition are given in sections 4.1 and 4.2, respectively. Conclusions are presented in section 5.

2. BACKGROUND

In sections 2.1 and 2.2, we review Prony's method and the complex exponential algorithm. Prony's method is a traditional technique that transforms $4n$ data points, known to be composed of exactly $2n$ complex exponentials, into the pole and residue parameters of

equation (1). The complex exponential algorithm is a generalization of Prony's method to data, which is overdetermined data and noisy.

2.1. PRONY'S METHOD

Prony's method transactions $N = 4n$ evenly sampled discrete data points, at this juncture taken to contain no noise, into the pole and residue parameters of $2n$ distinct exponentials as described in equation (1) [2, 4]. When a signal contains more than $2n$ complex exponentials, the data are said to be *underdetermined*, and Prony's method will often provide an exact map between the data and results. However, the results from an underdetermined data set may not match the data exactly in some instances. For example, $v(1:4) = [1, -0{\cdot}9, 0{\cdot}8, -0{\cdot}7]^{\mathrm{T}}$ does not conform with equation (1) for any choice of λ_i and A_i when $m \leqslant 2$. Unfortunately, there is no reliable physical interpretation or good resolution for such results. On the other hand, most signal evaluations involve quantities of data that far exceed the number of parameters desired. For these evaluations, the data are referred to as *overdetermined*, which will be readdressed through the complex exponential algorithm in section 2.2. For now, we assume that $4n$ data points are composed of exactly $2n$ distinct complex exponentials. In the absence of noise, the following expressions hold:

$$v(\ell) = \sum_{i=1}^{2n} A_i \exp(\lambda_i[\ell - 1]\Delta t) = \sum_{i=1}^{2n} A_i[z_i]^{\ell-1} \tag{5}$$

$$= [A_1, A_2, \ldots, A_{2n}] \begin{bmatrix} z_1 & & & 0 \\ & z_2 & & \\ & & \ddots & \\ 0 & & & z_{2n} \end{bmatrix}^{\ell-1} \begin{Bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{Bmatrix},$$

where $t_\ell = [\ell - 1]\Delta t$ is time in s. The parameters z_i and λ_i represent the ith pole in the z domain and frequency domain respectively. These parameters are related by

$$\lambda_i = \frac{1}{\Delta t} \ln(z_i), \tag{6}$$

where λ_i has units of rad/s. The problem reduces to finding z_i and A_i such that equation (5) holds for all $\ell \in \{1, 2, \ldots, 4n\}$.

The z-domain characteristic polynomial is given by

$$\prod_{i=1}^{2n} [z - z_i] = \sum_{k=0}^{2n} \alpha_k [z]^k, \tag{7}$$

where $\alpha_{2n} = 1$ and its roots z_i are the poles, which need to be uncovered. We note that the characteristic polynomial has the special property

$$\sum_{k=0}^{2n} \alpha_k [z]^k = 0 \tag{8}$$

for all $z \in \{z_1, z_2, \ldots, z_{2n}\}$.

Next, we introduce the Hankel matrix $V(i, k) \equiv v(i + k - 1)$, which can be expressed in verbose form as

$$V(1:2n, 1:2n+1) = \begin{bmatrix} v(1) & v(2) & \cdots & v(2n+1) \\ v(2) & v(3) & \cdots & v(2n+2) \\ \vdots & \vdots & \ddots & \vdots \\ v(2n) & v(2n+1) & \cdots & v(4n) \end{bmatrix}. \tag{9}$$

As a consequence of equation (5), every row in the Hankel matrix can be written as

$$V(\ell, 1:2n+1) = [A_1, A_2, \ldots, A_{2n}] \begin{bmatrix} z_1 & & & 0 \\ & z_2 & & \\ & & \ddots & \\ 0 & & & z_{2n} \end{bmatrix}^{\ell-1} \begin{bmatrix} 1 & z_1 & \cdots & z_1^{2n} \\ 1 & z_2 & \cdots & z_2^{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & z_{2n} & \cdots & z_{2n}^{2n} \end{bmatrix}. \tag{10}$$

However, it follows from equation (8) that

$$\begin{bmatrix} 1 & z_1 & \cdots & z_1^{2n} \\ 1 & z_2 & \cdots & z_2^{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & z_{2n} & \cdots & z_{2n}^{2n} \end{bmatrix} \begin{Bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_{2n} \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{Bmatrix}. \tag{11}$$

Therefore, $V(\ell, 1:2n+1)[\alpha_0, \alpha_1, \ldots, \alpha_{2n}]^{\mathrm{T}} = 0$ and $V(1:2n, 1:2n+1)[\alpha_0, \alpha_1, \ldots, \alpha_{2n}]^{\mathrm{T}}$ results in the null vector $[0, 0, \ldots, 0]^{\mathrm{T}}$. Hence, the entries in each row of the Hankel matrix are related to the coefficients of the characteristic polynomial. Since $\alpha_{2n} = 1$, the characteristic polynomial coefficients can be determined directly from the Hankel matrix. Partitioning the right-band column and inverting the remainder, we obtain

$$[\alpha_0, \alpha_1, \ldots, \alpha_{2n-1}]^{\mathrm{T}} = \cdots - V(1:2n, 1:2n)^{-1} V(1:2n, 2n+1), \tag{12}$$

where $[\,]^{-1}$ denotes the matrix inverse operator. The poles z_i can then be calculated by solving equation (8).

We emphasize that the above procedure is valid when the signal contains exactly $2n$ complex exponentials. If the signal contains less than $2n$ complex exponentials, the calculation of the inverse of the matrix will be ill-conditioned. If the signal contains more than $2n$ complex exponentials, the calculation will not respect the *true* spectral information and may not match the data exactly.

Having calculated the characteristic delays z_i, we determine the residues, A_i from

$$\begin{Bmatrix} A_1 \\ A_2 \\ \vdots \\ A_{2n} \end{Bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ z_1 & z_2 & \cdots & z_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ z_1^{2n-1} & z_2^{2n-1} & \cdots & z_{2n}^{2n-1} \end{bmatrix}^{-1} \begin{Bmatrix} v(1) \\ v(2) \\ \vdots \\ v(2n) \end{Bmatrix}. \tag{13}$$

We note that this computation requires only the first $2n$ data points. When dealing with overdetermined data sets in the following section, all of the data are utilized in establishing

the residues. Here, the first $2n$ data points produce results that are equivalent to those obtained with any other combination of $2n$ data points in $v(\ell)$. In the next section, we provide two generalizations of Prony's method: namely, the treatment of overdetermined data and noise.

2.2. COMPLEX EXPONENTIAL ALGORITHM

In this section, we review the complex exponential algorithm [2]; *which is known as the least-squares Prony's method, the autoregressive technique, or the maximum entropy method.* The complex exponential algorithm is an *estimation procedure* as opposed to a *transformation.* Prony's method is a transformation because $4n$ data points are represented by $4n$ parameters. The complex exponential algorithm represents $N \geqslant 4n$ data points with $4n$ parameters, as in equation (1) or (2), in the presence of noise. We assume that N evenly sampled data points are known to contain exactly $2n$ distinct complex exponentials in the presence of Gaussian white noise, as in equation (2). We intend to uncover the parameters of equation (1) and quantify the noise contribution.

The basic concept of the complex exponential algorithm is to utilize data redundancy not considered by Prony's method to provide an averaged or, at least, a more consistent result. A closed-form solution is established by breaking the problem into two linearized stages. The first stage uncovers the characteristic delays and the second one uncovers the associated residues. Again, we examine the structure of the Hankel matrix,

$$Y(1:N-2n,\,1:2n+1)=\begin{bmatrix} y(1) & y(2) & \cdots & y(2n+1) \\ y(2) & y(3) & \cdots & y(2n+2) \\ \vdots & \vdots & \ddots & \vdots \\ y(N-2n) & y(N-2n+1) & \cdots & y(N) \end{bmatrix}, \tag{14}$$

where $y(\ell)=v(\ell)+\varepsilon(\ell)$. In this case, the column length of Y is $N-2n$ and is chosen so that the index ℓ does not exceed N. Each row of the Hankel matrix $Y(\ell, 1:2n+1)$ can still be related by the coefficients of the characteristic polynomial, in equation (7), with an associated error $e(\ell)$ given by

$$e(\ell)=Y(\ell,1:2n+1)\begin{Bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_{2n} \end{Bmatrix}. \tag{15}$$

This error must be equal to $\sum_{k=0}^{2n}\alpha_k\varepsilon(\ell+k)$ according to the definition of $y(\ell)$ in equation (2). The first stage of the least-squares solution is obtained by summing the squares of the errors and minimizing the primary objective function

$$SSE_1(\alpha_0,\ldots,\alpha_{2n-1};\alpha_{2n})=\sum_{\ell=1}^{N-2n+1}e(\ell)^2. \tag{16}$$

The solution is expressed as

$$[\alpha_0,\alpha_1,\ldots,\alpha_{2n-1}]^{\mathrm{T}}=-Y(1:N-2n,1:2n)^{+}\,Y(1:N-2n,2n+1), \tag{17}$$

where $\alpha_{2n} = 1$ and the matrix pseudo-inverse operator $[\ ]^+ \equiv [[\ ]^{\mathrm{T}}[\ ]]^{-1}[\ ]^{\mathrm{T}}$ for real quantities and $[\ ]^+ \equiv [[\ ]^{\mathrm{H}}[\ ]]^{-1}[\ ]^{\mathrm{H}}$ for complex ones. The poles z_i are determined from equation (8).

In the second stage, the residues A_i are determined by defining a secondary objective function as

$$\mathrm{SSE}_2(A_1, \ldots, A_{2n}) = \sum_{\ell=1}^{N} \varepsilon(\ell)^2. \tag{18}$$

The minimum of this function is realized when

$$\begin{Bmatrix} A_1 \\ A_2 \\ \vdots \\ A_{2n} \end{Bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ z_1 & z_2 & \cdots & z_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ z_1^{N-1} & z_2^{N-1} & \cdots & z_{2n}^{N-1} \end{bmatrix}^+ \begin{Bmatrix} y(1) \\ y(2) \\ \vdots \\ y(N) \end{Bmatrix}. \tag{19}$$

We note also that this computation involves all N data points in contrast to equation (13). Prony's method and the complex exponential algorithm provide the same result if either the data matches equation (1) or $N = 4n$.

3. CIRCULAR-HYPERBOLIC DECOMPOSITION

In this section, we detail three implementations of the new circular–hyperbolic decomposition technique (CHD). In the first implementation, we utilize a two-stage least-squares approach that parallels the complex exponential algorithm. In the second one, we modify the technique incorporating a novel concept of noise reduction attributed to Pisarenko [1]. In the third implementation, we provide additional means of noise reduction through a covariance formulation that avoids zero-lag terms. These implementations are discussed in sections 3.1–3.3 respectively.

The basic concept of the CHD involves constraining the form of the characteristic polynomial coefficients α_k defined in equation (7). However, constraining these coefficients is not a new concept [1, 5]. In the Pisarenko harmonic decomposition (PHD) [1], we find the α_k as the eigenvector associated with the smallest eigenvalue μ_0 of

$$\begin{bmatrix} B(0) & B(1) & \cdots & B(2n) \\ B(1) & B(0) & \cdots & B(2n-1) \\ \vdots & \vdots & \ddots & \vdots \\ B(2n) & B(2n-1) & \cdots & B(0) \end{bmatrix} \begin{Bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_{2n} \end{Bmatrix} = \mu_0 \begin{Bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_{2n} \end{Bmatrix}.$$

Here, $B(k)$ is a known covariance function of a real stationary signal of order $2n$. For instance, we often take the covariance function to be the autocorrelation function. Observe that the first row reads exactly the same as the last row in reverse, the second row reads exactly the same as the second to least row in reverse, and so forth. Thus, the PHD implicitly enforces the symmetry requirement

$$\alpha_{n+i} = \alpha_{n-i} \quad \text{for } i \in \{1, 2, \ldots, n\}. \tag{20}$$

It can be shown that a stationary signal whose autocorrelation function satisfies equation (20) must be composed of circular sine and cosine functions. Pisarenko [1] and Marple [5] provide additional insights into this point.

We should stress that, as a cautionary note, we do not detail the implementation of the PHD. We utilize some concept from the PHD in Section 3.2, but the implementation of Section 3.2 should not be confused with he PHD. A detailed description of the PHD is given in reference [1].

The α_k can certainly be estimated from a covariance function, as is the case with the PHD. The process of estimating a covariance function has its own associated difficulties and errors, particularly for signals with a short duration. The phase information of a signal cannot generally be recovered from a covariance function. It is also obvious that covariance functions are also ill-defined for hyperbolic functions in the stationary sense. Hence, it may not be desirable to estimate a covariance function and subsequently estimate the signal parameters directly from the signal itself.

As detailed in section 2.1, the α_k also relate the columns of the Hankel matrix of a sampled signal. But, the Hankel matrix does not enjoy the symmetries observed above in the covariance matrix B. This section details different methods of enforcing the symmetries in the α_k. We note that the method that we use to enforce this symmetry is similar to the one used in the Prony spectral line (PSL) estimation technique [5]. However, we do not provide details of the PSL.

Another important issue to address, here, is that an estimation algorithm can neither recognize that a signal is stationary nor generally enforce the requirements of stationarity. Upon close examination, equation (20) can also be satisfied by a large class of non-stationary functions; namely, hyperbolic sine and cosine functions and the products of hyperbolic sine and cosine functions and circular sine and cosine functions. Enforcing the symmetry conditions in equation (20) is broader than just a means to uncover circular sine and cosine functions. It provides the means for decomposing a time-limited signal in terms of circular and hyperbolic sine and cosine functions.

3.1. THE LEAST-SQUARES PROCEDURE

In this section, we describe how to constrain the complex exponential algorithm to decompose a signal in terms of circular-hyperbolic functions in the presence of noise, as expressed in equation (4). We begin by examining the structure of the characteristic polynomial for circular and hyperbolic functions; that is, $p\cos(\omega t_\ell) + q\sin(\omega t_\ell)$, $a\cosh(\kappa t_\ell) + b\sinh(\kappa t_\ell)$, or $(-1)^{\ell-1}[c\cosh(\eta t_\ell) + d\sinh(\eta t_\ell)]$). We observe that each of these functions has two poles $\{z_1, z_2\}$ with the property

$$z_1 z_2 = 1. \tag{21}$$

Hence, the characteristic polynomial, for either a circular or a hyperbolic function, always takes the form

$$z^2 + (z_1 + z_2)z + 1 = z^2 + r_1 z + 1, \tag{22}$$

where r_1 is real. Similarly, we examine the structure of the characteristic polynomial for the terms involving the product of circular and hyperbolic functions in equation (3). We observe that each of these terms has four poles $\{z_1, z_2, z_3, z_4\} = \{z, 1/z, z^*, 1/z^*\}$ with the property

$$z_1 z_2 = z_3 z_4 = 1. \tag{23}$$

Therefore, the characteristic polynomial for the inner product of a circular and a hyperbolic function, always takes the form

$$[z^2 + (z_1 + z_2)z + 1][z^2 + (z_3 + z_4)z + 1] = \prod_{i=1}^{2} (z^2 + r_i z + 1), \tag{24}$$

where the r_i are real. Subsequently, the characteristic polynomial for n circular–hyperbolic functions, where the product of a circular and hyperbolic function counts as 2, can be expressed as

$$\begin{aligned}\prod_{i=1}^{n} [z^2 + r_i z + 1] &= z^{2n} + \sum_{i=1}^{n} r_i z^{2n-1} + \sum_{i=1}^{n} \sum_{k=1}^{n} r_i r_k z^{2n-2} + \cdots + \sum_{i=1}^{n} r_i z + 1 \\ &= z^{2n} + \alpha_{2n-1} z^{2n-1} + \alpha_{2n-2} z^{2n-2} + \cdots + \alpha_1 z + 1. \end{aligned} \tag{25}$$

Examining the structure of equation (25), we note that the polynomial coefficients exhibit the symmetries

$$\alpha_{n-i} = \alpha_{n+i} \quad \text{for } i \in \{1, 2, \ldots, n\} \tag{26}$$

which provides n constraints. This symmetry requirement is exactly the same as in equation (20).

Next, we choose the matrix

$$\tilde{Y}(1:N-2n,\, 1:n+1) = \left[\begin{array}{c|ccc} y(n+1) & y(n) + y(n+2) & \cdots & y(1) + y(2n+1) \\ y(n+2) & y(n+1) + y(n+3) & \cdots & y(2) + y(2n+2) \\ \vdots & \vdots & \ddots & \vdots \\ y(N-n) & y(N-n-1) & \cdots & y(N-2n) + y(N) \end{array}\right]. \tag{27}$$

Due to the symmetries in equation (26), the objective function, expressed in equation (16), has a constrained minimum when

$$[\alpha_n, \alpha_{n+1}, \ldots, \alpha_{2n-1}]^{\mathrm{T}} = -\tilde{Y}(1:N-2n, 1:n)^{+}\, \tilde{Y}(1:N-2n, n+1), \tag{28}$$

where $\alpha_{2n} = 1$. Here, we note that the minimum number of data points required is reduced; equation (28) can be executed when $N \geqslant 3n$, whereas equation (17) requires that $N \geqslant 4n$. The polynomial coefficients $\{\alpha_0, \alpha_1, \ldots, \alpha_{n-1}\}$ are given by equation (26), and the poles are calculated by solving equation (8) as before. Then, the residues can be calculated according to equation (19). If the results can be presented in exponential form, the decomposition is complete.

If we were using either the PHD or the PSL, we would not know how to contend with z domain pole estimates away from the unit circle because they are not expected. This can be a practical concern when implementing either the PHD or the PSL. One approach to resolve this dilemma, for spectral line estimation, is to ignore the modulus of the pole estimates and to recalculate them via $\exp(j\,\mathrm{Arg}(z))$, where $\mathrm{Arg}(z)$ denotes the argument of z (the polar angle on the z domain). Although, this approach may be appropriate for practical purposes, it is not correct in a rigorous sense. It ignores the basic fact that the solution set that satisfies the symmetry requirement in equation (20) is broader than just the circular sine and cosine functions.

We also observe that equation (3) actually extends to the entire z domain. One might now wonder if there is really any difference between equations (1) and (3). The difference is one of emphasis. This symmetry requirement encourages the pole estimates to occur on either the unit circle or the real line, where it takes only one degree of freedom $n = n_\omega$, n_κ, or n_η per frequency. Elsewhere, it takes two degrees of freedom $n = 2n_\lambda$ per frequency. This also suggests a potential for difficulties in large-order problems or problems where the model order is highly overspecified with the so-called computational degrees of freedom. The full benefits over the complex exponential algorithm may not be realized in such instances.

When the results are desired in the form of equation (3), some additional algebra is required. First, we collect the poles on the unit circle and pair complex conjugate poles and residues. The circular frequencies ω_i and amplitudes p_i and q_i associated with the pole pair $\{z, z^*\}$ and corresponding residues $\{A, A^*\}$ are

$$\omega = |\mathrm{Arg}(z)|/\Delta t, \qquad p = 2\ \mathrm{Real}(A), \qquad q = -2\ \mathrm{Imag}(A). \tag{29–31}$$

Next, we collect the positive poles and pair reciprocal poles and corresponding residues. The first hyperbolic frequencies κ_i and amplitudes a_i and b_i associated with the pole pair $\{z, 1/z\}$ ($|z| > 1$) and corresponding residues $\{A_z, A_{1/z}\}$ are

$$\kappa = \ln(z)/\Delta t, \qquad a = (A_z + A_{1/z}), \qquad b = (A_z - A_{1/z}). \tag{32–34}$$

Then, we collect the negative poles and pair reciprocal poles and corresponding residues. The second hyperbolic frequencies η_i and amplitudes c_i and d_i associated with the pole pair $\{z, 1/z\}$ ($|z| > 1$) and corresponding residues $\{A_z, A_{1/z}\}$ are

$$\eta = \ln(|z|)/\Delta t, \qquad c = (A_z + A_{1/z}), \qquad d = (A_z - A_{1/z}). \tag{35–37}$$

Finally, the remaining poles must correspond to the λ in equation (3). The frequencies λ_i and amplitude e_i, f_i, g_i, and h_i associated with the poles $\{z, 1/z, z^*, 1/z^*\}$ ($|z| > 1$ and $\mathrm{Im}(z) > 0$) and corresponding residues $\{A_z, A_{1/z}, A_{z^*}, A_{1/z^*}\}$ are

$$\lambda = \ln(z)/\Delta t, \qquad e = 2\ \mathrm{Re}(A_z + A_{1/z}), \qquad f = -2\ \mathrm{Im}(A_z - A_{1/z}), \tag{38–40}$$

$$g = 2\ \mathrm{Re}(A_z - A_{1/z}), \qquad h = -2\ \mathrm{Im}(A_z + A_{1/z}). \tag{41, 42}$$

Equations (29)–(42) provides the estimates of the parameters for the deterministic part of the signal in the form of equation (3). The noise part can be obtained by generating equation (3) from the parameters estimates and substracting the result from the original data. Equations (19), (26), and (28) represent the basic formulation of CHD.

3.2. NOISE REDUCTION VIA THE PISARENKO PROCEDURE

In this section, we utilize a noise reduction from Pisarenko to establish an unbiased estimate of the characteristic polynomial from the covariance matrix. Pisarenko [1] showed that this is possible if the effect of noise on the covariance matrix is known. However, the method presented here is not the PHD and should not be confused as such.

The expected value of the covariance of y can be expressed as

$$E[y(i)y(k)] = \begin{cases} v(i)v(i) + \mu_0, & \text{for } i = k, \\ v(i)v(k), & \text{for } i \neq k, \end{cases} \tag{43}$$

where $E[\ ]$ denotes the expected value. The expected value of the random part is zero whenever $i \neq k$. When $i = k$, the expected value of the random part approaches the noise variance

$$\mu_0 = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \varepsilon(i)^2 \tag{44}$$

which is a zero-lag term. Defining $\tilde{y}_i$ as the ith column of $\tilde{Y}$, we form a special time-dependent or transient covariance matrix as

$$C(1:n+1, 1:n+1) = \begin{bmatrix} \tilde{y}_1^{\mathrm{T}} \tilde{y}_1 & \tilde{y}_1^{\mathrm{T}} \tilde{y}_2 & \cdots & \tilde{y}_1^{\mathrm{T}} \tilde{y}_{n+1} \\ \tilde{y}_2^{\mathrm{T}} \tilde{y}_1 & \tilde{y}_2^{\mathrm{T}} \tilde{y}_2 & \cdots & \tilde{y}_2^{\mathrm{T}} \tilde{y}_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{y}_{n+1}^{\mathrm{T}} \tilde{y}_1 & \tilde{y}_{n+1}^{\mathrm{T}} \tilde{y}_2 & \cdots & \tilde{y}_{n+1}^{\mathrm{T}} \tilde{y}_{n+1} \end{bmatrix} \tag{45}$$

which is an unconventional expression due to our definition of $\tilde{Y}$ and subsequently $\tilde{y}_i$. The expected value of C can be expressed as

$$E[C(1:n+1, 1:n+1)] = \begin{bmatrix} \tilde{v}_1^{\mathrm{T}} \tilde{v}_1 + \mu & \tilde{v}_1^{\mathrm{T}} \tilde{v}_2 & \cdots & \tilde{v}_1^{\mathrm{T}} \tilde{v}_{n+1} \\ \tilde{v}_2^{\mathrm{T}} \tilde{v}_1 & \tilde{v}_2^{\mathrm{T}} \tilde{v}_2 + 2\mu & \cdots & \tilde{v}_2^{\mathrm{T}} \tilde{v}_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{v}_{n+1}^{\mathrm{T}} \tilde{v}_1 & \tilde{v}_{n+1}^{\mathrm{T}} \tilde{v}_2 & \cdots & \tilde{v}_{n+1}^{\mathrm{T}} \tilde{v}_{n+1} + 2\mu \end{bmatrix} \tag{46}$$

where $\tilde{v}_i$ is the noise-free part of $\tilde{y}_i$ and μ is μ_0 times the column length of $\tilde{Y}$. The eigenvalue decomposition

$$\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & \frac{1}{2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{2} \end{bmatrix} C(1:n+1, 1:n+1) = \Psi \Lambda \Psi^{-1} \tag{47}$$

provides the eigenvalues as the components of the diagonal matrix Λ and the eigenvectors as the column of Ψ. It follows from the structure of equation (46) that the eigenvector $\Psi(:, 1)$ associated with the minimum eigenvalue $\Lambda(1)$ is an unbiased estimate of the characteristic polynomial coefficients $\{\alpha_n, \alpha_{n+1}, \ldots, \alpha_{2n}\}^{\mathrm{T}}$ for which μ is a minimum. Additionally, the eigenvalue decomposition may provide improved numerical stability in comparison to equation (28). Furthermore, the value of $\Lambda(1)$ divided by the column length of $\tilde{Y}$ provides an estimate of the noise variance μ_0. We note that α_{2n} need not be set equal to unity. As before, the coefficients $\{\alpha_0, \alpha_1, \ldots, \alpha_{n-1}\}$ are determined according to equation (26). Then, the poles and residues are calculated using equations (8) and (13) respectively.

3.3. NOISE REDUCTION VIA THE COVARIANCE PROCEDURE

In the preceding section, we observed that only the zero-lag terms of the expected value of the covariance matrix C are affected by Gaussian white noise. Assuming that all of our other model assumptions are satisfied, the least-squares formulation of the complex exponential

algorithm biases results because the errors $e(\ell)$ in equation (15) are not statistically independent. The statistical independence is an important and an often violated assumption involved with the least-squares regression [7]. We begin with the observation that equation (28) can also be expressed in terms of the covariance matrix as

$$\begin{Bmatrix} \alpha_n \\ \alpha_{n+1} \\ \vdots \\ \alpha_{2n-1} \end{Bmatrix} = -C(1:n, 1:n)^{-1} C(1:n, n+1). \tag{48}$$

We observe from equation (46) that while $E[C(i, i)]$ contains a noise contribution, $E[C(i, k)]$ does not whenever $i \neq k$.

By letting

$$\tilde{C}(1:n, 1:n+1) = \tilde{Y}(1:N-2n, 1:n)^{\mathrm{T}} \tilde{Y}(2n+1:N, 1:n+1), \tag{49}$$

we can obtain an unbiased estimate of the characteristic polynomial coefficients via

$$\begin{Bmatrix} \alpha_n \\ \alpha_{n+1} \\ \vdots \\ \alpha_{2n-1} \end{Bmatrix} = -C(1:n, 1:n)^{-1} \tilde{C}(1:n, n+1). \tag{50}$$

because the α_k also relate the noise-free portions of the covariance matrix, as shown for the Hankel matrix in section 2.1. The key feature of this approach involves avoiding the zero-lag terms of the covariance matrix. The remaining coefficients $\{\alpha_0, \alpha_1, \ldots, \alpha_{n-1}\}$ are obtained according to equation (26). The poles and residues are determined according to equations (8) and (13), respectively.

We note that, in general, the data do not conform, with our other model assumption, and the assumption of statistical independence may or may not be significant in comparison with other potential violations of the model assumptions.

4. APPLICATIONS

In this section, we provide experimental and numerical examples of the application of the CHD. In section 4.1, we consider an experiment involving the motions of a shaker system with harmonic distortion. The data obtained from this experiment appear to be at least weakly stationary. For stationary data, only the sine and cosine terms are significant. In section 4.2, we uncover the model parameters from numerically generated data of a cable deformed under the effect of gravity.

4.1. EXPERIMENTAL EXAMPLE: HARMONIC DISTORTION OF A SHAKER SYSTEM

In this system, we implement the CHD described in sections 3.1–3.3 to uncover the harmonic distortion of a shaker system. A signal generator provides a sinusoidal command voltage to the shaker amplifier. An accelerometer is studded to a reaction mass attached to the shaker's armature. An acquistion system samples the analog signal from the accelerometer and associated hardware into a record of 160 s at the rate of 64 samples/s.

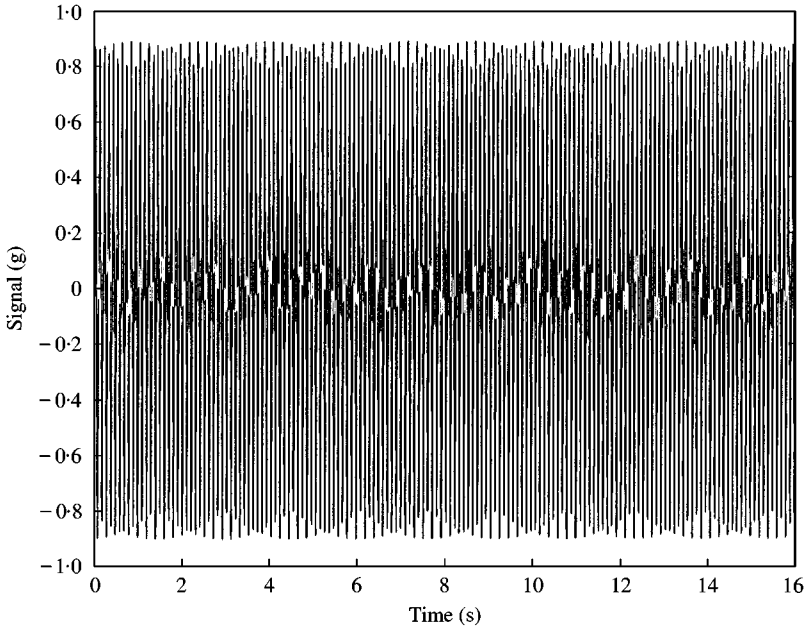

Figure 1. Experimental example of acceleration time series $y(t_i)$.

The record is carved into 37 ensembles with 75% overlap. The duration of each ensemble or sampled window is 16 s, as shown in Figure 1. The frequency of the sinusoidal command voltage is approximately equal to 10 Hz. However, the excitation frequency is not known *a priori* and periodicity within the sampled window is not ensured.

The first issue is to establish a model order m for the signal. Perhaps, the most common method is to examine the singular values of the singular value decomposition of the covariance matrix $Y^{\mathrm{T}}Y = QDQ^{\mathrm{T}}$ or Hankel matrix $Y = PD^{1/2}Q^{\mathrm{T}}$ normalized with the column length of Y. When $D(m+1)$ is small, the model order is m. We demonstrated in equation (35) that Gaussian white noise results in positive contributions to the zero-lag covariance terms $C(i, i)$. In low-noise situations, the singular values drop sharply to approximately μ once the appropriate model order has been exceeded, and they maintain the same order of magnitude for subsequent values. Since the smallest singular value does not approach zero ever for large m, the singular values of $Y^{\mathrm{T}}Y$ are termed biased singular values. Sinusoids with mean-squared values of the same order as μ are indistinguishable from noise. Conversely, the smaller singular values of $Y(:, 1:m)^{\mathrm{T}}Y(:, m+1:2m)$ do tend towards zero, since zero-lag terms are avoided; they are termed unbiased singular values. In moderate noise situations, the unbiased singular values may provide a valuable tool for choosing m when $N \gg 1$.

In Figure 2, we show the biased and unbiased singular values for the signal given in Figure 1. In this situation, the difference between the biased and unbiased singular values is small, yet noticeable, for the smaller values of m. The biased singular values drop-off sharply from 5 to -28 dB as m exceeds 2 and from -52 to -65 dB as m exceeds 7. Since $D(1)$ and $D(2)$ are much greater then $D(3)$, one is tempted to use $m = 2$. But the singular values above $m = 2$ continue to decrease with increasing m, implying that the next few terms are small deterministic signal components. However, the biased singular values $D(m)$ for $m > 8$ are roughly the order of magnitude of $D(8)$, implying that they may be attributed to noise. We

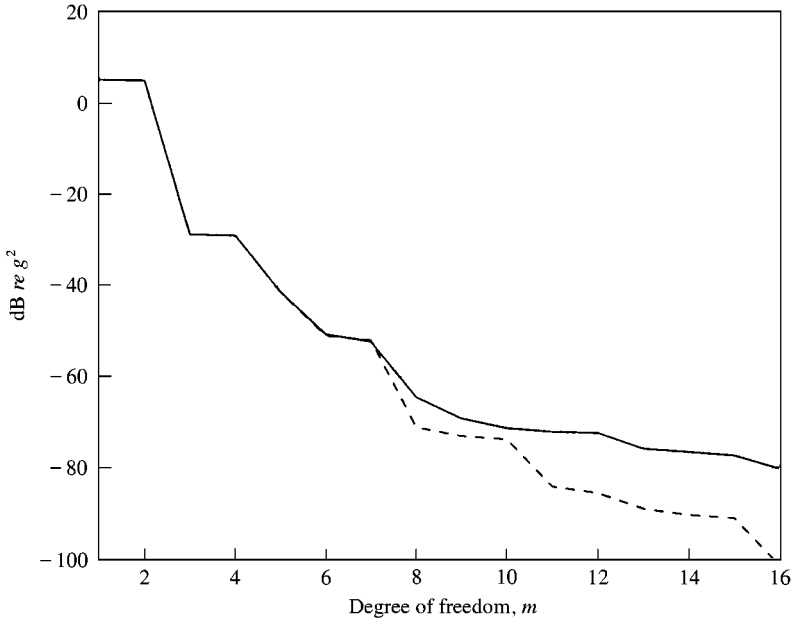

Figure 2. Experimental example of singular values: ——, biased singular values; - - -, unbiased singular values.

also notice that the value of the unbiased versus biased singular values diverge at $m = 8$, also implying a model order of $m = 8$. We choose the model orders of $m = 2$ and 8 for the signal. The first choice demonstrates the performance of CHD in the presence of biased noise due to an under-valued model order. The second choice demonstrates the performance of the CHD with a well-selected model order. The differences between the original signal $y(t_\ell)$ and its estimates $v(t_\ell)$ using the procedure implemented in section 3.1 with model orders $m = 2$ and 8 are shown in Figures 3(a) and 3(b) respectively. A comparison of the two figures shows that the error is reduced significantly by increasing the model order from $m = 2$ to 8. We should also note that the error is large near the start and the end of the record. This can be attributed to the error in the frequency estimates. This is akin to the problem of leakage or wrap around error associated with the DFT [8, 9].

Estimates of the frequency, amplitude, and associated errors with model order $m = 2$ for the three implementations are shown in Table 1. The results obtained by implementing the scheme in section 3.2 are quantitatively and qualitatively better than those obtained with the schemes in sections 3.1 and 3.3. However, this is not always the case. The model order is smaller than the order of the signal. Hence, the noise is deterministic and biased. The form of noise violates the basic assumptions used in the development of the algorithms. Surprisingly, the correlation γ^2 between measured and estimated values is near unity for the first two cases and reasonable for the third one, which implies that the performance of the estimation algorithms is fair.

Next, we assume that $m = 8$ and implement the CHD for each of the 37 ensembles. The three implementations provide about 30–45 dB better results than the results obtained with $m = 2$. This time the decomposition can explain approximately the first 55 dB of the signal and the results are more consistent among the various estimation schemes. The signal is composed of the primary frequency, its first and second harmonics, and an extremely small near-zero frequency component. A summary of the mean results is provided in Table 2. The components of the first and second harmonic are -34 and -57 dB below the primary

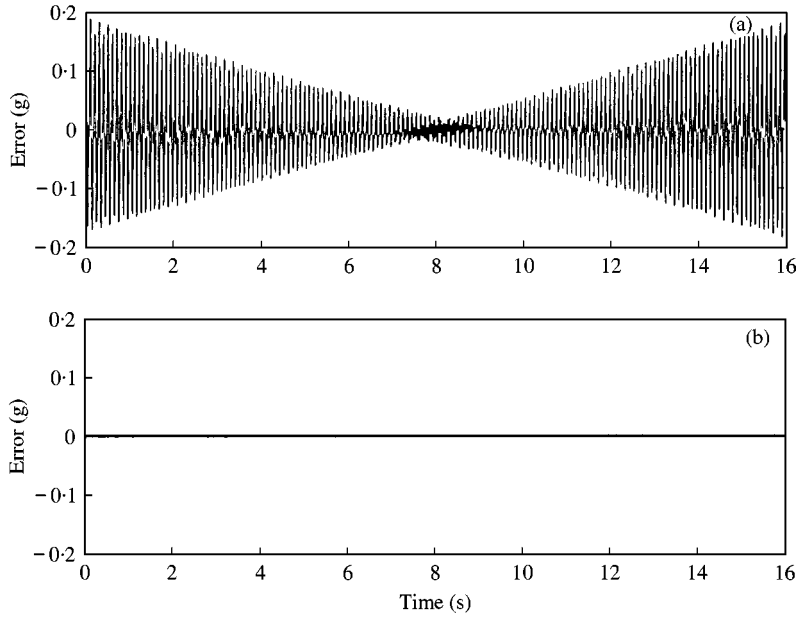

Figure 3. Experimental example of error $\varepsilon(t_\ell)$ using the technique of section 3.1: (a) $m = 2$ and (b) $m = 8$.

TABLE 1

Experimental example—mean parameter estimates of 37 *ensembles* $n = 1$

Method	ω_1 (Hz)	$\lvert p_1 + jq_1 \rvert$ (g)	γ^2 (dB)	(S/N)
Section 3.1	9·80155(0·00038)	0·8887(0·0013)	0·9811	17·2
Section 3.2	9·79857(0·00037)	0·8959(0·0005)	0·9971	25·3
Section 3.3	9·80699(0·00063)	0·8587(0·0048)	0·9161	10.4

tone, respectively. In other words, the total harmonic distortion is about 2.0%. In Figure 4, we show the power spectral density (PSD) of the singular using a boxcar window, the PSD using CHD (section 3.1 and $m = 8$), and the PSD of the noise. We note that the implementation procedure of section 3.3 consistently represents the zero or near-zero frequency component with a small hyperbolic term, rather than a circular one. This emphasizes our assertion in section 3 that the algorithm itself does not enforce an assumption of stationarity.

For comparison, the peak parameters of the DFT are provided in Table 3, which demonstrate the sensitivity of the amplitudes of phase estimates to the frequency. Here, the effects of leakage associated with the DFT are much more significant than the errors observed for either of the cases documented in Figure 3. Each of the three implementations of the CHD describes the signal well.

4.2. NUMERICAL EXAMPLE: CABLE DEFORMATION DUE TO GRAVITY

In this section, we show how a signal composed of a hyperbolic function, with a small Gaussian white noise component, can be analyzed with the techniques discussed in sections

TABLE 2

Experimental example—mean parameter estimates of 37 *ensembles* $n = 4$

Method	ω_1 or $\kappa_i^{\dagger}$ (Hz)	$\lvert p_i + jq_i \rvert$ or $\lvert a_i + jb_i \rvert^{\dagger}$ (g)	γ^2 (dB)	(S/N)
Section 3.1	0·699393(0·367055)	0·000101(0·000064)	0·99999	55·2
	9·796872(0·000002)	0·896984(0·000145)		
	19·593836(0·000091)	0·018090(0·000113)		
	29·376525(0·001472)	0·001241(0·000038)		
Section 3.2	0·529780(0·284751)	0·000100(0·000056)	0·99999	55·8
	9·796872(0·000002)	0·896984(0·000144)		
	19·593825(0·000092)	0·018090(0·000113)		
	29·392829(0·000997)	0·001359(0·000025)		
Section 3.3	0·257881(0·142207)†	†,‡	0·99999	55·5
	9·796872(0·000015)	0·896985(0·000144)		
	19·593734(0·000096)	0·018090(0·000113)		
	29·399722(0·002036)	0·001317(0·000028)		

Note: Parenthesis denotes the standard deviation of the preceding parameter estimate.
†Denotes hyperbolic term.
‡Near-zero and ill-defined.

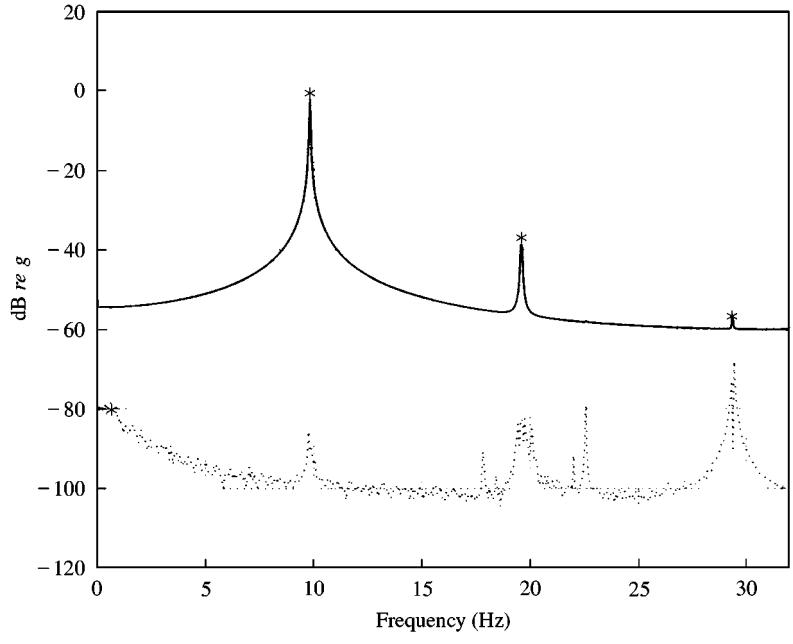

Figure 4. Experimental example of power spectrum: ——, DFT PSD; *, CHD PSD; - - -, error DFT PSD.

3.1–3.3. The frequency and amplitudes are given in Table 4. The signal, to be examined here, is generated numerically and corresponds to a cable supported at its ends by two pins and loaded by its own weight. The shape of this cable is hyperbolic. To conduct the proposed analysis, we normalize the signal to a unit length L and sample it given a sampling interval of $0{\cdot}025L$.

TABLE 3

Experimental example—peak amplitudes of the DFT and parameter estimates for a single ensemble $n = 4$

Method	ω_i or $\kappa_i^{\dagger}$ (Hz)	p_i or $a_i^{\dagger}$ (g)	q_i or $b_i^{\dagger}$ (g)	γ^2 (dB)	(S/N)
Section 3.1	0·424557	0·000011	0·000137	0·99999	53·6
	9·796870	0·604806	−0·661885		
	19·593810	−0·016996	0·006848		
	29·379217	0·000482	0·001168		
Section 3.2	0·332196	0·000094	0·000083	0·99999	53·9
	9·796870	0·604806	−0·661889		
	19·593802	−0·016993	0·006854		
	29·393740	−0·000444	0·001277		
Section 3.3	0·091571†	0·002675†	−0·002678†	0·99999	55·7
	9·796876	0·605007	−0·661712		
	19·593848	−0·017013	0·006810		
	29·402133	−0·000895	0·000942		
DFT	9·8125	0·805861	0·036122		
	19·625	−0·003412	0·010855		
	29·375	0·001227	−0·000963		

†Denotes hyperbolic term.

TABLE 4

Numerical example—mean parameter estimates and standard deviation $n = 1$

Method	κ_1 (L^{-1})	a_1 (L)	b_1 (L)	γ^2 (dB)	(S/N)
Actual	0·3	0·6	−0·4	0·999991	50·5
Section 3.1	0·2957(0·0134)	0·5965(0·0107)	−0·3923(0·0233)	0·999882	39·3
Section 3.2	0·2971(0·0132)	0·5962(0·0106)	−0·3946(0·0230)	0·999889	39·6
Section 3.3	0·2993(0·0133)	0·5994(0·0106)	−0·3984(0·0228)	0·999893	39·7

Note: Parenthesis denotes the standard deviation of the preceeding parameter estimate.

With the obvious assumption that $m = 2$, we conduct 25 trials of the CHD and obtain the mean parameter estimate and standard deviations in Table 4. The deterministic part of the signal is the same for all trials, whereas a different set of random numbers is utilized each time to simulate Gaussian white noise.

The performance of the implementations described in sections 3.1.–3.3 demonstrates a similar performance. In this example, the implementation described in section 3.3 provides marginally better results than the implementations described in sections 3.1 and 3.2. A typical example of the results for the implementation described in section 3.3. are graphically documented in Figures 5(a) and 5(b), representing one trial. The signal-to-noise ratio (S/N) and the correlation γ^2 were calculated on a power averaged basis among all of the trials. The implementations of the procedures in sections 3.1–3.3 were able to describe the first 39 dB of the signal. CHD describes the signal well and provides exceptional values for the correlation γ^2 between the measured values and the various estimates.

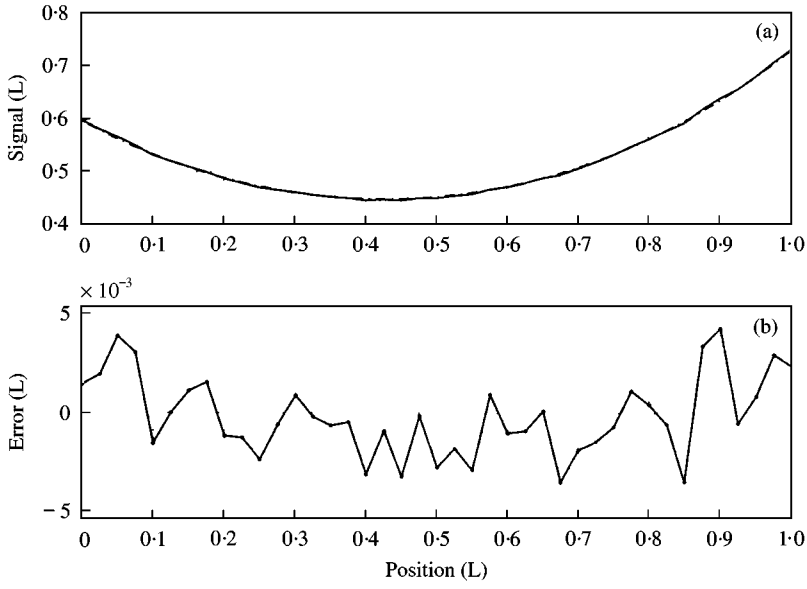

Figure 5. Numerical example of signal $y(t_\ell)$ and error $\varepsilon(t_\ell)$. (a) Signal: —, actual; - - -, model. (b) Error, section 3.3., with $n = 1$.

5. CONCLUSIONS

In this article, we establish constraints on the complex exponential algorithm to limit the estimation procedure to the special case of circular-hyperbolic functions. In sum, the algorithm and constraints are called the circular-hyperbolic decomposition (CHD). Three implementation procedures for this new algorithm are developed. The first is closely related to the least-squares approach of the complex exponential algorithm. The second and third utilize an eigenvalue approach and a covariance approach to avoid zero-lag components, respectively. Experimental and numerical examples of these implementation are provided. They demonstrate that the CHD is able to closely match data that can be represented as a linear combination of sines, cosines, hyperbolic sines, hyperbolic cosines, and their products. The CHD provides improved amplitude and phase estimates compared to the discrete Fourier transform when the frequencies of the circular functions are not integer multiples of each other in moderate noise environments. The present demonstration is applicable to a wide class of problems involving evenly sampled signals composed of circular and hyperbolic sine and cosine functions in the presence of Gaussian white noise.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research under Grant No. N00014-96-1-1123 and Air Force Office of Scientific Research under Grant No. F 49620-98-1-0393.

REFERENCES

1. V. F. PISARENKO 1973 *The Geophysical Journal of the Royal Astronomical Society* **33**, 347–366. The retrieval of harmonics from a covariance function.

2. F. R. SPITZNOGLE and A. H. QUAZI 1970 *Journal of the Acoustical Society of America* **47**, 1150–1155. Representation and analysis of time-limited signals using a complex exponential algorithm.
3. S. O'F. FAHEY and J. PRATT 1988 *Experimental Techniques*, *Society for Experimental Mechanics* **22**, 45–49. Time domain modal estimation techniques.
4. R. PRONY 1795 *Journal de l'École Polytechnique*, *Paris* **1**, 24–76. Essai Expérimental et Analytique: Sur les lois de la Dilatabilité des fluides élastiques et sur celles de la Force expensiver de la vapeur del'eau et de la vapeur de l'alkool, à différentes températures.
5. L. MARPLE 1979 *IEEE International Conference on Acoustics, Speech & Signal Processing* 159–161. Special line analysis by Pisarenko an Prony methods, Washington, D.C.
6. S. M. KAY and S. L. MARPLE JR. 1981 *Proceedings of IEEE* **68**, 1380–1419. Spectrum analysis—a random prespective.
7. D. G. KLEINBAUM, L. L. KUPPER and K. E. MULLER 1988 *Applied Regression Analysis and Other Multivariable Methods*, Chapter 5. Belmont, CA: Wadsworth Publishing; second edition.
8. F. J. HARRIS 1978 *Proceedings of IEEE* **66**, 51–84. On the use of windows for harmonic analysis with the discrete Fourier transform.
9. K. G. MCCOPNNELL 1995 *Vibration Testing*: *Theory and Practice*, Chapter 5. New York: Wiley.