Mechanical Systems and Signal Processing 40 (2013) 421–438

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

journal homepage: www.elsevier.com/locate/ymssp

# Signal decomposition and reconstruction using complex exponential models

Sau-Lon James Hu [a,*], Wen-Long Yang [b], Hua-Jun Li [b]

[a] Department of Ocean Engineering, University of Rhode Island, Narragansett, RI 02882-1197, USA
[b] Shandong Provincial Key Lab of Ocean Engineering, Ocean University of China, Qingdao 266100, PR China

ARTICLE INFO

*Article history:*
Received 17 September 2012
Received in revised form
2 May 2013
Accepted 30 June 2013
Available online 21 July 2013

*Keywords:*
Signal decomposition
Prony's method
State-space model
Numerical conditioning
Numerical stability

ABSTRACT

The theme of this paper is signal decomposition and reconstruction, not specific for or limited to system identification. In dealing with aperiodic damped signals, *Prony-based techniques* — which decompose a signal into real- and/or complex-valued exponential components — are often utilized. In essence, the derivation of Prony's method has been based on a high order homogeneous difference equation. In this paper, an alternative approach that uses a first-order matrix homogenous difference equation (state-space model) to replace the high order homogenous difference equation is advocated. Although the proposed method and Prony's method appear to be theoretically identical, this paper shows that they are drastically different over crucial numerical issues, including conditioning and stability. While Prony's method is very sensitive to sampling rate and round-off error, the proposed method is not. While Prony's method has trouble to deal with noise embedded in the signal, the proposed method can handle noisy signals properly because it has a build-in noise rejection mechanism via the usage of truncated singular value decomposition. While root-finding of a high order polynomial — a classical ill-conditioned problem — is a required step in Prony's method, the proposed method completely avoids it. The proposed method is also applicable to intermittent signals, and can recover the missing parts of intermittent signals nicely through reconstruction.



## 1. Introduction

For many physical and mathematical applications, under the assumption that a given signal is the superposition of many individual components, it is often desirable to decompose this signal into individual components. The most widely used signal decomposition techniques are Fourier-based, which decompose a signal into the sum of many, possibly infinite, simple sine and cosine functions. Fourier techniques involve the assumption that a signal is either infinite in duration or repetitive within some fundamental period over all time. However, a recorded signal is always finite in duration, most likely aperiodic, and even damped. To deal with an aperiodic damped signal, *Prony-based techniques* — which decompose a signal into real-valued and/or complex-valued exponential components — are often utilized. Other popular signal decomposition techniques include wavelet transform [1], empirical mode decomposition [2], etc.

From the signal decomposition perspective, the goal of this paper is identical to that of *Prony's method*, that is, to decompose a signal into several real/complex exponential components [3–5]. Prony's method is for fitting a model by using

* Corresponding author. Tel.: +1 401 874 6688; fax: +1 401 874 6837.
*E-mail addresses:* hu@egr.uri.edu (S.-L. James Hu), oucwillow@gmail.com (W.-L. Yang), huajun@ouc.edu.cn (H.-J. Li).

0888-3270/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ymssp.2013.06.037

equally spaced measured data points, and conventionally it has been referred to as a 3-step method. The three steps of Prony's method are: (1) determining the linear prediction parameters of a difference equation that fits the sampled data; (2) computing the roots, from which the estimates of damping factor and frequency of each exponential term are obtained, of a characteristic polynomial formed from the linear prediction coefficients; and (3) solving for a set of linear equations to yield the estimates of the exponential amplitude and sinusoidal initial phase of each component.

The fact that Prony's method is sensitive to the noise embedded in the signal has been well documented in the literature. Several computational techniques have been developed to deal with cases with noisy data, depending on the problem being attacked [6–9]. Three general methods which have been often employed to enhance identification of actual exponential signals in the data, and thus to improve the accuracy of the frequency and damping component estimates, are involving the use of (1) both forward and backward linear prediction polynomial zeros, (2) high prediction orders and (3) singular value decomposition (SVD). In the SVD approach, the number of exponentials in the signal is estimated by comparing the relative magnitudes of the singular values associated with a Hankel matrix obtained from the signal. As the signal-related singular values tend to be larger than the noise-related ones, the effect of using a *truncated singular value decomposition* (TSVD) is to increase the signal-noise ratio in the data prior to implementing Prony's method. The performance of parameter estimation can be further improved if a *structured low rank approximation* (SLRA) approach has been taken. In the SLRA approach, both low-rank (through TSVD approach) and Hankel properties of the prediction are achieved in the matrix approximation [10].

Even when a signal is noise free, implementing Prony's method on this signal may still pose an ill-conditioned problem if the sampling rate is not properly chosen [9,11,12]. In other words, a small variation of the sampling rate while conducting Prony's method might cause a significant difference in the estimation of component parameters. A selection scheme for the optimal sampling interval was studied by examining the condition number of the data matrix used in Prony's method, and it was found that the optimal sampling rate should be neither too fast nor too slow [9,12].

Also, the determination of the zeros of a polynomial, given the polynomial coefficients, is a classic example of an ill-conditioned problem [13,14]. Although root-finding of a polynomial is exactly what is required in the second step of Prony's method, published articles on studying improved Prony's methods seemly never mentioned this problem. While it is natural to expect that ill-conditioning occurs when the polynomial has zeros which are very close, the problem may also be extremely ill-conditioned for polynomials with well separated zeros [13,14]. A classical example is Wilkinson's polynomial [13].

To date, although many methods have been proposed to decompose a signal into complex exponential components, there is still a need for more efficient and reliable algorithms [15]. In this paper, an alternative approach which begins with a first-order matrix difference equation — a state-space model — to replace a high order difference equation used in Prony's method is advocated. The proposed method also involves three steps. While its third step is identical to that of Prony's method, the first two steps of the proposed method are: (1) obtaining a realization for the state matrix of first-order matrix difference equation from the Hankel matrices composed by the sampled data; and (2) computing the eigenvalues of the realization matrix and then obtaining the estimates of the frequency and damping factor of each exponential term. One obvious advantage of the newly proposed approach is completely avoiding the ill-conditioned problem of solving the zeros of a high order polynomial.

The central theme of this paper is the signal decomposition/reconstruction — not the system identification or parameter estimation — of a given signal. By definition, the system identification is to extract the embedded frequencies and damping factors from measured signals. In the classical system identification studies, one often assumes that the given signal is the output of a dynamic system subjected to either an impulse or a Gaussian white noise process. When the input is a Gaussian white noise process, it becomes a stochastic system identification [16,17]. An example of the stochastic system identification is based on an auto-regressive (AR) model which usually involves a step of computing the correlation function of the signal [18]. In contrast, the signal decomposition is deterministic in nature as it separates the given signal into a number of individual components. There are various reasons other than the system identification to perform the signal decomposition, including: data compression, noise removal, and others. *Because of the differences on their purposes, the best system identification method is not always the best signal decomposition method, and vice versa.* In a typical signal decomposition study, only a finite-length signal of interest is given, and an underlying model for the signal itself is often assumed while performing the decomposition. For example, the DFT (discrete Fourier transform) adopts a harmonic model with predetermined frequencies to decompose a signal into many harmonic components. For the model of the signal itself, the signal decomposition in this study is based on a complex exponential model. In the published literature, while state space models have been utilized in several studies for system identification or parameter estimation, the adopted models were often built upon the output of a time-invariant linear dynamic system subjected to the input being either an impulse or a Gaussian white noise process [18–20]. In this study, a simpler state space formulation based on the complex exponential form for the signal itself will be newly derived, without resort to the concept of input–output of a dynamic system. Another development of the mathematical derivation in this study is to extend the proposed technique to decomposing an intermittent signal—a signal with missing segments. Furthermore, the superiority about numerical conditioning for a state-space model over the corresponding high polynomial order model will be thoroughly investigated.

In this paper, four distinct examples are included in numerical studies. The first example illustrates the procedure of the proposed method by using only a few equally spaced data points; it also demonstrates the relationship between the minimal number of data points needed and the maximal number of exponential components extracted. The second

example, considering a signal composed of only undamped harmonic components, shows the difference in harmonic decompositions between the proposed method and the traditional DFT. The third example investigates the difference in performance between the proposed method and Prony's method; particularly the influences of sampling rate, round-off error and signal noise on both methods are examined. The last example demonstrates that the proposed method is also applicable to an intermittent signal, namely, a signal with missing segments.

## 2. Preliminaries

### 2.1. Prony series versus Fourier series

In this study, a signal $y(t)$ of finite duration $0 \le t < T$ is to be decomposed. When the signal $y(t)$ is decomposed into a complex Fourier series expansion

$$y(t) = \sum_{n=-\infty}^{\infty} \beta_n e^{i\omega_n t} \tag{1}$$

it automatically assumes that $y(t)$ is a periodic signal with period $T$, where $\omega_n = n\Delta\omega$ and $\Delta\omega = 2\pi/T$. Since $y(t)$ is a real-valued signal, the complex Fourier coefficients $\beta_{-n}$ must be equal to $\beta_n^*$, the complex conjugate of $\beta_n$.

In contrast to the Fourier series, the so-called *Prony series* is a linear combination of exponentials with real-valued and/or complex-valued exponents for the signal

$$y(t) = \sum_{n=1}^{p} \gamma_n e^{\lambda_n t} \quad 0 \le t < T \tag{2}$$

where $p$ is the number of terms. In Eq. (2), because $y(t)$ is a real-valued signal, $\lambda_n$ must either be real numbers or occur in complex conjugate pairs (not limited to be pure imaginary). Let $\lambda_n \equiv -\alpha_n + i\omega_n$, then $\alpha_n$ is the damping factor in seconds$^{-1}$ and $\omega_n$ is the frequency in radians. The coefficients $\gamma_n$ corresponding to complex exponents $\lambda_n$ must also appear in complex conjugate pairs. Let $\gamma_n \equiv A_n \exp(i\theta_n)$, then $A_n$ is the amplitude and $\theta_n$ is the sinusoidal initial phase in radians associated with $e^{\lambda_n t}$.

Eqs. (1) and (2) show the continuous time series. For digital signal analysis, the continuous $y(t)$ is not known and usually equally spaced samples are available. Denoting sampling interval $\Delta t$, $t_k = k\Delta t$, and $y_k \equiv y(t_k)$, where $k = 0, 1, \ldots, N-1$, one writes the discrete Fourier series as

$$y_k = \sum_{n=0}^{N-1} \beta_n x_n^k \tag{3}$$

where $x_n = e^{i(2\pi n/N)}$ and $N$ is the number of time steps.

Likewise, the Prony series corresponding to a discrete signal is expressed by

$$y_k = \sum_{n=1}^{p} \gamma_n z_n^k \tag{4}$$

where $z_n = e^{\lambda_n \Delta t}$.

The major difference between Eq. (1) (Fourier series) and Eq. (2) (Prony series), or between Eqs. (3) and (4), can be summarized below. First, because of its periodic assumption, a Fourier series is composed of many harmonic components, where each component is with a preset frequency which is an integer times the fundamental frequency. In contrast, a Prony series has its components being complex exponentials, which could be purely harmonic, damped harmonic or purely damped exponential components. Second, while Eq. (3) always consists of $N$ terms, Eq. (4) contains $p$ terms, where $p$ must be determined by properly analyzing the given signal, and often $p$ is much smaller than $N$. Third, each $x_n$ in Eq. (3) is known, *i.e.* $x_n = e^{i(2\pi n/N)}$, but each $z_n$ in Eq. (4) must be estimated first from the discrete signal $y_k$ before computing the complex coefficients $\gamma_n$.

### 2.2. Differential/difference equation for prony series

In Eq. (2), the exponential functions $e^{\lambda_n t}$, $n = 1, \ldots, p$ form a basis on the open interval $0 \le t < T$. Clearly, the Prony series (Eq. (2)) can also be viewed as the general solution of a $p$th-order homogenous linear ordinary differential equation with constant real-valued coefficients $a_n$ [21]

$$\sum_{n=0}^{p} a_n y^{(n)}(t) = 0, \quad \text{or } 0 \le t < T \tag{5}$$

where $y^{(n)}(t) = d^n y/dt^n$. And the exponents $\lambda_n$ $(n = 1, \ldots, p)$ in Eq. (2) are the roots of the characteristic equation associated with Eq. (5)

$$\sum_{n=0}^{p} a_n \lambda^n = 0 \tag{6}$$

If only the equally spaced sampled data $y_k$ ($k=0,1,\ldots,N-1$) are given, then Eq. (5) becomes a $p$th-order difference equation

$$\sum_{n=0}^{p} b_n y_{k+n} = 0 \quad \text{for } k=0,1,\ldots,N-p-1 \tag{7}$$

where $b_n$ are real-valued constants. Without losing the generality, let $b_p=1$. From Eqs. (4) and (7), the characteristic equation associated with Eq. (7) can be obtained to be [4]

$$\sum_{n=0}^{p} b_n z^n = 0 \tag{8}$$

Note that the linear difference equation and its characteristic equation have the same coefficients $b_n$.

### 2.3. Prony's method

In brief, Prony's method is a technique for fitting an exponential model to a few equally spaced measured data points. While the original Prony's method presented a method of exactly fitting as many purely damped exponentials as needed to fit the available data points, the modern version of Prony's method generalizes to damped sinusoidal models and also makes use of the least squares analysis to approximately fit an exponential model for cases where there are more data points than needed to fit to the assumed number of exponential terms. Three sequential steps, using Eqs. (7), (8), and (4), respectively, are implemented in Prony's method:

*Step 1:* One determines the linear prediction parameters $b_0, b_1, \ldots, b_{p-1}$ (noting $b_p=1$) that fit the given discrete signal by repeatedly using Eq. (7). In matrix form, it is written as

$$\mathbf{Yb} = -\mathbf{y}' \tag{9}$$

where

$$\mathbf{Y} = \begin{bmatrix} y_0 & y_1 & \cdots & y_{p-1} \\ y_1 & y_2 & \cdots & y_p \\ \vdots & \vdots & \ddots & \vdots \\ y_{N-p-1} & y_{N-p} & \cdots & y_{N-2} \end{bmatrix} \tag{10}$$

$$\mathbf{b} = [b_0 \;\; b_1 \;\; \ldots \;\; b_{p-1}]^T \tag{11}$$

$$\mathbf{y}' = [y_p \;\; y_{p+1} \;\; y_{N-1}]^T \tag{12}$$

*Step 2:* With the linear prediction coefficients **b** obtained from Step 1, the roots (denoted as $z_n$, $n=1,\ldots,p$) of the characteristic equation (Eq. (8)) are computed.

*Step 3:* Once $z_n$, $n=1,\ldots,p$, have been computed, the third step involves the solution of a second set of linear equations to yield the estimates of the exponential amplitude and sinusoidal initial phase. Based on Eq. (4), a matrix form is written as

$$\begin{bmatrix} z_1^0 & z_2^0 & \cdots & z_p^0 \\ z_1^1 & z_2^1 & \cdots & z_p^1 \\ \vdots & \vdots & \ddots & \vdots \\ z_1^{N-1} & z_2^{N-1} & \cdots & z_p^{N-1} \end{bmatrix} \begin{Bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_p \end{Bmatrix} = \begin{Bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{N-1} \end{Bmatrix} \tag{13}$$

Recall that the roots of Eq. (6), namely $\lambda_n$, are related to those of Eq. (8) through $\lambda_n = \ln z_n/\Delta t$. Appropriate linear least squares procedures for the first and third steps of the three-step Prony's method have sometimes been called the extended Prony's method. Since the second step of Prony's method is an ill-conditioned problem and round-off errors must exist for the linear prediction parameters **b** computed in the first step, the estimation of $z_n$ in the second step of Prony's method can have significant error.

## 3. First-order matrix equation model

To avoid dealing with an ill-conditioned problem on determining the roots of a high order polynomial encountered in Prony's method, a procedure to change the first two steps of Prony's method completely is taken. Instead of attacking at a

$p$th-order ordinary differential equation (ODE) directly, the $p$th-order equation is converted into a system of $p$ first-order ODEs.

### 3.1. First-order matrix differential equation

By introducing

$$x_1 = y^{(0)},\ x_2 = y^{(1)}, \ldots, x_p = y^{(p-1)} \tag{14}$$

and assuming $a_p = 1$, Eq. (5) is converted into a first-order matrix differential equation as

$$\dot{\mathbf{x}}(t) = \mathbf{F}\mathbf{x}(t) \tag{15}$$

where $\dot{\mathbf{x}}(t) = d\mathbf{x}(t)/dt$,

$$\mathbf{F} = \begin{bmatrix} 0 & 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & 0 & 1 \\ -a_0 & -a_1 & -a_2 & -a_3 & \ldots & -a_{p-2} & -a_{p-1} \end{bmatrix} \in \mathbb{R}^{p\times p} \tag{16}$$

and

$$\mathbf{x}(t) = \begin{Bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{Bmatrix} \in \mathbb{R}^p \tag{17}$$

The solution of Eq. (15) has the form $\mathbf{x}(t) = \mathbf{v}e^{\lambda t}$ where $\lambda$ is a constant and $\mathbf{v}$ is a time-independent $p$-element column vector. When $\mathbf{x}(t) = \mathbf{v}e^{\lambda t}$ is substituted into Eq. (15), it yields the algebraic eigenvalue equation

$$\mathbf{F}\mathbf{v} = \lambda\mathbf{v} \tag{18}$$

Theoretically, the $p$ eigenvalues of the matrix $\mathbf{F}$ are exactly the $p$ roots of the characteristic equation of Eq. (6) [22]. Instead of determining the roots of Eq. (6) required in Prony's method, the proposed method is to compute the eigenvalues of the matrix $\mathbf{F}$. In the area of numerical linear algebra, there are a number of efficient and accurate methods available to compute the eigenvalues and eigenvectors of arbitrary matrices. All these computational procedures avoid computing the coefficients of the corresponding characteristic polynomial [14,23].

The elements of vector $\mathbf{x}$ in Eq. (17) are often termed the state variables of the linear dynamic system, and they serve as the "coordinates" of the system. However, there are an infinite number of ways of selecting the coordinates. Denoting an arbitrary nonsingular transformation matrix by $\mathbf{\Phi}\in\mathbb{R}^{p\times p}$, and another set of coordinates by $\mathbf{q}\in\mathbb{R}^p$, one can formulate the transformation relationship between $\mathbf{x}$ and $\mathbf{q}$ as

$$\mathbf{x}(t) = \mathbf{\Phi}\mathbf{q}(t) \tag{19}$$

Now, substitute $\mathbf{x}(t) = \mathbf{\Phi}\mathbf{q}(t)$ and $\dot{\mathbf{x}}(t) = \mathbf{\Phi}\dot{\mathbf{q}}(t)$ into Eq. (15) and premultiply both sides by $\mathbf{\Phi}^{-1}$, the resulting equation becomes

$$\dot{\mathbf{q}}(t) = \mathbf{G}\mathbf{q}(t) \tag{20}$$

where $\mathbf{G} = \mathbf{\Phi}^{-1}\mathbf{F}\mathbf{\Phi}$. Note that Eqs. (20) and (15) share the same form. Because $\mathbf{G}$ is a similarity transformation of $\mathbf{F}$, they must have the same eigenvalues [24].

Let the eigenvalues and the corresponding eigenvectors of $\mathbf{G}$ be denoted by $\lambda_1, \lambda_2, \ldots, \lambda_p$ and $\mathbf{u}_{(1)}, \mathbf{u}_{(2)}, \ldots, \mathbf{u}_{(p)}$, respectively. Because $\mathbf{u}_{(1)}e^{\lambda_1 t}$, $\mathbf{u}_{(2)}e^{\lambda_2 t}, \ldots, \mathbf{u}_{(p)}e^{\lambda_p t}$ form a basis for solutions of Eq. (20), the general solution of Eq. (20) can be written as

$$\mathbf{q}(t) = \sum_{n=1}^{p} g_n \mathbf{u}_{(n)} e^{\lambda_n t} \tag{21}$$

where $g_n$ are arbitrary constants. In Eq. (21), each element of $\mathbf{q}(t)$ is a linear combination of $e^{\lambda_n t}$, $n = 1, \ldots, p$, so is $y(t)$ in Eq. (2). Therefore, $y(t)$ is a linear combination of the elements of $\mathbf{q}(t)$, that is,

$$y(t) = \sum_{n=1}^{p} c_n q_n(t) \tag{22}$$

or

$$y(t) = \mathbf{c}^T\mathbf{q} \tag{23}$$

where $q_n(t)$ is the $n$th element of $\mathbf{q}(t)$ and $\mathbf{c}\in\mathbb{R}^p$ is a column vector.

### 3.2. First order matrix difference equation

The discrete version of Eq. (20) is a first-order homogeneous matrix difference equation which can be written as [20]

$$\mathbf{q}_{k+1} = \mathbf{A}\mathbf{q}_k \tag{24}$$

where $\mathbf{q}_k \equiv \mathbf{q}(t_k)$ and $\mathbf{A} = \exp(\mathbf{G}\Delta t)$. From Eq. (23), at $t = t_k$, one has

$$y_k = \mathbf{c}^T\mathbf{q}_k \tag{25}$$

Here, in parallel to the first step of Prony's method, the specific goal is using the given discrete signal $y_k$, $k = 0, \ldots, N-1$, to obtain a realization of $\mathbf{A}$. Next, the eigenvalues of $\mathbf{A}$, and those of $\mathbf{F}$ (*i.e.* $\lambda_n$, $n = 1, \ldots, p$), can be computed without going through the procedure of determining polynomial coefficients and zeros. In the proposed method, the classical realization theory and algorithms are applied to compute $\lambda_n$, $n = 1, \ldots, p$. The basic development of the state-space realization is attributed to Ho and Kalman [25] who introduced the important principles of *minimum realization theory*. All minimum realizations have the same set of eigenvalues, which are modal parameters of the system itself. The methodology has been modified and substantially extended to develop the *eigensystem realization algorithm* [20] to identify modal parameters from noisy measurement data.

### 3.3. Proposed algorithm

By repeatedly using Eq. (24), one has $\mathbf{q}_k = \mathbf{A}^k\mathbf{q}_o$ for $k = 1, \ldots, N-1$. Substituting $\mathbf{q}_k = \mathbf{A}^k\mathbf{q}_o$ into Eq. (25) yields $y_k = \mathbf{c}^T\mathbf{A}^k\mathbf{q}_o$. Let the Hankel matrix $\mathbf{H}(k) \in \mathbb{R}^{\xi\times\eta}$ be defined as

$$\mathbf{H}(k) = \begin{bmatrix} y_k & y_{k+1} & \cdots & y_{k+\eta-1} \\ y_{k+1} & y_{k+2} & \cdots & y_{k+\eta} \\ \vdots & \vdots & \ddots & \vdots \\ y_{k+\xi-1} & y_{k+\xi} & \cdots & y_{k+\xi+\eta-2} \end{bmatrix} \tag{26}$$

Upon substituting $y_k = \mathbf{c}^T\mathbf{A}^k\mathbf{q}_o$ into Eq. (26), it can be shown that $\mathbf{H}(k)$ is the product of three matrices

$$\mathbf{H}(k) = \mathcal{P}_\xi \mathbf{A}^k \mathcal{Q}_\eta \tag{27}$$

where

$$\mathcal{P}_\xi = \begin{Bmatrix} \mathbf{c}^T \\ \mathbf{c}^T\mathbf{A} \\ \mathbf{c}^T\mathbf{A}^2 \\ \cdots \\ \mathbf{c}^T\mathbf{A}^{\xi-1} \end{Bmatrix} \in \mathbb{R}^{\xi\times p} \tag{28}$$

and

$$\mathcal{Q}_\eta = [\mathbf{q}_o \;\; \mathbf{A}\mathbf{q}_o \;\; \mathbf{A}^2\mathbf{q}_o \;\; \ldots \;\; \mathbf{A}^{\eta-1}\mathbf{q}_o] \in \mathbb{R}^{p\times\eta} \tag{29}$$

The proof of Eq. (27) could be easily done by carrying out the matrix multiplication of $\mathcal{P}_\xi$, $\mathbf{A}^k$ and $\mathcal{Q}_\eta$.

Furthermore, introducing weighting matrices $\mathbf{W}_1 \in \mathbb{R}^{\xi\times\xi}$ and $\mathbf{W}_2 \in \mathbb{R}^{\eta\times\eta}$ and defining

$$\tilde{\mathbf{H}}(k) = \mathbf{W}_1\mathbf{H}(k)\mathbf{W}_2 \tag{30}$$

one writes

$$\tilde{\mathbf{H}}(k) = \tilde{\mathcal{P}}_\xi \mathbf{A}^k \tilde{\mathcal{Q}}_\eta \tag{31}$$

where $\tilde{\mathcal{P}}_\xi = \mathbf{W}_1\mathcal{P}_\xi$ and $\tilde{\mathcal{Q}}_\eta = \mathcal{Q}_\eta\mathbf{W}_2$. The main purpose to extend from $\mathbf{H}(k)$ to $\tilde{\mathbf{H}}(k)$ is to mathematically show that the proposed algorithm is also applicable to an intermittent signal, that is, some part of a signal is missing. Specifically, when all the elements of a particular row of $\mathbf{W}_1$ are zero, pre-multiplying $\mathbf{W}_1$ amounts to deleting the corresponding row from $\mathbf{H}(k)$. Similarly, setting all the elements of a particular column of $\mathbf{W}_2$ to zero achieves the deletion of the corresponding column from $\mathbf{H}(k)$.

A realization of $\mathbf{A}$ can be obtained from $\tilde{\mathbf{H}}(0)$ and $\tilde{\mathbf{H}}(1)$. For simplicity, $\mathbf{H}(0)$ and $\mathbf{H}(1)$, instead of $\tilde{\mathbf{H}}(0)$ and $\tilde{\mathbf{H}}(1)$, are used in the following derivation. Substituting $k=0$ and $k=1$ into Eq. (31) yields

$$\mathbf{H}(0) = \mathcal{P}_\xi \mathcal{Q}_\eta \tag{32}$$

and

$$\mathbf{H}(1) = \mathcal{P}_\xi \mathbf{A} \mathcal{Q}_\eta \tag{33}$$

Applying the singular value decomposition of $\mathbf{H}(0)$, one writes [23]

$$\mathbf{H}(0)=[\mathbf{U}_1\ \ \mathbf{U}_2]\begin{bmatrix}\mathbf{S}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{V}_1^T\\ \mathbf{V}_2^T\end{bmatrix}=\mathbf{U}_1\mathbf{S}_1\mathbf{V}_1^T \tag{34}$$

In theory, the model order of the dynamic system, $p$, is equal to the number of non-zero singular values in Eq. (34), that is, $\mathbf{U}_1\in\mathbb{R}^{\xi\times p}$, $\mathbf{S}_1\in\mathbb{R}^{p\times p}$ and $\mathbf{V}_1\in\mathbb{R}^{p\times\eta}$, and the rank of $\mathbf{H}(0)$ is $p$. The singular values should go to zero when the rank of the matrix is exceeded, but for measured data due to random errors and small inconsistencies in the data, the singular values will not become zero but will become very small. While choosing the size of $\mathbf{H}(0)$, both $\xi$ and $\eta$ must be greater than $p$.

A direct comparison between Eqs. (32) and (34) suggests that $\mathcal{P}_\xi$ is related to $\mathbf{U}_1$, and $\mathcal{Q}_\eta$ to $\mathbf{V}_1^T$. A possible "balanced" choice is $\mathcal{P}_\xi=\mathbf{U}_1\mathbf{S}_1^{1/2}$ and $\mathcal{Q}_\eta=\mathbf{S}_1^{1/2}\mathbf{V}_1^T$. Substituting $\mathcal{P}_\xi=\mathbf{U}_1\mathbf{S}_1^{1/2}$ and $\mathcal{Q}_\eta=\mathbf{S}_1^{1/2}\mathbf{V}_1^T$ into Eq. (33), one obtains

$$\mathbf{H}(1)=\mathbf{U}_1\mathbf{S}_1^{1/2}\mathbf{A}\mathbf{S}_1^{1/2}\mathbf{V}_1^T \tag{35}$$

Premultipling $\mathbf{S}_1^{-1/2}\mathbf{U}_1^T$ and postmultiplying $\mathbf{V}_1\mathbf{S}_1^{-1/2}$ at Eq. (35) yields a realization of $\mathbf{A}$ as

$$\mathbf{A}=\mathbf{S}_1^{-1/2}\mathbf{U}_1^T\mathbf{H}(1)\mathbf{V}_1\mathbf{S}_1^{-1/2} \tag{36}$$

While the computed eigenvalues of $\mathbf{A}$ are $z_n$, $n=1,\ldots,p$, the corresponding $\lambda_n$ in Eq. (2) are $\lambda_n=\ln(z_n)/\Delta t$. Finally, the complex coefficients $\gamma_n$, $n=1,\ldots,p$ can be computed in the same way as that used in the third step of Prony's method (see Eq. (13)), but the original signal $y_k$ could have been replaced by a filtered signal $\widehat{y}_k$.

The filtered signal $\widehat{y}_k$ could be obtained by applying a *structured low rank approximation* (SLRA) technique on a given Hankel matrix $\mathbf{H}$. The optimal $L_2$-norm lower rank approximation to a given Hankel matrix $\mathbf{H}$ can be obtained from the truncated singular value decomposition (TSVD) of $\mathbf{H}$, but the resulting matrix will not possess a Hankel structure. For retrieving a filtered signal, the resulting matrix must maintain a Hankel structure. A simple engineering algorithm for the restoration of a Hankel matrix is the Cadzow's algorithm [26], which has been also called the anti-diagonal averaging method [27].

### 3.4. Conditioning: proposed method versus Prony's method

From equally spaced sampling signal data, both Prony's method and the proposed method are for estimating $\lambda_n$ and $\gamma_n$ in Eq. (2). Although they begin with the same continuous and discrete signal formulations, i.e., Eqs. (2) and (4), they proceed differently in mathematical model, as well as on the procedure how to calculate the exponents $\lambda_n$. Based on a $p$-th order difference equation, Prony's method calculates the polynomial coefficients $\mathbf{b}$ of the pertaining characteristic equation in step 1, and then computes the roots of the polynomial in step 2. In contrast, the proposed method using a state-space equation obtains a realization of state matrix $\mathbf{A}$ in step 1, and then computes the eigenvalues of $\mathbf{A}$ in step 2. Although these two methods, while solving for $\lambda_n$, seem identical mathematically, they are very different over the fundamental issue of numerical analysis: conditioning. Conditioning pertains to the perturbation behavior of a mathematical problem itself; it should not be confused with the numerical stability which pertains to the perturbation behavior of an algorithm used to solve that problem on a computer [14].

Conditioning of a mathematical problem often is quantified by a condition number, and the condition number is a property of the problem. For example, the condition number associated with the linear equation $\mathbf{D}\mathbf{x}=\mathbf{d}$ gives a bound on how inaccurate the solution $\mathbf{x}$ will be after approximate solution. Conditioning is a property of the matrix, not the algorithm or floating point accuracy of the computer used to solve the corresponding system. In the field of numerical linear algebra, the condition number of matrix $\mathbf{D}$ is defined as

$$\kappa(\mathbf{D})=\frac{\sigma_{\max}(\mathbf{D})}{\sigma_{\min}(\mathbf{D})} \tag{37}$$

where $\sigma_{\max}(\mathbf{D})$ and $\sigma_{\min}(\mathbf{D})$ are maximal and minimal singular values of $\mathbf{D}$ respectively. If the condition number of $\mathbf{D}$ is large, even a small error in $\mathbf{d}$ may cause a large error in $\mathbf{x}$.

For the proposed and Prony's methods, numerical conditioning related to Step 1 and Step 2 is addressed below:

*Step 1:* While Prony's method requires the computation of the pseudo-inverse of $\mathbf{Y}$ to obtain $\mathbf{b}$ (see Eq. (9)), the proposed method must use the truncated singular value decomposition of $\mathbf{H}(0)$ to get a realization of state-matrix $\mathbf{A}$. The conditions of the pertaining problems are thus characterized, respectively, by the condition number of matrix $\mathbf{Y}$ and that of the matrix $\widehat{\mathbf{H}}(0)=\mathbf{U}_1\mathbf{S}_1\mathbf{V}_1^T$, where $\widehat{\mathbf{H}}(0)$ is the rank-$p$ TSVD of $\mathbf{H}(0)$. If the singular values are ranked from the largest to smallest, i.e. $\sigma_1\geq\sigma_2\geq\ldots\geq\sigma_p$, then the condition number for both methods is $\sigma_1/\sigma_p$. Note that the formulation of $\mathbf{Y}\in\mathbb{R}^{n\times p}$ is unique, but there are many options for the formulation of $\mathbf{H}(0)\in\mathbb{R}^{\xi\times\eta}$, depending on the choice of $\xi$ and $\eta$. The choice of $\xi$ and $\eta$ is fairly arbitrary provided extremely rectangular Hankel matrices are avoided [19]. If the computational time is not a major concern, a better choice is a near square Hankel matrix, i.e., $\xi$ and $\eta$ are close to each other [28].

*Step 2:* The conditioning difference between the two methods in this step basically is computing eigenvalues in the proposed method against finding polynomial roots in Prony's method. It is a well established knowledge that the

eigenvalue problem is well-conditioned, but the problem of finding the roots from the polynomial coefficients is in general ill-conditioned. Because the eigenvalues of the companion matrix of the polynomial coincide with the roots of the polynomial, one can use any eigenvalue algorithm to find the roots of the polynomial. But finding polynomial roots from the companion matrix of the polynomial would not change the condition of the original polynomial root-finding problem. On the other hand, it is a bad strategy to compute eigenvalues of a matrix by first calculating the coefficients of the matrix's characteristic polynomial and then finding its roots, since using the coefficients as an intermediate step may introduce an extreme ill-conditioning even if the underlying eigenvalue problem is well-conditioned [14,23]. One of the most popular methods today to compute eigenvalues and eigenvectors of arbitrary matrices is the QR algorithm.

## 4. Numerical examples

Four numerical examples are given in this paper. The first example presents the detailed numerical operation of the proposed method by using only a few equally spaced data points. It also demonstrates the minimum number of data points needed for extracting a certain number of exponential components. The second example considers a periodic signal which is composed of multiple harmonic components. It shows the difference over the harmonic decomposition of a periodic signal between the proposed method and the discrete Fourier transform (DFT). The third example is to evaluate the proposed method against Prony's method. The evaluation includes the influence on numerical conditioning and stability due to sampling rate, round-off error and random noise. Finally, the applicability of the proposed method for an intermittent signal, i.e. a signal with missing segments, is demonstrated in the last example.

### 4.1. Proposed method: detailed numerical illustration

In this example, a signal is synthesized by using the following formula:

$$y_k = \sum_{n=1}^{3} A_n \exp(-\alpha_n t_k)\cos(2\pi f_n t_k + \theta_n) \tag{38}$$

where the parameters in the above equation are listed in Table 1. With $\Delta t = 0.5$ s, the first 10 data points computed from Eq. (38) are: 1.7876, 1.7738, 0.6631, −0.8089, −1.6255, −1.2520, −0.0137, 1.1881, 1.5743 and 1.0392; and the corresponding plot is shown in Fig. 1. Using this simple example allows the detailed numerical results being presented, also provides the reader an opportunity to verify the correctness of the proposed method independently.

**Table 1**
The parameters used in Eq. (38).

| $n$ | $f_n$ | $\alpha_n$ | $A_n$ | $\theta_n$ |
|---|---|---|---|---|
| 1 | 0 | −0.003 | 0.20 | 0 |
| 2 | 0.2 | 0.03 | 0.80 | $\pi/8$ |
| 3 | 0.3 | 0.04 | 1.20 | $-\pi/4$ |

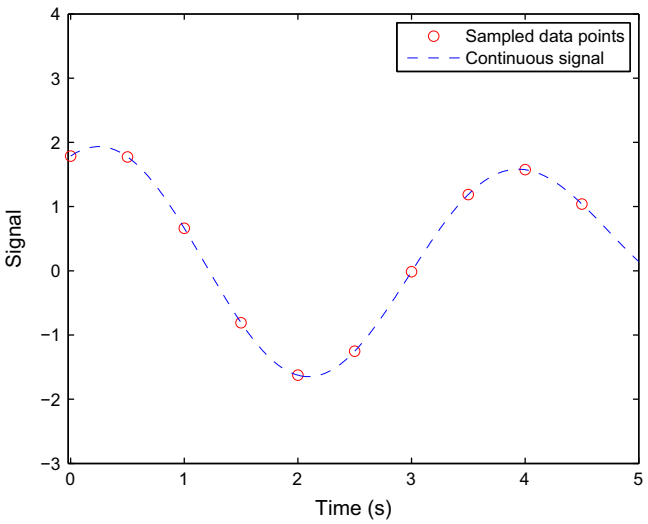


**Fig. 1.** First 10 data points computed from Eq. (38).

Using the 10 sampled data listed above, from Eq. (26), one formulates

$$\mathbf{H}(0)=\begin{bmatrix} 1.7876 & 1.7738 & 0.6631 & -0.8089 & -1.6255 \\ 1.7738 & 0.6631 & -0.8089 & -1.6255 & -1.2520 \\ 0.6631 & -0.8089 & -1.6255 & -1.2520 & -0.0137 \\ -0.8089 & -1.6255 & -1.2520 & -0.0137 & 1.1881 \\ -1.6255 & -1.2520 & -0.0137 & 1.1881 & 1.5743 \end{bmatrix}$$

and

$$\mathbf{H}(1)=\begin{bmatrix} 1.7738 & 0.6631 & -0.8089 & -1.6255 & -1.2520 \\ 0.6631 & -0.8089 & -1.6255 & -1.2520 & -0.0137 \\ -0.8089 & -1.6255 & -1.2520 & -0.0137 & 1.1881 \\ -1.6255 & -1.2520 & -0.0137 & 1.1881 & 1.5743 \\ -1.2520 & -0.0137 & 1.1881 & 1.5743 & 1.0392 \end{bmatrix}$$

Taking the singular value decomposition $\mathbf{H}(0)=\mathbf{USV}^T$ yields

$$\mathbf{U}=\begin{bmatrix} 0.5846 & -0.1972 & 0.7046 & -0.3291 & -0.1209 \\ 0.4830 & 0.4323 & 0.0633 & 0.5379 & 0.5352 \\ 0.0673 & 0.7249 & 0.0446 & 0.0317 & -0.6833 \\ -0.3701 & 0.4930 & 0.2609 & -0.5694 & 0.4772 \\ -0.5324 & -0.0754 & 0.6553 & 0.5264 & -0.0652 \end{bmatrix} \quad \mathbf{S}=\begin{bmatrix} 5.3224 & 0 & 0 & 0 & 0 \\ 0 & 3.1381 & 0 & 0 & 0 \\ 0 & 0 & 0.1777 & 0 & 0 \\ 0 & 0 & 0 & 0.0255 & 0 \\ 0 & 0 & 0 & 0 & 0.0016 \end{bmatrix}$$

and

$$\mathbf{V}=\begin{bmatrix} 0.5846 & 0.1972 & 0.7046 & -0.3291 & 0.1209 \\ 0.4830 & -0.4323 & 0.0633 & 0.5379 & -0.5352 \\ 0.0673 & -0.7249 & 0.0446 & 0.0317 & 0.6833 \\ -0.3701 & -0.4930 & 0.2609 & -0.5694 & -0.4772 \\ -0.5324 & 0.0754 & 0.6553 & 0.5264 & 0.0652 \end{bmatrix}$$

From Eq. (36), a realization of $\mathbf{A}$ is then computed to be

$$\mathbf{A}=\mathbf{S}^{-1/2}\mathbf{U}^T\mathbf{H}(1)\mathbf{V}\mathbf{S}^{-1/2}=\begin{bmatrix} 0.6665 & 0.7443 & -0.1737 & -0.1278 & 0.0909 \\ -0.7443 & 0.5916 & 0.0971 & -0.0469 & 0.0147 \\ -0.1737 & -0.0971 & 1.1951 & 1.0924 & -0.5756 \\ -0.1278 & 0.0469 & 1.0924 & 1.4446 & -1.4128 \\ -0.0909 & 0.0147 & 0.5756 & 1.4128 & -0.1500 \end{bmatrix}$$

The eigenvalues of $\mathbf{A}$, denoted by $z_n$, $n=1,\ldots,5$, turn out to be two complex conjugate pairs and one real value: $0.5761\pm0.7930i$, $0.7970\pm0.5790i$, and 1.0015. From $\lambda_n=\ln(z_n)/\Delta t$, it yields $\lambda_n=-0.0400\pm1.8850i$, $-0.0300\pm1.2566i$, and 0.0030. According to the definition $\lambda_n\equiv-\alpha_n+i2\pi f_n$, one then obtains three sets of $(f_n,\alpha_n)$ to be $(0.3,0.04)$, $(0.2,0.03)$ and $(0,-0.003)$. Next, using Eq. (13), one computes $\gamma_n$ to be $0.4243\pm0.4243i$, $0.3696\pm0.1531i$, and 0.2000. Following the definition $\gamma_n\equiv A_n\exp(i\theta_n)$, one obtains $(A_n,\theta_n)$ to be $(1.20,-0.25\pi)$, $(0.80,0.125\pi)$ and $(0.20,0)$.

The above estimated values are identical to the target values listed in Table 1. Thus, the two damped harmonic and one exponential components have been successfully extracted by using just 10 sampled data points. From the data fitting point of view, the current example is to estimate the 10 unknown parameters associated with the synthesized signal. Mathematically, at least 10 equations are required to estimate 10 unknowns. Because the signal value at each discrete time represents an equation, 10 equations are available.

### 4.2. Proposed method versus DFT

When a signal is composed of damped harmonic components, employing DFT for the signal decomposition certainly will not get the target components. For intentionally putting the DFT analysis in a favorable situation, a periodic signal is simulated in the present example. The signal to be investigated is composed of 5 harmonic components

$$y_k=\sum_{n=1}^{5}A_n\cos(2\pi f_n t_k+\theta_n),\quad k=0,1,\ldots,N-1 \tag{39}$$

The frequency $f_n$, amplitude $A_n$ and phase angle $\theta_n$ of each component are listed in Table 2. Because the five component frequencies $f_n$ (see Table 2) are all an integer times $\Delta f=0.02$ Hz, the simulated signal has period 50 s.

With sampling interval $\Delta t = 0.05$ s and $N = 1024$ data points, the total duration of the signal is 51.2 s. In order to apply the proposed method, first one needs to determine the model order associated with the signal, which is related to the rank of $\mathbf{H}(0)_{\xi\times\eta}$, or $\mathbf{H}(1)_{\xi\times\eta}$. Estimating the rank of a matrix could be done if the singular values of the matrix have been ordered sequentially from the largest to the smallest. A conventional way to choose the noise threshold is to find a significant gap of the normalized singular values. In this example, $\xi = 512$ and $\eta = 512$ are chosen. Displayed in Fig. 2 is the diagram for the first 100 normalized singular values of $\mathbf{H}(0)_{512\times512}$. From this diagram, it is clear that the rank of $\mathbf{H}(0)_{512\times512}$ is 10, which is in agreement with that there are 5 harmonic components in the simulated signal.

Referring to Eq. (34) and taking $p = 10$, one obtains $\mathbf{U}_1 \in \mathbb{R}^{512\times10}$, $\mathbf{S}_1 \in \mathbb{R}^{10\times10}$ and $\mathbf{V}_1 \in \mathbb{R}^{10\times512}$ from the SVD of $\mathbf{H}(0)_{512\times512}$, and then utilize Eq. (36) to compute $\mathbf{A}$. Next, one computes $z_n$, the eigenvalues of $\mathbf{A}$, and $\lambda_n = \ln(z_n)/\Delta t$. Finally, Eq. (13) is employed to compute $\gamma_n$. The estimated $f_n$, $\alpha_n$, $A_n$ and $\theta_n$ from $\lambda_n$ and $\gamma_n$ turn out to be exactly the same as the target values listed in Table 2.

The comparison of the estimated amplitude–frequency from DFT and that of the proposed method is shown in Fig. 3. In the DFT analysis, the frequency resolution is $\Delta f = 1/51.2$ Hz, and the signal is decomposed into 1024 components with the frequencies $f_n = n\Delta f$, $n = 0, 1, \ldots, 1023$. Clearly, the signal decomposition from DFT is not only limited by the frequency resolution, but also suffers a server leakage problem. It is particularly difficult to identify the three close target frequencies from the DFT plot: 2.00 Hz, 2.02 Hz and 2.04 Hz.

Although both DFT analysis and the newly proposed method can reconstruct the original 51.2-s signal from superposing their components, they predict completely different signals after 51.2 s. The periodic assumption of the DFT suggests the signal will repeat itself for every 51.2 s, but the signal actually has period 50 s. In contrast, the new method extends the signal based on the superposition of the estimated 5 components. Shown in Fig. 4 is the comparison of the two signals for the time between 49 and 54 s.

### 4.3. Proposed method versus Prony's method

Prony's method has been known to be sensitive to the signal sampling rate [11,12], as well as to the round-off error and random noise [6–9]. The focus of this example is on investigating these issues associated with the proposed method, comparing against those of Prony's method. Two tasks are carried out below for both methods: (1) numerical conditioning and stability associated with signal sampling rate; and (2) the effects on estimating parameters due to the round-off error and random noise of the signal.

**Table 2**
Component parameters used in Eq. (39).

| $n$ | $f_n$ | $A_n$ | $\theta_n$ |
|---|---|---|---|
| 1 | 2.00 | 1.60 | $\pi/4$ |
| 2 | 2.02 | 2.00 | $-\pi/8$ |
| 3 | 2.04 | 3.00 | $-3\pi/4$ |
| 4 | 2.40 | 1.40 | $\pi/2$ |
| 5 | 3.00 | 3.60 | $\pi/8$ |

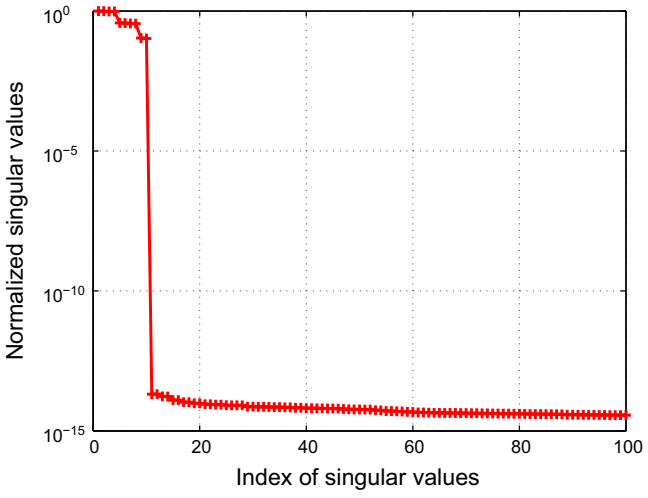


**Fig. 2.** Normalized singular values of $\mathbf{H}(0)_{512\times512}$.

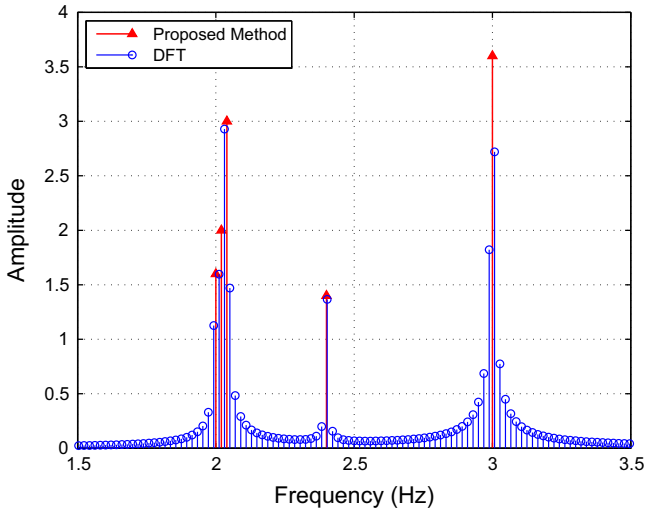


**Fig. 3.** Comparison of the result from the DFT and the proposed method.

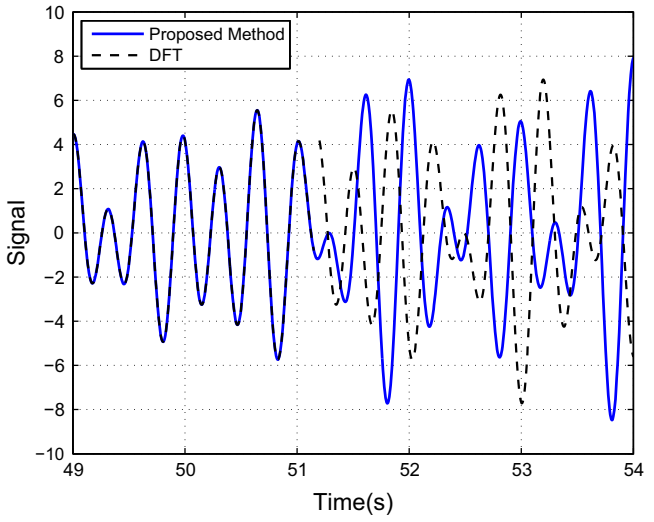


**Fig. 4.** Comparison of the two signals obtained from the DFT analysis and the proposed method.

The signal to be analyzed in this example consists of 4 components, with 3 damped sinusoidals and 1 undamped sinusoidal

$$y_k = \sum_{n=1}^{4} A_n \exp(-\alpha_n t_k) \cos(2\pi f_n t_k + \theta_n) \tag{40}$$

where the values for $A_n$, $\alpha_n$, $f_n$ and $\theta_n$ are given in Table 3; $t_k = k\Delta t$, $\Delta t$ is the sampling interval and $k = 0, 1, \ldots, 1023$.

*4.3.1. Sampling rate*

This section investigates the numerical conditioning and stability against the sampling interval, $\Delta t$. Because the maximum frequency of the signal is 3.2 Hz (see Table 3), the sampling frequency must be greater than 6.4 Hz, or the sampling interval less than 0.15625 s, to avoid aliasing problems. With sampling interval ranged from 0.01 s to 0.15 s, the condition number of the coefficient matrix $\mathbf{Y}$ (see Eq. (10)) versus sampling rate is plotted in Fig. 5. It reveals that $\kappa(\mathbf{Y})$ varies from greater than $10^9$ for $\Delta t = 0.01$ s to less than $10^2$ for $\Delta t = 0.1$ s. Qualitatively, the data exhibited in Fig. 5 — an optimal sampling rate is one that is neither too fast nor too slow — agrees well with that reported previously [12]. When the sampling interval is small, all values in each row of $\mathbf{Y}$ become very close to each other, and thus an ill-conditioned case, i.e. a large $\kappa(\mathbf{Y})$, occurs. The estimated frequencies and damping factors from Prony's method for different sampling intervals are shown in Fig. 6. When the sampling interval is small, implementing Prony's method results in poor estimation for component frequencies and damping factors.

**Table 3**
Values of the parameters in Eq. (40).

| $n$ | $f_n$ | $\alpha_n$ | $A_n$ | $\theta_n$ |
|---|---|---|---|---|
| 1 | 1.8 | 0.02 | 2.2 | $\pi/6$ |
| 2 | 2.2 | 0 | 1.0 | $\pi/2$ |
| 3 | 3.0 | 0.01 | 1.4 | $-\pi/4$ |
| 4 | 3.2 | 0.04 | 2.6 | $3\pi/8$ |

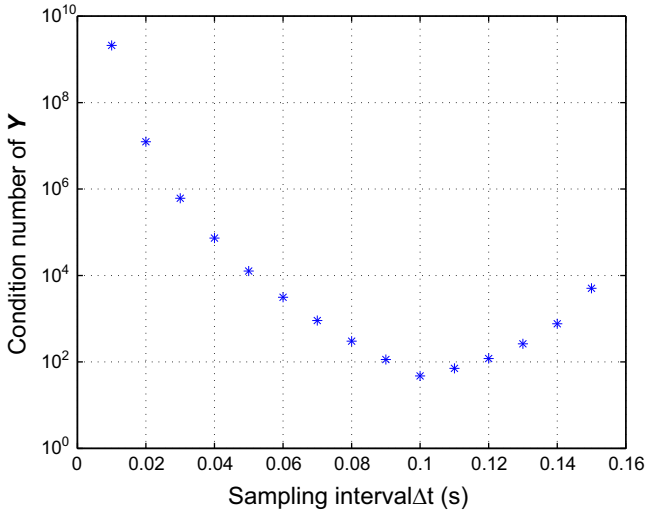


**Fig. 5.** Condition number of $\mathbf{Y}\in\mathbb{R}^{1016\times 8}$ with different sampling intervals.

For the proposed method, the condition number of $\widehat{\mathbf{H}}(0)_{512\times 512}$ versus sampling interval is shown in Fig. 7. Clearly, $\kappa(\widehat{\mathbf{H}}(0)_{512\times 512})$ is much smaller than the corresponding $\kappa(\mathbf{Y})$. The well-conditioned $\widehat{\mathbf{H}}(0)_{512\times 512}$ can be attributed to two factors: (1) spanning longer time for each row of $\widehat{\mathbf{H}}(0)_{512\times 512}$, and (2) using TSVD that only keeps the highest 8 singular values. Shown in Fig. 8 are the estimated frequencies and damping factors by the proposed method. In brief, the proposed method yields good estimation even when the sampling interval is small.

The results shown in Figs. 7 and 8 have been based on the specific Hankel matrix $\mathbf{H}(0)_{\xi\times\eta}$ with the size $512\times 512$. In the following,different sizes of $\mathbf{H}(0)_{\xi\times\eta}$ formed from the signal with $\Delta t=0.01$ s are considered. Without losing generality, one assumes $\xi\geq\eta$. Because $\xi+\eta=1024$ and the rank of $\mathbf{H}(0)$ must be at least 8,the value of $\eta$ can range from 8 to 512. Taking the values of $\eta$ equal to $2^{\ell}$,where $\ell$ is an integer from 3 to 8,the condition number of $\widehat{\mathbf{H}}(0)_{\xi\times\eta}$ against $\eta$ is plotted in Fig. 9, which shows that the condition number of $\widehat{\mathbf{H}}(0)_{\xi\times\eta}$ decreases with increasing value of $\eta$. Note that when $\eta=8$,the corresponding Hankel matrix $\widehat{\mathbf{H}}(0)_{\xi\times\eta}$ becomes $\mathbf{Y}$, that is, $\widehat{\mathbf{H}}(0)_{1016\times 8}=\mathbf{Y}$. Because using $\mathbf{Y}$ in Prony's method yields very poor estimation for both frequency and damping factor when $\Delta t=0.01$ (see Fig. 6),one would expect to obtain poor frequency and damping estimations if $\widehat{\mathbf{H}}(0)_{1016\times 8}$ was utilized in the proposed method. However, the estimated frequencies and damping factors by using the proposed method are surprisingly good even when $\widehat{\mathbf{H}}(0)_{1016\times 8}$ is utilized (see Fig. 10). This result suggests that the numerical procedure of the proposed method is superior to that of Prony's method.

### 4.3.2. Round-off error and signal noise

For the remaining part of this example, the signal utilized to evaluate the effects of round-off error and random noise is based on $\Delta t=0.05$ s. From Figs. 5 and 7, at $\Delta t=0.05$, the condition numbers of $\mathbf{Y}$ and $\widehat{\mathbf{H}}(0)_{512\times 512}$ are about $10^4$ and 1.4, respectively. Round-off errors of a measured signal are inevitable because it is always recorded with limited significant digits. With the precision of the measured signal being limited to 5 significant digits, the frequencies and damping factors by the proposed method and Prony's method are estimated. While those estimations obtained from the proposed method are exact up to 5 significant digits, those obtained from Prony's method are with significant error, especially estimation of the damping factors (see Table 4). In view of the condition number of $\mathbf{Y}$ equal to $10^4$, the erroneous estimation from Prony's method is not surprising.

For investigating the effects of random noise to the proposed method, two noisy signals are simulated by adding 5% and 20% random Gaussian white noise, respectively, to the clean signal. The level of the additive white noise is quantified by a stated percentage, which is defined as the ratio of the standard deviation of the white noise ($\sigma_{\text{noise}}$) to that of the noise-free signal ($\sigma_{\text{signal}}$): $\sigma_{\text{noise}}/\sigma_{\text{signal}}\times 100\%$. When the signal-to-noise, S/N, ratio is defined as $\text{S/N}=20\log_{10}(\sigma_{\text{signal}}/\sigma_{\text{noise}})$, 5% and 20% noisy signals are corresponding to S/N=26.02 dB and 13.98 dB, respectively. Similar to the way to obtain Fig. 2, shown

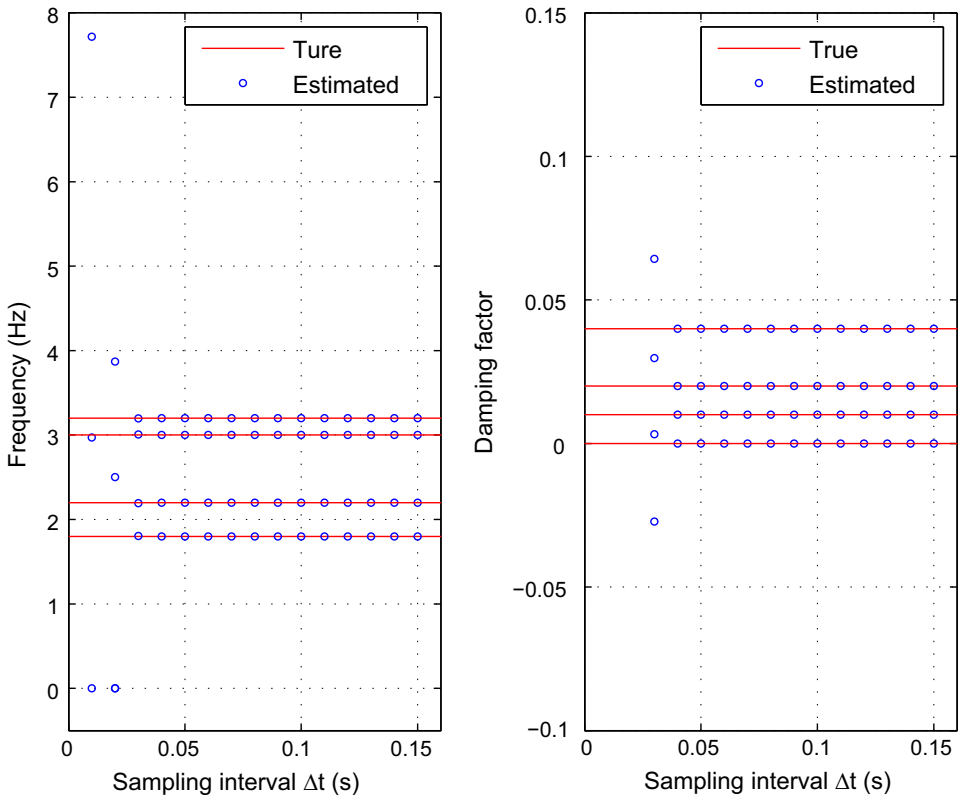


**Fig. 6.** Estimated frequencies and damping factors from Prony's method.

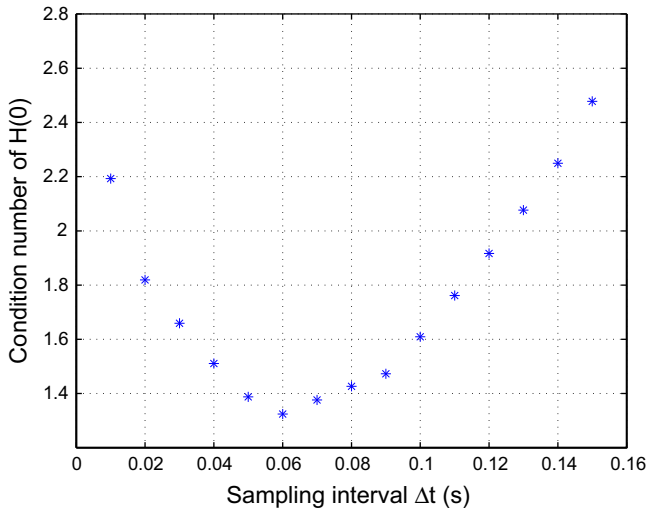


**Fig. 7.** Condition number of $\hat{\mathbf{H}}(0)$ with different sampling intervals.

in Fig. 11 are the normalized singular values of $\mathbf{H}(0)_{512\times512}$ for the two noisy signals, together with those for the clean signal. Clearly, all three singular value curves in Fig. 11 have a significant gap between their 8th and 9th singular values. As the gap represents the separation between true signal and noise, the model order for those signals is set equal to 8. Implementing the proposed method with model order 8, one obtains the estimated frequencies and damping factors for the clean, 5%-noise and 20%-noise signals in Table 5, and amplitudes and phase angles in Table 6. As expected, the estimated component parameters of the 20%-noise signal are less accurate than those of the 5%-noise signal. However, they are still close to those of the clean signal. In short, the proposed method is very robust on estimating component parameters, and insensitive to the random noise embedded in the signal. In contrast, when Prony's method was conducted for noisy signals,

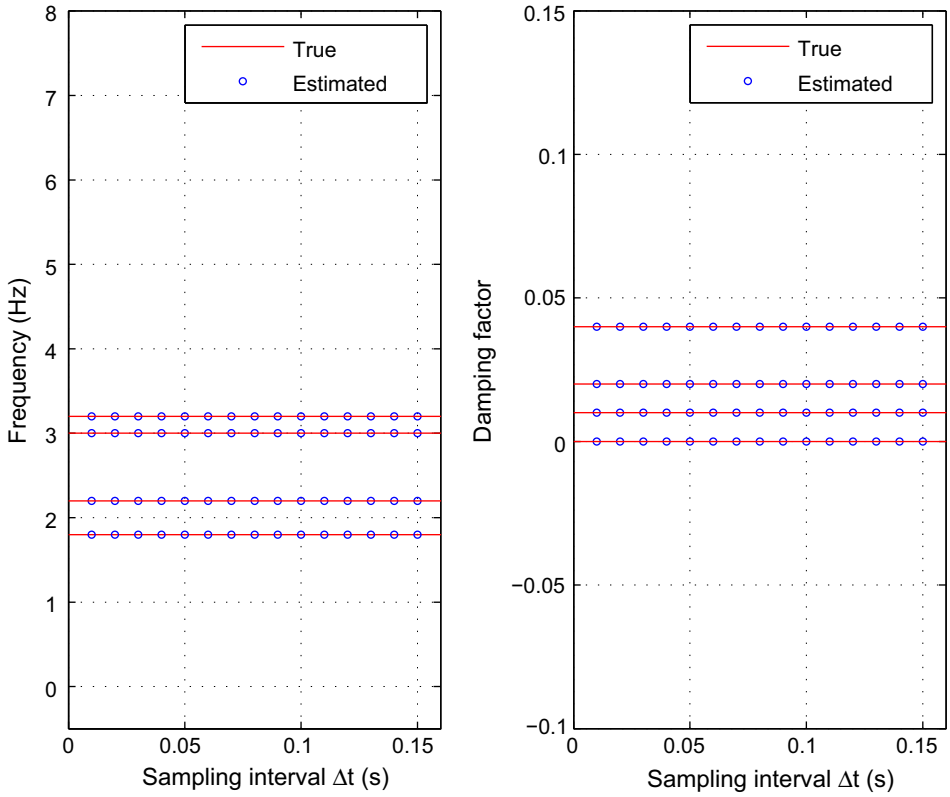


**Fig. 8.** Estimated frequencies and damping factors from the proposed method.

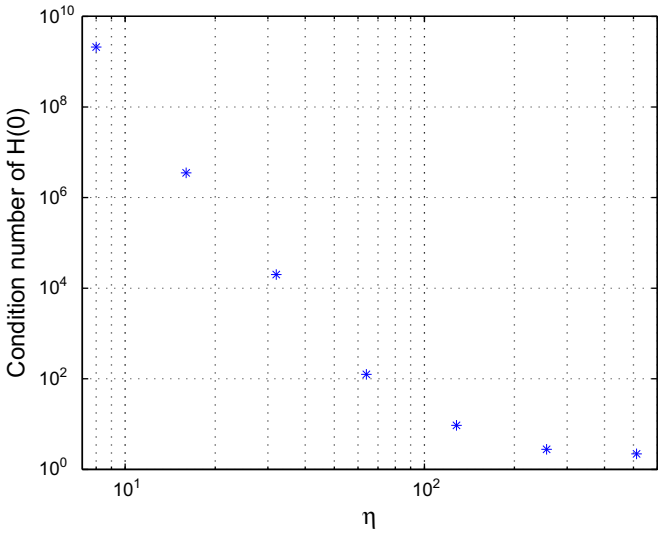


**Fig. 9.** Condition number of $\widehat{\mathbf{H}}(0)$ against $\eta$.

the estimated frequencies and damping factors were too far away from, and could not even be aligned with, the target values (see Tables 7 and 8). For the 5%-noise case, the obtained roots (poles) are three pairs of complex conjugates and two real values, and the estimated frequencies from the three pairs of complex conjugates are 1.8985, 3.1162, and 7.4259 respectively. As suggested by previous studies, a noise removal algorithm, such as SLRA technique, must be performed prior to implementing Prony's method [10].

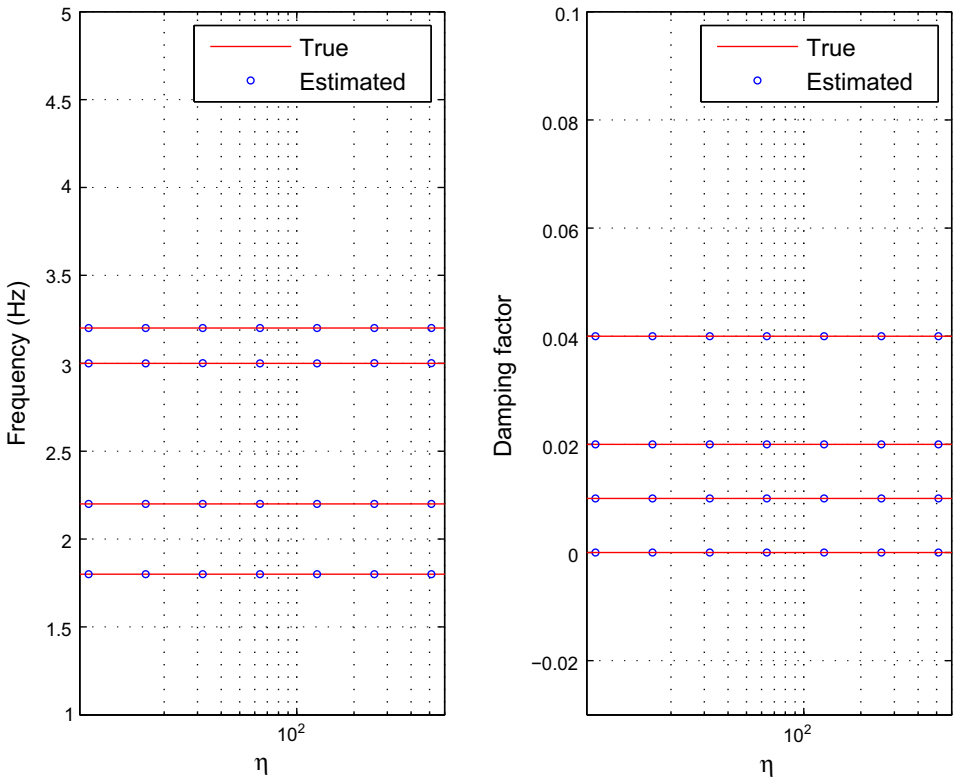


**Fig. 10.** Estimated frequencies and damping factors from the proposed method for $\Delta t = 0.01$ s.

**Table 4**
Estimated frequencies and damping factors from the signal with round-off error.

| $n$ | Frequency ($f_n$) | | Damping factor ($\alpha_n$) | |
|---|---|---|---|---|
| | Proposed | Prony's | Proposed | Prony's |
| 1 | 1.8000 | 1.8005 | 0.0200 | 0.0288 |
| 2 | 2.2000 | 2.2027 | 0.0000 | 0.0451 |
| 3 | 3.0000 | 3.0035 | 0.0100 | 0.0481 |
| 4 | 3.2000 | 3.2009 | 0.0400 | 0.0520 |

### 4.4. Proposed method on intermittent signal

Because of the interruption of strong ambient noise or sensor malfunction, measured signals could be intermittent. Shown in Fig. 12 is the intermittent signal to be tested in this example. This signal has been generated from using Eq. (40) with parameters in Table 3, $\Delta t = 0.05$ and $N = 1024$, but two segments of the signal are missing: from 5 s to 10 s, and from 30 s to 37.5 s. Implementing the proposed method to this intermittent signal uses $\tilde{\mathbf{H}}(0)_{\xi\times\eta}$ and $\tilde{\mathbf{H}}(1)_{\xi\times\eta}$ (see Eq. (30)) which are equivalent to deleting those rows or columns of $\mathbf{H}(0)_{\xi\times\eta}$ and $\mathbf{H}(1)_{\xi\times\eta}$ that correspond to the two missing segments.

By choosing $\xi = 200$ and $\eta = 824$ to construct $\mathbf{H}(0)_{200\times824}$ and $\mathbf{H}(1)_{200\times824}$ and then deleting the columns of them that involve missing data, the corresponding $\tilde{\mathbf{H}}(0)_{200\times273}$ and $\tilde{\mathbf{H}}(1)_{200\times273}$ are obtained. From which the frequency, damping factor, amplitude, and phase angle of each component are estimated following the same procedure as before. The results turn out to be the same as those listed in Table 3. Once signal parameters are properly estimated, the missing segments of the intermittent signal can be reconstructed (see Fig. 12).

## 5. Concluding remarks

A purpose of implementing Prony's method is to decompose a signal into the sum of $p$ real and/or complex exponential components, where $p$ must be predetermined. Mathematically, this sum is also the homogenous solution of a $p$th-order linear ordinary differential equation. Thus, given equally spaced discrete time signal data, the derivation of Prony's method has been based on a $p$th-order homogeneous difference equation. In this paper, an alternative method that uses a first-order

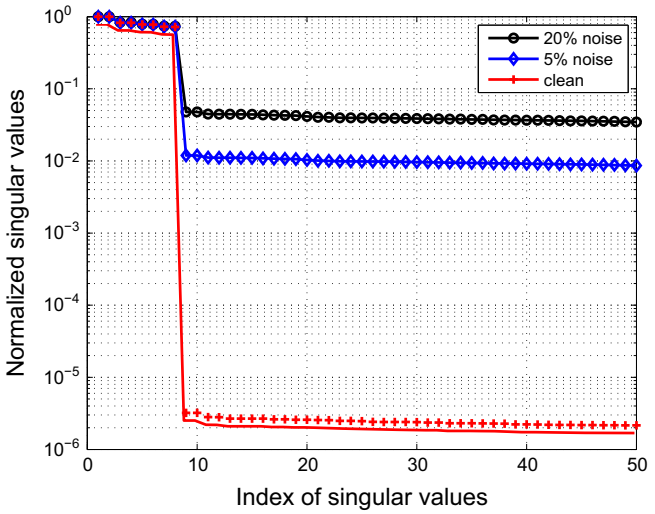


**Fig. 11.** Normalized singular values for signals with different levels of noise.

**Table 5**
Estimation of the frequencies and damping factors by the proposed method.

| $n$ | Frequency ($f_n$) | | | Damping factor ($\alpha_n$) | | |
|---|---|---|---|---|---|---|
| | Clean | 5% noise | 20% noise | Clean | 5% noise | 20% noise |
| 1 | 1.8000 | 1.8000 | 1.8000 | 0.0200 | 0.0199 | 0.0196 |
| 2 | 2.2000 | 2.1999 | 2.1997 | 0.0000 | −0.0003 | −0.0011 |
| 3 | 3.0000 | 3.0000 | 3.0000 | 0.0100 | 0.0100 | 0.0102 |
| 4 | 3.2000 | 3.2001 | 3.2004 | 0.0400 | 0.0400 | 0.0400 |

**Table 6**
Estimation of the amplitudes and phase angles by the proposed method.

| $n$ | Amplitude ($A_n$) | | | Phase angle ($\theta_n$) | | |
|---|---|---|---|---|---|---|
| | Clean | 5% noise | 20% noise | Clean | 5% noise | 20% noise |
| 1 | 2.2000 | 2.1964 | 2.1868 | $0.1667\pi$ | $0.1656\pi$ | $0.1625\pi$ |
| 2 | 1.0000 | 0.9949 | 0.9807 | $0.5000\pi$ | $0.5019\pi$ | $0.5081\pi$ |
| 3 | 1.4000 | 1.4005 | 1.4034 | $-0.2500\pi$ | $-0.2493\pi$ | $-0.2471\pi$ |
| 4 | 2.6000 | 2.6131 | 2.6532 | $0.3750\pi$ | $0.3714\pi$ | $0.3609\pi$ |

**Table 7**
Estimation of the frequencies and damping factors by Prony's method.

| $n$ | Frequency ($f_n$) | | Damping factor ($\alpha_n$) | |
|---|---|---|---|---|
| | 5% noise | 20% noise | 5% noise | 20% noise |
| 1 | 1.8985 | 1.8707 | 0.2226 | 0.4282 |
| 2 | 3.1162 | 3.1165 | 0.1680 | 0.4260 |
| 3 | 7.4259 | 6.4277 | 8.4767 | 5.3744 |
| 4 | – | 8.8485 | – | 6.2388 |

**Table 8**
Estimation of the amplitudes and phase angles by Prony's method.

| $n$ | Amplitude ($A_n$) | | Phase angle ($\theta_n$) | |
|---|---|---|---|---|
| | 5% noise | 20% noise | 5% noise | 20% noise |
| 1 | 1.0256 | 2.3556 | $-0.2609\pi$ | $-0.0576\pi$ |
| 2 | 1.4685 | 2.0817 | $0.8958\pi$ | $0.7359\pi$ |
| 3 | 4.4413 | 3.4221 | $-0.5279\pi$ | $-0.4299\pi$ |
| 4 | – | 2.7086 | – | $-0.1299\pi$ |

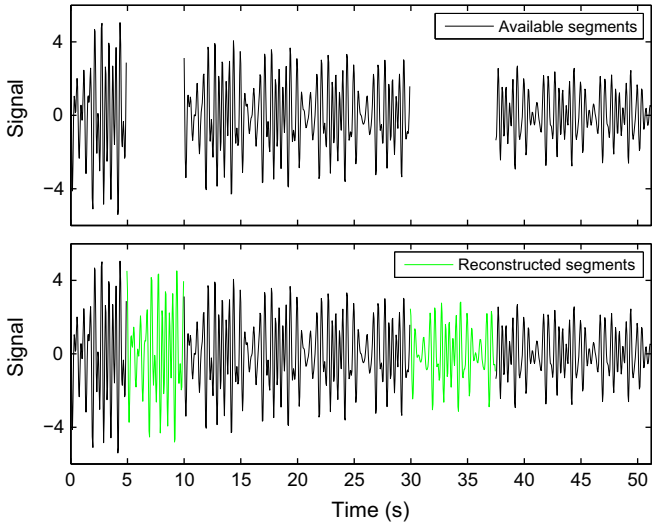


**Fig. 12.** An intermittent signal and its reconstruction.

matrix homogenous difference equation (state-space model), replacing the $p$th-order homogenous difference equation was proposed. Although the proposed method and Prony's method seem to be theoretically identical, they are dramatically different over several numerical issues, such as conditioning, stability and the accuracy on component parameter estimation. Through numerical examples, this paper confirmed the conventional wisdom that Prony's method is very sensitive to sampling rate, round-off error and random noise. This paper also demonstrated that the developed method is very robust to sampling rate and round-off error. Because the proposed method has a build-in noise rejection mechanism via the usage of truncated singular value decomposition, it can handle noisy signals properly as well. The proposed method is also applicable to intermittent signals, and can recover the missing parts of intermittent signals nicely through reconstruction. Another advantage of the proposed method is that it requires only a short duration signal to gain high frequency resolution in the decomposition process. This advantage would be particularly attractive for the time–frequency signal analysis [5].

## Acknowledgments

This research was financially supported by the National Natural Science Foundation of China (Grant no. 51079134), the NSFC Major International Joint Research Project (Grant no. 51010009) and the National Basic Research Program of China (973 Program) (Grant no. 2011CB013704). The second author was supported by China Scholarship Council while conducting his research in USA.

## References

[1] S. Mallat, A. Wavelet, Tour of Signal Processing, 2nd edition, Academic, San Diego, CA, 1999.
[2] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shin, Q. Zheng, N.-C. Yen, C.C. Tung, H.H. Liu, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, Proc. R. Soc. London Ser. A 454 (1998) 903–995.
[3] G.C.F.M.R. de Prony, Essai experimental et analytique: Sur les lois de la dilatabilite des fluids elastiques et sur celles de la force expansive de la vapeur deleau et de la vapeur de l, a differentes temperatures (in french), J. Ec. Polytech. 1 (2) (1795) 24–76.

[4] S.L. Marple, Digital Spectral Analysis with Applications, Prentice-Hall, Engelwood Cliffs, NJ, 1987.
[5] M.P. Ribeiro, D.J. Ewins, D.A. Robb, Non-stationary analysis and noise filtering using a technique extended from the original Prony method, Mech. Syst. Signal Process. 17 (3) (2003) 533–549.
[6] R. Kumaresan, R.W. Tufts, Estimating the parameters of exponentially damped sinusoids and pole-zero modelling in noise, IEEE Trans. Acoust. Speech Signal Process. 30 (1982) 833–840.
[7] D.J. Trudnowski, J.M. Johnson, J.F. Hauer, Making Prony analysis more accurate using multiple signals, IEEE Trans. Power Syst. 14 (1999) 226–231.
[8] L. Qi, L. Qian, S. Woodruff, D. Cartes, Prony analysis for power system transient harmonics, EURASIP J. Adv. Signal Process. (2007) 1–12.
[9] J.-H. Peng, N.-K. Nair, Adaptive sampling scheme for monitoring oscillations using prony analysis, IET Gener. Transm. Distrib. 13 (6) (2009) 1052–1060.
[10] Y. Li, K.J.R. Liu, J. Razavilar, Improved parameter estimation schemes for damped sinusoidal signals based on low-rank Hankel approximation, IEEE Trans. Signal Process. 45 (1997) 481–487.
[11] P. Dupuis, T. Sels, J. Driesen, R. Belmans, Exponential parameters measurement using a modified Prony method, in: IMTC, 2004, Como of Italy, pp. 1590–1594.
[12] J.-H. Lee, H.-T. Kim, Selection of sampling interval for least squares Prony method, Electron. Lett. 41 (2005) 47–49.
[13] J.H. Wilkinson, Rounding Errors in Algebraic Processes, Prentice Hall, Englewood Cliffs, New Jersey, 1963.
[14] L.N. Trefethen, D. Bau, Numerical Linear Algebra, SIAM, Philadelphia, PA, 1997.
[15] K. Holmstrom, J. Petersson, A review of the parameter estimation problem of fitting positive exponential sums to empirical data, Appl. Math. Comput. 126 (2002) 31–61.
[16] A. Cunha, E. Caetano, Experimental modal analysis of civil engineering structures, Sound Vib. (2006) 12–20.
[17] B. Peeters, G.D. Roeck, Reference-based stochastic subspace identification for output-only modal analysis, Mech. Syst. Signal Process. 13 (6) (1999) 855–878.
[18] S.Y. Kung, K.S. Arun, D.V.B. Rao, State-space and singular-value decomposition-based approximation methods for the harmonic retrieval problem, J. Opt. Soc. Am. 73 (12) (1983) 1799–1811.
[19] M. Kahn, A state space method for estimating frequencies and dampings, J. Stat. Comput. Simulation 48 (1993) 101–116.
[20] J.-N. Juang, R.S. Pappa, An eigensystem realization algorithm for modal parameter identification and model reduction, J. Guidance Control Dyn. 8 (1985) 620–627.
[21] E.O. Kreyszig, H. Kreyszig, E.J. Norminton, Advanced Engineering Mathematics, 10th edition, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2011.
[22] J.C. Polking, A. Boggess, D. Arnold, Differential Equations, Pearson Prentice Hall, 2006.
[23] G.H. Golub, C.F.V. Loan, Matrix Computations, 3rd edition, Johns Hopkins University Press, Baltimore, MD, 1996.
[24] G. Strang, Linear Algebra and Its Application, 3rd edition, Harcourth Brace Jovanovich, Inc., San Diego, CA, USA, 1988.
[25] B.L. Ho, R.E. Kalman, Effective construction of linear state-variable models from input/output functions, Regelungstechnik 14 (1966) 545–548.
[26] J.A. Cadzow, Signal enhancement—a composite property mapping algorithm, IEEE Trans. Acoust. Speech Signal Process. 36 (1988) 49–62.
[27] D.W. Tufts, Estimation of a signal waveform from noisy data using low-rank approximation to a data matrix, IEEE Trans. Signal Process. 41 (4) (1993) 1716–1721.
[28] S.-L.J. Hu, X. Bao, H. Li, Model order determination and noise removal for modal parameter estimation, Mech. Syst. Signal Process. 24 (6) (2010) 1605–1620.