EFFICIENT PARAMETER ESTIMATION OF MULTIPLE DAMPED SINUSOIDS BY COMBINING SUBSPACE AND WEIGHTED LEAST SQUARES TECHNIQUES

Weize Sun and H.C. So

Department of Electronic Engineering, City University of Hong Kong, Hong Kong SAR, China
weizesun2@student.cityu.edu.hk, hcs@ee.cityu.edu.hk

ABSTRACT
A new signal subspace approach for sinusoidal parameter estimation of multiple tones is proposed in this paper. Our main ideas are to arrange the observed data into a matrix without reuse of elements and exploit the principal singular vectors of this matrix for parameter estimation. Comparing with the conventional subspace methods which employ Hankel-style matrices with redundant entries, the proposed approach is more computationally efficient. Computer simulations are also included to compare the proposed methodology with the weighted least squares and ESPRIT approaches in terms of estimation accuracy and computational complexity.

Index Terms—frequency estimation, subspace method, singular value decomposition, linear prediction, weighted least squares

1. INTRODUCTION
Estimating the parameters of sinusoidal signals from noisy observations is an important research topic in science and engineering. The crucial step is to find the damping factors and frequencies which are nonlinear functions in the observed data. Once they have been estimated, computation of the remaining parameters reduces to a least squares (LS) fit.

Generally speaking, parameter estimation can be achieved by either nonparametric or parametric methodologies [1]. In most of the cases, the parametric approach, which assumes that the signal satisfies a generating model with known functional form, will have a higher resolution than the nonparametric ones. In estimating multiple nonlinear parameters, the maximum-likelihood based methodology requires extensive computations for a multi-dimensional search, which may not be suitable in many applications. The subspace methods that separate the data into signal and noise subspaces via eigenvalue decomposition of the sample covariance matrix or the singular value decomposition (SVD) of the raw data matrix, such as MUSIC [2] and ESPRIT [3], can achieve a high resolution with a moderate complexity. On the other hand, linear prediction (LP) approach, which includes the weighted least square (WLS) estimator [4], can attain optimum performance when signal-to-noise ratio (SNR) is sufficiently high. In this work, we propose to combine subspace and WLS techniques to achieve sinusoidal parameter estimation with low computational complexity and high accuracy.

The rest of this paper is organized as follows. The notation and formulation for sinusoidal parameter estimation are given in Section 2, then the proposed estimator is derived in Section 3. In Section 4, simulation results are included to evaluate the performance of the developed approach by comparing with the WLS [4] and ESPRIT [5] algorithms, as well as Cramér-Rao lower bound (CRLB). Finally, conclusions are drawn in Section 5.

2. NOTATION AND DATA MODEL
Throughout this paper, bold upper/lower case symbols denote matrices/vectors. The \( N_1 \times N_2 \) zero matrix, \( N \times N \) identity matrix and Kronecker product are represented by \( 0_{N_1 \times N_2} \), \( I_N \) and \( \otimes \), and superscripts \( T, H, *, -1 \) and \( {\dagger} \) denote transpose, Hermitian transpose, complex conjugation, matrix inversion and pseudo-inverse, respectively. Moreover, we use \( \hat{A} \) and \( \tilde{A} \) to represent the noise-free counterpart and estimate of \( A \).

The observed damped sinusoidal signal is:

\[
x_n = s_n + \xi_n, \quad n = 1, 2, \cdots, N
\]

(1)

where

\[
s_n = \sum_{k=1}^{K} \gamma_k \alpha_k e^{j \omega_k n}, \quad k = 1, 2, \cdots, K
\]

(2)

The \( \gamma_k, \alpha_k \in (0, 1) \), \( \omega_k \in [-\pi, \pi) \) are the complex amplitudes, damping factors and frequencies while \( \{\xi_n\} \) are zero mean complex white Gaussian noises with unknown variance \( \sigma^2 \). The number of sinusoids, denoted by \( K \), is assumed known \textit{a priori}. It is also assumed that \( N \), the length of the data, can be factorized as \( N = N_1 N_2 \), where \( N_1 > K \) and \( N_2 \) are integers. Note that even if \( N \) is not factorizable, we can simply discard a few samples and find \( N_1 \) and \( N_2 \) such that their product is closest to \( N \), and the performance loss will be negligible for a sufficiently large data length. Under this definition, stacking \( x_n \) into a matrix \( X \in \mathbb{C}^{N_1 \times N_2} \) yields:
\[ X = S + \Xi \]  
(3)

where
\[
S = [s_1 \ s_2 \ \cdots \ s_{N_2}] 
\]  
(4)

and \(s_{n_2} = [s_{(n_2-1)N_1 + 1} \ s_{(n_2-1)N_1 + 2} \ \cdots \ s_{n_2N_1}]\) for \(n_2 = 1, 2, \ldots, N_2\) while \(X\) and \(\Xi\) contain \(\{x_n\}\) and \(\{\xi_n\}\) accordingly.

### 3. PROPOSED ESTIMATOR

Following [7], \(S\) can be factorized as:
\[
S = G \Gamma \Omega_H^T
\]  
(5)

where
\[
\Gamma = \text{diag}([\gamma_1 \ \gamma_2 \ \cdots \ \gamma_K])
\]  
(6)
\[
G = [g_1 \ g_2 \ \cdots \ g_K]
\]  
(7)
\[
H = [h_1 \ h_2 \ \cdots \ h_K]
\]  
(8)
\[
g_k = [g_k^T \ \cdots \ g_k^{N_1}]^T
\]  
(9)
\[
h_k = [h_k^T \ \cdots \ h_k^{N_2}]^T
\]  
(10)
\[
g_{\ell_k} = \alpha_{\ell_k} e^{i\omega_{\ell_k}} \text{ and } h_{\ell_k} = \beta_{\ell_k} e^{i\phi_{\ell_k}}
\]  
(11)

Apparently, we have \(\alpha_{\ell_k} = \alpha_k, \omega_{\ell_k} = \omega_k, \beta_{\ell_k} = \beta_k^{N_1}\) and \(\mu_k = (N_1 \omega_k)\) mod \(2\pi\) for all \(k\). On the other hand, decomposing \(X\) using SVD gives:
\[
X = U A V^H = [U_s \ U_n] \begin{bmatrix} \Lambda_s & 0 \\ 0 & \Lambda_n \end{bmatrix} [V_s \ V_n]^H
\]  
(12)

where \(U_s \in \mathbb{C}^{N_1 \times l_K}, \Lambda_s \in \mathbb{C}^{l_K \times l_K}\) and \(V_s \in \mathbb{C}^{N_2 \times l_K}\), \(l_K = \min\{N_2, K\}\), are the signal subspace components. According to the decomposition in (5)-(11), the best rank-\(l_K\) approximation of \(S\) according to (12), denoted by \(\tilde{S}\), is
\[
\tilde{S} = U_s \Lambda_s V_s^H
\]  
(13)

As \(\text{span}(U_s) \subseteq \text{span}(G)\), we have
\[
\tilde{U}_s = G \Omega_G
\]  
(14)

where \(\Omega_G\) is an unknown \(K \times l_K\) matrix. Equation (14) shows that each column of \(\tilde{U}_s\), namely, \(\tilde{u}_k, k = 1, 2, \ldots, l_K\), is a sum of \(K\) damped cisoids such that the frequencies and damping factors in \(\{\tilde{u}_k\}\) are identical but having different amplitudes, which corresponds to a multi-channel spectral estimation problem [6]. For each \(\tilde{u}_k\), we have the following linear prediction (LP) property:
\[
\sum_{i=0}^{K} c_i [\tilde{u}_{k}]_{n_{1} - i} = 0
\]  
(15)

for \(k = 1, 2, \ldots, l_K, n_1 = K + 1, \ldots, N_1\) with \(c_0 = 1\) and \(c = [c_1 \ c_2 \ \cdots \ c_K]^T\) being the LP coefficient vector. By finding the roots of
\[
\sum_{i=0}^{K} c_i z^{K-i} = 0
\]  
(16)
says \(\tilde{g}_k, k = 1, 2, \ldots, K\). The frequency and damping factor estimates \(\hat{\omega}_{L,k}\) and \(\hat{\alpha}_{L,k}\) are:
\[
\hat{\omega}_{L,k} = \angle(\tilde{g}_k) \text{ and } \hat{\alpha}_{L,k} = |\tilde{g}_k|
\]  
(17)

Then the problem is reduced to finding the LP coefficient vector \(c\). By constructing the LP error vector \(e = D c - f\), \(e\) can be solved by the WLS technique [6]:
\[
\hat{c} = \arg \min_{c} e^H W e = (D^H WD)^{-1} D^H W f
\]  
(18)
\[
D = [D_1^T \ D_2^T \ \cdots \ D_K^T]^T
\]  
(19)
\[
f = [f_1^T \ f_2^T \ \cdots \ f_K^T]^T
\]  
(20)

\[
D_k = \text{Toeplitz} \left( [u_k^T]_K \ [u_k]_{K+1} \ \cdots \ [u_k]_{N_1-1}^T, [u_k]_K \ [u_k]_{K+1} \ \cdots \ [u_k]_{N_1-1}^T \right)
\]  
(21)
\[
f_k = - [u_k]_{K+1} \ [u_k]_{K+2} \ \cdots \ [u_k]_{N_1}^T
\]  
(22)

where \(W\) is a symmetric weighting matrix. Define \(A = \text{Toeplitz} \left( [c_K \ 0_{1 \times (N_1-K)}]^T, [c_{1\Pi} \ 0_{1 \times (N_1-K-1)}]^T \right)\) where \(c_{1\Pi} = [c_K \ c_{K-1} \ \cdots \ c_1 \ 1]^T\) and \(U_s = \tilde{U}_s + \Delta U_s\), we have the fact that \(A U_s = 0_{(N_1-K) \times N_1}\). It is shown in [8] that following the Gauss-Markov theorem [9], the optimal \(W\) is:
\[
W = \sigma^2 \left[ E \{ e e^H \} \right]^{-1}
\]  
(23)

\[
= \sigma^2 \left[ E \{ e \vec{A} U_s \vec{A}^H e \} \right]^{-1}
\]  
(23)
\[
= \sigma^2 \left[ E \{ e \vec{A} \Delta U_s \vec{A}^H e \} \right]^{-1}
\]  
(23)
\[
= \text{diag}(\lambda_1^2, \lambda_2^2, \ldots, \lambda_K^2) \otimes (A A^H)^{-1}
\]  
(23)
\[
\approx \text{diag}(\lambda_1^2, \lambda_2^2, \ldots, \lambda_K^2) \otimes (A A^H)^{-1}
\]  
(23)

by approximating \(\lambda_k\) using \(\lambda_k\). The estimation of \(c\) is done by an iterative procedure of (18) and (23) with an initial choice of \(W = \text{diag}(\lambda_1^2, \lambda_2^2, \ldots, \lambda_K^2) \), which can be obtained by applying (14) to (23) to \(V_s^*\). However, this will lead to two problems. First, although the \(\omega_k\)'s are identical naturally, \(\mu_k\)'s might not be identical as \(\mu_k = (N_1 \omega_k)\) mod \(2\pi\). Second, an extra pairing step between \(\omega_k\)'s and \(\mu_k\)'s is needed, which is costly. In order to overcome these problems, we employ another procedure for estimating \(\mu_k\)'s and \(\beta_k\)'s as follows. From (5), we have
\[
X \approx \tilde{G} H^T
\]  
(24)

where \(\tilde{G}\) is the estimate of \(G\) which is constructed by assigning \(g_k = \tilde{g}_k\) and
\[
H^T = \Gamma H^T = [h_1 \ h_2 \ \cdots \ h_K]^T
\]  
(25)
\[
h_k = \gamma_k h_k.
\]  
(26)

From (24), the LS estimate of \(H\) is
\[
\hat{H} = X^T (\tilde{G}^T G)^{-1}.
\]  
(27)
Noting that the elements of \( h_k \) possess the same LP property as in \( h_k \), we extract \( h_k \) from \( \hat{H} \) to construct the equations
\[
T_1 \hat{h}_k h_k \approx T_2 \hat{h}_k
\]
where \( T_1 = [I_{N_2 -1} \ 0_{(N_2 -1) \times 1}]^T \) and \( T_2 = [0_{(N_2 -1) \times 1} \ I_{N_2 -1}]^T \) are selection matrices. Considering sufficiently small error conditions such that \( \hat{g}_k \rightarrow g_k \), we have \( \hat{G} \rightarrow G \), and then \( X \) will be independent of \( G \), therefore the disturbances among each vector of \( \hat{h}_k \) can be assumed independent and identically distributed. Following [10], the WLS estimate of \( h_k, k = 1, 2, \cdots, K \), is computed as:
\[
\hat{h}_k = ( (T_1 \hat{h}_k)^H \Psi_k (T_2 \hat{h}_k)^{-1} (T_1 \hat{h}_k)^H \Psi_k T_2 \hat{h}_k
\]
The optimum weighting matrix \( \Psi_k \) has the form:
\[
\Psi_k = [E \left( (T_1 \hat{h}_k h_k - T_2 \hat{h}_k)^H (T_1 \hat{h}_k h_k - T_2 \hat{h}_k) \right)]^{-1} = (B_k B_k^H)^{-1}
\]
where
\[
B_k = \text{Toeplitz} \left( [h_k \ 0_{1 \times (N_2 -2)}]^T, [h_k \ 1 \ 0_{1 \times (N_2 -2)}]^T \right).
\]
We start with \( \Psi_k = I_{N_1 -1} \) in the iterations between (29) and (30) to obtain \( h_k, k = 1, 2, \cdots, K \). Finally, the frequencies \( \mu_k \) and damping factors \( \beta_k \) for \( k = 1, 2, \cdots, K \) are estimated as
\[
\hat{\mu}_k = \angle (\hat{h}_k) \quad \text{and} \quad \hat{\beta}_k = |\hat{h}_k|
\]
where \( \hat{h}_k \) and \( \hat{g}_k \) are automatically paired up.

As it is shown in [7], \( \hat{\mu}_k \) corresponds to \( 2|N_1/2| + 1 \) possible estimates of \( \omega_k \), where \( \lfloor \cdot \rfloor \) rounds the value to the nearest integer towards \( -\infty \), denoted by \( \hat{\omega}_{R,k,i} \), \( i = -\lfloor N_1/2 \rfloor, -\lfloor N_1/2 \rfloor + 1, \cdots, \lfloor N_1/2 \rfloor \):
\[
\omega_{R,k,i} = \frac{\hat{\mu}_k + 2\pi i}{N_1}
\]
Defining \( \hat{\omega}_{R,k} = \hat{\omega}_{R,k,f} \) where \( f \) is computed from
\[
f = \arg \min_{i \in \{-\lfloor N_1/2 \rfloor, -\lfloor N_1/2 \rfloor + 1, \cdots, \lfloor N_1/2 \rfloor \}} |\hat{\omega}_{R,k,i} - \hat{\omega}_{L,k}| \quad (34)
\]
and
\[
\hat{\alpha}_{R,k} = \hat{\beta}_k^{1/N_1} \quad (35)
\]
It will be shown in Section 4 that using \( \hat{\omega}_{R,k} \) and \( \hat{\alpha}_{R,k} \) has a much higher accuracy than that of \( \hat{\omega}_{L,k} \) and \( \hat{\alpha}_{L,k} \). Therefore, we assign
\[
\hat{\omega}_k \approx \hat{\omega}_{R,k} \quad \text{and} \quad \hat{\alpha}_k \approx \hat{\alpha}_{R,k}
\]
as the final estimates.

4. SIMULATION RESULTS

Computer simulations have been carried out to evaluate the performance of the proposed algorithm for multiple damped sinusoids in the presence of white Gaussian noise. The average mean square error (MSE) is assigned to evaluate the algorithm performance and the SNR in dB is defined as SNR = 10\log_{10}(\sum_{n=1}^{N}|s_n|^2/\sigma^2). All results provided are averages of 2000 independent runs.

In the first test, we study the performance of the proposed algorithm comparing with the WLS [4] and standard ESPRIT (SE) [5] approaches. The signal parameters are \( \gamma_1, \gamma_2 = [1 \ 2\gamma] \) \( \alpha_1, \alpha_2 = [0.99 \ 0.98] \), \( \omega_1, \omega_2 = [0.05\pi \ 0.36\pi] \). Figure 3 plots the MSEs for frequencies under different SNR. It is also shown that the proposed scheme performs almost the same as the SE algorithm, and has a better threshold performance than the GWLP method.

Figure 4 investigates the MSEs for frequencies when \( \omega_2 \) varies from \( -0.95\pi \) to \( 0.97\pi \) at \( \omega_1 = -0.99\pi \). The SNR is set to 40dB while other parameters remain unchanged. We see that for a larger \( N_1 \), the proposed method can correctly separate more closely spaced sources. It also indicates that with a properly chosen \( N_1 \) and \( N_2 \), the proposed scheme has almost the same frequency resolution performance as the SE and WLS methods.

5. CONCLUSION

A fast and accurate sinusoidal parameter estimation approach based on principal singular value decomposition of the data matrix has been derived. The key point is to employ the left singular vectors to get a rough estimation of the parameters at first, and then use the whole data matrix together with these values to get a better result. Furthermore, it is demonstrated that the proposed subspace scheme has an outstanding performance in terms of computational complexity and/or estimation accuracy.
6. REFERENCES


