Technical Research Report

Improved Parameter Estimation Schemes for Damped Sinusoidal Signals Based on Low-Rank Hankel Approximation

by Y. Li, K.J.R. Liu and J. Razavilar

T.R. 95-11

ISR
INSTITUTE FOR SYSTEMS RESEARCH

Sponsored by
the National Science Foundation
Engineering Research Center Program,
the University of Maryland,
Harvard University,
and Industry

Improved Parameter Estimation Schemes for Damped Sinusoidal Signals Based on Low-Rank Hankel Approximation

Ye Li, K. J. Ray Liu, and Javad Razavilar
Electrical Engineering Department
and Institute for Systems Research
University of Maryland
College Park, MD 20770
Fax: 1-301-405-6707
Email: liye(kjrliu, javad)@src.umd.edu

ABSTRACT

The parameter estimation of damped sinusoidal signals is an important issue in spectral analysis and many applications. The existing algorithms, such as the KT algorithm[8] and the TLS algorithm[13], are based on the low-rank approximation of prediction matrix, which ignores the Hankel property of the prediction matrix, We will prove in this paper that the performance of parameter estimation can be improved if both rank-deficient and Hankel properties of the prediction matrix are exploited in the matrix approximation. Based on this idea, a modified KT (MKT) algorithm and a super-resolution algorithm–damped MUSIC (DMUSIC) algorithm are proposed. Computer simulation results demonstrate that, compared with the original KT algorithm, the MKT and DMUSIC algorithms have about 5dB lower noise threshold and can estimate the parameters of signal with larger damping factors.

SP EDICS:

SP3.1.1 Spectral Analysis: Single-channel Time Series;

SP3.6.1 Parameter Estimation: Single-channel Time Series;

The work was supported in part by the NIH grant 1R01GM49707 and the NSF grants MIP9309506 and MIP9457397.

I Introduction

The problem of parameter estimation of damped sinusoidal signals in the presence of additive noise is very important in spectral analysis and many applications, such as magnetic resonance spectroscopy and radioastronomy. The parameter estimation of one-dimensional (1-D) sinusoidal signals also plays a key role in the parameter estimation of higher-dimensional (H-D) sinusoidal signals[1, 3, 16].

The difficulty of this problem stems from the fact that the damped sinusoidal signal is nonstationary and the correlation matrix can not be found. Hence, many efficient traditional approaches[15] are not applicable. There are several model-based algorithms being devised to cope with this problem. The Prony method[5] is one of the widely used algorithms, but it is sensitive to measurement noise. The forward-backward linear prediction algorithm[21] is one of the well-known algorithms for the frequency estimation of pure sinusoidal signals, and the backward linear prediction algorithm (or Kumaresan-Tufts (KT) algorithm)[8] for damped sinusoidal signals can attain the *Carmer-Rao* bound if the peak signal-to-noise ratio (SNR) is high and the damping factors of signals are small. But for the signals with lower SNR or large damping factor, the KT algorithm is unable to estimate the signal parameters effectively. Several algorithms have been proposed to improve the high noise threshold problem in the KT algorithm. Two of them are the total least square (TLS) algorithm [13] based on the rank approximation and the maximum likelihood (ML) algorithm [2] based on iterative optimization. A singular value decomposition (SVD)-based information theoretic criteria[14] have recently been presented to detect the number of damped/undamped sinusoids and parameter estimation.

The existing parameter estimation algorithms for damped sinusoidal signals use only the rank-deficient property of the prediction matrix and ignore its Hankel property. In this papper, we prove that the parameter estimation of sinusoidal signals from noisy data is equivalent to the low-rank Hankel matrix approximation of data matrix (or prediction matrix) and that the performance of parameter estimation will be improved significantly if both rank-deficiency and Hankel properties of the prediction matrix are exploited in matrix approximation. Based on this idea, a modified KT algorithm and a novel super-resolution algorithm–damped MUSIC (DMUSIC) algorithm are proposed, which use both the Hankel and the rank-deficiency properties of the prediction matrix. This paper is organized as follows. In Section II, the matrix approximation in the KT algorithm and TLS algorithm is analyzed. Then, a modified Kumaresan-Tufts (MKT) algorithm is developed in Section III. Next, a novel super-resolution algorithm–DMUSIC algorithm is presented in Section IV. Finally, computer simulation results are presented in Section V to demonstrate the performance of MKT and DMUSIC algorithms.

II Matrix Approximation in KT and TLS Algorithms

A sequence $x(n)$ consists of K damped sinusoidal signals can be expressed as

$$x(n) = \sum_{k=1}^{K} c_k e^{s_k n}, \tag{2.1}$$

where c_k's are nonzero complex amplitudes, $s_k = -\alpha_k + \jmath\omega_k$, and $\alpha_k \in \mathcal{R}^+,\ \omega \in [-\pi, \pi]$ for $k = 1,\ 2, \cdots,\ K$. α_k is called the *damping factor* of the damped sinusoid with angle frequency ω_k. The larger the damping factor, the faster the amplitude of the sinusoid decays. The observed sequence $y(n)$ is obtained from $x(n)$ corrupted by additive noise $w(n)$ which is assumed to be complex white Gaussian process. Normally, we have to make sure that $N(N \geq 2K)$. The observed data can expressed as

$$y(n) = x(n) + w(n) \quad for\ \ n = 0,\ 1,\ 2, \cdots\ N-1. \tag{2.2}$$

The *Kumaresan-Tufts* (KT) algorithm[8] and the *total least square* (TLS) algorithm[13] are two effective algorithms for parameter estimation of damped sinusoidal signals. We will briefly discuss the KT algorithm and the TLS algorithm in this section and then reveal how the matrix approximation is done in both algorithms.

A. KT algorithm and TLS algorithm

Due to the nonstationarity of damped sinusoidal signals, we first set up an $(N - L) \times L$

$(\min(N-L, L) \geq K)$ conjugate backward prediction matrix

$$\mathbf{A}_{N-L,L} = \begin{pmatrix} y^*(1) & y^*(2) & \cdots & y^*(L) \\ y^*(2) & y^*(3) & \cdots & y^*(L+1) \\ \vdots & \vdots & \vdots & \vdots \\ y^*(N-L) & y^*(N-L+1) & \cdots & y^*(N-1) \end{pmatrix}, \quad (2.3)$$

and an $(N-L)$-component column vector

$$\mathbf{h}_{N-L} = \begin{pmatrix} y^*(0) \\ y^*(1) \\ \vdots \\ y^*(N-L-1) \end{pmatrix}, \quad (2.4)$$

where "$*$" stands for the complex conjugate. From $\mathbf{A}$ and $\mathbf{h}$, we can get an $(N-L)\times(L+1)$ augmented matrix

$$\mathbf{B}_{N-L,L+1} = (\mathbf{h}_{N-L}, \mathbf{A}_{N-L,}). \quad (2.5)$$

To find the frequencies of the damped sinusoids, a L-component prediction coefficient vector $\mathbf{c}$ should be found such that

$$\mathbf{A}\mathbf{c} \approx -\mathbf{h}, \quad (2.6)$$

where

$$\mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_L \end{pmatrix}, \quad (2.7)$$

is the backward linear prediction coefficients. Then, $e^{-s_k^*}$ for $k = 1,\ 2, \cdots,\ K$ can be estimated by calculating the roots of the prediction polynomial

$$C(z) = 1 + c_1 z^{-1} + \cdots + c_L z^{-L}. \tag{2.8}$$

Hence, the performance of an algorithm relys on how accurate the estimation of the predication polynomial is.

The KT algorithm estimates $\mathbf{c}$ by

$$\mathbf{c} = -\sum_{k=1}^{M} \sigma_k^{-1}(\mathbf{u}_k^H \mathbf{h})\mathbf{v}_k, \tag{2.9}$$

where σ_k's ($\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_L$) are the singular values of $\mathbf{A}$, and $\mathbf{u}_k$ and $\mathbf{v}_k$ are the left singular vector and the right singular vector of $\mathbf{A}$ corresponding to the singular value σ_k, and "H" stands for the conjugate transposition. To obtain optimum performance, the L ($L \geq K$) is chosen to be larger than $N - L$, usually $3N/4$[15]. It has been proved in [8] and [6] that if $\mathbf{c}$ is estimated using (2.9), then only K of $C(z)$'s zeros are outside the unit circle, which are signal zeros $e^{-s_k^*}$ for $k = 1,\ 2, \cdots, K$. The rest of the $L - K$ zeros are inside the unit circle. By means of this property, the desired zeros can be easily identified to estimate the parameters.

The TLS algorithm, on the other hand, tries to find a $1 \times L$ vector $\mathbf{c}$ and an $(N - L) \times (L + 1)$ matrix $\mathbf{D}$ with minimum norm $\| \mathbf{D} \|_F$ subject to

$$(\mathbf{B} + \mathbf{D}) \begin{pmatrix} 1 \\ \mathbf{c} \end{pmatrix} = 0, \tag{2.10}$$

where $\| \,.\, \|_F$ denotes the Frobenius norm given by

$$\| \mathbf{D} \|_F = (\sum_i \sum_j \mid d_{ij}^2 \mid)^{1/2}. \tag{2.11}$$

For the TLS algorithm, one usually picks $(N - L) \geq (L + 1)$. Unlike the KT algorithm, some of the $L - K$ extraneous zeros of prediction polynomial may lie outside the unit circle, which makes it difficult to distinguish the signal zeros from those extraneous zeros outside the unit circle. The difference in estimating the prediction polynomial gives rise to their performance difference.

B. Matrix Approximation Point of View

If there is no noise, the rank of $\mathbf{A}_{N-L,L}$ or $\mathbf{B}_{N-L,L+1}$ will be K. However, $\mathbf{A}_{N-L,L}$ or $\mathbf{B}_{N-L,L+1}$ is usually full rank due to the measurement noise. Hence, a low-rank matrix approximation to $\mathbf{A}_{N-L,L}$ or $\mathbf{B}_{N-L,L+1}$ is used in both KT algorithm and TLS algorithm.

According to [10], the optimum rank K matrix approximation of $\mathbf{A}_{N-L,L}$ can be expressed as

$$\widehat{\mathbf{A}}_{N-L,L} = \sum_{k=1}^{K} \sigma_k \mathbf{u}_k \mathbf{v}_k^H. \tag{2.12}$$

To make the system equations

$$\widehat{\mathbf{A}}\mathbf{c} = -\mathbf{h} \tag{2.13}$$

have solution, either $\mathbf{h}$ must be in span$\{\mathbf{u}_1,\ \mathbf{u}_2, \cdots, \mathbf{u}_K\}$ or $\widehat{\mathbf{h}}$, the projection of $\mathbf{h}$ on span$\{\mathbf{u}_1,\ \mathbf{u}_2, \cdots, \mathbf{u}_K\}$, must be used instead of $\mathbf{h}$ in (2.13). In either case, (2.13) can be written as

$$\widehat{\mathbf{A}}\mathbf{c} = -\widehat{\mathbf{h}}, \tag{2.14}$$

where

$$\widehat{\mathbf{h}} = \sum_{k=1}^{K} (\mathbf{u}_k^H \mathbf{h}) \mathbf{u}_k. \tag{2.15}$$

Since rank$(\widehat{\mathbf{A}}) = K \leq L$, (2.14) is an underdetermined system of equations about $\mathbf{c}$, and there are multiple solutions. The solution minimizing $\| \mathbf{c} \|$ is given in (2.9), which is the $\mathbf{c}$ in the KT algorithm. The minimum norm solution of $\mathbf{c}$ makes the extraneous zeros of the prediction polynomial lie inside the unit circle[6].

If $(N-L) \geq (L+1)$ in $\mathbf{B}_{N-L,L+1}$, then rank$(\mathbf{B}_{N-L,L+1})$ is uaually equal to $L+1$ because of noise. Let $\widehat{\mathbf{B}}_{N-L,L+1}$ be the optimum rank L matrix approximation of $\mathbf{B}_{N-L,L+1}$, then $\widehat{\mathbf{B}}_{N-L,L+1}$ is the matrix that minimizes $\| \mathbf{B}_{N-L,L+1} - \widehat{\mathbf{B}}_{N-L,L+1} \|_F$ and makes (2.8) have a unique solution. Hence, the TLS algorithm first makes the low-rank matrix approximation to $\mathbf{B}_{N-L,L+1}$, then finds $\mathbf{D}$ by

$$\mathbf{D} = \widehat{\mathbf{B}}_{N-L,L+1} - \mathbf{B}_{N-L,L+1}, \tag{2.16}$$

and finally obtains the estimation $\mathbf{c}$ by solving

$$(\mathbf{B} + \mathbf{D}) \begin{pmatrix} 1 \\ \\ \mathbf{c} \end{pmatrix} = \mathbf{0} \tag{2.17}$$

where $\mathbf{0}$ is a zero matrix with an appropriate dimension.

From the above discussion, both the KT algorithm and the TLS algorithm use the low-rank matrix approximation to reduce the noise effect. When the signal to noise ratio (SNR) is high and enough data are available, the rank approximation in the KT algorithm will reduce the measurement noise significantly, hence KT algorithm in this case will almost attain the *Cramer-Rao* bound[8]. However, if the SNR is reduced to certain degree, the rank approximation in the KT algorithm is unable to reduce the noise effect efficiently and moreover, it may introduce an extra perturbation. In that case, the noise threshold appears. Since the TLS algorithm only makes a one-order lower rank matrix approximation to $\mathbf{B}_{N-L,L+1}$, it only gives rise to smaller perturbation under low SNR. Therefore, it has lower noise threshold than the KT algorithm does.

Since the noise threshold of the KT algorithm is due to the low-rank matrix approximation, to reduce the noise threshold, the matrix approximation approach employed in the KT algorithm must be improved. We will show in the next section that if the Hankel structure of the backward prediction matrix can be preserved while performing the low-rank approximation, the noise threshold can be significantly reduced.

III Modified Kumaresan-Tufts Algorithm

From (2.3), we can see that the prediction matrix of a data sequence is of Hankel form. Indeed, there is a very interesting property which can be summerized in the following theorem.

Theorem 3.1 : *If a data sequence $x(n)$ consists of K distinct sinusoids as in (2.1), then for any $L(L > K)$, the $L \times L$ prediction matrix $\mathbf{P}_L = [x(i+j)]_{i,j=0}^{L-1}$ is a singular Hankel matrix with rank K and full rank $K \times K$ principle minor $\mathbf{P}_K = [x(i+j)]_{i,j=0}^{K-1}$. Conversely, for any $L \times L$ singular Hankel matrix $\mathbf{P}_L = [x(i+j)]_{i,j=0}^{L-1}$ with rank K, if its $K \times K$ principle minor $\mathbf{P}_K = [x(i+j)]_{i,j=0}^{K-1}$ is full rank, then $x(n)$ for $n = 0,\ 1, \cdots, (2L-2)$ can be uniquely expressed as the summation of K distinct sinusoids as given by (2.1).*

Proof: The first part of the theorem is easy to prove and it is much more clear after the derivation of the DMUSIC algorithm in the next section. Its proof is thus omited. Only the proof of the second part is given here.

Let $\mathbf{P}_L = [x(i+j)]_{i,j=0}^{L-1}$ be a $L \times L$ singular Hankel matrix with rank K and $\mathbf{P}_K = [x(i+j)]_{i,j=0}^{K-1}$ be its $K \times K$ principle minor with full rank. Then $\mathbf{P}_L$ can be viewed as a singular extension of $\mathbf{P}_K$. According to [4], there exists a unique set of complex numbers $a_0,\ a_1, \cdots, a_{K-1}$ such that

$$x(n) = \sum_{i=0}^{K-1} a_i x(n-i) \quad \text{for} \quad n = K,\ K+1, \cdots,\ (2L-2). \tag{3.1}$$

Therefore, $x(n)$ for $n = 0,\ 1, \cdots,\ (2L-2)$ can be uniquely expressed as the summation of K distinct sinusoids if $x(0),\ x(1), \cdots,\ x(K-1)$ are given.

□

The above theorem reveals a one-to-one correspondence between a data sequence consisting of damped sinusoidal signals and rank-deficient Hankel matrix. Therefore, parameter estimation of damped sinusoidal signals from noisy data is equivalent to performing the low-rank Hankel matrix approximation. More specifically, let $\mathbf{P}_L$ be an $L \times L$ prediction matrix of noisy data $y(n)$,

$$\mathbf{P}_L = [y(i+j)]_{i,j=0}^{L-1}. \tag{3.2}$$

If we can find an $L \times L$ Hankel matrix $\bar{\mathbf{P}} = [\bar{y}(i+j)]_{i,j}^{L-1}$ with rank K and a full rank $K \times K$ principle minor, then the parameters of the signal can be uniquely determined from $\bar{\mathbf{P}}$.

For both the TLS and the KT algorithms, only the rank-deficiency characteristics of the prediction matrix is used in matrix approximation. The approximated matrix $\widehat{\mathbf{A}}_{N-L,L}$ in (2.12) or $\widehat{\mathbf{B}}_{N-L,L+1}$ in (2.17) unfortunately loses the Hankel property. If both the rank and Hankel properties of the matrix are used in the matrix approximation to reduce the noise effect, the performance of the estimation will be improved significantly. The modified KT algorithm introduced here will exploit both properties.

To use the low-rank Hankel matrix approximation to reduce the measurement noise, we first set up a square prediction matix from the observed noisy data:

$$\mathbf{P}_L = [y(i+j)]_{i,j=0}^{L-1}. \tag{3.3}$$

To make full use of the given data, let $L = \lceil N/2 \rceil$ here. Since there is no analytical low-rank Hankel matrix approximation approach available, an iterative approach for low-rank Hankel matrix approximation is used here. First, an optimum rank K matrix approximation to $\mathbf{P}_L$ is made using the SDV.

$$\bar{\mathbf{P}}_L = [\bar{y}_{i,j}]_{i,j=0}^{L-1} = \sum_{k=1}^{K} \sigma_k \mathbf{u}_k \mathbf{v}_k^H, \tag{3.4}$$

where σ_k for $k = 1, 2, \cdots, K$ are the K largest singular value of $\mathbf{P}_L$ and $\mathbf{u}_k$, $\mathbf{v}_k$ are corresponding left and right singular vectors. $\bar{\mathbf{P}}_L$ is uaually not Hankel. Then a Hankel matrix $\widehat{\mathbf{P}}_L$ is found to minimize $\| \widehat{\mathbf{P}}_L - \bar{\mathbf{P}}_L \|_F$, where $\widehat{\mathbf{P}}_L$ is given by

$$\widehat{\mathbf{P}}_L = [\widehat{y}(i+j)]_{i,j=0}^{L-1}, \tag{3.5}$$

and

$$\widehat{y}_{i+j} = \frac{1}{N} \sum_{0 \le n,m \le L-1,\ m+n=i+j} \bar{y}_{n,m}, \tag{3.6}$$

with N being the number of the elements in matrix $\bar{\mathbf{P}}_L$ satisfying $n+m = i+j$ in (3.4). After this step the rank of $\widehat{\mathbf{P}}_L$ is not necessarily K. A low-rank approximation is used again. The procedures are repeated until a Hankel matrix with only K dominate singular values is obtained. From the approximated Hankel matrix $\widehat{\mathbf{P}}_L$, a better noise-reduced data $\widehat{y}(n)$ can be found. Then by using the KT algorithm, the parameters of the signal can be

Table 1: Modified KT algorithm

Step 1	Form square prediction matrix $\mathbf{P}_{L_m}$
Step 2	Find $\bar{\mathbf{P}}_{L_m}$ by (3.4)
Step 3	Find $\widehat{\mathbf{P}}_{L_m}$ by (3.5)
Step 4	Repeat Step 2 and 3 to get estimation of $\widehat{y}(n)$
Step 5	Estimat parameters using the KT algorithm to $\widehat{y}(n)$

obtained from $\widehat{y}(n)$. The algorithm is summerized in Table 1. The proof of the convergence of the above iteration remains open although it always converges in our simulation. In what follows, we will prove that the proposed low-rank Hankel approximation can indeed achive better performance.

Theorem 3.2 *Let* $\mathbf{P}_{true} = [x(i+j)]_{i,j=0}$ *be the true prediction matrix, then*

$$\| \widehat{\mathbf{P}}_L - \mathbf{P}_{true} \|_F \leq \| \bar{\mathbf{P}}_L - \mathbf{P}_{true} \|_F . \tag{3.7}$$

The equality holds only if $\bar{\mathbf{P}}_L$ *is Hankel.*

Proof: From (3.6), we have

$$\begin{aligned}
& \frac{1}{N} \sum_{i+j=n,\ 0\leq i,j\leq n} | x(n) - \bar{y}_{i,j} |^2 \qquad (3.8)\\
= & \frac{1}{N} \sum_{i+j=n,\ 0\leq i,j\leq n} | (x(n) - \widehat{y}(n)) + (\widehat{y}(n) - \bar{y}_{i,j}) |^2 \\
= & | x(n) - \widehat{y}(n) |^2 + \frac{1}{N} \sum_{i+j=n,\ 0\leq i,j\leq n} | \widehat{y}(n) - \bar{y}_{i,j} |^2 \\
& + \frac{1}{N} \sum_{i+j=n,\ 0\leq i,j\leq n} \{(x(n) - \widehat{y}(n))(\widehat{y}(n) - \bar{y}_{i,j})^* + (x(n) - \widehat{y}(n))^*(\widehat{y}(n) - \bar{y}_{i,j})\} \\
= & | x(n) - \widehat{y}(n) |^2 + \frac{1}{N} \sum_{i+j=n,\ 0\leq i,j\leq n} | \widehat{y}(n) - \bar{y}_{i,j} |^2 \\
\geq & | x(n) - \widehat{y}(n) |^2,
\end{aligned}$$

where N is the number of elements in matrix $\bar{\mathbf{P}}_L$ satisfying $i+j=n$. The third equality is based upon the assumption in (3.6).

Using the above inequality, a direct calculation yields that

$$\begin{aligned}
& \| \bar{\mathbf{P}}_L - \mathbf{P}_{true} \|_F^2 \\
= & \| \bar{\mathbf{P}}_L - \widehat{\mathbf{P}}_L \|_F^2 + \| \widehat{\mathbf{P}}_L - \mathbf{P}_{true} \|_F^2 \\
\geq & \| \widehat{\mathbf{P}}_L - \mathbf{P}_{true} \|_F^2 .
\end{aligned} \tag{3.9}$$

□

The above theorem demonstrates that $\bar{\mathbf{P}}_L$ is always more accurate than $\widehat{\mathbf{P}}_L$. If the SVD in the iteration procedures can reduce the noise effect efficientely, a better estimation of $\mathbf{P}_{true}$ can be obtained by preserving the Hankel form after each iteration. Hence, the performance of the modified KT algorithm is better than that of the original KT algorithm.

Even though we emphasize the modified KT algorithm in this section, the similar procedures can also be used for the TLS algorithm.

IV Damped MUSIC Algorithm

The rank-deficiency and the Hankel properties of the prediction matrix can be directly used in the parameter estimation as in the DMUSIC algorithm developed in this section. Since there is no correlation matrix for damped sinusoidal signals, different from the conventional MUSIC algorithm, the DMUSIC algorithm can be derived only if the data matrix is set up in a structural way.

To derive the DMUSIC algorithm, we will set up a new $L \times L$ *prediction matrix*:

$$\mathbf{A} = \begin{pmatrix} y(0) & y(1) & \cdots & y(L-1) \\ y(1) & y(2) & \cdots & y(L) \\ \vdots & \vdots & \vdots & \vdots \\ y(L-1) & y(L) & \cdots & y(2L-2) \end{pmatrix} . \tag{4.1}$$

The prediction matrix to DMUSIC algorithm is as the correlation matrix to MUSIC algorithm. From (2.1) and (2.2), $\mathbf{A}$ can be written as

$$\mathbf{A} = \sum_{k=1}^{K} c_k \mathbf{r}(s_k)\mathbf{r}^T(s_k) + \mathbf{W} = \mathbf{SCS}^T + \mathbf{W}, \tag{4.2}$$

where $\mathbf{r}(s)$ and $\mathbf{S}$ are *signal vector* and *signal matrix* that are defined as

$$\mathbf{r}(s_k) = \begin{pmatrix} 1 \\ e^{s_k} \\ \vdots \\ e^{(L-1)s_k} \end{pmatrix} \quad \text{and} \quad \mathbf{S} = [\mathbf{r}(s_1), \mathbf{r}(s_2), \cdots, \mathbf{r}(s_K)], \tag{4.3}$$

respectively, $\mathbf{C}$ is a $K \times K$ diagonal matrix with $\text{diag}(\mathbf{C}) = (c_1,\ c_2, \cdots, c_K)$ and $\mathbf{W} = [w(i+j)]_{i,j=0}^{L-1}$ is the *noise matrix.*

If s_k's are distinct, then $\mathbf{r}(s_k)$ for $k = 1,\ 2, \cdots, K$ are linear independent and hence $\mathbf{S}$ is of full column rank. Since the rank of $\mathbf{C}$ is K, the rank of $\mathbf{A}$ is equal to K if there is no measurement noise. Now, assume that there is no noise. By means of singular value decomposition, $\mathbf{A}$ can be decomposed into the product of three matrices[10]

$$\mathbf{A} = \mathbf{UDV}^H, \tag{4.4}$$

where $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices, $\mathbf{D}$ is a diagonal matrix and

$$\text{diag}(\mathbf{D}) = (\sigma_1,\ \sigma_2, \cdots \sigma_K,\ 0, \cdots,\ 0), \qquad \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_K. \tag{4.5}$$

According to (4.4),

$$\mathbf{AV} = \mathbf{UD} \tag{4.6}$$

Denote $\mathbf{v}_i$ the i-th column of $\mathbf{V}$. span$\{\mathbf{v}_1, \cdots, \mathbf{v}_K\}$ is called *signal subspace* for

$$\text{span}\{\mathbf{v}_1, \cdots, \mathbf{v}_K\} = \text{span}\{\mathbf{r}(s_1), \cdots, \mathbf{r}(s_K)\}, \tag{4.7}$$

where span{} refers to the subspace that is defined by the set of all linear combinations of the vectors.

From (4.5) and (4.6), we have the following *orthogonality relations*

$$\mathbf{A}\mathbf{V_n} = \mathbf{0}, \quad \text{or} \quad \mathbf{A}\mathbf{v}_k = \mathbf{0} \quad \text{for } k = K+1, \cdots, L. \tag{4.8}$$

where $\mathbf{V_n} = [\mathbf{v}_{K+1}, \cdots, \mathbf{v}_L]$. From (2.4),

$$\mathbf{S}\mathbf{C}\mathbf{S}^T\mathbf{v}_k = \mathbf{0} \quad \text{for } k = K+1, \cdots, L. \tag{4.9}$$

Since both $\mathbf{S}$ and $\mathbf{C}$ are of full rank, $\mathbf{S}^T\mathbf{v_k} = \mathbf{0}$ for $k = K+1, \cdots, L$, i.e. $\mathbf{r}^T(s_n)\mathbf{v_k} = 0$ for $k = K+1, \cdots, L$ and $n = 1,\ 2, \cdots K$. Hence, ${\mathbf{V_n}}^T\mathbf{r}(s) = \mathbf{0}$ only when $s = s_1, \cdots, s_K$. Therefore, s_k can be obtained by finding s which makes $\| {\mathbf{V_n}}^T\mathbf{r}(s) \| = 0$.

When noise exists, the orthogonality relations (4.8) no longer hold. In this case, we can search for signal vectors that most closely orthogonal to the noise subspaces. Hence, s_k can be obtained by finding the peak of the following spectrum

$$\frac{1}{\bar{\mathbf{r}}^H(s){\mathbf{V_n}}^*{\mathbf{V_n}}^T\bar{\mathbf{r}}(s)}, \tag{4.10}$$

where

$$\bar{\mathbf{r}} = \frac{\mathbf{r}}{\| \mathbf{r} \|} \tag{4.11}$$

The algorithm is summerized in Table 2.

The algorithm discussed above is called the damped MUSIC (DMUSIC) algorithm for it looks like the MUSIC algorithm. But, there are several crucial differences between DMUSIC algorithm and MUSIC algorithm in that DMUSIC algorithm is for the parameter estimation of the damped sinusoidal signal which is nonstationary. Since the correlation matrix

Table 2: Damped MUSIC algorithm

Step 1	Forming data matrix **A** using (4.1)
Step 2	Finding $\mathbf{V_n}$ by making SVD to **A**
Step 3	Estimating s_k by find the peaks of (4.10)

is not available for nonstationary signals, the prediction matrix is used in DMUSIC algorithm. DMUSIC algorithm searchs on $(\alpha,\ \omega)$ so that these two parameters can be estimated simultaneously.

Remarks:

- In the derivation of the DMUSIC algorithm, the prediction matrix **A** in (4.1) is set up to be square. By similar argument, it is easy to show that the DMUSIC algorithm can also work when **A** is not square.

- In order to obtain the noise space and the signal space, the model order should be known in advance. We may determine the model order by observing the number of the dominate eigenvalues. Since the problem of order determination is an important topic in spectrum analysis itself, we shall simply note the existence of various effective methods [9], [14] and [19].

- At the first glance, the DMUSIC algorithm has nothing to do with the Hankel matrix approximation. But it exploits the rank-deficiency and Hankel properties implicitly. By searching $s = \alpha + \jmath\omega$ to maximize (4.10), the DMUSIC algorithm find the complex frequencies $s_1,\ s_2, \cdots, s_K$ such that

$$\mathbf{A} \approx \mathbf{SCS}^T, \tag{4.12}$$

where **S** is as in (4.3). $\mathbf{SCS}^T$ is a Hankel matrix with rank K. Hence, the DMUSIC algorithm directly finds the low-rank Hankel approximation of the prediction matrix **A**.

V Computer Simulation Examples

In this section, we will test the performance of the MKT and the DMUSIC algorithms and compare them with the KT algorithm by two computer simulation examples.

In our examples, the damped sinusoids is corrupted by complex white Gaussian noise with zero-mean and variance σ^2. The SNR used in the examples is the peak signal-to-noise ratio defined as

$$\text{SNR} = 10\log(\frac{1}{2\sigma^2}). \tag{5.1}$$

The performance of the algorithms is measured by the mean square error (MSE). For comparison, we also simulate the performance of Kumaresan-Tufts (KT) algorithm[8] and calculate the Cramer-Rao bound using the formula in [8]. In our simulation, $N = 24$, $L = 18$ for K-T algorithm and MKT algorithm and $L = 12$ for DMUSIC algorithm.

Example 1:

The simulated data are given by

$$x(n) = e^{sn} + w(n). \tag{5.2}$$

where $s = -\alpha + \jmath\omega$, and $w(n)$ is complex white Gaussian noise and with variance σ^2.

When we fix $s = -0.20 + \jmath 2\pi(0.52)$ and change the SNR, the simulation results are shown in Figure 1 (a) and (b), for the MSE of damping factor α and frequency ω, respectively. From this figure, the performance of the KT algorithm and the MKT algorithm are near Cramer-Rao (CR) bound if SNR is high. The MSE of DMUSIC is a little bit larger than the CR bound, but the noise threshold of DMUSIC algorithm is the lowest of the three algorithms. We can see that the performance of both MKT and DMUSIC algorithms follow closely to the CR bound in estimating damping factor α, and the noise threshold in estimating the frequency is about 6-7dB below that of the KT algorithm.

When we fix SNR $= 20dB$ and change the damping factor α. The MSE's of α and ω are shown in Figure 2 (a) and (b). From these, we can see that when the damping factor is less than 0.2, the performance of KT, MKT and DMUSIC algorithms is near CR bound. If $\alpha \geq 0.55$ (Figure 2 (b)), KT algorithm is unable to estimate the parameters while MKT and DMUSIC algorithms can still estimate the parameters effectively. In particular, the

MKT algorithm outperforms the DMUSIC algorithm for large damping factor.

Example 2:

The simulated data are generated by

$$y(n) = e^{s_1 n} + e^{s_2 n} + w(n), \tag{5.3}$$

where $s_1 = -0.2 + \jmath 2\pi(0.42)$, $s_2 = -0.1 + \jmath 2\pi(0.42 + \Delta)$, and $w(n)$ is complex white Gaussian noise with variance σ^2.

When SNR = 40dB, $\Delta = 0.1$. the spectrum of the DMUSIC is shown in Figure 3 (a) and the contour in Figure 3 (b). From the figures, the damping factors and frequencies of the signal can be easily estimated simultaneously by finding the peak on the spectrum. But if SNR = 40dB, $\Delta = 0$, i.e. two exponentially damped signals with the same frequency, the spectrum has just one peak (see Figure 3 (c) and (d)). Hence, the damping factors of the signals can not be correctly estimated by the DMUSIC algorithm. However, if SNR is increased to $60dB$, both the damping factors and the frequencies of the signals can be estimated again as demonstrated by Figure 3 (e) and (f).

The zeros of prediction polynomials are shown in Figure 4. From the figure, it is clear that the MKT algorithm has less bias and smaller variance than the KT algorithm.

If we fix $\Delta = 0.1$ (in this case, the example is the Experiment 1 of [8]), the MSE's of $\alpha_1, \omega_1, \alpha_2$ and ω_2 for KT algorithm, MKT algorithm and DMUSIC algorithm are shown in Figure 5 (a)-(d). From these figures, the noise threshold of MKT algorithm and DMUSIC algorithm is about $5dB$ lower than that of KT algorithm.

VI Conclusions

The reduced-rank matrix approximation has been an effective tool in many branches of signal processing. In this paper, we show that if we can also preserve the matrix structure, such as the Hankel structure for the case of parameter estimation of damped sinusoidal signals, the performance can be further improved. Specifically, we presented the MKT algorithm and the DMUSIC algorithm to estimate the parameters of damped sinusoidal signals. The MKT algorithm and DMUSIC algorithm exploit both reduced rank and Hankel properties of the prediction matrix. Compared with the original KT algorithm, they have about 5dB lower noise threshold and are able to estimate the parameters of signal with large damping factors. Hence, preserving the Hankel structure in reduced-rank matrix approximation does improve the performance significantly. The proposed approach and concept presented in this paper can also be extended to the general area of redeuced rank signal processing [17] where structural low-rank approximation can be very effective in performance improvement.

References

[1] M. M. Barbieri and P. Barone, "A two-dimensional Prony's method for spectral estimation" *IEEE Trans. on Signal Processing*, vol. 40, pp. 2747-2756, Nov. 1992.

[2] Y. Bresler and A. Macovski, "Exact maximum likelihood parameter estimation of superimposed exponential signals in noise," *IEEE Trans. on Acoust., Speech, Signal Processing*, vol. 34, pp. 1081-1089, Oct. 1986.

[3] Y. Hua, "Estimating two-dimensional frequencies by matrix enhancement and matrix pencil" *IEEE Trans. on Signal Processing*, vol. 40, pp. 2267-2280, Sept. 1992.

[4] I. S. Iohvidov, *Hankel and Toeplitz Matrices and Forms*, Birkhäuser Boston 1982.

[5] S. M. Kay and S. L. Marple, Jr., "Spectrum analysis-a modern perspective," *Proceeding of the IEEE*, vol. 69, pp. 1380-1419, Nov. 1981.

[6] R. Kumaresan, "On the zeros of the linear prediction-error filter for deterministic signals ," *IEEE Trans. on Acoust., Speech, Signal Processing*, vol. 31, pp. 217-220, Feb. 1983.

[7] R. Kumaresan, L. L. Scharf, and A. K. Shaw, "An algorithm for pole-zero modeling and spectral analysis," *IEEE Trans. on Acoust., Speech, Signal Processing*, vol. 34, pp. 637-640, June 1986.

[8] R. Kumaresan and R. W.Tufts, "Estimation the parameters of exponentially damped sinusoids and pole-zero modelling in noise," *IEEE Trans. on Acoust., Speech, Signal Processing*, vol. 30, pp. 833-840, Dec. 1982.

[9] K. Konstantinides and K. Yao, "Statistical analysis of effective singular values in matrix rank determination," *IEEE Trans. on Acoust., Speech, Signal Processing*, vol. 36, pp. 757-763, May 1988.

[10] C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*, Prentice-Hall, 1974.

[11] C. K. Papadopoulos and C. L. Nikias, "Parameter estimation of exponentially damped sinusoids using higher order statistics," *IEEE Trans. on Acoust., Speech, Signal Processing*, vol. 38, pp. 1424-1436, August 1990.

[12] B. Porat and B. Friedlander, "On the accuracy of the Kumaresan-Tufts method for estimating complex damped exponentials," IEEE Trans. on Acoust., Speech, Signal Processing, vol. 35, pp. 231-235, Feb. 1987.

[13] M. A. Rahman and K. B. Yu, "Total least squares approach for frequency estimation using linear prediction," *IEEE Trans. on Acoust., Speech, Signal Processing*, vol. 35, pp. 1440-1454, Oct. 1987.

[14] V. U. Reddy and L. B. Biradar, "SVD-based information theoretic criteria for detection of the number of Damped/Undamped sinusoids and their performance analysis," *IEEE Trans. on Signal Processing*, vol. 41, pp. 2872-2881, Sept. 1993.

[15] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, 1991.

[16] J. J. Sacchini, W. M. Steedly and R. L. Moses "Two-dimensional Prony modeling and parameter estimation," *IEEE Trans. on Signal Processing*, vol. 41, pp. 3127-3137 November 1993.

[17] L. L. Scharf "The SVD and reduced rank signal processing," *Signal Processing*, vol. 41, pp. 113-133 1991.

[18] D. S. Stephenson, "Linear prediction and maximum entropy methods in NMR spectroscopy," *Progress in NMR Spectroscopy*, vol. 20, pp 515-626, 1988.

[19] G. W. Stewart, "Perturbation Theory for the Singular Value Decomposition," *SVD and Signal Processing, II* (edited by R. J. Vaccaro), pp 99-109, ElsevierScience Publisher, 1991

[20] A. Swami and J. M. Mendel, "Cumulant-Based Approach to the Harmonic Retrieval and Related Problems," *IEEE Trans. on Signal Processing*, vol. 39, pp. 1099-1109, May 1991.

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

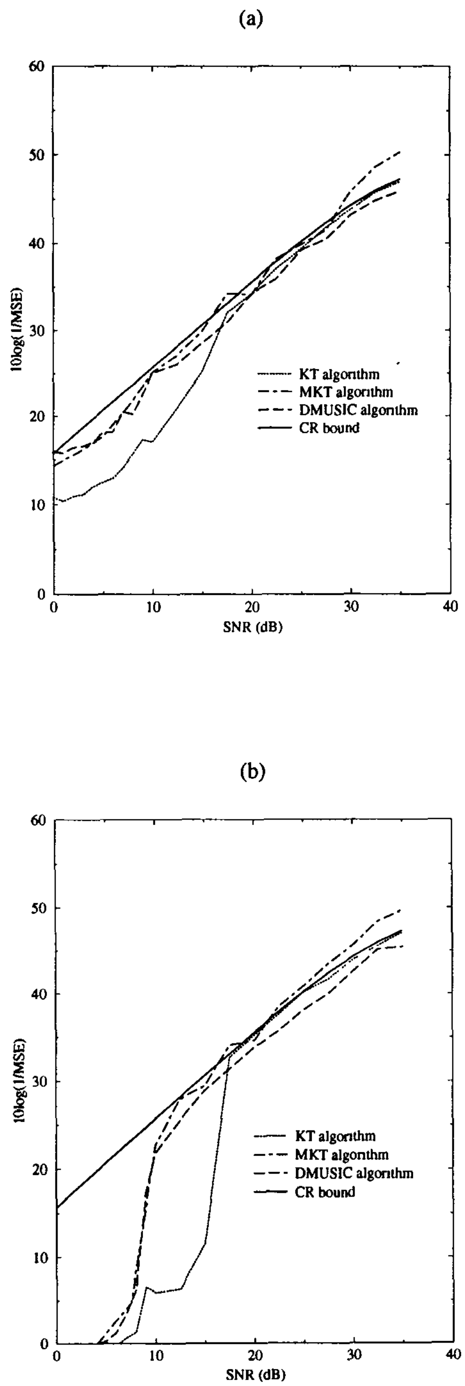


Figure 1: The MSE of (a) α and, (b) ω vs SNR of KT, MKT and DMUSIC algorithm obtained in 100 trails when $s = -0.2 + \jmath 2\pi 0.42$ and $N = 24$

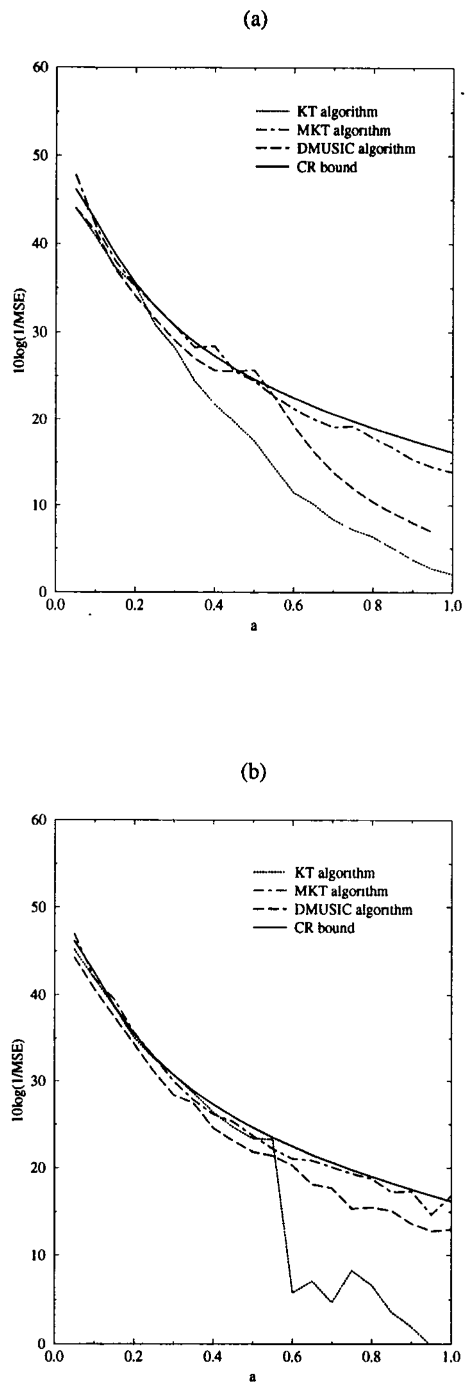
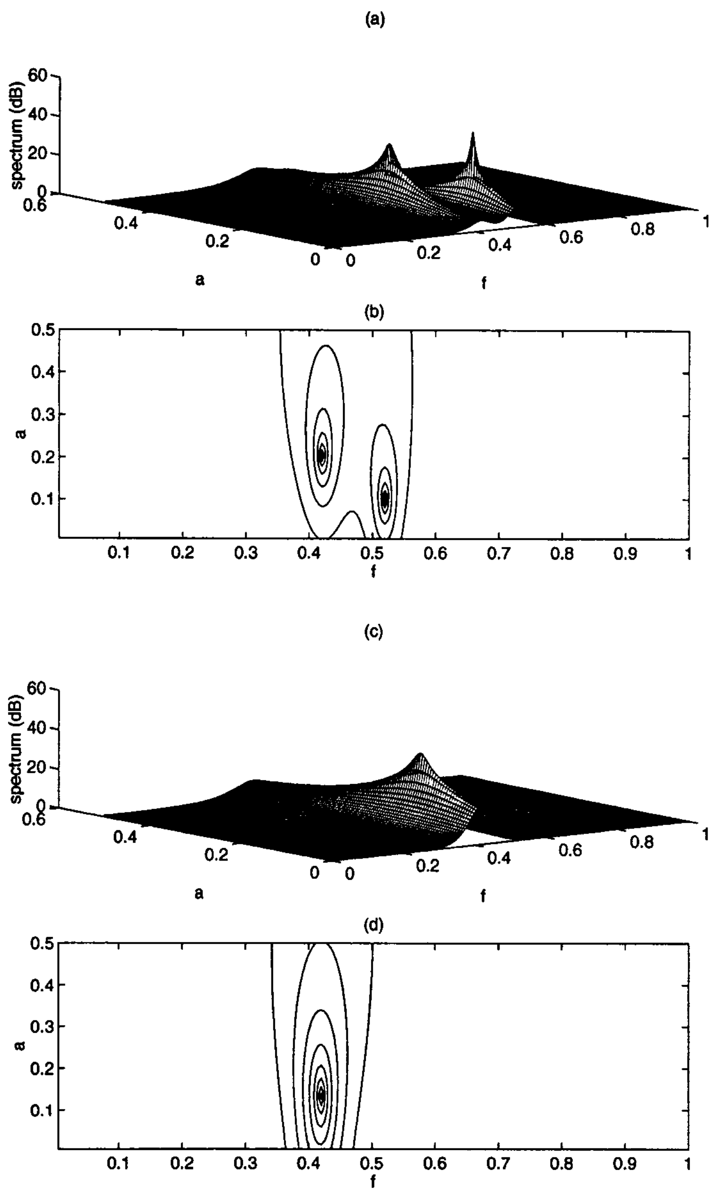


Figure 2: The MSE of (a) α and (b) ω vs α of KT, MKT and DMUSIC algorithm obtained in 100 trails when $s = -0.2 + \jmath 2\pi 0.42$ and $N = 24$



(a)
spectrum (dB)
60
40
20
0
0.6
0.4
0.2
0
a
0
0.2
0.4
0.6
0.8
1
f
(b)
0.5
0.4
0.3
0.2
0.1
a
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
f
(c)
spectrum (dB)
60
40
20
0
0.6
0.4
0.2
0
a
0
0.2
0.4
0.6
0.8
1
f
(d)
0.5
0.4
0.3
0.2
0.1
a
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
f

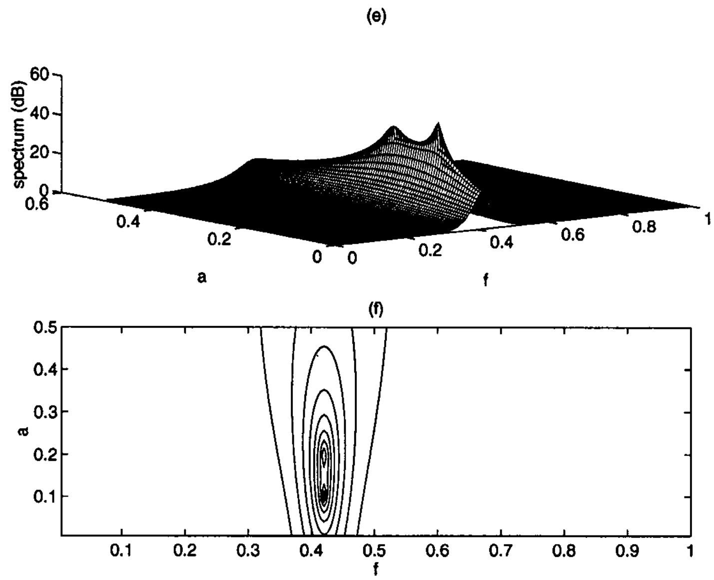


Figure 3: Spectrum and contour of DMUSIC algorithm (a) and (b) when $s_1 = -0.2 + j2\pi 0.42$, $s_2 = -0.1 + j2\pi 0.52$ and $SNR = 40dB$, (c) and (d) when $s_1 = -0.2 + j2\pi 0.42$, $s_2 = -0.1 + j2\pi 0.42$ and $SNR = 40dB$, (e) and (f) when $s_1 = -0.2 + j2\pi 0.42$, $s_2 = -0.1 + j2\pi 0.42$ and $SNR = 60dB$

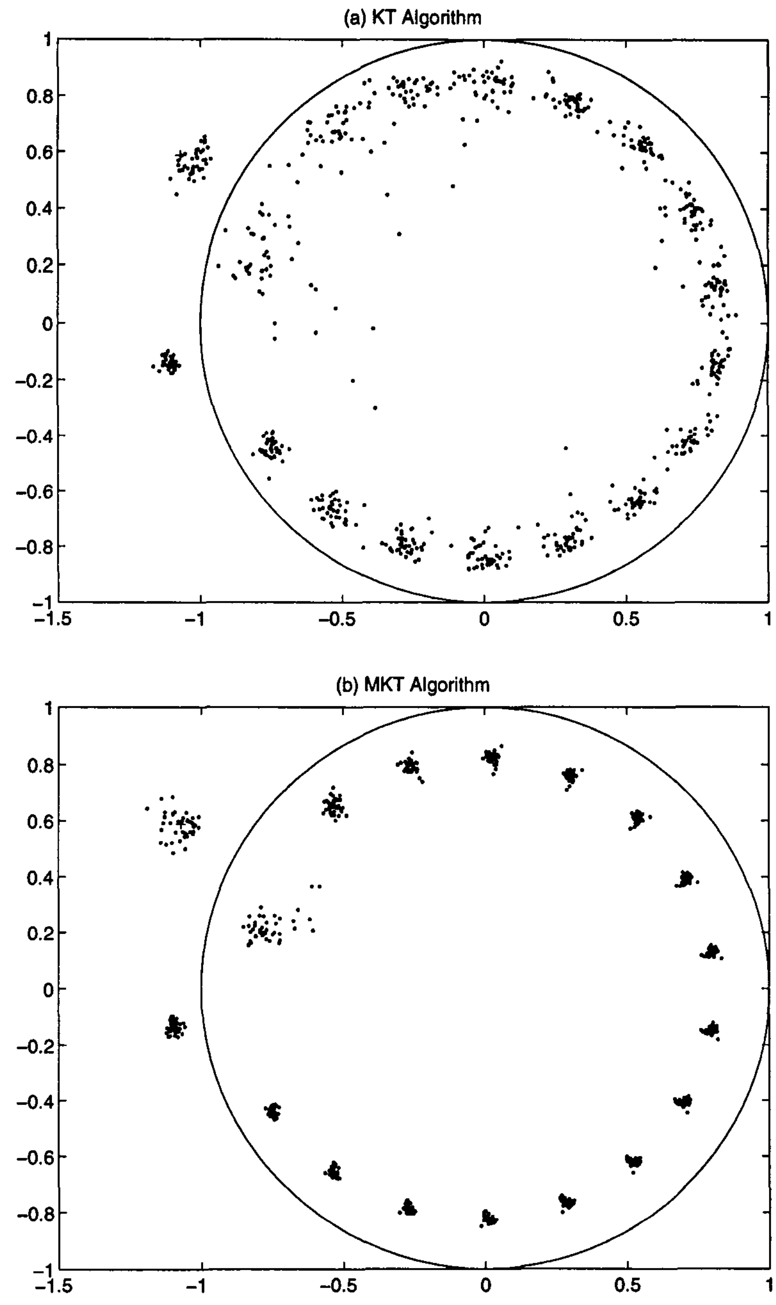
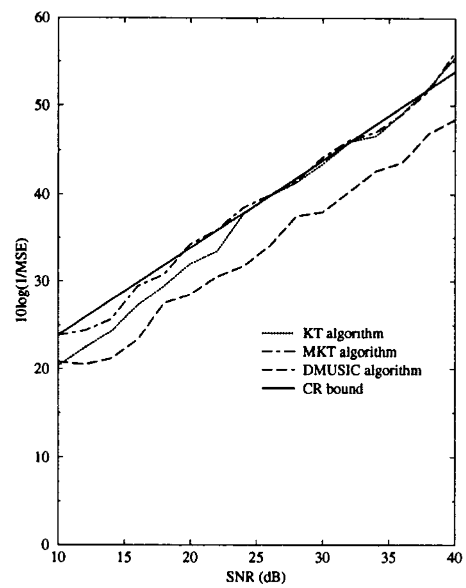


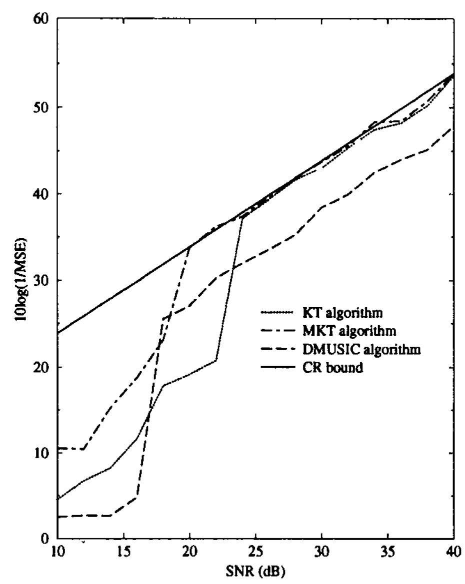
Figure 4: The zeros of $C(z)$ obtained in 40 trails of (a) KT algorithm and (b) MKT algorithm when $SNR = 15dB$

(a)



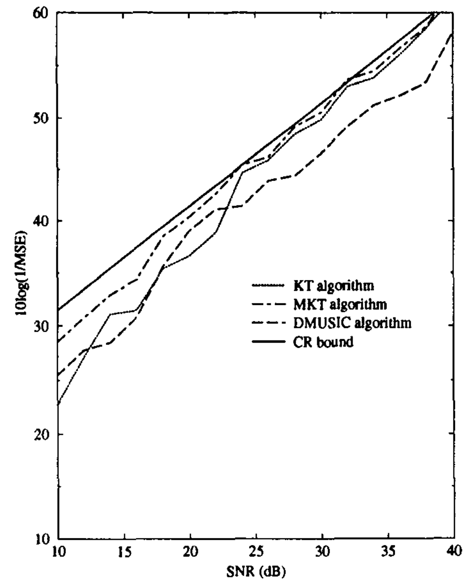
60
50
40
30
20
10
0
10
15
20
25
30
35
40
SNR (dB)
10log(1/MSE)
KT algorithm
MKT algorithm
DMUSIC algorithm
CR bound

(b)



60
50
40
30
20
10
0
10
15
20
25
30
35
40
SNR (dB)
10log(1/MSE)
KT algorithm
MKT algorithm
DMUSIC algorithm
CR bound

(c)



(d)

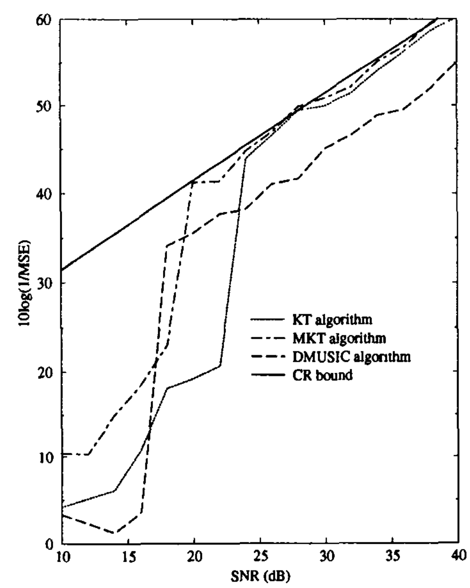


Figure 5: The MSE of (a) α_1, (b) ω_1, (c) α_2 and (d) ω_2 obtained in 100 trails of KT, MKT and DMUSIC algorithm when $s_1 = -0.2 + \jmath 2\pi 0.42$, $s_2 = -0.1 + \jmath 2\pi 0.52$ and $N = 25$