

## **UC Riverside**

### **UC Riverside Previously Published Works**

**Title**

Detection of cisoids using least square error function

**Permalink**

<https://escholarship.org/uc/item/5b95w3d7>

**Journal**

IEEE Transactions on Signal Processing, 45(6)

**ISSN**

1053-587X

**Authors**

Cheng, Qi  
Hua, Yingbo

**Publication Date**

1997-06-01

Peer reviewed

# Detection of Cisoids Using Least Square Error Function

Qi Cheng and Yingbo Hua, *Senior Member, IEEE*

**Abstract**—In this paper, we propose a new hypothesis testing method for detection of cisoids (complex sinusoids) from a single measurement of data. The testing is performed on the least square error. The least square error is shown to exhibit  $\chi^2$  distribution, which leads to an efficient threshold setup for the proposed method. The new method is a combined detection-estimation technique and provides improved performance over several existing techniques.

## I. INTRODUCTION

**D**ETECTION of cisoids from the following single measurement of data

$$\tilde{x}(n) = \sum_{i=1}^I v_i e^{j(n-1)\theta_i} + w(n), \quad n = 1, \dots, N \quad (1)$$

is commonly encountered in harmonic retrieval, sensor array processing, and nuclear magnetic resonance imaging. In (1),  $v_i$  ( $\neq 0$ ) and  $\theta_i$  ( $0 \leq \theta_i \leq 2\pi$ ) are the unknown complex amplitude and frequency associated with the  $i$ th complex sinusoid,  $I$  is the unknown number of cisoids, and  $w(1), \dots, w(N)$  are independent complex Gaussian noise with zero mean and the same variance  $\sigma^2$ . The real and imaginary parts of  $w(n)$  are also independent Gaussian variables with zero mean and variance  $\sigma^2/2$ .  $\sigma^2$  is assumed to be known (as in [5] and [6]). In practice,  $\sigma^2$  can be estimated from noise-only data, e.g., output of the same system without excitation. The estimation of the integer  $I$  is the focus of this paper.

Several methods have been developed for this detection problem. They include the methods by Kumaresan, Tufts, and Scharf (KTS) [1], Zhao, Krishnaiah, and Bai (ZKB) [2], Hwang and Chen (HC) [3], Djurić [4], Konstantinides and Yao (KY) [5], and Tufts and Shah (TS) [6]. The KTS method is based on a test of linear prediction errors but performs much worse than the recently developed TS method. The ZKB and HC methods apply the so-called efficient detection criterion but perform much worse than the Djurić method, especially for short data sequences. In the rest of this paper, the KTS, ZKB, and HC methods will not be further addressed. A short

Manuscript received November 27, 1995; revised December 9, 1996. This work was supported by the Australian Cooperative Research Center for Sensor Signal and Information Processing and the Australian Research Council. The associate editor coordinating the review of this paper and approving it for publication was Dr. Ananthram Swami.

Q. Cheng was with the Department of Electrical Engineering, University of Melbourne, Victoria 3052, Australia. He is now with the School of Electrical Engineering, Northern Territory University, Darwin, Australia (e-mail: chengq@darwin.ntu.edu.au).

Y. Hua is with the Department of Electrical Engineering, University of Melbourne, Victoria 3052, Australia (e-mail: yhua@ee.mu.oz.au).

Publisher Item Identifier S 1053-587X(97)04221-9.

review of the Djurić, KY, and TS methods will be shown in Section IV.

The new method that we will show in this paper applies a similar concept as used in the KY and TS methods. The KY and TS methods perform some hypothesis testing on the singular values of a Hankel data matrix. Instead of using the singular values, the new method performs a test on the least square error between the data and the reconstructed signal. As will be shown, the least square error asymptotically exhibits  $\chi^2$  distribution when the estimated number of cisoids is equal to the true one. This leads to an efficient threshold setup for testing on the least square error. As will be shown, the new method performs better than both the KY and TS methods. A comparison between the new method and the Djurić method will also be discussed.

In the next section, the new method is described. The theoretical basis for the new method is shown in Section III. In Section IV, the procedures of the Djurić, KY, and TS methods are first summarized, and their performances are then compared with that of the new method.

## II. DESCRIPTION OF THE NEW METHOD

Let  $H_k$  be a hypothesis that the number of cisoids is equal to  $k$ . Under  $H_k$ , the least square error of fitting the signal model to the measurement is

$$\|\tilde{\mathbf{x}} - \mathbf{A}_k \mathbf{b}\|^2 \quad (2)$$

where  $\tilde{\mathbf{x}} \triangleq [\tilde{x}(1), \dots, \tilde{x}(N)]^T$ ,  $\mathbf{b} = [b_1, \dots, b_k]^T$ ,  $\mathbf{A}_k \triangleq [\mathbf{a}_1, \dots, \mathbf{a}_k]$ , and  $\mathbf{a}_r \triangleq [1, e^{j\phi_r}, \dots, e^{j(N-1)\phi_r}]^T$  for  $r = 1, \dots, k$ .  $\phi_1, \dots, \phi_k$  are  $k$  unknown frequencies. It is easy to show that the least square amplitude estimates are

$$\begin{aligned} \hat{\mathbf{b}} &= \arg \min_{\mathbf{b}} \|\tilde{\mathbf{x}} - \mathbf{A}_k \mathbf{b}\|^2 \\ &= (\mathbf{A}_k^H \mathbf{A}_k)^{-1} \mathbf{A}_k^H \tilde{\mathbf{x}}. \end{aligned} \quad (3)$$

Substituting (3) into (2), we obtain the following least square error function in terms of frequencies:

$$\begin{aligned} L_k &\triangleq L(\phi_1, \dots, \phi_k) \\ &\triangleq \tilde{\mathbf{x}}^H [\mathbf{I}_N - \mathbf{A}_k (\mathbf{A}_k^H \mathbf{A}_k)^{-1} \mathbf{A}_k^H] \tilde{\mathbf{x}} \end{aligned} \quad (4)$$

$$L_0 \triangleq \tilde{\mathbf{x}}^H \tilde{\mathbf{x}}. \quad (5)$$

The least square frequency estimates  $\hat{\phi}_1^k, \dots, \hat{\phi}_k^k$  are defined as the coordinates of the smallest valley of the  $k$ -variate

function  $L_k$  (4) and calculated via a  $k$ -dimensional search. The minimum value of the least square error function  $L_k$ , i.e.,

$$\hat{L}_k \triangleq L_k(\hat{\phi}_1^k, \dots, \hat{\phi}_k^k) \quad (6)$$

will be called the least square error. Note that  $L_k$  is a monotonically decreasing function of the likelihood function of  $\tilde{\mathbf{x}}$  and is proportional to the maximum likelihood function of  $\tilde{\mathbf{x}}$  since the noise is white and Gaussian.

As will be shown by Theorem 1 in the next section, in the case of high SNR (or small  $\sigma^2$ ) and  $N > 3I/2$

$$\hat{L}_I \approx \frac{\sigma^2}{2} \chi_{2N-3I}^2 \quad (7)$$

where  $\approx$  denotes the first-order approximation (i.e., keeping the nonzero terms of the lowest order).<sup>1</sup> Equation (7) indicates that the least square error  $\hat{L}_I$  is  $\chi^2$  distributed with the  $2N - 3I$  degrees of freedom.  $2N$  is the number of (real) data points, and  $3I$  is the number of (real) unknowns (complex amplitude and frequency). It will also be shown by Theorem 2 in the next section that in the case of high SNR

$$\hat{L}_k \approx \text{positive value independent of noise} \quad (8)$$

for  $k = 0, \dots, I - 1$ . Equation (8) shows that the least square error  $\hat{L}_k$  for  $k = 0, \dots, I - 1$  is independent of noise. Therefore, for small  $\sigma^2$ ,  $\hat{L}_I$  is separable from  $\hat{L}_k$  for  $k = 0, \dots, I - 1$ . Based on this observation, our new method for finding  $I$  is as follows.

Define a threshold  $T_k$  such that

$$P(\mu_k \leq T_k) = \alpha \quad (9)$$

where  $\mu_k = (\sigma^2/2)\chi_{2N-3k}^2$  and  $\alpha$  ( $0 < \alpha < 1$ ) is a user chosen value of "confidence." Then, we estimate  $I$  by

$$\hat{I} = \min \{k | \hat{L}_k \leq T_k\}. \quad (10)$$

For the high SNR case (small  $\sigma^2$ ), the probability of underestimation is very small because of (8) and the fact that  $T_k$  is proportional to  $\sigma^2$  and the probability of correct detection

$$P(\hat{I} = I) \approx P(\hat{L}_I \leq T_I) = \alpha. \quad (11)$$

Note that  $T_k^2$  can be easily obtained using MATLAB or by just table lookup [9], as only the distribution of  $\chi_{2N-3k}^2$  is involved. In contrast, the computation of a threshold used in the TS method is a much more tedious task because the distribution of a weighted sum of  $\chi_2^2$  random variables is required there.

<sup>1</sup>This convention will be followed throughout this paper.

<sup>2</sup>The determination of  $T_k$  requires the knowledge of  $\sigma^2$ . If this information is not available, it can be estimated using the method proposed in [7] based on a different measurement or using another as suggested in [8] based on the same measurement.

### III. ASYMPTOTIC PROPERTIES OF THE LEAST SQUARE ERROR

In this section, we provide a number of properties of the least square error  $\hat{L}_k$  [which is defined in (6)]. These properties constitute the theoretical basis for the method described in Section II. For brevity, we need to show the following additional list of symbols:

Notations:

$$\bar{a}: \text{ the real part of } a \quad (12)$$

$$\tilde{a}: \text{ the imaginary part of } a \quad (13)$$

$$\mathbf{A} = \mathbf{A}_I |_{\phi_i=\theta_i, i=1, \dots, I} \quad (14)$$

$$\mathbf{P}_I^\perp = \mathbf{I}_N - \mathbf{A}_I(\mathbf{A}_I^H \mathbf{A}_I)^{-1} \mathbf{A}_I^H |_{\phi_i=\theta_i, i=1, \dots, I} \quad (15)$$

$$\mathbf{D} = \left[ \frac{\partial \mathbf{a}_1}{\partial \phi_1}, \dots, \frac{\partial \mathbf{a}_I}{\partial \phi_I} \right]_{\phi_i=\theta_i, i=1, \dots, I} \quad (16)$$

$$\mathbf{V} = \text{diag} [v_1, \dots, v_I] \quad (17)$$

$$\mathbf{w} = [w(1), \dots, w(N)]^T \quad (18)$$

$$\mathbf{G} = \mathbf{P}_I^\perp \mathbf{D} \mathbf{V} |_{\phi_i=\theta_i, i=1, \dots, I} \quad (19)$$

$$\mathbf{H} = \overline{\mathbf{G}^H \mathbf{G}}. \quad (20)$$

$\mathbf{H}$  is the Hessian matrix of  $L(\theta_1, \dots, \theta_I)$  [which is defined in (4)] for the noiseless case ( $\mathbf{w} = \mathbf{0}$ ). To ensure that  $(\theta_1, \dots, \theta_I)$  is a global minimum point of  $L_I$  for the noiseless case,  $\mathbf{H}$  should be nonsingular. Lemma A1 in Appendix A shows that if  $N \geq 3I/2$ ,  $\mathbf{H}$  is nonsingular.

The following Lemma 1 is concerned with the quadratic approximation of  $\hat{L}_I$ .

*Lemma 1:* If  $\sigma \ll \min \{|v_1|, \dots, |v_I|\}$  and  $N \geq 3I/2$ ,

$$\hat{L}_I \approx \boldsymbol{\mu}^T \{ \mathbf{I}_{2N} - \mathbf{B}_1(\mathbf{B}_1^T \mathbf{B}_1)^{-1} \mathbf{B}_1^T - \mathbf{B}_2(\mathbf{B}_2^T \mathbf{B}_2)^{-1} \mathbf{B}_2^T \} \boldsymbol{\mu} \quad (21)$$

where  $\boldsymbol{\mu} = [\bar{\mathbf{w}}^T, \tilde{\mathbf{w}}^T]^T$

$$\mathbf{B}_1 = \begin{bmatrix} \tilde{\mathbf{A}} & \tilde{\mathbf{A}} \\ \tilde{\mathbf{A}} & -\tilde{\mathbf{A}} \end{bmatrix}_{2N \times 2I} \quad (22)$$

$$\mathbf{B}_2 = \begin{bmatrix} \tilde{\mathbf{G}} \\ \tilde{\mathbf{G}} \end{bmatrix}_{2N \times I} \quad (23)$$

*Proof:* See Appendix A.

In the proof of Lemma 1, it is shown that

$$\boldsymbol{\mu}^T \{ \mathbf{I}_{2N} - \mathbf{B}_1(\mathbf{B}_1^T \mathbf{B}_1)^{-1} \mathbf{B}_1^T \} \boldsymbol{\mu} \approx L(\theta_1, \dots, \theta_I) \quad (24)$$

is the (first-order) fitting error given noisy data and the true frequencies, and

$$\boldsymbol{\mu}^T \{ \mathbf{B}_2(\mathbf{B}_2^T \mathbf{B}_2)^{-1} \mathbf{B}_2^T \} \boldsymbol{\mu} \approx \mathbf{x}^H \mathbf{P}_I^\perp \mathbf{x}^H |_{\phi_i=\hat{\theta}_i^I, i=1, \dots, I} \quad (25)$$

is the (first-order) fitting error given noise-free data and the least square frequency estimates. Thus, the fitting error (21) can be interpreted as the difference between these two kinds of errors.

It is shown in Lemma B2 that  $\mathbf{B}_1$  and  $\mathbf{B}_2$  are orthogonal to each other. Then, we have

$$\begin{aligned} & \mathbf{I}_{2N} - \mathbf{B}_1(\mathbf{B}_1^T \mathbf{B}_1)^{-1} \mathbf{B}_1^T - \mathbf{B}_2(\mathbf{B}_2^T \mathbf{B}_2)^{-1} \mathbf{B}_2^T \\ &= \mathbf{I}_{2N} - \begin{bmatrix} \mathbf{B}_1 & \mathbf{B}_2 \end{bmatrix} \left\{ \begin{bmatrix} \mathbf{B}_1^T \\ \mathbf{B}_2^T \end{bmatrix} \begin{bmatrix} \mathbf{B}_1 & \mathbf{B}_2 \end{bmatrix} \right\}^{-1} \begin{bmatrix} \mathbf{B}_1^T \\ \mathbf{B}_2^T \end{bmatrix} \\ &\triangleq \mathbf{P}_{\mathbf{B}_1 \mathbf{B}_2}^\perp. \end{aligned} \quad (26)$$

When  $N = 3I/2$ ,  $\mathbf{P}_{\mathbf{B}_1\mathbf{B}_2}^\perp = \mathbf{0}$ , and then, the first-order approximation of  $\hat{L}_I$  is zero, and the two kinds of errors cancel each other.

Let  $\mathbf{u}_1, \dots, \mathbf{u}_{2N-3I}$  be real unitary eigenvectors of  $\mathbf{P}_{\mathbf{B}_1\mathbf{B}_2}^\perp$  corresponding to its eigenvalue 1,  $\mathbf{U}_0 = [\mathbf{u}_1, \dots, \mathbf{u}_{2N-3I}]$ , and  $\xi_l = (2/\sigma^2)\boldsymbol{\mu}^T \mathbf{u}_l$ . Then, Lemma 1 leads to

$$\hat{L}_I \approx \boldsymbol{\mu}^T \mathbf{U}_0 \mathbf{U}_0^T \boldsymbol{\mu} = \frac{\sigma^2}{2} \sum_{l=1}^{2N-3I} \xi_l^2. \quad (27)$$

Since  $\xi_1, \dots, \xi_{2N-3I}$  are independent Gaussian random variables with zero mean and unit variance, the following theorem holds.

**Theorem 1:**  $\hat{L}_I \approx (\sigma^2/2)\chi_{2N-3I}^2$  when  $\sigma \ll \min\{|v_1|, \dots, |v_I|\}$  and  $N > 3I/2$ .

**Theorem 2:**  $\hat{L}_k \approx$  positive value independent of noise for  $k = 0, 1, \dots, I-1$  when  $\sigma \ll \min\{|v_1|, \dots, |v_I|\}$  and  $N > 3I/2$ .

*Proof:* Since  $\mathbf{x} = \mathbf{A}\mathbf{v}$ , where  $|v_i| \neq 0$  for  $i = 1, 2, \dots, I$ , then for any  $\phi_1, \dots, \phi_k$

$$\mathbf{x}^H [\mathbf{I}_N - \mathbf{A}_k (\mathbf{A}_k^H \mathbf{A}_k)^{-1} \mathbf{A}_k^H] \mathbf{x} > 0. \quad (28)$$

Since the noise variance is small, for any (deterministic)  $\phi_1, \dots, \phi_k$ , we know that

$$\begin{aligned} L(\phi_1, \dots, \phi_k) &= \mathbf{x}^H [\mathbf{I}_N - \mathbf{A}_k (\mathbf{A}_k^H \mathbf{A}_k)^{-1} \mathbf{A}_k^H] \mathbf{x} \\ &\quad + 2\text{Re} \{ \mathbf{x}^H [\mathbf{I}_N - \mathbf{A}_k (\mathbf{A}_k^H \mathbf{A}_k)^{-1} \mathbf{A}_k^H] \mathbf{w} \} \\ &\quad + \mathbf{w}^H [\mathbf{I}_N - \mathbf{A}_k (\mathbf{A}_k^H \mathbf{A}_k)^{-1} \mathbf{A}_k^H] \mathbf{w} \\ &\approx \mathbf{x}^H [\mathbf{I}_N - \mathbf{A}_k (\mathbf{A}_k^H \mathbf{A}_k)^{-1} \mathbf{A}_k^H] \mathbf{x} \end{aligned} \quad (29)$$

$$\begin{aligned} \hat{L}_k &= \min_{\phi_1, \dots, \phi_k} L(\phi_1, \dots, \phi_k) \\ &\approx \min_{\phi_1, \dots, \phi_k} \mathbf{x}^H [\mathbf{I}_N - \mathbf{A}_k (\mathbf{A}_k^H \mathbf{A}_k)^{-1} \mathbf{A}_k^H] \mathbf{x} \\ &> 0. \end{aligned} \quad (30)$$

The expectation of  $\hat{L}_k$  is generally a decreasing function of  $k$ . The statistical properties of  $\hat{L}_k$  for  $k > I$  seem hard to obtain as  $\hat{L}_k$  involves spurious frequency estimates. The lack of this information however does not significantly affect the procedure of the new method, when SNR is high and  $\alpha$  is close to 1, as can be seen from (11).

#### IV. SIMULATION

In this section, we will compare the performance of the new method shown in Section II to those of the Djurić [4], KY [5], and TS [6] methods. For the reader's convenience, these three methods are briefly described below.

The KY method [5] estimates the number of frequencies by determining a number  $\hat{I}_{KY}$  of "effective singular values," i.e.

$$\hat{I}_{KY} = \{k | \lambda_k > \epsilon_{KY} \geq \lambda_{k+1}\} \quad (31)$$

where  $\lambda_1, \lambda_2, \dots, \lambda_K$  are singular values of an  $K \times (N - K + 1)$  data matrix  $\tilde{\mathbf{X}}_e$  in descending order with

$$\tilde{\mathbf{X}}_e = \begin{bmatrix} \tilde{x}(1) & \tilde{x}(2) & \cdots & \tilde{x}(N-K+1) \\ \tilde{x}(2) & \tilde{x}(3) & \cdots & \tilde{x}(N-K+2) \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{x}(K) & \tilde{x}(K+1) & \cdots & \tilde{x}(N) \end{bmatrix} \quad (32)$$

and  $\epsilon_{KY}^3$  is given by

$$\epsilon_{KY} = \sqrt{c(N-K+1)}. \quad (33)$$

$c$  is determined by  $P\{S^2 \leq c\} = \alpha$ , where  $S^2 = |w(1)|^2 + \dots + |w(K)|^2 = 0.5\sigma^2\chi_{2K}^2$ . In simulation,  $K = 17$ .

The TS method [6] works on the forward-backward version of  $\tilde{\mathbf{X}}_e$ , i.e.,  $[\tilde{\mathbf{X}}_e, \mathbf{P}\tilde{\mathbf{X}}_e^*]$ , where  $\mathbf{P}$  is the permutation matrix obtained by reversing the column order of the identity matrix. The estimate of the number  $I$  of frequencies given by the TS method is obtained as

$$\hat{I}_{TS} = \min \left\{ k \left| \sum_{i=k+1}^K \mu_i^2 \leq \epsilon_{TS} \right. \right\} \quad (34)$$

where  $\mu_1, \mu_2, \dots, \mu_K$  are singular values of  $[\tilde{\mathbf{X}}_e, \mathbf{P}\tilde{\mathbf{X}}_e^*]$  in descending order, and  $\epsilon_{TS}$  is chosen such that  $P(S_0 \leq \epsilon_{TS}) = \alpha$  with  $S_0 = \sum_{k=1}^{N-K} k[|w(k)|^2 + |w(K-k+1)|^2] + (N-K+1) \sum_{k=N-K+1}^K |w(k)|^2$  (see [6, eq. (11)] for details).

The Djurić method [4] is based on the Bayesian principle. It computes  $\text{BAY}(k) = (N-L) \log \hat{L}_k + (5k/2) \log(N/L)$  and then estimates the number of frequencies by

$$\hat{I}_{BAY} = \arg \min_k \{\text{BAY}(k)\} \quad (35)$$

where  $0 \leq k \leq \bar{I}$ ,  $\bar{I}$  is the maximum possible value of  $I$ , and  $L$  is another user-chosen parameter [4].

In our simulation, the following signal parameters were used:

$$\begin{aligned} I &= 2, \\ N &= 25, \\ \theta_1/2\pi &= 0.5, \\ \theta_2/2\pi &= 0.52, \\ v_1 &= 1, \\ v_2 &= e^{j\pi/4}. \end{aligned}$$

SNR is defined as  $-10 \log \sigma^2$ . The probability of correct detection was computed from 3000 runs of simulations.

Fig. 1 shows the probability of correct detection as a function of SNR using the new method, the KY method, and the TS method. As seen from this figure, the KY method gives a very low probability of correct detection even for SNR at 14 dB, the TS method significantly improves the performance of the KY method for SNR larger than 10 dB, and the new method yields almost perfect detection for SNR as low as 4 dB. This is because bounds  $\epsilon_{KY}$  given by (33) and  $\epsilon_{TS}$  in [6] were so large that  $\epsilon_{KY} \geq \lambda_k$  and  $\epsilon_{TS} \geq \sum_{i=k+1}^{17} \mu_i^2$  for some  $k \leq I-1$  at medium SNR, whereas the threshold given by the new method was small enough to separate  $\hat{L}_I$  from  $\hat{L}_k$  for  $k = 1, \dots, I-1$ .

Fig. 2 presents a comparison of the new method with the Djurić method for  $\bar{I} = 3$  and  $\bar{I} = 9$ . It is not surprising to see that the Djurić method performed better for small  $\bar{I}$  (but larger than  $I$ ) than for large  $\bar{I}$ . For  $\bar{I} = 9$ , the Djurić method performed worse than the new method.

The Djurić method requires more computations than the new method because the Djurić method needs to compute  $\hat{L}_k$  for

<sup>3</sup>According to [5], this bound is the tightest among the three bounds derived therein for correlated noise in  $\tilde{\mathbf{X}}_e$ .

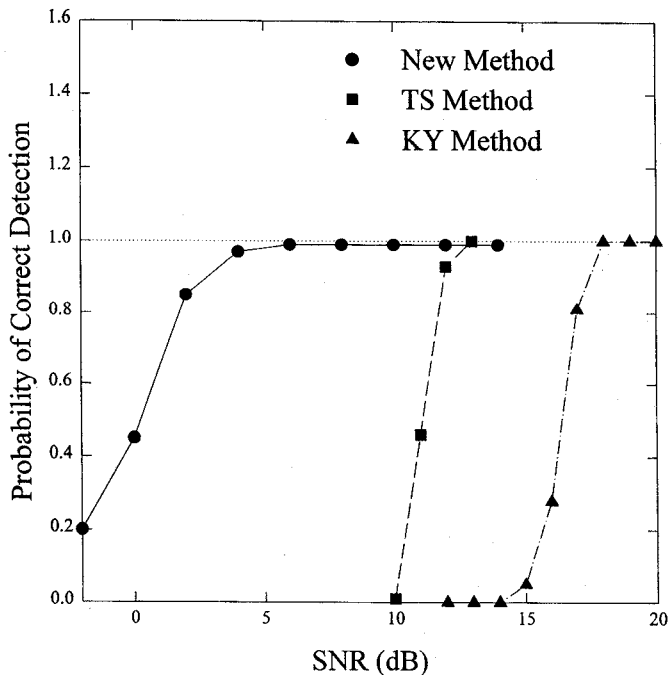


Fig. 1. Performance comparison of the new method, the TS method, and the KY method.  $\alpha = 0.99$  for all.

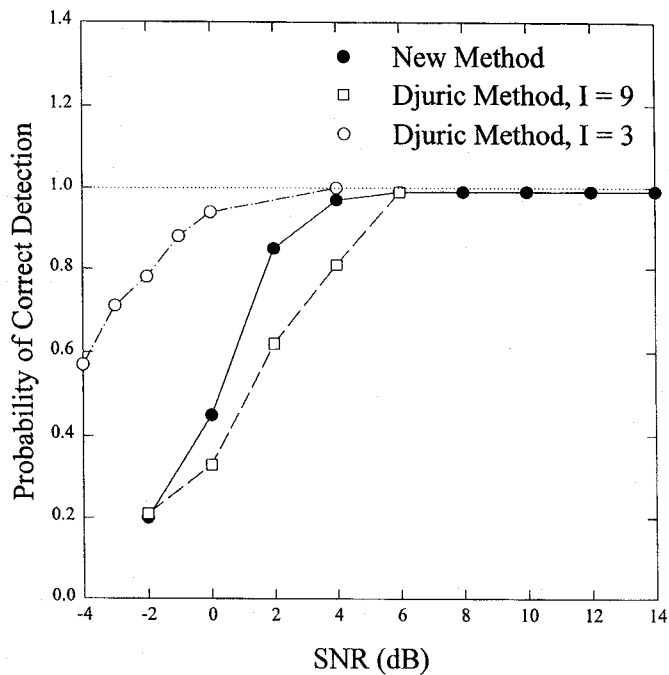


Fig. 2. Performance comparison of the new method and the Djurić method. For the new method,  $\alpha = 0.99$ , and for the Djurić method,  $L = 4$ .

$k = 0, 1, \dots, \bar{I}$ , whereas the new method computes  $\hat{L}_k$  for  $k = 0, 1, \dots, \hat{I}$ . Note that  $\hat{I}$  is, in general, smaller than  $\bar{I}$ .

Furthermore, the Djurić method requires the choice of  $L$ , for which there is no theoretical guideline. An advantage of the Djurić method is, however, that it does not assume the knowledge of  $\sigma^2$ .

In the above simulations, the probabilities of underestimation and overestimation using the new method are shown in Fig. 3(a) and (b), respectively. It can be seen that when

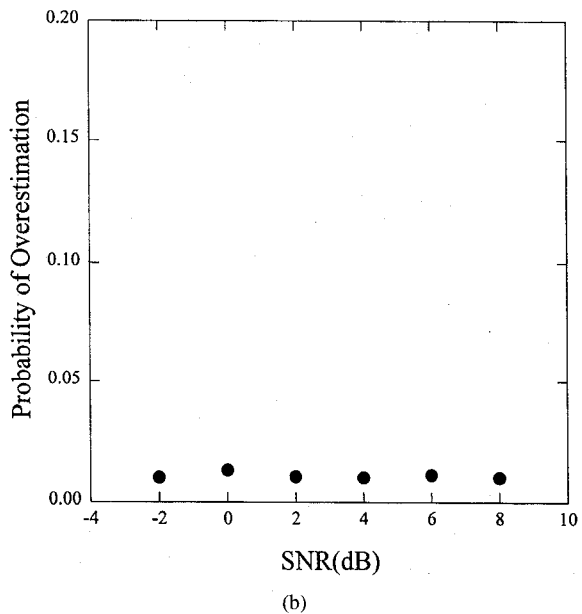
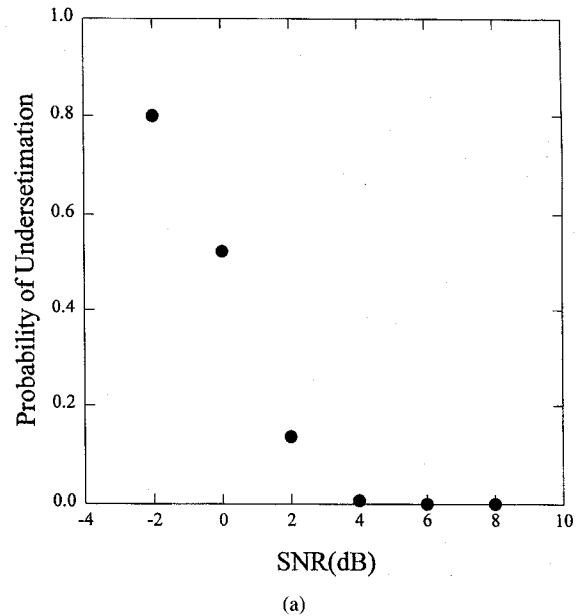


Fig. 3. (a) Probability of underestimation versus SNR for  $\alpha = 0.99$ . (b) Probability of overestimation versus SNR for  $\alpha = 0.99$ .

$\alpha = 0.99$ , the probability of underestimation is zero at SNR larger than 4 dB, and the probability of overestimation is around 0.01.

The relative phase  $[\triangleq \arg(v_2/v_1)]$  has some effect on the detection performance of the new method when SNR is relatively low. Fig. 4 shows the probability of correct detection as a function of the relative phase  $\psi$  for SNR = 4 dB (low), where  $v_2 = e^{j\psi}$ , and all other signal parameters remain the same as in the previous simulations. The poor performance of the new method for  $\psi \in [270^\circ, 315^\circ]$  is due to the fact that the high SNR condition is violated, and the least square error is no longer  $\chi^2$  distributed.

Finally, it is important to note that the new method does not necessarily depend on the least square frequency estimates. In fact, the computations required by the new method can be drastically reduced if the least square (optimum) frequency

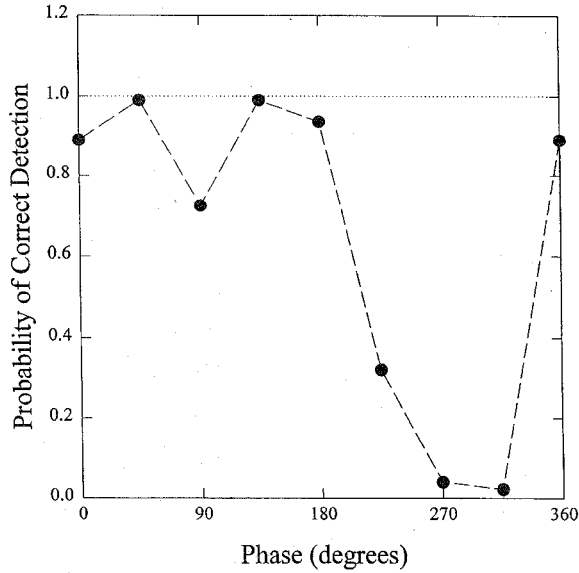


Fig. 4. Probability of correct detection versus phase for SNR = 4 dB.

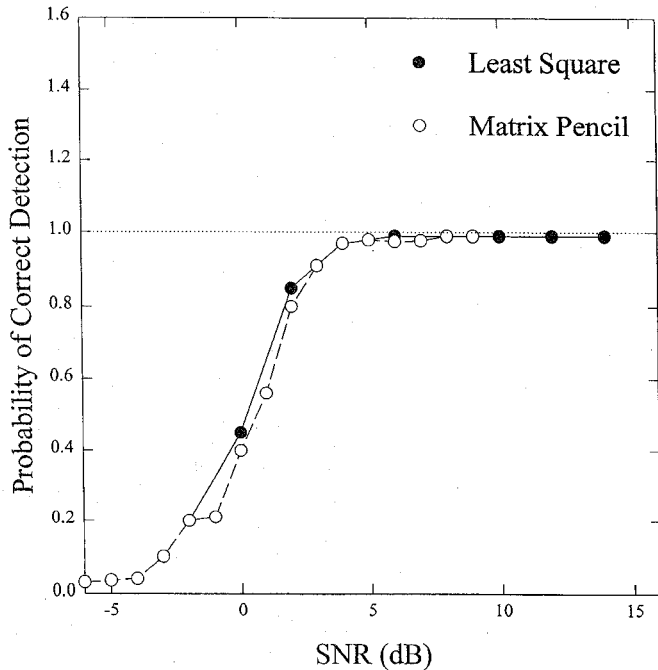


Fig. 5. Performance comparison of the new method using, respectively, the least square and matrix pencil frequency estimates.  $\alpha = 0.99$ .

estimates are approximated by using near-optimum methods such as the matrix pencil method [11], [12]. Fig. 5 shows that the detection performances using the least square frequency estimates and the matrix pencil frequency estimates are nearly the same.

#### APPENDIX A PROOF OF LEMMA 1

*Lemma A1:*  $N \geq 3I/2$  is a sufficient condition such that  $\mathbf{H}$  is nonsingular.

*Proof:* Note that

$$\mathbf{H} = \overline{\mathbf{G}^H \mathbf{G}} = [\tilde{\mathbf{G}}^T \quad \tilde{\mathbf{G}}^T] \begin{bmatrix} \mathbf{G} \\ \tilde{\mathbf{G}} \end{bmatrix} = \mathbf{B}_2^T \mathbf{B}_2 \quad (36)$$

where  $\mathbf{H}$ ,  $\mathbf{G}$ , and  $\mathbf{B}_2$  are defined in (20), (19), and (23), respectively. Lemma B1 will show that  $\mathbf{B}_2$  has the rank  $I$  when  $N \geq 3I/2$ , and hence,  $\mathbf{H}$  is nonsingular.  $\spadesuit$

*Lemma A2:* If  $\sigma \ll \min \{|v_1|, \dots, |v_I|\}$ , the least square frequency estimate errors are given as

$$\begin{bmatrix} \Delta\theta_1 \\ \vdots \\ \Delta\theta_I \end{bmatrix} \approx \mathbf{H}^{-1} \overline{\mathbf{G}^H \mathbf{w}} \quad (37)$$

where  $\Delta\theta_i \triangleq \hat{\phi}_i^I - \theta_i$  for  $i = 1, \dots, I$ ,  $\mathbf{H}$ ,  $\mathbf{G}$ , and  $\mathbf{w}$  are defined in (20), (19), and (18), respectively.

*Proof of Lemma A2:* For the noiseless case ( $\mathbf{w} = \mathbf{0}$ ),  $[\theta_1, \theta_2, \dots, \theta_I]$  is a global minimum point of  $\hat{L}_I$ . For small noise variance, the global minimum point of  $\hat{L}_I$  will be close to  $[\theta_1, \theta_2, \dots, \theta_I]$ . Expanding  $\hat{L}_I$  around the true frequency values in the Taylor series and retaining the zero and first-order terms result in

$$\begin{bmatrix} \Delta\theta_1 \\ \vdots \\ \Delta\theta_I \end{bmatrix} \approx - \begin{bmatrix} \frac{\partial^2 L_I}{\partial \phi_1 \partial \phi_1} & \cdots & \frac{\partial^2 L_I}{\partial \phi_1 \partial \phi_I} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 L_I}{\partial \phi_I \partial \phi_1} & \cdots & \frac{\partial^2 L_I}{\partial \phi_I \partial \phi_I} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial L_I}{\partial \phi_1} \\ \vdots \\ \frac{\partial L_I}{\partial \phi_I} \end{bmatrix} \Big|_{\phi_l = \theta_l, l=1, \dots, I} \quad (38)$$

Observe that

$$\begin{aligned} \frac{\partial L_I}{\partial \phi_l} \Big|_{\phi_l = \theta_l, l=1, \dots, I} &= -\tilde{\mathbf{x}}^H \frac{\partial \mathbf{P}_I}{\partial \phi_l} \tilde{\mathbf{x}} \Big|_{\phi_l = \theta_l, l=1, \dots, I} \\ &= \left[ -\mathbf{x}^H \frac{\partial \mathbf{P}_I}{\partial \phi_l} \mathbf{x} - 2\mathbf{x}^H \frac{\partial \mathbf{P}_I}{\partial \phi_l} \mathbf{w} - \mathbf{w}^H \frac{\partial \mathbf{P}_I}{\partial \phi_l} \mathbf{w} \right] \Big|_{\phi_l = \theta_l, l=1, \dots, I} \end{aligned} \quad (39)$$

$$\begin{aligned} \frac{\partial^2 L_I}{\partial \phi_i \partial \phi_j} \Big|_{\phi_l = \theta_l, l=1, \dots, I} &= -\tilde{\mathbf{x}}^H \frac{\partial^2 \mathbf{P}_I}{\partial \phi_i \partial \phi_j} \tilde{\mathbf{x}} \Big|_{\phi_l = \theta_l, l=1, \dots, I} \\ &\approx -\mathbf{x}^H \frac{\partial^2 \mathbf{P}_I}{\partial \phi_i \partial \phi_j} \mathbf{x} \Big|_{\phi_l = \theta_l, l=1, \dots, I} \end{aligned} \quad (40)$$

where

$$\mathbf{x} = \tilde{\mathbf{x}} - \mathbf{w} \quad (41)$$

$$\mathbf{P}_I = \mathbf{A}_I (\mathbf{A}_I^H \mathbf{A}_I)^{-1} \mathbf{A}_I^H. \quad (42)$$

By Theorem 4 of [10, p. 153], we know that

$$\begin{aligned} \frac{\partial \mathbf{P}_I}{\partial \phi_l} &= \left[ \mathbf{A}_I (\mathbf{A}_I^H \mathbf{A}_I)^{-1} \frac{\partial \mathbf{A}_I^H}{\partial \phi_l} \mathbf{P}_I^T \right. \\ &\quad \left. + \mathbf{P}_I^T \frac{\partial \mathbf{A}_I}{\partial \phi_l} (\mathbf{A}_I^H \mathbf{A}_I)^{-1} \mathbf{A}_I^H \right] \Big|_{\phi_l = \theta_l, l=1, \dots, I}. \end{aligned} \quad (43)$$

It is easy to prove that

$$\mathbf{x}^H \frac{\partial \mathbf{P}_I}{\partial \phi_l} \mathbf{x} \Big|_{\phi_l = \theta_l, l=1, \dots, I} = 0. \quad (44)$$

We then have

$$\begin{aligned} \frac{\partial L_I}{\partial \phi_l} \Big|_{\phi_l = \theta_l, l=1, \dots, I} &\approx -2 \mathbf{v}^H \frac{\partial \mathbf{A}_I^H}{\partial \phi_l} \mathbf{P}_I^\perp \mathbf{w} \Big|_{\phi_l = \theta_l, l=1, \dots, I} \\ &= -2 v_l^* \frac{\partial \mathbf{a}_l^H}{\partial \phi_l} \mathbf{P}_I^\perp \mathbf{w} \Big|_{\phi_l = \theta_l, l=1, \dots, I} \end{aligned} \quad (45)$$

or, in vector form

$$\begin{bmatrix} \frac{\partial L_I}{\partial \phi_1} \\ \vdots \\ \frac{\partial L_I}{\partial \phi_I} \end{bmatrix} \Big|_{\phi_l = \theta_l, l=1, \dots, I} \approx -2 \overline{\mathbf{G}^H \mathbf{w}}. \quad (46)$$

We also know that

$$\frac{\partial^2 \mathbf{P}_I}{\partial \phi_i \partial \phi_j} = \mathbf{R} + \mathbf{R}^H \quad (47)$$

where

$$\begin{aligned} \mathbf{R} = &\left\{ -\mathbf{P}_I^\perp \frac{\partial \mathbf{A}_I}{\partial \phi_i} (\mathbf{A}_I^H \mathbf{A}_I)^{-1} \mathbf{A}_I^H \frac{\partial \mathbf{A}}{\partial \phi_j} (\mathbf{A}_I^H \mathbf{A}_I)^{-1} \mathbf{A}_I^H \right. \\ &- \mathbf{A}_I (\mathbf{A}_I^H \mathbf{A}_I)^{-1} \frac{\partial \mathbf{A}_I^H}{\partial \phi_i} \mathbf{P}_I^\perp \frac{\partial \mathbf{A}}{\partial \phi_j} (\mathbf{A}_I^H \mathbf{A}_I)^{-1} \mathbf{A}_I^H \\ &+ \mathbf{P}_I^\perp \frac{\partial^2 \mathbf{A}_I}{\partial \phi_i \partial \phi_j} (\mathbf{A}_I^H \mathbf{A}_I)^{-1} \mathbf{A}_I^H \\ &\left. + \mathbf{P}_I^\perp \frac{\partial \mathbf{A}_I}{\partial \phi_j} \frac{\partial [(\mathbf{A}_I^H \mathbf{A}_I)^{-1} \mathbf{A}_I^H]}{\partial \phi_i} \right\} \Big|_{\phi_l = \theta_l, l=1, \dots, I} \end{aligned} \quad (48)$$

We again have

$$\frac{\partial^2 L_I}{\partial \phi_i \partial \phi_j} \Big|_{\phi_l = \theta_l, l=1, \dots, I} \approx 2 v_i^* v_j \frac{\partial \mathbf{a}_i^H}{\partial \phi_i} \mathbf{P}_I^\perp \frac{\partial \mathbf{a}_j}{\partial \phi_j} \Big|_{\phi_l = \theta_l, l=1, \dots, I} \quad (49)$$

or, in matrix form

$$\begin{aligned} &\begin{bmatrix} \frac{\partial^2 L_I}{\partial \phi_1 \partial \phi_1} & \dots & \frac{\partial^2 L_I}{\partial \phi_I \partial \phi_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 L_I}{\partial \phi_1 \partial \phi_I} & \dots & \frac{\partial^2 L_I}{\partial \phi_I \partial \phi_I} \end{bmatrix} \Big|_{\phi_l = \theta_l, l=1, \dots, I} \\ &\approx 2 \overline{\mathbf{V}^H \mathbf{D}^H \mathbf{P}_I^\perp \mathbf{D} \mathbf{V}} \\ &= 2 \overline{\mathbf{G}^H \mathbf{G}} = 2\mathbf{H}. \end{aligned} \quad (50)$$

Plugging (46) and (50) into (38) leads to the statement of the Lemma.  $\blacklozenge$

*Proof of Lemma 1:* Theorem 4 of [10, p. 153] shows another difference form

$$\Delta \mathbf{P}_I = \mathbf{A} (\mathbf{A}^H \mathbf{A})^{-1} \Delta \mathbf{A}^H \mathbf{P}_I^\perp + \mathbf{P}_I^\perp \Delta \mathbf{A} (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \quad (51)$$

where  $\mathbf{A}$  is defined in (14)

$$\begin{aligned} \Delta \mathbf{P}_I = &\mathbf{A}_I (\mathbf{A}_I^H \mathbf{A}_I)^{-1} \mathbf{A}_I^H \Big|_{\phi_l = \hat{\phi}_l^I, l=1, \dots, I} \\ &- \mathbf{A}_I (\mathbf{A}_I^H \mathbf{A}_I)^{-1} \mathbf{A}_I^H \Big|_{\phi_l = \theta_l, l=1, \dots, I} \end{aligned} \quad (52)$$

$$\begin{aligned} \Delta \mathbf{A} = &\mathbf{A}_I \Big|_{\phi_l = \hat{\phi}_l^I, l=1, \dots, I} - \mathbf{A} \\ \approx &\mathbf{D} \begin{bmatrix} \Delta \theta_1 & & \\ & \ddots & \\ & & \Delta \theta_I \end{bmatrix}. \end{aligned} \quad (53)$$

Making use of (51) and the high SNR condition, we have

$$\begin{aligned} \tilde{\mathbf{x}}^H \hat{\mathbf{P}}_I^\perp &\approx \mathbf{w}^H \mathbf{P}_I^\perp - \mathbf{x}^H \Delta \mathbf{P}_I^\perp \\ &= \mathbf{w}^H \mathbf{P}_I^\perp - \mathbf{x}^H \mathbf{A} (\mathbf{A}^H \mathbf{A})^{-1} \Delta \mathbf{A}^H \mathbf{P}_I^\perp \\ &= \mathbf{w}^H \mathbf{P}_I^\perp - \mathbf{v}^H \Delta \mathbf{A}^H \mathbf{P}_I^\perp \\ &= \mathbf{w}^H \mathbf{P}_I^\perp - \left( \mathbf{P}_I^\perp \mathbf{D} \mathbf{V} \begin{bmatrix} \Delta \theta_1 \\ \dots \\ \Delta \theta_I \end{bmatrix} \right)^H \\ &= \mathbf{w}^H \mathbf{P}_I^\perp - (\mathbf{G} \mathbf{H}^{-1} \overline{\mathbf{G}^H \mathbf{w}})^H \\ &= \mathbf{w}^H \mathbf{P}_I^\perp - \overline{\mathbf{w}^H \mathbf{G} \mathbf{H}^{-1} \mathbf{G}^H} \\ &= \boldsymbol{\mu}^T \begin{bmatrix} \mathbf{P}_I^\perp - \overline{\mathbf{G} \mathbf{H}^{-1} \mathbf{G}^H} \\ -j \mathbf{P}_I^\perp - \tilde{\mathbf{G}} \mathbf{H}^{-1} \mathbf{G}^H \end{bmatrix}. \end{aligned} \quad (54)$$

Let

$$\mathbf{E} = \begin{bmatrix} \mathbf{P}_I^\perp - \overline{\mathbf{G} \mathbf{H}^{-1} \mathbf{G}^H} \\ -j \mathbf{P}_I^\perp - \tilde{\mathbf{G}} \mathbf{H}^{-1} \mathbf{G}^H \end{bmatrix}. \quad (55)$$

Then, we have (56), shown at the bottom of the page, where we used  $\mathbf{P}_I^\perp \mathbf{G} = \mathbf{G}$ . Using  $\mathbf{H} = \overline{\mathbf{G}^H \mathbf{G}}$  [from (20)], we have

$$\begin{aligned} \overline{\mathbf{E} \mathbf{E}^H} &= \begin{bmatrix} \mathbf{P}_I^\perp - \overline{\mathbf{G} \mathbf{H}^{-1} \mathbf{G}^H} & -\tilde{\mathbf{P}}_I^\perp - \overline{\mathbf{G} \mathbf{H}^{-1} \tilde{\mathbf{G}}^T} \\ \mathbf{P}_I^\perp - \tilde{\mathbf{G}} \mathbf{H}^{-1} \mathbf{G}^T & \mathbf{P}_I^\perp - \overline{\mathbf{G} \mathbf{H}^{-1} \tilde{\mathbf{G}}^T} \end{bmatrix}_{2N \times 2N} \\ &= \mathbf{I}_{2N} - \begin{bmatrix} \mathbf{A} \\ -j \mathbf{A} \end{bmatrix} (\mathbf{A}^H \mathbf{A})^{-1} [\mathbf{A}^H \quad j \mathbf{A}^H] \\ &\quad - \begin{bmatrix} \mathbf{G} \\ \tilde{\mathbf{G}} \end{bmatrix} \mathbf{H}^{-1} [\mathbf{G}^T \quad \tilde{\mathbf{G}}^T] \\ &= \mathbf{I}_{2N} - \mathbf{B}_1 \mathbf{F}_1 \mathbf{B}_1^T - \mathbf{B}_2 \mathbf{H}^{-1} \mathbf{B}_2^T \end{aligned} \quad (57)$$

where  $\mathbf{B}_1$  is defined in (22),  $\mathbf{B}_2$  is defined in (23), and

$$\mathbf{F}_1 = \begin{bmatrix} (\mathbf{A}^H \mathbf{A})^{-1} & (\mathbf{A}^H \mathbf{A})^{-1} \\ -(\mathbf{A}^H \mathbf{A})^{-1} & (\mathbf{A}^H \mathbf{A})^{-1} \end{bmatrix}. \quad (58)$$

Note that

$$\begin{aligned} \mathbf{B}_1^T \mathbf{B}_1 &= \begin{bmatrix} \mathbf{A}^T & \tilde{\mathbf{A}}^T \\ \tilde{\mathbf{A}}^T & -\mathbf{A}^T \end{bmatrix} \begin{bmatrix} \mathbf{A} & \tilde{\mathbf{A}} \\ \tilde{\mathbf{A}} & -\mathbf{A} \end{bmatrix} \\ &= \begin{bmatrix} \overline{\mathbf{A}^H \mathbf{A}} & \overline{\mathbf{A}^H \tilde{\mathbf{A}}} \\ -\overline{\mathbf{A}^H \tilde{\mathbf{A}}} & \overline{\mathbf{A}^H \mathbf{A}} \end{bmatrix} \end{aligned} \quad (59)$$

$$\begin{aligned} \overline{\mathbf{E} \mathbf{E}^H} = &\begin{bmatrix} \mathbf{P}_I^\perp - \overline{\mathbf{G} \mathbf{H}^{-1} \mathbf{G}^H} - \mathbf{G} \mathbf{H}^{-1} \mathbf{G}^T + \overline{\mathbf{G} \mathbf{H}^{-1} \mathbf{G}^H \mathbf{G} \mathbf{H}^{-1} \mathbf{G}^T}, \\ -j \mathbf{P}_I^\perp - \tilde{\mathbf{G}} \mathbf{H}^{-1} \mathbf{G}^H + j \mathbf{G} \mathbf{H}^{-1} \tilde{\mathbf{G}}^T + \tilde{\mathbf{G}} \mathbf{H}^{-1} \mathbf{G}^H \mathbf{G} \mathbf{H}^{-1} \tilde{\mathbf{G}}^T, \\ j \mathbf{P}_I^\perp - j \overline{\mathbf{G} \mathbf{H}^{-1} \mathbf{G}^H} - \mathbf{G} \mathbf{H}^{-1} \tilde{\mathbf{G}}^T + \overline{\mathbf{G} \mathbf{H}^{-1} \mathbf{G}^H \mathbf{G} \mathbf{H}^{-1} \tilde{\mathbf{G}}^T} \\ \mathbf{P}_I^\perp - j \tilde{\mathbf{G}} \mathbf{H}^{-1} \mathbf{G}^H + \mathbf{G} \mathbf{H}^{-1} \tilde{\mathbf{G}}^T + \tilde{\mathbf{G}} \mathbf{H}^{-1} \mathbf{G}^H \mathbf{G} \mathbf{H}^{-1} \tilde{\mathbf{G}}^T \end{bmatrix}_{2N \times 2N} \end{aligned} \quad (56)$$

$$\begin{aligned} \mathbf{F}_1 \mathbf{B}_1^T \mathbf{B}_1 &= \begin{bmatrix} (\overline{\mathbf{A}^H \mathbf{A}})^{-1} & (\overline{\mathbf{A}^H \mathbf{A}})^{-1} \\ -(\overline{\mathbf{A}^H \mathbf{A}})^{-1} & (\overline{\mathbf{A}^H \mathbf{A}})^{-1} \end{bmatrix} \begin{bmatrix} \overline{\mathbf{A}^H \mathbf{A}} & \overline{\mathbf{A}^H \mathbf{A}} \\ -\overline{\mathbf{A}^H \mathbf{A}} & \overline{\mathbf{A}^H \mathbf{A}} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_N \end{bmatrix}. \end{aligned} \quad (60)$$

Then, using (36) and  $\mathbf{F}_1 = (\mathbf{B}_1^T \mathbf{B}_1)^{-1}$ , we obtain

$$\begin{aligned} \hat{\mathbf{L}}_I &\approx \boldsymbol{\mu}^T \mathbf{E} \mathbf{E}^H \boldsymbol{\mu} \\ &= \boldsymbol{\mu}^T \overline{\mathbf{E} \mathbf{E}^H} \boldsymbol{\mu} \\ &= \boldsymbol{\mu}^T [\mathbf{I}_{2N} - \mathbf{B}_1 (\mathbf{B}_1^T \mathbf{B}_1)^{-1} \mathbf{B}_1^T - \mathbf{B}_2 (\mathbf{B}_2^T \mathbf{B}_2)^{-1} \mathbf{B}_2^T] \boldsymbol{\mu}. \end{aligned} \quad (61)$$

#### APPENDIX B A FEW LEMMAS

*Lemma B1:* The rank of  $\mathbf{B}_2$  is  $I$  if  $N \geq 3I/2$ .

*Proof of Lemma B1:* Let  $\mathbf{U}$  contain the  $N-I$  unitary eigenvectors of  $\mathbf{P}_I^\perp$  corresponding to 1. Then,  $\mathbf{P}_I^\perp = \mathbf{U} \mathbf{I}_{N-I} \mathbf{U}^H = \mathbf{U} \mathbf{U}^H$ .  $\mathbf{B}_2$  can be rewritten as

$$\mathbf{B}_2 = \begin{bmatrix} \overline{\mathbf{U}} & -\tilde{\mathbf{U}} \\ \tilde{\mathbf{U}} & \overline{\mathbf{U}} \end{bmatrix} \begin{bmatrix} \overline{\mathbf{U}} & -\tilde{\mathbf{U}} \\ \tilde{\mathbf{U}} & \overline{\mathbf{U}} \end{bmatrix}^T \begin{bmatrix} \overline{\mathbf{D}\mathbf{V}} \\ \overline{\mathbf{D}\mathbf{V}} \end{bmatrix}. \quad (62)$$

We will first show that  $[\overline{\mathbf{D}\mathbf{V}}^T, \tilde{\mathbf{D}\mathbf{V}}^T]^T$  has the rank  $I$ . Let an  $I$ -element vector  $\mathbf{c}$  satisfy

$$\begin{bmatrix} \overline{\mathbf{D}\mathbf{V}} \\ \tilde{\mathbf{D}\mathbf{V}} \end{bmatrix} \mathbf{c} = \mathbf{0} \Rightarrow (\mathbf{D}\mathbf{V})\mathbf{c} = \mathbf{0}. \quad (63)$$

As  $\mathbf{D}\mathbf{V}$  has the rank  $I$ , we have

$$(\mathbf{D}\mathbf{V})\mathbf{c} = \mathbf{0} \Rightarrow \mathbf{c} = \mathbf{0}. \quad (64)$$

Then, the rank of  $[\overline{\mathbf{D}\mathbf{V}}^T, \tilde{\mathbf{D}\mathbf{V}}^T]^T$  is  $I$ .

Next, we will show that

$$\begin{bmatrix} \overline{\mathbf{U}} & -\tilde{\mathbf{U}} \\ \tilde{\mathbf{U}} & \overline{\mathbf{U}} \end{bmatrix} \quad (65)$$

has the rank  $2(N-I) \geq I$ . Let a  $2I$ -element real vector  $\mathbf{c} = [\mathbf{c}_1^T, \mathbf{c}_2^T]^T$  satisfy

$$\begin{bmatrix} \overline{\mathbf{U}} & -\tilde{\mathbf{U}} \\ \tilde{\mathbf{U}} & \overline{\mathbf{U}} \end{bmatrix} \mathbf{c} = \mathbf{0} \quad (66)$$

where  $\mathbf{c}_1$  and  $\mathbf{c}_2$  are two  $I$ -element subvectors of  $\mathbf{c}$ . Then

$$\overline{\mathbf{U}}\mathbf{c}_1 - \tilde{\mathbf{U}}\mathbf{c}_2 = \mathbf{0} \quad (67)$$

$$\tilde{\mathbf{U}}\mathbf{c}_1 + \overline{\mathbf{U}}\mathbf{c}_2 = \mathbf{0}. \quad (68)$$

Multiplying (68) by  $j$  and adding it to (67) leads to

$$\mathbf{U}(\mathbf{c}_1 - j\mathbf{c}_2) = \mathbf{0}. \quad (69)$$

Since  $\mathbf{U}$  has  $I$  independent columns, then  $\mathbf{c}_1 = \mathbf{0}$ , and  $\mathbf{c}_2 = \mathbf{0}$ . We now know that the only solution to (66) is  $\mathbf{c} = \mathbf{0}$ , and hence, the matrix in (65) has the rank  $2(N-I)$ .

Combining the above results completes the proof of Lemma B1.  $\spadesuit$

*Lemma B2—* $\mathbf{B}_1^T \mathbf{B}_2 = \mathbf{0}$ :

*Proof of Lemma B2:* Note that  $\mathbf{A}$  and  $\mathbf{U}$  are orthogonal. Then

$$\begin{aligned} \mathbf{B}_1^T \mathbf{B}_2 &= \begin{bmatrix} \overline{\mathbf{A}^H \mathbf{U}} & -\overline{\mathbf{A}^H \mathbf{U}} \\ \overline{\mathbf{A}^H \mathbf{U}} & \overline{\mathbf{A}^H \mathbf{U}} \end{bmatrix} \begin{bmatrix} \overline{\mathbf{U}} & -\tilde{\mathbf{U}} \\ \tilde{\mathbf{U}} & \overline{\mathbf{U}} \end{bmatrix}^T \begin{bmatrix} \overline{\mathbf{D}\mathbf{V}} \\ \overline{\mathbf{D}\mathbf{V}} \end{bmatrix} \\ &= \mathbf{0}. \end{aligned} \quad (70)$$

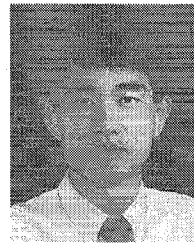
#### REFERENCES

- [1] R. Kumaresan, D. W. Tufts, and L. L. Scharf, "A Prony method for noisy data: Choosing the signal components and selecting the order in exponential signal models," *Proc. IEEE*, vol. 72, pp. 230-233, 1984.
- [2] L. C. Zhao, P. R. Krishnaiah, and Z. D. Bai, "On detection of the number of signals in presence of white noise," *J. Multivariate Anal.*, vol. 20, no. 1, pp. 1-25, Oct. 1986.
- [3] J. Hwang and Y. Chen, "A combined detection-estimation algorithm for the harmonic retrieval," *Signal Processing*, vol. 30, pp. 177-197, 1993.
- [4] P. M. Djurić, "Simultaneous detection and frequency estimation of sinusoidal signals," in *Proc. ICASSP'93*, 1993, vol. IV, pp. 53-56.
- [5] K. Konstantinides and K. Yao, "Statistical analysis of effective singular values in matrix rank determination," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 757-763, May 1988.
- [6] A. A. Shah and D. W. Tufts, "Rank determination in time-series analysis," in *Proc. ICASSP'94*, Adelaide, Australia, Apr. 1994, vol. IV, pp. 21-24.
- [7] P. Stoica, "On estimating the noise power in array processing," *Signal Processing*, vol. 26, pp. 205-220, 1992.
- [8] D. W. Tufts and A. A. Shah, "Determination of the dimension of a signal subspace from short data records," *IEEE Trans. Signal Processing*, vol. 42, pp. 2531-2535, 1994.
- [9] E. S. Pearson and H. O. Hartley, *Biometrika Tables for Statisticians*. Cambridge, U.K.: Cambridge Univ. Press, 1962, vol. I.
- [10] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus with Application in Statistics and Econometrics*. New York: Wiley, 1988.
- [11] Y. Hua and T. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 814-824, 1990.
- [12] Y. Hua and T. Sarkar, "On SUD for estimating generalized eigenvalues of singular matrix pencils in noise," *IEEE Trans. Signal Processing*, vol. 39, pp. 892-900, 1991.



**Qi Cheng** was born in Bengbu, China, on July 15, 1962. He received the B.E. degree from Anhui University, Anhui, China, in 1982, the master of engineering degree from the Institute of Electronics, Chinese Academy of Science, Beijing, in 1985, and the Ph.D. degree from the University of Melbourne, Parkville, Australia, in 1995.

He was initially an Associate Lecturer and then a Lecturer at the Graduate School of the University of Science and Technology of China from 1985 to 1990. From 1990 to 1991, he was a Visiting Scholar at the University of Tasmania, Hobart, Australia. From 1995 to 1996, he was a Research Fellow at the University of Melbourne. Since February 1996, he has been a Lecturer at the Northern Territory University, Darwin, Australia. His research interests include estimation and detection in the areas of radar, sonar, and communications.



**Yingbo Hua** (S'86-M'88-SM'92) was born in China in 1960. He received the B.S. degree from the Nanjing Institute of Technology (currently Southeast University), Nanjing, China, in February 1982 and the M.S. and Ph.D. degrees from Syracuse University, Syracuse, NY, in 1983 and 1988, respectively.

Since February 1990, he has been with the University of Melbourne, Parkville, Australia, where he was Lecturer from 1990 to 1992, Senior Lecturer from 1993 to 1995, and has been an Associate Professor and Reader since January 1996. He has published approximately 50 journal papers and 80 conference papers in the areas of spectral estimation, array processing, radar and NMR imaging, system identification, neural networks, and wireless communications.

Dr. Hua has been an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING since 1994 and is a member of a number of international technical committees.  $\spadesuit$