



## Professor C R Rao's contributions in statistical signal processing and its long-term implications

DEBASIS KUNDU

Department of Mathematics and Statistics, Indian Institute of Technology Kanpur,  
Kanpur 208 016, India  
E-mail: kundu@iitk.ac.in

Published online 25 July 2020

**Abstract.** Professor C R Rao has made significant contributions in different areas of statistics and in related fields particularly in inference, biometrics, design of experiments, linear models, variance components, econometrics, and most of his contributions are well known to the statistical community. But it may not be known to many statisticians that Professor Rao has worked quite extensively, for about six to seven years, on statistical signal processing and has made some fundamental contributions which has generated significant interest along that direction. He along with his collaborators have worked mainly on three classical signal processing problems, and provided theoretical foundations and efficient estimation procedures. In my opinion, the main contribution of Professor Rao is that he has provided new insights in all these problems, which has helped to bring new way of solving these and some related problems. He has guided two Ph.D. students in the area of Statistical Signal Processing. The main aim of this article is two-fold. First, we would like to introduce to the statistical community the contribution of Professor Rao in this area and our second aim is to provide a class of different related open problems which are of interest and which require sophisticated statistical tools to provide efficient solutions.

**Keywords.** Sinusoidal model; least squares estimator; Prony's method; asymptotic distribution; chirp model.

**Mathematics Subject Classification.** 62F10, 62F03, 62H12.

### 1. Introduction

Signal processing may broadly be considered as the recovery of information from different physical observations. The signals can be observed from different sources and they are usually corrupted by noise due to electrical, mechanical, atmospheric or intentional interferences. Since the observed signals are random in nature, statistical techniques are needed to extract the original signals. Statistics is used in the formulation of the proper models to describe the behavior of the underlying process, developing proper techniques to estimate the unknown model parameters, and assessment of model performances. Statistical signal processing broadly refers to the analysis of different random signals using appropriate statistical procedures.

---

This article is part of the "Special Issue in Honour of Professor C R Rao on His Birth Centenary".

Professor Rao has worked on statistical signal processing in early 90's for a period of six to seven years. He along with his collaborators have worked on three fundamental problems in this area, namely (a) one-dimensional superimposed exponential model, (b) two-dimensional superimposed exponential model and, (c) direction of arrival (DOA) model. These are classical problems, and they have applications in several areas including telecommunications, radio location of objects, seismic signal processing, image processing, computer-assisted medical diagnostics, etc. Although these problems have several applications, they are highly non-linear in nature. The first two problems can be seen as non-linear regression problems. Both the problems can be formulated as the estimation of unknown parameters of a non-linear model in the presence of additive noise. The third problem involves as the estimation of the unknown parameters in a non-linear random effect model in the presence of additive noise. It will be seen that finding efficient estimators and deriving properties of these estimators are quite challenging problems due to the nature of the problems.

It may be mentioned that an extensive amount of work has been done in standard non-linear regression models, see for example the books by Bates and Watts [8] and Seber and Wild [52] in this respect. Least squares estimators seem to be the natural choice in a non-linear regression problem. But finding least squares estimators and establishing their properties are quite non-trivial, in general. Jennrich [20] and later on, Wu [55] provided several sufficient conditions for establishing the consistency and asymptotic normality properties of the least squares estimators of a non-linear regression model. The one dimensional and two-dimensional superimposed exponential models do not satisfy those sufficient conditions. Therefore, it is not immediate whether the least squares estimators will be consistent or not in these cases. Professor Rao and his collaborators established the consistency and asymptotic normality properties in these cases quite differently than the methods developed by Jennrich [20] and Wu [55] for general non-linear regression models.

Although the least squares estimators are the most efficient estimators, finding the least squares estimators in case of superimposed exponential models is well known to be a numerically difficult problem. The standard Newton–Rahson or Gauss–Newton algorithms do not work well as the problems are highly non-linear in nature. The least squares surface has several local minima. Hence, most of the standard iterative procedure converges to the local minimum rather than the global minimum. Due to this reason, an extensive amount of work has been done to find efficient estimators which behave like least squares estimators. Professor Rao and his collaborators established an efficient algorithm which converges in a finite number of steps and it produces estimates which have the same asymptotic convergence rates as the least squares estimators.

The main aim of this paper is two-fold. First, we would like to define the three classical problems from statistical perspective and provide solutions developed by Professor Rao and his collaborators. Our second aim is to show how his work has influenced the future research in this area. There are several interesting open problems in statistical signal processing for which sophisticated statistical techniques are needed to provide efficient solutions. We will provide several open problems and relevant references for the statistical community who may not be very familiar in this area.

The rest of the paper is organized as follows. We need some preliminaries for basic understanding of these problems, and that will be provided in Section 2. In Section 3, we describe the one-dimensional superimposed exponential models and provide different estimation procedures as provided by Professor Rao and his collaborators. Two-dimensional

superimposed exponential model will be considered in Section 4 and DOA model will be described in Section 5. In Section 6, we provide several related problems which have been considered in the statistical signal processing literature in the last two to three decades and we conclude the paper.

## 2. Preliminaries

We provide one important result which has been used quite extensively in the statistical signal processing literature, and it is well known as Prony's equation. Prony proposed the method in 1795 mainly to estimate the unknown parameters of a sum of real exponential model. It is available in several numerical analysis text books, for example, in [15] or [18]. It can be described as follows. Suppose

$$\mu(t) = \alpha_1 e^{\beta_1 t} + \cdots + \alpha_M e^{\beta_M t}, \quad t = 1, \dots, n.$$

Here  $\alpha_1, \dots, \alpha_M$  are arbitrary real number and  $\beta_1, \dots, \beta_M$  are distinct real numbers. Then, for a given  $\{\mu_1, \dots, \mu_n\}$ , there exists  $(M + 1)$  constants  $\{g_0, \dots, g_M\}$ , such that

$$\mathbf{A}\mathbf{g} = \mathbf{0}, \quad (1)$$

where

$$\mathbf{A} = \begin{bmatrix} \mu(1) & \cdots & \mu(M+1) \\ \vdots & \ddots & \vdots \\ \mu(n-M) & \cdots & \mu(n) \end{bmatrix}, \quad \mathbf{g} = \begin{bmatrix} g_0 \\ \vdots \\ g_M \end{bmatrix} \quad \text{and} \quad \mathbf{0} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}.$$

It can be shown that the rank of the matrix  $\mathbf{A}$  is  $M$ . Hence, the null space of the matrix  $\mathbf{A}$  is of dimension one. Note that to make  $\mathbf{g}$  to be unique, we can put restrictions on  $g_0, \dots, g_M$  such that  $\sum_{j=0}^M g_j^2 = 1$  and  $g_0 > 0$ . The set of linear equations (1) is known as Prony's equations. It can be shown that the roots of the following polynomial equation

$$p(x) = g_0 + g_1 x + \cdots + g_M x^M = 0 \quad (2)$$

are  $e^{\beta_1}, \dots, e^{\beta_M}$ . Therefore, there is a one-to-one correspondence between  $\{\beta_1, \dots, \beta_M\}$  and  $\{g_0, \dots, g_M\}$ , such that  $\sum_{j=0}^M g_j^2 = 1$  and  $g_0 > 0$ . Moreover,  $\{g_0, \dots, g_M\}$  do not depend on  $\{\alpha_1, \dots, \alpha_M\}$ , but depend only on  $\{\beta_1, \dots, \beta_M\}$ .

One natural question is how to recover  $\alpha_1, \dots, \alpha_M$  and  $\beta_1, \dots, \beta_M$  for a given  $\mu(1), \dots, \mu(n)$ . Note that

$$\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\alpha},$$

where

$$\boldsymbol{\mu} = \begin{bmatrix} \mu(1) \\ \vdots \\ \mu(n) \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} e^{\beta_1} & \cdots & e^{\beta_M} \\ \vdots & \ddots & \vdots \\ e^{n\beta_1} & \cdots & e^{n\beta_M} \end{bmatrix} \quad \text{and} \quad \boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_M \end{bmatrix}.$$

It is immediate that for  $\beta_1 \neq \dots \neq \beta_M$ , the rank of the matrix  $\mathbf{X}$  is  $M$ . Hence, for  $n > M$ , the matrix  $\mathbf{X}^\top \mathbf{X}$  is of full rank, and

$$\boldsymbol{\alpha} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\mu}. \quad (3)$$

Therefore, for a given  $\mu(1), \dots, \mu(n)$ , first obtain the vector  $\mathbf{g}$  which satisfies (1). Obtain  $\beta_1, \dots, \beta_M$  from the roots of the polynomial equation  $p(x)$  as given in (2). Finally, obtain  $\alpha_1, \dots, \alpha_M$  from (3).

Note that although we have described Prony’s equations when  $\alpha$ ’s and  $\beta$ ’s are real valued, it is still valid when  $\alpha$ ’s and  $\beta$ ’s are complex-valued also. Due to this reason, Prony’s equations play an important role in the statistical signal processing literature. Several algorithms have been developed using Prony’s equations. It has been extended to two-dimensional superimposed exponential model also, see for example, [7].

As it has been mentioned above that Prony’s equations are valid even when  $\alpha$ ’s and  $\beta$ ’s are complex-valued, let us consider the following special case which is known as the undamped exponential model:

$$\mu(t) = A_1 e^{i\omega_1 t} + \dots + A_M e^{i\omega_M t}, \quad t = 1, \dots, n. \tag{4}$$

Here,  $A_1, \dots, A_M$  are arbitrary complex number,  $0 < \omega_1 \neq \dots \neq \omega_M < 2\pi$ , and  $i = \sqrt{-1}$ . In this case also, there exists  $\{g_0, \dots, g_M\}$ , which are complex-valued, such that they satisfy (1). Moreover, the roots of the polynomial equation  $p(z)$ , as in (2), are  $z_1 = e^{i\omega_1}, \dots, z_M = e^{i\omega_M}$ . Observe that

$$|z_1| = \dots = |z_M| = 1, \quad \bar{z}_1 = z_1^{-1}, \dots, \bar{z}_M = z_M^{-1}. \tag{5}$$

Here,  $\bar{z}_k$  denotes the complex conjugate of  $z_k$ , for  $k = 1, \dots, M$ . Define the new polynomial

$$Q(z) = z^{-M} \bar{p}(z) = \bar{g}_0 \bar{z}^{-M} + \dots + \bar{g}_M.$$

From (5), it is clear that  $p(z)$  and  $Q(z)$  have the same roots. Hence, by comparing the coefficients of the two polynomials  $p(z)$  and  $Q(z)$ , we obtain

$$\frac{g_k}{g_M} = \frac{\bar{g}_{M-k}}{\bar{g}_0}, \quad k = 0, \dots, M. \tag{6}$$

Therefore, if we denote

$$b_k = g_k \left( \frac{\bar{g}_0}{g_M} \right)^{-\frac{1}{2}}, \quad k = 0, \dots, M,$$

then

$$b_k = \bar{b}_{M-k}, \quad k = 0, \dots, M.$$

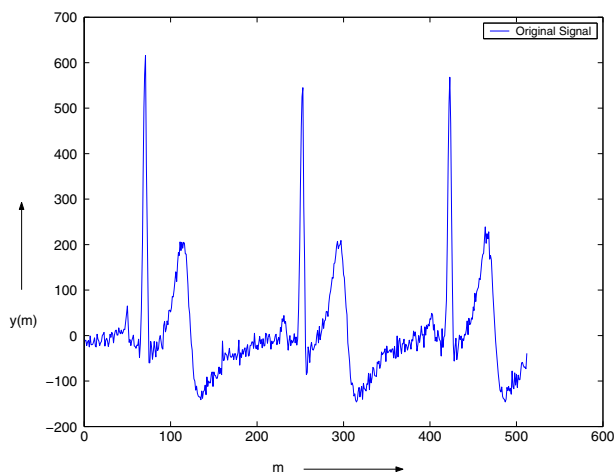
Hence, if we denote

$$\mathbf{b} = \begin{bmatrix} b_0 \\ \vdots \\ b_M \end{bmatrix}, \quad \bar{\mathbf{b}} = \begin{bmatrix} \bar{b}_0 \\ \vdots \\ \bar{b}_M \end{bmatrix}, \quad \mathbf{J} = \begin{bmatrix} 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 1 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 1 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \end{bmatrix},$$

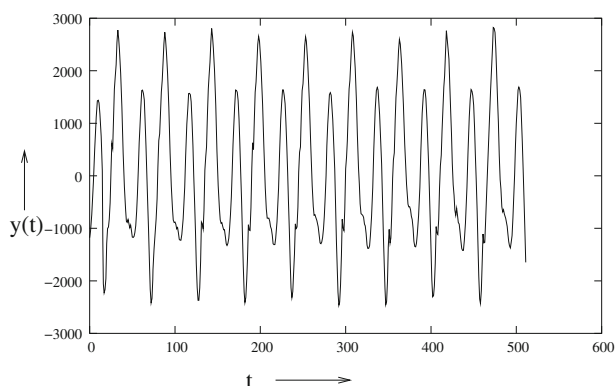
then  $\mathbf{b} = \bar{\mathbf{J}}\bar{\mathbf{b}}$ . Therefore, it is immediate that for a given  $\mu(1), \dots, \mu(n)$ , as in (4), there exists a vector  $\mathbf{g} = (g_0, \dots, g_M)^\top$ , such that  $\sum_{k=0}^M |g_k|^2 = 1$ , which satisfies (1) and also  $\mathbf{g} = \mathbf{J}\bar{\mathbf{g}}$ . For more details, interested readers are referred to [36].

### 3. One-dimensional superimposed exponential model

We observe periodic phenomena everyday in our lives. For example, the number of tourists visiting the famous Taj Mahal, the daily temperature of the capital, New Delhi, the ECG



**Figure 1.** ECG Plot of a normal human being.



**Figure 2.** UUU-sound of a male.

pattern of a normal human being or the musical sounds, clearly follows a periodic pattern. Sometimes it may not be exactly periodic but nearly periodic. Moreover, the periodic signal may not be in one dimension only, it can be in two- or three-dimensions also. One major problem in statistical signal processing is to analyze periodic or nearly periodic data when they are observed with noise. Some of the natural questions are: what do you mean by a periodic signal? and, why do we want to analyze a periodic data? We will not provide a formal definition of a periodic data, but informally speaking, a periodic data indicates a repeated pattern in one dimension and a symmetric pattern in two- or three-dimensions. See for example the following figures which indicate periodic nature. Figure 1 shows the ECG plot of a normal human being and Figure 2 shows 'UUU' vowel sound of a male.

Now the second question is: why somebody wants to analyze a periodic data? One reason might be purely theoretical in nature, otherwise providing a proper periodic model can be very useful for compression or prediction purposes.

Note that the simple periodic function is the sinusoidal function and it can be written in the following form:

$$y(t) = A \cos(\omega t) + B \sin(\omega t).$$

Clearly, the period of the function  $y(t)$  is the shortest time taken for  $y(t)$  to repeat itself, and it is  $2\pi/\omega$ . In general, a smooth periodic function (mean adjusted) with period  $2\pi/\omega$  can be written in the form

$$y(t) = \sum_{k=1}^{\infty} [A_k \cos(k\omega t) + B_k \sin(k\omega t)], \quad (7)$$

and it is well known as the Fourier expansion of  $y(t)$ . From  $y(t)$ ,  $A_k$  and  $B_k$  can be obtained uniquely for  $k \geq 1$  as

$$\int_0^{2\pi/\omega} \cos(k\omega t) y(t) dt = \frac{\pi A_k}{\omega} \quad \text{and} \quad \int_0^{2\pi/\omega} \sin(k\omega t) y(t) dt = \frac{\pi B_k}{\omega}.$$

Since, in practice,  $y(t)$  is always corrupted with noise, so it is reasonable to assume that we have the following observation:

$$y(t) = \sum_{k=1}^{\infty} [A_k \cos(\omega t) + B_k \sin(\omega t)] + e(t),$$

where  $e(t)$  is the noise component. It is impossible to estimate infinite number of parameters. Hence the model is approximated by the following model:

$$y(t) = \sum_{k=1}^p [A_k \cos(\omega_k t) + B_k \sin(\omega_k t)] + e(t), \quad (8)$$

for some  $p < \infty$ . The aim is to extract (estimate) the deterministic component  $\mu(t)$ , where

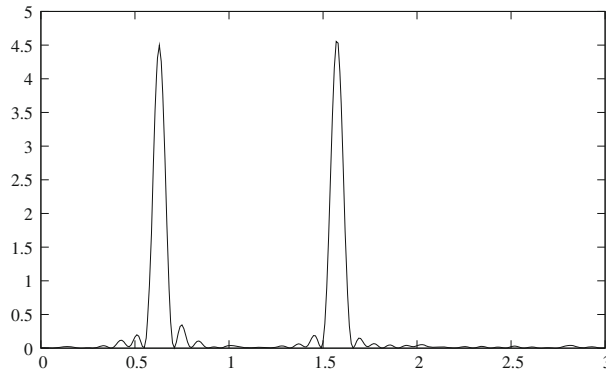
$$\mu(t) = \sum_{k=1}^p [A_k \cos(\omega_k t) + B_k \sin(\omega_k t)],$$

in the presence of the random error component  $e(t)$ , based on the available data  $y(t)$ ,  $t = 1, \dots, n$ . Hence, the problem becomes the estimation of  $p$ ,  $A_k$ ,  $B_k$  and  $\omega_k$ , for  $k = 1, \dots, p$ .

It should be mentioned that often, instead of working with the model (8), it might be more convenient to work with the associated complex-valued model. With the abuse of notation, we use the corresponding complex-valued model as

$$y(t) = \mu(t) + e(t) = \sum_{k=1}^p \alpha_k e^{i\omega_k t} + e(t). \quad (9)$$

In the model (9),  $y(t)$ 's,  $\alpha_k$ 's,  $e(t)$ 's are all complex valued and  $i = \sqrt{-1}$ . The model (9) can be obtained by taking the Hilbert transformation of (8). Therefore, the two models are equivalent. Although, any observed signal is always real-valued, by taking the Hilbert transformation of the signal, the corresponding complex-valued model can be used. Any analytical result or numerical procedure for model (9) can be used for model (8) and vice-versa. Although, Rao and his collaborators have mainly dealt with model (9), all the results are valid for model (8) also, as we have just mentioned. Due to this reason, in this paper,



**Figure 3.** Periodogram plot of  $y(t)$  obtained from Model (11).

we provide the results either for model (8) or for model (9) and it should be clear from the context.

Now consider the model (8), and the problem is to estimate the unknown parameters based on a sample  $\{y(t); t = 1, \dots, n\}$ . It is necessary to make certain assumptions on the random error component  $e(t)$ . At this moment, we simply make the assumptions that  $e(t)$  has mean zero and finite variance. We will make it more explicit later when it is needed. The most used and popular estimation procedure is known as the periodogram estimators (see for example, [53]). The periodogram at a particular frequency is defined as

$$I(\theta) = \left| \frac{1}{n} \sum_{t=1}^n y(t) \cos(\theta t) \right|^2 + \left| \frac{1}{n} \sum_{t=1}^n y(t) \sin(\theta t) \right|^2. \quad (10)$$

The main reason to use the periodogram function to estimate the frequencies and the number of components  $p$  is that  $I(\theta)$  has local maxima at the true frequencies if there is no noise in the data. Therefore, if the noise variance is not high, it is expected that  $I(\theta)$  can be used quite effectively to estimate the frequencies and  $p$  also. It may be observed from the following example. Let us consider the following signal:

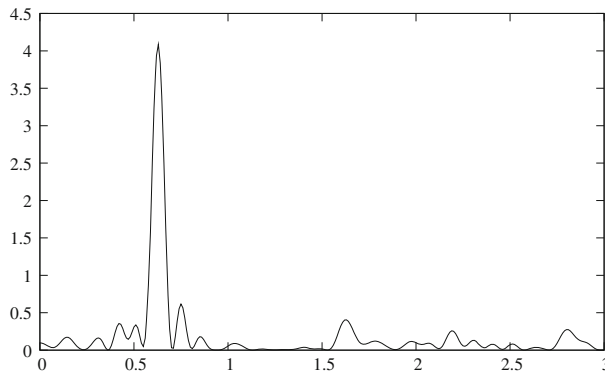
$$y(t) = 3.0 \cos(0.2\pi t) + 3.0 \sin(0.2\pi t) + 3.0 \cos(0.5\pi t) + 3.0 \sin(0.5\pi t) + e(t). \quad (11)$$

Here  $e(t)$ 's are assumed to be independent and identically distributed (i.i.d.) normal random variables with mean 0 and variance 2. The plot of the periodogram function  $I(\theta)$  for the model (11) is provided in Figure 3. From Figure 3, it is clear that  $p = 2$  and the local maxima are close to the true frequencies. But, the same may not be true always. For example, let us consider the following signal:

$$y(t) = 3.0 \cos(0.2\pi t) + 3.0 \sin(0.2\pi t) + 0.25 \cos(0.5\pi t) + 0.25 \sin(0.5\pi t) + e(t). \quad (12)$$

Here  $e(t)$ 's are i.i.d. normal random variables with mean 0 and variance 5.0. The periodogram plot of  $y(t)$  obtained from the model (12) is provided in Figure 4. From Figure 4, it is not clear that  $p = 2$  and the true location of the frequencies.

Therefore, the least squares estimators seems to be a natural choice. In case of model (9) for a given  $p$ , the least squares estimators can be obtained by the argument minimum



**Figure 4.** Periodogram plot of  $y(t)$  obtained from Model (12).

of the sum of residuals as follows:

$$Q(\boldsymbol{\alpha}, \boldsymbol{\omega}) = \sum_{t=1}^n \left| y(t) - \sum_{k=1}^p \alpha_k e^{i\omega_k t} \right|^2. \tag{13}$$

Here  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_p)^\top$  and  $\boldsymbol{\omega} = (\omega_1, \dots, \omega_p)^\top$ . Note that the model (9) is a complex valued non-linear regression model. There exists an extensive list of literature related to non-linear regression models (interested readers may refer to the book by Seber and Wild [52]). When  $e(t)$ 's are i.i.d. random variables with mean zero and finite variance, Jennrich [20] and later on, Wu [55] have provided several sufficient conditions for the least squares estimators to be consistent and asymptotically normally distributed. Some of these results have been generalized to the complex valued models also by Kundu [26]. It can be easily shown (see for example, [33]) that the model (9) does not satisfy the sufficient conditions of [20,55] or [26]. Hence, the consistency of the least squares estimators under the assumptions of i.i.d. errors is not guaranteed.

Rao and Zhao [48] provided the consistency and asymptotic normality properties of the maximum likelihood estimators of  $\boldsymbol{\alpha}$ ,  $\boldsymbol{\omega}$  and  $\sigma^2$ , under the following assumptions on  $e(t)$ , for known  $p$ . Let us denote  $\hat{\boldsymbol{\alpha}}$  and  $\hat{\boldsymbol{\omega}}$  as the least squares estimators of  $\boldsymbol{\alpha}$  and  $\boldsymbol{\omega}$ , respectively, and

$$\widehat{\sigma^2} = \frac{1}{n} Q(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\omega}}).$$

We need the following assumptions for further development.

*Assumption 3.1.*  $\{e(t)\}$ ,  $t = 1, 2, \dots$  are i.i.d. complex-valued random variables. If  $\text{Re}(e(t))$  and  $\text{Im}(e(t))$  denote the real and imaginary parts of  $e(t)$ , then it is assumed that  $\text{Re}(e(t))$  and  $\text{Im}(e(t))$  are independent normal random variables with mean zero and variance  $\sigma^2/2$ .

*Assumption 3.2.*  $\alpha_k \neq 0$ ,  $0 < \omega_k < 2\pi$ , and  $\omega_j \neq \omega_k$ , for  $j \neq k$ ,  $k = 1, \dots, p$ .

**Theorem 1.** *If the diagonal matrix  $\mathbf{D} = \text{diag}\{\alpha_1, \dots, \alpha_p\}$ , we denote  $\mathbf{A}_1 = \text{Im } \mathbf{D}$ ,  $\mathbf{A}_2 = \text{Re } \mathbf{D}$  and  $\mathbf{B} = (\mathbf{D}^H \mathbf{D})^{-1}$ . Here ‘ $H$ ’ denotes the complex conjugate transpose of a matrix or of a vector. If  $e(t)$ 's satisfy Assumption 3.1 and the model parameters satisfy Assumption 3.2, then*



- (i)  $\hat{\omega}$ ,  $\hat{\alpha}$  and  $\hat{\sigma}^2$  are strongly consistent estimators of  $\omega$ ,  $\alpha$  and  $\sigma^2$ , respectively.
- (ii) The limiting distribution of

$$(n^{3/2}(\hat{\omega} - \omega), n^{1/2} \text{Re}(\hat{\alpha} - \alpha), n^{1/2} \text{Im}(\hat{\alpha} - \alpha), \hat{\sigma}^2)^\top$$

is a  $(3p + 1)$ -variate normal distribution with mean vector zero, and covariance matrix  $\sigma^2 \Sigma$ , where

$$\Sigma = \begin{bmatrix} 6\mathbf{B} & 3\mathbf{A}_1\mathbf{B} & -3\mathbf{A}_2\mathbf{B} & \mathbf{0} \\ 3\mathbf{A}_1\mathbf{B} & \frac{1}{2}\mathbf{I} + \frac{3}{2}\mathbf{A}_1^2\mathbf{B} & -\frac{3}{2}\mathbf{A}_2\mathbf{A}_1\mathbf{B} & \mathbf{0} \\ -3\mathbf{A}_2\mathbf{B} & -\frac{3}{2}\mathbf{A}_2\mathbf{A}_1\mathbf{B} & \frac{1}{2}\mathbf{I} + \frac{3}{2}\mathbf{A}_2^2\mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \sigma^2 \end{bmatrix}.$$

Note that Theorem 1 provides the strong consistency and asymptotic normality of the least squares estimators under the assumptions that the errors are normally distributed even if the model does not satisfy some of the standard sufficient conditions which are available for a general non-linear regression model. The results can be used for constructing asymptotic confidence intervals of the unknown parameters or for testing of hypothesis purposes also. The most important point of Theorem 1 is that the rates of convergence of the linear parameters are of the order  $O_p(n^{-1/2})$ , which is quite common, whereas for the frequencies it is  $O_p(n^{-3/2})$ , which is quite uncommon for a general non-linear regression model. It implies that the least squares estimators of the frequencies are more efficient than the least squares estimators of the linear parameters. The results of Rao and Zhao [48] were extended by Kundu and Mitra [34] when the errors are i.i.d. random variables with mean zero and finite variance. It has been further extended in case of stationary errors also (for details, see for example the monograph by Kundu and Nandi [36]).

Now we will discuss how to compute the least squares estimators of the unknown parameters of the model (9). The model (9) can be written as follows:

$$\mathbf{Y} = \mathbf{X}\alpha + \mathbf{e}. \tag{14}$$

Here

$$\mathbf{Y} = \begin{bmatrix} y(1) \\ \vdots \\ y(n) \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} e^{i\omega_1} & \dots & e^{i\omega_p} \\ \vdots & \ddots & \vdots \\ e^{in\omega_1} & \dots & e^{in\omega_p} \end{bmatrix}, \quad \alpha = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_p \end{bmatrix}, \quad \mathbf{e} = \begin{bmatrix} e(1) \\ \vdots \\ e(n) \end{bmatrix}.$$

Therefore,  $Q(\alpha, \omega)$  as in (13) can be written as

$$Q(\alpha, \omega) = (\mathbf{Y} - \mathbf{X}(\omega)\alpha)^H (\mathbf{Y} - \mathbf{X}(\omega)\alpha). \tag{15}$$

For a given  $\omega$ , the least squares estimators of  $\alpha$ , say  $\hat{\alpha}(\omega)$ , can be obtained as

$$\hat{\alpha}(\omega) = (\mathbf{X}^H(\omega)\mathbf{X}(\omega))^{-1}\mathbf{X}^H(\omega)\mathbf{Y}.$$

Hence,

$$Q(\hat{\alpha}(\omega), \omega) = \mathbf{Y}^H(\mathbf{I} - \mathbf{P}_{\mathbf{X}(\omega)})^H (\mathbf{I} - \mathbf{P}_{\mathbf{X}(\omega)})\mathbf{Y} = \mathbf{Y}^H(\mathbf{I} - \mathbf{P}_{\mathbf{X}(\omega)})\mathbf{Y}.$$

Here  $\mathbf{P}_{\mathbf{X}(\omega)} = \mathbf{X}(\omega)(\mathbf{X}^H(\omega)\mathbf{X}(\omega))^{-1}\mathbf{X}^H(\omega)$  is the projection matrix on the column space of  $\mathbf{X}(\omega)$ . Hence, the least squares estimator of  $\omega$ , say  $\hat{\omega}$  can be obtained as

$$\hat{\omega} = \text{argmin } \mathbf{Y}^H(\mathbf{I} - \mathbf{P}_{\mathbf{X}(\omega)})\mathbf{Y} = \text{argmax } \mathbf{Y}^H\mathbf{P}_{\mathbf{X}(\omega)}\mathbf{Y}. \tag{16}$$

The problem to find  $\hat{\omega}$  from (16) is a non-linear optimization problem. It has been observed by Bresler and Macovski [10] that the standard Newton–Raphson algorithm or its variants

does not work in this case because the problem is a highly non-linear in nature. Using (1) it can be easily seen that there exists a  $n \times (n - p)$  matrix  $\mathbf{G}(\mathbf{g})$ , such that  $\mathbf{G}^H(\mathbf{g})\boldsymbol{\mu} = \mathbf{0}$ , where

$$\mathbf{G}^H(\mathbf{g}) = \begin{bmatrix} g_0 & g_1 & \dots & g_p & 0 & 0 & \dots & 0 \\ 0 & g_0 & \dots & g_p & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & \dots & g_0 & \dots & g_p \end{bmatrix},$$

and  $\mathbf{g} = (g_0, \dots, g_p)^\top$ . Since,  $\mathbf{G}^H(\mathbf{g})\boldsymbol{\mu} = \mathbf{0}$ , it implies  $\mathbf{G}^H(\mathbf{g})\mathbf{X}(\boldsymbol{\omega}) = \mathbf{0}$ . Hence,

$$\mathbf{Y}^H \mathbf{P}_{\mathbf{X}(\boldsymbol{\omega})} \mathbf{Y} = \mathbf{Y}^H (\mathbf{I} - \mathbf{P}_{\mathbf{G}(\mathbf{g})}) \mathbf{Y}.$$

Therefore, if

$$\hat{\mathbf{g}} = \operatorname{argmin} \mathbf{Y}^H \mathbf{P}_{\mathbf{G}(\mathbf{g})} \mathbf{Y},$$

and  $\hat{\boldsymbol{\omega}} = (\hat{g}_0, \dots, \hat{g}_p)^\top$ , then  $\hat{\boldsymbol{\omega}}$  can be obtained from the roots of the polynomial equation (2) by replacing  $g_i$ 's by  $\hat{g}_i$ , for  $i = 0, 1, \dots, p$ . Kundu [27] converted this problem as a non-linear eigenvalue problem and provided an efficient estimation procedures of the frequencies.

Since it is difficult to obtain the least squares estimators or the maximum likelihood estimators when the errors are normally distributed, an extensive amount of work has been done in the signal processing literature to develop several sub-optimal algorithms, which are non-iterative in nature and which behave like the maximum likelihood estimators. Most of these algorithms have used Prony's equations in some way or the other. It is difficult to list all the algorithms in a limited space, which are available till today. Among the different algorithms, the most popular ones are FBLP and MUSIC algorithms. Interested readers are referred to the monograph by Kundu and Nandi [36] in this respect. We provide a non-iterative procedure which was proposed by Bai *et al.* [2] based on Prony's equations and it is very easy to implement in practice. Moreover, it does not require the assumptions of normality on the error term. It only needs that the errors should be i.i.d. random variables with mean zero and finite variance. Consider the following  $(n - p) \times (p + 1)$  data matrix  $\mathbf{E}$  and  $(p + 1) \times (p + 1)$  data matrix  $\mathbf{F}$  as follows:

$$\mathbf{E} = \begin{bmatrix} y(1) & \dots & y(p + 1) \\ \vdots & \vdots & \vdots \\ y(n - p) & \dots & y(n) \end{bmatrix}, \quad \mathbf{F} = \frac{1}{n - p} \mathbf{E}^H \mathbf{E}.$$

Obtain the eigenvector corresponding to the minimum eigenvalue of  $\mathbf{F}$ . It gives an estimator  $\tilde{\mathbf{g}} = (\tilde{g}_0, \dots, \tilde{g}_p)^\top$  of  $\mathbf{g}$ . Construct a polynomial equation of the form

$$\tilde{g}_0 + \tilde{g}_1 z + \dots + \tilde{g}_p z^p = 0, \quad (17)$$

and obtain the roots of the form  $\tilde{\rho}_1 e^{-i\tilde{\omega}_1}, \dots, \tilde{\rho}_p e^{-i\tilde{\omega}_p}$ . Take  $\tilde{\omega}_1, \dots, \tilde{\omega}_p$  as the estimates of  $\omega_1, \dots, \omega_p$ , respectively. It has been shown by Bai *et al.* [2] that  $\tilde{\omega}_j$  is a consistent estimate of  $\omega_j$ , and  $\tilde{\omega}_j - \omega_j = O_p(n^{-1/2})$ , for  $j = 1, \dots, p$ . Using the conjugate symmetric properties of the polynomial coefficients as described in (6), the method has been modified by Kannan and Kundu [23].

Now we describe an efficient estimation procedure developed by Professor Rao and his colleagues, see Bai *et al.* [5], which is iterative in nature, but it is guaranteed that it stops in a fixed number of iteration. It also does not need the assumptions of normality on the error

terms similar to the EVLP method as described above. Moreover, the most salient feature of this algorithm is that it produces frequency estimators which have the same asymptotic efficiency as the least squares estimators. It is quite an unique algorithm in the sense that no other iterative procedure available today, guarantees convergence in a fixed number of steps. It does not use the whole data set at each step. It starts with a smaller data set, and then gradually increases to the complete data set. The algorithm can be described as follows:

Let  $\tilde{\omega}_j$  be a consistent estimator of  $\omega_j$ , for  $j = 1, \dots, p$ , and compute  $\hat{\omega}_j$  for  $j = 1, \dots, p$  as follows:

$$\hat{\omega}_j = \tilde{\omega}_j + \frac{12}{n^2} \text{Im} \left[ \frac{C_n}{D_n} \right],$$

where

$$C_n = \sum_{t=1}^n y(t)(t - n/2)e^{-i\tilde{\omega}_j t}, \quad D_n = \sum_{t=1}^n y(t)e^{-i\tilde{\omega}_j t}.$$

Then we have the following result whose proof can be found in Bai et al. [5].

**Theorem 2.** *If  $\tilde{\omega}_j - \omega_j = O_p(n^{-1-\delta})$  for  $\delta \in (0, 1/2]$ , and for  $j = 1, \dots, p$ , then*

- (1)  $\hat{\omega}_j - \omega_j = O_p(n^{-1-2\delta})$ , for  $\delta \leq 1/4$ .
- (2)  $n^{3/2}(\hat{\omega} - \omega) \rightarrow N_p(\mathbf{0}, 6\sigma^2(\mathbf{D}^H \mathbf{D})^{-1})$  if  $\delta > 1/4$ .

Here  $\mathbf{D}$  is  $p \times p$  diagonal matrix as  $\mathbf{D} = \text{diag}\{\alpha_1, \dots, \alpha_p\}$ .

We start with a consistent estimate of  $\omega_j$ , and improve it step by step by a recursive algorithm. The  $m$ -th stage estimate of  $\hat{\omega}_j^{(m)}$  is computed from the  $(m-1)$ -th stage estimate  $\hat{\omega}_j^{(m-1)}$ , by the formula

$$\hat{\omega}_j^{(m)} = \hat{\omega}_j^{(m-1)} + \frac{12}{n_m^2} \text{Im} \left[ \frac{C_{n_m}}{D_{n_m}} \right], \quad (18)$$

where  $C_{n_m}$  and  $D_{n_m}$  can be obtained from  $C_n$  and  $D_n$  by replacing  $n$  and  $\tilde{\omega}_j$  with  $n_m$  and  $\hat{\omega}_j^{(m-1)}$ , respectively. We apply the formula (18) repeatedly choosing  $n_m$  at each stage as follows:

*Step 1.* With  $m = 1$ , choose  $n_1 = n^{0.40}$  and  $\hat{\omega}_j^{(0)} = \tilde{\omega}_j$ , the FBLP estimator. Note that

$$\tilde{\omega}_j - \omega_j = O_p(n^{-1/2}) = O_p(n_1^{-1-1/4}).$$

Then substituting  $n_1 = n^{0.40}$ ,  $\hat{\omega}_j^{(0)} = \tilde{\omega}_j$  in (18) and applying Theorem 2, we obtain

$$\hat{\omega}_j^{(1)} - \omega_j = O_p(n_1^{-1-1/2}) = O_p(n^{-0.60}).$$

*Step 2.* With  $m = 2$ , choose  $n_2 = n^{0.48}$  and compute  $\hat{\omega}_j^{(2)}$  from  $\hat{\omega}_j^{(1)}$  using (18). Since  $\hat{\omega}_j^{(1)} - \omega_j = O_p(n^{-0.60}) = O_p(n_2^{-1-1/4})$  and using Theorem 2, we obtain

$$\hat{\omega}_j^{(2)} - \omega_j = O_p(n_2^{-1-1/2}) = O_p(n^{-0.72}).$$

Choosing  $n_3, \dots, n_6$  as given below and applying the main theorem in the same way as above, we have

Step 3.  $n_3 = n^{0.57}$  provides  $\hat{\omega}_j^{(3)} - \omega_j = O_p(n^{-0.87})$ .

Step 4.  $n_4 = n^{0.70}$  provides  $\hat{\omega}_j^{(4)} - \omega_j = O_p(n^{-1.04})$ .

Step 5.  $n_5 = n^{0.83}$  provides  $\hat{\omega}_j^{(5)} - \omega_j = O_p(n^{-1.25})$ .

Step 6.  $n_6 = n^{0.92}$ , provides  $\hat{\omega}_j^{(6)} - \omega_j = O_p(n^{-1.58})$ .

Finally take  $n_7 = n$ , and compute  $\hat{\omega}_j^{(7)}$  from  $\hat{\omega}_j^{(6)}$ . Now applying Theorem 2, we have

$$n^{3/2}(\hat{\omega}_j^{(7)} - \omega) \rightarrow N_p(\mathbf{0}, 6\sigma^2(\mathbf{D}^H \mathbf{D})^{-1}).$$

So far, all the methods we have discussed are based on the assumptions that the number of exponential components is known and also the data are equispaced, namely at  $t = 1, \dots, n$ . Professor Rao and his colleagues, see Bai *et al.* [6], provided a method to estimate the number of components and frequencies simultaneously when some observations are missing, i.e. the data need not be equispaced. Clearly, this method is applicable when the complete data are available. The method is developed based on the following assumptions on the error terms. The error random variables  $e(t)$ 's are i.i.d. and

$$E(e(1)) = 0, \quad E|e(1)|^2 = \sigma^2, \quad E|e(1)|^4 < \infty. \tag{19}$$

It is further assumed that  $p$  is also unknown and  $0 \leq p \leq P < \infty$ , where  $P$  is known. The method can be described as follows: For a given  $p$ , consider the  $(n - p) \times (p + 1)$  data matrix  $\mathbf{Z}$  as given below:

$$\mathbf{Z} = \begin{bmatrix} y(p+1) & y(p) & \dots & y(1) \\ y(p+2) & y(p+1) & \dots & y(2) \\ \vdots & \vdots & \vdots & \vdots \\ y(n) & y(n-1) & \dots & y(n-p) \end{bmatrix} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{n-p}]^\top.$$

It is assumed that the observations  $\{y(k); k \in \kappa_n\}$  are missing. Here  $\kappa_n \subset \{1, 2, \dots, n\}$ . Let  $\mathbf{A}_p$  be the matrix obtained from the matrix  $\mathbf{Z}$  after removing the rows having missing observations. Let  $r_p$  be the number of rows of  $\mathbf{A}_p$ . Let us define

$$\hat{\Gamma}^{(p)} = \frac{1}{r_p} \mathbf{A}_p^H \mathbf{A}_p = ((\hat{\gamma}_{m_j}^{(p)}))$$

and

$$S_p = \min\{r_p^{-1} \|\mathbf{A}_p \mathbf{g}^{(p)}\|^2 : \|\mathbf{g}^{(p)}\| = 1\}, \tag{20}$$

for  $p = 0, \dots, P$ , where  $\mathbf{g}^{(p)} = (g_0^{(p)}, g_1^{(p)}, \dots, g_p^{(p)})^\top$ . Note that  $S_p$  is the smallest eigenvalue of  $\hat{\Gamma}^{(p)}$ . Consider

$$R_p = S_p + pC_n, \quad p = 0, 1, \dots, P,$$

where  $C_n$  satisfies the following assumptions:

$$\lim_{n \rightarrow \infty} C_n = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{\sqrt{r_p} C_n}{\sqrt{\ln \ln r_p}} = \infty.$$

Then find  $\hat{p} \leq P$ , such that  $R_{\hat{p}} = \min_{0 \leq p \leq P} R_p$ , and  $\hat{p}$  is an estimate of  $p$ . Once  $\hat{p}$  is obtained  $\hat{\mathbf{g}}^{(\hat{p})}$  can be obtained from (20), by replacing  $p$  with  $\hat{p}$ . Once  $\hat{\mathbf{g}}^{(\hat{p})}$  is obtained the estimates of  $\omega_1, \dots, \omega_{\hat{p}}$  can be obtained from the polynomial equation (17). It has been shown by the authors that based on the error assumptions (19), the above method

provides consistent estimates for  $p$  as well as for  $\omega_1, \dots, \omega_p$ . If  $p$  is known, the consistent estimates of the frequencies when some observations are missing can be obtained in a slightly weaker conditions, namely

$$E(e(1)) = 0, \quad E|e(1)|^2 = \sigma^2 < \infty$$

(see for example, [32]).

So far, we have discussed mainly one-dimensional superimposed exponential model. Now we will discuss about the contributions of Professor Rao for two-dimensional superimposed exponential signals.

#### 4. Two-dimensional superimposed exponential model

Rao *et al.* [49] considered the two dimensional superimposed exponential signals which has the following form:

$$y(s, t) = \sum_{j=1}^p \sum_{k=1}^q \gamma_{jk} e^{i(s\mu_j + tv_k)} + w(s, t), \quad s = 1, \dots, m, \quad t = 1, \dots, n. \quad (21)$$

Here,  $y(s, t)$  is the observed signal,  $\mu_j \in (0, 2\pi)$  and  $v_k \in (0, 2\pi)$  are the unknown frequencies of the signals,  $\gamma_{jk}$  is the unknown complex-valued linear parameters,  $w(s, t)$  is an array of two-dimensional random variables with mean 0. The explicit assumptions of  $w(s, t)$  will be explained later. The problem is to extract the original signal from the noise corrupted data.

The problem had a long history. It was first considered by McClellan [38] to analyze two-dimensional symmetric images. Since then several other authors also considered this model (see for example, [7, 19, 24] and see the references cited therein). The corresponding real two-dimensional sum of sinusoidal model can be written as follows

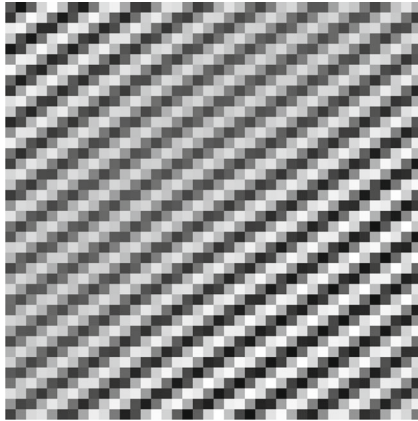
$$y(s, t) = \sum_{j=1}^p \sum_{k=1}^q \gamma_{jk} \cos(s\mu_j + tv_k) + \sin(s\mu_j + tv_k) + w(s, t), \quad s = 1, \dots, m, \quad t = 1, \dots, n. \quad (22)$$

In this case, similar to the one-dimensional model,  $y(s, t)$ ,  $\gamma_{jk}$  and  $w(s, t)$  are all real-valued. Some two-dimensional plots based on model (22) are provided in Figures 5 and 6.

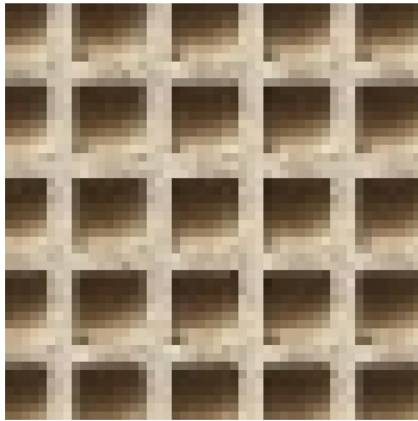
In this case also, the most popular method to estimate the frequencies is obtained through the periodogram function approach. The two-dimensional periodogram function can be written as follows:

$$I(\mu, \nu) = \left| \frac{1}{mn} \sum_{s=1}^m \sum_{t=1}^n e^{i(s\mu + t\nu)} \right|^2, \quad (23)$$

and the estimate of the frequencies can be obtained from the local maxima of  $I(\mu, \nu)$ . Similar to the one-dimensional model, it can be easily shown that when there is no noise, then the two-dimensional periodogram function has local maxima at the true frequencies. Therefore, when there is no noise or the noise variance is small, the true frequencies can be easily estimated, but the same is not true when the error variance is high. Due to this reason, several methods have been proposed in the literature and most of the methods are based on two-dimensional extension of the Prony's equations.



**Figure 5.** Two dimensional plot obtained from Model (22).



**Figure 6.** Another two dimensional plot obtained from Model (22).

In this case also, the most reasonable estimators will be the least squares estimators, and they can be obtained by minimizing

$$Q(\boldsymbol{\gamma}, \boldsymbol{\mu}, \mathbf{v}) = \frac{1}{mn} \sum_{s=1}^m \sum_{t=1}^n \left( y(s, t) - \sum_{j=1}^p \sum_{k=1}^q \gamma_{jk} e^{i(s\mu_j + t\nu_k)} \right)^2, \quad (24)$$

with respect to  $\boldsymbol{\gamma}$ ,  $\boldsymbol{\mu}$  and  $\mathbf{v}$ , where  $\boldsymbol{\gamma} = (\gamma_{11}, \dots, \gamma_{pq})^\top$ ,  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_p)^\top$  and  $\mathbf{v} = (\nu_1, \dots, \nu_q)^\top$ . If the least squares estimators of  $\gamma_{jk}$ ,  $\mu_j$  and  $\nu_k$  are denoted by  $\hat{\gamma}_{jk}$ ,  $\hat{\mu}_j$  and  $\hat{\nu}_k$ , respectively, then an estimator of  $\sigma^2$  becomes

$$\hat{\sigma}^2 = \frac{1}{mn} \sum_{s=1}^m \sum_{t=1}^n \left( y(s, t) - \sum_{j=1}^p \sum_{k=1}^q \hat{\gamma}_{jk} e^{i(s\hat{\mu}_j + t\hat{\nu}_k)} \right)^2.$$

The model (21) is also a non-linear regression model, but it can be shown (see for example, [31]) that the model does not satisfy some of the standard sufficient conditions which are available in the literature for the least squares estimators to be consistent and asymptotically normally distributed, similar to the one-dimensional model. Hence, although the

least squares estimators seem to be the most reasonable estimators, the consistency cannot be guaranteed. Rao *et al.* [49] provided the consistency and asymptotic normality results of the least squares estimators of the complex linear parameters and the frequencies under the following assumptions.

*Assumption 4.1.*  $\{w(s, t)\}$  is an array of complex-valued i.i.d. random variables. Moreover,  $\text{Re}(w(s, t))$  and  $\text{Im}(w(s, t))$  are normal random variables with mean 0 and variance  $\sigma^2/2$  each, and they are independently distributed.

*Assumption 4.2.* The frequencies  $\mu_1, \dots, \mu_p$  are different from each other and so are  $\nu_1, \dots, \nu_q$ .

*Assumption 4.3.*  $\|\mathbf{y}_j\|^2 = \sum_{k=1}^q |\gamma_{jk}|^2 > 0$ ,  $\|\mathbf{y}_{\cdot k}\|^2 = \sum_{j=1}^p |\gamma_{jk}|^2 > 0$ , where  $\mathbf{y}_j = (\gamma_{j1}, \dots, \gamma_{jq})^\top$ ,  $\mathbf{y}_{\cdot k} = (\gamma_{1k}, \dots, \gamma_{pk})^\top$  and  $\|\cdot\|$  denotes the  $L_2$  norm of a vector or a matrix.

*Assumption 4.4.* There are positive constants  $c_1, c_2, \alpha_1$  and  $\alpha_2$ , such that

$$c_1 n^{\alpha_1} \leq m \leq c_2 n^{\alpha_2}.$$

We need the following notations:

$$\begin{aligned} \Delta_j &= \hat{\mu}_j - \mu_j, \quad \delta_k = \hat{\nu}_k - \nu_k, \quad \mathbf{\Delta} = (\Delta_1, \dots, \Delta_p)^\top, \quad \mathbf{\delta} = (\delta_1, \dots, \delta_q)^\top, \\ \mathbf{\Gamma}_{\mathbf{\Delta}} &= \text{diag}(\|\gamma_{1\cdot}\|, \dots, \|\gamma_{p\cdot}\|)^\top; \quad p \times p, \\ \mathbf{\Gamma}_{\mathbf{\delta}} &= \text{diag}(\|\gamma_{\cdot 1}\|, \dots, \|\gamma_{\cdot q}\|)^\top; \quad q \times q, \\ \mathbf{G}_{\mathbf{\Delta}} &= \text{diag}(\mathbf{y}_1, \dots, \mathbf{y}_p)^\top; \quad pq \times p, \\ \mathbf{G}_{\mathbf{\delta}} &= (\text{diag}(\gamma_{11}, \dots, \gamma_{1q}), \dots, \text{diag}(\gamma_{p1}, \dots, \gamma_{pq}))^\top; \quad pq \times q. \end{aligned}$$

Now based on the above notations we can state the main result without the proof.

**Theorem 3.** *If Assumptions 4.1–4.4 are met, then  $\hat{\boldsymbol{\gamma}}$  and  $\hat{\sigma}^2$  are strongly consistent estimates of  $\boldsymbol{\gamma}$  and  $\sigma^2$ , respectively, as  $n \rightarrow \infty$ . Moreover, the limiting distribution of*

$$\begin{aligned} &(m^{3/2} n^{1/2} \mathbf{\Delta}^\top, m^{1/2} n^{3/2} \mathbf{\delta}^\top, \sqrt{mn} \text{Re}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})^\top, \\ &\sqrt{mn} \text{Im}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})^\top, \sqrt{mn} (\hat{\sigma}^2 - \sigma^2))^\top \end{aligned}$$

*is a  $(2pq + p + q + 1)$ -variate normal distribution with mean vector zero, and the covariance matrix  $\sigma^2 \boldsymbol{\Sigma} = \sigma^2 (\boldsymbol{\Sigma}_{kl})$ ,  $k, l = 1, \dots, 5$ , where*

$$\begin{aligned} \boldsymbol{\Sigma}_{11} &= 6\boldsymbol{\gamma}_{\mathbf{\Delta}}^{-2}, \quad \boldsymbol{\Sigma}_{13} = \boldsymbol{\Sigma}_{31}^\top = 3 \text{Im}(\boldsymbol{\Gamma}_{\mathbf{\Delta}}^{-2} \mathbf{G}_{\mathbf{\Delta}}^\top), \quad \boldsymbol{\Sigma}_{14} = \boldsymbol{\Sigma}_{41}^\top = -3 \text{Re}(\boldsymbol{\Gamma}_{\mathbf{\Delta}}^{-2} \mathbf{G}_{\mathbf{\Delta}}^\top), \\ \boldsymbol{\Sigma}_{22} &= 6 \boldsymbol{\Gamma}_{\mathbf{\delta}}^{-2}, \quad \boldsymbol{\Sigma}_{23} = \boldsymbol{\Sigma}_{32}^\top = 3 \text{Im}(\boldsymbol{\gamma}_{\mathbf{\delta}}^{-2} \mathbf{G}_{\mathbf{\delta}}^\top), \quad \boldsymbol{\Sigma}_{24} = \boldsymbol{\Sigma}_{42}^\top = 3 \text{Re}(\boldsymbol{\Gamma}_{\mathbf{\delta}}^{-2} \mathbf{G}_{\mathbf{\delta}}^\top), \\ \boldsymbol{\Sigma}_{33} &= \frac{3}{2} (\text{Im}(\mathbf{G}_{\mathbf{\Delta}}) \boldsymbol{\Gamma}_{\mathbf{\Delta}}^{-2} \text{Im}(\mathbf{G}_{\mathbf{\Delta}}^\top)) + (\text{Im}(\mathbf{G}_{\mathbf{\delta}}) \boldsymbol{\Gamma}_{\mathbf{\delta}}^{-2} \text{Im}(\mathbf{G}_{\mathbf{\delta}}^\top)) + \frac{1}{2} \mathbf{I}_{pq}, \end{aligned}$$

$$\begin{aligned}\Sigma_{34} &= \Sigma_{43}^\top = -\frac{3}{2}(\text{Im}(\mathbf{G}_\Delta)\Gamma_\Delta^{-2}\text{Re}(\mathbf{G}_\Delta^\top)) + (\text{Im}(\mathbf{G}_\delta)\Gamma_\delta^{-2}\text{Re}(\mathbf{G}_\delta^\top)), \\ \Sigma_{44} &= \frac{3}{2}(\text{Re}(\mathbf{G}_\Delta)\Gamma_\Delta^{-2}\text{Re}(\mathbf{G}_\Delta^\top)) + (\text{Re}(\mathbf{G}_\delta)\Gamma_\delta^{-2}\text{Re}(\mathbf{G}_\delta^\top)) + \frac{1}{2}\mathbf{I}_{pq}, \\ \Sigma_{55} &= \sigma^2,\end{aligned}$$

and the remaining  $\Sigma_{kl}$ 's are zeros.

In this case also, the most interesting feature is that the maximum likelihood estimators of the linear parameters have the convergence rate as  $O_p(m^{-1/2}n^{-1/2})$ , whereas the maximum likelihood estimators of the frequencies have the convergence rate as  $O_p(m^{-3/2}n^{-1/2})$  and  $O_p(m^{-1/2}n^{-3/2})$ . Rao *et al.* [49] provided the results when the errors follow normal distributions. The results have been extended by Kundu and Gupta [31] for i.i.d. errors and further it has been extended by Zhang and Mandrekar [56] and Kundu and Nandi [35] for stationary random fields.

Although the theoretical properties of the least squares estimators have been established, finding the least squares estimates is a numerically challenging problem. Due to this reason, several sub-optimal estimators have been proposed by several authors (see for example, [7, 12, 13, 19, 24] and the references cited therein). But none of these authors discussed the convergence rates of the estimators theoretically. Miao *et al.* [39] and Nandi *et al.* [40] provided two efficient algorithms which produce estimators, which have the same convergence rates as the least squares estimators.

Professor Rao has worked on another important signal processing model and it is known as the direction of arrival (DOA) model. Now we will provide his contribution briefly in the next section.

## 5. Direction of arrival (DOA) model

The DOA model can be written as follows:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t), \quad t = 1, \dots, n. \quad (25)$$

Here at time  $t$ ,  $\mathbf{x}(t)$  is a  $p \times 1$  complex vector of observations received by  $p$  sensors uniformly spaced,  $\mathbf{s}(t)$  is a  $q \times 1$  complex vector of unobservable signals emitted from  $q$  sources, and  $\mathbf{n}(t)$  is a  $p \times 1$  complex-valued noise vector. The number of sources is less than the number of sensors. The matrix  $\mathbf{A}$  has a special structure with its  $k$ -th column  $\mathbf{a}_k = \mathbf{a}(\tau_k)$  of the form

$$\mathbf{a}_k^\top = \mathbf{a}(\tau_k)^\top = (1, e^{-i\omega_0\tau_k}, \dots, e^{-i\omega_0(p-1)\tau_k}),$$

here as before,  $i = \sqrt{-1}$ ,  $\tau_k = c^{-1}\Delta \sin \theta_k$ ,  $c$  = speed of propagation,  $\theta_k$  is the DOA from the  $k$ -th source, and  $\Delta$  is the inter sensor distance. Without loss of generality,  $\omega_0$  is taken to be unity. It is assumed that  $\{\mathbf{s}(t)\}$  and  $\{\mathbf{n}(t)\}$  are independent sequences of i.i.d. random variables with

$$\mathbb{E}[\mathbf{n}(t)\mathbf{n}(t)^H] = \sigma^2\mathbf{I}_p, \quad \mathbb{E}[\mathbf{s}(t)\mathbf{s}(t)^H] = \mathbf{\Gamma} \quad \text{and} \quad \text{Rank}(\mathbf{\Gamma}) = q. \quad (26)$$

There are two important problems connected with this model. One is the estimation of  $q$ , the number of sources and the other is the estimation of  $\tau_1, \dots, \tau_q$ , providing the estimates of  $\theta_1, \dots, \theta_q$ , the DOA of signals.



This model has been used in censor array processing [11,21,25], in harmonic analysis [41], in retrieving the poles of a system from natural responses [54] and also in retrieving overlapping echoes from radar backscatter [43]. This model has received considerable amount of attention in the signal processing literature. Estimation of  $\tau_1, \dots, \tau_q$  assuming  $q$  known is usually solved by some eigen-decomposition methods available in the literature (for example, MUSIC [9,51], ESPRIT, TLS-ESPRIT [50], GESE [43]). For detailed discussions of the different eigen-decomposition methods, interested readers are referred to the Ph.D. thesis of Kannan [22] or the review article of Paulraj et al. [42]. The estimation of  $q$  is obtained by using some information theoretic criterion (see for example, [57,58] in this respect).

First, let us address the estimation of  $q$ , and then we provide the estimation of the DOA of signals. In addition to the assumption (26), if it is further assumed that  $\{\mathbf{n}(t)\}$  has a  $p$ -variate complex normal distribution, then the sample variance covariance matrix

$$\mathbf{S} = \frac{1}{n} \sum_{t=1}^n \mathbf{x}(t)\mathbf{x}(t)^H$$

has a complex Wishart distribution with  $n$  degrees of freedom and the covariance matrix

$$\mathbf{\Sigma} = \mathbf{A}\mathbf{\Gamma}\mathbf{A}^H + \sigma^2\mathbf{I}.$$

Since  $\mathbf{\Gamma}$  is non-singular, then the number of signals is  $q = p - s$ , where  $s$  is the multiplicity of the smallest eigenvalue of  $\mathbf{\Sigma}$ . Hence, the problem of estimation of  $q$  can be studied within the framework of testing the equality of the given number of smallest eigenvalues of  $\mathbf{\Sigma}$ .

The log-likelihood function of the observed data without the additive constant can be written as

$$l_n^{(q)} = -\frac{n}{2} (\ln |\mathbf{\Sigma}| + \text{tr } \mathbf{\Sigma}^{-1}\mathbf{S}).$$

The maximum of  $l_n^{(q)}$  for a given  $q$  is

$$\hat{l}_n^{(q)} = -\frac{n}{2} \left( \sum_{i=1}^q \ln \hat{\lambda}_i + (p - q) \ln \frac{\hat{\lambda}_{q+1} + \dots + \hat{\lambda}_p}{p - q} \right),$$

where  $\hat{\lambda}_1 > \dots > \hat{\lambda}_p$  are the ordered eigenvalues of  $\mathbf{S}$  and they are distinct with probability 1. The likelihood ratio test statistic for testing the equality of the last  $p - q$  eigenvalues of  $\mathbf{\Sigma}$  is

$$G_n^{(q)} = 2(l_n^{(p)} - l_n^{(q)}) = n(p - q) \ln \frac{\hat{\lambda}_{q+1} + \dots + \hat{\lambda}_p}{p - q} - \ln(\hat{\lambda}_{q+1} + \dots + \hat{\lambda}_p)$$

and it is asymptotically distributed as  $\chi^2$  with  $(p - q)^2 - 1$  degrees of freedom. Let

$$\hat{q} = \max\{q : G_n^{(q)} \leq c_\alpha\},$$

where  $c_\alpha$  is the upper  $\alpha$ -th percentile point of a  $\chi^2$  distribution with  $(p - q)^2 - 1$  degrees of freedom, then an upper  $100(1 - \alpha)\%$  confidence limit to  $q$  is  $\hat{q}$ .

Further, to get a point estimate of  $q$ , a general information theoretic criterion can be used as follows. Consider

$$\text{GIC}(k) = 2 \ln_n^{(k)} - \nu(k)c_n,$$

where  $\nu(k)$  is the number of free parameters in the model and  $c_n$  is a function of  $n$ , such that as  $n \rightarrow \infty$ ,

$$\frac{c_n}{n} \rightarrow 0 \quad \text{and} \quad \frac{c_n}{\ln \ln n} \rightarrow \infty. \quad (27)$$

In this case,  $\nu(k) = q(2p - q) + 1$  and the estimate of  $q$  is  $\hat{q}$ , where

$$\text{GIC}(\hat{q}) = \max_k \text{GIC}(k).$$

It has been shown that  $\hat{q}$  is a strongly consistent estimate of  $q$ . Note that the choice of  $c_n$  is quite arbitrary except that it satisfies (27). Moreover, the performance of the method depends on the choice of  $c_n$ . Not enough studies have been performed to choose  $c_n$ . Some attempt has been made by Kundu [30] in this direction.

Now we discuss about the estimation of DOA of signals. From now on, it is assumed that  $q$  is known. Note that for a given  $q$ , the eigenvalues of  $\Sigma$  are of the form

$$\lambda_1 \geq \dots \geq \lambda_q > \lambda_{q+1} = \dots = \lambda_p = \sigma^2.$$

Let  $\mathbf{e}_1, \dots, \mathbf{e}_q, \mathbf{e}_{q+1}, \dots, \mathbf{e}_p$  be the corresponding eigenvectors and define matrices

$$\mathbf{E}_s = (\mathbf{e}_1, \dots, \mathbf{e}_q) \quad \text{and} \quad \mathbf{E}_n = (\mathbf{e}_{q+1}, \dots, \mathbf{e}_p).$$

The vector space generated by  $\mathbf{E}_s$  is called the signal space and by  $\mathbf{E}_n$  is called the noise space. Observe that

$$\mathbf{a}(\tau)^\top \mathbf{E}_n = \mathbf{0} \Leftrightarrow D(\tau) = \mathbf{a}(\tau)^\top \mathbf{E}_n \mathbf{E}_n^\top \mathbf{a}(\tau) = 0,$$

for  $\tau = \tau_i, i = 1, 2, \dots, q$ . In practice we do not have  $\Sigma$ , we only have an estimate of  $\Sigma$ , i.e.  $\mathbf{S}$ . Let the eigenvalues and the corresponding eigenvectors of  $\mathbf{S}$  be

$$\hat{\lambda}_1 > \dots > \hat{\lambda}_q > \dots > \hat{\lambda}_p, \\ \hat{\mathbf{e}}_1, \dots, \hat{\mathbf{e}}_q, \dots, \hat{\mathbf{e}}_p$$

and they provide consistent estimates of the corresponding eigenvalues and eigenvectors of  $\Sigma$ . Moreover, an estimate of  $\mathbf{E}_n$ , spanning the noise eigenspace is

$$\hat{\mathbf{E}}_n = (\hat{\mathbf{e}}_{q+1} \dots \hat{\mathbf{e}}_p).$$

The function

$$\hat{D}(\tau) = \mathbf{a}(\tau)^\top \hat{\mathbf{E}}_n \hat{\mathbf{E}}_n^\top \mathbf{a}(\tau)$$

may not vanish at any  $\tau$ , but in the neighborhood of  $\tau_1, \dots, \tau_q$ , it is expected to be small. This is the main idea of the MUSIC algorithm as proposed by Bienvenu and Kopp [9]. Using the conjugate symmetry property of the polynomial coefficients, as described in Section 2, the MUSIC algorithm has been modified by Kundu [28]. It is important to note that in both MUSIC or modified MUSIC algorithms, the solutions are obtained by a search method. Hence, although the consistency property of the estimators can be established, the asymptotic distributions are difficult to study. The following approach may be used to study the asymptotic properties of the estimators.

Note that there exists a matrix  $\mathbf{G}$  of order  $p \times (p - q)$ ,

$$\mathbf{G} = (\mathbf{G}_1, \dots, \mathbf{G}_{p-q}), \tag{28}$$

where the  $p \times 1$  vector  $\mathbf{G}_k$  for  $k = 1, \dots, (p - q)$  is of the form

$$\mathbf{G}_k = (0, \dots, 0, g_0, \dots, g_q, 0, \dots, 0)^\top,$$

with the first  $k$  and the last  $(p - k - q - 1)$  elements are zeros,  $g_q > 0, g_0 \bar{g}_0 + \dots + g_q \bar{g}_q = 1$ , such that  $\mathbf{a}(\tau)^\top \mathbf{G} = \mathbf{0}$ , for  $\tau = \tau_1, \dots, \tau_q$ , and  $e^{i\tau_1}, \dots, e^{i\tau_q}$  are the solutions of the polynomial equation

$$g_0 + g_1 z + \dots + g_q z^q = 0.$$

Therefore, the statistical problem is to estimate the vector  $\mathbf{g} = (g_0, \dots, g_q)^\top$  from the given observation. Since the columns of  $\mathbf{G}$  generate the noise eigenspace of  $\hat{\mathbf{E}}_n$ , the problem of estimating  $\mathbf{G}$  can be thought of fitting a basis of the type  $\mathbf{G}_1, \dots, \mathbf{G}_{p-q}$  to the estimated noise eigenspace  $\hat{\mathbf{E}}_n$ . Mathematically, it can be formulated as minimizing the Euclidean norm

$$\|\hat{\mathbf{E}}_n - \hat{\mathbf{G}}\mathbf{B}\|$$

with respect to a matrix  $\mathbf{G}$  of the form (28) and an arbitrary matrix  $\mathbf{B}$  of order  $(p - q) \times (p - q)$ . Since this optimization problem is quite complicated numerically and explicit solutions do not exist, Bai and Rao [4] proposed the following method. By Householder transformation, i.e. multiplying by an unitary matrix  $\mathbf{O}$  of order  $(p - q) \times (p - q)$ , convert  $\hat{\mathbf{E}}_n$  into the form

$$\hat{\mathbf{E}}_n\mathbf{O} = (\mathbf{u}_{q+1}, \dots, \mathbf{u}_p),$$

where  $\mathbf{u}_{q+i} = (u_{0,q+i}, \dots, u_{q+i-1,q+i}, 0, \dots, 0)^\top$ , for  $i = 1, \dots, (p - q)$  with  $u_{q+i-1,q+i} \neq 0$  (with probability one). Solve the equation

$$u_{0,q+1} + \dots + u_{q,q+1}z^q = 0,$$

obtain the roots of the form  $\hat{\rho}_1 e^{i\hat{\tau}_1}, \dots, \hat{\rho}_q e^{i\hat{\tau}_q}$  and choose  $\hat{\tau}_1, \dots, \hat{\tau}_q$  as estimates of  $\tau_1, \dots, \tau_q$ .

We need the following assumptions to establish the asymptotic properties of the proposed estimators:

*Assumption 5.1.* The second moment of  $\mathbf{x}(t)$  exists.

*Assumption 5.2.* The fourth moment of  $\mathbf{x}(t)$  exists.

*Assumption 5.3.*  $\text{Var}[\text{Re } \mathbf{s}(t)] = \text{Var}[\text{Im } \mathbf{s}(t)] = 2^{-1} \text{Re } \mathbf{\Gamma}$ ,  $\text{E}[(\text{Re } \mathbf{s}(t))(\text{Im } \mathbf{s}(t))^\top] = -\text{E}[(\text{Im } \mathbf{s}(t))(\text{Re } \mathbf{s}(t))^\top] = 2^{-1} \text{Im } \mathbf{\Gamma}$ .

*Assumption 5.4.*  $\text{Var}[\text{Re } \mathbf{n}(t)] = \text{Var}[\text{Im } \mathbf{n}(t)] = 2^{-1} \sigma^2 \mathbf{I}_p$ ,  $\text{Cov}[(\text{Re } \mathbf{n}(t), \text{Im } \mathbf{n}(t))] = \mathbf{0}$ .

*Assumption 5.5.*  $\text{E}[\text{Re } n_k(t)]^4 = \text{E}[\text{Im } n_k(t)]^4 = \frac{3}{4} \sigma^4$ ,  $k = 1, \dots, p$ ;  $\text{E}(\text{Re } n_k(t))^2 (\text{Re } n_h(t))^2 = \text{E}(\text{Im } n_k(t))^2 (\text{Im } n_h(t))^2 = 4^{-1} \sigma^4$ ,  $k \neq h = 1, \dots, p$ ;  $\text{E}(\text{Re } n_k(t))^2 (\text{Im } n_h(t))^2 = 4^{-1} \sigma^4$ ,  $k, h = 1, \dots, p$ , all other kinds of fourth moments are zero.

The following results regarding the asymptotic properties of the proposed estimators have been obtained by Bai *et al.* [3].

**Theorem 4.** Under Assumption 5.1,  $(\hat{\tau}_1, \dots, \hat{\tau}_q)^\top$  is a strongly consistent estimator of  $(\tau_1, \dots, \tau_q)^\top$ .

**Theorem 5.** Under Assumptions 5.2–5.4, the limiting distribution of  $\sqrt{n}(\hat{\tau}_1 - \tau_1, \dots, \hat{\tau}_q - \tau_q)^\top$  is a  $q$ -variate normal with mean vector zero and covariance matrix

$$2^{-1} \text{Re}[\mathbf{G}^{-1}(\sigma^4 \mathbf{\Gamma}^{-1}(\mathbf{A}^H \mathbf{A})^{-1} \mathbf{\Gamma}^{-1} + \sigma^2 \mathbf{\Gamma}^{-1})\mathbf{G}^{-1}],$$

where

$$\mathbf{G} = \text{diag}(D(\tau_1), \dots, D(\tau_q)),$$

$$D(\tau_k) = g_{q+1} \prod_{j=1, j \neq k}^q (e^{i\tau_k} - e^{i\tau_j}) e^{i\tau_k}, \quad k = 1, \dots, q.$$

Using the conjugate symmetric properties of the polynomial coefficients as discussed in Section 2, the method has been modified and the associated asymptotic results have been established by Kundu [29]. The method has been extended for the multiple moving targets in a sequence of papers by Rao *et al.* [44–47].

## 6. Some recent developments and conclusions

Professor Rao has worked in this area of statistical signal processing for about 6–7 years. But it has left a long lasting impact among some of the statisticians who are working in this field. In recent time one related model has received a considerable attention in the area of signal processing and it is known as the chirp model. A one dimensional multiple chirp model can be expressed as

$$y(t) = \sum_{k=1}^p \alpha_k e^{i(\omega_k t + \theta_k t^2)} + e(t). \quad (29)$$

Here also as before,  $y(t)$ 's,  $\alpha_k$ 's and  $e(t)$ 's are complex valued. As the multiple exponential model  $\omega_k$ 's are known as frequency, and  $\theta_k$ 's are known as frequency rate. This model appears in many areas of signal processing, one of the most important one being the radar problem. It can also appear in active sonar and passive sonar system. The problem of parameter estimation of chirp signal has received a considerable amount of attention in the signal processing literature (see for example, [1, 14, 17] and the references cited therein for different estimation procedures). Several statistical aspects related to this model, asymptotic properties of different estimators of this model and efficient estimation procedures have been developed by Lahiri [37] and Grover [16]. It should be mentioned that both Lahiri [37] and Grover [16] have successfully extended many of the methods proposed by Rao and his collaborators to establish their results.

Grover [16] has proposed a chirp-like model which has the following form:

$$y(t) = \sum_{k=1}^p \alpha_k e^{i\omega_k t} + \sum_{k=1}^q \beta_k e^{i\theta_k t^2} + e(t). \quad (30)$$

It is observed that the proposed chirp-like model (30) behaves very similarly to the traditional chirp model (29). On the other hand, estimation of the parameters for the model (30) is less challenging compared to the model (29). Efficient estimation procedure has been developed by Grover [16] for the model (30). In principle, it should be possible to generalize for two-dimensional case also, which has several applications in image processing, particularly in finger printing. More work is needed along this direction.

## Acknowledgements

The author is thankful to Professor Partha Pratim Mazumdar, the President of Indian Academy of Sciences, for organizing the birth centenary symposium of the legendary

scientist Professor Rao on December 11–12, 2019 at Bengaluru and also for inviting the author to give a talk. The paper is based on the presentation given at that symposium. Part of the work has been supported by a grant from SERB, Government of India.

## References

- [1] Abatzoglou T, Fast maximum likelihood joint estimation of frequency and frequency rate, *IEEE Trans. Aerosp. Electron. Syst.* **22** (1986) 708–715
- [2] Bai Z D, Krishnaiah R and Zhao L C, On the simultaneous estimation of the number of signals and frequencies under a model with multiple sinusoids, Technical Report, 86–37 (1986) (Center for Multivariate Analysis, University of Pittsburgh)
- [3] Bai Z D, Miao B Q and Rao C R, Estimation of direction of arrival of signals: asymptotic results, *Advances in Spectrum Analysis and Array Processing*, edited by S Kaykin (1991) (NJ: Prentice Hall, Englewood, Cliffs) vol. II, Chapter 9
- [4] Bai Z D and Rao C R, Spectral analytic methods for the estimation of signals and direction of arrival, *Spectral Analysis in One or Two dimensions*, edited by S Prasad and R L Kashyap (1989) (New Delhi: Oxford and IBH Publications Co.) pp. 493–507
- [5] Bai Z D, Rao C R, Chow M and Kundu D, An efficient algorithm for estimating the parameters of superimposed exponential signal, *J. Stat. Plan. Inference* (2003) vol. 110, pp. 23–34
- [6] Bai Z D, Rao C R, Wu Y, Zen M-M and Zhao L, The simultaneous estimation of the number of signals and frequencies of multiple sinusoids when some observations are missing: I. Asymptotic, *Proc. Natl. Acad. Sci.* **96** (1999) 11106–11110
- [7] Barbieri M M and Barone P, A two dimensional Prony’s method for spectral estimation, *IEEE Trans. Signal Process.* **40(11)** (1992) 2747–2756
- [8] Bates D M and Watts G W, *Nonlinear Regression Analysis and Its Applications* (1988) (New York: Wiley)
- [9] Bienvenu G and Kopp L, Adaptivity of background noise spatial coherence for high resolution passive methods, *Proceedings of the ICASSP 80* (1980) (CO: Denver) pp. 307–310
- [10] Bresler Y and Macovski A, Exact maximum likelihood parameter estimation of superimposed exponential signals in noise, *IEEE Trans. Acoust. Speech Signal Process.* **A55(10)** (1986) pp. 1081–1089
- [11] Bronez T P and Cadzow J A, An algebraic approach to super-resolution array processing, *IEEE Trans. Aerosp. Electron. Syst.* **19** (1983) 123–133
- [12] Cabrera S D and Bose N K, Prony’s method for two-dimensional complex exponential modelling, (Chapter 15), *Applied Control Theory*, edited by S G Tzafestas (1993) (New York: Marcel and Dekker) pp. 401–411
- [13] Chun J and Bose N K, Parameter estimation via signal selectivity of signal subspaces (PESS) and its applications, *Digital Signal Process.* **5** (1995) 58–76
- [14] Djurić P M and Kay S M, Parameter estimation of chirp signals, *IEEE Trans. Acoust. Speech Signal Process.* **38** (1990) 2118–2126
- [15] Froberg C E, *Introduction to Numerical Analysis*, 2nd edition (1969) (Boston: Addison-Wesley Publ. Co.)
- [16] Grover R, *Frequency and Frequency Rate Estimation of Some Non-stationary Signal Processing Models*, Ph.D. thesis (2020) (Kanpur: Indian Institute of Technology Kanpur)
- [17] Guo J, Zou H, Yang X and Liu G, Parameter estimation of multicomponent chirp signals via sparse representation, *IEEE Trans. Aerosp. Electron. Syst.* **47** (2011) 2261–2268
- [18] Hilderband F B, *An Introduction to Numerical Analysis* (1956) (New York: McGraw Hill)
- [19] Hua Y, Estimating two-dimensional frequencies by matrix enhancement and matrix pencil, *IEEE Trans. Signal Process.* **40** (1992) 2267–2280
- [20] Jennrich R I, Asymptotic properties of the non-linear least squares estimation, *Ann. Math. Stat.* **40** (1969) 633–643

- [21] Johnson D H and Degra S R, Improving the resolution of bearing in passive sonar arrays by eigenanalysis, *IEEE Trans. Acoust. Speech Signal Process* **29** (1982) 401–413
- [22] Kannan N, Estimation of Direction of Arrival of Signals in Signal Processing, Ph.D. thesis (1992) (USA: Department of Statistics, The Pennsylvania State University, University Park)
- [23] Kannan N and Kundu D, On modified EVLP and ML methods for estimating superimposed exponential signals, *Signal Process.* **39** (1994) 223–233
- [24] Kay S M and Nekovei R, An efficient two-dimensional frequency estimator, *IEEE Trans. Acoust. Speech Signal Process.* **38** (1990) 1807–1809
- [25] Kumaresan R and Tufts D, Estimation of arrival of multiple plane waves, *IEEE Trans. Aerosp. Electron. Syst.* **19** (1983) 134–139
- [26] Kundu D, Asymptotic properties of the complex valued non-linear regression model, *Commun. Stat. Theory Methods* **20** (1991) 3793–3803
- [27] Kundu D, Estimating the parameters of undamped exponential signals, *Technometrics* **35** (1993) 215–218
- [28] Kundu D, Modified MUSIC algorithm for estimating DOA of signals, *Signal Process.* **48** (1996) 85–90
- [29] Kundu D, Estimating DOA of signals: some asymptotic results *Sankhya Ser. A* **61** (1999) 120–138
- [30] Kundu D, Estimating the number of signals in the presence of white noise, *J. Stat. Plan. Inference* **90** (2000) 57–68
- [31] Kundu D and Gupta R D, Asymptotic properties of the least squares estimators of a two dimensional texture model, *Metrika* **48** (1998) 83–97
- [32] Kundu D and Kundu R, Consistent estimates of super imposed exponential signals when some observations are missing, *J. Stat. Plan. Inference* **44** (1995) 205–218
- [33] Kundu D and Mitra A, A note on the consistency of the undamped exponential signals model, *Statistics* **28** (1996) 25–33
- [34] Kundu D and Mitra A, On asymptotic properties and confidence intervals for exponential signals, *Signal Process.* **72** (1999) 129–139
- [35] Kundu D and Nandi S, Determination of discrete spectrum in a stationary random field, *Stat. Neerl.* **57** (2003) 258–284
- [36] Kundu D and Nandi S, *Statistical Signal Processing: Frequency Estimation* (2012) (New Delhi: Springer)
- [37] Lahiri A, Estimators of Parameters of Chirp Signals and their Properties, Ph.D. thesis (2012) (Kanpur: Indian Institute of Technology Kanpur)
- [38] McClellan J H, Multidimensional spectral estimation, *Proc. IEEE* **70** (1982) 1029–1039
- [39] Miao B Q, Wu Y and Zhao L C, On strong consistency of a two-dimensional frequency estimation algorithm, *Stat. Sin.* **8** (1998) 559–570
- [40] Nandi S, Prasad A and Kundu D, An efficient and fast algorithm for estimating the parameters of two-dimensional sinusoidal signals, *J. Stat. Plan. Inference* **140** (2010) 153–168
- [41] Owsley N L, Spectral signal set extraction, *Aspects of Signal Processing, Part II*, edited by G Eacconi (1977) pp. 469–475
- [42] Paulraj A, Otteresten B, Roy R, Swindlehurst A, Xu G and Kailath T, Subspace methods for direction of arrival estimation, *Handbook of Statistics*, edited by N K Bose and C R Rao (1993) (Amsterdam: Elsevier Science Publishers) vol. 10, pp. 693–739 (Chapter 16)
- [43] Pillai S U, *Array Signal Processing* (1989) (New York: Springer)
- [44] Rao C R, Sastry C R and Zhou B, Tracking the direction of arrival of multiple moving targets, *IEEE Trans. Signal Process.* **42(5)** (1994) 1133–1144
- [45] Rao C R, Zhang L and Zhao L C, New algorithm for multitarget angle tracking, *IEE Proc. F* **140(5)** (1993) 335–338
- [46] Rao C R, Zhang L and Zhao L C, Multiple target angle tracking using sensor array outputs, *IEEE Trans. Aerosp. Electron. Syst.* **29** (1993) 268–271
- [47] Rao C R, Zhang L and Zhao L C, Multitarget angle tracking an algorithm for data association, *IEEE Trans. Signal Process.* **42(2)** (1994) 459–462

- [48] Rao C R and Zhao L C, Asymptotic behavior of maximum likelihood estimators of superimposed exponential signals, *IEEE Trans. Signal Process.* **41(3)** (1993) 1461–1464
- [49] Rao C R, Zhao L and Zhou B, Maximum likelihood estimation of 2-D superimposed exponential signals, *IEEE Trans. Signal Process.* **42(7)** (1994) 1795–1802
- [50] Roy R, Paulraj A and Kailath T, ESPRIT—a subspace rotation approach to estimation of parameters of cisoids in noise, *IEEE Trans. Acoust. Speech Signal Process.* **34** (1986) 1340–1342
- [51] Schmidt R O, A Signal Subspace Approach to Multiple Emitter Location and Spectral Estimation, Ph.D. thesis (1981) (USA: Stanford University)
- [52] Seber G A F and Wild C J, *Nonlinear Regression* (1989) (New York: Wiley)
- [53] Stoica P and Moses R, *Spectral Analysis of Signals* (2005) (New Jersey, USA: Prentice Hall)
- [54] Wax M, Schmidt R O and Kailath T, Eigen-structure method for retrieving the poles from the natural response, *Proceedings of Second CDC, TX* (1983) pp. 1343–1344
- [55] Wu C F J, Asymptotic theory of non-linear least squares estimation, *Ann. Stat.* **9** (1981) 501–513
- [56] Zhang H and Mandrekar V, Estimation of hidden frequencies for 2D stationary processes, *J. Time Ser. Anal.* **22(5)** (2001) 613–629
- [57] Zhao L C, Krishnaiah P R and Bai Z D, On detection of the number of signals in presence of white noise, *J. Multivar. Anal.* **20** (1986a) 1–25
- [58] Zhao L C, Krishnaiah P R and Bai Z D, On detection of the number of signals when the noise covariance matrix is arbitrary, *J. Multivar. Anal.* **20** (1986b) 26–49