A Review on Chirp and Some Other Related Signal Processing Models

Debasis Kundu *and Rhythm Grover †and Swagata Nandi ‡

Abstract

Analyzing periodic or nearly periodic signals is a fundamental problem in Signal Processing. If the signal is periodic, then the sum of sinusoidal model has been used quite extensively to analyze the signal. But if the signal is nearly periodic, then sinusoidal model may not be very effective. Different alternative models have been proposed in the Signal Processing literature to analyze nearly periodic signal. The Chirp model is one such model which has been used quite effectively to analyze nearly periodic signal. The Chirp model has received a considerable amount of attention in the Signal Processing literature and an extensive amount of work has been published in this topic during the last three decades. The main aim of this chapter is to provide a comprehensive review of chirp and various related models which are available in the literature and which have been used to analyze periodic and nearly periodic data both in one and two dimensions. We would like to provide some new results and several open problems for future research.

1 Introduction

We observe periodic phenomenon everyday in our daily lives. For example, the number of tourists visits the famous Taj Mahal in India, the daily average temperature in Toronto, Canada, the number passengers travel daily between New York and London, the Electro

*Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Pin 208016, India. E-mail: kundu@iitk.ac.in, Phone no. 91-512-2597141, Fax no. 91-512-2597500.

†Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Pin 208016, India. E-mail: rhythm@iitk.ac.in

‡Division of Theoretical Statistics and Mathematics, Indian Statistical Institute, Delhi Center, 7 SJS Sansanwal Marg, New Delhi, Pin 110016, India. E-mail: nandi@isid.ac.in

Cardio Graph (ECG) signal of a normal human being are all periodic in nature. Sometimes the data or signals may not be exactly periodic but it may be nearly periodic. See for example the following figures which indicate periodic nature of a signal. Figure 1: the ECG plot of a normal human being and Figure 2: 'UUU' vowel sound of a male. Both of them indicate periodic/ nearly periodic nature of the signal.

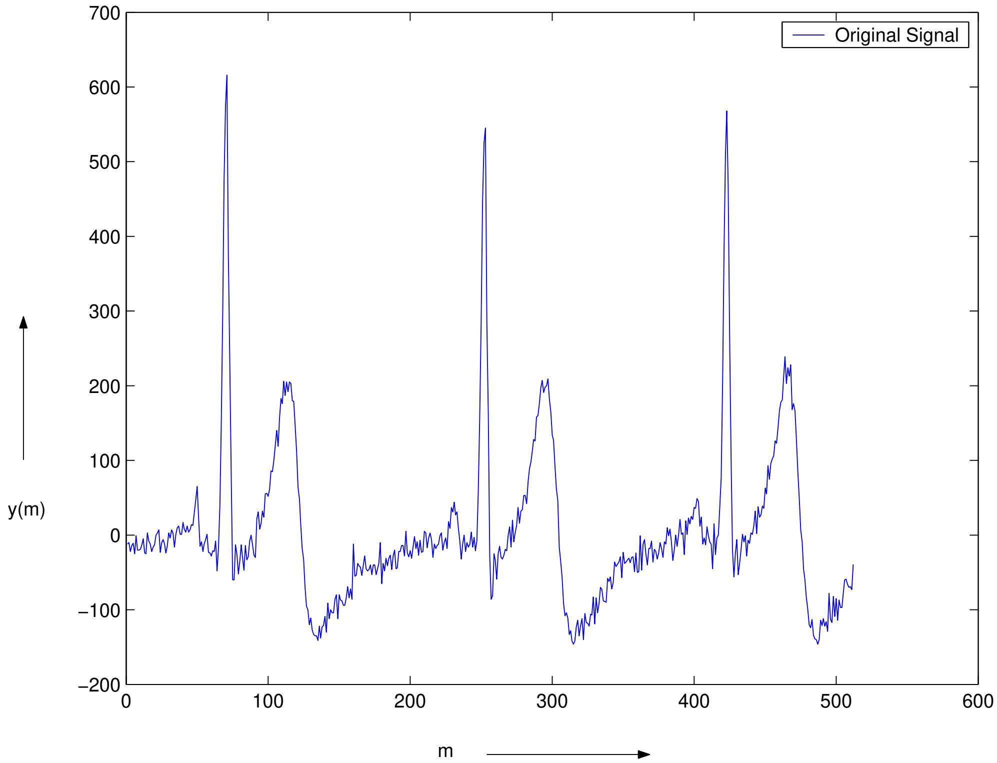

Figure 1: ECG Plot of a normal human being

The periodic or nearly periodic phenomena may be observed even in higher dimension also. See for example a black-white image plot of a symmetric image in Figure 3 and a colored symmetric image plot in Figure 3. Both of them indicate periodic nature in higher dimension.

The received signal is usually corrupted with noise. Signal Processing may broadly be considered as the recovery of information from noisy signal/data. Due to random nature of

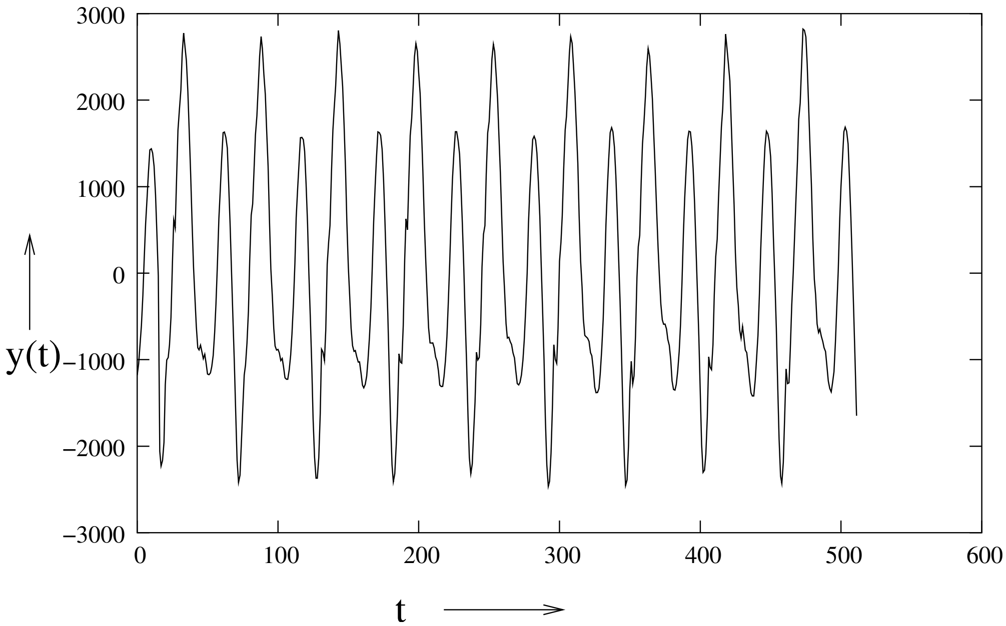

Figure 2: UUU-sound of a male.

the noise corrupted signal, statistics plays an important role in Signal Processing. Different statistical techniques may be used quite effectively to recover the original signal from the noisy signal. The main aim of this chapter is to provide different methods which are available till date to analyze periodic or nearly periodic signal in presence of additive noise. One natural question arises why somebody wants to analyze a periodic or nearly periodic signal. The first reason might be purely academic curiosity. But the main practical reason might be that a good periodic or nearly periodic model can be used quite effectively for compression or for prediction purposes.

A simple periodic function can be represented as follows:

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

Here $0 < \omega < 2\pi$. In this case the period of the function $y(t)$, i.e. the shortest time it needs to repeat itself is $2\pi/\omega$. Any mean zero smooth periodic function can be written in general

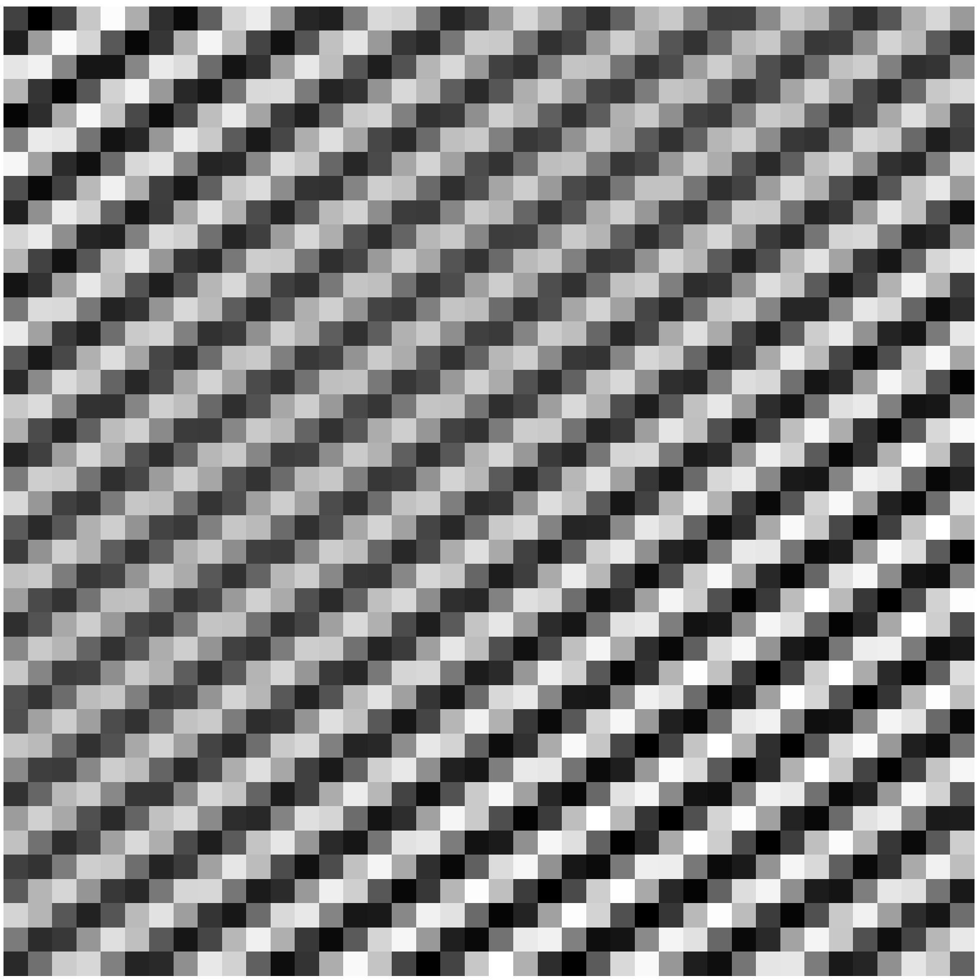

Figure 3: Black and white symmetric image plot

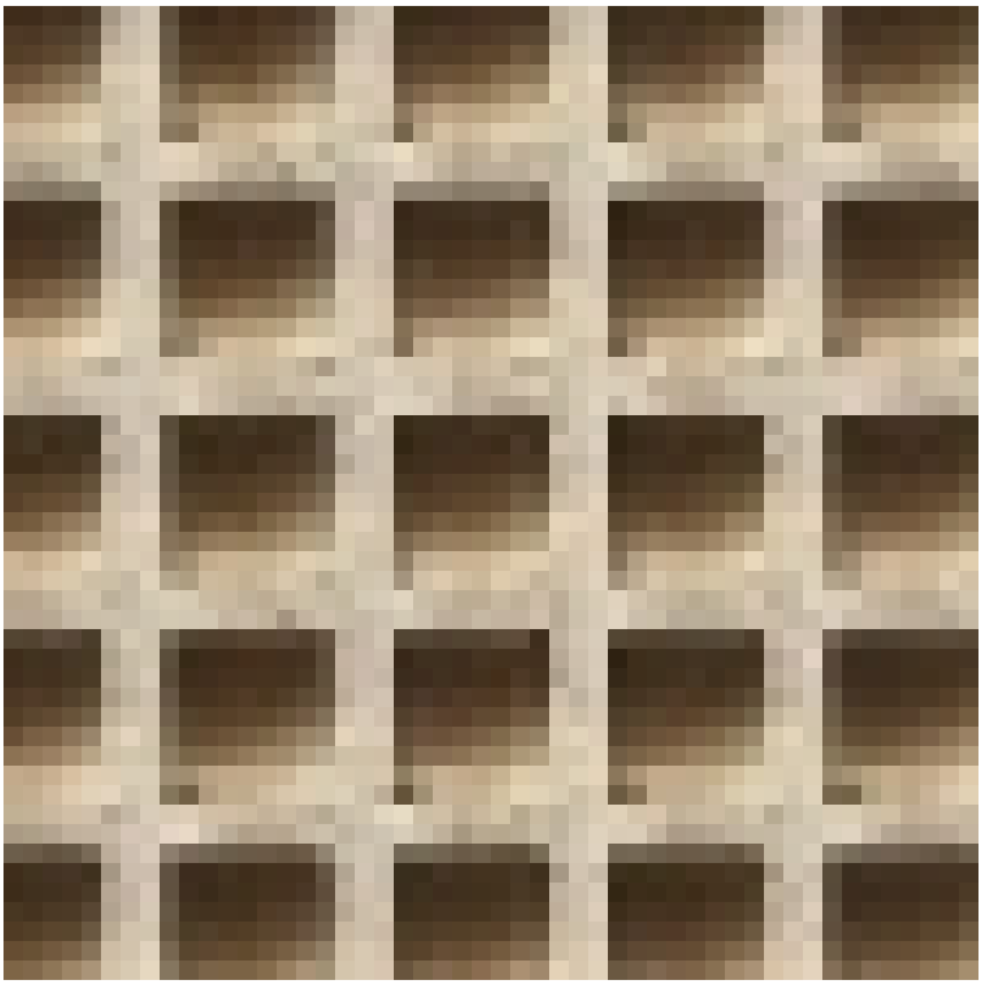

Figure 4: Colored symmetric image plot

as

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

and it mainly follows from the Fourier expansion of a periodic function. For any given smooth periodic function $y(t)$, the coefficients A_k's and B_k's are unique and they can be recovered from $y(t)$ as follows:

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

Since an observed signal is usually corrupted with noise, it is more reasonable to write it in a general form as

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

Here $e(t)$ is the noise component of the signal with mean zero. It is difficult to estimate infinite number of parameters, hence, (1) is approximated as

$$y(t) = \sum_{k=1}^{p} \{A_k \cos(\omega_k t) + B_k \sin(\omega_k t)\} + e(t). \tag{2}$$

Here $0 < \omega_1, \ldots, \omega_k < 2\pi$. The sum of sinusoidal model (2) plays an important role in the Signal Processing. The problem is to extract the deterministic component of the signal

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

based on the noisy signal at the discrete time points $\{y(1), \ldots, y(N)\}$. Hence, the problem becomes the estimation of 'p' as well as A_k, B_k, ω_k, for $k = 1, \ldots, p$. Here, 'p' is the order of sinusoidal component, $A_K^2 + B_k^2$ is known as amplitude and ω_k is the frequency, for $k = 1, \ldots, p$.

Sometime it is more convenient to work with the associated complex version of the model (2), and with the abuse of notation it can be written as

$$y(t) = \sum_{k=1}^{p} A_k e^{i\omega_k t} + e(t). \tag{3}$$

In model (3), $y(t)$ is complex valued signal, $e(t)$ is the complex valued error, A_k is a complex number, and $|A_k|^2$ is the amplitude and $i = \sqrt{-1}$. The model (3) can be obtained from model (2) by taking the Hilbert transformation. Both the models have been used quite extensively to analyze noisy periodic signal. The two models are equivalent, and results which have developed for model (2) can be modified suitably for model (3) and vice versa. An extensive amount of work has been done on models (2) and (3), a brief review will be presented in Section 3.

If the signal is nearly periodic several models have been used to analyze nearly periodic signal. Among them, one of the most used one is known as chirp model and it can be described as follows:

$$y(t) = A\cos(\alpha t + \beta t^2) + B\sin(\alpha t + \beta t^2) + e(t). \tag{4}$$

Here also A and B are real number, and $|A|^2 + |B|^2$ is the amplitude of the signal, α and β are known as frequency and frequency rate, respectively, $y(t)$ is the observed signal and $e(t)$ is a mean zero noise. This one component chirp model has been used quite extensively in different applications in signal processing. One most important application can be found in modeling radar signals. For example, suppose a radar is illuminating a target, then the transmitted signal will undergo a phase shift induced by the distance and relative speed between target and the receiver. Based on the assumption that the speed is continuous and differentiable, the phase shift can be approximated as $\phi(t) = a_0 + a_1 t + a_2 t^2$. Here the parameters a_1 and a_2 are either related to speed and acceleration or range and speed depending on what the radar is intended for, and on the kind of waveform transmitted. For explicit details see Rihaczek [108] (page 56 - 65). A more general form of the chirp signal can be written as follows:

$$y(t) = \sum_{k=1}^{p}\{A_k\cos(\alpha_k t + \beta_k t^2) + B_k\sin(\alpha_k t + \beta_k t^2)\} + e(t). \tag{5}$$

The model (5) can be found in several engineering applications also, see for example Wehner [121].

In case of chirp signals also, sometimes it is more convenient to work with the associated complex model. With the abuse of notation, it can be written as follows:

$$y(t) = \sum_{k=1}^{p} A_k e^{i(\alpha_k t + \beta_k t^2)} + e(t). \tag{6}$$

Similar to the complex exponential model, in this case also, for the model (6) $y(t)$'s are complex valued signals, A_k's are complex numbers, and $e(t)$'s are complex valued error. The models (5) and (6) are equivalent models. The methodologies and results which have been developed for model (6) can be used for model (5) and vice versa.

This problem is a challenging problem both from the theoretical and computational views points, since it is a highly non-linear model. Although the model (5) or (6) can be seen as a non-linear regression problem, but it does not satisfy some of the standard sufficient conditions needed to establish the basic properties like consistency and asymptotic normality of least squares estimators (LSEs) or maximum likelihood estimators (MLEs). Hence, it is not guaranteed whether the LSEs or the MLEs will be consistent or not. Moreover, even if it can be established that the LSEs or the MLEs will be consistent, finding them is not a trivial issue. Most of the standard algorithms do not work in this case. During the last fifteen years an extensive amount of work can be found both in the Signal Processing and in Statistics literature developing different efficient estimation procedures and establishing their properties. We provide a comprehensive review of the different methods which have been developed in the last fifteen to twenty years related to one dimensional one component chirp model in Section 4 and one dimensional multicomponent chirp model in Section 5.

Several related models have been proposed in the literature. A more general model than the one component chirp model was proposed by Djurić and Kay [24], it is known as the polynomial phase model and it can be written as follows:

$$y(t) = Ae^{i(\alpha_1 t + \alpha_2 t^2 + \ldots + \alpha_p t^p)} + e(t). \tag{7}$$

Here also $y(t)$, A^0 and $e(t)$ are complex valued. The corresponding real valued model can be written as

$$y(t) = A\cos(\alpha_1 t + \alpha_2 t^2 + \ldots + \alpha_p t^p) + B\sin(\alpha_1 t + \alpha_2 t^2 + \ldots + \alpha_p t^p) + e(t). \tag{8}$$

Here $y(t)$, A, B and $e(t)$ are all real valued, and (7) and (8) are equivalent models. Clearly, the model (8) is more flexible than the model (4), and it has also been used quite extensively in different signal processing applications. Model (7) or (8) is a highly non-linear model. In Section 6 we have provided various applications, theoretical developments and different estimations procedures related to this model.

Besson and Stoica [9] considered the random amplitude sinusoidal model and it can be defined as follows:

$$y(t) = \alpha(t)e^{i(a_0 + a_1 t)} + e(t). \tag{9}$$

Here $\alpha(t)$ is assumed to be real valued Gaussian stationary process and $e(t)$ is a complex valued noise. The random amplitude sinusoidal model (9) has several applications in different signal processing applications, for example when a radar on-board a train emits a continuous wave towards the track, the echo received is known to have a slowly varying amplitude and it can be modeled by using (9), see Besson and Castanié [8]. Similarly, in underwater systems, it has been observed that the acoustic signals propagating through the ocean are perturbed by multiplicative noise, and [30] observed that model (9) can be used for this purpose.

Along the same line, Besson, Ghogho and Swami [10] proposed random amplitude chirp signal model, and it can be represented as follows:

$$y(t) = \alpha(t)e^{i(a_0 + a_1 t + a_2 t^2)} + e(t). \tag{10}$$

Here $\alpha(t)$ and $e(t)$ are same as defined above. Clearly, the random amplitude chirp model (10) is more flexible than the one component chirp model. If $\alpha(t)$ is constant, then the random amplitude chirp model becomes the one component chirp model. Recently Nandi and Kundu [94] proposed a multicomponent random amplitude chirp model and it can be described as follows:

$$y(t) = \sum_{k=1}^{p} \alpha_k(t) e^{i(\theta_{1k} t + \theta_{2k} t^2)} + e(t). \tag{11}$$

Here $\{\alpha_k(t)\}$ is a sequence of independent and identically distributed real valued random variables and $e(t)$ is the complex valued noise. Clearly the multicomponent random amplitude chirp model (11) is more flexible than the multiple chirp model (6), and the model (6) can be obtained from the model (11), when $\alpha_k(t)$'s are assumed to be constant. In Section 7 we provide different applications, theoretical development and various estimation procedures related to one component and multicomponent random amplitude chirp models.

Recently, Grover [46] proposed a model which is a combination of the sinusoidal and the chirplet component and it can be expressed as follows:

$$y(t) = A\cos(\alpha t) + B\sin(\alpha t) + C\sin(\beta t^2) + D\cos(\beta t^2) + e(t). \tag{12}$$

Here, A, B, C and D are real numbers, $|A|^2 + |B|^2$ is the amplitude corresponds to the sinusoidal component, $|C|^2 + |D|^2$ is the amplitude corresponds to the chirplet component, $|A|^2 + |B|^2 + |C|^2 + |D|^2$ is the amplitude corresponds to the signal $y(t)$, α is the sinusoidal component, β is the chirp component and $e(t)$ is the real valued noise component. It is observed that model (12) behaves very similarly as the one component chirp model. In her thesis the author proposed a more general multicomponent chirp like model as follows:

$$y(t) = \sum_{k=1}^{p} \{A_k \cos(\alpha_k t) + B_k \sin(\alpha_k t)\} + \sum_{k=1}^{q} \left\{C_k \sin(\beta_k t^2) + D_k \cos(\beta_k t^2)\right\} + e(t). \tag{13}$$

The multicomponent chirp like model (13) behaves very similar to the multicomponent chirp model. Any signal which can be modelled using the multicomponent chirp model can be modelled using the multicomponent chirp like model also and vice versa. The main advantage about the chirp like model is that it is more convenient to use it in practice than the chirp model. In Section 8 we provide theoretical developments about one component and multicomponent chirp like models, different estimation procedures and the properties of these estimators have been discussed in details.

Hannan [49] first proposed the fundamental frequency sinusoidal model as follows:

$$y(t) = \sum_{k=1}^{p} \{A_k \cos(k\alpha t) + B_k \sin(k\alpha t)\} + e(t). \tag{14}$$

Model (14) has been used quite extensively for analyzing speech data. In short span periodic speech signal often it is observed that there is one fundamental frequency, and it has several harmonics. Due to the particular structure, it reduces the number of non-linear parameters of the model. Along the same line, Jensen et al. [59] proposed harmonic chirp signal, and it can be written as follows:

$$y(t) = \sum_{k=1}^{p} \{A_k \cos(k\alpha t + k\beta t^2) + B_k \sin(k\alpha t + k\beta t^2)\} + e(t). \tag{15}$$

Here, α and β are the fundamental frequency and fundamental frequency rate, respectively. Due to presence of only two non-linear parameters, numerically it is easier to handle than the multiple chirp model. Therefore, if a non-stationary signal can be modelled using the harmonic chirp model, it is always preferable. Jensen et al. [59] proposed estimation procedures of the unknown parameters of the model (15) and recently Grover [42] provided theoretical properties of different estimators. In Section 9 we describe different estimation procedure and their theoretical properties in details under various error assumptions.

Barbieri and Barone [6] proposed two dimensional sinusoidal model to model symmetric texture data. A p component sinusoidal frequency model can be written as follows:

$$y(m,n) = \sum_{k=1}^{p} \left[A_k^0 \cos(m\lambda_k^0 + n\mu_k^0) + B_k^0 \sin(m\lambda_k^0 + n\mu_k^0)\right] + e(m,n), \tag{16}$$

Here, A_k^0, B_k^0 are real numbers, λ_k^0, μ_k^0 are the frequencies, and $e(m,n)$ is the real valued error component. A typical texture image as shown in Figure 3 can be obtained using model (16). This is a basic model in many fields, such as antenna array processing, gephysical perception, biomedical spectral analysis etc. Since the introduction of the model, an extensive amount of work has been done developing different estimation procedures and establishing their properties for the two dimensional sinusoidal model. Along that line Friedlander and Francos [38] considered two dimensional polynomial phase signal and it has found several applications in image processing and finger print imaging. The two dimensional polynomial phase model can be expressed as follows:

$$y(m,n) = Ae^{i(\alpha_1 m + \beta_1 n + \alpha_2 m^2 + \beta_2 n^2 + \gamma mn)} + e(m,n). \tag{17}$$

Here, A is a complex number, $e(m,n)$ is the complex valued two dimensional noise and $y(m,n)$ is the two dimensional signal.

The most general two dimensional model has been considered by Lahiri and Kundu [75], and it can be expressed as follows:

$$y(m,n) = A\cos\left(\sum_{p=1}^{r}\sum_{j=0}^{p}\alpha(j,p-j)m^j n^{p-j}\right) + B\cos\left(\sum_{p=1}^{r}\sum_{j=0}^{p}\alpha(j,p-j)m^j n^{p-j}\right) + e(m,n). \tag{18}$$

Here as before A and B are real valued amplitudes, $\alpha(j,p-j)$'s are distinct frequency rates of order $(j,p-j)$, respectively. Different forms of (18) have been considered in the literature. A significant amount of work has been done in developing efficient estimation procedures of the unknown parameters of polynomial phase two dimensional signal. Lahiri and Kundu [75] discussed about the theoretical properties of the least squares estimators of the unknown parameters of the most general two dimensional polynomial phase model (18) under a fairly general error assumptions. In Section 10 we provide different theoretical developments and numerical procedures available related to this model. In all the sections we provide several open problems for future research, and finally we conclude the chapter in Section 11.

2 Notations and Preliminaries

In this Chapter, the scalar quantities are usually denoted by lower or uppercase letters. The lower and uppercase bold typefaces are used for vectors and matrices. For a real matrix $\boldsymbol{A}$, $\boldsymbol{A}^\top$ denotes the transpose. Similarly, for a complex matrix $\boldsymbol{A}$, $\boldsymbol{A}^H$ denotes the complex conjugate transpose. If c is a complex number $\bar{c}$ denotes the complex conjugate of c. A $m \times m$ diagonal matrix, with diagonal elements $a_1, \ldots, a_m$, are denoted by $\text{diag}\{a_1, \ldots, a_m\}$. If $\boldsymbol{A}$ is a real or a complex square matrix with full rank, the projection matrix on the column space of $\boldsymbol{A}$ is denoted by $\boldsymbol{P}_{\boldsymbol{A}} = \boldsymbol{A}(\boldsymbol{A}^\top\boldsymbol{A})^{-1}\boldsymbol{A}^\top$ or $\boldsymbol{P}_{\boldsymbol{A}} = \boldsymbol{A}(\boldsymbol{A}^H\boldsymbol{A})^{-1}\boldsymbol{A}^\top$. If $\{X_n; n \geq 1\}$ is a sequence of random variables, and X is also a random variable, then if X_n convergence to X almost surely, we will denote it by $X_n \xrightarrow{a.s.} X$. If X_n convergence to X in distribution, we will denote it by $X_n \xrightarrow{d} X$. Suppose X is a m-variate normal random vector with mean vector $\boldsymbol{\mu}$ and dispersion matrix $\boldsymbol{\Sigma}$, then it will be denoted by $N_m(\boldsymbol{0}, \boldsymbol{\Sigma})$. Throughout this chapter we have used the same notation in many places, for example the parameter vector $\boldsymbol{\theta}$ has been used in general to denote the set of unknown parameters of a model, similarly we have used the notation $\widehat{\boldsymbol{\theta}}$ in general to define different estimators of $\boldsymbol{\theta}$, the identity matrix of different orders are denoted by $\boldsymbol{I}$, the diagonal matrix of different orders and different diagonal entries are denoted by $\boldsymbol{D}$ etc., but they should be clear from the context.

2.1 Prony's Equation

In the Statistical Signal Processing, particularly developing different estimation procedures in case of sum of sinusoidal or sum of complex exponential model, Prony's equations have played an important role. Prony proposed this method in 1795, more than two hundred years back, to fit sum of real valued exponential model to real life data. Since then, different variants of Prony's equations have been developed. Now a days, it is available in most of the standard numerical analysis text books, see for example, Froberg [39] or Hilderband [51] in this respect. It can be briefly described as follows.

Suppose $\alpha_1, \ldots, \alpha_M$ are arbitrary non-zero real numbers and $\beta_1, \ldots, \beta_M$ are distinct real numbers. If

$$\mu(t) = \alpha_1 e^{\beta_1 t} + \ldots + \alpha_M e^{\beta_M t}; \quad t = 1, \ldots, N, \tag{19}$$

and $M < N$, then there exists $(M+1)$ constants $\{g_0, \ldots, g_M\}$, such that

$$\boldsymbol{A}\boldsymbol{g} = \boldsymbol{0}, \tag{20}$$

where

$$\boldsymbol{A} = \begin{bmatrix} \mu(1) & \ldots & \mu(M+1) \\ \vdots & \ddots & \vdots \\ \mu(N-M) & \ldots & \mu(N) \end{bmatrix}, \quad \boldsymbol{g} = \begin{bmatrix} g_0 \\ \vdots \\ g_M \end{bmatrix}, \quad \boldsymbol{0} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}.$$

The matrix $\boldsymbol{A}$ is of rank M, and $\boldsymbol{g}$ is the eigen vector corresponds to the zero eigen value. It is always possible to put restrictions on $g_0, \ldots, g_M$, such as

$$\sum_{i=0}^{M} g_j^2 = 1 \quad \text{and} \quad g_0 > 0, \tag{21}$$

such that $\boldsymbol{g}$ becomes unique. The set of linear equations (20) is known as Prony's equations. It is possible to recover $\beta_1, \ldots, \beta_M$ from $g_0, \ldots, g_M$ as follows. The following polynomial

$$p(x) = g_0 + g_1 x + \ldots + g_M x^M = 0 \tag{22}$$

has M distinct roots, and they are $e^{\beta_1}, \ldots, e^{\beta_M}$. Therefore, there is a one to one correspondence between $\{\beta_1, \ldots, \beta_M\}$ and $\{g_0, \ldots, g_M\}$, such that (21) hold. Further, $\{g_0, \ldots, g_M\}$ does not depend on $\{\alpha_1, \ldots, \alpha_M\}$. Once, $\beta_1, \ldots, \beta_M$ are obtained from $\mu(1), \ldots, \mu(N)$, $\alpha_1, \ldots, \alpha_M$ can be obtained from the linear equation

$$\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\alpha}, \tag{23}$$

where $\boldsymbol{\mu} = (\mu(1), \ldots, \mu(N))^\top$, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_M)^\top$ and

$$\boldsymbol{X} = \begin{bmatrix} e^{\beta_1} & \ldots & e^{\beta_M} \\ \ldots & \ddots & \ldots \\ e^{N\beta_1} & \ldots & e^{N\beta_M} \end{bmatrix}.$$

The matrix $\boldsymbol{X}$ is a of rank M, hence, $\boldsymbol{\alpha} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{\mu}$.

The idea of Prony has been extended for the complex case also. Consider the following complex exponential model

$$\mu(t) = \alpha_1 e^{i\omega_1 t} + \ldots + \alpha_M e^{i\omega_M t}; \qquad t = 1, \ldots, N. \tag{24}$$

Here, $i = \sqrt{-1}$, $\alpha_1, \ldots, \alpha_M$ are non-zero complex numbers, $0 < \omega_1, \ldots, \omega_M < 2\pi$ and they are distinct. Then, there exists $\{g_0, \ldots, g_M\}$ such that (20) satisfies. Moreover, the complex polynomial equation

$$p(z) = g_0 + g_1 z + \ldots + g_M z^M = 0,$$

has the roots $z_1 = e^{i\omega_1}, \ldots, z_M = e^{i\omega_M}$ and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_M)^\top = (\boldsymbol{X}^H \boldsymbol{X})^{-1} \boldsymbol{X}^H \boldsymbol{\mu}$. Note that the roots of the complex polynomial $p(z)$ satisfy

$$|z_1| = \ldots = |z_M| = 1, \quad \bar{z}_k = z_k^{-1}; \quad k = 1, \ldots, M. \tag{25}$$

Define the new complex polynomial

$$q(z) = z^{-M} \bar{p}(z) = \bar{g}_0 z^{-M} + \ldots + \bar{g}_M. \tag{26}$$

Due to (25), the polynomials $p(z)$ and $q(z)$ have same roots. Therefore,

$$\frac{g_k}{g_M} = \frac{\bar{g}_{M-k}}{\bar{g}_0}; \qquad k = 0, \ldots, M,$$

can be obtained by comparing the coefficients of $p(z)$ and $q(z)$. Hence,

$$b_k = \bar{b}_{M-k}; \qquad k = 0, \ldots, M, \tag{27}$$

where

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

The condition (27) is known as the conjugate symmetry property and in the matrix notation it can be written as

$$\boldsymbol{b} = \boldsymbol{J} \bar{\boldsymbol{b}},$$

here $\boldsymbol{b} = (b_0, \ldots, b_M)^\top$, and $\boldsymbol{J}$ is an exchange matrix as follows:

$$\boldsymbol{J} = \begin{bmatrix} 0 & 0 & \ldots & 0 & 1 \\ 0 & 0 & \ldots & 1 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 1 & \ldots & 0 & 0 \\ 1 & 0 & \ldots & 0 & 0 \end{bmatrix}.$$

2.2 Number Theoretic Results and Conjecture

We need some number theoretic results and a conjecture in establishing the asymptotic results related to chirp signal model. We present it here, and we will refer it whenever they are used. The following trigonometric results have been used quite extensively in establishing different results related to sinusoidal model. These results can be easily established, see for example Mangulis [86].

$$\lim_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N} \cos(\omega t) = o\left(\frac{1}{N}\right) \tag{28}$$

$$\lim_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N} \sin(\omega t) = o\left(\frac{1}{N}\right) \tag{29}$$

$$\lim_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N} \cos^2(\omega t) = \frac{1}{2} + o\left(\frac{1}{N}\right) \tag{30}$$

$$\lim_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N} \sin^2(\omega t) = \frac{1}{2} + o\left(\frac{1}{N}\right) \tag{31}$$

$$\lim_{N\to\infty} \frac{1}{N^{k+1}}\sum_{t=1}^{n} t^k \cos^2(\omega t) = \frac{1}{2} + o\left(\frac{1}{N}\right) \tag{32}$$

$$\lim_{N\to\infty} \frac{1}{N^{k+1}}\sum_{t=1}^{n} t^k \sin^2(\omega t) = \frac{1}{2} + o\left(\frac{1}{N}\right) \tag{33}$$

$$\lim_{N\to\infty} \frac{1}{N^{k+1}}\sum_{t=1}^{n} t^k \cos(\omega t)\sin(\omega t) = o\left(\frac{1}{N}\right), \tag{34}$$

here $a_N = o\left(\frac{1}{N}\right)$ means $Na_N \to 0$ as $N \to \infty$. We need similar results in case of chirp models also.

LEMMA 2.1 *Let $\theta_1, \theta_2 \in (0, \pi)$ and $k = 0, 1, \ldots$. Then except for countable number of points the followings are true:*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{n} \cos(\theta_1 t + \theta_2 t^2) = 0 \tag{35}$$

$$\lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{n} \sin(\theta_1 t + \theta_2 t^2) = 0 \tag{36}$$

$$\lim_{N\to\infty} \frac{1}{N^{k+1}} \sum_{t=1}^{n} t^k \cos^2(\theta_1 t + \theta_2 t^2) = \frac{1}{2(k+1)} \tag{37}$$

$$\lim_{N\to\infty} \frac{1}{N^{k+1}} \sum_{t=1}^{n} t^k \sin^2(\theta_1 t + \theta_2 t^2) = \frac{1}{2(k+1)} \tag{38}$$

$$\lim_{N\to\infty} \frac{1}{N^{k+1}} \sum_{t=1}^{n} t^k \cos(\theta_1 t + \theta_2 t^2) \sin(\theta_1 t + \theta_2 t^2) = 0. \tag{39}$$

PROOF: Using the number theoretic results of Vinogradov [115], the above results can be established, see Lahiri, Kundu and Mitra [76] for details. ∎

The following conjecture was proposed by Vinogradov [115], and it still remains an open problem. Although, extensive simulation results suggest the validity of the conjecture. This conjecture also has been used to establish some of the crucial results related to chirp model.

CONJECTURE 2.1 *Let $\theta_1, \theta_2, \theta_1', \theta_2' \in (0, \pi)$ and $k = 0, 1, \ldots$. Then except for countable number of points the followings are true:*

$$\lim_{N\to\infty} \frac{1}{\sqrt{N} N^k} \sum_{t=1}^{N} t^k \cos(\theta_1 t + \theta_2 t^2) \sin(\theta_1' t + \theta_2' t^2) = 0 \tag{40}$$

In addition if $\theta_2 \neq \theta_2'$, then

$$\lim_{N\to\infty} \frac{1}{\sqrt{N} N^k} \sum_{t=1}^{N} t^k \cos(\theta_1 t + \theta_2 t^2) cos(\theta_1' t + \theta_2' t^2) = 0$$

$$\frac{1}{\sqrt{N} N^k} \sum_{t=1}^{N} t^k \sin(\theta_1 t + \theta_2 t^2) sin(\theta_1' t + \theta_2' t^2) = 0.$$

3 Sinusoidal Model

In this section we briefly provide different estimation procedures of the frequencies of a periodic signal and their properties. It should be mentioned that there exists a vast amount of literature dealing with this problem, see for example the books by Kay [60] or Stoice and Mosses [112]. Therefore, it will not be possible in a limited space to provide all the methods till date. Therefore, we provide those procedures which may be useful in dealing with the chirp signal model also.

In this section we mainly consider the following model

$$y(t) = \sum_{k=1}^{p} \{A_k^0 \cos(\omega_k^0 t) + B_k^0 \sin(\omega_k^0 t)\} + e(t); \qquad 1, 2, \ldots, N, \tag{41}$$

where A_k^0, B_k^0, ω_k^0 for $k = 1, \ldots, p$ are true parameter values and they are unknown. In practice p is also usually unknown. Our problem is to estimate the unknown parameters based on the sample $\{y(1), \ldots, y(N)\}$. First it is assumed that p is known and later we discuss different estimation procedures of p also. Before progressing further, we need to make some assumptions on the error components $e(t)$. The error component $e(t)$ has mean zero and finite variance, and it can have either one of the two following forms:

Assumption 3.1 *$\{e(t)\}$ is a sequence of independent identically distributed (i.i.d.) normal random variables with mean zero and variance σ^2.*

Assumption 3.2 *$\{e(t)\}$ is a sequence of i.i.d. random variables with $E(e(t)) = 0$ and $V(e(t)) = \sigma^2$.*

Assumption 3.3 *$\{e(t)\}$ is a stationary linear process with the following form:*

$$e(t) = \sum_{j=0}^{\infty} a(j) X(t-j),$$

here $\{X(t); t = 1, 2, \ldots\}$ *are i.i.d. random variables with* $E(X(t)) = 0$ *and* $V(X(t) = \sigma^2$*, and* $\sum_{j=0}^{\infty} |a(j)| < \infty$.

3.1 Periodogram Estimators

The most popular estimator of the frequencies is the periodogram estimator. The periodogram function $I(\theta)$ at a particular frequency of the signal $y(t)$ is defined as follows:

$$I(\theta) = \frac{1}{N} \left| \sum_{t=1}^{N} y(t) e^{i\theta t} \right|^2 . \tag{42}$$

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, then $I(\theta)$ can be used to detect the number of signals and also the associated frequencies. Let us look at the following example of $y(t)$;

$$y(t) = 3.0 \cos(0.2\pi t) + 3.0 \sin(0.2\pi t) + 3.0 \cos(0.5\pi t) + 3.0 \sin(0.5\pi t) + e(t). \tag{43}$$

Here $e(t)$ satisfies Assumption 3.2 with $\sigma^2 = 2$. The periodogram plot of the signal (43) is provided in Figure 5. It is clear from Figure 5 that the number of sinusoidal components of the signal $y(t)$ is 2, and the location of the frequencies also can be estimated from the periodogram function. But it may not be true always. Now consider the following signal

$$y(t) = 3.0 \cos(0.2\pi t) + 3.0 \sin(0.2\pi t) + 0.25 \cos(0.5\pi t) + 0.25 \sin(0.5\pi t) + e(t). \tag{44}$$

Here also $e(t)$ satisfies Assumption 3.2, but $\sigma^2 = 5$. In this case, the periodogram plot of $y(t)$ is provided in Figure 6. From Figure 6 it is not clear that $p = 2$ and the true location of the frequencies. Therefore, it is clear that the periodogram estimators may not work always, particularly if the sample size is not large or the error variance is high. Hannan [48] and Walker [117] independently established the strong consistency and asymptotic normality

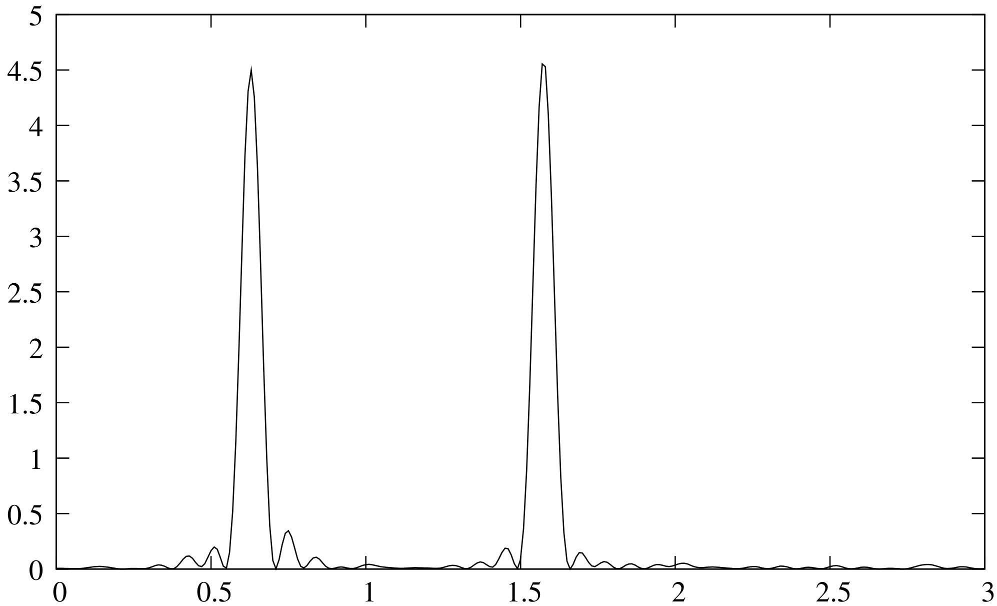

Figure 5: Periodogram plot of $y(t)$ obtained from Model (43)

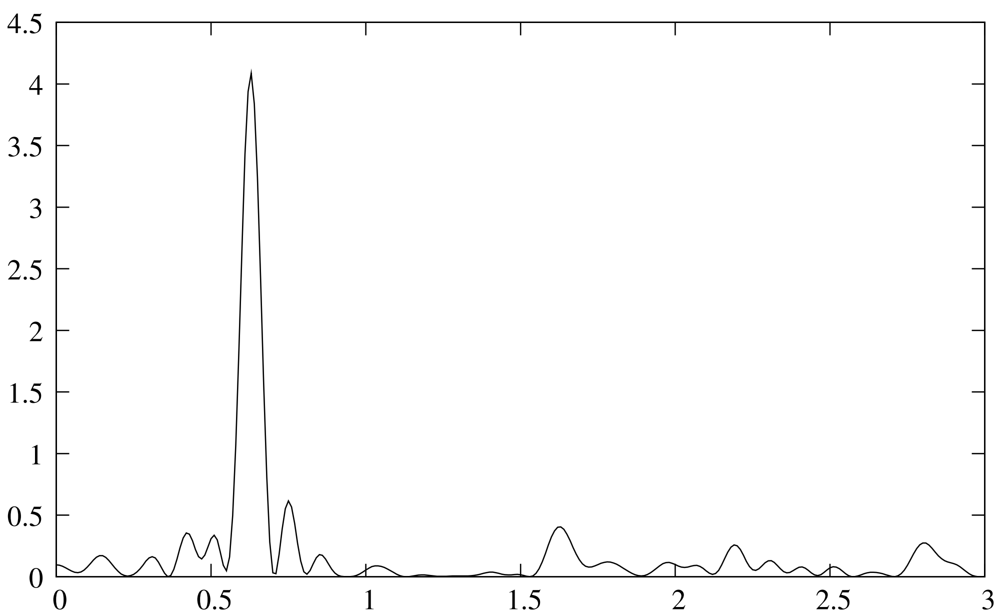

Figure 6: Periodogram plot of $y(t)$ obtained from Model (44)

results of the periodogram estimators of the frequencies based on Assumption 3.3. The theoretical results of Hannan [48] and Walker [117] justify the use of periodogram estimators at least for large sample sizes. Since the model (2) is a non-linear regression model, one natural estimator is the least squares estimators, and they coincide with the maximum likelihood (ML) estimators, if the error satisfies Assumption 3.1.

3.2 Least Squares Estimators

The least squares estimators (LSEs) or the maximum likelihood estimators (MLEs) under the assumption of i.i.d. Gaussian error of the error random variables, can be obtained by minimizing the error sum of squares as defined below

$$Q(\boldsymbol{\omega}, \boldsymbol{\theta}) = \sum_{t=1}^{N} \left(y(t) - \sum_{k=1}^{p} \{A_k \cos(\omega_k t) + B_k \sin(\omega_k t)\} \right)^2 . \tag{45}$$

Here $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_p)^\top$, $\boldsymbol{\theta} = (A_1, B_1, \ldots, A_p, B_p)^\top$. Now (45) can be written as

$$Q(\boldsymbol{\omega}, \boldsymbol{\theta}) = (\boldsymbol{Y} - \boldsymbol{Z}(\boldsymbol{\omega})\boldsymbol{\theta})^\top (\boldsymbol{Y} - \boldsymbol{Z}(\boldsymbol{\omega})\boldsymbol{\theta}), \tag{46}$$

where $\boldsymbol{Y} = (y(1), \ldots, y(N))^\top$, $\boldsymbol{e} = (e(1), \ldots, e(N))^\top$ and

$$\boldsymbol{Z}(\boldsymbol{\omega}) = \begin{bmatrix} \cos(\omega_1) & \sin(\omega_1) & \ldots & \cos(\omega_p) & \sin(\omega_p) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \cos(n\omega_1) & \sin(n\omega_1) & \ldots & \cos(n\omega_p) & \sin(n\omega_p) \end{bmatrix}.$$

For a given $\boldsymbol{\omega}$, the LSE of $\boldsymbol{\theta}$ can be obtained as

$$\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega}) = (\boldsymbol{Z}(\boldsymbol{\omega})^\top \boldsymbol{Z}(\boldsymbol{\omega}))^{-1} \boldsymbol{Z}(\boldsymbol{\omega})^\top \boldsymbol{Y}. \tag{47}$$

Substituting (47) in (46) we obtain

$$R(\boldsymbol{\omega}) = Q(\boldsymbol{\omega}, \widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})) = \boldsymbol{Y}^\top (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{Z}(\boldsymbol{\omega})}) \boldsymbol{Y}. \tag{48}$$

Here $\boldsymbol{P}_{\boldsymbol{Z}(\boldsymbol{\omega})}$ is the projection matrix on the column space of $\boldsymbol{Z}(\boldsymbol{\omega})$. Therefore, the LSE of $\boldsymbol{\omega}$ can be obtained by minimizing $R(\boldsymbol{\omega})$ with respect to $\boldsymbol{\omega}$. Bresler and Macovski [12] proposed a special purpose algorithm named as iterative quadratic maximum likelihood (IQML) to maximize $R(\boldsymbol{\omega})$. Using Prony's equation it can be observed that $\boldsymbol{I}-\boldsymbol{P}_{\boldsymbol{Z}(\boldsymbol{\omega})} = \boldsymbol{P}_{\boldsymbol{B}(\boldsymbol{g})}$, where $\boldsymbol{B}(\boldsymbol{g})$ is a $(N+1)\times(N+1-2p)$ matrix of rank $(N+1-2p)$ as follows:

$$\boldsymbol{B}(\boldsymbol{g}) = \begin{bmatrix} g_0 & 0 & \dots & 0 \\ g_1 & g_0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ g_{2p} & g_{2p-1} & \dots & 0 \\ 0 & g_{2p} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & g_0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & g_{2p} \end{bmatrix}, \tag{49}$$

and $\boldsymbol{g} = (g_0, g_1, \dots, g_{2p})^\top$. Note that minimization of $\boldsymbol{Y}^\top(\boldsymbol{I}-\boldsymbol{P}_{\boldsymbol{Z}(\boldsymbol{\omega})})\boldsymbol{Y}$ with respect to $\boldsymbol{\omega}$ is same as the minimization of $\boldsymbol{Y}^\top\boldsymbol{P}_{\boldsymbol{B}(\boldsymbol{g})}\boldsymbol{Y}$ with respect to $\boldsymbol{g}$. Now observe that

$$\boldsymbol{Y}^\top\boldsymbol{P}_{\boldsymbol{B}(\boldsymbol{g})}\boldsymbol{Y} = \boldsymbol{g}^\top\boldsymbol{Y}_D^\top(\boldsymbol{B}(\boldsymbol{g})^\top\boldsymbol{B}(\boldsymbol{g}))^{-1}\boldsymbol{Y}_D\boldsymbol{g},$$

where $\boldsymbol{Y}_D$ is a $(N-2p)\times(2p+1)$ matrix as given below:

$$\boldsymbol{Y}_D = \begin{bmatrix} y(1) & \dots & y(2p+1) \\ \vdots & \ddots & \vdots \\ y(n-2p) & \dots & y(N) \end{bmatrix}.$$

Bresler and Macovski [12] suggested the following. If at the $k-th$ stage of the iteration the estimate of $\boldsymbol{g}$ is $\boldsymbol{g}^{(k)}$, then obtain $\boldsymbol{g}^{(k+1)}$ by minimizing $\boldsymbol{g}^\top\boldsymbol{Y}_D^\top(\boldsymbol{B}(\boldsymbol{g}^{(k)})^\top\boldsymbol{B}(\boldsymbol{g}^{(k)}))^{-1}\boldsymbol{Y}_D\boldsymbol{g}$ with respect to $\boldsymbol{g}$. The iteration continues until convergence takes place. No proof of convergence has been provided by Bresler and Macovski [12]. Kundu [65] converted the minimization of $\boldsymbol{Y}^\top\boldsymbol{P}_{\boldsymbol{B}(\boldsymbol{g})}\boldsymbol{Y}$ as a non-linear eigenvalue problem, and provided an iterative procedure based on solving a standard non-linear eigen value problem. Further, he has provided the proof of convergence of the proposed iterative procedure.

Note that the multiple sinusoidal model (41) is a non-linear regression problem. An extensive amount of work has been done in the statistical literature developing different properties of the least squares estimators for different non-linear regression model, see for example the book by Seber and Wild [111] or Bates and Watts [7] in this respect. Jennrich [58] and Wu [122] developed different sufficient conditions which ensure the consistency and asymptotic normality of the least squares estimators of the parameters of a nonlinear regression model. Unfortunately, the multiple sinusoidal model (41) does not satisfy the sufficient conditions of Jennrich [58] or Wu [122]. Hence, it is not guaranteed that the LSEs will be consistent in this case. Kundu [66] proved the consistency and asymptotic normality of the LSEs of the unknown parameters of the multiple sinusoidal model. The following results have been obtained by Kundu [66]. Extensive simulations suggest that the iterative procedure proposed by Kundu [66] performs quite well in general.

THEOREM 3.1 *If the error $e(t)$ of model (41) satisfies Assumption 3.3, and for $k = 1, \ldots, p$, if we denote $(\widehat{A}_k, \widehat{B}_k, \widehat{\omega}_k)$, as the LSE of $(A_k^0, B_k^0, \omega_k^0)$, then*

$$(\widehat{A}_k, \widehat{B}_k, \widehat{\omega}_k)^\top \xrightarrow{a.s.} (A_k^0, B_k^0, \omega_k^0)^\top$$

and

$$(N^{1/2}(\widehat{A}_k - A_k^0), N^{1/2}(\widehat{B}_k - B_k^0), N^{3/2}(\widehat{\omega}_k - \omega_k^0)) \xrightarrow{d} N_3(\mathbf{0}, 2\sigma^2 c(\omega_k^0)\mathbf{\Sigma}_k).$$

Here $c(\omega_k^0) = \left| \sum_{j=0}^{\infty} a(j) e^{-ij\omega_k^0} \right|^2$ *and*

$$\mathbf{\Sigma}_k^{-1} = \frac{1}{A_k^{0^2} + B_k^{0^2}} \begin{bmatrix} A_k^{0^2} + 4B_k^{0^2} & -3A_k^0 B_k^0 & -3B_k^0 \\ -3A_k^0 B_k^0 & 4A_k^{0^2} + B_k^{0^2} & 3A_K^0 \\ -3B_k^0 & 3A_k^0 & 6 \end{bmatrix}.$$

Moreover, $(\widehat{A}_k, \widehat{B}_k, \widehat{\omega}_k)$ for $k = 1, \ldots, p$ are asymptotically independent.

Some of the points are quite interesting to observe in Theorem 3.1. It can be seen that although the multiple sinusoidal model does not satisfy some of the existing sufficient conditions for the LSEs to be consistent and asymptotically normal, still in this case the LSEs are consistent and they are asymptotically normally distributed. It should be mentioned that the asymptotic variances of the linear parameters are of the order $O(N^{-1})$, where as the asymptotic variances of the frequencies are of the order $O(N^{-3})$. Hence, the LSEs of the frequencies converge to the true frequency much faster than the LSEs of the amplitudes. Asymptotic confidence intervals of the amplitudes and frequencies can be easily constructed using Theorem 3.1. Note that although the LSEs have certain desirable properties, computing the LSEs may not be a trivial issue. It has been observed that the computation of the LSEs involves solving a p dimensional optimization problem and finding the roots of a p degree polynomial. Therefore, if p is small it is tractable, but if p is large it is a numerically challenging problem. In fact it has been observed in case of an ECG signal the number of components can be as high as 60. Therefore, it becomes very difficult to implement the above methods in such a case. Prasad, Kundu and Mitra [102] introduced a sequential estimation procedure which can be used quite effectively for large p and it has the same asymptotic properties as the LSEs.

3.3 Efficient Estimators

Nandi and Kundu [93] proposed an algorithm which is computationally very efficient. The most important feature of this algorithm is that it is computationally efficient but it produces estimators which have the asymptotic properties as the LSEs. It is an iterative algorithm, but it is known that it is going to converge in three steps when it starts with the initial estimators as the periodogram estimators. It produces estimators which are equivalent with the least squares estimators. More over the algorithm works even when the errors are from a stationary distribution. Interestingly, it does not use the whole sample at the first two stages, it uses only a fraction of the whole sample and only at the last step it uses the whole sample. It starts with $p = 1$, and then using the sequential procedure as described by Prasad, Kundu and Mitra [102] can be used quite effectively for any general p. Therefore, we describe the procedure only for $p = 1$.

If at the j-th stage the estimator of ω is denoted by $\omega^{(j)}$, then $\omega^{(j+1)}$ is calculated as

$$\omega^{(j+1)} = \omega^{(j)} + \frac{12}{N_j}\text{Im}\left[\frac{P(j)}{Q(j)}\right], \tag{50}$$

where

$$\begin{aligned} P(j) &= \sum_{t=1}^{N_j} y(t)\left(t - \frac{N_j}{2}\right) e^{-i\omega^{(j)}t}, \\ Q(j) &= \sum_{t=1}^{N_j} y(t) e^{-i\omega^{(j)}t}, \end{aligned}$$

and N_j denotes the sample size used at the j-th iteration. If $\omega^{(0)}$ denotes the periodogram estimator of ω as defined in Section 3.1, then the algorithm takes the following form:

ALGORITHM 3.1 *Step 1: Compute $\omega^{(1)}$ from $\omega^{(0)}$ using (50) with $N_1 = N^{0.8}$.*

Step 2: Compute $\omega^{(2)}$ from $\omega^{(1)}$ using (50) with $N_2 = N^{0.9}$.

Step 3: Compute $\omega^{(3)}$ from $\omega^{(2)}$ using (50) with $N_3 = N$.

It has been indicated by Nandi and Kundu [93] that those choices of N_1 and N_2 are not unique. There are several other choices for which the same asymptotic results hold. They have shown that the estimators obtained by the above iterative procedures are consistent and they have the same asymptotic distribution as the least squares estimators.

3.4 Super Efficient Estimators

Kundu et al. [68] proposed a modified Newton-Raphson method to obtain super efficient estimators of the frequencies of the model and it can produce estimators of the frequencies which have lower asymptotic variances than the least squares estimators. It is well known that the usual Newton-Raphson method often does not converge for this particular problem. But the authors have used a proper step factor modification which not only guarantees the convergence, but it also produces estimators which have smaller asymptotic variances than the LSEs, which is quite counterintuitive. The method works even when the errors are from a stationary distribution. It is assumed that $p = 1$, and one can easily use the sequential procedure to get the result for general p. The method can be described as follows. Let us denote

$$S(\omega) = \boldsymbol{Y}^\top \boldsymbol{Z}(\omega)(\boldsymbol{Z}^\top(\omega)\boldsymbol{Z}(\omega))^{-1}\boldsymbol{Z}^\top(\omega)\boldsymbol{Y}, \tag{51}$$

where $\boldsymbol{Y}$ and $\boldsymbol{Z}(\omega)$ are same as defined before. Therefore, the LSE of ω can be obtained by maximizing $S(\omega)$ with respect to ω. The maximization of $S(\omega)$ using Newton-Raphson method can be performed as follows:

$$\omega^{(j+1)} = \omega^{(j)} - \frac{S'(\omega^{(j)})}{S''(\omega^{(j)})}, \tag{52}$$

here $\omega^{(j)}$ is the estimate of ω at the j-th stage, $S'(\omega^{(j)})$ and $S''(\omega^{(j)})$ denote the first derivative and second derivative, respectively, of $S(\omega)$ evaluated at $\omega^{(j)}$. The standard Newton-Raphson algorithm is modified with a smaller correction factor as follows

$$\omega^{(j+1)} = \omega^{(j)} - \frac{1}{4} \times \frac{S'(\omega^{(j)})}{S''(\omega^{(j)})}.$$

Suppose $\omega^{(0)}$ denotes the periodogram estimator of ω, then the algorithm can be described as follows:

Algorithm 3.2 *(1) Take $N_1 = N^{6/7}$, and calculate*

$$\omega^{(1)} = \omega^{(0)} - \frac{1}{4} \times \frac{S'_{N_1}(\omega^{(0)})}{S''_{N_1}(\omega^{(0)})},$$

here $S'_{n_1}(\omega^{(0)})$ and $S''_{n_1}(\omega^{(0)})$ are same as $S'(\omega^{(0)})$ and $S''(\omega^{(0)})$, respectively, computed using a subsample of size N_1.

(2) With $N_j = N$, repeat

$$\omega^{(j+1)} = \omega^{(j)} - \frac{1}{4} \times \frac{S'_{N_1}(\omega^{(j)})}{S''_{N_1}(\omega^{(j)})}, \quad j = 1, 2, \ldots,$$

until a suitable stopping criterion is satisfied.

It is observed by the authors based on extensive simulations that any consecutive N_1 data points can be used at Step (1) to start the algorithm, and the choice of the initial estimators do not have any visible effect on the final estimators. Moreover, as it has been seen before in some of the other estimators also that the factor 6/7 in the exponent of Step (1) is not unique and there are several other ways Step (1) can be initiated. It is observed that the iteration converges very quickly and it produces frequency estimator which has lower variance than the LSE.

3.5 Estimation of the Number of Components

So far we have discussed about different estimation of the frequencies and amplitudes, assuming the number of components 'p' to be known. In this section we briefly discuss some of the well known methods of estimation of p. The estimation of 'p' can be considered as a model selection problem. Consider the class of models

$$\mathcal{M}_k = \left\{ \mu_k : \mu_k(t) = \sum_{j=1}^{k} \left[A^0_{j,k} \cos(\omega^0_{j,k} t) + B^0_{j,k} \sin(\omega^0_{j,k} t) \right] \right\}; \quad k = 1, 2, \ldots. \tag{53}$$

Based on the data $\{y(t); t = 1, \ldots, N\}$ one needs to choose $\widehat{p}$, an estimate of 'p', such that the model $\mathcal{M}_{\widehat{p}}$ fits the data 'best'. Therefore, any model selection procedure can be used to estimate 'p'.

One important point to observe here that the models are nested here, i.e.

$$\mathcal{M}_1 \subset \mathcal{M}_2 \subset \mathcal{M}_3 \subset \ldots.$$

Hence, one of the natural procedures is to use a test of significance for each additional term as it is introduced in the model. Fisher [33] first considerd this problem as a simple testing of hypothesis problem and it is based on the well known *likelihood ratio test*. This is based on the ratio of the maximimized likelihood for the k-th term to the maximized likelihood for $(k-1)$-th term of the model (53). If this quantity is large, it indicates that the k-th term is needed in the model, otherwise not.

The problem can be formulated as follows:

$$H_0 : p = p_0 \quad \text{vs.} \quad H_1 : p = p_1, \tag{54}$$

where $p_1 > p_0$. Based on the assumptions that the errors are i.i.d. normal random variables with mean zero and variance σ^2, it can be shown after some calculations, see Kundu and Nandi [73], that the likelihood ratio test takes the following form: rejects H_0, if L is large, where

$$L = \frac{\sum_{t=1}^{N}\left(y(t) - \sum_{k=1}^{p_0}\left\{\widehat{A}_{k,p_0}\cos(\widehat{\omega}_{k,p_0}t) + \widehat{B}_{k,p_0}\sin(\widehat{\omega}_{k,p_0}t)\right\}\right)^2}{\sum_{t=1}^{N}\left(y(t) - \sum_{k=1}^{p_0}\left\{\widehat{A}_{k,p_0}\cos(\widehat{\omega}_{k,p_0}t) + \widehat{B}_{k,p_0}\sin(\widehat{\omega}_{k,p_0}t)\right\}\right)^2}, \tag{55}$$

here $\widehat{A}_{k,p_0}$, $\widehat{B}_{k,p_0}$, $\widehat{\omega}_{k,p_0}$ are MLEs of A_{k,p_0}, B_{k,p_0} and ω_{k,p_0}, respectively. Similarly, $\widehat{A}_{k,p_1}$, $\widehat{B}_{k,p_1}$, $\widehat{\omega}_{k,p_1}$ are also defined. Now to apply this likelihood ratio test in practice, one needs to obtain the exact/ asymptotic distribution of L under the null hypothesis. It seems fining the exact/ asymptotic distribution of L is not an easy problem. Quinn and Hannan [104] provided an approximate distribution of L under certain restrictions on the form of the frequencies, which is quite complicated and may not have much practical importance.

Rao [106] proposed to use the cross validation to estimate the number of components in a sinusoidal model. It may be mentioned the cross validation method is a model selection technique and it can be used in a fairly general set up. The basic assumption of cross validation technique is that there exists an M, such that $1 \le k \le M$ for the models defined in (53). The cross validation method can be described as follows. For a given k, such that

$1 \leq k \leq M$, remove the j-th observation from $\{y(1), \ldots, y(N)\}$, and estimate '$y(j)$', say $\widehat{y}_k(j)$ based on the model assumption $\mathcal{M}_k$ and the observation $\{y(1), \ldots, y(j-1), y(j+1), \ldots, y(N)\}$. Obtain the cross validatory error for the model $\mathcal{M}_k$ as

$$\mathrm{CV}(k) = \sum_{t=1}^{N} \left(y(t) - \widehat{y}_k(t)\right)^2;$$

for $k = 1, \ldots, M$. Choose $\widehat{p}$ as the estimate of p, if

$$\widehat{p} = \arg\min\{\mathrm{CV}(1), \ldots, \mathrm{CV}(M)\}.$$

One major issue in implementing the cross validation method is to estimate $y(j)$ based on $\{y(1), \ldots, y(j-1), y(j+1), \ldots, y(N)\}$ based on the model assumption $\mathcal{M}_k$. Rao [106] did not provide any explicit method to compute $\widehat{y}_k(j)$. Later Kundu and Kundu [69] provided a non-iterative method to estimate $y(j)$ based on the missing observation. The method has been further modified by Kundu and Mitra [70] and this can be used quite effectively to estimate 'p' based on cross validation method. Extensive simulations indicate that for sample the method works quite well.

Rao [106] also proposed to use different information theoretic criteria to detect the number of components based on the assumptions that the errors are i.i.d. normal random variables with mean zero and finite variance. Based on the above error assumptions, the Akaike Information Criterion (AIC) takes the following form:

$$\mathrm{AIC}(k) = n \ln R_k + 2(7k),$$

here R_k denotes the minimum value of

$$Q(\boldsymbol{\theta}_k) = \sum_{t=1}^{N} \left(y(t) - \sum_{j=1}^{k} \left(A_j \cos(\omega_j t) + B_j \sin(\omega_j t)\right) \right)^2, \tag{56}$$

where $\boldsymbol{\theta}_k = (A_1, \ldots, A_k, B_1, \ldots, B_k, \omega_1, \ldots, \omega_k)^\top$ and the minimization of $Q(\boldsymbol{\theta}_k)$ as defined in(56) is performed with respect to $\boldsymbol{\theta}_k$. Here '$7k$' denotes the number of unknown parameters when the number of sinusoids is k. Choose $\widehat{p}$ an estimate of p, where

$$\widehat{p} = \arg\min\{\text{AIC}(1), \ldots, \text{AIC}(M)\}.$$

Under the same assumption, the Bayesian Information Criterion (BIC) takes the following form

$$\text{BIC}(k) = N \ln R_k + \frac{1}{2}(7k) \ln N,$$

here R_k is same as defined above. In this case also, choose $\widehat{p}$ an estimate of p, where

$$\widehat{p} = \arg\min\{\text{BIC}(1), \ldots, \text{BIC}(M)\}.$$

Although, Rao [106] suggested to use different Information Theoretic Criteria to estimate p, he did not provide any practical implementation procedure particularly, the computation of R_k. Kundu [64] suggested a practical implementation procedure of the method suggested by Rao [106] and performed extensive simulation experiments for different models, for different error variances, and for different sample sizes. It is observed that AIC does not provide consistent estimate of the model order, although small sample sizes, performances of AIC are quite good. In general it is observed that BIC performs quite well.

Kundu [67] suggested the following simple estimation procedure of 'p'. If 'M' denotes the maximum possible order, then for some fixed $L > 2M$, consider the data matrix $\boldsymbol{A}_L$ as follows:

$$\boldsymbol{A}_L = \begin{bmatrix} y(1) & \ldots & y(L) \\ \vdots & \ddots & \vdots \\ y(N-L+1) & \ldots & y(N) \end{bmatrix}.$$

Let $\widehat{\sigma}_1^2 > \ldots > \widehat{\sigma}_L^2$ be the L eigen values of the $L \times L$ matrix $\frac{1}{N}\boldsymbol{A}_L^\top \boldsymbol{A}_L$. Consider,

$$\text{KIC}(k) = \widehat{\sigma}_{2k+1}^2 + kc(N),$$

here, $c(N) > 0$, and satisfies the following two conditions

$$\lim_{N\to\infty} c(N) = 0 \quad \text{and} \quad \lim_{N\to\infty} \frac{c(N)\sqrt{N}}{(\ln\ln N)^{1/2}} = \infty.$$

Choose $\widehat{p}$ an estimate of p, where

$$\widehat{p} = \arg\min\{\text{KIC}(1), \ldots, \text{KIC}(M)\}.$$

Under the assumption of i.i.d. errors, Kundu [67] proved the strong consistency of the above procedure. The probability of wrong detection has also been obtained in terms of linear combination of chi-square variables. Extensive simulations have been performed to check the effectiveness of the proposed method and to find proper $c(N)$. It is observed that $C(N) = (\ln N)^{-1/2}$ performs quite well, although no theretical justification has been provided. There are some other methods which are available in the literature, see for example Sakai [110], Wang [118], Quinn [103] and the references cited therein. The details are avoided.

4 One Dimensional One Component Chirp Model

In this section first we consider one component chirp model as discussed before, and it can be written as follows.

$$y(t) = A^0 \cos(\alpha^0 t + \beta^0 t^2) + B^0 \sin(\alpha^0 t + \beta^0 t^2) + e(t); \quad t = 1, \ldots, N. \tag{57}$$

Here A^0, B^0 are unknown real numbers, $A^{0^2} + B^{0^2}$ is known as amplitude, α^0 is the frequency and β^0 is the frequency rate. Here the linear parameters $0 < |A_0|, |B_0| < \infty$, and $0 < \alpha^0, \beta^0 < \pi$. The error $e(t)$ is an additive error with mean zero and finite variance. At this moment let us assume that $e(t)$'s are i.i.d. random variables. The more general structure will be considered later. The problem is to estimate the unknown parameters $A^0, B^0, \alpha^0, \beta^0$ based on a sample of size N, namely $\{y(1), \ldots, y(N)\}$.

4.1 Least Squares Estimators

Chirp signals are encountered in many different areas of engineering applications, particularly in radar, active sonar and also in passive sonar systems. The problem of parameter estimation of the chirp parameters has received considerable amount of attention in the Signal Processing literature. The one component chirp model (57) is a non-linear regression model. Therefore, the least squares estimators (LSEs) or equivalently the maximum likelihood estimators when the errors are i.i.d. Gaussian random variables are the natural choice. The LSEs of the unknown parameters of the model (57) can be obtained by minimizing the residual sums of squares, as given below:

$$Q(\boldsymbol{\theta}) = \sum_{t=1}^{N} (y(t) - A\cos(\alpha t + \beta t^2) - B\sin(\alpha t + \beta t^2))^2. \tag{58}$$

Here $\boldsymbol{\theta} = (A, B, \alpha, \beta)^\top$, and $\widehat{\boldsymbol{\theta}} = (\widehat{A}, \widehat{B}, \widehat{\alpha}, \widehat{\beta})^\top$ minimizes $Q(\boldsymbol{\theta})$, and it is the LSE of $\boldsymbol{\theta}^0 = (A^0, B^0, \alpha^0, \beta^0)^\top$. Now to minimize $Q(\boldsymbol{\theta})$ let us rewrite (58) in a matrix notation.

$$Q(\boldsymbol{\theta}) = (\boldsymbol{Y} - \boldsymbol{W}(\alpha, \beta)\boldsymbol{\delta})^\top (\boldsymbol{Y} - \boldsymbol{W}(\alpha, \beta)\boldsymbol{\delta}), \tag{59}$$

here $\boldsymbol{Y} = (y(1), \ldots, y(N))^\top$, $\boldsymbol{\delta} = (A, B)^\top$ and the matrix $\boldsymbol{W}(\alpha, \beta)$ is

$$\boldsymbol{W}(\alpha, \beta) = \begin{bmatrix} \cos(\alpha + \beta) & \sin(\alpha + \beta) \\ \cos(2\alpha + 4\beta^2) & \sin(2\alpha + 4\beta^2) \\ \vdots & \vdots \\ \cos(N\alpha + N^2\beta) & \sin(N\alpha + N^2\beta) \end{bmatrix}. \tag{60}$$

Therefore, for fixed α and β,

$$\begin{bmatrix} \widehat{A}(\alpha, \beta) \\ \widehat{B}(\alpha, \beta) \end{bmatrix} = (\boldsymbol{W}^\top(\alpha, \beta)\boldsymbol{W}(\alpha, \beta))^{-1}\boldsymbol{W}^\top(\alpha, \beta)\boldsymbol{Y}$$

minimizes $Q(\boldsymbol{\theta})$. Hence, the LSEs of α and β, say $\widehat{\alpha}$ and $\widehat{\beta}$, respectively, can be obtained as

$$(\widehat{\alpha}, \widehat{\beta})^\top = \arg\min \; Q(\widehat{A}(\alpha, \beta), \widehat{B}(\alpha, \beta), \alpha, \beta). \tag{61}$$

Note that

$$\begin{aligned} Q(\widehat{A}(\alpha, \beta), \widehat{B}(\alpha, \beta), \alpha, \beta) &= \boldsymbol{Y}^\top(\boldsymbol{I} - \boldsymbol{W}(\alpha, \beta)(\boldsymbol{W}^\top(\alpha, \beta)\boldsymbol{W}(\alpha, \beta))^{-1}\boldsymbol{W}^\top(\alpha, \beta)\boldsymbol{Y} \\ &= \boldsymbol{Y}^\top(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{W}(\alpha,\beta)})\boldsymbol{Y}, \end{aligned}$$

here $\boldsymbol{I}$ is the diagonal matrix of order $N \times N$ and

$$\boldsymbol{P}_{\boldsymbol{W}(\alpha,\beta)} = \boldsymbol{W}(\alpha,\beta)(\boldsymbol{W}^\top(\alpha,\beta)\boldsymbol{W}(\alpha,\beta))^{-1}\boldsymbol{W}^\top(\alpha,\beta)$$

is the projection matrix on the column space of $\boldsymbol{W}^\top(\alpha,\beta)$. Therefore,

$$(\widehat{\alpha},\widehat{\beta})^\top = \arg\max \ \boldsymbol{Y}^\top \boldsymbol{P}_{\boldsymbol{W}(\alpha,\beta)}\boldsymbol{Y}.$$

Once $\widehat{\alpha}$ and $\widehat{\beta}$ are obtained, the LSEs of A and B can be obtained as

$$\widehat{A} = \widehat{A}(\widehat{\alpha},\widehat{\beta}) \quad \text{and} \quad \widehat{B} = \widehat{B}(\widehat{\alpha},\widehat{\beta}),$$

respectively. Therefore, it is observed that for one component chirp model the LSEs of the unknown parameters can be obtained by solving a two dimensional optimization problem. Some of the standard numerical technique like Newton-Raphson method or Gauss-Newton method may be used to compute the LSEs. It is observed that the least squares surface of α and β, namely $\boldsymbol{Y}^\top(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{W}(\alpha,\beta)})\boldsymbol{Y}$ is a highly non-linear surface.

Therefore, a very good set of initial values of α and β is needed for any iterative process to converge. Otherwise, the iterative process may not converge, or some times even if it converges it might converge to a local minimum rather than the global minimum.

Saha and Kay [109] proposed a Monte Carlo importance sampling procedure which does not need any initial guesses. It is a purely Monte Carlo based method, and uses the basic result of Pincus [100]. The implementation of the method is very simple and it can be described as follows. Based on the result of Pincus [100] it follows that

$$\widehat{\alpha} = \lim_{c\to\infty} \frac{\int_0^\pi \int_0^\pi \alpha \exp(c\boldsymbol{Z}(\alpha,\beta))d\beta d\alpha}{\int_0^\pi \int_0^\pi \exp(c\boldsymbol{Z}(\alpha,\beta))d\beta d\alpha} \quad \text{and} \quad \widehat{\beta} = \lim_{c\to\infty} \frac{\int_0^\pi \int_0^\pi \beta \exp(c\boldsymbol{Z}(\alpha,\beta))d\beta d\alpha}{\int_0^\pi \int_0^\pi \exp(c\boldsymbol{Z}(\alpha,\beta))d\beta d\alpha}, \tag{62}$$

where

$$\boldsymbol{Z}(\alpha,\beta) = \boldsymbol{Y}^\top \boldsymbol{P}_{\boldsymbol{W}(\alpha,\beta)}\boldsymbol{Y}.$$

Now to compute $\widehat{\alpha}$ and $\widehat{\beta}$ based on (62) Monte Carlo importance procedure can be used. The following algorithm can be used to compute $\widehat{\alpha}$ and $\widehat{\beta}$.

Algorithm 4.1 *Step 1: Generate a random sample of size M, say $\{\alpha_1, \ldots, \alpha_M\}$ from uniform $(0, \pi)$. Similarly, generate a random sample of size M, say $\{\beta_1, \ldots, \beta_M\}$ also from uniform $(0, \pi)$.*

Step 2: Choose a sequence $\{c_1, c_2, \ldots, \}$, such that $c_n \to \infty$, and $c_1 < c_2 < c_3 < \ldots$. For a fixed c_m, compute

$$\widehat{\alpha}_{c_m} = \frac{\frac{1}{M}\sum_{k=1}^{M} \alpha_k \exp(c_m \boldsymbol{Z}(\alpha_k, \beta_k))}{\frac{1}{M}\sum_{k=1}^{M} \exp(c_m \boldsymbol{Z}(\alpha_k, \beta_k))} \quad \text{and} \quad \widehat{\beta}_{c_m} = \frac{\frac{1}{M}\sum_{k=1}^{M} \beta_k \exp(c_m \boldsymbol{Z}(\alpha_k, \beta_k))}{\frac{1}{M}\sum_{k=1}^{M} \exp(c_m \boldsymbol{Z}(\alpha_k, \beta_k))}$$

Step 3: Stop the iteration if $\left|\widehat{\alpha}_{c_{m-1}} - \widehat{\alpha}_{c_m}\right| < \epsilon$ and $\left|\widehat{\beta}_{c_{m-1}} - \widehat{\beta}_{c_m}\right| < \epsilon$. Otherwise change c_m to c_{m+1} and go back to Step 2.

This may be mentioned that the same generated sample namely $\{\alpha_1, \ldots, \alpha_M\}$ and $\{\beta_1, \ldots, \beta_M\}$ can be used in each iteration. Therefore, the proposed method is very efficient. We will see later that the method can be easily generalized for multiple chirp model also.

Although, Saha and Kay [109] proposed a very efficient estimation procedure of the unknown parameters, they did not discuss any properties of the LSEs or the MLEs based on the assumption that the error are i.i.d. Gaussian random variables. The one component chirp model as described in (57) is a non-linear regression model. An extensive amount of work has been done in the statistical literature developing the properties of the LSEs of the unknown parameters of a non-linear regression model. Interested readers are referred to two books by Bates and Watts [7] and Seber and Wild [111] which deal with different aspects of non-linear regression models. Jennrich [58] and Wu [122] provided different sufficient conditions of non-linear regression models so that the LSEs become consistent and asymptotically normally distributed. Note that both the properties are quite important for any estimator. Unfortunately, the one component chirp model (57) does not satisfy the sufficient conditions of Jennrich [58] or Wu [122] as mentioned above. Hence, in case of chirp model, it is not immediate whether LSEs will be consistent or not. Nandi and Kundu [91] established the consistency and asymptotic normality properties of the LSEs of the unknown parameters of a one component chirp model (57) as given below.

THEOREM 4.1 *If there exists a* K*, such that* $0 < |A^0| + |B^0| < K$*,* $0 < \alpha^0, \beta^0 < \pi$*, and* $\sigma > 0$*, then* $\widehat{\boldsymbol{\theta}} = (\widehat{A}, \widehat{B}, \widehat{\alpha}, \widehat{\beta})^\top$ *is a strongly consistent estimate of* $\boldsymbol{\theta}^0 = (A^0, B^0, \alpha^0, \beta^0)^\top$.

THEOREM 4.2 *Under the same assumptions as in Theorem 4.1,*

$$\left(N^{1/2}(\widehat{A} - A^0), N^{1/2}(\widehat{B} - B^0), N^{3/2}(\widehat{\alpha} - \alpha^0), N^{5/2}(\widehat{\beta} - \beta^0)\right)^\top \xrightarrow{d} N_4(\mathbf{0}, 2\sigma^2\boldsymbol{\Sigma}), \tag{63}$$

here

$$\boldsymbol{\Sigma} = \frac{2}{A^{0^2} + B^{0^2}} \begin{bmatrix} \frac{1}{2}(A^{0^2} + 9B^{0^2}) & -4A^0B^0 & -18B^0 & 15B^0 \\ -4A^0B^0 & \frac{1}{2}(9A^{0^2} + B^{0^2}) & 18A^0 & -15A^0 \\ -18B^0 & 18A^0 & 96 & -90 \\ 15B^0 & -15A^0 & -90 & 90 \end{bmatrix}. \tag{64}$$

The proof of Theorem 4.1 can be found in Nandi and Kundu [91]. Theorem 4.1 establishes the consistency of the LSEs, whereas Theorem 4.2 states the asymptotic distribution along with the rate of convergence of the LSEs. When the errors are i.i.d. Gaussian random variables then the LSEs become the MLEs also. Therefore, the same results hold true.

It is interesting to observe that the rates of the convergence of the LSEs of the linear parameters are significantly different than the corresponding non-linear parameters as we have observed in case of sinusoidal model parameters also in Section 3. The non-linear parameters are estimated more efficiently than the linear parameters for a given sample size. The variance of the frequency estimator goes to zero with the rate N^{-3}, whereas the variance of the frequency rate estimator goes to zero with the rate N^{-5}. Moreover, as $A^{0^2} + B^{0^2}$ decreases, the asymptotic variances of the LSEs of the frequency and frequency rate increase. Therefore, it may be noted that the asymptotic distribution of the LSEs can be easily used to construct asymptotic confidence intervals of the unknown parameters, or developing testing of hypotheses also.

Nandi and Kundu [91] first time established formally the asymptotic properties of the LSEs of the parameters of the one component chirp model (57) based on the assumptions that the additive errors are i.i.d. with mean zero and finite variance. Later it is observed by the same authors, see Kundu and Nandi [72], that the results are true for a more general class of error distributions with stationary errors. The following assumptions on $e(t)$ are needed.

Assumption 4.1

$$e(t) = \sum_{j=-\infty}^{\infty} a(j)X(t-j), \tag{65}$$

where $\{X(t)\}$ is a sequence of i.i.d. random variables with mean zero and finite fourth moment, and

$$\sum_{j=-\infty}^{\infty} |a(j)| < \infty. \tag{66}$$

The above assumption is a standard assumption of a linear process. The stationary autoregressive process or the moving average process are special cases of the above linear process. Kundu and Nandi [72] established the following two results similar to the i.i.d. errors.

Theorem 4.3 *If there exists a K, such that $0 < |A^0| + |B^0| < K$, $0 < \alpha^0, \beta^0 < \pi$, $\sigma^2 > 0$ and $e(t)$ satisfies Assumption 4.1, then $\widehat{\boldsymbol{\theta}} = (\widehat{A}, \widehat{B}, \widehat{\alpha}, \widehat{\beta})^\top$ is a strongly consistent estimate of $\boldsymbol{\theta}^0 = (A^0, B^0, \alpha^0, \beta^0)^\top$.* ∎

Theorem 4.4 *Under the same assumptions as in Theorem 4.3,*

$$\left(N^{1/2}(\widehat{A} - A^0), N^{1/2}(\widehat{B} - B^0), N^{3/2}(\widehat{\alpha} - \alpha^0), N^{5/2}(\widehat{\beta} - \beta^0)\right)^\top \xrightarrow{d} \mathcal{N}_4(\mathbf{0}, 2c\sigma^2\boldsymbol{\Sigma}), \tag{67}$$

here $\boldsymbol{\Sigma}$ is same as defined in Theorem 4.2, and $c = \sum_{j=-\infty}^{\infty} a^2(j)$. ∎

4.2 Periodogram Estimator

Periodogram estimator as described in Section 3.1 is one of the most popular estimators for the sinusoidal model. Along the same line the periodogram estimators of the frequency and frequency rate of the one component chirp model (57) also can be defined as follows. First let us define the periodogram function associated with the chirp model as follows:

$$\begin{aligned} I(\alpha,\beta) &= \frac{1}{N}\left|\sum_{t=1}^{N} y(t)e^{i(\alpha t+\beta t^2)}\right|^2 \\ &= \frac{1}{N}\left\{\left(\sum_{t=1}^{N} y(t)\cos(\alpha t+\beta t^2)\right)^2 + \left(\sum_{t=1}^{N} y(t)\sin(\alpha t+\beta t^2)\right)^2\right\}. \end{aligned} \tag{68}$$

Grover, Kundu and Mitra [43] proposed periodogram estimators (PEs) or approximate least squares estimators (ALSEs) of α and β as

$$(\widehat{\alpha},\widehat{\beta}) = \arg\max\ \ I(\alpha,\beta). \tag{69}$$

In this case also the idea is same as the periodogram estimator of the frequency of a sinusoidal model. If the error variance is small then the maximum of the periodogram function $I(\alpha,\beta)$ occurs near the true values of the frequency and frequency rate, i.e. at (α^0,β^0). The PEs of α and β are obtained by solving a two dimensional optimization problem. Newton-Raphson or Gauss-Newton method may be used to compute the PEs. Periodogram surface is also highly non-linear in nature. It has several local maxima. Hence, a very good initial estimators are needed for any algorithm to converge to the global maximum rather than a local maximum. Among the points $\left\{\left(\frac{\pi i}{N},\frac{\pi j}{N^2}\right); i=1,\ldots,N-1, j=1,\ldots,(N-1)^2\right\}$, find (i_0,j_0), such that

$$I\left(\frac{\pi i_0}{N},\frac{\pi j_0}{N^2}\right) \geq I\left(\frac{\pi i}{N},\frac{\pi j}{N^2}\right),$$

for all i,j. This can be used as a good initial estimators of (α,β). Alternatively, Algorithm 4.1 also can be used to maximize $I(\alpha,\beta)$ simply by replacing $\boldsymbol{Z}(\alpha_k,\beta_k)$ with $I(\alpha_k,\beta_k)$ at each step.

Once $\widehat{\alpha}$ and $\widehat{\beta}$ are obtained the estimators of A and B can be obtained as

$$\widehat{A} = \frac{1}{N}\sum_{t=1}^{N} y(t)\cos(\widehat{\alpha}t + \widehat{\beta}t^2) \quad \text{and} \quad \widehat{B} = \frac{1}{N}\sum_{t=1}^{N} y(t)\sin(\widehat{\alpha}t + \widehat{\beta}t^2). \tag{70}$$

Grover, Kundu and Mitra [43] established the asymptotic properties of the PEs of the unknown parameters of one component chirp model. It has been shown that under the assumption of stationary errors as in Assumption 4.1 the PEs are strongly consistent and they have the same asymptotic distribution as the LSEs.

4.3 Finite Step Algorithm

So far we have discussed about the LSEs and PEs and although both of them have some computational issues, they are the most efficient estimators in the sense if the errors are i.i.d. Gaussian then the error variances achieve the Cramer-Rao lower bound. Lahiri, Kundu and Mitra [76] proposed an iterative procedure which guarantees to converge in four steps provided it starts with the initial guesses as it has been used in case of PEs in Section 4.2 and at the same time it produces estimators which have the same convergence rates as the LSEs or PEs. It is observed that if we start the initial guesses of α^0 and β^0 with convergence rates $O_p(N^{-1})$ and $O_p(N^{-2})$, respectively, then after four iterations, the algorithm produces an estimate of α^0 with convergence rate $O_p(N^{-3/2})$, and an estimate of β^0 with convergence rate $O_p(N^{-5/2})$. Before providing the algorithm in details first we show how to improve the estimators of α^0 and β^0. If $\widetilde{\alpha}$ is an estimator of α^0, such that for $\delta_1 > 0$, $\widetilde{\alpha} - \alpha^0 = O_p(N^{-1-\delta_1})$, and $\widetilde{\beta}$ is an estimator of β^0, such that for $\delta_2 > 0$, $\widetilde{\beta} - \beta^0 = O_p(N^{-2-\delta_2})$, then the improved estimators of α^0 and β^0, can be obtained as

$$\widetilde{\widetilde{\alpha}} = \widetilde{\alpha} + \frac{48}{N^2} Im\left(\frac{P_N^{\alpha}}{Q_N}\right) \tag{71}$$

$$\widetilde{\widetilde{\beta}} = \widetilde{\beta} + \frac{45}{N^4} Im\left(\frac{P_N^{\beta}}{Q_N}\right), \tag{72}$$

respectively, where

$$\begin{aligned}
P_N^{\alpha} &= \sum_{t=1}^{N} y(t)\left(t - \frac{N}{2}\right) e^{-i(\widetilde{\alpha}t + \widetilde{\beta}t^2)}, \\
P_N^{\beta} &= \sum_{t=1}^{N} y(t)\left(t^2 - \frac{N^2}{3}\right) e^{-i(\widetilde{\alpha}t + \widetilde{\beta}t^2)}, \\
Q_N^{\alpha} &= \sum_{t=1}^{N} y(t) e^{-i(\widetilde{\alpha}t + \widetilde{\beta}t^2)}.
\end{aligned}$$

The justification of the algorithm comes from the following two theorems. Please see Lahiri, Kundu and Mitra [76] for the detailed proof.

THEOREM 4.5 *If* $\widetilde{\alpha} - \alpha^0 = O_p(N^{-1-\delta_1})$ *for* $\delta_1 > 0$, *then*

$$\begin{array}{lll}
(a)\ (\widetilde{\widetilde{\alpha}} - \alpha^0) = O_p(N^{-1-2\delta_1}) & if & \delta_1 \le 1/4, \\
(b)\ N^{3/2}(\widetilde{\widetilde{\alpha}} - \alpha^0) \xrightarrow{d} \mathcal{N}(0, \sigma_1^2) & if & \delta_1 > 1/4,
\end{array}$$

where $\sigma_1^2 = \dfrac{384\sigma^2}{A^{0^2} + B^{0^2}}$. ∎

THEOREM 4.6 *If* $\widetilde{\beta} - \beta^0 = O_p(N^{-2-\delta_2})$ *for* $\delta_2 > 0$, *then*

$$\begin{array}{lll}
(a)\ (\widetilde{\widetilde{\beta}} - \beta^0) = O_p(N^{-2-2\delta_2}) & if & \delta_2 \le 1/4, \\
(b)\ N^{5/2}(\widetilde{\widetilde{\beta}} - \beta^0) \xrightarrow{d} \mathcal{N}(0, \sigma_2^2) & if & \delta_1 > 1/4,
\end{array}$$

where $\sigma_2^2 = \dfrac{360\sigma^2}{A^{0^2} + B^{0^2}}$. ∎

Now we show that starting from initial guesses $\widetilde{\alpha}$, $\widetilde{\beta}$ with convergence rates $\widetilde{\alpha} - \alpha^0 = O_p(N^{-1})$ and $\widetilde{\beta} - \beta^0 = O_p(N^{-2})$, respectively, how the above procedure can be used to obtain efficient estimators. It may be noted that finding initial guesses with the above convergence rates is not difficult. We can use the same initial guesses as it has been used in case of PEs in Section 4.2. The main idea is not to use the whole sample at the beginning, as it was originally suggested by Bai et al. [3]. A part of the sample is used at the beginning, and we gradually proceed towards the complete sample. With varying sample size, more and more data points are used with the increasing number of iteration. The algorithm can be described as follows. Denote the estimates of α^0 and β^0 obtained at the j-th iteration as $\widetilde{\alpha}^{(j)}$ and $\widetilde{\beta}^{(j)}$, respectively.

Algorithm 4.2 *Step 1: Choose* $N_1 = N^{8/9}$. *Therefore,* $\widetilde{\alpha}^{(0)} - \alpha^0 = O_p(N^{-1}) = O_p(N_1^{-1-1/8})$ *and* $\widetilde{\beta}^{(0)} - \beta^0 = O_p(N^{-2}) = O_p(N_1^{-2-1/4})$. *Perform steps (71) and (72). Therefore, after 1-st iteration, we have*

$$\widetilde{\alpha}^{(1)} - \alpha^0 = O_p(N_1^{-1-1/4}) = O_p(N^{-10/9}) \quad \textit{and} \quad \widetilde{\beta}^{(1)} - \beta^0 = O_p(N_1^{-2-1/2}) = O_p(N^{-20/9}).$$

Step 2: Choose $N_2 = N^{80/81}$. *Therefore,* $\widetilde{\alpha}^{(1)} - \alpha^0 = O_p(N_2^{-1-1/8})$ *and* $\widetilde{\beta}^{(1)} - \beta^0 = O_p(N_2^{-2-1/4})$. *Perform steps (71) and (72). Therefore, after 2-nd iteration, we have*

$$\widetilde{\alpha}^{(2)} - \alpha^0 = O_p(N_2^{-1-1/4}) = O_p(N^{-100/81}) \quad \textit{and} \quad \widetilde{\beta}^{(2)} - \beta^0 = O_p(N_2^{-2-1/2}) = O_p(N^{-200/81}).$$

Step 3: Choose $N_3 = N$. *Therefore,* $\widetilde{\alpha}^{(2)} - \alpha^0 = O_p(N_3^{-1-19/81})$ *and* $\widetilde{\beta}^{(2)} - \beta^0 = O_p(N_3^{-2-38/81})$. *Perform steps (71) and (72). Therefore, after 3-rd iteration, we have*

$$\widetilde{\alpha}^{(3)} - \alpha^0 = O_p(N^{-1-38/81}) \quad \textit{and} \quad \widetilde{\beta}^{(3)} - \beta^0 = O_p(N^{-2-76/81}).$$

Step 4: Choose $N_4 = N$ *and perform steps (71) and (72). Now we obtain the required convergence rates, i.e.*

$$\widetilde{\alpha}^{(4)} - \alpha^0 = O_p(N^{-3/2}) \quad \textit{and} \quad \widetilde{\beta}^{(4)} - \beta^0 = O_p(N^{-5/2}).$$

The above algorithm can be used quite efficiently to compute estimators in four steps which are equivalent to the LSEs. It may be mentioned that the fraction of the sample sizes which have been used in each step is not unique. It is possible to obtain equivalent estimators with different choices. Although they are asymptotically equivalent, the finite sample performances might be different. It might be interesting to compare the finite sample performances of the different estimators by extensive Monte Carlo simulations.

4.4 Least Absolute Deviation Estimators

Although LSEs and PEs are the most natural estimators in case of chirp model, it is well known that they are not robust, in the sense if there are outliers in the data or if the error variances are very high they may not work very well. Therefore, more robust estimators are usually preferred. Among different robust estimators the L_1 norm estimators or least absolute deviation (LAD) estimators are quite popular for general linear and non-linear regression problems, see for example Dielman [23] or Dodge [28]. There are two major issues which are associated with the LAD estimators, (i) computational one, i.e. how to compute the LAD estimators in an efficient manner, and (ii) theoretical issue, i.e. how to develop properties of the LAD estimators.

In case of linear regression model in the famous paper by Charnes, Cooper and Ferguson [17] it has been shown that the LAD estimators can be obtained very effectively by solving a linear programming problem. Although, it has not been explored, similar method may be tried even for non-linear regression model also. In this section we provide an alternative method to compute the LAD estimators. Regarding developing the theoretical properties of the LAD estimators, it is well known to be a difficult problem. Even in case of linear regression model it has been observed that one needs stronger assumptions on the error distribution than what is needed in case of LSEs. Some assumption on the PDF of the error distribution is usually needed to develop the properties of LAD estimators. Lahiri, Kundu and Mitra [78] developed the properties of LAD estimators under similar assumptions and details will be presented here.

Note that LAD estimators of the unknown parameters of the one component chirp model (57) can be obtained by minimizing

$$R(\boldsymbol{\theta}) = \sum_{n=1}^{N} \left| y(n) - A\cos(\alpha n + \beta n^2) - B\sin(\alpha n + \beta n^2) \right|. \quad (73)$$

Here $\boldsymbol{\theta} = (A, B, \alpha, \beta)^\top$ is same as defined before. Let us denote the LAD estimators of A^0, B^0, α^0 and β^0, by $\widehat{A}, \widehat{B}, \widehat{\alpha}$ and $\widehat{\beta}$, respectively. They can be obtained as

$$(\widehat{A}, \widehat{B}, \widehat{\alpha}, \widehat{\beta}) = \arg\min R(\boldsymbol{\theta}). \tag{74}$$

Clearly, the computation of the LAD estimators is a challenging issue. Any optimization algorithm which does not require derivative information may be used for this purpose. Alternatively, one can use the importance sampling technique as it has been suggested by Saha and Kay [109] in Section 4.1 with the obvious modifications. Developing efficient LAD estimators is an interesting challenging problem.

Efficient technique of computing the LAD estimators is not available till date, but Lahiri, Kundu and Mitra [78] established the asymptotic properties of the LAD estimators under some regularity conditions. The results are stated below. The proofs are available in Lahiri, Kundu and Mitra [78].

ASSUMPTION 4.2 *The error random variable $\{e(t)\}$ is a sequence of i.i.d. random variables with mean zero, variance σ^2, and it has a probability density function (PDF) $f(x)$. The PDF $f(x)$ is symmetric and differentiable in $(0, \epsilon)$ and $(-\epsilon, 0)$, for some $\epsilon > 0$, and $f(0) > 0$.*

THEOREM 4.7 *If there exists an K, such that $0 < |A^0| + |B^0| < K$, $0 < \alpha^0, \beta^0 < \pi$, $\sigma^2 > 0$, and $\{e(t)\}$ satisfies Assumption 4.2, then $\widehat{\boldsymbol{\theta}} = (\widehat{A}, \widehat{B}, \widehat{\alpha}, \widehat{\beta})^\top$ is a strongly consistent estimate of $\boldsymbol{\theta}^0$.* ∎

THEOREM 4.8 *Under the same assumptions as in Theorem 4.7,*

$$\Big(N^{1/2}(\widehat{A} - A^0), N^{1/2}(\widehat{B} - B^0), N^{3/2}(\widehat{\alpha} - \alpha^0), N^{5/2}(\widehat{\beta} - \beta^0)\Big)^\top \xrightarrow{d} \mathcal{N}_4(\mathbf{0}, \frac{1}{f^2(0)}\boldsymbol{\Sigma}), \tag{75}$$

here $\boldsymbol{\Sigma}$ is same as defined in (64). ∎

4.5 Iterative Approach

So far all the methods we have discussed are computationally quite demanding, but all of them provide efficient estimators in the sense they achieve the best convergence rates possible and it has been shown theoretically also. In all these cases the main computational time involves finding the initial guess which requires a $O(N^3)$ search. Therefore, if N is large this can take a significant amount of time. To avoid that several heuristics methods are available in the literature. Out of several heuristic methods, the iterative method suggested by Ikarm, Abed-Meraim and Hua [54], see also Ikarm, Abed-Meraim and Hua [53], is an important one and we will briefly discuss this method now.

To implement the iterative method Ikarm, Abed-Meraim and Hua [54] used the complex one parameter chirp model, i.e. without abuse of notation we can write the model as follows:

$$y(t) = Ae^{i(\alpha t + \beta t^2)} + e(t); \quad t = 1, \ldots, N. \tag{76}$$

Here A is the complex valued amplitude, and $e(t)$ is a complex valued random variable with mean zero and finite variance. The frequency α and the frequency rate β are same as defined before. The basic idea of the proposed iterative algorithm of Ikarm, Abed-Meraim and Hua [54] is the following. Suppose

$$\mu(t) = Ae^{i(\alpha t + \beta t^2)}; \quad t = 1, \ldots, N.$$

Then for any fixed integer τ, consider

$$m(t) = \mu^H(t)\mu(t+\tau) = Ce^{i(2\beta\tau t)}; \quad t = 1, \ldots, N - \tau. \tag{77}$$

Here $\mu^H(t)$ is the conjugate of $\mu(t)$ and $C = |A|^2 e^{i(\alpha + \beta\tau^2)}$. The equation (77) indicates that the sequence $m(t)$ represents a sinusoidal signal with frequency $2\beta\tau$. Therefore, the frequency β can be estimated from $m(t)$ using one of the frequency estimation technique.

Once β is estimated say by $\widehat{\beta}$, then the signal $\mu(t)$ can be demodulated as

$$z(t) = \mu(t)e^{-i(\widehat{\beta}t^2)} \approx Ae^{i(\alpha t)}.$$

Hence, α can be estimated from the demodulated signal $z(t)$ by using again the same frequency estimation technique. Note that in estimating α, it is assumed that the effect of β has been removed in the demodulated signal $z(t)$. Hence, Ikarm, Abed-Meraim and Hua [54] proposed the following iterative technique to estimate β. Once in the first phase an accurate estimate of β is obtained then in the second phase α is estimated by using the demodulation technique as described above. Let us use the subscript k as the iteration number and τ_k as the lag parameter at the k-th iteration.

The first step of the iterative algorithm depends on the following transformation

$$m_k(t) = z_{k-1}(n + \tau_{k-1})z_{k-1}^H(t); \quad t = 1, \ldots, N - \tau_{k-1}, \tag{78}$$

where

$$z_{k-1}(t) = y(t)e^{-i\widehat{\beta}_{k-1}t^2}; \qquad t = 1, \ldots, N - 1.$$

Here, $\widehat{\beta}_{k-1}$ is the estimate of β at the $(k-1)$-th iterate. Let us start the iteration as follows. For the first iteration $(k = 1)$ the initial values of $\widehat{\beta}_0 = 0$ and $\tau_0 = 1$ are chosen. Now from the transformed data

$$m_1(t) = y(t + 1)y^H(t); \qquad t = 1, \ldots, N - 1,$$

obtain $\widehat{\beta}_1$. At the kth $(k = 1, 2, \ldots)$ iteration, the demodulated signal $z_{k-1}(t)$ is considered to be a chirp signal with the same frequency as the original signal $y(t)$ but with different frequency rate $\Delta\beta_{k-1} = \beta - \widehat{\beta}_{k-1}$. The parameter $\Delta\beta_{k-1}$ is estimated using the transform $m_k(t)$, say as $\widehat{\Delta}\beta_{k-1}$. The estimate of β is than updated using

$$\widehat{\beta}_k = \widehat{\beta}_{k-1} + \widehat{\Delta}\beta_{k-1}.$$

It is very clear that the selection of the lag parameter τ_k affects the accuracy in estimating β. Different τ_k's can be chosen at different iteration steps. Ikarm, Abed-Meraim and Hua [54] conducted extensive simulation experiments and they have taken fixed lag size $\tau_k = 9$ and it is observed that the above method provides a very good result, although theoretically the properties of the estimators could not be established.

4.6 Bayes Estimates

Lin and Djurić [81] and recently, Mazumder [87] considered the Bayes estimators of the unknown parameters of the model (57) when $e(t)$'s are i.i.d. normal random variables. Mazumder [87] provided a more general solution than Lin and Djurić [81], hence, we are providing the methodology provided by Mazumder [87].

The following transformations and assumptions have been made for mathematical convenience.

$$A^0 = r^0 \cos(\theta^0), \quad B^0 = r^0 \sin(\theta^0), \quad r^0 \in (0, K], \quad \theta^0 \in [0, 2\pi], \quad \alpha, \beta \in (0, \pi).$$

The following prior assumptions distributions have been made on the above unknown parameters.

$$r \sim \text{uniform}(0, K) \quad (79)$$

$$\theta \sim \text{uniform}(0, 2\pi) \quad (80)$$

$$\alpha \sim \text{vonMises}(a_0, a_1) \quad (81)$$

$$\beta \sim \text{vonMises}(b_0, b_1) \quad (82)$$

$$\sigma^2 \sim \text{inverse gamma}(c_0, c_1). \quad (83)$$

It may be mentioned that r and θ have non-informative priors, σ^2 has a conjugate prior. In this case, 2α and 2β are circular random variables, that is why, von Misses distribution, the natural analog of the normal distribution in circular data has been considered. Let us denote the prior densities of r, θ, α, β, σ^2 as $[r]$, $[\theta]$, $[\alpha]$, $[\beta]$, $[\sigma^2]$, respectively, and $\boldsymbol{Y}$ is the data vector as defined in Section 2.1.1. If it is assumed that the priors are independently distributed, then the joint posterior density function of r, θ, α, β, σ^2 can be obtained as

$$[r, \theta, \alpha, \beta, \sigma^2 | \boldsymbol{Y}] \propto [r][\theta][\alpha][\beta][\sigma^2][\boldsymbol{Y} | r, \theta, \alpha, \beta, \sigma^2].$$

Now to compute the Bayes estimates of the unknown parameters using the Gibbs sampling technique one needs to compute conditional distribution of each parameter given all the parameters, known as the full conditional distribution, denoted by $[\cdot | \ldots]$, and they are given by

$$[r | \ldots] \propto [r][\boldsymbol{Y} | r, \theta, \alpha, \beta, \sigma^2]$$

$$[\theta | \ldots] \propto [\theta][\boldsymbol{Y} | r, \theta, \alpha, \beta, \sigma^2]$$

$$[\alpha | \ldots] \propto [\alpha][\boldsymbol{Y} | r, \theta, \alpha, \beta, \sigma^2]$$

$$[\beta | \ldots] \propto [\beta][\boldsymbol{Y} | r, \theta, \alpha, \beta, \sigma^2]$$

$$[\sigma^2 | \ldots] \propto [\sigma^2][\boldsymbol{Y} | r, \theta, \alpha, \beta, \sigma^2].$$

The closed form expression of the full conditionals cannot be obtained. Mazumder [87] proposed to use the random walk Markov Chain Monte Carlo (MCMC) technique to update these parameters, which can be easily implemented in practice. The author has used this method to predict future observation also.

4.7 Testing of Hypothesis

So far we have mainly talked about different estimation procedures of the unknown parameters for one component chirp model. But along with the estimation problem the associated testing of hypothesis problem is of significant practical importance. The testing of hypothesis can be very useful in identifying a source, mainly to detect whether a particular signal is coming from a known object or not. Recently, Dhar, Kundu and Das [22] considered the following testing of hypothesis problem for the one component chirp model. The authors considered the model (57) and the additive errors $e(t)$'s are assumed to be i.i.d. random variables with mean zero and finite variance σ^2. To develop the testing procedure some more technical assumptions on the density functions of the error random variables are needed, and it will be spelt out later.

If we denote the vector $\boldsymbol{\theta}^0 = (A^0, B^0, \alpha^0, \beta^0)^\top$, then we would like to test the following hypothesis

$$H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}^0 \qquad vs. \qquad H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}^0. \tag{84}$$

It can be interpreted this way that whether the observed signal is coming from a known source which emits chirp model with parameter $\boldsymbol{\theta}^0$.

Dhar, Kundu and Das [22] proposed four different tests to test the hypothesis (84) based on the following test statistics:

$$T_{N,1} = ||\mathbf{D}^{-1}(\widehat{\boldsymbol{\theta}}_{N,LSE} - \boldsymbol{\theta}^0)||_2^2, \tag{85}$$

$$T_{N,2} = ||\mathbf{D}^{-1}(\widehat{\boldsymbol{\theta}}_{N,LAD} - \boldsymbol{\theta}^0)||_2^2, \tag{86}$$

$$T_{N,3} = ||\mathbf{D}^{-1}(\widehat{\boldsymbol{\theta}}_{N,LSE} - \boldsymbol{\theta}^0)||_1^2, \tag{87}$$

$$T_{N,4} = ||\mathbf{D}^{-1}(\widehat{\boldsymbol{\theta}}_{N,LAD} - \boldsymbol{\theta}^0)||_1^2, \tag{88}$$

Here, $\widehat{\boldsymbol{\theta}}_{N,LSE}$ and $\widehat{\boldsymbol{\theta}}_{N,LAD}$ denote the LSE of $\boldsymbol{\theta}$ and LAD estimate of $\boldsymbol{\theta}$ as discussed in Section 4.1 and Section 4.4, respectively. Further, $||\cdot||_2$ and $||\cdot||_1$ denote the Euclidean and L_1 distances, respectively, and the 4×4 diagonal matrix $\mathbf{D}$ is as follows:

$$\mathbf{D} = diag\{N^{-1/2}, N^{-1/2}, N^{-3/2}, N^{-5/2}\}.$$

All these test statistics are based on some normalized values of the distances between the estimates and the true parameter value under the null hypothesis. The authors have chosen two specific distances, but in principle, any other distance can be considered. In all these cases, it is expected that the null hypothesis should be rejected if the values of the test statistics are large.

The natural question is how large is large? For this purpose we need to compute the critical value of the tests for a given significance level. The hypothesis will be rejected for the given significance level, if the value of the test statistic exceeds the critical value. For a given error distribution based on extensive simulations, it is possible to compute the critical value of a test, for a given significance level. To calculate it theoretically, the following error assumptions are made by the authors.

ASSUMPTION 4.3 *Let F_n be the distribution function of $y(n)$ with the probability density function $f_n(y, \boldsymbol{\theta})$, which is twice continuously differentiable with respect to $\boldsymbol{\theta}$. It is assumed that $E\left[\frac{\partial}{\partial\theta_i} f_n(y, \boldsymbol{\theta})\right]^{2+\delta}_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} < \infty$, for some $\delta > 0$ and $E\left[\frac{\partial^2}{\partial\theta_i \partial\theta_j} f_n(y, \boldsymbol{\theta})\right]^{2}_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} < \infty$, for all $n = 1, 2, \ldots, N$. Here θ_i and θ_j, for $1 \leq i, j \leq 4$, are the i-th and j-th component of $\boldsymbol{\theta}$.*

Note that the above Assumption is not very unnatural. The above Assumption holds for most of the well known probability density functions, e.g. normal, Lalace and Cauchy probability density functions. Based on the above assumption the critical values for all the four tests can be calculated, they are quite involved and they are not presented here. Interested readers are refereed to the original article of Dhar, Kundu and Das [22] for details. It can be shown that all the tests are consistent, i.e. as the sample size tends to infinity, the power of the tests go to one.

In this section we have mainly addressed different inferential issues associated with the one component chirp model. Now in the next section we will address the multicomponent chirp model.

5 One Dimensional Multicomponent Chirp Model

In this section we address different estimation problems associated with the multicomponent chirp model, namely the following model

$$y(t) = \sum_{k=1}^{p}\{A_k^0 \cos(\alpha_k^0 t + \beta_k^0 t^2) + B_k^0 \sin(\alpha_k^0 t + \beta_k^0 t^2)\} + e(t); \quad t = 1, \ldots, N. \tag{89}$$

Here, similar to the one component model, A_k^0, B_k^0 are real amplitudes, α_k^0 and β_k^0 are the frequency and frequency rate, respectively. The additive error $e(t)$ has mean zero and the other conditions will be explicitly stated whenever it is needed. At this moment, A_k^0, B_k^0, α_k^0 and β_k^0 are assumed to be unknown, and the number of components p is assumed to be known. Based on the sample $\{y(t); t = 1, \ldots, N\}$, the problem is to estimate the unknown parameters. With the abuse of notations, the associated complex model can be written as follows:

$$y(t) = \sum_{k=1}^{p} A_k^0 e^{(\alpha_k^0 t + \beta_k^0 t^2)} + e(t); \quad t = 1, \ldots, N. \tag{90}$$

Here, A_k^0 are complex valued amplitudes, $e(t)$ is a complex valued error with mean zero. The problem remains the same to estimate the unknown parameters based on the complex valued signal $\{y(t); t = 1, \ldots, N\}$. Most of the estimation methods which have been developed for the one component model, can be extended for the multicomponent model also. We mainly discuss the estimation procedures for the real multicomponent model only (89).

Before proceeding further, let us define the following notations: $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)^\top$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$, $\boldsymbol{A} = (A_1, \ldots, A_p)^\top$, $\boldsymbol{B} = (B_1, \ldots, B_p)^\top$, and $\boldsymbol{\Gamma}_j = (A_j, B_j, \alpha_j, \beta_j)^\top$. Similarly, $\boldsymbol{\alpha}^0$, $\boldsymbol{\beta}^0$, $\boldsymbol{A}^0$, $\boldsymbol{B}^0$ and $\boldsymbol{\Gamma}_j^0$ are also defined. Based on the assumption that additive error $e(t)$s are i.i.d. random variables with mean 0 and variance σ^2, the most natural estimator of the unknown parameters will be the least squares estimators. Now we will provide the estimation procedures of the LSEs and derive their theoretical properties.

5.1 Least Squares Estimators

The LSEs of the unknown parameters of the model (89) can be obtained by minimizing

$$Q(\mathbf{\Gamma}_1, \ldots, \mathbf{\Gamma}_p) = \sum_{t=1}^{N} \left(y(t) - \sum_{k=1}^{p} \{A_k \cos(\alpha_k t + \beta_k t^2) + B_k \sin(\alpha_k t + \beta_k t^2)\} \right)^2. \tag{91}$$

Note that if $e(t)$s are i.i.d. Gaussian random variables, then the LSEs become MLEs also. The minimization of (91) can be performed along the same line as the one component model. We can write $Q(\mathbf{\Gamma}_1, \ldots, \mathbf{\Gamma}_p)$ as follows It may be observed that $Q(\mathbf{\Gamma}_1, \ldots, \mathbf{\Gamma}_p)$ can be written as follows:

$$Q(\mathbf{\Gamma}_1, \ldots, \mathbf{\Gamma}_p) = \left[\boldsymbol{Y} - \sum_{j=1}^{p} \boldsymbol{W}(\alpha_j, \beta_j) \boldsymbol{\theta}_j \right]^\top \left[\boldsymbol{Y} - \sum_{j=1}^{p} \boldsymbol{W}(\alpha_j, \beta_j) \boldsymbol{\theta}_j \right], \tag{92}$$

where the $N \times 2$ matrix $\boldsymbol{W}(\alpha, \beta)$ is same as defined in (60) and $\boldsymbol{\theta}_j = (A_j, B_j)^\top$ is a 2×1 vector for $j = 1, \ldots, p$. The data vector $\boldsymbol{Y} = (y(1), \ldots, y(N))^\top$. The LSEs of the unknown parameters can be obtained by minimizing (92) with respect to the unknown parameters. Now define the $N \times 2p$ matrix $\widetilde{\boldsymbol{W}}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ as

$$\widetilde{\boldsymbol{W}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = [\boldsymbol{W}(\alpha_1, \beta_1) : \cdots : \boldsymbol{W}(\alpha_p, \beta_p)],$$

then, for fixed $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, the LSEs of $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_p$, the linear parameter vectors, can be obtained as

$$[\widehat{\boldsymbol{\theta}}_1^\top(\alpha_1, \beta_1) : \cdots : \widehat{\boldsymbol{\theta}}_p^\top(\alpha_p, \beta_p)]^\top = \left[\widetilde{\boldsymbol{W}}^\top(\boldsymbol{\alpha}, \boldsymbol{\beta}) \widetilde{\boldsymbol{W}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \right]^{-1} \widetilde{\boldsymbol{W}}^\top(\boldsymbol{\alpha}, \boldsymbol{\beta}) \boldsymbol{Y}.$$

Observe that because $\widetilde{\boldsymbol{W}}^\top(\boldsymbol{\alpha}, \boldsymbol{\beta}) \widetilde{\boldsymbol{W}}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is a diagonal matrix for large N due to Lemma 2.1, then $\widehat{\boldsymbol{\theta}}_j = \widehat{\boldsymbol{\theta}}_j(\alpha_j, \beta_j)$, the LSEs of $\boldsymbol{\theta}_j$, $j = 1, \ldots, p$ can also be expressed as

$$\widehat{\boldsymbol{\theta}}_j(\alpha_j, \beta_j) = \left[\boldsymbol{W}^\top(\alpha_j, \beta_j) \boldsymbol{W}(\alpha_j, \beta_j) \right]^{-1} \boldsymbol{W}^\top(\alpha_j, \beta_j) \boldsymbol{Y}.$$

Using similar techniques as in one component case, the LSEs of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ can be obtained as the argument maximum of

$$\boldsymbol{Y}^\top \widetilde{\boldsymbol{W}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) [\widetilde{\boldsymbol{W}}^\top(\boldsymbol{\alpha}, \boldsymbol{\beta}) \widetilde{\boldsymbol{W}}(\boldsymbol{\alpha}, \boldsymbol{\beta})]^{-1} \widetilde{\boldsymbol{W}}^\top(\boldsymbol{\alpha}, \boldsymbol{\beta}) \boldsymbol{Y}. \tag{93}$$

The criterion function, given in (93), is a highly non-linear function of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, therefore, the LSEs of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ cannot be obtained in closed form. One needs to solve a $2p$ dimensional optimization problem to compute the LSEs of the unknown parameters. Since, the function defined in (93) is a highly non-linear function, very good initial estimates are needed to find the global maximum. The problem becomes more complicated for larger p. Saha and Kay [109] suggested to use the method of Pincus [100] to maximize (93). Before describing other efficient methods, we provide the properties of the LSEs.

Similar to the one component model, Kundu and Nandi [72] obtained the following consistency result for the LSEs of the multicomponent model.

THEOREM 5.1 *Suppose there exists a K, such that for $j = 1, \ldots, p$, $0 < |A_j^0| + |B_j^0| < K$, $0 < \alpha_j^0, \beta_j^0 < \pi$, α_j^0 are distinct, similarly β_j^0 are also distinct. If $e(t)$s are i.i.d. random variables with mean zero and finite variance σ^2, then $\widehat{\boldsymbol{\Gamma}}_j = (\widehat{A}_j, \widehat{B}_j, \widehat{\alpha}_j, \widehat{\beta}_j)^\top$, the LSE of $\boldsymbol{\Gamma}_j^0$ $= (A_j^0, B_j^0, \alpha_j^0, \beta_j^0)^\top$ is strongly consistent, for $j = 1, \ldots, p$.* ∎

Kundu and Nandi [72] also obtained the asymptotic normality properties of $\widehat{\boldsymbol{\Gamma}}_j$ along with the consistency results, but the elements of the asymptotic variance covariance matrix are quite messy. Later Lahiri, Kundu and Mitra [79] using the result Lemma 2.1 and the Conjecture 2.1 simplified the entries of the asymptotic variance covariance matrix and it can be presented as follows.

THEOREM 5.2 *Under the same assumptions as in Theorem 5.1, for $j = 1, \ldots, p$,*

$$\Big(N^{1/2}(\widehat{A}_j - A_j^0), N^{1/2}(\widehat{B}_j - B^0), N^{3/2}(\widehat{\alpha}_j - \alpha_j^0), N^{5/2}(\widehat{\beta}_j - \beta_j^0)\Big)^\top \xrightarrow{d} N_4(\mathbf{0}, 2\sigma^2\boldsymbol{\Sigma}_j), \quad (94)$$

where $\boldsymbol{\Sigma}_j$ can be obtained from the matrix $\boldsymbol{\Sigma}$ defined in (64), by replacing A^0 and B^0 with A_j^0 and B_j^0, respectively. Moreover, $\widehat{\boldsymbol{\Gamma}}_j$ and $\widehat{\boldsymbol{\Gamma}}_k$, for $j \neq k$ are asymptotically independently distributed. ∎

The consistency and asymptotic normality of the LSEs hold even when the errors are from a stationary linear process.

5.2 Sequential Least Squares Estimators

Even though, the LSEs are the most natural estimators for multicomponent chirp model, finding the LSEs is a numerically challenging problem, particularly if p is large. Due to this reason, Lahiri, Kundu and Mitra [79] proposed a sequential procedure for the multiple chirp model similar to the sequential procedure as suggested by Prasad, Kundu and Mitra [102] in case of multiple sinusoidal model. Based on the number theoretic Lemma 2.1 and the Conjecture 2.1, it has been shown that the LSEs and the sequential estimators are asymptotically equivalent, i.e. the sequential estimators are also strongly consistent and they have the same asymptotic distributions as the LSEs. The sequential estimators as suggested by Lahiri, Kundu and Mitra [79] can be described as follows. Let us assume the following, without loss of generality:

$$A_1^{0^2} + B_1^{0^2} > \ldots > A_p^{0^2} + B_p^{0^2}.$$

Then first obtain the estimates of A_1^0, B_1^0, α_1^0 and β_1^0 by minimizing

$$Q(\mathbf{\Gamma}_1) = \left[\boldsymbol{Y} - \boldsymbol{W}(\alpha_1, \beta_1)\boldsymbol{\theta}_1\right]^\top \left[\boldsymbol{Y} - \boldsymbol{W}(\alpha_1, \beta_1)\boldsymbol{\theta}_1\right], \tag{95}$$

Note that the minimization of (95) can be performed by solving a two dimensional optimization problem. Suppose $\widehat{\alpha}_1$, $\widehat{\beta}_1$, $\widehat{\boldsymbol{\theta}}_1 = (\widehat{A}_1, \widehat{B}_1)^\top$, are the estimates of α_1^0, β_1^0 and $(A_1^0, B_1^0)^\top$, respectively, then obtain the new data set $\boldsymbol{Y}^{(1)}$ from the original data set $\boldsymbol{Y}$, by removing the effect of the first component as follows:

$$\boldsymbol{Y}^{(1)} = \boldsymbol{Y} - \boldsymbol{W}(\widehat{\alpha}_1, \widehat{\beta}_1)\widehat{\boldsymbol{\theta}}_1 \tag{96}$$

Then at the second step obtain the estimates of A_2^0, B_2^0, α_2^0 and β_2^0 by minimizing

$$Q(\mathbf{\Gamma}_2) = \left[\boldsymbol{Y}^{(1)} - \boldsymbol{W}(\alpha_2, \beta_2)\boldsymbol{\theta}_2\right]^\top \left[\boldsymbol{Y}^{(1)} - \boldsymbol{W}(\alpha_2, \beta_2)\boldsymbol{\theta}_2\right]. \tag{97}$$

In this case also the minimization of (97) can be performed by solving a two dimensional optimization problem. Repeating this process we can obtain the estimates of $(A_3^0, B_3^0, \alpha_3^0, \beta_3^0)$, ..., $(A_p^0, B_p^0, \alpha_p^0, \beta_p^0)$, sequentially. We call these estimators as the sequential least squares estimators. It has been shown by Lahiri, Kundu and Mitra [79] that under the assumption of stationary errors, the sequential estimators are strongly consistent and they have the same asymptotic distribution as the LSEs. It has also been shown by Lahiri, Kundu and Mitra [79] that if the process is repeated beyond p times, then the corresponding amplitude estimates converge to zero almost surely. Hence, in practice the sequential procedure can be used to estimate the number of components in a chirp signal also.

Note that the sequential procedure is a very powerful tool. It mainly works due to the fact that as the sample size increases the two chirp components become orthogonal because of the number theoretic Lemma 2.1 and the Conjecture 2.1. Hence, this sequential procedure works for other estimators also. For example Grover, Kundu and Mitra [44] provided the sequential periodogram estimators which are strongly consistent and which have the same asymptotic distribution as the LSEs. It can be easily shown that the sequential method works for the finite step algorithm as described in Section 4.3. It will be interesting to develop the theoretical properties of the sequential LAD estimators and also sequential testing of hypothesis for multicomponent chirp model.

6 Polynomial Phase Model

6.1 One Component Polynomial Phase Model

So far we have discussed about one component and multicomponent chirp model. But a more general model than a one component chirp model is the one component polynomial phase model. The one component polynomial phase signal was first introduced by Djurić and Kay [24] and it can be described as follows:

$$y(t) = A^0 e^{i(\alpha_1^0 t + \alpha_2^0 t^2 + \ldots + \alpha_p^0 t^p)} + e(t); \qquad t = 1, \ldots, N. \tag{98}$$

Here also A^0 is a complex number, and $0 < \alpha_1^0, \ldots, \alpha_p^0 < \pi$ are the coefficients of the polynomial and $e(t)$s are noise random variables with mean zero and finite variance. The explicit assumptions on $e(t)$s will be mentioned later. Here also the problem remains the same, based on the observations $y(1), \ldots, y(N)$, estimate the unknown parameters namely A^0, $\alpha_1^0, \ldots, \alpha_p^0$. Throughout it is assumed that the degree of the polynomial p is known.

The associated real valued model can be written as follows:

$$y(t) = A^0 \cos(\alpha_1^0 t + \alpha_2^0 t^2 + \ldots + \alpha_p^0 t^p) + B^0 \cos(\alpha_1^0 t + \alpha_2^0 t^2 + \ldots + \alpha_p^0 t^p) + e(t); \qquad t = 1, \ldots, N. \tag{99}$$

In case of model (99) A^0, B^0 are real valued and $e(t)$s are real valued random variables with mean zero and finite variance. Clearly, the polynomial phase model is more general than the chirp model.

One of the primary motivations for studying polynomial phase signals comes from Doppler radar applications. Although the continuous time transmitted radar signal does not have polynomial phase, the samples taken at the matched filter output of a pulsed radar system give rise to a discrete time polynomial phase signal, when the target is moving, see for example Rihaczek [108]. The polynomial coefficients are then related to the kinetic parameters of the target. The polynomial phase models are often used in analyzing synthetic aperture radar (SAR) images, see for example Porchia et al. [101].

Djurić and Kay [24] considered the MLEs of the unknown parameters based on the assumptions that the error random variables are complex valued Gaussian random variables. Hence, for the model (98) the MLEs of the unknown parameters can be obtained by minimizing

$$Q(\boldsymbol{\theta}_p) = \sum_{t=1}^{N} \left| y(t) - Ae^{i(\alpha_1 t + \ldots, \alpha_p t^p)} \right|^2, \tag{100}$$

with respect to $\boldsymbol{\theta}_p$, where $\boldsymbol{\theta}_p = (A_R, A_I, \alpha_1, \ldots, \alpha_p)^\top$ and $A = A_R + iA_I$. The minimization of (100) can be performed by solving a p dimensional optimization problem. Any standard non-linear optimization method may be used, but one needs a very good initial guesses to attend the global optimum solution. Alternatively, the importance sampling as described in Section 4.1 can be used for this purpose. Although, Djurić and Kay [24] proposed the MLEs of the unknown parameters of the model (98), they did not provide any properties of the MLEs. Clearly many desirable properties of the MLEs are not guaranteed as the non-linear model (98) does not satisfy the sufficient conditions of Jennrich [58], Wu [122] or Kundu [63] which are needed for the MLEs, under the assumptions of Gaussian error, or the LSEs to be consistent and asymptotically normally distributed. Later Nandi and Kundu [91] established the consistency and asymptotic normality properties of the LSEs under the assumptions that the errors are i.i.d. complex valued random variables with mean zero and finite variance. The following results are obtained, for detailed proofs, the readers are referred to Nandi and Kundu [91]. We need the following assumption before stating the main result.

Assumption 6.1 *$A^0 = A_R^0 + iA_I^0$ is an arbitrary complex number, $0 < \alpha_1^0 < \ldots < \alpha_p^0, \pi$ and $e(t)$s are i.i.d. complex valued random variables. Let us write $e(t) = e_R(t) + ie_I(t)$, where $e_R(t)$ and $e_I(t)$ are the real and imaginary parts of $e(t)$. It is assumed $E(e_R(t)) = E(e_I(t)) = 0$ and $V(e_R(t)) = V(e_I(t)) = \frac{\sigma^2}{2}$, and $e_R(t)$ and $e_I(t)$ are independently distributed.*

Theorem 6.1 *If $e(t)$s are i.i.d. complex valued random variables satisfying Assumption 6.1, then the LSEs of $(A_R, A_I, \alpha_1, \ldots, \alpha_p)^\top$, say $(\widehat{A}_R, \widehat{A}_I, \widehat{\alpha}_1, \ldots, \widehat{\alpha}_p)^\top$, are strongly consistent and*

$$\left(\sqrt{N}(\widehat{A}_R - \widehat{A}_R^0), \sqrt{N}(\widehat{A}_I - \widehat{A}_I^0), N^{\frac{3}{2}}(\widehat{\alpha}_1 - \widehat{\alpha}_1^0), \ldots, N^{\frac{2p+1}{2}}(\widehat{\alpha}_p - \widehat{\alpha}_p^0)\right)^\top \xrightarrow{d} N_{p+2}(\mathbf{0}, \sigma^2 \boldsymbol{\Sigma}_{p+2}),$$

where $\mathbf{\Sigma}_{p+2}$ is a $(p+2)\times(p+2)$ positive definite matrix, and it is defined through its inverse as follows:

$$\mathbf{\Sigma}_{p+2}^{-1} = \begin{bmatrix} 1 & 0 & -\frac{1}{2}A_I^0 & -\frac{1}{3}A_I^0 & \cdots & -\frac{1}{p+1}A_I^0 \\ 0 & 1 & \frac{1}{2}A_R^0 & \frac{1}{3}A_R^0 & \cdots & \frac{1}{p+1}A_R^0 \\ -\frac{1}{2}A_I^0 & \frac{1}{2}A_R^0 & \frac{1}{3}|A^0|^2 & \frac{1}{4}|A^0|^2 & \cdots & \frac{1}{p+2}|A^0|^2 \\ -\frac{1}{3}A_I^0 & \frac{1}{3}A_R^0 & \frac{1}{4}|A^0|^2 & \frac{1}{5}|A^0|^2 & \cdots & \frac{1}{p+3}|A^0|^2 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{p+1}A_I^0 & \frac{1}{p+1}A_R^0 & \frac{1}{p+2}|A^0|^2 & \frac{1}{p+3}|A^0|^2 & \cdots & \frac{1}{2p+1}|A^0|^2 \end{bmatrix}.$$

Similar to the periodogram estimators as proposed by Grover, Kundu and Mitra [44] described in Section 4.2, in this case also the periodogram estimators can be used. The periodogram function for the polynomial phase model (98) can be described as follows:

$$I(\alpha_1,\ldots,\alpha_p) = \left| \frac{1}{N}\sum_{t=1}^{N} y(t)e^{-i(\alpha_1 t+\ldots+\alpha_p t^p)} \right|^2. \tag{101}$$

Therefore, the estimators of $\alpha_1,\ldots,\alpha_p$ can be obtained by maximizing $I(\alpha_1,\ldots,\alpha_p)$ with respect to the unknown parameters. If

$$(\widehat{\alpha}_1,\ldots,\widehat{\alpha}_p) = \arg\max I(\alpha_1,\ldots,\alpha_p),$$

then $\widehat{\alpha}_1,\ldots,\widehat{\alpha}_p$ are called the periodogram estimators of $\alpha_1^0,\ldots,\alpha_p^0$, respectively. Once we obtain the periodogram estimators of $\alpha_1^0,\ldots,\alpha_p^0$, the periodogram estimator of A^0 can be obtained as

$$\widehat{A} = \frac{1}{N}\sum_{t=1}^{N} y(t)e^{-i(\widehat{\alpha}_1 t+\ldots+\widehat{\alpha}_p t^p)}.$$

Note that the periodogram estimators of the polynomial chirp model also can be obtained by solving a p dimensional optimization problem. If the error random variables $e(t)$s satisfy Assumption 6.1, then following similar approach of Grover, Kundu and Mitra [44], even in this case also it can be shown that the periodogram estimators are strongly consistent and they have the same asymptotic distribution as the least squares estimators.

6.2 Multicomponent Polynomial Phase Model

Although, one component polynomial phase model has been used quite extensively in modeling SAR images and in many other signal processing applications, in number of practical situations such as analysis of non-stationary signal in the presence of another non-stationary jamming signal, the multicomponent model is more relevant. A multicomponent polynomial phase model in presence of additive noise can be described as follows:

$$y(t) = \sum_{k=1}^{m} A_k^0 e^{i(\alpha_{1k}^0 t + \alpha_{2k}^0 t^2 + \ldots + \alpha_{pk}^0 t^p)} + e(t); \qquad t = 1, \ldots, N. \tag{102}$$

Here $A_1^0, \ldots, A_m^0$ are complex numbers, $e(t)$s are complex valued random variables as before, and $0 < \alpha_{1k}^0, \ldots, \alpha_{pk}^0 < \pi$, for $k = 1, \ldots, m$. The problem remains the same, i.e. to estimate the unknown parameters namely $A_1^0, \ldots, A_m^0$ and $\alpha_{1k}^0, \ldots, \alpha_{pk}^0$, for $k = 1, \ldots, m$. Note that a more general model then the model (102) will be when the degrees of the polynomials for each components are different. Although, it is a more general model, analytically it has the same challenges as the model (102). Hence, we restrict to the model (102) mainly for notational simplicity.

This problem has been considered by several authors, see for example Friedlander and Francos [37], Barbarossa, Scaglione and Giannakis [5], Ticahvsky and Handel [113], Pham and Zoubir [99] and see the references cited therein. There are mainly two issues related to multicomponent polynomial phase model. First, the interaction between the two components, these are often called cross-terms, and they give rise to undesired sinusoids in the higher order instantaneous moment, which strongly affects algorithms based on frequency estimation. Secondly, the principle of demodulation of mono component polynomial phase signals no longer works with multicomponent polynomial phase signals.

Pham and Zoubir [99] proposed to use the maximum likelihood method to estimate the unknown parameters when the errors are i.i.d. complex Gaussian random variables with mean zero and finite variance. The MLEs of the unknown parameters can be obtained by solving non-linear optimization problem. If the errors are not complex Gaussian random variables but i.i.d. with mean zero and finite variance, then the least squares method can be applied and LSEs can be obtained by minimizing

$$Q(\boldsymbol{A}, \boldsymbol{\alpha}) = \sum_{t=1}^{n} \left| y(t) - \sum_{k=1}^{m} A_k e^{i(\alpha_{1k} t + \alpha_{2k} t^2 + \ldots + \alpha_{pk} t^p)} \right|^2 \tag{103}$$

with respect to $\boldsymbol{A} = (A_1^0, \ldots, A_m^0)^\top$ and $\boldsymbol{\alpha} = (\alpha_{11}, \ldots, \alpha_{p1}, \ldots, \alpha_{1m}, \ldots, \alpha_{pm})^\top$. The properties of the MLEs or the LSEs have not yet been established. It will be interesting to develop the consistency and asymptotic normality properties of the MLEs and LSEs under different error assumptions. More work is needed along that direction.

7 Random Amplitude Chirp Model

7.1 One Component Random Amplitude Chirp Model

So far we have discussed one component and multicomponent chirp model when the amplitude is constant. Along the same line as the random amplitude sinusoidal model as proposed by Besson and Stoica [9], Besson, Ghogho and Swami [10] proposed random amplitude chirp signal model. The random amplitude chirp model can be described as follows:

$$y(t) = \alpha(t) e^{i(\theta_0^0 + \theta_1^0 t + \theta_2^0 t^2)} + e(t); \quad t = 1, \ldots, N. \tag{104}$$

Here, $e(t)$s are the additive noise random variables with mean zero and finite variance, and $\alpha(t)$ is the random time varying amplitude. The explicit structure of $\alpha(t)$ will be defined later. This kind of signal arises in many applications signal processing. It is being used quite extensively in different radar problems. Let us assume that a radar illuminating a target. The transmitted signal will then be affected by two different phenomena. First, it will usually undergo a phase shift induced by the distance and relative motion between the target and the receiver. Based on some smoothness assumption on the motion, the phase shift can be modeled as quadratic function of time. The coefficients of the quadratic functions are usually depend on the speed and acceleration of the radar, and kind of waveforms transmitted by the target. The second phenomenon is due to amplitude distortion caused either by target fluctuation or scattering of the medium. Due to these reasons, the random amplitude chirp signal (104) can be used quite effectively in dealing with different radar problems.

Besson, Ghogho and Swami [10] made the following assumptions on the random amplitude $\alpha(t)$ and the error component $e(t)$. It is assumed that $\alpha(t)$ is a real valued stationary mixing process with non-zero mean. The error component $e(t)$ is a white complex Gaussian process with mean zero and finite variance. Moreover, $e(t)$ and $\alpha(t)$ are assumed to be independent. Recently Nandi and Kundu [94] considered the random amplitude chirp model (104), and provided a theoretical justification that if the mean of the stationary mixing process $\alpha(t)$ is unknown and θ_0^0 is also unknown, then both are not identifiable. Therefore, if both are present then one has to be known. Therefore, without loss of generality, we consider the model

$$y(t) = \alpha(t)e^{i(\theta_1^0 t + \theta_2^0 t^2)} + e(t); \qquad t = 1, \ldots, N. \tag{105}$$

The problem remains the same, i.e., estimate the frequency θ_1^0 and the frequency rate θ_2^0 based on a random sample $\{y(t); t = 1, \ldots, N\}$ from the model (105). Besson, Ghogho and Swami [10] proposed the following estimators of θ_1^0 and θ_2^0. Consider the following function

$$Q(\boldsymbol{\theta}) = \frac{1}{N}\left|\sum_{t=1}^{N} y^2(t)e^{-i2(\theta_1 t + \theta_2 t^2)}\right|^2. \tag{106}$$

Note that the function $Q(\boldsymbol{\theta})$ as defined in (106) is the periodogram function of the transformed signal $y^2(t)$ with exponent replaced by twice the usual periodogram component. The unknown parameters θ_1 and θ_2 are estimated by maximizing $Q(\boldsymbol{\theta})$. Let $\widehat{\theta} = (\widehat{\theta}_1, \widehat{\theta}_2)^\top$ be the maximizer of $Q(\boldsymbol{\theta})$, then

$$\widehat{\boldsymbol{\theta}} = (\widehat{\theta}_1, \widehat{\theta}_2)^\top = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}). \tag{107}$$

In this case also, the maximization of $Q(\boldsymbol{\theta})$ can be performed by solving a two dimensional optimization problem. Once $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are obtained, then the random amplitude $\alpha(t)$ at the time point t may be estimated by

$$\widehat{\alpha}(t) = \mathrm{Re}\left\{y(t) \times e^{-(\widehat{\alpha}_1 t + \widehat{\alpha}_2 t^2)}\right\}.$$

Besson, Ghogho and Swami [10] provided the Cramer-Rao lower bound, and performed extensive simulations to show that the above estimators perform very well in practice. Besson, Ghogho and Swami [10] did not provide any theoretical properties of the estimators. Nandi and Kundu [94] provided the theoretical properties of $\widehat{\theta}_1$ and $\widehat{\theta}_2$ based on the following assumptions of $\alpha(t)$ and $e(t)$.

Assumption 7.1 *The random amplitude $\{\alpha(t)\}$ is a sequence of i.i.d. real-valued random variables with mean μ_α, variance σ_α^2, $\mu_\alpha \neq 0$ and $\sigma_\alpha^2 > 0$. The fourth moment of $\{\alpha(t)\}$ exists.*

Assumption 7.2 *The additive error $\{e(t)\}$ is a sequence of complex-valued i.i.d. random variables with mean zero and variance σ^2. Write $e(t) = e_R(t) + ie_I(t)$, then $\{e_R(t)\}$ and $\{e_I(t)\}$ are i.i.d. random variables with mean 0 and variance $\frac{\sigma^2}{2}$, have fourth moment γ, and are independently distributed.*

Assumption 7.3 *$\{e(t)\}$ is assumed to be independent of $\{\alpha(t)\}$ and $0 < \theta_1^0, \theta_2^0 < \pi$.*

It may be noted that the assumptions used by Nandi and Kundu [94] in case of the random amplitude are stronger than it has been mentioned by Besson, Ghogho and Swami [10], where as the assumptions on the error components are weaker than it has been used by Besson, Ghogho and Swami [10]. Based on the above assumptions, the following consistency and asymptotic normality results have been obtained. For detailed proofs, interested readers are referred to Nandi and Kundu [94].

Theorem 7.1 *Under Assumptions 7.1-7.3, $\widehat{\theta}_1$ and $\widehat{\theta}_2$ as defined in (107) are strongly consistent estimators of $\widehat{\theta}_1^0$ and $\widehat{\theta}_2^0$, respectively.*

Theorem 7.2 *Under Assumptions 7.1-7.3, as $N \to \infty$,*

$$\boldsymbol{D}^{-1}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) \xrightarrow{d} N_2(\mathbf{0}, 4(\sigma_\alpha^2 + \mu_\alpha^2)^2 \boldsymbol{\Sigma}^{-1}\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-1}),$$

where with $C_\alpha = 8(\sigma_\alpha^2 + \mu_\alpha^2)\sigma^2 + \frac{1}{2}\gamma + \frac{1}{8}\sigma^4$,

$$\boldsymbol{D} = \begin{pmatrix} N^{-\frac{3}{2}} & 0 \\ 0 & N^{-\frac{5}{2}} \end{pmatrix}, \qquad \boldsymbol{\Sigma} = \frac{2(\sigma_\alpha^2 + \mu_\alpha^2)^2}{3} \begin{pmatrix} 1 & 1 \\ 1 & \frac{16}{15} \end{pmatrix}, \qquad \boldsymbol{\Gamma} = C_\alpha \begin{bmatrix} \frac{1}{3} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{5} \end{bmatrix}.$$

From Theorem 7.2 it follows that the asymptotic variances of $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are

$$\begin{aligned} \mathrm{Var}(\widehat{\theta}_1) &= \frac{93}{N^3(\sigma_\alpha^2 + \mu_\alpha^2)^2} \left[8(\sigma_\alpha^2 + \mu_\alpha^2)\sigma^2 + \frac{1}{2}\gamma + \frac{1}{8}\sigma^4 \right], \\ \mathrm{Var}(\widehat{\theta}_2) &= \frac{135}{N^5(\sigma_\alpha^2 + \mu_\alpha^2)^2} \left[8(\sigma_\alpha^2 + \mu_\alpha^2)\sigma^2 + \frac{1}{2}\gamma + \frac{1}{8}\sigma^4 \right] \end{aligned}$$

respectively. In case the additive error is Gaussian with mean zero and variance σ^2, then the fourth moment is $3\sigma^4$ and the asymptotic variances reduce to

$$\begin{aligned} \mathrm{Var}(\widehat{\theta}_1) &= \frac{93}{N^3(\sigma_\alpha^2 + \mu_\alpha^2)^2} \left[8(\sigma_\alpha^2 + \mu_\alpha^2)\sigma^2 + \frac{13}{8}\sigma^4 \right], \\ \mathrm{Var}(\widehat{\theta}_2) &= \frac{135}{N^5(\sigma_\alpha^2 + \mu_\alpha^2)^2} \left[8(\sigma_\alpha^2 + \mu_\alpha^2)\sigma^2 + \frac{13}{8}\sigma^4 \right]. \end{aligned}$$

It should be mentioned that the asymptotic variances of $\widehat{\theta}_1$ and $\widehat{\theta}_2$ in case of Gaussian errors, obtained heuristically by Besson, Ghogho and Swami [10] are slightly different than the above and they are

$$\begin{aligned} \mathrm{Var}(\widehat{\theta}_1) &= \frac{96}{N^3(\sigma_\alpha^2 + \mu_\alpha^2)^2} \left[(\sigma_\alpha^2 + \mu_\alpha^2)\sigma^2 + \frac{1}{2}\sigma^4 \right], \\ \mathrm{Var}(\widehat{\theta}_2) &= \frac{90}{N^5(\sigma_\alpha^2 + \mu_\alpha^2)^2} \left[(\sigma_\alpha^2 + \mu_\alpha^2)\sigma^2 + \frac{1}{2}\sigma^4 \right]. \end{aligned}$$

Comment 7.1 *The random amplitude It may be sinusoidal model $y(t) = \alpha(t)e^{i\theta^0 t} + e(t)$ is a special case of model (105). In this case, the effective frequency does not change over time* *and it constant since the frequency rate is zero. The unknown frequency can be estimated by maximizing $Q(\theta)$, where*

$$Q(\theta) = \frac{1}{N}\left|\sum_{t=1}^{N} y^2(t)e^{-i2\theta t}\right|^2.$$

The consistency and asymptotic normality of the estimator of θ^0 follow along the same way.

In Section 6.1 we have discussed polynomial phase signal. A more general random amplitude polynomial phase signal can be defined as follows:

$$y(t) = \alpha(t)e^{i(\theta_1^0 t+\ldots+\theta_p^0 t^p)} + e(t); \qquad t = 1,\ldots,N. \tag{108}$$

Under proper conditions on the random amplitude $\alpha(t)$ and the error random variables $e(t)$, it will be interesting to develop consistency and asymptotic normality properties of the estimators of the unknown parameters. More work is need along that direction.

7.2 Multi Component Random Amplitude Chirp Model

Nandi and Kundu [94] proposed the multi component chirp model, where instead of a single frequency and chirp rate pair, p such pairs are present. The model can be written as follows:

$$y(t) = \sum_{k=1}^{p} \alpha_k(t)e^{i(\theta_{1k}^0 t+\theta_{2k}^0 t^2)} + e(t); \qquad t = 1,\ldots,N. \tag{109}$$

Here, the number of component p is assumed to be known. The additive error $e(t)$ satisfies Assumption 7.2, the random amplitudes $\alpha_1(t),\ldots,\alpha_p(t)$ and the parameters satisfy the following assumptions

Assumption 7.4 *The random amplitude corresponds to k-th component $\{\alpha_k(t)\}$ is a sequence of i.i.d. real-valued random variables with mean $\mu_{k\alpha} \neq 0$, variance $\sigma_{k\alpha}^2 > 0$ and with finite fourth moment, for $k = 1,\ldots,p$. Moreover, $\{\alpha_j(t)\}$ and $\{\alpha_k(t)\}$ for $j \neq k$, are independent.*

Assumption 7.5 *The additive error component $\{e(t)\}$ is independent of the random amplitudes $\{\alpha_1(t)\}, \ldots, \{\alpha_p(t)\}$.*

Assumption 7.6 *The parameters satisfy $0 < \theta_{11}^0, \theta_{21}^0, \ldots, \theta_{1p}^0, \theta_{2p}^0 < \pi$, and $(\theta_{1j}^0, \theta_{2j}^0) \neq (\theta_{1k}^0, \theta_{2k}^0)$, for $j \neq k$, $j, k = 1, \ldots, p$.*

In this case the unknown parameters are estimated by maximizing $Q(\boldsymbol{\theta})$, for $\boldsymbol{\theta} = (\theta_1, \theta_2)^\top$ as defined in (106) locally. Let us write $\boldsymbol{\theta}_k = (\theta_{1k}, \theta_{2k})^\top$ and $\boldsymbol{\theta}_k^0$ be the true value of $\boldsymbol{\theta}_k$. Let us further use the notation N_k, as a neighborhood of $\boldsymbol{\theta}_k^0$, such that for $j \neq k$, $\boldsymbol{\theta}_j^0 \notin N_k$. Estimate $\boldsymbol{\theta}_k$ as

$$\widehat{\boldsymbol{\theta}}_k = (\widehat{\theta}_{1k}, \widehat{\theta}_{2k})^\top = \arg\max_{(\theta_1,\theta_2)\in N_k} \frac{1}{N} \left| \sum_{t=1}^{N} y^2(t) e^{-i2(\theta_1 t + \theta_2 t^2)} \right|^2. \tag{110}$$

Nandi and Kundu [94] obtained the following results.

Theorem 7.3 *Under Assumptions 7.2 and 7.4 to 7.6, $\widehat{\theta}_k$ is a strongly consistent estimator of θ_k^0, for $k = 1, \ldots, p$.*

Theorem 7.4 *Under Assumptions 7.2 and 7.4 to 7.6, as $N \to \infty$,*

$$\boldsymbol{D}^{-1}(\widehat{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_k^0) \xrightarrow{d} N_2(\mathbf{0}, 4(\sigma_{k\alpha}^2 + \mu_{k\alpha}^2)^2 \boldsymbol{\Sigma}_k^{-1} \boldsymbol{\Gamma}_k \boldsymbol{\Sigma}_k^{-1}),$$

where with $C_{k\alpha} = 8(\sigma_{k\alpha}^2 + \mu_{k\alpha}^2)\sigma^2 + \frac{1}{2}\gamma + \frac{1}{8}\sigma^4$,

$$\boldsymbol{D} = \begin{pmatrix} N^{-\frac{3}{2}} & 0 \\ 0 & N^{-\frac{5}{2}} \end{pmatrix}, \qquad \boldsymbol{\Sigma}_k = \frac{2(\sigma_{k\alpha}^2 + \mu_{k\alpha}^2)}{3} \begin{pmatrix} 1 & 1 \\ 1 & \frac{16}{15} \end{pmatrix}, \qquad \boldsymbol{\Gamma}_k = C_{k\alpha} \begin{bmatrix} \frac{1}{3} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{5} \end{bmatrix}.$$

Theorem 7.5 *Under Assumptions 7.2 and 7.4 to 7.6, $\boldsymbol{D}^{-1}(\widehat{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_k)$ and $\boldsymbol{D}^{-1}(\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j)$ for $k \neq j$ are asymptotically independently distributed.*

8 Sum of Sinusoidal and Chirplet Model

So far we have discussed about some of the well known models namely chirp model and some related models which have been introduced in the literature for some times. Now we will be discussing a new model which has recently been introduced by Grover [42] in her Ph.D. thesis and which is very closely associated with the multicomponent chirp model and it has been named as 'chirp like model'. The 'chirp like model' can be described as follows;

$$y(t) = \sum_{j=1}^{p} \left\{ A_j^0 \cos(\alpha_j^0 t) + B_j^0 \sin(\alpha_j^0 t^2) \right\} + \sum_{k=1}^{q} \left\{ C_k^0 \cos(\beta_k^0 t^2) + D_k^0 \sin(\beta_k^0 t^2) \right\} + e(t); \quad t = 1, \ldots, N. \quad (111)$$

Here $p \geq 1, q \geq 1, A_j^0 s, B_j^0 s, C_k^0 s, D_k^0$ are real numbers, α_j^0s and β_j^0s are the frequencies, and frequency rates, respectively. It has been observed by extensive simulations that the model (111) behaves very closely with a multicomponent chirp model (89), in the sense if a signal has been generated from the model (89) then the model (111) fits the generated signal very well and vice versa. Similarly, it has been observed while analyzing several data sets that if the multicomponent chirp model (89) fits a data set well, then the 'chirp-like-model' also fits the data set well and it is very difficult to distinguish the two fitted models. On the other hand the model (111) is easier to handle both analytically as well as computational point of view. That is the main motivation of this model. While developing theoretical properties of different estimators it has been assumed that p and q are known, but Grover [42] has discussed estimation of p and q while analyzing real data set. So here, p and q are assumed to be known.

It is interesting to note that the sinusoidal frequency model can be obtained as a special case of model (111) when $C_j = D_j = 0$, for $j = 1, \ldots, q$. Similarly, a simple elementary chirp model namely

$$y(t) = C^0 \cos(\beta^0 t^2) + D^0 \sin(\beta^0 t^2) + e(t); t = 1, \ldots, N. \quad (112)$$

can be obtained as a special case of the model (111) and it has been available in the signal processing literature, see for example Casazza and Fickus [15] and Mboup and Adali [88]. Now we consider different estimation procedures of the 'chirp like model'. First we consider one component 'chirp like model' and then we will discuss about the multicomponent ''chirp like model'.

8.1 One Component Chirp Like Model

The one-component 'chirp like model' can be described as follows:

$$y(t) = A^0 \cos(\alpha^0 t) + B^0 \sin(\alpha^0 t^2) + C^0 \cos(\beta^0 t^2) + D^0 \sin(\beta^0 t^2) + e(t); \qquad t = 1, \ldots, N. \quad (113)$$

Here A^0, B^0, C^0 and D^0 are real numbers and $0 < \alpha^0, \beta^0 < \pi$, $e(t)$s are i.i.d. random variables with mean zero and finite variance. Note that the one component chirp like model has six parameters, where as one component chirp model has four parameters. Due to presence of extra parameters it is expected that the one component chirp like model (113) becomes more flexible than the one component chirp model (57).

Now the natural question is how to estimate the unknown parameters given the observed signal $\{y(t); t = 1, \ldots, N\}$. The most natural estimators of the unknown parameters will be the least squares estimators. The LSEs of $\boldsymbol{\theta}^0 = (A^0, B^0, \alpha^0, C^0, D^0, \beta^0)^\top$ can be obtained by minimizing

$$Q(\boldsymbol{\theta}) = \sum_{t=1}^{N} \left(y(t) - A^0 \cos(\alpha^0 t) - B^0 \sin(\alpha^0 t^2) - C^0 \cos(\beta^0 t^2) - D^0 \sin(\beta^0 t^2) \right)^2, \qquad (114)$$

with respect to the unknown parameters $\boldsymbol{\theta} = (A, B, \alpha, C, D, \beta)^\top$. If we write (114) in matrix notation, it becomes

$$Q(\boldsymbol{\theta}) = (\boldsymbol{Y} - \boldsymbol{Z}(\alpha, \beta)\boldsymbol{\mu})^\top (\boldsymbol{Y} - \boldsymbol{Z}(\alpha, \beta)\boldsymbol{\mu}), \qquad (115)$$

here $\boldsymbol{Y} = (y(1), \ldots, y(N))^\top$, $\boldsymbol{\mu} = (A, B, C, D)^\top$ and

$$\boldsymbol{Z}(\alpha, \beta) = \begin{bmatrix} \cos(\alpha) & \sin(\alpha) & \cos(\beta) & \sin(\beta) \\ \vdots & \vdots & \vdots & \vdots \\ \cos(N\alpha) & \sin(N\alpha) & \cos(N^2\beta) & \sin(N^2\beta) \end{bmatrix}.$$

Since $\boldsymbol{\mu}$ is a linear parameter vector, then using the method of separable regression technique of Richards [107], the LSE of $\boldsymbol{\mu}$ for a given α and β becomes

$$\widehat{\boldsymbol{\mu}}(\alpha, \beta) = [\boldsymbol{Z}(\alpha, \beta)^\top \boldsymbol{Z}(\alpha, \beta)]^{-1} \boldsymbol{Z}(\alpha, \beta)^\top \boldsymbol{Y}. \tag{116}$$

Hence, if we denote the LSEs of A, B, C and D as $\widehat{A}(\alpha, \beta)$, $\widehat{B}(\alpha, \beta)$, $\widehat{C}(\alpha, \beta)$ and $\widehat{D}(\alpha, \beta)$, respectively, then

$$\begin{aligned} R(\alpha, \beta) &= Q(\widehat{A}(\alpha, \beta), \widehat{B}(\alpha, \beta), \widehat{C}(\alpha, \beta), \widehat{D}(\alpha, \beta), \alpha, \beta) \\ &= \boldsymbol{Y}^\top (\boldsymbol{I} - \boldsymbol{Z}(\alpha, \beta)[\boldsymbol{Z}^\top(\alpha, \beta)\boldsymbol{Z}(\alpha, \beta)]^{-1} \boldsymbol{Z}^\top(\alpha, \beta))\boldsymbol{Y}. \end{aligned}$$

Hence, the LSEs of α^0 and β^0 can be obtained by minimizing $R(\alpha, \beta)$ with respect to α and β. It is a two dimensional optimization process. As we have shown before, once the LSEs of α^0 and β^0 are obtained, say $\widehat{\alpha}$ and $\widehat{\beta}$, then the LSEs of A^0, B^0, C^0 and D^0 can be easily obtained as

$$\widehat{A} = \widehat{A}(\widehat{\alpha}, \widehat{\beta}), \quad \widehat{B} = \widehat{B}(\widehat{\alpha}, \widehat{\beta}), \quad \widehat{C} = \widehat{C}(\widehat{\alpha}, \widehat{\beta}), \quad \widehat{D} = \widehat{D}(\widehat{\alpha}, \widehat{\beta}),$$

respectively. Therefore, it is observed that although the one component 'chirp like model' has six parameters, the LSEs of the unknown parameters can be obtained by solving a two dimensional optimization problem similar to the one component chirp model only. The following theoretical properties of $\widehat{\boldsymbol{\theta}} = (\widehat{A}, \widehat{B}, \widehat{\alpha}, \widehat{C}, \widehat{D}, \widehat{\beta})^\top$, the LSE of $\boldsymbol{\theta}^0$ can be obtained, see Grover [42] in this respect.

THEOREM 8.1 *Under the assumptions that there exists a K such that $0 < |A^0|, |B^0|, |C^0|, |D^0| < K$ and $e(t)$s are i.i.d. random variables with mean zero and finite variance $\sigma^2 > 0$,*

$$\widehat{\boldsymbol{\theta}} \xrightarrow{a.s.} \boldsymbol{\theta}^0.$$

THEOREM 8.2 *Under the same set of assumptions as of Theorem 8.1*

$$\boldsymbol{D}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) \xrightarrow{d} N_6(\mathbf{0}, \sigma^2\boldsymbol{\Sigma}),$$

where

$$\boldsymbol{D} = diag\{\sqrt{N}, \sqrt{N}, N\sqrt{N}, \sqrt{N}, \sqrt{N}, N^2\sqrt{N}\}, \qquad \boldsymbol{\Sigma} = \left[\begin{array}{c|c} \boldsymbol{\Sigma}^{(1)} & \mathbf{0} \\ \hline \mathbf{0} & \boldsymbol{\Sigma}^{(2)}, \end{array}\right]$$

and

$$\boldsymbol{\Sigma}^{(1)} = \frac{1}{(A^{0^2} + B^{0^2})} \begin{bmatrix} 2(A^{0^2} + 4B^{0^2}) & -6A^0B^0 & -12B^0 \\ -6A^0B^0 & 2(4A^0 + B^0) & 12A^0 \\ -12B^0 & 12A^0 & 24 \end{bmatrix},$$

$$\boldsymbol{\Sigma}^{(1)} = \frac{1}{2(C^{0^2} + D^{0^2})} \begin{bmatrix} 4C^{0^2} + 9D^{0^2} & -5C^0D^0 & -15D^0 \\ -5C^0D^0 & 9C^0 + 4D^0 & 15C^0 \\ -15D^0 & 15C^0 & 45 \end{bmatrix}.$$

Therefore, the behavior of the LSEs of the linear, frequency and chirp parameters are very similar with the cone component chirp model. It is observed that the LSEs of frequency and chirp parameters of the 'chirp like model' can be obtained by solving a two dimensional optimization problem. As we have seen in case of chirp model, the main issue about computing the LSEs is to find proper initial guesses and that involves a search of the order $O(N^3)$. If N is large, it is computationally quite challenging. Due to this reason Grover [42] proposed a sequential method which involves search of the order $O(N^2 + N)$, and which produce estimators which have the same asymptotic properties as the LSEs. We call these estimators as the sequential least squares estimators. The sequential procedure can be described as follows, we use the following notations for that purpose. Let us partition the matrix $\boldsymbol{Z}(\alpha, \beta)$ as follows:

$$\boldsymbol{Z}(\alpha, \beta) = \left[\begin{array}{c|c} \boldsymbol{Z}^{(1)}(\alpha) & \boldsymbol{Z}^{(2)}(\beta) \end{array}\right],$$

where

$$\boldsymbol{Z}^{(1)}(\alpha) = \begin{bmatrix} \cos(\alpha) & \dots & \cos(N\alpha) \\ \sin(\alpha) & \dots & \sin(N\alpha) \end{bmatrix}^\top \quad \text{and} \quad \boldsymbol{Z}^{(2)}(\beta) = \begin{bmatrix} \cos(\beta) & \dots & \cos(N^2\beta) \\ \sin(\beta) & \dots & \sin(N^2\beta) \end{bmatrix}^\top.$$

Similarly, the vectors

$$\boldsymbol{\mu} = \left(\boldsymbol{\mu}^{(1)^\top} \mid \boldsymbol{\mu}^{(2)^\top} \right)^\top, \qquad \boldsymbol{\theta} = \left(\boldsymbol{\theta}^{(1)^\top} \mid \boldsymbol{\theta}^{(2)^\top} \right)^\top$$

where $\boldsymbol{\mu}^{(1)} = (A, B)^\top$, $\boldsymbol{\mu}^{(2)} = (C, D)^\top$, $\boldsymbol{\theta}^{(1)} = (A, B, \alpha)^\top$ and $\boldsymbol{\theta}^{(2)} = (C, D, \beta)^\top$. Similarly, $\boldsymbol{\mu}^0$ and $\boldsymbol{\theta}^0$ are also partitioned. First obtain estimators of $\boldsymbol{\mu}^{(1)^0}$ and α^0 by minimizing

$$Q_1(\boldsymbol{\theta}^{(1)}) = (\boldsymbol{Y} - \boldsymbol{Z}(\alpha)\boldsymbol{\mu}^{(1)})^\top(\boldsymbol{Y} - \boldsymbol{Z}(\alpha)\boldsymbol{\mu}^{(1)}). \tag{117}$$

Note that the minimization of (117) can be obtained by solving a one dimensional optimization problem. Let the sequential least squares estimators be denoted by $\widehat{\boldsymbol{\mu}}^{(1)}$ and $\widehat{\alpha}$, i.e.

$$\widehat{\boldsymbol{\theta}}^{(1)} = (\widehat{\boldsymbol{\mu}}^{(1)^\top}, \widehat{\alpha})^\top = (\widehat{A}, \widehat{B}, \widehat{\alpha})^\top = \arg\min \ Q_1(\boldsymbol{\theta}^{(1)}).$$

Then at the second stage obtain the new data vector:

$$\boldsymbol{Y}_1 = \boldsymbol{Y} - \boldsymbol{Z}^{(1)}(\widehat{\alpha})\widehat{\boldsymbol{\mu}}^{(1)}.$$

Compute the estimators of $\boldsymbol{\mu}^{(2)^0}$ and β^0, by minimizing

$$Q_2(\boldsymbol{\theta}^{(2)}) = (\boldsymbol{Y}_1 - \boldsymbol{Z}^{(2)}(\beta)\boldsymbol{\mu}^{(2)})^\top(\boldsymbol{Y}_1 - \boldsymbol{Z}^{(2)}(\beta)\boldsymbol{\mu}^{(2)}). \tag{118}$$

In this case also the sequential least squares estimators of $\boldsymbol{\mu}^{(2)^0}$ and β^0 can be obtained by solving a one-dimensional optimization problem. As it has been mentioned before that the sequential least squares estimators also have the same asymptotic properties as the LSEs.

It will be interesting to look at the approximate least squares estimators of α and β and they can be obtained by maximizing $I_1(\alpha)$ and $I_2(\beta)$, respectively, where

$$I_1(\alpha) = \frac{1}{N}\left|\sum_{t=1}^{N} y(t)e^{-i\alpha t}\right|^2 \quad \text{and} \quad I_1(\beta) = \frac{1}{N}\left|\sum_{t=1}^{N} y(t)e^{-i\beta t^2}\right|^2.$$

Once the approximate least squares estimators of α and β are obtained the corresponding linear parameters can be obtained in explicit form. Therefore, the approximate least squares estimators also can be obtained by solving two one dimensional optimization problems, and in this case also finding the initial guesses involves search of the order $O(N^2 + N)$. It is expected that the asymptotic properties of the approximate least squares estimators should be same as the least squares estimators, although it has not been established yet. More work is needed along that direction.

8.2 Multicomponent Chirp Like Model

Now we will be discussing about different estimation procedures of the multicomponent chirp model as defined in (111). The most intuitive estimators will be the LSEs and they can be obtained by minimizing

$$Q(\boldsymbol{\theta}) = \sum_{t=1}^{N} \left(y(t) - \sum_{j=1}^{p} \left\{ A_j \cos(\alpha_j t) + B_j \sin(\alpha_j t^2) \right\} - \sum_{k=1}^{q} \left\{ C_k \cos(\beta_k t^2) + D_k \sin(\beta_k t^2) \right\} \right)^2 \tag{119}$$

with respect to the unknown parameters

$$\boldsymbol{\theta} = (A_1, B_1, \alpha_1, \ldots, A_p, B_p, \alpha_p, C_1, D_1, \beta_1, \ldots, C_q, D_q, \beta_q)^\top.$$

It can be easily seen as before, that the LSEs of $\boldsymbol{\vartheta}$ can be obtained by solving a $p + q$ dimensional optimization problem. Grover [42] provided the asymptotic properties of $\widehat{\boldsymbol{\vartheta}}$, the LSE of $\boldsymbol{\vartheta}^0$, the true parameter value of the model.

Theorem 8.3 *Under the assumptions that there exists a K such that*

$$0 < |A_1^0|, |B_1^0|, \ldots, |A_p|^0, |B_p|^0, |C_1^0|, |D_1^0|, \ldots, |C_q^0|, |D_q^0| < K$$

and $e(t)$s are i.i.d. random variables with mean zero and finite variance $\sigma^2 > 0$, then

$$\widehat{\boldsymbol{\theta}} \xrightarrow{a.s.} \boldsymbol{\theta}^0.$$

The following results provide the asymptotic distribution of $\widehat{\boldsymbol{\theta}}$.

Theorem 8.4 *Under the same set of assumptions as in Theorem 8.3,*

$$\boldsymbol{D}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0\right) \xrightarrow{d} N_{3(p+q)}(\mathbf{0}, \boldsymbol{\Sigma}).$$

Here

$$\boldsymbol{D} = diag(\underbrace{\boldsymbol{D}_1, \ldots, \boldsymbol{D}_1}_{p\ times}, \underbrace{\boldsymbol{D}_2, \ldots, \boldsymbol{D}_2}_{q\ times})$$

with

$$\boldsymbol{D}_1 = diag\left(\sqrt{N}, \sqrt{N}, N\sqrt{N}\right), \boldsymbol{D}_2 = diag\left(\sqrt{N}, \sqrt{N}, N^2\sqrt{N}\right)$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_1^{(1)} & \mathbf{0} & \ldots & \ldots & \ldots & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} & \ldots & \ldots & \mathbf{0} \\ \vdots & \vdots & \boldsymbol{\Sigma}_p^{(1)} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \ldots & 0 & \boldsymbol{\Sigma}_1^{(2)} & \ldots & \mathbf{0} \\ \mathbf{0} & \ldots & \ldots & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \ldots & \ldots & \ldots & \mathbf{0} & \boldsymbol{\Sigma}_q^{(2)} \end{bmatrix}$$

and for $j = 1, \ldots, p,\ k = 1, \ldots, q,$

$$\boldsymbol{\Sigma}_j^{(1)} = \begin{bmatrix} \frac{1}{2} & 0 & \frac{B_j^0}{4} \\ 0 & \frac{1}{2} & -\frac{A_j^0}{4} \\ \frac{B_j^0}{4} & -\frac{A_j^0}{4} & \frac{{A_j^0}^2 + {B_j^0}^2}{6} \end{bmatrix}, \quad and \quad \boldsymbol{\Sigma}_k^{(2)} = \begin{bmatrix} \frac{1}{2} & 0 & \frac{D_k^0}{6} \\ 0 & \frac{1}{2} & -\frac{C_k^0}{6} \\ \frac{D_k^0}{6} & -\frac{C_k^0}{6} & \frac{{C_k^0}^2 + {D_k^0}^2}{10} \end{bmatrix}.$$

In this case also to avoid solving $(p+q)$ optimization problem, Grover [42] proposed sequential least squares estimators with the obvious modification of the one component sequential least squares procedure as it has been described in details in Section 8.1 and based on the number theoretic Conjecture 2.1, it has been shown that the LSEs and sequential LSEs are asymptotically equivalent. Note that the sequential LSEs can be obtained by solving $(p+q)$ one dimensional optimization problems.

In case of the multicomponent 'chirp like model' also, the approximate least squares estimators also can be obtained along the same line as the one component 'chirp like model' as we have described in Section 8.1. The approximate LSEs also can be obtained by solving $(p+q)$ one dimensional optimization problems. It is expected that the approximate LSEs should be equivalent asymptotically with the LSEs, although it has not been established. It will be interesting to see the performance of the approximate LSEs and develop its properties. It is an interesting and important open problem.

9 Harmonic Chirp Model

In this chapter, we consider a special case of the chirp model, where the frequencies and frequency rates instead of being arbitrary, are harmonics of a common constant fundamental frequency and common constant fundamental frequency rate, respectively. The model can be expressed mathematically as follows:

$$y(t) = \sum_{k=1}^{p} \left\{ A_k^0 \cos(k\alpha^0 t + k\beta^0 t^2) + B_k^0 \sin(k\alpha^0 t + k\beta^0 t^2) \right\} + e(t); \qquad t = 1, \ldots, N. \qquad (120)$$

Here A_k^0s and B_k^0s are arbitrary real numbers, $0 < \alpha^0 < \pi$ is the fundamental frequency and $0 < \beta^0 < \pi$ is the fundamental frequency rate of the observed signal $y(t)$, and $e(t)$s are real valued additive error with mean zero and finite variance.

It may be mentioned that the problem of estimating the fundamental frequency of harmonic time stationary sinusoids has several applications in speech processing, communications, radar and sonar, biomedical systems, electric power, and semiconductor devices, see for example Christensen et al. [18, 19], Eddy [31] and Li, Stoica and Li [80]. An extensive amount of work has been done in developing efficient estimators and developing their properties both in the Statistical and Signal Processing Literature, see for example Quinn and Thomson [105], Irizarry [56], Nandi and Kundu [92, 95], Chang and Chen [16], Jain and Singh [57] and the references cited therein.

But in many other applications it is observed that the signal is more appropriately modeled as a sum of harmonic components of a non-stationary signal, i.e. a signal that its frequency content is varied with respect time. In active transmission used in tissue harmonic imaging in ultrasound or by mammals, namely bats, dolphins, whales, the signal is deliberately transmitted as a sum of harmonic linear frequency modulated chirps to increase the detectability of the source of interest, i.e. an organ in ultrasound or a prey in case of mammals, see for example Vespe, Jones and Baker [114], Kopsinis et al. [62].

Doweck, Amar and Cohen [29] considered the model (120 when the errors are i.i.d. Gaussian random variables with mean zero and finite variance. They have discussed about the maximum likelihood estimators of the unknown parameters assuming the number of components p to be known, and then using BIC criterion they have estimated p also. They have obtained the Cramer-Rao lower bound and discussed several computational issues associated with this problem. They have also used product high order ambiguity function method to estimate the unknown parameters, but did not provide any theoretical results. It will be interesting to develop theoretical properties of these estimators.

Grover [42] considered the LSEs and Approximate LSEs of the unknown parameters under a fairly general set of assumptions on the additive errors $e(t)$s and provided the theoretical properties of these estimators. Before providing the exact results let us define the following notations: $\boldsymbol{\theta} = (A_1, B_1, \ldots, A_P, B_p, \alpha, \beta)^\top$. The true parameter value will be denoted by $\boldsymbol{\theta}^0$. The LSEs of the unknown parameters can be obtained by minimizing

$$Q(\boldsymbol{\theta}) = \sum_{t=1}^{N} \left(y(t) - \sum_{k=1}^{p} \left\{ A_k \cos(k\alpha t + k\beta t^2) + B_k \sin(k\alpha t + k\beta t^2) \right\} \right)^2 \qquad (121)$$

with respect to $\boldsymbol{\theta}$. Using matrix notation (121) can be written as

$$Q(\boldsymbol{\theta}) = \left(\boldsymbol{Y} - \boldsymbol{Z}(\alpha, \beta)\boldsymbol{\phi}\right)^\top \left(\boldsymbol{Y} - \boldsymbol{Z}(\alpha, \beta)\boldsymbol{\phi}\right). \qquad (122)$$

Here $\boldsymbol{Y} = (y(1), \ldots, y(N))^\top$ is the observed data vector, $\boldsymbol{\phi} = (A_1, B_1, \ldots, A_p, B_P)^\top$ is the vector of linear parameters and the matrix $\boldsymbol{Z}(\alpha,\beta)$ is defined as

$$\boldsymbol{Z}(\alpha,\beta) = \begin{bmatrix} \cos(\alpha+\beta) & \sin(\alpha+\beta) & \dots & \cos(p\alpha+p\beta) & \sin(p\alpha+p\beta) \\ \vdots & \vdots & \dots & \vdots & \vdots \\ \cos(N\alpha+N^2\beta) & \sin(N\alpha+N^2\beta) & \dots & \cos(pN\alpha+pN^2\beta) & \sin(pN\alpha+pN\beta) \end{bmatrix}. \tag{123}$$

Since $\boldsymbol{\phi}$ is a vector of linear parameters, for given α and β, they can be estimated using simple linear regression technique as discussed before, and it will be

$$\widehat{\boldsymbol{\phi}}(\alpha,\beta) = \left[\boldsymbol{Z}^\top(\alpha,\beta)\boldsymbol{Z}(\alpha,\beta)\right]^{-1}\boldsymbol{Z}^\top(\alpha,\beta)\boldsymbol{Y}. \tag{124}$$

Therefore, the LSEs of α and β can be obtained by minimizing

$$\begin{aligned} R(\alpha,\beta) &= \left(\boldsymbol{Y} - \boldsymbol{Z}(\alpha,\beta)\widehat{\boldsymbol{\phi}}(\alpha,\beta)\right)^\top \left(\boldsymbol{Y} - \boldsymbol{Z}(\alpha,\beta)\widehat{\boldsymbol{\phi}}(\alpha,\beta)\right) \\ &= \boldsymbol{Y}^\top(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{Z}}(\alpha,\beta))\boldsymbol{Y}, \end{aligned} \tag{125}$$

where $\boldsymbol{P}_{\boldsymbol{Z}}(\alpha,\beta) = \boldsymbol{Z}(\alpha,\beta)[\boldsymbol{Z}^\top(\alpha,\beta)\boldsymbol{Z}(\alpha,\beta)]^{-1}\boldsymbol{Z}^\top(\alpha,\beta)$ is the projection matrix on the column space of the matrix $\boldsymbol{Z}(\alpha,\beta)$. Hence, the LSEs of α^0 and β^0 can be obtained by minimizing $R(\alpha,\beta)$ with respect to α and β and let us denote them as $\widehat{\alpha}$ and $\widehat{\beta}$, respectively, i.e.

$$(\widehat{\alpha},\widehat{\beta})^\top = \arg\min R(\alpha,\beta). \tag{126}$$

Hence, the LSE of $\boldsymbol{\phi}$ becomes

$$\widehat{\boldsymbol{\phi}} = \widehat{\boldsymbol{\phi}}(\widehat{\alpha},\widehat{\beta}) = \left[\boldsymbol{Z}^\top(\widehat{\alpha},\widehat{\beta})\boldsymbol{Z}(\widehat{\alpha},\widehat{\beta})\right]^{-1}\boldsymbol{Z}^\top(\widehat{\alpha},\widehat{\beta})\boldsymbol{Y}. \tag{127}$$

Grover [42] obtained the following consistency result of the LSEs.

THEOREM 9.1 *It is assumed that there exits a K, such that $0 < |A_1|^2, |B_1|^2, \dots, |A_p|^2, |B_p|^2 < K$, $0 < \alpha^0, \beta^0 < \dfrac{\pi}{p}$ and the error random variables $e(t)$s satisfy Assumption 4.1. Then, $\widehat{\boldsymbol{\theta}}$, the LSE of $\boldsymbol{\theta}^0$, is strongly consistent i.e.*

$$\widehat{\boldsymbol{\theta}} \xrightarrow{d} \boldsymbol{\theta}^0.$$

The asymptotic distribution of the LSEs also has been obtained by Grover [42], and it is as follows:

Theorem 9.2 *Under the same set of assumptions as in Theorem 9.1*

$$\boldsymbol{D}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) \xrightarrow{d} N_{2p+2}(\mathbf{0}, 2c\sigma^2\boldsymbol{\Sigma}).$$

Here $\boldsymbol{D} = diag\{\underbrace{\sqrt{N}, \ldots, \sqrt{N}}_{2p\ times}, N\sqrt{N}, N^2\sqrt{N}\}$, $c = \sum_{j=-\infty}^{\infty} a^2(j)$, *the matrix*

$$\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix},$$

where $\boldsymbol{I}$ *is a* $2p \times 2p$ *identity matrix,* $\boldsymbol{\Sigma}_{12}$ *is a* $2p \times 2$ *matrix,* $\boldsymbol{\Sigma}_{22}$ *is a* $2p \times 2p$ *matrix as given below:*

$$\boldsymbol{\Sigma}_{12} = \begin{bmatrix} \frac{B_1^0}{2} & \ldots & \frac{pB_p^0}{2} & -\frac{A_1^0}{2} & \ldots & -\frac{A_p^0}{2} \\ \frac{B_1^0}{3} & \ldots & \frac{pB_p^0}{3} & -\frac{A_1^0}{3} & \ldots & -\frac{A_p^0}{3} \end{bmatrix},$$

$\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21}^\top$ *and*

$$\boldsymbol{\Sigma}_{22} = \begin{bmatrix} \frac{\sum_{k=1}^p k^2({A_k^0}^2 + {B_k^0}^2)}{3} & \frac{\sum_{k=1}^p k^2({A_k^0}^2 + {B_k^0}^2)}{4} \\ \frac{\sum_{k=1}^p k^2({A_k^0}^2 + {B_k^0}^2)}{4} & \frac{\sum_{k=1}^p k^2({A_k^0}^2 + {B_k^0}^2)}{5} \end{bmatrix}.$$

Grover [42] also proposed approximate LSEs of the unknown parameters, and they can be obtained by maximizing $I(\alpha, \beta)$, where

$$\begin{aligned} I(\alpha, \beta) &= \frac{1}{N} \left| \sum_{t=1}^{N} y(t) e^{-i(\alpha t + \beta t^2)} \right|^2 \\ &= \frac{1}{N} \boldsymbol{Y}^\top \boldsymbol{Z}(\alpha, \beta) \boldsymbol{Z}^\top(\alpha, \beta) \boldsymbol{Y}. \end{aligned} \tag{128}$$

The maximization of $I(\alpha, \beta)$ can be obtained by solving a two dimensional optimization problem. If $\widehat{\alpha}$ and $\widehat{\beta}$ maximize $I(\alpha, \beta)$, i.e.

$$(\widehat{\alpha}, \widehat{\beta})^\top = \text{arg max} \;\; I(\alpha, \beta),$$

then the corresponding approximate LSE of $\boldsymbol{\phi}$ can be obtained as

$$\widehat{\phi} = \frac{2}{N} \boldsymbol{Z}^\top(\widehat{\alpha}, \widehat{\beta}) \boldsymbol{Y}.$$

Under the same set of assumptions as in Theorem 9.1 Grover [42] showed that the LSEs and approximate LSEs are asymptotically equivalent, i.e. approximate LSEs are also consistent and they have the same asymptotic distribution as the LSEs.

10 Two Dimensional Chirp Model

So far we have mainly talked about one dimensional chirp model and some of its variants. In this section we will be discussing about the two dimensional (2-D) chirp model in presence of additive errors and some of its variants. Mathematically a 2-D chirp model can be expressed as follows:

$$\begin{aligned} y(m,n) &= \sum_{k=1}^{p} \left\{ A_k^0 \cos(\alpha_k^0 m + \beta_k^0 m^2 + \gamma_k^0 n + \delta_k^0 n^2) + B_k^0 \cos(\alpha_k^0 m + \beta_k^0 m^2 + \gamma_k^0 n + \delta_k^0 n^2) \right\} \\ & \qquad + e(m,n); \quad m = 1, \ldots, M, \quad n = 1, \ldots, N. \end{aligned} \tag{129}$$

Here, $y(m,n)$ is the observed signal in two dimension, A_k^0s, B_k^0s are real numbers, α_k^0s, γ_k^0s are frequencies, and β_k^0s, δ_k^0s are frequency rates. The random component $e(m,n)$s are real valued random variables with mean zero and finite variance. The more explicit assumptions on $e(m,n)$s will be stated later.

The 2-D chirp model is a natural generalization of the 2-D sinusoidal model which can be described as follows:

$$\begin{aligned} y(m,n) &= \sum_{k=1}^{p} \left\{ A_k^0 \cos(\alpha_k^0 m + \gamma_k^0 n) + B_k^0 \cos(\alpha_k^0 m + \gamma_k^0 n) \right\} \\ & \qquad + e(m,n); \quad m = 1, \ldots, M, \quad n = 1, \ldots, N. \end{aligned} \tag{130}$$

The 2-D sinusoidal model (130) has received considerable amount of attention in the signal processing literature. This model has different applications in 'Multidimensional Signal Processing'. This is basic model in many fields such as antenna array processing, geophysical perception, biomedical spectral analysis etc. see for example the work by Barbieri and Barone

[6], Cabrera and Bose [13], Chun and Bose [20], Hua [52] and see the references cited therein. This problem has a special interest in spectrography, and it has bees studied quite rigorously by Malliavan [84, 85].

The above 2D-chirp model (129) is a natural generalization of the 2-D sinusoidal model and 1-D chirp model. Many 2-D signal processing applications require modeling and analysis of non-homogeneous signals. In fact, almost any application where image interpretation is required, has to face challenge of analyzing non-homogeneous in the observed image. Model (129) and some of its variants have been used in modeling and analyzing magnetic resonance imaging (MRI), optical imaging and different texture imaging. It has been also used in modeling black and white 'gray' images, and to analyze finger print images data. This model has a wide applications in modeling Synthetic Aperture Radar (SAR) data and in particular Interferometric SAR data. See, for example, Pelag and Porat [98], Hedley and Rosenfeld [50], Friedlander and Francos [38], Francos and Friedlander [35, 36], Cao, Wang and Wang [14], Zhang and Liu [129], Zhang, Wang and Cao [130], and see the references cited therein.

We mainly discuss different estimation procedures and their properties for single component 2-D chirp model. All the results can be easily generalized for multiple 2-D chirp model using sequential procedures as described before for 1-D chirp model.

10.1 Least Squares Estimators

The one component 2-D chirp model can be written as follows:

$$\begin{aligned} y(m,n) &= A^0\cos(\alpha^0 m + \beta^0 m^2 + \gamma^0 n + \delta^0 n^2) + B^0 \sin(\alpha^0 m + \beta^0 m^2 + \gamma^0 n + \delta^0 n^2) \\ &\quad + e(m,n); \quad m = 1,\ldots,M, n = 1,\ldots,N. \end{aligned} \tag{131}$$

Here one component 2-D chirp model (131) can be obtained from model (129) when $p = 1$. The problem is to find the estimators of the unknown parameters based on the observed data $y(m,n)$, under a suitable error assumption. First let us assume that the error components $e(m,n)$s are i.i.d. random variables with mean zero and finite variance σ^2, for $m = 1, \ldots, M$ and $n = 1, \ldots, N$. More general error assumptions will be considered in subsequent sections.

Based on the assumptions that the errors are i.i.d. random variables the most reasonable estimators will be the LSEs, and they can be obtained by minimizing the residual sums of squares, i.e.

$$Q(\boldsymbol{\theta}) = \sum_{m=1}^{M}\sum_{n=1}^{N}\left(y(m,n) - A\cos(\alpha m + \beta m^2 + \gamma n + \delta n^2) - B\sin(\alpha m + \beta m^2 + \gamma n + \delta n^2)\right)^2, \tag{132}$$

where $\boldsymbol{\theta} = (A, B, \alpha, \beta, \gamma, \delta)^\top$. The LSEs of the unknown parameters $\boldsymbol{\theta}$, say $\widehat{\boldsymbol{\theta}}$, can be obtained as the argument minimum of $Q(\boldsymbol{\theta})$, i.e.

$$\widehat{\boldsymbol{\theta}} = (\widehat{A}, \widehat{B}, \widehat{\alpha}, \widehat{\beta}, \widehat{\gamma}, \widehat{\delta})^\top = \text{arg min} \;\; Q(\boldsymbol{\theta}).$$

As expected, the LSEs cannot be obtained in explicit forms. One needs to use some numerical techniques to compute the MLEs. Newton-Raphson, Gauss-Newton, Genetic algorithm or simulated annealing method may be used for this purpose. Alternatively, the method suggested by Saha and Kay [109] as described in the Section 4.1 may be used to find the LSEs of the unknown parameters. The details are avoided.

Lahiri [74], see also Lahiri and Kundu [75] in this respect, established the asymptotic properties of the LSEs under a fairly general set of conditions. It has been shown that the LSEs are strongly consistent and they are asymptotically normally distributed. The results are provided in details below.

THEOREM 10.1 *If there exists a K, such that $0 < |A^0| + |B^0| < K$, $0 < \alpha^0, \beta^0, \gamma^0, \delta^0 <$*

π, and $\sigma^2 > 0$, then $\widehat{\boldsymbol{\theta}} = (\widehat{A}, \widehat{B}, \widehat{\alpha}, \widehat{\beta}, \widehat{\gamma}, \widehat{\delta})^\top$ is a strongly consistent estimate of $\boldsymbol{\theta}^0 = (A^0, B^0, \alpha^0, \beta^0, \gamma^0, \delta^0)^\top$. ∎

Theorem 10.2 *Under the same assumptions as in Theorem 10.1, if we denote $\boldsymbol{D}$ as a 6×6 diagonal matrix as*

$$\boldsymbol{D} = diag\left\{M^{1/2}N^{1/2}, M^{1/2}N^{1/2}, M^{3/2}N^{1/2}, M^{5/2}N^{1/2}, M^{1/2}N^{3/2}, M^{1/2}N^{5/2}\right\},$$

then

$$\boldsymbol{D}(\widehat{A} - A^0, \widehat{B} - B^0, \widehat{\alpha} - \alpha^0, \widehat{\beta} - \beta^0, \widehat{\gamma} - \gamma^0, \widehat{\delta} - \delta^0)^\top \xrightarrow{d} N_6(\mathbf{0}, 2\sigma^2\boldsymbol{\Sigma}),$$

where

$$\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} 1 & 0 & \frac{B^0}{2} & \frac{B^0}{3} & \frac{B^0}{2} & \frac{B^0}{3} \\ 0 & 1 & -\frac{A^0}{2} & -\frac{A^0}{3} & -\frac{A^0}{2} & -\frac{A^0}{3} \\ \frac{B^0}{2} & -\frac{A^0}{2} & \frac{{A^0}^2+{B^0}^2}{3} & \frac{{A^0}^2+{B^0}^2}{4} & \frac{{A^0}^2+{B^0}^2}{4} & \frac{{A^0}^2+{B^0}^2}{4.5} \\ \frac{B^0}{3} & -\frac{A^0}{3} & \frac{{A^0}^2+{B^0}^2}{4} & \frac{{A^0}^2+{B^0}^2}{5} & \frac{{A^0}^2+{B^0}^2}{4.5} & \frac{{A^0}^2+{B^0}^2}{5} \\ \frac{B^0}{2} & -\frac{A^0}{2} & \frac{{A^0}^2+{B^0}^2}{4} & \frac{{A^0}^2+{B^0}^2}{4.5} & \frac{{A^0}^2+{B^0}^2}{3} & \frac{{A^0}^2+{B^0}^2}{4} \\ \frac{B^0}{3} & -\frac{A^0}{3} & \frac{{A^0}^2+{B^0}^2}{4.5} & \frac{{A^0}^2+{B^0}^2}{5} & \frac{{A^0}^2+{B^0}^2}{4} & \frac{{A^0}^2+{B^0}^2}{5} \end{bmatrix}.$$

∎

Note that the above result Theorem 10.2 can be used to construct confidence intervals of the unknown parameters. They can be used for testing of hypothesis problem also. Since, the LSEs are MLEs under the assumption of i.i.d. Gaussian errors, the above LSEs are MLEs also when the errors are i.i.d. Gaussian random variables. In fact, the LSEs are consistent and asymptotically normally distributed even for a more general error assumptions. Let us make the following assumption on the error component $e(m,n)$.

Assumption 10.1 *The error component $e(m,n)$ has the following form:*

$$e(m,n) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} a(j,k) X(m-j, n-k),$$

with

$$\sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} |a(j,k)| < \infty,$$

where, $\{X(m,n)\}$ is a double array sequence of i.i.d. random variables with mean zero, variance σ^2, and with finite fourth moment.

Lahiri [74] established that under Assumption 10.1, the LSE of $\boldsymbol{\theta}$ is strongly consistent under the same assumption as in Theorem 10.1. Moreover, the LSEs are asymptotically normally distributed as provided in the following theorem.

THEOREM 10.3 *Under the same assumptions as in Theorem 10.1 and Assumption 10.1,*

$$\boldsymbol{D}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) \xrightarrow{d} N_6(\mathbf{0}, 2c\sigma^2\boldsymbol{\Sigma}),$$

where matrices $\boldsymbol{D}$ and $\boldsymbol{\Sigma}$ are same as defined in Theorem 10.2 and

$$c = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} a^2(j,k).$$

■

10.2 Approximate Least Squares Estimators

As an alternative to the LSEs, Grover, Kundu and Mitra [44] considered approximate LSEs which can be obtained by maximizing a 2-D periodogram type function defined as follows:

$$I(\alpha,\beta,\gamma,\delta) = \frac{2}{MN}\left\{\left(\sum_{m=1}^{M}\sum_{n=1}^{N} y(m,n)\cos(\alpha m + \beta m^2 + \gamma n + \delta n^2)\right)^2 + \left(\sum_{m=1}^{M}\sum_{n=1}^{N} y(m,n)\sin(\alpha m + \beta m^2 + \gamma n + \delta n^2)\right)^2\right\}. \qquad (133)$$

The main idea about the above 2-D periodogram type function has been obtained from the periodogram estimator in case the 2-D sinusoidal model. It has been observed by Kundu and Nandi [71] that the 2-D periodogram type estimators for a 2-D sinusoidal model are consistent and asymptotically equivalent to the corresponding LSEs. The 2-D periodogram type estimators or the ALSEs of the 2-D chirp model can be obtained by the argument maximum of $I(\alpha, \beta, \gamma, \delta)$ given in (133) over the range $(0, \pi) \times (0, \pi) \times (0, \pi) \times (0, \pi)$. The explicit solutions of the argument maximum of $I(\alpha, \beta, \gamma, \delta)$ cannot be obtained analytically. Numerical methods are required to compute the ALSEs. Extensive simulation experiments have been performed by Grover et al. [44], and it has been observed that the Downhill-Simplex method performs quite well to compute the ALSEs, provided the initial guesses are quite close to the true values. It has been observed in simulation studies that although both LSEs and ALSEs involve solving 4-D optimization problem, computational time of the ALSEs is significantly lower than the LSEs. It has been established that the LSEs and ALSEs are asymptotically equivalent. Therefore, the ALSEs are strongly consistent and the asymptotic distribution of the ALSEs is same as the LSEs.

Since the computation of the LSEs or the ALSEs is a challenging problem, several other computationally efficient methods are available in the literature. But unfortunately in most the cases, either the asymptotic properties are unknown or they may not have the same efficiency as the LSEs or ALSEs. Now we provide two estimators which have the same asymptotic efficiency as the LSEs or ALSEs and at the same time both of them can be computed more efficiently than the LSEs or the ALSEs.

10.3 2-D Finite Step Efficient Algorithm

Lahiri, Kundu and Mitra [77] proposed finite step efficient algorithm for 2-D one component chirp signal model. It is observed that if we start with the initial guesses of α^0 and γ^0 having convergence rates $O_p(M^{-1}N^{-1/2})$ and $O_p(N^{-1}M^{-1/2})$, respectively, and β^0 and δ^0 having convergence rates $O_p(M^{-2}N^{-1/2})$ and $O_p(N^{-2}M^{-1/2})$, respectively, then after four iterations the algorithm produces estimates of α^0 and γ^0 having convergence rates $O_p(M^{-3/2}N^{-1/2})$ and $O_p(N^{-3/2}M^{-1/2})$, respectively, and β^0 and δ^0 having convergence rates $O_p(M^{-5/2}N^{-1/2})$ and $O_p(N^{-5/2}M^{-1/2})$, respectively. Therefore, the efficient algorithm produces estimates which have the same rates of convergence as the LSEs or the ALSEs. Moreover, it is guaranteed that the algorithm stops after four iterations.

Before providing the algorithm in details, we introduce the following notation and some preliminary results which we need for further development. If $\widetilde{\alpha}$ is an estimator of α^0 such that $\widetilde{\alpha} - \alpha^0 = O_p(M^{(-1-\lambda_{11})}N^{-\lambda_{12}})$, for some $0 < \lambda_{11}, \lambda_{12} \leq 1/2$, and $\widetilde{\beta}$ is an estimator of β^0 such that $\widetilde{\beta} - \beta^0 = O_p(M^{(-2-\lambda_{21})}N^{-\lambda_{22}})$, for some $0 < \lambda_{21}, \lambda_{22} \leq 1/2$, then an improved estimator of α^0 can be obtained as

$$\widetilde{\widetilde{\alpha}} = \widetilde{\alpha} + \frac{48}{M^2}\text{Im}\left(\frac{P_{MN}^{\alpha}}{Q_{MN}^{\alpha,\beta}}\right), \tag{134}$$

with

$$P_{MN}^{\alpha} = \sum_{n=1}^{N}\sum_{m=1}^{M} y(m,n)\left(m - \frac{M}{2}\right)e^{-i(\widetilde{\alpha}m+\widetilde{\beta}m^2)} \tag{135}$$

$$Q_{MN}^{\alpha,\beta} = \sum_{n=1}^{N}\sum_{m=1}^{M} y(m,n)e^{-i(\widetilde{\alpha}m+\widetilde{\beta}m^2)}. \tag{136}$$

Similarly, an improved estimator of β^0 can be obtained as

$$\widetilde{\widetilde{\beta}} = \widetilde{\beta} + \frac{45}{M^4}\text{Im}\left(\frac{P_{MN}^{\beta}}{Q_{MN}^{\alpha,\beta}}\right), \tag{137}$$

with

$$P_{MN}^{\beta} = \sum_{n=1}^{N}\sum_{m=1}^{M} y(m,n)\left(m^2 - \frac{M^2}{3}\right)e^{-i(\widetilde{\alpha}m+\widetilde{\beta}m^2)} \tag{138}$$

and $Q_{MN}^{\alpha,\beta}$ is same as defined above in (136).

The following two results provide the justification for the improved estimators, whose proofs can be obtained in Lahiri, Kundu and Mitra [77].

THEOREM 10.4 *If* $\widetilde{\alpha} - \alpha^0 = O_p(M^{-1-\lambda_{11}} N^{-\lambda_{12}})$ *for* $\lambda_{11}, \lambda_{12} > 0$*, then*

$$\begin{array}{ll} (a)\ (\widetilde{\widetilde{\alpha}} - \alpha^0) = O_p(M^{-1-2\lambda_{11}} N^{-\lambda_{12}}) & if \quad \lambda_{11} \leq 1/4, \\ (b)\ M^{3/2} N^{1/2} (\widetilde{\widetilde{\alpha}} - \alpha^0) \stackrel{d}{\longrightarrow} N(0, \sigma_1^2) & if \quad \lambda_{11} > 1/4, \lambda_{12} = 1/2, \end{array}$$

where $\sigma_1^2 = \dfrac{384 c \sigma^2}{A^{0^2} + B^{0^2}}$*, the asymptotic variance of the LSE of* α^0*, and c is same as defined in Theorem 10.3.* ∎

THEOREM 10.5 *If* $\widetilde{\beta} - \beta^0 = O_p(M^{-2-\lambda_{21}} N^{-\lambda_{22}})$ *for* $\lambda_{21}, \lambda_{22} > 0$*, then*

$$\begin{array}{ll} (a)\ (\widetilde{\widetilde{\beta}} - \beta^0) = O_p(M^{-2-\lambda_{21}} N^{-\lambda_{22}}) & if \quad \lambda_{21} \leq 1/4, \\ (b)\ M^{5/2} N^{1/2} (\widetilde{\widetilde{\beta}} - \beta^0) \stackrel{d}{\longrightarrow} N(0, \sigma_2^2) & if \quad \lambda_{21} > 1/4, \lambda_{22} = 1/2, \end{array}$$

where $\sigma_2^2 = \dfrac{360 c \sigma^2}{A^{0^2} + B^{0^2}}$*, the asymptotic variance of the LSE of* β^0*, and c is same as defined in the previous theorem.* ∎

In order to find estimators of γ^0 and δ^0, interchange the roles of M and N. If $\widetilde{\gamma}$ is an estimator of γ^0 such that $\widetilde{\gamma} - \gamma^0 = O_p(N^{(-1-\kappa_{11})} N^{-\kappa_{12}})$, for some $0 < \kappa_{11}, \kappa_{12} \leq 1/2$, and $\widetilde{\delta}$ is an estimator of δ^0 such that $\widetilde{\delta} - \delta^0 = O_p(M^{(-2-\kappa_{21})} N^{-\kappa_{22}})$, for some $0 < \kappa_{21}, \kappa_{22} \leq 1/2$, then an improved estimator of γ^0 can be obtained as

$$\widetilde{\widetilde{\gamma}} = \widetilde{\gamma} + \frac{48}{N^2} \text{Im} \left(\frac{P_{MN}^{\gamma}}{Q_{MN}^{\gamma,\delta}} \right), \tag{139}$$

with

$$P_{MN}^{\gamma} = \sum_{n=1}^{N} \sum_{m=1}^{M} y(m,n) \left(n - \frac{N}{2} \right) e^{-i(\widetilde{\gamma} n + \widetilde{\delta} n^2)} \tag{140}$$

$$Q_{MN}^{\gamma,\delta} = \sum_{n=1}^{N} \sum_{m=1}^{M} y(m,n) e^{-i(\widetilde{\gamma} n + \widetilde{\delta} n^2)}, \tag{141}$$

and an improved estimator of δ^0 can be obtained as

$$\widetilde{\widetilde{\delta}} = \widetilde{\delta} + \frac{45}{N^4} \text{Im} \left(\frac{P_{MN}^{\delta}}{Q_{MN}^{\gamma,\delta}} \right), \tag{142}$$

with

$$P^{\delta}_{MN} = \sum_{n=1}^{N}\sum_{m=1}^{M} y(m,n)\left(n^2 - \frac{N^2}{3}\right) e^{-i(\widetilde{\gamma}n + \widetilde{\delta}n^2)} \tag{143}$$

and $Q^{\gamma,\delta}_{MN}$ is same as defined in (141).

The following two results provide the justification for the improved estimators, the details can be obtained in Lahiri, Kundu and Mitra [77].

THEOREM 10.6 *If* $\widetilde{\gamma} - \gamma^0 = O_p(N^{-1-\kappa_{11}}M^{-\kappa_{12}})$ *for* $\kappa_{11}, \kappa_{12} > 0$, *then*

$$\begin{array}{lll} (a)(\widetilde{\widetilde{\gamma}} - \gamma^0) = O_p(N^{-1-2\kappa_{11}}M^{-\kappa_{12}}) & if & \kappa_{11} \leq 1/4, \\ (b) N^{3/2}M^{1/2}(\widetilde{\widetilde{\gamma}} - \gamma^0) \xrightarrow{d} N(0, \sigma_1^2) & if & \kappa_{11} > 1/4, \kappa_{12} = 1/2. \end{array}$$

Here σ_1^2 *and* c *are same as defined in Theorem 10.3.* ∎

THEOREM 10.7 *If* $\widetilde{\delta} - \delta^0 = O_p(N^{-2-\kappa_{21}}M^{-\kappa_{22}})$ *for* $\kappa_{21}, \kappa_{22} > 0$, *then*

$$\begin{array}{lll} (a)(\widetilde{\widetilde{\delta}} - \delta^0) = O_p(N^{-2-\kappa_{21}}M^{-\kappa_{22}}) & if & \kappa_{21} \leq 1/4, \\ (b) N^{5/2}M^{1/2}(\widetilde{\widetilde{\delta}} - \delta^0) \xrightarrow{d} N(0, \sigma_2^2) & if & \kappa_{21} > 1/4, \kappa_{22} = 1/2. \end{array}$$

Here σ_2^2 *and* c *are same as defined in Theorem 10.5.* ∎

Now we show that starting from the initial guesses $\widetilde{\alpha}$, $\widetilde{\beta}$, with convergence rates $\widetilde{\alpha} - \alpha^0 = O_p(M^{-1}N^{-1/2})$ and $\widetilde{\beta} - \beta^0 = O_p(M^{-2}N^{-1/2})$, respectively, how the above results can be used to obtain efficient estimators, which have the same rate of convergence as the LSEs. It may be noted that finding initial guesses with the above convergence rates are not difficult. It can be obtained by finding the minimum of $Q_1(\alpha, \beta)$, where

$$Q_1(\alpha,\beta) = \sum_{m=1}^{M}\left(\sum_{n=1}^{N} y(m,n) - \mathcal{A}\cos(\alpha m + \beta m^2) - \mathcal{B}\sin(\alpha m + \beta m^2)\right)^2.$$

Here also the main idea is not to use the whole sample at the beginning. We use part of the sample at the beginning and gradually proceed towards the complete sample. The algorithm is described below. We denote the estimates of α^0 and β^0 obtained at the j-th iteration as $\widetilde{\alpha}^{(j)}$ and $\widetilde{\beta}^{(j)}$, respectively.

Algorithm 4:

Step 1: Choose $M_1 = M^{8/9}$, $N_1 = N$. Therefore,

$$\begin{aligned}\widetilde{\alpha}^{(0)} - \alpha^0 &= O_p(M^{-1}N^{-1/2}) = O_p(M_1^{-1-1/8}N_1^{-1/2}) \quad \text{and} \\ \widetilde{\beta}^{(0)} - \beta^0 &= O_p(M^{-2}N^{-1/2}) = O_p(M_1^{-2-1/4}N_1^{-1/2}).\end{aligned}$$

Perform steps (134) and (137). Therefore, after 1-st iteration, we have

$$\begin{aligned}\widetilde{\alpha}^{(1)} - \alpha^0 &= O_p(M_1^{-1-1/4}N_1^{-1/2}) = O_p(M^{-10/9}N^{-1/2}) \quad \text{and} \\ \widetilde{\beta}^{(1)} - \beta^0 &= O_p(M_1^{-2-1/2}N_1^{-1/2}) = O_p(M^{-20/9}N^{-1/2}).\end{aligned}$$

Step 2: Choose $M_2 = M^{80/81}$, $N_1 = N$. Therefore,

$$\begin{aligned}\widetilde{\alpha}^{(1)} - \alpha^0 &= O_p(M_2^{-1-1/8}N_2^{-1/2}) \quad \text{and} \\ \widetilde{\beta}^{(1)} - \beta^0 &= O_p(M_2^{-2-1/4}N_2^{-1/2}).\end{aligned}$$

Perform steps (134) and (137). Therefore, after 2-nd iteration, we have

$$\begin{aligned}\widetilde{\alpha}^{(2)} - \alpha^0 &= O_p(M_2^{-1-1/4}N_2^{-1/2}) = O_p(M^{-100/81}N^{-1/2}) \quad \text{and} \\ \widetilde{\beta}^{(2)} - \beta^0 &= O_p(M_2^{-2-1/2}N_2^{-1/2}) = O_p(M^{-200/81}N^{-1/2}).\end{aligned}$$

Step 3: Choose $M_3 = M$, $N_3 = N$. Therefore,

$$\begin{aligned}\widetilde{\alpha}^{(2)} - \alpha^0 &= O_p(M_3^{-1-19/81}N_3^{-1/2}) \quad \text{and} \\ \widetilde{\beta}^{(2)} - \beta^0 &= O_p(M_2^{-2-38/81}N_3^{-1/2}).\end{aligned}$$

Again, performing steps (134) and (137), after 3-rd iteration, we have

$$\begin{aligned}\widetilde{\alpha}^{(3)} - \alpha^0 &= O_p(M^{-1-38/81}N^{-1/2}) \quad \text{and} \\ \widetilde{\beta}^{(3)} - \beta^0 &= O_p(M^{-2-76/81}N^{-1/2}).\end{aligned}$$

Step 4: Choose $M_4 = M$, $N_4 = N$, and after performing steps (134) and (137) we obtain the required convergence rates, i.e.

$$\begin{aligned}\widetilde{\alpha}^{(4)} - \alpha^0 &= O_p(M^{-3/2}N^{-1/2}) \quad \text{and} \\ \widetilde{\beta}^{(4)} - \beta^0 &= O_p(M^{-5/2}N^{-1/2}).\end{aligned}$$

Similarly, interchanging the role of M and N, we can get the algorithm corresponding to γ^0 and δ^0. Extensive simulation experiments have been carried out by Lahiri [74], and it is observed that the performance of the finite step algorithm is quite good in terms of MSEs and biases. In initial steps, the part of the sample is selected in such a way that the dependence structure is maintained in the subsample. The MSEs and biases of the finite step algorithm are very similar with the corresponding performance of the LSEs. Therefore, the finite step algorithm can be used quite efficiently in practice.

10.4 Efficient Algorithm Based on Dimension Reduction

Recently, Grover, Kundu and Mitra [46] proposed an efficient estimators of the unknown parameters of a 2-D chirp model (131) when the errors are i.i.d. random variables with mean zero and finite variance σ^2. It is a numerically efficient method and it uses the reduction of 2-D chirp model to 1-D chirp models. The main advantage of the proposed estimators is that although they can be obtained in a more computationally efficient manner, they have the same convergence rates as the LSEs.

Suppose we fix $n = n_0$, then (131) can be rewritten for $m = 1, \ldots, M$ as follows

$$\begin{aligned}y(m, n_0) &= A^0 \cos(\alpha^0 m + \beta^0 m^2 + \gamma^0 n_0 + \delta^0 n_0^2) + B^0 \sin(\alpha^0 m + \beta^0 m^2 + \gamma^0 n_0 + \delta^0 n_0^2) \\ &\qquad + e(m, n_0); \\ &= A^0(n_0) \cos(\alpha^0 m + \beta^0 m^2) + B^0(n_0) \sin(\alpha^0 m + \beta^0 m^2) + e(m, n_0). \qquad (144)\end{aligned}$$

Cleary, (144) represents a 1-D chirp model with $A^0(n_0)$, $B^0(n_0)$ as amplitudes, α^0 as the frequency parameter and β^0 as the frequency rate parameter. Here,

$$\begin{aligned} A^0(n_0) &= \quad A^0 \cos(\gamma^0 n_0 + \delta^0 n_0^2) + B^0 \sin(\gamma^0 n_0 + \delta^0 n_0^2), \\ B^0(n_0) &= -A^0 \sin(\gamma^0 n_0 + \delta^0 n_0^2) + B^0 \cos(\gamma^0 n_0 + \delta^0 n_0^2). \end{aligned}$$

Therefore, for any fixed $n_0 \in \{1, \ldots, N\}$, it represents a 1-D chirp model with the same frequency and frequency rate parameters, though different amplitudes. Thus each column of the 2-D data matrix can be thought of as a data vector coming from 1-D chirp model with the same frequency and frequency rate.

Therefore, each column of the data matrix can be used to estimate α^0 and β^0 and for this purpose least squares method may be used. Hence, the estimators of α^0 and β^0 can be obtained by minimizing the following function:

$$R_M(\alpha, \beta, n_0) = \boldsymbol{Y}_{n_0}^\top (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{Z}_M}(\alpha, \beta))\, \boldsymbol{Y}_{n_0}$$

for each n_0. Here, $\boldsymbol{Y}_{n_0} = \big[y[1, n_0], \ \ldots, \ y[M, n_0]\big]^\top$ is the $n_0 th$ column of the original data matrix, $\boldsymbol{P}_{\boldsymbol{Z}_M}(\alpha, \beta) = \boldsymbol{Z}_M(\alpha, \beta)(\boldsymbol{Z}_M(\alpha, \beta)^\top \boldsymbol{Z}_M(\alpha, \beta))^{-1} \boldsymbol{Z}_M(\alpha, \beta)^\top$ is the projection matrix on the column space of the matrix $\boldsymbol{Z}_M(\alpha, \beta)$ and

$$\boldsymbol{Z}_M(\alpha, \beta) = \begin{bmatrix} \cos(\alpha + \beta) & \sin(\alpha + \beta) \\ \vdots & \vdots \\ \cos(M\alpha + M^2\beta) & \sin(M\alpha + M^2\beta) \end{bmatrix}. \tag{145}$$

Therefore, estimates of α^0 and β^0 can be obtained from each column of the data vector and they can be combined. This process involves minimizing N 2-D functions corresponding to the N columns of the matrix. To avoid that, Grover, Kundu and Mitra [46] propose to minimize the following function instead:

$$R_{MN}^{(1)}(\alpha, \beta) = \sum_{n_0=1}^{N} R_M(\alpha, \beta, n_0) = \sum_{n_0=1}^{N} \boldsymbol{Y}_{n_0}^\top (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{Z}_M}(\alpha, \beta))\, \boldsymbol{Y}_{n_0} \tag{146}$$

with respect to α and β simultaneously and obtain $\widehat{\alpha}$ and $\widehat{\beta}$. It reduces the computational burden significantly. It may be mentioned that since the errors are assumed to be i.i.d., replacing these N functions by their sum is justifiable.

Similarly, we can obtain the estimates, $\widehat{\gamma}$ and $\widehat{\delta}$, of γ^0 and δ^0, by minimizing the following criterion function:

$$R^{(2)}_{MN}(\gamma,\delta) = \sum_{m_0=1}^{M} R_N(\gamma,\delta,m_0) = \sum_{m_0=1}^{M} \boldsymbol{Y}_{m_0}^{\top}(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{Z}_N}(\gamma,\delta))\,\boldsymbol{Y}_{m_0} \tag{147}$$

with respect to γ and δ simultaneously. The data vector $\boldsymbol{Y}_{m_0} = \big[y[m_0,1],\ \ldots,\ y[m_0,N]\big]^{\top}$, is the $m_0 th$ row of the data matrix, $m_0 = 1,\ldots,M$, $\boldsymbol{P}_{\boldsymbol{Z}_N}(\gamma,\delta)$ is the projection matrix on the column space of the matrix $\boldsymbol{Z}_N(\gamma,\delta)$ and the matrix $\boldsymbol{Z}_N(\gamma,\delta)$ can be obtained by replacing α, β, N by γ, δ and M, respectively in the matrix $\boldsymbol{Z}_M(\alpha,\beta)$, defined in (145).

Once the non-linear parameters have been estimated, the estimates of the linear parameters can be obtained as given below:

$$\begin{bmatrix}\widehat{A}\\ \widehat{B}\end{bmatrix} = [\,\boldsymbol{W}(\hat{\alpha},\hat{\beta},\hat{\gamma},\hat{\delta})^{\top}\,\boldsymbol{W}(\hat{\alpha},\hat{\beta},\hat{\gamma},\hat{\delta})]^{-1}\,\boldsymbol{W}(\hat{\alpha},\hat{\beta},\hat{\gamma},\hat{\delta})^{\top}\,\boldsymbol{Y}_{MN\times 1}.$$

Here, $\boldsymbol{Y}_{MN\times 1} = \big[y(1,1),\ \ldots,\ y(M,1),\ \ldots,\ y(1,N),\ \ldots,\ y(M,N)\big]^{\top}$ is the observed data vector, and

$$\boldsymbol{W}(\alpha,\beta,\gamma,\delta)_{MN\times 2} = \begin{bmatrix} \cos(\alpha+\beta+\gamma+\delta) & \sin(\alpha+\beta+\gamma+\delta)\\ \cos(2\alpha+4\beta+\gamma+\delta) & \sin(2\alpha+4\beta+\gamma+\delta)\\ \vdots & \vdots\\ \cos(M\alpha+M^2\beta+\gamma+\delta) & \sin(M\alpha+M^2\beta+\gamma+\delta)\\ \vdots & \vdots\\ \cos(\alpha+\beta+N\gamma+N^2\delta) & \sin(\alpha+\beta+N\gamma+N^2\delta)\\ \cos(2\alpha+4\beta+N\gamma+N^2\delta) & \sin(2\alpha+4\beta+N\gamma+N^2\delta)\\ \vdots & \vdots\\ \cos(M\alpha+M^2\beta+N\gamma+N^2\delta) & \sin(M\alpha+M^2\beta+N\gamma+N^2\delta) \end{bmatrix}. \tag{148}$$

The following assumptions are needed to establish the asymptotic properties of the estimators.

Assumption 10.2 *The error random variable $\{e(m,n)\}$ is a double array sequence of i.i.d. random variables with mean zero, variance σ^2 and finite fourth order moment.*

Assumption 10.3 *The non-linear parameters satisfy* $0 < \alpha^0, \beta^0, \gamma^0, \delta^0 < \pi$ *and there exists a* K*, such that* $0 < {A^0}^2 + {B^0}^2 < K$.

The results obtained on the consistency of the proposed estimators are presented in the following theorems:

Theorem 10.8 *Under assumptions 10.2 and 10.3,* $\widehat{\alpha}$ *and* $\widehat{\beta}$ *are strongly consistent estimators of* α^0 *and* β^0 *respectively, that is,*

$$\widehat{\alpha} \xrightarrow{a.s.} \alpha^0 \text{ as } M \to \infty.$$

$$\widehat{\beta} \xrightarrow{a.s.} \beta^0 \text{ as } M \to \infty.$$

Theorem 10.9 *Under assumptions 10.2 and 10.3,* $\widehat{\gamma}$ *and* $\widehat{\delta}$ *are strongly consistent estimators of* γ^0 *and* δ^0 *respectively, that is,*

$$\widehat{\gamma} \xrightarrow{a.s.} \gamma^0 \text{ as } N \to \infty.$$

$$\widehat{\delta} \xrightarrow{a.s.} \delta^0 \text{ as } N \to \infty.$$

The following theorems provide the asymptotic distributions of the proposed estimators:

Theorem 10.10 *If the assumptions, 10.2 and 10.3 are satisfied, then*

$$\boldsymbol{D}_1 \left[(\widehat{\alpha} - \alpha^0),\ (\widehat{\beta} - \beta^0)\right]^\top \xrightarrow{d} N_2(\boldsymbol{0}, 2\sigma^2\boldsymbol{\Sigma}) \text{ as } M \to \infty.$$

Here, $\boldsymbol{D}_1 = \operatorname{diag}(M^{\frac{3}{2}}N^{\frac{1}{2}}, M^{\frac{5}{2}}N^{\frac{1}{2}})$ *and*

$$\boldsymbol{\Sigma} = \frac{2}{{A^0}^2 + {B^0}^2}\begin{bmatrix} 96 & -90 \\ -90 & 90 \end{bmatrix} \tag{149}$$

Theorem 10.11 *If the assumptions, 10.2 and 10.3 are satisfied, then*

$$\boldsymbol{D}_2 \left[(\widehat{\gamma} - \gamma^0),\ (\widehat{\delta} - \delta^0)\right]^\top \xrightarrow{d} \boldsymbol{\mathcal{N}}_2(\boldsymbol{0}, 2\sigma^2\boldsymbol{\Sigma}) \text{ as } N \to \infty.$$

Here, $\boldsymbol{D}_2 = \operatorname{diag}(M^{\frac{-1}{2}}N^{\frac{-3}{2}}, M^{\frac{-1}{2}}N^{\frac{-5}{2}})$ *and* $\boldsymbol{\Sigma}$ *is as defined in (149).*

The asymptotic distributions of $(\widehat{\alpha}, \widehat{\beta})$ and $(\widehat{\gamma}, \widehat{\delta})$ are observed to be the same as those of the corresponding LSEs. Thus, we get the same efficiency as that of the LSEs without going through the process of actually computing the LSEs.

10.5 Two-Dimensional Polynomial Phase Signal Model

So far we have discussed about 1-D and 2-D chirp models in details. But 2-D polynomial phase signal model also has received significant amount of attention in the signal processing literature. The most general form of the 2-D polynomial phase signal model was introduced by Francos and Friedlander [35]. In presence of additive error the model can be described as follows:

$$\begin{aligned} y(m,n) &= A^0 \cos\left(\sum_{p=1}^{r}\sum_{j=0}^{p} \alpha^0(j,p-j) m^j n^{p-j}\right) + B^0 \sin\left(\sum_{p=1}^{r}\sum_{j=0}^{p} \alpha^0(j,p-j) m^j n^{p-j}\right) \\ &\quad + e(m,n); \quad m = 1,\ldots,M; \quad n = 1,\ldots,N, \qquad (150) \end{aligned}$$

here $e(m,n)$ is the additive error with mean 0, A^0 and B^0 are non zero real numbers, and for $j = 0,\ldots,p, = 1,\ldots,r$, $\alpha^0(j,p-j)$'s are distinct frequency rates of order $(j,p-j)$, respectively, and lie strictly between 0 and π. Here $\alpha^0(0,1)$ and $\alpha^0(1,0)$ are called frequencies. The explicit assumptions on the error $X(m,n)$ will be provided later.

Different specific forms of model (150) have been used quite extensively in the literature. For example Friedlander and Francos [38] used the following 2-D polynomial phase signal model

$$\begin{aligned} y(m,n) = \quad & A^0 \cos(\alpha^0 m + \beta^0 m^0 + \gamma^0 n + \delta^0 n^2 + \nu^0 mn) \\ & + B^0 \sin(\alpha^0 m + \beta^0 m^0 + \gamma^0 n + \delta^0 n^2 + \nu^0 mn) \\ & + e(m,n); \quad m = 1,\ldots,M; \quad n = 1,\ldots,N, \end{aligned}$$

to analyze finger print type data and Djurović et al. [27] used a specific 2-D cubic phase signal model due to its applications in modelling Synthetic Aperture Radar (SAR) data and in particular Interferometric SAR data. Further, 2-D polynomial phase signal model also has been used in modeling and analyzing magnetic resonance imaging (MRI), optical imaging and different texture imaging. For some of the other specific applications of this model, one may refer to Djukanović and Djurović [21], Ikram and Zhou [55], Wang and Zhou [120], Tichavsky and Handel [113], Amar, Leshem and van der Veen [2] and see the references cited therein.

Different estimation procedures have been suggested in the literature based on the assumption that the error components are i.i.d. random variables with mean zero and finite variance. For example, Farquharson et al. [32] provided a computationally efficient estimation procedures of the unknown parameters of a polynomial phase signals. Djurović and Stanković [26] proposed the quasi maximum likelihood estimators based on the normality assumption of the error random variables. Recently, Djurović et al. [25] considered an efficient estimation method of the polynomial phase signal parameters by using a cubic phase function. Some of the refinement of the parameter estimation of the polynomial phase signals can be obtained in O'Shea [97].

Interestingly, a significant amount of work has been done in developing different estimation procedures to compute parameter estimates of the polynomial phase signal, but not much work has been done in developing the properties of the estimators. In most of the cases, the mean squared errors or the variances of the estimates are compared with the corresponding Cramer-Rao lower bound, without establishing formally that the asymptotic variances of the maximum likelihood estimators attain the lower bound in case the errors are i.i.d. normal random variables. Recently, Lahiri and Kundu [75] established formally the asymptotic properties of the LSEs of parameters in model (150) under a fairly general assumptions on the error random variables. From the results of Lahiri and Kundu [75] it can be easily obtained that when the errors are i.i.d. normally distributed the asymptotic variance of the MLEs attain the Cramer-Rao lower bound. Lahiri and Kundu [75] established the consistency and the asymptotic normality properties of the LSEs of the parameters of model (150) under a more general error assumptions. Although, the LSEs are the most efficient estimators, finding the LSEs is a computationally challenging problem. Moreover, computationally efficient algorithm which provides estimators which have the same rate of convergence as the LSEs does not exist. More work is needed along that direction.

11 Conclusions

In this chapter we have provided a brief review of different 1-D and 2-D chirp models which have received a considerable amount of attention in recent years in the Statistical Signal Processing literature. These models have been used quite effectively in analyzing 1-D and 2-D non-stationary signals. Several numerical algorithms and sophisticated statistical tools are needed to develop efficient estimators and derive their properties. Since the models are non-linear in nature, most of the properties are asymptotic. Extensive simulations have been performed by different researchers to verify the effectiveness of the different estimators. We have provided several open problems for future research. We hope that this review chapter will generate enough interests among your researchers to come forward and solve some of these challenging problems.

Acknowledgements

Part of the work of the first author has been supported by a grant from the Science and Engineering Research Board, Department of Science and Technology, Government of India.

References

[1] Abatzoglou, T. (1986), "Fast maximum likelihood joint estimation of frequency and frequency rate", *IEEE Transactions on Aerospace and Electronic Systems*, vol. 22, 708 – 715.

[2] Amar, A., Leshem, A. and van der Veen, A. J. (2010), "A low complexity blind estimator of narrow band polynomial phase signals", *IEEE Transactions on Signal Processing*, vol. 58, 4674 – 4683.

[3] Bai, Z.D., Rao, C.R., Chow, M. and Kundu, D. (2003), "An efficient algorithm for estimating the parameters of superimposed exponential signals", *Journal of Statistical Planning and Inference*, vol. 110, 23 – 34.

[4] Barbarossa, S. and Petrone, V. (1997). Analysis of polynomial-phase signals by the integrated generalized ambiguity function. *IEEE Transactions on Signal Processing*, **45**, 316-327.

[5] Barbarossa, S., Scaglione, A. and Giannakis, G. (1998). Product high-order ambiguity function for multicomponent polynomial-phase signal modeling. *IEEE Transactions on Signal Processing*, **46**, 691-708.

[6] Barbieri, M. M. and Barone, P. (1992) "A two dimensional Prony's method for Spectral estimation", *IEEE Transactions on Signal Processing*, vol. 40, 2747 – 2756.

[7] Bates, D.M. and Watts, D.G. (1988), *Nonlinear regression analysis and its applications*, John Wiley & Sons, New York.

[8] Besson, O. and Castanié F. (1993), "On estimating the frequency of a sinusoid in autoregressive multiplicative noise", *Signal Processing*, vol. 30, 65–83.

[9] Besson, O. and Stoica, P. (1995), "Sinusoidal signals with random amplitude: Least- squares estimators and their statistical analysis", *IEEE Transactions on Signal Processing*, vol. 43, 2733 – 2744.

[10] Besson, O., Ghogho, M. and Swami, A. (1999), "Parameter estimation for random amplitude chirp signals", *IEEE Transactions on Signal Processing*, vol. 47, 3208 – 3219.

[11] Besson, O., Giannakis, G.B. and Gini, F. (1999), "Improved estimation of hyperbolic frequency modulated chirp signals", *IEEE Transactions on Signal Processing*, vol. 47, 1384 – 1388.

[12] Bresler, Y. and Macovski, A. (1986), "Exact maximum likelihood parameter estimation of superimposed exponential signals in noise", *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 34, 1081 – 1089.

[13] Cabrera, S.D. and Bose, N.K. (1993), "Prony's method for two-dimensional complex exponential modeling", Chapter 15, In S.G. Tzafestas (Ed.), *Applied COntrol Theory*, pp 401–411, New York, NY, Marcel and Dekker.

[14] Cao, F.,Wang, S., Wang, F. (2006), "Cross-Spectral Method Based on 2-D Cross Polynomial Transform for 2-D Chirp Signal Parameter Estimation", *ICSP2006 Proceedings*, DOI:10.1109/ICOSP.2006.344475.

[15] Casazza, P.G. and Fickus, M. (2006) "Fourier transforms of finite chirps", *EURASIP Journal on Applied Signal Processing*, vol. 2006, Article ID 70204, pp. 1-7.

[16] Chang, G.W. and Chen, C. (2010), "An accurate time domain procedure for harmonics and inter-harmonics detection", *IEEE Transactions of Power Del.*, vol. 25, 1787–1795.

[17] Charnes A., Cooper,W.W. and Ferguson, R (1955), "Optimal estimation of executive compensation by linear programing", *Management Science*, 1, 138–151.

[18] Christensen, M.G., Stoica, P., Jakobsson, A. and Jensen, H.S. (2008), "Multi-pitch estimation", *Signal Processing*, vol. 88, 972–983.

[19] Christensen, M.G., Jensen, S.H., Jakobsson, A. and Jensen, S.H. (2008), "Optimal filter designs for fundamental frequency estimation", *IEEE Signal Processing Letters*, vol.15, 745–748.

[20] Chun, J. and Bose, N.K. (1995), "Parameter estimation via signal selectivity of signal subspaces (PESS) and its applications", *Digital Signal Processing*, vol. 5, 58 – 76.

[21] Djukanović, S. and Djurović, I. (2012) "Aliasing detection and resolving in the estimation of polynomial-phase signal parameters", *Signal Processing*, vol. 92, 235 – 239.

[22] Dhar, S.S., Kundu, D. and Das, U. (2019), "On testing parameters of chirp signal model", *IEEE Transactions on Signal Processing*, vol. 67, 4291 – 4301.

[23] Dielman, T.E. (1984), "Least absolute value estimation in regression models: An annotated bibliography", *Communication in Statistics - Theory and Methods*, vol. 4, 513 – 541.

[24] Djurić, P.M. and Kay, S.M. (1990), "Parameter estimation of chirp signals", *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 38, 2118 – 2126.

[25] Djurović, I., Simeunović, M. and Wang, P. (2017), "Cubic phase function: A simple solution for polynomial phase signal analysis", *Signal Processing*, vol. 135, 48 – 66.

[26] Djurović, I. and Stanković L. J. (2014), "Quasi maximum likelihood estimators of polynomial phase signals", *IET Signal Processing*, vol. 13, 347 – 359.

[27] Djurović, I., Wang, P. and Ioana, C. (2010) "Parameter estimation of 2-D cubic phase signal function using genetic algorithm", *Signal Processing*, vol. 90, 2698 – 2707.

[28] Dodge, Y. (1987), " An introduction to L_1-norm bases statistical data analysis", *Computational Statistics and Data Analysis*, vol. 5, 239 - 253.

[29] Doweck, Y., Amar, A. and Cohen, I. (2015), "Joint model order selection and parameter estimation of chirps with harmonic components", *IEEE Transactions on Signal Processing*, vol. 63, 1765 – 1778.

[30] Dwyer, R.F. (1991), "Fourth-order spectra of Gaussian amplitude modulated sinusoids", *Journal of the Acoustics Society of America*, vol. 90, 918–926.

[31] Eddy, T.W. (1980), "Maximum likelihood detection and estimation for harmonic sets", *Journal of Acoustic Society of America*, vol. 68, 149–155.

[32] Farquharson, M., O'Shea, P. and Ledwich, G. (2005), "A computationally efficient technique for estimating the parameters phase signals from noisy observations", *IEEE Transactions on Signal Processing*, vol. 53, 3337 – 3342.

[33] Fisher, R.A. (1929), "Tests of signifcance in harmonic analysis", *Proceedings of the Royal Society of London, Series A*, vol. 125, 54 – 59.

[34] Fourier, D., Auger, F., Czarnecki, K. and Meignen, S. (2017), "Chirp rate and instantaneous frequency estimation: application to recursive vertical synchrosqueezing", *IEEE Signal Processing Letters*, 1-1, PP, 99.

[35] Francos, J.M. and Friedlander, B. (1998), "Two-dimensional polynomial phase signals: parameter estimation and bounds", *Multidimensional Systems and Signal Processing*, vol. 9, 173 – 205.

[36] Francos, J.M. and Friedlander, B. (1999), "Parameter estimation of 2-D random amplitude polynomial phase signals", *IEEE Transactions on Signal Processing*, vol. 47, 1795 – 1810.

[37] Friedlander, B. and Francos, J.M. (1995), "Estimation of amplitudes and phase parameters of multicomponent signals", *IEEE Transactions on Signal Processing*, vol. 43, pp. 917 – 926.

[38] Friedlander, B. and Francos, J.M. (1996), "An estimation algorithm for 2-D polynomial phase signals", *IEEE Transactions on Image Processing*, vol. 5, 1084 – 1087.

[39] Froberg, C.E. (1969), *Introduction to numerical analysis*, 2nd edition, Addison-Wesley, Pub. Co.

[40] Gabor, D. (1946), "Theory of communication. Part 1: The analysis of information" *Journal of the Institution of Electrical Engineers - Part III: Radio and Communication Engineering*, vol. 93, 429 – 441.

[41] Gini, F., Montanari, M. and Verrazzani, L. (2000), "Estimation of chirp signals in compound Gaussian clutter: a cyclostationary approach", *IEEE Transactons on Acoustics, Speech and Signal Processing*, vol. 48, 1029 – 1039.

[42] Grover, R. (2019), *Frequency and frequency rate estimation of some non-stationary signal processing models*, Ph.D. thesis at Indian Institute of Technology, Kanpur, India.

[43] Grover, R., Kundu, D. and Mitra, A. (2018a), "On approximate least squares estimators of parameters of one-dimensional chirp signal", *Statistics*, vol. 52, 1060 – 1085.

[44] Grover, R, Kundu, D. and Mitra, A. (2018b), "Asymtotic of approximate least squares estimators of parameters of two-dimensional chirp signal", *Journal of Multivariate Analysis*, (to appear).

[45] Grover, R, Kundu, D. and Mitra, A. (2018c), "A chirp like model to analyze periodic and nearly periodic signals", *arXiv*.

[46] Grover, R., Kundu, D. and Mitra, A. (2020), "An efficient methodology to estimate the parameters of a two-dimensional chirp signal model", *Multidimensional Systems and Signal Processing*, (to appear).

[47] Guo, J., Zou, H., Yang, X. and Liu, G. (2011), "Parameter estimation of multicomponent chirp signals via sparse representation", *IEEE Transactions on Aerospace and Electronic Systems*, vol. 47, 2261 – 2268.

[48] Hannan, E. J. (1971), "Non-linear time series regression", *Journal of Applied Probability*, vol. 8, 767 – 780.

[49] Hannan, E.J. (1974), "Time series analysis", *IEEE Transactions on Automatic Control*, vol. 19, 706 – 715.

[50] Hedley, M. and Rosenfeld, D. (1992). A new two-dimensional phase unwrapping algorithm for MRI images. *Magnetic Resonance in Medicine*, **24**, 177-181.

[51] Hilderband, F.B. (1956), *An introduction to numerical analysis*, McGraw Hill, New York.

[52] Hua, Y. (1992), "Estimating two-dimensional frequencies by matrix enhancement and matrix Pencil", *IEEE Transactions on Signal Processing*, vol. 40, 2267 – 2280.

[53] Ikram, M.Z., Abed-Meraim, K., Hua, Y. (1997), "Iterative parameter estimation of multiple chirp signals", *Electronics Letters*, vol. 33, 657 – 659.

[54] Ikram, M.Z., Abed-Meraim, K., Hua, Y. (1998), "Estimating the parameters of chirp signals: an iterative approach", *IEEE Transactions on Signal Processing*, vol. 46, 3436 –3441.

[55] Ikram, M.Z. and Zhou, G.T. (2001), "Estimation of multicomponent phase signals of mixed orders", *Signal Processing*, vol. 81, 2293 – 2308.

[56] Irizarry, R. A. (2000), "Asymptotic distribution of estimates for a time-varying parameter in a harmonic model with multiple fundamentals", *Statistica Sinica*, vol. 10, 1041 – 1067.

[57] Jain, S.K. and Singh, S.N. (2012), "Exact model order ESPRIT techniquefor harmonics and inter-harmonics estimation", *IEEE Transactions on Instrumental Measurement*, vol. 61, 1915–1923.

[58] Jennrich, R.I. (1969), "Asymptotic properties of the nonlinear least squares estimators", *Annals of Mathematical Statistics*, vol. 40, 633 – 643.

[59] Jensen, T.L., Nielsen, J.K., Jensen, J.R., Christensen, M.G. and Jensen, S.H. (2017), "A fast algorithm for maximum likelihood estimation of harmonic chirp parameters", *IEEE Transactions on Signal Processing*, vol. 65, 5137 – 5152.

[60] Kay, S. M.(1988), *Modern Spectral Estimation, Theory and Application*, Englewood Cliffs, NJ: Prentice Hall.

[61] Kennedy, W.J. Jr. and Gentle, J.E. (1980), *Statistical Computing*, Marcel Dekker, Inc., New York.

[62] Kopsinis, Y., Aboutanios, E., Waters, D.A. and McLaughlin, S. (2010), "Time-frequency and advanced frequency estimation techniques for the investigation of bat echolocation calls", *Journal of Acoustics Society of America*, vol. 127, 1124–1134.

[63] Kundu, D. (1991), "Asymptotic properties of the complex valued non-linear regression model", *Communications in Statistics - Theory and Methods*, vol 20, 3793 –3803.

[64] Kundu, D. (1992), "Estimating the number of signals using information theoretic criterion", *Journal of Statistical Computation and Simulation*, vol. 44, 117–131.

[65] Kundu, D. (1993), "Estimating the parameters of complex valued exponential signals", *Technometrics*, vol. 35, 215 – 218.

[66] Kundu, D. (1997), "Asymptotic properties of the least squares estimators of sinusoidal signals", *Statistics*, vol. 30, 221–238.

[67] Kundu, D. (1997), "Estimating the number of sinusoids in additive white noise", *Signal Processing*, vol. 56, 103–110.

[68] Kundu, D., Bai, Z.D., Nandi, S., Bai, L. (2011), "Super efficient frequency estimation", *Journal of Statistical Planning and Inference*, vol. 141, 2576 – 2588.

[69] Kundu, D. and Kundu, R. (1995), "Consistent estimates of superimposed exponential signals when some observations are missing", *Journal of Statistical Planning and Inference*, vol. 44, 205 – 218.

[70] Kundu, D. and Mitra, A. (1995), "Consistent method of estimating the superimposed exponential signals", *Scandinavian Journal of Statistics*, vol. 22, 73–82.

[71] Kundu, D. and Nandi, S. (2003), "Determination of discrete spectrum in a random field", *Statistica Neerlandica*, vol. 57, 258 – 283.

[72] Kundu, D. and Nandi, S. (2008), "Parameter estimation of chirp signals in presence of stationary Christensen, noise", *Statistica Sinica*, vol. 18, 187 – 201.

[73] Kundu, D. and Nandi, S. (2012), *Statistical Signal Processing: Frequency Estimation*, Springer, New Delhi.

[74] Lahiri, A. (2011), *Estimators of parameters of chirp signals and their properties*, Ph.D. thesis, Department of Mathematics and Statistics, Indian Institute of Technology, Kanpur, India.

[75] Lahiri, A. and Kundu, D. (2017), "On parameter estimation of two-dimensional polynomial phase signal model", *Statistica Sinica*, vol. 27, 1779 – 1792.

[76] Lahiri, A., Kundu, D. and Mitra, A. (2012), "Efficient algorithm for estimating the parameters of chirp signal", *Journal of Multivariate Analysis*, vol. 108, 15 – 27.

[77] Lahiri, A., Kundu, D. and Mitra, A. (2013), "Efficient algorithm for estimating the parameters of two dimensional chirp signal", *Sankhya, Ser. B*, vol. 75, 65 – 89.

[78] Lahiri, A., Kundu, D. and Mitra, A. (2014), "On least absolute deviation estimator of one dimensional chirp model", *Statistics*, vol. 48, 405 – 420.

[79] Lahiri, A., Kundu, D. and Mitra, A. (2015), "Estimating the parameters of multiple chirp signals", *Journal of Multivariate Analysis*, vol. 139, 189 – 205.

[80] Li, H., Stoica, P. and Li, J. (2000), "Computationally efficient parameter estimation for harmonic sinusoidal signals", *Signal Processing*, vol. 80, 1937–1944.

[81] Lin, C-C. and Djurić, P.M. (2000), "Estimation of chirp signals by MCMC", *ICASSP-1998*, vol. 1, 265 – 268.

[82] Liu, X. and Yu, H. (2013), "Time-domain joint parameter estimation of chirp signal based on SVR", *Mathematical Problems in Engineering*, Article ID 952743, 1 – 9.

[83] Lu, Y., Demirli, R., Cardoso, G. and Saniie, J. (2006), "A successive parameter estimation algorithm for chirplet signal decomposition", *IEEE Transactions on Ultrasonic, Ferroelectrics and Frequency Control*, vol. 53,

[84] Malliavan, P. (1994),"Sur la norme d'une matrice circulate Gaussienne Serie I", *C.R. Acad. Sc. Paris t*, vol. 319, 45 – 49.

[85] Malliavan, P. (1994),"Estimation d'un signal Lorentzien Serie I", *C.R. Acad. Sc. Paris t*, vol. 319, 991–997.

[86] Mangulis, V. (1965), *Handbook of series for scientists and engineers*, New York, Academic Press.

[87] Mazumder, S. (2017), "Single-step and multiple-step forecasting in one-dimensional chirp signal using MCMC-based Bayesian analysis", *Communications in Statistics - Simulation and Computation*, vol. 46, 2529 – 2547.

[88] Mboup, M. and Adali, T. (2012), "A generalization of the Fourier transform andits application to spectral analysis of chirp-like signals", *Applied and Computational Harmonic Analysis*, vol. 32, Issue 2, pp. 305-312.

[89] Montgomery, H.L. (1990), *Ten Lectures on the Interface Between Analytic Number Theory and Harmonic Analysis*, American Mathematical Society, 196.

[90] Nandi, S. and Kundu, D. (2003), "Estimating the fundamental frequency of a periodic function", *Statistical Methods and Applications*, vol. 12, 341 – 360.

[91] Nandi, S. and Kundu, D. (2004), "Asymptotic properties of the least squares estimators of the parameters of the chirp signals", *Annals of the Institute of Statistical Mathematics*, vol. 56, 529 – 544.

[92] Nandi, S. and Kundu, D. (2006), "Analyzing non-stationary signals using a cluster type model", *Journal of Statistical Planning and Inference*, vol. 136, 3871 – 3903.

[93] Nandi, S. and Kundu, D. (2006), "A fast and efficient algorithm for estimating the parameters of sinusoidal signals", *Sankhya*, vol. 68, 283 – 306.

[94] Nandi, S. and Kundu, D. (2020), "Random amplitude chirp model", *Signal Processing*, (to appear).

[95] Nandi, S. and Kundu, D. (2019), "Estimating the fundamental frequency using modified Newton-Raphson algorithm", *Statistics*, vol. 53, 440 - 458.

[96] Nandi, S. and Kundu, D. (2019), "Estimation of Parameters of Multiple Chirp Signal in presence of Heavy Tailed Errors", submitted for publication.

[97] O'Shea, P. (2010), "On refining polynomial phase signal parameter estimates", *IEEE Transactions on Aerospace, Electronic Systems*, vol. 4, 978 – 987.

[98] Pelag, S. and Porat, B. (1991), "Estimation and classification of polynomial phase signals", *IEEE Transactions on Information Theory*, vol. 37, 422 – 430.

[99] Pham, D.S. and Zoubir, A.M. (2007), "Analysis of multicomponent polynomial phase signals", *IEEE Transations on Signal Processing*, vol. 55, 56 – 65.

[100] Pincus, M. (1968), "A closed form solution of certain programming problems", *Operation Research*, vol. 16, 690 – 694.

[101] Porchia, A., Barbarossa, S., Scaglione, A. and Giannakis, G.B. (1996), "Auto focusing techniques for SAR Imaging using the multilag high-order ambiguity function", *Proc. ICASSP*, Atlanta, GA, May 7–11, vol. IV, pp. 2086 – 2089.

[102] Prasad, A., Kundu, D. and Mitra, A. (2008), "Sequential estimation of the sum of sinusoidal model parameters", *Journal of Statistical Planning and Inference*, vol. 138, 1297 – 1313.

[103] Quinn, B.G. (1989), "Estimating the number of terms in a sinusoidal regression", *Journal of Time Series Analysis*, vol. 10, 71 – 75.

[104] Quinn, B.G. and Hannan, E.J. (2001), *The estimation and tracking of frequency*, Cambridge Universty Press, New York.

[105] Quinn, B. G. and Thomson, P. J. (1991), "Estimating the frequency of a periodic function", *Biometrika*, vol. 78, 65 – 74.

[106] Rao, C.R. (1988), "Some results in signal detection", In *Decision Theory and Related Topics*, Eds. Gupta, S.S. and Berger, J.O., pp 319 – 332, Springer, New York.

[107] Richards, F.S.G. (1961), "A method of maximum likelihood estimation", *Journal of the Royal Statistical Society, Ser. B*, vol. 23, 469 - 475.

[108] Rihaczek, A.W. (1969), *Principles of high resolution radar*, McGraw-Hill, New York.

[109] Saha, S. and Kay, S.M. (2002), "Maximum likelihood parameter estimation of superimposed chirps using Monte Carlo importance sampling", *IEEE Transactions on Signal Processing*, vol. 50, 224 – 230.

[110] Sakai, H. (1990), "An application of a BIC-type method to harmonic analysis and a new criterion for order determination of an error process", *IEEE Transactions of Acoustics Speech and Signal Processing*, vol. 38, 999 – 1004.

[111] Seber, G. A. F. and Wild, C.J. (1989), *Nonlinear Regression*, John Wiley and Sons, New York.

[112] Stoica, P. and Moses, R. (2005), *Spectral Analysis of Signals*, Prentice Hall, Upper Saddle River, New Jersey.

[113] Ticahvsky, P. and Handel, P. (1999), "Multicomponent polynomial phase signal analysis using a tracking algorithm", *IEEE Transactions on Signal Processing*, vol. 47, 1390 – 1395.

[114] Vespe, M., Jones, G. and Baker, C. (2009), "Lessons for radar: Waveform diversity in echolocating mammals", *IEEE Transactions on Signal Processing Magazine*, vol.26, 65–75

[115] Vinogradov, I. M. (1954), "The method of trigonometrical sums in the theory of numbers", *Interscience*, Translated from Russian, Revised and annoted by K.F. Roth and A. Devenport. Reprint of the 1954 translation. Dover publications, Inc., Mineola, New York, U.S.A., 2004.

[116] Volcker, B. and Ottersten, B. (2001), "Chirp parameter estimation from a sample covariance matrix", *IEEE Transactions on Signal Processing*, vol. 49, 603 – 612.

[117] Walker, A.M. (1971), "On the estimation of a harmonic component in a time series with stationary independent residuals", *Biometrika*, vol. 58, 21 - 36.

[118] Wang, X. (1993), "An AIC type estimator for the number of cosinusoids", *Journal of Time Series Analysis*, vol. 14, 433 – 440.

[119] Wang, P. and Yang, J. (2006), "Multicomponent chirp signals analysis using product cubic phase function", *Digital Signal Processing*, vol. 16, 654 – 669.

[120] Wang, Y. and Zhou, Y.G.T. (1998), "On the use of high-order ambiguity function for multi-component polynomial phase signals", *Signal Processing*, vol. 5, 283 – 296.

[121] Wehner, D.R. (1995), *High-Resolution Radar*, 2nd edition, Norwell, MA: Artech House, USA.

[122] Wu, C.F.J. (1981), "Asymptotic theory of the nonlinear least squares estimation", *Annals of Statistics*, vol. 9, 501-513.

[123] Wang, J. Z., Su, S. Y., Chen, Z. . (2015), "Parameter estimation of chirp signal under low SNR", *Science China: Information Sciences*, vol. 58, 020307:1 – 020307:13.

[124] Wu, Y., So, H.C. and Liu, H. (2008). Subspace-Based Algorithm for Parameter Estimation of Polynomial Phase Signals. *IEEE Transactions on Signal Processing*, **56**, 4977 - 4983.

[125] Xinghao, Z., Ran, T., Siyong, Z. (2003), "A novel sequential estimation algorithm for chirp signal parameters", *IEEE Conference in Neural Networks and & Signal Processing*, Nanjing, China, December 14-17, 2003, pp 628 – 631.

[126] Yang, P., Liu, Z. and Jiang, W-L. (2015), "parameter estimation of multicomponent chirp signals based on discrete chirp Fourier transform and population Monte Carlo", *Signal, Image and Video Processing*, vol. 9, 1137 – 1149.

[127] Yaron, D., Alon, A. and Israel, C. (2015), "Joint model order selection and parameter estimation of chirps with harmonic components", *IEEE Transactions on Signal Processing*, vol. 63, 1765 – 1778.

[128] Zhang, H., Liu, H., Shen, S., Zhang, Y., Wang, X. (2013), "Parameter estimation of chirp signals based on fractional Fourier transform", *The Journal of China Universities of Posts and Telecommunications*, vol. 20, (Suppl. 2), 95 – 100.

[129] Zhang, H. and Liu, Q. (2006). Estimation of instantaneous frequency rate for multicomponent polynomial phase signals. *ICSP2006 Proceedings*, 498- 502, DOI: 10.1109/ICOSP.2006.344448.

[130] Zhang, K., Wang, S., Cao, F. (2008). Product Cubic Phase Function Algorithm for Estimating the Instantaneous Frequency Rate of Multicomponent Two-dimensional Chirp Signals. *2008 Congress on Image and Signal Processing*, DOI: 10.1109/CISP.2008.352.