Sequential Estimation of the Sum of Sinusoidal Model Parameters

Anurag Prasad† & Debasis Kundu† & Amit Mitra†

Abstract

Estimating the parameters of the sum of a sinusoidal model in presence of additive noise is a classical problem. It is well known to be a difficult problem when the two adjacent frequencies are not well separated or when the number of components is very large. In this paper we propose a simple sequential procedure to estimate the unknown frequencies and amplitudes of the sinusoidal signals. It is observed that if there are p components in the signal then at the k-th ($k \leq p$) stage our procedure produces strongly consistent estimators of the k dominant sinusoids. For $k > p$, the amplitude estimators converge to zero almost surely. Asymptotic distribution of the proposed estimators is also established and it is observed that it coincides with the asymptotic distribution of the least squares estimators. Numerical simulations are performed to observe the performance of the proposed estimators for different sample sizes and for different models. One ECG data and one synthesized data are analyzed for illustrative purpose.

Keywords: Sinusoidal signals; least squares estimators; asymptotic distribution; over and under determined models, strongly consistent estimators.

† Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Kanpur, Pin 208016, INDIA.

Corresponding Author: Debasis Kundu, Phone no. 91-512-2597141, FAX no. 91-512-2597500, e-mail: kundu@iitk.ac.in.

1 Introduction

The problem of estimating the parameters of sinusoidal signals is a classical problem. The sum of a sinusoidal model has been used quite effectively in different signal processing applications, and time series data analysis. Starting with the work of Fisher [3] this problem has received a considerable attention because of its widespread applicability. Brillinger [1] discussed some of the very important real life applications from different areas and provided solutions using the sum of a sinusoidal model. See the expository article of Kay and Marple [7] from the Signal processors point of view. Stoica [16] provided a list of references of that particular problem up to that time and see Kundu [9] for some recent references.

The basic problem can be formulated as follows;

$$y(n) = \sum_{j=1}^{p} \left(A_j^0 \cos(\omega_j^0 n) + B_j^0 \sin(\omega_j^0 n) \right) + X(n); \qquad n = 1, \ldots, N. \tag{1}$$

Here A_j^0s and B_j^0s are unknown amplitudes and none of them is identically equal to zero. The ω_j^0s are unknown frequencies lying strictly between 0 and π and they are distinct. The error random variables $X(n)$s are from a stationary linear process with mean zero and finite variance. The explicit assumptions of $X(n)$s will be defined later. The problem is to estimate the unknown parameters A_j^0s, B_j^0s and ω_j^0s, given a sample of size N.

The problem is well known to be numerically difficult. It becomes particularly more difficult if $p \geq 2$ and the separation of the two frequencies is small, see Kay [6]. Several methods are available in the literature for estimating the parameters of the sinusoidal signals. Of course the most efficient estimators are the least squares estimators. The rates of convergence of the least squares estimators are $O_p(N^{-\frac{3}{2}})$ and $O_p(N^{-\frac{1}{2}})$ respectively for the frequencies and amplitudes, see Hannan [5], Walker [18] or Kundu [8]. But it is well known that finding the least squares estimators is not a trivial task, since there are several local minima of the least squares surface. The readers are referred to the article of Rice and Rosenblatt [14] for a nice discussion on this issue. It is observed that if two frequencies are very close to each other or if the number of components is very large then finding the initial guesses itself is very difficult and therefore starting any iterative process to find the least squares estimators is not a trivial task. One of the standard methods to find the initial guesses of the frequencies is to find the maxima at the Fourier frequencies of the periodogram function $I(\omega)$, where

$$I(\omega) = \left| \frac{1}{n} \sum_{t=1}^{n} y(t) e^{-i\omega t} \right|^2. \tag{2}$$

Asymptotically the periodogram function has local maxima at the true frequencies. But unfortunately, if the two frequencies are very close to each other then this method may not work properly. Let us consider the following synthesized signal for $n = 1, \ldots, 75$;

$$y(n) = 3.0\cos(0.20\pi n) + 3.0\sin(0.20\pi n) + 0.25\cos(0.19\pi n) + 0.25\sin(0.19\pi n) + X(n). \tag{3}$$

Here $X(n)$s are independent identically distributed (*i.i.d.*) normal random variables with mean 0 and variance 0.5. The periodogram function is plotted in Figure 1. In this case clearly the two frequencies are not resolvable. Therefore, it is not clear how to choose the initial estimates in this case to start any iterative process for finding the least squares estimators.

Several other techniques are available which attempt to find computationally efficient estimators and are non-iterative in nature. Therefore, they do not require any initial guess. See for example Pisarenko [11], Chan, Lavoie and Plant [2], Tufts and Kumaresan [17] etc. But unfortunately the frequency estimators produced by the methods proposed by Pisarenko [11] and Chan, Lavoi and Plant [2] are of the order $O_p(N^{-\frac{1}{2}})$ not $O_p(N^{-\frac{3}{2}})$ and the frequency estimators produced by the method of Tufts and Kumaresan [17] may not be even consistent.

Another practical problem occurs while using the least squares estimators when p is very large. It was observed recently (see Nandi and Kundu [13]) that for some of the speech signals the value of p can be 7 or 8 and it is observed in this paper that for some of the ECG signals

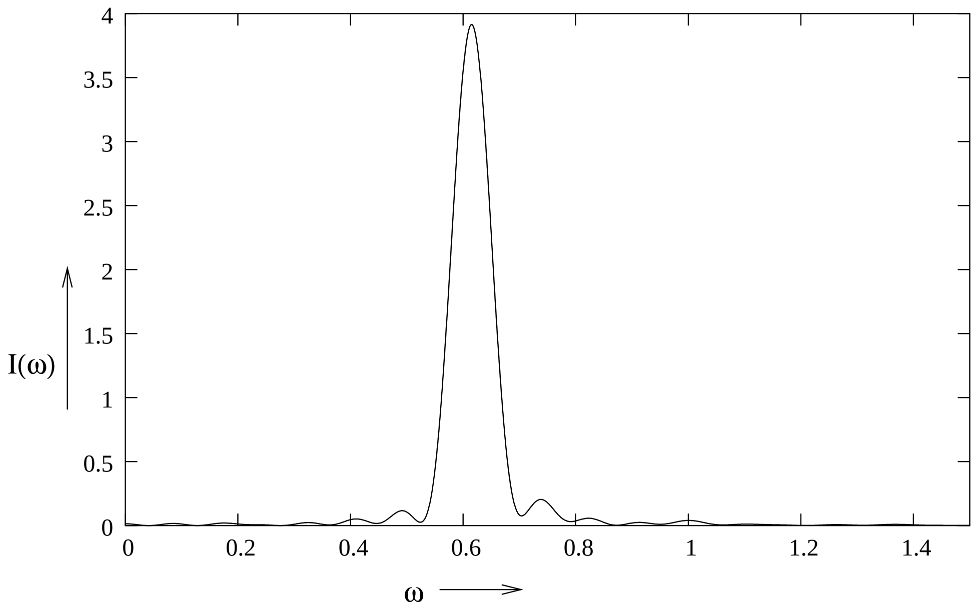

Figure 1: Periodogram plot of the synthesized signal.

the value of p can be even between 80 to 90. Therefore, in a high dimensional optimization problem the choice of initial guess can be very crucial and because of the presence of several local minima often the iterative process will converge to a local optimum point rather than the global optimum point.

The aim of this paper is twofold. First of all if p is known, then we propose a step-by-step sequential procedure to estimate the unknown parameters. It is observed that the p-dimensional optimization problem can be reduced to p one-dimensional optimization problems. Therefore, if p is large then the proposed method can be very useful. Moreover, it is observed that the estimators obtained by the proposed method have the same rate of convergence as the least squares estimators.

The second aim of this paper is to study the properties of the estimators if p is not known. If p is not known and we want to fit a lower order model, it is observed that the proposed estimators are consistent estimators of the dominant components with the same convergence rate as the least squares estimators. If we fit a higher order model, then it is observed that the estimators obtained up to p-th step are consistent estimators of the unknown parameters with the same convergence rate as the least squares estimators and the amplitude estimates after the p-th step converge to zero almost surely. We perform some numerical simulations to study the behavior of the proposed estimators. One synthesized data and one ECG data have been analyzed for illustrative purpose.

It should be mentioned that, although estimation of p is not the aim of this paper but it is well known to be a difficult problem. Extensive work has been done in this area, see for example the article by Kundu and Nandi [10] and the references cited there. It is observed that most of the methods work well if the noise variance is low but the performances are not satisfactory when the noise variance is high. In this paper, we have seen the performances of BIC, and it is observed that if strong autoregressive peaks are present then BIC can detect the number of components correctly if the error variance is low, but if the error variance is high large sample size is needed for correct detection of the number of components.

The rest of the paper is organized as follows. In section 2, we provide the necessary assumptions and also the methodology. Consistency of the proposed estimators are obtained in section 3. Asymptotic distributions or the convergence rates are provided in section 4. Numerical examples are provided in section 5. Data analysis results are presented in section 6 and finally we conclude the paper in section 7.

2 Model Assumptions and Methodology

It is assumed that we observe the data from the model in (1). We make the following assumptions. The additive error $\{X(n)\}$ is from a stationary linear process with mean zero and finite variance. It satisfies the following Assumption 1. From now on we denote the set of positive integers as $\mathcal{Z}$.

ASSUMPTION 1: $\{X(n); n \in \mathcal{Z}\}$ can be represented as

$$X(n) = \sum_{j=-\infty}^{\infty} a(j)e(n-j), \tag{4}$$

where $\{e(n)\}$ is a sequence of independent identically distributed (*i.i.d.*) random variables with mean zero and finite variance σ^2. The real valued sequence $\{a(n)\}$ satisfies

$$\sum_{n=-\infty}^{\infty} |a(n)| < \infty. \tag{5}$$

ASSUMPTION 2: The frequencies ω_k^0 s are distinct and $\omega_k^0 \in (0, \pi)$ for $k = 1, \ldots, p$.

ASSUMPTION 3: The amplitudes satisfy the following restrictions;

$$\infty > S^2 \geq A_1^{0^2} + B_1^{0^2} > \ldots > A_p^{0^2} + B_p^{0^2}. \tag{6}$$

METHODOLOGY: We propose the following simple procedure to estimate the unknown parameters. The method can be applied even when p is unknown. Consider the following $N \times 2$ matrix;

$$X(\omega) = \begin{bmatrix} \cos(\omega) & \sin(\omega) \\ \vdots & \vdots \\ \cos(N\omega) & \sin(N\omega) \end{bmatrix}, \tag{7}$$

and use the following notation; $\alpha = (A, B)^T$, $Y = (y(1), \ldots, y(N))^T$. First minimize

$$Q_1(A, B, \omega) = [Y - X(\omega)\alpha]^T [Y - X(\omega)\alpha], \tag{8}$$

with respect to (*w.r.t.*) A, B and ω. Therefore, by using the separable regression technique of Richards [15], it can be seen that for fixed ω,

$$\widehat{\alpha}(\omega) = \left[X^T(\omega)X(\omega)\right]^{-1} X^T(\omega)Y \tag{9}$$

minimizes $Q_1(A, B, \omega)$. Replacing α by $\widehat{\alpha}(\omega)$ in (8), we obtain

$$R_1(\omega) = Q_1(\widehat{A}(\omega), \widehat{B}(\omega), \omega) = Y^T(I - P_{X(\omega)})Y, \tag{10}$$

where

$$P_{X(\omega)} = X(\omega)\left[X^T(\omega)X(\omega)\right]^{-1}X^T(\omega)$$

is the projection matrix of the column space of the matrix $X(\omega)$. Therefore, if $\widehat{\omega}$ minimizes (10), then $(\widehat{A}(\widehat{\omega}), \widehat{B}(\widehat{\omega}), \widehat{\omega})$ minimize (8). We will denote these estimators as $(\widehat{A}_1, \widehat{B}_1, \widehat{\omega}_1)$. Now we consider the following sequence

$$Y^{(1)} = Y - X(\widehat{\omega}_1)\widehat{\alpha}_1, \tag{11}$$

where $\widehat{\alpha}_1 = (\widehat{A}_1, \widehat{B}_1)^T$. Now we replace Y by $Y^{(1)}$ and define

$$Q_2(A, B, \omega) = \left[Y^{(1)} - X(\omega)\alpha\right]^T \left[Y^{(1)} - X(\omega)\alpha\right]. \tag{12}$$

We minimize $Q_2(A, B, \omega)$ *w.r.t.* A, B and ω as before and denote the estimators obtained at the 2-nd step by $(\widehat{A}_2, \widehat{B}_2, \widehat{\omega}_2)$.

If p is the number of sinusoids and it is known, we continue the process up to p steps. If p is not known then we fit sequentially an order q model where q may not be equal to p. In the next section we provide the properties of the proposed estimators in both cases when p is known/ unknown.

3 Consistency of the Proposed Estimators

In this section we provide the consistency results for the proposed estimators when the number of components is unknown. We consider two cases separately, when the number of components of the fitted model (q) is less than the actual number of components (p) and when it is more. We need the following lemma for further development.

Lemma 1: Let $\{X(n); n \in \mathcal{Z}\}$ be a sequence of stationary random variables satisfying Assumption 1, then as $N \longrightarrow \infty$

$$\sup_{\alpha}\left|\frac{1}{N}\sum_{n=1}^{N} X(n)e^{in\alpha}\right| \longrightarrow 0 \quad a.s..$$

Proof of Lemma 1: See Kundu [8].

Lemma 2: Consider the set $S_c = \{\theta; \theta \in \Theta, \quad \text{and} \quad |\theta - \theta_1^0| \geq c\}$; where $\theta = (A, B, \omega)$ and $\theta_1^0 = (A_1^0, B_1^0, \omega_1^0)$, $\Theta = [-S, S] \times [-S, S] \times [0, \pi]$. If for any $c \geq 0$,

$$\liminf \inf_{\theta \in S_c} \frac{1}{N} \left\{ Q_1(\theta) - Q_1(\theta_1^0) \right\} > 0, \quad a.s. \tag{13}$$

then $\widehat{\theta}_1$ which minimizes $Q_1(\theta)$, is a strongly consistent estimator of θ_1^0.

Proof of Lemma 2: Suppose $\widehat{\theta_1}^{(N)}$ is not consistent estimator of θ_1^0, this implies, if

$$\Omega_0 = \{\omega : \widehat{\theta_1}^{(N)}(\omega) \to \theta_1^0\},$$

then $P(\Omega_0) < 1$. Since $P(\Omega_0') > 0$, there exists a sequence $\{N_k\}_{k \geq 1}$, a constant $c > 0$ and a set $\Omega_1 \subset \Omega_0'$, such that $P(\Omega_1) > 0$ and

$$\widehat{\theta_1}^{(N_k)}(\omega) \in S_c, \tag{14}$$

for all $k = 1, 2, \ldots$ and for all $\omega \in \Omega_1$. Since $\widehat{\theta_1}^{(N_k)}$ is the LSE of θ_1^0, we have for all k and for all $\omega \in \Omega_1$

$$\frac{1}{N_k} \left\{ Q_1^{(N_k)}(\widehat{\theta}_1^{(N_k)}(\omega)) - Q_1^{(N_k)}(\theta_1^0) \right\} < 0.$$

This implies for all $\omega \in \Omega_1$,

$$\liminf_k \frac{1}{N_k} \left\{ Q_1^{(N_k)}(\widehat{\theta}_1^{(N_k)}(\omega)) - Q_1^{(N_k)}(\theta_1^0) \right\} \leq 0.$$

Note that for all $\omega \in \Omega_1$

$$\liminf \inf_{\theta \in S_c} \frac{1}{N} \left\{ Q_1(\theta) - Q_1(\theta_1^0) \right\} \leq \liminf_k \frac{1}{N_k} \left\{ Q_1^{(N_k)}(\widehat{\theta}_1^{(N_k)}(\omega)) - Q_1^{(N_k)}(\theta_1^0) \right\} \leq 0,$$

because of (14). It is a contradiction of (13). ∎

Theorem 1: If the Assumptions 1-3 are satisfied, then $\widehat{\theta}_1$ is a strongly consistent estimator of θ_1^0.

Proof of Theorem 1: Consider the following expression:

$$\frac{1}{N}\left[Q_1(\theta) - Q_1(\theta_1^0)\right] = f(\theta) + g(\theta),$$

where

$$\begin{aligned} f(\theta) &= \frac{1}{N}\sum_{n=1}^{N}\left[A_1^0\cos(\omega_1^0 n) + B_1^0\sin(\omega_1^0 n) - A\cos(\omega n) - B\sin(\omega n)\right]^2 \\ &+\frac{2}{N}\sum_{n=1}^{N}\left[A_1^0\cos(\omega_1^0 n) + B_1^0\sin(\omega_1^0 n) - A\cos(\omega n) - B\sin(\omega n)\right] \\ &\times\left[\sum_{k=2}^{p}\left\{A_k^0\cos(\omega_k^0 n) + B_k^0\sin(\omega_k^0 n)\right\}\right] \end{aligned}$$

and

$$g(\theta) = \frac{2}{N}\sum_{n=1}^{N} X(n)\left[A_1^0\cos(\omega_1^0 n) + B_1^0\sin(\omega_1^0 n) - A\cos(\omega n) - B\sin(\omega n)\right].$$

Now using Lemma 1, it immediately follows that

$$\sup_{\theta\in S_c}|g(\theta)| \longrightarrow 0 \quad a.s.$$

Using lengthy but straightforward calculations and splitting the set S_c similar to Kundu [8], it can be easily shown that

$$\liminf \inf_{\theta\in S_c} f(\theta) > 0 \quad a.s.,$$

therefore,

$$\liminf \inf_{\theta\in S_c}\frac{1}{N}\left[Q_1(\theta) - Q_1(\theta_1^0)\right] > 0, \quad a.s.$$

This proves the result.

Now we want to prove that at the second step also the proposed estimators are consistent. We need the following lemma for that.

Lemma 3: If the Assumptions 1-3 are satisfied, then

$$N(\widehat{\omega}_1 - \omega_1^0) \longrightarrow 0 \quad a.s.$$

Proof of Lemma 3: The proof is provided in Appendix A.

Now we can state the result for the consistency of the estimates at the second step.

Theorem 2: If the Assumptions 1-3 are satisfied and $p \geq 2$, then $\widehat{\theta}_2$ obtained by minimizing $Q_2(A, B, \omega)$, as defined in (12), is a strongly consistent estimator of θ_2^0.

Proof of Theorem 2: Using Theorem 1 and Lemma 3, we obtain

$$\begin{aligned} \widehat{A}_1 &= A_1^0 + o(1) \quad a.s. \\ \widehat{B}_1 &= B_1^0 + o(1) \quad a.s. \\ \widehat{\omega}_1 &= \omega_1^0 + o(N) \quad a.s. \end{aligned}$$

Here a random variable $U = o(1)$ means $U \longrightarrow 0$ $a.s.$ and $U = o(N)$ means $NU \longrightarrow 0$ $a.s.$. Therefore for any fixed n as $N \to \infty$,

$$\widehat{A}_1 \cos(\widehat{\omega}_1 n) + \widehat{B}_1 \sin(\widehat{\omega}_1 n) = A_1^0 \cos(\omega_1^0 n) + B_1^0 \sin(\omega_1^0 n) + o(1) \quad a.s. \tag{15}$$

Now the result follows using (15) and similar technique as in Theorem 1.

The result can be extended up to the k-th step for $1 \leq k \leq p$. We can formally state the result as follows.

Theorem 3: If the Assumptions 1-3 are satisfied for $k \leq p$, then $\widehat{\theta}_k$, the estimator obtained by minimizing $Q_k(A, B, \omega)$, where $Q_k(A, B, \omega)$ is defined analogously to $Q_2(A, B, \omega)$ for the k-th step, is a consistent estimator of θ_k^0.

It will be interesting to investigate the properties of the estimators if the sequential process is continued even after p-th step. For this we need the following lemma.

Lemma 4: If $\{X(n)\}$ is same as defined in Assumption 1, and $\widehat{A}$, $\widehat{B}$ and $\widehat{\omega}$ are obtained by minimizing

$$\frac{1}{N} \sum_{n=1}^{N} \left(X(n) - A\cos(\omega n) - B\sin(\omega n) \right)^2,$$

then

$$\widehat{A} \longrightarrow 0 \quad a.s. \quad \text{and} \quad \widehat{B} \longrightarrow 0 \quad a.s.$$

Proof of Lemma 4: Using the similar steps as in Walker [18], it easily follows that

$$\begin{aligned}
\widehat{A} &= \frac{2}{N}\sum_{n=1}^{N} X(n)\cos(\widehat{\omega}n) + o(1) \quad a.s. \\
\widehat{B} &= \frac{2}{N}\sum_{n=1}^{N} X(n)\sin(\widehat{\omega}n) + o(1) \quad a.s.
\end{aligned}$$

Now using Lemma 1, the result follows. Therefore, we have the following result.

Theorem 4: If the Assumptions 1-3 are satisfied, then for $k > p$, if $\widehat{\theta}_k = (\widehat{A}_k, \widehat{B}_k, \widehat{\omega}_k)$ minimizes $Q_k(A, B, \omega)$, then

$$\widehat{A}_k \longrightarrow 0 \quad a.s. \quad \text{and} \quad \widehat{B}_k \longrightarrow 0 \quad a.s..$$

4 Asymptotic Distribution of the Estimators

In this section we obtain the asymptotic distributions of the proposed estimators at each step. In this section we denote $Q_1(A, B, \omega)$ as $Q_1(\theta)$, *i.e.*,

$$Q_1(\theta) = \sum_{n=1}^{N} \left(y(n) - A\cos(\omega n) - B\sin(\omega n)\right)^2. \tag{16}$$

Now if we denote 3×3 diagonal matrix D as follows;

$$D = \begin{bmatrix} N^{-\frac{1}{2}} & 0 & 0 \\ 0 & N^{-\frac{1}{2}} & 0 \\ 0 & 0 & N^{-\frac{3}{2}} \end{bmatrix}, \tag{17}$$

then from (23) we can write

$$(\widehat{\theta}_1 - \theta_1^0)D^{-1}\left[DQ_1''(\bar{\theta})D\right] = -Q_1'(\theta_1^0)D. \tag{18}$$

Now observe that $\bar{\omega} \longrightarrow \omega_1^0$ $a.s.$, and $N(\bar{\omega} - \omega_1^0) \longrightarrow 0$ $a.s.$, therefore,

$$\lim_{N \longrightarrow \infty} DQ_1''(\bar{\theta})D = \lim_{N \longrightarrow \infty} DQ_1''(\theta_1^0)D. \tag{19}$$

It has been shown in the Appendix B that

$$Q_1'(\theta_1^0)D \xrightarrow{d} N_3(0, 4\sigma^2 c_1 \Sigma_1), \tag{20}$$

and

$$\lim_{N \longrightarrow \infty} DQ_1''(\theta_1^0)D \longrightarrow 2\Sigma_1, \tag{21}$$

where Σ_1 is same as defined in (28) and

$$c_1 = \left| \sum_{j=-\infty}^{\infty} a(j) \cos(\omega_1^0 j) \right|^2 + \left| \sum_{j=-\infty}^{\infty} a(j) \sin(\omega_1^0 j) \right|^2.$$

Here '$\xrightarrow{d}$' means convergence in distribution. Therefore, we have the following result.

THEOREM 5: If the Assumption 1-3 are satisfied, then

$$\left(N^{\frac{1}{2}}(\widehat{A}_1 - A_1^0), N^{\frac{1}{2}}(\widehat{B}_1 - B_1^0), N^{\frac{3}{2}}(\widehat{\omega}_1 - \omega_1^0)\right) \xrightarrow{d} N_3\left(0, \sigma^2 c_1 \Sigma_1^{-1}\right),$$

where

$$\Sigma_1^{-1} = \frac{4}{A_1^{0^2} + B_1^{0^2}} \begin{bmatrix} \frac{1}{2}A_1^{0^2} + 2B_1^{0^2} & -\frac{3}{2}A_1^0 B_1^0 & -3B_1^0 \\ -\frac{3}{2}A_1^0 B_1^0 & \frac{1}{2}B_1^{0^2} + 2A_1^{0^2} & 3A_1^0 \\ -3B_1^0 & 3A_1^0 & 6 \end{bmatrix}.$$

Proceeding exactly in the same manner, and using Theorem 2, it can be shown that similar result holds for any $k \leq p$ and it can be stated as follows.

THEOREM 6: If the Assumptions 1-3 are satisfied, then

$$\left(N^{\frac{1}{2}}(\widehat{A}_k - A_k^0), N^{\frac{1}{2}}(\widehat{B}_k - B_k^0), N^{\frac{3}{2}}(\widehat{\omega}_k - \omega_k^0)\right) \xrightarrow{d} N_3\left(0, \sigma^2 c_k \Sigma_k^{-1}\right),$$

here c_k and Σ_k^{-1} can be obtained from c_1 and Σ_1^{-1} by replacing A_1^0, B_1^0, ω_1^0 by A_k^0, B_k^0, ω_k^0.

5 Numerical Results

We performed several numerical experiments to check how the asymptotic results work for different sample sizes and for different models. All the computations were performed at the Indian Institute of Technology Kanpur, using the random number generator RAN2 of Press *et al.* [12]. All the programs are written in FORTRAN 77 and they can be obtained from the authors on request. We have considered the following three models:

Model 1: $y(n) = \sum_{j=1}^{2}[A_j \cos(\omega_j n) + B_j \sin(\omega_j n)] + X(n)$,

Model 2: $y(n) = \sum_{j=1}^{3}[A_j \cos(\omega_j n) + B_j \sin(\omega_j n)] + X(n)$,

Model 3: $y(n) = A_1 \cos(\omega_1 n) + B_1 \sin(\omega_1 n) + X(n)$.

Here $A_1 = 1.2$, $B_1 = 1.1$, $\omega_1 = 1.8$, $A_2 = 0.9$, $B_2 = 0.8$, $\omega_2 = 1.5$, $A_3 = 0.5$, $B_3 = 0.4$, $\omega_3 = 1.2$ and

$$X(n) = e(n) + 0.25\, e(n-1), \tag{22}$$

where $e(n)$'s are i.i.d. normal random variables with mean 0 and variance 1.0. For each model we considered different k values and different sample sizes. Mainly the following cases have been considered:

Case 1: Model 1, k = 1; Case 2: Model 2, k = 1;

Case 3: Model 2, k = 2; Case 4: Model 3, k = 2.

Therefore, in Case 1, 2 and 3, we have considered under-estimation, and in Case 4, over-estimation. For each p and N, we have generated the sample using the model parameters and the error structure (22). Then for fixed k we estimate the parameters using the sequential procedure provided in section 2. At each step the optimization has been performed using the downhill simplex method as described in Press *et al.* [12]. In each case we repeat the procedure 1000 times and report the average estimates and mean squared errors of all the unknown parameters. The results are reported in Tables 1 - 4.

Table 1: Model 1 is considered with $k = 1^*$.

		$A_1 = 1.20$	$B_1 = 1.10$	$\omega_1 = 1.8$
	AE	1.0109	1.1983	1.7980
N=100	MSE	(0.890E-01)	(0.534E-01)	(0.129E-02)
	ASYV	(0.450E-01)	(0.499E-01)	(0.859E-05)
	AE	1.1423	1.1287	1.7992
N=200	MSE	(0.289E-01)	(0.278E-01)	(0.272E-03)
	ASYV	(0.225E-01)	(0.250E-01)	(0.107E-05)
	AE	1.1686	1.1293	1.8001
N=300	MSE	(0.175E-01)	(0.177E-01)	(0.337E-06)
	ASYV	(0.150E-01)	(0.166E-01)	(0.318E-06)
	AE	1.1525	1.1449	1.8002
N=400	MSE	(0.142E-01)	(0.144E-01)	(0.170E-06)
	ASYV	(0.112E-01)	(0.125E-01)	(0.134E-06)

* The average estimates and the MSEs are reported for each parameter. The first row represents the true parameter values. In each box corresponding to each sample size, the first row represents the average estimates, the corresponding MSEs and the asymptotic variances (ASYV) are reported below within brackets.

Table 2: Model 2 is considered with $k = 1^*$.

		$A_1 = 1.20$	$B_1 = 1.10$	$\omega_1 = 1.8$
	AE	0.9743	1.2145	1.7979
N=100	MSE	(0.106E+00)	(0.558E-01)	(0.138E-02)
	ASYV	(0.450E-01)	(0.499E-01)	(0.859E-05)
	AE	1.1224	1.1537	1.7994
N=200	MSE	(0.319E-01)	(0.273E-01)	(0.272E-03)
	ASYV	(0.225E-01)	(0.250E-01)	(0.107E-05)
	AE	1.1556	1.1381	1.8001
N=300	MSE	(0.189E-01)	(0.183E-01)	(0.353E-06)
	ASYV	(0.150E-01)	(0.166E-01)	(0.318E-06)
	AE	1.1430	1.1574	1.8002
N=400	MSE	(0.154E-01)	(0.156E-01)	(0.186E-06)
	ASYV	(0.112E-01)	(0.125E-01)	(0.134E-06)

* The average estimates and the MSEs are reported for each parameter. The first row represents the true parameter values. In each box corresponding to each sample size, the first row represents the average estimates, the corresponding MSEs and the asymptotic variances (ASYV) are reported below within brackets.

Table 3: Model 2 is considered with $k = 2^*$.

		$A_2 = 0.9$	$B_2 = 0.8$	$\omega_2 = 1.5$
	AE	0.7757	0.8473	1.5052
N=100	MSE	(0.730E-01)	(0.567E-01)	(0.176E-02)
	ASYV	(0.510E-01)	(0.588E-01)	(0.182E-04)
	AE	0.8617	0.8172	1.5011
N=200	MSE	(0.281E-01)	(0.285E-01)	(0.273E-03)
	ASYV	(0.255E-01)	(0.294E-01)	(0.227E-05)
	AE	0.8813	0.8151	1.5001
N=300	MSE	(0.191E-01)	(0.201E-01)	(0.736E-06)
	ASYV	(0.170E-01)	(0.196E-01)	(0.673E-06)
	AE	0.8691	0.8215	1.5002
N=400	MSE	(0.144E-01)	(0.141E-01)	(0.315E-06)
	ASYV	(0.128E-01)	(0.147E-01)	(0.284E-06)

* The average estimates and the MSEs are reported for each parameter. The first column represents the parameter values and the corresponding asymptotic variance is reported within brackets. Below each sample size the average estimates and the corresponding MSEs and the asymptotic variances (ASYV) are also reported.

Table 4: Model 3 is considered with $k = 2^*$.

		$A_2 = 0.0$	$B_2 = 0.0$
	AE	0.7757	0.8473
N=100	VAR	(0.576E-01)	(0.545E-01)
	AE	0.8617	0.8172
N=200	VAR	(0.267E-01)	(0.282E-01)
	AE	0.8813	0.8151
N=300	VAR	(0.188E-01)	(0.199E-01)
	AE	0.8691	0.8215
N=400	VAR	(0.135E-01)	(0.136E-01)

* The average estimates and the variances (VARs) are reported for each parameter. In each box the first row represents the true parameter values which are zeros. In each box for each sample size, the first row represents the average estimates and the corresponding variances are reported below within brackets.

Some of the points are quite clear from these experiments. It is observed in all the cases that as the sample size increases the biases and the MSEs decrease. It verifies the consistency property of all the estimators. The biases of the linear parameters are much more than the non-linear parameters as expected. The MSEs match quite well with the asymptotic variances in most of the cases. Comparing Tables 1 and 2, it is observed that the biases and MSEs are more for most of the cases in Table 2. It indicates that even if we estimate the same number of parameters, if the number of parameters in the original model is less, then the estimates are better in terms of MSE.

For comparison purposes (one referee has suggested) we have approximated our method as indicated below. Let us estimate ω at each step as follows. At the k-th step instead of minimizing $R_k(\omega)$ for $0 < \omega < \pi$, minimize over Fourier frequencies only. The approximation is exactly equivalent to attributing in each step the maximum of the periodogram of the residual of the previous step to a sinusoidal model. Thus the possibility that such a maximum is due to a peak in the spectrum of the noise series (e.g. an autoregressive process with a root of its polynomial close to the unit circle) is ruled out. This is justified by the fact that the sinusoids will be asymptotically dominant in the periodogram of the data: in fact if there is a sinusoidal of frequency ω, then the periodogram at ω will be of order $O(N)$, while at all other frequencies it will of order $O(1)$. We have obtained the results for all the cases, but we have reported the results for Model 1 only in Table 5. Now comparing Table 1 and Table 5 it is clear that the MSEs of the frequencies are larger in the approximation method. It is not very surprising, because in the original method the variance of the frequency estimates are of the order $O(N^{-3})$, where as in the approximation method they are of the order $O(N^{-2})$, see Rice and Rosenblatt [14].

Table 5: Model 1 is considered with $k = 1^*$.

		$A_1 = 1.20$	$B_1 = 1.10$	$\omega_1 = 1.8$
N=100	AE	1.3219	0.5601	1.7890
	MSE	(0.219E+00)	(0.353E+00)	(0.134E-02)
	ASYV	(0.450E-01)	(0.499E-01)	(0.859E-05)
N=200	AE	0.4146	1.1465	1.8029
	MSE	(0.921E+00)	(0.459E+00)	(0.329E-03)
	ASYV	(0.225E-01)	(0.250E-01)	(0.107E-05)
N=300	AE	0.8085	1.3894	1.8019
	MSE	(0.160E+00)	(0.903E-01)	(0.361E-05)
	ASYV	(0.150E-01)	(0.166E-01)	(0.318E-06)
N=400	AE	1.3362	0.9144	1.7993
	MSE	(0.234E-01)	(0.394E-01)	(0.516E-06)
	ASYV	(0.112E-01)	(0.125E-01)	(0.134E-06)

* The average estimates and the MSEs are reported for each parameter. The first row represents the true parameter values. In each box corresponding to each sample size, the first row represents the average estimates, the corresponding MSEs and the asymptotic variances (ASYV) are reported below within brackets.

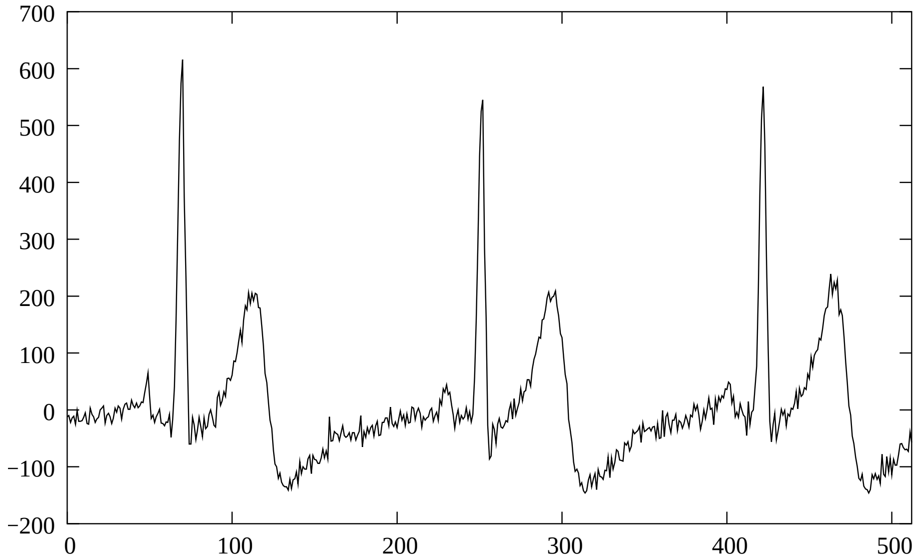

Figure 2: ECG signal.

6 Data Analysis

In this section we present two data analysis mainly for illustrative purpose. One is the original ECG signal and the other is the synthesized signal.

ECG Data: We want to model the following ECG signal, see Figure 2, using the model in (1). But in this case p is not known. We have plotted the periodogram function as defined in (2),in Figure 3 , to have an idea about the number of sinusoidal components present in the ECG signal. The number of components is not clear from the periodogram plot. Since it does not give any idea about the number of components, we have fitted the model sequentially for $k = 1, \ldots, 100$ and use the BIC to choose the number of components. The BIC takes the following form in this case

$$BIC(k) = N \ln \widehat{\sigma}_k^2 + \frac{1}{2}\ (3k + ar_k + 1)\ \ln N,$$

here $\widehat{\sigma}_k^2$ is the innovative variance, when the number of sinusoids is k. In this case, the number of parameters to be estimated is $3k + ar_k + 1$, where ar_k denotes the number of autoregressive parameters of an AR model when fitted to the residual. We plot the $BIC(k)$

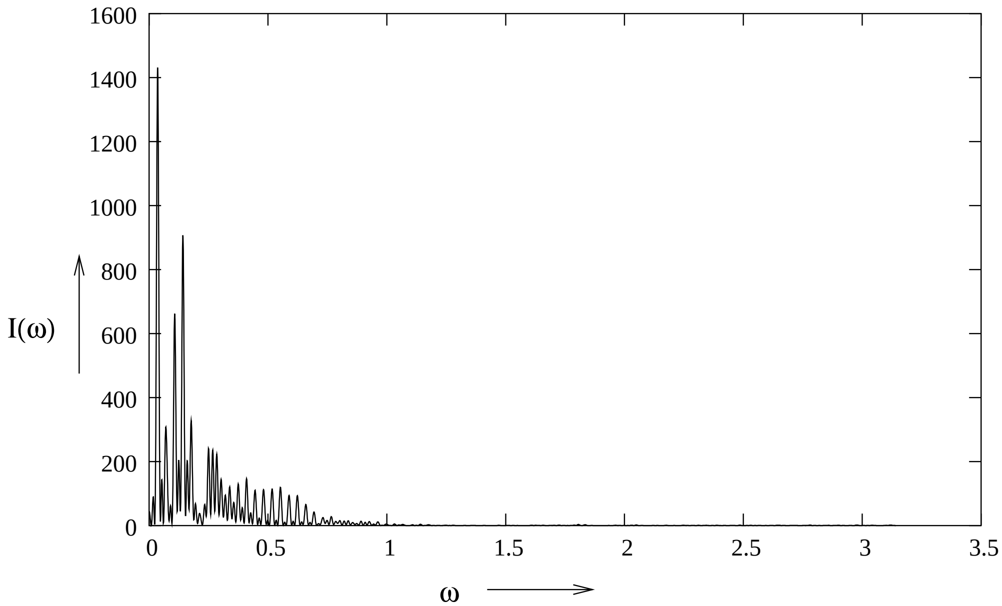

Figure 3: Periodogram function of the ECG signal.

as a function of k, in Figure 4. It is observed that for $k = 85$, the BIC takes the minimum value, therefore, in this case the estimate of p, say $\widehat{p} = 85$. So, we have fitted the model in (1) to the ECG data with $\widehat{p} = 85$. We estimate the parameters sequentially as described in section 2. The predicted and the actual signal are presented in Figure 5. They match quite well. We have also plotted the residuals in Figure 6. It is observed that the stationary AR(2) model fits the residuals. Therefore, the model assumptions are satisfied in this case. Note that it is possible to fit such a large order model because it has been done sequentially, otherwise it would have been a difficult task to estimate all the parameters simultaneously.

Synthesized Data: Now we analyze the synthesized signal which was presented in section 1. The data were generated from the model (3) and it is presented in Figure 7. Its periodogram function has been presented already in Figure 1. Although there are two frequencies - one at 0.20π and the other at 0.19π - present in the original signal, they are not evident from inspection of the periodogram.

We estimate the parameters using the sequential estimation technique as proposed in

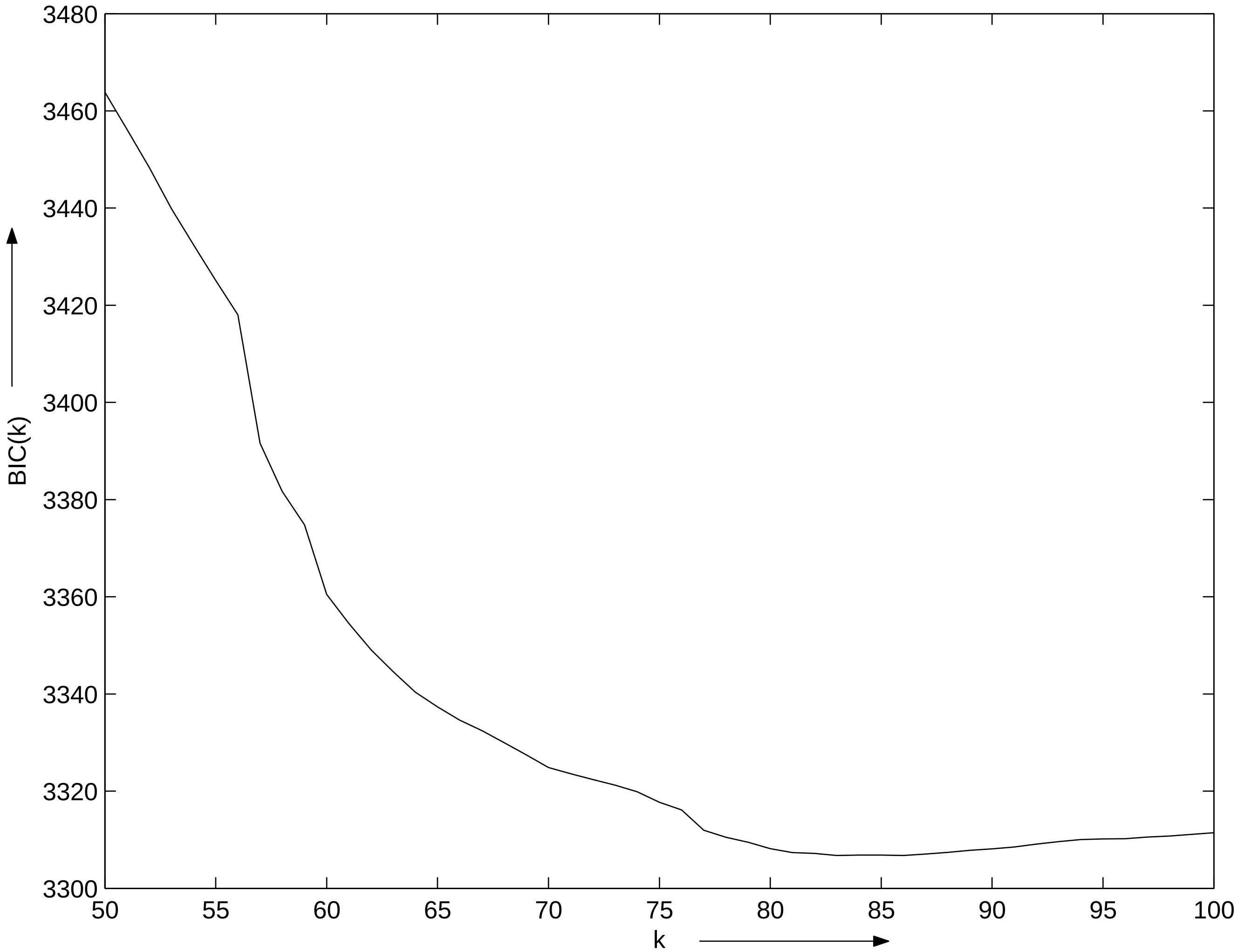

Figure 4: $BIC(k)$ values for different k.

section 2 and the estimates are as follows;

$$\widehat{A}_1 = 3.0513, \quad \widehat{B}_1 = 3.1137, \quad \widehat{\omega}_1 = 0.1996,$$

$$\widehat{A}_2 = 0.2414, \quad \widehat{B}_2 = -0.0153, \quad \widehat{\omega}_2 = 0.1811.$$

Therefore, it is observed that the estimates are quite good and they are quite close to the true values except for the parameter B_2^0. We provide the plot of predicted signal in Figure 8. We obtained the predicted plot just by replacing A's, B's and ω's by their estimates. It is easily observed that the predicted values match very well with the original data. Therefore,

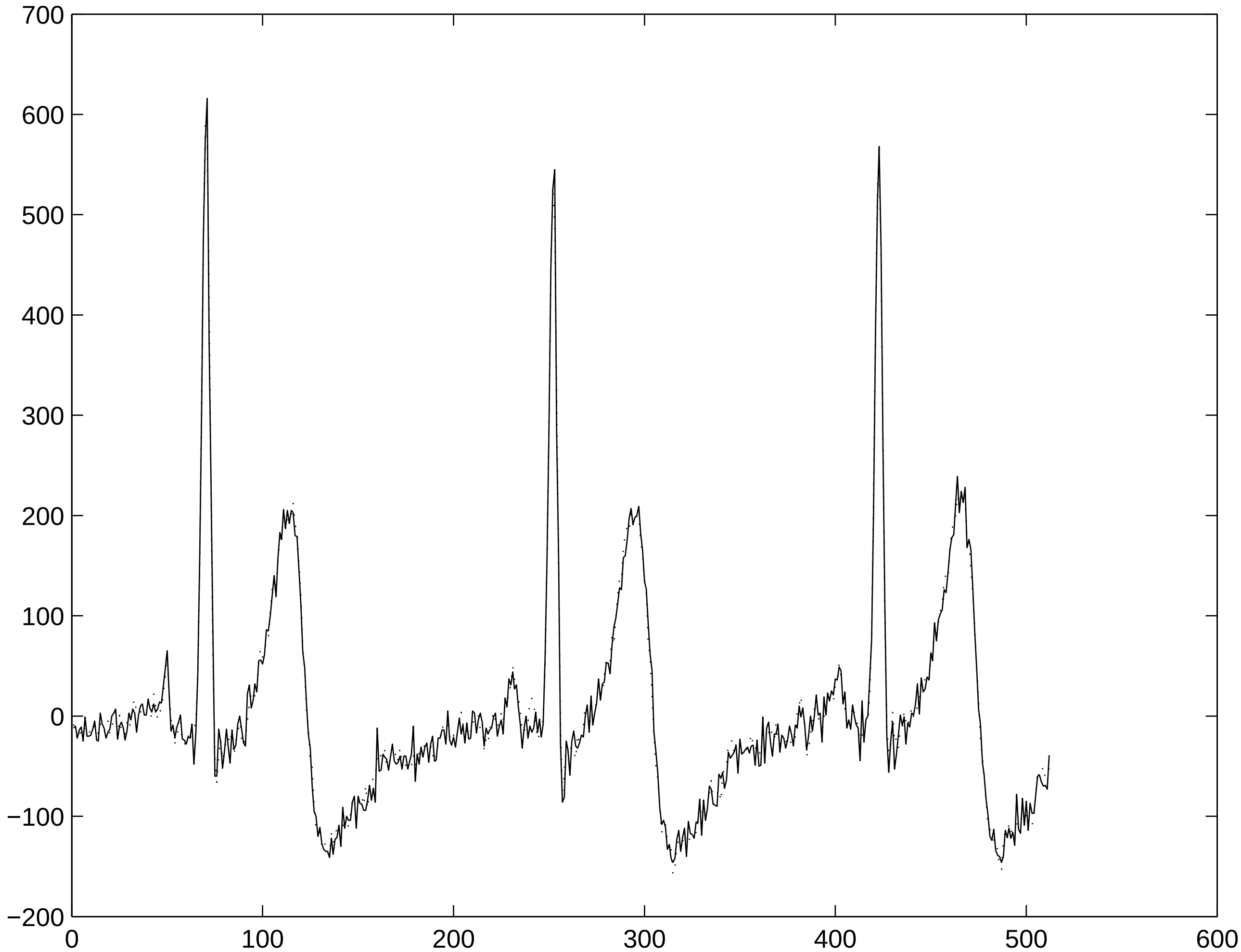

Figure 5: The original and the predicted ECG signal.

it is clear that the effect of the parameter B_2^0 is negligible in the data and that might be a possible explanation why $\widehat{B}_2$ is not close to B_2^0.

To see the effectiveness of BIC, we have simulated from the synthesized data with the *i.i.d.* error and with the stationary AR(2) error when the roots are very close to unity, for two different sample sizes namely $N = 75$ and $N = 750$, and have computed the percentage of times BIC detects the true model (*i.e.* $p = 2$). The results are presented in Table 6 and in Table 7. We have used the following AR(2) model; $X_t = 0.99X_{t-1} - 0.9801X_{t-2} + Z_t$. Note the when the roots of AR(2) model are very close to unity, then its periodogram has strong

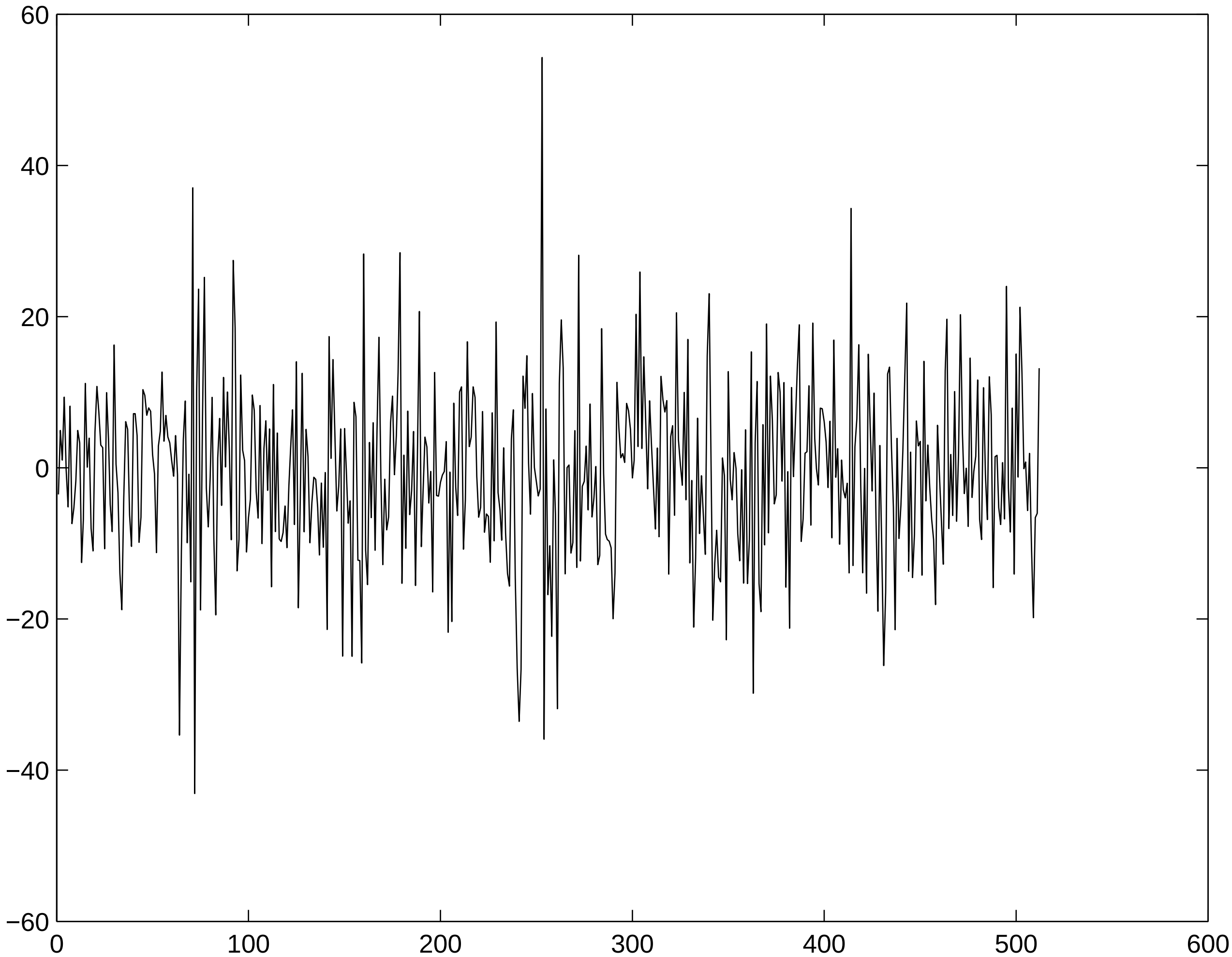

Figure 6: Residual plot after fitting the model to the ECG signal.

peaks.

7 Conclusions

In this paper we have provided a sequential estimation procedure for estimation of the unknown parameters of the sum of sinusoidal model. It is well known that this is a difficult problem from the numerical point of view. Although the least squares estimators are the most efficient estimators, it is difficult to use them when the number of components is large or when

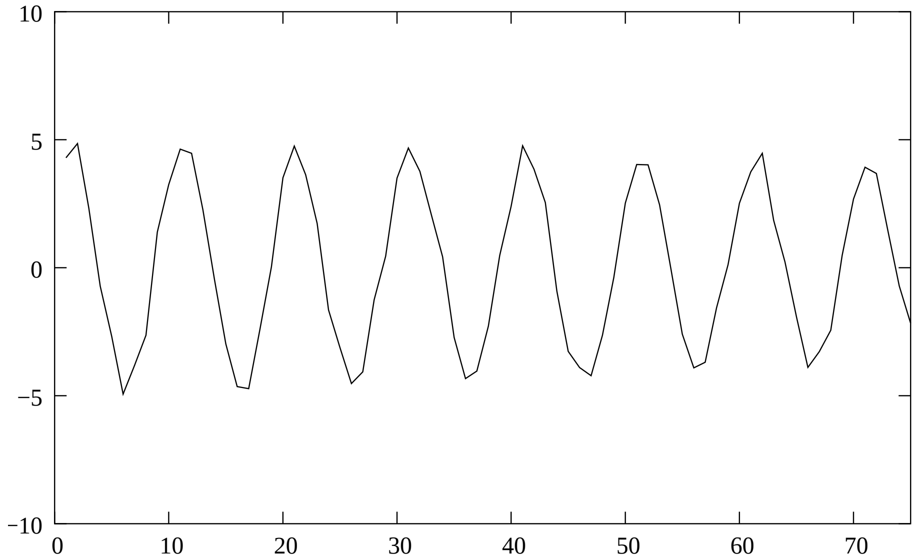

Figure 7: The synthesized signal.

Table 6: Percentage of Samples chosen by BIC for i.i.d. error

Sample Size →	75	750
Components	% of Samples	% of Samples
1	99	0
2	1	100
3	0	0
>= 4	0	0

two frequencies are very close to each other. It is observed that when we use the sequential procedure we solve several one dimensional minimization problems which are much easier to solve and also it is possible to detect two closely spaced frequencies. Interestingly, although the sequential estimates are different from the least squares estimators yet they have the same asymptotic efficiency as the least squares estimators. The proposed sequential method is very easy to implement and performs quite satisfactorily.

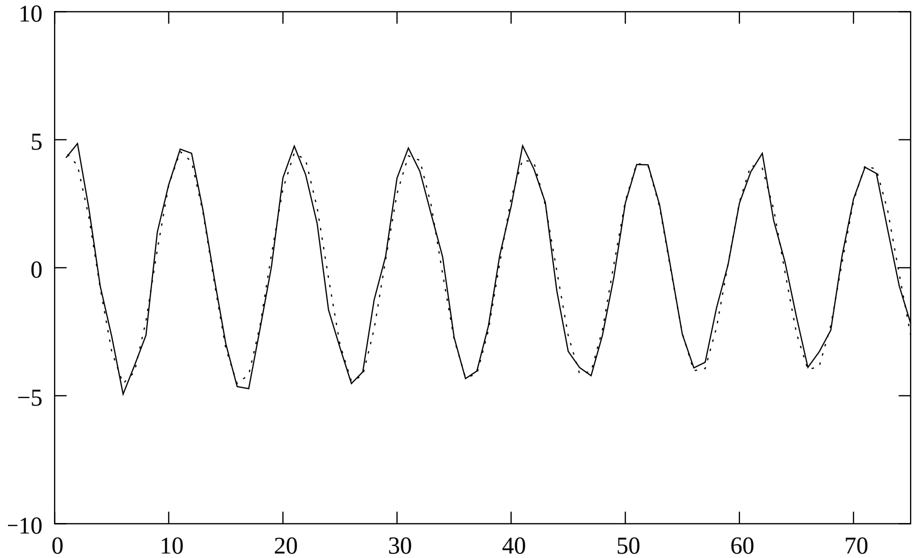

Figure 8: The synthesized signal and the predicted signal.

Table 7: Percentage of Samples chosen by BIC for AR(2) error

Sample Size →	75	750
Components	% of Samples	% of Samples
1	100	0
2	0	63
3	0	35
>= 4	0	2

Acknowledgements

The authors would like to thank two referees for their constructive suggestions. They would also like to thank the past editor-in-chief Professor John Stufken for his encouragements. The first author wants to thank CSIR for research funding.

Appendix A

Proof of Lemma 3

Suppose $\widehat{\theta}_1 = (\widehat{A}_1, \widehat{B}_1, \widehat{\omega}_1)$, $\widehat{\theta}_1^0 = (\widehat{A}_1^0, \widehat{B}_1^0, \widehat{\omega}_1^0)$ and $\bar{\theta} = (\bar{A}, \bar{B}, \bar{\omega})$. Let us denote $Q_1'(\theta_1)$ as the 3×1 first derivative matrix and $Q_1''(\theta_1)$ as the 3×3 second derivative matrix of $Q_1(\theta_1)$. Now from multivariate Taylor series expansion, we obtain

$$Q_1'(\widehat{\theta}_1) - Q_1'(\theta_1^0) = (\widehat{\theta}_1 - \theta_1^0) \;\; Q_1''(\bar{\theta}), \tag{23}$$

where $\bar{\theta} = (\bar{A}, \bar{B}, \bar{\omega})$ is a point on the line joining $\widehat{\theta}_1$ and θ_1^0. Note that $Q_1'(\widehat{\theta}_1) = 0$. Consider the following 3×3 diagonal matrix D_1 as follows;

$$D_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & N^{-1} \end{bmatrix}. \tag{24}$$

Now (23) can be written as

$$\left[(\widehat{\theta}_1 - \theta_1^0) D_1^{-1}\right] \left[\frac{1}{N} D_1 Q''(\bar{\theta}) D_1\right] = -\left[\frac{1}{N} Q'(\theta_1^0) D_1\right] \tag{25}$$

Let us consider the elements of $\frac{1}{N} Q'(\theta_1^0) D_1$,

$$\begin{aligned}
\frac{1}{N}\frac{\partial Q_1(\theta_1^0)}{\partial A} &= -\frac{2}{N}\sum_{n=1}^{N} X(n)\cos(\omega_1^0 n) \longrightarrow 0 \;\; a.s. \\
\frac{1}{N}\frac{\partial Q_1(\theta_1^0)}{\partial B} &= -\frac{2}{N}\sum_{n=1}^{N} X(n)\sin(\omega_1^0 n) \longrightarrow 0 \;\; a.s. \\
\frac{1}{N^2}\frac{\partial Q_1(\theta_1^0)}{\partial \omega} &= \frac{2}{N^2}\sum_{n=1}^{N} nX(n)A_1^0\sin(\omega_1^0) - \frac{2}{N^2}\sum_{n=1}^{N} nX(n)B_1^0\cos(\omega_1^0) \longrightarrow 0 \;\; a.s.
\end{aligned}$$

Therefore,

$$\frac{1}{N} Q'(\theta_1^0) D_1 \longrightarrow 0 \;\; a.s.$$

Observe that $\bar{\omega} \longrightarrow \omega^0$ $a.s.$, therefore

$$\lim_{N\to\infty} \frac{1}{N} D_1 Q_1''(\bar{\theta}) D_1 = \lim_{N\to\infty} \frac{1}{N} D_1 Q_1''(\theta_1^0) D_1 \qquad a.s.. \tag{26}$$

Now consider the elements of $\lim_{N\to\infty} \frac{1}{N} D_1 Q''(\theta^0) D_1$. By straight forward but routine calculations it easily follows that

$$\lim_{N\to\infty} \frac{1}{N} D_1 Q''(\theta^0) D_1 = 2\Sigma_1 \tag{27}$$

where

$$\Sigma_1 = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{4}\ B_1^0 \\ 0 & \frac{1}{2} & -\frac{1}{4}\ A_1^0 \\ \frac{1}{4}\ B_1^0 & -\frac{1}{4}\ A_1^0 & \frac{1}{6}(A_1^{0^2} + B_1^{0^2}) \end{bmatrix}, \tag{28}$$

which is a positive definite matrix. Therefore, $(\widehat{\theta}_1 - \theta_1^0)D_1^{-1} \longrightarrow 0$ *a.s.* Hence the lemma. ∎

Appendix B

First we show here that

$$Q_1'(\theta_1^0)D \xrightarrow{d} N_3(0, 4\sigma^2 c_1 \Sigma_1), \tag{29}$$

To prove (29), we need different elements of $Q_1'(\theta_1^0)$. Note that

$$\begin{aligned} \frac{\partial Q_1(\theta_1^0)}{\partial A} &= -2\sum_{n=1}^N \cos(\omega_1^0 n)\left[\sum_{j=2}^p \left[A_j^0\cos(\omega_j^0 n) + B_j^0 \sin(\omega_j^0 n)\right] + X(n)\right] \\ \frac{\partial Q_1(\theta_1^0)}{\partial B} &= -2\sum_{n=1}^N \sin(\omega_1^0 n)\left[\sum_{j=2}^p \left[A_j^0\cos(\omega_j^0 n) + B_j^0 \sin(\omega_j^0 n)\right] + X(n)\right] \\ \frac{\partial Q_1(\theta_1^0)}{\partial \omega} &= -2\sum_{n=1}^N n\left(A_1^0 \sin(\omega_1^0 n) - B_1^0\cos(\omega_1^0 n)\right)\left[\sum_{j=2}^p \left[A_j^0\cos(\omega_j^0 n) + B_j^0 \sin(\omega_j^0 n)\right] + X(n)\right]. \end{aligned}$$

Since for $0 < \alpha \neq \beta < \pi$,

$$\lim_{N\longrightarrow\infty} \frac{1}{\sqrt{N}}\sum_{n=1}^N \cos(\alpha n)\cos(\beta n) = 0, \quad \lim_{N\longrightarrow\infty} \frac{1}{\sqrt{N}}\sum_{n=1}^N \sin(\alpha n)\sin(\beta n) = 0 \tag{30}$$

$$\lim_{N\longrightarrow\infty} \frac{1}{N^{\frac{3}{2}}}\sum_{n=1}^N n\sin(\alpha n)\sin(\beta n) = 0, \lim_{N\longrightarrow\infty} \frac{1}{N^{\frac{3}{2}}}\sum_{n=1}^N n\cos(\alpha n)\cos(\beta n) = 0, \tag{31}$$

therefore,

$$Q_1'(\theta_1^0)D \overset{a.eq.}{=} -2\begin{bmatrix} N^{-\frac{1}{2}}\sum_{n=1}^N \cos(\omega_1^0 n)X(n) \\ N^{-\frac{1}{2}}\sum_{n=1}^N \sin(\omega_1^0 n)X(n) \\ N^{-\frac{3}{2}}\sum_{n=1}^N nX(n)\left(A_1^0\sin(\omega_1^0 n) - B_1^0\cos(\omega_1^0 n)\right) \end{bmatrix}. \tag{32}$$

Here $\overset{a.eq.}{=}$ means asymptotically equivalent. Now using the Central Limit Theorem (CLT) of the stochastic processes (see Fuller [4]), the right hand side of (32) tends to a 3-variate normal distribution with mean vector 0 and dispersion matrix $4\sigma^2 c_1 \Sigma_1$. Therefore, the result follows.

To prove

$$\lim_{N \longrightarrow \infty} DQ_1''(\theta_1^0)D \longrightarrow 2\Sigma_1, \tag{33}$$

we use the following results in addition to (30) and (31), for $0 < \alpha \neq \beta < \pi$,

$$\lim_{N \longrightarrow \infty} \frac{1}{N}\sum_{n=1}^{N} \sin^2(\alpha n) = \frac{1}{2}, \quad \lim_{N \longrightarrow \infty} \frac{1}{N}\sum_{n=1}^{N} \sin(\alpha n)\sin(\beta n) = 0,$$
$$\lim_{N \longrightarrow \infty} \frac{1}{N^2}\sum_{n=1}^{N} n\sin^2(\alpha n) = \frac{1}{4}, \lim_{N \longrightarrow \infty} \frac{1}{N^3}\sum_{n=1}^{N} n^2\sin^2(\alpha n) = \frac{1}{6},$$

similar results for cosine function also. Now the results can be obtained by routine calculations mainly considering each element of the $Q_1''(\theta_1^0)$ matrix and using the above equalities.

References

[1] Brillinger, D. (1987), “Fitting cosines: Some procedures and some physical examples”, Applied Probability, Stochastic Process and Sampling Theory, Ed. I.B. MacNeill and G. J. Umphrey, 75-100, dordrecht: Reidel.

[2] Chan, Y. T., Lavoie, J. M. M. and Plant, J. B. (1981), “A parameter estimation approach to estimation of frequencies of sinusoids”, *IEEE Trans. Acoust. Speech and Signal Processing*, ASSP-29, 214-219.

[3] Fisher, R.A. (1929), “Tests of significance in Harmonic analysis”, *Proc. Royal Soc. London, Ser A*, 125, 54-59.

[4] Fuller, W. A. (1976), *Introduction to Statistical Time Series*, John Wiley and Sons, New York.

[5] Hannan, E. J. (1971), "Non-linear time series regression", *Jour. Appl. Prob.*, Vol. 8, 767-780.

[6] Kay, S.M. (1988), *Modern Spectral Estimation: Theory and Applications*, Prentice Hall, Englewood Cliffs, New Jersey.

[7] Kay, S. M. and Marple, S. L. (1981), "Spectrum analysis- a modern perspective", *Proc. IEEE*, vol. 69, 1380-1419.

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

[9] Kundu, D. (2002), "Estimating parameters of sinusoidal frequency; some recent developments", *National Academy of Sciences Letters*, vol. 25, 53-73, 2002.

[10] Kundu, D. and Nandi, S. (2005), "Estimating the number of components of the fundamental frequency model", *Journal of the Japan Statistical Society*, vol. 35, no. 1, 41 - 59.

[11] Pisarenko, V. F. (1973), "The retrieval of harmonics from a covariance function", *Jour. Royal Astr. Soc.*, 33, 347-366.

[12] Press, W.H., Teukolsky, S.A., Vellerling, W.T., Flannery, B.P. (1992), *Numerical recipes in FORTRAN, the art of scientific computing*, 2nd. Edition, Cambridge University Press, Cambridge.

[13] Nandi, S. and Kundu, D. (2006), "Analyzing non-stationary signals using generalized multiple fundamental frequency model", *Journal of the Statistical Planning and Inference*, vol. 136, 3871-3903.

[14] Rice, J. A. and Rosenblatt, M. (1988), "On frequency estimation", *Biometrika*, vol. 75, 477-484.

[15] Richards, F. S. G. (1961), "A method of maximum likelihood estimation", *Journal of the Royal Statistical Society*, B, 469-475.

[16] Stoica, P. (1993), "List of references on spectral analysis", *Signal Processing*, vol. 31, 329-340.

[17] Tufts, D.W. and Kumaresan, R. (1982), "Estimation of frequencies of multiple sinusoids; making linear prediction perform like maximum likelihood", *Proceedings of IEEE*, 70, 975 - 989.

[18] Walker, A.M. (1971), "On the estimation of a Harmonic components in a time series with stationary residuals", *Biometrika*, 58, 21-36.