Robust Spectral Analysis of Sinusoidal Signal in Heavy Tail Impulsive Noise Environment

Debasis Kundu *

Abstract

Robust spectral analysis of sinusoidal signals in presence of impulsive noise environment is an important problem in the Statistical Signals Processing. In this paper we address the issue of robust estimation of the parameters of multicomponent sinusoidal model in presence of heavy tail impulsive noise. It is well known that although the least squares estimators are the most efficient estimators, in presence of outliers they perform quite poorly. On the other hand some of the standard robust estimators like the least absolute deviation estimators or Huber-M estimators are very difficult to implement in practice. Moreover, establishing the properties of the least absolute deviation estimators or Huber-M estimators in presence of heavy tail impulsive error is quite challenging. In this paper we have proposed to use sequential weighted least squares estimators, which are easy to implement in practice even when the number of components is quite large. Theoretical properties of the proposed estimators namely the consistency and the asymptotic normality have been established. Extensive simulations indicate that the performances of the proposed estimators are quite comparable with the other robust estimators like the least absolute deviation estimators and Huber-M estimators. It is observed that the performances of the proposed estimators depend on the choice of the weight function, and we have proposed two methods how the proper weight function can be chosen in practice. An illustrative example has been provided to show how the proposed method can be used in practice.

Key Words and Phrases: Sinusoidal model; least squares estimators; weighted least squares estimators; robust method of estimation; heavy tail random variable; consistency; asymptotic distribution.

AMS Subject Classifications: 62F10, 62F03, 62H12.

*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.

1 Introduction

Spectral analysis of sinusoidal signal in presence of additive noise is an important problem in the Statistical Signal processing. It has been used quite extensively in different areas of science and technology; for example in Physics, in bio-sciences, in bio-medical engineering, in finance, in music and voiced speech signal processing, in radar and sonar data analysis etc. An extensive amount of work has been done mainly to extract the original signal in presence of noise. Several estimation procedures and efficient algorithms have been developed during the last few decades, see for example Stoica and Moses [11] and Nandi and Kundu [8].

In practice, in many situations the measurements are contaminated with Gaussian and non-Gaussian noise, of which impulsive noise is the most common. An impulsive noise is an important class of disturbance in signal measurement which follows a non-Gaussian and heavy tailed distribution, and it can occur randomly with a value several times larger than the standard deviation of the background noise. It may not have finite variance also. Hence, the standard spectral analysis methods which work quite efficiently in presence of Gaussian background noise cannot be used in this case. One needs to develop robust spectral analysis method for this purpose. Robust spectral analysis of sinusoidal signal in presence of impulsive noise environment has received a considerable amount of attention in the Statistical Signal Processing. An extensive amount of work can be found in the literature dealing with the estimation of the signal parameters in presence of impulsive noise, see for example Tzagkarakis et al. [12], Wang et al. [13], Abrar et al. [1], Nandi et al. [6, 7], Zhou et al. [15], Lee and Kashyap [3], Zoubir et al. [16] and see the references cited there in.

In this paper we address the issue of the robust estimation of the parameters of multi-component sinusoidal model in presence of heavy tail impulsive noise. We have assumed the definition of Mandelbrot [5] that a heavy tail noise has infinite variance and the impulsive noise can be modeled as the mixture of a Gaussian noise and a heavy tail noise. To be more specific, it is assumed that the impulsive noise has a alpha stable distribution with infinite variance. It may be mentioned that the assumption of alpha stable error distribution is not very uncommon in the signal processing literature, see for example Lin et al. [4], Yue and Peng [14], Nandi et al. [6, 7], Nikias and Shao [9] and see the references cited there in.

Although, the least squares estimators (LSEs) or the maximum likelihood estimators (MLEs), in presence of Gaussian noise, are the most efficient estimators where the asymptotic variances attend the Cramer-Rao lower bound, it is well known that they perform quite poorly in presence of heavy tail error. In presence of impulsive errors, some robust estimator for example least absolute deviation estimators (LADEs) or Huber-M estimators (HMEs) may be preferred. But the main problem about the LADEs or HMEs is that they are very difficult to implement in practice. Particularly, it becomes quite challenging if the number of components is large. It needs to solve higher dimensional optimization problem, which require very accurate initial estimators, for any iterative procedure to converge. Moreover, establishing the theoretical properties of these robust estimators in presence of impulsive noise is quite challenging. So far it is not available in the literature. Hence, one needs to rely only on the simulation experiments to judge the performances of these robust estimators. Moreover, if there is no outliers in the data, their performances are worse than the LSEs.

In this paper we propose a very novel weighted least squares estimators (WLS estimators) as an alternative to the LSEs. First we will show how the WLS estimators can be implemented for one component model and then for multicomponent model. It is observed that the implementation of the WLS estimators in case of multicomponent model is computationally quite demanding particularly when the number of component is large. Hence, we have proposed a very efficient sequential WLS estimators which can be obtained quite conveniently even when the number of components is large. Our second aim is to establish the theoretical properties of the WLS estimators and the sequential WLS estimators. It is observed that the WLS estimators and the sequential WLS estimators have the same rate of convergence, and they can be established under the same set of assumptions as what are needed in case of the LSEs. Extensive simulations have been performed to see the performances of the proposed WLS estimators. It has been observed that the performances of the WLS estimators depend very much on the choice of the weight function. Hence, the choice of the weight function is an important issue in practice. We have provided two methods how the weight functions can be chosen in practice. The first one is to choose a proper weight function from a class of different weight functions and second one is based on an adaptive technique. An illustrative example has been provided to show how the proposed methods can be used in practice. The results are quite satisfactory.

The rest of the paper is organized as follows. In Section 2 we have provided the basic formulation of the model and also the necessary assumptions of the impulsive error. The WLS estimators of the one component model and multicomponent models are discussed in Section 3 and Section 4, respectively. In Section 5 the numerical results have been presented. The choice of weight function and an illustrative example have been presented in Section 6. Finally in Section 7 we have concluded the paper.

2 Model Formulation and Basic Assumptions

The model can be written as follows:

$$y(t) = \sum_{k=1}^{p} \{A_k^0 \cos(\theta_k^0 t) + B_k^0 \sin(\theta_k^0 t)\} + e(t); \qquad t = 1, \ldots, n. \tag{1}$$

Here, $y(t)$'s are the observed signal, θ_j's are the unknown frequencies lying between $(0, \pi)$, A_j's and B_j's are amplitude parameters and they are also unknown. In this paper the number of components p is assumed to be known and the error random variables $e(t)$'s have mean zero, and the explicit assumptions on $e(t)$'s will be provided now.

Before progressing further, let us recall that a symmetric (around 0) random variable S, i.e. X and -X have the same distribution, is said to have a symmetric α-stable ($S\alpha S$) distribution with the scale parameter $\beta > 0$ and the stability index $0 < \alpha \leq 2$, if the characteristic function of the random variable X is $E(e^{itX}) = e^{-\beta^\alpha |t|^\alpha}$. From now on it will be denoted by $S\alpha S(\alpha, \beta)$. The $S\alpha S$ distribution is a special case of the general stable distribution with non-zero shift and a skewness parameter. Note that for $\alpha = 2$, the $S\alpha S$ distribution becomes a normal distribution and for $\alpha = 1$, it becomes a Cauchy distribution. From now on we will always take $1 + \delta < \alpha < 2$, for $\delta > 0$. The distributions have undefined variance for $0 < \alpha < 2$ and undefined mean for $0 < \alpha \leq 1$. For different properties of the stable and α-stable distributions, one is referred to Samorodnitsky and Taqqu [10].

We need the following assumptions for further development.

Assumption 1: The error random variables $\{e(t)\}$ is a sequence of independent and identically distributed (i.i.d.) random variables and $e(t)$ has the following representation

$$e(t) = (1 - \pi)Z(t) + \pi X(t); \quad t = 1, \ldots n.$$

Here $Z(t)$'s are i.i.d. Gaussian random variables with mean 0 and variance σ^2, $X(t)$'s are i.i.d. random variables, with $E(X(t)) = 0$ and $E|X(t)|^{1+\delta} < \infty$, for $0 < \delta < 1$, $X(t)$ and $Z(t)$ are independently distributed and $0 < \pi \leq 1$.

Assumption 2: The error random variables $\{e(t)\}$ is a sequence of independent and identically distributed (i.i.d.) random variables and $e(t)$ has the following representation

$$e(t) = (1 - \pi)Z(t) + \pi X(t); \quad t = 1, \ldots n.$$

Here $Z(t)$'s are i.i.d. Gaussian random variables with mean 0 and variance σ^2, $X(t)$'s are i.i.d. $S\alpha S(\alpha, \beta)$ random variables, with $\beta > 0$, $1 + \delta < \alpha < 2$, $\delta > 0$, $X(t)$ and $Z(t)$ are independently distributed and $0 < \pi \leq 1$.

Assumption 3: The frequencies θ_k^0's are distinct, and $\theta_k^0 \in (0, \pi)$, for $k = 1, \ldots, p$.

Assumption 4: The amplitude parameters A_k^0 and B_k^0 satisfy the following order restrictions:

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

First observe that Assumption 1 indicates that the noise is a mixture of mean zero Gaussian error with probability $(1 - \pi)$ and and an error with mean zero and infinite variance with probability π. The $X(t)$ part behaves like an impulsive error which mean zero and infinite variance. Assumption 2 is a special case of Assumption 1 it will be needed to derive the asymptotic distribution of the proposed weighted least squares estimators. Assumption 3 is needed because if any two frequencies are same then the components can be combined and p will change. Further, the frequencies should not be in the boundaries. Note that Assumption 4 is a standard one which is needed for the identifiability purposes while estimating the frequencies, see for example Nandi and Kundu [8].

We also need to introduce the weight function $w(t)$ which will be used to construct the weighted least squares estimators (WLS estimators) in the subsequent sections. It is assumed that $w(t)$ is a m-degree polynomial defined on [0,1],i.e.

$$w(t) = a_0 + a_1 t + \ldots + a_m t^m, \tag{2}$$

where the weights $a_0, \ldots, a_m$ are such that

$$\min_{0 \le t \le 1} w(t) > \gamma > 0. \tag{3}$$

Suppose K is such that $\max\limits_{0 \le t \le 1} w(t) \le K$. Note that K depends on $a_0, \ldots, a_m$, but we do not make it explicit. Based on the above weight function we will define the WLS estimators of the unknown parameters for one and multicomponent sinusoidal model. It should be mentioned that although we have considered the polynomial weight function as provided in (3), all the results can be established even for any continuous $w(t)$ such that it satisfies (3).

3 WLS estimators: One Component Model

In this section we define the WLS estimators of the one component model and derive its asymptotic properties under appropriate conditions. The one component sinusoidal model is defined as follows:

$$y(t) = A^0 \cos(\theta^0 t) + B^0 \sin(\theta^0 t) + e(t). \tag{4}$$

Here, $e(t)$ is same as defined in (1). We use the following notation $\boldsymbol{\theta} = (A, B, \theta)$, and the true value of $\boldsymbol{\theta}$ as $\boldsymbol{\theta}^0$. If we define the weighted residual sums of squares as

$$Q(A, B, \theta) = Q(\boldsymbol{\theta}) = \sum_{t=1}^{n} w\left(\frac{t}{n}\right)\left(y(t) - A\cos(\theta t) - B\sin(\theta t)\right)^2, \tag{5}$$

then the WLS estimators of A^0, B^0 and θ^0, say $\widehat{A}$, $\widehat{B}$ and $\widehat{\theta}$, respectively, can be obtained by minimizing (5) with respect to A, B and θ. For fixed θ, the WLS estimators of A^0 and B^0 can be obtained as

$$\widehat{A}(\theta) = \frac{1}{a_{11}a_{22} - (a_{12})^2}[a_{22}d_1 - a_{12}d_2] \quad \text{and} \quad \widehat{B}(\theta) = \frac{1}{a_{11}a_{22} - (a_{12})^2}[a_{11}d_2 - a_{12}d_1], \tag{6}$$

where

$$a_{11} = \frac{1}{n}\sum_{t=1}^{n} w\left(\frac{t}{n}\right)\cos^2(t\theta), \quad a_{22} = \frac{1}{n}\sum_{t=1}^{n} w\left(\frac{t}{n}\right)\sin^2(t\theta),$$

$$a_{12} = a_{21} = \frac{1}{n}\sum_{t=1}^{n} w\left(\frac{t}{n}\right)\cos(t\theta)\sin(t\theta),$$

$$d_1 = \frac{1}{n}\sum_{t=1}^{n} w\left(\frac{t}{n}\right)y(t)\cos(t\theta), \quad d_2 = \frac{1}{n}\sum_{t=1}^{n} w\left(\frac{t}{n}\right)y(t)\sin(t\theta).$$

Once, $\widehat{A}(\theta)$ and $\widehat{B}(\theta)$ are obtained, the WLS estimator $\widehat{\theta}$ can be obtained by minimizing

$$Q(\widehat{A}(\theta), \widehat{B}(\theta), \theta) = \sum_{t=1}^{n} w\left(\frac{t}{n}\right)\left(y(t) - \widehat{A}(\theta)\cos(\theta t) - \widehat{B}(\theta)\sin(\theta t)\right)^2, \tag{7}$$

with respect to θ. Hence, it becomes a one-dimensional optimization problem. Once $\widehat{\theta}$, the WLS estimator of θ is obtained as the argument minimum of (7), the WLS estimators of A and B, say $\widehat{A}$ and $\widehat{B}$, respectively, can be obtained from (6), by replacing θ with $\widehat{\theta}$. The following theorem provides the consistency properties of the WLS estimators under appropriate conditions. It indicates that as the sample size increases, the WLS estimators converge to the true parameter values.

Theorem 1: Under Assumptions 1, 3 and 4, if the weight function $w(\cdot)$ satisfies (3), then $\widehat{\boldsymbol{\theta}} = (\widehat{A}, \widehat{B}, \widehat{\theta})$, the WLS estimator of $\boldsymbol{\theta}$, is a consistent estimator.

Proof: See in the Appendix A.

Now to derive the asymptotic properties of the WLS estimators, we need to introduce the following notations. For $k = 0, 1, \ldots,$

$$u_k = \int_0^1 t^k w(t) dt \quad \text{and} \quad v_k = \int_0^1 t^k w^2(t) dt.$$

The 3×3 diagonal matrices $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ as follows;

$$\boldsymbol{D}_1^{-1} = \text{diag}\left\{\pi n^{1/\alpha}, \pi n^{1/\alpha}, \pi n^{(1+\alpha)/\alpha}\right\}, \boldsymbol{D}_2^{-1} = \text{diag}\left\{\pi n^{-(\alpha-1)/\alpha}, \pi n^{-(\alpha-1)/\alpha}, \pi n^{-(2\alpha-1)/\alpha}\right\},$$

and the matrices

$$\boldsymbol{\Sigma} = \begin{bmatrix} v_0 & 0 & B^0 v_1 \\ 0 & v_0 & -A^0 v_1 \\ B^0 v_1 & -A^0 v_1 & (A^{0^2} + B^{0^2}) v_2 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Gamma} = \begin{bmatrix} u_0 & 0 & B^0 u_1 \\ 0 & u_0 & -A^0 u_1 \\ B^0 u_1 & -A^0 u_1 & (A^{0^2} + B^{0^2}) u_2 \end{bmatrix}.$$

Further, let $g_n(\boldsymbol{t}) = \frac{1}{n} \sum_{j=1}^{n} |K_j(\boldsymbol{t})|^\alpha$, where

$$K_j(\boldsymbol{t}) = w\left(\frac{j}{n}\right)\left(-t_1 \cos(j\theta^0) - t_2 \sin(j\theta^0) + \frac{jt_3}{n}(A^0 \sin(j\theta^0) - B^0 \cos(j\theta^0))\right),$$

and $\lim_{n\to\infty} g_n(\boldsymbol{t}) = g_0(\boldsymbol{t})$. Now we can state the asymptotic properties of the WLS estimators.

Theorem 2: Under Assumptions 2, 3 and 4, if the weight function $w(\cdot)$ satisfies (3), then $(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0)\boldsymbol{D}_2$ converges to a multivariate stable distribution with the characteristic function $\phi_0(\boldsymbol{t}) = e^{-2^\alpha \beta^\alpha g_0(\boldsymbol{t})}$.

Proof: See Appendix B.

4 Sequential WLS estimators: Multicomponent Model

In this section we consider the multicomponent sinusoidal model (1). The WLS estimators of the multicomponent model (1) can be obtained by minimizing

$$Q(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_p) = \sum_{t=1}^{n} w\left(\frac{t}{n}\right)\left(y(t) - \sum_{k=1}^{p}\{A_k \cos(\theta_k t) + B_k \sin(\theta_k t)\}\right)^2, \tag{8}$$

with respect to $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_p$, where $\boldsymbol{\theta}_k = (A_k, B_k.\theta_k)$, for $k = 1, \ldots, p$. It is a p dimensional optimization problem, and it becomes quite a challenging problem if p is large. Hence, to avoid that we propose a sequential WLS estimators which have the same asymptotic properties as the WLS estimators, and they can be obtained by solving p one dimensional optimization problems. Hence, the implementation of the proposed method is quite straight forward in practice even when p is large. It may be mentioned that the proposed method works based on the fact that the p sinusoidal components are asymptotically orthonormal when the frequencies are distinct (Assumption 3) and Assumption 4. The sequential method can be described as follows.

Obtain the estimates of $(A_1^0, B_1^0, \theta_1^0)$ by minimizing

$$Q_1(A_1, B_1, \theta_1) = \sum_{t=1}^{n} w\left(\frac{t}{n}\right)\left(y(t) - \{A_1 \cos(\theta_1 t) + B_1 \sin(\theta_1 t)\}\right)^2.$$

Let the estimates be $\widehat{\boldsymbol{\theta}}_1 = (\widehat{A}_1, \widehat{B}_1, \widehat{\theta}_1)$. Note that these estimates can be obtained by solving a one dimensional optimization problem. Next we remove the effect of the first component from the data and obtain the modified data as follows:

$$\widetilde{y}(t) = y(t) - \widehat{A}_1 \cos(\widehat{\theta}_1 t) - \widehat{B}_1 \sin(\widehat{\theta}_1 t).$$

Now obtain the estimates of $(A_2^0, B_2^0, \theta_2^0)$ by minimizing

$$Q_2(A_2, B_2, \theta_2) = \sum_{t=1}^{n} w\left(\frac{t}{n}\right)\left(\widetilde{y}(t) - \{A_2 \cos(\theta_2 t) + B_2 \sin(\theta_2 t)\}\right)^2,$$

and obtain the estimates as $\widehat{\boldsymbol{\theta}}_2 = (\widehat{A}_2, \widehat{B}_2, \widehat{\theta}_2)$. Continue the process p times and obtain the estimates as $(\widehat{\boldsymbol{\theta}}_1, \ldots, \widehat{\boldsymbol{\theta}}_p)$. They will be called as sequential WLS estimators (SWLS estimators). We have the following properties of the proposed SWLS estimators.

Theorem 3: Under Assumptions 1, 3 and 4, if the weight function $w(\cdot)$ satisfies (3), then $\widehat{\boldsymbol{\theta}}_k = (\widehat{A}_k, \widehat{B}_k, \widehat{\theta}_k)$, the WLS estimator of $\boldsymbol{\theta}_k$, for $k = 1, \ldots, p$, are consistent estimators.

Proof: See in the Appendix C.

Theorem 4: Under Assumptions 2, 3 and 4, if the weight function $w(\cdot)$ satisfies (3), then for $(\widehat{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_k^0)\boldsymbol{D}_2$ converges to a multivariate stable distribution with the same form of the characteristic function $\phi_0(\boldsymbol{t})$ as in Theorem 2, where $\boldsymbol{\theta}^0$ has been replaced by $\boldsymbol{\theta}_k^0$, for $k = 1, \ldots, p$. Further, for $1 \le k \ne m \le p$, as $n \to \infty$, $(\widehat{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_k^0)\boldsymbol{D}_2$ and $(\widehat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_k^0)\boldsymbol{D}_2$ are independently distributed.

Proof: Note that the first part of the proof follows along the same line as the proof of Lemma 2. The second part of the proof follows using Lemma 1. ∎

Comments: It should be mentioned that the sequential WLSEs provide consistent estimators because for a p component sinusoidal model, asymptotically the components are orthonormal when the frequencies are different. Due to the same reason, asymptotically the sequential WLSEs and WLSEs behave in a similar manner and for $k \ne m$, $(\widehat{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_k^0)\boldsymbol{D}_2$ and $(\widehat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_k^0)\boldsymbol{D}_m$ are independently distributed as $n \to \infty$.

5 Numerical Results

In this section we have presented some simulation results to show how the proposed methods perform in practice. The main aim of this section is to see how the weighted and the sequential weighted least squares estimators perform in presence of impulsive noise compared to the well known LAD and Huber-M estimators. We have considered two different models, and different sample sizes. We have also considered different proportions of outliers to see the effect of the proportion of outliers in the data set. The following two models have been considered.

Model 1:

$$y(t) = A^0 \cos(\theta^0 t) + B^0 \sin(\theta^0 t) + \epsilon(t), \tag{9}$$

Model 2:

$$y(t) = A_1^0 \cos(\theta_1^0 t) + B_1^0 \sin(\theta_1^0 t) + A_2^0 \cos(\theta_2^0 t) + B_2^0 \sin(\theta_2^0 t) + \epsilon(t), \tag{10}$$

In case of Model 1, we have considered $A^0 = B^0 = 2$, and $\theta^0 = 0.15\pi$. In case of Model 2, we have considered $A_2^0 = B_2^0 = 2$, $A_1^0 = B_1^0 = 1$, $\theta_1^0 = 0.15\pi$ and $\theta_2^0 = 1.25\pi$. For the error components we have taken $\sigma^2 = 1$, $\beta = 1$, $\alpha = 1.5$, and different π values. Moreover, it is assumed that the outliers are present in the middle portion of the data set. It may be mentioned that there is nothing special about the choice of these parameter values. Similar phenomena have been observed for other choices also.

We have considered LSEs, LADEs, HMEs and two different WLSEs. Note that to compute LADEs and HMES we have used Nelder-Mead algorithm, and one needs to solve three and six dimensional optimization problems for Model 1 and Model 2, respectively. In case of sequential WLSEs, one needs to solve one dimensional optimization problem in each case. We have considered the following two weight functions for the computation of the WLSEs, for $0 \le t \le 1$;

$$w_1(t) = 1 + \left(t - \frac{1}{2}\right)^2 \quad \text{and} \quad w_2(t) = \begin{cases} 2 - 4t & \text{if} \quad 0 \le t \le \frac{1}{4} \\ 1 & \text{if} \quad \frac{1}{4} \le t \le \frac{3}{4} \\ 4t - 2 & \text{if} \quad \frac{3}{4} \le t \le 1. \end{cases}$$

Observe that here $w_1(t)$ is a polynomial of degree 2, and it has the minimum at the point 1/2 and the minimum value is 1. The function $w_2(t)$ is not a polynomial, but it is a non-negative continuous function with the minimum value 1. It is a decreasing function from 0 to 1/4, constant between 1/4 to 3/4 and then it is an increasing function from 3/4 to 1. Both the weight functions put less weights in the middle portion of the data set as it has been assumed that the outliers are in the middle portion of the data set.

For each generated sample and we have computed the LSEs, LADEs, HMEs and WLSEs. We have reported the average estimates and the associated mean absolute errors (MAEs) based on 10,000 replications. It may be mentioned that since outliers are present it is better to provide MAEs instead of mean squared errors. The results for Model 1 are presented in Table 1 and for Model 2 in Tables 2 and 3.

Some of the points are quite clear from these numerical results. It is observed as the sample size increases the average biases and the MAEs decreases for a given proportion of outliers. It verifies the consistency properties of all the different estimators. Now comparing the different estimators, it is observed that when there is no outliers, the performances of the LSEs are the best. The two WLSEs also behave very similar to the LSEs, where as the LADEs and HMEs do not perform that well in terms of the MAEs. When there are outliers in the data, the performances of the LSEs are the worst compared to the other robust estimators. The performances of the LADEs, HMEs, and the two WLSEs are very similar in presence of outliers. Since the implementation of the WLSEs is more convenient compared to the LADEs or HMEs, we recommend to use WLSEs in practice.

π	n	LSE	WLSE (w_1)	WLSE (w_2)	MAE	HME
	25	0.4747	0.4751	0.4764	0.4762	0.4771
		(3.0754E-02)	(3.1326E-02)	(3.1541E-02)	(3.2537E-02)	(3.2722E-02)
	50	0.4711	0.4722	0.4692	0.4755	0.4745
0.00		(8.9790E-03)	(1.0240E-02)	(1.1573E-02)	(1.8315E-02)	(1.2745E-02)
	75	0.4714	0.4714	0.4721	0.0.4721	0.4715
		(4.7195E-03)	4.9913E-03)	(4.8934E-03)	(5.1123E-03)	(5.1234E-03)
	100	0.4713	0.4713	0.4713	0.4729	0.4931
		(2.8765E-03)	(2.9998E-03)	(3.0245E-03)	(3.1321E-03)	(3.1765E-03)
	25	0.4867	0.4819	0.4762	0.4768	0.4749
		(4.7212E-02)	(4.3808E-02)	(4.3812E-02)	(4.3799E-02)	(3.3809E-02)
	50	0.4770	0.4715	0.4721	0.4770	0.4713
0.20		(2.0723E-02)	(1.1240E-02)	(1.1743E-02)	(1.0843E-02)	(1.1267E-02)
	75	0.4747	0.4716	0.4726	0.0.4717	0.4716
		(9.5390E-03)	5.2454E-03)	(5.6741E-03)	(5.2953E-03)	(5.2655E-03)
	100	0.4755	0.4715	0.4710	0.4719	0.4718
		(8.0090E-03)	(3.2817E-03)	(3.3679E-03)	(3.3273E-03)	(3.2755E-03)
	25	0.4941	0.4823	0.4756	0.4779	0.4751
		(5.7350E-02)	(4.6376E-02)	(4.7123E-02)	(4.6299E-02)	(4.6366E-02)
	50	0.4800	0.4719	0.4724	0.4773	0.4722
0.30		(2.5228E-02)	(1.2004E-02)	(1.2115E-02)	(1.2076E-02)	(1.2111E-02)
	75	0.4747	0.4716	0.4726	0.0.4717	0.4716
		(9.5390E-03)	5.2454E-03)	(5.6741E-03)	(5.2953E-03)	(5.2655E-03)
	100	0.4762	0.4717	0.4711	0.4712	0.4717
		(1.0024E-02)	(3.5509E-03)	(3.5871E-03)	(3.5423E-03)	(3.5517E-03)
	25	0.5097	0.4862	0.4758	0.4821	0.4784
		(8.1128E-02)	(5.6141E-02)	(5.5921E-02)	(5.5911E-02)	(5.6021E-02)
	50	0.4860	0.4739	0.4735	0.4731	0.4735
0.40		(3.2942E-02)	(1.5464E-02)	(1.5732E-02)	(1.5412E-02)	(1.5527E-02)
	75	0.4820	0.4729	0.4727	0.0.4727	0.4726
		(2.0128E-03)	7.0357E-03)	(7.1114E-03)	(7.0446E-03)	(7.1003E-03)
	100	0.4788	0.4721	0.4711	0.4715	0.4715
		(1.3515E-02)	(4.4127E-03)	(4.4336E-03)	(4.3992E-03)	(4.4117E-03)

Table 1: Average estimates, MAEs of the unknown parameters based on LSEs, LADEs, HMEs and WLSEs of the frequency θ^0, for Model 1. Here $\theta^0 = 0.15\pi = 0.4712$, $\sigma^2 = 1$ and $\alpha = 1.5$.

π	n	LSE	WLSE (w_1)	WLSE (w_2)	MAE	HME
	25	0.4723	0.4709	0.4703	0.4679	0.4682
		(1.6845E-02)	(1.6248E-02)	(1.6139E-02)	(1.6352E-02)	(1.6211E-02)
	50	0.4725	0.4727	0.4731	0.4732	0.4731
0.00		(4.3661E-03)	(4.4603E-03)	(4.7826E-03)	(6.2764E-03)	(5.3554E-03)
	75	0.4716	0.4716	0.4717	0.4716	0.4716
		(2.2819E-03)	2.3093E-03)	(2.3746E-03)	(3.1990E-03)	(2.6997E-03)
	100	0.4716	0.4716	0.4717	0.4715	0.4715
		(1.4221E-03)	(1.4384E-03)	(1.4635E-03)	(1.9271E-03)	(1.6353E-03)
	25	0.4722	0.4709	0.4703	0.4710	0.4709
		(1.6845E-02)	(1.6245E-02)	(1.6231E-02)	(1.6576E-02)	(1.6214E-02)
	50	0.4738	0.4728	0.4736	0.4729	0.4729
0.20		(6.8556E-03)	(6.2743E-03)	(6.1118E-03)	(6.3781E-03)	(6.2715E-03)
	75	0.4726	0.4719	0.4723	0.0.4717	0.4716
		(3.6937E-03)	3.1120E-03)	(3.0402E-03)	(3.2405E-03)	(3.0976E-03)
	100	0.4722	0.4724	0.4719	0.4718	0.4717
		(2.5648E-03)	(2.2817E-03)	(2.1878E-03)	(2.1988E-03)	(2.2755E-03)
	25	0.4738	0.4714	0.4711	0.4712	0.4713
		(1.9393E-02)	(1.7793E-02)	(1.7664E-02)	(1.7784E-02)	(1.7795E-02)
	50	0.4734	0.4734	0.4742	0.4735	0.4737
0.30		(7.8163E-03)	(6.9315E-03)	(6.9345E-03)	(6.9411E-03)	(6.9337E-03)
	75	0.4729	0.4722	0.4727	0.0.4729	0.4722
		(4.6256E-03)	3.7458E-03)	(3.7623E-03)	(3.7554E-03)	(3.7321E-03)
	100	0.4718	0.4726	0.4718	0.4716	0.4715
		(3.1165E-03)	(2.9876E-03)	(2.9778E-03)	(2.8987E-03)	(2.9453E-03)
	25	0.5097	0.4862	0.4758	0.4821	0.4784
		(8.1128E-02)	(5.6141E-02)	(5.5921E-02)	(5.5911E-02)	(5.6021E-02)
	50	0.4860	0.4739	0.4735	0.4731	0.4735
0.40		(3.2942E-02)	(1.5464E-02)	(1.5732E-02)	(1.5412E-02)	(1.5527E-02)
	75	0.4820	0.4729	0.4727	0.0.4727	0.4726
		(9.0128E-03)	7.0357E-03)	(7.1114E-03)	(7.0446E-03)	(7.1003E-03)
	100	0.4788	0.4721	0.4711	0.4715	0.4715
		(8.3515E-03)	(4.4127E-03)	(4.4336E-03)	(4.3992E-03)	(4.4117E-03)

Table 2: Average estimates, MAEs of the unknown parameters based on LSEs, LADEs, HMEs and WLSEs of the frequency θ^0, for Model 2. Here $\theta^0 = 0.15\pi = 0.4712$, $\sigma^2 = 1$ and $\alpha = 1.5$.

π	n	LSE	WLSE (w_1)	WLSE (w_2)	MAE	HME
0.00	25	3.9301	3.9308	3.9327	3.9451	3.9339
		(3.1741E-02)	(3.3276E-02)	(3.4694E-02)	(5.3383E-02)	(4.0373E-02)
	50	3.9277	3.9274	3.9278	3.9278	3.9274
		(8.5684E-03)	(4.4830E-03)	(8.6446E-03)	(8.7057E-03)	(8.7226E-03)
	75	3.9270	3.9269	3.9270	3.9270	3.9270
		(3.9834E-03)	(3.9843E-03)	(4.1131E-03)	(4.2315E-03)	(4.2221E-03)
	100	3.9270	3.9270	3.9270	3.9270	3.9270
		(2.4513E-03)	(2.4413E-03)	(2.5016E-03)	(2.5531E-03)	(2.5725E-03)
0.20	25	3.9176	3.9220	3.9381	3.9495	3.9339
		(1.0694E-01)	(7.4561E-02)	(7.2657E-02)	(7.4119E-02)	(7.4220E-02)
	50	3.9204	3.9266	3.9307	3.9392	3.9280
		(4.9677E-02)	(2.2456E-02)	(2.4531E-02)	(2.2665E-02)	(2.3154E-02)
	75	3.9262	3.9243	3.9288	3.9269	3.9269
		(2.5892E-02)	1.8664E-02)	(1.7656E-02)	(1.7751E-02)	(1.7659E-02)
	100	3.9216	3.9263	3.9257	3.9271	3.9269
		(1.8561E-02)	(1.2546E-02)	(1.2580E-02)	(1.2567E-02)	(1.2552E-02)
0.30	25	3.9181	3.9267	3.9389	3.9487	3.9362
		(1.3556E-01)	(1.1193E-01)	(1.1182E-01)	(1.1213E-01)	(1.1191E-01)
	50	3.9221	3.9193	3.9316	3.9299	3.9289
		(6.5183E-02)	(4.3054E-02)	(4.1123E-02)	(4.2678E-02)	(4.2897E-02)
	75	3.9236	3.9268	3.9323	3.9270	3.9269
		(3.6136E-02)	2.3392E-02)	(3.0875E-02)	(3.1156E-02)	(3.1789E-02)
	100	3.9242	3.9242	3.9244	3.9270	3.9270
		(2.3172E-02)	(1.3123E-02)	(1.3021E-02)	(1.3321E-02)	(1.3002E-02)
0.40	25	3.9183	3.9312	3.9358	3.9576	3.9402
		(1.9119E-01)	(1.7064E-01)	(1.6754E-01)	(1.7113E-01)	(1.7041E-01)
	50	3.9335	3.9262	3.9323	3.9331	3.9338
		(8.9901E-02)	(5.8843E-02)	(5.8765E-02)	(5.9013E-02)	(5.9101E-02)
	75	3.9236	3.9229	3.9263	3.9272	3.9269
		(5.1209E-02)	3.8226E-02)	(3.7787E-02)	(3.8001E-02)	(3.7910E-02)
	100	3.9230	3.9270	3.9271	3.9270	3.9270
		(3.3832E-02)	(2.3415E-02)	(2.3212E-02)	(2.3192E-02)	(2.3212E-02)

Table 3: Average estimates, LAEs of the unknown parameters based on LSEs, MAEs, HMEs and WLSEs of the frequency θ^0, for Model 2. Here $\theta^0 = 1.25\pi = 3.9270$, $\sigma^2 = 1$ and $\alpha = 1.5$.

6 Choice of Weight Function and an Illustrative Example

6.1 Choice of Weight Function

In the previous section we have performed an extensive simulation experiments to show the performances of the WLSEs. It is observed that the performances of the WLSEs are better than the LSEs in presence of outliers and they perform very similar to the LSEs when there is no outliers in the data. But it is also true that the performances of the WLSEs depend on the choice of the weight function. We may use the following criterion to choose the optimal weight function. Choose the optimal weight function as $\widehat{w}(\cdot)$ which minimizes

$$\sum_{t=1}^{n} w\left(\frac{t}{n}\right)\left(y(t)-\sum_{k=1}^{p}\{\widehat{A}_k\cos(\widehat{\theta}_k t)+\widehat{B}_k\sin(\widehat{\theta}_k t)\}\right)^2, \tag{11}$$

among the class of all weight function $w(\cdot)$ which satisfies (3). Here, the estimators $\{(\widehat{A}_k, \widehat{B}_k, \widehat{\theta}_k); k = 1, \ldots, p\}$ clearly depend on the weight function $w(\cdot)$, but we do not make it explicit for brevity. Unfortunately, analytically it becomes a very difficult problem, hence we suggest two heuristic methods which can be used in practice.

If it is known before hand in which portion of the data set, the outliers might be present then one can choose the weight function accordingly. But in practice it may not be known, in that case we suggest the following two methods to choose an appropriate weight function.

Method 1:

Choose a class of non-negative continuous functions in the domain $[0, 1]$, satisfying (3) with different shapes. For example choose functions of different shapes like increasing, decreasing, unimodal, bathtub shaped etc. Calculate WLSEs in each case and also compute the corresponding residual sums of squares. Now choose that particular weight function which has the minimum residual sums of squares. In practice we have seen that if we take five to six weight functions in the class it serves the purpose. We also suggest to keep $w(t) = 1$, for all $0 \leq t \leq 1$ as one of the member in the class. We illustrate the proposed method with the following example.

Method 2: Since we do not have any idea whether there exists any out liers in the data set or not, further even is there are out liers, there position is not known, we first find the least squares estimators of the known parameters. Based on the LSEs, we obtain the residuals as $r(1), \ldots, r(n)$. We obtain the best fitted m degree polynomial to $r(1)^{-2}, \ldots, r(n)^{-2}$. Let the best fitted m degree polynomial be denoted by

$$\widehat{w}(t) = \widehat{a}_0 + \widehat{a}_1 t + \ldots + \widehat{a}_m t^m.$$

Here m is chosen based on BIC criterion. Now use $\widehat{w}(t)$ as the weight function. Intuitively, $\widehat{w}(t)$ puts less weight where $r^{-2}(t)$ is small, i.e. $r^2(t)$ is large.

6.2 Illustrative Example

In this section we provide an illustrative example how the above method can be used in practice. We will also illustrate how difficult it will be to implement other standard robust estimators like LADEs or HMEs. We have generated a sample from the following model parameters:

$$A_1 = B_1 = 5.0, A_2 = B_2 = 4.0, A_3 = B_3 = 3.0, A_2 = B_2 = 2.0, A_5 = B_5 = 1.0$$

$$\omega_1 = 1.25\pi, \omega_2 = 0.25\pi, \omega_3 = 0.75\pi, \omega_4 = 0.15\pi, \omega_5 = 0.95\pi.$$

The signal is added with standard Gaussian error with 20% outliers with SαS error with β =1 and α = 1.5. The signal has been generated based on five components. The signal $y(t)$ and the corresponding periodogram function have been plotted in Figure 1 and Figure 2, respectively. Although the signal has five components, the periodogram function cannot identify the five components, hence it becomes very difficult to identify all the initial guesses to start the iterative process for LADEs or HMEs. On the other hand we can easily identify the initial guesses for the sequential WLSEs.

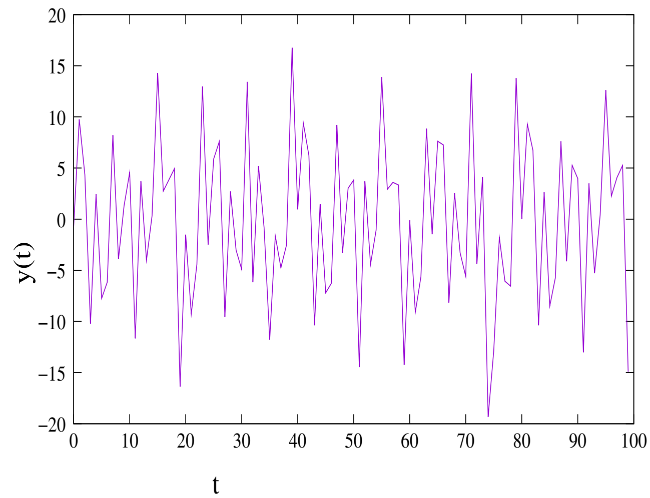

Figure 1: Noisy signal with outliers.

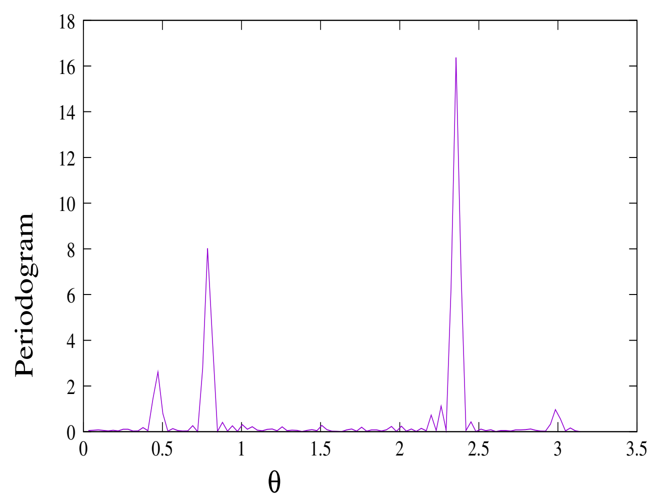

Figure 2: Periodogram plot of the signal plotted in Figure 1.

Since we have no idea about the position of the outliers, we have taken the following six

WLSEs	w_1	w_2	w_3	w_4	w_5	w_6
RSS	3.9600	3.9531	3.9385	3.9359	3.9501	3.9942

Table 4: Residual sum of squares based on different WLSEs.

different weight functions with various shapes;

$$(i)\quad w_1(t) = 1 + (t - 0.5)^2, \quad (ii)\quad w_2(t) = 1 + (t - 0.25)^2, \quad (iii)\quad w_3(t) = 1 - (t - 0.5)^2,$$

$$(iv)\quad w_4(t) = \frac{1}{1 + (t - 0.75)^2}, \quad (v)\quad w_5(t) = 1$$

and

$$(vi)\quad w_6(t) == \begin{cases} 2 - 4t & \text{if} \quad 0 \le t \le \frac{1}{4} \\ 1 & \text{if} \quad \frac{1}{4} \le t \le \frac{3}{4} \\ 4t - 2 & \text{if} \quad \frac{3}{4} \le t \le 1. \end{cases}$$

Different weight functions have been plotted in Figure 3. We have calculated the WLSEs

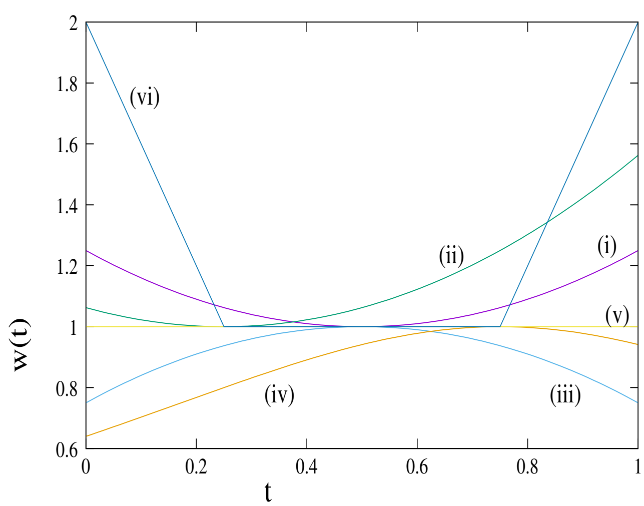

Figure 3: Different weight functions.

based on all these weight functions, and we have computed the residual sum of squares (RSS) in all these cases. The results are presented in Table 4. Among the different WLSEs, the estimators based on the weight function $w_4(t)$ provides the best fit in terms of smallest residual sum of squares (RSS).

Now based on the second method, we have obtained the weight function as the following four degree polynomial;

$$w_7(t) = 0.64 + 0.61t + 0.57t^2 - 1.17t^3 + 30t^4, \tag{12}$$

and the corresponding RSS is 3.9312, and it is quite close to the corresponding RSS of $w_4(t)$. We have provided the plots of $w_4(t)$ and $w_7(t)$ in Figure 4, and they are very similar in nature. It shows that both the methods can choose very similar weight functions in practice.

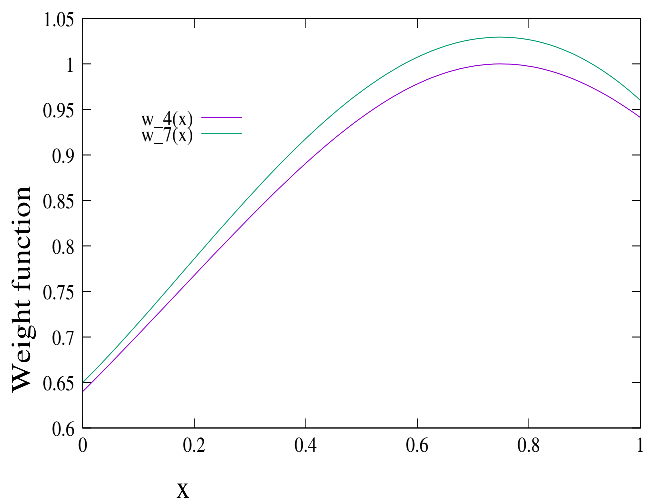

Figure 4: The weight functions based on two different methods.

We have provided the observed signal and estimated signal in Figure 5. It provides a good fit to the noisy data with outliers.

7 Conclusions

In this paper we have considered multiple sinusoidal model with impulsive noise with infinite variance. We have used a very novel weighted least squares method as a robust estimators of the amplitudes and frequencies. We have established the consistency properties of the proposed estimators. The proposed method can be used very conveniently even when the number of components is large, where the traditional robust methods like least absolute

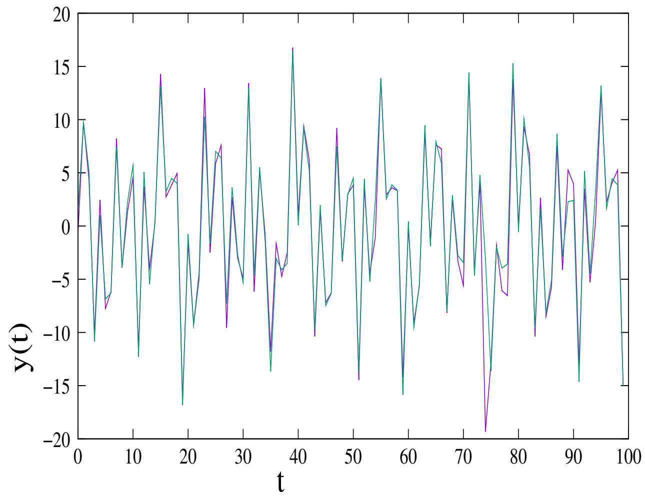

Figure 5: The observed and estimated signals.

deviation method of Huber's method cannot be applied very conveniently. Moreover, the asymptotic properties of these traditional methods are not known when the noise variance is not finite. We have shown how our method can be used in practice. Hence, for all practical purposes the proposed method can be used quite conveniently.

Acknowledgements:

The author would like to thank the unknown reviewers for their constructive comments which have helped to improve the manuscript significantly.

Data Availability and Conflict of Interest Statement:

No real data has been used in this manuscript. The author does not have any conflict of interest in the preparation of this manuscript.

Funding Statement:

No funding has been received to prepare this manuscript.

Appendix A

We need the following lemma for further development.

Lemma 1: If $\theta \in (0, \pi)$, then for $k = 0, 1, 2, \ldots$

$$\lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} \cos(\theta t) = \lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} \sin(\theta t) = 0$$

$$\lim_{n\to\infty} \frac{1}{n^{k+1}}\sum_{t=1}^{n} n^k \cos^2(\theta t) = \lim_{n\to\infty} \frac{1}{n^{k+1}}\sum_{t=1}^{n} n^k \sin^2(\theta t) = \frac{1}{2(k+1)}$$

$$\lim_{n\to\infty} \frac{1}{n^{k+1}}\sum_{t=1}^{n} n^k \cos(\theta t)\sin(\theta t) = 0.$$

We need the following lemmas to prove Theorem 1.

Lemma 2: Let $\{X(t)\}$ be a sequence of i.i.d. random variables with mean zero and $E|X(t)|^{1+\delta} < \infty$, for $0 < \delta < 1$, then for $m = 0, 1, 2, \ldots,$

$$\lim_{n\to\infty} \sup_{\theta} \left| \frac{1}{n^{m+1}} \sum_{t=1}^{n} X(t) t^m \cos(t\theta) \right| = 0 \quad a.s. \tag{13}$$

The same holds when $\cos(t\theta)$ is replaced by $\sin(t\theta)$.

Proof: In this proof the constant '$C > 0$' and it may not be same at different places. Let us define

$$V(t) = \begin{cases} X(t) & \text{if} \quad |X(t)| \le t^{1/(1+\delta)} \\ 0 & \text{if} \quad |X(t)| > t^{1/(1+\delta)}. \end{cases}$$

Then

$$\sum_{t=1}^{\infty} P(V(t) \neq X(t)) = \sum_{t=1}^{\infty} P(|X(t)| > t^{1/(1+\delta)}) = \sum_{t=1}^{\infty} \sum_{2^{t-1} \le k < 2^t} P(|X(1)| > k^{1/(1+\delta)})$$

$$
\begin{aligned}
&\le \sum_{t=1}^{\infty} 2^t P(|X(1)| > 2^{(t-1)/(1+\delta)}) = \sum_{t=1}^{\infty} 2^t \sum_{j=t}^{\infty} P(2^{(j-1)/(1+\delta)} \le |X(1)| < 2^{j/(1+\delta)}) \\
&= \sum_{j=1}^{\infty} P(2^{(j-1)/(1+\delta)} \le |X(1)| < 2^{j/(1+\delta)}) \sum_{t=1}^{j} 2^t \\
&\le C \sum_{j=1}^{\infty} 2^{j-1} P(2^{(j-1)/(1+\delta)} \le |X(1)| < 2^{j/(1+\delta)}) \\
&\le CE|X(1)|^{1+\delta} < \infty.
\end{aligned}
$$

Hence, $P(V(t) \neq X(t)\ i.o.) = 0$. Therefore

$$
\sup_{\theta} \left| \frac{1}{n^{m+1}} \sum_{t=1}^{n} X(t) t^m \cos(t\theta) \right| = 0 \quad a.s. \quad \Leftrightarrow \quad \sup_{\theta} \left| \frac{1}{n^{m+1}} \sum_{t=1}^{n} V(t) t^m \cos(t\theta) \right| = 0 \quad a.s.
$$

Let $W(t) = V(t) - E(V(t))$. Note that if $F(\cdot)$ denotes the distribution function of $X(t)$, then

$$
\begin{aligned}
\sup_{\theta} \left| \frac{1}{n^{m+1}} \sum_{t=1}^{n} E(V(t)) t^m \cos(t\theta) \right| &\le \frac{1}{n^{m+1}} \sum_{t=1}^{n} |E(V(t))| t^m \\
&\le \frac{1}{n} \sum_{t=1}^{n} |E(V(t))| = \frac{1}{n} \sum_{t=1}^{n} \left| \int_{-t^{1/(1+\delta)}}^{t^{1/(1+\delta)}} x dF(x) \right| \to 0.
\end{aligned}
$$

Thus to prove (13), it is enough to show that

$$
\sup_{\theta} \frac{1}{n^{m+1}} \sum_{t=1}^{n} W(t) t^m \cos(t\theta) = 0 \quad a.s. \tag{14}
$$

For any fixed θ and $\epsilon > 0$, let $0 \le h \le \dfrac{1}{2n^{1/(1+\delta)}}$, then we have

$$
\begin{aligned}
P\left\{ \left| \frac{1}{n^{m+1}} \sum_{t=1}^{n} W(t) t^m \cos(t\theta) \right| \ge \epsilon \right\} &= P\left\{ \left| \frac{1}{n} \sum_{t=1}^{n} W(t) \left(\frac{t}{n}\right)^m \cos(t\theta) \right| \ge \epsilon \right\} \\
&\le 2\exp(-hn\epsilon) \prod_{t=1}^{n} E\left(\exp\left\{ hW(t) \left(\frac{t}{n}\right)^m \cos(t\theta) \right\} \right).
\end{aligned}
$$

Since, $\left| hW(t) \left(\dfrac{t}{n}\right)^m \cos(t\theta) \right| \le \dfrac{1}{2}$, $\exp(x) \le 1 + x + 2|x|^{1+\delta}$ for $|x| \le \dfrac{1}{2}$ and $E|W(t)|^{1+\delta} \le C$, for some $C > 0$, therefore,

$$
\prod_{t=1}^{n} E\left(\exp\left\{ hW(t) \left(\frac{t}{n}\right)^m \cos(t\theta) \right\} \right) \le \prod_{t=1}^{n} (1 + 2Ch^{1+\delta}) \le \exp\left(2nCh^{1+\delta}\right).
$$

Choose $h = \dfrac{1}{2n^{1/(1+\delta)}}$, then for large n,

$$P\left\{\left|\frac{1}{n^{m+1}}\sum_{t=1}^{n} W(t)t^m\cos(t\theta)\right| \geq \epsilon\right\} \leq 2\exp\left(-\frac{\epsilon}{2}n^{\delta/(1+\delta)} + C\right) \leq C\exp\left(-\frac{\epsilon}{2}n^{\delta/(1+\delta)}\right).$$

Let $K = n^2$, choose $\theta_1, \ldots, \theta_K$, such that for each $\theta \in (0, \pi)$, we have a θ_j satisfying $|\theta_j - \theta| \leq \dfrac{\pi}{n^2}$. Therefore,

$$\begin{aligned}\frac{1}{n^{m+1}}\left|\sum_{t=1}^{n} W(t)t^m(\cos(t\theta) - \cos(t\theta_j))\right| &\leq \frac{1}{n}\sum_{t=1}^{n}|W(t)|\left(\frac{t}{n}\right)^m|\cos(t\theta) - \cos(t\theta_j)| \\ &\leq C\frac{1}{n}\sum_{t=1}^{n} t^{1/(1+\delta)}t\left(\frac{\pi}{n^2}\right) \leq C\pi n^{-\delta/(1+\delta)} \to 0.\end{aligned}$$

Therefore, for large n

$$\begin{aligned}P\left\{\sup_{0<\theta<\pi}\left|\frac{1}{n^{m+1}}\sum_{t=1}^{n} W(t)t^m\cos(t\theta)\right| \geq 2\epsilon\right\} &\leq P\left\{\max_{1\leq j\leq n^2}\left|\frac{1}{n^{m+1}}\sum_{t=1}^{n} W(t)t^m\cos(t\theta_j)\right| \geq \epsilon\right\} \\ &\leq Cn^2\exp\left(-\frac{\epsilon}{2}n^{\delta/(1+\delta)}\right).\end{aligned}$$

Since $\displaystyle\sum_{n=1}^{\infty} Cn^2\exp\left(-\frac{\epsilon}{2}n^{\delta/(1+\delta)}\right) < \infty$, (14) is proved using Borel-Cantelli lemma.

Lemma 3: Let $\{Z(t)\}$ be a sequence of i.i.d. Gaussian random variables with mean zero and finite variance, then for $m = 0, 1, 2, \ldots$,

$$\lim_{n\to\infty}\sup_{\theta}\left|\frac{1}{n^{m+1}}\sum_{t=1}^{n} Z(t)t^m\cos(t\theta)\right| = 0 \quad a.s.$$

Proof: It follows from Lemma 2.

Lemma 4: Let $\{e(t)\}$ be a sequence of i.i.d. random variables satisfying Assumption 2, and $w(t)$ is a m-degree polynomial satisfying (3), then

$$\lim_{n\to\infty}\sup_{\theta}\left|\frac{1}{n}\sum_{t=1}^{n} e(t)w\left(\frac{t}{n}\right)\cos(t\theta)\right| = 0 \quad a.s.$$

Proof: It follows from Lemma 2 and Lemma 3.

Lemma 5: Let $\{e(t)\}$ be a sequence of i.i.d. random variables satisfying Assumption 2, and $w(t)$ is a m-degree polynomial satisfying (3), then for any $\delta > 0$,

$$\liminf_{n\to\infty} \inf_{|\boldsymbol{\theta}^0 - \boldsymbol{\theta}| > \delta} \frac{1}{n}\left\{Q(\boldsymbol{\theta}) - Q(\boldsymbol{\theta}^0)\right\} > 0, \qquad a.s,$$

here $\boldsymbol{\theta} = (A, B, \theta)$, $\boldsymbol{\theta}^0 = (A^0, B^0, \theta^0)$ and $Q(\boldsymbol{\theta})$ is same as defined in (5).

Proof: Using Lemma 4, and following the same procedure as in the proof of Lemma 3 in Kundu [2], it can be proved. The details are avoided.

Proof of Theorem 1: In this proof only let us denote $\widehat{\boldsymbol{\theta}}$ by $\widehat{\boldsymbol{\theta}}_n$, just to emphasize that it is a function of n. Suppose $\widehat{\boldsymbol{\theta}}_n$ is not a consistent estimator $\boldsymbol{\theta}^0$, then there exists a subsequence $\{n_k\}$ of $\{n\}$, such that $\widehat{\boldsymbol{\theta}}_{n_k} \to \boldsymbol{\theta}' \neq \boldsymbol{\theta}^0$. Since, for all $k \geq 1$

$$\frac{1}{n_k} Q\left(\widehat{\boldsymbol{\theta}}_{n_k}\right) \leq \frac{1}{n_k} Q\left(\widehat{\boldsymbol{\theta}}^0\right),$$

there exists a $\delta > 0$, such that

$$\liminf_{n\to\infty} \inf_{|\boldsymbol{\theta}^0 - \boldsymbol{\theta}| > \delta} \frac{1}{n}\left\{Q(\boldsymbol{\theta}) - Q(\boldsymbol{\theta}^0)\right\} \leq 0, \qquad a.s.$$

It contradicts Lemma 5. Hence, $\widehat{\boldsymbol{\theta}}_n$ is a consistent estimator $\boldsymbol{\theta}^0$.

Appendix B

Proof of Theorem 2:

Let us consider $Q(\boldsymbol{\theta})$ as defined in (5), and we use the following notations:

$$Q'(\boldsymbol{\theta}) = \left(\frac{\partial Q(\boldsymbol{\theta})}{\partial A}, \frac{\partial Q(\boldsymbol{\theta})}{\partial B}, \frac{\partial Q(\boldsymbol{\theta})}{\partial \omega}\right),$$

and $Q''(\boldsymbol{\theta})$ is the 3×3 matrix of the second derivatives of $Q(\boldsymbol{\theta})$. Now expanding $Q'(\widehat{\boldsymbol{\theta}})$ around $\boldsymbol{\theta}^0$ by multivariate Taylor series, we obtain

$$Q'(\widehat{\boldsymbol{\theta}}) - Q'(\boldsymbol{\theta}_0) = (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) Q''(\overline{\boldsymbol{\theta}}), \tag{15}$$

where $\overline{\boldsymbol{\theta}}$ is a point on the line joining $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}^0$. Since $Q'(\widehat{\boldsymbol{\theta}}) = \mathbf{0}$, (15) can be written as

$$-Q'(\boldsymbol{\theta}_0)\boldsymbol{D}_1 = (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\boldsymbol{D}_2^{-1}\boldsymbol{D}_2 Q''(\overline{\boldsymbol{\theta}})\boldsymbol{D}_1. \tag{16}$$

Since, $\widehat{\boldsymbol{\theta}}$ converges to $\boldsymbol{\theta}^0$ a.s. from Theorem 1, and $Q''(\boldsymbol{\theta})$ is a continuous function of $\boldsymbol{\theta}$, using continuous mapping theorem and Lemma 1, it follows that

$$\lim_{n\to\infty} \boldsymbol{D}_2 Q''(\overline{\boldsymbol{\theta}})\boldsymbol{D}_1 = \lim_{n\to\infty} \boldsymbol{D}_2 Q''(\boldsymbol{\theta}^0)\boldsymbol{D}_1 = \boldsymbol{\Gamma}. \tag{17}$$

Since $\boldsymbol{\Gamma}$ is a full rank matix, for large n, (16) can be written as

$$-Q'(\boldsymbol{\theta}_0)\boldsymbol{D}_1[\boldsymbol{D}_2 Q''(\overline{\boldsymbol{\theta}})\boldsymbol{D}_1]^{-1} = (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\boldsymbol{D}_2^{-1}. \tag{18}$$

We want to show that $Q'(\boldsymbol{\theta}_0)\boldsymbol{D}_1$ converges to a multivariate stable distribution if $0 < \pi \le 1$. It is assumed that, $e(t)$ follows Assumption 2. Let us write $Q'(\boldsymbol{\theta}_0)\boldsymbol{D}_1 = (U_{1n}, U_{2n}, U_{3n})$, where

$$\begin{aligned}
U_{1n} &= \frac{2}{\pi n^{1/\alpha}} \sum_{t=1}^{n} w\left(\frac{t}{n}\right)((1-\pi)Z(t) + \pi X(t))\cos(\theta^0 t) \\
U_{2n} &= \frac{2}{\pi n^{1/\alpha}} \sum_{t=1}^{n} w\left(\frac{t}{n}\right)((1-\pi)Z(t) + \pi X(t))\sin(\theta^0 t) \\
U_{3n} &= \frac{2}{\pi n^{(1+\alpha)/\alpha}} \sum_{t=1}^{n} w\left(\frac{t}{n}\right) t((1-\pi)Z(t) + \pi X(t))(A^0\sin(\theta^0 t) - B^0\cos(\theta^0 t)).
\end{aligned}$$

First observe that if we denote

$$\begin{aligned}
V_{1n} &= \frac{2}{n^{1/\alpha}} \sum_{t=1}^{n} w\left(\frac{t}{n}\right) X(t)\cos(\theta^0 t) \\
V_{2n} &= \frac{2}{n^{1/\alpha}} \sum_{t=1}^{n} w\left(\frac{t}{n}\right) X(t)\sin(\theta^0 t) \\
V_{3n} &= \frac{2}{n^{(1+\alpha)/\alpha}} \sum_{t=1}^{n} w\left(\frac{t}{n}\right) tX(t)(A^0\sin(\theta^0 t) - B^0\cos(\theta^0 t)),
\end{aligned}$$

then $(U_{1n} - V_{1n}, U_{2n} - V_{2n}, U_{3n} - V_{3n})$ converges to $(0,0,0)$ in probability. Hence, asymptotically the joint distribution of (U_{1n}, U_{2n}, U_{3n}) is same as the joint distribution of (V_{1n}, V_{2n}, V_{3n}). If $\boldsymbol{t} = (t_1, t_2, t_3)$, then the joint characteristic function of (V_{1n}, V_{2n}, V_{3n}) is

$$\phi_n(\boldsymbol{t}) = Ee^{it_1V_{1n} + it_2V_{2n} + it_3V_{3n}} = e^{-2^\alpha\beta^\alpha g_n(\boldsymbol{t})},$$

where $g_n(\boldsymbol{t}) = \frac{1}{n}\sum_{j=1}^{n} |K_j(\boldsymbol{t})|^{\alpha}$, and

$$K_j(\boldsymbol{t}) = w\left(\frac{j}{n}\right)\left(-t_1\cos(j\theta^0) - t_2\sin(j\theta^0) + \frac{jt_3}{n}(A^0\sin(j\theta^0) - B^0\cos(j\theta^0))\right).$$

Let $\lim_{n\to\infty} g_n(\boldsymbol{t}) = g_0(\boldsymbol{t})$. We will show that $g_0(\boldsymbol{t}) > 0$. First observe that for $\boldsymbol{t} \neq \mathbf{0}$ and $1 \le j \le n$,

$$|K_j(\boldsymbol{t})| \le K\left(|t_1| + |t_2| + |t_3|(|A^0| + |B^0|)\right) = S \quad \text{(say)}.$$

Here 'K' is same as defined in Section 2. Therefore, $|K_j(\boldsymbol{t})/S| \le 1$, for all $1 \le j \le n$. Hence, for $1 < \alpha < 2$,

$$|K_j(\boldsymbol{t})/S|^{\alpha} \ge |K_j(\boldsymbol{t})/S|^2 \quad \Rightarrow \quad |K_j(\boldsymbol{t})|^{\alpha} \ge S^{\alpha-2}|K_j(\boldsymbol{t})|^2.$$

Using Lemma 1, it follows that $\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}|K_j(\boldsymbol{t})|^2 > 0$. Hence, $g_0(\boldsymbol{t}) > 0$. Since, any linear combination of V_{1n}, V_{2n} and V_{3n} converges to a stable distribution, therefore, (V_{1n}, V_{2n}, V_{3n}) converges to a multivariate stable distribution, which has the characteristic function $\phi_0(\boldsymbol{t}) = e^{-2^{\alpha}\beta^{\alpha}g_0(\boldsymbol{t})}$. Hence, (U_{1n}, U_{2n}, U_{3n}) also converges to a multivariate stable distribution, which has the same characteristic function $\phi_0(\boldsymbol{t}) = e^{-2^{\alpha}\beta^{\alpha}g_0(\boldsymbol{t})}$.

Appendix C

To prove Theorem 3, we need the following Lemma.

Lemma 6: Suppose $\{e(t)\}$ be a sequence of i.i.d. random variables satisfying Assumption 2, and $w(t)$ is a m-degree polynomial satisfying (3). Let us denote

$$S_{1c} = \{\boldsymbol{\theta} : \boldsymbol{\theta} = (A, B, \theta)^{\top}, |\boldsymbol{\theta}^0 - \boldsymbol{\theta}| > 3c\}.$$

If there exists $c > 0$, such that

$$\liminf \inf_{\boldsymbol{\theta}\in S_{1c}} \frac{1}{n}[Q_1(\boldsymbol{\theta}) - Q_1(\boldsymbol{\theta}^0)] > 0, \qquad a.s,$$

$\widehat{\boldsymbol{\theta}}_1$ is a strongly consistent estimator of $\boldsymbol{\theta}_1^0$.

Proof: Using Lemma 4, following the same procedure as in Lemma 5, it can be proved by contradiction. The details are avoided.

Proof of Theorem 3: It follows similarly as the proof of Theorem 1, where $Q(\cdot)$ is replaced by $Q_1(\cdot)$.

References

[1] Abrar, S., Zerguine, A., Abed-Meraim, K. (2022), "Adaptive algorithms for blind channel equalization in impulsive noise", *Signal Processing*, vol. 201, 108626.

[2] Kundu, D. (2024), "On robust method of estimation in polynomial sinusoidal regression", *Signal Processing*, vol. 217, Paper No. 109328.

[3] Lee, D.D. and Kashyap, R.L. (1992), "Robust maximum likelihood bearing estimation in contaminated Gaussian noise", *IEEE Transactions of Signal Processing*, vol. 40, 1983 - 1986.

[4] Lin, L., Yu, X, Jiang, Q., Zang, B. and Jiang. L. (2022), "Synchrosqueezing transform meets α-stable distribution: An adaptive fractional lower-order SST for instantaneous frequency estimation and non-stationary signal recovery", *Signal Proccssing*, vol. 201, Paper no. 108683.

[5] Mandelbrot, B.B. (1963), "The variation of certain speculative prices", *Journal of Business*, Vol. 36, 394-419.

[6] Nandi, S., Iyer, S. K. and Kundu, D. (2002), "Estimating the Frequencies in Presence of Heavy Tail Errors", *Statistics and Probability Letters*, vol. 58, 265–282.

[7] Nandi, S., Grover, R., Kundu, D. (2021), "Estimation of Parameters of Multiple Chirp Signal in Presence of Additive Alpha-Stable Errors", *Signal Processing*, vol.189, Paper no. 108232.

[8] Nandi, S. and Kundu, D. (2020), *Statistical Signal Processing: Frequency Estimation*, Second edition, Springer, Singapore, ISBN: 978-981-15-6279-2; 978-981-15-6280-8.

[9] Nikias, C.L. and Shao, M. (1994), "Recent Advances in Signal Processing with α-Stable Distributions", *IFAC Proceedings Volumes*, Vol. 27, 65–70.

[10] Samorodnitsky, G and Taqqu, M.S. (1994), *Stable Non-Gaussian Random Processes; Stochastic Models with Infinite Variance*, Chapman and Hall, New York.

[11] Stoica,P. and Moses, R.L. (2005), *Spectral Analysis of Signals*, Pearson Prentice Hall, Upper Saddle River, New Jersey, U.S.A.

[12] Tzagkarakis, G., Nolan, J.P., Tsakalides, P. (2021), "Robust nonlinear compressive sampling using symmetric alpha-stable distributions", *Signal Processing*, vol. 182, paper no. 107944.

[13] Wang, M., Feng, D-Z., Chen, M-H., Su, T-T., Zhang, X-J., (2023), "A robust direction-of-arrival estimation method for impulsive noise environments", *Signal Processing*, vol. 212, paper no. 109175.

[14] Yue, B. and Peng, Z. (2015), "A validation study of α-stable distribution characteristic for seismic data", *Signal Processing*, vol. 106, 1-9.

[15] Zhou, Z., Huang, L., Christensen, M.G. and Zhang, S. (2022), "Robust spectral analysis of multi-channel sinusoidal signals in impulsive noise environment", *IEEE Transactions of Signal Processing*, vol. 70, 919 - 935.

[16] Zoubir, A.M., Koivunen, V., Chakhchoukh, Y. and Muma, M., (2012), "Robust estimation in signal processing: A tutorial type treatment of fundamental concepts", *IEEE Signal Processing Magazine*, vol. 29, 61 – 80.