RC24446 Revised (W0809-045) September 12, 2008
Mathematics

IBM Research Report

Estimation of the Parameters of Sinusoidal Signals in Non-Gaussian Noise

Ta-Hsin Li
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598
USA
(thl@us.ibm.com)

Kai-Sheng Song
Department of Mathematics
University of North Texas
Denton, TX 76203-1430
USA
(ksong@unt.edu)

Abstract

Accurate estimation of the amplitude and frequency parameters of sinusoidal signals from noisy observations is an important problem in many signal processing applications. In this paper, the problem is investigated under the assumption of non-Gaussian noise in general and Laplace noise in particular. It is proven mathematically that the maximum likelihood estimator derived under the condition of Laplace white noise is able to attain an asymptotic Cramér-Rao lower bound which is one half of that achieved by periodogram maximization and nonlinear least squares. It is also proven that when applied to non-Laplace situations, the Laplace maximum likelihood estimator, which may also be referred to as the nonlinear least-absolute-deviations estimator, can achieve an even higher statistical efficiency especially when the noise distribution has heavy tails. A computational procedure is proposed to overcome the difficulty of local extrema in the likelihood function. Simulation results are provided to validate the analytical findings.

Keywords: Frequency estimation, harmonic retrieval, heavy tail, impulsive noise, Laplace distribution, least absolute deviation, robust, spectral analysis.

1 Introduction

Consider the problem of estimating the parameter $\boldsymbol{\theta} := [A, B, \omega]^T$ of a sinusoidal signal from a time series of noisy observations

$$y_t := A\cos(\omega t) + B\sin(\omega t) + \varepsilon_t \quad (t = 1, \ldots, n), \tag{1}$$

where $A \in \mathbb{R}$, $B \in \mathbb{R}$, and $\omega \in \Omega := (0, \pi)$ are unknown constants and $\{\varepsilon_t\}$ is a white noise process with zero mean and unknown finite variance $\sigma^2 > 0$. The literature on this subject is extensive [1] [2]. Among the most popular approaches to this problem are Fourier transform (periodogram) [3]–[6], Gauss maximum likelihood (also known as nonlinear least squares) [7]–[9], autoregression [10]–[13], and eigen-decomposition (signal/noise subspace) [14]–[19].

It is well known that if $\{\varepsilon_t\}$ is Gaussian white noise (GWN), then the asymptotic Cramér-Rao lower bound (CRLB) for the frequency parameter ω can be expressed as $(12/\gamma)\,n^{-3}$, where $\gamma := \frac{1}{2}(A^2 + B^2)/\sigma^2$ is the signal-to-noise ratio (SNR). We shall refer to this bound as the Gauss CRLB. A number of analytical and simulation studies suggest [5] [7] that the Gauss CRLB can be attained asymptotically by the maximum likelihood estimator (MLE) that maximizes the Gauss likelihood function, or equivalently, that minimizes the sum of squared errors

$$\ell_2(\boldsymbol{\vartheta}) := \sum_{t=1}^{n} |y_t - (\vartheta_1 \cos(\vartheta_3 t) + \vartheta_2 \sin(\vartheta_3 t))|^2, \tag{2}$$

where $\boldsymbol{\vartheta} := [\vartheta_1, \vartheta_2, \vartheta_3]^T \in \Theta_0 := \mathbb{R} \times \mathbb{R} \times \Omega$. Rigorous proofs of this assertion are provided recently in [20] and [21]. Several numerical procedures have been proposed to compute the estimator [7]–[9].

The minimizer of $\ell_2(\boldsymbol{\vartheta})$ in (2) is known as the nonlinear least-squares (NLS) estimator when applied to non-Gaussian situations. It can be shown [20] [21] that the NLS estimator attains the Gauss CRLB asymptotically for any white noise process, Gaussian or non-Gaussian, with zero mean and finite variance. Typical estimators in the literature either reach the Gauss CRLB asymptotically (e.g., NLS and periodogram maximization) or fall short of it (e.g., the signal/noise subspace methods). An interesting question is: can we do better than the Gauss CRLB under non-Gaussian conditions?

In this paper, we provide an affirmative answer to this question. Toward that end, we first examine the CRLB under non-Gaussian noise and show that the Gauss CRLB is the worse-case performance limit, namely, the largest lower bound, among a large family of noise distributions. We then focus on the special case of Laplace white noise (LWN) and prove that the Laplace MLE attains the Laplace CRLB asymptotically which is only one half of the Gauss CRLB.

The Gaussian assumption is often made in practice not because of its fitness to noise data but because of its mathematical tractability. In reality, departures from the Gaussian assumption can occur in many different forms, one of which is in the form of heavy tails. A heavy-tailed distribution has greater tail probabilities than suggested by the Gaussian model. It manifests itself in practice as impulsive errors and outliers in the observations, capable of causing algorithms developed under the Gaussian assumption to malfunction. The Laplace (or double exponential) distribution is an example of heavy-tailed distributions. This model has been employed to describe impulsive noise as well as to serve as a surrogate in developing robust algorithms against outliers and in solving problems that have no solution under the Gaussian assumption [22]–[24]. The alpha-stable distribution considered in [25]–[27] is another popular model for heavy-tailed noise.

Like the Gauss MLE, the Laplace MLE is difficult to compute without an extremely good initial guess, because the likelihood function has numerous local extrema in the vicinity of the desired solution. To obtain the initial value, we use an iterative filtering method called the three-step algorithm (TSA) [28] [29]. In addition to its unified architecture suitable for practical implementation, the TSA has the analytically proven property of fast and virtually global convergence to an estimate as accurate as the Gauss MLE. We present some simulation results to confirm the validity of the TSA as an initialization procedure for the Laplace MLE.

We also provide an asymptotic analysis of the Laplace MLE when it is applied to non-Laplace situations, where the estimator will be referred to as nonlinear least absolute deviations (NLAD) in analogy to the Gauss MLE being referred to as nonlinear least squares when applied to non-Gaussian cases. It is shown that the NLAD estimator has an asymptotic normal distribution, just like the NLS estimator, but the asymptotic variance of the NLAD estimator can be much smaller than the Gauss CRLB as well as the Laplace CRLB, especially for heavy-tailed noise. This result justifies the NLAD estimator as a superior alternative to the NLS estimator under non-Laplace heavy-tailed conditions.

The rest of the paper is organized as follows. Section II discusses the CRLB under non-Gaussian conditions; Section III contains the result regarding the asymptotic distribution of the Laplace MLE; Section IV focuses on the computational issues; Section V provides an asymptotic analysis of the Laplace MLE under non-Laplace conditions; and Section VI demonstrates an application of the asymptotic results in constructing confidence regions for the amplitude and frequency parameters. Concluding remarks are given in Section VII.

2 CRLB Under Non-Gaussian Noise

In this paper, we always assume that $\{\varepsilon_t\}$ is a sequence of independent and identically distributed (i.i.d.) random variables with a common probability distribution function $F(x)$ such that $E(\varepsilon_t) = \int x\,dF(x) = 0$ and $0 < \sigma^2 := \mathrm{Var}(\varepsilon_t) = \int x^2\,dF(x) < \infty$.

Proposition 1. *If F is twice differentiable everywhere except a finite number of points and satisfies $0 < \lambda_F := \sigma^2 \int \{\ddot{F}(x)/\dot{F}(x)\}^2\,dF(x) < \infty$, where $\dot{F}$ and $\ddot{F}$ denote the first and second derivatives of F, then the Fisher information matrix of $\boldsymbol{y} := [y_1, \ldots, y_n]^T$ with respect to the parameter $\boldsymbol{\theta}$ can be expressed as $I_F(\boldsymbol{\theta}) = \lambda_F\, I_G(\boldsymbol{\theta})$, where $I_G(\boldsymbol{\theta}) := (1/\sigma^2) X^T X$ is the Fisher information matrix under the GWN assumption, with $X := [\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3]$, $\boldsymbol{x}_1 := \mathrm{vec}[\cos(\omega t)]$, $\boldsymbol{x}_2 := \mathrm{vec}[\sin(\omega t)]$, and $\boldsymbol{x}_3 := \mathrm{vec}[-At\sin(\omega t) + Bt\cos(\omega t)]$.*

Proof. The log-likelihood function of $\boldsymbol{y}$ can be written as $L(\boldsymbol{\theta}|\boldsymbol{y}) = \sum \log \dot{F}(y_t - s_t(\boldsymbol{\theta}))$, where $s_t(\boldsymbol{\theta}) := A\cos(\omega t) + B\sin(\omega t)$. Therefore, $\partial L/\partial A = -\sum \cos(\omega t)\ddot{F}(\varepsilon_t)/\dot{F}(\varepsilon_t)$, $\partial L/\partial B = -\sum \sin(\omega t)\ddot{F}(\varepsilon_t)/\dot{F}(\varepsilon_t)$, and $\partial L/\partial \omega = -\sum\{-At\sin(\omega t) + Bt\cos(\omega t)\}\,\ddot{F}(\varepsilon_t)/\dot{F}(\varepsilon_t)$. This yields $I_F(\boldsymbol{\theta}) := E\{\nabla L(\boldsymbol{\theta}|\boldsymbol{y})\,\nabla^T L(\boldsymbol{\theta}|\boldsymbol{y})\} = E\{(\ddot{F}(\varepsilon_1)/\dot{F}(\varepsilon_1))^2\}\,(X^T X) = (\lambda_F/\sigma^2)\,(X^T X)$. If F is Gaussian, then $\lambda_F = 1$. □

From Proposition 1, the following result can be obtained.

Proposition 2. *Let the conditions in Proposition 1 be satisfied. Assume further that the support of $\dot{F}$ is equal to $\mathbb{R}$ and $\ddot{F}(x)$ is continuous for almost every $x \in \mathbb{R}$. Then, the Cramér-Rao inequality holds for any unbiased estimator of $\boldsymbol{\theta}$ on the basis of $\boldsymbol{y}$, and the CRLB can be expressed as $CRLB_F(\boldsymbol{\theta}) = \lambda_F^{-1}\,CRLB_G(\boldsymbol{\theta})$, where $CRLB_G(\boldsymbol{\theta}) := I_G^{-1}(\boldsymbol{\theta}) = \sigma^2(X^T X)^{-1}$ is the CRLB under the GWN assumption, satisfying $D_n^{-1} CRLB_G(\boldsymbol{\theta})\,D_n^{-1} = \boldsymbol{\Sigma}^{-1} + \mathcal{O}(n^{-1})$ for large n, with*

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{\gamma}\begin{bmatrix} A^2+4B^2 & -3AB & -6B \\ & 4A^2+B^2 & 6A \\ \textit{symmetry} & & 12 \end{bmatrix}$$

and $D_n := \mathrm{diag}(n^{-1/2}, n^{-1/2}, n^{-3/2})$.

Proof. By the fundamental theorem of calculus for Lebesgue integration, we can write $\sqrt{\dot{F}(x+h)} - \sqrt{\dot{F}(x)} = \int_0^1 \frac{1}{2} h\ddot{F}(x+uh)/\sqrt{\dot{F}(x+uh)}\,du$. It follows that $\sqrt{\dot{F}(x)}$ is differentiable in quadratic mean ([30], pp. 95–96). Moreover, by Proposition 2.29 in [30] (p. 22), $\int |\ddot{F}(x+h)/\sqrt{\dot{F}(x+h)} - \ddot{F}(x)/\sqrt{\dot{F}(x)}\,|^2\,dx \to 0$ as $h \to 0$. Finally, $(1/\sigma^2)X^T X = D_n^{-1}\{\boldsymbol{\Sigma} + \mathcal{O}(n^{-1})\}\,D_n^{-1}$, where

$$\boldsymbol{\Sigma} := \frac{1}{\sigma^2}\begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{4}B^2 \\ & \frac{1}{2} & -\frac{1}{4}A^2 \\ \text{symmetry} & & \frac{1}{6}(A^2+B^2) \end{bmatrix}. \tag{3}$$

Combining these results with the assumption that the support of $\dot{F}(x)$ is equal to $\mathbb{R}$ proves that the regularity conditions in [32] (p. 65) are satisfied. The assertion follows from Proposition 1 and the Cramér-Rao inequality ([32], Theorem 7.3, p. 73). □

The next proposition asserts that the Gauss CRLB is the worse-case performance limit among a large family of noise distributions.

Proposition 3. *Let $\mathcal{F}$ be the collection of probability distributions that satisfy the assumptions in Proposition 2. Then, it follows that $\lambda_F \geq 1$, hence $CRLB_F(\boldsymbol{\theta}) \leq CRLB_G(\boldsymbol{\theta})$, for all $F \in \mathcal{F}$, where the equality holds if and only if F is Gaussian.*

Proof. Consider the problem of estimating $\theta \in \mathbb{R}$ on the basis of $Y \sim F(y-\theta)$. It is easy to show that the Fisher information of Y equals λ_F/σ^2 and Y is an unbiased estimator of θ with $\mathrm{Var}(Y) = \sigma^2$. So, the Cramér-Rao inequality can be written as $\lambda_F \geq 1$, with "=" iff $a(y-\theta) = (d/d\theta)\log \dot{F}(y-\theta)$ for $y \in \mathbb{R}$, where a is nonzero and independent of y ([32], p. 77). With $x := y-\theta$, this condition can be rewritten as $(d/dx)\log \dot{F}(x) = -ax$, leading to $\dot{F}(x) = \exp(-\frac{1}{2}ax^2 + b)$. Imposing the constraint on the variance and the fact that $\dot{F}(x)$ integrates to unity completes the proof. □

Remark 1. A similar conclusion was drawn in [31] for estimating the parameters of non-Gaussian autoregressive processes. For more results under the general signal-plus-noise models, see [32] (Chapt. II).

Remark 2. According to Proposition 3, th quantity λ_F can be regarded as a measure of deviation from the Gaussian distributions. It is a dimensionless quantity, invariant to the scale of the noise.

Remark 3. Because $\mathrm{CRLB}_G(\boldsymbol{\theta}) = \sigma^2(X^TX)^{-1}$, Proposition 3 can be easily extended to cover a larger family of distributions in which σ^2 is the upper bound of the variance. Proposition 3 is related to Huber's minimax approach in robust statistics [33], where the Gaussian distribution with mean zero and variance σ^2 is known to be the least favorable distribution, namely, one that minimizes the Fisher information [34].

An example of F that satisfies the assumptions is the Laplace distribution with

$$\dot{F}(x) = (2c)^{-1}\exp\big(-|x|/c\big), \tag{4}$$

where $c := \sigma/\sqrt{2}$. The Laplace distribution has higher concentration around zero as well as heavier tails than the Gaussian distribution with the same variance. For the Laplace distribution, $\lambda_F = 2$. This, combined

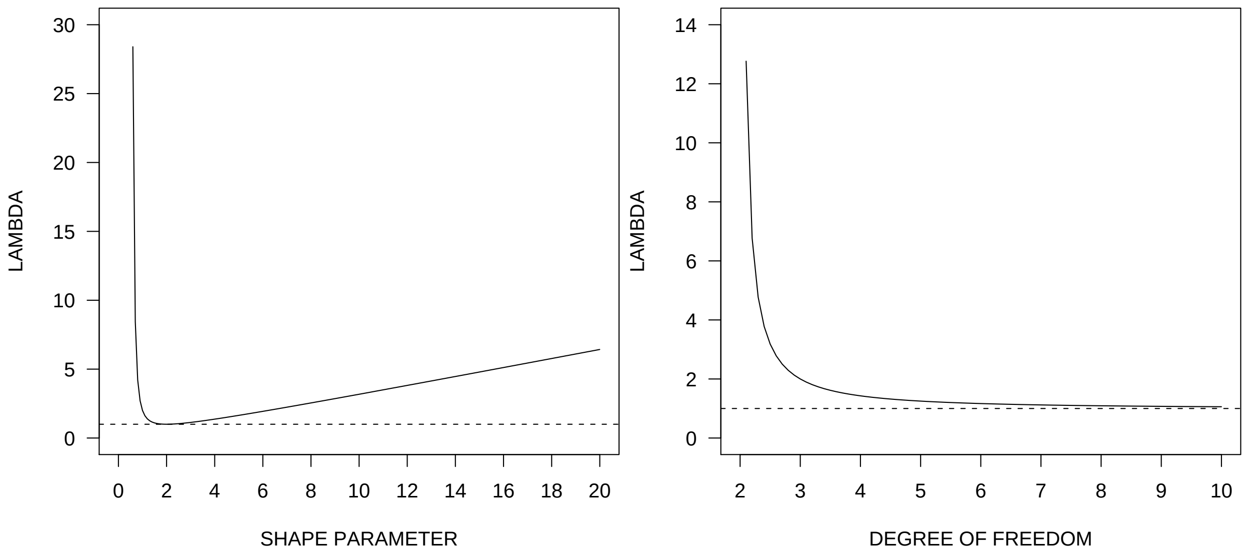

Figure 1: (a) Plot of λ_F as a function of the shape parameter q for the generalized Gaussian distributions in (5). (b) Plot of λ_F as a function of the degree of freedom ν for the T distributions in (6). Dashed line represents the lower bound $\lambda = 1$.

with Proposition 2, implies that $\mathrm{CRLB}_L(\boldsymbol{\theta}) = \frac{1}{2}\mathrm{CRLB}_G(\boldsymbol{\theta})$, where $\mathrm{CRLB}_L(\boldsymbol{\theta})$ denotes the CRLB under the Laplace assumption, or the Laplace CRLB.

The Laplace distribution in (4) belongs to the family of generalized Gaussian distributions with

$$\dot{F}(x) = \frac{q}{2c\,\Gamma(1/q)} \exp(-|x/c|^q) \tag{5}$$

where $q > 0$ and $c > 0$ are the shape and scale parameters. For these distributions, it can be shown that $\lambda_F = q^2\,\Gamma(2-1/q)\,\Gamma(3/q)/\{\Gamma(1/q)\}^2$. Fig. 1(a) depicts λ_F as a function of q. As can be seen, λ_F is minimized at $q = 2$ which corresponds to the Gaussian distribution.

Another example is the family of Student's T distributions (denoted by T_ν) with

$$\dot{F}(x) = \frac{\Gamma((\nu+1)/2)}{c\,\Gamma(\nu/2)\sqrt{\pi\nu}}\{1+(x/c)^2/\nu\}^{-(\nu+1)/2}, \tag{6}$$

where ν is the degree-of-freedom parameter. To ensure a finite variance, it is required that $\nu > 2$. In this case, one can show that $\lambda_F = \nu(\nu+1)/\{(\nu-2)(\nu+3)\}$. Fig. 1(b) depicts λ_F as a function of ν. Note that

$\lambda_F > 1$ for all $\nu > 2$ and $\lambda_F \to 1$ as $\nu \to \infty$. This limiting value is a manifestation that the T_ν distribution, if properly scaled, converge to the standard Gaussian distribution as $\nu \to \infty$.

3 Maximum Likelihood Estimation

In this section, we focus on the estimation problem under the condition of Laplace white noise (LWN). As mentioned in the Introduction, typical frequency estimators in the literature can only achieve the Gauss CRLB at best regardless of the noise distribution. The Laplace CRLB discussed in Section II suggests the possibility of reducing the estimation error by 50% when the noise actually has a Laplace distribution. The key question is whether this reduced CRLB can be achieved at all, and if so, by what means.

To answer this question, we turn to the maximum likelihood method, which typically produces asymptotically efficient estimates. Under the LWN assumption, maximizing the Laplace likelihood function is equivalent to minimizing the sum of absolute deviations

$$\ell_1(\boldsymbol{\vartheta}) := \sum_{t=1}^{n} |y_t - (\vartheta_1 \cos(\vartheta_3 t) + \vartheta_2 \sin(\vartheta_3 t))|. \tag{7}$$

Therefore, instead of minimizing the ℓ_2 error in (2), we propose to minimize the ℓ_1 error in (7).

To prove the efficiency of this estimator mathematically is not a simple exercise. We cannot adopt the standard argument used to prove the asymptotic normality of the maximum likelihood estimators from i.i.d. observations [30], not only because the y_t do not have the same distribution for each t, but also because the Laplace likelihood function is not everywhere differentiable and the rate of convergence is not the same for all parameters. Fortunately, by employing more sophisticated mathematical tools, we are able to prove the following result (see Appendix I for proof).

Theorem 1. *Let $\{y_t\}$ be given by (1), where $\{\varepsilon_t\}$ is an LWN process with zero mean and finite variance $\sigma^2 > 0$. Let $\Theta_n := \{\boldsymbol{\vartheta} \in \Theta_0 : \|D_n^{-1}(\boldsymbol{\vartheta} - \boldsymbol{\theta})\| \le \kappa n^\alpha\}$ be a neighborhood of $\boldsymbol{\theta}$, where $\kappa > 0$ and $\alpha \in (0, \frac{1}{6})$ are some constants. Then, as $n \to \infty$, $\hat{\boldsymbol{\theta}}_n := \arg\min\{\ell_1(\boldsymbol{\vartheta}) : \boldsymbol{\vartheta} \in \Theta_n\}$ is asymptotically distributed as*

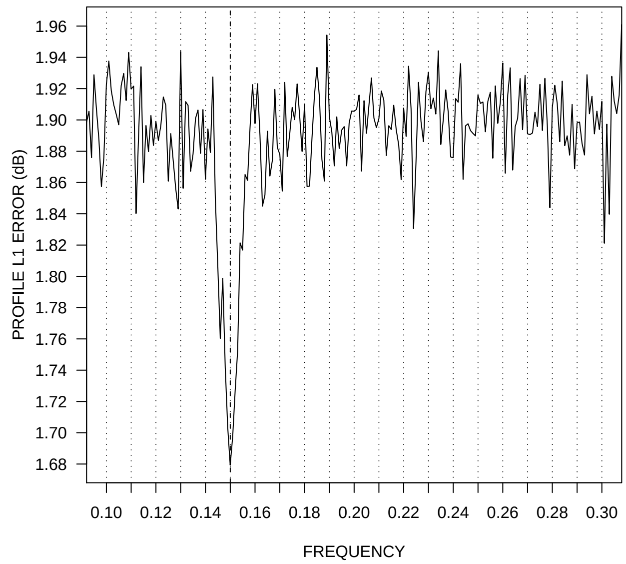

Figure 2: Plot of profile ℓ_1 error as a function of f for a noisy sinusoid of length $n = 100$. True frequency is $f = 0.15$ (dashed line). Dotted vertical lines represent the Fourier frequencies k/n $(k = 1, 2, \ldots)$.

$N(\boldsymbol{\theta}, CRLB_L(\boldsymbol{\theta}))$, *where* $CRLB_L(\boldsymbol{\theta}) = \frac{1}{2} CRLB_G(\boldsymbol{\theta})$.

Theorem 1 shows that by minimizing the ℓ_1 error, the Laplace MLE is able to attain the Laplace CRLB, thus reducing the asymptotic variance of the NLS estimator by 50%. Note that Theorem 1 not only asserts the asymptotic normality of the Laplace MLE but also explicitly specifies the neighborhood in which the Laplace MLE should be obtained as a minimizer of $\ell_1(\boldsymbol{\vartheta})$. The neighborhood shrinks as the sample size increases, an indication of stringent requirements on initial values.

4 A Computational Procedure

Finding the minimizer of $\ell_1(\boldsymbol{\vartheta})$ numerically is a challenging problem, not only because the objective function is not everywhere differentiable, but most importantly because it contains many local minima in the vicinity of the desired solution. The seriousness of this problem can be appreciated by examining the profile ℓ_1 error plot shown in Fig. 2, where $\ell_1(\boldsymbol{\vartheta})$ is plotted against the normalized frequency $f := \vartheta_3/(2\pi)$ with ϑ_1 and ϑ_2 replaced by their minimizing values (obtained numerically) for each given f.

The methods of NLS and periodogram maximization face a similar problem of local extrema. It has been demonstrated that an initial value of accuracy $\mathcal{o}(n^{-1})$ in standard error for the frequency parameter is required to guarantee the convergence of standard iterative algorithms [20] [35]. A typical remedy to the problem consists of a coarse search of the *n*-point DFT periodogram followed by a fine search based on local interpolation. This procedure usually generates sufficiently accurate initial values for standard optimization routines such as the Newton-Raphson algorithm [3]–[9].

Theorem 1 suggests that the Laplace MLE should be obtained in a neighborhood of radius $\mathcal{o}(n^{-4/3})$ of the true frequency. To produce an initial value in this neighborhood, we propose to use a unified procedure derived from the inverse filtering method discussed in [13] [28] [29], rather than to stitch together various techniques of different flavors.

The gist of the method is as follows. For any given $a \in (-2\eta/(1+\eta^2), 2\eta/(1+\eta^2))$ with fixed $\eta \in (0,1)$, let $y_t(a) := H_a(z^{-1})y_t$, where $H_a(z^{-1}) := \{1-(1+\eta^2)az^{-1}+\eta^2 z^{-2}\}^{-1}$ is a second-order IIR filter. Let $\rho_n(a) := (1+\eta^2)^{-1}\sum y_{t-1}(a)\{y_t(a)+\eta^2 y_{t-2}(a)\}/\sum y_{t-1}^2(a)$, which is the minimizer of the weighted sum of forward and backward prediction error sums of squares $\sum\{y_t(a)-\rho y_{t-1}(a)\}^2+\eta^2\sum\{y_{t-2}(a)-\rho y_{t-1}(a)\}^2$. Using this function of a, a sequence $\{\hat{a}_n^{(m)}\}$ is produced by the fixed-point iteration

$$\hat{a}_n^{(m)} := 2\rho_n\big(\hat{a}_n^{(m-1)}\big) - \hat{a}_n^{(m-1)} \quad (m = 1,2,\ldots). \tag{8}$$

It can be shown that with a suitable initial value the sequence converges to a fixed point $\hat{a}_n$ of $\rho_n(a)$, i.e., $\hat{a}_n^{(m)} \to \hat{a}_n = \rho_n(\hat{a}_n)$ as $m \to \infty$, and the fixed point leads to a frequency estimator $\hat{\omega}_n := \arccos(\hat{a}_n)$. The consistency and asymptotic normality of $\hat{\omega}_n$ as an estimator of ω can be established.

The bandwidth parameter η plays an important role in determining the required accuracy of the initial values and the final accuracy of the frequency estimator. By taking advantage of this relationship, we proposed a three-step algorithm (TSA) in [28] to bring an initial value of accuracy $\mathcal{O}(1)$ to a final estimate of accuracy arbitrarily close to $\mathcal{O}(n^{3/2})$ at the cost of computational complexity $\mathcal{O}(n\log n)$:

1. Take $1-\eta_1 = \mathcal{O}(1)$ and iterate $\mathcal{O}(n\log n)$ times with an initial value of accuracy $\mathcal{O}(1)$. This step produces an estimate of accuracy $\mathcal{O}(n^{-1/2})$.

2. Take $1-\eta_2 = \mathcal{O}(n^{-1/3})$ and iterate $\mathcal{O}(1)$ times using the result from Step 1 as initial value. This step produces an estimate of accuracy $\mathcal{O}(n^{-1})$.

3. Take $1-\eta_3 = \mathcal{O}(n^{-\nu})$ with $\nu = 1^-$ and iterate $\mathcal{O}(1)$ times using the result from Step 2 as initial value. This yields an estimate of accuracy arbitrarily close to $\mathcal{O}(n^{-3/2})$.

While the TSA produces a frequency estimate for initializing a general-purpose optimization routine to minimize $\ell_1(\boldsymbol{\vartheta})$, the required initial value of A and B can be obtained by linear regression with the TSA estimates in place of the true frequency values.

Note that a fourth step with $\eta_4 = 1$ is able to yield an estimator which is mathematically proven to attain the Gauss CRLB [36]. In practice, this step can be omitted because it does not offer any appreciable improvement in the estimation accuracy. Note also that Prony's estimator can be used as the initial value in Step 1. Prony's estimator equals $\arccos(\rho_n(0))$ with $\eta = 0$ (no filtering) and has accuracy $\mathcal{O}(1)$ (due to bias). With this estimator as the initial value instead of other alternatives such as SVD-based estimates, the entire procedure becomes unified in architecture, thus simplifying the hardware/software implementation.

Because $\ell_1(\boldsymbol{\vartheta})$ is not everywhere differentiable, gradient-based algorithms, such as the Newton-Raphson algorithm suitable for NLS, cannot be used to compute the Laplace MLE. Fortunately, there are many general-purpose algorithms that do not require the differentiability. The simplex algorithm of Nelder and Mead [37], available in software packages such as Mathematica and `R`, is such an example. This algorithm is the default choice in the `R` function `optim`; it is employed with all other default options in our simulation. The interior point algorithm discussed in [38] can also be used, but its `R` implementation with the default options does not seem to be as reliable as the simplex algorithm in `optim`.

Fig. 3 shows the result of a simulation where the MSE of frequency estimates for a sinusoid in LWN is

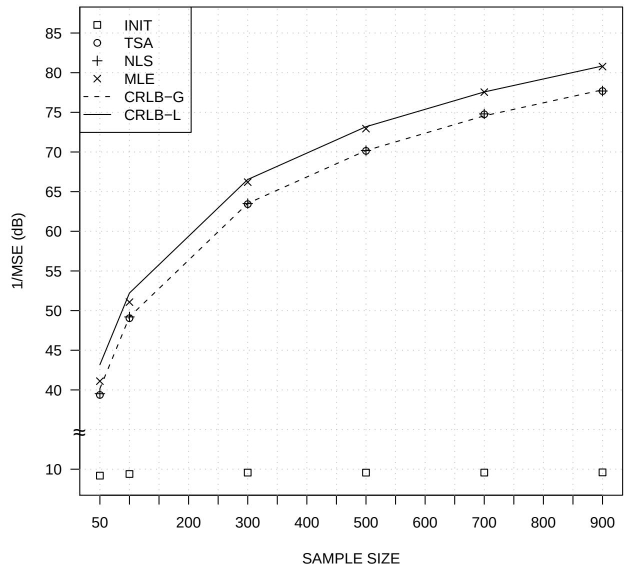

Figure 3: Reciprocal MSE of the estimates of $\omega/(2\pi)$ for a sinusoid plus LWN. Solid line, Laplace CRLB; dashed line, Gauss CRLB; $\times$, Laplace MLE initialized by TSA; $+$, NLS initialized by TSA; $\circ$, TSA initialized by Prony's estimator; $\square$, Prony's estimator. Results are based on 1,000 Monte Carlo runs.

calculated by the Monte Carlo method for various sample sizes. The signal and noise parameters are: $A = 1$, $B = 0$, $\omega = 0.15 \times 2\pi$, and $\gamma = 0$ dB. The three values of η in the TSA are 0.85, $1 - n^{-0.6}$, and $1 - n^{-0.9}$. With Prony's estimates as initial values, (8) is iterated 6, 3, and 11 times respectively to produce the final TSA estimates which in turn serve as initial values for the Laplace MLE and the NLS.

As can be seen, for all sample sizes considered in the simulation, the TSA is able to take the $\mathcal{O}(1)$-accurate Prony's estimator as input and produce a frequency estimator whose MSE follows the Gauss CRLB closely. The NLS estimates, obtained by minimizing $\ell_2(\boldsymbol{\vartheta})$ using the Nelder-Mead algorithm with the TSA estimates as initial values, do not offer any improvement in the MSE. By minimizing $\ell_1(\boldsymbol{\vartheta})$ instead of $\ell_2(\boldsymbol{\vartheta})$, the accuracy is improved considerably: except for the smallest sample size $n = 50$, the MSE of the resulting Laplace MLE is well approximated by the Laplace CRLB (3 dB improvement).

5 Performance Under Non-Laplace Noise

The Laplace MLE, which minimizes $\ell_1(\boldsymbol{\vartheta})$, can also be used to estimate the signal parameters when the noise is non-Laplace or when the noise distribution is unknown. In these cases, we refer to the resulting estimator as NLAD in analogy to NLS. The following theorem establishes the asymptotic normality of the NLAD estimator under a mild assumption of the noise distribution (see Appendix II for proof).

Theorem 2. *Let $\{y_t\}$ be given by (1), where $\{\varepsilon_t\}$ is a white noise process with zero mean, finite variance $\sigma^2 > 0$, and marginal distribution function $F(x)$. Assume that $F(0) = \frac{1}{2}$, $\dot{F}(0) > 0$, and $F(x) - F(0) = \dot{F}(0)x + \mathcal{O}(x^{d+1})$ for some constant $d > 0$ and all $|x| \ll 1$. Let $\Theta_n := \{\boldsymbol{\vartheta} \in \Theta_0 : \|D_n^{-1}(\boldsymbol{\vartheta} - \boldsymbol{\theta})\| \le \kappa n^{\alpha}\}$, where $\kappa > 0$ and $\alpha \in (0, \alpha_0)$ are constants, with $\alpha_0 := \min\{\frac{1}{6}, (d - \frac{1}{2})/(2d+1)\}$ if $d > \frac{1}{2}$ and $\alpha_0 := d/(2d+4)$ otherwise. Then, as $n \to \infty$, $\hat{\boldsymbol{\theta}}_n := \arg\min\{\ell_1(\boldsymbol{\vartheta}) : \boldsymbol{\vartheta} \in \Theta_n\}$ is asymptotically distributed as $N(\boldsymbol{\theta}, \beta_F^{-2} CRLB_L(\boldsymbol{\theta}))$, where $CRLB_L(\boldsymbol{\theta}) = \frac{1}{2} CRLB_G(\boldsymbol{\theta})$, $\beta_F := 2c\dot{F}(0)$, and $c := \sigma/\sqrt{2}$.*

Remark 4. If $F(x)$ has a continuous second derivative in a neighborhood of $x = 0$, then $d = 1$ (and $\alpha_0 = \frac{1}{6}$) by Taylor's expansion. In general, the smoothness of $F(x)$ near $x = 0$ determines the size of the neighborhood in which the objective function can be approximated by a quadratic form and thus the neighborhood in which the NLAD estimator should be obtained in order to have the guaranteed asymptotic properties.

Remark 5. The quantity β_F is the ratio of probability densities at zero: that of the true noise distribution divided by that of the Laplace distribution which equals $1/(2c)$. This ratio is dimensionless and invariant to the scale of the noise.

As a pleasant surprise, Theorem 2 reveals that the NLAD estimator can do better than the Laplace CRLB if the noise distribution satisfies $\beta_F > 1$. This is in complete contrast with the NLS estimator which in no circumstances can outperform the Gauss CRLB. Consider the T_ν distribution for example. It is easy to see that $\beta_F > 1$ if and only if $\nu \in (2, \nu_0)$, where $\nu_0 \approx 2.724$ is the solution to $\Gamma((\nu+1)/2) = \Gamma(\nu/2)\sqrt{\pi(\nu-2)/2}$. Note that for $\nu \in (2, \nu_0)$ the noise has a finite variance but an infinite third moment, an indication of heavy

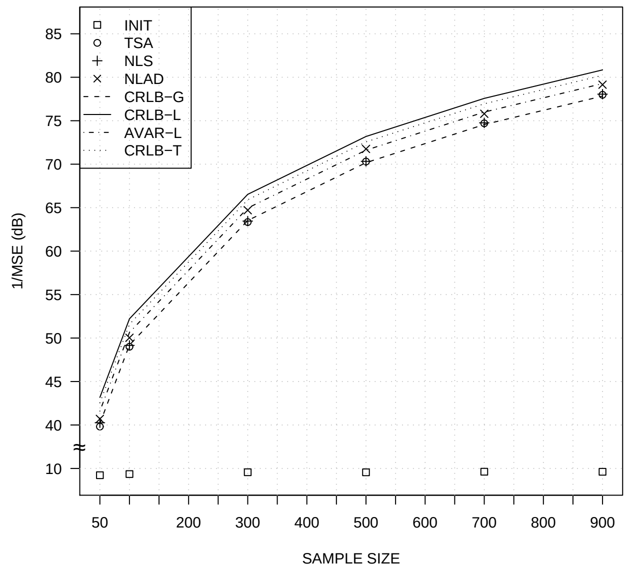

Figure 4: Same as Fig. 3 except the noise has a Student's T distribution with 3.3 degrees of freedom. Dash-dotted line, asymptotic variance in Theorem 2; dotted line, CRLB under the T distribution. Results are based on 1,000 Monte Carlo runs.

tails. In general, $\beta_F > 1$ if and if $\dot{F}(0) > 1/(\sqrt{2}\sigma)$. This is a condition that generally favors heavy-tailed distributions, because more outliers are needed to balance an increased concentration of probability mass near the origin in order to maintain a constant variance. It can be shown [39] [40] that the Laplace distribution in (4) is the least favorable distribution among symmetric unimodal distributions with $\beta_F \geq 1$.

Of course, it is also possible that $\beta_F < 1$ for some F, in which case the accuracy of NLAD cannot reach the Laplace CRLB. For Gaussian noise in particular, $\beta_F = 1/\sqrt{\pi}$. This means that the asymptotic variance of the NLAD estimator under GWN is equal to $\frac{1}{2}\pi \approx 1.57$ times the Gauss CRLB. Therefore, the efficiency gain of NLAD under heavy-tailed noise is achieved at the expense of slight loss of efficiency under the Gaussian and other light-tailed noise. The NLAD estimator exceeds the Gauss CRLB if and only if $\beta_F > 1/\sqrt{2}$, or equivalently, $\dot{F}(0) > 1/(2\sigma)$. Because $\dot{F}(x) = (1/\sigma)p(x/\sigma)$ for some unit-variance probability density $p(x)$, the condition that favors NLAD over NLS can be rewritten as $p(0) > 1/2$.

Figs. 4–6 present some simulation results of the NLAD estimator for frequency estimation under non-

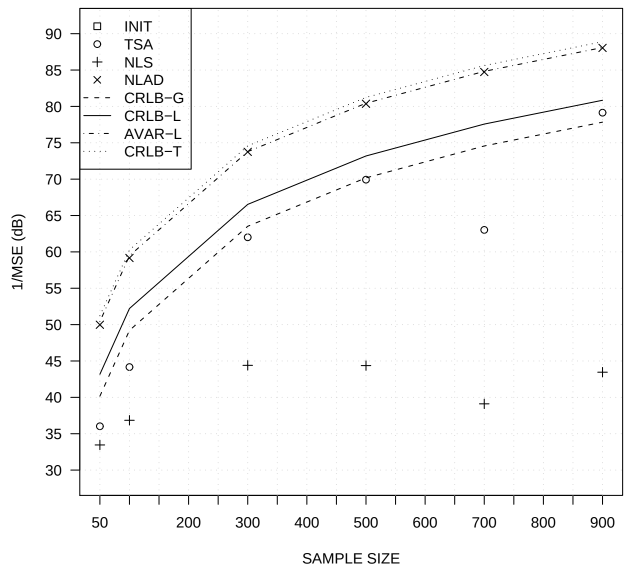

Figure 5: Same as Fig. 4 except the noise has a Student's T distribution with 2.1 degrees of freedom and all algorithms are initialized by the true parameter value. The MSE of Prony's estimator is not presented. Results are based 10,000 Monte Carlo runs.

Laplace conditions. In Figs. 4 and 5, the noise has a T distribution with $\nu = 3.3$ and $\nu = 2.1$, respectively. The first case represents a situation of moderately heavy tailed noise for which the third moment is finite but the fourth moment is infinite; the second case represents a situation where the noise has very heavy tails with an infinite third moment. In both cases, the simulated MSE of the NLAD estimator is closely approximated by the asymptotic variance in Theorem 2 for all sample sizes considered. The same is true for the Gaussian case shown in Fig. 6. It is important to observe that the MSE of the NLAD estimator exceeds the Gauss CRLB in both Figs. 4 and 5, and it does so by a large margin in Fig. 5 where the Laplace CRLB is also exceeded. Of course, the MSE falls short of the Gauss CRLB in Fig. 6 when the noise is Gaussian.

The case shown in Fig. 5 deserves some special comments. In this case, the TSA may fail to reach the Gauss CRLB, even with the true parameter value as the initial guess. A lack of robustness to very heavy tailed noise is largely responsible for the failure, giving rise to a handful of outliers in the estimates that inflate the resulting MSE considerably. This is not too surprising because the TSA is made of linear

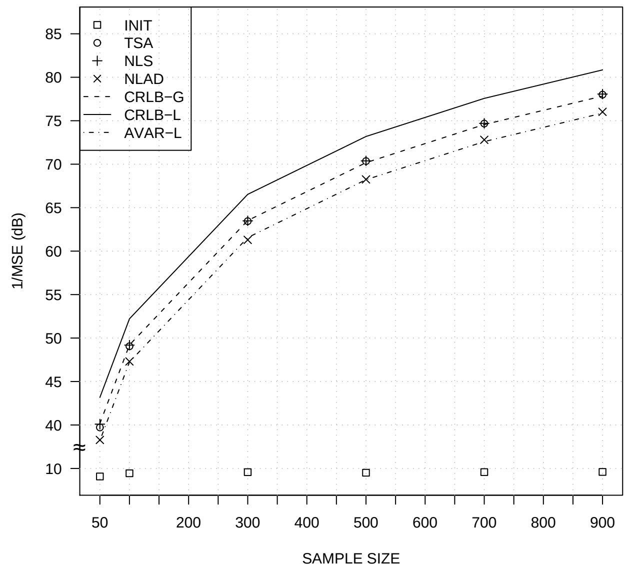

Figure 6: Same as Fig. 4 except the noise is Gaussian.

filtering and linear least-squares estimation, none of which is robust to heavy-tailed noise. The results of NLS shown in Fig. 5 are much worse. This may be explained by the compounded problem of robustness and convergence for the optimization routine in calculating the NLS estimates.

Fig. 7 gives an example that illustrates the lack of robustness of NLS and TSA in contrast with the robustness benefit of NLAD. Shown in Fig. 7(a) is a time series of 100 observations with the noise generated from the T distribution with $\nu = 2.1$. This particular time series contains a large outlier, so large that the periodogram reverses its typical behavior by producing a valley rather than a peak at the signal frequency, making the method of periodogram maximization completely invalid. The outlier has a similar impact on the mapping $\rho_n(a)$ shown in Fig. 7(b), where the fixed point closest to the signal frequency becomes a repeller instead of an attractor, causing the TSA to diverge from the signal frequency. The ℓ_2 error that the NLS estimator minimizes also exhibits reversed characteristics in this case: Fig. 7(d) shows that the signal frequency is closer to a maximizer rather than a minimizer of the ℓ_2 error. In complete contrast to the TSA and NLS, the NLAD estimator remains intact: Fig. 7(c) shows that the ℓ_1 error continues to provide a well-defined minimizer at the signal frequency in spite of the outlier contamination.

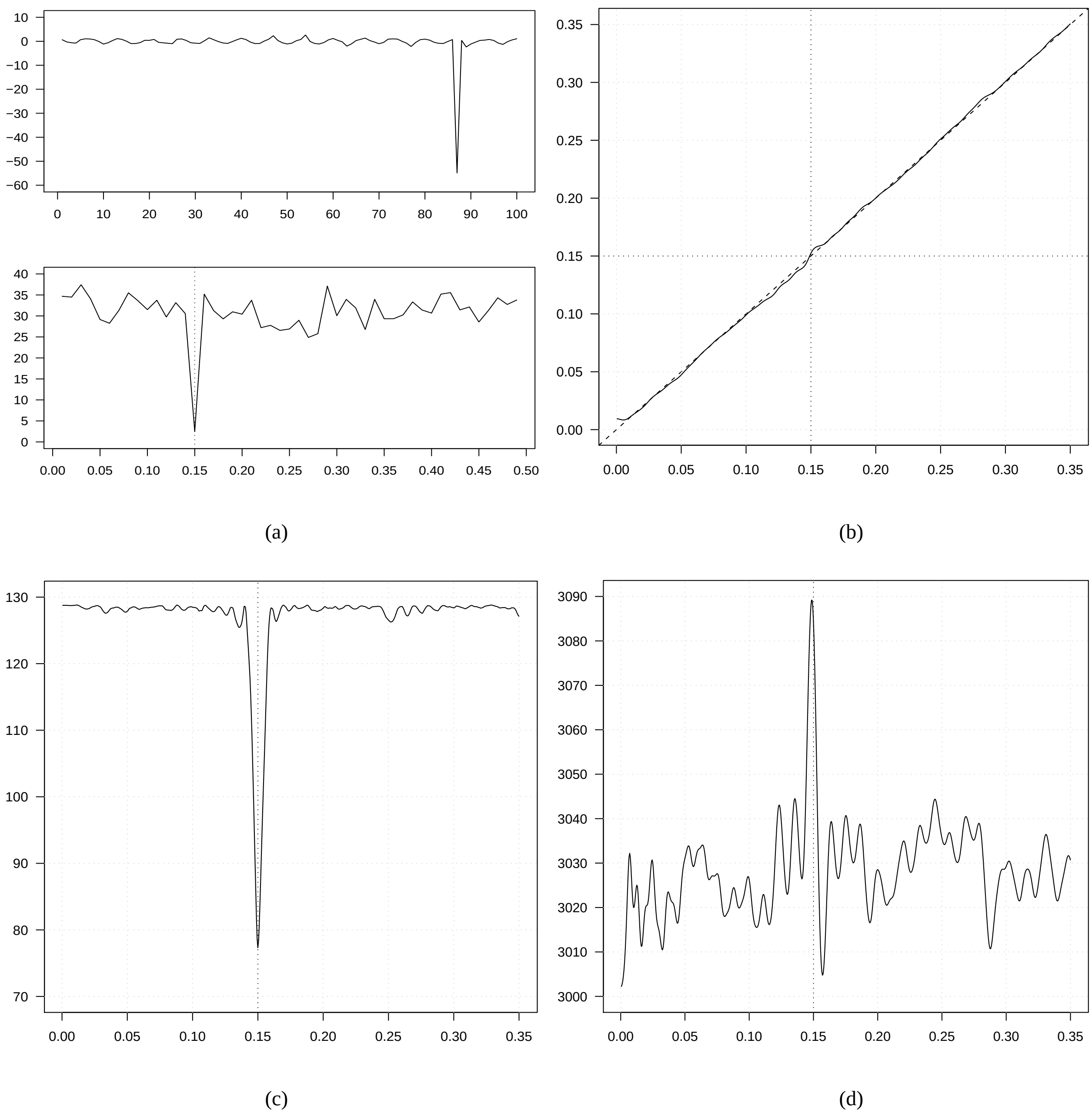

Figure 7: (a) Time series (top) of a sinusoid plus T-distributed noise with 2.1 degrees of freedom and its periodogram (bottom). (b) TSA mapping $f \mapsto \arccos\{\rho_n(\cos(2\pi f))\}/(2\pi)$ ($\eta = 0.984$). (c) Profile ℓ_1 error, similar to Fig. 2. (d) Profile ℓ_2 error. Dotted lines indicate the true frequency $f = 0.15$.

Note that an alternative method to deal with impulsive noise is to clip large observations whose magnitude exceeds a certain threshold [41]. This is a simple and effective technique if the range of "normal" observations is known a priori so that the right threshold can be set in advance. Selection of the threshold ultimately depends on the magnitude of the signal as well as the tail behavior of the noise. Such a priori knowledge is not assumed in our problem.

To end this section, we should point out that whereas the simulation is focused on the frequency estimation, similar remarks concerning the efficiency and robustness benefit of NLAD can be made about the amplitude estimation as well.

6 Confidence Regions

Although useful in practice and common in the literature, directly comparing the MSE with the asymptotic variance is not entirely appropriate in general, because convergence in distribution does not necessarily imply convergence in the first and second moments. A more proper way of using the results in Theorems 1 and 2 is to construct confidence regions for the parameters. In this application, a smaller asymptotic variance will yield a tighter confidence region for the same coverage probability.

For example, let v_f be the $(3,3)$ entry of $2\beta_F^{-2}\text{CRLB}_G(\boldsymbol{\theta})$ and $\hat{\omega}_n$ be the 3rd element in $\hat{\boldsymbol{\theta}}_n$. By Theorem 2, $(\omega-\hat{\omega}_n)^2/v_f \xrightarrow{D} \chi_1^2$ as $n\to\infty$, meaning that $\Pr\{(\omega-\hat{\omega}_n)^2/v_f \le \chi_1^2(p)\} \to p$ for any $p\in(0,1)$, where $\chi_1^2(p)$ is the pth quantile of χ_1^2. If $\hat{v}_f$ is a consistent estimator of v_f, then, by substituting $\hat{v}_f$ for v_f, we obtain a confidence interval for the frequency parameter: $\Omega_n(p) := \{\omega\in(0,\pi) : (\omega-\hat{\omega}_n)^2/\hat{v}_f \le \chi_1^2(p)\}$. By Theorem 2, the coverage of $\Omega_n(p)$, defined as $\Pr\{\omega\in\Omega_n(p)\}$, is approximately equal to p for large n.

Similarly, a confidence elliptic region for the amplitude parameters is given by $\mathcal{A}_n(p) := \{\boldsymbol{a}\in\mathbb{R}^2 : (\boldsymbol{a}-\hat{\boldsymbol{a}}_n)^T\hat{V}_a^{-1}(\boldsymbol{a}-\hat{\boldsymbol{a}}_n) \le \chi_2^2(p)\}$, where $\hat{\boldsymbol{a}}_n := [\hat{A}_n,\hat{B}_n]^T$ consists of the amplitude estimates in $\hat{\boldsymbol{\theta}}_n$ and $\hat{V}_a$ is a consistent estimator of the corresponding 2×2 submatrix of $2\beta_F^{-2}\text{CRLB}_G(\boldsymbol{\theta})$. A confidence ellipsoidal region for all three parameters is given by $\mathcal{C}_n(p) := \{\boldsymbol{\vartheta}\in\Theta_0 : (\boldsymbol{\vartheta}-\hat{\boldsymbol{\theta}}_n)^T\hat{V}^{-1}(\boldsymbol{\vartheta}-\hat{\boldsymbol{\theta}}_n) \le \chi_3^2(p)\}$, with $\hat{V}$

being a consistent estimator of $2\beta_F^{-2}\text{CRLB}_G(\boldsymbol{\theta})$.

To obtain $\hat{v}_f$, $\hat{V}_a$, and $\hat{V}$, one can plug in the NLAD estimator $\hat{\boldsymbol{\theta}}_n := [\hat{A}_n, \hat{B}_n, \hat{\omega}_n]^T$ for $\boldsymbol{\theta}$ in $\text{CRLB}_G(\boldsymbol{\theta})$ and replace $\dot{F}(0)$ in β_F by a density estimator such as $(2nh_n)^{-1}\sum I(|\hat{\varepsilon}_t| \le h_n)$, where $h_n > 0$ is a bandwidth parameter satisfying $h_n \to 0$ as $n \to \infty$ and $\hat{\varepsilon}_t := y_t - \{\hat{A}_n \cos(\hat{\omega}_n t) + \hat{B}_n \sin(\hat{\omega}_n t)\}$ $(t = 1, \ldots, n)$ are the residuals from the NLAD fit (the unknown σ^2 in β_F^{-2} cancels out with that in $\text{CRLB}_G(\boldsymbol{\theta})$). If the noise distribution is known *a priori* except for the variance, which is the scenario of our simulation study in this section, then β_F is known and the unknown σ^2 in $\text{CRLB}_G(\boldsymbol{\theta})$ can be substituted by the MLE of σ^2 on the basis of $\hat{\varepsilon}_t$ acting like ε_t $(t = 1, \ldots, n)$, which, in the Laplace case for example, equals $2(n^{-1}\sum|\hat{\varepsilon}_t|)^2$.

Tables I–III contain the results of a simulation study where the coverage probabilities of $\Omega_n(p)$, $\mathcal{A}_n(p)$, and $\mathcal{C}_n(p)$ are computed on the basis of 10,000 Monte Carlo runs for the nominal value $p = 0.95$ and for various sample sizes and noise distributions. The signal and noise parameters are: $A = 1$, $B = 0$, $\omega = 0.15 \times 2\pi$, and $\gamma = 0$ dB. In addition to the finite-sample CRLB in Proposition 1 that requires the inversion of $X^T X$, the asymptotic CRLB, defined as $D_n \boldsymbol{\Sigma}^{-1} D_n$ with $\boldsymbol{\Sigma}^{-1}$ given by Proposition 2, is also used as a computationally simpler alternative in the construction of the confidence regions.

As can be seen, the simulated coverage does approach the nominal value in all cases as the sample size grows; but the speed of convergence depends on the noise distribution: for example, it requires a larger sample size to reach the nominal value under the Laplace distribution than under the $T_{2.1}$ distribution. Moreover, for small sample sizes (e.g., $n = 50, 100$), the finite-sample CRLB is superior to the asymptotic CRLB, but the superiority diminishes as the sample size grows. The plug-in estimates of CRLB are effective as a substitute for the exact CRLB, especially for larger sample sizes.

7 Concluding Remarks

In this paper we have demonstrated analytically and numerically that under non-Gaussian noise conditions the amplitude and frequency parameters of sinusoidal signals can be estimated more accurately than the

TABLE I

COVERAGE OF 95% CONFIDENCE REGION FOR FREQUENCY

Noise	CRLB	$n=50$	$n=100$	$n=300$	$n=500$	$n=700$	$n=900$
Laplace	Estim (A)	0.83	0.87	0.91	0.92	0.93	0.93
	Estim (F)	0.88	0.90	0.92	0.93	0.93	0.94
	Exact (F)	0.89	0.91	0.92	0.93	0.93	0.94
$T_{2.1}$	Estim (A)	0.92	0.93	0.94	0.95	0.95	0.95
	Estim (F)	0.93	0.93	0.94	0.95	0.95	0.95
	Exact (F)	0.94	0.94	0.94	0.95	0.95	0.95
$T_{3.3}$	Estim (A)	0.87	0.90	0.93	0.94	0.94	0.94
	Estim (F)	0.92	0.94	0.94	0.95	0.95	0.95
	Exact (F)	0.94	0.94	0.95	0.95	0.95	0.95
Gauss	Estim (A)	0.86	0.89	0.92	0.93	0.94	0.94
	Estim (F)	0.94	0.94	0.95	0.95	0.95	0.95
	Exact (F)	0.94	0.95	0.95	0.95	0.95	0.95

Results are based on 10,000 Monte Carlo runs. (A) = Asymptotic CRLB, (F) = Finite-sample CRLB, Estim = Plug-in estimate of CRLB, Exact = True CRLB.

TABLE II

COVERAGE OF 95% CONFIDENCE REGION FOR AMPLITUDE

Noise	CRLB	$n=50$	$n=100$	$n=300$	$n=500$	$n=700$	$n=900$
Laplace	Estim (A)	0.81	0.86	0.90	0.92	0.93	0.93
	Estim (F)	0.84	0.88	0.92	0.92	0.93	0.93
	Exact (F)	0.87	0.90	0.92	0.93	0.93	0.93
$T_{2.1}$	Estim (A)	0.91	0.92	0.94	0.95	0.95	0.95
	Estim (F)	0.92	0.93	0.94	0.95	0.95	0.95
	Exact (F)	0.93	0.94	0.95	0.95	0.95	0.95
$T_{3.3}$	Estim (A)	0.87	0.90	0.93	0.94	0.94	0.94
	Estim (F)	0.91	0.93	0.94	0.95	0.95	0.95
	Exact (F)	0.93	0.94	0.95	0.95	0.95	0.95
Gauss	Estim (A)	0.86	0.89	0.92	0.93	0.94	0.94
	Estim (F)	0.91	0.93	0.94	0.95	0.95	0.95
	Exact (F)	0.93	0.94	0.95	0.95	0.95	0.95

TABLE III

COVERAGE OF 95% JOINT CONFIDENCE REGION FOR ALL PARAMETERS

Noise	CRLB	$n=50$	$n=100$	$n=300$	$n=500$	$n=700$	$n=900$
Laplace	Estim (A)	0.77	0.84	0.90	0.91	0.92	0.93
	Estim (F)	0.80	0.86	0.91	0.92	0.92	0.93
	Exact (F)	0.84	0.88	0.91	0.92	0.93	0.93
$T_{2.1}$	Estim (A)	0.89	0.92	0.94	0.95	0.95	0.95
	Estim (F)	0.91	0.92	0.94	0.95	0.95	0.95
	Exact (F)	0.93	0.93	0.94	0.95	0.95	0.95
$T_{3.3}$	Estim (A)	0.85	0.90	0.93	0.94	0.94	0.94
	Estim (F)	0.89	0.93	0.94	0.95	0.95	0.95
	Exact (F)	0.92	0.94	0.95	0.95	0.95	0.95
Gauss	Estim (A)	0.85	0.89	0.92	0.93	0.94	0.94
	Estim (F)	0.90	0.93	0.94	0.94	0.95	0.95
	Exact (F)	0.92	0.94	0.94	0.95	0.95	0.95

Gauss CRLB suggests. We have shown in particular that the maximum likelihood estimator derived under the Laplace white noise assumption, which minimizes the sum of absolute deviations, is able to attain the Laplace CRLB asymptotically, which is 50% lower than the Gauss CRLB attained by the most efficient methods available such as nonlinear least squares and periodogram maximization. We have also provided an asymptotic analysis of the Laplace MLE for non-Laplace noise and shown that it produces highly efficient estimates that outperform the nonlinear least squares considerably for heavy-tailed noise.

Furthermore, we have proposed a computational procedure for the maximum likelihood estimation. The procedure employs a simple iterative algorithm called TSA to obtain sufficiently accurate initial values for standard optimization routines. Owing to the global convergence property of the TSA, the proposed procedure is able to accommodate poor initial values of accuracy $\mathcal{O}(1)$ for the frequency parameter and produce a final frequency estimator of accuracy $\mathcal{O}(n^{-3/2})$ that attains the Laplace CRLB asymptotically.

It is straightforward to generalize the analytical results in this paper to the case of multiple sinusoids with $y_t = \sum_{k=1}^{q} \{A_k \cos(\omega_k t) + B_k \sin(\omega_k t)\} + \varepsilon_t$, where q is a known integer and $\boldsymbol{\theta} := [A_1, B_1, \omega_1, \ldots, A_q, B_q, \omega_q]^T$ is the unknown parameter vector. In fact, all propositions and theorems remain unchanged except that X and $\mathrm{CRLB}_G(\boldsymbol{\theta})$ are interpreted as $X := [X_1, \ldots, X_q]$ and $\mathrm{CRLB}_G(\boldsymbol{\theta}) := \mathrm{diag}\{\mathrm{CRLB}_G(\boldsymbol{\theta}_1), \ldots, \mathrm{CRLB}_G(\boldsymbol{\theta}_q)\}$, where X_k and $\mathrm{CRLB}_G(\boldsymbol{\theta}_k)$ are the matrices for the single sinusoid corresponding to $\boldsymbol{\theta}_k := [A_k, B_k, \omega_k]^T$. As discussed in [29], the TSA remains useful in providing initial values as long as the frequencies have an adequate separation from each other. Alternatively, one can use the multivariate version of the algorithm discussed in [13] and [42] to deal with closely spaced frequencies.

Our simulation studies indicate that whereas the TSA works well in providing initial values for the Laplace MLE under moderately heavy-tailed noise, more robust initialization alternatives are still needed in situations where the noise has very heavy tails. One possibility is the robust subspace methods in [26] and [27]. This topic deserves further investigation.

Appendix I: Proof of Theorem 1

Let $I_L(\boldsymbol{\theta})$ denote the Fisher information matrix under LWN. Then, $D_n I_L(\boldsymbol{\theta}) D_n = 2 D_n I_G(\boldsymbol{\theta}) D_n = 2\boldsymbol{\Sigma} + \mathcal{O}(n^{-1})$, where $\boldsymbol{\Sigma}$ is given by (3). We call $\tilde{I}_L(\boldsymbol{\theta}) := 2\boldsymbol{\Sigma}$ the normalized asymptotic Fisher information matrix under the LWN assumption. Define $Z_n(\boldsymbol{\delta}) := c^{-1}\{\ell_1(\boldsymbol{\theta} + D_n\boldsymbol{\delta}) - \ell_1(\boldsymbol{\theta})\}$, where $c := \sigma/\sqrt{2}$. The key to the proof of Theorem 1 is to establish that

$$Z_n(\boldsymbol{\delta}) = -\boldsymbol{\delta}^T \boldsymbol{\zeta}_n + \tfrac{1}{2}\boldsymbol{\delta}^T \tilde{I}_L(\boldsymbol{\theta})\boldsymbol{\delta} + R_n(\boldsymbol{\delta}), \tag{9}$$

where $R_n(\boldsymbol{\delta}) = \mathcal{o}_p(1)$ uniformly in $\boldsymbol{\delta} \in \Delta_n := \{\boldsymbol{\delta} : \boldsymbol{\theta} + D_n\boldsymbol{\delta} \in \Theta_n\} = \{\boldsymbol{\delta} : \boldsymbol{\theta} + D_n\boldsymbol{\delta} \in \Theta_0, \|\boldsymbol{\delta}\| \leq \kappa n^{\alpha}\}$ and $\boldsymbol{\zeta}_n \xrightarrow{D} N(\mathbf{0}, \tilde{I}_L(\boldsymbol{\theta}))$. The quadratic function of $\boldsymbol{\delta}$ in (9) has a unique minimum $\tilde{\boldsymbol{\delta}}_n := \tilde{I}_L^{-1}(\boldsymbol{\theta})\boldsymbol{\zeta}_n$. The assertion follows if we can show that $\hat{\boldsymbol{\delta}}_n := \arg\min\{Z_n(\boldsymbol{\delta}) : \boldsymbol{\delta} \in \Delta_n\} = D_n^{-1}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})$ is $\mathcal{o}_p(1)$ away from $\tilde{\boldsymbol{\delta}}_n$, i.e., $\hat{\boldsymbol{\delta}}_n - \tilde{\boldsymbol{\delta}}_n \xrightarrow{P} 0$. Toward that end, we rewrite (9) as $Z_n(\boldsymbol{\delta}) = Z_n(\tilde{\boldsymbol{\delta}}_n) + \frac{1}{2}(\boldsymbol{\delta} - \tilde{\boldsymbol{\delta}}_n)^T \tilde{I}_L(\boldsymbol{\theta})(\boldsymbol{\delta} - \tilde{\boldsymbol{\delta}}_n) + R_n(\boldsymbol{\delta}) - R_n(\tilde{\boldsymbol{\delta}}_n)$ and define $R_n := \max\{|R_n(\boldsymbol{\delta})| : \boldsymbol{\delta} \in \Delta_n\}$. For any constant $\mu > 0$, if $\tilde{\boldsymbol{\delta}}_n \in \Delta_n$, then $\inf\{Z_n(\boldsymbol{\delta}) : \boldsymbol{\delta} \in \Delta_n, \|\boldsymbol{\delta} - \tilde{\boldsymbol{\delta}}_n\| > \mu\} \geq Z_n(\tilde{\boldsymbol{\delta}}_n) + \frac{1}{2}a\mu^2 - 2R_n$, where $a > 0$ is the smallest eigenvalue of $\tilde{I}_L(\boldsymbol{\theta})$. Since $\tilde{\boldsymbol{\delta}}_n$ converges in distribution and $\Delta_n \to \mathbb{R}^3$, we have $\Pr(\tilde{\boldsymbol{\delta}}_n \notin \Delta_n) \to 0$. This, combined with $R_n \xrightarrow{P} 0$, implies that $\Pr(\|\hat{\boldsymbol{\delta}}_n - \tilde{\boldsymbol{\delta}}_n\| > \mu, \tilde{\boldsymbol{\delta}}_n \in \Delta_n) \to 0$, which, in turn, leads to $\Pr(\|\hat{\boldsymbol{\delta}}_n - \tilde{\boldsymbol{\delta}}_n\| > \mu) \to 0$.

To prove (9), consider the Taylor expansion $s_t(\boldsymbol{\theta} + D_n\boldsymbol{\delta}) = s_t(\boldsymbol{\theta}) + v_t + \tilde{r}_t$ with $v_t := (D_n\boldsymbol{\delta})^T \boldsymbol{g}_t(\boldsymbol{\theta})$, $\tilde{r}_t := (D_n\boldsymbol{\delta})^T H_t(\boldsymbol{\theta}_{nt})(D_n\boldsymbol{\delta})$, where $\boldsymbol{g}_t(\boldsymbol{\vartheta})$ and $H_t(\boldsymbol{\vartheta})$ denote the gradient vector and Hessian matrix of $s_t(\boldsymbol{\vartheta})$ and $\boldsymbol{\theta}_{nt}$ is a point between $\boldsymbol{\theta} + D_n\boldsymbol{\delta}$ and $\boldsymbol{\theta}$. Since $y_t = s_t(\boldsymbol{\theta}) + \varepsilon_t$, we can write $Z_n(\boldsymbol{\delta}) = c^{-1}\sum\{|\varepsilon_t - v_t - \tilde{r}_t| - |\varepsilon_t|\}$. Let $\phi(u,s) := I(u \leq s) - I(u \leq 0)$. Then, using Knight's identity [43],

$$|u - v| - |u| = -v\,\mathrm{sgn}(u) + 2\int_0^v \phi(u,s)\,ds,$$

with $u = \varepsilon_t$ and $v = v_t + \tilde{r}_t$, we obtain $Z_n(\boldsymbol{\delta}) = Z_{n1} + Z_{n2} + Z_{n3} + Z_{n4}$, where $Z_{n1} := -c^{-1}\sum v_t\,\mathrm{sgn}(\varepsilon_t)$, $Z_{n2} := 2c^{-1}\sum\int_0^{v_t}\phi(\varepsilon_t,s)\,ds$, $Z_{n3} := -c^{-1}\sum\tilde{r}_t\,\mathrm{sgn}(\varepsilon_t)$, and $Z_{n4} := 2c^{-1}\sum\int_{v_t}^{v_t+\tilde{r}_t}\phi(\varepsilon_t,s)\,ds$.

First, we show that $Z_{n2} = \frac{1}{2}\boldsymbol{\delta}^T \tilde{I}_L(\boldsymbol{\theta})\boldsymbol{\delta} + \mathcal{o}_p(n^{-r})$ uniformly in $\boldsymbol{\delta} \in \Delta_n$ for any fixed $r \in [0, \frac{1}{4} - \frac{3}{2}\alpha)$. Toward that end, let $\xi_t := \int_0^{v_t}\phi(\varepsilon_t,s)\,ds$ so that $Z_{n2} = 2c^{-1}\sum\xi_t$. Since $F(x) = \frac{1}{2}\exp(x/c)$ for $x < 0$ and

$F(x) = 1 - \frac{1}{2}\exp(-x/c)$ for $x \geq 0$, we obtain $2c^{-1}E(\xi_t) = 2c^{-1}\int_0^{v_t}\{F(s) - F(0)\}\,ds = \exp(-|v_t|/c) - 1 + |v_t|/c = \frac{1}{2}(v_t/c)^2 + \mathcal{O}(v_t^3)$. This, combined with $\sum(v_t/c)^2 = \boldsymbol{\delta}^T\tilde{I}_L(\boldsymbol{\theta})\boldsymbol{\delta} + \mathcal{O}(n^{-1+2\alpha})$ and $\sum|v_t|^3 = \mathcal{O}(n^{-1/2+3\alpha})$, leads to $E(Z_{n2}) = 2c^{-1}\sum E(\xi_t) = \frac{1}{2}\boldsymbol{\delta}^T\tilde{I}_L(\boldsymbol{\theta})\boldsymbol{\delta} + \mathcal{O}(n^{-1/2+3\alpha})$. Moreover, since $0 \leq \xi_t \leq |v_t|$, we have $\mathrm{Var}(\xi_t) \leq E(\xi_t^2) \leq |v_t|E(\xi_t)$. Combining this with $\max|v_t| = \mathcal{O}(n^{-1/2+\alpha})$ yields $\mathrm{Var}(Z_{n2}) = 4c^{-2}\sum\mathrm{Var}(\xi_t) \leq 2c^{-1}(\max|v_t|)E(Z_{n2}) = \mathcal{O}(n^{-1/2+3\alpha})$. Now, substitute $\boldsymbol{\delta}$ with $n^\alpha\boldsymbol{\eta}$ in $n^{-2\alpha}Z_{n2}$ and denote the resulting function of $\boldsymbol{\eta}$ as $\tilde{Z}_{n2}(\boldsymbol{\eta})$. Then, the previous results can be restated as $E\{\tilde{Z}_{n2}(\boldsymbol{\eta})\} = \frac{1}{2}\boldsymbol{\eta}^T\tilde{I}_L(\boldsymbol{\theta})\boldsymbol{\eta} + \mathcal{O}(n^{-1/2+\alpha})$ and $\mathrm{Var}\{\tilde{Z}_{n2}(\boldsymbol{\eta})\} = \mathcal{O}(n^{-1/2-\alpha})$ for any fixed $\boldsymbol{\eta}$. Citing Chebyshev's inequality proves $\tilde{Z}_{n2}(\boldsymbol{\eta}) = \frac{1}{2}\boldsymbol{\eta}^T\tilde{I}_L(\boldsymbol{\theta})\boldsymbol{\eta} + \mathcal{O}_p(n^{-1/4-\alpha/2})$ for any fixed $\boldsymbol{\eta}$. Because ξ_t is a convex function of v_t and v_t is a linear function of $\boldsymbol{\eta}$, it follows that $\tilde{Z}_{n2}(\boldsymbol{\eta})$ is a convex function of $\boldsymbol{\eta}$, so the Convexity Lemma in [44] guarantees that $\tilde{Z}_{n2}(\boldsymbol{\eta}) = \frac{1}{2}\boldsymbol{\eta}^T\tilde{I}_L(\boldsymbol{\theta})\boldsymbol{\eta} + o_p(n^{-r})$ uniformly in $\|\boldsymbol{\eta}\| \leq \kappa$ for any fixed $r \in [0, \frac{1}{4} + \frac{1}{2}\alpha)$. The assertion is proved upon noting that $\|n^{-\alpha}\boldsymbol{\delta}\| \leq \kappa$ for any $\boldsymbol{\delta} \in \Delta_n$.

Next, we show $Z_{n3} = \mathcal{O}_p(n^{-1/2+3\alpha})$ uniformly in $\boldsymbol{\delta} \in \Delta_n$. This can be done by using the Taylor expansion $\tilde{r}_t = r_t + u_t$, where $r_t := (D_n\boldsymbol{\delta})^T H_t(\boldsymbol{\theta})(D_n\boldsymbol{\delta})$. In this expansion, u_t is a linear combination of the third partial derivatives of $s_t(\boldsymbol{\vartheta})$ evaluated at an intermediate point which may depend on t. These partial derivatives are either zero (for those involving differentiation with respect to the amplitudes more than once) or can be expressed as $\mathcal{O}(t^\beta)$ for some $\beta = 2, 3$. Moreover, owing to the presence of $D_n\boldsymbol{\delta}$ and to the fact that the components in $\boldsymbol{\theta}_{nt} - \boldsymbol{\theta}$ that correspond to the amplitudes can be expressed as $\mathcal{O}(n^{-1/2+\alpha})$ and those that correspond to the frequency as $\mathcal{O}(n^{-3/2+\alpha})$, the coefficients of the third partial derivatives of the form $\mathcal{O}(t^\beta)$ take the form $\mathcal{O}(n^{-(\beta+1)-1/2+3\alpha})$. This leads to $\sum|u_t| = \mathcal{O}(n^{-1/2+3\alpha})$ and hence $\sum u_t\,\mathrm{sgn}(\varepsilon_t) = \mathcal{O}(n^{-1/2+3\alpha})$. Furthermore, because $\sum r_t\,\mathrm{sgn}(\varepsilon_t) = \boldsymbol{\delta}^T\Xi_n\boldsymbol{\delta}$, where $\Xi_n := D_n(\sum H_t(\boldsymbol{\theta})\,\mathrm{sgn}(\varepsilon_t))D_n$ is a random matrix whose elements have mean zero and variance $\mathcal{O}(n^{-1})$, it follows from Chebyshev's inequality that $\Xi_n = \mathcal{O}_p(n^{-1/2})$ and hence $\sum r_t\,\mathrm{sgn}(\varepsilon_t) = \mathcal{O}_p(n^{-1/2+2\alpha})$. Combing these results proves the assertion.

Write $Z_{n4} = Z'_{n4} + Z''_{n4}$, where $Z'_{n4} := 2c^{-1}\sum\psi_t$, $Z''_{n4} := 2c^{-1}\sum\rho_t$, $\psi_t := \int_{v_t}^{v_t+r_t}\phi(\varepsilon_t, s)\,ds$, and $\rho_t := \int_{v_t+r_t}^{v_t+r_t+u_t}\phi(\varepsilon_t, s)\,ds$. Since $|\rho_t| \leq 2|u_t|$ and $\sum|u_t| = \mathcal{O}(n^{-1/2+3\alpha})$, we have $Z''_{n4} = \mathcal{O}(n^{-1/2+3\alpha})$ uniformly in $\boldsymbol{\delta} \in \Delta_n$. We also have $E(Z'_{n4}) = \mathcal{O}(n^{-1/2+3\alpha})$, because $2c^{-1}E(\psi_t) = \int_{v_t}^{v_t+r_t}\{F(s)-F(0)\}\,ds = v_t r_t/c^2 + \frac{1}{2}r_t^2/c^2 + \mathcal{O}((v_t+r_t)^3) + \mathcal{O}(v_t^3)$, $\max(|v_t|) = \mathcal{O}(n^{-1/2+\alpha})$, $\max(|r_t|) = \mathcal{O}(n^{-1+2\alpha})$, $\sum v_t r_t = \mathcal{O}(n^{-1/2+3\alpha})$, and $\sum r_t^2 = \mathcal{O}(n^{-1+4\alpha})$. Furthermore, by the same technique employed to bound $\mathrm{Var}(Z_{n2})$, we can show that the variance of $Z'_{n4} = 2c^{-1}\sum(\int_0^{v_t+r_t} - \int_0^{v_t})\phi(\varepsilon_t, s)\,ds$ takes the form $\mathcal{O}(n^{-1+4\alpha})$. Let $\boldsymbol{\delta}$ be substituted by $n^{\alpha}\boldsymbol{\eta}$ in $n^{-2\alpha}Z'_{n4}$ and denote the resulting function of $\boldsymbol{\eta}$ by $\tilde{Z}_{n4}(\boldsymbol{\eta})$. Then, the previous evaluations imply $n^r\tilde{Z}_{n4}(\boldsymbol{\eta}) = \mathcal{O}_p(n^{-1/2+\alpha+r}) \xrightarrow{P} 0$ for any fixed $\boldsymbol{\eta}$ and $r \in [0, \frac{1}{2}-\alpha)$. It remains to show that the convergence is also uniform in $\|\boldsymbol{\eta}\| \le \kappa$. According to Lemma 2.2 in [45] (see also Theorem 1 in [46]), it suffices to show that for any $\tau > 0$,

$$\lim_{h\to 0}\limsup_{n\to\infty} \Pr\left\{\max_{(\boldsymbol{\eta},\boldsymbol{\eta}')\in\mathcal{B}(h)} n^r\,|\tilde{Z}_{n4}(\boldsymbol{\eta}) - \tilde{Z}_{n4}(\boldsymbol{\eta}')| \ge \tau\right\} = 0, \tag{10}$$

where $\mathcal{B}(h) := \{(\boldsymbol{\eta},\boldsymbol{\eta}') : \|\boldsymbol{\eta}\| \le \kappa, \|\boldsymbol{\eta}'\| \le \kappa, \|\boldsymbol{\eta}-\boldsymbol{\eta}'\| \le h\}$. This so-called stochastic equicontinuity condition is proved in Appendix III. Thus, $Z'_{n4} = o_p(n^{-r})$ uniformly in $\boldsymbol{\delta} \in \Delta_n$ for any $r \in [0, \frac{1}{2}-3\alpha)$.

Finally, let $\boldsymbol{\zeta}_n := c^{-1}D_n\sum \boldsymbol{g}_t(\boldsymbol{\theta})\,\mathrm{sgn}(\varepsilon_t)$. It suffices to show that $Z_{n1} = -\boldsymbol{\delta}^T\boldsymbol{\zeta}_n \xrightarrow{D} N(0, \boldsymbol{\delta}^T\tilde{I}_L(\boldsymbol{\theta})\boldsymbol{\delta})$ for any fixed $\boldsymbol{\delta} \ne \mathbf{0}$. Toward that end, we note that $E(Z_{n1}) = 0$ and $\mathrm{Var}(Z_{n1}) = (D_n\boldsymbol{\delta})^T I_L(\boldsymbol{\theta})(D_n\boldsymbol{\delta}) = \boldsymbol{\delta}^T\tilde{I}_L(\boldsymbol{\theta})\boldsymbol{\delta} + \mathcal{O}(n^{-1})$. Furthermore, since $\Gamma_n := c^{-3}\sum E\{|v_t\,\mathrm{sgn}(\varepsilon_t)|^3\} = c^{-3}\sum|v_t|^3 = \mathcal{O}(n^{-1/2})$, we have $\{\mathrm{Var}(Z_{n1})\}^{-3/2}\Gamma_n = \mathcal{O}(n^{-1/2}) \to 0$. Citing Liapounov's central limit theorem [47] proves the assertion.

Appendix II: Proof of Theorem 2

Similar to the proof of Theorem 1, the goal is to show that $Z_n(\boldsymbol{\delta}) = -\boldsymbol{\delta}^T\boldsymbol{\zeta}_n + \frac{1}{2}\beta_F\boldsymbol{\delta}^T\tilde{I}_L(\boldsymbol{\theta})\boldsymbol{\delta} + R_n(\boldsymbol{\delta})$, where $R_n(\boldsymbol{\delta}) = o_p(1)$ uniformly in $\boldsymbol{\delta} \in \Delta_n$ and $\boldsymbol{\zeta}_n \xrightarrow{D} N(\mathbf{0}, \tilde{I}_L(\boldsymbol{\theta}))$. By following the steps in the proof of Theorem 1, it suffices to show that $Z_{n2} - \frac{1}{2}\beta_F\boldsymbol{\delta}^T\tilde{I}_L(\boldsymbol{\theta})\boldsymbol{\delta}$ and Z'_{n4} can be expressed as $o_p(n^{-r})$ for any fixed $\boldsymbol{\delta}$ and $0 \le r < \min\{\frac{1}{4}-\frac{3}{2}\alpha, \frac{1}{2}d-(d+2)\alpha\}$. Toward that end, we note that $E(\xi_t) = \int_0^{v_t}\{\dot{F}(0)s + \mathcal{O}(s^{d+1})\}\,ds = \frac{1}{2}\dot{F}(0)v_t^2 + \mathcal{O}(v_t^{d+2})$. This, combined with $\sum|v_t|^{d+2} = \mathcal{O}(n^{-d/2+(d+2)\alpha})$, yields $E(Z_{n2}) = \frac{1}{2}\beta_F\boldsymbol{\delta}^T\tilde{I}_L(\boldsymbol{\theta})\boldsymbol{\delta} + \mathcal{O}(n^{-d/2+(d+2)\alpha})$. Similarly, $E(\psi_t) = \int_{v_t}^{v_t+r_t}\{\dot{F}(0)s + \mathcal{O}(s^{d+1})\}\,ds = \dot{F}(0)(v_t r_t + \frac{1}{2}r_t^2) +$

$\mathcal{O}((v_t+r_t)^{d+2})+\mathcal{O}(v_t^{d+2})$, so $E(Z'_{n4})=\mathcal{O}(n^{-d/2+(d+2)\alpha})$. Furthermore, both $\mathrm{Var}(Z_{n2})$ and $\mathrm{Var}(Z'_{n4})$ take the form $\mathcal{O}(n^{-1/2+3\alpha})$. Combining these results proves the assertion. Finally, because $\frac{1}{4}-\frac{3}{2}\alpha<\frac{1}{2}d-(d+2)\alpha$ if and only if $\alpha<(d-\frac{1}{2})/(2d+1)$, taking $0<\alpha<\min\{\frac{1}{6},(d-\frac{1}{2})/(2d+1)\}$ in the first case and $0<\alpha<d/(2d+4)$ in the second case ensures $R_n(\boldsymbol{\delta})=o_p(1)$ uniformly in $\boldsymbol{\delta}\in\Delta_n$.

Appendix III: Proof of (10)

Lemma 1. *Let $\phi(u,s):=I(u\leq s)-I(u\leq 0)$. Then, $|\int_a^b\phi(u,s)\,ds|\leq|b-a|\,I(|u|\leq\max(|a|,|b|))$.*

Proof. Let $\psi(u;a,b):=\int_a^b\phi(u,s)\,ds$ and consider the case of $b>a$. It is easy to show that

$$\text{for } u>0: \quad \psi(u;a,b)=\begin{cases}0 & \text{if } b<u,\\ b-u & \text{if } a<u\leq b,\\ b-a & \text{if } u\leq a,\end{cases}$$

$$\text{for } u\leq 0: \quad \psi(u;a,b)=\begin{cases}-(b-a) & \text{if } u>b,\\ a-u & \text{if } a<u\leq b,\\ 0 & \text{if } u\leq a.\end{cases}$$

Combining these results yields the following expressions:

$$\text{for } b>a\geq 0: \quad \psi(u;a,b)=\begin{cases}0 & \text{if } u>b,\\ b-u & \text{if } a<u\leq b,\\ b-a & \text{if } 0<u\leq a,\\ 0 & \text{if } u\leq 0.\end{cases}$$

$$\text{for } b>0>a: \quad \psi(u;a,b)=\begin{cases}0 & \text{if } u>b,\\ b-u & \text{if } 0<u\leq b,\\ a-u & \text{if } a<u\leq 0,\\ 0 & \text{if } u\leq a.\end{cases}$$

$$\text{for } 0 \geq b > a: \quad \psi(u;a,b) = \begin{cases} 0 & \text{if } u > 0, \\ a-b & \text{if } b < u \leq 0, \\ a-u & \text{if } a < u \leq b, \\ 0 & \text{if } u \leq a. \end{cases}$$

In all three cases we have $|\psi(u;a,b)| \leq |b-a|\, I(|u| < \max(|a|,|b|))$. This inequality remains true in the case of $a > b$ because $\psi(u;a,b) = -\psi(u;b,a)$. □

By definition, $\psi_t = \int_{v_t}^{w_t} \phi(\varepsilon_t, s)\, ds$, where $w_t := v_t + r_t$. With $\boldsymbol{\delta}$ replaced by $n^\alpha \boldsymbol{\eta}$, we regard ψ_t, v_t, r_t, and w_t as functions of $\boldsymbol{\eta}$ and denote their values at $\boldsymbol{\eta}'$ by ψ_t', v_t', r_t', and w_t'. Then, we can write $\psi_t - \psi_t' = (\int_{v_t}^{w_t} - \int_{v_t'}^{w_t'})\phi(\varepsilon_t, s)\, ds = (\int_{w_t'}^{w_t} - \int_{v_t'}^{v_t})\phi(\varepsilon_t, s)\, ds := \Delta\psi_{t1} - \Delta\psi_{t2}$. By Lemma 1, we have $|\Delta\psi_{t1}| \leq |w_t - w_t'|\, I(|\varepsilon_t| < \max(|w_t|, |w_t'|)$ and $|\Delta\psi_{t2}| \leq |v_t - v_t'|\, I(|\varepsilon_t| < \max(|v_t|, |v_t'|)$. It is shown in Appendix I that $\max|v_t| = \mathcal{O}(n^{-1/2+\alpha})$ and $\max|r_t| = \mathcal{O}(n^{-1+2\alpha})$ uniformly in $\|\boldsymbol{\eta}\| \leq \kappa$. Therefore, we obtain $|\Delta\psi_{t1}| \leq |w_t - w_t'|\, I(|\varepsilon_t| < c_n)$ and $|\Delta\psi_{t2}| \leq |v_t - v_t'|\, I(|\varepsilon_t| < c_n)$, where $c_n := c_1 n^{-1/2+\alpha}$ for some constant $c_1 > 0$. Moreover, for any $(\boldsymbol{\eta}, \boldsymbol{\eta}') \in \mathcal{B}(h)$, there exists a constant $\kappa_1 > 0$, which may depend on κ, such that $|v_t - v_t'| = n^\alpha |(\boldsymbol{\eta} - \boldsymbol{\eta}')^T D_n \mathbf{g}_t(\boldsymbol{\theta})| \leq \kappa_n h$, $|r_t - r_t'| = n^{2\alpha} |\boldsymbol{\eta}^T D_n H_t(\boldsymbol{\theta}) D_n \boldsymbol{\eta} - \boldsymbol{\eta}'^T D_n H_t(\boldsymbol{\theta}) D_n \boldsymbol{\eta}'| \leq \kappa_n h$, and $|w_t - w_t'| \leq |v_t - v_t'| + |r_t - r_t'| \leq \kappa_n h$, where $\kappa_n := \kappa_1 n^{-1/2+\alpha}$. Because $\tilde{Z}_{n4}(\boldsymbol{\eta}) = 2c^{-1} n^{-2\alpha} \sum \psi_t$, we have $|\tilde{Z}_{n4}(\boldsymbol{\eta}) - \tilde{Z}_{n4}(\boldsymbol{\eta}')| \leq 2c^{-1} n^{-2\alpha} \{\sum(|w_t - w_t'| + |v_t - v_t'|)\} I(\max|\varepsilon_t| \leq c_n) \leq d_n h\, I(\max|\varepsilon_t| \leq c_n)$, where $d_n := 4c^{-1} n^{-2\alpha+1} \kappa_n = 4c^{-1} \kappa_1 n^{1/2-\alpha}$. Therefore, for large n such that $n^r d_n h \geq \tau$, we have

$$\Pr\left\{ \max_{(\boldsymbol{\eta}, \boldsymbol{\eta}') \in \mathcal{B}(h)} n^r\, |\tilde{Z}_{n4}(\boldsymbol{\eta}) - \tilde{Z}_{n4}(\boldsymbol{\eta}')| \geq \tau \right\} \leq e_n,$$

where $e_n := \{F(c_n) - F(-c_n)\}^n \to 0$ as $n \to \infty$ because $0 \leq F(c_n) - F(-c_n) \leq \frac{1}{2}$ for large n.

References

[1] P. Stoica, "List of references on spectral line analysis," *Signal Processing*, vol. 31, no. 3, pp. 329–340, 1993.

[2] B. G. Quinn and E. J. Hannan, *The Estimation and Tracking of Frequency*, Cambridge, UK: Cambridge University Press, 2001.

[3] A. M. Walker, “On the estimation of a harmonic component in a time series with stationary independent residuals,” *Biometrika*, vol. 58, no. 1, pp. 21–36, 1971.

[4] L. C. Palmer, “Coarse frequency estimation using the discrete Fourier transform,” *IEEE Trans. Inform. Theory*, vol. 20, no. 1, pp. 104–109, 1974.

[5] D. C. Rife and R. R. Boorstyn, “Single-tone parameter estimation from discrete-time observations,” *IEEE Trans. Inform. Theory*, vol. 20, no. 5, pp. 591–598, 1974.

[6] E. Aboutanios and B. Mulgrew, “Iterative frequency estimation by interpolation on Fourier coefficients,” *IEEE Trans. Signal Processing*, vol. 53, no. 4, pp. 1237–1242, 2005.

[7] P. Stoica, R. L. Moses, B. Friedlander, and T. Söderström, “Maximum likelihood estimation of the parameters of multiple sinusoids from noisy measurements,” *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. 37, no. 3, pp. 378–392, 1989.

[8] D. Starer and A. Nehorai, “Newton algorithm for conditional and unconditional maximum likelihood estimation of the parameters of exponential signals in noise,” *IEEE Trans. Signal Processing*, vol. 40, no. 6, pp. 1528–1534, 1992.

[9] H. Van Hamme, “Maximum likelihood estimation of superimposed complex sinusoids in white Gaussian noise by reduced effort coarse search (RECS),” *IEEE Trans. Signal Processing*, vol. 39, no. 2, pp. 536–538, 1991.

[10] S. M. Kay, “Accurate frequency estimation at low signal-to-noise ratio,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 32, no. 3, pp. 540–547, 1984.

[11] M. S. Mackisack and D. S. Poskitt, "Autoregressive frequency estimation," *Biometrika*, vol. 76, no. 3, pp. 565–575, 1989.

[12] P. Stoica, T. Söderström, and F. Ti, "Asymptotic properties of the high-order Yule-Walker estimates of sinusoidal frequencies," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, no. 11, pp. 1721–1734, 1989.

[13] T. H. Li and B. Kedem, "Iterative filtering for multiple frequency estimation," *IEEE Trans. Signal Processing*, vol. 42, no. 5, pp. 1120–1132, 1994.

[14] V. F. Pisarenko, "The retrieval of harmonics from a covariance function," *Geophys. J. Roy. Astronom. Soc.*, vol. 33, pp. 347–366, 1973.

[15] D. W. Tufts and R. Kumaresan, " Estimation of frequencies of multiple sinusoids: Making linear prediction performance like maximum likelihood," *Proc. IEEE*, vol. 70, no. 9, pp. 975–989, 1982.

[16] R. Roy and T. Kailath, "ESPRIT – estimation of signal parameters via rotation invariance techniques," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, no. 7 pp. 984–995, 1989.

[17] P. Stoica and T. Söderström, "Statistical analysis of MUSIC and subspace rotation estimates of sinusoidal frequencies," *IEEE Trans. Signal Processing*, vol. 39, no. 8, pp. 1836–1847, 1991.

[18] K. W. Chan and H. C. So, "An exact analysis of Pisarenko's single-tone frequency estimation algorithm," *Signal Processing*, vol. 83, no. 3, pp. 685–690, 2003.

[19] K. Mahata, "Subspace fitting approaches for frequency estimation using real-valued data," *IEEE Trans. Signal Processing*, vol. 53, no. 8, pp. 3099–3110, 2005.

[20] T. H. Li and K. S. Song, "On asymptotic normality of nonlinear least squares for sinusoidal parameter estimation," *IEEE Trans. Signal Processing*, vol. 56, no. 9, pp. 4511–4515, 2008.

[21] K. S. Song and T. H. Li, “A note on asymptotics of least squares for nonlinear harmonic regression,” IBM Research Report RC24481, 2008.

[22] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, Wiley, New York, 2001.

[23] G. Shevlyakov and K. Kim, “Robust minimax detection of a weak signal in noise with a bounded variance and density value at the center of symmetry,” *IEEE Trans. Inform. Theory*, vol. 52, no. 3, pp. 1206–1211, 2006.

[24] M. W. Thompson, D. R. Halverson, and G. L. Wise, “Robust detection in nominally Laplace noise,” *IEEE Trans. Commun.*, vol. 42, no. 2/3/4, pp. 1651–1660, 1994.

[25] J. Friedman, H. Messer, and J.-F. Cardoso, “Robust parameter estimation of a deterministic signal in impulsive noise,” *IEEE Trans. Signal Processing*, vol. 48, no. 4, pp. 935–942, 2000.

[26] T. Liu and J. M. Mendel, “A subspace-based direction finding algorithm using fractional lower order statistics,” *IEEE Trans. Signal Processing*, vol. 49, no. 8, pp. 1605–1613, 2001.

[27] M. Altinkaya, H. Delic, B. Sankur, and E. Anarim, “Subspace-based frequency estimation of sinusoidal signals in alpha-stable noise,” *Signal Processing*, vol. 82, no. 12, pp. 1807–1827, 2002.

[28] K. S. Song and T. H. Li, “A statistically and computationally efficient method for frequency estimation,” *Stochastic Processes Appl.*, vol. 86, pp. 29–47, 2000.

[29] T. H. Li and K. S. Song, “Asymptotic analysis of a fast algorithm for efficient multiple frequency estimation,” *IEEE Trans. Inform. Theory*, vol. 48, no. 10, pp. 2709–2720, 2002. (Errata: vol. 49, no. 2, p. 529, 2003).

[30] A. W. van der Vaart, *Asymptotic Statistics*, Cambridge, UK: Cambridge University Press, 1998.

[31] D. Sengupta and S. M. Kay, "Efficient estimation of parameters for non-Gaussian autoregressive processes," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, no. 6, pp. 789–794, 1989.

[32] I. A. Ibragimov and R. Z. Has'minskii, *Statistical Estimation: Asymptotic Theory*, New York: Springer-Verlag, 1981.

[33] P. J. Huber, *Robust Statistics*, Chapt. 4, New York: Wiley, 1981.

[34] A. M. Kagan, Y. V. Linnik, and C. R. Rao, *Characterization Problems in Mathematical Statistics*, Theorem 13.1.1, p. 406, New York: Wiley, 1973.

[35] J. A. Rice and M. Rosenblatt, "On frequency estimation," *Biometrika*, vol. 75, no. 3, pp. 477–484, 1988.

[36] B. G. Quinn and J. M. Fernandes, "A fast efficient technique for the estimation of frequency," *Biometrika*, vol. 78, no. 3, pp. 489–497, 1991.

[37] J. A. Nelder and R. Mead, "A simplex algorithm for function minimization," *Comput. J.*, vol. 7, pp. 308–313, 1965.

[38] R. Keonker and B. Park, "An interior point algorithm for nonlinear quantile regression," *J. Econometrics*, vol. 71, pp. 265–283, 1996.

[39] B. T. Poljak and J. Z. Tsypkin, "Robust identification," *Automatica*, vol. 16, pp. 53–63, 1980.

[40] H. Delic, P. Papantoni-Kazakos, and D. Kazakos, "Fundamental structures and asymptotic performace criteria in decentralized binary hypothesis testing," *IEEE Trans. Commun.*, vol. 43, no. 1, pp. 32–43, 1995.

[41] A. Swami and B. M. Sadler, "On some detection and estimation problems in heavy-tailed noise," *Signal Processing*, vol. 82, pp. 1829–1846, 2002.

[42] K. S. Song and T. H. Li, “On convergence and bias correction of a joint estimation method for multiple sinusoidal frequencies,” *J. Amer. Statist. Assoc.*, vol. 101, no. 474, pp. 830–842, 2006.

[43] K. Knight, “Limiting distributions for L_1 regression estimators under general conditions,” *Ann. Stat.*, vol. 26, pp. 755–770, 1998.

[44] D. Pollard, “Asymptotics for least absolute deviation regression estimators,” *Econometric Theory*, vol. 7, pp. 186–199, 1991.

[45] M. A. Arcones, “Large deviations of empirical processes,” in *High Dimensional Probability III*, J. Hoffmann-Jorgensen, M. B. Marcus, and J. A. Wellner, Eds., pp. 205–224, Boston, MA: Birkhäuser, 2004.

[46] D. W. K. Andrews, “Generic uniform convergence,” *Econometric Theory*, vol. 8, pp. 241–257, 1992.

[47] K. L. Chung, *A Course in Probability Theory*, 3rd Edn., Theorem 7.1.2, p. 209, New York: Academic Press, 2001.