Estimation of the Frequency of Sinusoidal Signals in Laplace Noise

Ta-Hsin Li
Department of Mathematical Sciences
IBM T. J. Watson Research Center
Yorktown Heights, NY 10598-0218, USA
thl@us.ibm.com

Kai-Sheng Song
Department of Statistics
Florida State University
Tallahassee, FL 32306-4330, USA
kssong@stat.fsu.edu

***Abstract*— Accurate estimation of the frequency of sinusoidal signals from noisy observations is an important problem in signal processing applications such as radar, sonar, and telecommunications. In this paper, we study the problem under the assumption of non-Gaussian noise in general and Laplace noise in particular. We prove that the Laplace maximum likelihood estimator is able to attain the asymptotic Cramér-Rao lower bound under the Laplace assumption which is one half of the Cramér-Rao lower bound in the Gaussian case. This provides the possibility of improving the currently most efficient methods such as nonlinear least-squares and periodogram maximization in non-Gaussian cases. We propose a computational procedure that overcomes the difficulty of local extrema in the likelihood function when computing the maximum likelihood estimator. We also provide some simulation results to validate the proposed approach.**

I. Introduction

Consider the problem of estimating the frequency, ω, of a sinusoidal signal from noisy observations

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

where $A \in \mathbb{R}$, $B \in \mathbb{R}$, $\omega \in \Omega := (0,\pi)$ are unknown constants and $\{\varepsilon_t\}$ is a white noise process with mean zero and unknown variance $\sigma^2 > 0$. The literature on this subject is extensive [1]. The most popular approaches include Fourier transform (periodogram) [2]–[5], Gaussian maximum likelihood (or nonlinear least-squares) [6]–[8], autoregression [9]–[12], and eigendecomposition (signal/noise subspace) [13]–[18].

It is well known that under the assumption that $\{\varepsilon_t\}$ is Gaussian white noise, the asymptotic Cramér-Rao lower bound (CRLB) with respect to 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). In the following we shall refer to this bound as the Gaussian CRLB. Analytical as well as simulation studies suggest [4] [6] that the Gaussian CRLB can be attained asymptotically by the maximum likelihood estimator (MLE) which maximizes the Gaussian likelihood function, or equivalently, minimizes the sum of squared errors

$$\ell_2(\boldsymbol{\theta}) := \sum_{t=1}^{n} |y_t - (A\cos(\omega t) + B\sin(\omega t))|^2$$

as a function of $\boldsymbol{\theta} := [A, B, \omega]^T$. Several numerical procedures have been proposed to compute this estimator [6] [8].

When the noise is not Gaussian, the minimizer of $\ell_2(\boldsymbol{\theta})$ is known as the nonlinear least-squares (NLS) estimator. Under fairly general assumptions about $\{\varepsilon_t\}$, it has been shown [2] that the NLS estimator of ω is asymptotically equivalent to the maximizer of the continuous periodogram, $I_n(\omega) := n^{-1}|\sum y_t \exp(it\omega)|^2$, with $i := \sqrt{-1}$, and the asymptotic variance of this estimator coincides with the Gaussian CRLB.

Analytical and simulation studies show that typical frequency estimation procedures that exist in the literature either attain the Gaussian CRLB asymptotically (e.g., the NLS method [2]) or fall short of it (e.g., the signal/noise subspace methods [16]). The question is: can we do better than the Gaussian CRLB in non-Gaussian cases?

In this paper, we provide a positive answer to the question. Toward that end, we first examine the CRLB in non-Gaussian cases generally and show that the Gaussian CRLB is the worse-case performance limit, i.e., the largest lower bound, among a large family of noise distributions. Then, we focus on the case of Laplace noise and show that the Laplace MLE attains asymptotically the Laplace CRLB which is only one half of the Gaussian CRLB.

The Gaussian assumption of the noise is often made in practice for its mathematical tractability rather than its goodness of fit to the data. In reality, departures from the Gaussian assumption can occur in many different forms, one of which is heavy tails. A heavy-tailed distribution has greater tail probabilities than what the Gaussian model suggests. It manifests itself as outliers in the observations that can cause algorithms developed under the Gaussian assumption to malfunction. The Laplace distribution has heavier tails than the Gaussian distribution and therefore is often used to model heavy-tailed data in the statistical literature. The Laplace distribution can also be used as a surrogate in developing more robust algorithms against outliers or in solving problems that do not have a solution under the Gaussian assumption (i.e., blind deconvolution of non-minimum phase systems).

As with the methods of Gaussian maximum likelihood and periodogram maximization, it is very difficult to compute the Laplace MLE without an extremely good initial guess, because the likelihood function is full of local extrema in

† Kai-Sheng Song was supported in part by an IBM Faculty Award. The authors thank Dr. Tailen Hsing of the Ohio State University for helpful discussions concerning Theorem 3.

the vicinity of the desired solution. To solve this problem, we use a well-understood iterative filtering method, called the three-step algorithm (TSA), discussed in [19] [20], to provide the necessary initial guess for a general-purpose optimization routine that computes the final estimates. In addition to its unified architecture suitable for practical implementation, the TSA initialization procedure has the analytically proven property of fast and virtually global convergence to an estimate as accurate as the Gaussian MLE and the periodogram maximizer. Simulation confirms the validity of this approach.

II. CRLB for Non-Gaussian Noise

Let the ε_t be i.i.d. random variables with mean zero, variance $\sigma^2 > 0$, and probability density function $p(x)$. Assume that $p(x)$ is almost everywhere differentiable with derivative $\dot{p}(x)$, except at a finite number of points, such that

$$0 < \lambda := \sigma^2 \int_{p(x)>0} \frac{\{\dot{p}(x)\}^2}{p(x)}\, dx < \infty.$$

Theorem 1. *Under these assumptions, the Fisher information matrix of* $\mathbf{y} := [y_1, \ldots, y_n]^T$ *with respect to* $\boldsymbol{\theta} := [A, B, \omega]^T$ *can be expressed as*

$$I(\boldsymbol{\theta}) = \lambda\, I_g(\boldsymbol{\theta}).$$

In this expression, $I_g(\boldsymbol{\theta})$ *is the Fisher information matrix under the Gaussian assumption, i.e.,*

$$I_g(\boldsymbol{\theta}) := \frac{1}{\sigma^2} \mathbf{X}^T \mathbf{X},$$

where $\mathbf{X} := [\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3]$, $\mathbf{x}_1 := \mathrm{vec}[\cos(\omega t)]$, $\mathbf{x}_2 := \mathrm{vec}[\sin(\omega t)]$, *and* $\mathbf{x}_3 := \mathrm{vec}[-At\sin(\omega t) + Bt\cos(\omega t)]$.

Proof. The log-likelihood function of $\mathbf{y}$ can be written as

$$L(\boldsymbol{\theta} \mid \mathbf{y}) = \sum_{t=1}^{n} \log p(y_t - s_t(\boldsymbol{\theta}))$$

where $s_t(\boldsymbol{\theta}) := A\cos(\omega t) + B\sin(\omega t)$. Therefore,

$$\begin{aligned}
\frac{\partial L}{\partial A} &= \sum_{t=1}^{n} \frac{\dot{p}(\varepsilon_t)}{p(\varepsilon_t)} \cos(\omega t) \\
\frac{\partial L}{\partial B} &= \sum_{t=1}^{n} \frac{\dot{p}(\varepsilon_t)}{p(\varepsilon_t)} \sin(\omega t) \\
\frac{\partial L}{\partial \omega} &= \sum_{t=1}^{n} \frac{\dot{p}(\varepsilon_t)}{p(\varepsilon_t)} \{-At\sin(\omega t) + Bt\cos(\omega t)\}
\end{aligned}$$

so that

$$\begin{aligned}
I(\boldsymbol{\theta}) &:= E\{\nabla L(\boldsymbol{\theta} \mid \mathbf{y}) \nabla^T L(\boldsymbol{\theta} \mid \mathbf{y})\} \\
&= E\left\{ \left(\frac{\dot{p}(\varepsilon)}{p(\varepsilon)} \right)^2 \right\} (\mathbf{X}^T \mathbf{X}) \\
&= (\lambda/\sigma^2)(\mathbf{X}^T \mathbf{X}).
\end{aligned}$$

The proof is complete upon noting that $\lambda = 1$ in the Gaussian case where $p(x) = (2\pi\sigma^2)^{-1/2} \exp(-\frac{1}{2}x^2/\sigma^2)$. □

From Theorem 1, the following result about the CRLB can be obtained immediately.

Theorem 2. *Under the assumptions of Theorem 1 and the standard regularity conditions, the asymptotic CRLB for estimating* ω *on the basis of* $\mathbf{y}$ *can be expressed as*

$$CRLB(\boldsymbol{\theta}) = \lambda^{-1}\, CRLB_g(\boldsymbol{\theta}),$$

where

$$CRLB_g(\boldsymbol{\theta}) := \frac{1}{\gamma} \begin{bmatrix} (A^2 + 4B^2)/n & -3AB/n & -6B/n^2 \\ & (4A^2 + B^2)/n & 6A/n^2 \\ \textit{symmetry} & & 12/n^3 \end{bmatrix}$$

is the asymptotic CRLB under the Gaussian assumption.

Proof. The assertion follows from Theorem 1 combined with the fact that $\mathbf{X}^T\mathbf{X} = \mathbf{D}_n^{-1}\{\boldsymbol{\Sigma} + o(1)\}\mathbf{D}_n^{-1}$, where

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

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

In the next result, we show that the Gaussian CRLB is the worse-case performance limit among a large family of distributions.

Theorem 3. *Let* P *be the set of probability density functions satisfying the assumptions of Theorem 1 and having the entire real line* $\mathbb{R}$ *as their support. Then,* $\lambda \geq 1$ *for any* $p(x) \in P$, *with "=" if and only if* $\{\varepsilon_t\}$ *is Gaussian.*

Proof. Consider the problem of estimating $\theta \in \mathbb{R}$ on the basis of $Y \sim p(y - \theta)$. It is easy to show that the Fisher information of Y equals λ/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 $\lambda \geq 1$, with "=" iff $a(y - \theta) = (d/d\theta)\log p(y - \theta)$ for some constant $a \neq 0$. With $x := y - \theta$, this condition can be rewritten as $d\log p(x)/dx = -ax$, leading to $p(x) = \exp(-\frac{1}{2}ax^2 + bx + c)$. Imposing the constraints on the mean and variance completes the proof. □

An example of $p(x)$ that satisfies the assumptions is the Laplace distribution with

$$p(x) = (2c)^{-1} \exp\left(-|x|/c\right),$$

where $c := \sigma/\sqrt{2}$. It is easy to show that $\lambda = 2$. This implies that the CRLB under the Laplace assumption is one half of the CRLB under the Gaussian assumption.

III. Maximum Likelihood Estimation

Whereas typical procedures of frequency estimation only achieve the Gaussian CRLB at best, the CRLB under the Laplace assumption suggests the possibility of reducing the error by 50% when the noise has a Laplace distribution. The question is: can we get there?

To answer this question, we turn to the maximum likelihood method, which typically produces the most efficient estimates for sufficiently large sample sizes. Note that maximizing the

Laplace likelihood function is equivalent to minimizing the sum of absolute errors

$$\ell_1(\boldsymbol{\theta}) := \sum_{t=1}^{n} |y_t - (A\cos(\omega t) + B\sin(\omega t))|.$$

Therefore, instead of minimizing the ℓ_2 error, which only achieves the Gaussian CRLB, we minimize the ℓ_1 error. However, to prove the efficiency of this estimator is not a simple exercise. We cannot adopt the standard argument of asymptotic normality of the maximum likelihood estimators from i.i.d. observations, which can be found in many textbooks such as [21], 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.

Nonetheless, by employing more sophisticated mathematical tools and the concept of local asymptotic normality (LAN), we are able to prove the following result.

Theorem 4. *Assume that $\{\varepsilon_t\}$ is a Laplace white noise process with mean zero and variance σ^2. Let $\Theta_n := \{(a,b,c)\colon |a-A| \le \kappa_1 n^{-1/2}, |b-B| \le \kappa_2 n^{-1/2}, |c-\omega| \le \kappa_3 n^{-3/2}\} \subset \mathbb{R}\times\mathbb{R}\times\Omega$ be a neighborhood of the true parameter $\boldsymbol{\theta} := [A,B,\omega]^T$ for some positive constants κ_1, κ_1, and κ_3. Then, there exists a sequence $\{\hat{\boldsymbol{\theta}}_n\} \subset \Theta_n$ of local maxima of the Laplace likelihood function such that $\hat{\boldsymbol{\theta}}_n$ is asymptotically distributed as $N(\boldsymbol{\theta}, CRLB(\boldsymbol{\theta}))$.*

Proof. We omit the proof here, except to say that the key in proving the assertion is to establish that asymptotically,

$$R_n(\boldsymbol{\delta}) := \ell_1(\boldsymbol{\theta} + \mathbf{D}_n\boldsymbol{\delta}) - \ell_1(\boldsymbol{\theta})$$

can be approximated in distribution by the random process

$$R(\boldsymbol{\delta}) := \boldsymbol{\delta}^T \boldsymbol{z} + (\sqrt{2}/\sigma)\,\boldsymbol{\delta}^T \boldsymbol{\Sigma} \boldsymbol{\delta}$$

uniformly for all $\boldsymbol{\delta}$ in any compact subset of $\mathbb{R}^3$, where $\boldsymbol{z} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, with $\boldsymbol{\Sigma}$ defined in the proof of Theorem 2. □

Theorem 4 shows that by minimizing the ℓ_1 error instead of the ℓ_2 error, the Laplace MLE is able to reduce the variance of the NLS estimator by 50% in the case of Laplace noise.

IV. A Computational Procedure

Finding the minimizer of $\ell_1(\boldsymbol{\theta})$ 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 small vicinity of the true parameter. A similar problem was also experienced with the methods of NLS and periodogram maximization, where it has been demonstrated that an initial value of accuracy $o(n^{-1})$ in standard error is required for iterative search algorithms to converge to the desired solution [22]. A typical remedy to the problem is to use a DFT-based coarse search followed by a local interpolation refinement to generate suitable initial values for standard optimization routines such as the Newton-Raphson algorithm [2]–[8].

Rather than stitching together various methods of different flavors, we propose a unified procedure to produce the required initial values, using the simple parametric filtering (PF) method discussed in [12] [19] [20].

TABLE I

Role of Bandwidth Parameter

Bandwidth $1-\eta = \mathcal{O}(n^{-\nu})$	Initial Accuracy $\hat{\omega}_n^{(0)} - \omega$	Final Accuracy $\hat{\omega}_n - \omega$
$\nu = 0$	$\mathcal{O}(1)$	$\mathcal{O}(n^{-1/2})$
$\nu \in (\frac{1}{3}, \frac{1}{2})$	$\mathcal{O}(n^{-\nu\delta})$	$\mathcal{O}(n^{-(1+3\nu)/2})$
$\nu \in (\frac{1}{2}, 1)$	$\mathcal{O}(n^{-\nu\delta})$	$\mathcal{O}(n^{-(2+\nu)/2})$
$\nu = \infty$	$o(n^{-1})$	$\mathcal{O}(n^{-3/2})$

The gist of the method is the following [19]. For any given $\alpha \in (-2\eta/(1+\eta^2), 2\eta/(1+\eta^2))$ with fixed $\eta \in (0,1)$, let $y_t(\alpha)$ be obtained by filtering the data with the 2nd-order IIR filter $H(z^{-1}) := \{1-(1+\eta^2)\alpha z^{-1} + \eta^2 z^{-2}\}^{-1}$, i.e.,

$$y_t(\alpha) := H(z^{-1})\, y_t.$$

Given $\{y_t(\alpha)\}$, let $\rho_n(\alpha)$ be the minimizer of the weighted sum of forward and backward prediction error sums of squares $\sum\{y_t(\alpha) - \rho y_{t-1}(\alpha)\}^2 + \eta^2 \sum\{y_{t-2}(\alpha) - \rho y_{t-1}(\alpha)\}^2$, which turns out be

$$\rho_n(\alpha) = \frac{\sum_{t=1}^{n} y_{t-1}(\alpha)\,\{y_t(\alpha) + \eta^2 y_{t-2}(\alpha)\}^2}{(1+\eta^2)\sum_{t=1}^{n}\{y_{t-1}(\alpha)\}^2}.$$

Using this function of α, a sequence $\{\hat{\alpha}_n^{(m)}\}$ can be obtained from the accelerated version of fixed-point iteration

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

which can be regarded as an iteration of linear filtering followed by least squares. It can be shown [19] that with a suitable initial value the sequence converges to a fixed point $\hat{\alpha}_n$ of $\rho_n(\alpha)$, i.e.,

$$\lim_{m\to\infty} \hat{\alpha}_n^{(m)} = \hat{\alpha}_n = \rho_n(\hat{\alpha}_n)$$

almost surely for sufficiently large n. From this fixed point, a frequency estimator,

$$\hat{\omega}_n := \arccos(\hat{\alpha}_n),$$

is produced. The consistency and asymptotic normality of $\hat{\omega}_n$ as an estimator of ω can be established [19].

In this algorithm 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. The relationship is summarized in Table I.

Based on these results, the following three-step algorithm (TSA) was proposed in [19] 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 $1-\eta_1 = \mathcal{O}(1)$ to accommodate initial values of accuracy $\mathcal{O}(1)$; iterate $\mathcal{O}(n\log n)$ times to obtain an estimate of accuracy $\mathcal{O}(n^{-1/2})$.
2. Take $1-\eta_2 = \mathcal{O}(n^{-1/3})$ and use the result from Step 1 as the initial value; iterate $\mathcal{O}(1)$ times to get an estimate of accuracy $\mathcal{O}(n^{-1})$.

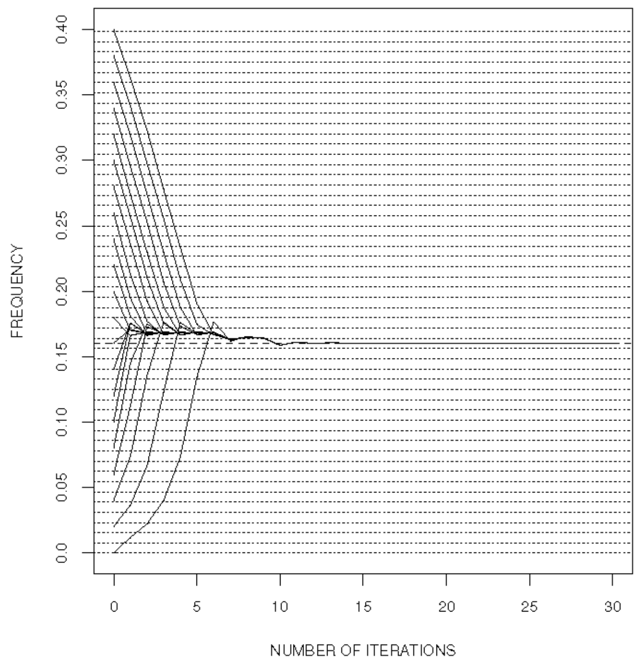

Fig. 1. Trajectory of normalized frequency estimates from the three-step algorithm for different initial values. Sample size is 128. Frequency normalization is defined as $\omega \mapsto f := \omega/(2\pi) \in (0, 0.5)$.

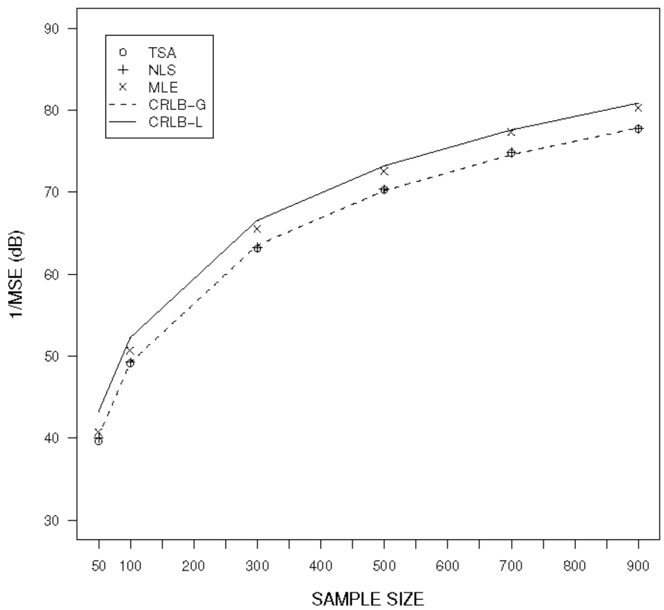

Fig. 2. Reciprocal MSE of the normalized frequency estimates and reciprocal CRLB for a sinusoidal signal in Laplace white noise. Solid line, Laplace CRLB; dotted line, Gaussian CRLB; $\times$, Laplace MLE initialized by the three-step algorithm; $+$, NLS initialized by the three-step algorithm; $\circ$, estimates from the three-step algorithm.

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

The global convergence property of this algorithm can be appreciated from Fig. 1, where all initial values lead to a convergence to the desired solution after about 12 iterations (6 iterations with $\eta = 0.75$, 3 more iterations with $\eta = 0.84$, and all remaining iterations with $\eta = 0.98$).

In practice, one can use Prony's estimator to initialize Step 1. Prony's estimator corresponds to $\rho_n(\alpha)$ with $\alpha = 0$ and $\eta = 0$ (no filtering) and has accuracy $\mathcal{O}(1)$ (because it is biased). 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.

It can be shown [23] that with $\eta = 1$ the estimator $\hat{\omega}_n$ is able to attain the Gaussian CRLB asymptotically. Simulation indicates that even with $\eta < 1$ (but sufficiently close to 1) the Gaussian CRLB can be attained by the estimator for samples sizes as small as 50. Therefore, starting with virtually any initial guess as the input, the TSA procedure should be able to produce an output accurate enough to initialize general-purpose optimization routines for minimizing $\ell_1(\boldsymbol{\theta})$. Initial values of A and B can be obtained easily by least-squares regression using the frequency estimate from the TSA procedure.

Because $\ell_1(\boldsymbol{\theta})$ is not everywhere differentiable, gradient-based algorithms, such as the Gauss-Newton algorithm for NLS, cannot be used to compute the Laplace MLE. Fortunately, there are a plenty of general-purpose algorithms that do not require the differentiability. The simplex algorithm of Nelder and Mead [24], available in many software packages such as Mathematica and `R`, is such an example. There are also special algorithms designed for nonlinear regression with the ℓ_1 norm. For example, the interior point algorithm proposed in [25] for nonlinear quantile regression is readily available for `R` (the function `nlrq` in the `quantreg` package).

Fig. 2 shows the result of a simulation study in the case of Laplace noise based on 1,000 Monte Carlo trials for each sample size. The signal and noise parameters are: $A = 1$, $B = 0$, $\omega = 0.15 \times 2\pi$, and $\gamma = 0$ dB (normalized for each trial). The three values of η in the TSA procedure are 0.85, $1-n^{-0.6}$, and $1-n^{-0.9}$. The numbers of iterations with these values are 6, 3, and 11 for each trial. The TSA procedure is initialized with Prony's estimator which is known to be biased regardless of the sample size and have a standard error $\mathcal{O}(n^{-1/2})$, so the initial MSE is merely $\mathcal{O}(1)$. Fig. 2 shows that the estimates from the TSA procedure attain the Gaussian CRLB for all the sample sizes. It also shows that using the TSA estimates as initial values to minimize $\ell_2(\boldsymbol{\theta})$ by a standard optimization routine (in this case the Nelder-Mead algorithm) does not lead to an improved accuracy. However, by replacing $\ell_2(\boldsymbol{\theta})$ with $\ell_1(\boldsymbol{\theta})$, the initial TSA estimates are improved considerably, and the final estimates (the Laplace MLE) closely follow the Laplace CRLB, as predicted by Theorem 2, except for the smallest sample size $n = 50$. As expected, the improvement is about 3 dB, or a 50% reduction in MSE.

V. Concluding Remarks

In this paper we have demonstrated the possibility of achieving more accurate frequency estimates than the Gaussian CRLB suggests for sinusoidal signals in non-Gaussian noise. In particular, we have shown that in the case of Laplace noise the maximum likelihood estimator, which minimizes the sum of absolute errors, is able to attain asymptotically the Laplace CRLB which is 50% smaller than the Gaussian CRLB attained by nonlinear least squares and periodogram maximization.

In addition to the theoretical findings, we have also proposed a computational procedure to obtain the maximum likelihood estimates numerically. The procedure utilizes an iterative algorithm proposed in [19] [20] to produce sufficiently accurate initial values for standard optimization routines. Owing to the global convergence property of the initialization algorithm, the proposed procedure is able to accommodate poor initial values of accuracy $\mathcal{O}(1)$ and produce a final estimator of accuracy $\mathcal{O}(n^{-3/2})$ that attains the Laplace CRLB for sufficiently large sample sizes.

The proposed procedure is also generalized to the case of multiple sinusoids, provided that the frequencies satisfy a separation condition [20]. In principle, one can also generalize the problem by assuming that the noise has a generalized Gaussian distribution of the form $p(x) \propto \exp(-|x/c|^{\alpha})$, where $\alpha > 0$ is a predetermined constant and $c > 0$ is the scale parameter. Note that for Gaussian noise, $\alpha = 2$ and for Laplace noise, $\alpha = 1$. Under this assumption, the maximum likelihood estimator of $\boldsymbol{\theta}$ can be found by minimizing $\ell_{\alpha}(\boldsymbol{\theta}) := \sum |y_t - (A\cos(\omega t) + B\sin(\omega t))|^{\alpha}$. Because special mathematical and computational tools are needed to handle the general case of $\alpha \neq 1, 2$, this problem deserves further investigation.

References

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

[2] 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.

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

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

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

[11] 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.

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

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

[14] 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.

[15] 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.

[16] 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.

[17] 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.

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

[19] 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.

[20] 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.

[21] E. L. Lehman, *Theory of Point Estimation*, Wiley, New York, 1983.

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

[23] 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.

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

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