Iterative Quadratic Maximum Likelihood based Estimator for a Biased Sinusoid

Frankie K. W. Chan, H. C. So, Md. Tawfiq Amin, C. F. Chan and W. H. Lau

Department of Electronic Engineering

City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong

Email : ckwf@hkexperts.com

Index terms : Parameter estimation, sinusoid, offset, linear prediction, maximum likelihood

Abstract

The problem of parameter estimation of a single sinusoid with unknown offset in additive Gaussian noise is addressed. After deriving the linear prediction property of the noise-free signal, the maximum likelihood estimator for the frequency parameter is developed. The optimum estimator is relaxed according to the iterative quadratic maximum likelihood technique. The remaining parameters are then solved in a linear least squares manner. Theoretical variance expression of the frequency estimate based on high signal-to-noise ratio assumption is also derived. Simulation results show that the proposed approach can give optimum estimation performance and is superior to the nonlinear least squares.

I. Introduction

Parameter estimation of sinusoidal signals in additive noise has been an active research topic [1]–[5] because of its numerous application areas in power delivery [6], signal processing [7], digital communications [8], instrumentation and measurement [9] and so on. In this paper, we tackle the parameter estimation problem for a real biased sinusoid. The observed signal, which is also known as the four-parameter sine wave model [9]–[10], is:

$$x(n) = s(n) + q(n), \qquad n = 1, 2, ..., N \tag{1}$$

where

$$s(n) = A\cos(\omega n + \phi) + B \tag{2}$$

where $A > 0$, $\omega \in (0, \pi)$, $\phi \in [0, 2\pi)$ and B are deterministic but unknown constants which denote the tone amplitude, frequency, phase and offset, respectively, while $q(n)$ is the additive zero-mean Gaussian noise. Our objective is to estimate A, ω, ϕ and B, from the N discrete-time noisy measurements $\{x(n)\}$.

Although there are numerous sinusoidal parameter estimation schemes in the literature such as maximum likelihood (ML), nonlinear least squares (NLS) [4],[10], iterative quadratic maximum likelihood (IQML) [11]–[12], linear prediction (LP) [2], most of them assume $B = 0$. In fact, it is of interest to estimate the non-zero offset or DC value as well [9]–[10]. In this paper, we contribute to the development of an accurate and computationally attractive parameter estimation approach for single tone with non-zero offset. We first derive the LP property of (2) and then produce the ML estimator for the frequency parameter. As the ML cost function is multi-modal, IQML-based relaxation is utilized to yield a simple iterative algorithm. The parameters of amplitude, phase and offset are then obtained in a linear least squares (LLS) manner. Furthermore, the variance of the frequency estimate in high signal-to-noise ratio (SNR) conditions is theoretically analyzed. The effectiveness of the proposed scheme is demonstrated by comparing with the NLS approach and Cramér-Rao lower bound (CRLB) [10].

II. Algorithm Development

It is well known that when $B = 0$, $s(n)$ obeys the LP property of $s(n)+s(n-2) = 2\cos(\omega)s(n-1)$. For nonzero B, its extension is:

$$\begin{aligned} s(n) + s(n-2) - 2B &= 2\cos(\omega)\left(s(n-1) - B\right) \\ \Rightarrow s(n) + s(n-2) - 2\cos(\omega)s(n-1) &= 2B\left(1 - \cos(\omega)\right) \end{aligned} \tag{3}$$

where we note that $2B\left(1-\cos(\omega)\right)$ is independent of the index n. Substituting n with $n-1$ in (3) yields another equality and equating with (3), we have:

$$\begin{aligned} s(n)+s(n-2)-2\cos(\omega)s(n-1) &= s(n-1)+s(n-3)-2\cos(\omega)s(n-2) \\ \Rightarrow s(n)-s(n-3)-(2\cos(\omega)+1)(s(n-1)-s(n-2)) &= 0 \end{aligned} \tag{4}$$

which is the LP property of single real tone with offset. As in conventional sinusoidal parameter estimation, finding ω is the first and crucial step because it is a nonlinear function in the received data sequence. The remaining parameters, namely, A, ϕ and B can then be estimated in a more straightforward manner after its determination.

Let $\rho = 2\cos(\omega)+1$ and define $\mathbf{x}_i = [x(i), x(i+1), \cdots, x(i+N-4)]^T$, $i = 1,2,3,4$, where ^{T} denotes the transpose operator. Using (4) and following the development in [13], it is shown that the ML estimate for ρ, denoted by $\hat{\rho}$, in the presence of Gaussian noise can be determined from the following minimization problem:

$$\hat{\rho} = \arg\min_{\tilde{\rho}} \left(\mathbf{x}_4-\mathbf{x}_1-\tilde{\rho}\left(\mathbf{x}_3-\mathbf{x}_2\right)\right)^T \mathbf{\Sigma}\left(\tilde{\rho}\right)^{-1}\left(\mathbf{x}_4-\mathbf{x}_1-\tilde{\rho}\left(\mathbf{x}_3-\mathbf{x}_2\right)\right) \tag{5}$$

where $\tilde{\rho}$ is the optimization variable for ρ and $^{-1}$ represents the matrix inverse. The covariance matrix $\mathbf{\Sigma}(\tilde{\rho})$ is also a function of $\tilde{\rho}$ and has the form of $\mathrm{E}\{\mathbf{p}\mathbf{p}^T\}$ with $\mathbf{p} = [p(1), p(2), \cdots, p(N-3)]^T$ whose element is $p(n) = q(n+3)-q(n)-\tilde{\rho}\left(q(n+2)-q(n+1)\right)$, $n = 1,2,\cdots,N-3$. For zero-mean white Gaussian noise, $\mathbf{\Sigma}\left(\tilde{\rho}\right)$ is expressed as:

$$\begin{aligned} \mathbf{\Sigma}\left(\tilde{\rho}\right) &= \mathrm{Toeplitz}\left(\begin{bmatrix} 2\left(\tilde{\rho}^2+1\right) & -\tilde{\rho}^2-2\tilde{\rho} & 2\tilde{\rho} & -1 & 0\cdots 0 \end{bmatrix}\right)\sigma^2 \\ &= \begin{bmatrix} 2\left(\tilde{\rho}^2+1\right) & -\tilde{\rho}^2-2\tilde{\rho} & 2\tilde{\rho} & -1 & 0 & \cdots & 0 \\ -\tilde{\rho}^2-2\tilde{\rho} & 2\left(\tilde{\rho}^2+1\right) & -\tilde{\rho}^2-2\tilde{\rho} & 2\tilde{\rho} & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \cdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -\tilde{\rho}^2-2\tilde{\rho} & 2\left(\tilde{\rho}^2+1\right) \end{bmatrix}\sigma^2 \end{aligned} \tag{6}$$

It is noteworthy that this ML estimator can be extended to more general Gaussian process of $q(n)$ as long as $\mathbf{\Sigma}\left(\tilde{\rho}\right)$ is known up to a scalar. However, the objective function in (5) is multi-modal and thus there is no guarantee that the globally optimum point can be obtained.

Utilizing the idea of the IQML technique [11]–[12], we relax (5) into a quadratic function by considering $\mathbf{\Sigma}\left(\tilde{\rho}\right)$ is independent of $\tilde{\rho}$ so that global optimization is attained and the estimate of ρ, denoted by $\hat{\rho}$, is easily computed as:

$$\hat{\rho} = \frac{(\mathbf{x}_3 - \mathbf{x}_2)^T \boldsymbol{\Sigma}(\hat{\rho})^{-1} (\mathbf{x}_4 - \mathbf{x}_1)}{(\mathbf{x}_3 - \mathbf{x}_2)^T \boldsymbol{\Sigma}(\hat{\rho})^{-1} (\mathbf{x}_3 - \mathbf{x}_2)} \tag{7}$$

Note that σ^2 of $\boldsymbol{\Sigma}(\tilde{\rho})$ in the numerator and denominator cancel each other and hence its value is not required to be known. We iterate (7) with an initial guess of $\boldsymbol{\Sigma}(\hat{\rho})$ while the estimated ρ is then employed to update $\boldsymbol{\Sigma}(\hat{\rho})$. The iterative procedure of the proposed estimator for ω is summarized as follows:

(i) Set $\boldsymbol{\Sigma}(\hat{\rho}) = \mathbf{I}_{N-3}$ which is the $(N-3) \times (N-3)$ identity matrix.

(ii) Compute $\hat{\rho}$ using (7).

(iii) Use $\hat{\rho}$ to construct $\boldsymbol{\Sigma}(\hat{\rho})$ of (6).

(iv) Repeat Steps (ii) and (iii) until a stopping criterion is reached. In this study, we terminate for a fixed number of iterations.

(v) Compute $\hat{\omega}$ using:

$$\hat{\omega} = \cos^{-1}\left(\frac{\hat{\rho} - 1}{2}\right) \tag{8}$$

Employing $\hat{\omega}$, the estimates of A, ϕ and B, denoted by $\hat{A}$, $\hat{\phi}$ and $\hat{B}$, respectively, are obtained by minimizing the following LLS cost function:

$$(\boldsymbol{\Xi}\boldsymbol{\kappa} - \mathbf{x})^T (\boldsymbol{\Xi}\boldsymbol{\kappa} - \mathbf{x}) \tag{9}$$

where

$$\boldsymbol{\Xi} = \begin{bmatrix} \cos(\hat{\omega}) & \cos(2\hat{\omega}) & \cdots & \cos(N\hat{\omega}) \\ -\sin(\hat{\omega}) & -\sin(2\hat{\omega}) & \cdots & -\sin(N\hat{\omega}) \\ 1 & 1 & \cdots & 1 \end{bmatrix}^T$$

$$\boldsymbol{\kappa} = \begin{bmatrix} A\cos(\phi) & A\sin(\phi) & B \end{bmatrix}^T$$

and

$$\mathbf{x} = \begin{bmatrix} x(1) & x(2) & \cdots & x(N) \end{bmatrix}^T$$

From (9), the LLS estimate of $\boldsymbol{\kappa}$ is

$$\hat{\boldsymbol{\kappa}} = \begin{bmatrix} [\hat{\boldsymbol{\kappa}}]_1 & [\hat{\boldsymbol{\kappa}}]_2 & [\hat{\boldsymbol{\kappa}}]_3 \end{bmatrix}^T = \left(\boldsymbol{\Xi}^T\boldsymbol{\Xi}\right)^{-1} \boldsymbol{\Xi}^T \mathbf{x} \tag{10}$$

which gives

$$\hat{A} = \sqrt{[\hat{\boldsymbol{\kappa}}]_1^2 + [\hat{\boldsymbol{\kappa}}]_2^2} \tag{11}$$

$$\hat{\phi} = \tan^{-1}\left(\frac{[\hat{\boldsymbol{\kappa}}]_2}{[\hat{\boldsymbol{\kappa}}]_1}\right) \tag{12}$$

and

$$\hat{B} = [\hat{\boldsymbol{\kappa}}]_3 \tag{13}$$

III. Variance Analysis

In this section, the variance of the $\hat{\omega}$ is analyzed based on high SNR assumption. Let $\mathbf{y} = \mathbf{x}_3 - \mathbf{x}_2 = \bar{\mathbf{y}} + \boldsymbol{\Delta}\mathbf{y}$ and $\mathbf{z} = \mathbf{x}_4 - \mathbf{x}_1 = \bar{\mathbf{z}} + \boldsymbol{\Delta}\mathbf{z}$ where $\bar{\mathbf{g}}$ and $\boldsymbol{\Delta}\mathbf{g}$ are the noise-free version and perturbation of $\mathbf{g}$, respectively. Upon parameter convergence, (7) implies

$$f(\hat{\rho}) = \mathbf{y}^T \boldsymbol{\Sigma}(\hat{\rho})^{-1} (\mathbf{y}\hat{\rho} - \mathbf{z}) = 0 \tag{14}$$

The $f(\hat{\rho})$ can be linearized using Taylor's series as:

$$0 = f(\hat{\rho}) \approx f(\rho) + f'(\rho)\,\boldsymbol{\Delta}\rho \tag{15}$$

where

$$\begin{aligned} f(\rho) &= \mathbf{y}^T \boldsymbol{\Sigma}(\rho)^{-1} (\mathbf{y}\rho - \mathbf{z}) \\ f'(\rho) &= \mathbf{y}^T \boldsymbol{\Sigma}(\rho)^{-1} \mathbf{y} + \mathbf{y}^T \boldsymbol{\Sigma}(\rho)^{-1} \boldsymbol{\Sigma}(\rho)' \boldsymbol{\Sigma}(\rho)^{-1} (\mathbf{z} - \mathbf{y}\rho) \\ \boldsymbol{\Delta}\rho &= \hat{\rho} - \rho \end{aligned}$$

Here, $f'(\rho)$ and $\boldsymbol{\Sigma}(\rho)'$ are the first derivatives of $f(\rho)$ and $\boldsymbol{\Sigma}(\rho)$, respectively. Although $f(\hat{\rho})$ has multiple roots and we cannot guarantee that our obtained root from the iterative procedure corresponds to ρ, it is expected that $\hat{\rho}$ will be located at a reasonable proximity of ρ when SNR is sufficiently large. By using $\bar{\mathbf{z}} = \bar{\mathbf{y}}\rho$ and neglecting second-order perturbation terms, $f(\rho)$ can be rewritten as

$$(\bar{\mathbf{y}} + \boldsymbol{\Delta}\mathbf{y})^T \boldsymbol{\Sigma}(\rho)^{-1} ((\bar{\mathbf{y}} + \boldsymbol{\Delta}\mathbf{y})\rho - (\bar{\mathbf{z}} - \boldsymbol{\Delta}\mathbf{z})) \approx \bar{\mathbf{y}}^T \boldsymbol{\Sigma}(\rho)^{-1} (\boldsymbol{\Delta}\mathbf{y}\rho - \boldsymbol{\Delta}\mathbf{z}) \tag{16}$$

As only first-order terms are retained, $f'(\rho)\,\boldsymbol{\Delta}\rho$, using $\bar{\mathbf{z}} = \bar{\mathbf{y}}\rho$, is approximated as

$$f'(\rho)\,\boldsymbol{\Delta}\rho \approx \bar{\mathbf{y}}^T \boldsymbol{\Sigma}(\rho)^{-1} \bar{\mathbf{y}}\boldsymbol{\Delta}\rho \tag{17}$$

Based on (15)–(17), we have

$$\boldsymbol{\Delta}\rho \approx \frac{\bar{\mathbf{y}}^T \boldsymbol{\Sigma}(\rho)^{-1} (\boldsymbol{\Delta}\mathbf{y}\rho - \boldsymbol{\Delta}\mathbf{z})}{\bar{\mathbf{y}}^T \boldsymbol{\Sigma}(\rho)^{-1} \bar{\mathbf{y}}} \tag{18}$$

The variance of $\hat{\rho}$, denoted by var($\hat{\rho}$), is obtained by squaring both sides of (18) and taking expectation

$$\begin{aligned}\text{var}(\hat{\rho}) &= E\left\{(\boldsymbol{\Delta}\rho)^2\right\} \\ &\approx \frac{\bar{\mathbf{y}}^T\boldsymbol{\Sigma}(\rho)^{-1}E\left\{(\boldsymbol{\Delta}\mathbf{y}\rho-\boldsymbol{\Delta}\mathbf{z})(\boldsymbol{\Delta}\mathbf{y}\rho-\boldsymbol{\Delta}\mathbf{z})^T\right\}\boldsymbol{\Sigma}(\rho)^{-1}\bar{\mathbf{y}}}{\left(\bar{\mathbf{y}}^T\boldsymbol{\Sigma}(\rho)^{-1}\bar{\mathbf{y}}\right)^2} \\ &= \frac{\bar{\mathbf{y}}^T\boldsymbol{\Sigma}(\rho)^{-1}\left(\sigma^2\boldsymbol{\Sigma}(\rho)\right)\boldsymbol{\Sigma}(\rho)^{-1}\bar{\mathbf{y}}}{\left(\bar{\mathbf{y}}^T\boldsymbol{\Sigma}(\rho)^{-1}\bar{\mathbf{y}}\right)^2} \\ &= \frac{\sigma^2}{\bar{\mathbf{y}}^T\boldsymbol{\Sigma}(\rho)^{-1}\bar{\mathbf{y}}}\end{aligned} \tag{19}$$

where E denotes the expectation operator. By using (8) and (19) as well as the relationship of var($\hat{\omega}$) $\approx$ $(h'(\rho))^2$var($\hat{\rho}$) if $\hat{\omega} = h(\hat{\rho})$ where var($\hat{\omega}$) denotes the variance of $\hat{\omega}$, we get:

$$\text{var}(\hat{\omega}) \approx \frac{\text{var}(\hat{\rho})}{4\sin^2(\omega)} = \frac{\sigma^2}{4\bar{\mathbf{y}}^T\boldsymbol{\Sigma}(\rho)^{-1}\bar{\mathbf{y}}\sin^2(\omega)} \tag{20}$$

Note that the variance expression (20) is only valid for high SNR scenarios and may have deviation if the SNR is low. Although there is no closed-form expression for (20), Section IV shows that (20) is numerically identical to that of the CRLB for ω [10].

IV. Simulation Results

Computer simulations have been carried out to evaluate the estimation accuracy of the proposed approach for the signal model of (1)−−(2). We compare the mean square error (MSE) performance of the IQML-based method with that of the NLS approach as well as CRLB [10]. As the mean of sinusoidal signal is zero, the offset is approximately equal to the average of the data. Hence, we can obtain other parameters by solving the NLS problem $\min\limits_{A,\omega,\phi}\sum_{n=1}^{N}\left(\zeta(n)-A\cos(\omega n+\phi)\right)^2$ where $\zeta(n)=x(n)-\frac{\sum_{n=1}^{N}x(n)}{N}$. We refer this algorithm to as offset-removal NLS method. We use 3 iterations in the proposed method as no obvious improvement is observed for more iterations while the NLS and the offset-removal NLS algorithms are initialized by (7)−−(8) with $\boldsymbol{\Sigma}(\hat{\rho})=\mathbf{I}_{N-3}$. The parameters of interest have the following values: $A=\sqrt{2}$, $\omega=0.35\pi$, $\phi=1$ and $B=2$. The data length is assigned as $N=20$. We scale the zero-mean white Gaussian noise $q(n)$ to produce different SNR conditions with SNR $=(B^2+A^2/2)/\sigma^2$. All results provided are averages of 1000 independent runs. Figures 1 to 4 plot the MSEs at different SNRs for ω, A, ϕ and B, respectively. All four figures indicate that the proposed algorithm can achieve optimum estimation performance at sufficiently high SNRs, that is, when SNR $\geq$ 18 dB. In particular, Figure 1 shows that it is able to attain the CRLB when SNR $\geq$ 14 dB, which agrees with the variance expression for ω in (20). Although the MSEs of the proposed and NLS estimators can approach the corresponding optimum benchmarks, the former is superior to the latter in terms of threshold SNR performance. It is also seen in Figure 4 that the offset-removal NLS scheme produces biased offset estimates in higher SNR scenarios. This is because averaging of a real tone will produce a bias of order $O(1)$ even in the absence of noise, and its effect dominates compared with σ^2/N when SNR is sufficiently large. Estimation of the amplitudes, frequency and phase, based on this biased offset estimate, will then be suboptimal, which is demonstrated in Figures 1 to 3. Finally, we study the MSE performance of the crucial parameter of frequency versus ω at SNR = ?? dB. All other parameter settings are identical to the first test and the results are shown in Figure 5. Although (20) indicates that the variance of $\hat{\omega}$ is large when ω approaches 0 or π, the CRLB also has larger values in these conditions. Moreover, the equivalence of (20) and CRLB for the admissible range of $\omega \in (0, \pi)$ is demonstrated.

V. Conclusion

To conclude, we have devised an accurate parameter estimation approach for a single sinusoid with unknown offset parameter. The basic steps in the algorithm development are deriving the LP property of the noise-free signal, establishing the ML estimator, and performing relaxation to produce a simple iterative algorithm. The optimality of the proposed scheme is demonstrated by comparing with the NLS approach and CRLB via computer simulations. It is noteworthy that the LP property development can be utilized to the complex and/or multiple tone scenario in a straightforward manner.

References

[1] S. L. Marple, *Digital Spectral Analysis with Applications*, Englewood Cliffs, NJ : Prentice-Hall, 1987

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

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

[4] P. Stoica and R. Moses, *Spectral Analysis of Signals*, Upper Saddle River, NJ : Prentice-Hall, 2005

[5] F. Castani, Ed., *Spectral Analysis: Parametric and Non-Parametric Digital Methods*, Wiley-ISTE, 2006

[6] R. Chudamani, K. Vasudevan and C. S. Ramalingam, "Real-time estimation of power system frequency using nonlinear least squares," *IEEE Transactions on Power Delivery*, vol.24, no.3, pp.1021-1028, Jul. 2009

[7] M. G. Christensen, P. Stoica, A. Jakobsson and S. H. Jensen, "Multi-pitch estimation," *Signal Processing*, vol.88, no.4, pp.972-983, Apr. 2008

[8] S.M. Sameer and R. V. Raja Kumar, "An efficient technique for the integer frequency offset estimation in OFDM systems," *Signal Processing*, vol.89, no.2, pp.252-256, Feb. 2009

[9] *IEEE Standard for Digitizing Waveform Recorders*, IEEE Std. 1057, 1994

[10] P. Handel, "Properties of the IEEE-STD-1057 four-parameter sine wave fit algorithm," *IEEE Trans. Instrum. Meas.*, vol.49, no.6, pp.1189-1193, Dec. 2000

[11] R. Kumaresan, L. L. Scharf and A. K. Shaw, "An algorithm for pole-zero modeling and spectral analysis," *IEEE Transactions on Acoustic, Speech and Signal Processing*, vol.34, pp.637-640, June 1986

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

[13] M. Aoki and P. C. Yue, "On a priori error estimates of some identification methods," *IEEE Transactions on Automatic Control*, vol.15, no.5, pp.541-548, Oct. 1970

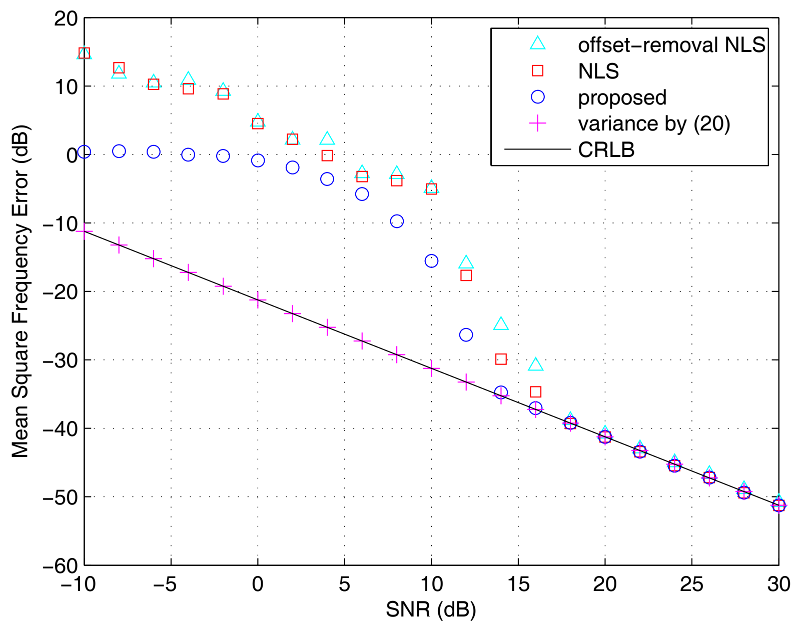

Fig. 1. Mean square error for ω versus SNR

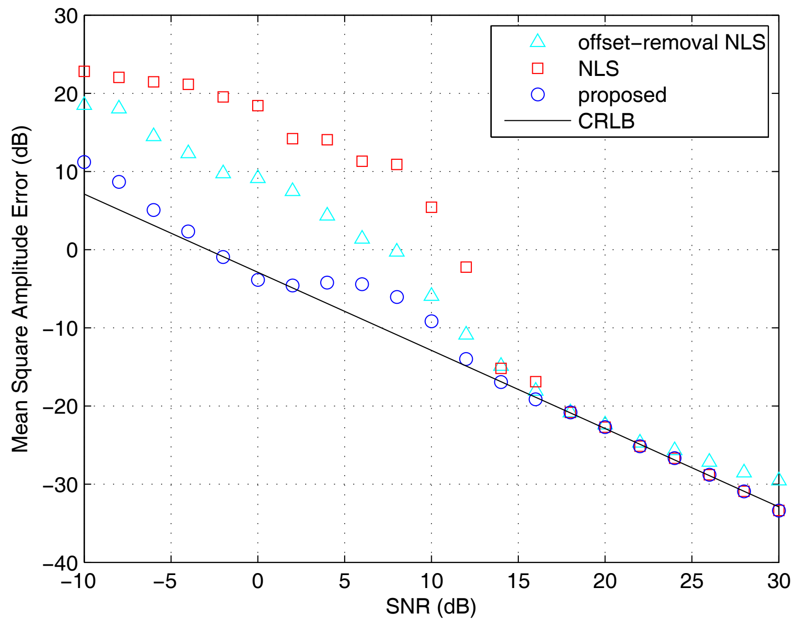

Fig. 2. Mean square error for A versus SNR

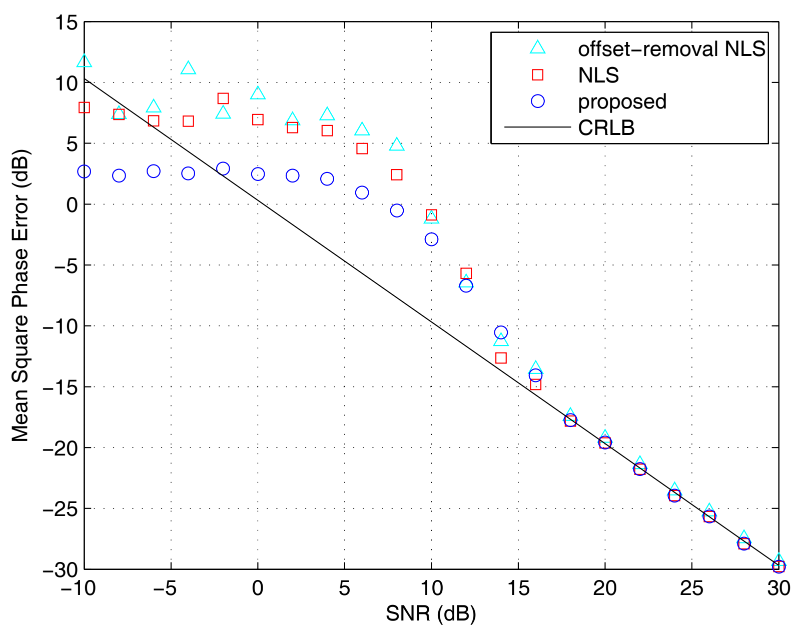

Fig. 3. Mean square error for ϕ versus SNR

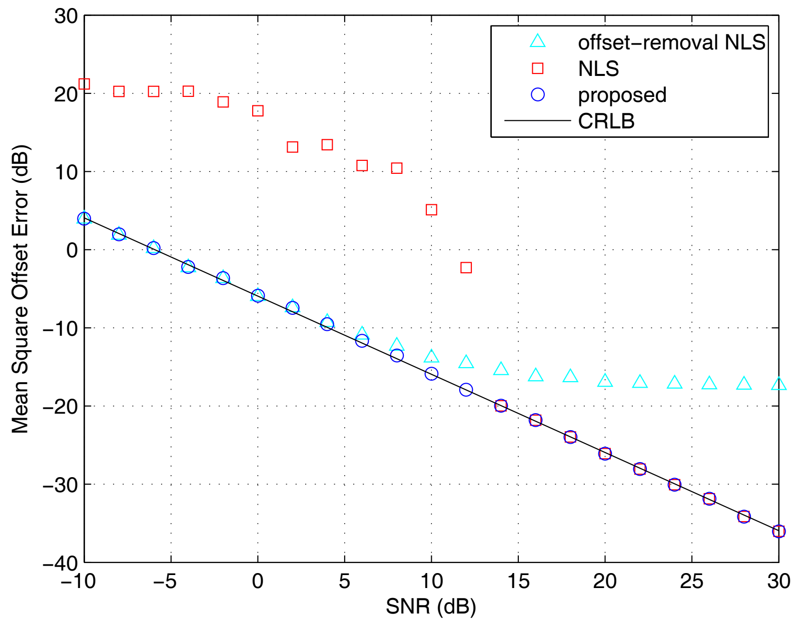

Fig. 4. Mean square error for B versus SNR

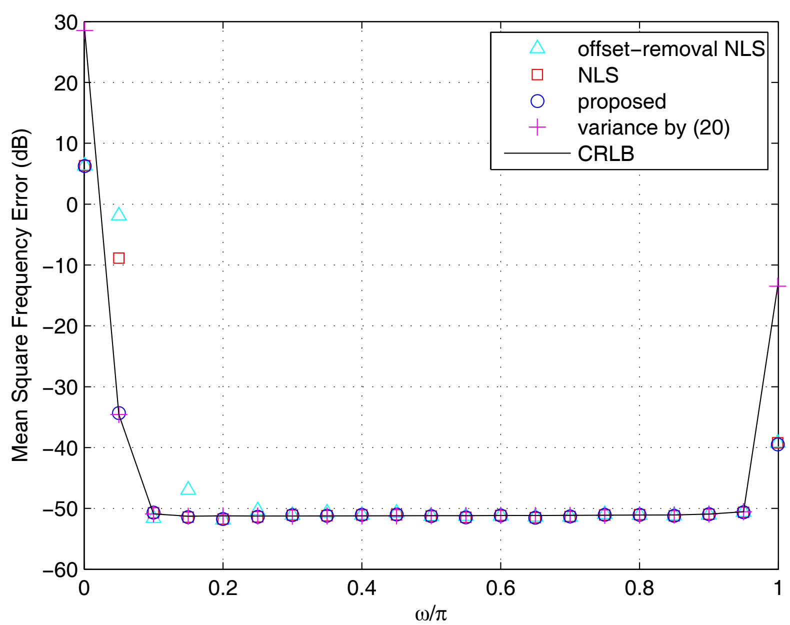

Fig. 5. Mean square error for ω versus ω

We would like to express our gratitude to the anonymous reviewers for their useful suggestions and criticism. The comments are well taken and the manuscript has been revised accordingly. Our responses to the comments are given as follows.

Responses to the Comments of Reviewer 1

1) *On page 6, equation 20, ω should have the hat.*

   Equation (20) is correct but there is a typo in the original manuscript. The sentence before (20) has been corrected as "... var$(\hat{\omega}) \approx (h'(\rho))^2var(\hat{\rho})$ if $\hat{\omega} = h(\hat{\rho})$...".

---

2) *According to (20), the variance of $\hat{\omega}$ may become very large when ω approaches zero. Does this require the caution in implementing the proposed estimator? How to avoid it if this may cause a problem?*

   Although (20) indicates that the variance of $\hat{\omega}$ is large when ω approaches 0 or π, this is the best we can do as (20) is equal to the CRLB. In the revised manuscript, we have included Figure 5 which illustrates the equivalence of (20) and CRLB for the admissible range of $\omega \in (0, \pi)$ and provided some elaborations on (20).

---

3) *The proposed method works well for high SNR situations as seen from simulation results. Two questions: (1) how SNR plays in the derivation of the proposed estimator is not clear. Please elaborate. (2) Suggestion on the SNR value to be sufficiently high?*

   Point taken. (1) The development of the algorithm in Section II is not affected by the SNR. While the variance analysis in Section III assumes sufficiently high SNR conditions as $f(\hat{\rho})$ has multiple roots and we cannot guarantee that our obtained root from the iterative procedure corresponds to ρ. Nevertheless, it is expected that $\hat{\rho}$ will be located at a reasonable proximity of ρ when SNR is sufficiently large. This elaboration has been included in Section III of the revised manuscript.
   (2) We have suggested in Section IV that SNR ≥ 18 dB corresponds to sufficiently high SNR conditions for the proposed algorithm to achieve optimum estimation performance at $N = 20$.

4) *Criterion for terminating the iterative procedure?*

Point taken. We have clearly stated that a fixed number of iterations is used for the stopping criterion on page 4 of the revised manuscript. In Section IV, we use 3 iterations in the proposed method as no obvious improvement is observed for more iterations.

Responses to the Comments of Reviewer 2

1) *In Section I, line 12, the authors is better to add some more recent references to support the statement that this problem is active area of research.*

Point taken. More recent references, namely, [4]–[8], have been included in the revised manuscript.

2) *Figure 4 presents the performance of different methods to estimate B against different SNRs. The question is that why the performance of offset removal NLS, which according to section IV estimates B by averaging, deteriorates when SNR increases. Averaging is an estimator whose CRLB is* $(\sigma^2)/N$ *where* σ^2 *is the variance of additive white Gaussian noise. The estimator's performance changes with N and* σ^2 *and doesn't change with SNR.*

If the signal only consists of the offset and zero-mean white noise, averaging gives an unbiased offset estimate with variance σ^2/N. However, in our case, the signal is a sum of the offset, noise as well as a real tone which will produce a bias of order $O(1)$ in the averaging even in the absence of noise. In high SNR scenarios, this bias effect dominates compared with σ^2/N, which explains the phenomenon of MSFE does not decrease with higher SNR in Figure 4. The estimation of the amplitudes, frequency and phase, based on this biased offset estimate, will certainly be suboptimal. We have included some elaboration on this issue in the revised manuscript.