Aalborg Universitet

AALBORG UNIVERSITY

DENMARK

On The Estimation of Low Fundamental Frequencies

Christensen, Mads Græsbøll

ON THE ESTIMATION OF LOW FUNDAMENTAL FREQUENCIES

Mads Græsbøll Christensen

Dept. of Architecture, Design and Media Technology
Aalborg University, Denmark
mgc@create.aau.dk

ABSTRACT

In this paper, we analyze the difficult problem of estimating low fundamental frequencies from periodic signals, like those produced by musical instruments. The problem arises when the fundamental frequency is low for a given number of samples as this causes the harmonics to overlap in the frequency domain. Moreover, we demonstrate how the performance of estimators can generally be improved by avoiding the asymptotic approximations that are commonly used in, for example, the harmonic summation method.

Index Terms— Pitch estimation, fundamental frequency estimation, spectral estimation

1. INTRODUCTION

The problem of estimating the fundamental frequency of a periodic (or approximately periodic) signal is one of the classical problems in speech and audio processing, as the fundamental frequency can be used in a myriad of applications, including coding, analysis, tuning, transcription, enhancement, and separation. Among the methodologies that have been employed to this problem maximum likelihood, least-squares, auto-/cross-correlation and related methods, linear prediction, filtering, and subspace methods can be mentioned [1–8]. For an overview, we refer the interested reader to [9]. Most of the commonly employed methods are based the assumption that the harmonics of periodic waveforms are well-separated in the spectrum, meaning that they do not overlap significantly. This assumption is, however, not accurate when the fundamental frequency is low for a given number of samples. It is, though, the case for any non-zero fundamental frequency for an infinite number of samples, and the assumption can hence be seen as an asymptotic approximation.

In this paper, we will analyze the problem of estimating low fundamental frequencies by avoiding the aforementioned asymptotic approximation in a) the computation of estimation bounds, more specifically the Cramér-Rao lower bound (CRLB) for the problem at hand, and in b) a specific estimator, namely the nonlinear least-squares (NLS) method. Both can then be said to be exact. The CRLB reveals the nature

of the problem and quantifies what is to be expected while the NLS method demonstrates what can be achieved with an actual method.

The remainder of the paper is organized as follows: In Section 2, we introduce the signal model, define the problem of interest and derive an expression for the CRLB for it. In Section 3, we then derive an exact estimator based on the NLS method. Finally, we present some simulation results in Section 4 before concluding on our work in Section 5.

2. MODEL, PROBLEM, AND BOUND

We will now proceed to define the problem of interest and the associated signal model. The observed real signal $x(n)$ is composed of a set of L sinusoids having frequencies that are integer multiples of a fundamental frequency $\omega_0 > 0$, real amplitude $A_l > 0$, and phases $\phi_l \in [0, 2\pi)$. Aside from the sinusoids, we assume that an additive noise source $e(n)$ is present. This noise source represents all stochastic signal components, even those that are inherent and important parts of natural signals. It is here assumed to be white and Gaussian distributed having variance σ^2. Mathematically, the observed signal can be expressed for $n = 0, \ldots, N - 1$ as

$$x(n) = \sum_{l=1}^{L} A_l \cos\left(\omega_0 l n + \phi_l\right) + e(n). \qquad (1)$$

The problem is then to estimate ω_0 from $x(n)$. For a given L, the fundamental frequency can be in the range $\omega_0 \in (0, \frac{\pi}{L})$. Regarding the remaining unknown parameters, some comments are in order. The model order, L, (also referred to as the number of harmonics) can be found in a variety of ways (see [9] for a review of these), and it is possible to solve jointly for the fundamental frequency and the model order. Once the fundamental frequency and the model order L has been found, the corresponding phases and amplitudes can be found using one of the existing amplitude estimators [10]. Compared to the problem of estimating the fundamental frequency, this is fairly easy, as these parameters are linear.

An estimate $\hat{\theta}_i$ of the ith real parameter θ_i of the parameter vector $\boldsymbol{\theta}$ is unbiased if its expected value the is identical to the true value. The CRLB is a lower bound on the variance

of such an estimate, and it is given by $\text{var}(\hat{\theta}_i) \geq \left[\mathbf{I}^{-1}(\boldsymbol{\theta}) \right]_{ii}$. Here, the notation $[\mathbf{I}(\boldsymbol{\theta})]_{il}$ means the ilth entry of the matrix $\mathbf{I}(\boldsymbol{\theta})$ and $\text{var}(\cdot)$ denotes the variance. Furthermore, $\mathbf{I}(\boldsymbol{\theta})$ is the Fisher information matrix. For the case of Gaussian signals with $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}(\boldsymbol{\theta}), \mathbf{Q})$ where $\mathbf{Q}$ is the noise covariance matrix (which is not parametrized by any of the parameters in $\boldsymbol{\theta}$) and $\boldsymbol{\mu}(\boldsymbol{\theta})$ is the mean, the likelihood function is given by

$$p(\mathbf{x}; \boldsymbol{\theta}) = \frac{1}{\det\left(2\pi\mathbf{Q}\right)^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}(\boldsymbol{\theta}))^T \mathbf{Q}^{-1}(\mathbf{x}-\boldsymbol{\mu}(\boldsymbol{\theta}))}. \quad (2)$$

For this case, Slepian-Bang's formula can be used for determining a more specific expression for the Fisher information matrix. More specifically, it is given by

$$[\mathbf{I}(\boldsymbol{\theta})]_{nm} = \frac{\partial \boldsymbol{\mu}^T(\boldsymbol{\theta})}{\partial \theta_n} \mathbf{Q}^{-1} \frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \theta_m}. \quad (3)$$

For the problem and signal model considered here, the involved quantities are $\mathbf{Q} \triangleq \sigma^2 \mathbf{I}$, $\boldsymbol{\mu}(\boldsymbol{\theta}) \triangleq \mathbf{Za}$ and

$$\mathbf{x} \triangleq \left[\, x(0) \; \cdots \; x(N-1) \, \right]^T \quad (4)$$
$$\boldsymbol{\theta} \triangleq \left[\, \omega_0 \; A_1 \; \phi_1 \; \cdots \; A_L \; \phi_L \, \right]^T$$
$$\mathbf{Z} \triangleq \left[\, \mathbf{z}(\omega_0) \; \mathbf{z}^*(\omega_0) \; \cdots \; \mathbf{z}(\omega_0 L) \; \mathbf{z}^*(\omega_0 L) \, \right],$$
$$\mathbf{a} \triangleq \frac{1}{2} \left[A_1 e^{j\phi_1} \; A_1 e^{-j\phi_l} \; \cdots \; A_L e^{j\phi_L} \; A_L e^{-j\phi_L} \right]^T$$
$$\mathbf{z}(\omega_0 l) \triangleq \left[\, 1 \; e^{j\omega_0 l 1} \; \cdots \; e^{j\omega_0 l(N-1)} \, \right]^T.$$

Note that for colored noise, the present derivations and results still hold provided that pre-whitening is applied. In relation to the problem at hand, some observations about the nature of the matrix $\mathbf{Z}$ can be made: Firstly, for $\omega_0 \neq 0$ and $\omega_0 \in (0, \frac{\pi}{L})$, $\mathbf{Z}$ has full rank. However, for $\omega_0 = 0$, it will be rank deficient and as $\omega_0 \to 0$, the condition number of $\mathbf{Z}$ will tend to infinity and the involved estimation problem is basically ill-conditioned.

With the above in place, we now have to determine the following derivatives:

$$\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \omega_0} = \frac{\partial \mathbf{Z}}{\partial \omega_0} \mathbf{a}, \; \frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial A_l} = \mathbf{Z} \frac{\partial \mathbf{a}}{\partial A_l}, \; \frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \phi_l} = \mathbf{Z} \frac{\partial \mathbf{a}}{\partial \phi_l}, \quad (5)$$

which are given by

$$\frac{\partial \mathbf{Z}}{\partial \omega_0} = \left[\, \frac{\partial \mathbf{z}(\omega_0)}{\partial \omega_0} \; \frac{\partial \mathbf{z}^*(\omega_0)}{\partial \omega_0} \; \cdots \; \frac{\partial \mathbf{z}(\omega_0 L)}{\partial \omega_0} \; \frac{\partial \mathbf{z}^*(\omega_0 L)}{\partial \omega_0} \, \right]$$
$$\frac{\partial \mathbf{z}(\omega_0 l)}{\partial \omega_0} = \left[\, 0 \; jle^{j\omega_0 l} \; \cdots \; j(N-1)le^{j\omega_0 l(N-1)} \, \right]^T. \quad (6)$$

and $\frac{\partial \mathbf{Z}}{\partial \omega_0} \mathbf{a} \triangleq \boldsymbol{\alpha}_0$, $\mathbf{Z} \frac{\partial \mathbf{a}}{\partial \phi_l} = -A_l \, \text{Im}\left\{ e^{j\phi_l} \mathbf{z}(\omega_0 l) \right\} \triangleq \boldsymbol{\gamma}_l$, and $\mathbf{Z} \frac{\partial \mathbf{a}}{\partial A_l} = \text{Re}\left\{ e^{j\phi_l} \mathbf{z}(\omega_0 l) \right\} \triangleq \boldsymbol{\beta}_l$. Here, $\text{Re}\left\{ \cdot \right\}$ and

$\text{Im}\left\{ \cdot \right\}$ denote the real and imaginary values, respectively. The Fisher information matrix can now be expressed as

$$\mathbf{I}(\boldsymbol{\theta}) = \frac{1}{\sigma^2} \begin{bmatrix} \boldsymbol{\alpha}_0^T \boldsymbol{\alpha}_0 & \boldsymbol{\alpha}_0^T \boldsymbol{\beta}_1 & \boldsymbol{\alpha}_0^T \boldsymbol{\gamma}_1 & \cdots \\ \boldsymbol{\beta}_1^T \boldsymbol{\alpha}_0 & \boldsymbol{\beta}_1^T \boldsymbol{\beta}_1 & \boldsymbol{\beta}_1^T \boldsymbol{\gamma}_1 & \cdots \\ \boldsymbol{\gamma}_1^T \boldsymbol{\alpha}_0 & \boldsymbol{\gamma}_1^T \boldsymbol{\beta}_1 & \boldsymbol{\gamma}_1^T \boldsymbol{\gamma}_1 & \cdots \\ \vdots & \vdots & \vdots & \ddots \\ \boldsymbol{\beta}_L^T \boldsymbol{\alpha}_0 & \boldsymbol{\beta}_L^T \boldsymbol{\beta}_1 & \boldsymbol{\beta}_L^T \boldsymbol{\gamma}_1 & \cdots \\ \boldsymbol{\gamma}_L^T \boldsymbol{\alpha}_0 & \boldsymbol{\gamma}_L^T \boldsymbol{\beta}_1 & \boldsymbol{\gamma}_L^T \boldsymbol{\gamma}_1 & \cdots \end{bmatrix}. \quad (7)$$

The CRLB can now be determined numerically by computing the inverse of this matrix and inspecting its diagonal elements. The simple closed form expressions for CRLBs obtained in [6, 9] can be found using the asymptotic orthogonality of complex sinusoids in computing the inner products above. However, we here do not employ this approximation, and we therefore refer to this CRLB as the exact CRLB. For reference, the asymptotic CRLB for the problem at hand is given by $\text{var}(\hat{\omega}_0) \geq 24\sigma^2 / \left(N^3 \sum_{l=1}^{L} A_l^2 l^2 \right)$. The lower bound can be seen to be determined by the so-called pseudo signal-to-noise ratio (PSNR) defined (in dB) as $SNR = 20 \log_{10} \sum_{l=1}^{L} A_l^2 l^2 / \sigma^2 \, [\text{dB}]$.

3. AN EXACT ESTIMATOR

We will now move on to deriving an estimator for solving the problem of interest without making use of the commonly used asymptotic approximations. The method is the NLS method, which is based on the principle of maximum likelihood estimation. The maximum likelihood estimator for the parameters $\boldsymbol{\theta}$ is given by

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \ln p(\mathbf{x}; \boldsymbol{\theta}). \quad (8)$$

Under the assumption that $\mathbf{x}$ is Gaussian distributed and the noise is white, i.e., $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}(\boldsymbol{\theta}), \sigma^2 \mathbf{I})$, the likelihood function is given by (2). By inserting (2) into (8), we obtain:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \frac{N}{2} \ln\left(2\pi\sigma^2\right) + \frac{1}{2\sigma^2} \left\| \mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\theta}) \right\|^2, \quad (9)$$

where $\|\cdot\|^2$ denotes the vector 2-norm. Dropping all constant terms and multipliers, we are left with

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left\| \mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\theta}) \right\|^2. \quad (10)$$

Using the definitions in Section 2, this results in the following estimator:

$$(\hat{\omega}_0, \hat{\mathbf{a}}) = \arg\min_{\omega_0, \mathbf{a}} \left\| \mathbf{x} - \mathbf{Za} \right\|^2. \quad (11)$$

Substituting the amplitudes by their maximum likelihood estimate, we obtain the following estimator, which depends only on ω_0:

$$\hat{\omega}_0 = \arg\min_{\omega_0} \left\| \mathbf{x} - \mathbf{Z} \left(\mathbf{Z}^H \mathbf{Z} \right)^{-1} \mathbf{Z}^H \mathbf{x} \right\|^2, \quad (12)$$

which can be written more compactly using the orthogonal projection matrix for the space spanned by the columns of $\mathbf{Z}$ given by $\mathbf{\Pi} = \mathbf{Z} \left(\mathbf{Z}^H \mathbf{Z} \right)^{-1} \mathbf{Z}^H$ as

$$\hat{\omega}_0 = \arg \min_{\omega_0} \| \mathbf{x} - \mathbf{\Pi}\mathbf{x} \|^2 = \arg \max_{\omega_0} \mathbf{x}^T \mathbf{\Pi} \mathbf{x}. \qquad (13)$$

This is the estimator that we will here refer to as the NLS estimator. For each fundamental frequency candidate it involves operations of complexity $\mathcal{O}(L^2 N + L^3 + L N^2 + N^2)$. The harmonic summation method [1] follows from this by using that the columns of $\mathbf{Z}$ are orthogonal asymptotically in N [9]. Although this leads to a fast implementation based on the fast Fourier transform, this ultimately also leads to the failure of this method for low ω_0 and N. We will refer to this method as approximate NLS (ANLS).

4. SOME RESULTS

We will now proceed to report some simulation results. First, we will investigate the dependency of the performance on the number of samples, N, and the fundamental frequency, ω_0. We will compare the performance of the exact NLS method to a number of state-of-the-art methods from [4, 9], namely the ANLS method, a method based on optimal filtering (OPTFILT), and a subspace method (MUSIC), and the weighted least-squares (WLS) method of [3], all of which are based on asymptotic approximations. To assess the performance, we use the mean squared estimation error (MSE) and use Monte Carlo simulations in which signals are generated using (1) with the following details: white Gaussian noise is added at a PSNR of 40 dB, Rayleigh distributed amplitudes along with uniformly distributed phases with five harmonics are used, and 100 trials are generated for each data point. The results are depicted in Figures 1(a)-1(b) (along with the exact CRLB) as functions of N for $\omega_0 = 0.3129$ (a) and as functions of the fundamental frequency with $N = 100$ (b). It can clearly be seen from the figures that the methods based on asymptotic approximations fail when the number of samples and/or the fundamental frequency is low. It can also clearly be seen that the NLS method performs much better under these circumstances. Regarding the CRLB, an interesting observation can be made from Figure 1(b): the asymptotic CRLB does not depend on ω_0, but the exact CRLB can clearly be seen from the Figure to vary with ω_0 when it approaches zero. This is another clear indication that the asymptotic approximation used to derived the asymptotic CRLB is incorrect in this case, and it should hence not be used as a benchmark for low fundamental frequencies.

Next, we will illustrate the problems associated with low fundamental frequencies using a recorded signal, namely 100 ms of a tone played by a contra bassoon sampled at 44.1 kHz with noise added at a signal-to-noise ratio of 40 dB. The signal is shown in Figure 2(a). In studying the effect of the low

fundamental frequency on the ability to obtain accurate estimates, the segment length is varied from 5 ms to 100 ms in steps of 5 ms, and the various estimators are then run on the resulting segments. We note that the number of harmonics was determined by visual inspection of the spectrum. The results are shown in Figure 2(b) for the same estimators as before. A number of interesting observations can be made from the figures. Firstly, all estimators converge to the same result when the segment length is increased. It can also be seen that all the methods break down eventually when the segment lengths get extremely short, but the increased robustness of the exact NLS towards this is also evident.

5. CONCLUSION

In this paper, we have analyzed the problem of estimating low fundamental frequencies. We have derived an exact expression for the Cramér-Rao lower bound and an exact estimator for finding the fundamental frequency. These avoid the asymptotic approximations that are commonly used in fundamental frequency estimators. Simulations clearly demonstrate the importance of this, as the exact estimator outperforms state-of-the-art methods for low fundamental frequencies and a low number of observations, and the exact bound can be observed to depend on the fundamental frequency. Future work includes avoiding such assumptions in other estimators, as these too may benefit from the principles introduced here.

6. REFERENCES

[1] M. Noll, "Pitch determination of human speech by harmonic product spectrum, the harmonic sum, and a maximum likelihood estimate," in *Proc. Symposium on Computer Proc. Com.*, 1969, pp. 779–797.

[2] B. G. Quinn and E. J. Hannan, *The Estimation and Tracking of Frequency*, ser. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2001.

[3] H. Li, P. Stoica, and J. Li, "Computationally efficient parameter estimation for harmonic sinusoidal signals," *Signal Processing*, vol. 80, pp. 1937–1944, 2000.

[4] M. G. Christensen, P. Stoica, A. Jakobsson and S. H. Jensen, "Multi-pitch estimation," *Signal Processing*, vol. 88(4), pp. 972–983, Apr. 2008.

[5] K. W. Chan and H. C. So, "Accurate frequency estimation for real harmonic sinusoids," *IEEE Signal Process. Lett.*, vol. 11(7), pp. 609–612, July 2004.

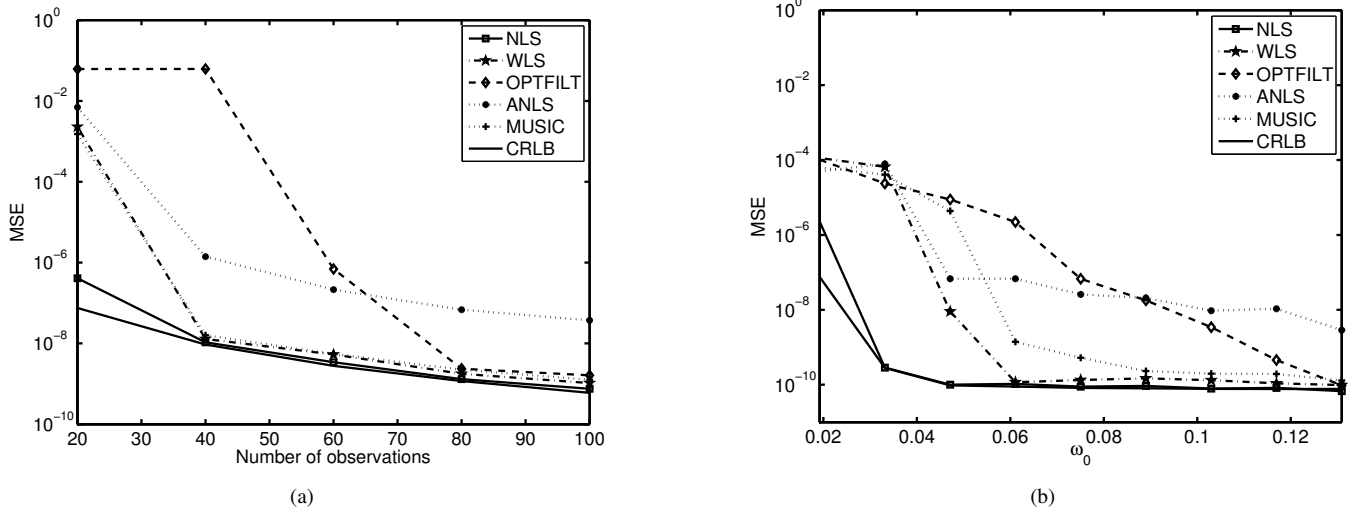[6] A. Nehorai and B. Porat, "Adaptive comb filtering for harmonic signal enhancement," *IEEE Trans. Acoust.*,

Figure 1: Performance measured in terms of the mean square estimation error (MSE) as a function of (a) the number of samples, N, and (b) the fundamental frequency, ω_0.
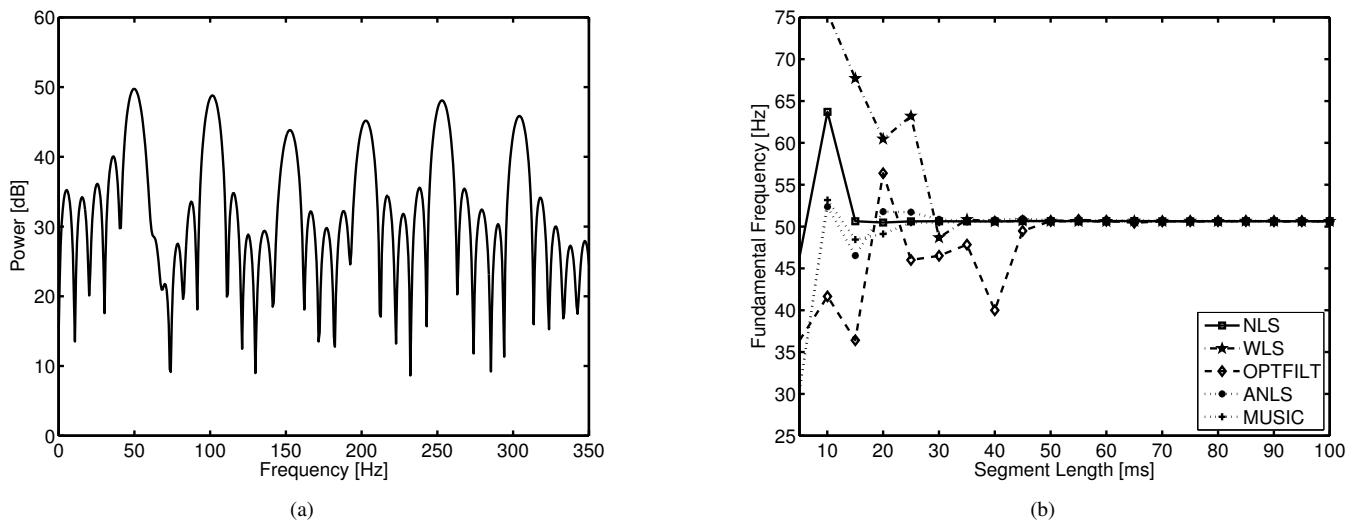

Figure 2: Example of a signal having a low frequency, here a tone played by a contra bassoon. Shown are (a) its spectrum for low frequencies computed from 100 ms, and (b) the obtained estimates as functions of the segment length.

Speech, Signal Process., vol. 34(5), pp. 1124–1138, Oct. 1986.

[7] A. de Cheveigné and H. Kawahara, "YIN, a fundamental frequency estimator for speech and music," *J. Acoust. Soc. Am.*, vol. 111(4), pp. 1917–1930, Apr. 2002.

[8] D. Talkin, "A robust algorithm for pitch tracking (RAPT)," in *Speech Coding and Synthesis*, W. B. Kleijn and K. K. Paliwal, Eds. Elsevier Science B.V., 1995, ch. 5, pp. 495–518.

[9] M. G. Christensen and A. Jakobsson, *Multi-Pitch Estimation*, ser. Synthesis Lectures on Speech & Audio Processing. Morgan & Claypool Publishers, 2009, vol. 5.

[10] P. Stoica, H. Li, and J. Li, "Amplitude estimation of sinusoidal signals: Survey, new results and an application," *IEEE Trans. Signal Process.*, vol. 48(2), pp. 338–352, Feb. 2000.