ESTIMATION OF AUTOREGRESSIVE PARAMETERS BY THE CONSTRAINED TOTAL LEAST SQUARE ALGORITHM USING A BOOTSTRAP METHOD

Ning Zhou *John W. Pierre*

Electrical and Computer Engineering Department
University of Wyoming, Laramie, WY 82071, USA
E-mail: {zn, pierre}@uwyo.edu

ABSTRACT

Modified Yule-Walker (MYW) equations are often used to estimate autoregressive parameters of an Auto-Regressive Moving Average (ARMA) model. Commonly used algorithms, i.e. the Least Square (LS) algorithm, the Total Least Square (TLS) algorithm, cannot give an optimal estimate because they do not exploit the Toeplitz property and covariance of the perturbation matrix. In this paper, a Constrained Total Least Squares (CTLS) algorithm is applied to solve modified Yule-Walker equations. The perturbation covariance matrix of the autocorrelation functions required by the CTLS algorithm is estimated by introducing the Bootstrap method. By utilizing Toeplitz property and the covariance of perturbation matrix, a Newton method based CTLS algorithm is shown to outperform TLS and LS solutions.

1. INTRODUCTION

Modified Yule-Walker (MYW) equations are often used to estimate the autoregressive parameters of an AutoRegressive Moving Average (ARMA) model [1]. Because the Least Square (LS) and Total Least Square (TLS) algorithms do not need prior knowledge about the perturbation covariance and are easy to implement, they are commonly used to solve MYW equations. Yet, failing to address the Toeplitz structure and covariance properties of the perturbation matrix of the MYW equations impairs the estimation accuracy of LS and TLS algorithms. In contrast, a Constrained Total Least Square (CTLS) algorithm [2] can help address these properties and provide the potential to improve the estimation accuracy. Yet, during its implementation, it requires the estimation of perturbation covariance, which is very difficult to obtain using conventional methods. Thus, according to authors' knowledge, the CTLS algorithm has not been used to solve MYW equations before. Recently, with the rapid growth of computation speed, some computation intensive methods, e.g. the Bootstrap method, are introduced to help assess statistical properties of parameter estimations. This provides a possible solution to disturbance covariance estimation. In this study, we apply the Bootstrap method to estimate the perturbation covariance and then employ CTLS algorithm to solve MYW equations using the estimated perturbation covariance.

The organization of this paper is listed as follows: In section 2, the difficulties with LS, TLS solution for MYW equations are discussed. In section 3, we will introduce the formulation of CTLS algorithm. In section 4, the Bootstrap method will be used to estimate the covariance matrix. In section 5, calculation procedure of CTLS algorithm based on estimated covariance matrix is described. In section 6, simulation is performed to show the estimation improvement by using CTLS algorithm. In section 7, some conclusions are discussed.

This work was supported by the U.S. Department of Energy and EPSCoR under grant DE-FG02-03ER46044.

2. MYW EQUATIONS AND LS/TLS ALGORITHMS

An (n^{th}, m^{th}) order ARMA model can be defined as

$$y(t)+\sum_{i=1}^{n}a_i y(t-i)=\sum_{j=0}^{m}b_j e(t-j) \tag{1}$$

where e(t) is a zero mean white noise sequence with variance of σ_e^2, a_i's for i=1, 2, … , n are the autoregressive (AR) parameters to be estimated. b_0=1. b_j's for j=1, 2, … , m are the moving average (MA) parameters to be estimated.

According to [1], we have

$$r_y(k)+\sum_{i=1}^{n}a_i r_y(k-i)=0 \quad for \quad k>m\cdot \tag{2}$$

This leads to the well-known MYW equations for AR parameters

$$\begin{bmatrix} r_y(m) & r_y(m-1) & \cdots & r_y(m-n+1) \\ r_y(m+1) & r_y(m) & \cdots & r_y(m-n+2) \\ \vdots & \vdots & \ddots & \vdots \\ r_y(m+M-1) & r_y(m+M-2) & \cdots & r_y(m+M-n) \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = -\begin{bmatrix} r_y(m+1) \\ r_y(m+2) \\ \vdots \\ r_y(m+M) \end{bmatrix}$$

or $R\cdot\vec{a}=\vec{r}$. (3)

Here, M is the number of equations. If $r_y(k)$'s were accurately known, (3) would give the exact solution for AR parameters with $M\geq n$. However, what is usually available is the data sample sequence of $\{y(1), y(2), \cdots, y(N)\}$, The autocorrelation function of y(t) is estimated through

$$\hat{r}_y(k)=\frac{1}{N}\sum_{t=k+1}^{N}y(t)\cdot y^*(t-k) \ for\ 0\leq k\leq N-1\ , \tag{4}$$

and it contains some estimation errors. As such, we can only obtain an estimation of the AR parameters through

$$\hat{R}\cdot\vec{a}\approx\vec{\hat{r}}\ . \tag{5}$$

In order to get a good estimation of AR parameters, we often set M>n to obtain a set of over-determined equations.
There are several algorithms to obtain the estimation solution for (5). The first algorithm is called LS method [1], which gives the solution for (5) as

$$\vec{a}_{LS} = \arg\min_{\vec{a}} \left\| \hat{R}\cdot\vec{a} - \vec{\hat{r}} \right\|^2, \quad (6)$$

or equivalently, obtain a minimal perturbation $\Delta\vec{\hat{r}}$ which makes the equation (5) consistent,

$$\min \left\|\Delta\vec{\hat{r}}\right\|^2 \text{ subject to } \hat{R}\cdot\vec{a}_{LS} = \vec{\hat{r}} + \Delta\vec{\hat{r}}. \quad (7)$$

This algorithm is optimal in the most likelihood sense when errors are only with $\vec{\hat{r}}$ and are independently normally distributed. In our application, there are estimation errors in both $\hat{R}$ and $\vec{\hat{r}}$. The LS solution is not optimal since it does not address the estimation errors in $\hat{R}$.
The second algorithm is called TLS algorithm [1], which gives the solution for (5) as

$$\min \left\| [\Delta\vec{\hat{r}} \quad \Delta\hat{R}] \right\|^2 \text{ subject to}$$

$$(\hat{R} + \Delta\hat{R})\cdot\vec{a}_{TLS} = \vec{\hat{r}} + \Delta\vec{\hat{r}} \quad (8)$$

The TLS algorithm includes the disturbance in both $\hat{R}$ and $\vec{\hat{r}}$ to address the estimation error in them.
Yet, in our MYW equations, the disturbance not only stays in both $\hat{R}$ and $\vec{\hat{r}}$ but also has a Toeplitz structure. To make it clear, notice that following equation holds.

$$[-\vec{r} \quad R] = [-\vec{\hat{r}} \quad \hat{R}] + [-\Delta\vec{r} \quad \Delta R], \quad (9)$$

in which the disturbance matrix has a Toeplitz structure. The TLS algorithm does not provide an optimal solution for MYW equations because it does not address the Toeplitz structure of this disturbance matrix.

3. THE CONSTRAINED TOTAL LEAST SQUARE ALGORITHM

The CTLS algorithm [2] has been successfully applied to damped sinusoids parameters estimation [3] and random amplitude sinusoid detection [4]. With the ability to address the Toeplitz structure and covariance of disturbance matrix, CTLS may provide an optimal solution to MYW equations. It is a generalization of TLS technique and has a maximum-likelihood interpretation [2]. To be consistent with notation in [2], we write (5) in the following form.

$$C\cdot\begin{bmatrix}\vec{\hat{a}}\\-1\end{bmatrix} \approx 0 \text{ where } C = [\hat{R} \quad \vec{\hat{r}}] \in R^{M\times(n+1)}.$$

We need to find a disturbance matrix $\Delta C = [\Delta R \quad \Delta\vec{r}] \in R^{M\times(n+1)}$ such that the whitened disturbance elements within ΔC are minimized and $[C+\Delta C]\cdot\begin{bmatrix}\vec{a}\\-1\end{bmatrix} = 0$. According to [2], in order to solve the (5) using CTLS algorithm, the disturbance matrix ΔC needs to be written in the form of its disturbance elements

$$\{\Delta r_y(m-n+1), \Delta r_y(m-n+2), \cdots, \Delta r_y(m+M)\}.$$

Also we have $\hat{r}_y(k) = \hat{r}_y(-k)$ for a real ARMA process. Thus the problem can be formulated in two situations,

Situation #1: $m-n+1 \geq 0$

The minimal disturbance elements are

$$\{\Delta r_y(m-n+1), \Delta r_y(m-n+2), \cdots, \Delta r_y(m+M)\}.$$

The number of minimal disturbance elements is

$$K \overset{\Delta}{=} M + n.$$

Define a disturbance vector as

$$\vec{v} \overset{\Delta}{=} [\Delta r_y(m-n+1) \quad \Delta r_y(m-n+2) \quad \cdots \quad \Delta r_y(m+M)]^T.$$

Let ΔC be written in terms of its columns as

$$\Delta C = [\Delta C_1 \quad \Delta C_2 \quad \cdots \quad \Delta C_{n+1}].$$

We have

$$\Delta C_i = F_i\cdot\vec{v} \quad \textit{for} \quad i = 1,2,\cdots,n+1$$

where $F_i \in R^{M\times K}$ can be deduced from the structure of ΔC_i as

$$F_i = [\mathrm{O}_{M\times(n-i)} \quad I_{M\times M} \quad \mathrm{O}_{M\times i}] \quad \textit{for} \quad i = 1,2,\cdots,n$$

$$F_{n+1} = [\mathrm{O}_{M\times n} \quad -I_{M\times M}],$$

where $\mathrm{O}_{r\times l}$ stands for zero matrix of dimension $r\times l$. $I_{r\times r}$ stands for identity matrix of r dimension.

Situation #2: $m-n+1 < 0$

The minimal disturbance elements are

$$\{\Delta r_y(0), \Delta r_y(1), \cdots, \Delta r_y(m+M)\}.$$

The number of minimal disturbance elements is

$$K \overset{\Delta}{=} M + m + 1.$$

Define a disturbance vector as

$$\vec{v} \overset{\Delta}{=} [\Delta r_y(0) \quad \Delta r_y(1) \quad \cdots \quad \Delta r_y(m+M)]^T.$$

Similarly, we have

$$\Delta C_i = F_i\cdot\vec{v} \quad \textit{for} \quad i = 1,2,\cdots,n+1$$

where $F_i \in R^{M\times K}$ can be deduced from the structure of ΔC_i as

$$F_i = [\mathrm{O}_{M\times(m-i+1)} \quad I_{M\times M} \quad \mathrm{O}_{M\times i}] \quad \textit{for} \quad i = 1,2,\cdots,m+1$$

$$F_i = \begin{bmatrix} \mathrm{O}_{(i-m-1)\times 1} & J_{(i-m-1)\times(i-m-1)} & \mathrm{O}_{(i-m-1)\times(M+2m-i+1)} \\ I_{(M-i+m+1)\times(M-i+m+1)} & & \mathrm{O}_{(M-i+m+1)\times i} \end{bmatrix} \quad \textit{for} \quad i = m+2,\cdots,n$$

$$F_{n+1} = [\mathrm{O}_{M\times(m+1)} \quad -I_{M\times M}],$$

where $J_{r\times r}$ stands for backward identity matrix of r dimension or

$$J_{r\times r} = \begin{bmatrix} 0 & & 1 \\ & \iddots & \\ 1 & & 0 \end{bmatrix}_{r\times r}.$$

For both situations, define the covariance of $\vec{v} \in R^{K\times 1}$ as $R_{\vec{v}} \overset{\Delta}{=} E\{\vec{v}\cdot\vec{v}^T\}$. Perform the Cholesky factorization on the perturbation covariance matrix $R_{\vec{v}} = PP^T$. Then we may define $\vec{u}$ as

$$\vec{u} \overset{\Delta}{=} P^{-1}\cdot v \in R^{K\times 1},$$

so that $\vec{u}$ is a white vector. And

$$\Delta C_i = F_i\cdot\vec{v} = F_i P\cdot\vec{u} \overset{\Delta}{=} G_i\cdot\vec{u} \quad \textit{for} \quad i = 1,2,\cdots,n+1.$$

As such, CTLS can be formulated as

$$\min_{\vec{u},\vec{a}} \|\vec{u}\|^2 \quad \text{subject to}$$

$$\left(C + \begin{bmatrix} G_1\vec{u} & G_2\vec{u} & \cdots & G_{n+1}\vec{u} \end{bmatrix}\right) \cdot \begin{bmatrix} \vec{a} \\ -1 \end{bmatrix} = 0 .$$

According to [2], we need to estimate $R_{\vec{v}}$, the perturbation covariance matrix of $\vec{v} \in R^{K\times 1}$ to do the CTLS calculation. It's very difficult, if not impossible, to get the covariance matrix using traditional methods. In our application, we will apply the Bootstrap method to estimate this covariance matrix.

4. ESTIMATE PERTURBATION COVARIANCE MATRIX USING THE BOOTSTRAP METHOD

The central idea about the Bootstrap method is re-sampling, which is well described in [5][6]. In our application, we are going to follow the following procedures of the bootstrapping residual method to estimate the perturbation covariance matrix $R_{\vec{v}}$.

1) Calculate parameter estimation $\{\hat{a}_1, \hat{a}_2, \cdots, \hat{a}_n, \hat{b}_1, \hat{b}_2, \cdots, \hat{b}_m\}$ by the two-stage least square method described in [1] based on a block of observation $\{y(1), y(2), \cdots, y(N)\}$.

2) Estimate a sequence of residuals $e(t)'s$ based on $\{\hat{a}_1, \hat{a}_2, \cdots, \hat{a}_n, \hat{b}_1, \hat{b}_2, \cdots, \hat{b}_m\}$ and $\{y(1), y(2), \cdots, y(N)\}$.

3) Treat the sequence $e(t)'s$ as a group of samples for a random variable E and estimate the probability density function (pdf) of E, $\hat{f}_E(x)$, based on $e(t)'s$. Note that if the $\hat{f}_E(x)$ appears to be close to a normal distribution, we only need to estimate the variance of E to obtain the pdf.

4) Set i=1.

5) Based on this pdf estimation $\hat{f}_E(x)$, we may generate a sequence of residuals $\hat{e}_{b(i)}(t)'s$, which are usually called bootstrapping residuals.

6) Pass this newly generated residual sequence through the ARMA model defined by $\{\hat{a}_1, \hat{a}_2, \cdots, \hat{a}_n, \hat{b}_1, \hat{b}_2, \cdots, \hat{b}_m\}$. We get a set of bootstrapping outputs $\{\hat{y}_{b(i)}(1), \hat{y}_{b(i)}(2), \cdots, \hat{y}_{b(i)}(N)\}$.

7) Estimate $\hat{y}_{b(i)}(t)$'s autocorrelation functions $\hat{r}_{\hat{y}_b(i)}(k)$ and form the corresponding disturbance vector, i.e.,

$if \ m - n + 1 \geq 0$

$$\hat{R}_i = [\hat{r}_{\hat{y}_{b(i)}}(m-n+1) \quad \hat{r}_{\hat{y}_{b(i)}}(m-n+2) \quad \cdots \quad \hat{r}_{\hat{y}_{b(i)}}(m+M)\ ]^T$$

$if \ m - n + 1 < 0$

$$\hat{R}_i = [\hat{r}_{\hat{y}_{b(i)}}(0) \quad \hat{r}_{\hat{y}_{b(i)}}(1) \quad \cdots \quad \hat{r}_{\hat{y}_{b(i)}}(m+M)\ ]^T$$

8) Set i=i+1 and repeat 5,6,7 for L times.

9) Define $\hat{R}_{mean} \stackrel{\Delta}{=} \frac{1}{L}\sum_{i=1}^{L} \hat{R}_i$. Construct the matrix

$$\hat{V} = [\hat{R}_1 - \hat{R}_{mean} \quad \hat{R}_2 - \hat{R}_{mean} \quad \cdots \quad \hat{R}_L - \hat{R}_{mean}\ ]$$

10) Estimate the perturbation covariance as $\hat{R}_{\vec{v}} = \frac{1}{L}\hat{V}\cdot\hat{V}^T$

Note that the Bootstrap method is a computation intensive method. There is a trade-off between the estimation accuracy and L, number of calculation repeats.

5. CALCULATION PROCEDURE OF CTLS USING NEWTON'S METHOD

A general closed-form CTLS solution is almost impossible. Thus the authors of [2] developed a Newton's method to address this quadratic minimization problem. With the Bootstrap estimation of the covariance matrix of $\vec{v} \in R^{K\times 1}$ discussed in previous section, we may calculate CTLS solution using Newton Method as follows.

1) Use the TLS solution of MYW equations as the initial value $\vec{a}(1) = \vec{a}_{TLS}$. Set j=1.

2) Get P from the Cholesky factorization of the perturbation covariance matrix $\hat{R}_{\vec{v}} = PP^T \in R^{K\times K}$.

3) Calculate $G_i \stackrel{\Delta}{=} F_i \cdot P \in R^{M\times K} \quad with \quad i = 1,2,\cdots,n+1$.

4) Do the following Calculation

$$H = \sum_{i=1}^{n} \hat{a}_i(j)\cdot G_i - G_{n+1} \in R^{M\times K}$$

$$u = (H\cdot H^T)^{-1} C \begin{bmatrix} \vec{a}(j) \\ -1 \end{bmatrix} \in R^{M\times 1}$$

$$\widetilde{B} = C\cdot\begin{bmatrix} I_{n\times n} \\ O_{1\times n} \end{bmatrix} - [G_1 H^T u \quad G_2 H^T u \quad \cdots \quad G_n H^T u] \in R^{M\times n}$$

$$\widetilde{G} = [G_1{}^T u \quad G_2{}^T u \quad \cdots \quad G_n{}^T u] \in R^{K\times n}$$

$$a = \left(u^T \widetilde{B}\right)^T \in R^{n\times 1}$$

$$A = -\widetilde{G}^T H^T (H\cdot H^T)^{-1}\widetilde{B} - (\widetilde{G}^T H^T (H\cdot H^T)^{-1}\widetilde{B})^T \in R^{n\times n}$$

$$B = [\widetilde{B}^T (H\cdot H^T)^{-1}\widetilde{B}]^T + \widetilde{G}^T [H^T (H\cdot H^T)^{-1} H - I_{K\times K}]\widetilde{G} \in R^{n\times n}$$

5) Revise of the solution as

$$\vec{a}(j+1) = \vec{a}(j) + (AB^{-1}A - B)^{-1}(a - AB^{-1}a)$$

and set j=j+1.

6) Repeat steps 4-5 until $\vec{a}(j)$ converges.

Notice that the Newton algorithm cannot guarantee the convergence. If divergence happens, we'll retain the $\vec{a}_{TLS}$ as the solution. Readers are referred to [2] for some details.

6. SIMULATION RESULTS

Two ARMA models, which were used on page 135 of [1], were employed in this paper to compare the performance of LS, TLS, CTLS algorithms in solving MYW equations. The broadband ARMA model takes form as

$$y(t) - 1.3817y(t-1) + 1.5632y(t-2) - 0.8843y(t-3) + 0.4096y(t-4)$$
$$= e(t) + 0.3544e(t-1) + 0.3508e(t-2) + 0.1736e(t-3) + 0.2401e(t-4)$$

which has poles at $0.8\cdot\exp(\pm i0.45\pi)$, $0.8\cdot\exp(\pm i0.25\pi)$, zeros at $0.7\cdot\exp(\pm i0.35\pi)$, $0.7\cdot\exp(\pm i0.75\pi)$ with n=m=4.

Another model is a narrowband ARMA model.

$$y(t) - 1.6408y(t-1) + 2.2044y(t-2) - 1.4808y(t-3) + 0.8145y(t-4)$$
$$= e(t) + 1.5857e(t-1) + 0.9604e(t-2)$$

which has poles at $0.95\cdot\exp(\pm i0.25\pi)$, $0.95\cdot\exp(\pm i0.45\pi)$ and zeros at $0.98\cdot\exp(\pm i0.8\pi)$ with n=4, m=2.

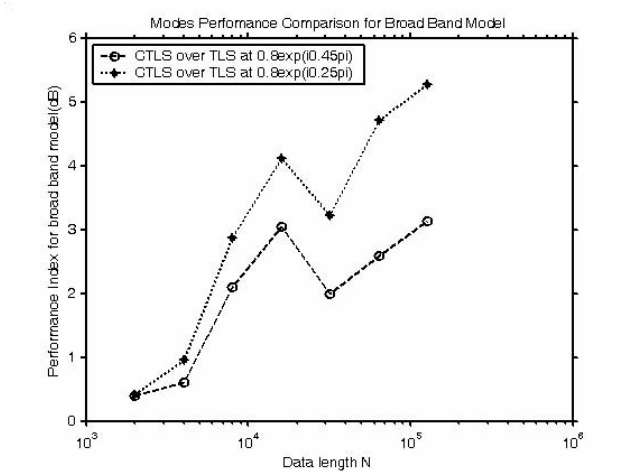

Fig 1. Performance comparison between the CTLS and TLS algorithms for the simulation data from the broad-band model.

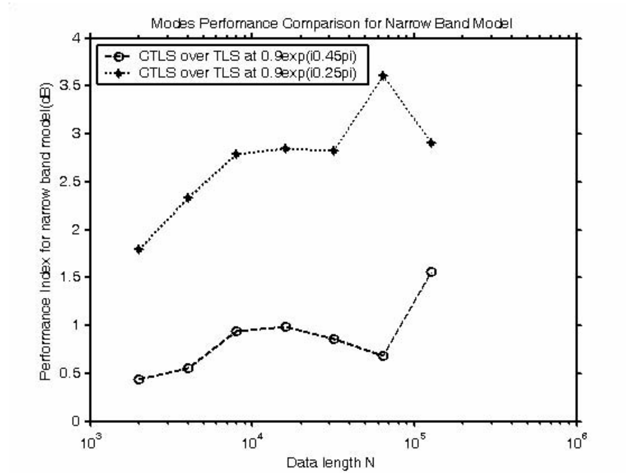

Fig 2. Performance comparison between the CTLS and TLS algorithms for the simulation data from the narrow-band model.

One hundred sets of simulation data were generated. The length of each data set was N=8000. First 1500 data were discarded to make ARMA model response enter into a steady state. MYW equations were constructed using the autocorrelation estimation. LS, TLS, CTLS algorithms were used to solve the MYW equations for AR parameters with M=2n=8. For CTLS algorithm, the Bootstrap with L=500 was used to estimate the perturbation covariance matrix. The Newton's method is used to get the recursive solution for the CTLS algorithm. The poles of models were calculated from estimated AR parameters to compare the performance among different algorithms.
Table 1 shows the mean square errors (MSE) of poles estimation results. It can be observed from the table that LS and TLS algorithms give the estimation with similar performance. The CTLS algorithm outperforms LS and TLS algorithms in the sense of MSE when it is used to do poles estimation for ARMA model.

TABLE 1. MEAN SQUARE ERROR (MSE) OF POLES ESTIMATION USING DIFFERENT ALGORITHMS

Poles	MSE of LS(10^{-4})	MSE of TLS(10^{-4})	MSE of CTLS(10^{-4})
$0.8\exp(\pm i0.45\pi)$	15.7	14.8	6.4
$0.8\exp(\pm i0.25\pi)$	7.7	7.7	4.5
$0.95\exp(\pm i0.45\pi)$	0.231	0.231	0.128
$0.95\exp(\pm i0.25\pi)$	0.169	0.169	0.155

In order to evaluate the influence of data length (N) on the performance improvement of the CTLS algorithm, we apply CTLS on the data sets of various length generated from both ARMA simulation models. In addition, we introduce the performance improvement index as

$$PI = 10 \cdot \log_{10}\left(MSE_{LS\ or\ TLS} \big/ MSE_{CTLS} \right)$$

As the performance of LS and TLS algorithms are similar, to make the plots concise, we only show the performance improvement of the CTLS over TLS in Fig 1 and 2 for both models. It can be observed from the figures that the performance improvement of CTLS over TLS is consistent and has an increase trend with the increase of data length.

7. CONCLUSIONS

In this study, a CTLS algorithm is applied to estimate the autoregressive parameters of an ARMA model. By utilizing the Bootstrap method, the perturbation covariance of the autocorrelation functions is estimated. The performance of CTLS method, which is based on a Newton method and this perturbation covariance estimation, is shown to outperform TLS and LS solution for both narrowband and broadband simulation models. Future works may include the investigation of the influence of relative pole position on the estimation results. Also, the application can be extended to address ARMA problem with some added measurement noises.

6. REFERENCES

[1] P. Stoica, and R. Moses, *Introduction to Spectral Analysis*, Prentice Hall, 1997.

[2] T. J. Abatzoglou, and J. M. Mendel, "The Constrained Total Least Squares Technique and Its Applications to Harmonic Supperresolution," *IEEE Transactions on Signal Processing*, vol. 39, no. 5, pp. 1070-1086, May, 1991.

[3] T. J. Abatzoglou, "Optimal Estimation of Parameters of Exponentially Damped Sinusoids by the Constrained Total Least Squares Method," *Conference Record - Asilomar Conference on Circuits, Systems & Computers*, vol. 1, pp. 447-451, 1991.

[4] O. Besson, "Improved Detection of a Random Amplitude Sinusoid by Constrained Least-square Technique," *Signal Processing*, pp. 347-356, vol. 45, no. 3, 1995.

[5] A.M. Zoubir and B. Boashash, "The Bootstrap and its Application in Signal Processing," *IEEE Signal Processing Magazine*, pp. 56-76, Jan. 1998.

[6] D. N. Politis, "Computer-Intensive Methods in Statistical Analysis," *IEEE Signal Processing Magazine*, pp. 39-55, Jan 1998.