Research Article

Detection of Wideband Signal Number Based on Bootstrap Resampling

Jiaqi Zhen

College of Electronic Engineering, Heilongjiang University, Harbin 150080, China

Correspondence should be addressed to Jiaqi Zhen; zhenjiaqi2011@163.com

Received 7 December 2015; Revised 22 February 2016; Accepted 22 February 2016

Academic Editor: Herve Aubert

Copyright © 2016 Jiaqi Zhen. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Knowing source number correctly is the precondition for most spatial spectrum estimation methods; however, many snapshots are needed when we determine number of wideband signals. Therefore, a new method based on Bootstrap resampling is proposed in this paper. First, signals are divided into some nonoverlapping subbands; apply coherent signal methods (CSM) to focus them on the single frequency. Then, fuse the eigenvalues with the corresponding eigenvectors of the focused covariance matrix. Subsequently, use Bootstrap to construct the new resampling matrix. Finally, the number of wideband signals can be calculated with obtained vector sequences according to clustering technique. The method has a high probability of success under low signal to noise ratio (SNR) and small number of snapshots.

1. Introduction

Detecting wideband source number is one of the hot and difficult problems in the area of array signal processing; meanwhile, it is widely used in many fields, such as radio monitoring [1–5], acoustic signal processing [6–10], and Internet of Things [11–14]; the corresponding methods are also being studied by a lot of scholars all the time. Generally speaking, knowing number of signals is also the premier condition of estimating direction of arrival (DOA). The initial method is based on hypothesis test [15]; as the detecting threshold needs to be set artificially, it is easy to be influenced by individual subjective factor. In order to avoid this problem, Wax and Kailath introduce the information theoretic criteria [16] into the question; then minimum description length (MDL) [17] and Akaike information criterion (AIC) [18] were put forward, improving the detecting performance to some extent. Liu et al. [19] estimated signal number and directions simultaneously by spatial difference method based on uniform linear array. Han and Nehorai [20] increased the array aperture with folding technology; thus, more signals can be processed. Liu et al. [21] set the detection threshold by proposing the concept of effective source number and distinguished the signals which were close to one another. Goldstein et al. [22] proposed an effective filtering technology based on reducing dimension in the late 1990s; it does not need to compute the covariance matrix and has a preferable responsiveness to time-varying signals. However, there are few published literatures for wideband signals; one of them is based on coherent signal method (CSM) [23] proposed by Wang and Kaveh, which focuses the signals on the reference frequency, and then employs MDL or AIC to obtain the final result. All the methods above need many snapshots, or their performance will deteriorate seriously. Unfortunately, they are difficult to be implemented in actual system: firstly, sampling time is strictly confined; for example, the impulse of radar is often narrow, and the data can be limited, especially for wideband signal; secondly, the signal can not be stable for a long time. So it is difficult to acquire more effective samples, which restricts the application of the methods above.

Bootstrap is another kind of technique arising in recent years which can be used with only a few samples [24–30]; some scholars highlighted the motivation for using the Bootstrap in signal processing [31, 32]. For signal number estimation, it needs to calculate the number of the minimum equal eigenvalues of the sampling covariance matrix based on the multiple hypothesis test, and it has a higher precision comparing with information theoretic criteria under the same circumstance. In 2000, Brcich used Bootstrap for estimating source number for the first time [27]; after a few
years, he improved the testing threshold by the Bonferroni multiple hypothesis test, enhancing the performance further [28, 29]. Huang et al. [30] used the eigenvalues of one frequency to form the testing threshold in combination with Bootstrap resampling to estimate wideband source number; as they only exploited the information of only one frequency, the estimation accuracy is limited.

In this paper, a new method for detecting wideband source number is proposed. The approach is based on Bootstrap technique and makes full use of the information of every frequency. First, the signals are divided into some nonoverlapping subbands, following focusing them on the same frequency. Then, the eigenvalues are fused with corresponding eigenvectors of the focused covariance matrix. Subsequently, Bootstrap is employed to construct the new resampling matrix. Finally, the number of wideband signals can be calculated with obtained vector sequences according to clustering technique.

2. Array Signal Model

Consider a uniform linear array (ULA) composed by $M$ omnidirectional sensors with spacing $d$; the first sensor is defined as the phase reference and the array signal model is shown in Figure 1.

Assume that $K$ far-field wideband signals $s_k(t)$ ($k = 1, 2, \ldots, K$) impinge on the array, their directions are $\theta = [\theta_1, \theta_K]$, observed time is $T$, and the received data by the array at time $t$ are expressed as

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_m(t) \\ \vdots \\ x_M(t) \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^{K} s_k(t) \\ \vdots \\ \sum_{k=1}^{K} s_k(t) \end{bmatrix}$$

for $t \in (0, T)$, where $x_m(t)$ is the output of the $m$th sensor, $c$ is the propagation speed of the signals, and $n_m(t)$ is the additive white noise subjected to Gaussian distribution $\text{CN}(0, \sigma^2)$.

Suppose that the frequency range of all signals is $[f_L, f_H]$, $KP$ data are collected at every frequency, and a $G$-point discrete Fourier transform (DFT) is applied to (1); the array output of the $g$th frequency can be expressed as

$$\mathbf{X}(f_g) = \mathbf{A}(f_g, \mathbf{\theta}) \mathbf{S}(f_g) + \mathbf{N}(f_g)$$

$$= \begin{bmatrix} x_1(f_g, 1) & \cdots & x_1(f_g, kp) & \cdots & x_1(f_g, KP) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_m(f_g, 1) & \cdots & x_m(f_g, kp) & \cdots & x_m(f_g, KP) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_M(f_g, 1) & \cdots & x_M(f_g, kp) & \cdots & x_M(f_g, KP) \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{X}(f_g, 1) & \cdots & \mathbf{X}(f_g, kp) & \cdots & \mathbf{X}(f_g, KP) \end{bmatrix}$$

$$g = 1, 2, \ldots, G.$$}

Here $f_L \leq f_g \leq f_H$ for $g = 1, 2, \ldots, G$, $\mathbf{X}(f_g, kp) = [x_1(f_g, kp), \cdots, x_M(f_g, kp)]^\top$, and $x_m(f_g, kp)$ is the sample data on the $m$th sensor of $f_g$ at snapshot $kp$, and $\mathbf{A}(f_g, \mathbf{\theta}) = [\mathbf{a}(f_g, \theta_1), \ldots, \mathbf{a}(f_g, \theta_K)]^\top$ is the array steering matrix with $M \times K$ dimension, where

$$\mathbf{a}(f_g, \theta_k) = \begin{bmatrix} 1, \ldots, \exp \left( -jm2\pi f_g \frac{d}{c} \sin \theta_k \right), \ldots, \\
\exp \left( -j(M - 1)2\pi f_g \frac{d}{c} \sin \theta_k \right) \end{bmatrix}^\top.$$  

The signals $\mathbf{S}(f_g)$ ($g = 1, 2, \ldots, G$), within different frequency bins, are assumed to be Gaussian distributed, and they are uncorrelated with the noise $\mathbf{N}(f_g)$ ($g = 1, 2, \ldots, G$).

3. Proposed Method

3.1. Focusing. The covariance matrix at $f_g$ can be solved by

$$\mathbf{\hat{R}}(f_g) = \frac{1}{KP} \mathbf{X}(f_g) \mathbf{X}^H(f_g) \quad g = 1, 2, \ldots, G.$$  

Then, we can acquire the covariance matrix $\mathbf{\hat{R}}(f_0)$ with the single frequency by transforming the signals at each frequency into a covariance matrix at a focusing frequency $f_0$ based on two-sided correlation transformation (TCT)
method [33], and so forth we can choose the center frequency as \( f_0 \). Apply eigendecomposition to \( \mathbf{R}(f_0) \); we have

\[
\mathbf{R}(f_0) = \sum_{k=1}^{K} \lambda_k(f_0) \mathbf{e}_k(f_0) \mathbf{e}_k^H(f_0)
\]

\[
+ \sum_{k=K+1}^{M} \lambda_k(f_0) \mathbf{e}_k(f_0) \mathbf{e}_k^H(f_0).
\]

(5)

Then, the eigenvalues \( \lambda_1(f_0) > \cdots > \lambda_K(f_0) > \lambda_{K+1}(f_0) = \cdots = \lambda_M(f_0) \) can be obtained, and the corresponding eigenvector is \( \mathbf{E}(f_0) = [\mathbf{e}_1(f_0), \ldots, \mathbf{e}_K(f_0), \mathbf{e}_{K+1}(f_0), \ldots, \mathbf{e}_M(f_0)] \), where \( \mathbf{E}_k(f_0) = [\mathbf{e}_1(f_0), \ldots, \mathbf{e}_k(f_0)] \) corresponds to signal subspace and \( \mathbf{E}_k(f_0) = [\mathbf{e}_{k+1}(f_0), \ldots, \mathbf{e}_M(f_0)] \) corresponds to the noise subspace.

3.2. Fusion. Since both \( \mathbf{a}(f_0, \theta) \) and \( \mathbf{E}_k(f_0) \) belong to the signal subspace, we have

\[
\mathbf{a}(f_0, \theta) = \sum_{k=1}^{K} \alpha_k(f_0) \mathbf{e}_k(f_0),
\]

where \( \alpha_k(f_0) \) is the corresponding coefficient, and the inner product of the eigenvector is defined by

\[
\mu_m(f_0) = \left| \mathbf{a}^H(f_0, \theta) \mathbf{e}_m(f_0) \right|,
\]

(7)

The signal and noise subspace are orthogonal to each other, so

\[
\mathbf{a}^H(f_0, \theta) \mathbf{e}_k(f_0) = 0, \quad k = K + 1, \ldots, M.
\]

(8)

According to (6) and (8), (7) can be represented as

\[
\mu_m(f_0) = \left| \sum_{k=1}^{K} \alpha_k(f_0) \mathbf{e}_k(f_0) \right|^2
\]

\[
= \left| \alpha_m \right|, \quad m = 1, 2, \ldots, K
\]

\[
= 0, \quad m = K + 1, \ldots, M.
\]

(9)

Traditional information theory criterion only exploits the information of eigenvalues, but when the snapshots are small, we should make full use of the sampling data; since the eigenvectors also carry information of the signal, let us fuse \( \mu_m(f_0) \) with \( \zeta_m(f_0) \):

\[
\xi_m(f_0) = \sqrt{\mu_m(f_0)}, \quad m = 1, 2, \ldots, M - 1.
\]

(10)

Here, \( \xi_m(f_0) \) is defined as the fusion inner product, and \( \xi(f_0) = [\xi_1(f_0), \ldots, \xi_m(f_0), \ldots, \xi_{M-1}(f_0)] \) is the fusion inner product vector. According to [34], the optimal \( \zeta_m(f_0) (m = 1, 2, \ldots, M - 1) \) are chosen as

\[
\zeta_m(f_0) = \frac{\left( \lambda_m(f_0) - \lambda_M(f_0) \right)^2}{\lambda_m(f_0)},
\]

(11)

\[
m = 1, 2, \ldots, M - 1.
\]

3.3. Bootstrap Resampling. We can employ Bootstrap technique to determine number of signals; on the one hand, there is no need to know the probability density distribution of the testing statistics; on the other hand, it is still effective under the circumstance of small number of snapshots. Suppose that \( \tilde{\mathbf{X}}(f_g) = [\tilde{\mathbf{X}}(f_0, 1) \cdots \tilde{\mathbf{X}}(f_0, k_0) \cdots \tilde{\mathbf{X}}(f_0, KP)] \) is the Bootstrap resampling matrix of \( \mathbf{X}(f_g) \), where \( \tilde{\mathbf{X}}(f_g, k_0) \mid k_0 = 1, 2, \ldots, KP \) is a random column selected from \( \mathbf{X}(f_g) \) and each selected column is uncorrelated to one another, execute the same operation to the signals of the other frequencies, and we can acquire \( \tilde{\mathbf{X}}(f_1), \ldots, \tilde{\mathbf{X}}(f_g), \ldots, \tilde{\mathbf{X}}(f_c) \), the corresponding focused covariance matrix \( \tilde{\mathbf{R}}(f_0) \), and the fusion inner product vector \( \tilde{\xi}(f_0) = [\tilde{\xi}_1(f_0), \ldots, \tilde{\xi}_m(f_0), \ldots, \tilde{\xi}_{M-1}(f_0)] \). Similarly, repeat the Bootstrap resampling, focusing, and fusion for \( T \) times; we will calculate \( T \) groups of focused covariance matrices \( \tilde{\mathbf{R}}(f_0, 1), \ldots, \tilde{\mathbf{R}}(f_0, t), \ldots, \tilde{\mathbf{R}}(f_0, T) \) and the corresponding inner product vector sequences \( \tilde{\xi}(f_0, t) = [\tilde{\xi}_1(f_0, t), \ldots, \tilde{\xi}_m(f_0, t), \ldots, \tilde{\xi}_{M-1}(f_0, t)] \mid t = 1, 2, \ldots, T \); average them to obtain the new vector \( \tilde{\xi} = (1/T) \sum_{t=1}^{T} \tilde{\xi}(f_0, t) = [\tilde{\xi}_1, \ldots, \tilde{\xi}_m, \ldots, \tilde{\xi}_{M-1}] \), where the new element \( \tilde{\xi}_m = (1/T) \sum_{t=1}^{T} \tilde{\xi}_m(f_0, t) (m = 1, 2, \ldots, M - 1) \).

3.4. Clustering. Signals and noise can be distinguished by clustering the elements in the vector \( \tilde{\xi} \); the larger ones belong to the signals, as there are large differences among them. While the others belong to the noise, they are smaller and close to one another. Here, clustering algorithm [35] is employed to estimate the source number; it is summarized as follows:

1. Select the initial class mean of the signal and noise:

\[
\tilde{\xi}_1^{(1)} = \tilde{\xi}_1,
\]

\[
\tilde{\xi}_2^{(1)} = \tilde{\xi}_{M-1}.
\]

(12)

2. In the \( rth \) iteration, every element is endowed with the signal or the noise class according to the criterion below, to the \( \tilde{\xi}_m \) \((m = 1, 2, \ldots, M - 1)\):

\[
\tilde{\xi}_m \in Q_2^{(r)} \quad \text{if} \quad Z_{1m}^{(r-1)} < Z_{m2}^{(r-1)},
\]

(13)

where

\[
Z_{1m}^{(r)} = \frac{\tilde{\xi}_m^{(r-1)}}{\tilde{\xi}_1^{(r-1)}},
\]

\[
Z_{m2}^{(r)} = \frac{\tilde{\xi}_m^{(r-1)}}{\tilde{\xi}_2^{(r-1)}},
\]

(14)

\( Q_1 \) denotes the noise class and \( Q_2 \) denotes the signal class; that is to say, if the ratio of a certain element and noise class mean is larger than that of the signal class mean and the certain element, then the element belongs to the signal; otherwise, it belongs to the noise; this means that we assign every element to the nearest class.
(3) Calculate the new value of each clustering center:

$$\tilde{\xi}_\alpha^{(r+1)} = \frac{1}{D_\alpha} \sum_{\beta \in \mathcal{Q}^r} \tilde{\xi}_\beta \quad \alpha = 1, 2,$$

(15)

where $D_\alpha$ is the number of the element in $\mathcal{Q}^r$.

(4) If $\tilde{\xi}_\alpha^{(r+1)} = \tilde{\xi}_\alpha^{(r)} \ (\alpha = 1, 2)$, there will be no new elements turning over each class, the algorithm is convergent, and the iteration is over; otherwise, turn to step (2) and go on to next iteration; thus, the number of signals $K$ can be decided.

It is seen from the deduction above that we can estimate $M - 2$ signals at most. Since the proposed method employs fusion, Bootstrap resampling, and clustering, it can be called FBC method for short.

Or we can omit the process of fusion, then only use Bootstrap resampling and focusing to construct the new covariance sequences $\mathbf{R}(f_0, 1), \ldots, \mathbf{R}(f_0, t), \ldots, \mathbf{R}(f_0, T)$, and subsequently average them; we have

$$\mathbf{\bar{R}} = \frac{1}{T} \sum_{t=1}^{T} \mathbf{R}(f_0, t).$$

(16)

Then, apply clustering above to the eigenvalues of $\mathbf{\bar{R}}$ to decide the signal number, but this method does not utilize the eigenvectors; similarly, we can call it BC for short.

4. Simulations

In order to verify the effectiveness of the method, some simulations are presented with Matlab below; consider some binary phase coded wideband signals impinging on a uniform linear array with 8 omnidirectional sensors, the center frequency is 1 GHz, the width of the band is 1 GHz × 20% = 0.2 GHz, signals are sampled with 20 frequency bins in the frequency domain, the spacing between adjacent sensors $d = 3 \times 10^2/(2 \times 1 \times 10^9)$ m, FBC, BC, and MDL based on CSM (CSM-MDL) are, respectively, employed for the simulations, 200 Monte-Carlo trials have run for each variable, and, in FBC and BC, $\alpha = 0.02$.

In the first simulation, five far-field wideband uncorrelated signals with the same power impinge on the array in $5^\circ, 12^\circ, 20^\circ, 26^\circ, \text{and} 33^\circ$; the noise is assumed to be Gaussian-distributed; it is obtained through an autoregressive (AR) filter [36] of coefficient 0.9; the resample times $T = 90$; Figure 2 presents the probability of detection as a function of SNR when the number of snapshots is 30, while Figure 3 shows that of the number of snapshots when SNR is 3 dB.

It is seen from Figure 2 that all the probabilities of success of these methods are improving with the increasing of SNR; in comparison, FBC is better than the other two methods; it is 100% when SNR reaches $-6$ dB and that of the BC and CSM-MDL can reach 100% when SNR is 0 dB and 3 dB, respectively. And it is seen from Figure 3 that when the snapshots are small, FBC is better than BC and CSM-MDL, and all the probabilities of success of these methods are improving with increasing of the snapshots; in comparison, the probabilities of success of FBC can reach 100% when the snapshots are 22 and that of the BC and CSM-MDL can reach 100% when the snapshots are 28 and 32, respectively.

In the second simulation, five far-field wideband uncorrelated signals with the same power impinge on the array in $5^\circ, 12^\circ, 20^\circ, 26^\circ$, and $33^\circ$; the noise is assumed to be non-Gaussian-distributed; it is obtained through an autoregressive (AR) filter [36] of coefficient 0.9; the resample times $T = 90$; Figure 4 presents the probability of detection as a function of SNR when the number of snapshots is 30, while Figure 5 shows that of the number of snapshots when SNR is 3 dB.

It is seen from Figure 4 that when the SNR is low, all the probabilities of success of these methods are improving with...
Table 1: Average calculating time of the four methods.

<table>
<thead>
<tr>
<th>Method</th>
<th>Computational complexity</th>
<th>Average calculating time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSM-MDL</td>
<td>$O((M^3 \times KP \times G + M^2 \times KP \times G + M^3 + M^4)/2)$</td>
<td>6.8</td>
</tr>
<tr>
<td>BC</td>
<td>$O(T \times (M^3 \times KP \times G + M^2 \times KP \times G + M^3 + M^4))$</td>
<td>310</td>
</tr>
<tr>
<td>FBC</td>
<td>$O(T \times (M^3 \times KP \times G + M^2 \times KP \times G + M^3) + T \times (M^3 \times G + M \times G) + M^2)$</td>
<td>312</td>
</tr>
</tbody>
</table>

Figure 4: Probability of success versus SNR for uncorrelated signals in non-Gaussian white noise.

Figure 5: Probability of success versus number of snapshots for uncorrelated signals in non-Gaussian white noise.

its increasing; when SNR reaches a certain extent, CSM-MDL is fluctuating and unstable, but all of them can reach 100% at last. And it is seen from Figure 5 that when the snapshots are small, all the probabilities of success of these methods are improving with the snapshots increasing. In comparison, FBC is better than BC and CSM-MDL; when snapshots reach a certain extent, CSM-MDL is fluctuating and unstable, while FBC and BC are still available relatively, but they still cannot reach 100%, so their estimations are not consistent for the non-Gaussian white noise.

In the third simulation, five far-field wideband coherent signals with the same power impinge on the array in $5^\circ$, $12^\circ$, $20^\circ$, $26^\circ$, and $33^\circ$; the noise is assumed to be Gaussian-distributed; the resample times $T=90$; Figure 6 presents the probability of detection as a function of SNR when the number of snapshots is 30, while Figure 7 shows that of the number of snapshots when SNR is 3 dB.

It is seen from Figure 6 that these methods can estimate the number of wideband coherent signals successfully when SNR is high enough, and all the probabilities of success of these methods are improving with the SNR increasing. In comparison, FBC is better than BC and CSM-MDL.

In the fourth simulation, five far-field wideband uncorrelated signals with the same power impinge on the array in $5^\circ$, $12^\circ$, $20^\circ$, $26^\circ$, and $33^\circ$; the noise is assumed to be Gaussian-distributed; Figure 8 presents the probability of detection as a function of resampling times when the number of snapshots is 20 and SNR is 3 dB, while we use digital signal processor ADSP-TS201S to implement these methods; the computational complexities and average calculating time are shown in Table 1.

It is seen from Figure 8 that the probabilities of success of the two methods are improving with the resampling times increasing. In comparison, as the FBC makes use of not only the eigenvalues but also the eigenvectors, it yields better detection performance than BC, so we should fully exploit the information of the eigenvectors and resampling under the circumstance of small number of snapshots.

It is not difficult to deduce that CSM-MDL needs the process of focusing and formula computation, so its computational complexity is $O(M^3 \times KP \times G + M^2 \times KP \times G + M^3 + M^4)/2)$; BC needs focusing, resampling, and clustering, and the complexity is $O(T \times (M^3 \times KP \times G + M^2 \times KP \times G + M^3) + M^2)$; FBC needs focusing, fusion, resampling, and clustering, and...
5. Conclusions

The paper proposes a method for determining the number of wideband signals based on Bootstrap resampling; after focusing the signal on the reference frequency, we make use of fusing the eigenvectors with the corresponding eigenvalues, subsequently we employ the Bootstrap to construct the resampling matrix, and then we use clustering to decide signal number. The simulations have shown that the performance of the proposed method is better than that of information theory criterion; it adapts to uncorrelated and coherent signals. Moreover, it also has an effect on the non-Gaussian noise.

Bootstrap resampling is used to estimate wideband signal number in this paper; however, due to the process of the resampling, its calculation is very complicated. Our work will be committed to optimizing the method to lower the computation complexity in future.

Competing Interests

The author declares that there are no competing interests.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (61501176) and the Education Department of Heilongjiang Province Science and Technology Research Projects (12541638).

References


thecomplexityisO(T×(M^3×KP×G+M^2×KP×G+M^2)+T×(M^2×G+M×G)+M^2).Anditcanbeseenfromthecalculating
timethattheresamplingspendsalongtime;itiswillincrease
the computation time by multiples, while focusing and fusion
are not obvious relatively; we can reduce the resampling times
properly and employ multiple digital signal processors to
increase efficiency in practical applications.

Figure 6: Probability of success versus SNR for coherent signals.

Figure 7: Probability of success versus number of snapshots for coherentsignals.


