Digital Signal Processing 20 (2010) 887–899

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Joint estimation of source number and DOA using simulated annealing algorithm

Junli Liang [a,*], Ding Liu [b]

[a] School of Automation & Information Engineering and the School of Computer Science & Engineering, Xi'an University of Technology, Xi'an, China
[b] School of Automation & Information Engineering, Xi'an University of Technology, Xi'an, China

ARTICLE INFO

Article history:
Available online 6 August 2009

Keywords:
Array signal processing
Simulated annealing (SA) algorithm
Directions-of-arrival (DOA)
Cumulant

ABSTRACT

The array signal processing problem has to do with two main issues: 1) the determination of source number; and 2) estimation of the directions-of-arrival (DOA) of these sources. In the classical methods, the determination of source number and the estimation of DOA parameters are executed independently rather than jointly. To jointly estimate source number and DOA, this paper proposes a simulated annealing-based algorithm. The key points of this paper are: i) the prior choices for some unknown parameters are calibrated to compute more efficiently; ii) a penalty term for the parameter number is added to the corresponding objective function Additionally to avoid overfitting; and iii) in order to move to the regions of interest quickly, some cumulants of properly chosen sensor outputs from the uniform linear array (ULA) is computed and their moduluses of Fourier transform (MFT) are mapped trickly to the uniform distribution. Finally simulation results are presented to validate the performance of the proposed method.

1. Introduction

Estimation of directions-of-arrival (DOA) has received a significant amount of attention over the last several decades [1–4]. It is a key problem in array signal processing fields such as radar, sonar, radio astronomy, and mobile communication systems. The maximum likelihood (ML) method, despite its theoretical optimality, has extremely demanding computational complexity due to the search computation and iteration process. To alleviate the computational overhead, several simpler algorithms have been developed, such as the propagator method [1], the ESPRIT method [2], and the MUSIC method [3]. However, these algorithms are implemented under the assumptions that the source number is known. To determine the source number, some information theoretic criteria are used [5], such as Akaike Information Criterion (AIC) and Bayes Information Criterion (BIC). However, these classical approaches usually involve maximizing the likelihood function that is associated with each possible source number and then calculating the corresponding criteria values. When large number of sources are possible, this quickly becomes infeasible unless a method that simultaneously determines the source number and corresponding DOAs is available.

Simulated annealing (SA) algorithm derives from the physical process of heating and then slowly cooling a crystalline substance [6–8]. If the structure is cooled sufficiently slowly, the molecules will line up in a rigid pattern corresponding to a state of minimum energy. The SA algorithm mimics this process by producing a sequence of draws from a series statistical distributions that move toward a point mass at the minimum of a chosen objective function as the 'temperature' is lowered. The classical SA algorithm is used to deal with the simple problem of parameter estimation under a fixed model. Recently,

* Corresponding author.
E-mail address: heery_2004@hotmail.com (J. Liang).

1051-2004/$ – see front matter
doi:10.1016/j.dsp.2009.08.007

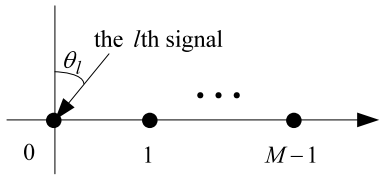


Fig. 1. Array configuration.

Brooks et al. developed a trans-dimensional simulated annealing algorithms to determine model orders [8–10], and Green et al. developed a reversible jump Markov Chain Monte Carlo-based algorithms for Bayesian model determination [11–16].

To jointly estimate source number and DOA of multiple sources, a simulated annealing (SA)-based algorithm is developed in this paper. *The key points of this paper are as follows*: i) combing modulus of Fourier transform (MFT) of some cumulants (virtual harmonic signal) [17–20] of properly chosen sensor outputs and uniform distribution, a novel and effective cumulant-based sampling method is developed to estimate DOA, which mainly relies on the effective aperture increase capacity of cumulant; ii) least squares method is used to decrease the number of unknown sampling parameter so as to alleviate computational complexity; and iii) the penalty term is added to the objective function for avoiding overfitting.

The rest of this paper is organized as follows. The problem is introduced in Section 2. A simulated annealing-based algorithm is developed in Section 3. Simulation results are presented in Section 4. Conclusions are drawn in Section 5.

2. Problem formulation

Suppose that L narrow-band, independent radiating sources impinge on the uniform linear array (ULA) of Msensors shown in Fig. 1. Let the 0th sensor be the phase reference point. After being sampled, the signal received by the ith sensor can be expressed as

$$x_i(k) = \sum_{l=1}^{L} s_l(k) \exp\left(\frac{-j2\pi id \sin\theta_l}{\lambda}\right) + n_i(k), \quad 0 \leqslant i \leqslant M-1,\ 0 \leqslant k \leqslant K-1, \tag{1}$$

where $s_l(k)$ is the lth incoming signal, K is the snapshot number, λ is the wavelength of these sources, and $n_i(k)$ is the additive white Gaussian noise (AWGN) of the ith sensor. In addition, d is the distance between the sensors, and θ_l is the direction of the lth signal.

In a matrix form, (1) can be written as

$$\mathbf{x}(k) = \mathbf{A}(\boldsymbol{\theta})\mathbf{s}(k) + \mathbf{n}(k), \quad k = 0, \ldots, K-1, \tag{2}$$

where

$$\mathbf{A}(\theta) = \left[\mathbf{a}(\theta_1)\ \ldots\ \mathbf{a}(\theta_l)\ \ldots\ \mathbf{a}(\theta_L)\right], \tag{3}$$

$$\mathbf{a}(\theta_l) = \left[1, e^{j\phi_l}, e^{j2\phi_l}, \ldots, e^{j(M-1)\phi_l}\right]^T, \tag{4}$$

$$\phi_l = \frac{-2\pi d \sin\theta_l}{\lambda}, \tag{5}$$

$$\mathbf{x}(k) = \left[x_0(k), x_1(k), \ldots, x_{M-1}(k)\right]^T, \tag{6}$$

$$\mathbf{n}(k) = \left[n_0(k), n_1(k), \ldots, n_{M-1}(k)\right]^T, \tag{7}$$

and

$$\mathbf{s}(k) = \left[s_1(k), s_2(k), \ldots, s_L(k)\right]^T. \tag{8}$$

The objective of this paper is to jointly estimate source number L and directions $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_L]$. Throughout the paper, the following hypotheses are assumed to hold:

1) the source signals are statistically mutually independent, non-Gaussian ones with nonzero kurtosis;
2) the sensor noise is the AWGN with zero-mean and independent of the source signals;
3) the directions of these sources are different from each other, that is $\theta_i \neq \theta_j$ for $i \neq j$;
4) for unique estimation, we require $L < M$.

3. Proposed algorithm

3.1. Simplified Bayesian model

The likelihood function comes directly from (2) and is given by

$$p\big(\mathbf{x}\big|\sigma_n^2,\boldsymbol{\theta},L,\mathbf{s}(k)\big)=\sum_{k=0}^{K-1}\frac{1}{\sqrt{2\pi\,|\sigma_n^2\mathbf{I}|}}\exp\!\left(-\frac{1}{2\sigma_n^2}\big\|\mathbf{x}(k)-\mathbf{A}(\boldsymbol{\theta})\mathbf{s}(k)\big\|^2\right) \tag{9}$$

where $\mathbf{x}=[\mathbf{x}(0),\mathbf{x}(1),\ldots,\mathbf{x}(K-1)]$, and $\mathbf{I}$ is an $M\times M$ identity matrix.

The Bayesian approach treats these unknown parameters $[\sigma_n^2,\boldsymbol{\theta},L,\mathbf{s}(k)]$ as random variables, and draws them from appropriate prior distributions. These priors reflect our degree of belief in the relevant values of these quantities. A hierarchical prior structure is used to treat the priors' parameters (hyper-parameters) as random variables drawn from suitable distributions (hyper-priors).

The prior distribution for these parameters can be represented as

$$p\big(\sigma_n^2,\boldsymbol{\theta},L,\mathbf{s}(k)\big)=p\big(\sigma_n^2\big)p(\boldsymbol{\theta}|L)p\big(\mathbf{s}(k)|L\big)p(L). \tag{10}$$

In the above-mentioned hierarchical prior structure, $\sigma_n^2,\boldsymbol{\theta},L$, and $\mathbf{s}(k)$, $k=0,\ldots,K-1$, are required to draw from appropriate prior distributions in every iteration.

Based on the maximum likelihood method, the conventional estimates of source signal $\mathbf{s}(k)$ can be given by

$$\hat{\mathbf{s}}(k)=\big[\mathbf{A}^H(\theta)\mathbf{A}(\theta)\big]^{-1}\mathbf{A}^H(\theta)\mathbf{x}(k). \tag{11}$$

Define

$$\mathbf{P}_{\mathbf{A}(\theta)}=\mathbf{A}(\theta)\big[\mathbf{A}^H(\theta)\mathbf{A}(\theta)\big]^{-1}\mathbf{A}^H(\theta), \tag{12}$$

and

$$\mathbf{P}_{\mathbf{A}(\theta)}^{\perp}=(\mathbf{I}-\mathbf{P}_{\mathbf{A}(\theta)})^H(\mathbf{I}-\mathbf{P}_{\mathbf{A}(\theta)})=\mathbf{I}-\mathbf{A}(\theta)\big[\mathbf{A}^H(\theta)\mathbf{A}(\theta)\big]^{-1}\mathbf{A}^H(\theta) \tag{13}$$

where the superscript $\perp$ stands for orthogonal complement.

Thus, the estimate of the variance for Gaussian distribution can be given by

$$\begin{aligned}\hat{\sigma}_n^2&=\frac{1}{MK}\sum_{k=0}^{K-1}\big\|\mathbf{x}(k)-\mathbf{A}(\theta)\big[\mathbf{A}^H(\theta)\mathbf{A}(\theta)\big]^{-1}\mathbf{A}^H(\theta)\mathbf{x}(k)\big\|^2\\&=\frac{1}{MK}\sum_{k=0}^{K-1}\big\|\mathbf{x}(k)-\mathbf{P}_{\mathbf{A}(\theta)}\mathbf{x}(k)\big\|^2\\&=\frac{1}{MK}\sum_{k=0}^{K-1}\big[(\mathbf{I}-\mathbf{P}_{\mathbf{A}(\theta)})\mathbf{x}(k)\big]^H\big[(\mathbf{I}-\mathbf{P}_{\mathbf{A}(\theta)})\mathbf{x}(k)\big]\\&=\frac{1}{MK}\sum_{k=0}^{K-1}\mathbf{x}^H(k)\mathbf{P}_{\mathbf{A}(\theta)}^{\perp}\mathbf{x}(k)\end{aligned} \tag{14}$$

or

$$\begin{aligned}\hat{\sigma}_n^2&=\frac{1}{MK}\sum_{k=0}^{K-1}\big\|\big(\mathbf{x}(k)-\mathbf{P}_{\mathbf{A}(\theta)}\mathbf{x}(k)\big)\big\|^2\\&=\frac{1}{MK}\sum_{k=0}^{K-1}\operatorname{trace}\big\{\big(\mathbf{x}(k)-\mathbf{P}_{\mathbf{A}(\theta)}\mathbf{x}(k)\big)\big(\mathbf{x}(k)-\mathbf{P}_{\mathbf{A}(\theta)}\mathbf{x}(k)\big)^H\big\}\\&=\frac{1}{MK}\sum_{k=0}^{K-1}\operatorname{trace}\big\{(\mathbf{I}-\mathbf{P}_{\mathbf{A}(\theta)})\mathbf{x}(k)\mathbf{x}^H(k)(\mathbf{I}-\mathbf{P}_{\mathbf{A}(\theta)})\big\}\\&=\frac{1}{MK}\operatorname{trace}\left\{(\mathbf{I}-\mathbf{P}_{\mathbf{A}(\theta)})\sum_{k=0}^{K-1}\big(\mathbf{x}(k)\mathbf{x}^H(k)\big)(\mathbf{I}-\mathbf{P}_{\mathbf{A}(\theta)})\right\}\end{aligned}$$

$$= \frac{1}{M}\operatorname{trace}\left\{(\mathbf{I}-\mathbf{P}_{\mathbf{A}(\theta)})\left(\frac{1}{K}\sum_{k=0}^{K-1}\mathbf{x}(k)\mathbf{x}^H(k)\right)(\mathbf{I}-\mathbf{P}_{\mathbf{A}(\theta)})\right\}$$
$$= \frac{1}{M}\operatorname{trace}\big(\mathbf{P}^{\perp}_{\mathbf{A}(\theta)}\hat{\mathbf{R}}\mathbf{P}^{\perp}_{\mathbf{A}(\theta)}\big). \tag{15}$$

From (14), it can be derived

$$\sum_{k=0}^{K-1}\big\|\big(\mathbf{x}(k)-\mathbf{P}_{\mathbf{A}(\theta)}\mathbf{x}(k)\big)\big\|^2 \cong MK\hat{\sigma}_n^2. \tag{16}$$

Using estimates $\hat{\mathbf{s}}(k)$ and $\hat{\sigma}_n^2$ in (11) and (15), $p(\mathbf{x}|\sigma_n^2,\boldsymbol{\theta},L,\mathbf{s}(k))$ can be approximated by

$$\begin{aligned} p\big(\mathbf{x}|\sigma_n^2,\boldsymbol{\theta},L,\mathbf{s}(k)\big) &= \sum_{k=0}^{K-1}\frac{1}{\sqrt{2\pi|\sigma_n^2\mathbf{I}|}}\exp\left(-\frac{1}{2\sigma_n^2}\big\|\mathbf{x}(k)-\mathbf{P}_{\mathbf{A}(\theta)}\mathbf{x}(k)\big\|^2\right)\\ &\cong \left(\frac{2\pi}{M}\operatorname{trace}\big(\mathbf{P}^{\perp}_{\mathbf{A}(\theta)}\hat{\mathbf{R}}\mathbf{P}^{\perp}_{\mathbf{A}(\theta)}\big)\right)^{-MK/2}\exp\left(-\frac{MK\hat{\sigma}_n^2}{2\sigma_n^2}\right)\\ &\cong \left(\frac{2\pi}{M}\right)^{-MK/2}\exp\left(-\frac{MK}{2}\right)\big(\operatorname{trace}\big(\mathbf{P}^{\perp}_{\mathbf{A}(\theta)}\hat{\mathbf{R}}\mathbf{P}^{\perp}_{\mathbf{A}(\theta)}\big)\big)^{-MK/2}\\ &\propto \big(\operatorname{trace}\big(\mathbf{P}^{\perp}_{\mathbf{A}(\theta)}\hat{\mathbf{R}}\mathbf{P}^{\perp}_{\mathbf{A}(\theta)}\big)\big)^{-MK/2}. \end{aligned} \tag{17}$$

From (17), it can be seen that the approximated likelihood function $p(\mathbf{x}|\sigma_n^2,\boldsymbol{\theta},L,\mathbf{s}(k))$ is merely related to $\boldsymbol{\theta}$ and L and can be rewritten as $\hat{p}(\mathbf{x}|\boldsymbol{\theta},L)$. Therefore, it isn't necessary to sample parameters σ_n^2 and $\mathbf{s}(k)$.

3.2. Penalized likelihood source number selection

Traditionally, penalized likelihood model order selection strategies, based on standard information criteria, require the evaluation of the maximum likelihood (ML) estimates for each possible model order. The number of required evaluations can be prohibitively expensive unless appropriate heuristics are available. Subsequently, a particular model is selected if it is the one that minimizes the sum of the log-likelihood and a penalty term that depends on the model dimension. In mathematical terms, this estimate is given by

$$MO(L_{\text{opt}}) = \arg\min\big\{-\log\big(\hat{p}(\mathbf{x}|\boldsymbol{\theta},L)\big)+\mathbf{P}\big\}, \tag{18}$$

where $\mathbf{P}$ is penalty term that depends on the model order. ML penalties include the well-known AIC and BIC information criteria. The expressions for these criteria are

$$\mathbf{P}_{\text{AIC}} = L, \quad \text{and} \quad \mathbf{P}_{\text{BIC}} = \frac{1}{2}L\log(K). \tag{19}$$

These criteria are motivated by different factors: AIC is based on expected information, and BIC is an asymptotic Bayes factor.

Based on (17) and (18), we have

$$\begin{aligned} MO(L_{\text{opt}}) &= \arg\min\big\{-\log\big(\hat{p}(\mathbf{x}|\boldsymbol{\theta},L)\big)+\mathbf{P}\big\}\\ &= \arg\min\big\{-\log\big\{\big[\operatorname{trace}\big(\mathbf{P}^{\perp}_{A(\theta)}\hat{\mathbf{R}}\mathbf{P}^{\perp}_{A(\theta)}\big)\big]^{-MK/2}\big\}+\mathbf{P}\big\}\\ &= \arg\min\left\{\frac{MK}{2}\log\big[\operatorname{trace}\big(\mathbf{P}^{\perp}_{A(\theta)}\hat{\mathbf{R}}\mathbf{P}^{\perp}_{A(\theta)}\big)\big]+\mathbf{P}\right\}\\ &= \arg\max\left\{-\frac{MK}{2}\log\big[\operatorname{trace}\big(\mathbf{P}^{\perp}_{A(\theta)}\hat{\mathbf{R}}\mathbf{P}^{\perp}_{A(\theta)}\big)\big]-\mathbf{P}\right\}\\ &= \arg\max\big\{\big[\operatorname{trace}\big(\mathbf{P}^{\perp}_{A(\theta)}\hat{\mathbf{R}}\mathbf{P}^{\perp}_{A(\theta)}\big)\big]^{-MK/2}e^{-\mathbf{P}}\big\}. \end{aligned} \tag{20}$$

3.3. The equivalent relationship between maximum penalized likelihood and maximum posterior distribution

Note that ϕ_l in Eq. (5) is the function of θ_l. It is obvious that θ_l can be easily estimated from $\hat{\phi}_l$, i.e. $\hat{\theta}_l = \operatorname{asin}(-\hat{\phi}_l\lambda/2\pi d)$. Thus, we transform the above-mentioned problem into an equivalent one, i.e. joint estimating L and $\hat{\phi}_l$, then estimating $\hat{\theta}_l$ from $\hat{\phi}_l$.

Based on $\hat{p}(\mathbf{x}|\boldsymbol{\theta},L)$, via Bayes' theorem, the posterior distribution $\hat{p}(\boldsymbol{\theta},L|\mathbf{x})$ can be expressed as [8–10]

$$\hat{p}(\boldsymbol{\theta}, L|\mathbf{x}) \propto \hat{p}(\boldsymbol{\theta}, L|\mathbf{x}) p(\boldsymbol{\theta}|L) p(L) \tag{21}$$

or

$$\hat{p}(\boldsymbol{\Phi}, L|\mathbf{x}) \propto \hat{p}(\mathbf{x}|\boldsymbol{\Phi}, L) p(\boldsymbol{\Phi}|L) p(L), \tag{22}$$

or

$$MO(L_{\text{opt}}) = \arg\max\left\{\left[\operatorname{trace}\left(\mathbf{P}_{A(\Phi)}^{\perp} \hat{\mathbf{R}} \mathbf{P}_{A(\Phi)}^{\perp}\right)\right]^{-MK/2} e^{-\mathbf{P}}\right\} \tag{23}$$

where $\boldsymbol{\Phi} = [\phi_1, \phi_2, \ldots, \phi_L]$.

Since $\theta_l \in [-\pi/2, \pi/2]$ and $d \leqslant \lambda/2$, $\phi_l \in [-\pi, \pi]$. Note that the source signal can come from arbitrary direction, thus ϕ_l is uniformly distributed on $[-\pi, \pi]$, i.e. $p(\boldsymbol{\Phi}|L) = (\frac{1}{2\pi})^L$.

Comparing Eqs. (20) and (22), we can find that these expressions agree whenever

$$p(L) \propto \exp(-\mathbf{P}) \propto \exp(-CL). \tag{24}$$

This equation shows that the expression for the calibrated posterior $\hat{p}(\boldsymbol{\theta}, L|\mathbf{x})$ corresponds to the term that needs to be maximized in the penalized likelihood framework. Note that for the purposes of optimization, we only need the proportionality condition with $C = 1$ for the AIC criterion and $C = \frac{1}{2}\log(K)$ for the BIC criterion [5]. It has thus been shown that by calibrating the priors in the Bayesian formulation, one can obtain the expression that needs to be maximized in the classical penalized likelihood formulation with AIC and BIC model selection criteria. Consequently, the penalized likelihood framework can be interpreted as a problem of maximizing the joint posterior distribution $\hat{p}(\boldsymbol{\theta}, L|\mathbf{x})$.

3.4. Design sampling functions for the unknown parameters $[\phi_1, \phi_2, \ldots, \phi_L]$

To estimate unknown parameters $[\phi_1, \phi_2, \ldots, \phi_L]$, we need to dram samples from the conditional joint distribution. Since sampling from $p(\boldsymbol{\Phi}|L)$ would be difficult due to it nonlinearity, we instead cycle through and update these 'frequencies' one at a time. Therefore, at every iteration i, we sample $\phi_j^{(i)}$, in turn from $p(\phi_j^{(i)}|\boldsymbol{\Phi}_{-j,L}^{(i)}, L^{(i)})$, where $\boldsymbol{\Phi}_{-j,L}^{(i)} = (\phi_1^{(i)}, \phi_2^{(i)}, \ldots, \phi_{j-1}^{(i)}, \phi_{j+1}^{(i-1)}, \ldots, \phi_L^{(i-1)})$.

We have chosen here to sample ϕ_j, $j = 1, 2, \ldots, L$, one-at-a-time using a mixture of Metropolis-Hastings (MH) steps. At each individual update of ϕ_j, $j = 1, 2, \ldots, L$, we choose randomly between two possible types of proposal distribution. The first one aims to move to the regions of interest quickly, and the second proposal is a random walk one. For each iteration, we choose between the two types with the fixed probabilities λ_0 and $1 - \lambda_0$ respectively to sample each ϕ_j, $j = 1, 2, \ldots, L$.

To develop the first proposal distribution, we begin with the fourth-order cumulants of some properly chosen sensor outputs [17–20], which can be expressed as

$$\begin{aligned}\operatorname{cum}\left\{x_m(k), x_n^*(k), x_p^*(k), x_q(k)\right\} &= \sum_{l=1}^{L} e^{j[m\phi_l - n\phi_l - p\phi_l + q\phi_l]} \operatorname{cum}\left\{s_l(k), s_l^*(k), s_l^*(k), s_l(k)\right\} \\ &= \sum_{l=1}^{L} c_{4,sl} e^{j[(m-n)-(p-q)]\phi_l}, \quad m, n, p, q \in [0, M-1],\end{aligned} \tag{25}$$

where $c_{4,sl} = \operatorname{cum}\{s_l(k), s_l^*(k), s_l^*(k), s_l(k)\}$is the kurtosis of the lth signal, and the superscript * denotes the complex conjugate. Because $m, n, p, q \in [0, M-1]$, both $(m-n)$ and $(p-q)$ lie in $[-M+1, M-1]$. Eq. (25) shows that the fourth-order cumulant of the output from real physical sensors is similar to a cross-covariance

$$\begin{aligned}E\left\{x_{m-n}(k) x_{p-q}^*(k)\right\} &= \sum_{l=1}^{L} E\left(\left\{s_l(k) e^{j[(m-n)\phi_l]}\right\}\left\{s_l(k) e^{j[(p-q)\phi_l]}\right\}^*\right) \\ &= \sum_{l=1}^{L} r_{sl} e^{j[(m-n)-(p-q)]\phi_l} = \sum_{l=1}^{L} r_{sl}\left(e^{j(m-n)\phi_l}\right)\left(e^{j(p-q)\phi_l}\right)^*, \quad m-n, p-q \in [-M+1, M-1],\end{aligned} \tag{26}$$

which were from the *noiseless* outputs of two real or virtual sensors $m-n$ and $p-q$. In Eq. (26), $r_{sl} = E\{s_l(t) s_l^*(t)\}$ is the power of the lth signal. Eq. (26) can be explained by virtue of the virtual array shown in Fig. 2 [formed from the array of Fig. 1, the shaded and clear circles denote real and virtual sensors].

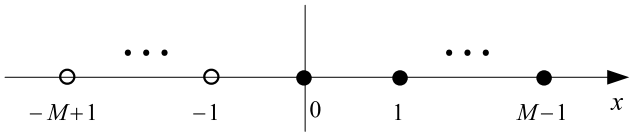


Fig. 2. Virtual array (with noiseless output) formed from the real array of Fig. 1. The shaded and clear circles represent real and virtual sensors, respectively.

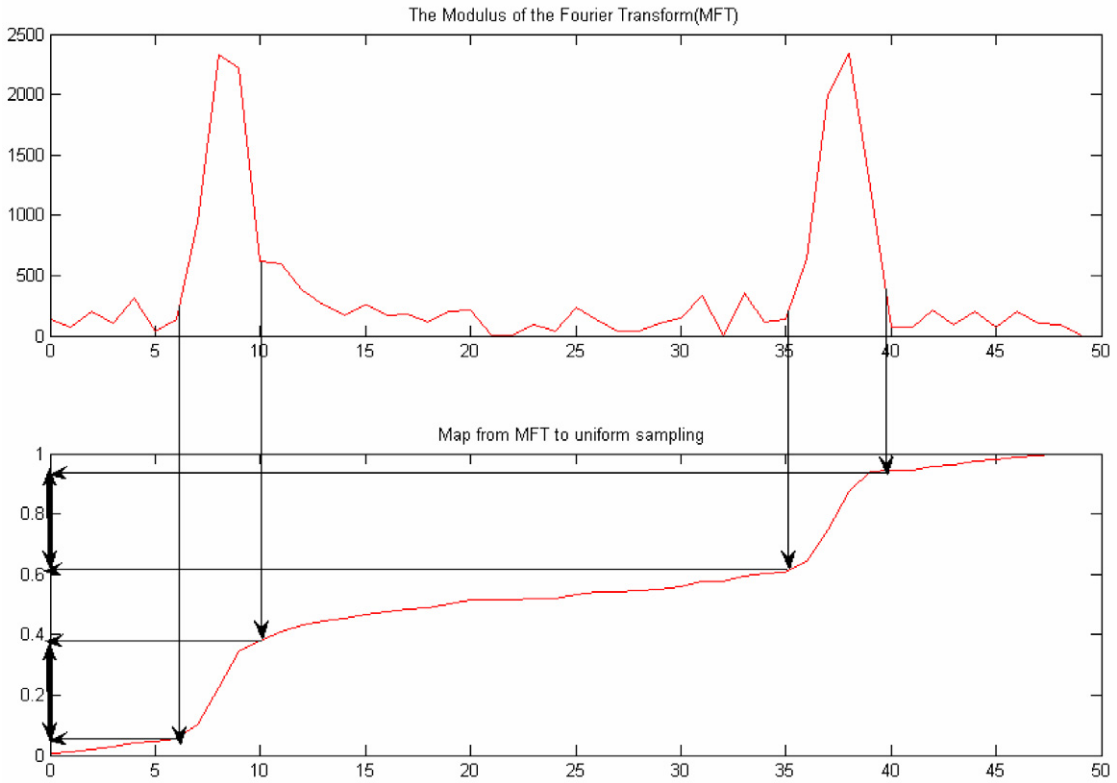


Fig. 3. Mapping from the MFT to the uniform distribution.

Since different (m,n) pairs may cause the equal $m-n$ (the same reason for p and q), we define the unique mapping relationship in Eq. (27).

$$\begin{cases} m=0;\ n=M-1,\ldots,2,1 & \Leftrightarrow \quad m-n\in[-M+1,-1], \\ m=0,1,\ldots,M-1;\ n=0 & \Leftrightarrow \quad m-n\in[0,M-1], \\ p=0;\ q=M-1,\ldots,2,1 & \Leftrightarrow \quad p-q\in[-M+1,-1], \\ p=0,1,\ldots,M-1;\ q=0 & \Leftrightarrow \quad p-q\in[0,M-1]. \end{cases} \tag{27}$$

Since $m-n,\ p-q\in[-M+1,M-1]$, $(m-n)-(p-q)\in[-2M+2,2M-2]$, i.e. $(m-n)-(p-q)$ has possible $4M-3$ values. Therefore, based on (25), we can obtain a sequence of $4M-3$ data, i.e. $y(i+2M-2)=\sum_{l=1}^{L} c_{4,sl}e^{ji\phi_l}$, $i=-2M+2,\ldots,2M-2$. Define that $z(i),\ i=0,\ldots,4M-3$, is the Fourier transform of $y(i)$, $i=0,\ldots,4M-3$. It is obvious that there are high values at the region of 'frequency' ϕ_j, $j=1,2,\ldots,L$, in the modulus $|z(i+2M-2)|$ of Fourier transform (MFT). Define $w(i)=\sum_{j=0}^{i}|z(i)|$, then normalize $w(i)$, i.e.

$$w(i)=\frac{w(i)}{w(4M-3)},\quad i=0,\ldots,4M-3.$$

It is obvious that $0\leqslant w(0)\leqslant w(1)\leqslant\cdots\leqslant w(4M-3)=1$. Sample u from uniform distribution $U[0,1]$. If $w(i)\leqslant u\leqslant w(i+1)$, frequency sample $i\pi/(4M-3)$ will be obtained indirectly from uniform distribution. The basic idea can be described in Fig. 3. The sub-figure above shows the MFT of two signals. From it, we can see that there are two peaks in the corresponding 'frequencies' ϕ_1 and ϕ_2. We cumulate these moduluses and normalize the maximum value and thus establish the mapping relationship between the signals' MFT and uniform distribution. It is obvious that when we sample

using uniform distribution, the sampling value will be prompted to lie in the approximate region of true value, which can be seen from the sub-figure down of Fig. 3.

The second proposal distribution is a perturbation about the current value. We propose an update from a Gaussian distribution with mean being equal to the current value and standard deviation σ_ϕ^2 fixed prior to running the algorithm, i.e., $\phi_j^{(i)} \sim N(\phi_j^{(i-1)}, \sigma_\phi^2)$.

Having sampled a value from the appropriate proposal distribution, the acceptance probability for the move is given by the standard MH formula

$$\min\left\{1, \left[\frac{\operatorname{trace}(\mathbf{P}^{\perp}_{A(\Phi)}\hat{\mathbf{R}}\mathbf{P}^{\perp}_{A(\Phi)})}{\operatorname{trace}(\mathbf{P}^{\perp}_{A(\Phi')}\hat{\mathbf{R}}\mathbf{P}^{\perp}_{A(\Phi')})}\right]^{-MK/2} \frac{q(\phi_j^{(i)}|\phi_j^{(i-1)})}{q(\phi_j^{(i-1)}|\phi_j^{(i)})}\right\} \tag{28}$$

where Φ' is similar to Φ, and the only difference lies in that $\phi_j^{(i-1)}$ is replaced with $\phi_j^{(i)}$ in Φ.

3.5. Design sampling functions for the unknown parameter L

Now, let us consider the case where L is unknown. Here, the Bayesian computation for the estimation of the joint posterior distribution $\hat{p}(\boldsymbol{\Phi}, L|\mathbf{x})$ is even more complex. One obvious solution would consist of running $L_{\max}$ independent Markov Chain Monte Carlo (MCMC) samplers, each being associated with a fixed number $L = 1, 2, \ldots, L_{\max}$. However, this approach suffers from severe drawbacks: computationally expensive and the same computational effort for some values with no interest. Standard MCMC method is not able to "jump" between different model orders L. However, recently, Green et al. have introduced a new flexible class of MCMC samplers (the so-called reversible jump MCMC) [10–16] that is capable of jumping between different model orders L. Based on this idea, for our problem, the following moves have to be implemented:

1) Birth Move: birth of a new signal, i.e. proposed a new ϕ at random on $[-\pi, \pi]$;
2) Death Move: death of an existing signal, i.e. removed a ϕ chosen randomly;
3) Update Move: update of the parameters of all ϕ.

The birth and death moves perform dimension changes, respectively, from L to $L+1$and L to $L-1$. For $L = 0$, the death move is impossible; for $L = L_{\max}$, the birth move is impossible.

Hence, the acceptance probabilities corresponding to the birth and death moves are:

$$\min\left\{1, \left[\frac{\operatorname{trace}(\mathbf{P}^{\perp}_{A(\Phi)}\hat{\mathbf{R}}\mathbf{P}^{\perp}_{A(\Phi)})}{\operatorname{trace}(\mathbf{P}^{\perp}_{A(\Phi')}\hat{\mathbf{R}}\mathbf{P}^{\perp}_{A(\Phi')})}\right]^{-MK/2} \frac{2\pi \exp(-C)}{L+1}\right\} \tag{29}$$

and

$$\min\left\{1, \left[\frac{\operatorname{trace}(\mathbf{P}^{\perp}_{A(\Phi)}\hat{\mathbf{R}}\mathbf{P}^{\perp}_{A(\Phi)})}{\operatorname{trace}(\mathbf{P}^{\perp}_{A(\Phi')}\hat{\mathbf{R}}\mathbf{P}^{\perp}_{A(\Phi')})}\right]^{-MK/2} \frac{L \exp(C)}{2\pi}\right\}, \tag{30}$$

respectively.

3.6. Description of the proposed algorithm

The simulated annealing algorithm is based on the following idea. Given an objective function $f(\boldsymbol{\Phi}, L) = -\log(p(\mathbf{x}|\boldsymbol{\Phi}, L))$ that we wish to minimize over the parameter subspace$(\boldsymbol{\Phi}, L)$, the corresponding Boltzmann distribution [6–10] admits a density $b_T(\boldsymbol{\Phi}, L) \propto \exp(-f(\boldsymbol{\Phi}, L)/T)$. This algorithm begins by simulating a Markov chain with stationary density $b_T(\boldsymbol{\Phi}, L)$ at some initial temperature $T = T_0$ and allow the chain to reach equilibrium. We then decrease the temperature and continue the chain until equilibrium is achieved at the new temperature. We repeat this process with the decrease of T, and the corresponding stationary density for the chain moves increasingly close to a point mass at the minimum of $f(\boldsymbol{\Phi}, L)$ and the system essentially 'freezes' in this minimal state.

Based on the above Sections 3.1, 3.2, 3.3, 3.4 and 3.5, the proposed algorithm can be described as follows:

Step 1 *(Initialization).* Choose some initial temperature T_0, some initial starting configuration $(\boldsymbol{\Phi}_0, L_0)$, and some cooling schedule.

Step 2 *(Iteration i).*

- Sample $u \sim U[0, 1]$ and set the temperature T_i according to the cooling schedule;
- If $u < p_b$ [where $p_b + p_d + p_u = 1$]
  - then Birth Move;

- else if $u < p_b + p_d$ then Death Move;
- else if $u > p_b + p_d$ then Update Move;

- End if.
- Perform an MH step with the annealed acceptance ratio corresponding to Birth Move, Death Move, and Update Move are:

$$\min\left\{1, \left(\left[\frac{\operatorname{trace}(\mathbf{P}^{\perp}_{A(\Phi)}\hat{\mathbf{R}}\mathbf{P}^{\perp}_{A(\Phi)})}{\operatorname{trace}(\mathbf{P}^{\perp}_{A(\Phi')}\hat{\mathbf{R}}\mathbf{P}^{\perp}_{A(\Phi')})}\right]^{-MK/2} \frac{2\pi\exp(-C)}{L+1}\right)^{1/T_i-1}\right\}, \tag{31}$$

$$\min\left\{1, \left(\left[\frac{\operatorname{trace}(\mathbf{P}^{\perp}_{A(\Phi)}\hat{\mathbf{R}}\mathbf{P}^{\perp}_{A(\Phi)})}{\operatorname{trace}(\mathbf{P}^{\perp}_{A(\Phi')}\hat{\mathbf{R}}\mathbf{P}^{\perp}_{A(\Phi')})}\right]^{-MK/2} \frac{L\exp(C)}{2\pi}\right)^{1/T_i-1}\right\} \tag{32}$$

and

$$\min\left\{1, \left(\left[\frac{\operatorname{trace}(\mathbf{P}^{\perp}_{A(\Phi)}\hat{\mathbf{R}}\mathbf{P}^{\perp}_{A(\Phi)})}{\operatorname{trace}(\mathbf{P}^{\perp}_{A(\Phi')}\hat{\mathbf{R}}\mathbf{P}^{\perp}_{A(\Phi')})}\right]^{-MK/2} \frac{q(\phi_j^{(i)}|\phi_j^{(i-1)})}{q(\phi_j^{(i-1)}|\phi_j^{(i)})}\right)^{1/T_i-1}\right\}, \tag{33}$$

respectively.

Step 3. Repeat Step 2 for a specified number of iterations N_0 until the chain is deemed to have reached an equilibrium state.

3.7. Discussion

The array signal processing problem has to do with 1) determine the number of signal components and 2) estimate the directions of arrival of the signals incident onto array of sensors. The proposed algorithm can simultaneously estimate the number and DOAs of narrow-band sources. Some algorithms, such as information theoretic criteria-type algorithm addressed by M. Wax [5], involve maximizing the likelihood function that is associated with each competing source number and then calculating the corresponding criteria values. Others, such as MUSIC and ESPRIT, estimate source DOAs from the EVD of covariance or cumulant matrices [2,3].

The prior choices are calibrated (see Eqs. (11) and (14)) in this paper so as to the problem of model selection within the penalized likelihood context (see Eq. (17)) can be mapped exactly to a problem of model selection via posterior probabilities (see Eq. (22)).

The proposed algorithm computes some cumulants of the properly chosen sensor outputs from the ULA, then maps trickly from MFT of these cumulants to the uniform distribution, and thus designs sampling function for the unknown parameters $[\phi_1, \phi_2, \ldots, \phi_L]$ so as to move to the regions of interest quickly. Although the samples from the MFT-based distribution merely leads to suboptimal values, the second one based on Gaussian distribution is perturbation about suboptimal values and results in the optimal solutions.

4. Simulation results

To verify the effectiveness of the proposed algorithm, we consider a 7-element array with element spacing $d = \frac{1}{2}\lambda$. Two equi-power statistically independent sources ($e^{j0.2\pi t}$ and $e^{j0.3\pi t}$), respectively with DOAs of 20° and 60°, impinge on this array. The received signals are polluted by zero-mean additive Gaussian noises.

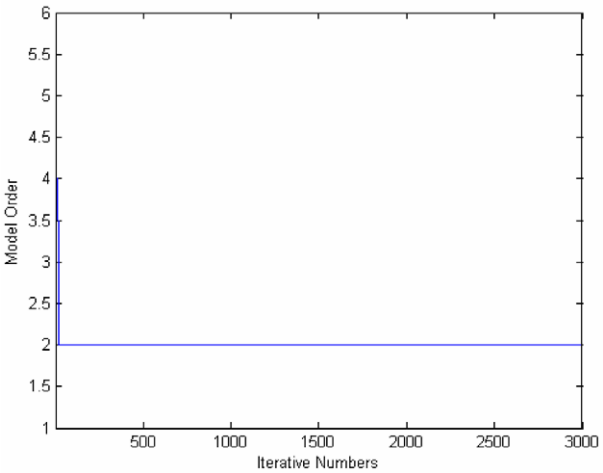
In the first experiment, the snapshot number and the SNR are set to 400 and 10 dB, respectively. The parameters of the algorithm are given in Table 1. Some initial starting configurations are given in Table 2. The proposed algorithm, respectively with AIC and BIC penalty terms, is implemented in a Monte Carlo run to jointly estimate source number and DOAs. The corresponding simulation results with AIC and BIC penalty terms are given in Figs. 4 and 5, respectively. Note that the results in the (c)–(e) subfigures are obtained after burin-in period of 500 samples. From these figures, we can see that our algorithm with the AIC penalty term has similar performance as that with the BIC penalty term.

Table 1
Algorithm parameters.

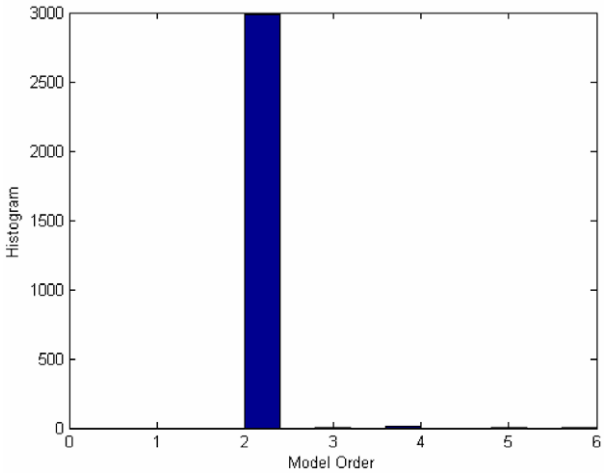
λ_0	σ_ϕ^2	p_b	p_d	p_u	$L_{\min}$	$L_{\max}$
0.2	$(1/250)^2$	0.35	0.4	0.25	1	6

Table 2
Initial starting configurations.

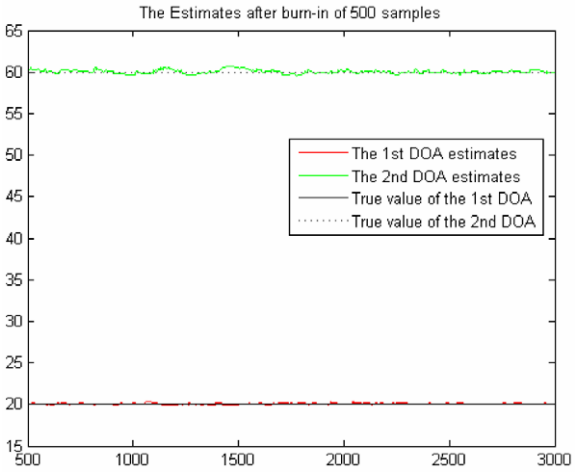
Chain length N_0	Cooling schedule	Initial $\boldsymbol{\Phi}_0$	Initial L_0
3000	$T_i = 1 - \frac{1-0.001}{3000}i$	Randomly sample from $[-\pi, \pi]$	6



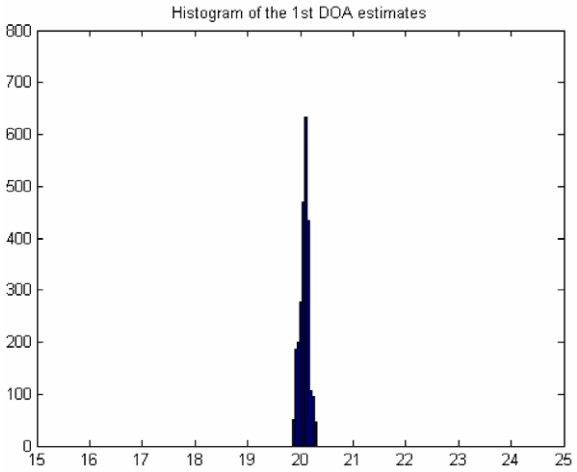
(a) Model order estimates versus annealing time.



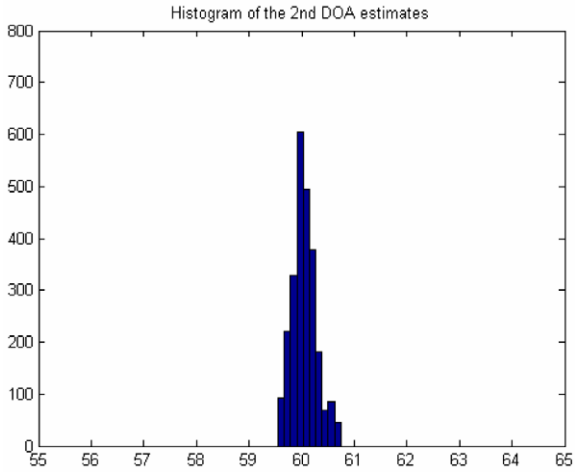
(b) Histogram of the estimates in (a).



(c) The estimates after burin-in of 500 samples.

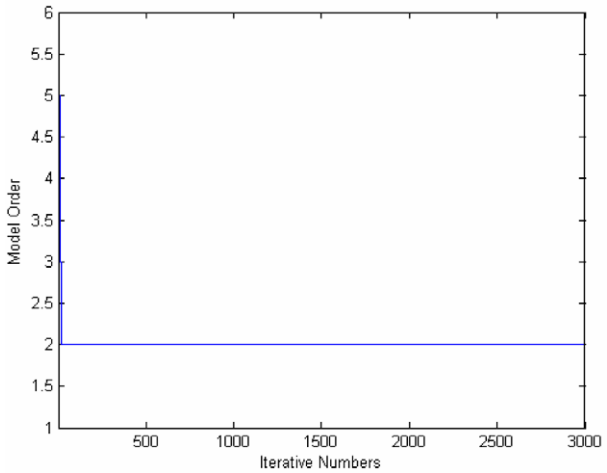


(d) Histogram of the 1st DOA estimates.

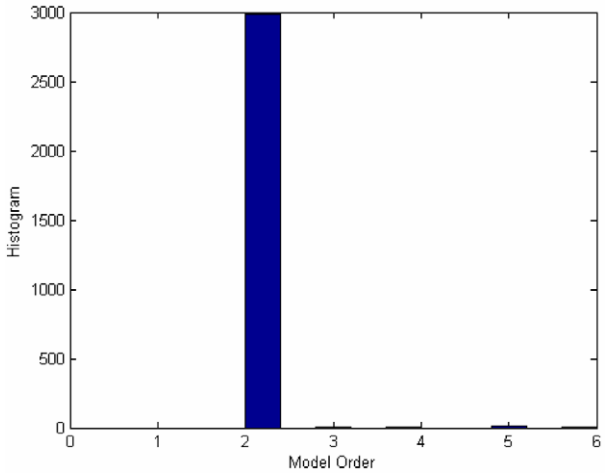


(e) Histogram of the 2nd DOA estimates.

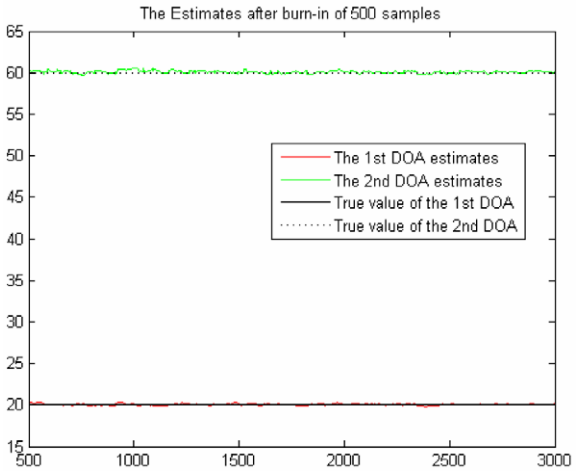
Fig. 4. Simulation results using the AIC penalty term.



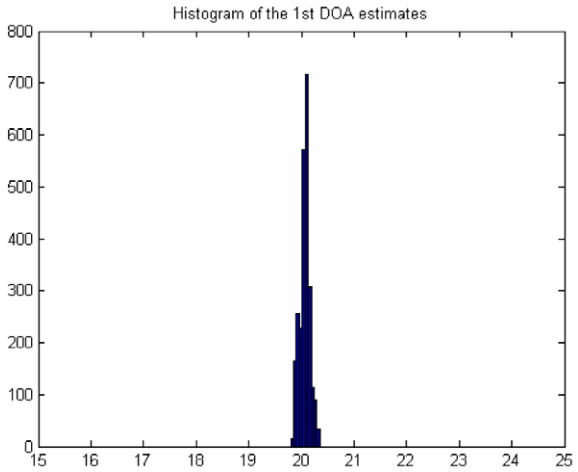
(a) Source number estimates versus annealing time.



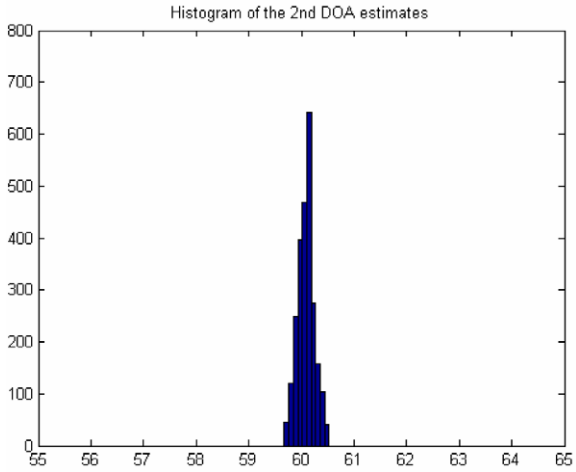
(b) Histogram of the estimates in (a).



(c) The estimates after burin-in of 500 samples.



(d) Histogram of the 1st DOA estimates.



(e) Histogram of the 2nd DOA estimates.

Fig. 5. Simulation results using the BIC penalty term.

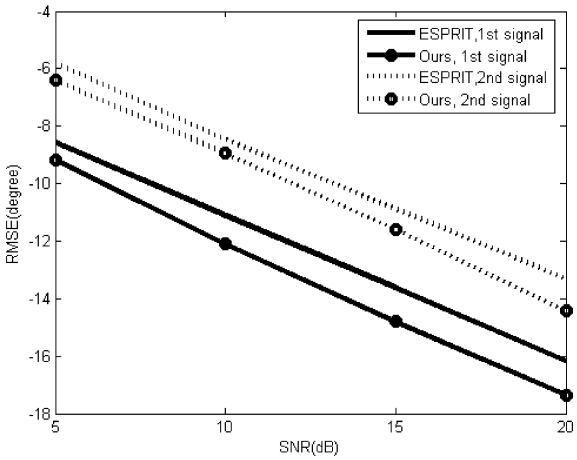
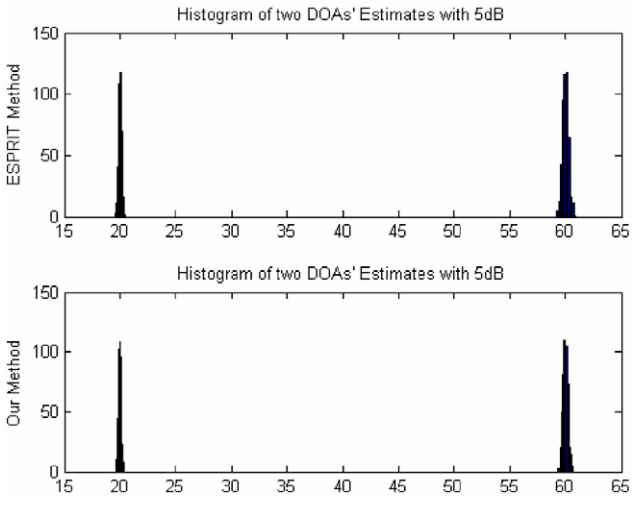
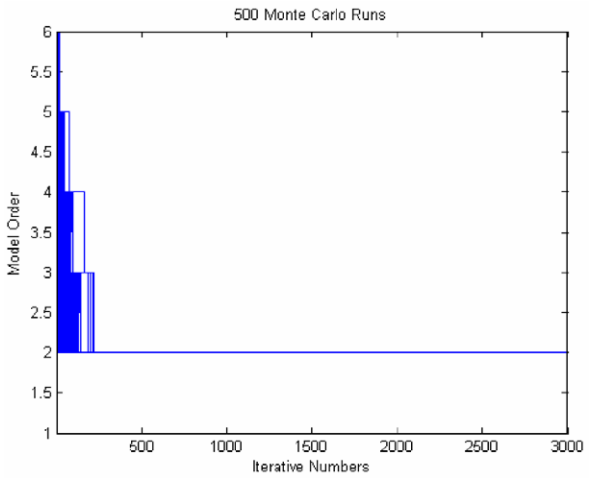


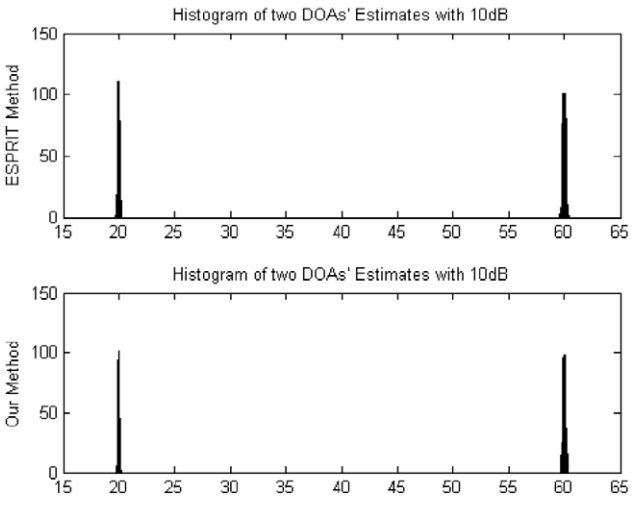
Fig. 6. RMSE of the DOAs' estimates versus SNR.



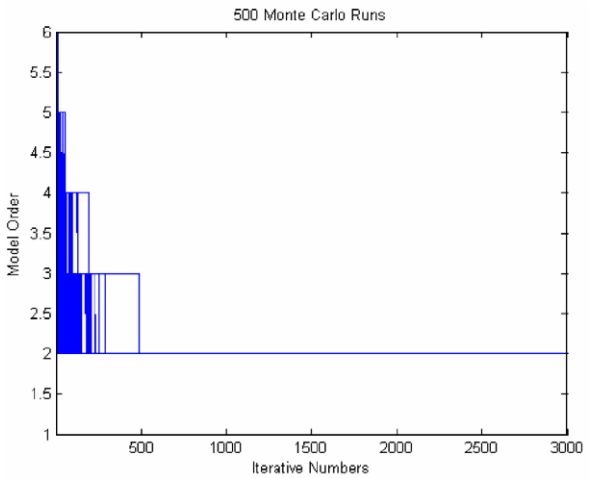
(a) Histogram of DOA estimates with 5 dB.



(b) 500 runs with 5 dB versus annealing time.

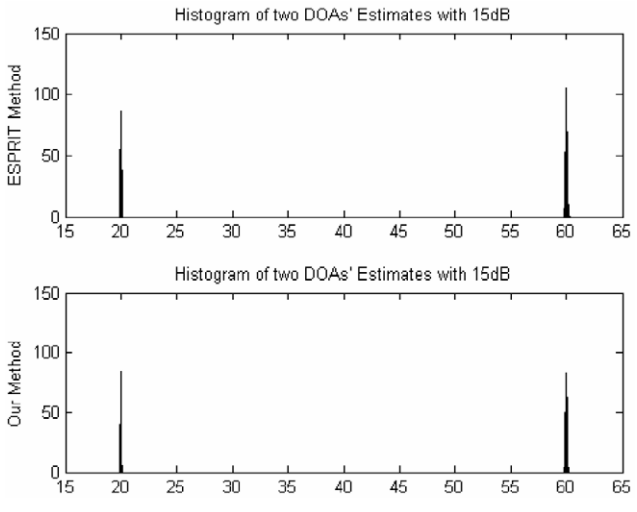


(c) Histogram of DOA estimates with 10 dB.

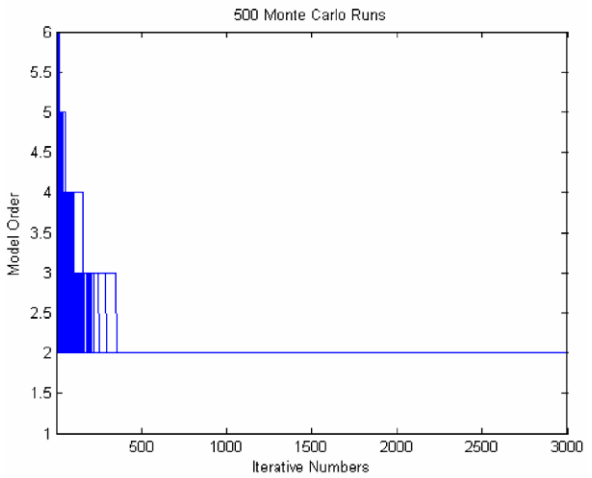


(d) 500 runs with 10 dB versus annealing time.

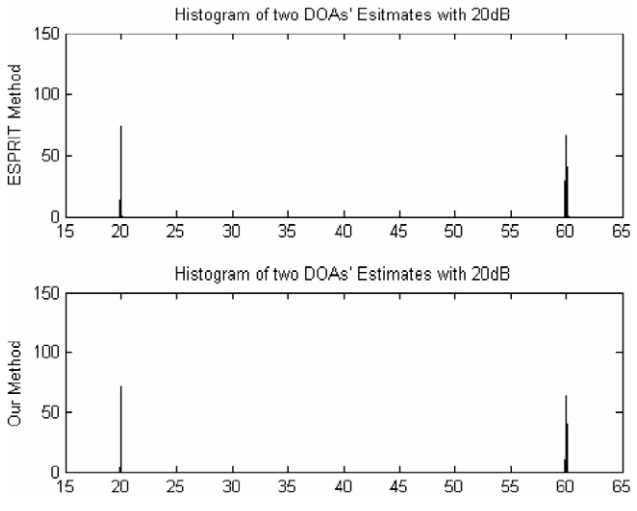
Fig. 7. Histogram of these estimates with different SNRs.



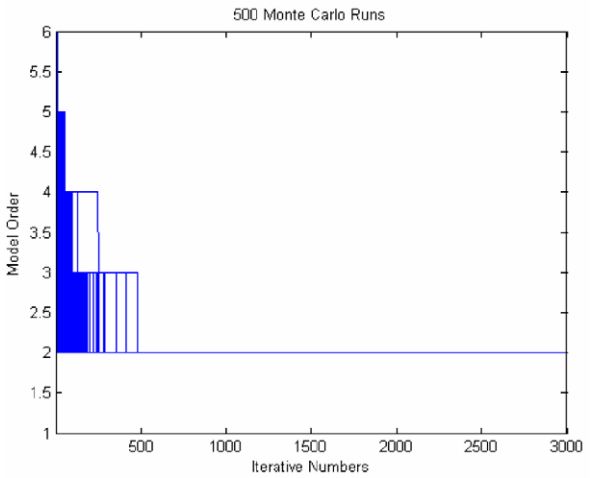
(e) Histogram of DOA estimates with 15 dB.



(f) 500 runs with 15 dB versus annealing time.



(g) Histogram of DOA estimates with 20 dB.



(h) 500 runs with 20 dB versus annealing time.

Fig. 7. (*continued*)

In the second experiment, the effect of high signal-to-noise (SNR) on the performance of the proposed algorithm with the AIC penalty term is investigated. The snapshot number is set to 400 and the SNR varies from 5 dB to 20 dB. 500 independent trials in total are run. The root mean square error (RMSE) for DOA estimation is defined as

$$\text{RMSE}(\text{the } l\text{th signal}) = \sqrt{\frac{1}{500}\sum_{i=1}^{500}(\hat{\theta}_{i,l} - \theta_l)^2}, \tag{34}$$

in which $\hat{\theta}_{i,l}$ stands for the estimation of the lth DOAθ_l in the ith trial. For comparison, we simultaneously execute the algorithm in [2], called "**ESPRIT**" in the following figures under the assumption that the source number is known. The averaged performance (RMSE of DOAs estimates versus SNR for two sources) of over 500 Monte Carlo runs for DOA estimates are shown in Fig. 6. In Fig. 7, the histograms of these estimates with different SNRs are given. From these figures, we can see that RMSE of the estimated parameters decrease as SNR increases. In addition, we observe that our method improves the performance significantly compared to the ESPRIT method.

5. Conclusion

A simulated annealing-based algorithm is proposed in this paper to jointly estimate the source number and DOAs of multiple signals received by sensor array. A novel and effective cumulant-based sampling method is developed to indirectly

sample DOA, which mainly relies on the effective aperture increase and removing AWGN capacities of cumulant. And these cumulant are obtained from some properly chosen sensor outputs from the uniform linear array. In addition, the Fourier Transform moduluses of these cumulants are mapped trickly to the uniform distribution.

Acknowledgments

The author would like to thank editors and anonymous reviewers for their valuable comments and suggestions. This work was supported by National Natural Foundation (60901059/F0103), Educational Department Foundation of Shaanxi Province (09JK629), Doctor Research Start Fund of Xi'an University of Technology (116-210903) of China, and Discipline Union Fund of Xi'an University of Technology (116-210905).

References

[1] S. Marcos, A. Marsal, M. Benidir, The propagator method for source bearing estimation, Signal Process. 42 (2) (1995) 121–138.
[2] R. Roy, T. Kailath, ESPRIT-estimation of signal parameters via rotational invariance techniques, IEEE Trans. Acoust. Speech Signal Process. 37 (7) (1989) 984–995.
[3] R. Schmidt, Multiple emitter location and signal parameter estimation, IEEE Trans. Antennas Propag. 34 (3) (1986) 276–280.
[4] Junli Liang, X. Zeng, B. Ji, J. Zhang, F. Zhao, A computational efficient algorithm for joint range-DOA-frequency estimation of near-field sources, Digital Signal Process. 19 (4) (2009) 596–611.
[5] M. Wax, T. Kailath, Detection of signals by information theoretic criteria, IEEE Trans. Acoust. Speech Signal Process. 33 (1985) 387–392.
[6] S. Kirkpatrick, C.D. Gelatt, M.P. Vecchi, Optimisation using simulated annealing, Science 220 (1983) 671–680.
[7] S. Kirkpatrick, Optimisation using simulated annealing:quantitative studies, J. Statist. Phys. 34 (1984) 975–986.
[8] S.P. Brooks, N. Friel, R. King, Classical model selection via simulated annealing, J. Roy. Statist. Soc. Ser. B 65 (2) (2003) 503–520.
[9] S. Bandyopadhyay, Simulated annealing using a reversible jump Markov chain Monte Carlo algorithm for fuzzy clustering, IEEE Trans. Knowl. Data Eng. 17 (4) (2005) 479–490.
[10] C. Andrieu, N. de Freitas, A. Doucet, Reversible jump MCMC simulated annealing for neural networks, in: Proceedings of the 16th Conference on Uncertainty Artificial Intelligence. San Francisco, CA, 2000, pp. 11–18.
[11] P.J. Green, Reversible jump Markov chain Monte Carlo computation Bayesian model determination, Biometrika 82 (1995) 711–732.
[12] K. Copsey, N. Gordon, A. Marrs, Bayesian analysis of generalized frequency-modulated signals, IEEE Trans. Signal Process. 50 (3) (2002) 725–735.
[13] C. Andrieu, A. Doucet, Joint Bayesian model selection estimation of noisy sinusoids via reversible jump MCMC, IEEE Trans. Signal Process. 47 (10) (1999) 2667–2676.
[14] C. Andrieu, N. de Freitas, A. Doucet, Robust full Bayesian learning for radial basis networks, Neural Comput. 13 (11) (2001) 2359–2407.
[15] C. Andrieu, A. Doucet, Joint Bayesian detection estimation of harmonics signals via reversible jump MCMC, Technical Report CUED-F-TR318, Cambridge University Engineering Department, Cambridge, 1998.
[16] C. Andrieu, N. de Freitas, A. Doucet, Robust full Bayesian learning for neural networks, Technical Report CUED/FINFENG/TR 343, Cambridge University Engineering Department, 1999.
[17] M.C. Dogan, J.M. Mendel, Applications of cumulants to array processing—part I: Aperture extension and array calibration, IEEE Trans. Signal Process. 435 (1995) 1200–1216.
[18] P. Chevalier, A. Ferreol, On the virtual array concept for the fourth order direction finding problem, IEEE Trans. Signal Process. 47 (9) (1999) 2592–2595.
[19] Junli Liang, Joint azimuth and elevation direction finding using cumulant, IEEE Sensors J. 9 (4) (2009) 390–398.
[20] Junli Liang, Ding Liu, Passive localization of near-field sources using cumulant, IEEE Sensors J. 9 (8) (2009) 953–960.

Junli Liang was born in China, in 1978. He received his M.S. degrees in Computer Science and Technology in Xidian University in 2004. In 2007, he received his Ph.D. degree in Signal and Information Processing from the Institute of Acoustics, Chinese Academy of Sciences. Currently, he is working in the School of Automation and Information Engineering and the School of Computer Science and Engineering, Xi'an University of Technology, Xi'an, China. So far, 20 papers of his have been accepted or published by IEEE Trans. Signal Processing, IEEE Trans. Antennas and Propagation, Digital Signal Processing, EURASIP Journal on Advances in Signal Processing, etc. During 2006–2008, he was serving as a reviewer of IEEE Trans. Signal Processing. His current research interests include neural network, array signal processing, adaptive filtering, and intelligent signal processing.

Ding Liu was born in China, in 1957. He received the B.S. and M.S. degrees in the School of Automation & Information Engineering, Xi'an University of Technology, Xi'an, China, in 1982 and 1987, respectively. In 1997, he received the Ph.D. degree from Xi'an Jiaotong University, Xi'an, China. From 1991 to 1993, he was a visiting scholar in the University of Fukui, Japan. Currently, he is the professor and president in Xi'an University of Technology. He has published more than 100 papers in IEEE Trans. Circuits & Systems II, and IEEE Sensors Journal, etc. His research interests include complex system's modeling & control, intelligent robot control, digital signal processing, and intelligent control theory.