Wideband Array Signal Processing Using MCMC Methods In Colored Noise

William Ng, James P. Reilly*, and Thia Kirubarajan
Department of Electrical and Computer Engineering,
McMaster University, 1280 Main St. W.,
Hamilton, Ontario,
Canada L8S 4K1

Abstract

This paper proposes a novel Bayesian solution to a difficult problem of joint detection and estimation of *wideband* sources impinging on a linear array of sensors in spatially colored noise with arbitrary covariance matrix. The wideband signals are approximated by an interpolation function and the signal samples. With the appropriate choices of prior distribution functions of the parameters, nuisance parameters, including the unknown noise covariance matrix, are analytically integrated out, allowing the use of the reversible jump Markov chain Monte Carlo (MCMC) procedure to simulateneously detect the unknown model order and estimate the inter-sensor delays (ISDs) of the sources.

Advantages of the proposed method include joint detection of model order and estimation of the ISD parameters, relaxation of the assumption of white Gaussian observation noise, and the fact that reliable performance can be obtained using significantly fewer observations than previous wideband methods. The ISD estimation performance of the proposed method is compared with the theoretical Cramér Rao lower bound (CRLB) for this problem. Simulation results demonstrate the effectiveness and robustness of the method.

I. Introduction

Array signal processing, a mature and specialized branch of signal processing, has found use in radar, sonar, communications, geophysical exploration, astrophysical exploration, biomedical signal processing, and acoustics [1] [2]. It focuses on extracting as much information as possible from an environment using an array of independent sensors. These sensors are placed at different points in space to "listen" to the received signal. In effect, the sensors provide a means of sampling the received signal in space.

Array signal processing has to do with 1) detecting the number of incident sources (model order), 2) estimating parameters, like direction-of-arrival (DOA) or inter-sensor delay (ISD) of the sources impinging onto the array, and 3) recovering the incident source waveforms. As the methods in [3] [4] solve for these problems jointly under the assumption that the incident signals are narrowband, the algorithm in [5] proposes the joint detection and estimation in the wideband scenario but assumes that the observation noise is white Gaussian. Array signal processing methods that can accomodate the unknown colored noise case are desirable since

Permission to publish abstract separately is granted.
J. Reilly, corresponding author: ph: 905 525 9140 x22895, fax: 905 521 2922, email: reillyj@mcmaster.ca

they offer improved performance in colored noise over methods developed assuming white noise. In addition, the colored background noise case is often encountered in the practical scenario.

In this paper, we propose an extension of the wideband array signal processing algorithm in [5], based on Markov chain Monte Carlo (MCMC) methods [3] [6] [7] [8], for joint and seamless detection and estimation of wideband sources in arbitrarily colored noise, using a single sensor array. In other words, no knowledge and assumption is required for the noise covariance matrix.

Since there exists no competing method that solves joint detection and estimation problem for wideband scenario in colored noise environment, we evaluate the performance of the proposed method with the theoretical Cramér Rao lower bound (CRLB) as well as the goodness-of-fit test [9].

There are several advantages offered by this approach in array signal processing. Firstly, by virtue of the *reversible jump Metropolis-Hastings algorithm* [10], the proposed MCMC approach *jointly* detects the model order and estimates all parameters of interest for wideband scenarios. This procedure is more efficient and accurate than other common approaches that would perform each process independently. Secondly, the method benefits from requiring only real arithmetic, resulting in significant reductions in hardware complexity, since the need for quadrature mixing to IF frequencies is alleviated. Also, like other wideband methods, the sources can be partially or fully correlated.

This paper is organized as follows. Section II presents a data model based on interpolation techniques to represent wideband signals as described in [5]. Section III describes the derivation of the necessary probability distributions. A description of the reversible jump MCMC algorithm used for model order detection is given in Section IV, followed by the simulation results and discussion in Section V. Conclusions are given in Section VI.

Notation: Bold upper case symbols denote matrices, bold lower case symbols denote vectors. The superscript ^{T} denotes the transpose operation, and the symbol "$\sim$" means "distributed as." The quantity $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ represents a normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, whereas $\mathcal{U}_{(a,b]}$ represents a uniform distribution over the open interval $(a, b]$. The quantity $p(\cdot)$ denotes a prior probability distribution, $l(\cdot)$ denotes a likelihood, and $\pi(\cdot)$ denotes a posterior distribution.

II. The Data Model

The development of the data model in this section is based on [5], so only a brief description is provided and interested readers should refer to [5].

Consider a uniform linear array with M elements is assumed, where K wideband signals are incident onto the array at angles ϕ_k, $k = 0, ..., K-1$, with respect to the axis of the array. Each signal $s_k(t)$ is bandlimited by its own upper and lower frequencies, f_k^u and f_k^l, such that

$$f_k^u = f_k^l + \Delta f_k, \tag{1}$$

where Δf_k is the bandwidth of the kth signal. Denote the inter-sensor delay (ISD) of the kth

signal by τ_k, defined by [1]

$$\tau_k \triangleq \frac{\Delta}{C} \sin \phi_k, \tag{2}$$

where Δ is the interspacing of the sensors and C is the speed of propagation. It can be shown [5] that the ISD τ_k can be bounded by $|\tau_k| \leq \frac{1}{2f_k^u}$. To define the maximum allowable inter-sensor delay, T_{max}, in the array for K bandlimited signals, we have

$$T_{max} = \min_{k=0,...,K-1} \left\{ \frac{1}{2f_k^u} \right\}. \tag{3}$$

Let $\boldsymbol{y}(n) \in \mathcal{R}^M$ be an observation vector, which represents the data received by a linear array of M sensors at the nth snapshot. The data vector is composed of K incident wideband signals $s_k(t)$, $k = 0, ..., K-1$, each of which is embedded in Gaussian noise. The received vector $\boldsymbol{y}(n)$ can be written as [5]

$$\boldsymbol{y}(n) = \sum_{k=0}^{K-1} \boldsymbol{s}_k(t - \tau_k) + \boldsymbol{w}(n), \quad n = 1, ..., N_t, \tag{4}$$

where N_t is the number of snapshots, $\boldsymbol{s}_k(t - \tau_k) \in \mathcal{R}^M$ is a vector of delayed signal for the kth source, defined as

$$\boldsymbol{s}_k(t - \tau_k) = [s_k(t), s_k(t - \tau_k), ..., s_k(t - m\tau_k), ..., s_k(t - (M-1)\tau_k)]^T, \quad m \in [0, M-1], \tag{5}$$

and $\boldsymbol{w}(n) \in \mathcal{R}^M$, a spatially colored noise vector, is an *iid* normally distributed according to $\mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_w)$, where $\boldsymbol{\Sigma}_w \in \mathcal{R}^{M \times M}$ is an unknown and arbitrary covariance matrix. According to [5], we can define a set of interpolation matrices, $\tilde{\boldsymbol{H}}(\tau_k) \in \mathcal{R}^{M \times L}$, where L is the number of taps in the interpolation function[1], as follows

$$\tilde{\boldsymbol{H}}(\tau_k) = \begin{cases} \boldsymbol{H}(\tau_k), & \text{if } \tau_k \geq 0, \\ \boldsymbol{E}_M \boldsymbol{H}(\tau_k), & \text{if } \tau_k < 0, \end{cases} \tag{6}$$

where the matrix $\boldsymbol{E}_M$ is an exchange matrix [11], and

$$\boldsymbol{H}(\tau_k) = \begin{bmatrix} h_0(0) & h_1(0) & \dots & h_{L-1}(0) \\ h_0(\tau_k) & h_1(\tau_k) & \dots & h_{L-1}(\tau_k) \\ \vdots & \vdots & \ddots & \vdots \\ h_0((M-1)\tau_k) & h_1((M-1)\tau_k) & \dots & h_{L-1}((M-1)\tau_k) \end{bmatrix}. \tag{7}$$

As a result, we can approximate the observation vector $\boldsymbol{y}(n)$ in terms of the interpolation matrices and the signal samples as follows

$$\boldsymbol{y}(n) \approx \sum_{k=0}^{K-1} \tilde{\boldsymbol{H}}(\tau_k) \boldsymbol{s}_k(n) + \boldsymbol{w}(n), \tag{8}$$

[1]For example, in the case of the uniform linear array, the interpolation sequence can be a windowed $sinc(\cdot)$ function.

where $\boldsymbol{s}_k(n) \in \mathcal{R}^L$ is a vector of the discrete signal samples of the kth source, given by

$$\boldsymbol{s}_k(n) = [s_k(n), s_k(n-1), ..., s_k(n-L+1)]^T. \tag{9}$$

Rearranging the columns in the interpolation matrices in (6) and (7), we have a new set of interpolation matrices, $\tilde{\boldsymbol{H}}_l(\boldsymbol{\tau}) \in \mathcal{R}^{M\times K}$, $l = 0, ..., L-1$, which are a collection of the lth columns in $\tilde{\boldsymbol{H}}(\tau_k)$ for K sources $\boldsymbol{\tau} = [\tau_0, \tau_1, ..., \tau_{K-1}]^T$. That is,

$$\tilde{\boldsymbol{H}}_l(\boldsymbol{\tau}) = [\tilde{\boldsymbol{h}}_l(\tau_0), \tilde{\boldsymbol{h}}_l(\tau_1), ..., \tilde{\boldsymbol{h}}_l(\tau_{K-1})]. \tag{10}$$

Accordingly, we may re-write (8) as follows

$$\boldsymbol{y}(n) = \sum_{l=0}^{L-1} \tilde{\boldsymbol{H}}_l(\boldsymbol{\tau})\boldsymbol{a}(n-l) + \boldsymbol{w}(n), \tag{11}$$

where $\boldsymbol{a}(n) = [s_0(n), s_1(n), ..., s_{K-1}(n)]^T$. Note that in (11) for notational convenience we replace the approximation sign by the equality and we adopt the equality from this point onwards. Denote $\boldsymbol{z}(n)$ as a transformed data vector as follows

$$\boldsymbol{z}(n) = \boldsymbol{y}(n) - \sum_{l=1}^{L-1} \tilde{\boldsymbol{H}}_l(\boldsymbol{\tau})\boldsymbol{a}(n-l), \tag{12}$$

which is the estimation error between the observation $\boldsymbol{y}(n)$ and its estimate using the $L-1$ past signal samples. Finally, the desired data model is expressed as follows

$$\boldsymbol{z}(n) = \tilde{\boldsymbol{H}}_0(\boldsymbol{\tau})\boldsymbol{a}(n) + \boldsymbol{w}(n). \tag{13}$$

There are several interesting features regarding the model in (13). It is similar to the familiar narrowband model [1], except that the data is modified and the steering matrix takes the form of an interpolation matrix. Further, the same model in (13) can accomodate either narrowband or wideband sources, without change of structure or parameters.

The signal and noise in (13) are assumed uncorrelated such that

$$E[\boldsymbol{a}^T(n)\tilde{\boldsymbol{H}}_0^T(\boldsymbol{\tau})\boldsymbol{w}(n)] = 0, \tag{14}$$

where $\boldsymbol{w}(n)$ and $\boldsymbol{a}(n)$ are the noise and signal vectors defined earlier. For a hypothesized number of signals k, we denote by $\boldsymbol{\mathcal{S}}(k) \in \mathcal{R}^{M\times k}$ the signal subspace spanned by the k column vectors in $\tilde{\boldsymbol{H}}_0(\boldsymbol{\tau})$. Further, we denote by $\boldsymbol{\mathcal{W}}(k) \in \mathcal{R}^{M\times(M-k)}$ the orthogonal complement of $\boldsymbol{\mathcal{S}}(k)$, such that

$$\boldsymbol{\mathcal{S}}^T(k)\boldsymbol{\mathcal{W}}(k) = \boldsymbol{0}. \tag{15}$$

III. Development of The Marginal Posterior Distribution

Here, it is assumed 1) that the noise vectors $\boldsymbol{w}(n)$ are *iid* Gaussian with unknown covariance matrix, and 2) that all the parameters describing the received signal are stationary throughout the entire observation interval. Using (13), we may define a set of N_t snapshots as

$$\boldsymbol{Z} = [\boldsymbol{z}(1), \ldots, \boldsymbol{z}(N_t)]\,. \tag{16}$$

Similarly, we create a set of amplitude vector $\boldsymbol{A}$ as follows

$$\boldsymbol{A} = [\boldsymbol{a}(1), \ldots, \boldsymbol{a}(N_t)]\,. \tag{17}$$

Then, the total likelihood function of all data can be expressed as follows

$$\ell(\boldsymbol{Z}|\boldsymbol{A}, \boldsymbol{\tau}, \Sigma_w, k) = \prod_{n=1}^{N_t} \frac{1}{(2\pi)^{M/2}|\boldsymbol{\Sigma}_w|^{1/2}} \exp\left\{ \frac{-1}{2} \left(\boldsymbol{z}(n) - \tilde{\boldsymbol{H}}_0(\boldsymbol{\tau})\boldsymbol{a}(n)\right)^T \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{z}(n) - \tilde{\boldsymbol{H}}_0(\boldsymbol{\tau})\boldsymbol{a}(n)\right)\right\}, \tag{18}$$

where k represents an estimate of the true number of sources, K.

Following the similar procedures in [12], we define an orthonormal matrix $\boldsymbol{Q}(\boldsymbol{\tau}, k) \in \mathcal{R}^{M\times M}$ as follows

$$\boldsymbol{Q}(\boldsymbol{\tau}, k) = [\boldsymbol{Q}_s(\boldsymbol{\tau}, k), \boldsymbol{Q}_w(\boldsymbol{\tau}, k)], \quad \boldsymbol{Q}_s(\boldsymbol{\tau}, k) \in \mathcal{R}^{M\times k}, \boldsymbol{Q}_w(\boldsymbol{\tau}, k) \in \mathcal{R}^{M\times(M-k)} \tag{19}$$

where $\boldsymbol{Q}_s(\boldsymbol{\tau}, k) \in \boldsymbol{\mathcal{S}}(k)$ and $\boldsymbol{Q}_w(\boldsymbol{\tau}, k) \in \boldsymbol{\mathcal{W}}(k)$. For notational convenience, we simplify the notation $\boldsymbol{Q}_s$ and $\boldsymbol{Q}_w$ instead of writing $\boldsymbol{Q}_s(\boldsymbol{\tau}, k)$ and $\boldsymbol{Q}_w(\boldsymbol{\tau}, k)$, respectively.

Let a set transformed data $\boldsymbol{x}(n)$, $n = 1, ..., N_t$ be defined as follows

$$\begin{aligned} \boldsymbol{x}(n) &= \boldsymbol{Q}^T(\boldsymbol{\tau}, k)\boldsymbol{z}(n), \\ &= \boldsymbol{x}_s(n) + \boldsymbol{x}_w(n), \end{aligned} \tag{20}$$

where $\boldsymbol{x}_s(n) \in \boldsymbol{\mathcal{S}}(k)$ and $\boldsymbol{x}_w(n) \in \boldsymbol{\mathcal{W}}(k)$ represent the signal and the noise components, respectively, given by

$$\boldsymbol{x}_s(n) = \boldsymbol{Q}_s^T \boldsymbol{z}(n), \tag{21}$$

$$\boldsymbol{x}_w(n) = \boldsymbol{Q}_w^T \boldsymbol{z}(n). \tag{22}$$

In the following development, we consider only the neighborhood around the true value where $\boldsymbol{\tau} \approx \boldsymbol{\tau}_o$ such that $\boldsymbol{\mathcal{S}}^T(k)\boldsymbol{Q}_w \approx \boldsymbol{0}$. We denote by $\boldsymbol{R}_{xx}$ the covariance matrix of the transformed data $\{\boldsymbol{x}(n), n = 1, ..., N_t\}$, defined as

$$\begin{aligned} \boldsymbol{R}_{xx} &\approx E[\boldsymbol{z}(n)\boldsymbol{z}^T(n)], \\ &= \begin{bmatrix} \boldsymbol{R}_{x_s x_s} & \boldsymbol{Q}_s^T \boldsymbol{\Sigma} \boldsymbol{Q}_w \\ \boldsymbol{Q}_w^T \boldsymbol{\Sigma} \boldsymbol{Q}_s & \boldsymbol{\Sigma}_{x_w} \end{bmatrix}. \end{aligned} \tag{23}$$

The block $\boldsymbol{R}_{x_s x_s}$ is the covariance matrix of the transformed signal component $\boldsymbol{x}_s(n)$ in (21), given by

$$\begin{aligned} \boldsymbol{R}_{x_s x_s} &= E[\boldsymbol{x}_s(n)\boldsymbol{x}_s^T(n)], \\ &= E[\tilde{\boldsymbol{a}}(n)\tilde{\boldsymbol{a}}^T(n)] + \boldsymbol{Q}_s^T \boldsymbol{\Sigma} \boldsymbol{Q}_s, \\ &= \boldsymbol{R}_{\tilde{a}\tilde{a}} + \boldsymbol{C} \end{aligned} \tag{24}$$

where $\tilde{\boldsymbol{a}}(n) = \boldsymbol{Q}_s^T \tilde{\boldsymbol{H}}_0(\boldsymbol{\tau})\boldsymbol{a}(n)$ and $\boldsymbol{C} = \boldsymbol{Q}_s^T \boldsymbol{\Sigma} \boldsymbol{Q}_s$. The block $\boldsymbol{\Sigma}_{x_w}$ is the covariance matrix of the transformed noise vector $\boldsymbol{x}_w(n)$ in (22), given by

$$\boldsymbol{\Sigma}_{x_w} = \boldsymbol{Q}_w^T \boldsymbol{\Sigma} \boldsymbol{Q}_w. \tag{25}$$

It can be shown [12] that locally arounf $\boldsymbol{\tau} \approx \boldsymbol{\tau}_o$, the vectors $\boldsymbol{x}_s(n)$ and $\boldsymbol{x}_w(n)$ are independent such that we have $\boldsymbol{x}_s(n)|\tilde{\boldsymbol{a}}(n) \sim \mathcal{N}(\tilde{\boldsymbol{a}}(n), \boldsymbol{C})$ and $\boldsymbol{x}_w(n) \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{x_w})$. In order to make the following development tractable, we neglect the off-diagonal blocks in (23), leading to a somewhat optimal solutiion [12]. Accordingly, we may re-write the likelihood function in (18) as follows

$$\begin{aligned}\ell(\boldsymbol{X}_s, \boldsymbol{X}_w|\tilde{\boldsymbol{A}}, \boldsymbol{\tau}, k, \boldsymbol{\Sigma}_{x_w}^{-1}) \approx \prod_{n=1}^{N_t} \frac{|\boldsymbol{C}|^{-1/2}}{(2\pi)^{k/2}} \exp\left\{ -\frac{1}{2}(\boldsymbol{x}_s(n) - \tilde{\boldsymbol{a}}(n))^T \boldsymbol{C}^{-1}(\boldsymbol{x}_s(n) - \tilde{\boldsymbol{a}}(n))\right\} \\ \times \frac{|\boldsymbol{\Sigma}_{x_w}|^{-1/2}}{(2\pi)^{(M-k)/2}} \exp\left\{ -\frac{1}{2}\boldsymbol{x}_w(n)^T \boldsymbol{\Sigma}_{x_w}^{-1}\boldsymbol{x}_w(n)\right\},\end{aligned} \tag{26}$$

where $\boldsymbol{X}_s = \{\boldsymbol{x}_s(n), n = 1, ..., N_t\}$, and similar definition is applied to $\boldsymbol{X}_w$ and $\tilde{\boldsymbol{A}}$, respectively.

The desired posterior distribution function $\pi(\boldsymbol{\tau}, k, \tilde{\boldsymbol{A}}, \boldsymbol{\Sigma}_{x_w}^{-1}) = p(\boldsymbol{\tau}, k, \tilde{\boldsymbol{A}}, \boldsymbol{\Sigma}_{x_w}^{-1}|\boldsymbol{X}_s, \boldsymbol{X}_w)$ can be expressed using Bayes' theorem in terms of the total likelihood function (26) and the prior distributions of the unknown parameters as

$$\pi(\boldsymbol{\tau}, k, \tilde{\boldsymbol{A}}, \boldsymbol{\Sigma}_{x_w}) \propto \ell(\boldsymbol{X}_s, \boldsymbol{X}_w|\tilde{\boldsymbol{A}}, \boldsymbol{\tau}, k, \boldsymbol{\Sigma}_{x_w}^{-1}) p(\tilde{\boldsymbol{A}}|\boldsymbol{\tau}, k, \boldsymbol{\Sigma}_{x_w}^{-1}) p(\boldsymbol{\tau}|k) p(\boldsymbol{\Sigma}_{x_w}^{-1}|\boldsymbol{\tau}, k) p(k). \tag{27}$$

We will discuss the assignment of the prior functions in (27) as follows.

The prior distribution function for $\tilde{\boldsymbol{A}}$ is selected as a normal distribution with zero mean and covariance matrix $\boldsymbol{D} = \delta^2 \boldsymbol{C}$, where δ^2 is a hyperparameter proportional to the signal-to-noise ratio (SNR). That is,

$$p(\tilde{\boldsymbol{A}}|\boldsymbol{\tau}, k, \boldsymbol{\Sigma}_{x_w}^{-1}) = \prod_{n=1}^{N_t} \mathcal{N}(\boldsymbol{0}, \boldsymbol{D}). \tag{28}$$

As a result, the corresponding joint confidence region at a specified probability level is significantly larger than that of a normal distribution with zero mean and covariance matrix $\boldsymbol{\Sigma}_{x_w}$. The prior distribution function for $\boldsymbol{\tau}$ is selected to be uniform such that

$$p(\boldsymbol{\tau}|k) = U[-T_{max}, T_{max}]^k. \tag{29}$$

The prior distribution function for k is chosen to be Poisson with expected number of sources Ξ, defined by

$$p(k) = \frac{\Xi^k}{k!} \exp(-\Xi). \tag{30}$$

Finally, the prior distribution function for $\boldsymbol{\Sigma}_{x_w}^{-1}$ is selected according to a noninformative multidimensional Jeffrey's prior [12] [13] for unknown covariance matrix. Accordingly, we have

$$p(\boldsymbol{\Sigma}_{x_w}^{-1}|\boldsymbol{\tau}, k) \propto |\boldsymbol{\Sigma}_{x_w}^{-1}|^{-(M-k)/2}. \tag{31}$$

Substituing (28) to (31) into (27), we have

$$\begin{aligned}\pi(\boldsymbol{\tau},k,\tilde{\boldsymbol{A}},\boldsymbol{\Sigma}_{x_w}^{-1}) \propto & \frac{|\boldsymbol{C}^{-1}|^{N_t/2}}{(2\pi)^{kN_t/2}}\exp\left\{-\frac{1}{2}\sum_{n=1}^{N_t}(\boldsymbol{x}_s(n)-\tilde{\boldsymbol{a}}(n))^T\boldsymbol{C}^{-1}(\boldsymbol{x}_s(n)-\tilde{\boldsymbol{a}}(n))\right\}\\ & \times\frac{|\boldsymbol{\Sigma}_{x_w}^{-1}|^{N_t/2}}{(2\pi)^{(M-k)N_t/2}}\exp\left\{-\frac{1}{2}\sum_{n=1}^{N_t}\boldsymbol{x}_w(n)^T\boldsymbol{\Sigma}_{x_w}^{-1}\boldsymbol{x}_w(n)\right\}\\ & \times\frac{|\boldsymbol{D}^{-1}|^{N_t/2}}{(2\pi)^{kN_t/2}}\exp\left\{-\frac{1}{2}\sum_{n=1}^{N_t}\tilde{\boldsymbol{a}}^T(n)\boldsymbol{D}^{-1}\tilde{\boldsymbol{a}}(n)\right\}\\ & \times\frac{1}{(2T_{max})^k}\times|\boldsymbol{\Sigma}_{x_w}^{-1}|^{-(M-k)/2}\times\frac{\Xi^k}{k!}\exp(-\Xi).\end{aligned} \tag{32}$$

Collecting like terms in (32) yields

$$\begin{aligned}\pi(\boldsymbol{\tau},k,\tilde{\boldsymbol{A}},\boldsymbol{\Sigma}_{x_w}^{-1}) \propto & \frac{|\boldsymbol{D}^{-1}|^{N_t/2}}{(2\pi)^{kN_t/2}}\frac{|\boldsymbol{C}^{-1}|^{N_t/2}}{(2\pi)^{kN_t/2}}\exp\left\{-\frac{1}{2}\sum_{n=1}^{N_t}\left[(\boldsymbol{x}_s(n)-\tilde{\boldsymbol{a}}(n))^T\boldsymbol{C}^{-1}(\boldsymbol{x}_s(n)-\tilde{\boldsymbol{a}}(n))\right.\right.\\ & \left.\left.+\tilde{\boldsymbol{a}}^T(n)\boldsymbol{D}^{-1}\tilde{\boldsymbol{a}}(n)\right]\right\}\times\frac{|\boldsymbol{\Sigma}_{x_w}^{-1}|^{N_t/2}}{(2\pi)^{(M-k)N_t/2}}\exp\left\{-\frac{1}{2}\sum_{n=1}^{N_t}\boldsymbol{x}_w(n)^T\boldsymbol{\Sigma}_{x_w}^{-1}\boldsymbol{x}_w(n)\right\}\\ & \times\frac{1}{(2T_{max})^k}\times|\boldsymbol{\Sigma}_{x_w}^{-1}|^{-(M-k)/2}\times\frac{\Xi^k}{k!}\exp(-\Xi).\end{aligned} \tag{33}$$

With the assumption that the matrix $\boldsymbol{D}$ is much larger than $\boldsymbol{\Sigma}_{x_w}$, for sufficiently large number N_t of observations, the term $\tilde{\boldsymbol{a}}^T(n)\boldsymbol{D}^{-1}\tilde{\boldsymbol{a}}(n)$ in the first exponential on the right in (33) is comparatively smaller than the term $\tilde{\boldsymbol{a}}^T(n)\boldsymbol{C}^{-1}\tilde{\boldsymbol{a}}(n)$ in the first exponential. As a result, we have the following simplified posterior distribution function

$$\begin{aligned}\pi(\boldsymbol{\tau},k,\tilde{\boldsymbol{A}},\boldsymbol{\Sigma}_{x_w}^{-1}) \propto & |\boldsymbol{D}^{-1}|^{N_t/2}|\boldsymbol{C}^{-1}|^{N_t/2}\exp\left\{-\frac{1}{2}\sum_{n=1}^{N_t}(\boldsymbol{x}_s(n)-\tilde{\boldsymbol{a}}(n))^T\boldsymbol{C}^{-1}(\boldsymbol{x}_s(n)-\tilde{\boldsymbol{a}}(n))\right\}\\ & \times|\boldsymbol{\Sigma}_{x_w}^{-1}|^{(N_t-M+k)/2}\exp\left\{-\frac{1}{2}\sum_{n=1}^{N_t}\boldsymbol{x}_w(n)^T\boldsymbol{\Sigma}_{x_w}^{-1}\boldsymbol{x}_w(n)\right\}\\ & \times\frac{\Xi^k}{(2T_{max})^k k!}.\end{aligned} \tag{34}$$

In the problem at hand, the parameters we are interested in are $(\boldsymbol{\tau},k)$. To further simplify the posterior distribution function $\pi(\boldsymbol{\tau},k,\tilde{\boldsymbol{A}},\boldsymbol{\Sigma}_{x_w})$, we can analytically integrate out the nuisance parameters $(\tilde{\boldsymbol{a}}(n),\boldsymbol{\Sigma}_{x_w}^{-1})$, i.e.,

$$\pi(\boldsymbol{\tau},k)=\int\int\pi(\boldsymbol{\tau},k,\tilde{\boldsymbol{A}},\boldsymbol{\Sigma}_{x_w}^{-1})d\tilde{\boldsymbol{A}}d\boldsymbol{\Sigma}_{x_w}^{-1}. \tag{35}$$

It can be shown that after the parameter $\tilde{\boldsymbol{a}}(n)$ has been integrated out, we have

$$\begin{aligned}\pi(\boldsymbol{\tau},k,\boldsymbol{\Sigma}_{x_w}^{-1}) &\propto |\boldsymbol{\Sigma}_{x_w}^{-1}|^{(N_t/2-(M-k)/2)}\exp\left\{-\frac{1}{2}\sum_{n=1}^{N_t}\boldsymbol{x}_w(n)^T\boldsymbol{\Sigma}_{x_w}^{-1}\boldsymbol{x}_w(n)\right\}\times\frac{\Xi^k}{(2T_{max})^k k!}\frac{1}{(\delta^2)^{kN_t/2}},\\ &=\left[|\boldsymbol{\Sigma}_{x_w}^{-1}|^{(N_t/2-(M-k)/2)}\exp\left\{-\frac{1}{2}\mathrm{Tr}N_t\widehat{\boldsymbol{R}}_{x_wx_w}(\boldsymbol{\tau},k)\boldsymbol{\Sigma}_{x_w}^{-1}\right\}\right]\times\frac{\Xi^k}{(2T_{max})^k k!}\frac{1}{(\delta^2)^{kN_t/2}},\end{aligned} \tag{36}$$

where $\mathrm{Tr}(\cdot)$ is the trace operator, $\widehat{\boldsymbol{R}}_{x_w x_w}(\boldsymbol{\tau}, k)$ is the sample covariance matrix of $\boldsymbol{x}_w(n)$, defined as

$$\widehat{\boldsymbol{R}}_{x_w x_w}(\boldsymbol{\tau}, k) = \sum_{n=1}^{N_t} \boldsymbol{x}_w(n)\boldsymbol{x}_w^T(n), \tag{37}$$

and in the last term on the right in (36), we use for $\boldsymbol{D} = \delta^2 \boldsymbol{I}_k$. To integrate out the noise covariance $\boldsymbol{\Sigma}_{x_w}^{-1}$, one can compare the term inside the square brackets in (36) with the Wishart distribution [14] on $\boldsymbol{\Sigma}_{x_w}^{-1}$, with order $\rho = M - k$ and parameter $N_t \widehat{\boldsymbol{R}}_{x_w x_w}(\boldsymbol{\tau}, k)$. This term integrates to $I(\widehat{\boldsymbol{R}}_{x_w x_w}(\boldsymbol{\tau}, k))$ as follows,

$$I(\widehat{\boldsymbol{R}}_{x_w x_w}(\boldsymbol{\tau}, k)) = \pi^{(1/4)(M-k)(M-k-1)} \prod_{i=1}^{M-k} \Gamma((N_t - i + 1)/2) |N_t \widehat{\boldsymbol{R}}_{x_w x_w}(\boldsymbol{\tau}, k)|^{-N_t/2}, \tag{38}$$

where $\Gamma(\cdot)$ is the Gamma function. Accordingly, the simplified posterior distribution function $\pi(\boldsymbol{\tau}, k)$ becomes

$$\pi(\boldsymbol{\tau}, k) \propto \frac{\pi^{(1/4)(M-k)(M-k-1)}}{(2T_{max}/\Xi)^k k! (\delta^2)^{kN_t/2}} \prod_{i=1}^{M-k} \Gamma((N_t - i + 1)/2) |N_t \widehat{\boldsymbol{R}}_{x_w x_w}(\boldsymbol{\tau}, k)|^{-N_t/2}. \tag{39}$$

The objective is to estimate the parameters of this highly nonlinear function, as the following *maximum a posteriori* (MAP) estimate

$$\left\{ \widehat{\boldsymbol{\tau}}, \widehat{k} \right\} = \arg\max_{k, \boldsymbol{\tau}} \pi(\boldsymbol{\tau}, k). \tag{40}$$

To recover the signal amplitude $\boldsymbol{a}(n)$, one can adopt a *maximum a posteriori* estimation on the expression in (33), given all other parameters, as follows

$$\boldsymbol{a}_{MAP}(n) \triangleq (1 + \delta^2)^{-1} \left[\tilde{\boldsymbol{H}}_0^T(\boldsymbol{\tau}) \boldsymbol{Q}_s \boldsymbol{Q}_s^T \tilde{\boldsymbol{H}}_0(\boldsymbol{\tau}) \right]^{-1} \tilde{\boldsymbol{H}}_0^T(\boldsymbol{\tau}) \boldsymbol{Q}_s^T \boldsymbol{x}_s(n), \tag{41}$$

where $\boldsymbol{D}^{-1} = \delta^{-2} \boldsymbol{C}^{-1}$ and $\boldsymbol{x}_s(n) = \boldsymbol{Q}_s^T \boldsymbol{z}(n)$.

A. Signal Recovery

According to the signal model, it is clear that the source sample $\boldsymbol{a}(n - L + 1)$ contributes to L successive snapshots $\boldsymbol{y}(n - L + 1)$ to $\boldsymbol{y}(n)$. Efficient estimation of the source sample $\boldsymbol{a}(n - L + 1)$ requires use of all these snapshots. From (11), we have

$$\boldsymbol{y}(n) = \sum_{l=0}^{L-1} \tilde{\boldsymbol{H}}_l(\boldsymbol{\tau}) \boldsymbol{a}(n - l) + \sigma_w \boldsymbol{w}(n), \tag{42}$$

The natural log of the posterior in (32) can therefore be written as

$$L(\boldsymbol{\tau}, \boldsymbol{A}, \boldsymbol{C}^{-1}) \propto \kappa - \frac{1}{2} \sum_{n=1}^{N_t} (\boldsymbol{x}_s(n) - \tilde{\boldsymbol{a}}(n))^T \boldsymbol{C}^{-1} (\boldsymbol{x}_s(n) - \tilde{\boldsymbol{a}}(n)) - \frac{1}{2} \sum_{n=1}^{N_t} \tilde{\boldsymbol{a}}^T(n) \boldsymbol{D}^{-1} \tilde{\boldsymbol{a}}(n), \tag{43}$$

where κ is a term independent of $\tilde{\boldsymbol{a}}(n)$ and $\boldsymbol{D}^{-1} = \delta^{-2} \boldsymbol{C}^{-1}$. The desired estimate $\hat{\boldsymbol{a}}(n - L + 1)$ is then obtained by maximizing (43) with respect to $\boldsymbol{a}(n - L + 1)$. The result is

$$\hat{\boldsymbol{a}}(n - L + 1) = (1 + \delta^{-2})^{-1} \left[\tilde{\boldsymbol{\mathcal{H}}}^T(\boldsymbol{\tau}) \tilde{\boldsymbol{\mathcal{H}}}(\boldsymbol{\tau}) \right]^{-1} \tilde{\boldsymbol{\mathcal{H}}}^T(\boldsymbol{\tau}) \boldsymbol{\epsilon}(n), \tag{44}$$

where

$$\tilde{\boldsymbol{\mathcal{H}}}(\boldsymbol{\tau}) = \left[\tilde{\boldsymbol{H}}_{L-1}^T(\boldsymbol{\tau})\boldsymbol{Q}_s^T, \ldots, \tilde{\boldsymbol{H}}_1^T(\boldsymbol{\tau})\boldsymbol{Q}_s^T, \tilde{\boldsymbol{H}}_0^T(\boldsymbol{\tau})\boldsymbol{Q}_s^T\right]^T. \tag{45}$$

and

$$\boldsymbol{\epsilon}(n) = \left[\boldsymbol{\epsilon}_0(n), \boldsymbol{\epsilon}_1(n), \ldots, \boldsymbol{\epsilon}_{L-1}(n)\right]^T. \tag{46}$$

The quantity $\boldsymbol{\epsilon}_p(n), p = 0, \ldots, L-1$, removes the contribution from samples other than $\boldsymbol{a}(n - L + 1)$ in the observation $\boldsymbol{y}(n-p)$. It is defined as

$$\begin{aligned} \boldsymbol{\epsilon}_p(n) &= \boldsymbol{Q}_s^T(\boldsymbol{y}(n-p) - \boldsymbol{x}_p(n)), \\ &= \boldsymbol{Q}_s^T(\tilde{\boldsymbol{H}}_{L-1-p}(\boldsymbol{\tau})\boldsymbol{a}(n-L+1) + \sigma_w \boldsymbol{w}(n)), \end{aligned} \tag{47}$$

where

$$\boldsymbol{x}_p(n) \stackrel{\Delta}{=} \sum_{l \neq p}^{L-1} \tilde{\boldsymbol{H}}_l(\boldsymbol{\tau})\boldsymbol{a}(n-p-l). \tag{48}$$

The $\boldsymbol{\epsilon}_p(n)$ are determined by evaluating the $\boldsymbol{x_p}(n)$ in (48), where the unknown quantities $\boldsymbol{a}(n - p - l)$ for $n - p - l > n - L + 1$ in (48) are tentatively estimated suboptimally using (41). The estimation of $\boldsymbol{a}(n)$ according to (44) results in significantly improved performance compared to the other, more direct methods for the estimation of this parameter.

The MCMC procedure described in the next section proposes a candidate sample $(\boldsymbol{\tau}^\star, k^\star)$ and requires evaluation of the posterior distribution given by (39) for that candidate. The following schema summarizes the process used for this evaluation.

Evaluation of the Posterior Density

1. Given a candidate sample $(\boldsymbol{\tau}^\star,\ k^\star)$ from the MCMC procedure, and a suitable interpolation function such as a windowed $sinc(\cdot)$, compute $\tilde{\boldsymbol{H}}_l(\boldsymbol{\tau}^\star), l = 0, \ldots, L-1$ of order $k^\star$ from (10).
2. For sample index $n = L-1, \ldots, N$
   - Follow the steps described in Section III-A to obtain $\hat{\boldsymbol{a}}(n - L + 1)$, in (44).
   - Given the source amplitudes, evaluate $\boldsymbol{z}(n-L+1)$ according to (13) and hence $\boldsymbol{x}_w(n-L+1) = \boldsymbol{Q}_w^T\boldsymbol{z}(n-L+1)$.
3. Given the $\boldsymbol{x}_w(n)$, $\boldsymbol{R}_{x_w x_w}$ in (37) can be computed, which allows evaluation of the posterior density $\pi(\boldsymbol{\tau}, k)$ in (39). This quantity is then used by the MCMC procedure to determine whether the candidate $(\boldsymbol{\tau}^\star,\ k^\star)$ is accepted as a sample.

The computational requirements of the above algorithm are mitigated by the fact that the matrix $\left[\tilde{\boldsymbol{\mathcal{H}}}^T(\boldsymbol{\tau})\tilde{\boldsymbol{\mathcal{H}}}(\boldsymbol{\tau})\right]^{-1}\tilde{\boldsymbol{\mathcal{H}}}^T(\boldsymbol{\tau})$ in (44) is independent of n and therefore need only be computed once per MCMC iteration.

B. Model Order Determination

In this section, we present an analysis on the hyper-parameter δ^2 in (39) for consistent determination of model order. According to (39), we can obtain the marginal posterior distribution for model order k as

$$\pi(k|\boldsymbol{X}_w) \quad \propto \quad \int_{\boldsymbol{\Phi}_k} \pi(\boldsymbol{\tau}, k|\boldsymbol{X}_w) d\boldsymbol{\tau}. \tag{49}$$

Denoting the true model order and delay values by k_0 and $\boldsymbol{\tau_0}$ respectively, we perform the following eigendecomposition, at $\boldsymbol{\tau} = \boldsymbol{\tau_0}$

$$N_t \widehat{\boldsymbol{R}}_{x_w x_w}(\boldsymbol{\tau}, k) = \boldsymbol{U}(\boldsymbol{\tau_0}) \Lambda(\boldsymbol{\tau_0}) \boldsymbol{U}^H(\boldsymbol{\tau_0}), \tag{50}$$

where $\boldsymbol{U}(\boldsymbol{\tau_0})$ is an orthonormal matrix which contains the $M-k$ eigenvectors associated with the $M-k$ smallest (noise) eigenvalues of the matrix $N_t \widehat{\boldsymbol{R}}_{x_w x_w}(\boldsymbol{\tau}, k)$, which are placed in the diagonal matrix, $\boldsymbol{\Lambda}(\boldsymbol{\tau_0}) = \text{diag}\,[\lambda_1, \lambda_2, \ldots, \lambda_{\text{M-k}}]$. For convenience, these eigenvalues λ_i, $i = 1, 2, ..., M-k$ are arranged in *assending* order. Assuming that N is large, and that the SNR level is moderate, then the posterior distribution function $\pi(\boldsymbol{\tau}, k|\boldsymbol{X}_w)$ concentrates around the true value $\boldsymbol{\tau_0}$. As a result, we can approximate the integral in (49) as follows:

$$\pi(k) \propto \frac{\pi^{(1/4)(M-k)(M-k-1)}}{(2T_{max}/\Xi)^k k! (\delta^2)^{kN_t/2}} \prod_{i=1}^{M-k} \Gamma((N_t - i + 1)/2) \Big(\prod_{i=1}^{M-k} \lambda_i \Big)^{-N_t/2}. \tag{51}$$

Let us define the event E_i as the declaration of a model order in error by i signals. Thus, the event E_i will occur if we declare $\hat{k} = k_0 + i$ or $\hat{k} = k_0 - i$. We assume $P(E_1) > P(E_2) > \cdots > P(E_{M-1})$. Accordingly, sufficient conditions which must be satisfied for consistent detection of the model order are

$$\lim_{N\to\infty} \frac{\pi(k_0 + 1|\boldsymbol{X}_w)}{\pi(k_0|\boldsymbol{X}_w)} \quad \to \quad 0, \tag{52}$$

$$\lim_{N\to\infty} \frac{\pi(k_0 - 1|\boldsymbol{X}_w}{\pi(k_0|\boldsymbol{X}_w} \quad \to \quad 0. \tag{53}$$

From (39), we have

$$\frac{\pi(k_0 + 1|\boldsymbol{X}_w)}{\pi(k_0|\boldsymbol{X}_w)} = \left(\frac{M - k_0 - 2}{M - k_0}\right) \left(\frac{\Xi}{2T_{max}}\right) \left(\frac{\Gamma^{-1}((N_t - M + k_0 + 1)/2)}{k_0 (\delta^2)^{N_t/2}}\right) (\lambda_{M-k_0})^{N_t/2}, \tag{54}$$

and

$$\frac{\pi(k_0 - 1|\boldsymbol{X}_w)}{\pi(k_0|\boldsymbol{X}_w)} = \left(\frac{M - k_0 + 1}{M - k_0 - 1}\right) \left(\frac{2T_{max}}{\Xi}\right) \Big(\Gamma((N_t - M + k_0)/2)\Big) \Big(k_0 (\delta^2)^{N_t/2}\Big) (\lambda_{M-k_0+1})^{-N_t/2}. \tag{55}$$

Given that parameters other than δ^2 and λ_{M-k_0} in (54) are scaling factors, we can see that (52) is satisfied if

$$\delta^2 > \lambda_{M-k_0}. \tag{56}$$

Similarly, from (55) we can see that (53) is satisfied if

$$\delta^2 < \lambda_{M-k_0+1}. \tag{57}$$

Therefore, combining (56) and (57) we have

$$\lambda_{M-k_0} < \delta^2 < \lambda_{M-k_0+1}. \tag{58}$$

Therefore, the specific range of the hyper-parameter δ^2 is dependent on the number of sensors, M, and the current SNR level. If δ^2 is set too small, the algorithm tends to overestimate the model order. Likewise, the model will be underestimated if δ^2 is set too large.

Referring to (58), the determination of the correct value of δ^2 requires the knowledge of the true model order k_0. However, we can still obtain useful information from (58) by having an estimate of the SNR level. This is a reasonable assumption, as we can obtain an estimate of the noise level by listening for the power level when it is assumed there is no signal present. Similarly, we can get an estimate of signal power by listening when the signal is transmitting. Thus, the right limit in (58) can be approximated if there is some knowledge of the SNR level. In practice, the left–hand term in (58) is close to unity. With this knowledge, we can obtain a reasonable estimate of the range within which δ^2 must fall.

IV. Reversible Jump MCMC

We use the *reversible jump* MCMC algorithm [10] to perform the Bayesian computation in jointly detecting the desired model order and extracting the other parameters of interest from the posterior distribution. The reversibile jump MCMC algorithm itself is similar to the Metropolis-Hastings (MH) [15] algorithm, but it allows the sampling process to jump between subspaces of different dimensions, which facilitates the detection of model order. As the procedures are similar to those in [5] [12], the following presentation is therefore brief. A further treatment on model order detection using MCMC methods is given in [7].

Denote the whole parameter space by $\cup_{k=0}^{k_{max}} k \times \mathbf{\Phi}_k$, where $\mathbf{\Phi}_k$ is the space of the parameters of the model of order k, and k_{max} is the maximum allowable model order. At each iteration, candidate samples are chosen from a set of proposal distributions $q(\cdot|\cdot)$ corresponding to different model orders, which are randomly accepted according to an acceptance ratio, given by

$$r(\boldsymbol{\tau}^\star, k^\star), (\boldsymbol{\tau}, k)) = \frac{\pi(\boldsymbol{\tau}^\star, k^\star) q(\boldsymbol{\tau}, k|\boldsymbol{\tau}^\star, k^\star)}{\pi(\boldsymbol{\tau}, k) q(\boldsymbol{\tau}^\star, k^\star|\boldsymbol{\tau}, k)} \times \boldsymbol{J}((\boldsymbol{\tau}^\star, k^\star), (\boldsymbol{\tau}, k)), \tag{59}$$

where it can be shown [16] that in our case, $\boldsymbol{J}((\boldsymbol{\tau}^\star, k^\star), (\boldsymbol{\tau}, k)) = 1$. This quantity is the Jacobian of the transformation, required to reconcile the total probability between spaces of different dimensions so that the reversibility condition is satisfied. The probability of accepting the candidate as the new state is given by $\alpha = \min\{1, r\}$. If the candidate is accepted, the chain takes the new state; otherwise the chain remains at the current state.

In the reversible jump algorithm, we choose our set of proposal distributions to correspond to the following set of moves

1. the *birth* move, valid for $k \leq k_{max}$. Here, a new τ is proposed at random on $[-T_{max}, T_{max}]$.

2. the *death* move, valid for $k > 0$. Here, a τ is randomly chosen to be removed.
3. the *update* move. Here, the order of the model is held fixed and the parameters describing the sources are updated.

The probabilities for choosing each move are denoted by u_k, b_k, and d_k, respectively, such that $u_k + b_k + d_k = 1$ for all k. In accordance with [3], we choose

$$b_k = c \min\left\{\frac{p(k+1)}{p(k)}, 1\right\}, \quad d_{k+1} = c \min\left\{\frac{p(k)}{p(k+1)}, 1\right\}, \tag{60}$$

where $p(\cdot)$ is the prior distribution of the kth model according to (30), and c is a tuning parameter which determines the ratio of update moves to jump moves. We choose $c = 0.5$ so that the probability of a jump is between 0.5 and 1 at each iteration [10] . The overall description of the reversible jump MCMC algorithm is determined by the choice of move at each iteration. This description is summarized as follows

---

Reversible Jump MCMC

1. Initialization: set $\mathbf{\Phi}^{(i=0)} = \left(\boldsymbol{\tau}^{(i=0)}, k^{(i=0)}\right)$, where i is the iteration index.
2. Iteration i
   - Sample $u \sim U_{[0,1]}$,
   - Compute $b_{k^{(i)}}$ and $d_{k^{(i)}}$ according to (60), using the value of $k^{(i)}$, which is the model order of the ith iteration.
   - if $(u < b_{k^{(i)}})$ then execute a "birth move" (see Section IV-B),
   - else if $(u < b_{k^{(i)}} + d_{k^{(i)}})$ then execute a "death move" (see Section IV-C),
   - else, execute an update move (see Section IV-A).
3. $i \leftarrow i + 1$, goto step 2

We now give details for each move type.

A. Update Move

Here, we assume that the current state of the algorithm is in $\{\mathbf{\Phi}_k, k\}$. When the update move is selected, the algorithm samples only on the space of $\mathbf{\Phi}_k$ for a fixed k, according to a proposal distribution function $q(\cdot|\cdot) = U_{[-T_{max}, T_{max}]}$. Substituting the relevant terms into (59), we obtain the acceptance ratio for the update move as follows

$$r_{update} = \frac{|N_t \widehat{\boldsymbol{R}}_{x_w x_w}(\boldsymbol{\tau}^\star, k^\star)|^{-N_t}}{|N_t \widehat{\boldsymbol{R}}_{x_w x_w}(\boldsymbol{\tau}, k)|^{-N_t}}. \tag{61}$$

The candidate $\boldsymbol{\tau}_k^\star$ is then accepted as the current state $\boldsymbol{\tau}_k^{(i+1)} = \boldsymbol{\tau}_k^\star$, with probability

$$\alpha_{update}\left(\boldsymbol{\tau}_k, \boldsymbol{\tau}_k^\star\right) = \min\left\{1, r_{update}\right\}. \tag{62}$$

In each iteration, a candidate $\boldsymbol{\tau}_k^\star$ is first sampled according to $U_{[-T_{max}, T_{max}]}$, and then the schema in Sect. IIIB is executed which then allows evaluation of r_{update} and α_{update}.

B. Birth and Death Moves

In the birth move, we assume the state of the algorithm is in $\{\boldsymbol{\Phi}_k, k\}$ at the present ith iteration, and we wish to determine whether the state is in $\{\boldsymbol{\Phi}_{k+1}, k+1\}$ at the next iteration. The acceptance function of the birth move is therefore defined as

$$r_{birth} = \frac{\pi(\boldsymbol{\tau}^{\star}_{k+1}, k+1|\boldsymbol{Z})q(\boldsymbol{\tau}_k, k|\boldsymbol{\tau}^{\star}_{k+1}, k+1)}{\pi(\boldsymbol{\tau}_k, k|\boldsymbol{Z})q(\boldsymbol{\tau}^{\star}_{k+1}, k+1|\boldsymbol{\tau}_k, k)}. \tag{63}$$

We propose a delay vector $\boldsymbol{\tau}^{\star}$ as

$$\boldsymbol{\tau}^{\star}_{k+1} = \left[\boldsymbol{\tau}^{(i)}_k, \tau_c\right], \tag{64}$$

where $\boldsymbol{\tau}^{(i)}_k$ is the delay vector at the ith iteration, and τ_c is a new time delay candidate selected uniformly on $[-T_{max}, T_{max}]$. Note that the prior for model order k in (30) is independent of that for $\boldsymbol{\tau}$ in (29). In the birth move, only the one new source is a random variable; the remaining sources are treated as constants. Accordingly, the proposal distribution $q\left(\boldsymbol{\tau}^{\star}_{k+1}, k+1|\boldsymbol{\tau}_k, k\right)$ from (63) is then

$$q\left(\boldsymbol{\tau}^{\star}_{k+1}, k+1|\boldsymbol{\tau}_k, k\right) = p(k+1) \times \frac{1}{2T_{max}}. \tag{65}$$

In contrast, the the distribution $q(k, \boldsymbol{\tau}_k|k+1, \boldsymbol{\tau}^{\star}_{k+1})$ in (63) refers to the proposal distribution when one source of $k+1$ is randomly removed. Note that the prior for model order k is independent of that for $\boldsymbol{\tau}$. Thus, we have

$$q(\boldsymbol{\tau}_k, k|\boldsymbol{\tau}^{\star}_{k+1}, k+1) = p(k) \div \binom{k+1}{1}. \tag{66}$$

Substituting all relevant functions into (63) yields

$$r_{birth} = \frac{|N_t \widehat{\boldsymbol{R}}_{x_w x_w}(\boldsymbol{\tau}^{\star}, k+1)|^{-N_t}}{|N_t \widehat{\boldsymbol{R}}_{x_w x_w}(\boldsymbol{\tau}, k)|^{-N_t}} \times \frac{\pi^{-(M-k-1)/2}\Gamma^{-1}((2N_t - M + k + 1)/2)}{(k+1)(\delta^2)^{N_t/2}} \tag{67}$$

The probability of accepting a birth move is therefore defined as

$$\alpha_{birth} = \min\left\{1, r_{birth}\right\}. \tag{68}$$

As in the update move, in each iteration a candidate $\boldsymbol{\tau}^{\star}_{k^{\star}}$ for $k^{\star} = k+1$ is first sampled according to a proposal distribution function $q(\cdot|\cdot) = U_{[-T_{max}, T_{max}]}$. The acceptance ratio is evaluated by executing the schema of Sect. III-B, and then evaluating (67).

In order to maintain the invariant distribution of the reversible jump MCMC algorithm with respect to model order, the Markov chain must be *reversible* with respect to moves across subspaces of different model orders. That is, the probability of moving from model order k to $k+1$ must be equal to that of moving from $k+1$ to k. Therefore we propose a death move in which a source in the current state $(\boldsymbol{\tau}_{k+1}, k+1)$ is randomly selected to be removed such that the next state becomes $(\boldsymbol{\tau}_k, k)$ at the next iteration. A sufficient condition for reversibility with respect to model order [10] is that the acceptance ratio for the death move be defined as

$$r_{death} = \frac{1}{r_{birth}}, \tag{69}$$

Root pair	Value
1	$0.2934 \pm j0.5850$
2	$-0.3934 \pm j0.2805$

TABLE 5.1

List of the AR roots used in generating the spatially correlated noise.

Parameter	SNR (dB)	M	K	L	N	δ^2	F_s (Hz)	$\boldsymbol{\theta}$ (deg)	$\boldsymbol{\tau}$ (sec)
Value	14	8	2	8	30	25.12	$1,000$	$[-3.44, 3.44]$	$[-7.5, 7.5] \times 10^{-5}$

TABLE 5.2

Parameters for the experiment.

and the new candidate of dimension k is accepted with probability

$$\alpha_{death} = \min\left\{1, r_{death}\right\}. \tag{70}$$

gg

V. Simulation Results

The proposed algorithm is now applied solve a joint detection and estimation problem for wideband signals in colored noise environment. The spatially colored noise is generated with an 4th order AR process whose roots are listed in Table 5.1 The observations in the simulation are generated using (11) with the parameters described in Tables 5.2 to jointly detect and estimate the relevant parameters $(\boldsymbol{\tau}, k)$ and the source amplitudes $\boldsymbol{s}_k(n)$. In these experiments, the model order k and the delay parameters $\boldsymbol{\tau}$ are kept constant throughout the entire observation period. To set up the corresponding interpolation matrix for a set of $\boldsymbol{\tau}$, a *sinc* function is chosen such that the (m,l)th element of the interpolation matrix for the kth source is given by

$$\left[\tilde{\boldsymbol{H}}\right]_{m,l}(\tau_k) = \frac{\sin\left(\pi f_c\left(lT_s - m\tau_k\right)\right)}{\pi f_c\left(lT_s - m\tau_k\right)} \times W\left(lT_s - m\tau_k\right), \tag{71}$$

where T_s is the sampling interval, f_c is the cutoff frequency and $W(\cdot)$ is a Hamming window function.

In this experiment, we generate $K = 2$ Gaussian processes for the sources that are zero mean with variance $\delta^2\sigma_w^2$, and bandlimited as follows

$$f \in [100, 400]\,\text{Hz}, \tag{72}$$

where the bandwidth of the signals is 300 Hz. The interspacing of two adjacent sensors, Δ, can be determined as

$$\Delta = \frac{1}{2}\lambda_{min} = \frac{C}{800}. \tag{73}$$

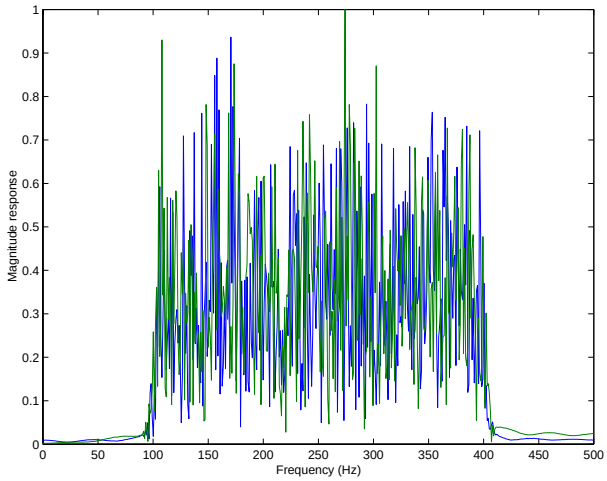

Fig. 1. The magnitude responses of the wideband signals that are bandlimited within 100 and 400 Hz.

The incident angles are -3.44 and 3.44 degrees, respectively, which are separated by an angle less than a standard half-beamwidth. Using the definition in (2), we can obtain the corresponding delay parameters as follows

$$\boldsymbol{\tau} = \frac{\Delta}{C} \sin \boldsymbol{\phi} = [-7.5, 7.5] \times 10^{-5}. \tag{74}$$

We can then generate $N_t = 30$ snapshots, according to the other parameters in Table 5.2. Figure 1 exhibits the magnitude spectrum of the generated wideband signals. The hyperparameter δ^2 is assumed known and is set to $\delta^2 = 25.12$.

The proposed algorithm randomly initializes all unknown parameters, and randomly assigns the initial model order k uniformly in $[1, k_{max}]$, where $k_{max} = M - 1 = 7$, is the maximum allowable model order. The number of MCMC iterations used in the algorithm is 10,000. The top portion of Fig. 2 exhibits the estimate of the number of sources for each iteration as the algorithm proceeds. The algorithm takes about 100 iterations to converge to the correct order. The bottom portion of Fig. 2 displays the histogram of the number of sources, from which the algorithm predicts the correct number of sources, $k = 2$.

Fig. 3 shows that histograms of the estimates. The algorithm takes about 1,000 iterations for a burn-in before the chain centres on the true delay values. It can be seen that the peaks of the histograms centre around the true value $\pm 7.5 \times 10^{-5}$. The MAP estimates of the ISDs in this particular simulation from these histograms are shown in Table 5.3. Fig. 4 shows a comparison between the true and reconstructed signal amplitudes $\boldsymbol{s}(n)$. It is clearly seen that given the values of other parameters $\{\boldsymbol{\tau}, k\}$, the signal amplitudes $\boldsymbol{s}(n)$ are separated and restored using the MAP estimation. Table 5.4 lists the mean-squared error of the restored amplitudes.

As seen in the simulation, the proposed method can perform joint detection and estimation for wideband signals in spatially correlated noise environment. Now we present comparisions with the theoretical Cramér-Roa lower bound (CRLB) derived in [17] with a slight modification from white Gaussian noise to colored noise, and goodness-of-fit and efficiency tests [5] [9].

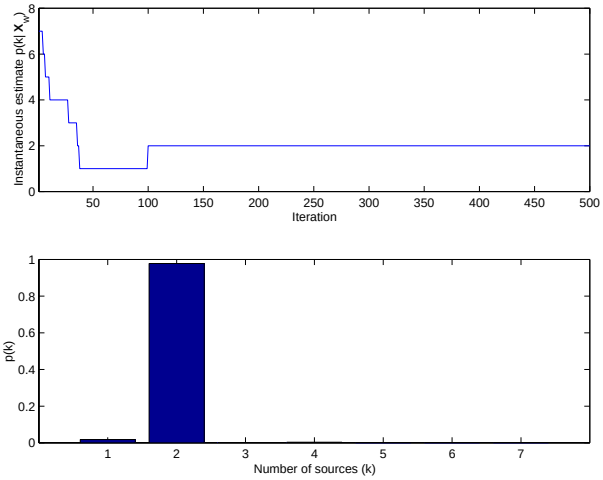

Fig. 2. (Top) Instantaneous estimate of the model order versus iteration index, for the first 500 iterations of the chain. (Bottom) Histogram of the number of sources after burn-in.

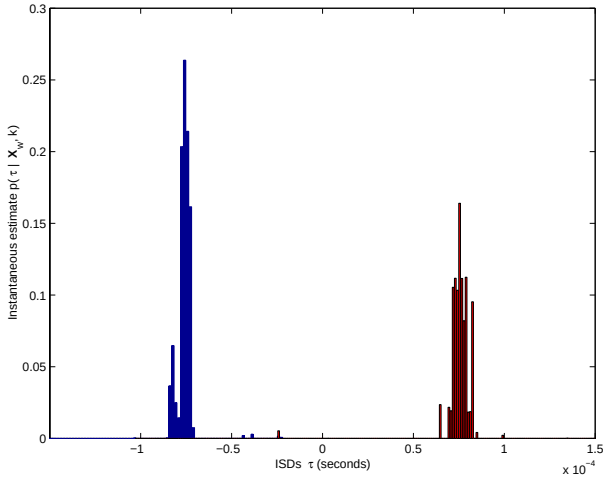

Fig. 3. Histogram of the delays of arrival of the sources after burn-in. The true values are $\pm 7.5 \times 10^{-5}$ seconds.

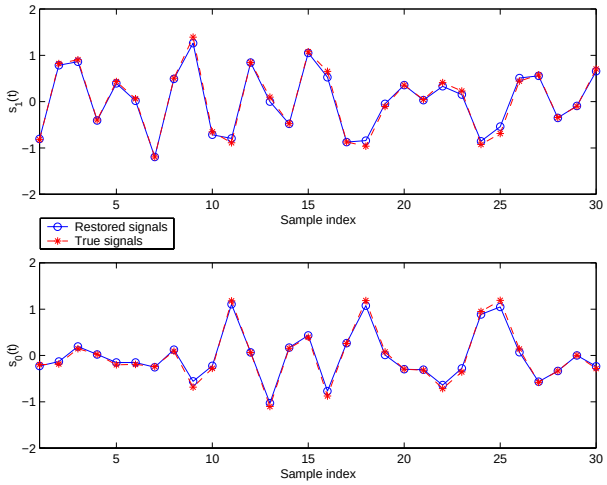

Fig. 4. A comparison between the true and restored signal amplitudes, $\boldsymbol{s}$(n).

True value (seconds)	MAP Estimate (seconds)	Relative Difference (%)
-7.5×10^{-5}	-7.23×10^{-5}	3.6
7.5×10^{-5}	7.93×10^{-5}	5.73

TABLE 5.3

A comparison between the true values and the estimates in one realization.

Source 1	Source 2
-26.19 dB	-25.97 dB

TABLE 5.4

The MSE of the restored source amplitudes relative to the true signal amplitudes for the experiment.

A. Comparison with the CRLB

To evaluate the performance of the proposed method, a total of 100 independent trials are run on the parameters listed in Table 5.5 for a range of SNR from -5dB to 18dB. In the evaluation, the model order is given. The estimate variances are computed and plotted against the CRLB.

The variances of the estimated $\boldsymbol{\tau}$ are plotted in Fig. 5 along with the respective theoretical CRLBs. As shown in Fig. 5, for SNR levels lower than -2dB, the algorithm starts to break down (i.e., departs rapidly from the CRLB. However, it is seen that the variances approach the CRLB closely, above this level. The reasons why the variances do not come closer to the theoretical CRLB are: 1) interpolation errors due to a non-ideal interpolation function being used and 2) the suboptimal procedure for estimating the source amplitudes. This procedure has an impact on the estimation accuracy of the DOA parameters.

B. Goodness-of-fit and efficiency tests

In addition, we use the goodness-of-fit and efficiency tests to evaluate the proposed method. Denote the normalized estimation error squared (NEES) for $\boldsymbol{\tau}$ by

$$\epsilon_{\boldsymbol{\tau}} \triangleq \tilde{\boldsymbol{\tau}}^T \boldsymbol{R}_{\tilde{\boldsymbol{\tau}}}^{-1} \tilde{\boldsymbol{\tau}}, \tag{75}$$

Parameter	M	K	L	N	$\lvert f\rvert/F_s$	$\boldsymbol{\tau}$ (sec)
Value	8	2	8	30	$[0.1, 0.4]$	$[5, 5] \times 10^{-4}$

TABLE 5.5

Parameters for the performance evaluation.

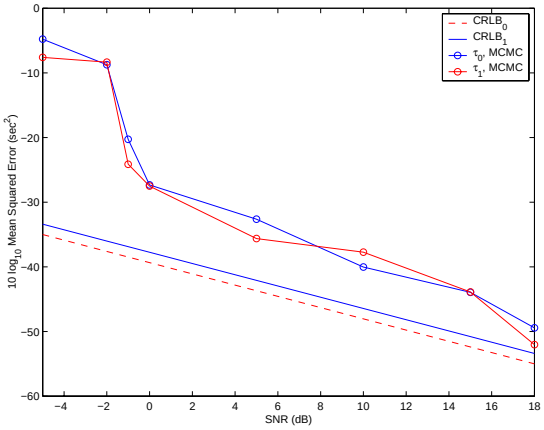

Fig. 5. Mean squared error of $\boldsymbol{\tau}$ versus the CRLB.

where

$$\tilde{\boldsymbol{\tau}} \triangleq \boldsymbol{\tau} - \hat{\boldsymbol{\tau}}, \tag{76}$$

$$\boldsymbol{R}_{\tilde{\boldsymbol{\tau}}} \triangleq E\left[\tilde{\boldsymbol{\tau}}\tilde{\boldsymbol{\tau}}^T\right]. \tag{77}$$

The quantity $\epsilon_{\boldsymbol{\tau}}$ is chi-squared distributed with K degrees of freedom, that is

$$\epsilon_{\boldsymbol{\tau}} \sim \chi^2_K. \tag{78}$$

Given a confidence level, say 95%, a set of *g-sigma ellipsoids* for $\epsilon_{\boldsymbol{\tau}}$ [9] can be obtained with different SNR values. These ellipsoids can be used to evaluate how well the MCMC method is performing in estimating $\boldsymbol{\tau}$. Fig. 6 shows the impact of different SNR values on the estimation of $\boldsymbol{\tau}$ with a 95% confidence interval. These figures are plotted in the same scale in order to reveal the impact of SNR levels on the estimation performance of the algorithm. It is clear that as SNR levels are larger than 10dB, the size of the ellipsoids is small and the estimates are clustered closely around the true value. However, as SNR is getting smaller and below 0dB, it is clear from Fig. 6 that the ellipsoid is getting bigger and the estimates distribute loosely around the center. Accordingly, we can conclude that the proposed method would start to break down in estimating $\boldsymbol{\tau}$ when SNR is below 0dB under the conditions used for this set of experiments.

In addition, the proposed method was tested to determine whether it is an efficient estimator. Using the 95% confidence level, the two-sided probability region for the NEES [9] for $\boldsymbol{\tau}$ is

$$[\epsilon_1, \epsilon_2] = [1.6273, 2.4106]. \tag{79}$$

Fig. 7 shows the NEES for $\epsilon_{\boldsymbol{\tau}}$ versus different SNR values. For SNR values above 0dB, the normalized errors fall inside the regions, that is, the method is consistent. However, as the SNR falls below 0dB, the errors are outside the bounds, indicating that the method starts with break down when the SNR is below 0dB. These findings are indeed consistent with those in Figs. 5 and 6.

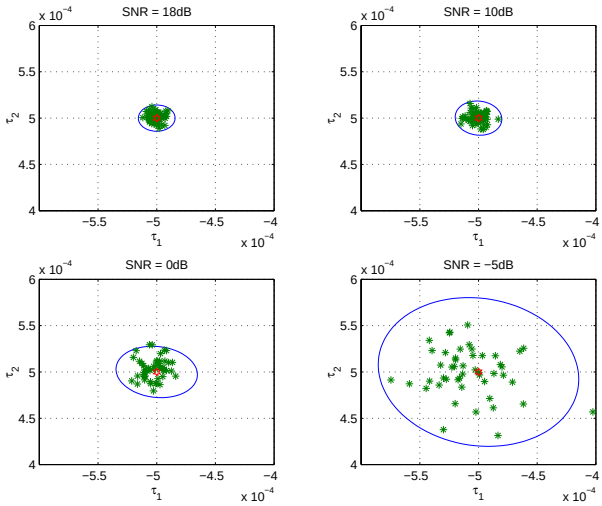

Fig. 6. The ellipsoids of the NEES with 95% confidence interval on MCMC with different SNR's: the triangle inside each ellipse corresponds the true value of $\boldsymbol{\tau}$, and the asterisks represent the distribution of the estimates of $\boldsymbol{\tau}$ of the 100 independent trials for different SNR levels.

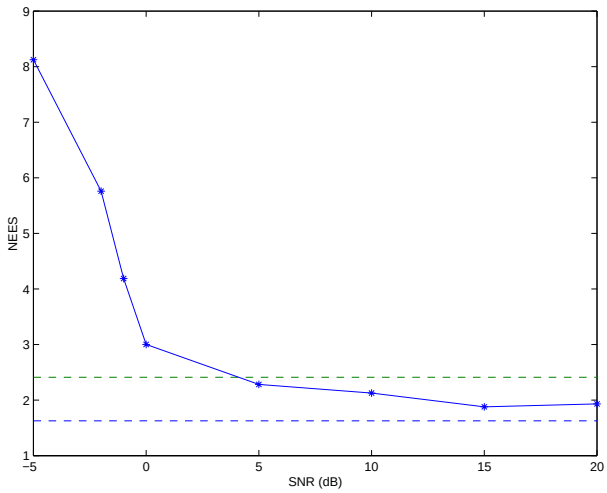

Fig. 7. Normalized estimation error squared from different SNR values with its 95% probability region.

C. Detection performance

Finally, we present the performance evaluation on the detection capability of the algorithm with different SNR levels and number of observations, as shown in Fig. 8. It can be seen that the detection probability is a linear function of the SNR levels for different numbers of obeservations when SNR levels are larger than 0dB. To obtain the probability of detection higher than 60%, the SNR levels need to be larger than 10dB, regardless of the number of observations. As the SNR levels are lower than 10dB, more observations are required to maintain a high detection probability. In addition, the probability of an error in detection of the model order tends to diminish toward zero with increasing number of snapshots, N_t, with moderate SNR values.

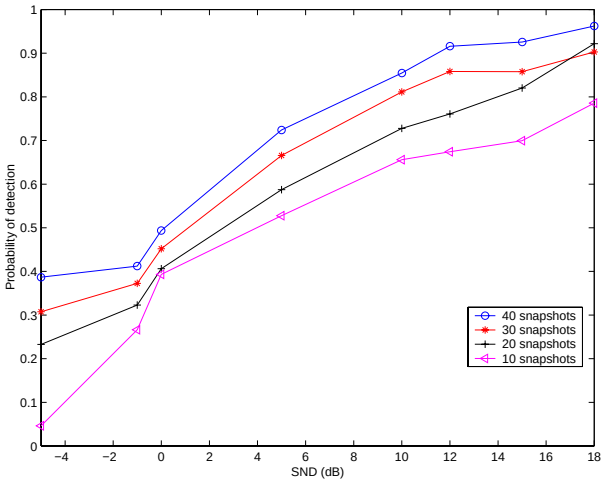

Fig. 8. Probability of detection as a function of number of snapshots for different SNR values.

VI. Conclusion

A novel structure for wideband array signal processing is proposed for joint detection and estimation in colored noise environment. A Bayesian approach is used, where a posterior density function which has the nuisance parameters, including the unknown, arbitrary colored noise covariance matrix, integrated out is formulated. The desired model order and ISD estimation parameters are determined through a reversible jump MCMC procedure. Simulation results support the effectiveness of the method, and demonstrate reliable detection of the number of sources and estimation of their inter-sensor delays with a single linear array, when the noise is spatially correlated in white noise environment.

References

[1] D. Johnson, "The application of spectral estimation to bearing estimation problem," *Proceedings of the IEEE*, vol. 70, pp. 1018–1028, Sept. 1982.

[2] B. D. V. Veen and K. M. Buckley, "Beamforming: A Versatile Approach to Spatial Filtering," in *IEEE ASSP MAGAZINE*, Apr. 1988.

[3] C. Andrieu and A. Doucet, "Joint Bayesian model selection and estimation of noisy sinusoids via reversible jump MCMC," *IEEE Transactions on Signal Processing*, vol. 47, pp. 2667–2676, Oct. 1999.

[4] Q. Wu and D. Fuhrmann, "A parametric method for determining the number of signals in narrow-band direction finding," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 39, pp. 1848–1857, Aug. 1991.

[5] W. Ng, J. P. Reilly, T. Kirubarajan, and J.-R. Larocque, "Wideband array signal processing using MCMC methods," 2002. Submitted to IEEE Transactions on Signal Processing, Jan. 2003, also available at http://www.ece.mcmaster.ca/∼reilly.

[6] C. Andrieu, A. Doucet, W. Fitzgerald, and S. Godsill, "An introduction to the theory and applications of simulation based computational methods in Bayesian signal processing," in *Tutorial from Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, (Seattle, WA), 1998.

[7] C. Andrieu, P. Djuric, and A. Doucet, "Model selection by MCMC computation," *Signal Processing*, vol. 81, pp. 19–37, Jan. 2001.

[8] W. Gilks, S. Richardson, and D. Spiegelhalter, *Markov Chain Monte Carlo in Practice*. New York: Chapman and Hall, 1998.

[9] X.-R. Li, Y. Bar-Shalom, and T. Kirubarajan, *Estimation, Tracking and Navigation: Theory, Algorithms and Software*. New York: John Wiley & Sons, June, 2001.

[10] P. Green, "Reversible jump Markov Chain Monte Carlo computation and Bayesian model determination," in *Biometrika*, vol. 82, pp. 711–732, 1995.

[11] G. H. Golub and C. F. V. Loan, *MATRIX Computations, 2nd Edition*. Baltimore, Maryland: The Johns Hopkins University Press, 1993.

[12] J.-R. Larocque and J. P. Reilly, "Reversible jump MCMC for joint detection and estimation of directions of arrival in coloured noise," *IEEE Transactions on Signal Processing*, vol. 50, pp. 231–240, Feb. 2002.

[13] H. Jeffreys, *Theory of Probability, 3rd ed*. London, U.K.: Oxford University Press, 1961.

[14] J. Bernardo and A. Smith, *Bayesian Theory*. New York: John Wiley and Sons, 1994.

[15] W. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," in *Biometrika*, vol. 57, pp. 97–109, 1970.

[16] S. Godsill, "On the relationship between MCMC model uncertainty methods," *J. Comp. Graph. Stats.*, vol. 10(2), 2001.

[17] W. Ng, J. P. Reilly, and T. Kirubarajan, "The derivation of the theoretical CRLB for array signal processing for wideband signal," May 2002. Please download the document at http://www.ece.mcmaster.ca/∼reilly.