ADVANCED BAYESIAN METHODS FOR ARRAY SIGNAL PROCESSING

BY

JEAN-RENÉ LAROCQUE

SEPTEMBER 2001

A THESIS

SUBMITTED TO THE DEPARTMENT OF ELECTRICAL & COMPUTER ENGINEERING

AND THE SCHOOL OF GRADUATE STUDIES

OF MCMASTER UNIVERSITY

IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Advanced Bayesian Methods for Array Signal Processing

Doctor of Philosophy (2001) McMaster University

(Electrical & Computer Engineering) Hamilton, Ontario

TITLE: Advanced Bayesian Methods for Array Signal Processing

AUTHOR: Jean-René Larocque

B.Sc.A., M.Sc. (Electrical Engineering)

Université Laval, Québec, Canada

SUPERVISOR: Dr. James P. Reilly

NUMBER OF PAGES: xv, 128

To my most wonderful wife, Kati

Abstract

This thesis focuses on the joint detection of the model order and the estimation of the parameters of interest, with applications to array signal processing in both the off-line and on-line contexts. In the off-line mode, Markov Chain Monte Carlo methods are applied to obtain a numerical approximation of the joint posterior distribution of the parameters. The on-line approach uses the sequential implementation of Monte Carlo methods applied to probabilistic dynamic systems.

Three problems were addressed in the course of this thesis.

1. A method for joint detection of the number of sources and estimation of their respective directions of arrival in coloured noise with unknown arbitrary covariance was developed.

2. The second algorithm represents an extension of the first one with the addition of the joint estimation of the times of arrival of the pulses, in the spirit of channel sounding for characterization of multipath channels.

   Both methods were successfully applied to real data, acquired on campus with a channel sounder during an extensive measurement campaign.

3. The final part of this thesis focuses on the sequential implementation of the Monte Carlo methods, i.e. particle filters, for probabilistic dynamic systems.

   The algorithm recursively estimates the posterior distribution of the evolving parameters of interest, allowing for the on-line detection of the number of sources and the estimation of their respective directions of arrival.

Acknowledgements

The last four years have been both very challenging and very enjoyable; I had the opportunity to work on a very interesting topic about which I had no prior knowledge, but for which the likelihood of success was quite well defined, making for a very promising posterior situation.

I would like to first thank Dr. Jim Reilly for supervising and supporting my work. His enthusiasm, approachability and patience were essential to the completion of this project. I am also very grateful to Dr. Tim Davidson and Dr. Christophe Andrieu (from Cambridge, now at Bristol) for their most insightful comments and discussions.

The efforts and assistance of my fellow ECE graduate students is also recognized. Kaywan, Matt, Kamran, Sabrina, William - Thank you!

Further thanks are also due to the ECE department and staff (in particular Cheryl Gies who solved all of my non-research related problems - what would this department do without you?).

My parents and my in-laws - the four of you are a great source of inspiration!

Last but not least, I would like to thank my lovely wife Kati for her uninterrupted love and encouragement which played an instrumental role in me completing this project.

To all you involved in this chapter of my life - THANK YOU!

The research reported in this thesis has been made possible by the financial support from the National Sciences and Engineering Research Council of Canada, the Canadian Wireless Telecommunications Association, the ECE Department of McMaster University, CITO (Communications and Information Technology Ontario) and Mitel Corporation.

Notation and Acronyms

Symbol	Definition
x	Scalar
$\boldsymbol{x}$	Vector
$\boldsymbol{Y}$	Matrix
$\boldsymbol{Y}^T$	Matrix transpose
$\boldsymbol{Y}'$	Hermitian transpose
$\lvert\boldsymbol{Y}\rvert$	Determinant
$\mathrm{Tr}(\boldsymbol{Y})$	Trace
Array Processing	
k	Model order (Number of sources)
N_t	Number of snapshots (observations)
$\boldsymbol{\phi}$	Physical directions of arrival
$\boldsymbol{S}(\boldsymbol{\phi})$	Steering matrix
$\boldsymbol{s}(\boldsymbol{\phi})$	One column of the steering matrix
ω_0	Carrier frequency in rad/s
f	Carrier frequency in Hz
d	Spacing (for a uniform linear array)
R	Radius (for a circular array)
v	Velocity of propagation
$\boldsymbol{R}_{yy}$	Covariance matrix of the observations $\boldsymbol{y}$
λ	Eigenvalue

e_l	Eigenvector corresponding to the lth eigenvalue
$\mathcal{N}$	Noise subspace
$\mathcal{S}$	Signal subspace
Statistics	
N	Number of samples (particles)
$P[A]$	Probability of the event A
$p(x)$ or $p(\boldsymbol{x})$	Probability density function (pdf) of x or $\boldsymbol{x}$
$\boldsymbol{x} \sim p(\boldsymbol{x})$	$\boldsymbol{x}$ is distributed or drawn from $p(\boldsymbol{x})$
$p(\boldsymbol{x}\|\mu;\sigma^2)$	Density of $\boldsymbol{x}$ conditional on μ,
	with the knowledge of the parameter σ^2
$\pi(\boldsymbol{x})$	Posterior density of interest $p(\boldsymbol{x}\|\boldsymbol{y})$
$q(\cdot)$	Importance (or candidate) function
r	Acceptance ratio
α	Acceptance probability
$w_t^{(i)}$	The weight of the (i)th particle at time t
Acronyms	
BIS	Bayesian Importance Sampling
iid	independant and identically distributed
IS	Importance Sampling
MC	Monte Carlo
MCMC	Markov Chain Monte Carlo
pdf	Probability Density Function
SIS	Sequential Importance Sampling
CRLB	Cramér-Rao Lower Bound
SNR	Signal-to-Noise Ratio
DOA	Direction of Arrival
TOA	Relative Time of Arrival

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Statistical methods for signal processing have a wide range of different applications, such as radar, sonar, wireless communications, telephony, geophysics and many more. The objectives are however the same, namely the extraction of parameters of interest from noisy observations.

In array signal processing, there are a number of possible objectives, significant ones being,

- *Detection of number of sources.* Most classical methods assume knowledge of this parameter, or assume that it was previously estimated. The knowledge of this parameter is often instrumental to methods focusing on the remaining objectives, below.

- *Source localization.* The azimuthal locations of the sources, i.e. the direction of arrival (DOAs) of the propagating waves relative to the axis of the array, are the parameters of interest. It is also possible to estimate other parameters, such as the range or the velocity of the sources.

- *Channel characterization.* The objective is to estimate the space-time manifold describing the propagation between the source and the receiver. The directions of arrival and the relative delays of propagation characterize the channel.

Depending on the application, the processing might be "batch mode" or "off-line",

when the data has been collected before processing, or "sequential" or "adaptive", when the algorithm proceeds as the observations are collected. There are many classical methods of either type, addressing each of the objectives.

1.1 Array signal processing

In an array signal processing scenario, we have N_t vector observations $\boldsymbol{y}(n) \in \mathcal{C}^M, \quad n = 1, \ldots, N_t$ from an array of M sensors, which is illuminated by k_o plane waves incident onto the array from angles $\phi_1, \ldots, \phi_{k_o}$ relative to the normal of the array. The objective is, given the observations, to estimate the parameter k_o and the corresponding directions of arrival.

1.1.1 Array geometries

In the context of array signal processing, several assumptions are necessary to support the development of the algorithms.

- *Plane waves.* The sources are assumed far enough from the array so that the incident signals have planar wave fronts.
- *Narrow band.* Under this assumption, the propagation over the length of the array is a function of only the phase term, as a simple delay factor.
- *Calibration.* The sensors are assumed calibrated, namely their radiation pattern is assumed known (Tranter *et al.*, 1999). That is, the array steering (response) vector $\boldsymbol{s}(\phi)$ is known for all ϕ within the field of view.
- *Uncorrelated noise.* The noise samples are assumed uncorrelated with the signals.
- $k_o < M$. For the covariance matrix of the observations to be full column rank, the true number of sources impinging the array must be less than the number of elements composing the array.

The structure of the steering matrix varies with the geometry of the array. The specific geometry of the array, e.g. whether the array is linear or circular, depends on the application. In the case of a linear array composed of M equally spaced sensors (see Figure 1.1), the manifold in the presence of k_o incident signals, is described by, $S(\phi) \in \mathcal{C}^{M \times k_o}$, the kth column of which is

$$s(\phi_k) = [1, e^{j\omega_0 d \sin(\phi_k)/v}, \ldots, e^{j(M-1)\omega_0 d \sin(\phi_k)/v}]^T \quad k = 1, \ldots, k_o, \tag{1.1}$$

and $\phi \in \mathcal{R}^{k_o}$. In particular, when the array elements are uniformly spaced, we assume

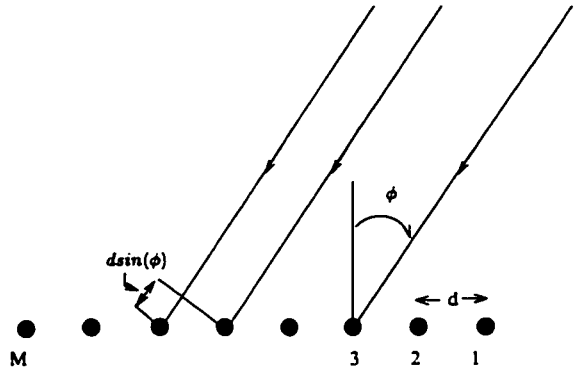

Figure 1.1: Definition of symbols for a linear antenna array scenario.

$d \leq \frac{\pi v}{\omega_0}$, where d is the distance between the sensors, ω_0 is the operating frequency, and v is the velocity of propagation. This structure is commonly used in the literature for its simplicity, but it suffers from an ambiguity problem, as one cannot distinguish whether the wave is impinging from the front or from the back of the array: two different plane waves impinging from mirror directions would give the same steering matrix.

For this reason, the circular array is most often used in practice and this thesis therefore adopts this geometry. The steering matrix of a circular array of radius R and made of M

sensors is defined as (with the angles shown in Figure 1.2)

$$\mathbf{S}(\phi) = \begin{bmatrix} e^{-j\kappa\cos(\phi_1-\gamma_1)} & e^{-j\kappa\cos(\phi_2-\gamma_1)} & \dots & e^{-j\kappa\cos(\phi_{k_o}-\gamma_1)} \\ e^{-j\kappa\cos(\phi_1-\gamma_2)} & e^{-j\kappa\cos(\phi_2-\gamma_2)} & \dots & e^{-j\kappa\cos(\phi_{k_o}-\gamma_2)} \\ \vdots & & & \vdots \\ e^{-j\kappa\cos(\phi_1-\gamma_M)} & e^{-j\kappa\cos(\phi_2-\gamma_M)} & \dots & e^{-j\kappa\cos(\phi_{k_o}-\gamma_M)} \end{bmatrix}, \tag{1.2}$$

where $\kappa = \frac{\omega_0 R}{v}$, and R the radius of the circular array.

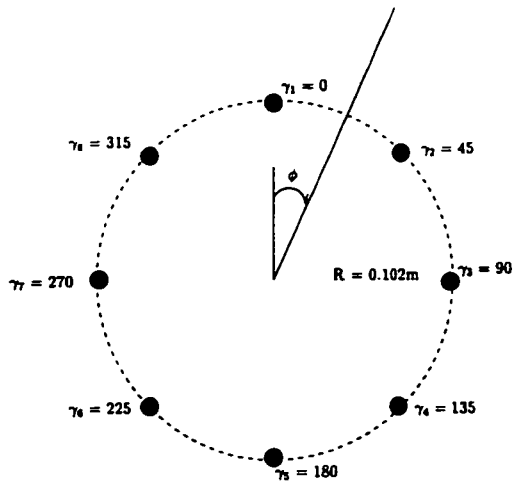

Figure 1.2: Definition of symbols for an 8 element circular antenna array scenario.

The vector $\boldsymbol{y} \in \mathcal{C}^M$ of observed data is the output of an array of M sensors. The output of the array at the sampling time n, i.e. at the nth *snapshot*, is expressed in term of its steering matrix $\boldsymbol{S}(\boldsymbol{\phi})$

$$\boldsymbol{y}(n) = \boldsymbol{S}(\boldsymbol{\phi})\boldsymbol{a}(n) + \boldsymbol{\nu}(n) \quad n = 1, 2, \dots, N_t, \tag{1.3}$$

where $\boldsymbol{\nu}(n)$ is the observation noise.

1.1.2 Traditional array processing

Traditional array processing methods for model selection and direction finding purposes often use the information contained in the covariance matrix of the data

$$E\{\boldsymbol{y}(n)\boldsymbol{y}'(n)\} = E\{(\boldsymbol{S}(\boldsymbol{\phi})\boldsymbol{a}(n) + \boldsymbol{\nu}(n))(\boldsymbol{S}(\boldsymbol{\phi})\boldsymbol{a}(n) + \boldsymbol{\nu}(n))'\},$$
$$\boldsymbol{R}_{yy} = \boldsymbol{S}\boldsymbol{\Phi}_a\boldsymbol{S}' + \boldsymbol{\Sigma}_\nu, \tag{1.4}$$

where $\boldsymbol{\Sigma}_\nu$ is the spatial covariance of the noise process and $\boldsymbol{\Phi}_a$ is the signal covariance matrix. The expression (1.4) is obtained with the assumption that the signal and the noise are uncorrelated. The rank of the matrix $\boldsymbol{S}\boldsymbol{\Phi}_a\boldsymbol{S}'$ in eq. (1.4) is the number of sources. However, this matrix is only available through the estimate of $\boldsymbol{R}_{yy}$, given as

$$\hat{\boldsymbol{R}}_{yy} = \frac{1}{N_t}\sum_{i=1}^{N_t}\boldsymbol{y}(n)\boldsymbol{y}'(n). \tag{1.5}$$

The signal subspace and the complementary noise subspace can be estimated with an eigen decomposition of the data covariance matrix. Let the set $[\hat{\boldsymbol{e}}_1, \ldots, \hat{\boldsymbol{e}}_M]$ be the eigenvectors of the data covariance matrix (1.5). The estimated signal subspace $\hat{\mathcal{S}}$ and the estimated noise subspace $\hat{\mathcal{N}}$ can be defined as

$$\hat{\mathcal{S}} \approx \text{span}[\hat{\boldsymbol{E}}_S] = \text{span}[\hat{\boldsymbol{e}}_1, \ldots, \hat{\boldsymbol{e}}_{k_o}],$$
$$\hat{\mathcal{N}} \approx \text{span}[\hat{\boldsymbol{E}}_N] = \text{span}[\hat{\boldsymbol{e}}_{k_o+1}, \ldots, \hat{\boldsymbol{e}}_M].$$

The above quantities with a hat denote those derived from the estimated covariance $\hat{\boldsymbol{R}}_{yy}$.

1.1.2.1 Direction finding

Subspace methods With the knowledge of the number of sources, the signal subspace and noise subspace can be correctly estimated with the corresponding eigenvectors. In the event that the noise is not white, generalized eigenvectors are required to describe the subspaces. It is possible to estimate the directions of arrival by projecting the steering vectors onto either space. If the signal vector is in the signal DOA's subspace, the magnitude

of the projection onto $\hat{\mathcal{S}}$ should be relatively large. Correspondingly, a projection onto $\hat{\mathcal{N}}$ will result in a small magnitude.

Denoting $s(\phi)$ to be a column of the steering matrix and Θ the set of the true directions of arrival,

$$||s(\phi)' * \hat{E}_N|| \to 0 \quad \text{when } \phi \in \Theta.$$

One method which uses this principle to estimate DOAs is the Pisarenko pseudo spectrum (Pisarenko, 1973), that projects a steering vector onto a single noise eigenvector. The Pisarenko pseudo spectrum is defined as

$$\hat{P}_{Pis}(\phi) = \frac{1}{s'(\phi)\hat{e}_M\hat{e}'_M s(\phi)}.$$

A more commonly used method, the MUSIC algorithm (Schmidt, 1986) involves a projection of a steering vector onto the whole noise subspace. The MUSIC pseudo spectrum is defined as

$$\hat{P}_{MU}(\phi) = \frac{1}{s'(\phi)\hat{E}_N\hat{E}'_N s(\phi)},$$

where $\hat{E}_N = [\hat{e}_{k_o+1}, \ldots, \hat{e}_M]$. The MUSIC method will perform well when the number of sources is over estimated. In the worst case, when the estimated numbers of sources is $M-1$, it degenerates to the Pisarenko pseudo spectrum. However, in this case, it is difficult to localize the extra artificial directions of arrival.

These pseudo spectra show high peaks at the directions of arrival that are represented in the data considered. The term 'pseudo spectrum' reflects the fact that the height of the peaks does not provide any information on the amplitudes of the signals.

Maximum Likelihood methods For a typical signal processing problem, a model can be assumed to describe the observation data as a function of some parameters. This signal model, along with the noise model, defines the likelihood function of the data, given the parameters. For example, using the data model described in eq. (1.3), the likelihood

function, assuming Gaussian noise, is [1]

$$p(\boldsymbol{y}(n)|\boldsymbol{\phi}) = l(\boldsymbol{\phi};\boldsymbol{y}(n)) \sim N_{\boldsymbol{y}}(\boldsymbol{S}(\boldsymbol{\phi})\boldsymbol{a}(n), \boldsymbol{\Sigma}_{\nu}).$$

This is a function of the parameter $\boldsymbol{\phi}$ and not of $\boldsymbol{y}(n)$, thus the notation $l(\boldsymbol{\phi};\boldsymbol{y}(n))$. The value of the parameter that maximizes this function is called the *maximum-likelihood* estimate.

$$\hat{\boldsymbol{\phi}}_{ML} = \arg\max_{\boldsymbol{\phi}\in\boldsymbol{\Phi}} p(\boldsymbol{y}|\boldsymbol{\phi}). \tag{1.6}$$

From the model of the observation in eq. (1.3), the least-squares estimate of the amplitudes is

$$\hat{\boldsymbol{a}}(n) = (\boldsymbol{S}'(\boldsymbol{\phi})\boldsymbol{S}(\boldsymbol{\phi}))^{-1}\boldsymbol{S}'(\boldsymbol{\phi})\boldsymbol{y}(n),$$

which, after substituting back into eq. (1.3), gives

$$\begin{aligned}\hat{\boldsymbol{y}}(n) &= \boldsymbol{S}(\boldsymbol{\phi})(\boldsymbol{S}'(\boldsymbol{\phi})\boldsymbol{S}(\boldsymbol{\phi}))^{-1}\boldsymbol{S}'(\boldsymbol{\phi})\boldsymbol{y}(n) + \boldsymbol{\nu}(n),\\ \hat{\boldsymbol{y}}(n) &= \boldsymbol{P}_{\boldsymbol{S}}(\boldsymbol{\phi})\boldsymbol{y}(n) + \boldsymbol{\nu}(n),\end{aligned}$$

with $\boldsymbol{P}_{\boldsymbol{S}}(\boldsymbol{\phi}) = \boldsymbol{S}(\boldsymbol{\phi})(\boldsymbol{S}'(\boldsymbol{\phi})\boldsymbol{S}(\boldsymbol{\phi}))^{-1}\boldsymbol{S}'(\boldsymbol{\phi})$.

The optimization problem of eq. (1.6) can be written as

$$\begin{aligned}\hat{\boldsymbol{\phi}}_{ML} &= \arg\min_{\boldsymbol{\phi}\in\boldsymbol{\Phi}} \sum_{n} ||(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{S}}(\boldsymbol{\phi})\boldsymbol{y}(n)||,\\ &= \arg\min_{\boldsymbol{\phi}\in\boldsymbol{\Phi}} \mathrm{Tr}(\boldsymbol{P}_{\boldsymbol{S}}^{\perp}(\boldsymbol{\phi})\hat{\boldsymbol{R}}_{\boldsymbol{yy}}),\\ &= \arg\max_{\boldsymbol{\phi}\in\boldsymbol{\Phi}} \mathrm{Tr}(\boldsymbol{P}_{\boldsymbol{S}}(\boldsymbol{\phi})\hat{\boldsymbol{R}}_{\boldsymbol{yy}}),\end{aligned}$$

where $\mathrm{Tr}(\cdot)$ is the trace operator, $\hat{\boldsymbol{R}}_{\boldsymbol{yy}}$ is defined in eq. (1.5), and $\boldsymbol{P}_{\boldsymbol{S}}(\boldsymbol{\phi})$ is a model dependent projector matrix. The same result can be obtained in a Bayesian context (Reilly, 1981; Haykin, 1982; Wu and Wong, 1994; Viberg *et al.*, 1991).

Using these approaches to estimate parameters has serious limitations. This optimization problem is well known to be difficult, as the function to optimize shows many saddle

[1] The definitions of the density functions can be found in appendix A.

points and local extrema. Any gradient based method would need a good initialization in order to succeed. Also, the likelihood function does not make use of any information other than the observed data itself. Any knowledge that might have been acquired during previous experiments is not utilized.

Secondly, the maximization of the likelihood function jointly with the model order, i.e. the number of incident signals, does not limit the number of parameters involved in modeling the data. Therefore, a more complex model would result in a higher likelihood. This limitation is well recognized and results in difficulty determining the model order k_o. Many attempts have been made to resolve this deficiency. Methods such as AIC and MDL (see section 1.1.2.2) use information theory - the result is an extra term that penalizes models of higher complexity.

1.1.2.2 Model selection

In the case where the noise is spatially white, the noise covariance matrix is given as

$$\boldsymbol{\Sigma}_{\boldsymbol{w}} = \sigma_w^2 \boldsymbol{I}_M.$$

Then, in the eigen decomposition of $\boldsymbol{R}_{\boldsymbol{yy}}$, there are exactly $M - k_o$ eigenvalues that are equal to σ_w^2 while the remainder are all larger, i.e.

$$\lambda_1 > \lambda_2 > \cdots > \lambda_{k_o} > \lambda_{k_o+1} = \lambda_{k_o+2} = \cdots = \lambda_M.$$

Unfortunately, depending on the number of snapshots used to estimate the covariance matrix, the estimated eigenvalues $\hat{\lambda}$ and eigenvectors $\hat{e}$ will show some variance, which will complicate the count of significant eigenvalues, especially at low signal-to-noise ratios.

The model order can be estimated using information theoretic criteria, such as the Akaike information criteria (AIC) (Akaike, 1974), the minimum description length criteria (MDL) (Rissanen, 1978) or the more recent D-MAP (Djuric, 1996). These approaches maximize the log-likelihood function of the data over all the model orders available, but with increasing penalties for higher dimension models. The AIC approach is not consistent, i.e. there remains a finite probability of error at high SNR levels and/or when the number of

snapshots approaches infinity. The other two approaches are known to be consistent (Zhao *et al.*, 1987).

These methods are based on maximizing the following functional form,

$$L(\hat{\phi}) = -\log f(\boldsymbol{y}|\hat{\phi}) + g(k, M, N_t),$$

where $\hat{\phi}$ is the maximum-likelihood estimate of the parameters for a specific model order, and $g(k, M, N_t)$ is a penalty term, which is a function of the model order, and other parameters. The AIC, MDL and D-MAP criteria for estimation of the number of sources can be written respectively as

$$\hat{k}_{AIC} = \min_k -\log f(\boldsymbol{y}|\hat{\phi}) + 2k(2M - k),$$

$$\hat{k}_{MDL} = \min_k -\log f(\boldsymbol{y}|\hat{\phi}) + \frac{k(2M-k)}{2}\log(N_t),$$

$$\hat{k}_{DMAP} = \min_k -\log f(\boldsymbol{y}|\hat{\phi}) + \frac{5k}{2}\log(N_t),$$

where N_t is the number of snapshots used to estimated the covariance matrix, M is the number of elements in the array and k is the model order considered. The maximum-likelihood estimate for a model of order k can be written in terms of eigenvalues (Wax and Kailath, 1985), and this term is common to all three methods,

$$f(\boldsymbol{y}|\hat{\phi}) = \left(\frac{(\prod_{i=k+1}^{M} \hat{\lambda}_i)^{1/(M-k)}}{\frac{1}{M-k}\sum_{i=k+1}^{M} \hat{\lambda}_i} \right)^{2(M-k)N_t},$$

where the eigenvalues $\hat{\lambda}_i$ are arranged in descending order.

These criteria suffer from severe limitations due to the assumptions supporting their development. The most limiting assumption is the noise, which is assumed additive, white and Gaussian. Experiments have shown that these criteria are very sensitive to the white noise assumption. A performance analysis of the criteria for varying level of colours can be found in (Chen, 1991).

There are a few methods that address this problem in coloured noise. If the noise characteristics are known, various pre-whitening approaches might be used. Another alternative

is the use of multiple arrays (Wong *et al.*, 1992; Fuchs, 1992; Wu and Wong, 1994; Wu and Fuhrmann, 1991; Nagesha and Kay, 1996; Chen *et al.*, 1996).

1.2 The Bayesian approach

The limitations and deficiencies of the maximum-likelihood method are very elegantly addressed in the Bayesian framework, by the involvement of prior distributions. This gives analysts the ability to use prior knowledge, and the option of marginalizing with respect to the parameters of interest, for evaluation of confidence regions, etc.

Until recently, Bayesian methods, which rely on the use of probabilistic models and the application of probability theory, were ignored mainly for three reasons. Firstly, these methods demand a very high implementation complexity. Most of the scenarios involve optimization or integration that cannot be carried out analytically. Secondly, numerical approximation of these optimizations or integrations often leads to discouragingly computationally intensive processing. Finally, the Bayesian methods make heavy use of the Bayes' theorem, which involves the use of prior distributions. This subject has been a constant source of controversy in the signal processing community.

Although the supporting theory is quite simple, the resulting posterior distributions to manipulate are often highly non-linear and complex, preventing any analytical analysis or manipulations. To solve any optimization or integration problems, numerical methods must be employed.

Markov Chain Monte Carlo (MCMC) batch methods were first proposed in the early fifties, but were ignored due to their intense computational requirements. With the ever growing power of personal computers, they have just recently matured into very powerful algorithms, providing a potential solution to difficult problems, when no other solution would exist. The whole MCMC theory could be summarized in one sentence as:

The idea of MCMC methods is to draw samples distributed according to an arbitrary probability density function by observing the states of a Markov chain, with the objective of computing a histogram that will approximate the probability density function, allowing

numerical inference on parameters.

When the parameters of interest evolve in time, the time-series of their posterior distributions constitute a probabilistic dynamic system. When new observations become available, the objective is to recursively update the weights of a set of samples to describe the updated posterior distribution. This approach is the basis for *particle filters.*

The next chapter will focus on these topics and will present detailed explanations of the different numerical methods. It is sufficient at this point to say that, when a large number of samples drawn from a distribution are available, then three types of numerically-based inference can be made using this approach.

- *Optimization.* The objective is to locate the extrema of the function. A complicated optimization procedure is replaced by a simple location searching approach.
- *Integration.* Marginalizing parameters would require complicated non-linear and most often prohibitive analytical integrations. Numerical marginalization of a parameter is simplified to summing bins of the histogram.
- *Simulation.* In some applications, the objective is simply to draw samples from an arbitrary density function, which is only possible in general when this function is either standard or integrable. MCMC methods proved powerful in this context.

1.2.1 Bayes' theorem

Let y be a random variable (e.g an observation) whose conditional probability density function $p(y|\phi)$ depends on the parameter ϕ. Then the function $p(y|\phi)$ defines the *likelihood* function.

Any information about the parameter can be included in a *prior* distribution $p(\phi)$. Then the Bayes theorem states that

$$p(y|\phi)p(\phi) = p(\phi, y) = p(\phi|y)p(y).$$

Given the observation y, the conditional distribution of the parameter ϕ defines the *posterior*

distribution

$$p(\phi|y) = \frac{p(y|\phi)p(\phi)}{p(y)} \propto p(y|\phi)p(\phi).$$

This is Bayes' theorem.

The total probability of the data, $p(y)$, for two different types of probability distribution, is

$$p(y) = \begin{cases} \int p(y|\phi)p(\phi)d\phi & \text{Continuous,} \\ \sum p(y|\phi)p(\phi) & \text{Discrete.} \end{cases}$$

This distribution only plays the role of a normalization constant and is often not of interest in parameter estimation, as here we are only interested in the location of the extrema, not their actual values. Furthermore, for most realistic models, this expression cannot be evaluated analytically, limiting the specification of the posterior distribution up to that constant.

The value of the parameter that maximizes the posterior distribution is referred to as the *Maximum A Posteriori* estimate (MAP). Statistical inferences can now be obtained from the posterior distribution, which was not possible in the maximum-likelihood framework.

It should be noted that the maximum-likelihood approach and the MAP are not mutually exclusive. If a mean-square cost function is selected to estimate a parameter in the maximum-likelihood sense, the same result would be obtained by calculating the mean of the posterior distribution with uniform prior (the use of a linear (absolute value) cost function is similar to finding the median of the posterior distribution; a uniform cost function, with a notch, the MAP estimate) (Van Trees, 1968; Kay, 1993).

1.2.2 Prior distributions

When observations from previous experiments are available, the posterior distribution of the parameters for these experiments could be used as a prior distribution for the current experiment. For example, some physical characteristics of the system might preclude some values of the parameters.

The notion of prior distributions has been a point of debate for decades between rigorous statisticians adhering to the classical philosophy, where the maximum-likelihood estimates are obtained with more or less well defined risk functions (e.g.mean-square or uniform), and the Bayesian analysts, that use prior distributions to include prior knowledge.

This debate seems to be fading away, now that the impact of the prior distributions on the end results is better understood. Also, when prior distributions are used, more inference can be obtained from the posterior distribution than from the maximum-likelihood estimate. For example, evaluation of confidence intervals for the MAP estimate, and marginalization of the posterior distribution for a parameter of interest are only possible in the Bayesian context.

1.2.2.1 Non-informative prior

When no information is available, a non-informative prior distribution should be used. A distribution is non-informative if it can be approximated by a constant over the range where the likelihood function is important, so that it does not contribute discrimination to the likelihood function. In other words, the prior distribution must be wider than the likelihood function, to "let the data talk" (Box and Tiao, 1973). Estimates obtained using non-informative prior distributions are identical to those obtained from maximizing the likelihood function.

1.2.2.2 Proper prior

As critics often argue, for these quantities (e.g. MAP estimates, confidence intervals) to exist, the prior distributions must be proper, i.e. summable. In some cases, improper prior distributions might be acceptable, such as a uniform prior over an infinite range. However, the application of MCMC methods to be considered later is more sensitive and requires a proper prior distribution.

1.2.2.3 Conjugate prior

To improve analytical tractability, e.g. regarding the analytical marginalization of undesired parameters from the posterior distribution, prior distributions are often chosen for their analytical forms, so that the posterior distribution follows the same distribution as the likelihood function, but with different moments. This is possible when the prior distribution is the *conjugate* prior. For example, let's assume that $p(y|\sigma^2;\mu)$ is a Gaussian distribution with known first moment. The conjugate prior for the variance is the Inverted Gamma distribution (see the table in appendix A for their expressions). Bernardo and Smith (1994) present an extensive table for most distributions.

1.2.2.4 Jeffrey's prior

Another form of prior distribution is derived using Jeffrey's rule (Jeffreys, 1961, original paper, 1939). In this case, the prior distribution for a set of parameters is taken to be proportional to the square root of the determinant of the Fisher information matrix (Box and Tiao, 1973). This approach proposes a prior distribution that is *data translated*, and is given as

$$p(\phi) \propto |\mathcal{J}(\phi)|^{1/2},$$

where

$$\mathcal{J}(\phi) = -\mathrm{E}_{y|\phi}\left[\frac{\partial^2 \log p(y|\phi)}{\partial \phi^2}\right].$$

This is the general form of the non-informative prior distribution.

1.2.3 Ockham's razor

When the model order is to be jointly estimated along with other parameters, a simpler model must be selected, as a compromise of complexity and data fit. This principle is known as Ockham's razor.

In the Bayesian framework, this principle favours models of lower dimensions and models of higher complexity are penalized because of the influence of the prior distribution on the

model. As model order increases, the heights of the prior distributions decrease faster, thus penalizing higher-order models.

In a sense, the effect of Ockham's razor is similar to the behaviour of an information theoretic criterion (such as AIC or MDL). However, the use of different forms of prior distributions results in different types of penalty function. It has been experimentally shown that criteria obtained from the Ockham's razor principle often perform better (Pope, 1993).

This is the supporting principle explaining the convergence of the MCMC methods toward the correct model order, as we discuss in later chapters.

1.3 Campaign for experimental measurement of multipath channels

During the fall of 1997, an extensive campaign for measurements of various propagation environments was undertaken, on McMaster University campus, with objective of providing a data set for testing any future algorithms. The complex channel impulse response, in time and space, was measured directly in the time domain by transmitting a wide-band spread-spectrum signal and correlating the received signal with the known transmitted sequence, at each element of the receiving array.

The receiving base station is a circular antenna array made of 8 mono-pole antennae. The transmitted signal is a 255 chip pseudo-noise (PN) sequence at 5 MHz. The received signal of each element was I-Q demodulated, converted to baseband, sampled at 10 MHz, and then stored for further processing. The measurements were conducted on campus, with the receiving base station at different locations and different heights in a pico-cell scenario that offered rich multipath characteristics with severe fading.

An initial calibration of the receiver array was based on measurements with the antenna array inside an anechoic chamber, illuminated from 64 different angles. This calibration data is necessary to extract accurate phase information from the data.

1.4 Scope of the thesis

This thesis discusses new approaches to various classical problems in array signal processing, using modern numerical Bayesian methods. We show how the *Markov Chain Monte Carlo* (MCMC) and the *Sequential Importance Sampling* (SIS) methods present new outlooks and offer many advantages to problems in the field.

Most array signal processing approaches require the knowledge of model order to estimate other parameters. Either it is assumed known, or it has been previously estimated separately. But for most real problems, this parameter is amongst the parameters of interest and should therefore be estimated jointly.

Very few methods tackle this joint problem of detection of the model order and estimation of the parameters, particularly in coloured noise with unknown and arbitrary covariance. A few methods proposed variations of the white noise methods. There is however a rich literature about the separate problems and many methods have been published in recent years. These will be presented throughout the thesis, mostly at the beginning of Chapters 3, 4 and 5.

This thesis presents the application of the Markov Chain Monte Carlo methods to the *joint problem* of detection and estimation in unknown coloured noise with arbitrary covariance and the sequential implementation of Monte Carlo methods applied to probabilistic dynamic systems, again in the context of joint detection and estimation.

Three problems were addressed in the course of this thesis:

1. A method for joint detection of the number of sources and estimation of their respective directions of arrival in coloured noise with unknown and arbitrary covariance using MCMC methods was developed.

   The method uses a single array of sensors and analytically integrates out the unknown noise covariance matrix to leave a marginalized posterior distribution that is only a function of the parameters of interest. The numerical optimization is accomplished with the Reversible Jump MCMC method.

2. The second algorithm represents an extension of the first one with the addition of the joint estimation of the times of arrival of the pulses.

   In the spirit of channel sounding or channel characterization, the received signal is composed of the superposition of multiple delayed and attenuated copies of the transmitted signal. The objective is to detect the number of multipath components and estimate their respective directions of arrival and excess propagation delays.

   Both methods were successfully applied to real data, acquired on campus with a channel sounder during the extensive measurement campaign previously mentioned.

3. The final part of this thesis focuses on the sequential implementation of the Monte Carlo methods in particle filters for probabilistic dynamic systems. This algorithm proposes a solution to the problem of tracking an unknown number of sources in white noise with an array of sensors. At each snapshot, the joint posterior distribution of the number of sources and their directions of arrival is recursively updated and optimized.

   This method allows for the instantaneous estimate of the number of sources and their evolving directions of arrival. In addition, by obtaining the marginal posterior distribution of the nuisance parameters, one could perform a very simple, yet quite effective, form of data association.

The selected numerical methods suffer from some practical difficulties. They are very computationally intense. However, these difficulties are offset by the following consideration:

- *Flexibility.* Once the samples are available, the integration and optimization objectives are very easily achieved.
- *Convergence.* In optimization, the probability of convergence to the global optimum is achieved with very high probability.
- *Adaptive.* For the SIS algorithm, there is no requirement of quasi-stationarity.
- *Performance.* They inherit the properties of Bayesian methods, which generally satisfy equality with the Cramér-Rao bound.

1.5 Outline of thesis

The thesis is divided into four main chapters.

- This first chapter was intended as an introduction of the knowledge required to understand the subsequent chapters and to help put the work in perspective.
- Chapter 2 introduces the fundamental theory supporting Markov Chain Monte Carlo methods and the sampling theory supporting particle filters.
- Chapter 3 presents the first of two methods developed for the joint detection and estimation of an unknown number of sources in coloured noise with unknown and arbitrary covariance. This algorithm can be applied in a frequency division duplex system for beamforming processing at the base station. As the weights are frequency dependent, the weights computed for the up-link channel cannot be reused for the down-link transmission. However, if the directions of arrival are estimated, both sets of weights can be calculated, which would allow for beamforming in both directions.
- Chapter 4 presents the second algorithm, which constitutes an extension of the first one, where the times of arrival of the plane waves are also jointly estimated. This method will be applied for channel sounding. It can also be applied at a base station, allowing for characterization of the multipath channel. This approach presents a potential solution to the 911-problem, where the location of users can be estimated.
- Chapter 5 presents the sequential implementation of the MC methods to a particle filter for the tracking of an unknown number of sources. This algorithm could be applied in a cellular communication scenario that utilizes smart antenna technology, where users can come in and out of the cell, to track the number of users and their angular locations.
- Finally, the conclusion section and appendices complete this work.

Chapter 2

Monte Carlo Methods

2.1 Introduction

As described in the previous chapter, the Bayesian approach allows for the consideration of prior knowledge included in posterior distributions, but often leads to highly non-linear and analytically non-tractable posterior distributions.

As most realistic data models will lead to multi-variate and highly non-linear posterior distributions that are not analytically workable, numerical methods present attractive alternatives.

If one was given a large set of samples drawn from a posterior distribution $\pi(\boldsymbol{x}) \triangleq p(\boldsymbol{x}|\boldsymbol{y})$, where $\boldsymbol{x}$ represents the parameter of interest, and $\boldsymbol{y}$ is the vector of observations, then one could easily get a numerical approximation of that posterior distribution by a histogram. This would allow for easy numerical integration, maximization or marginalization.

Assuming that a large number $N \gg 1$ of samples distributed according to a distribution of interest $p(\boldsymbol{x})$ are available, the Monte Carlo numerical approximation of this distribution is given by

$$\hat{P}_N(d\boldsymbol{x}) = \frac{1}{N}\sum_{i=1}^{N}\delta_{\boldsymbol{x}^{(i)}}(d\boldsymbol{x}), \tag{2.1}$$

where the indicator function is defined as

$$\delta_{\boldsymbol{x}^{(i)}}(d\boldsymbol{x}) = \begin{cases} 1 & \boldsymbol{x}^{(i)} \in d\boldsymbol{x} \\ 0 & \text{otherwise,} \end{cases}$$

and where $d\boldsymbol{x}$ is a small finite region surrounding a value of $\boldsymbol{x}$ of interest, e.g. a histogram bin.

Using this numerical approximation of the distribution, numerical estimates can readily be obtained. For example, the integration of any function $f(\boldsymbol{x})$ can be solved as

$$\mathrm{E}_{p(\boldsymbol{x})}\{f(\boldsymbol{x})\} \approx f_N \triangleq \int f(\boldsymbol{x})\hat{P}_N(d\boldsymbol{x}) = \frac{1}{N}\sum_{i=1}^{N} f(\boldsymbol{x}^{(i)}), \tag{2.2}$$

and the optimization of the posterior probability density function is simply

$$\hat{\boldsymbol{x}}_{MAP} = \arg \max_{\boldsymbol{x}^{(i)};i=1,\ldots,N} p(\boldsymbol{y}|\boldsymbol{x}^{(i)})p(\boldsymbol{x}^{(i)}).$$

These estimates are unbiased, in the sense that $\hat{P}_N(d\boldsymbol{x})$ converges to the true distribution. Furthermore, if the samples $\boldsymbol{x}^{(i)}$ are statistically independent, then the estimates converge to the true estimates, from the strong law of large numbers, e.g.

$$\lim_{N\to+\infty} f_N \to \mathrm{E}_{p(\boldsymbol{x})}\{f(\boldsymbol{x})\}.$$

Furthermore,

$$\lim_{N\to+\infty} \sqrt{N}(f_N - \mathrm{E}_{p(\boldsymbol{x})}\{f(\boldsymbol{x})\}) \stackrel{\sim}{\to} N(0,\sigma^2),$$

with $\sigma^2 < \infty$ and $\stackrel{\sim}{\to}$ denotes convergence in distribution.

Generally, it is not possible to draw samples directly from any arbitrary probability density function. In some cases, for standard distributions such as Gaussian, or uniform, many techniques exist to perform this task. For example, most classical methods are based on the inversion of the cumulative probability density function. However, this approach assumes that the probability density function can be integrated. Many other methods exist for drawing samples from standard distributions (Ripley, 1987). However, in practical cases of interest, these methods are found to be unsuitable.

The challenge is to generate *iid* samples, from any arbitrary multi-variate non-standard probability density functions. We now present a variety of methods for accomplishing this task.

2.2 Sampling techniques

2.2.1 Direct sampling

In the event that the constant of proportionality is known, or can at least be evaluated approximately, one classic algorithm is the Accept-Reject procedure. This algorithm assumes that

$$\pi(\boldsymbol{x}) \leq Mq(\boldsymbol{x}) \quad M < \infty \quad \forall \boldsymbol{x} \in \boldsymbol{E},$$

where $q(\boldsymbol{x})$ is some candidate distribution on the space $\boldsymbol{E}$ that is easy to sample from, and M is the "blanket" factor. The procedure is described below (Andrieu *et al.*, 1998).

Accept-Reject Algorithm

1. Sample $\boldsymbol{x} \sim q(\boldsymbol{x})$ and $u \sim U_{[0,1]}$.
2. if $u < \frac{\pi(\boldsymbol{x})}{Mq(\boldsymbol{x})}$ then return $\boldsymbol{x}$, otherwise return to step 1.

The set of returned $\boldsymbol{x}$ is then distributed according to $\pi(\boldsymbol{x})$. This procedure suffers from one major drawback: the "blanketing" factor M needs to be estimated, which is not always possible. Also, the probability of accepting a proposed $\boldsymbol{x}$ is $\frac{1}{M}$, making the algorithm inefficient in practice.

2.2.2 Importance sampling

In scenarios where the normalizing constant of the distribution of interest is unknown or can not be estimated, importance sampling constitutes a more efficient approach (Rubin,

1988)

Again, a candidate distribution $q(\boldsymbol{x})$ is used. This distribution, now named *importance function*, must have the same support as $\pi(\boldsymbol{x})$. Also, it must be easy to draw a large number of statistically independent samples from $q(\boldsymbol{x})$.

The Monte Carlo approximations defined in eqs. (2.1) and (2.2) can be redefined as

$$\hat{\pi}_N(d\boldsymbol{x}) = \frac{\sum_{i=1}^{N} w(\boldsymbol{x}^{(i)})\delta_{\boldsymbol{x}^{(i)}}(d\boldsymbol{x})}{\sum_{i=1}^{N} w(\boldsymbol{x}^{(i)})}, \tag{2.3}$$

$$f_N = \frac{\sum_{i=1}^{N} w(\boldsymbol{x}^{(i)}) f(\boldsymbol{x}^{(i)})}{\sum_{i=1}^{N} w(\boldsymbol{x}^{(i)})}, \tag{2.4}$$

with the importance weights defined as

$$w(\boldsymbol{x}^{(i)}) = \frac{\pi(\boldsymbol{x}^{(i)})}{q(\boldsymbol{x}^{(i)})}.$$

In effect, this algorithm converts samples from the distribution $q(\boldsymbol{x})$ to the desired distribution $\pi(\boldsymbol{x})$.

In the event that the desired distribution $\pi(\boldsymbol{x})$ is completely known, the weights would already be normalized. The algorithm is then formally referred to as the importance sampling algorithm.

In most applications, however, it is impossible to evaluate the total probability of $\pi(\boldsymbol{x})$, which remains known up to a constant and thus the normalization of the weights is required, as done in the denominator of eqs. (2.3) and (2.4). We refer to this algorithm as the Bayesian importance sampling procedure.

Using the approximation of eq. (2.3) for $\pi(\boldsymbol{x})$ with a finite number of samples, the estimate from eq. (2.4) is biased. It is however asymptotically unbiased and

$$\lim_{N\to\infty} f_N \to \mathrm{E}_{\pi(\boldsymbol{x})}\{f(\boldsymbol{x})\}.$$

This approach is very easy to implement and is well adapted for sequential processing. On the other hand, it can be quite inefficient if the importance function does not represent the target distribution closely. Furthermore, if the system is multivariate, it will be very difficult to propose good candidates. It is rarely used in practice if the system has more than a few parameters.

2.3 Markov Chain Monte Carlo Methods

Another strategy to obtain the samples is to observe the states of a Markov chain whose limiting distribution is our distribution of interest $\pi(\boldsymbol{x})$. Each state of the chain represents a bin of the histogram. These algorithms have nice properties, such as guaranteed convergence and insensitivity to the initial values, which, in some circumstances, might outweigh the computational burden, which can be significant in some cases.

The supporting theory is summarized in the following sections. The reader is referred to Ruanaidh and Fitzgerald (1996) for a good introductory presentation and to Gilks *et al.* (1998), Robert and Casella (1999) and Gamerman (1997) for a more complete treatment.

2.3.1 General theory of Markov chains

A Markov chain is a special type of stochastic process defined in terms of its states. Let the possible values that the process $\boldsymbol{x}_i$ can take be denoted by the set $A = \{S_1, S_2, \ldots, S_q\}$. When $\boldsymbol{x}_i = S_k$, the Markov chain is said to be in state k.

For a first order Markov process, the probability of the next state only depends on the present state, and not on any previous states, that is

$$P[\boldsymbol{x}_{i+1} \in A | \boldsymbol{x}_i, \ldots, \boldsymbol{x}_o] = P[\boldsymbol{x}_{i+1} \in A | \boldsymbol{x}_i].$$

Let's define the one-step transition probability from another state as

$$P_{k|j}[\boldsymbol{x}_i] = P[\boldsymbol{x}_i = S_k | \boldsymbol{x}_{i-1} = S_j].$$

Let's assume that the transition probabilities are stationary over time, then the complete set of these probabilities can be conveniently represented in a stochastic matrix named transition matrix, which has the form

$$\boldsymbol{\Pi} = \begin{bmatrix} P_{1|1} & P_{1|2} & \cdots & P_{1|q} \\ P_{2|1} & P_{2|2} & \cdots & P_{2|q} \\ & \vdots & & \\ P_{q|1} & P_{q|2} & \cdots & P_{q|q} \end{bmatrix}. \tag{2.5}$$

The same system can be represented in a space diagram. The following figure, 2.1, shows a simplified version of a three state Markov chain.

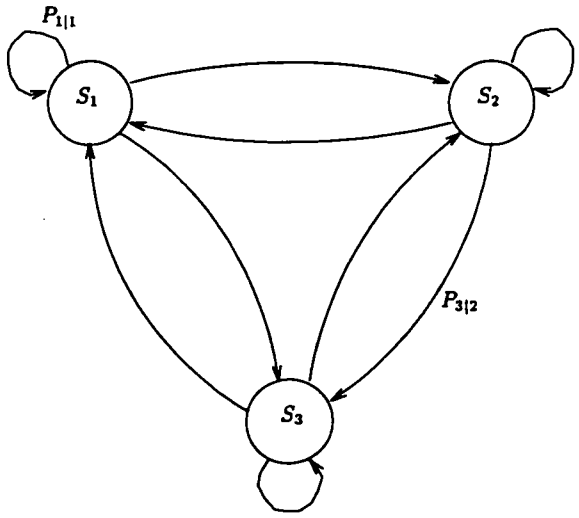

Figure 2.1: State transition diagram for a three state Markov chain.

From our definition of the transition matrix in eq. (2.5), it is easy to derive a general expression for the probability of $\boldsymbol{x}_i$ being in state S_k, from state S_j, after n steps as

$$P^{(n)}_{k|j}[\boldsymbol{x}_i] = P[\boldsymbol{x}_i = S_k | \boldsymbol{x}_{i-n} = S_j] = \Pi^n P[\boldsymbol{x}_{i-n} = S_j].$$

When the kernel satisfies certain conditions, the Markov chain will converge toward a special limiting distribution, denoted π, which will be independent of the initial state,

$$\pi = \lim_{n\to\infty} P[\boldsymbol{x}_i] = \lim_{n\to\infty} \Pi^{n-n_o} P[\boldsymbol{x}_{n_o}].$$

The distribution of the states then represents the limiting distribution of the chain. At that point, the states of the chain are all distributed according to the limiting distribution,

$$\Pi\pi = \pi. \tag{2.6}$$

Thus, the limiting distribution is unchanged after a state transition. We see that if the limiting distribution π of a Markov chain is the posterior distribution of interest, observing

the states of the chain provides samples from the distribution. The first samples are not distributed as desired, as the chain has yet to reach the equilibrium state, during a process called "burn-in". These samples are discarded.

Methods to determine whether or not a particular chain has reached its limiting distribution is a field of research on its own. In this thesis, the time-series of states are observed and the end of the "burn-in" periods will be empirically estimated.

2.3.2 Properties of Markov chains

Although not all Markov chains have a limiting distribution, many algorithms exist to set up Markov chains that will converge to the desired density function. For these algorithms to perform as intended, some conditions must be satisfied: invariance, reversibility, irreductibility, recurrence, and aperiodicity.

- Invariance

  This condition is necessary as it implies that the Markov chain has a limiting distribution. It means that if a state $\boldsymbol{x}_i$ is distributed according to $\pi(d\boldsymbol{x})$, then all the following states are distributed according to the same distribution, as described by eq. (2.6).

  To ensure that a Markov chain is π invariant, a sufficient condition is the reversibility condition.

- Reversibility

  A Markov chain is reversible if the distribution of $\boldsymbol{x}$ conditioned on $\boldsymbol{y}$ is the same as the distribution of $\boldsymbol{y}$ conditioned on $\boldsymbol{x}$. In simpler words, the direction of time does not influence the dynamic of the chain,

  $$P[\boldsymbol{x}|\boldsymbol{y}]\pi(\boldsymbol{x}) = P[\boldsymbol{y}|\boldsymbol{x}]\pi(\boldsymbol{y}).$$

- Irreductibility

  The irreductibility condition means that all the states can be reached with a non-null

probability in a finite number of iterations. A sufficient condition for a kernel P to be irreductible with respect to a distribution ψ is that P^n has a positive density with respect to ψ (Gilks *et al.*, 1998).

- Recurrence

  While irreductibility means that all states can be reached, recurrence means that the states can be reached infinitely often.

- Aperiodicity

  This condition prevents the use of kernels which would induce a periodic behaviour in the states.

All the MCMC algorithms have been designed to satisfy these constraints. The predominant methods are the Gibbs sampler and the Metropolis-Hasting algorithm. This thesis relies heavily on the use of the M-H algorithm and its variants.

2.3.3 Gibbs sampler

The Gibbs sampler was first introduced for image processing applications (Geman and Geman, 1984). Here, it is only discussed briefly since it will not be considered further in this thesis. Given a random vector $\boldsymbol{x}$ of length k, the Gibbs sampler samples each parameter, one at the time, according to the conditional distributions when all the other parameters are fixed. Instead of sampling from a complex k dimensional distribution, the problem is reduced to sampling k times from one dimensional conditional distributions.

To improve the rate of convergence, it is recommended that highly correlated variables be sampled jointly, creating p partitions. Also, it might be beneficial to randomly vary the order of the components.

Gibbs Sampling Algorithm

1. Initialization $\boldsymbol{x}_k^{(0)}$ at iteration 0

2. Iteration i, $i \geq 1$,

- Sample $\boldsymbol{x}_1^{(i)} \sim \pi(\boldsymbol{x}_1|\boldsymbol{x}_{-1}^{(i)})$
- Sample $\boldsymbol{x}_2^{(i)} \sim \pi(\boldsymbol{x}_2|\boldsymbol{x}_{-2}^{(i)})$
- Sample $\boldsymbol{x}_p^{(i)} \sim \pi(\boldsymbol{x}_p|\boldsymbol{x}_{-p}^{(i)})$

with, $\boldsymbol{x}_{-m}^{(i)} = (x_1^{(i)}, x_2^{(i)}, ..., x_{m-1}^{(i)}, x_{m+1}^{(i-1)} ...)$

■

For this algorithm to be a viable option, all the full conditional posterior distributions must be available in their analytical form.

2.3.4 Metropolis-Hasting algorithm

Another sampling scheme is the Metropolis-Hastings algorithm (M-H) (Hastings, 1970), which uses a *candidate function* $q(\cdot)$ to sample from $\pi(\cdot)$. This candidate function is chosen to be easy to sample from. One major advantage of this algorithm is that the knowledge of the normalizing constant of the posterior distribution is not required. The posterior distribution is only present in ratios, where this unknown normalizing constant will cancel out, assuming it remains constant.

Let's assume that the chain is in state $\boldsymbol{x}$. A candidate $\boldsymbol{x}^\star$ for the next state is obtained by sampling $q(\cdot)$, which in the general case is conditional on $\boldsymbol{x}$. This candidate will be accepted with probability α defined as

$$\alpha(\boldsymbol{x}^\star, \boldsymbol{x}) = \min\{r(\boldsymbol{x}^\star, \boldsymbol{x}), 1\}, \tag{2.7}$$

with the acceptance ratio r defined as

$$r(\boldsymbol{x}^\star, \boldsymbol{x}) = \frac{\pi(\boldsymbol{x}^\star)q(\boldsymbol{x}|\boldsymbol{x}^\star)}{\pi(\boldsymbol{x})q(\boldsymbol{x}^\star|\boldsymbol{x})}. \tag{2.8}$$

If the candidate is accepted, the chain takes the new state $\boldsymbol{x}^\star$; otherwise the chain remains at the current state $\boldsymbol{x}$.

Metropolis-Hasting Algorithm

1. Initialization $\boldsymbol{x}^{(0)}$ at iteration 0

2. Iteration i, $i \geq 1$,

   - Sample a candidate $\boldsymbol{x}^{\star} \sim q(\boldsymbol{x}|\boldsymbol{x}^{(i-1)})$.
   - Evaluate the acceptance probability

$$r(\boldsymbol{x}^{\star}, \boldsymbol{x}^{(i-1)}) = \frac{\pi(\boldsymbol{x}^{\star}) q(\boldsymbol{x}^{(i-1)}|\boldsymbol{x}^{\star})}{\pi(\boldsymbol{x}^{(i)}) q(\boldsymbol{x}^{\star}|\boldsymbol{x}^{(i-1)})}.$$

$$\alpha(\boldsymbol{x}^{\star}, \boldsymbol{x}^{(i-1)}) = \min\{r(\boldsymbol{x}^{\star}, \boldsymbol{x}^{(i-1)}), 1\}$$

   - Sample $u \sim U_{[0,1]}$
   - if $u \leq \alpha(\boldsymbol{x}^{\star}, \boldsymbol{x}^{(i-1)})$
     then the chain takes the state $\boldsymbol{x}^{(i)} = \boldsymbol{x}^{\star}$, otherwise it remains at $\boldsymbol{x}^{(i)} = \boldsymbol{x}^{(i-1)}$.

■

A simple condition that ensures the irreductibility and the aperiodicity of the M-H algorithm requires that the candidate function $q(\cdot)$ be continuous and strictly positive on the support of $\pi(\boldsymbol{x})$. The choice of the acceptance probability defined in eq. (2.7) with eq. (2.8) can be shown to satisfy the reversibility requirement.

For good mixing properties, the candidate function should allow the chain to move considerably from its actual state, but with substantial probability of being accepted.

In the case where $\boldsymbol{x}$ is of high dimension, it becomes very difficult to select a good candidate function that would lead to a reasonable acceptance rate and allow the chain to mix. To address this problem, the *Metropolis-Hasting one-at-the-time* algorithm (Andrieu *et al.*, 1998), in a similar fashion to the Gibbs sampler, samples each component (or partition), conditionally on the other components, using a set of candidate functions. Obviously, this algorithm includes the Gibbs sampler as a special case for which the candidate functions are the full conditional distributions and the candidates are always accepted.

2.3.5 Reversible jump MCMC

The methods described earlier work well in practice. They however require the dimension of the parameter space to be explored (i.e. *model order*) be fixed or known.

One could obviously sample the subspace corresponding to a range of model orders independently, but the computational cost of such a scheme can be very large. Also this approach is very inefficient since the same effort is allocated to all model orders, even though some models will have a very low posterior probability.

A more general MCMC framework was developed by Green (1995) for model order uncertainty. This method directly samples the model order, k, as a parameter. The whole parameter space $\bigcup_{k=0}^{k_{max}} k \times \Phi_k$, where Φ_k is the space of the parameters of the model of order k, is visited by moves designed to preserve the reversibility condition. This method allows for the joint detection of the model order and the sampling of the parameters from its posterior distribution. The samples concentrate on models with high posterior probability.

The algorithm itself is similar to the M-H algorithm. At each iteration, the algorithm proposes a candidate, this time from a set of candidate functions. These functions are designed to change the model order and explore the different subspaces.

At each iteration, a candidate distribution $q_m(\cdot)$ is randomly chosen, and a candidate $\boldsymbol{x}^\star$ of size $k^\star$ is obtained by sampling $q_m(\cdot)$. This candidate will be accepted with probability α defined as

$$\alpha((\boldsymbol{x}^\star, k^\star), (\boldsymbol{x}, k)) = \min\{r((\boldsymbol{x}^\star, k^\star), (\boldsymbol{x}, k)), 1\}, \tag{2.9}$$

with the acceptance ratio r defined as

$$r((\boldsymbol{x}^\star, k^\star), (\boldsymbol{x}, k)) = \frac{\pi(\boldsymbol{x}^\star, k^\star) q_m(\boldsymbol{x}, k)}{\pi(\boldsymbol{x}, k) q_m(\boldsymbol{x}^\star, k^\star)} J((\boldsymbol{x}^\star, k^\star), (\boldsymbol{x}, k)). \tag{2.10}$$

If the candidate is accepted, the chain takes the new state; otherwise the chain remains at the current state.

The term $J((\boldsymbol{x}^\star, k^\star), (\boldsymbol{x}, k))$ is the Jacobian of the transformation, required to reconcile the total probability between spaces of different dimensions so that the reversibility

condition is satisfied, and is defined (see Godsill (1998)) as

$$J((\boldsymbol{x}^{\star}, k^{\star}), (\boldsymbol{x}, k)) = \left| \frac{\partial \boldsymbol{x}^{\star}}{\partial \boldsymbol{x}} \right|. \tag{2.11}$$

The reversibility condition is satisfied by the appropriate choice of the candidate functions and the definition of the acceptance probability eq. (2.9) and eq. (2.10).

The most widely used candidate functions are the birth/death complementary moves. When the death move is selected, the algorithm proposes a candidate in the model of lower dimension, as opposed to the birth move, which proposes a candidate of higher dimension. The methods was first applied to engineering problems by Andrieu and Godsill (Andrieu and Doucet, 1999; Andrieu, 1997; Troughton and Godsill, 1997, 1998) for the detection and estimation of parameters in white noise.

The probabilities for choosing each move are denoted u_k, b_k and d_k, respectively, such that $u_k + b_k + d_k = 1$ for all k, and so that the probability of a jump is between 0.5 and 1 at each iteration (Green, 1995). The overall description of the reversible jump MCMC algorithm is determined by the choice of moves at each iteration. This description is summarized as follows.

Reversible Jump MCMC

1. Initialization: set $\Phi^{(0)} = (\boldsymbol{x}^{(0)}, k^{(0)})$

2. Iteration i,

   - Sample $u \sim U_{[0,1]}$
   - if $(u < b_{k^{(i)}})$ then execute a "birth move" ,
     - else if $(u < b_{k^{(i)}} + d_{k^{(i)}})$ then execute a "death move" ,
     - else, execute an update move .

3. $i \leftarrow i + 1$, goto step 2

2.3.5.1 Update move

When the update move is selected, a candidate of same dimension is proposed. One of the previously described MCMC algorithms can be used to sample the space of fixed and known dimension.

2.3.5.2 Birth and death moves

In the death move, we assume the current state is in $(\Phi_{k+1}, \{k+1\})$ and we wish to determine whether the state is in $(\Phi_k, \{k\})$ at the next iteration. The subscript k now indicates the size of the parameter space. This involves the removal of an element of the parameter vector, which is chosen randomly amongst the $(k+1)$ existing elements.

The proposal distribution $q(\boldsymbol{x}_k^\star, k|\boldsymbol{x}_{k+1}, k+1)$ for the death move is therefore chosen as

$$q(\boldsymbol{x}_k^\star, k|\boldsymbol{x}_{k+1}, k+1) = p(k) \div \binom{k+1}{1} \propto p(k)\frac{1}{(k+1)},$$

where $p(k)$ is the prior distribution of the model order. The candidate state $(\boldsymbol{x}_k^\star, k)$ is then accepted with probability $\alpha_{death} = \min\{r, 1\}$ with r defined by eq. (2.10).

Similarly, in the birth move, we assume the current state is in $(\Phi_k, \{k\})$ and we wish to determine whether the next state is in $(\Phi_{k+1}, \{k+1\})$. This involves the addition of a element x_c, selected from a candidate distribution, to the existing vector parameter, to create

$$\boldsymbol{x}_{k+1}^\star = [\boldsymbol{x}_k, x_c].$$

The proposal distribution $q(\boldsymbol{x}_{k+1}^\star, (k+1)|\boldsymbol{x}_k, k)$ for the birth move is therefore

$$q(\boldsymbol{x}_{k+1}^\star, (k+1)|\boldsymbol{x}_k, k) = p(k+1)p(x_c),$$

where $p(x_c)$ is the candidate distribution of an individual parameter x_i. The candidate state $(\boldsymbol{x}_{k+1}^\star, k+1)$ is then accepted with probability $\alpha_{birth} = \min\{r, 1\}$ with r defined by eq. (2.10).

In the case of the birth/death moves, the Jacobian of the transformation is unity. However, this is not always the case, as it will be demonstrated with the introduction of new moves described in chapter 5.

The following block describes the algorithm for the death move.

Death Move

- Select randomly the jth element to form the candidate,

$$\boldsymbol{x}_k^{(i+1)} = [\boldsymbol{x}_{1:(j-1)}^{(i)}, \boldsymbol{x}_{(j+1):(k+1)}^{(i)}]$$

- Evaluate α_{death} with eq. (2.9).
- Sample $u \sim U_{[0,1]}$.
- if $(u \leq \alpha_{death})$ then the state of the Markov Chain becomes $(\boldsymbol{x}_k^{(i+1)}, k)$, else it remains at $(\boldsymbol{x}_{k+1}^{(i)}, k+1)$.

The following block describes the algorithm for the birth move.

Birth Move

- Propose a new element x_c and a candidate,

$$\boldsymbol{x}_{k+1}^{(i+1)} = [\boldsymbol{x}_k^{(i)}, x_c]$$

- Evaluate α_{birth} with eq. (2.9).
- Sample $u \sim U_{[0,1]}$.
- if $(u \leq \alpha_{birth})$ then the state of the Markov Chain becomes $(\boldsymbol{x}_{k+1}^{(i+1)}, k+1)$, else it remains at $(\boldsymbol{x}_k^{(i)}, k)$.

This sampling scheme constitutes the main engine for Chapters 3, for DOA estimation, and 4, for multipath channel characterization, both problems defined in unknown and arbitrary coloured noise.

The problems that will be addressed are formulated from the array signal processing model and have the following structure, as in Chapter 1,

$$\boldsymbol{y}(n) = \boldsymbol{S}(\boldsymbol{\phi})\boldsymbol{a}(n) + \boldsymbol{\nu}(n) \quad n = 1, \ldots, N_t,$$

where the amplitudes $\boldsymbol{a}(n)$ are unknown and *iid* and the statistics of the Gaussian noise samples $\boldsymbol{\nu}(n)$ are also unknown. The DOAs $\boldsymbol{\phi}$ become the parameters of interest previously defined as $\boldsymbol{x}$ in the MCMC context.

The objective is to find the MAP estimate of the parameter $\boldsymbol{\phi}$, based on the observations $\boldsymbol{y}(n)$, as

$$(\hat{\boldsymbol{\phi}}_{MAP}, \hat{k}_{MAP}) = \arg \max_{\boldsymbol{\phi}^{(i)}, k^{(i)}; i=1,\ldots,N} \pi(\boldsymbol{\phi}, k).$$

2.4 Sequential Monte Carlo methods

The algorithms presented so far are all off-line methods. They require the accumulation of a number of observations before any processing can be done and assume that the parameters of interest remain constant over the window of observation. In many applications, sequential processing is desired which would allow for the tracking of varying parameters.

When the parameter $\boldsymbol{x}$ evolves in time, the sequence of evolving posterior density functions $\pi(\boldsymbol{x}_t)$ is called a *probabilistic dynamic system.*

To implement Monte Carlo methods for a dynamic system, a set of samples (*particles*) distributed according to the $\pi(\boldsymbol{x}_t)$ is required. As it is usually not possible to sample directly from this posterior distribution, the samples are drawn from another distribution $q(\cdot)$ and weighted according to the Bayesian importance sampling procedure.

Also, in most applications, the difference between $\pi(\boldsymbol{x}_t)$ and $\pi(\boldsymbol{x}_{t+1})$ is small and is caused by the incorporation of new information, which could allow for the recycling of the

samples. The samples used at time t can be recursively re-weighted to construct the set at time $t+1$.

A family of algorithms based on the sequential importance sampling (SIS) method have accomplished this task. This area of research is very recent and only a few books are available on the subject, e.g. Doucet *et al.* (2001b). Many key papers are available on this new subject (Liu and Chen, 1993; Andrieu *et al.*, 1999; Doucet *et al.*, 2001a; Djuric, 2000; Gordon *et al.*, 1993).

2.4.1 Particle filtering

This section presents in detail the implementation of a sequential importance sampling procedure for the particle filter. The objective is to sequentially and recursively estimate the posterior distribution of interest as time evolves.

Let's assume that at time t a set of particles $\boldsymbol{x}_{1:t}^{(i)}, i = 1, \ldots, N$ distributed according to the joint distribution $p(\boldsymbol{x}_{1:t}|\boldsymbol{y}_{1:t})$ are available. Then, as new data $\boldsymbol{y}_{t+1}$ becomes available, the objective is update the numerical histogram of the particles to give $p(\boldsymbol{x}_{1:t+1}|\boldsymbol{y}_{1:t+1})$, without resampling all the particles from time 1 to $t+1$. The importance sampling approximation can be recursively updated as new data arrives, in order to keep the previously simulated trajectories of the particles, and to avoid sampling the increasingly long parameter vector as a whole.

The joint posterior distribution of all parameters $\boldsymbol{x}$ from time 1 to $t+1$ can be written using Bayes' theorem as

$$\pi(\boldsymbol{x}_{1:t+1}) = \frac{p(\boldsymbol{y}_{1:t+1}|\boldsymbol{x}_{1:t+1})p(\boldsymbol{x}_{1:t+1})}{p(\boldsymbol{y}_{1:t+1})}. \tag{2.12}$$

It can be shown (Doucet, 1998), using the Markov properties of the model and the *iid* assumptions on the noise variables, that (2.12) can be written in the recursive, time-update form as (see appendix B)

$$p(\boldsymbol{x}_{1:t+1}|\boldsymbol{y}_{1:t+1}) = p(\boldsymbol{x}_{1:t}|\boldsymbol{y}_{1:t})\frac{p(\boldsymbol{y}_{t+1}|\boldsymbol{x}_{t+1})p(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t)}{p(\boldsymbol{y}_{t+1}|\boldsymbol{y}_{1:t})}. \tag{2.13}$$

In principle, this recursion would allow us to sequentially and recursively compute the

posterior distribution. However, it is not useful in its present form, since the normalizing constant $p(\boldsymbol{y}_{t+1}|\boldsymbol{y}_{1:t})$ and the desired marginal distributions require the evaluation of complex, multi-dimensional integrals, which are generally difficult or impossible to evaluate analytically. Also, it is generally not possible to directly generate samples directly from the posterior distribution. We therefore resort to the Bayesian importance sampling scheme to generate the particles and estimate the posterior distribution of interest.

Assuming at the tth time instant a set of weights $\tilde{w}_{1:t}^{(i)}, i = 1, \ldots, N$ is available, the approximation $\hat{\pi}_N(d\boldsymbol{x}_{1:t})$ of the joint distribution $\pi(\boldsymbol{x}_{1:t})$ can be determined as

$$\hat{\pi}_N(d\boldsymbol{x}_{1:t}) = \frac{\sum_{i=1}^{N} \tilde{w}^{(i)}(\boldsymbol{x}_{1:t}^{(i)})\delta_{\boldsymbol{x}_{1:t}^{(i)}}(d\boldsymbol{x}_{1:t})}{\sum_{i=1}^{N} \tilde{w}^{(i)}(\boldsymbol{x}_{1:t}^{(i)})}.$$

In order to implement the sequential MC procedure, the importance function must satisfy the following recursivity condition (Doucet, 1998)

$$q(\boldsymbol{x}_{1:t+1}|\boldsymbol{y}_{1:t+1}) = q(\boldsymbol{x}_{1:t}|\boldsymbol{y}_{1:t})q(\boldsymbol{x}_{t+1}|\boldsymbol{x}_{1:t}, \boldsymbol{y}_{1:t+1}). \tag{2.14}$$

Also, in order to minimize the variance of the resulting weights as discussed below, the optimal importance function that satisfies the recurrence requirement eq. (2.14) is (Doucet *et al.*, 2000)

$$q_{optimal}(\cdot) = q(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t, \boldsymbol{y}_{t+1}). \tag{2.15}$$

These constraints on the importance function along with eq. (2.13) allow us to evaluate the unnormalized importance weights recursively as (see appendix B)

$$\tilde{w}(t+1) = w(t)\frac{p(\boldsymbol{y}_{t+1}|\boldsymbol{x}_{t+1})p(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t)}{q(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t, \boldsymbol{y}_{t+1})}. \tag{2.16}$$

The vector of weights must be normalized by dividing by the sum of its components.

$$w(t+1) = \frac{\tilde{w}(t+1)}{\sum_{i=1}^{N} \tilde{w}^{(i)}(t+1)}. \tag{2.17}$$

2.4.1.1 Resampling

Even though the importance function was chosen carefully according to eq. (2.15), in practice the recursive algorithm involving eq. (2.16) still quickly degenerates. Only a handful

of particles have meaningful weights after a few iterations of the importance sampling algorithm.

It is desirable to minimize the variance of the weights to get the most out of the samples. In the limit, if only one particle has all the weight, then any inference would be done on one sample. In the ideal case, all the particles have equal weights (with zero variance), therefore they are *iid*.

Therefore, any estimates based on these very few particles would show a large variance. It is thus necessary to resample the particles according to their importance weights, by multiplying/suppressing the particles according to their importance weights. This resampling operation can be done very efficiently, with $\mathcal{O}(N)$ operations, where N is the number of particles (Ripley, 1987; Doucet, 1998).

However, with the resampling process the trajectories with high importance weights are statistically selected many times, which limits the true statistical diversity amongst the particles. One approach would consist of adding an arbitrary perturbation to the child particles. However, this does not add true statistical diversity and only increases the variance of the estimates.

A more clever approach, proposed by Andrieu and Doucet (Andrieu *et al.*, 1999; Doucet *et al.*, 2000; MacEachern *et al.*, 1999; Gilks and Berzuini, 1998) uses an MCMC step on each particle. At time t, the particles are marginally distributed as $p(\boldsymbol{x}_{1:t}|\boldsymbol{y}_{1:t})$. An MCMC engine with a kernel of invariant distribution $p(\boldsymbol{x}_{1:t}|\boldsymbol{y}_{1:t})$ can be used to generate a new set of particles that will also be distributed according to this posterior distribution. This approach provides a valid way to re-introduce the diversity amongst the particles. Any of the MCMC methods described earlier can be used, including the reversible jump MCMC method. It is this MCMC step that allows the algorithm to converge toward the right model order.

The following schema describes the SIS procedure applied at each time step $t = 1, 2, \ldots$ for tracking changing parameters.

Generic Sequential Importance Sampling for Tracking an Unknown Number of Parameters

1. The Importance Sampling Step:

   - For $i = 1, \ldots, N$, Get the particles from the importance function

   $$\boldsymbol{x}^{(i)}(t) \sim q(\boldsymbol{x}(t)|\boldsymbol{x}^{(i)}(t-1), \boldsymbol{y}(t), k^{(i)}(t)) \tag{2.18}$$

   - For $i = 1, \ldots, N$, Evaluate the un-normalized importance weights $\tilde{w}^{(i)}(t)$ from eq. (2.16)

   $$\tilde{w}^{(i)}(t) = \tilde{w}^{(i)}(t-1) \times \frac{p(\boldsymbol{y}(t)|\boldsymbol{x}^{(i)}(t), k^{(i)}(t)) p(\boldsymbol{x}^{(i)}(t)|\boldsymbol{x}^{(i)}(t-1), k^{(i)}(t))}{q(\boldsymbol{x}^{(i)}(t)|\boldsymbol{x}^{(i)}(t-1), \boldsymbol{y}(t), k^{(i)}(t))} \tag{2.19}$$

   - For $i = 1, \ldots, N$, Normalize the weights: $w^{(i)}(t) = \frac{\tilde{w}^{(i)}(t)}{\sum_{j=1}^{N} \tilde{w}^{(j)}(t)}$

   Eqns (2.18) and (2.19) represent generalized forms of eqs. (2.15) and (2.16).

2. The Resampling/Selection of the Particles:

   - Sample a vector of index $\boldsymbol{l}$, with pdf described by the weights

   $$P(l(j) = i) = w^{(i)}(t)$$

   - Resample the particles with the index vector

   $$\boldsymbol{x}_{0:t}^{(i)} = \boldsymbol{x}_{0:t}^{(l(i))}$$

   - Re-assign all the weights to $w^{(i)}(t) = \frac{1}{N}$.

3. Proceed with the Reversible Jump MCMC Step to introduce diversity in the particles and to facilitate detection of model order.

■

The generic problem addressed in this thesis follows the state-space model

$$\begin{aligned} k(t) &= k(t-1) + \epsilon(t), \\ \boldsymbol{\phi}(t) &= \boldsymbol{\phi}(t-1) + \sigma_v \boldsymbol{v}(t), \\ \boldsymbol{a}(t) &\sim \mathcal{N}(\mathbf{0}, \sigma_a^2 \boldsymbol{I}_{k(t)}), \\ \boldsymbol{y}(t) &= \boldsymbol{S}(\boldsymbol{\phi}(t))\boldsymbol{a}(t) + \sigma_w \boldsymbol{\omega}(t), \end{aligned}$$

where the amplitudes $\boldsymbol{a}(t)$ are unknown and *iid* between observations, and the parameter of interest $\boldsymbol{\phi}$ evolves in time, with the possibility that its the size, k, might change at any time during the period of observation.

The random variables $\boldsymbol{v}$, $\boldsymbol{\omega}$ are Gaussian with zero mean and unit variance. The model order is allowed to change by at most one between snapshots, with ϵ defined as

$$\begin{aligned} P(\epsilon(t) = -1) &= h/2 \\ P(\epsilon(t) = 0) &= 1 - h \\ P(\epsilon(t) = 1) &= h/2, \end{aligned}$$

where $h \in [0, 1]$.

The objective is to construct sequentially the numerical approximation of $p(\boldsymbol{\phi}_{1:t}, k_{1:t} | \boldsymbol{y}_{1:t})$ and to estimate

$$\{\hat{\boldsymbol{\phi}}_{1:t}, \hat{k}_{1:t}\} = \arg \max_{\boldsymbol{\phi}^{(i)}, k^{(i)}; i=1,\ldots,N} p(\boldsymbol{\phi}_{1:t}, k_{1:t} | \boldsymbol{y}_{1:t})$$

A more complete treatment of this problem is presented in Chapter 5, where once more in the context of array signal processing, the parameter $\boldsymbol{x}$ is the vector of DOAs $\boldsymbol{\phi}$.

Chapter 3

Joint Detection and Estimation in Coloured Noise

This chapter presents a novel Bayesian solution to the difficult problem of joint detection of the number of sources and estimation of the corresponding DOAs of sources impinging on an array in spatially coloured noise with arbitrary covariance structure. Robustness to the noise covariance is achieved by integrating out the unknown covariance matrix in an appropriate posterior distribution. The proposed solution uses the Reversible Jump Markov Chain Monte Carlo method to perform the numerical optimization.

We show that the determination of model order is consistent provided a particular hyperparameter is within a specified range. Simulation results support the effectiveness of the method.

3.1 Introduction

Current algorithms for DOA estimation commonly make assumptions regarding the characteristics of the noise field and the incident signals; in particular, that the noise is white (Schmidt, 1986; Andrieu and Doucet, 1999; Wu and Fuhrmann, 1991) or of known covariance. Further, in the typical scenario, the determination of model order (detection) and the

estimation of desired signal parameters are executed independently rather than jointly.

The direction of arrival problem in the case where the noise has unknown covariance has been addressed by Nagesha and Kay (1996); Wong *et al.* (1992); Agrawal and Prasad (2000). In these papers, the model order is assumed known, and the associated detection problem is not addressed. Established methods, e.g., using information theoretic criteria (Wax and Kailath, 1985; Rissanen, 1978), have been developed under the assumption of white noise, and exhibit considerable sensitivity to violations in whiteness.

Joint detection and estimation has been addressed by Andrieu and Doucet (1999) and Wu and Fuhrmann (1991) under the white noise assumption, and later generalized by Cho and Djuric (1995) to the autoregressive noise case.

Determination of model order for the unknown arbitrarily spatially coloured noise case has long been regarded as a difficult problem. Methods for detection in coloured noise have been proposed (LeCadre, 1989; Chen *et al.*, 1996; Fuchs, 1992), but these techniques require in effect two separate antenna arrays with restrictive assumptions on the configuration of the signal and noise fields. In (Wax, 1992), this problem was addressed from the information theoretic stand point, and a solution to the joint detection/estimation problem in arbitrary unknown noise was proposed. The method however involves the minimization of a highly nonlinear objective function and is therefore prone to being trapped in local minima.

A method for joint detection and estimation for array signal processing using Markov Chain Monte Carlo (MCMC) methods for the white noise case has been proposed by Andrieu and Doucet (1999). In this chapter (and in Larocque and Reilly (2000a,b,c)), we extend the work of Andrieu and Doucet (1999) to the case where the noise covariance is unknown and arbitrary. A significant consequence of this extension is a procedure for determination of model order in unknown coloured noise with a single array, with relaxed assumptions on the signal and noise fields.

This algorithm can be applied in a frequency division duplex system for beamforming processing at the base station. As the weights are frequency dependent, the weights computed for the up-link channel cannot be reused for the down-link transmission. However, if the directions of arrival are estimated, both sets of weights can be calculated, which would

allow for beamforming in both directions (Pahlavan and Levesque, 1995).

The signal model we consider here consists of an M-dimensional complex data vector $\boldsymbol{y}(n) \in \mathcal{C}^M$ which represents the data received by a linear array of sensors at the nth snapshot. The data vector is composed of incident narrow-band plane wave signals each at centre frequency ω_0 from k_o distinct sources embedded in Gaussian noise. The received vector at the nth snapshot can be written as

$$\boldsymbol{y}(n) = \boldsymbol{S}(\boldsymbol{\phi})\boldsymbol{a}(n) + \boldsymbol{\nu}(n), \qquad n = 1, \ldots, N_t, \tag{3.1}$$

where N_t is the number of observed snapshots. Each of these incident plane-wave signals impinges on the array of sensors at an angle $\theta_k, k = 1, \ldots, k_o$, to the normal of the array. Then $S(\boldsymbol{\phi})$ is the $M \times k_o$ matrix, defined as in eq. (1.1) or eq. (1.2).

The quantity $\boldsymbol{a}(n) \in \mathcal{C}^{k_o}$ represents the complex amplitudes of the incident signals at the nth snapshot. We assume the amplitudes are *iid* Gaussian between snapshots, with unknown and arbitrary mean and covariance. The spatially coloured noise vector $\boldsymbol{\nu}(n)$ is an *iid* normally distributed noise vector distributed according to $\mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_\nu)$, where $\boldsymbol{\Sigma}_\nu \in \mathcal{C}^{M \times M}$ is an unknown and arbitrary covariance matrix. The signal and noise are assumed uncorrelated such that

$$E(\boldsymbol{a}'\boldsymbol{S}'\boldsymbol{\nu}) = \mathbf{0}. \tag{3.2}$$

For a hypothesized number of signals k, we designate by $\mathcal{S}(k)$ the signal subspace spanned by the vectors $[\boldsymbol{s}(\phi_1), \ldots, \boldsymbol{s}(\phi_k)]$. The $(M-k)$-dimensional orthogonal complement noise subspace of $\mathcal{S}(k)$ is denoted by $\mathcal{N}(k)$. In the following discussion, we simplify notation by writing $\mathcal{S}$ and $\mathcal{N}$ instead of $\mathcal{S}(k)$ and $\mathcal{N}(k)$ respectively.

In this chapter, we jointly estimate k_o and $\boldsymbol{\phi}$ using MCMC techniques. Determination of k_o using MCMC methods involves sampling a posterior distribution $p(k, \boldsymbol{\phi}|\boldsymbol{y})$ of varying model order. This consideration requires the use of the reversible jump MCMC method (Green, 1995), which was presented in Chapter 2 and is treated in further detail in section 3.3.

We treat the unknown quantities $\boldsymbol{a}_n, n = 1, \ldots, N_t$ and $\boldsymbol{\Sigma}_\nu$ as nuisance parameters, which are integrated out. It is the elimination of $\boldsymbol{\Sigma}_\nu$ in this way which enables us to

handle correlated noise with unknown covariance. Even though these parameters can be eliminated by the sampling process inherent in the MCMC technique, we choose to integrate them analytically since the analysis is tractable, to produce a more efficient sampler for the parameters of interest.

After integration of the nuisance parameters, the parameter space Φ of interest for joint detection of the model order and estimation of the incident signal angles, includes k and the corresponding angles $\phi_1 \dots \phi_k$. We denote the permissible set of k to be $\mathcal{K} = \{0, \dots, M-1\}$. To reflect the fact that the number of parameters changes with model order, we note that Φ can be written as a finite union of subspaces as $\Phi = \cup_{k=0}^{M-1}\{k\} \times \Phi_k$, where $\Phi_k \triangleq \{(0, 2\pi)^k : |\phi_i - \phi_j| > \epsilon\}$, where ϵ is a small number > 0, and $(i, j) = \{0, \dots, k\}$. The set Φ_k is defined in this way so that $\boldsymbol{S}$ in eq. (3.1) is always full column rank. The set $\Phi_0 = \emptyset$.

In section 3.2, we develop the desired marginal posterior distribution $p(k, \boldsymbol{\phi}|\boldsymbol{y})$. We perform the integration of $\boldsymbol{\Sigma}_\nu$ by projecting the observed data onto the noise subspace $\mathcal{N}$. After assigning a Jeffreys' prior, the resulting posterior may then be integrated by comparison to a complex Wishart distribution. In section 3.3, we apply the reversible jump MCMC algorithm to sample the posterior distribution and so obtain joint estimates of the desired parameters k and $\boldsymbol{\phi}$. In section 3.4 we discuss conditions for consistency of the estimation of model order. Simulation results demonstrating the satisfactory performance of the proposed method are given in section 3.5, and the conclusions are presented in section 3.6.

3.2 Development of the marginal posterior distribution

Since the N_t snapshots are *iid*, the total likelihood function of all the data can be expressed as

$$p(\boldsymbol{Y}|\boldsymbol{\phi}_k, \boldsymbol{A}, \boldsymbol{\Sigma}_\nu, k) = \frac{1}{\pi^{MN_t}|\boldsymbol{\Sigma}_\nu|^{N_t}} e^{-\sum_{n=1}^{N_t}\left(\boldsymbol{y}(n)-\boldsymbol{S}(\boldsymbol{\phi}_k)\boldsymbol{a}(n)\right)'\boldsymbol{\Sigma}_\nu^{-1}\left(\boldsymbol{y}(n)-\boldsymbol{S}(\boldsymbol{\phi}_k)\boldsymbol{a}(n)\right)},$$

where $\boldsymbol{Y}, \boldsymbol{A}$ are all the data and amplitudes, respectively, $'$ denotes Hermitian transpose, and $|\cdot|$ denotes determinant.

To proceed with the integration of the nuisance parameters $\boldsymbol{\Sigma}_\nu$ and $\boldsymbol{A}$, we first define an orthonormal matrix $\boldsymbol{U}(\boldsymbol{\phi},k) \in \mathcal{C}^{M\times M}$ as in (Wong *et al.*, 1992; Cho and Djuric, 1995)

$$\underset{M\times M}{\boldsymbol{U}(\boldsymbol{\phi},k)} = \underset{M\times k}{[\boldsymbol{U}_s(\boldsymbol{\phi},k)} \quad \underset{M\times(M-k)}{\boldsymbol{U}_\nu(\boldsymbol{\phi},k)}],$$

where $\boldsymbol{U}_s(\boldsymbol{\phi},k) \in \mathcal{S}$ and $\boldsymbol{U}_\nu(\boldsymbol{\phi},k) \in \mathcal{N}$. We now transform the received data $\boldsymbol{y}(n)$ into $\boldsymbol{z}(n) \triangleq \boldsymbol{U}'\boldsymbol{y}(n)$, with a signal component $\boldsymbol{z}_s(n) \in \mathcal{S}$ and a noise component $\boldsymbol{z}_\nu(n) \in \mathcal{N}$ defined respectively as

$$\boldsymbol{z}_s(n) = \boldsymbol{U}_s'(\boldsymbol{\phi},k)\boldsymbol{y}(n),$$

and

$$\boldsymbol{z}_\nu(n) = \boldsymbol{U}_\nu'(\boldsymbol{\phi},k)\boldsymbol{y}(n).$$

In the neighbourhood around the true value where $\boldsymbol{\phi} \approx \tilde{\boldsymbol{\phi}}$ we have $\boldsymbol{S}'\boldsymbol{U}_\nu \approx \boldsymbol{0}$. Then using eq. (3.2), we can write

$$\boldsymbol{R}_{zz} = \mathrm{E}\left[\boldsymbol{z}\boldsymbol{z}'\right] \approx \begin{bmatrix} \boldsymbol{R}_{\tilde{a}\tilde{a}} + \boldsymbol{U}_s'\boldsymbol{\Sigma}_\nu\boldsymbol{U}_s & \boldsymbol{U}_s\boldsymbol{\Sigma}_\nu\boldsymbol{U}_\nu \\ \boldsymbol{U}_\nu'\boldsymbol{\Sigma}_\nu\boldsymbol{U}_s & \boldsymbol{U}_\nu'\boldsymbol{\Sigma}_\nu\boldsymbol{U}_\nu \end{bmatrix}, \tag{3.3}$$

where $\boldsymbol{R}_{\tilde{a}\tilde{a}} \triangleq \mathrm{E}\{\tilde{\boldsymbol{a}}(n)\tilde{\boldsymbol{a}}'(n)\}$ and $\tilde{\boldsymbol{a}}(n) = \boldsymbol{U}_s'\boldsymbol{S}(\boldsymbol{\phi})\boldsymbol{a}(n)$. If we neglect the effect of the off-diagonal blocks above, then locally around $\boldsymbol{\phi} \approx \tilde{\boldsymbol{\phi}}$ we can assume that $\boldsymbol{z}_s$ is independent of $\boldsymbol{z}_\nu$. Then using eq. (3.3) it follows that $\boldsymbol{z}_s \sim \mathrm{N}(\tilde{\boldsymbol{a}}, \boldsymbol{C})$, and $\boldsymbol{z}_\nu \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{W})$, where $\boldsymbol{C} \triangleq \boldsymbol{U}_s'\boldsymbol{\Sigma}_\nu\boldsymbol{U}_s$, $\boldsymbol{W} \triangleq \boldsymbol{U}_\nu'\boldsymbol{\Sigma}_\nu\boldsymbol{U}_\nu$. Under these conditions the joint likelihood function of $\boldsymbol{z}_s$ and $\boldsymbol{z}_\nu$ is given as

$$\begin{aligned} p(\boldsymbol{Z}_s, \boldsymbol{Z}_\nu | \tilde{\boldsymbol{A}}, \boldsymbol{\phi}, k, \boldsymbol{W}^{-1}) \approx & \\ \pi^{-N_t k}|\boldsymbol{C}^{-1}|^{N_t} \exp\left\{-\sum_{n=1}^{N_t}(\boldsymbol{z}_s(n) - \tilde{\boldsymbol{a}}(n))'\boldsymbol{C}^{-1}(\boldsymbol{z}_s(n) - \tilde{\boldsymbol{a}}(n))\right\} & \\ \times\, \pi^{-N_t(M-k)}|\boldsymbol{W}^{-1}|^{N_t} \exp\left\{-\sum_{n=1}^{N_t}\boldsymbol{z}_\nu'(n)\boldsymbol{W}^{-1}\boldsymbol{z}_\nu(n)\right\}. & \end{aligned} \tag{3.4}$$

The desired posterior distribution can then be written using Bayes' theorem as

$$\begin{aligned} p(\bar{\boldsymbol{A}}, \boldsymbol{\phi}, k, \boldsymbol{W}^{-1} | \boldsymbol{Z}_s, \boldsymbol{Z}_\nu) \propto & \\ & p(\boldsymbol{Z}_s, \boldsymbol{Z}_\nu | \bar{\boldsymbol{A}}, \boldsymbol{\phi}, k, \boldsymbol{W}^{-1}) \\ & \times p(\bar{\boldsymbol{A}} | \boldsymbol{\phi}, k, \boldsymbol{W}^{-1}) \cdot p(\boldsymbol{W}^{-1} | \boldsymbol{\phi}, k) \cdot p(\boldsymbol{\phi} | k) \cdot p(k). \end{aligned} \tag{3.5}$$

We now discuss the assignment of each of the prior distributions in eq. (3.5). These priors are chosen to be non-informative where possible. When convenient, we also choose the structural form of these distributions for their desirable conjugate properties. The prior distributions are described as follows.

- $\bar{\boldsymbol{A}}$ is assigned a non-informative prior distribution described as a Gaussian function with a large covariance matrix $\boldsymbol{D}$ (compared to $\boldsymbol{C}$), and zero mean. Thus,

$$p(\bar{\boldsymbol{A}} | \boldsymbol{\phi}_k, k, \boldsymbol{W}^{-1}) = \prod_{n=1}^{N_t} \mathrm{N}(\boldsymbol{0}, \boldsymbol{D}), \tag{3.6}$$

  where

$$\boldsymbol{D} = d^2 \boldsymbol{I}_k, \tag{3.7}$$

  and where $\boldsymbol{I}_k$ is the identity matrix of size k. The above assumes that the amplitudes of the projected signals are independent with the same large variance. We discuss the choice of the hyper-parameter d^2 in section 3.4.

- The prior distribution for $\boldsymbol{\phi}$ is chosen to be uniform

$$p(\boldsymbol{\phi}_k | k) = \mathrm{U}[0, 2\pi]^k.$$

- The prior on k is chosen to be Poisson with expectation Λ. Although this choice is not strictly non-informative, it results in a more efficient MCMC sampling procedure, and further, simulation results have shown that performance is robust to the choice of prior on k.

$$p(k) = \Lambda^k e^{-\Lambda} / k!. \tag{3.8}$$

- $\boldsymbol{W}^{-1}$: We use a non-informative multi-dimensional Jeffreys' prior (Jeffreys, 1961) for the unknown transformed noise covariance matrix.

We now apply this rule to obtain the prior distribution $p(\boldsymbol{W}^{-1} \mid \boldsymbol{\phi}, k)$. The log likelihood is given from eq. (3.4) as

$$\begin{aligned}\mathcal{L}(\tilde{\boldsymbol{A}}, \boldsymbol{\phi}, k, \boldsymbol{W}^{-1} \mid \boldsymbol{Z}_\nu) = & \\ & - N_t(M-k)\log(\pi) \\ & + N_t \log\left|\boldsymbol{W}^{-1}\right| \\ & - \sum_{n=1}^{N_t} \boldsymbol{z}'_\nu(n)\boldsymbol{W}^{-1}\boldsymbol{z}_\nu(n) + f(\boldsymbol{Z}_s(n)),\end{aligned} \tag{3.9}$$

where $f(\boldsymbol{Z}_s(n))$ is the log value of the Gaussian block in $\boldsymbol{z}_s(n)$ in eq. (3.4). The Fisher information matrix is then given by

$$\boldsymbol{J}_{W^{-1}} = -E\left[\nabla \cdot \nabla \mathcal{L}(\cdot)\right], \tag{3.10}$$

where the matrix operator ∇ is defined such that the (l,m)th element is given by

$$\nabla_{w^{lm}} = \frac{1}{2}\left(\frac{\partial}{\partial w_R^{lm}} - j\frac{\partial}{\partial w_I^{lm}}\right), \tag{3.11}$$

with w^{lm} being the (l,m)th element of $\boldsymbol{W}^{-1}$ having w_R^{lm} and w_I^{lm} as the real and imaginary parts. We note that $\boldsymbol{J}_{W^{-1}}$ is an $(M-k)^2 \times (M-k)^2$ matrix. By applying eq. (3.10) to eq. (3.9), it can be shown (Wong *et al.*, 1992) that

$$\det(\boldsymbol{J}_{W^{-1}}) = N_t \mid \boldsymbol{W}^{-1} \mid^{-2(M-k)},$$

so that when Jeffreys' rule is applied, the non-informative prior distribution of the transformed noise covariance matrix can be written as

$$p(\boldsymbol{W}^{-1} \mid \boldsymbol{\phi}, k) \propto \mid \boldsymbol{W}^{-1} \mid^{-(M-k)}. \tag{3.12}$$

The posterior distribution from eq. (3.5) is then

$$\begin{aligned} p(k,\boldsymbol{\phi},\boldsymbol{W}^{-1},\tilde{\boldsymbol{A}}|\boldsymbol{Z}_s,\boldsymbol{Z}_\nu) \propto & \\ & |\boldsymbol{C}^{-1}|^{N_t} e^{-\sum_{n=1}^{N_t}(\tilde{\boldsymbol{a}}(n)-\boldsymbol{z}_s(n))'\boldsymbol{C}^{-1}(\tilde{\boldsymbol{a}}(n)-\boldsymbol{z}_s(n))} \\ & \times \pi^{-kN_t}|\boldsymbol{D}^{-1}|^{N_t} e^{-\sum_{n=1}^{N_t}\tilde{\boldsymbol{a}}(n)'\boldsymbol{D}^{-1}\tilde{\boldsymbol{a}}(n)} \\ & \times |\boldsymbol{W}^{-1}|^{N_t-(M-k)} e^{-\sum_{n=1}^{N_t}\boldsymbol{z}_\nu'(n)\boldsymbol{W}^{-1}\boldsymbol{z}_\nu(n)} \frac{\Lambda^k}{k!(2\pi)^k}, \end{aligned} \tag{3.13}$$

where superfluous constants independent of $\boldsymbol{\phi}$ or k have been absorbed into the constant of proportionality.

Remark 1 *The hyper-parameter Λ could either be estimated ahead of time, or could be considered known as part of the design parameters. It could also be treated as a random variable, with its own prior distribution, to make the algorithm more robust. However, this complicates the problem and does not prove necessary.* □

We now proceed to integrate out the nuisance parameters analytically. Since we assume the matrix $\boldsymbol{D}$ is large enough to dominate the matrix $\boldsymbol{C}$ (for a non-informative prior distribution), the term $\tilde{\boldsymbol{a}}'\boldsymbol{D}^{-1}\tilde{\boldsymbol{a}}$ in the second exponential above is small in comparison to the term $\tilde{\boldsymbol{a}}'\boldsymbol{C}^{-1}\tilde{\boldsymbol{a}}$ in the first exponential. This simplifies the posterior to

$$\begin{aligned} p(k,\boldsymbol{\phi},\boldsymbol{W}^{-1},\tilde{\boldsymbol{A}}|\boldsymbol{Z}_s,\boldsymbol{Z}_\nu) \propto & \\ & \pi^{-kN_t}|\boldsymbol{C}^{-1}|^{N_t} e^{-\sum_{n=1}^{N_t}(\tilde{\boldsymbol{a}}(n)-\boldsymbol{z}_s(n))'\boldsymbol{C}^{-1}(\tilde{\boldsymbol{a}}(n)-\boldsymbol{z}_s(n))} \\ & \times |\boldsymbol{W}^{-1}|^{N_t-(M-k)} e^{-\sum_{n=1}^{N_t}\boldsymbol{z}_\nu'(n)\boldsymbol{W}^{-1}\boldsymbol{z}_\nu(n)} |\boldsymbol{D}^{-1}|^{N_t} \frac{\Lambda^k}{k!(2\pi)^k}. \end{aligned}$$

The first nuisance parameter (the amplitude of the sources $\tilde{\boldsymbol{a}}(n)$) can now be easily integrated out by comparison, since it only appears in an isolated Gaussian distribution. The posterior distribution can then be simplified to

$$\begin{aligned} p(k,\boldsymbol{\phi},\boldsymbol{W}^{-1}|\boldsymbol{Z}_\nu) &\propto |\boldsymbol{W}^{-1}|^{N_t-(M-k)} \exp\left\{-\sum_{n=1}^{N_t}\boldsymbol{z}_\nu'(n)\boldsymbol{W}^{-1}\boldsymbol{z}_\nu(n)\right\} \frac{\Lambda^k}{k!(2\pi)^k d^{2kN_t}} \\ &= \left[|\boldsymbol{W}^{-1}|^{N_t-(M-k)} \exp\left\{-\mathrm{tr}N_t\hat{\boldsymbol{W}}\boldsymbol{W}^{-1}\right\}\right] \frac{\Lambda^k}{k!(2\pi)^k (d^2)^{kN_t}}, \end{aligned}$$

where tr(·) is the trace operator, $N_t\hat{\boldsymbol{W}}((\boldsymbol{\phi},k)) \triangleq \sum_{n=1}^{N} \boldsymbol{z}_\nu(n)\boldsymbol{z}_\nu'(n)$, and we have used eq. (3.7) for $\boldsymbol{D}$. The noise covariance matrix can now be integrated out by comparing the term inside the square brackets above to a complex Wishart distribution on $\boldsymbol{W}^{-1}$, with order $p = M - k$ and parameter $N_t\hat{\boldsymbol{W}}$, i.e. the conventional roles of $\boldsymbol{W}$ and $\hat{\boldsymbol{W}}$ are reversed. This term integrates to $I(\hat{\boldsymbol{W}})$, given as (Goodman, 1963)

$$I(\hat{\boldsymbol{W}}) = \pi^{\frac{1}{2}(M-k)(M-k-1)} \prod_{i=1}^{M-k} \Gamma(N_t - i + 1) \left| N_t\hat{\boldsymbol{W}} \right|^{-N_t},$$

where Γ is the Gamma function. The posterior distribution, after carrying out the integration and some manipulation is then

$$p(k,\boldsymbol{\phi}|\boldsymbol{Z}_\nu) \propto \frac{\pi^{\frac{1}{2}(M-k)(M-k-1)} \prod_{i=1}^{M-k} \Gamma(N_t - i + 1)}{(2\pi/\Lambda)^k k!(d^2)^{kN_t}} \left| N_t\hat{\boldsymbol{W}} \right|^{-N_t}. \tag{3.14}$$

The objective is to estimate the parameters of this highly non-linear function, as the Maximum A Posteriori (MAP) estimates

$$\{\hat{k},\hat{\boldsymbol{\phi}}\} = \arg \max_{k,\boldsymbol{\phi}\in\mathcal{K}\times[-\pi,\pi]^k} p(k,\boldsymbol{\phi}|\mathbf{Z}_v). \tag{3.15}$$

3.3 The reversible jump MCMC algorithm

We now propose the reversible jump MCMC algorithm (Green, 1995) to perform the Bayesian computation in extracting the parameters of interest from the posterior distribution eq. (3.14).

In the reversible jump algorithm, candidate samples are chosen from a *set* of proposal distributions, which are randomly accepted according to an acceptance ratio that ensures reversibility, and therefore the invariance of the Markov chain with respect to the desired posterior distribution. Here, we choose our set of proposal distributions to correspond to the following set of moves, as described in Chapter 2:

1. the *birth* move, valid for $k < M$. Here, a new incident plane wave is proposed at random on $(-\pi,\pi]$.

2. the *death* move, valid for $k > 0$. Here, a randomly chosen incident plane wave is removed.

3. the *update* move. Here, the parameters describing the incident plane wave are updated for a fixed value of k.

The probabilities for choosing each move are denoted u_k, b_k and d_k, respectively, such that $u_k + b_k + d_k = 1$ for all k. In accordance with (Green, 1995), we choose

$$b_k = c\min\{\frac{p(k+1)}{p(k)}, 1\}, \quad d_{k+1} = c\min\{\frac{p(k)}{p(k+1)}, 1\}, \tag{3.16}$$

where $p(\cdot)$ is the prior probability of the kth model according to eq. (3.8), and c is the tuning parameter for 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 (Green, 1995). The overall description of the reversible jump MCMC algorithm is determined by the choice of move at each iteration. This description is summarised as follows.

Reversible Jump MCMC

1. Initialization: set $\Phi^{(0)} = (\phi^{(0)}, k^{(0)})$

2. Iteration i,

   - Sample $u \sim U_{[0,1]}$
   - if $(u < b_{k^{(i)}})$ then execute a "birth move" (see section 3.3.2)
     - else if $(u < b_{k^{(i)}} + d_{k^{(i)}})$ then execute a "death move" (see section 3.3.2)
     - else, execute an update move (see section 3.3.1) .

3. $i \leftarrow i + 1$, goto step 2

3.3.1 Update move

Here, we assume that the current state of the algorithm is in $(\Phi_k, \{k\})$. When the update move is selected, the algorithm samples only on the space of Φ_k, for k fixed. The acceptance ratio $r = r_{update}$ from eqs. (2.8) and (3.14) for the update move, in the case of coloured noise, is therefore

$$r_{update}(\phi_k^\star, k, \phi_k, k) = \frac{\left|N_t \hat{W}(\phi^\star, k)\right|^{-N_t}}{\left|N_t \hat{W}(\phi, k)\right|^{-N_t}}, \tag{3.17}$$

$$\alpha_{update} = \min[r_{update}, 1]. \tag{3.18}$$

The candidate $\phi^\star$ is then accepted as the current state ($\phi_k^{(i+1)} = \phi_k^\star$), with probability α_{update}. The performance of the proposed method is enhanced by selecting randomly between two types of proposal distributions for the update case: one involves a global exploration of the parameter space, while the other involves a local exploration.

The global exploration is realized by proposing candidates according to their prior distributions, which is the whole horizon. The local exploration uses a Gaussian distribution, with mean ϕ_k and known covariance Σ_ϕ, around the present values with a reasonable variance as candidate distribution. Therefore, the candidates are located in the vicinity of the present values. This strategy offers the chain more freedom to mix, resulting in better characteristics of the samples.

Update Move (hybrid M-H)

Sample $u_1 \sim U_{[0,1]}$

- if $u_1 < 0.5$, Propose a global exploration

$$\phi_k^\star \sim U_{[0,2\pi]^k}$$

- else, propose a local exploration

$$\phi_k^\star \sim \mathrm{N}(\phi_k^{(i)}, \Sigma_\phi)$$

- Evaluate α_{update} with eq. (3.18).
- Sample $u \sim U_{[0,1]}$.
- if $(u \leq \alpha_{update})$ then the state of the Markov Chain becomes $(\boldsymbol{\phi}_k^\star)$, else it remains at $(\boldsymbol{\phi}_k^{(i)})$.

■

3.3.2 Birth and death moves

In the death move case, we assume the current state is in $(\boldsymbol{\Phi}_{k+1}, \{k+1\})$ and we wish to determine whether the state is in $(\boldsymbol{\Phi}_k, \{k\})$ at the next iteration. This involves the removal of an incident signal, which is chosen randomly amongst the $(k+1)$ existing incident signals. The proposal distribution $q(\boldsymbol{\phi}_k^\star, k|\boldsymbol{\phi}_{k+1}, k+1)$ for the death move is therefore chosen as

$$q(\boldsymbol{\phi}_k^\star, k|\boldsymbol{\phi}_{k+1}, k+1) = p(k) \div \binom{k+1}{1} \propto \frac{\Lambda^k}{k!}\frac{1}{(k+1)}. \tag{3.19}$$

Similarly, in the birth move case, we assume the current state is in $(\boldsymbol{\Phi}_k, \{k\})$ and we wish to determine whether the next state is in $(\boldsymbol{\Phi}_{k+1}, \{k+1\})$. This involves the addition of a new incident signal, which is proposed uniformly over $(0, 2\pi]$. The proposal distribution $q(\boldsymbol{\phi}_{k+1}^\star, (k+1)|\boldsymbol{\phi}_k, k)$ for the birth move is therefore

$$q(\boldsymbol{\phi}_{k+1}^\star, (k+1)|\boldsymbol{\phi}_k, k) = p(k+1)\frac{1}{2\pi} \propto \frac{\Lambda^{k+1}}{(k+1)!}\frac{1}{2\pi}.$$

For the death move, a candidate state $(\boldsymbol{\phi}_k^\star, k)$ is then sampled from eq. (3.19). The acceptance ratio $r = r_{death}$ from eqs. (2.10) and (3.14) is then given as

$$\begin{aligned} r_{death}(\boldsymbol{\phi}_k^\star, k, \boldsymbol{\phi}_{k+1}, k+1) &= \frac{\left|N_t \hat{\boldsymbol{W}}(\boldsymbol{\phi}^\star, k)\right|^{-N_t}}{\left|N_t \hat{\boldsymbol{W}}(\boldsymbol{\phi}, k+1)\right|^{-N_t}} \\ &\times \pi^{M-k-1}\Gamma(N_t - M + k + 1)(k+1)d^{2N_t}. \end{aligned}$$

The quantity α_{death} is then defined according to

$$\alpha_{death} = \min[r_{death}, 1]. \tag{3.20}$$

The acceptance ratio $r = r_{birth}$ for the birth move in the case of coloured noise can be verified to be

$$\alpha_{birth} = \min[1, \frac{1}{r_{death}}]. \tag{3.21}$$

The following block describes the algorithm for the birth move.

Birth Move

- Propose a new direction of arrival, uniform on $(0, 2\pi)$
  $\phi_{k+1}^{(i+1)} = [\phi_k^{(i)}, \quad U_{[0,2\pi]}]$
- Evaluate α_{birth} with eq. (3.21).
- Sample $u \sim U_{[0,1]}$.
- if $(u \leq \alpha_{birth})$ then the state of the Markov Chain becomes $(\phi_{k+1}^{(i+1)}, k+1)$, else it remains at $(\phi_k^{(i)}, k)$.

The description for the death move is similar, with appropriate modifications.

Death Move

- Select at random one direction of arrival, n_k
  $\phi_k^{(i+1)} = [\phi_{1:(n_k-1)}^{(i)}, \phi_{n_k+1:(k+1)}^{(i)}]$
- Evaluate α_{death} with eq. (3.20).
- Sample $u \sim U_{[0,1]}$.
- if $(u \leq \alpha_{death})$ then the state of the Markov Chain becomes $(\phi_k^{(i+1)}, k)$, else it remains at $(\phi_{k+1}^{(i)}, k+1)$.

Remark 2 *At this point, it is enlighting to take the logarithm of the previously obtained posterior distribution eq. (3.14),*

$$\log p(k,\phi_k|\boldsymbol{Z}_\nu) \propto -N_t \log(\left|N_t\hat{\boldsymbol{W}}\right|) + kN_t\log(d^2) + k\log(\Lambda/2\pi) + \log(k!) + \frac{1}{2}(M-k)(M-k-1)\log(\pi) + \log(\prod_{i=1}^{M-k}\Gamma(N_t-i+1)).$$

In this form, the similarities with previous model selection criteria such as AIC, MDL, D-MAP of Djuric (Djuric, 1996) or W-MDL of Wax (Wax, 1992) are made apparent. The first term represents the likelihood term, while the remaining ones jointly constitute a "penalty term", which is a result of the use of prior distributions for $\tilde{a}$, ϕ, and k. □

3.4 Model order determination

In this section, we discuss conditions which must apply on the hyper-parameter d^2 in eqs. (3.6) and (3.7) for consistent determination of model order.

The marginal posterior distribution for k is given as

$$p(k|\boldsymbol{Z}_\nu) \propto \int_{\boldsymbol{\Phi}_k} p(k,\phi|\boldsymbol{Z}_\nu)d\phi. \tag{3.22}$$

Let the eigenvalues of $N_t\hat{\boldsymbol{W}}(\phi,k)$ in eq. (3.14) at $\phi=\tilde{\phi}$ [1] be given as $\lambda_1,\lambda_2,\ldots,\lambda_{M-k}$ arranged in *ascending*, rather than the usual *descending* order. For the case where N_t is large, for moderate values of SNR, the joint posterior distribution $p(\phi,k|\boldsymbol{Z}_\nu)$ concentrates around the true value $\phi=\tilde{\phi}$. Thus eq. (3.22) can be written as

$$p(k|\boldsymbol{Z}_\nu) \propto \left[\frac{\pi^{\frac{1}{2}(M-k)(M-k-1)}}{(2\pi/\Lambda)^k k!}\right]\left[\frac{\prod_{i=1}^{M-k}\Gamma(N_t-i+1)}{d^{2kN_t}}\left(\prod_{i=1}^{M-k}\lambda_i\right)^{-N_t}\right] \tag{3.23}$$

where $\propto$ indicates "approximately proportional to".

Let us define the event E_i as the declaration of a model order in error by i signals; i.e., E_i occurs when we declare $\hat{k}=k_o+i$ or $\hat{k}=k_o-i$. In the following analysis, we assume

[1] Since k is variable in this context but $\tilde{\phi}$ is a constant with fixed dimension, there is an apparent conflict with the dimension of $\tilde{\phi}$. This can be resolved by adding or removing appropriate elements from $\tilde{\phi}$ as required.

$P(E_1) > P(E_2) > \ldots > P(E_{M-1})$. This implies $p(k|\boldsymbol{Z}_\nu)$ is unimodal in k. From this assumption, a necessary and sufficient condition for consistent detection of model order is therefore

$$\lim_{N_t \to \infty} \frac{p(k_o + 1|\boldsymbol{Z}_\nu)}{p(k_o|\boldsymbol{Z}_\nu)} \to 0, \tag{3.24}$$

and

$$\lim_{N_t \to \infty} \frac{p(k_o - 1|\boldsymbol{Z}_\nu)}{p(k_o|\boldsymbol{Z}_\nu)} \to 0. \tag{3.25}$$

Using eq. (3.23) we can write

$$\begin{aligned} \frac{p(k_o + 1|\boldsymbol{Z}_\nu)}{p(k_o|\boldsymbol{Z}_\nu)} &\approx \frac{\left[\frac{\pi^{\frac{1}{2}(M-k_o-1)(M-k_o-2)}}{(2\pi/\Lambda)^{k_o+1}(k_o+1)!}\right]\left[\frac{\prod_{i=1}^{M-k_o-1}\Gamma(N_t-i+1)}{(d)^{2N_t(k_o+1)}}\left(\prod_{i=1}^{M-k_o-1}\lambda_i\right)^{-N_t}\right]}{\left[\frac{\pi^{\frac{1}{2}(M-k_o)(M-k_o-1)}}{(2\pi/\Lambda)^{k_o}k_o!}\right]\left[\frac{\prod_{i=1}^{M-k_o}\Gamma(N_t-i+1)}{(d)^{2k_oN_t}}\left(\prod_{i=1}^{M-k_o}\lambda_i\right)^{-N_t}\right]} \\ &= \frac{\pi^{-(M-k_o-1)}(\lambda_{M-k_o})^{N_t}(\frac{\Lambda}{2\pi})}{\Gamma(N_t - M + k_o + 1)d^{2N_t}(k_o+1)}, \end{aligned} \tag{3.26}$$

where we have used the property $\Gamma(n+1) = (n+1)\Gamma(n)$ and that the first $(M - k_o - 1)$ smallest eigenvalues are common to both the numerator and to the denominator. Since for $k = k_o$ and $\phi = \tilde{\phi}$, and large N_t, we have $\hat{\boldsymbol{W}} \to \boldsymbol{U}'_\nu \boldsymbol{\Sigma}_\nu \boldsymbol{U}_\nu \approx \boldsymbol{U}'_\nu \boldsymbol{R}_{yy} \boldsymbol{U}_\nu$, where $\boldsymbol{R}_{yy} \triangleq \mathrm{E}\{\boldsymbol{y}\boldsymbol{y}'\}$. Since the eigenvalues of $\boldsymbol{U}'_\nu \boldsymbol{R}_{yy} \boldsymbol{U}_\nu$ are the same as the nonzero eigenvalues of $\boldsymbol{P}_{\mathcal{N}} \boldsymbol{R}_{yy} \boldsymbol{P}_{\mathcal{N}}$, where $\boldsymbol{P}_{\mathcal{N}}$ is the projector onto $\mathcal{N}$, the eigenvalue λ_{M-k_o} above is the largest eigenvalue of the covariance matrix formed from the data projected onto $\mathcal{N}$.

Using a similar development we also have

$$\frac{p(k_o - 1|\boldsymbol{Z}_\nu)}{p(k_o|\boldsymbol{Z}_\nu)} \approx \frac{\Gamma(N_t - M + k_o)d^{2N_t}(k_o)}{\pi^{-(M-k_o)}(\lambda_{M-k_o+1})^{N_t}(\frac{\Lambda}{2\pi})}. \tag{3.27}$$

In this case, $k = k_o + 1$, and since the estimated dimension of the noise subspace is too large, it incorporates part of $\mathcal{S}$. The quantity λ_{M-k_o+1} may thus be associated with the smallest signal eigenvalue of $\boldsymbol{R}_{yy}$ projected into $\mathcal{S}$, which is the same as the smallest signal eigenvalue of $\boldsymbol{R}_{yy}$.

We can now evaluate the limits of eqs. (3.26) and (3.27). To do so, we require the *Stirling approximation* (Abramowitz and Stegun, 1965) to the Γ-function, which is valid for

large values of the argument

$$\Gamma(x) \approx \sqrt{2\pi} e^{-x} x^{(x-\frac{1}{2})}. \tag{3.28}$$

We also note from the definition of $N\hat{\boldsymbol{W}}$ that the eigenvalues λ_i are directly proportional to N_t, and thus they can be written as

$$\lambda_i = N_t \tilde{\lambda}_i, \qquad i = 1, \ldots, M-k, \tag{3.29}$$

where $\tilde{\lambda}_i$ is the normalized version of λ_i. Substituting eqs. (3.28) and (3.29) into eq. (3.26) we find that eq. (3.24) is satisfied if

$$\tilde{\lambda}_{M-k_o} < \frac{d^2}{e}. \tag{3.30}$$

In a similar way, using eq. (3.27), eq. (3.25) is satisfied if

$$\tilde{\lambda}_{M-k_o+1} > \frac{d^2}{e}. \tag{3.31}$$

Therefore, by combining eq.(3.30) and eq. (3.31), we see that detection of model order is consistent if the hyper-parameter d^2 is chosen so that

$$\tilde{\lambda}_{M-k_o} < \frac{d^2}{e} < \tilde{\lambda}_{M-k_o+1}, \tag{3.32}$$

i.e., the quantity $\frac{d^2}{e}$ must lie in the gap between the largest projected normalized noise eigenvalue and the smallest normalized signal eigenvalue.

Strictly speaking, this procedure for determining d^2 cannot be used for consistent detection because eq. (3.32) depends on the unknown k_o. However, it should be possible in the practical scenario to propose an *ad hoc* scheme to approximate eq. (3.32). For example, it is usually possible to form an estimate of $\boldsymbol{\Sigma}_\nu$ during periods where it is known with reasonable certainty that there are no signals present. The largest eigenvalue of this matrix could then be used as an upper bound on $\tilde{\lambda}_{M-k_o}$. If some *a priori* knowledge on the DOAs were available, then an estimate of $\boldsymbol{U}_\nu$ can be evaluated and a better estimate of $\tilde{\lambda}_{M-k_o}$ could be determined. In either case, d^2 could be given as an empirically determined constant times this eigenvalue estimate.

3.5 Simulation results

In this section, we present simulation results of the sampling scheme developed in sections 3.2 and 3.3. The estimates are obtained as the MAP estimators of the histogram of the samples, as defined in eq. (3.15).

The spatially coloured noise is generated with an AR process of order 2 with roots $0.95e^{-j1.07\pi}$ and $0.95e^{-j0.88\pi}$ as in (Cho and Djuric, 1995), with excitation from complex white noise samples of equal variance $\sigma_w^2/2$ for both the real and imaginary parts. Figure 3.1 shows the directional spectrum of the spatially coloured noise.

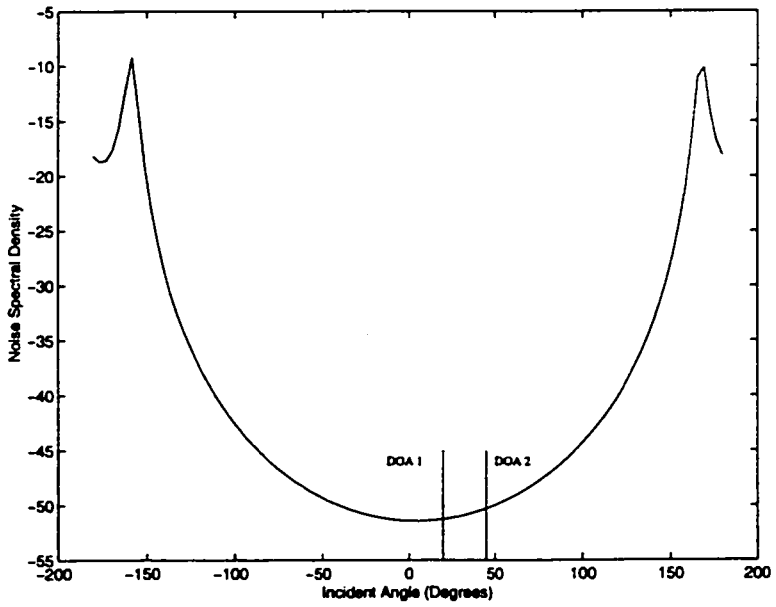

Figure 3.1: Spectrum of the spatially coloured noise used in simulations. The directions of arrival for the first scenario are indicated on the figure.

Two scenarios are presented. First, the directions of arrival to estimate are located in a somewhat favourable zone of the noise spectrum, as shown on Figure 3.1. A second scenario will locate the DOAs in a steeper zone of the spectrum, around 120°.

3.5.1 First scenario

The characteristics of the signals and parameters used to obtain the first simulation results are summarized in Table 3.1. This scenario is very difficult, using only a small number of snapshots at low SNR and presenting two sources well within a beamwidth of the receiver array. The hyperparameter of d^2 was assigned the value of 1000, in accordance to the

Amplitude	DOA	Sensors	Snapshots	SNR	MCMC iterations
10	20°	5	30	2 dB	10000
10	45°				

Table 3.1: Characteristics of the signals and parameters for simulation #1 .

criterion eq. (3.32). The Figures 3.2 and 3.3 show typical results for 10000 iterations (after a sufficient burn-in period of 5000 iterations, based on the observation of the chain behaviour) of the Reversible Jump Sampler with $N_t = 30$ observations of a circular array made of 5 equi-spaced sensors (with a radius of $0.102m$ at 1.86GHz), when the SNR is 2 dB. The SNR is defined as

$$SNR = \frac{a_1^2}{2\sigma_w^2}.$$

It is interesting to observe the evolution of the instantaneous model probabilities (top portion of Figure 3.2) and how they reach an equilibrium value. The bottom portion of Figure 3.2 shows the posterior histogram of the estimated number of sources after burn-in. Figure 3.3 shows the histograms of $\hat{P}(\phi|k_o = 2)$. Clearly, from visual inspection of the figures, we see the algorithm detected the right number of sources and has estimated their respective DOAs as 19.5° and 46.0°, which are close to the true values.

The next figure, Figure 3.4, shows the behavior of the algorithm when the hyperparameter Λ is initialized at $\Lambda = 5$. This represents the scenario where no a priori or wrong information is available. As expected, the algorithm takes a little longer to converge to the proper model order, but the correct a posteriori estimate is obtained. This supports the claim that the algorithm is robust to this hyperparameter.

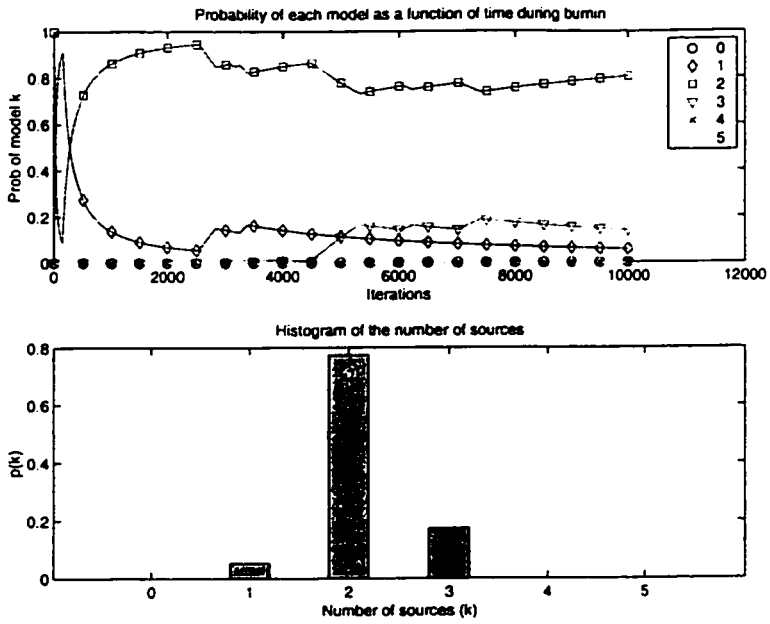

Figure 3.2: Coloured noise with unknown covariance matrix, Simulation #1: Instantaneous estimate of the model probability (top half); Histogram of the number of sources after burn-in (bottom half).

3.5.2 Second scenario

The characteristics of the signals and parameters used to obtain the second simulation results are summarized in Table 3.2. For this scenario, the sources are located in a region where the spectrum of the coloured noise is steeper. The hyperparameter d^2 was assigned

Amplitude	DOA	Sensors	Snapshots	SNR	MCMC iterations
10	110°	5	30	2 dB	10000
10	130°				

Table 3.2: Characteristics of the signals and parameters for simulation #2 .

the value of 140, in accordance to the criterion eq. (3.32). The Figures 3.5 and 3.6 show typical results for the same type of simulation, e.g. 10000 iterations (after a sufficient burn-in period of 5000 iterations, based on the observation of the chain behaviour) of the

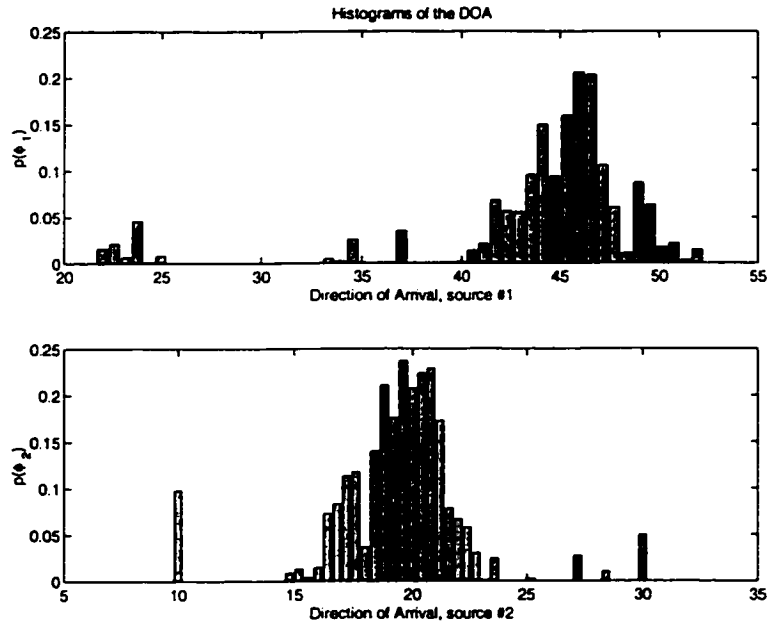

Figure 3.3: Coloured noise with unknown covariance matrix, Simulation #1: Histogram of the DOAs after burn-in: Source 1 (top) and Source 2 (bottom).

Reversible Jump Sampler with $N_t = 30$ observations.

Again, the Figure 3.5 shows the posterior histogram of the estimated number of sources after burn-in. Figure 3.6 shows the histograms of $\hat{P}(\boldsymbol{\phi}|k_o = 2)$. Clearly, from visual inspection of the figures, we see the algorithm detected the right number of sources and has estimated their respective DOAs as 133.0° and 110.0°, which are close to the true values.

3.5.3 Performance of the method

In order to assess the performance of the algorithm for joint detection/estimation in terms of variance of the estimates as a function of the SNR, the algorithm was applied to 50 Monte Carlo noise realizations of scenario #1, for a range of values of SNR, from −4dB to 16dB, with the other parameter values given as before in Table 3.1. Specifically, the value of d^2 was held at the value 1000, which was verified to satisfy (3.32) over the entire range of SNR values considered.

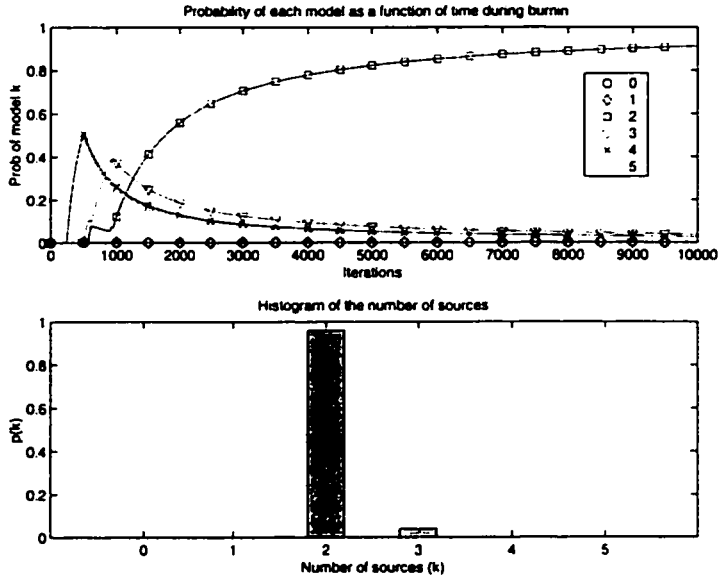

Figure 3.4: The hyperparameter Λ was set at 5. Instantaneous estimate of the model probability (top half); Histogram of the number of sources after burn-in (bottom half).

The results are compared with the algorithm developed by Wax (Wax, 1992) and implemented by the alternating projection (Ziskind and Wax, 1988) algorithm, where a steepest descent gradient method was used for the one dimensional optimization. For SNR levels lower than -2 dB, the detection part of both the Wax algorithm and the MCMC algorithm is not reliable. The Wax method was initialized at the *true* value of the DOA parameters. Without this procedure, the performance curves for the Wax method as shown in Figure 3.7 would be severely degraded, due to convergence of the algorithm to local minima. In contrast, initial DOA values for the MCMC method were assigned random values uniformly distributed over $[0, 2\pi]$, for large number of iterations of the MCMC algorithm. The probability of detection, and the variances of the estimated DOAs are shown in Table 3.3 and Figures 3.7 and 3.8 respectively for both methods. The Cramér-Rao lower bounds were evaluated from the results in (Gershman *et al.*, 2001; Ye and DeGroat, 1995).

With reference to Figure 3.7, it may be observed that the performance of the MCMC

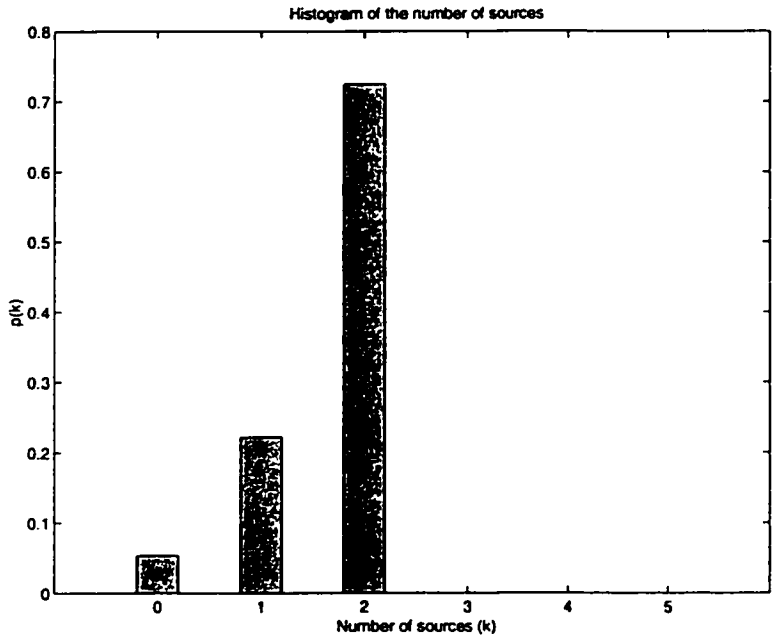

Figure 3.5: Coloured noise with unknown covariance matrix, Simulation #2: Histogram of the number of sources after burn-in.

method is comparable to that of the Wax method above the threshold value of approximately -2 dB, but at a much higher computational cost. However, the computational cost is seen as a relatively insignificant advantage compared to the problems of initialisation. The colored noise method proposed clearly outperforms the classical maximum-likelihood approach developed under white noise assumption.

For reasons of clarity, the curves for the signal at DOA_2 are omitted of Figure 3.7. The omitted curve behaves in a similar way. The MCMC method achieves comparable performance levels for both DOAs, close to the Cramér-Rao lower bound. It may be observed that the slight degradation in performance of the MCMC may be caused by neglecting the off-diagonal terms of the covariance matrix in the development of the posterior distribution eq. (3.4).

Figure 3.8 shows the same performance curves, but in comparison with their respective Cramér-Rao bound curves. We note that the performance of the signal associated with

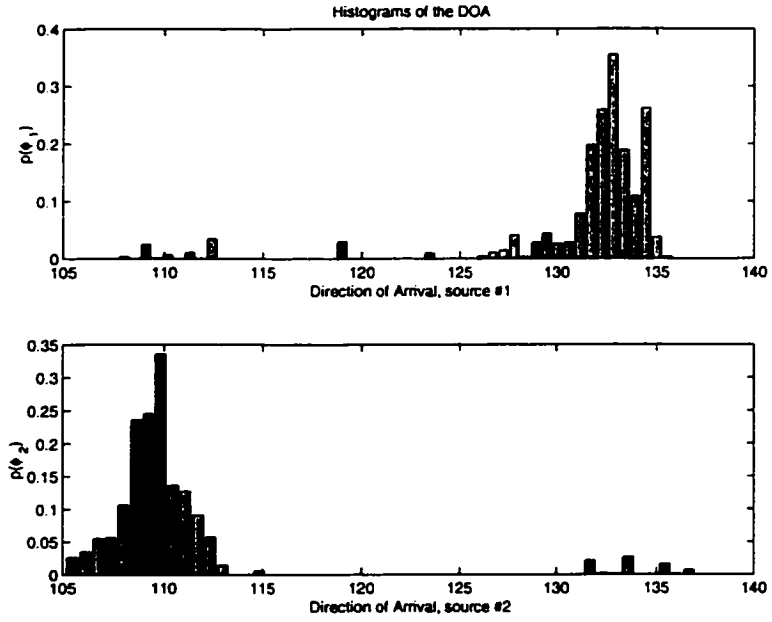

Figure 3.6: Coloured noise with unknown covariance matrix, Simulation #2: Histogram of the DOAs after burn-in: Source 1 (top) and Source 2 (bottom).

DOA_2 is degraded over that for DOA_1. This is because, with reference to Figure 3.1, the signal with DOA_2 receives a higher noise level than that for DOA_1. For this reason, Figure 3.7 shows variances for both DOA values. It is shown in (Wong *et al.*, 1992) that the variances of the DOA estimates do not approach zero as $N_t \to \infty$. However, it is seen from Figures 3.7 and 3.8 that the variances approach the Cramér-Rao bound closely.

Both detection procedures provide similar performance, as indicated in Table 3.3. Further simulation results, as shown in Figure 3.9, demonstrate that the probability of an error in detection of the model order diminishes towards zero with increasing N_t, thus verifying the development of section 3.4.

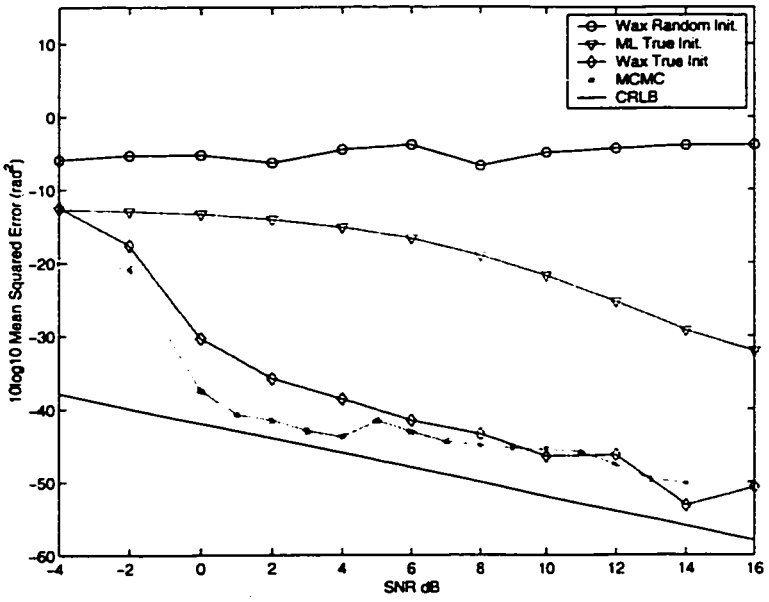

Figure 3.7: Coloured noise with unknown covariance matrix: Mean Squared Error. 50 Monte Carlo runs of 20000 MCMC iterations and W-MDL of Wax. The Wax method was initialized to the true values of the parameters, whereas the MCMC method was initialized arbitrarily. The curve for the second DOA angle is omitted for reasons of clarity.

3.6 Application to real-life measurements

To demonstrate the robustness of the discussed methods, we apply the method developed in this chapter to a typical data set collected on campus through the experimental measurement campaign, discussed in section 1.3. We collected 60 snapshots, from a position, shown in Figure 3.10, where the multipath characteristics have been frequently observed to be 2 rays well separated in angle.

Figure 3.11 shows typical results for 75000 iterations of the Reversible Jump Sampler with 60 observations of the array output of the first leg of the ground scenario. Two rays impinge the array with approximated angles of arrival of 30° and −80°. The transmitter was moving slowly along the parking lot known as Annex 3, from behind the teacher's college toward Main street. The array was located on the front lawn of the CRL. The direct path shows a DOA of about $20° - 30°$ and the first multipath arrives at the array at an angle of

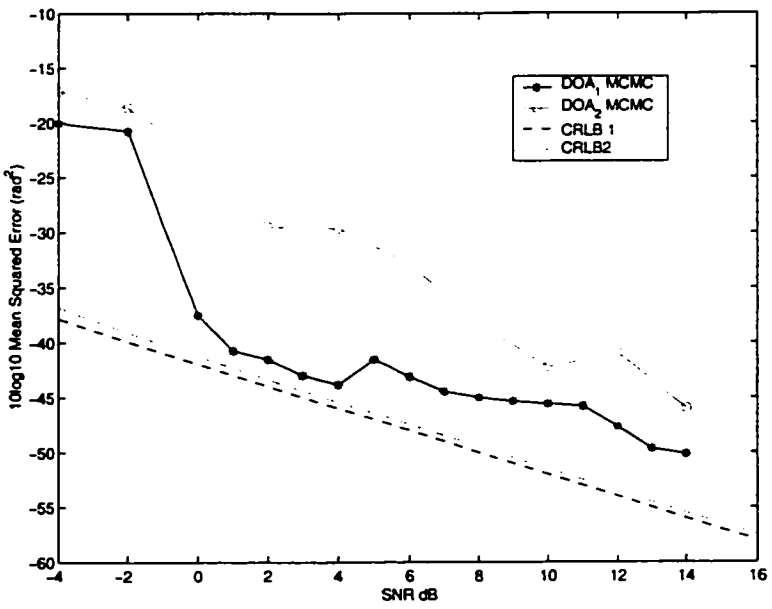

Figure 3.8: Coloured noise with unknown covariance matrix: mean squared error versus the CRLB.

about $-80°$, after bouncing off the JHE building. The expected signal to noise ratio is at least 20dB, but the noise is highly correlated as there is a lot of interference. This data set is adequate to test the robustness of the method.

In reference to the Figure 3.11, one can notice that the posterior distribution of the direction of arrival of the line of sight is wider. As the transmitter was moving slowly across the horizon of the array, the direction of arrival of the line of sight ray varies in time, which violates the assumptions of the algorithm. During the collection of the 60 snapshots, the transmitter travelled about $100m$, which changed the direction of arrival from the initial value of $20°$ slowly toward $30°$. On the other hand, as the transmitter was walking directly away from the JHE building, the direction of arrival of the multipath ray seen from the array remained constant, thus the narrower posterior distribution.

			SNR (dB)						
			-4	-3	-2	-1	0	1	2
M		1	6	14	28	26	40	18	30
C		2	64	82	72	74	60	82	70
M	k	3	28	4					
C		4	2						
		5							
W		1	2						2
M		2	74	82	66	78	72	78	78
D	k	3	18	10	20	12	20	20	10
L		4	4	8	12	2	8	2	6
		5	2		2	8			4

Table 3.3: Probability of detection (in %) VS SNR

3.7 Conclusion

A new application of the Reversible Jump MCMC method was developed and presented for the problem of joint detection/estimation of sources impinging on an array of sensors in spatially unknown coloured noise with arbitrary covariance.

This method is based on the formulation of a posterior density function which has all the nuisance parameters integrated out. Consistent detection has been verified, for values of d^2 in the range given by eq. (3.32). Simulation results support the effectiveness of the method, and demonstrate reliable detection of the number of sources and estimation of their directions of arrival in coloured noise with a single array. This is reinforced with the successful application to real data.

Although MCMC approaches are computationally intensive, a significant advantage as demonstrated in this work is that they provide the global solution with robustness to the initial guess of the parameter values.

In comparison, classical approaches have been shown to suffer degradations in DOA estimation performance due to local solutions, unless *a priori* knowledge of the DOA estimates is available.

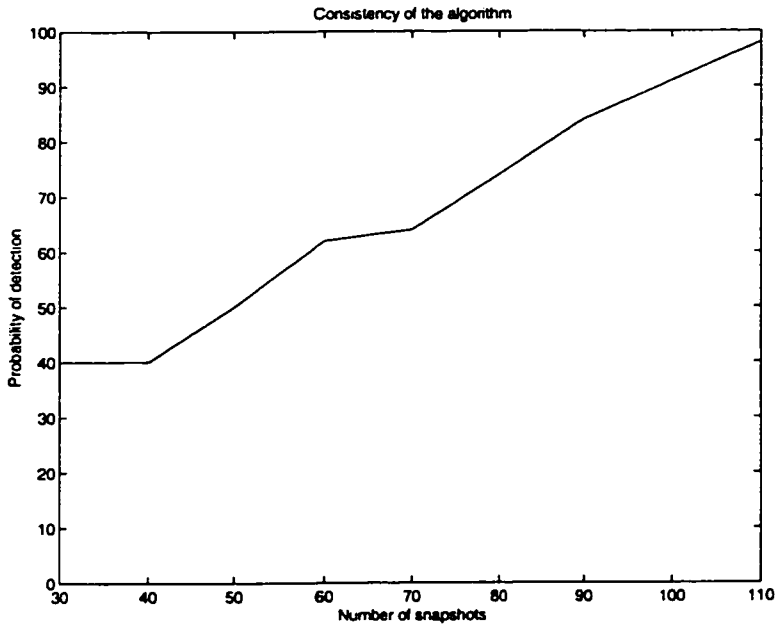

Figure 3.9: Probability of detection as a function of number of observations (2dB SNR).

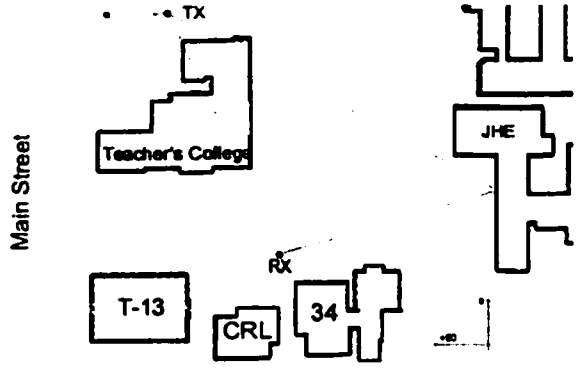

Figure 3.10: Layout of the setup for the first leg of the ground scenario.

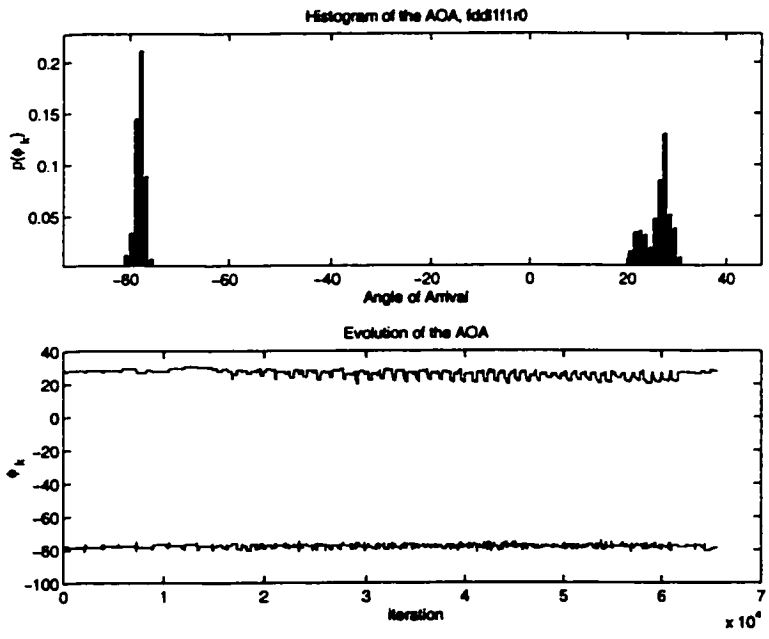

Figure 3.11: Measurements: Histogram of the DOA (top); Instantaneous estimates (bottom) (after the burn-in period)

Chapter 4

Wide Band Channel Characterization

In this chapter, we extend the treatment of the previous chapter to include estimation of the times of arrival (TOA) of multipath components onto arrays of sensors in unknown and arbitrary coloured noise. Thus, in this chapter, we consider the joint detection and estimation of the times of arrival and directions of arrival of the multipath components characterizing a multipath channel.

The developed method is applied to simulations and, more interestingly, with emphasis to real-life propagation measurements to show the performance and robustness of the approach.

4.1 Introduction

In recent years there has been a considerable amount of work done in the estimation of time delays (or times of arrival (TOAs)) and directions of arrival (DOAs) of the components of a multipath channel (Vanderveen *et al.*, 1997; Wax and Leshem, 1997; Wax, 1992; Larocque and Reilly, 2000b; Blanz *et al.*, 2000; Logothetis and Carlemalm, 2000). This is a central problem in many fields including radar, sonar, and wireless communications, for both

the indoor and outdoor scenarios. In the wireless case, knowledge of the statistics of the TOA/DOA profile of typical multipath channels is necessary to determine the complexity of proposed receivers during the design process, and to evaluate the performance of a given wireless configuration. The knowledge of these parameters could also be used to locate a user within a wireless network, thus providing a solution to the 911 problem.

In this chapter, we propose the use of the reversible jump Markov chain Monte Carlo (RJMCMC) technique (Green, 1995) for joint blind[1] estimation of the number of incident multipath components (model order), and the channel TOA/DOA information in the presence of noise with arbitrary covariance structure, using arrays of sensors. This approach is an extension to algorithm developed in Chapter 3 for the purpose of channel characterization.

The proposed solution differs from (Vanderveen *et al.*, 1997) in that the number of multipath components (reflections) is not assumed known in our case. In (Wax and Leshem, 1997), again, the number of reflections is assumed known, or is *a priori* estimated with the MDL approach. This limits the application of the proposed algorithm to white noise scenarios. Although another algorithm was developed by Wax (Wax, 1992) for the coloured noise case, it appears that it is very sensitive to initialization and local minima (Larocque and Reilly, 2000b). More recently, other methods have been proposed to address this difficult problem. In (Blanz *et al.*, 2000), the number of multipath components is assumed known. Lastly, the application of the SAGE algorithm by Logothetis and Carlemalm (2000) is limited to the white noise case. In many other papers, where the joint detection of the number of multipath components and the estimation of the parameters is addressed, only the directions of arrival are estimated, or only the temporal channel impulse response is estimated, but never jointly. It has been shown by Wax and Leshem (1997) that individual estimation of either the TOA or DOA parameters is suboptimal, and that significantly improved performance is obtained from *joint* TOA/DOA estimation, as is proposed in this chapter.

[1]By "blind" we mean no training sequence is required nor is the transmitted source sequence assumed known.

This chapter is organized in the following manner. In section 4.2, we describe the formulation of the problem, state the necessary assumptions and define the notation. In section 4.3, we develop the statistical model for the received signal in terms of the parameters of interest, where the undesired nuisance parameters are integrated out analytically. In section 4.4, we then describe the reversible jump MCMC procedure for joint determination of model order and TOA/DOA estimation. Results from simulated data are presented in section 4.5. In section 4.6, we describe the measurement technique used to gather the real propagation measurements and we apply the proposed method to these measurements. Conclusions are given in section 4.7.

4.2 Problem formulation

Consider an array composed of M sensors with arbitrary locations and arbitrary spatial response. The array is immersed in a multipath propagation environment where the corresponding channel consists of k_o discrete reflections with time delays (TOAs) $\tau = [\tau_1, \ldots, \tau_{k_o}]^T$, incident on the array from directions $\phi = [\phi_1, \ldots, \phi_{k_o}]^T$. We assume the sensors and the reflection points are coplanar and that the reflecting points are in the far field of the array, so that ϕ_k is the DOA of the k_{th} reflection. The signal vector $y(n) \in \mathcal{C}^M$ received by the array at time t can be written as

$$y(t) = \sum_{k=1}^{k_o} a_k(t - \tau_k)s(\phi_k) + \nu(t), \tag{4.1}$$

where

- $a_k(t)$ is the amplitude parameter associated with the k_{th} reflection
- τ_k is the delay (TOA) of the k_{th} reflection
- $s(\phi_k) \in \mathcal{C}^M$ is the array response vector from direction ϕ_k
- $\nu(t) \in \mathcal{C}^M$ is the noise vector at time t. ν is assumed to be uncorrelated with the signal, zero mean, distributed as $\mathcal{N}(\mathbf{0}, \Sigma_\nu)$, where $\Sigma_\nu \in \mathcal{C}^{M \times M}$ is unknown, positive definite and Hermitian, but otherwise arbitrary.

We choose to express the problem in a CDMA scenario[2], where the number of chips in one transmitted symbol is denoted by P, and the chip duration is ΔT. After sampling $\boldsymbol{y}(t)$ in eq. (4.1) in time, we may now express the received signal over the P chips comprising the n_{th} symbol in the form

$$\boldsymbol{Y}(n) = \boldsymbol{S}(\boldsymbol{\phi})\boldsymbol{A}(n)\boldsymbol{T}(\boldsymbol{\tau}) + \boldsymbol{V}(n), \qquad n = 1, \ldots, N_t, \tag{4.2}$$

where

- $\boldsymbol{Y}(\cdot)$ and $\boldsymbol{V}(\cdot)$ have P columns consisting of their lower case counterparts
- $\boldsymbol{S}(\boldsymbol{\phi}) \in \mathcal{C}^{M \times k_o} = [\boldsymbol{s}(\phi_1), \ldots, \boldsymbol{s}(\phi_{k_o})]$
- $\boldsymbol{A}(n) \in \mathcal{C}^{k_o \times k_o}$ is a diagonal matrix. The diagonal elements contain the signal amplitudes at the nth symbol
- $\boldsymbol{T}(\boldsymbol{\tau}) \in \mathcal{C}^{k_o \times P}$. Each row of $\boldsymbol{T}(\boldsymbol{\tau})$ consists of zero elements, except for a single one in the pth position. This element indicates that the relative delay of the corresponding scattering component is $p\Delta T$.
- N_t is the number of observed snapshots.

Given only observations $[\boldsymbol{Y}(1), \ldots, \boldsymbol{Y}(N_t)]$, our objective is to jointly estimate the number of scatterers, k_o, their directions of arrival, $\boldsymbol{\phi} \in [0, 2\pi]^{k_o}$ and their times of arrival within the resolution ΔT of one chip, represented by the vector of integers $\boldsymbol{\tau} \in [0, P]^{k_o}$.

Note that since the number of scatterers k_o is unknown, the dimensions of the parameters $\boldsymbol{\phi}$ and $\boldsymbol{\tau}$ are also unknown. We therefore denote them as $\boldsymbol{\phi}_k$ and $\boldsymbol{\tau}_k$ respectively, where k is the hypothesized number of signals.

We make the set of assumptions on the model:

- $\boldsymbol{S}(\boldsymbol{\phi})$ is full rank
- the noise $\boldsymbol{\nu}(\cdot)$ is *iid* between symbols

[2]This is done solely for ease of presentation and is not necessary for the development of the central ideas of this chapter. It allows us to consider a single user, isolated by its characteristic CDMA code.

- ϕ and τ are stationary over the entire N observations.
- the symbol amplitudes a_k are *iid* between symbols.

4.3 Development of the posterior distribution

The model described by eq. (4.2) can now be rearranged to a more familiar form using Kronecker algebra,

$$y(n) = Z(\tau, \phi)b(n) + \nu(n), \qquad n = 1, \ldots, N_t,$$

where

$$\begin{aligned} y(n) &= \text{vec}(Y(n)), \\ b(n) &= \text{vec}(A(n)), \\ Z(\tau, \phi) &= T^T(\tau) \otimes S(\phi), \end{aligned}$$

where vec$(\cdot)$ is the *vectorization* operator and $\otimes$ is the Kronecker matrix product. Furthermore, noting that the matrix $A(n)$ is diagonal, the vector $b(n)$ as defined only operates on a few columns of Z. Regrouping these useful columns into a new matrix $H(\tau, \phi)$ we have

$$y(n) = H(\tau, \phi)a(n) + \nu(n) \quad n = 1, \ldots, N_t,$$

where $a(n)$ holds the diagonal elements of $A(n)$, and the matrix $H(\tau, \phi)$ defines the space-time structure of the multipath. This form is more familiar and can now easily be analyzed in the Bayesian framework, as done in the previous chapter. The presentation in this chapter is therefore intentionally briefer.

Since the N_t snapshots are *iid*, the total likelihood function of all the data can be expressed in the form of

$$\begin{aligned} p(Y|\phi, \tau, A, \Sigma_\nu, k) &= \frac{1}{\pi^{N_t MP} |\Sigma_\nu|^{N_t}} \times \\ & e^{-\sum_{n=1}^{N_t} (y(n) - H(\phi,\tau)a(n))' \Sigma_\nu^{-1} (y(n) - H(\phi,\tau)a(n))}, \end{aligned}$$

where the matrices $\boldsymbol{Y}$ and $\boldsymbol{A}$ represent the sets $\{\boldsymbol{y}(n)\}_{n=1}^{N_t}$ and $\{\boldsymbol{a}(n)\}_{n=1}^{N_t}$, respectively. This likelihood function can be simplified by integrating out the undesired parameters $\boldsymbol{\Sigma}_\nu$, and $\boldsymbol{A}$. To proceed with this step, we first define an orthonormal matrix $\boldsymbol{U}(\boldsymbol{\phi}, \boldsymbol{\tau}, k) \in \mathcal{C}^{MP \times MP}$ for a hypothesized model order k, as in Chapter 3 (Wong *et al.*, 1992; Cho and Djuric, 1995),

$$\underset{MP \times MP}{\boldsymbol{U}(\boldsymbol{\phi}, \boldsymbol{\tau}, k)} = \underset{MP \times k}{[\boldsymbol{U}_s(\boldsymbol{\phi}, \boldsymbol{\tau}, k)} \quad \underset{MP \times (MP-k)}{\boldsymbol{U}_\nu(\boldsymbol{\phi}, \boldsymbol{\tau}, k)],}$$

where $\boldsymbol{U}_s(\boldsymbol{\phi}, \boldsymbol{\tau}, k) \in \mathcal{H}$, the signal subspace, and $\boldsymbol{U}_\nu(\boldsymbol{\phi}, \boldsymbol{\tau}, k) \in \mathcal{N}$, the noise subspace. We now transform the received data $\boldsymbol{y}(n)$ into $\mathcal{H}$ and $\mathcal{N}$ to form a signal component $\boldsymbol{z}_s(n)$ and a noise component $\boldsymbol{z}_\nu(n)$ respectively as

$$\boldsymbol{z}_s(n) = \boldsymbol{U}_s'(\boldsymbol{\phi}, \boldsymbol{\tau}, k)\boldsymbol{y}(n),$$

and

$$\boldsymbol{z}_\nu(n) = \boldsymbol{U}_\nu'(\boldsymbol{\phi}, \boldsymbol{\tau}, k)\boldsymbol{y}(n).$$

The new parameters $\boldsymbol{z}_s(n)$ and $\boldsymbol{z}_\nu(n)$ are both Gaussian.

Following the same treatment as in Chapter 3, the joint likelihood function of $\boldsymbol{z}_s$ and $\boldsymbol{z}_\nu$ is then given as

$$\begin{aligned} p(\boldsymbol{Z}_s, \boldsymbol{Z}_\nu | \tilde{\boldsymbol{A}}, \boldsymbol{\phi}, \boldsymbol{\tau}, k, \boldsymbol{W}^{-1}) &\approx \pi^{-N_t k} |\boldsymbol{C}^{-1}|^{N_t} \\ &\times \exp\left\{ -\sum_{n=1}^{N_t} (\boldsymbol{z}_s(n) - \tilde{\boldsymbol{a}}(n))' \boldsymbol{C}^{-1} (\boldsymbol{z}_s(n) - \tilde{\boldsymbol{a}}(n)) \right\} \\ &\times \pi^{-N_t(MP-k)} |\boldsymbol{W}^{-1}|^{N_t} \exp\left\{ -\sum_{n=1}^{N_t} \boldsymbol{z}_\nu'(n) \boldsymbol{W}^{-1} \boldsymbol{z}_\nu(n) \right\}. \end{aligned} \tag{4.3}$$

To complete the model, prior distributions are chosen to be non-informative where possible. When convenient, we also choose the structural form of these distributions for their desirable conjugate properties. Using the same prior distributions as in Chapter 3, with the addition of the prior distribution of the times of arrival

$$p(\boldsymbol{\tau}|k) = \frac{1}{P^k},$$

the posterior distribution, after carrying out the integration of the nuisance parameters and ignoring the constant terms, is then

$$p(k,\phi,\tau|Z_\nu) \propto \frac{\pi^{\frac{1}{2}(MP-k)(MP-k-1)}\prod_{i=1}^{MP-k}\Gamma(N_t-i+1)}{(2\pi P/\Lambda)^k k!(d^2)^{kN_t}} \times \left|N_t\hat{W}(\phi,\tau,k)\right|^{-N_t}. \tag{4.4}$$

with $N_t\hat{W}(\phi,\tau,k) \triangleq \sum_{n=1}^{N_t} z_\nu(n)z_\nu(n)^H$. Take note that this function depends only on the slowly varying parameters of interest. The objective is to estimate the parameters of this highly non-linear function, as the Maximum A Posteriori (MAP) estimates.

$$\{\hat{k},\hat{\phi},\hat{\tau}\} = \arg\max_{k,\phi,\tau\in\Theta} p(k,\phi,\tau|\mathbf{Z}_v). \tag{4.5}$$

4.3.1 White noise hypothesis

In the event that the noise is known to be spatially white, the integration of the nuisance parameters is straightforward. The detailed development is presented in (Andrieu and Doucet, 1999). The resulting posterior distribution is given as

$$p_{white}(\phi,\tau,k|\{y(n)\}) \propto \left(\sum_{n=1}^{N_t} y'(n)P_H^{\perp}(\phi,\tau,k)y(n)\right)^{N_tMP} \times \frac{\Lambda^k}{(2\pi)^k P^k k!(1+\delta^2)^{kN_t}},$$

where

$$P_H^{\perp}(\phi,\tau,k) = I - HMH^H,$$
$$M^{-1} = H^H H(1+\delta^{-2}).$$

and the hyperparameter δ^2 now represents the estimated SNR.

However, this form remains quite sensitive to the degree of colour in the noise. Our experiences show that the performance of the detection of model order degrades very quickly in the presence of even mildly coloured noise.

4.4 The reversible jump MCMC algorithm for channel characterization

The general reversible jump MCMC algorithm was completely defined in the previous chapters. To avoid repetition, only the key elements particular to this chapter are presented. The description is summarized as follows.

Reversible Jump MCMC

1. Initialization: set $\Phi^{(0)} = (\phi^{(0)}, \tau^{(0)}, k^{(0)})$ according to the prior distributions.
2. Iteration i,
   - Sample $u \sim U_{[0,1]}$
   - if $(u < b_{k^{(i)}})$ then execute a "birth move" (see section 4.4.2)
     - else if $(u < b_{k^{(i)}} + d_{k^{(i)}})$ then execute a "death move" (see section 4.4.2)
     - else, execute an update move (see section 4.4.1) .
3. $i \leftarrow i + 1$, goto step 2

4.4.1 Update move

Here, we assume that the current state of the algorithm is $(\phi_k, \tau_k, \{k\})$. When the update move is selected, the algorithm samples all parameters for k fixed. The proposal distributions in the general case of global exploration of the space are given as

$$q(\phi^{(i)}, \tau^{(i)}, k) = q(\phi^{\star}, \tau^{\star}, k) = \frac{1}{(2\pi P)^k}. \tag{4.6}$$

The acceptance ratio $r = r_{update}$ for the update move, in the case of coloured noise is obtained by substituting eqs. (4.6) and (4.4) into the expression for the acceptance ratio

defined in Chapter 2, eq. (2.8) to give

$$r_{update}(\phi_k^*, \tau_k^*, k, \phi_k, \tau_k, k) = \frac{\left|N_t \hat{W}(\phi^*, \tau^*, k)\right|^{-N_t}}{\left|N_t \hat{W}(\phi, \tau, k)\right|^{-N_t}}, \tag{4.7}$$

$$\alpha_{update} = \min[r_{update}, 1]. \tag{4.8}$$

The candidate (ϕ^*, τ^*) is then accepted as the current state $(\phi_k^{(i+1)} = \phi_k^*)$ and $(\tau_k^{(i+1)} = \tau_k^*)$, with probability α_{update}.

For the specific case at hand, the performance of the proposed method is again enhanced by selecting randomly between two types of proposal distributions for the update case rather than the one given by eq. (4.6): one type involves a global exploration of the parameter space, while the other involves a local exploration. The method is summarized as follows.

Update Move (Metropolis-one-at-the-time)

At iteration i, DO

- **Sampling the directions of arrival:**
  - Propose new directions of arrival, maintaining fixed values for $\tau^{(i)}$. The choice of proposal distribution is described as follows:
  - Sample $u_1 \sim U_{[0,1]}$
    - if $u_1 < 0.5$, then propose a global exploration

$$\phi_k^* \sim U_{[0,2\pi]^k}$$

    - else, propose a local exploration

$$\phi_k^* \sim \mathcal{N}(\phi_k^{(i)}, \Sigma_\phi) \tag{4.9}$$

      where Σ_ϕ is a covariance matrix such that samples from (4.9) are closely clustered around $\phi_k^{(i)}$.

- Evaluate α_{update} with eq. (4.8).
- Sample $u \sim U_{[0,1]}$.
- if $(u \leq \alpha_{update})$ then the state of the Markov Chain at iteration $i+1$ becomes (ϕ_k^*, k), else it remains at $(\phi_k^{(i)}, k)$.

- **Sampling the times of arrival:** (This process is analogous to sampling the DOAs, as above).
  - Propose new times of arrival, maintaining the directions of arrival fixed at their current values.
  - Sample $u_1 \sim U_{[0,1]}$
    - if $u_1 < 0.5$, then propose a global exploration

$$\tau_k^* \sim U_{[0,P]^k}$$

    - else, propose a local exploration

$$\tau_k^* \sim \mathcal{N}(\tau_k^{(i)}, \Sigma_\tau)$$

  - Evaluate α_{update} with eq. (4.8).
  - Sample $u \sim U_{[0,1]}$.
  - if $(u \leq \alpha_{update})$ then the state of the Markov Chain becomes (τ_k^*, k), else it remains at $(\tau_k^{(i)}, k)$.

■

4.4.2 Birth and death moves

In the death move case, we assume the current state is $(\phi_{k+1}, \tau_{k+1}, k+1)$ and we wish to determine whether the state is (ϕ_k, τ_k, k) at the next iteration. This involves the removal of an incident signal, which is chosen randomly amongst the $(k+1)$ existing incident signals.

The proposal distribution $q(\phi_k^\star, \tau_k^\star, k|\phi_{k+1}, \tau_{k+1}, k+1)$ for the death move is therefore chosen as

$$q(\phi_k^\star, \tau_k^\star, k|\phi_{k+1}, \tau_{k+1}, k+1) = p(k) \div \binom{k+1}{1} \propto \frac{\Lambda^k}{k!} \frac{1}{(k+1)},$$

where $p(k)$ is the prior distribution defined in eq. (3.8). The remaining values of $\phi_k^\star$ and $\tau_k^\star$ are set to the corresponding values of $\phi_k^{(i)}$ and $\tau_k^{(i)}$.

Similarly, in the birth move case, we assume the current state is (ϕ_k, τ_k, k) and we wish to determine whether the next state is $(\phi_{k+1}, \tau_{k+1}, k+1)$. This involves the addition of a new incident signal, which is proposed uniformly over $(0, 2\pi]$ and over $(0, P]$. The proposal distribution $q(\phi_{k+1}^\star, \tau_{k+1}^\star, (k+1)|\phi_k, \tau_k, k)$ for the birth move is therefore

$$q(\phi_{k+1}^\star, \tau_{k+1}^\star, (k+1)|\phi_k, \tau_k, k) = p(k+1)\frac{1}{2\pi}\frac{1}{P} \propto \frac{\Lambda^{k+1}}{(k+1)!}\frac{1}{2\pi}\frac{1}{P}$$

Using these proposal distributions with eq. (4.4), the acceptance ratio for the death move $r = r_{death}$ is

$$r_{death}(\phi_k^\star, \tau_k^\star, k, \phi_{k+1}, \tau_{k+1}, \quad k+1 \quad) = \frac{\left|N_t \hat{W}(\phi^\star, \tau^\star, k)\right|^{-N_t}}{\left|N_t \hat{W}(\phi, \tau, k+1)\right|^{-N_t}}$$
$$\times \quad \pi^{MP-k-1}\Gamma(N_t - MP + k + 1)(k+1)d^{2^{N_t}}. \quad (4.10)$$

The quantity α_{death} is then defined according to

$$\alpha_{death} = \min[r_{death}, 1]. \quad (4.11)$$

The acceptance ratio $r = r_{birth}$ for the birth move in the case of coloured noise can be verified to be

$$\alpha_{birth} = \min[1, \frac{1}{r_{death}}]. \quad (4.12)$$

The following block describes the algorithm for the birth move.

Birth Move

- Propose a new direction of arrival and time of arrival,

$$\boldsymbol{\phi}^{\star}_{k+1} = [\boldsymbol{\phi}^{(i)}_{k} : U_{[0,2\pi]}] \qquad \boldsymbol{\tau}^{\star}_{k+1} = [\boldsymbol{\tau}^{(i)}_{k} : U_{[0,P]}]$$

- Evaluate α_{birth} with eq. (4.12).
- Sample $u \sim U_{[0,1]}$.
- if $(u \leq \alpha_{birth})$ then the state of the Markov Chain becomes $(\boldsymbol{\phi}^{\star}_{k+1}, \boldsymbol{\tau}^{\star}_{k+1}, k+1)$, else it remains at $(\boldsymbol{\phi}^{(i)}_{k}, \boldsymbol{\tau}^{(i)}_{k}, k)$.

■

The description for the death move is similar, with appropriate modifications.

4.5 Simulation results

The proposed algorithm for coloured noise is now applied to simulation data, generated for $k_o = 2$ scatterers with the parameters described in Table 4.1. The receiver array is composed of 5 elements. The amplitudes are *iid* Rayleigh distributed over $N_t = 125$ received symbols

Scatterers	DOA (deg)	TOA (bins)	Amplitude (dB)
S1	65°	8	10
S2	20°	2	10

Table 4.1: Parameters of the multipath environment for simulated data

(or snapshots) for $P = 25$ chips, with an SNR of 5dB. The noise is once more coloured with an AR filter, the poles of which are $0.95e^{-j1.07\pi}$ and $0.95e^{-j0.88\pi}$. The corresponding spatial spectrum is shown in Figure 3.1. The hyper-parameter d^2 was set to 10, according to the procedure described earlier, in section 3.4. The hyper-parameter Λ was set to the

true number of multipath components. As discussed in the previous chapter, this parameter does not influence the maximum a posteriori estimates.

The Reversible Jump MCMC scheme goes through 1500 iterations after a burn-in period of 300 iterations. The results, as found by the algorithm, are summarized in Figure 4.1. It is clear from the histograms that the DOA and TOA estimates concentrate around their true values. The posterior probability of the number of scatterers $\hat{k}$ being equal to the true value of two was evaluated at 75% (due to a high noise level), as summarized in Table 4.2. The bottom line, "measurements", describes the detection performance using real measured data, as presented in section 4.6. Clearly, the algorithm correctly identified the parameters of the simulated multipath scenario. Further simulation results for the DOA case only are presented in (Larocque and Reilly, 2000c). There are omitted in this document as the present work focuses on the *joint* detection and estimation of the parameters. Further

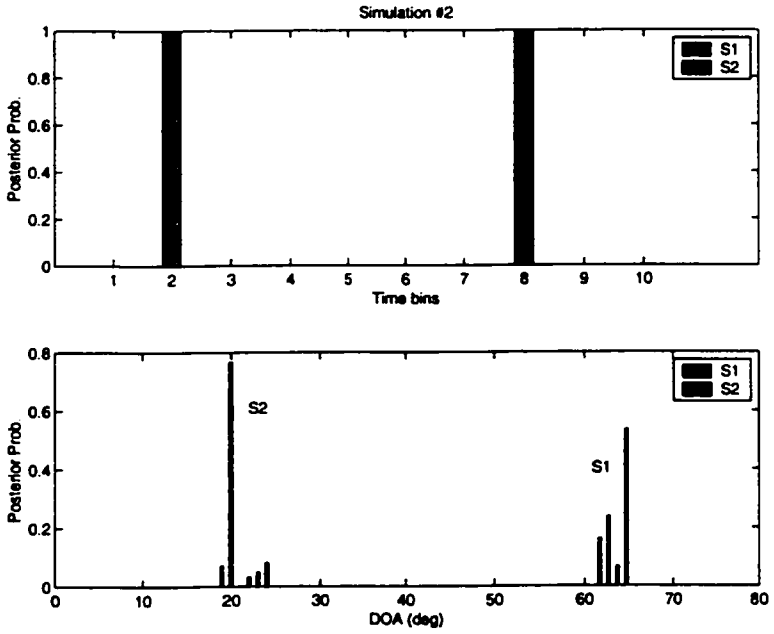

Figure 4.1: Simulations: Histogram of the TOA (top); Histogram of DOA (bottom).

simulation results of the joint problem can be found in (Larocque and Reilly, 2001a,b).

$p(\hat{k}=i)$ %	1	2	3
Simulations	0	75	25
Measurements	1	98	1

Table 4.2: Posterior estimate of the number of paths using MCMC, for both the simulated data and the real data.

4.5.1 Performance of the method

The TOA parameter is discrete; therefore the estimate may be considered fixed for reasonable values of SNR. Thus the joint TOA/DOA estimation performace is approximately determined by examining the DOA estimation performance alone. Thus, the performance measures showing the DOA-only case provided in Chapter 3 are still meaningful in providing joint DOA/TOA performance in the present context.

4.6 Application to real-life problem

In this section, we apply our proposed scheme for the coloured noise case to real-life outdoor propagation measurements with a typical data set collected on McMaster University campus.

4.6.1 The measurement scenario

The complex channel impulse response, in time and space, is measured directly in the time domain by transmitting a wide-band spread-spectrum signal and correlating the received signal with the known transmitted sequence, at each element of the receiving array.

The receiving base station is a circular antenna array made of 8 mono-pole antennae. The transmitted signal is a 255 chip pseudonoise (PN) sequence at 5 MHz. The received signal of each element is I-Q demodulated, converted to baseband, sampled at 10 MHz, and then stored for further processing. One time bin ΔT therefore corresponds to 100ns. The measurements were conducted on the McMaster University campus, with the receiving base station at different locations and different heights in a pico-cell scenario that offered rich

multipath characteristics with severe fading. An initial calibration is based on measurements with the antenna array inside an anechoic chamber and illuminated from 64 different angles.

4.6.2 Processing

For the purpose of demonstrating the proposed algorithm, $N_t = 20$ received data symbols (or snapshots) are used, measured from a position where the multipath characteristics have been geographically observed to be 2 rays incident approximatively from angles of arrival of 30° and 135°, as shown in Figure 4.2.

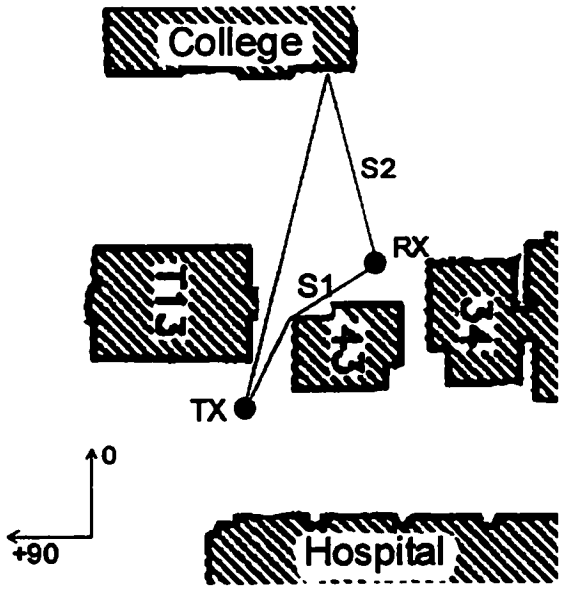

Figure 4.2: Map describing the geometry of the setup.

Using traditional beamforming techniques, one can easily obtain an angular power spectrum. The temporal impulse response is directly obtained from the output of the correlator. Both results are presented in Figure 4.3, where the estimates found by the proposed algorithm are indicated. As the figure shows, it is not easy to first determine the number of multipath components; it can easily be over-estimated. Furthermore, there is ambiguity in which DOA estimate corresponds with which TOA estimate. With the proposed algorithm, since the optimization is done jointly for the two sets of parameters, the ambiguity

is resolved.

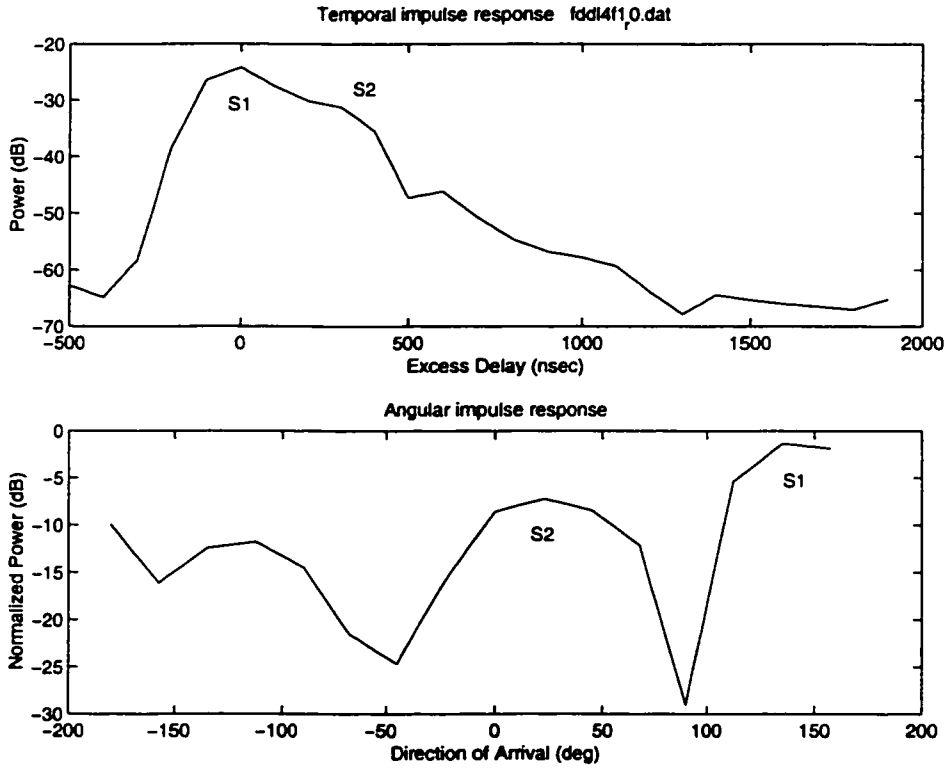

Figure 4.3: Channel impulse response: Temporal (top); Angular (bottom).

The results from application of the proposed algorithm are now presented. Figure 4.4 shows typical results for 1000 iterations of the Reversible Jump MCMC Sampler. The initialization of the algorithm was totally random, with no prior information being used. It is clear that the algorithm identifies the two major multipath components, in time and in direction. These estimates are presented in Table 4.3. From the geometry of the system, the first multipath ray goes around the corner of building 43 (or over the top), while the second ray impinges from another direction, after bouncing off the Teachers' College building across the street. The receiving array is located approximately mid-way between the College and the transmitter. The extra propagation time is thus simply the path from the transmitter to the College, which was measured to be approximately 100m which corresponds to a time

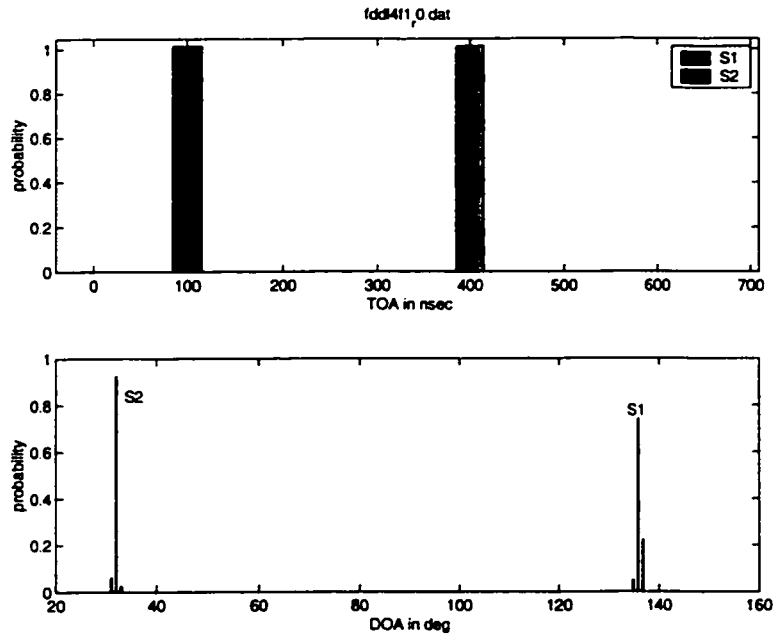

Figure 4.4: Measurements: Histogram of the TOA (top); Histogram of the DOA (bottom).

delay of 330ns. Within the resolution of the 100ns time bins, the excess delay measurement of S2 obtained from geographical truthing is therefore 300ns.

Sources	DOA	TOA (Excess delay)
S1	136°	0 ns
S2	32°	300 ns

Table 4.3: Estimate of the multipath components using MCMC

It is seen that the parameters extracted by the algorithm are in accordance with the physical characteristics of the scenario, as described on the map. The performance of the detection procedure for the real measurement case is shown in the bottom line of Table 4.2.

4.7 Conclusion

In this chapter, a new and innovative approach to channel characterization in unknown coloured noise with arbitrary covariance is presented in the Bayesian framework. We used a Markov Chain Monte Carlo method to perform the joint estimation of the model order, times of arrival and directions of arrival, which jointly characterize the multipath channel. The nuisance parameters (unknown noise variance and amplitudes) are integrated out analytically.

Simulation results support the effectiveness of the method, and demonstrate reliable detection of the number of sources and estimation of their directions and times of arrival in coloured noise with a single array. The performance is reinforced with the successful application to real data and it shows how the algorithm provides a potential solution to a practical problem of localization.

Chapter 5

Sequential Monte Carlo: Particle Filters

This chapter addresses the application of sequential importance sampling (SIS) schemes to tracking DOAs of an unknown number of sources, using a passive array of sensors. This proposed technique has significant advantages in this application, including the ability to detect a changing number of signals at arbitrary times throughout the observation period, and that the requirement for quasi-stationarity over a limited interval may be relaxed, relative to other methods.

We propose the use of a reversible jump MCMC step to enhance the statistical diversity of the particles. This step also enables us to introduce two novel moves which significantly enhance the performance of the algorithm when the DOA tracks cross. The superior performance of the method is demonstrated by examples of application of the particle filter to sequential tracking of the DOAs of an unknown and non-stationary number of sources, and to a scenario where the targets cross. Our results are compared to the PASTd method of Yang (1995a).

5.1 Introduction

The problem of tracking the directions of arrival of multiple targets in background noise using passive arrays of sensors is of great interest to the signal processing community, with applications in communications, radar, sonar, acoustics and others. For example, in a beamforming application, one is typically interested in extracting a signal of interest arriving onto an array of sensors, from multiple interfering sources arriving from different DOAs. For this approach to be effective, the DOA of the desired source must be estimated from the received data. In many scenarios, the desired source is moving, necessitating *target tracking* of the desired DOA.

Recently, many high resolution DOA estimation techniques have been proposed. These include beamforming methods (Johnson, 1982; Capon, 1969), subspace-based methods (Schmidt, 1986; Viberg and Ottersen, 1991) and maximum likelihood methods (Reilly and Haykin, 1982; Ligget, 1973). Since these high-resolution methods incorporate the benefits of temporal averaging and knowledge of the model order, the target must be assumed stationary over the period of observation. Thus, these methods fail or suffer performance degradations when the DOAs of the targets exhibit significant motion during the observation period.

In recent years there have been several methods developed for estimating or tracking the DOAs of moving targets using passive sensors or arrays of sensors , e.g., (Zhou *et al.*, 1999; Yang, 1995b; Molnar and Modestino, 1998), etc. Like the high-resolution methods, these approaches also assume the targets are stationary over a limited time interval. The approach in (Yang, 1995b) is based on adaptively estimating a noise subspace basis from the received signal covariance matrix. These methods then rely on a high-resolution technique such as MUSIC (Schmidt, 1986) to estimate the desired DOAs. In (Zhou *et al.*, 1999), a method based on maximum likelihood estimation of a novel state-space representation for tracking is presented.

An important consideration in target tracking problems is the data association problem; i.e., the association of tracks with measurements. In the case where passive arrays of sensors are used, the data association problem reduces to the association of targets before and after

their DOA tracks cross each other. In (Molnar and Modestino, 1998), a method for DOA tracking for disparately-spaced sensors using the EM algorithm is presented. This method treats the DOAs as unknown parameters and the data associations as the missing data.

In this chapter, we discuss the use of sequential MC (Monte Carlo) methods for target tracking. MC and MCMC (Markov chain Monte Carlo) methods (Gilks *et al.*, 1998; Andrieu *et al.*, 1998; Larocque *et al.*, 2001a,b) have been capturing the attention of researchers in the field of statistics throughout the past decade and have more recently emerged as useful methods in the signal processing arena. However, conventional MC methods are not well suited to problems where data arrive sequentially, due to excessive computational requirements. This consideration has motivated the development of *sequential* MC methods, also known as *particle filters* (Doucet, 1998; Andrieu *et al.*, 2001; Djuric, 2000), which are capable of recursively updating the probability distributions of interest as new data become available.

In this chapter, we propose the application of particle filters to joint detection, estimation and tracking of an unknown and time–varying number of sources. There are several advantages offered by this approach.

Firstly, previous methods require prior determination of model order. The MDL (Rissanen, 1978) and AIC criteria (Wax and Kailath, 1985) are often used for this purpose. These methods require the assumption of stationarity and are highly sensitive to the white noise assumption. The proposed approach offers robust estimation of the model order *jointly* with other parameters of interest, and furthermore can accommodate changes in model order occurring arbitrarily throughout the observation interval.

Secondly, the particle filtering approach estimates the posterior distribution of the parameters given all past data. This distribution can then be marginalized to yield the "instantaneous" posterior distribution of the desired parameters at the current time sample. Thus we need not assume stationarity. This is in contrast to most other methods which involve estimation of second- or higher-order statistics by temporal averaging, a process which requires stationarity over an appropriate interval.

Thirdly, with the particle filtering approach, the joint posterior distribution of the target

amplitudes given the received data is readily available. This greatly facilitates high-accuracy data association.

Finally, in contrast to other methods, because any form of MCMC technique produces an approximation of the entire distribution of interest, one can easily calculate confidence intervals, marginalize with respect to desired parameters, or make inferences on the parameters, etc.

One of the difficulties with particle filtering is the loss of statistical diversity in the recursive update of the importance weights (Doucet, 1998). To mitigate this difficulty, we have introduced a new form of the Reversible Jump MCMC (Green, 1995; Andrieu and Doucet, 1999) process as a novel resampling engine. We propose the use of two new moves, called the *split/merge* (Richardson and Green, 1997; Andrieu, 1997) moves, which are specifically designed to handle the crossing of the signal tracks, and allow for the joint detection and tracking of the number of sources.

The chapter is organized as follows. Section 5.2 presents the state-space model. In sections 5.3 and 5.4, we discuss and extend the particle filtering approach as developed in (Doucet, 1998) to the target tracking problem and provide insights on the methods. Results from simulations are presented in section 5.5 where our results are compared with the method presented in (Yang, 1995a,b). Conclusions are given in section 5.6.

5.2 The State-Space model

The problem of interest is the sequential detection of the number of sources impinging an array and the estimation of their corresponding directions of arrival.

The signal model we consider consists of a complex vector of observations $\boldsymbol{y}(t) \in \mathcal{C}^M$ which represents the data received by an array of M sensors at the tth snapshot. The observation vector is composed of incident narrow-band plane wave signals from $k(t)$ distinct sources embedded in Gaussian noise. Each of these incident plane-wave signals impinges on the array of sensors at a physical angle $\phi_k, k = 1, \ldots, k(t)$, relative to the normal of the array. The amplitudes of the sources at the tth time instant are denoted by the vector

$\boldsymbol{a}(t) \in \mathcal{C}^{k(t)}$.

The sequential sampling approach we adopt admits a first order state-space hidden Markov model. The states $[\boldsymbol{\phi}(t), \boldsymbol{a}(t)]$ evolve according to

$$\boldsymbol{\phi}(t) = \boldsymbol{\phi}(t-1) + \sigma_v \boldsymbol{v}(t), \tag{5.1}$$

$$\boldsymbol{a}(t) \sim \mathcal{N}(\mathbf{0}, \sigma_a^2 \mathrm{I}_{k(t)}), \tag{5.2}$$

whereas the observation is given by

$$\boldsymbol{y}(t) = \boldsymbol{S}(\boldsymbol{\phi}(t))\boldsymbol{a}(t) + \sigma_w \boldsymbol{w}(t). \tag{5.3}$$

The noise variables $\boldsymbol{v}(t) \in \mathcal{R}^M$, $\boldsymbol{w}(t) \in \mathcal{C}^M$ are *iid* Gaussian variables with zero mean and unit variance, independent of the parameters. The respective variances of the scaled noise terms are σ_v^2 and σ_w^2. The variance of the prior distribution of the amplitudes will be defined shortly. The dimension $k(t)$ of the model is described by the following stochastic relationship at time t,

$$k(t) = k(t-1) + \epsilon_k(t), \tag{5.4}$$

where the $\epsilon_k(t)$ are discrete *iid* random variables such that

$$\begin{aligned} P(\epsilon_k(t) = -1) &= h/2 \\ P(\epsilon_k(t) = 0) &= 1 - h \\ P(\epsilon_k(t) = 1) &= h/2, \end{aligned} \tag{5.5}$$

where $h \in [0, 1]$. In eq. (5.5), it is tacitly assumed that the model order changes by no more than one in each sample period.

In the proposed system of equations, the noise variances σ_v^2 and σ_w^2 are assumed unknown but constant over time. The unknown vectors of amplitudes $\boldsymbol{a}(t)$ are assumed *iid* between snapshots.

We introduce a vector $\boldsymbol{\theta}$ of all the parameters describing the model

$$\boldsymbol{\theta}_{1:t} \triangleq (\{\boldsymbol{\phi}_{k(t)}\}_{1:t}, \{\boldsymbol{a}_{k(t)}\}_{1:t}, k_{1:t}, \sigma_v^2, \sigma_w^2),$$

where the notation $(\cdot)_{1:t}$ indicates all the elements from time 1 to time t, and the subscript $k(t)$ indicates the size of the corresponding vector. The posterior distribution of interest is then given by $\pi(\boldsymbol{\theta}_{1:t}) \triangleq p(\boldsymbol{\theta}_{1:t}|\boldsymbol{y}_{1:t})$, and can be specified within a normalizing constant, using Bayes' theorem, as

$$\pi(\boldsymbol{\theta}_{1:t}) \propto p(\boldsymbol{y}_{1:t}|\boldsymbol{\theta}_{1:t})p(\boldsymbol{\theta}_{1:t}), \tag{5.6}$$

where $p(\boldsymbol{y}_{1:t}|\boldsymbol{\theta}_{1:t})$ is the likelihood function and $p(\boldsymbol{\theta}_{1:t})$ is the prior distribution of the parameters. From the model description, it is clear that the prior distributions for some of the parameters are conditional on $\boldsymbol{k}_{1:t}$ and also on other parameters. Thus, we expand the posterior distribution in eq. (5.6) to give

$$\begin{aligned}\pi(\boldsymbol{\theta}_{1:t}) \propto\ & p(\boldsymbol{y}_{1:t}|\boldsymbol{\phi}_{1:t}, \boldsymbol{a}_{1:t}, \boldsymbol{k}_{1:t}, \sigma_v^2, \sigma_w^2) \times \\ & p(\boldsymbol{\phi}_{1:t}|\sigma_v^2, \boldsymbol{k}_{1:t})p(\boldsymbol{a}_{1:t}|\boldsymbol{\phi}_{1:t}, \sigma_w^2, \boldsymbol{k}_{1:t})p(\boldsymbol{k}_{1:t})p(\sigma_v^2)p(\sigma_w^2).\end{aligned} \tag{5.7}$$

We now assign distributions for each of the terms in eq. (5.7). It is assumed that the observations, given the states, are *iid* and that the conditional update likelihoods of the states are also *iid*. Therefore, assuming the distribution of the initial states to be uniform, and using the Markov properties of the model, the distributions of eq. (5.7) can be written in the form of

$$p(\boldsymbol{y}_{1:t}|\boldsymbol{\phi}_{1:t}, \boldsymbol{a}_{1:t}, \boldsymbol{k}_{1:t}, \sigma_w^2) = \prod_{l=1}^{t} \mathcal{N}(\boldsymbol{S}(\boldsymbol{\phi}_l)\boldsymbol{a}_l, \sigma_w^2 \boldsymbol{I}_M), \tag{5.8}$$

$$p(\boldsymbol{\phi}_{1:t}|\boldsymbol{k}_{1:t}, \sigma_v^2) = \prod_{l=1}^{t} \mathcal{N}(\boldsymbol{\phi}_{l-1}, \sigma_v^2 \boldsymbol{I}_{k_l}), \tag{5.9}$$

$$p(\boldsymbol{a}_{1:t}|\boldsymbol{\phi}_{1:t}, \boldsymbol{k}_{1:t}, \sigma_w^2) = \prod_{l=1}^{t} \mathcal{N}(\boldsymbol{0}, \delta^2 \sigma_w^2 (\boldsymbol{S}'(\boldsymbol{\phi}_l)\boldsymbol{S}(\boldsymbol{\phi}_l))^{-1}), \tag{5.10}$$

$$p(\boldsymbol{k}_{1:t}) = \prod_{l=1}^{t} p(k_l|k_{l-1}) = \prod_{l=1}^{t} \epsilon_k(l). \tag{5.11}$$

The prior distribution defined in eq. (5.10) is the maximum entropy prior (Andrieu and Doucet, 1999) with the parameter δ^2 set to an estimate of the SNR. The prior distribution on the variances σ_v^2 and σ_w^2 are both assumed to follow the inverse Gamma distribution, which is the conjugate distribution for the Normal distribution,

$$p(\sigma_v^2) \sim \mathcal{IG}(\frac{\nu_0}{2}, \frac{\gamma_0}{2}), \tag{5.12}$$

$$p(\sigma_w^2) \sim \mathcal{IG}(\nu_1, \gamma_1). \tag{5.13}$$

Since $p(\sigma_v^2)$ is later combined with a real Normal distribution (instead of complex), the factors of $\frac{1}{2}$ in eq. (5.12) are required to maintain the conjugate property. The above priors are noninformative when the hyperparameters ν and γ are set to zero.

The model is now clearly defined. In the application addressed in this chapter, the parameters of interest are primarily the DOAs $\boldsymbol{\phi}_{1:t}$ and the model order $k_{1:t}$. The amplitudes $\boldsymbol{a}_{1:t}$, along with the state update noise variances σ_w^2 and σ_v^2, may be considered nuisance parameters. Even though it is straightforward to numerically marginalize the posterior density to eliminate these undesired parameters using the proposed Monte-Carlo based estimation methods, the resulting procedure is more efficient if the nuisance parameters can be integrated out analytically. Such is the case with the signal amplitudes. We now proceed to eliminate the amplitudes from the posterior distribution $\pi(\boldsymbol{\theta}_{1:t})$ by marginalization.

Using the *iid* Normal distribution of the noise variables, and the model structure given by eqs. (5.1) to (5.3), and eqs. (5.8) to (5.13), the posterior distribution $\pi(\boldsymbol{\theta}_{1:t})$ of eq. (5.7)

can be written as

$$\begin{aligned}\pi(\boldsymbol{\theta}_{1:t}) \propto & \prod_{l=1}^{t} \frac{1}{\sigma_w^{2M}\pi^M} \exp\left[-\frac{1}{\sigma_w^2}(\boldsymbol{y}_l - \boldsymbol{S}(\boldsymbol{\phi}_l)\boldsymbol{a}_l)'(\boldsymbol{y}_l - \boldsymbol{S}(\boldsymbol{\phi}_l)\boldsymbol{a}_l)\right] \\ & \times \prod_{l=1}^{t} \frac{|\boldsymbol{S}'(\boldsymbol{\phi}_l)\boldsymbol{S}(\boldsymbol{\phi}_l)|}{(\delta\sigma_w)^{2k_l}\pi^{k_l}} \exp\left[-\frac{1}{\delta^2\sigma_w^2}(\boldsymbol{a}_l)'(\boldsymbol{S}'(\boldsymbol{\phi}_l)\boldsymbol{S}(\boldsymbol{\phi}_l))(\boldsymbol{a}_l)\right] \\ & \times \prod_{l=1}^{t} \frac{1}{\sigma_v^{2k_l/2}(2\pi)^{k_l/2}} \exp\left[-\frac{1}{2\sigma_v^2}(\boldsymbol{\phi}_l - \boldsymbol{\phi}_{l-1})'(\boldsymbol{\phi}_l - \boldsymbol{\phi}_{l-1})\right] \\ & \times \sigma_v^{2(-\frac{\alpha_0}{2}-1)} \exp\left[\frac{-\gamma_0}{2\sigma_v^2}\right] \times \sigma_w^{2(-\alpha_1-1)} \exp\left[\frac{-\gamma_1}{\sigma_w^2}\right] \\ & \times \prod_{l=1}^{t} p(k_l|k_{l-1}).\end{aligned}$$

The terms relating to the amplitudes can be collected together to give the following expression, as in (Andrieu and Doucet, 1999),

$$\begin{aligned}\pi(\boldsymbol{\theta}_{1:t}) \propto & \prod_{l=1}^{t} \frac{1}{\sigma_w^{2k_l}\pi^{k_l}} \exp\left[\frac{-1}{\sigma_w^2}(\boldsymbol{a}_l - \boldsymbol{m}_{a_l})'\boldsymbol{\Sigma}_{k_l}^{-1}(\boldsymbol{a}_l - \boldsymbol{m}_{a_l})\right] \\ & \times \prod_{l=1}^{t} \frac{1}{\sigma_w^{2M}\delta^{2k_l}} \exp\left[\frac{-1}{\sigma_w^2}\boldsymbol{y}_l'\boldsymbol{P}_S^{\perp}(\boldsymbol{\phi}_l)\boldsymbol{y}_l\right] \\ & \times \prod_{l=1}^{t} \frac{|\boldsymbol{S}'(\boldsymbol{\phi}_l)\boldsymbol{S}(\boldsymbol{\phi}_l)|}{\sigma_v^{2(k_l/2)}(2\pi)^{(k_l/2)}} \exp\left[\frac{-1}{2\sigma_v^2}(\boldsymbol{\phi}_l - \boldsymbol{\phi}_{l-1})'(\boldsymbol{\phi}_l - \boldsymbol{\phi}_{l-1})\right] \\ & \times \sigma_v^{2(-\frac{\nu_0}{2}-1)} \exp\left[\frac{-\gamma_0}{2\sigma_v^2}\right] \times \sigma_w^{2(-\nu_1-1)} \exp\left[\frac{-\gamma_1}{\sigma_w^2}\right] \\ & \times \prod_{l=1}^{t} p(k_l|k_{l-1}),\end{aligned} \tag{5.14}$$

where

$$\begin{aligned}\boldsymbol{\Sigma}_{k_l}^{-1} &= \boldsymbol{S}'(\boldsymbol{\phi}_l)\boldsymbol{S}(\boldsymbol{\phi}_l)(1+\delta^{-2}), \\ \boldsymbol{m}_{a_l} &= \boldsymbol{\Sigma}_{k_l}\boldsymbol{S}'(\boldsymbol{\phi}_l)\boldsymbol{y}_l,\end{aligned}$$

and

$$\boldsymbol{P}_S^{\perp}(\boldsymbol{\phi}_l) = \boldsymbol{I}_M - \frac{\boldsymbol{S}(\boldsymbol{\phi}_l)(\boldsymbol{S}'(\boldsymbol{\phi}_l)\boldsymbol{S}(\boldsymbol{\phi}_l))^{-1}\boldsymbol{S}'(\boldsymbol{\phi}_l)}{(1+1/\delta^2)}.$$

From eq. (5.14) a *maximum a posteriori* estimate of the amplitudes, knowing the other parameters is readily available as

$$\hat{a}_{MAP}(l) = m_{a_l}. \tag{5.15}$$

Thus, the amplitude parameters need not be included in the particle filter. Instead, they can be estimated at each iteration, after the sampling of the other parameters as discussed in sections 5.3 and 5.4.

It then becomes straightforward to integrate out the amplitudes in eq. (5.14) to yield a simpler definition of the posterior distribution in terms of the remaining parameters. The posterior distribution can then be simplified to

$$\begin{aligned}\pi(\phi_{1:t}, \sigma_v^2, \sigma_w^2, k_{1:t}) \propto & \prod_{l=1}^{t} \frac{1}{\sigma_w^{2M}(1+\delta^2)^{k_l}} \exp\left[\frac{-1}{\sigma_w^2} y_l' P_S^{\perp}(\phi_l) y_l\right] \\ & \times \prod_{l=1}^{t} \frac{1}{\sigma_v^{2(k_l/2)}(2\pi)^{(k_l/2)}} \exp\left[\frac{-1}{2\sigma_v^2}(\phi_l - \phi_{l-1})'(\phi_l - \phi_{l-1})\right] \\ & \times \sigma_v^{2(-\frac{\nu_0}{2}-1)} \exp\left[\frac{-\gamma_0}{2\sigma_v^2}\right] \times \sigma_w^{2(-\nu_1-1)} \exp\left[\frac{-\gamma_1}{\sigma_w^2}\right] \\ & \times \prod_{l=1}^{t} p(k_l|k_{l-1}).\end{aligned} \tag{5.16}$$

The MAP estimators of the nuisance parameters of the variances can be readily obtained by comparing the previous distribution with a product of Inverted Gamma distributions. Using the fact that the mode of the Inverted Gamma distribution is $\frac{\gamma}{\nu+1}$, it follows that

$$\sigma_{v_{MAP}}^2(t) = \frac{\frac{\gamma_0}{2} + \frac{1}{2}\sum_{l=1}^{t}(\phi_l - \phi_{l-1})'(\phi_l - \phi_{l-1})}{\frac{\nu_0}{2} + \frac{1}{2}\sum_{l=1}^{t} k(l) + 1}, \tag{5.17}$$

$$\sigma_{w_{MAP}}^2(t) = \frac{\gamma_1 + \sum_{l=1}^{t}(y_l' P_S^{\perp}(\phi_l) y_l)}{\nu_1 + Mt + 1}. \tag{5.18}$$

We choose however to keep these parameters in the expression of the posterior distribution in eq. (5.16). This simplifies the derivation of the acceptance probabilities of the moves as discussed in section 5.4. Since the nuisance parameters can be estimated, we now define a new vector α of parameters to sample with the particle filter, as

$$\alpha_{1:t} \triangleq (\phi_{1:t}, k_{1:t}).$$

5.3 Sequential Importance Sampling

This section describes the SIS procedure introduced in Chapter 2, which is used to extract the DOA estimates for tracking. In this section, the background treatment on the SIS methodology is necessarily brief. The reader is referred to Chapter 2, (Doucet, 1998; Liu and Chen, 1993; Doucet *et al.*, 2001b) and the references therein for a more complete coverage of this topic.

As described in Chapter 2, the optimal importance function that satisfies the recurrence requirement eq. (2.14) and minimizes the variance of the weights generated by the recursion eq. (2.16), is given by

$$q_{optimal}(\cdot) = q(\alpha_t^{(i)}|\alpha_{t-1}^{(i)}, y_t).$$

Unfortunately, this distribution is not easily evaluated directly for the problem at hand. However, an suboptimal approximation is readily obtained by means of a local linearization (Taylor expansion) of the observation equation. The observation equation (5.3) and the state update equation (5.1) for ϕ are reproduced here for convenience

$$\begin{aligned}\phi(t) &= \phi(t-1) + \sigma_v v(t),\\ y(t) &= S(\phi(t))a(t) + \sigma_w w(t).\end{aligned}$$

These equations yield

$$y(t) \approx S(\phi(t-1))a(t) + \left.\frac{\partial S(\phi(t))a(t)}{\partial \phi(t)}\right|_{\begin{pmatrix}\phi(t)=\phi(t-1)\\ a(t)=a(t-1)\end{pmatrix}} \times \big(\phi(t) - \phi(t-1)\big) + \sigma_w w(t).$$

After solving the above for $\phi(t)$, from the assumptions on the model noise, we see the resulting distribution for $\phi(t)|\phi(t-1), y(t)$, i.e., the optimal importance function, is linear and Gaussian and can be expressed as

$$q(\phi^{(i)}(t)|\phi^{(i)}(t-1), y(t)) \sim \mathcal{N}(m^{(i)}(t), \Sigma^{(i)}(t)), \tag{5.19}$$

where, for each particle,

$$\Sigma^{-1}(t) = \sigma_v^{-2}(t) I_{k(t)} + G'(\sigma_w^{-2}(t) I_M) G,$$

$$\boldsymbol{m}(t) = \boldsymbol{\Sigma}(t)\Big(\sigma_v^{-2}(t)\boldsymbol{I}_k(t)\boldsymbol{\phi}(t-1) + \boldsymbol{G}'(\sigma_w^{-2}(t)\boldsymbol{I}_M)[\boldsymbol{y}(t) - \boldsymbol{S}(\boldsymbol{\phi}(t-1))\boldsymbol{a}(t-1) + \boldsymbol{G}\boldsymbol{\phi}(t-1)]\Big),$$

and the matrix $\boldsymbol{G}$ is the gradient of the observation equation

$$\boldsymbol{G} = \frac{\partial \boldsymbol{S}(\boldsymbol{\phi}(t))\boldsymbol{a}(t)}{\partial \boldsymbol{\phi}(t)}.$$

In summary, the recursive update for the weights is obtained using the following form of eq. (2.16)

$$\tilde{w}^{(i)}(t) = w^{(i)}(t-1) \times \frac{p\big(\boldsymbol{y}(t)|\boldsymbol{\phi}^{(i)}(t), k^{(i)}(t), \boldsymbol{a}^{(i)}(t), \sigma_w^{2(i)}\big)p\big(\boldsymbol{\phi}^{(i)}(t)|\boldsymbol{\phi}^{(i)}(t-1), k^{(i)}(t), \sigma_v^{2(i)}\big)}{q(\boldsymbol{\phi}^{(i)}(t)|\boldsymbol{\phi}^{(i)}(t-1), \boldsymbol{y}(t))}. \tag{5.20}$$

Equations (5.8) and (5.9) are used for the respective terms on the numerator, and eq. (5.19) is used for the denominator.

As discussed in section 2.4.1, it is convenient to use a reversible jump MCMC procedure to provide diversity amongst the particles. The MCMC procedure samples $\pi(\boldsymbol{\alpha})$ directly, thus introducing statistical diversity amongst the particles. Also, the reversible jump process is capable of exploring parameter spaces of varying dimension, which as we see in section 5.4 is the key to detection of model order (Andrieu and Doucet, 1999; Larocque and Reilly, 2000b).

The above procedure is summarized in the following schema.

Sequential Importance Sampling for Tracking an Unknown Number of DOAs

For time $t = 1$, initialize the weights $w^{(i)}(1), i = 1, \ldots, N$

For each time step $t = 2, 3, \ldots$, DO

1. The Importance Sampling Step:

   - For $i = 1, \ldots, N$, generate the particles by sampling from the distribution $q(\cdot|\cdot)$, as follows (see eq. (5.19)):

$$q(\boldsymbol{\phi}^{(i)}(t)|\boldsymbol{\phi}^{(i)}(t-1), \boldsymbol{y}(t)) \sim \mathcal{N}(\boldsymbol{m}^{(i)}(t), \boldsymbol{\Sigma}^{(i)}(t))$$

- For $i = 1, \ldots, N$, Evaluate the un-normalized importance weights from eq. (5.20).
- For $i = 1, \ldots, N$, normalize the weights:

$$w^{(i)}(t) = \frac{\tilde{w}^{(i)}(t)}{\sum_{j=1}^{N} \tilde{w}^{(j)}(t)} \tag{5.21}$$

2. The Resampling/Selection of the Particles:
   - Sample a vector of index $\boldsymbol{l}$ distributed as:

$$P(l(j) = i) = w^{(i)}(t) \tag{5.22}$$

   - Resample the particles with the index vector:

$$\boldsymbol{\phi}_{0:k}^{(i)} = \boldsymbol{\phi}_{0:k}^{(l(i))} \tag{5.23}$$

   - Re-assign all the weights to $w^{(i)}(t) = \frac{1}{N}$.
3. The Reversible Jump MCMC Step:
   - Apply the sampler to be described in section 5.4 to enhance diversity amongst the particles and facilitate detection of model order.

The SIS procedure is now completely described, in order to use this procedure to track the DOAs. Our objective is to estimate the parameters of interest $\boldsymbol{\phi}(t)$ and $k(t)$ given all past observations, at each time instant. This can be achieved by forming the marginal distribution corresponding only to the specific parameters of interest from $\pi(\boldsymbol{\alpha}_{1:t})$. One of the primary advantages of using a numerical Bayesian procedure for parameter estimation is that this implicit integration is readily performed directly from the histogram $\hat{\pi}_N(d\boldsymbol{\alpha}_{1:t})$.

5.4 The reversible jump MCMC diversity step

In our application, since the dimension of the parameter space $\boldsymbol{\phi}_{1:t}$ varies with $k_{1:t}$, we use the *reversible jump* MCMC method which samples directly from the joint distribution over

all model orders of interest. In effect, the process jumps between subspaces of different dimensions, thus visiting all relevant model orders. In the reversible jump case, candidate samples are chosen from a *set* of proposal distributions, which are randomly accepted according to an acceptance ratio that ensures reversibility, and therefore the invariance of the Markov chain with respect to the desired posterior distribution. Here, we choose our set of proposal distributions to correspond to the following set of moves, which includes new moves:

- the *birth move*, chosen with probability b_k, for which a new source is proposed at random; i.e., $k(t) = k(t-1) + 1$.
- the *death move*, chosen with probability d_k, for which one of the existing sources is proposed to be removed; i.e., $k(t) = k(t-1) - 1$.

These moves, in conjunction with the update move described below, enable us to sample the parameter $k_{1:t}$. By forming the marginal of $\pi(\alpha_{1:t})$ with respect to $k_{1:t}$, we can detect the most likely number of sources vs. time. In addition to these moves, we propose two further novel moves, which we have shown in simulations to improve performance when two neighbouring DOA tracks cross, by proposing better candidates. These additional moves are

- *the split move*, for which an existing source is proposed to be split into two sources. This move is chosen with probability s_k .
- *the merge move*, for which two neighbouring sources are proposed to be merged into one. This move is chosen with probability m_k .

We also have the update move:

- with the *update move*, all the parameters are updated with fixed dimension; i.e., $k(t) = k(t-1)$. This move is executed with probability $1 - b_k - d_k - s_k - m_k$. With the update, birth and death moves, $\boldsymbol{J}$ in eq. (2.10) is readily shown (Andrieu and Doucet, 1999) to be unity.

It is shown in (Doucet *et al.*, 2001a) that the proposed MCMC sampling procedure requires no burn–in period in this application. This is a consequence of the fact the particles before the MCMC step are already distributed according to the limiting distribution of the chain. Thus, in the interest of computational efficiency, only one MCMC iteration need be applied to each particle at each time step. Furthermore, this allows for the use of moves that are not all mutually reversible.

The selection of moves is described by the following schema.

Reversible Jump MCMC

1. Current state of the chain = current state of the particles $(k(t), \phi^{(i)}(t))$.
2. Iteration j for the ith particle, $i = 1, \ldots, N$:
   - Sample $u \sim U_{[0,1]}$
   - if $(u < b_k)$ then "birth move"
   - else if $(u < b_k + d_k)$ then "death move"
   - else if $(u < b_k + d_k + s_k)$ then "split move"
   - else if $(u < b_k + d_k + s_k + m_k)$ then "merge move"
   - else update all the parameters
3. $j \leftarrow j + 1$, goto step 2

5.4.1 Update move

If the update move is selected, all the parameters are resampled, with fixed model order $k(t)$. The proposal distribution $q(\phi_k^\star | \phi_k(t))$ for the candidate $\phi_k^\star$ is given using eq. (5.1) as

$$q(\phi_k^\star | \phi_k(t)) = \mathcal{N}(\phi_k(t), \sigma_v^2(t) I_{k(t)}). \tag{5.24}$$

This candidate function is different from the optimal function described by eq. (5.19). When the optimal importance function is used in the MCMC step with wrong parameters, the Markov chain does not mix efficiently and most candidates are rejected. By using a less informative candidate function, more candidates are accepted, and a higher degree of diversity amongst the particles is achieved.

By substituting eqs. (5.24) and (5.16) into eq. (2.10), after some algebraic manipulation we obtain the following expression for the acceptance ratio for the update move,

$$r_{update} = \frac{\exp^{\frac{-1}{\sigma_w^2}\boldsymbol{y}'(t)\boldsymbol{P}_S^{\perp}(\boldsymbol{\phi}^{\star}_{k(t)})\boldsymbol{y}(t)}}{\exp^{\frac{-1}{\sigma_w^2}\boldsymbol{y}'(t)\boldsymbol{P}_S^{\perp}(\boldsymbol{\phi}_{k(t)})\boldsymbol{y}(t)}}. \tag{5.25}$$

The candidate $\boldsymbol{\phi}_k^{\star}$ is accepted as the ith particle at time t, with probability

$$\xi_{update} = \min(r_{update}, 1). \tag{5.26}$$

The amplitude parameters, which are required for data association in section V, are estimated directly from eq. (5.15). The noise variance parameters σ_w^2 and σ_v^2, which are required in eqs. (5.24) and (5.25), are estimated directly from eqs. (5.17) and (5.18).

The update move is summarized with the following schema.

Update Move

- Propose a candidate $\boldsymbol{\phi}^{\star}$ from eq. (5.24).
- Evaluate ξ_{update} with eq. (5.26)
- Sample $u \sim U_{[0,1]}$
- if $(u \leq \xi_{update})$ then
  - The state of the Markov Chain becomes $(k, \boldsymbol{\phi}^{\star})$,
  - If desired, estimate $\boldsymbol{a}|\boldsymbol{\phi}$ from eq. (5.15)
  - update σ_v^2 and σ_w^2 from eqs. (5.17) and (5.18) respectively.

  else it remains at $(k, \boldsymbol{\phi})$

5.4.2 Birth and death moves

The birth move proposes a candidate in a higher dimension model, as opposed to the death move, which in turn proposes a candidate in a lower dimension model.

For the birth move, a new source ϕ_c is proposed at random from the prior distribution for the directions of arrival

$$\phi^*_{k(t)+1} = [\phi_{k(t)}, \quad \phi_c].$$

After straightforward algebra, the acceptance ratio for the birth move is (using eq. (5.16))

$$r_{birth} = \frac{\exp^{\frac{-1}{\sigma^2_w}\mathbf{y}'(t)\mathbf{P}^{\perp}_S(\phi^*_{k(t)+1})\mathbf{y}(t)}}{\exp^{\frac{-1}{\sigma^2_w}\mathbf{y}'(t)\mathbf{P}^{\perp}_S(\phi_{k(t)})\mathbf{y}(t)}} \times \frac{1}{(1+\delta^2)(k+1)},$$

with corresponding acceptance probability given by

$$\xi_{birth} = \min\{r_{birth}, 1\}. \tag{5.27}$$

If the move is accepted, then the amplitudes and noise variance parameters are then updated in the same manner as described for the update move process.

The death move is just the reverse. A source, amongst the $(k+1)$ sources is randomly selected to be removed. It is straightforward to show the new candidate, of dimension k, is then accepted with probability

$$\xi_{death} = \min\{\frac{1}{r_{birth}}, 1\}. \tag{5.28}$$

The schemas for the birth and death moves are similar to that for the update move with appropriate changes. However, for the birth move, if the candidate is accepted, the new state becomes $(k+1, \phi_{k+1})$, otherwise, it remains at (k, ϕ_k). For the death move, if the candidate is accepted, the new state becomes (k, ϕ_k), otherwise, it remains at $(k+1, \phi_{k+1})$.

Birth Move

- Propose a candidate vector of directions of arrival as follows:

$$\boldsymbol{\phi}_{k+1}^{\star} = [\boldsymbol{\phi}_k, \quad \phi_c]$$

- Evaluate ξ_{birth} with eq. (5.27)
- Sample $u \sim U_{[0,1]}$
- if $(u \leq \xi_{birth})$ then
  - The state of the Markov Chain becomes $(k+1, \boldsymbol{\phi}_{k+1})$,
  - If desired, estimate $\boldsymbol{a}_{k+1}|k+1, \boldsymbol{\phi}_{k+1}$ from eq. (5.15)
  - update σ_v^2 and σ_w^2 from eqs. (5.17) and (5.18) respectively.

  else it remains at $(k, \boldsymbol{\phi}_k)$.

Death Move

- Pick at random one direction of arrival among the $(k+1)$ existing DOAs
- Evaluate ξ_{death} with eq. (5.28)
- Sample $u \sim U_{[0,1]}$
- if $(u \leq \xi_{death})$ then
  - The state of the Markov Chain becomes $(k, \boldsymbol{\phi}_k)$,
  - If desired, estimate $\boldsymbol{a}_k|k, \boldsymbol{\phi}_k$ from eq. (5.15)
  - update σ_v^2 and σ_w^2 from eqs. (5.17) and (5.18) respectively.

  else it remains at $(k+1, \boldsymbol{\phi}_{k+1})$.

■

5.4.3 Split and merge moves

The split move proposes a candidate in a higher dimension model, as opposed to the merge move, which in turn proposes a candidate in a lower dimension model. The split move is designed to handle the situation where two DOA tracks separate after crossing. The merge move corresponds to the case where two adjacent DOA tracks coalesce before crossing.

For the split move, two new sources $\phi_j^\star$ and $\phi_{j+1}^\star$ are proposed as a replacement of the source $\phi_j \in \boldsymbol{\phi}_k$, selected at random amongst the existing k sources

$$\phi_j^\star = \phi_j - uW,$$
$$\phi_{j+1}^\star = \phi_j + uW,$$

where W is some fixed and known parameter and $u \sim U_{[0,1]}$. These new angles are inserted in the parameter vector, replacing the jth element, to produce a candidate vector $\boldsymbol{\phi}_{k+1}^\star$ for the split move as follows

$$\boldsymbol{\phi}_{k+1}^\star = \left[\boldsymbol{\phi}_k\big(1:(j-1)\big), \phi_j^\star, \phi_{j+1}^\star, \boldsymbol{\phi}_k\big((j+1):k\big)\right].$$

The merge move is just the reverse of the split move. A source, ϕ_j, amongst the first k of the $(k+1)$ sources is randomly selected. A candidate source $\phi_j^\star$ is proposed as the superposition of two adjoining sources

$$\phi_j^\star = (\phi_j + \phi_{j+1})/2.$$

This combined angle is inserted in the parameter vector, replacing elements j and $(j+1)$ to produce the candidate vector for the merge move as follows

$$\boldsymbol{\phi}_k^\star = \left[\boldsymbol{\phi}_k\big(1:(j-1)\big), \phi_j^\star, \boldsymbol{\phi}_k\big((j+2):(k+1)\big)\right].$$

The proposal functions for this pair of moves are defined as

$$q(\boldsymbol{\phi}_{k+1}^\star, k+1|\boldsymbol{\phi}_k, k) = p(k+1)p(\boldsymbol{\phi}_{k+1}^\star|\boldsymbol{\phi}_k) \propto \frac{\Lambda^k}{k!}\frac{1}{(Wk)},$$
$$q(\boldsymbol{\phi}_k^\star, k|\boldsymbol{\phi}_{k+1}, k+1) = p(k) \div \binom{k}{1} \propto \frac{\Lambda^k}{k!}\frac{1}{(k)}.$$

As opposed to the previously defined moves, the split/merge moves require the evaluation of the Jacobian term in eq. (2.10). The Jacobian can be evaluated as

$$J = \left\| \begin{matrix} \frac{\partial u}{\partial \phi_1} & \frac{\partial \phi}{\partial \phi_1} \\ \frac{\partial u}{\partial \phi_2} & \frac{\partial \phi}{\partial \phi_2} \end{matrix} \right\| = \frac{1}{2W}.$$

After some straightforward algebra, using eq. (5.16) and the candidate functions defined above, the acceptance ratio for the split move is given as

$$\begin{aligned} r_{split} = & \frac{\exp\left[\frac{-1}{\sigma_w^2}\boldsymbol{y}'(t)\boldsymbol{P}_S^{\perp}(\boldsymbol{\phi}^{\star}_{k(t)+1})\boldsymbol{y}(t)\right]}{\exp\left[\frac{-1}{\sigma_w^2}\boldsymbol{y}'(t)\boldsymbol{P}_S^{\perp}(\boldsymbol{\phi}_{k(t)})\boldsymbol{y}(t)\right]} \\ & \times \frac{\exp\left[\frac{-1}{2\sigma_v^2}\left(\phi^{\star}_{k(t)+1} - \phi_{k(t)+1}(t-1)\right)'\left(\phi^{\star}_{k(t)+1} - \phi_{k(t)+1}(t-1)\right)\right]}{\exp\left[\frac{-1}{2\sigma_v^2}(\boldsymbol{\phi}_{k(t)}(t) - \boldsymbol{\phi}_{k(t)}(t-1))'(\boldsymbol{\phi}_{k(t)}(t) - \boldsymbol{\phi}_{k(t)}(t-1))\right]} \\ & \times \frac{1}{2\sqrt{2\pi}\sigma_v(1+\delta^2)}, \end{aligned} \tag{5.29}$$

$$\xi_{split} = \min\{r_{split}, 1\}. \tag{5.30}$$

This acceptance probability does not depend on the hyper-parameter W. This split move is attempted only if no original sources fall between the two proposed candidates, such that the reverse move, the merge move, makes the sampling reversible. This measure is necessary to satisfy the reversibility condition (Green, 1995), which in turn is sufficient for the invariant distribution of the Markov chain to converge to the desired density.

Similar to the death move case, it is straightforward to show that the candidate vector $\boldsymbol{\phi}^{\star}_{k}$, of dimension k is accepted with probability

$$\xi_{merge} = \min\{\frac{1}{r_{split}}, 1\}.$$

For both the split and merge moves, the amplitude and noise variance parameters are also updated in the manner described for the update move.

Split Move

- Pick at random one DOA element ϕ_j, amongst the first k existing directions. Evaluate:

$$\phi_j^\star = \phi_j - uW$$
$$\phi_{j+1}^\star = \phi_j + uW$$

- Evaluate ξ_{split} with eq. (5.30)
- Sample $u \sim U_{[0,1]}$
- if $(u \leq \xi_{split})$then
  - The state of the Markov Chain becomes $(k+1, \phi_{k+1}^\star)$,
  - If desired, estimate $a_{k+1}|k+1, \phi_{k+1}^\star$ from eq. (5.15)
  - update σ_v^2 and σ_w^2 from eqs. (5.17) and (5.18) respectively.

  else it remains at (k, ϕ_k) .

■

Merge Move

- Pick at random two adjacent directions of arrival among the $(k+1)$ existing DOA

$$\phi_j^\star = (\phi_j + \phi_{j+1})/2$$

- Evaluate ξ_{merge} with eq. (5.28)
- Sample $u \sim U_{[0,1]}$
- if $(u \leq \xi_{merge})$ then
  - The state of the Markov Chain becomes $(k, \phi_k^\star)$,
  - If desired, estimate $a_k|k, \phi_k^\star$ from eq. (5.15)
  - update σ_v^2 and σ_w^2 from eqs. (5.17) and (5.18) respectively.

else it remains at $(k+1, \phi_{k+1})$.

■

5.5 Simulation results

The proposed algorithm is now verified with simulated data, generated for the true number of sources $k_o = 2$, with parameters described in Table 5.1. The received array is circular and composed of $M = 8$ elements. The parameters and observations evolve according to the state space model eqs. (5.1) - (5.3), with an initial SNR of 20dB, which is defined as

$$SNR = \frac{\boldsymbol{a}'(1)\boldsymbol{a}(1)}{\sigma_w^2}. \tag{5.31}$$

Parameter	σ_v^2	σ_w^2	$\boldsymbol{\phi}(0)$	$\boldsymbol{a}(1)$	σ_a^2
Value	5deg^2.	0.15	$[70°, 110°]$	$[2-2j, 4+j]$	0.0707

Table 5.1: Parameters of the state-space model for simulated data

The hyperparameter is set to $\delta^2 = 100$ in accordance with the initial SNR value, and the hyperparameters $\nu_0, \nu_1, \gamma_0, \gamma_1$ are all set to zero, corresponding to a non-informative prior on the respective variances. (Brief experiments have verified performance is robust to the values of the hyperparameters). The DOAs are simulated using a first–order random walk, with variance σ_v^2, thus generating a nonstationary DOA environment. The parameter k for the model order is initialized at $k(1) = 1$ or $k(1) = k_{max}$ (where k_{max} is the maximum allowable model order) and all the other parameters are initialized at random over their respective parameter space. The initialization is therefore done blindly. The particle filter uses $N = 300$ particles. We compare the performance of the SIS method with that of the PASTd method (Yang, 1995a) with joint rank estimation.

5.5.1 First scenario: change point in the number of sources

In this first scenario, the number of sources is initialized to $k = 1$, and is complicated by a change point at $t = 50$, when one of the sources vanishes. Figure 5.1 shows the results obtained with the particle filter. It shows that the directions of arrival are well traced by

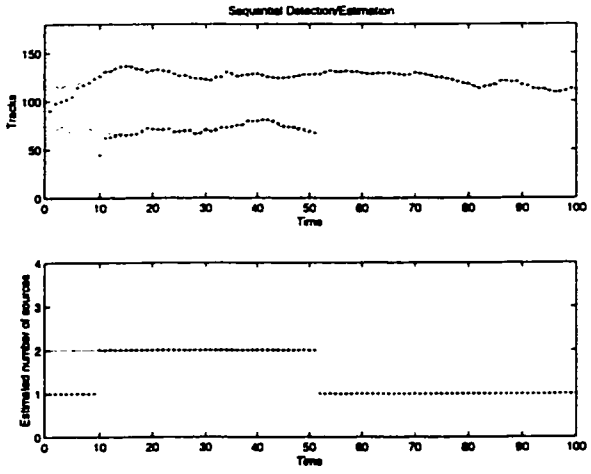

Figure 5.1: Top: Sequential MAP estimates of the directions of arrival, and Bottom: the number of detected signals, each vs. time, using the particle filter (Scenario A), initialized to $k(1) = 1$. The finely dotted line shows the true values, and the coarsely dotted line gives the estimated values.

their estimates throughout the entire tracking process and that the number of sources is correctly estimated. Initially, since $k = 1$, the algorithm tracks towards only one source, until a birth move is accepted. The second source is then detected and later on correctly estimated. The change-point was detected within one sample period and the estimates of the parameters, following the change point, quickly adapt to the true values. The same scenario is used in Figure 5.2, but the model order is initially set to $k_{max} = 5$. Within 35 observations, the correct model order is determined, and the DOA trajectories follow the true values. Again, the change point is detected within one sample period. It is therefore seen that the proposed particle filter approach performs well under nonstationary conditions and variations in the initial values of the parameters.

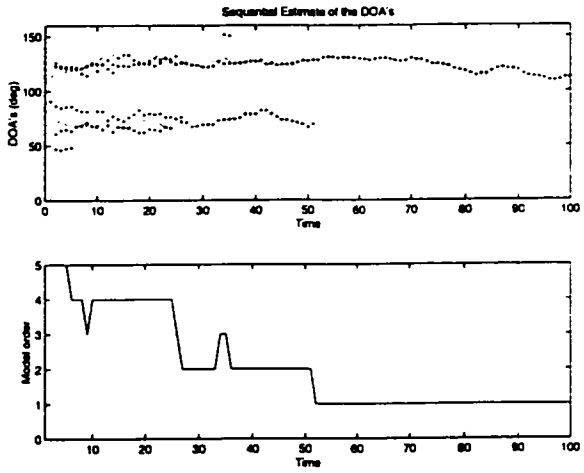

Figure 5.2: Same as Figure 5.1, except the filter is initialized to $k(1) = k_{max} = 5$. The finely dotted line shows the true values, and the coarsely dotted line gives the estimated values.

We now discuss the comparison of the particle filter results with those from the PASTd algorithm. Due to the nonstationarity, the PASTd algorithm fails to give meaningful results in this environment. In the nonstationary case, rank estimation fails as the number of sources is always over-estimated and the DOA estimates obtained are smoothed versions of the true values. This behaviour is typical of any algorithm which is based on time-averaged statistics.

We therefore consider a simulation scenario which is more favourable to the PASTd algorithm. In this case, the source DOAs are held steady at $\pm 30°$ for the first 500 observations. As shown on Figure 5.3, the PASTd algorithm with joint rank estimation is much slower to converge to the true number of sources than the particle filter case, both initially and after the change point. For this favourable case, the DOA estimates produced by the algorithm include values which are close to the true values.

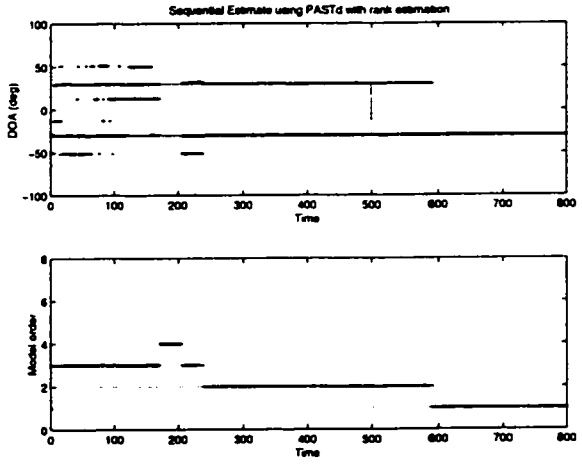

Figure 5.3: Sequential estimates of directions of arrival using PASTd and root-MUSIC (Scenario A). The finely dotted line shows the true values, and the coarsely dotted line gives the estimated values.

5.5.2 Second scenario: sources crossing

In this subsection, we apply the previously developed algorithm to a scenario where the source DOAs cross, thus verifying the performance of the split/merge move combinations. In this scenario, the variance σ_v^2 of the update equation is reduced to $\sigma_v^2 = 1$ deg^2 .

The same initial parameters as for the previous case are used. As verified in Figure 5.4, the algorithm performs well under these adverse conditions. As is evident from the figure, the number of detected sources varies cleanly from $2-1$ and back again in the region where the tracks cross. Also, the algorithm shows no apparent tendency towards outliers in the DOA estimates in the cross region, as is commonly exhibited with other algorithms.

When the sources cross, the steering matrix $\boldsymbol{S}(\cdot)$ becomes rank one and hence the two targets are seen as a single source, which explains the apparent miss-detection of a second source during the period of time when they are very close. This scenario is more difficult than the ones presented in Yang (1995b), as the sources here follow steep trajectories and the variance between snapshots is high, making estimation of statistics by time-averaging very difficult. As expected under these conditions, the performance of the PASTd method

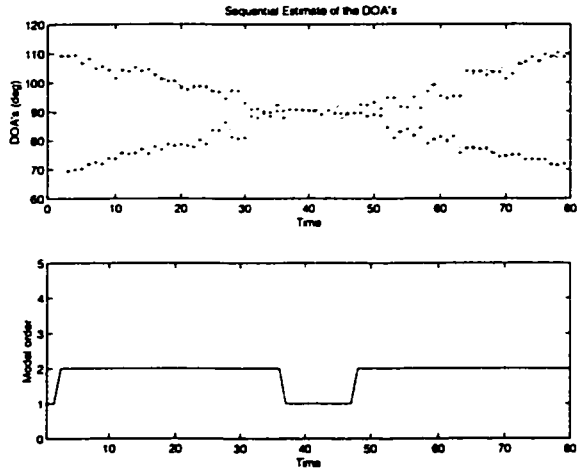

Figure 5.4: Top: Sequential estimates of the directions of arrival, using the particle filter (Scenario B), and Bottom: corresponding number of detected sources. The finely dotted line shows the true values, and the coarsely dotted line gives the estimated values.

is significantly degraded, as shown in Figure 5.5.

5.5.3 Data association via matching of probability distributions

The data association problem in our case becomes determining the most likely correspondence of trajectories with DOA enumeration, particularly after the crossing of trajectories. In our numerical framework, we can use the estimated posterior distributions of the parameters to perform the data association task. For example, we can compare first and second moments of the approximate marginal distributions of the amplitude parameters before and after crossing of the targets. Many other possibilities exist. Figure 5.6 shows an example of data association using the (unnormalized) marginal distribution of the amplitudes of the two sources for the above scenario. The first column shows the histograms of the two amplitudes before the targets cross, while the second column shows the histograms after crossing. In this case, from visual inspection, it is clear the diagonally opposite distributions match.

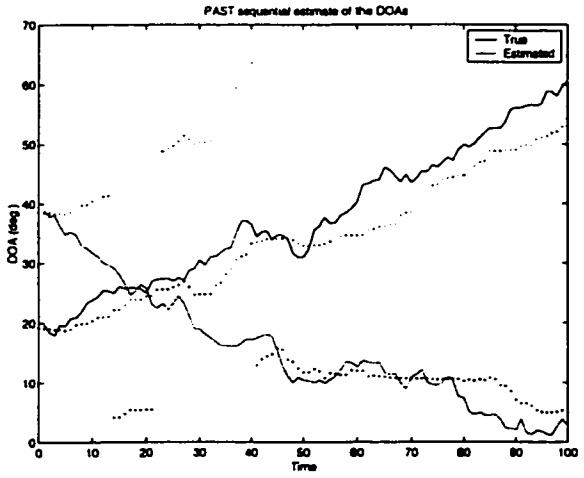

Figure 5.5: Sequential estimates of directions of arrival using PASTd and root-MUSIC (Scenario B). The finely dotted line shows the true values, and the coarsely dotted line gives the estimated values.

5.5.4 Approximate joint confidence regions

An advantage of numerical approaches to parameter estimation is that an approximation to the joint confidence region of the parameters is readily established from the histogram approximating the joint posterior distribution of the parameters. Figure 5.7 shows a contour plot for the joint histogram of the DOAs for the second scenario, at the 7th sample. The probability level associated with the joint confidence region is determined by integrating inside the respective contour of the normalized distribution.

5.5.5 Performance of the method

In this subsection, the proposed algorithm was applied to 50 different scenarios of 50 observations, for different values of SNR, in order to estimate the variance of the estimate as a function of the SNR. A single source is tracked, and this information is used to help the implementation of the PASTd algorithm. For each run, the first 25 sequential estimates (considered in the acquisition mode) were discarded and the following 25 estimates were used to get one sample of the variance. The results are compared with those obtained with

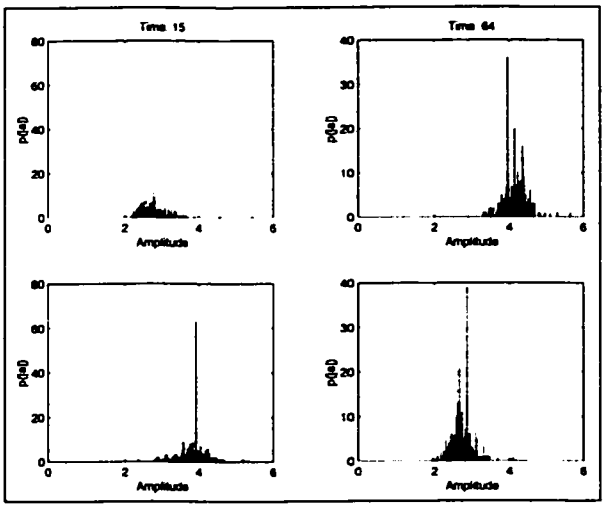

Figure 5.6: Data association via the marginal posterior distributions of the amplitudes.

the PASTd method in Figure 5.8.

The limiting factor of the particle filter at high SNR is the number of particles. The estimates are obtained as the MAP estimates of a histogram. Clearly, using more particles will improve the variance of the estimates. The following figure, Figure 5.9, shows the results. The same data was used to obtain each point of the figure, with the parameters set as described in Table 5.1.

5.5.6 Further discussion

It is easily shown that as $N \rightarrow \infty$, the global optimum of the desired posterior distribution coincides with the most heavily-weighted histogram bin corresponding to the particles. In practice, the global optimum is achieved within a histogram bin-width with finite N with high probability. Thus, the global optimum can be attained by a simple search, instead of a complicated global optimization over what is shown in Figure 5.7 to be a multi-modal surface.

The computational expense of the particle filter approach is fairly high, requiring $\mathcal{O}(N)$ function evaluations each time step. However, the evaluation of the particles is easily

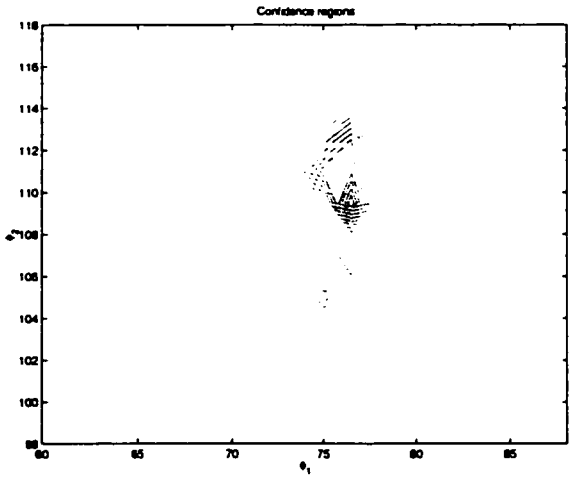

Figure 5.7: Contour lines of the approximate posterior distribution of the DOAs at the 7th sample.

parallelizable, and this order of computation does not necessarily compare unfavourably with that of a global optimization procedure. Further, the relative computational expense of the method is offset by its advantages; namely, a joint detection capability and improved performance in nonstationary environments.

5.6 Conclusion

In this chapter, a particle filter that includes a reversible jump MCMC with two new components, the merge and split moves, is used for sequential joint detection and estimation of an unknown number of sources and their corresponding directions of arrival.

The algorithm compares favourably to an established approach in computer simulations. The algorithm proved robust to changes in initial values and shows robust convergence to the global minimum. The superior performance of the particle filter over conventional methods which use time-averaged statistics in nonstationary environments has been clearly indicated. Examples of data association of tracks before and after crossing using histogram matching, and of the joint confidence region of the parameter estimates, have been given.

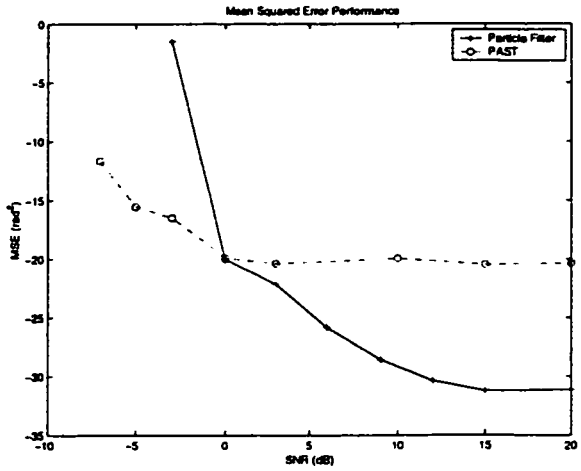

Figure 5.8: Performance of the tracking versus SNR.

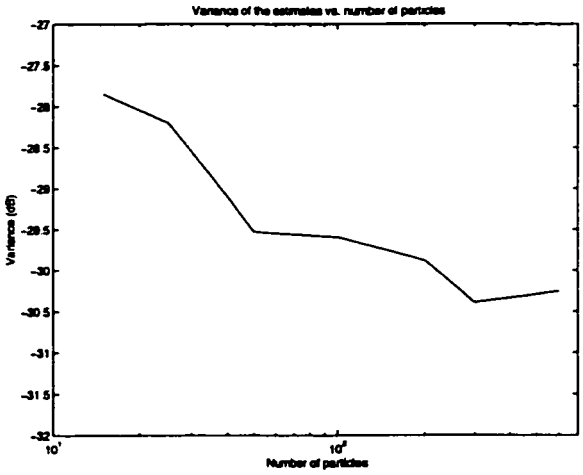

Figure 5.9: Performance of the tracking versus the number of particles.

Chapter 6

Conclusion

This thesis discussed new approaches to various classical problems in array signal processing, using modern numerical Bayesian methods. We showed how the *Markov Chain Monte Carlo* (MCMC) and the *Sequential Importance Sampling* (SIS) methods present new outlooks and offer many advantages to problems in the field.

This thesis presented the application of the Markov Chain Monte Carlo methods to the *joint problem* of detection and estimation in coloured noise and the sequential implementation of Monte Carlo methods applied to probabilistic dynamic systems.

Very few methods tackled this joint problem of detection of the model order and estimation of the parameters, particularly in unknown coloured noise with arbitrary covariance and none are as flexible and adaptive.

Three problems were addressed in the course of this thesis.

1. A method for joint detection of the number of sources and estimation of their respective directions of arrival in coloured noise using MCMC methods was developed.

   The method uses a single array of sensors and analytically integrates out the unknown noise covariance matrix to leave a marginalized posterior distribution that is only a function of the parameters of interest. The numerical optimization is accomplished with the Reversible Jump MCMC method.

   A criteria to ensure the consistency of the method was developed and its performance

was demonstrated by simulations. Also, the performance of the estimation of the parameters was compared successfully with the Cramér-Rao bound.

2. The second algorithm represents an extension of the first one with the addition of the joint estimation of the times of arrival of the pulses. The emphasis lies on the verification of the method using real propagation measurements taken around the McMaster University campus.

   In the spirit of channel sounding or channel characterization, the received signal is composed of the superposition of multiple delayed and attenuated copies of the transmitted signal. The objective is to detect the number of multipath components and estimate their respective directions of arrival and excess propagation delays.

   Both methods were successfully applied to real data, acquired on campus with a channel sounder during the extensive measurement campaign previously mentioned. It shows how these methods work in real-life.

3. The final part of this thesis focuses on the sequential implementation of the Monte Carlo methods in particle filters for probabilistic dynamic systems. A lot of insight on how to successfully implement such methods are provided throughout the chapter.

   This algorithm proposes a solution to the problem of tracking an unknown number of sources in white noise. At each snapshot, the joint posterior distribution of the number of sources and their directions of arrival is recursively updated and optimized.

   This method allows for the instantaneous estimate of the number of sources and their evolving directions of arrival. In addition, by obtaining the marginal posterior distribution of the nuisance parameters, one could perform a very simple, yet quite effective, form of data association.

6.1 Contributions to the scientific literature

All the work presented in Chapters 3, 4 and 5 was published in various conferences and three journal papers were submitted for publication.

- Journal Papers
  - *Particle Filters for Tracking an Unknown Number of Sources.* (Larocque *et al.*, 2001b) Submitted to IEEE Transactions on Signal Processing, May, 2001.
  - *Reversible Jump MCMC for Joint Detection and Estimation of Directions of Arrival in Coloured Noise.* (Larocque and Reilly, 2000b) Submitted to IEEE Transactions on Signal Processing for the Special Issue on MCMC, October, 2000.
  - *Wideband Channel Characterization using the Reversible Jump MCMC.* (Larocque and Reilly, 2001b) Submitted to IEEE Transactions on Vehicular Technology, June 2001.
- Conference Papers
  - *Sequential Monte Carlo for Spatial Signal Separation and Restoration.* (Ng, Reilly and Larocque) 2001 Workshop on Maximum Entropy and Bayesian Methods in Science and Engineering.
  - *Wideband Channel Characterization in Coloured Noise Using the Reversible Jump MCMC.* (Larocque and Reilly, 2001a) ICASSP 2001.
  - *On the Implementation of Particle Filters for DOA Tracking.* (Larocque *et al.*, 2001a) Lecture at ICASSP 2001.
  - *Reversible Jump MCMC for Joint Detection and Estimation of Directions of Arrival in Coloured Noise.* (Larocque and Reilly, 2000c) IMA 2000, Nominated for best student paper award.
  - *Application of the Reversible Jump Markov Chain Monte Carlo Method to Real-Life Propagation Measurements.* (Larocque and Reilly, 2000a) ICASSP 2000.
  - *On the Calibration of an Antenna Array.* (Tranter *et al.*, 1999) Symposium at Virginia Tech, 1998.

Appendix A

Probability Density Functions

Definitions of selected probability functions that appear in the thesis (Bernardo and Smith, 1994).

Name	Symbol	Functional Form	c
Beta	$Be(\alpha,\beta)$	$cz^{\alpha-1}(1-z)^{\beta-1}\mathbb{I}_{[0,1]}(z)$	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$
Gamma	$G(\alpha,\beta)$	$cz^{\alpha-1}\exp(-\beta z)\mathbb{I}_{[0,+\infty)}(z)$	$\frac{\beta^\alpha}{\Gamma(\alpha)}$
Inverted Gamma	$IG(\alpha,\beta)$	$cz^{-\alpha-1}\exp(-\beta/z)\mathbb{I}_{[0,+\infty)}(z)$	$\frac{\beta^\alpha}{\Gamma(\alpha)}$
Real Normal	$N(\mathbf{m},\Sigma)$	$c\exp(-\frac{1}{2}(\mathbf{z}-\mathbf{m})^T\Sigma^{-1}(\mathbf{z}-\mathbf{m}))$	$\lvert 2\pi\Sigma\rvert^{-1/2}$
Complex Normal	$N(\mathbf{m},\Sigma)$	$c\exp(-(\mathbf{z}-\mathbf{m})'\Sigma^{-1}(\mathbf{z}-\mathbf{m}))$	$\lvert\pi\Sigma\rvert^{-1}$
Complex Wishart	$W(\mathbf{A})$	$\frac{1}{c}\lvert\mathbf{A}\rvert^{N-p}\exp(-\mathrm{tr}\Sigma^{-1}\mathbf{A})$	$\pi^{\frac{1}{2}p(p-1)}\lvert\Sigma\rvert^N$ $\times[\prod_{i=1}^p\Gamma(N-i+1)]$
Poisson	$P(\lambda)$	$c\frac{\lambda^z}{z!}\mathbb{I}_{\mathbf{N}}(z)$	$\exp(-\lambda)$
Exponential	$Exp(\lambda)$	$c\exp(-z)$	λ

Table A.1: Definition of selected probability density functions

Appendix B

Recursivity of Particle Filters

This appendix shows how to obtain the recursive form of the posterior equation (2.13)

$$p(\boldsymbol{x}_{1:t+1}|\boldsymbol{y}_{1:t+1}) = p(\boldsymbol{x}_{1:t}|\boldsymbol{y}_{1:t})\frac{p(\boldsymbol{y}_{t+1}|\boldsymbol{x}_{t+1})p(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t)}{p(\boldsymbol{y}_{t+1}|\boldsymbol{y}_{1:t})}.$$

We start with the application of Bayes' theorem,

$$p(\boldsymbol{x}_{1:t+1}|\boldsymbol{y}_{1:t+1}) = \frac{p(\boldsymbol{y}_{1:t+1}|\boldsymbol{x}_{1:t+1})p(\boldsymbol{x}_{1:t+1})}{p(\boldsymbol{y}_{1:t+1})}.$$

Since the observation noise is assumed *iid*, the conditional terms are assumed independant and the total likelihood term can be expanded one step into

$$p(\boldsymbol{x}_{1:t+1}|\boldsymbol{y}_{1:t+1}) = \frac{p(\boldsymbol{y}_{t+1}|\boldsymbol{x}_{t+1})p(\boldsymbol{y}_{1:t}|\boldsymbol{x}_{1:t})p(\boldsymbol{x}_{1:t+1})}{p(\boldsymbol{y}_{1:t+1})}.$$

Applying once more Bayes' theorem on the diminished total likelihood term, we get,

$$p(\boldsymbol{x}_{1:t+1}|\boldsymbol{y}_{1:t+1}) = p(\boldsymbol{x}_{1:t}|\boldsymbol{y}_{1:t})p(\boldsymbol{y}_{t+1}|\boldsymbol{x}_{t+1})\frac{p(\boldsymbol{y}_{1:t})p(\boldsymbol{x}_{1:t+1})}{p(\boldsymbol{y}_{1:t+1})p(\boldsymbol{x}_{1:t})}.$$

Expanding the total density function of $\boldsymbol{x}$ as (also for $\boldsymbol{y}$),

$$p(\boldsymbol{x}_{1:t+1}) = p(\boldsymbol{x}_{t+1}|\boldsymbol{x}_{1:t})p(\boldsymbol{x}_t|\boldsymbol{x}_{1:t-1})\cdots p(\boldsymbol{x}_o),$$

and simplifying the common terms in the numerator and denominator, we obtain

$$p(\boldsymbol{x}_{1:t+1}|\boldsymbol{y}_{1:t+1}) = p(\boldsymbol{x}_{1:t}|\boldsymbol{y}_{1:t})p(\boldsymbol{y}_{t+1}|\boldsymbol{x}_{t+1})\frac{p(\boldsymbol{x}_{t+1}|\boldsymbol{x}_{1:t})}{p(\boldsymbol{y}_{t+1}|\boldsymbol{y}_{1:t})}.$$

Applying the first order Markovian property of the model, the previous equation simplifies to

$$p(\boldsymbol{x}_{1:t+1}|\boldsymbol{y}_{1:t+1}) = p(\boldsymbol{x}_{1:t}|\boldsymbol{y}_{1:t})\frac{p(\boldsymbol{y}_{t+1}|\boldsymbol{x}_{t+1})p(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t)}{p(\boldsymbol{y}_{t+1}|\boldsymbol{y}_t)},$$

which is the desired recursive form. QED.

When the proposal function satisfies the following recursivity condition

$$q(\boldsymbol{x}_{1:t+1}|\boldsymbol{y}_{1:t+1}) = q(\boldsymbol{x}_{1:t}|\boldsymbol{y}_{1:t})q(\boldsymbol{x}_{t+1}|\boldsymbol{x}_{1:t}, \boldsymbol{y}_{1:t+1}),$$

the importance weights can be evaluated recursively.

The importance weights are defined as

$$\tilde{w}(t+1) = \frac{p(\boldsymbol{x}_{1:t+1}|\boldsymbol{y}_{1:t+1})}{q(\boldsymbol{x}_{1:t+1}|\boldsymbol{y}_{1:t+1})}.$$

Replacing the distributions by their corresponding definitions, we get

$$\tilde{w}(t+1) = \frac{p(\boldsymbol{x}_{1:t}|\boldsymbol{y}_{1:t})p(\boldsymbol{y}_{t+1}|\boldsymbol{x}_{t+1})p(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t)}{p(\boldsymbol{y}_{t+1}|\boldsymbol{y}_t)q(\boldsymbol{x}_{1:t}|\boldsymbol{y}_{1:t})q(\boldsymbol{x}_{t+1}|\boldsymbol{x}_{1:t}, \boldsymbol{y}_{1:t+1})}.$$

Recognizing that

$$\tilde{w}(t) = \frac{p(\boldsymbol{x}_{1:t}|\boldsymbol{y}_{1:t})}{q(\boldsymbol{x}_{1:t}|\boldsymbol{y}_{1:t})},$$

we can obtain the recursive form of the unnormalized importance weights as,

$$\tilde{w}(t+1) = w(t)\frac{p(\boldsymbol{y}_{t+1}|\boldsymbol{x}_{t+1})p(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t)}{q(\boldsymbol{x}_{t+1}|\boldsymbol{x}_{1:t}, \boldsymbol{y}_{1:t+1})}.$$

Bibliography

Abramowitz, M. and Stegun, I. (1965). *Handbook of Mathematical Functions.* Dover Publications, New York.

Agrawal, M. and Prasad, S. (2000). A modified likelihood function approach to DOA estimation in the presence of unknown spatially correlated Gaussian noise using a uniform linear array. *IEEE Transactions on Signal Processing*, **48**(10), 2743–2749.

Akaike, H. (1974). A new look at statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716–723.

Andrieu, C. (1997). *Méthodes MCMC pour l'analyse Bayesienne de modèles de régression paramétrique non-linéaire. Application à l'analyse de raies et à la déconvolution impulsionnelle.* Ph.D. thesis, Université de Cergy-Pontoise, France.

Andrieu, C. and Doucet, A. (1999). Joint Bayesian model selection and estimation of noisy sinusoids via reversible jump MCMC. *IEEE Transactions on Signal Processing*, **47**(10), 2667–2676.

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

Andrieu, C., Freitas, N. D., and Doucet, A. (1999). Sequential MCMC for Bayesian model selection. In *Proceedings of the Internationl Workshop on Higher Order Statistics*, Ceasarea, Isreal.

Andrieu, C., Djuric, P. M., and Doucet, A. (2001). Model selection by MCMC computation. *Signal Processing*, **81**(1), 19–37.

Bernardo, J. and Smith, A. (1994). *Bayesian Theory*. John Wiley and Sons, New York.

Blanz, J. J., Papathanassiou, A., Haardt, M., Furio, I., and Baier, P. W. (2000). Smart antennas for combined DOA and joint channel estimation in time-slotted CDMA mobile radio systems with joint detection. *IEEE Transactions on Vehicular Technology*, **49**(2), 293–306.

Box, G. and Tiao, G. (1973). *Bayesian Inference in Statistical Analysis*. Addison Wesley, Reading, M.A.

Capon, J. (1969). High resolution frequency-wavenumber spectrum analysis. *Proceedings of the IEEE*, **57**, 1408–1418.

Chen, W., Reilly, J., and Wong, K. (1996). Detection of the number of signals in noise with banded covariance matrices. *IEE Proceedings on Radar, Sonar and Navigation*, **143**, 289–294.

Chen, W.-G. (1991). *Detection of the number of signals in array signal processing*. Ph.D. thesis, Department of Electrical and Computer Engineering, McMaster University, Hamilton, Ontario, Canada.

Cho, C.-M. and Djuric, P. M. (1995). Bayesian detection and estimation of cisoids in colored noise. *IEEE Transactions on Signal Processing*, **43**(12), 2943–2951.

Djuric, P. M. (1996). A model selection rule for sinusoids in white Gaussian noise. *IEEE Transactions on Signal Processing*, **44**(7), 1744–1751.

Djuric, P. M. (2000). Sequential estimation of random parameters under model uncertainty. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, Istanbul, Turkey.

Doucet, A. (1998). On sequential simulation-based methods for Bayesian filtering. Technical Report TR.310, University of Cambridge, Department of Engineering, Signal Processing Group, England.

Doucet, A., Godsill, S., and Andrieu, C. (2000). On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, **10**, 197–208.

Doucet, A., Gordon, N., and Krishnamurthy, V. (2001a). Particle filters for state estimation of jump Markov linear systems. *IEEE Transactions on Signal Processing*, **49**(3), 613–624.

Doucet, A., de Freitas, N., and Gordon, N., editors (2001b). *Sequential Monte Carlo in Practice*. Springer-Verlag, New York. To appear.

Fuchs, J. (1992). Estimation of the number of signals in the presence of unknown correlated sensor noise. *IEEE Transactions on Signal Processing*, **40**, 1053–1061.

Gamerman, D. (1997). *Markov Chain Monte Carlo. Stochastic Simulation for Bayesian Inference*. Chapman and Hall, Suffolk, England.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**(6), 721–732.

Gershman, A. B., Stoica, P., Pesavento, M., and Larsson, E. (2001). Stochastic Cramér-Rao bound for direction estimation in unknown noise fields. *IEEE Transactions on Signal Processing*. To appear.

Gilks, W. and Berzuini, C. (1998). Following a moving target - Monte Carlo inference for dynamic Bayesian models. Technical report, University of Pavia, Department of Computer Sciences and Systems, Italy.

Gilks, W., Richardson, S., and Spiegelhalter, D. (1998). *Markov Chain Monte Carlo in Practice.* Chapman and Hall, New York.

Godsill, S. (1998). On the relationship between MCMC model uncertainty methods. Technical Report TR.305, Cambridge University, Engineering Department, England.

Goodman, N. (1963). Statistical analysis based on a certain multivariate complex Gaussian distribution. *Annals of Mathematical Statistics*, **34**, 152–171.

Gordon, N., Salmond, D., and Smith, A. (1993). Novel approach to non-linear/non-Gaussian Bayesian state estimation. *IEE Proceedings-F*, **140**(2), 107–113.

Green, P. (1995). Reversible jump Markov Chain Monte Carlo computation and Bayesian model determination. *Biometrika*, **82**(4), 711–732.

Hastings, W. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**(1), 97–109.

Haykin, S. (1982). *Adaptive Filter Theory.* Prentice Hall, Englewood Cliffs, NJ.

Jeffreys, H. (1961). *Theory of Probability.* Oxford University Press, London, 3rd edition.

Johnson, D. (1982). The application of spectral estimation to bearing estimation problem. *Proceedings of the IEEE*, **70**, 1018–1028.

Kay, S. M. (1993). *Fundamentals of Statistical Signal Processing: Estimation Theory.* Prentice Hall, Englewood Cliffs, NJ.

Larocque, J.-R. and Reilly, J. P. (2000a). Application of the reversible jump MCMC method to real-life propagation measurements. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, Istanbul, Turkey.

Larocque, J.-R. and Reilly, J. P. (2000b). Reversible jump MCMC for joint detection and estimation of directions of arrival in coloured noise. *IEEE Transactions on Signal Processing.* Submitted for publication, also available at http://www.ece.mcmaster.ca/~reilly.

Larocque, J.-R. and Reilly, J. P. (2000c). Reversible jump MCMC for joint detection and estimation of directions of arrival in coloured noise. In *Proceedings of the IMA International Conference on Mathematics in Signal Processing*, University of Warwick, England.

Larocque, J.-R. and Reilly, J. P. (2001a). Wideband channel characterization in coloured noise using the reversible jump MCMC algorithm. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, Salt Lake City, UT.

Larocque, J.-R. and Reilly, J. P. (2001b). Wideband channel characterization in coloured noise using the reversible jump MCMC algorithm. *IEEE Transactions on Vehicular Technology*. Submitted for publication, also available at http://www.ece.mcmaster.ca/~reilly.

Larocque, J.-R., Reilly, J. P., and Ng, W. (2001a). Particle filter for the tracking of an unknown number of sources. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, Salt Lake City, UT.

Larocque, J.-R., Reilly, J. P., and Ng, W. (2001b). Particle filter for the tracking of an unknown number of sources. *IEEE Transactions on Signal Processing*. Submitted for publication, also available at http://www.ece.mcmaster.ca/~reilly.

LeCadre, J. (1989). Parametric methods for spatial signal processing in the presence of unknown coloured noise fields. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **37**, 965–983.

Ligget, W. (1973). *Passive Sonar: Fitting Models To Multiple Time Series*. Academic Press, New York.

Liu, J. and Chen, R. (1993). Sequential Monte Carlo methods for dynamic systems. *Journal of the American Statistical Association*, **93**, 1032–1044.

Logothetis, A. and Carlemalm, C. (2000). SAGE algorithms for multipath detection and parameter estimation in asynchronous CDMA systems. *IEEE Transactions on Signal Processing*, **48**(11), 3162–3174.

MacEachern, S., Clyde, M., and Liu, J. (1999). Sequential importance sampling for non-parametric Bayes models: the next generation. *The Canadian Journal of Statistics*, **27**, 251–267.

Molnar, K. and Modestino, J. (1998). Application of the EM algorithm for the multi-target/multisensor tracking problem. *IEEE Transactions on Signal Processing*, **46**(1), 115–129.

Nagesha, V. and Kay, S. (1996). Maximum likelihood estimation for array processing in coloured noise. *IEEE Transactions on Signal Processing*, **44**(2), 169–180.

Pahlavan, K. and Levesque, A. H. (1995). *Wireless Information Networks*. John Wiley and Sons, New York.

Pisarenko, V. F. (1973). The retrieval of harmonics from a covariance function. *Geophysical Journal Royal of Astronomical Society*, **33**, 347–366.

Pope, K. (1993). *Time series analysis*. Ph.D. thesis, University of Cambridge, Department of Engineering, England.

Reilly, J. and Haykin, S. (1982). Maximum likelihood receiver for low-angle tracking radar Part 2: the nonsymmetric case. *IEE Proceedings*, **129 Pt.F**(5), 331–340.

Reilly, J. P. (1981). *Nonlinear array processing techniques with applications to correlated multipath*. Ph.D. thesis, McMaster University, Hamilton, Ontario, Canada.

Richardson, S. and Green, P. (1997). On Bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society*, **B**(59), 731–792.

Ripley, B. (1987). *Stochastic Simulation*. John Wiley and Sons, New York.

Rissanen, J. (1978). Modeling by shortest data description. *Automatica*, **14**, 465–471.

Robert, C. and Casella, G. (1999). *Monte Carlo Statistical Methods*. Springer-Verlag, New York.

Ruanaidh, J. O. and Fitzgerald, W. (1996). ***Numerical Bayesian Methods Applied to Signal Processing.*** Springer-Verlag, New York.

Rubin, B. (1988). ***Bayesian Statistics 3: Using the SIR Algorithm to Simulate Posterior Distributions (with discussions).*** Oxford University Press, London.

Schmidt, R. (1986). Multiple emitter location and signal parameter estimation. ***IEEE Transactions on Antennas and Propagation*, 34**, 284–280.

Tong, L. and Perreau, S. (1998). Multichannel blind identification: From subspace to maximum likelihood methods. ***Proceedings of the IEEE*, 86**(10), 1951–1968.

Tranter, W. H., Rappaport, T., Woerner, B., and Reed, J., editors (1999). ***Wireless Personal Communications. Emerging Technologies for Enhanced Communications***, chapter 24. Kluwer Academic Publishers, Norwell, MA.

Troughton, P. and Godsill, S. (1997). Bayesian model selection for time series using Markov Chain Monte Carlo. In ***Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing***, Munich, Germany.

Troughton, P. and Godsill, S. (1998). A reversible jump sampler for autoregressive time series. In ***Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing***, Seattle, WA.

Van Trees, H. L. (1968). ***Detection, Estimation and Modulation Theory***, volume 1. John Wiley and Sons, New York.

Vanderveen, M. C., Papadias, C. B., and Paulraj, A. (1997). Joint angle and delay estimation (JADE) for multipath signals arriving at an antenna array. ***IEEE Communications Letters*, 1**(1), 12–14.

Viberg, M. and Ottersen, B. (1991). Sensor array processing based on subspace fitting. ***IEEE Transactions on Signal Processing*, 39**, 1110–1121.

Viberg, M., Ottersten, B., and Kailath, T. (1991). Detection and estimation in sensor arrays using weighted subspace fitting. *IEEE Transactions on Signal Processing*, **39**, 2436–2449.

Wax, M. (1992). Detection and localization of multiple sources in noise with unknown covariance. *IEEE Transactions on Signal Processing*, **40**(1), 245–249.

Wax, M. and Kailath, T. (1985). Detection of signals by information theoretic criteria. *IEEE Transactions on Acoustics, Speech and Signal Processing*, **33**(2), 387–392.

Wax, M. and Leshem, A. (1997). Joint estimation of time delays and directions of arrival of multiple reflections of a known signal. *IEEE Transactions on Signal Processing*, **45**(10), 2477–2484.

Wong, K.-M., Reilly, J. P., Wu, Q., and Qiao, S. (1992). Estimation of the directions of arrival of signals in unknown correlated noise, Part 1: The MAP approach and its implementation. *IEEE Transactions on Signal Processing*, 40(8), 2007–2017.

Wu, Q. and Fuhrmann, D. (1991). A parametric method for determining the number of signals in narrow-band direction finding. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **39**, 1848–1857.

Wu, Q. and Wong, K.-M. (1994). UN-MUSIC and UN-CLE: An application of generalized correlation analysis to the estimation of the direction of arrival of signals in unknown correlated noise. *IEEE Transactions on Signal Processing*, **42**, 2331–2343.

Yang, B. (1995a). An extension of the PASTd algorithm to both rank and subspace tracking. *IEEE Signal Processing Letters*, **2**(9), 179–182.

Yang, B. (1995b). Projection approximation subspace tracking. *IEEE Transactions on Signal Processing*, **43**(1), 95–107.

Ye, H. and DeGroat, R. (1995). Maximum likelihood DOA estimation and asymptotic Cramér-Rao bounds for additive unknown colored noise. *IEEE Transactions on Signal Processing*, **43**(4), 938–949.

Zhao, L., Krishnaiah, P., and Bai, Z. (1987). Remarks on certain criteria for detection of the number of signals. *IEEE Transactions on Acoustics, Speech and Signal Processing*, **35**(2), 129–132.

Zhou, Y., Yip, P., and Leung, H. (1999). Tracking the direction-of-arrival of multiple moving targets by passive arrays: Algorithm. *IEEE Transactions on Signal Processing*, **47**(10), 2655–2666.

Ziskind, I. and Wax, M. (1988). Maximum likelihood localization of multiple sources by alternating projection. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **36**(10), 1553–1560.