Joint Detection and Estimation of Noisy Sinusoids using Bayesian Inference with Reversible Jump MCMC Algorithm

D. ÜSTÜNDAĞ
The Department of Mathematics, the Faculty of Science and Arts,
Marmara University, Istanbul, Turkey.
dustundag@marmara.edu.tr

Abstract: - In this paper, we consider a problem of detecting and estimating of sinusoids corrupted by random noise within a Bayesian framework. Unfortunately, all Bayesian inference draw from posterior probability distributions of parameters requires evaluation of some complicated high-dimensional integrals. Therefore, an attempt for performing the Bayesian computation is made to improve an efficient stochastic algorithm based on reversible jump Markov chain Monte Carlo (RJMCMC) methods. This algorithm, coded in Mathematica programming language is evaluated in simulation studies on synthetic data sets. All the simulations results support the effectiveness of the method.

Keywords: -Bayesian Inference, Model Selection, Parameter Estimation, Reversible Jump MCMC.

1 Introduction

An important task in science is to understand and predict the nature of modeling. Except in artificial problems, a real model is an approximation that, up to some degree, describes the process which generates a particular signal or set of observations. It comes out naturally in the analysis of trying to determine which of the models best describes the data at hand. Estimating values of its free parameters is also very interesting problem to tackle. Therefore, a fundamental problem in signal and data analysis is to develop models which are observed and to determine whether the model function, used to describe the data, is actually appropriate for the particular problem under investigation.

Our goal is to concerned with a problem of simultaneous model selection and parameter estimation within a Bayesian framework. In Bayesian analysis, statistical inference is obtained in the form of posterior distributions, which incorporate both the scientist’s beliefs and the observations, in a probabilistic framework. However, one often faces with problems for establishing prior beliefs, in the form of prior probability distributions on the models under consideration and the computing the quantities that lead to Bayesian model selection-typically integrals of large dimension that do not admit any closed form of analytical solution. In a few cases, for example when the sinusoids are well-separated and many samples are available, suitable analytic approximations to these integrals can possibly be performed [1]. These approximations are difficult to quantify and not valid in the interesting cases where the amount of available data is small, and some sinusoids are close to each other. If we want to perform Bayesian inference in these important cases, it is necessary to numerically approximate these integrals. Some early attempts [1,2,3] to solve this computational problem using classical deterministic multiple integration and Monte Carlo methods are presented in [10,14]. The main problems of these methods are that they are not flexible and are difficult to use when the dimension of the integrand becomes large. Recent advances in the Bayesian literature and, in particular, the advent of the reversible jump Markov chain Monte Carlo (RJMCMC) algorithm, which is initially proposed by Green (1995), have greatly simplified Bayesian model determination problem. These methods are based upon the construction of Markov chains that traverse both parameter and model space so that the best of models can be identified.

In this paper, we make an attempt to develop the RJMCMC algorithm to solve the problem of the joint detection and estimation of sinusoids in white Gaussian noise. The algorithm is coded in Mathematica, which is a state-of-the-art and powerful systems for doing mathematics by computer. The remainder of the paper is organized as follows. In section 2, the signal model is specified. Section 3 presents prior distributions and Section 4 introduces detection and estimation process. The reversible jump MCMC algorithm to simulate from the resulting posterior distribution is developed in Section 5. Section 6 reports the results of computer simulations on synthetic data sets. In Section 7, some conclusions are finally drawn.

2 Problem of Signal Processing
Let \( \mathbf{d} = \{d_0, d_1, ..., d_{N-1}\}^T \) be an observed vector of \( N \) real data samples, which may be represented by different models \( M_k \), corresponding either to samples of noise only or to the superposition of \( k \) sinusoids corrupted by the noise:

\[
M_k: \mathbf{d}(t) = \sum_{j=1}^{\pi} a_{j,k} \cos(w_{j,k} t) + a_{j,k} \sin(w_{j,k} t) + \eta(t) \quad k = 0
\]

where \( w_{j,k} \neq w_{j',k} \) for \( j \neq j' \) and \( a_{j,k}, a_{j,k}, w_{j,k} \) are the amplitudes and the radial frequencies of the \( j \)th sinusoid of the \( k \)th model signal, respectively. The noise vector \( \eta(t) \) is assumed to be drawn from a zero mean white Gaussian of variance \( \sigma_k^2 \). In vector-matrix form, we have

\[
\mathbf{d} = \mathbf{G}(\omega_k) \mathbf{a}_k + \mathbf{n}_k, \tag{2}
\]

where \( \mathbf{a}_k \) are \( 2 \times 1 \) vectors of \( a_{j,k}, a_{j,k}, w_{j,k} \) and \( \omega_k \) is the matrix \( \mathbf{G}(\omega_k) \) defined as

\[
\mathbf{G}(\omega_k) \equiv \begin{bmatrix} \cos(\omega_{1,k} t) & \cos(\omega_{2,k} t) & \cdots & \cos(\omega_{N,k} t) \\ \sin(\omega_{1,k} t) & \sin(\omega_{2,k} t) & \cdots & \sin(\omega_{N,k} t) \end{bmatrix}, \tag{3}
\]

The subscript \( k \) is added here to emphasize that these parameters depend on the model \( M_k \). We assume here that the number \( k \) of sinusoids and their parameters \( \theta_k = \{a_{k,i}, \omega_k, \sigma_k^2\} \) are unknown. Given the data set \( \mathbf{d} \), our goal is to determine the number \( k \) of sinusoids and estimate their associated \( \theta_k \) parameters.

### 3 Prior Distributions

In order to perform Bayesian inference, we first set up the parameter space, which consists of a finite union of subspaces:

\[
\Theta = \bigcup_{k=0}^{K} \{k\} \times \Omega_k, \tag{4}
\]

where \( \theta_0 \not\in \mathbb{R}^* \) and \( \theta_k \not\in \mathbb{R}^2 \times \Omega_k \times \mathbb{R}^* \) for \( k \in \{0,1,...,K\} \) with \( \Omega_k \not\in \{0, \pi\} \) and \( K \in (N-1)/2 \).

On this parameter space, we secondly specify prior probability distributions for the Bayesian analysis. Let us set up a natural hierarchical structure, which we formalize by modeling the joint distribution of all variables as

\[
p(\psi, \theta_k | \mathbf{d}) = p(\mathbf{d} | \theta_k, \psi) p(\theta_k | \psi) p(\psi), \tag{5}
\]

where \( p(k | \psi) \) is a prior model probability condition on hyper parameter vector \( \psi \); \( p(\theta_k | \psi) \) is a prior probability for the parameters, which is conditional on \( (\psi, \theta_k) \) and \( p(\mathbf{d} | \theta_k, \psi) \) is the likelihood, which does not depend on \( \psi \). For a given model in (2), it is defined by the following form:

\[
p(\mathbf{d} | \theta_k, \psi) = -\frac{1}{2\sigma^2} \exp \left( -\frac{\mathbf{G}(\omega_k) \mathbf{a}}{\sigma^2} \right)^2 \tag{6}
\]

Given hyper parameters \( \psi \) the joint conditional distribution of \( (k, \theta_k) \) is defined by the following structure:

\[
p(k, \theta_k | \psi) = p(k | \psi) p(\theta_k | \psi), \tag{7}
\]

where \( \sigma_k^2 \) is a scale parameter that is assumed to be distributed according to a conjugate inverse-Gamma prior distribution, for instant \( \sigma_k^2 \sim \Gamma(v_0/2, \gamma_0/2) \). When \( v_0 = 0 \) and \( \gamma_0 = 0 \), Jeffrey’s**` uninformative prior is obtained[11]. For all parameters \( (k, \omega_k, \mathbf{a}_k) \) we introduce the following prior distributions:

\[
p(k, \omega_k, \mathbf{a}_k | \psi) \propto \Lambda_k^{\frac{1}{k}} \exp(-\Lambda) \left( \frac{2 \sigma^2 \Sigma_k}{\mathbf{a}_{k}^T \Sigma_k^{-1} \mathbf{a}} \right)^{\frac{1}{2}}, \tag{8}
\]

where

\[
\Sigma_k^{-1} = \delta^2 \mathbf{G}^T(\omega_k) \mathbf{G}(\omega_k), \tag{9}
\]

and

\[
I_{\Omega_k}(k, \omega_k) = \begin{cases} 1, & \text{if } (k, \omega_k) \in \Omega_k \\ 0, & \text{otherwise} \end{cases} \tag{10}
\]

The prior probability model distribution \( p(k | \psi) \) is given by a truncated Poisson distribution. Conditional on \( k \), the frequencies are assumed uniformly distributed in \( \Omega_k \). Finally, conditional on \( (k, \omega_k) \) the amplitudes are drawn from a zero mean Gaussian distribution with covariance \( \sigma_k^2 \Sigma_k \). The hyper parameter vector \( \psi \) consists of \( \delta^2 \) and \( \Lambda \) which can be, respectively, interpreted as an expected signal-to-noise ratio and the expected number of sinusoids.

### 4 Detection and Estimation

In a fully Bayesian approach, the unknown \( k \) and the set of unknown parameters \( \theta_k \) are regarded as being drawn
from an appropriate prior distribution \( p(k, \theta_k | \psi) \). This prior distribution reflects our degree of belief on the relevant values of the parameters. The Bayesian inference of \( k \) and \( \theta_k \) is therefore based on the joint posterior distribution \( p(k, \theta_k | d) \) obtained from Bayes’ theorem. Under weak additional assumptions, the \( M \times 1 \) samples generated by the Markov chain are asymptotically distributed according to the posterior distribution and thus allow easy evaluation of all posterior features of interest. In particular, it allows us to evaluate the posterior model probability \( p(k | d) \), defined by

\[
p(k = j | d) = \frac{1}{M} \sum_{i=1}^{M} I_{ij}(k_i, \omega_{k_i}) .
\]  

(11)

This can be used to perform model selection by choosing the model order as \( \hat{k} = \arg \max_{k=0,...,K} p(k | d) \).

In addition, it allows us to perform the estimation of the parameters by computing the conditional expectation:

\[
E[\theta_i | d, k = j] = \left( \sum_{i} \theta_i I_{ij}(k_i, \omega_{k_i}) \right) / \left( \sum_{i} I_{ij}(k_i) \right) .
\]

(12)

However, it is usually impossible to obtain these quantities analytically. Indeed, it requires the evaluation of high-dimensional integrals of nonlinear functions in the parameters. For our problem some integrations can be taken analytically and do not require any Monte Carlo integration scheme. Consequently, according to Bayes’ theorem, we have

\[
p(k = 1 | a, \sigma^2 | d) \propto p(a | \sigma^2) p(k = 1 | \sigma^2) p(d | a, \sigma^2, k = 1) \propto \exp \left( \frac{1}{2\sigma^2} (a - \bar{a})^T M_k (a - \bar{a}) \right) \propto \exp \left( \frac{1}{2\sigma^2} (\gamma_0 + d^T P_k d)^{-\frac{N_k}{2}} \right) \]

(13)

\[
\propto \exp \left( -\frac{1}{2\sigma^2} M_k^T \Sigma_k^{-1} M_k \right)
\]

where

\[
M_k^{-1} = G^T (\omega_k) G(\omega_k) + \Sigma_k^{-1}
\]

\[
\hat{a}_k = M_k^{-1} G^T (\omega_k) d
\]

\[
P_k = I_{k} - G^T (\omega_k) M_k G(\omega_k)
\]

Then we can integrate Equation (13) with respect to the amplitudes and the variance of the noise, conditional upon the hyper parameters to obtain the following conditional posterior distribution for the number of sinusoids and their frequencies:

\[
p(k, \omega_k | d) \propto (\gamma_0 + d^T P_k d)^{-\frac{N_k}{2}} \frac{\Lambda^k}{k!(1+\delta^2)^2 \pi^k} I_{k}(k, \omega_k).
\]

(15)

## 5 Reversible Jump MCMC Algorithm

The reversible jump MCMC sampler [3] is an extension of the standard MCMC method in that it allows for jumps between models and their parameter spaces of different dimensions as the sampling proceeds. It is based on the Metropolis- Hastings (MH) algorithm described in [4] and it involves moves that represent changes of models or updating of the current model parameters. The key point is to build an ergodic Markov chain \( \{k, \theta_{k_i}\} \), whose equilibrium distribution is the desired posterior distribution. Briefly, this procedure samples directly from different model orders from the joint posterior distribution \( p(k, \omega_k | d) \). In effect, the process jumps between subspaces of different dimensions, thus visiting all model orders for \( k_i \in \{0, 1, ..., K\} \). The resulting transition kernel of the simulated Markov chain is then a mixture of the different transition kernels associated with the move (birth, death or update), which is randomly chosen at each iteration \( i \). The probabilities of the birth (proposing a new sinusoid) and death (removing a sinusoid) moves are given by

\[
b_k = \min \left\{ \frac{p(k+1)}{p(k)} \right\},
\]

\[
d_k = \min \left\{ \frac{p(k)}{p(k+1)} \right\}.
\]

(16)

and the probability of update move is defined in the form: \( u_k = 1 - (b_k + d_k) \) for all \( 0 < k < K \). Here \( p(k) \) is the prior probability of the model \( M_k \) and \( c = 0.5 \) is a parameter which tunes the proportion of dimension or update move. There is no death move when \( k = 0 \) and no birth move when \( k = K \).

**Birth Move:** Let us suppose that the current state of the Markov chain is in \( \{k \} \times \Omega_k \). We now propose a new frequency at random on \( (0, \pi) \). This can be done by generating \( \omega \) randomly from an uniform distribution \( U_{(0, \pi)} \) and setting up \( \omega_{k+1} = \{\omega_k, \omega\} \). We then evaluate the acceptance ratio:

\[
\alpha_{k,k+1} = \min \left\{ 1, \frac{r_{\text{post}}}{r_{\text{prop}}} \right\}
\]

(17)

where

\[
r_{\text{post}} = (\text{posterior distribution ratio}) \times (\text{proposal ratio})
\]

\[
= \frac{\gamma_0 + d^T P_k d}{\gamma_0 + d^T P_{k+1} d} \frac{(k+1)^{N_{k+1}}}{k^N}
\]

(18)

We choose a random number \( u \) from an uniform distribution \( U_{(0,1)} \) on the interval \( (0,1) \). If \( u \leq \alpha_{k,k+1} \) then the state of the Markov chain becomes \( \{k+1, \omega_{k+1}\} \), otherwise we stay at \( \{k, \omega_k\} \).

**Death Move:** Let us suppose that the current state of the Markov chain is in \( \{k+1\} \times \Omega_{k+1} \). We choose a
sinusoid with label at random among the \((k+1)\) existing sinusoids and remove it. Then we evaluate \( \alpha_{k+1,k} = \left\{ 1, r_{k+1,k} \right\} \) and choose a random number \( u \) from a uniform distribution \( U(0,1) \). If \( u \leq \alpha_{k+1,k} \), then the state of the Markov chain becomes \((k, \omega_k)\), otherwise we stay at \((k+1, \omega_{k+1})\).

**Update Move:** It does not involve changing the dimension of the model and require an iteration of the hybrid MCMC sampler. In our application, the target distribution is the full conditional distribution of a frequency:

\[
p(\omega'_{j,k} | \omega_{j,k}) \propto \left( \gamma_0 + d^T p_j d \right)^{\frac{(N+N_0)}{2}} I_\Omega(k, \omega_k). \tag{19}
\]

With a probability \(0 < \lambda < 1\), we perform a MH step with proposal distribution \( q_1(\omega'_{j,k} | \omega_{j,k}) \), independent from the current state \( \omega_{j,k} \):

\[
q_1(\omega'_{j,k} | \omega_{j,k}) \propto \sum_{i=0}^{N-1} p_i I_{(\pi/N_\omega,i+1)}(k, \omega_k) \tag{20}
\]

where \( p_i \) is the value of the squared modulus of the Fourier Transform (FT) of the observations \( d \) at frequency \((i\pi/N_\omega)\). We will take \( N_p = N \) but, \( N_p > N \) can be also used to improve the interpolation of the FT via zero padding. Choosing a new frequency \( \omega'_{j,k} \) independent from \( \omega_{j,k} \) in the regions where the modulus of the FT has high values helps us to build the regions of interest of the posterior distribution quickly. With probability \((1-\lambda)\), we perform a MH step with a proposal distribution:

\[
q_2(\omega'_{j,k} | \omega_{j,k}) \propto N(\omega'_{j,k}, \sigma_{RW}^2). \tag{21}
\]

This random walk leads to perform a local exploration of the posterior distribution and to ensure irreducibility of the Markov chain. In both cases, the acceptance probability is given by

\[
\min \left\{ \frac{\gamma_0 + d^T P_j d}{\gamma_0 + d^T P_{j'} d} \right\}^{\frac{N_0}{2}} q(\omega_{j,k} | \omega'_{j,k}) q(\omega'_{j,k} | \omega_{j,k}) I_\Omega(k, \omega_k). \tag{22}
\]

where \( P_{j'}, M_{j'} \) and \( \Sigma_{j'} \) are similar to \( P_j, M_j \) and \( \Sigma_j \) respectively by replacing \( \omega_{j,k} \) with \( \omega'_{j,k} \).

It is easy to prove that the algorithm converges geometrically to the required posterior distribution depending on the starting point. For detail information, we refer to papers \([5,6,13]\).

6. **Computer simulations**

The algorithm described previously requires the specification of parameters, \( \lambda \) and \( \sigma_{RW}^2 \), that have no influence on the posterior distribution. Therefore, we used here \( \lambda = 0.2 \) and \( \sigma_{RW} = 1/(5N) \), but their influence on the speed of convergence of the algorithm will be given in a different paper. On the other hand, the Bayesian model is specified by \( \nu, \gamma_0, \Lambda \) and \( \delta^2 \). For the values of \( \Lambda \) and \( \delta^2 \), we used the procedure introduced in papers \([5-9]\). The algorithm was coded in Mathematica and the simulations were performed on a workstation, which contains four processors.

To demonstrate the approach with an example, we have created \( N = 512 \) data samples according to signal model with two close harmonic frequencies:

\[
d_i = 0.5403 \cos(0.3t_i) - 0.8415 \sin(0.3t_i) - 0.4161 \cos(0.31t_i) + 0.9093 \sin(0.31t_i) + n(t_i), \tag{23}
\]

where \( n(t_i) \) is generated at \( t_i \in \{0, 1, ..., N-1\} \) from zero mean Gaussian distribution with \( \sigma^2 = 1 \). After obtaining simulated data, we carried out Bayesian analysis, assuming neither we know the number of sinusoids nor values of their parameters. We give starting values to the parameters are clearly estimated within the calculated accuracy. It is also noted that the separation of the sinusoids is less than the Rayleigh resolution\([12]\).

The maximum value of \( p(k | d) \) indicates that the number of sinusoids in data is more likely equal to \( k = 2 \). The results, illustrated in Table 1, show that all values of the parameters are clearly estimated within the calculated accuracy. There is also noted that the separation of the sinusoids is less than the Rayleigh resolution\([12]\).

In general, we consider a multiple harmonic frequency model signal:

\[
d_i = \sum_{j=1}^{k} A_j \cos(\omega_j t_i + \varphi_j) + n(t_i), \quad (i = 1, ..., N) \tag{24}
\]

where \( A_j = \sqrt{a_{c,j}^2 + a_{s,j}^2} \) and \( \varphi_j = - \arctan \left( a_{s,j} / a_{c,j} \right) \) are the amplitudes and phases, respectively. In a similar way, we produced \( N = 512 \) sample simulated data and ran the Mathematica code for this experiment again. After 50 thousand iterations, Figure 2 shows the histogram of the posterior probability \( p(k | d) \) and indicates that there is more likely five sinusoids \( (k = 5) \). The best estimates of parameters are tabulated in Table 2. Once again, all the frequencies have been well resolved. These results are also similar to that of
Bretthorst [1] and demonstrate the advantages of Bayesian RJMCMC.

It is interesting to point out that the number of hypothesized sinusoids does not have to be limited. For as long as we provide a scheme which allows reachability of any model from any point in our state space, the RJMCMC sampler will move through it and spend most of the time with the model or set of models which are the most likely. Figure 3 shows the maximizing process of \( \log(p(k,o|d)) \) as the number of iteration increases. It is clearly seen that the marginal probability density reaches an equilibrium after a burn-in period[12].

7 Conclusions
We have given what we consider to be a very simple and direct way of handling the difficult problem of detecting the number of sinusoids and estimating their parameters simultaneously from noisy data, using a Bayesian RJMCMC approach. It is also provided a brief summary of the theory and examples of how one might apply it. Over all the results presented here indicate that RJMCMC sampling, although computationally intensive, is a powerful methodology for signal processing. On the other hand, RJMCMC analyses are computationally well suited for Mathematica type environments but there is a need for them to be embedded in more end-user friendly computing environments.

Acknowledgements
This work is a part of the projects, whose names are "Bayesian Model Selection and Parameter Estimation" with a number FEN - DKR - 200407 - 0082 and "Bayesian Spectrum Estimation of Harmonic Signals" with a number FEN-BGS-060907-0191, supported by the University of Marmara, Istanbul, Turkey.

References
Figure 1. Histogram of the posterior probability \( p(k|d) \) for two closed frequency signal model is obtained after 50 000 iterations.

Table 1. Computer simulations for two closed frequencies signal model.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>True</th>
<th>Estimated</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \omega_1 )</td>
<td>0.01</td>
<td>0.0998</td>
</tr>
<tr>
<td>( \omega_2 )</td>
<td>0.15</td>
<td>0.1498</td>
</tr>
<tr>
<td>( \omega_3 )</td>
<td>0.30</td>
<td>0.3002</td>
</tr>
<tr>
<td>( \omega_4 )</td>
<td>0.31</td>
<td>0.3088</td>
</tr>
<tr>
<td>( \omega_5 )</td>
<td>1.00</td>
<td>0.9999</td>
</tr>
</tbody>
</table>

Table 2. Computer simulations for a multiple harmonic signal model

<table>
<thead>
<tr>
<th>Parameters</th>
<th>True</th>
<th>Estimated</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \varphi_1 )</td>
<td>1.00</td>
<td>1.0600</td>
</tr>
<tr>
<td>( \varphi_2 )</td>
<td>2.00</td>
<td>2.0456</td>
</tr>
<tr>
<td>( \varphi_3 )</td>
<td>3.00</td>
<td>2.9487</td>
</tr>
<tr>
<td>( \varphi_4 )</td>
<td>5.00</td>
<td>5.0418</td>
</tr>
<tr>
<td>( \varphi_5 )</td>
<td>5.00</td>
<td>5.2565</td>
</tr>
<tr>
<td>( A_1 )</td>
<td>5.00</td>
<td>5.2276</td>
</tr>
<tr>
<td>( A_2 )</td>
<td>5.00</td>
<td>5.2276</td>
</tr>
<tr>
<td>( A_3 )</td>
<td>5.00</td>
<td>5.2565</td>
</tr>
<tr>
<td>( A_4 )</td>
<td>3.00</td>
<td>2.9631</td>
</tr>
</tbody>
</table>

Figure 2. Histogram of the posterior probability \( p(k|d) \) for five frequencies signal model is obtained after 50 000 iterations.

Table 2. Computer simulations for a multiple harmonic signal model

\[
\log(p(\omega|d))
\]