BAYESIAN MODEL SELECTION AND ESTIMATION OF MULTIPLE CISOID MODELS

Roland Jonsson
Ericsson Microwave Systems AB,
S-431 84 Mölndal, Sweden,
Tel: +46 31 7472672,
`Roland.Jonsson@emw.ericsson.se`

ABSTRACT

In this paper we address the problem of Bayesian model selection and estimation, of signals that consists of a sum of complex sinusoids ("cisoids"). This kind of signal models are abundant in a wide range of engineering applications, but has usually been treated in a non Bayesian way. The recent development of Markov Chain Monte Carlo methods (MCMC) have opened up the possibility to use Bayesian methods to analyze this kind of signals. Here we present a new combined model selection and estimation method for the case of signals with additive white Gaussian noise and known variance. We demonstrate its use on cisoids closely spaced in frequency, using the Jeffrey prior.

1 INTRODUCTION

The recent development of the reversible jump Markov Chain Monte Carlo method [3] has opened a possibility to perform Bayesian model selection and estimation for complex models in a number of fields. In the signal processing community, this methodology has been applied to the problem of sinusoidal estimation and model selection in the work of Andrieu and Doucet [1]. In this paper, the problem is approached under somewhat different assumptions. The main difference is that in [1] the noise variance was assumed unknown, and that a rather particular choice of prior distributions made it possible to analytically marginalize the amplitude and the noise variance parameters. That resulted in a reduction of the dimension of the problem, that is beneficial to the MCMC simulation, but the computation of acceptance probabilities required the computation of a projection matrix on the signal subspace.

In many engineering applications (e.g. radar) the noise variance is known, or can be estimated from auxiliary data. In this paper we show that, in the case of known noise variance, it is possible to choose a proposal distribution for the amplitudes in such a way that the complexity of the computation of acceptance probabilities is greatly reduced. The sampling of the amplitudes also enable us to freely choose prior distributions, and we specifically use the Jeffreys prior. We demonstrate that, with this prior, a combined model selection and estimation procedure can be designed which is "well behaved" for closely spaced frequencies at a low signal to noise ratio.

2 SIGNAL MODEL

The type of signals that we are considering, is a sum of complex sinusoids (cisoids) in white Gaussian noise of known variance σ^2. The signal is then an N-dimensional complex vector $\mathbf{x}$, that is complex Gaussian distributed $\mathcal{CN}(\mathbf{s}, \sigma^2\mathbf{I}_N)$, where $\mathbf{s} = \sum_{l=1}^{k} a_l \mathbf{m}(\omega_l)$ with $\mathbf{m}(\omega) = \left[1 \; e^{j\omega} \ldots e^{j(N-1)\omega}\right]^T$. The quantities that we are estimating are $\boldsymbol{\theta} = (k, \boldsymbol{\theta}_k)$ where $\boldsymbol{\theta}_k = [a_1 \ldots a_k, \omega_1 \ldots \omega_k]^T$, and the likelihood is:

$$f(\mathbf{x}|\boldsymbol{\theta}_k, k) = (\pi\sigma^2)^{-N} \exp\left[-\frac{1}{\sigma^2}\left\|\mathbf{x} - \sum_{l=1}^{k} a_l \mathbf{m}(\omega_l)\right\|^2\right]$$

As we are using a Bayesian methodology we seek the posterior distribution:

$$f(k, \boldsymbol{\theta}_k|\mathbf{x}) = \frac{f(\mathbf{x}|\boldsymbol{\theta}_k, k)\, f(\boldsymbol{\theta}_k, k)}{f(\mathbf{x})}$$

and estimators such as the Maximum Aposteriori (MAP):

$$\widehat{k} = \operatorname{Arg\,max}_k \; f(k|\mathbf{x}) \quad f(k|\mathbf{x}) = \int f(k, \boldsymbol{\theta}_k|\mathbf{x}) d\boldsymbol{\theta}_k$$

and the conditional mean:

$$\widehat{\boldsymbol{\theta}}_k = \int \boldsymbol{\theta}_k f(\boldsymbol{\theta}_k|\mathbf{x},\, k) d\boldsymbol{\theta}_k$$

The Reversible Jump MCMC sampler derived in section 4 draws samples from $f(k, \boldsymbol{\theta}_k|\mathbf{x})$ and are then used to approximate the integrals above. The choice of the prior distribution $f(\boldsymbol{\theta}_k, k)$ will be discussed in the next section. We will make the assumption that the frequencies are ordered: $0 \leq \omega_1 < \omega_2 < \ldots < \omega_k < 2\pi$. The parameters $\boldsymbol{\theta}_k$ of order k are points in $\Theta_k = \mathbb{C}^{K} \times [0..2\pi)^K$, and the overall parameter space is $\Theta = \bigcup_{k=0}^{k_{max}} \{k\} \times \Theta_k$.

3 PRIOR DISTRIBUTIONS

The choice of prior distributions is central to the Bayesian paradigm. The design of a Bayesian algorithm is determined by this choice, and by the choice of the decision procedure. We will in this section describe the choice of prior distributions used in the simulations. Note that the MCMC sampler described in the next section is not dependent on this choice.

Because the problem at hand does not contain any particular prior information, we seek a "noninformative" prior distribution. The Jeffreys prior, i.e. $|\mathbf{I}(\boldsymbol{\theta})|^{1/2}$ where $\mathbf{I}(\boldsymbol{\theta})$ is the Fisher information matrix, can for estimation problems be motivated as "noninformative" by information geometric arguments [4]. For the current signal model, it has the intuitively appealing behavior of going to zero when the frequency difference between two cisoids tends to zero, or when the magnitude of an amplitude tends to zero. Thus it formalizes the idea of distinguishable signals. The Jeffreys prior, being an improper prior, can appear to be problematic for the model selection problem; but following the reasoning in [2] one can justify a choice: $f(\boldsymbol{\theta}_k, k) = \kappa^k \, |\mathbf{I}(\boldsymbol{\theta}_k)|^{1/2}$ where κ is a dimensionless constant. We use an approach similar to classical detection theory, where a threshold is determined from a predetermined false alarm rate. In our case, we consider a MAP choice between a model with one cisoid or no cisoids, and a signal that contain noise only. In this case, it is possible to analytically evaluate an approximation to the false alarm rate as a function of κ. We use this to find the value of κ for a suitable false alarm rate (one finds that κ is proportional to $1/N$ and has a quite complex dependence on the false alarm rate).

We can unfortunately not, due to space limitations of this paper present the expressions for the prior and the false alarm rate analysis.

4 REVERSIBLE JUMP MCMC SAMPLING

The reversible jump Markov Chain Monte Carlo method introduced by Green in [3] makes it possible to sample distributions defined over complex spaces such as Θ. This method is similar to the well known Metropolis-Hastings method, in that the algorithm propose a move to a candidate state, then evaluate an acceptance probability, and accept the move with that probability or otherwise remain at the current state. In this case, the moves can be moves between subspaces of different dimension, and there can be a number of types of moves.

In our sampler, we use for each k: k updating moves that only change the parameters for one cisoid, and k birth and k death moves which add or remove one cisoid. We denote the move type by a triple $m \in \bigcup_{k=0}^{k_{max}} \{k\} \times \{1..k\} \times \{0, +, -\}$, and denote by a tilde all proposed quantities. The acceptance probability for a move of type m taking $\boldsymbol{\theta}$ into $\widetilde{\boldsymbol{\theta}}$ are denoted by $\alpha_m(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}})$. According to [3] one should, in order to sample from a distribution p, chose

$$\alpha_m(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}}) = \min \left\{ 1, \frac{p(\widetilde{\boldsymbol{\theta}}) \, q_m(\widetilde{\boldsymbol{\theta}}, \boldsymbol{\theta})}{p(\boldsymbol{\theta}) \, q_m(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}})} J \right\}$$

where $q_m(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}})$ is the probability to propose $\widetilde{\boldsymbol{\theta}}$ by move m when currently in $\boldsymbol{\theta}$. The J in the formula above is the determinant of the Jacobian of a bijective function which takes $(\boldsymbol{\theta}, u_1)$ into $(\widetilde{\boldsymbol{\theta}}, u_2)$, where u_1 and u_2 are the random variables that goes into the construction of the proposals. In our case, where we update the parameters for one cisoid, draw new parameters for a new cisoid, or remove the parameters for one cisoid, we always have $J = 1$ [3]. The acceptance probability for the sampling of our posterior distribution can now be written as :

$$\alpha_m(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}}) = \min \left\{ 1, \frac{f(\widetilde{\boldsymbol{\theta}}_{\widetilde{k}}, \widetilde{k}) \, f(\mathbf{x}|\widetilde{\boldsymbol{\theta}}_{\widetilde{k}}, \widetilde{k}) \, q_m(\widetilde{\boldsymbol{\theta}}, \boldsymbol{\theta})}{f(\boldsymbol{\theta}_k, k) \, f(\mathbf{x}|\boldsymbol{\theta}_k, k) \, q_m(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}})} \right\}$$

As always in MCMC methods, a high acceptance probability is desirable, and a good choice of proposal distributions becomes important. It is also, from a practical point of view, important to have an acceptance probability that is not to demanding to compute. We will see that the choice to sample the parameters for one cisoid at the time, together with a particular choice of proposal distributions will result in a simple expression for the acceptance probability.

Introducing $\boldsymbol{\theta}_{k,-l}$ for $\boldsymbol{\theta}_k$ with ω_l and a_l removed, and $\mathbf{y}(\boldsymbol{\theta}_{k,-l}) = \mathbf{x} - \sum_{m=1, m\neq l}^{k} a_m \mathbf{m}(\omega_m)$, we can write the likelihood of $\mathbf{x}$ as:

$$\begin{aligned} f(\mathbf{x}|k, \boldsymbol{\theta}_k) &= (\pi\sigma^2)^{-N} \exp\left[-\tfrac{1}{\sigma^2} \|\mathbf{y}(\boldsymbol{\theta}_{k,-l})\|^2 \right] \\ &\times \exp\left[-\tfrac{N}{\sigma^2} \left| a_l - \tfrac{1}{N} \mathbf{y}(\boldsymbol{\theta}_{k,-l})^H \mathbf{m}(\omega_l) \right|^2 \right] \\ &\times \exp\left[-\tfrac{1}{N\sigma^2} \left| \mathbf{y}(\boldsymbol{\theta}_{k,-l})^H \mathbf{m}(\omega_l) \right|^2 \right] \end{aligned}$$

Alternatively we write $\mathbf{y}(\boldsymbol{\theta}_k) = \mathbf{x} - \sum_{m=1}^{k} a_m \mathbf{m}(\omega_m)$ and

$$f(\mathbf{x}|k, \boldsymbol{\theta}_k) = (\pi\sigma^2)^{-N} \exp\left[-\tfrac{1}{\sigma^2} \|\mathbf{y}(\boldsymbol{\theta}_k)\|^2 \right]$$

We now chose the proposal distribution in the following way. For an updating move that changes cisoid l, we first draw a new frequency $\widetilde{\omega}_l$ according to the distribution $\mathcal{N}(\omega_l, \sigma_\omega^2)$, i.e. we use a random walk chain for the frequencies. We then draw the amplitude $\widetilde{a}_l$ from $\mathcal{CN}(\frac{1}{N}\mathbf{y}(\boldsymbol{\theta}_{k,-l})^H \mathbf{m}(\widetilde{\omega}_l), \frac{\sigma^2}{N})$, i.e. we use a "Gibbs proposal" for the amplitudes. The acceptance probability can now be computed as:

$$\alpha_{\{k,l,0\}}(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}}) = \min \left\{ 1, \frac{f(\widetilde{\boldsymbol{\theta}}_k, k) \exp\left[\frac{1}{N\sigma^2} \left| \mathbf{y}(\boldsymbol{\theta}_{k,-l})^H \mathbf{m}(\widetilde{\omega}_l) \right|^2 \right]}{f(\boldsymbol{\theta}_k, k) \exp\left[\frac{1}{N\sigma^2} \left| \mathbf{y}(\boldsymbol{\theta}_{k,-l})^H \mathbf{m}(\omega_l) \right|^2 \right]} \right\}$$

Choosing a σ_ω less then $\frac{2\pi}{N}$ and an initial ω_l near the maximum of $\left| \mathbf{y}(\boldsymbol{\theta}_{k,-l})^H \mathbf{m}(\omega) \right|$ will ensure a low rejection rate and a effective sampling of the high probability region of the posterior distribution. In the simulations, σ_ω was set to $0.3\frac{2\pi}{N}$. The only computationally demanding quantity that has to be evaluated, besides the prior probability, is $\mathbf{y}(\boldsymbol{\theta}_{k,-l})^H\mathbf{m}(\omega_l)$. To further simplify the computations, we restrict the frequencies to a number of discrete sample points, which enable us to make the computations in the frequency domain.

The birth moves is constructed in such a way that proposals for a new frequency, $\widetilde{\omega}_l$, has high probability close to the maximum of $\left|\mathbf{y}(\boldsymbol{\theta}_k)^H\mathbf{m}(\omega)\right|$ in the interval $\omega_l < \omega < \omega_{l+1}$ (where ω_{k+1} should be understood as ω_1). In the following we denote this maximum with $\widehat{\omega}_{k,l}$. Let $q^{\omega}_{\{k,l,+\}}(\boldsymbol{\theta},\widetilde{\omega}_l)$ be the proposal distribution for $\widetilde{\omega}_l$, and then draw the amplitude $\widetilde{a}_l$ from $\mathcal{CN}(\frac{1}{N}\mathbf{y}(\boldsymbol{\theta}_k)^H\mathbf{m}(\widetilde{\omega}_l), \frac{\sigma^2}{N})$. The acceptance probability can now be computed as:

$$\alpha_{\{k,l,+\}}(\boldsymbol{\theta},\widetilde{\boldsymbol{\theta}}) = \min\left\{1, \frac{f(\widetilde{\boldsymbol{\theta}}_{k+1},k+1)\,\exp\left[\frac{1}{N\sigma^2}\left|\mathbf{y}(\boldsymbol{\theta}_k)^H\mathbf{m}(\widetilde{\omega}_l)\right|^2\right]}{f(\boldsymbol{\theta}_k,k)\,\frac{N}{\pi\sigma^2}\,q^{\omega}_{\{k,l,+\}}(\boldsymbol{\theta},\widetilde{\omega}_l)}\frac{1}{Q_k}\right\}$$

where Q_k is the ratio between the probability to propose a birth move at order k and the probability to propose a death move at order $k+1$. The corresponding death move just removes cisoid l, with acceptance probability:

$$\alpha_{\{k,l,-\}}(\boldsymbol{\theta},\widetilde{\boldsymbol{\theta}}) = \min\left\{1, \frac{f(\boldsymbol{\theta}_{k,-l},k-1)\,\frac{N}{\pi\sigma^2}\,q^{\omega}_{\{k-1,l,+\}}(\boldsymbol{\theta}_{k,-l},\omega_l)}{f(\boldsymbol{\theta}_k,k)\,\exp\left[\frac{1}{N\sigma^2}\left|\mathbf{y}(\boldsymbol{\theta}_{k,-l})^H\mathbf{m}(\omega_l)\right|^2\right]}Q_{k-1}\right\}$$

It remains to chose $q^{\omega}_{\{k,l,+\}}(\boldsymbol{\theta},\widetilde{\omega}_l)$. Here we use a mixture between a uniform distribution and a Gaussian centered at $\widehat{\omega}_{k,l}$. The addition of the uniform distribution is needed to hold up the acceptance ratio for the death moves in the case when k are greater then the actual number of cisoids. In this case spurious samples occurs (with low probability) for any ω, and the acceptance ratio for death moves should not be weighted down to much when they are far away from $\widehat{\omega}_{k,l}$. We chose the variance of the Gaussian as $\frac{\sigma^2}{|\widehat{a}_{k,l}|^2}\frac{6}{N^3}$ where $\widehat{a}_{k,l} = \frac{1}{N}\mathbf{y}(\boldsymbol{\theta}_k)^H\mathbf{m}(\widehat{\omega_{k,l}})$. With this choice, we get an acceptance probability that is constant for $\widetilde{\omega}_l$ close to $\widehat{\omega}_{k,l}$ for $\mathbf{y}(\boldsymbol{\theta}_k)$ equal to a pure cisoid. We can now write:

$$q^{\omega}_{\{k,l,+\}}(\boldsymbol{\theta},\widetilde{\omega}_l) = \frac{P_u}{2\pi} + (1-P_u)\frac{1}{\sqrt{2\pi}}\frac{|\widehat{a}_{k,l}|N^{\frac{3}{2}}}{\sigma\sqrt{6}}\exp\left[-\frac{N^3|\widehat{a}_{k,l}|^2}{12\sigma^2}(\widetilde{\omega}_l-\widetilde{\omega}_l)^2\right]$$

where P_u is the probability for the uniform part of the mixture.

We implemented the sampler in such a way that a new sample was recorded after each "sweep". Where a sweep consists of first doing a deterministic cycle through all the updating moves; and then at random propose, either a birth move, a death move, or nothing at all. In the simulations the probability for proposing either a birth move or a death move was set to 0.2.

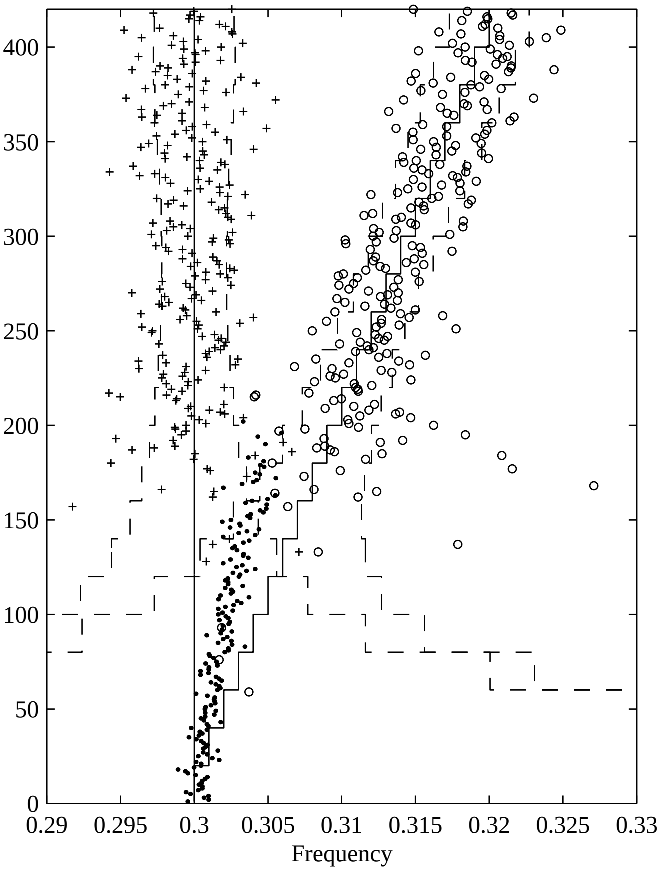

Figure 1: Estimates of the frequencies of two cisoids with varying frequency separation. ($a_1 = a_2 = 1$)

5 SIMULATION RESULTS

In all the simulations we have used the noise level $\sigma = 1$ and $N = 50$ samples. We have generated signals that contain two cisoids of equal amplitude $a_1 = a_2 = 1$, and various frequency separations less then a Fourier resolution. We are thus in a region of low signal to noise ratios, where traditional methods of order selection, such as MDL, over-estimates the model order [1]. We used 5000 sweeps of the MCMC sampler for each estimation, which is considerably less then many reported uses of MCMC samplers. However, we believe that our sampler achieves adequate mixing in 5000 sweeps. Doing an experiment running 50000 sweeps, we found that the estimates did not change much from the estimates computed using 5000 sweeps. The change was small compared to the estimation errors.

We used a MAP estimator of the number of cisoids and a conditional mean estimator for the frequencies and amplitudes. Changing the frequency separation in steps, running 20 different noise realizations at each, a total of 420 estimation experiments where done. The results are illustrated in figure 1. We have plotted the frequency estimates in cases when one cisoid was determined as dots, and the case when two cisoids was determined as

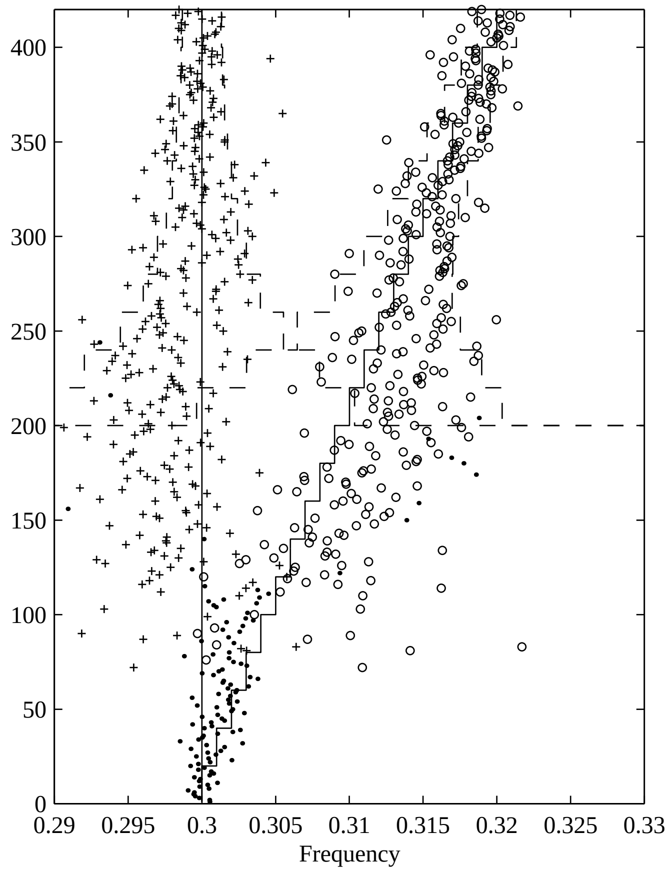

Figure 2: Estimates of the frequencies of two cisoids with varying frequency separation. ($a_1 = 1$, $a_2 = j$)

'+' and 'o'. The different estimation experiments are plotted along the vertical axis. In the background, the true frequencies are plotted as solid lines, and the standard deviation computed from the Cramer Rao bound are plotted with dashed lines. There were only six cases of estimating $k = 3$, and there is seven estimates at $k = 2$ that fell outside the limits of the plot. We see in the figure, that the estimator are in agreement with the Cramer Rao bound, until the separation is about half a Fourier resolution, it then starts to estimate one cisoid halfway between the true locations. The results for the amplitudes are not shown here, but they are also in agreement with Cramer Rao bound for the $k = 2$ case, and an estimate around 2 for the $k = 1$ case.

We repeated the simulations above with the only change that one amplitude was chosen to be (one unit) imaginary. This is a much harder case then the previous one, as is indicated by the Cramer Rao bond, and we get 90 cases of estimating $k = 3$ and 44 of estimating $k = 4$. This problem of estimating a too high model order can be mitigated by using a Bayesian decision procedure for model selection with a more complex cost function. We have chosen a procedure that does not count posterior samples that have frequencies falling outside an interval of ± 1.5 Fourier resolutions around the mean of the samples. This will decrease chance of choosing an over parameterized model, which typically have a broad posterior distribution of the frequencies. This reduced the estimated cases of $k = 3$ to 11 and of $k = 4$ to 1. The results of frequency estimates are presented in figure 2 in the same manner as before, and we can note an orderly behavior in the problematic region.

6 CONCLUSIONS

We have, in this paper, described a reversible jump Markov Chain Monte Carlo method for the problem of combined model selection and estimation of multiple cisoids in additive white Gaussian noise of known variance. The method, that can be efficiently implemented, sample all the parameters and enable us to use any prior distribution.

We argued that the Jeffrey prior is a natural choice, and showed with simulations that good results could be obtained for cisoids closely spaced in frequency. The estimator, using a MAP rule for order selection and conditional mean estimates for frequencies and amplitudes, produced estimates whose variance attained the Cramer Rao bound until the separation was decreased to less then half a Fourier resolution. It then started to estimate a one cisoid model with frequency half way between the two, and after a brief transition interval with some spurious estimates, firmly chose this one cisoid model with low variance. In the case of a 90 degree phase difference between the amplitudes of the cisoids, the MAP selection rule produced many erroneous results. However, this was mitigated by a modified rule.

The method described in this paper could easily be extended to a more complex, and thus more application specific, signal models as long as the signal can be written as a sum of components scaled by amplitudes.

7 REFERENCES

[1] C. Andrieu and A. Doucet. Joint Bayesian model selection and estimation of noisy sinusoids via reverible jump MCMC. *IEEE Trans. Signal Processing*, 47(10):2667–2676, October 1999.

[2] A. P. Dawid. The trouble with bayes factors. Technical Report 202, University College London, May 1999.

[3] P. J. Green. Reversible jump MCMC computation and Bayesian model determination. *Biometrica*, 82:711–732, 1995.

[4] R. E. Kass. The geometry of asymptotic inference. *Statistical Science*, 4(3):188–234, 1989.