Joint Bayesian Model Selection and Estimation of Noisy Sinusoids via Reversible Jump MCMC

Christophe Andrieu and Arnaud Doucet

***Abstract*—In this paper, the problem of joint Bayesian model selection and parameter estimation for sinusoids in white Gaussian noise is addressed. An original Bayesian model is proposed that allows us to define *a posterior* distribution on the parameter space. All Bayesian inference is then based on this distribution. Unfortunately, a direct evaluation of this distribution and of its features, including posterior model probabilities, requires evaluation of some complicated high-dimensional integrals. We develop an efficient stochastic algorithm based on reversible jump Markov chain Monte Carlo methods to perform the Bayesian computation. A convergence result for this algorithm is established. In simulation, it appears that the performance of detection based on posterior model probabilities outperforms conventional detection schemes.**

***Index Terms*— Bayesian methods, MCMC, model selection, spectral analysis.**

I. Introduction

MODEL selection is a fundamental data analysis task. It has many applications in various fields of science and engineering. Over the past two decades, many of these problems have been addressed using information criteria such as Akaike information criterion (AIC) [1] or Rissanen's principle of minimum description length (MDL) [23]. The widespread use of these criteria is mainly due to their intrinsic simplicity. AIC and MDL are applied by evaluating two terms: a data term that requires the maximization of the likelihood and a penalty term of the complexity of the model. Within a Bayesian framework, model selection appears more difficult as it involves the evaluation of Bayes factors, which typically requires the computation of high-dimensional integrals with no closed-form analytical expression. These computational problems have limited the use of Bayesian model selection, except for the cases for which asymptotic expansions of the Bayes factors are valid [5].

In this paper, we address the problem of joint detection and estimation of sinusoids in white Gaussian noise. This problem is of great interest in many fields, including seismology, nuclear magnetic resonance, and radar. Under the assumption of a known number of sinusoids, several algorithms have been proposed to obtain the maximum likelihood (ML) frequency estimator [22], [24]. However, in practice, there are numerous applications where this number is unknown and has to be estimated [13]. The use of AIC and MDL requires reliable procedures for ML parameter estimation for each possible model and the evaluation of the criteria. Experimental evidence shows that AIC and MDL, which are criteria designed using asymptotic arguments, do indeed tend to estimate a wrong number of components for a small sample size and a low signal-to-noise ratio; see [13, Sec. VI].

We follow a Bayesian approach whereby the unknown parameters, including the amplitudes, the radial frequencies, and the noise variance, together with the number of sinusoids, are regarded as random quantities with known prior distribution. Several previous works have already addressed this problem, in some restricted scenarios, following the Bayesian approach. Bayesian parameter estimation and model selection for such signals have been addressed in a series of papers by Bretthorst [8]–[10] and, more recently, in [14] and [15]. In [12], the problem of power spectrum estimation in the case of harmonic signals is treated. Bayesian model selection for such signals based on posterior model probabilities has also been investigated by Djurić [13]. The main problem of the Bayesian approach is that it typically requires the evaluation of high-dimensional integrals that do not admit any closed-form analytical expression. In a few cases, for example when the sinusoids are well-separated and many samples are available, suitable analytic approximations to these integrals can be performed [8]. In [13], an asymptotic (in the number of data) expansion around an ML estimate of the frequencies is performed. 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 to solve this computational problem using classical deterministic multiple integration and Monte Carlo methods are presented in [10] and [12]; see also [7] in the context of damped sinusoids. The main problem of these methods is that they are not flexible and are difficult to use when the dimension of the integrand is large. Recently, Dou and Hodgson derived a Markov Chain Monte Carlo (MCMC) method [14], [15], but their algorithm suffers from several severe drawbacks; see Section IV-D for a discussion. Another MCMC method for parameter estimation of damped sinusoids is presented in [4].

To the best of our knowledge, the joint detection/estimation problem of harmonic signals has never been addressed in a

Manuscript received May 19, 1998; revised March 22, 1999. C. Andrieu was supported by AT&T Laboratories, Cambridge, U.K. A Doucet was supported by EPSRC, UK. The associate editor coordinating the review of this paper and approving it for publication was Dr. Jean-Jacques Fuchs.

The authors are with the Signal Processing Group, Department of Engineering, University of Cambridge, Cambridge, U.K. (e-mail: ca226@eng.cam.ac.uk; ad2@eng.cam.ac.uk).

Publisher Item Identifier S 1053-587X(99)07658-8.

Bayesian framework. In this paper, a new approach to jointly solve these problems is proposed. An original Bayesian model is proposed that allows us to define a posterior probability distribution over the space of possible structures of the signal. Similar to previous related works [8], [10], [12], [14], the posterior distribution appears highly nonlinear in its parameters, thus precluding analytical calculations. The case treated here is even more complex. Indeed, since the number of sinusoids is assumed random, the posterior distribution to be evaluated is defined on a finite disconnected union of subspaces of various dimensions. Each subspace corresponds to a model with a fixed number of sinusoids. To evaluate the joint posterior distribution of the number of sinusoids and their parameters, we propose an efficient stochastic algorithm based on reversible jump MCMC methods [17]. MCMC's are powerful stochastic algorithms that have revolutionized applied statistics; see [6] and [25] for some reviews.

The paper is organized as follows. In Section II, the signal model is given. In Section III, we formalize the Bayesian model and specify the prior distributions. Section IV is devoted to Bayesian computation. We first propose an MCMC sampler to perform Bayesian inference when the number of sinusoids is given. Then, a reversible jump MCMC algorithm is derived when the number of sinusoids is unknown. The uniform geometric convergence of this algorithm is established. The performance of these algorithms is illustrated by computer simulations and compared with classical detection methods in Section V. Finally, some conclusions are drawn in Section VI. Appendix A defines the notation used in the paper. The proof of convergence of the algorithm is given in Appendix B.

II. Problem Statement

Let $\mathbf{y} = (y[0], y[1], \cdots, y[N-1])^t$ be an observed vector of N real data samples. The elements of $\mathbf{y}$ may be represented by different models $\mathcal{M}_k$ corresponding either to samples of noise only ($k = 0$) or to the superposition of k ($k \geq 1$) sinusoids corrupted by noise; more precisely for $k = 0$ and $k \geq 1$, respectively

$$\begin{aligned} \mathcal{M}_0&: y[i] = n_0[i] \\ \mathcal{M}_k&: y[i] = \Sigma_{j=1}^{k}(a_{c_{j,k}} \cos[\omega_{j,k} i] + a_{s_{j,k}} \sin[\omega_{j,k} i]) \\ &\quad + n_k[i] \end{aligned} \tag{1}$$

where $\omega_{j_1,k} \neq \omega_{j_2,k}$ for $j_1 \neq j_2$ and $a_{c_{j,k}}$, $a_{s_{j,k}}$, $\omega_{j,k}$ are, respectively, the amplitudes and the radial frequency of the jth sinusoid for the model with k sinusoids. The noise sequence $\mathbf{n}_k \triangleq (n_k[0], \cdots, n_k[N-1])^t$ is assumed zero-mean white Gaussian of variance σ_k^2. In vector–matrix form, we have

$$\mathbf{y} = \mathbf{D}(\boldsymbol{\omega}_k)\mathbf{a}_k + \mathbf{n}_k \tag{2}$$

where $[\mathbf{a}_k]_{2i-1,1} \triangleq a_{c_{i,k}}$, $[\mathbf{a}_k]_{2i,1} \triangleq a_{s_{i,k}}$, and $[\boldsymbol{\omega}_k]_{i,1} \triangleq \omega_{i,k}$ for $i = 1, \cdots, k$. The $N \times 2k$ matrix $\mathbf{D}(\boldsymbol{\omega}_k)$ is defined as

$$\begin{aligned} [\mathbf{D}(\boldsymbol{\omega}_k)]_{i+1,2j-1} &= \cos[\omega_{j,k} i] \\ [\mathbf{D}(\boldsymbol{\omega}_k)]_{i+1,2j} &= \sin[\omega_{j,k} i] \end{aligned} \tag{3}$$

for $i = 0, \cdots, N-1$ and $j = 1, \cdots, k$. We assume here that the number k of sinusoids and their parameters $\boldsymbol{\theta}_k \triangleq (\mathbf{a}_k^t, \boldsymbol{\omega}_k^t, \sigma_k^2)^t$ are unknown. Given the data set $\mathbf{y}$, our objective is to estimate k and $\boldsymbol{\theta}_k$.

III. Bayesian Model and Aims

We follow a Bayesian approach where the unknowns k and $\boldsymbol{\theta}_k$ are regarded as being drawn from appropriate prior distributions. These priors reflect our degree of belief of the relevant values of the parameters [5]. We first propose a model that sets up a probability distribution over the space of possible structures of the signal. Subsequently, we specify the detection/estimation aims. Finally, we exploit the analytical properties of the model to obtain an expression, up to a normalizing constant, of the posterior distribution $p(k, \boldsymbol{\omega}_k|\mathbf{y})$.

A. Prior Distributions

The overall parameter space $\boldsymbol{\Theta}$ can be written as a finite union of subspaces $\boldsymbol{\Theta} = \cup_{k=0}^{k_{\max}} \{k\} \times \boldsymbol{\Theta}_k$, where $\boldsymbol{\Theta}_0 \triangleq \mathbb{R}^+$ and $\boldsymbol{\Theta}_k \triangleq \mathbb{R}^{2k} \times \boldsymbol{\Omega}_k \times \mathbb{R}^+$ for $k \in \{1, \cdots, k_{\max}\}$ with $\boldsymbol{\Omega}_k \triangleq (0, \pi)^k$ and $k_{\max} \triangleq \lfloor (N-1)/2 \rfloor$.[1] We also define $\boldsymbol{\Omega} \triangleq \cup_{k=0}^{k_{\max}} \{k\} \times \boldsymbol{\Omega}_k$, where $\boldsymbol{\Omega}_0 \triangleq \emptyset$. There is a natural hierarchical structure to this setup [17], which we formalize by modeling the joint distribution of all variables as

$$p(k, \boldsymbol{\theta}_k, \mathbf{y}) = p(\mathbf{y}|k, \boldsymbol{\theta}_k) p(\boldsymbol{\theta}_k|k) p(k) \tag{4}$$

where $p(k)$ is the prior model probability, $p(\boldsymbol{\theta}_k|k)$ is the parameter prior, and $p(\mathbf{y}|k, \boldsymbol{\theta}_k)$ is the likelihood. From the model given in Section II, the likelihood is

$$\begin{aligned} p(\mathbf{y}|k, \boldsymbol{\theta}_k) = (2\pi\sigma_k^2)^{-N/2} \exp\Bigg(&-\frac{1}{2\sigma_k^2}(\mathbf{y} - \mathbf{D}(\boldsymbol{\omega}_k)\mathbf{a}_k)^t \\ &\times (\mathbf{y} - \mathbf{D}(\boldsymbol{\omega}_k)\mathbf{a}_k)\Bigg). \end{aligned} \tag{5}$$

For $(k, \boldsymbol{\theta}_k)$, we assume the structure

$$p(k, \boldsymbol{\theta}_k) = p(k, \mathbf{a}_k, \boldsymbol{\omega}_k|\sigma_k^2) p(\sigma_k^2) \tag{6}$$

where σ_k^2 is a scale parameter that is assumed to be distributed according to a conjugate inverse-Gamma prior distribution, i.e., $\sigma_k^2 \sim \mathcal{IG}(\upsilon_0/2, \gamma_0/2)$. When $\upsilon_0 = 0$ and $\gamma_0 = 0$, we obtain Jeffreys' uninformative prior $p(\sigma_k^2) \propto 1/\sigma_k^2$ [5]. For $(k, \boldsymbol{a}_k, \boldsymbol{\omega}_k)$, we introduce the prior distribution

$$\begin{aligned} p(k, \mathbf{a}_k, \boldsymbol{\omega}_k|\sigma_k^2) \propto \frac{\Lambda^k}{k!} \exp(-\Lambda) &\frac{1}{|2\pi\sigma_k^2 \boldsymbol{\Sigma}_k|^{1/2}} \\ &\times \exp\left[-\frac{\mathbf{a}_k^t \boldsymbol{\Sigma}_k^{-1} \mathbf{a}_k}{2\sigma_k^2}\right] \frac{\mathbb{I}_{\boldsymbol{\Omega}}(k, \boldsymbol{\omega}_k)}{\pi^k} \end{aligned} \tag{7}$$

where $\boldsymbol{\Sigma}_k^{-1} = \delta^{-2} \mathbf{D}^t(\boldsymbol{\omega}_k)\mathbf{D}(\boldsymbol{\omega}_k)$. The prior probability model distribution $p(k)$ is a truncated Poisson distribution. Conditional on k, the frequencies are assumed uniformly distributed in $\boldsymbol{\Omega}_k$. Finally, conditional on $(k, \boldsymbol{\omega}_k)$, the amplitudes are

[1] The constraint $2k < N$ is added because otherwise, the columns of $\mathbf{D}(\boldsymbol{\omega}_k)$ are linearly dependent, and the parameters $\boldsymbol{\omega}_k$ may not be uniquely defined from the data [see (2)].

assumed zero-mean Gaussian with covariance $\sigma_k^2\mathbf{\Sigma}_k$. Proportionality in (7) comes from the fact that $k \leq k_{\max}$. The terms δ^2 and Λ can be, respectively, interpreted as an expected signal-to-noise ratio and the expected number of sinusoids. The influence of Λ can, of course, be removed by computing Bayes factors, namely, $p(\mathbf{y}|k_1)/p(\mathbf{y}|k_2)$. The prior distribution $p(\mathbf{a}_k, \sigma_k^2|k, \boldsymbol{\omega}_k)$ corresponds to the popular g-prior distribution; see [26] for motivation. We can also obtain (7) using a maximum entropy method [3]. In the case $k = 0$, we adopt the following conventions: $\mathbf{a}_0^t\mathbf{\Sigma}_0^{-1}\mathbf{a}_0 \triangleq 0$ and $|2\pi\sigma_0^2\mathbf{\Sigma}_0|^{1/2} \triangleq 1$.

B. Estimation/Detection Aims

The Bayesian inference of k and $\boldsymbol{\theta}_k$ is based on the joint posterior distribution $p(k, \boldsymbol{\theta}_k|\mathbf{y})$ obtained from Bayes' theorem. Our aim is to estimate this joint distribution from which, by standard probability marginalization and transformation techniques, we can "theoretically" obtain all posterior features of interest. In particular, it allows us to evaluate the posterior model probability $p(k|\mathbf{y})$, which can be used to perform model selection by selecting the model order as $\arg\max_{k\in\{0,\cdots,k_{\max}\}} p(k|\mathbf{y})$. In addition, it allows us to perform parameter estimation by computing the conditional expectation $\mathbb{E}(\boldsymbol{\theta}_k|\mathbf{y}, k)$. However, it is clear that it is not possible to obtain these quantities analytically. Indeed, it requires the evaluation of high-dimensional integrals of nonlinear functions in the parameters, as we shall see in Section III-C. We propose here to use an MCMC method to perform Bayesian computation. MCMC techniques were introduced in the mid 1950's in statistical physics but have only been introduced in applied statistics in the early 1990's and, more recently, in signal processing [6], [25]. The key idea is to build an ergodic Markov chain $(k^{(i)}, \boldsymbol{\theta}_{k^{(i)}})_{i\in\mathbb{N}}$ whose equilibrium distribution is the desired posterior distribution. Under weak additional assumptions, the $P \gg 1$ 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. For example

$$\hat{p}(k = j|\mathbf{y}) = \frac{1}{P}\sum_{i=1}^{P} \mathbb{I}_{\{j\}}\left(k^{(i)}\right)$$

and

$$\hat{\mathbb{E}}(\boldsymbol{\theta}_k|\mathbf{y}, k = j) = \frac{\sum_{i=1}^{P} \boldsymbol{\theta}_{k^{(i)}}^{(i)} \mathbb{I}_{\{j\}}(k^{(i)})}{\sum_{i=1}^{P} \mathbb{I}_{\{j\}}(k^{(i)})}. \tag{8}$$

However, for our problem, some integrations can be performed analytically and do not require any Monte Carlo integration scheme.

C. Integration of the Nuisance Parameters

The proposed Bayesian model allows for the integration of the so-called nuisance parameters $\mathbf{a}_k$ and σ_k^2 and, subsequently, to obtain an expression for $p(k, \boldsymbol{\omega}_k|\mathbf{y})$ up to a normalizing constant. According to Bayes theorem

$$\begin{aligned} p(\mathbf{a}_k, \boldsymbol{\omega}_k, \sigma_k^2, k|\mathbf{y}) &\propto p(\mathbf{y}|\mathbf{a}_k, \boldsymbol{\omega}_k, \sigma_k^2, k)p(k, \mathbf{a}_k, \boldsymbol{\omega}_k|\sigma_k^2)p(\sigma_k^2) \\ &\propto (2\pi\sigma_k^2)^{-N/2} \exp\left[\frac{-1}{2\sigma_k^2}(\mathbf{a}_k - \mathbf{m}_k)^t\mathbf{M}_k^{-1}(\mathbf{a}_k - \mathbf{m}_k)\right] \\ &\quad \times \exp\left[\frac{-1}{2\sigma_k^2}(\gamma_0 + \mathbf{y}^t\mathbf{P}_k\mathbf{y})\right](\sigma_k^2)^{-\upsilon_0/2-1} \\ &\quad \times |2\pi\mathbf{\Sigma}_k\sigma_k^2|^{-1/2}\frac{(\Lambda/\pi)^k}{k!}\mathbb{I}_{\mathbf{\Omega}}(k, \boldsymbol{\omega}_k) \end{aligned} \tag{9}$$

with

$$\begin{aligned} \mathbf{M}_k^{-1} &= \mathbf{D}^t(\boldsymbol{\omega}_k)\mathbf{D}(\boldsymbol{\omega}_k) + \mathbf{\Sigma}_k^{-1} \\ \mathbf{m}_k &= \mathbf{M}_k\mathbf{D}^t(\boldsymbol{\omega}_k)\mathbf{y} \\ \mathbf{P}_k &= \mathbf{I}_N - \mathbf{D}(\boldsymbol{\omega}_k)\mathbf{M}_k\mathbf{D}^t(\boldsymbol{\omega}_k). \end{aligned} \tag{10}$$

The integration of $\mathbf{a}_k$ (normal distribution) and then of σ_k^2 (inverse gamma distribution) yields

$$\begin{aligned} p(k, \boldsymbol{\omega}_k|\mathbf{y}) &\propto (\gamma_0 + \mathbf{y}^t\mathbf{P}_k\mathbf{y})^{-(N+\upsilon_0)/2} \\ &\quad \times \frac{(\Lambda/((\delta^2+1)\pi))^k}{k!}\mathbb{I}_{\mathbf{\Omega}}(k, \boldsymbol{\omega}_k). \end{aligned} \tag{11}$$

It is worth noticing that this posterior distribution is highly nonlinear in the angular frequencies $\boldsymbol{\omega}_k$ and that an expression of $p(k|\mathbf{y})$ cannot be obtained in closed form. We develop in the next sections MCMC methods to estimate the required posterior distribution $p(k, \boldsymbol{\omega}_k|\mathbf{y})$ or, if needed, $p(k, \mathbf{a}_k, \boldsymbol{\omega}_k, \sigma_k^2|\mathbf{y})$.

IV. Bayesian Computation

For the sake of clarity, we first assume in Section IV-A that k is given, and for notational convenience, we will not include k in the probability distributions in this subsection. In this case, the posterior distribution of interest is given by $p(\boldsymbol{\vartheta}_k|\mathbf{y}, k) \propto p(\mathbf{y}|k, \boldsymbol{\vartheta}_k)p(\boldsymbol{\vartheta}_k|k)$, where $\boldsymbol{\vartheta}_k \triangleq \{\boldsymbol{\omega}_k\}$ or $\boldsymbol{\vartheta}_k \triangleq \{\mathbf{a}_k, \boldsymbol{\omega}_k, \sigma_k^2\}$, depending on the posterior distribution from which we want to sample. In Section IV-B, we present an algorithm for the case when k is unknown.

A. Hybrid MCMC Sampler for a Fixed Dimension Model

We propose to use an hybrid MCMC sampler that combines Gibbs steps and Metropolis–Hastings (MH) steps, see [6] and [25, Sec. 2.4]. λ is a real number satisfying $0 < \lambda < 1$.

MCMC algorithm for spectral analysis

1. Initialization. Set $\boldsymbol{\theta}^{(0)} = (\mathbf{a}_k^{(0)}, \boldsymbol{\omega}_k^{(0)}, \sigma_k^{2(0)})$ and $i = 1$.
2. Iteration i
   - For $j = 1, \cdots, k$
     - — Sample $u \sim \mathcal{U}_{[0,1]}$.
     - — If $u < \lambda$, perform an MH step with $p(\omega_{j,k}|\mathbf{y}, \boldsymbol{\omega}_{-j,k}^{(i)})$ as invariant distribution and $q_1(\omega'_{j,k}|\omega_{j,k})$ as proposal distribution, see (Section IV-A1).
     - — Else perform an MH step with $p(\omega_{j,k}|\mathbf{y}, \boldsymbol{\omega}_{-j,k}^{(i)})$ as invariant distribution

and $q_2(\omega'_{j,k}|\omega_{j,k})$ as proposal distribution see (Section IV-A1).

End For.

3. Optional step: sample the nuisance parameters $(\sigma_k^{2(i)}, \mathbf{a}_k^{(i)}) \sim p(\sigma_k^2, \mathbf{a}_k|\mathbf{y}, \boldsymbol{\omega}_k^{(i)})$, see (Section IV-A2).
4. $i \leftarrow i+1$ and go to 2.

These different steps are detailed in the following subsections. In order to simplify notation, we drop the superscript $\cdot^{(i)}$ from all variables at iteration i.

1) Updating of the Frequencies: Sampling the frequencies is difficult because the distribution is nonlinear in these parameters. We have chosen here to sample the frequencies one-at-a-time using a mixture of MH steps. This well-known algorithm is described in [6] and [25, Sec. 2.4]. An MH step of invariant distribution, say, $\pi(\mathbf{x})$, and proposal distribution, say, $q(\mathbf{x}'|\mathbf{x})$, consists of sampling a candidate value $\mathbf{x}'$ given the current value $\mathbf{x}$ according to $q(\mathbf{x}'|\mathbf{x})$. The Markov chain moves toward $\mathbf{x}'$ with probability $\alpha(\mathbf{x}, \mathbf{x}') = \min\{1, (\pi(\mathbf{x})q(\mathbf{x}'|\mathbf{x}))^{-1}\pi(\mathbf{x}')q(\mathbf{x}|\mathbf{x}')\}$; otherwise, it remains equal to $\mathbf{x}$. This algorithm is very general, but to perform well in practice, it is necessary to use a clever proposal distribution to avoid rejecting too many candidates. In our application, the target distribution is the full conditional distribution of a frequency

$$p(\omega_{j,k}|\mathbf{y}, \boldsymbol{\omega}_{-j,k}) \propto [\gamma_0 + \mathbf{y}^t\mathbf{P}_k\mathbf{y}]^{-(N+\upsilon_0)/2}\mathbb{I}_{\boldsymbol{\Omega}}(k, \boldsymbol{\omega}_k). \quad (12)$$

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

$$q_1(\omega'_{j,k}|\omega_{j,k}) \propto \sum_{l=0}^{N_p-1} p_l \mathbb{I}_{[l\pi/N_p, (l+1)\pi/N_p)}(\omega'_{j,k}) \quad (13)$$

where p_l is the value of the squared modulus of the Fourier transform (FT) of the observations $\mathbf{y}$ at frequency $l\pi/N_p$. Typically, we will take $N_p = N$, but we can use $N_p > N$ to improve the interpolation of the FT via zero padding. The basic idea is to propose a frequency $\omega'_{j,k}$ independent of $\omega_{j,k}$ in the regions where the modulus of the FT has high values. The motivation for using such a proposal distribution is that the regions of interest of the posterior distribution are reached quickly. A related idea involves performing a piecewise constant approximation of the target distribution. However, the results are similar in simulations. With probability $1-\lambda$, we perform a MH step with proposal distribution $q_2(\omega'_{j,k}|\omega_{j,k})$

$$\omega'_{j,k}|\omega_{j,k} \sim \mathcal{N}(\omega_{j,k}, \sigma^2_{RW}). \quad (14)$$

This proposal distribution yields a candidate $\omega'_{j,k}$ that is a perturbation of the current frequency. The perturbation is a zero-mean Gaussian random variable with variance σ^2_{RW}. This random walk is introduced 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 for $i = 1, 2$ by

$$\min\left\{1, \left(\frac{\gamma_0 + \mathbf{y}^t\mathbf{P}_k\mathbf{y}}{\gamma_0 + \mathbf{y}^t\mathbf{P}'_k\mathbf{y}}\right)^{(N+\upsilon_0)/2} \frac{q_i(\omega_{j,k}|\omega'_{j,k})}{q_i(\omega'_{j,k}|\omega_{j,k})} \mathbb{I}_{\boldsymbol{\Omega}}(k, \boldsymbol{\omega}'_k)\right\} \quad (15)$$

where $\mathbf{P}'_k$, $\mathbf{M}'_k$, and $\boldsymbol{\Sigma}'_k$ are similar to $\mathbf{P}_k$, $\mathbf{M}_k$, and $\boldsymbol{\Sigma}_k$ with $\boldsymbol{\omega}_k$ replaced by $\boldsymbol{\omega}'_k \triangleq (\omega_{1,k}, \cdots, \omega_{j-1,k}, \omega'_{j,k}, \omega_{j+1,k}, \cdots, \omega_{k,k})$. Several other proposal distributions for the MH steps can be used, but we have found the combination of the two MH steps we propose to be very efficient in simulations.

2) Updating the Nuisance Parameters:

$$p(\sigma_k^2, \mathbf{a}_k|\mathbf{y}, \boldsymbol{\omega}_k) = p(\sigma_k^2|\mathbf{y}, \boldsymbol{\omega}_k)p(\mathbf{a}_k|\mathbf{y}, \boldsymbol{\omega}_k, \sigma_k^2). \quad (16)$$

By straightforward calculations, we obtain, using (9)

$$\begin{aligned} \sigma_k^2|(\mathbf{y}, \boldsymbol{\omega}_k) &\sim \mathcal{IG}\left(\frac{\upsilon_0+N}{2}, \frac{\gamma_0 + \mathbf{y}^t\mathbf{P}_k\mathbf{y}}{2}\right) \\ \mathbf{a}_k|(\mathbf{y}, \boldsymbol{\omega}_k, \sigma_k^2) &\sim \mathcal{N}(\mathbf{m}_k, \sigma_k^2\mathbf{M}_k) \end{aligned} \quad (17)$$

with $\mathbf{P}_k$, $\mathbf{m}_k$, and $\mathbf{M}_k$ defined in (10).

B. Bayesian Computation for an Unknown Model Dimension

Now, let us consider the case where k is unknown. Here, the Bayesian computation for the estimation of the joint posterior distribution $p(k, \boldsymbol{\theta}_k|\mathbf{y})$ is even more complex. One obvious solution would consist of running $k_{\max}+1$ independent MCMC samplers, each being associated with a fixed number $k = 0, \cdots, k_{\max}$. However, this approach suffers from severe drawbacks. First, it is computationally very expensive since $k_{\max}$ can be large. Second, the same computational effort is attributed to each value of k. In fact, some of these values are of no interest in practice because they have a very weak posterior model probability $p(k|\mathbf{y})$. Another solution would be to construct an MCMC sampler that would be able to sample directly from the joint distribution on $\boldsymbol{\Theta} = \cup_{k=0}^{k_{\max}}\{k\} \times \boldsymbol{\Theta}_k$. Standard MCMC methods are not able to "jump" between subspaces $\boldsymbol{\Theta}_k$ of different dimensions. However, recently, Green has introduced a new flexible class of MCMC samplers (the so-called reversible jump MCMC) that are capable of jumping between subspaces of different dimensions [17]. This is a general state-space MH algorithm. We propose candidates according to a set of proposal distributions. These candidates are randomly accepted according to an acceptance ratio that ensures reversibility and, thus, invariance of the Markov chain with respect to the posterior distribution. Here, the chain must move across subspaces of different dimensions, and therefore, the proposal distributions are more complex; see [17] for details. For our problem, the following moves have been selected:

1) birth of a new sinusoid, i.e., proposing a new sinusoid at random on $(0, \pi)$;
2) death of an existing sinusoid, i.e., removing a sinusoid chosen randomly;
3) update of the parameters of all the sinusoids, when $k \neq 0$, and the variance of the observation noise.

The birth and death moves perform dimension changes, respectively, from k to $k+1$ and k to $k-1$. These moves are defined by heuristic considerations, the only condition to be fulfilled being to maintain the correct invariant distribution. A particular choice will only have influence on the convergence rate of the algorithm. Other moves may be proposed, but we have found that the ones suggested here lead to satisfactory results.

The resulting transition kernel of the simulated Markov chain is then a mixture of the different transition kernels associated with the moves described above. This means that at each iteration, one of the candidate moves (birth, death, or update) is randomly chosen. The probabilities for choosing these moves are b_k, d_k, and u_k, respectively, such that $b_k + d_k + u_k = 1$ for all $0 \leq k \leq k_{\max}$. The move is performed if the algorithm accepts it. For $k = 0$, the death move is impossible; therefore, $d_0 \triangleq 0$. For $k = k_{\max}$, the birth move is impossible, and thus, $b_{k_{\max}} \triangleq 0$. Except in the cases described above, we take the probabilities

$$b_k \triangleq c \min\left\{1, \frac{p(k+1)}{p(k)}\right\}$$
$$d_{k+1} \triangleq c \min\left\{1, \frac{p(k)}{p(k+1)}\right\} \tag{18}$$

where $p(k)$ is the prior probability of model $\mathcal{M}_k$, and c is a parameter that tunes the proportion of dimension/update move. As pointed out in [17, pp. 719], this choice ensures that $b_k p(k)[d_{k+1}p(k+1)]^{-1} = 1$, which means that an MH algorithm on the sole dimension in the case of no observation would have 1 as acceptance probability. We take $c = 0.5$ and then $b_k + d_k \in [0.5, 1]$ for all k [17]. We can then describe the main steps of the algorithm.

Reversible jump MCMC algorithm

1. Initialization: set $(k^{(0)}, \boldsymbol{\theta}_k^{(0)}) \in \boldsymbol{\Theta}$.
2. Iteration i.
   - Sample $u \sim \mathcal{U}_{[0,1]}$.
   - If $(u \leq b_{k^{(i)}})$
     - then "birth" move (see Section IV-B1).
     - else if $(u \leq b_{k^{(i)}} + d_{k^{(i)}})$ then "death" move (see Section IV-B1).
     - else update the parameters (see Section IV-B2).

     End If.
3. $i \leftarrow i+1$ and go to 2.

We describe more precisely these different moves below. In what follows, in order to simplify notation, we drop the superscript $\cdot^{(i)}$ from all variables at iteration i.

1) Birth Move/Death Move: Suppose that the current state of the Markov chain is in $\{k\} \times \boldsymbol{\Theta}_k$. Then, we have the following.

Birth move

- Propose a new frequency at random on $(0, \pi)$: $\omega \sim \mathcal{U}_{(0,\pi)}$.
- Evaluate α_{birth}, see (21), and sample $u \sim \mathcal{U}_{[0,1]}$.
- If $u \leq \alpha_{\text{birth}}$ then the state of the Markov chain becomes $(k+1, \boldsymbol{\omega}_{k+1})$, else it remains at $(k, \boldsymbol{\omega}_k)$.
- or (**Optional**) If $u \leq \alpha_{\text{birth}}$ then the state of the Markov chain becomes $(k+1, \mathbf{a}_{k+1}, \boldsymbol{\omega}_{k+1}, \sigma^2_{k+1})$, else stay at $(k, \mathbf{a}_k, \boldsymbol{\omega}_k, \sigma^2_k)$ where $(\sigma^2_{k+1}, \mathbf{a}_{k+1})$ are sampled according to $p(\sigma^2_{k+1}, \mathbf{a}_{k+1}|k+1, \boldsymbol{\omega}_{k+1}, \mathbf{y})$, (see Section IV-A2).

Assume that the current state of the Markov chain is in $\{k+1\} \times \boldsymbol{\Theta}_{k+1}$. Then, we have the following.

Death move

- Choose a sinusoid at random among the $k+1$ existing sinusoids: $l \sim \mathcal{U}_{\{1,\ldots,k+1\}}$.
- Evaluate α_{death}, see (21), and sample $u \sim \mathcal{U}_{[0,1]}$.
- If $u \leq \alpha_{\text{death}}$ then the state of the Markov chain becomes $(k, \boldsymbol{\omega}_k)$, else it remains $(k+1, \boldsymbol{\omega}_{k+1})$.
- or (**Optional**) If $u \leq \alpha_{\text{death}}$ then the state of the Markov chain becomes $(k, \mathbf{a}_k, \boldsymbol{\omega}_k, \sigma^2_k)$, else it remains $(k+1, \mathbf{a}_{k+1}, \boldsymbol{\omega}_{k+1}, \sigma^2_{k+1})$ where $(\sigma^2_k, \mathbf{a}_k)$ are sampled according to $p(\sigma^2_k, \mathbf{a}_k|k, \boldsymbol{\omega}_k, \mathbf{y})$, (see Section IV-A2).

The acceptance ratio for the proposed moves are deduced from the following expression [17]:

$$r_{\text{birth}} \triangleq (\text{posterior distributions ratio}) \times (\text{proposal ratio}). \tag{19}$$

After simplifications

$$r_{\text{birth}} = \left(\frac{\gamma_0 + \mathbf{y}^t \mathbf{P}_k \mathbf{y}}{\gamma_0 + \mathbf{y}^t \mathbf{P}_{k+1} \mathbf{y}}\right)^{(N+\upsilon_0)/2} \frac{1}{(k+1)(1+\delta^2)}. \tag{20}$$

Then, the acceptance probabilities corresponding to the described moves are

$$\alpha_{\text{birth}} = \min\{1, r_{\text{birth}}\}$$
$$\alpha_{\text{death}} = \min\{1, r^{-1}_{\text{birth}}\}. \tag{21}$$

2) Update Move: The update move does not involve changing the dimension of the model. It requires an iteration of the hybrid MCMC sampler presented in Section IV-A.

C. Uniform Geometric Convergence of the Algorithm

It is easy to prove that the algorithm converges, i.e., that the Markov chain $(k^{(i)}, \boldsymbol{\omega}_k^{(i)})_{i \in \mathbb{N}}$ is ergodic. We prove here a stronger result by showing that $(k^{(i)}, \boldsymbol{\omega}_k^{(i)})_{i \in \mathbb{N}}$ converges to the required posterior distribution uniformly geometrically, i.e., at a geometric rate independent of the starting point. We have the following result.

Theorem 1: Let $(k^{(i)}, \boldsymbol{\omega}_k^{(i)})_{i\in\mathbb{N}}$ be the Markov chain whose transition kernel has been described in Section III. If $\mathbf{y} \notin \text{span}\{[\mathbf{D}(\boldsymbol{\omega}_k)]_{1:N,j}; j = 1, \cdots, 2k\}$ for any $(k, \boldsymbol{\omega}_k) \in \boldsymbol{\Omega}$, then this Markov chain converges to the probability distribution $p(k, \boldsymbol{\omega}_k|\mathbf{y})$. Furthermore, this convergence occurs at a uniform geometric rate, i.e., there exist a $C_0 > 0$ and $\rho \in [0, 1)$ such that whatever the initial point $(k^{(0)}, \boldsymbol{\omega}_k^{(0)}) \in \boldsymbol{\Omega}$ is

$$||p^{(i)}(k, \boldsymbol{\omega}_k) - p(k, \boldsymbol{\omega}_k|\mathbf{y})||_{TV} \leq C_0\rho^{\lfloor i/k_{\max}\rfloor} \tag{22}$$

where $p^{(i)}(\boldsymbol{\omega}_k, k)$ is the distribution of $(k^{(i)}, \boldsymbol{\omega}_k^{(i)})$, and $||\cdot||_{TV}$ is the total variation norm [25].

Proof: See Appendix B. ■

Corollary 1: If for each iteration i, one simulates the nuisance parameters $(\mathbf{a}_k, \sigma_k^2)$, then the distribution of the series $(k^{(i)}, \mathbf{a}_k^{(i)}, \boldsymbol{\omega}_k^{(i)}, \sigma_k^{2(i)})_{i\in\mathbb{N}}$ converges uniformly geometrically toward $p(k, \mathbf{a}_k, \boldsymbol{\omega}_k, \sigma_k^2|\mathbf{y})$ at the same rate ρ.

In other words, independent of the starting point of the Markov chain, the distribution of the Markov chain converges at least at a geometric rate to the required equilibrium distribution $p(k, \boldsymbol{\vartheta}_k|\mathbf{y})$.

Remark 1: In practice, we cannot evaluate ρ, but Theorem 1 proves its existence. This type of convergence ensures that a central limit theorem for ergodic averages is valid [25]. Moreover, in practice, there is empirical evidence that the Markov chain converges quickly.

D. Discussion

In the case where the number of sinusoids is known, an MCMC algorithm based on the Gibbs sampler has been proposed by Dou and Hodgson [14], [15]. Our sampler has several major differences to theirs and is more efficient. Their algorithm samples the amplitudes $\mathbf{a}_k$ and the noise σ_k^2 one at a time. In our algorithm, these simulation steps can be avoided if we are not interested in these parameters, and in this case, all these parameters are simulated jointly. Theoretical results established by Liu *et al.* [18] suggest that our sampling scheme is more efficient as it leads to a higher mixing rate for the simulated Markov chain. To simulate the radial frequencies, they sample for each j from the full conditional density $p(\omega_{j,k}|\mathbf{y}, k, \mathbf{a}_k, \boldsymbol{\omega}_{-j,k}, \sigma_k^2)$. As a direct simulation is impossible, they approximate this distribution by an expansion at its maximum. This yields a univariate Student-t distribution from which we can easily sample. However, this method requires a maximization that can be difficult if the SNR is low. Moreover, the algorithm is no longer a theoretically valid MCMC method as samples are drawn from an approximation of $p(\omega_{j,k}|\mathbf{y}, k, \mathbf{a}_k, \boldsymbol{\omega}_{-j,k}, \sigma_k^2)$. Thus, the associated convergence results of MCMC algorithms are not applicable. Our algorithm based on a MH step avoids such a maximization and remains theoretically valid. Furthermore, the proposal distribution based on the FT of the observations enables the Markov chain to avoid getting stuck in local maxima of the distribution.

TABLE I
PARAMETERS FOR THE FIRST EXPERIMENT

i	E_i	$-\arctan(a_{s_i}/a_{c_i})$	$\omega_i/2\pi$
1	20	0	0.2
2	6.3246	$\pi/4$	$0.2 + 1/N$
3	20	$\pi/3$	$0.2 + 2/N$

TABLE II
PARAMETERS FOR THE SECOND EXPERIMENT

i	E_i	$-\arctan(a_{s_i}/a_{c_i})$	$\omega_i/2\pi$
1	20	0	0.2
2	20	$\pi/4$	$0.2 + 1/(lN)$

V. SIMULATION RESULTS

In this section, we present an extensive Monte Carlo study of the performance of the method and algorithm that we have proposed to solve the problem of detection/estimation of sinusoids embedded in noise. It is not possible to theoretically evaluate such a performance as the quantities required (correct detection and over and under estimation probabilities) are not available in closed form. We thus propose a Monte Carlo simulation study for two experiments.

A. Description of the Data

The parameters for the first experiment are as follows: $N = 64$, $k = 3$. We define $E_i \triangleq a_{c_i}^2 + a_{s_i}^2$, and the parameters of the three sinusoids are given in Table I. The signal-to-noise ratio is defined as SNR $\triangleq 10\log_{10} E_1/(2\sigma^2)$. For the weaker sinusoid, the SNR is 5 dB less than the SNR for the other sinusoids. The parameters of the second experiment are $N = 64$, $k = 2$. The two sinusoids are defined in Table II. In the simulation we performed, we used $l = 1, 2, 4$. The signal-to-noise ratio for this experiment has the same definition as in the first one.

B. Parameters of the Algorithm

The developed algorithms require the specification of parameters that have no influence on the posterior distribution. These parameters only have an influence on the speed of convergence of the algorithm. The parameters we used are equal to $\lambda = 0.2$ and $\sigma_{RW} = 1/(5N)$. Parameter λ has been determined in a rather heuristic way. They are the first values we tried, and they provide the Markov chain with very satisfactory properties. Parameter σ_{RW} was set so that the mean acceptance probability was 0.4/0.5, which is often considered as a good indicator for a random walk with few parameters [16].

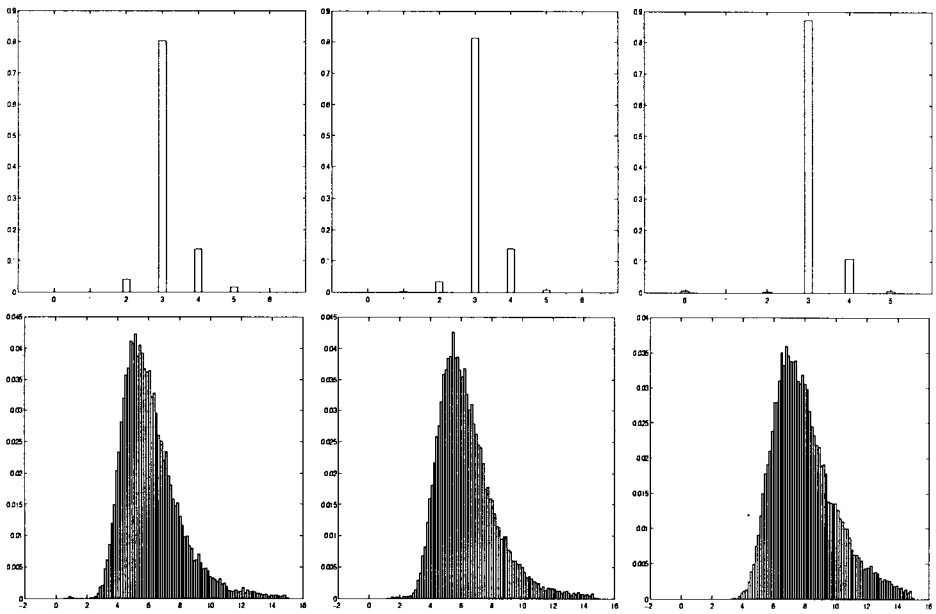

Fig. 1. Estimation of the posterior distributions $p(k|\mathbf{y})$ (top) and $p(\delta|\mathbf{y})$ (bottom) for SNR $= 3$ dB and $\beta_{\delta^2} = 1, 10, 100$ (left to right).

C. Extended Bayesian Model and Hyper-Hyperparameters

The Bayesian model is specified by υ_0, γ_0, Λ, and δ^2. For the parameters of the noise, we set $\upsilon_0 = \gamma_0 = 0$, i.e., we select the uninformative Jeffreys' prior. In practice, it is necessary to estimate Λ and δ^2 to given values. Typically, rough values of these parameters give good results. However, for sake of rigour, we have opted for a "blind" strategy, which is certainly more convincing. Note that the way we proceed is closely related to [21], for example, in a totally different context. The method consists of considering δ^2 and Λ as random, that is, we define an extended hierarchical model. The Bayesian model is thus slightly modified, and the posterior distribution we aim to estimate is now

$$p(\delta^2, \Lambda, k, \boldsymbol{\theta}_k|\mathbf{y}) \propto p(\mathbf{y}|k, \boldsymbol{\theta}_k)p(k, \boldsymbol{\theta}_k) \\ p(\delta^2|k, \boldsymbol{a}_k, \boldsymbol{\omega}_k, \sigma_k^2)p(\Lambda|k). \quad (23)$$

As δ^2 is a scale parameter, we ascribe a vague conjugate prior density to it: $\delta^2 \sim \mathcal{IG}(\alpha_{\delta^2}, \beta_{\delta^2})(\alpha_{\delta^2} = 2$ and $\beta_{\delta^2} > 0)$. We do not use Jeffreys' prior as $\delta^2 = 0$ can then be an absorbing state of the Markov chain.[2] We apply the same method to Λ by setting an uninformative conjugate prior [5] $\Lambda \sim \mathcal{G}a(1/2 + \varepsilon_1, \varepsilon_2)(\varepsilon_i \ll 1\ i = 1, 2)$. To estimate the distribution $p(\delta^2, \Lambda, k, \boldsymbol{\theta}_k|\mathbf{y})$, we need now to simulate the parameters (δ^2, Λ) each time an update is performed (Section IV-A). The probability densities allowing this update require the simulation of the nuisance parameters $\mathbf{a}_k$ and σ_k^2. We obtain

$$\delta^2|(k, \boldsymbol{\theta}_k) \sim \mathcal{IG}\left(k + \alpha_{\delta^2}, \frac{\mathbf{a}_k^t \mathbf{D}^t(\boldsymbol{\omega}_k)\mathbf{D}(\boldsymbol{\omega}_k)\mathbf{a}_k}{2\sigma_k^2} + \beta_{\delta^2}\right) \quad (24)$$

and Λ is updated using a MH step with $\mathcal{G}a(1/2+k+\varepsilon_1, 1+\varepsilon_2)$ as proposal distribution. In practice, estimates of $p(\delta^2|\mathbf{y})$ and $p(\Lambda|\mathbf{y})$ are quickly obtained, from which values of δ^2 and Λ can be estimated. Then, the original algorithm can be implemented with these values. We have chosen to numerically integrate δ^2 and Λ out, as in [21]. This approach was also suggested in [26, Remark 1, p. 237].[3]

D. Bayesian Robustness of the Prior on δ^2

The strategy we propose requires the specification of hyper-hyperparameters α_{δ^2} and β_{δ^2} for the random variable δ^2 (the prior for Λ is uninformative). $\alpha_{\delta^2} = 2$ ensures an infinite variance, and the crucial parameter is then β_{δ^2}.[4] Here, we experimentally demonstrate its weak influence on the posterior distributions $p(\delta^2|\mathbf{y})$ and $p(k|\mathbf{y})$. This general problem, namely, Bayesian robustness, still remains an unsolved problem. This stems from the fact that it is, in general, not possible to obtain any closed-form expression for the quantities of interest in realistic models. Simulation methods allow for numerical study of the effect of such a prior on the results.

We demonstrate the weak influence of the choice of β_{δ^2} for a wide range of values. Note, however, that an exhaustive study is impossible. For both models, we have applied the algorithm described in Section IV to the first experiment (Table I). Fig. 1 displays the estimates of $p(\delta|\mathbf{y})$ and $p(k|\mathbf{y})$ for a SNR of 3 dB and the following values of $\beta_{\delta^2} = 1, 10, 100$, thus covering a large range of values (even inconsistent values). It is experimentally observed that for all the signal-to-noise ratios and values of β_{δ^2}, results are very stable. Additional simulation results are presented in [3]. The detection procedure appears rather insensitive to the specification of hyper-hyperparameters.

E. Performance of the Detection

We do not present here estimation results because of a lack of space. Estimation is a by-product of the algorithm

[2] We could of course avoid this problem by defining a reversible jump between $\boldsymbol{\Theta}_0$ and $\boldsymbol{\Theta}_1$ that would sample δ^2 from any instrumental probability density.

[3] Simple convergence is still true, but we have not established a uniform geometric convergence result. There is empirical evidence that the algorithm still converges quickly.

[4] Recall that the mean of $\mathcal{IG}(\alpha, \beta)$ is $\beta/(\alpha - 1)$.

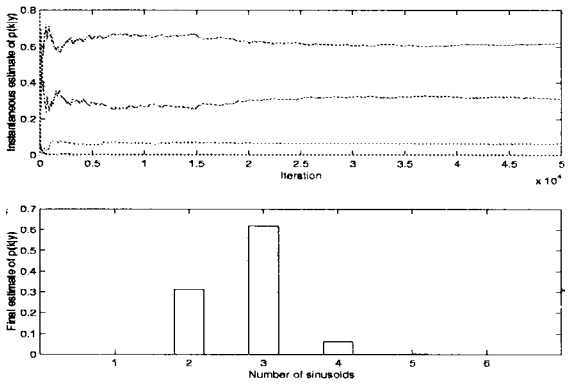

Fig. 2. Instantaneous estimation of $p(k|\mathbf{y})$ (top)—Final estimation of $p(k|\mathbf{y})$ (bottom) for one realization of the noise at 0 dB for experiment 1.

as it provides us with samples $\{(k^{(i)}, \boldsymbol{\theta}_k^{(i)}), i = 1, \cdots, P\}$ from the joint distribution $p(k, \boldsymbol{\theta}_k|\mathbf{y})$. However, as there is no identifiability constraint on the radial frequencies, the posterior distribution for a given dimension k is a mixture of $k!$ similar distributions up to a permutation of labels. A way to eliminate these artificial modes to ensure identifiability and perform practical estimation is to postprocess the samples. More precisely, we sort the radial frequencies of the samples; see the discussion in [21] for related approaches.

In this subsection, we compare the results for model order selection based on the maximum of the posterior model probability $p(k|\mathbf{y})$, i.e., $\arg\max_{k\in\{0,\cdots,k_{\max}\}} p(k|\mathbf{y})$, and those obtained on the same set of data using MDL and the criterion introduced by Djurić denoted D-MAP [13]. To perform efficiently the maximizations needed to evaluate MDL and D-MAP, we used a simulated annealing version of reversible jump MCMC algorithms; see [2] for details.

1) First Experiment: We have applied the algorithm presented in Section IV to 100 realizations of the first experiment. The number of iterations was 50 000, which was shown to be sufficient (histograms of the posterior model probability $p(k|\mathbf{y})$ were stabilized; see Fig. 2).

The algorithm was coded using Matlab©, and the simulation was performed on a Cyrix 200+®. Each processing of a realization required on average 130 s. The criterion we used was the *maximum* of $\hat{p}(k|\mathbf{y})$. Obtained results are presented in Table III.

2) Second Experiment: We applied the algorithm presented in Section IV for 100 realizations of the second experiment. The number of iterations of the algorithm was 50 000. The detection criterion we used was still the *maximum* of $\hat{p}(k|\mathbf{y})$. The obtained results are presented in Table IV.

F. Discussion

The results obtained using MDL and D-MAP are consistent with those of [13]. AIC was also implemented but tends to systematically overestimate the model order. Compared with results obtained using the posterior model probability $p(k|\mathbf{y})$, we observe that in all cases, this model order estimate outperforms other criteria for all signal-to-noise ratios. Depending on the signal-to-noise ratio and l values, correct detection rates can be improved by 10 to 30%.

TABLE III
PERFORMANCE OF DETECTION FOR THE FIRST EXPERIMENT

Criterion	*SNR*	$k \leq 1$	$k = 2$	$k = 3$	$k \geq 4$
	0dB	0	39	**43**	18
MDL	1dB	0	24	**53**	23
	2dB	0	13	**59**	28
	3dB	0	6	**63**	31
	0dB	1	**92**	7	0
$D-MAP$	1dB	0	**72**	28	0
	2dB	0	**80**	20	0
	3dB	0	**51**	49	0
	0dB	15	33	**52**	0
$\arg\max_{k\in\{0,\ldots,k_{\max}\}} p(k\|\mathbf{y})$	1dB	0	35	**63**	2
	2dB	0	19	**81**	0
	3dB	0	6	**93**	1

VI. CONCLUSIONS

In this paper, joint Bayesian model selection and parameter estimation of sinusoids in white Gaussian noise have been addressed. An original Bayesian model was proposed that allows us to define a posterior distribution over the space of possible structures of the signal. The evaluation of this posterior distribution and of its features of interest requires numerical methods. An efficient computational algorithm based on reversible jump MCMC methods was derived to estimate this posterior distribution. An extensive simulation study is carried out, and results show that model selection based on the posterior model probabilities $p(k|\mathbf{y})$ performs better than other classical criteria. This method is of great interest when addressing scenarios for which a low SNR, small sample size, or closely spaced frequencies are encountered. Of course, in more favorable cases, computationally cheaper methods are a good alternative.

We would like to point out that there are several possible extensions to this work. Indeed, the framework we propose is flexible enough to allow the inclusion of any additional prior knowledge. It could be adapted to address the cases of damped sinusoids [4], non-Gaussian noise, or the more challenging problem of sinusoids in colored noise [20].

APPENDIX A
NOTATION

- $[\mathbf{A}]_{i,j}$ is the ith row, jth column of matrix $\mathbf{A}$.
- $|\mathbf{A}|$ is the determinant of matrix $\mathbf{A}$.
- If $\mathbf{z} \triangleq (z_1, \cdots, z_{j-1}, z_j, z_{j+1}, \cdots, z_k)^t$, then $\mathbf{z}_{-j} \triangleq (z_1, \cdots, z_{j-1}, z_{j+1}, \cdots, z_k)^t$.

TABLE IV
PERFORMANCE OF DETECTION FOR THE SECOND EXPERIMENT

Criterion	*SNR*	l	$k \leq 1$	$k = 2$	$k \geq 3$
MDL	3dB	1	0	**69**	31
		2	23	**66**	11
		4	**57**	29	14
	10dB	1	0	**72**	28
		2	0	**69**	31
		4	0	**70**	30
$D-MAP$	3dB	1	0	**97**	3
		2	13	**86**	1
		4	**93**	7	0
	10dB	1	0	**100**	0
		2	0	**100**	0
		3	5	**91**	0
$\arg\max_{k\in\{0,\ldots,k_{\max}\}} p(k\|\mathbf{y})$	3dB	1	0	**99**	1
		2	0	**95**	4
		4	**81**	19	0
	10dB	1	0	**100**	0
		2	0	**100**	2
		4	1	**99**	0

- $\mathbf{0}_{n\times p}$ is the null matrix of dimension $n \times p$.
- $\mathbf{I}_n$ is the identity matrix of dimension $n \times n$.
- $\mathbb{I}_E(\mathbf{z})$ is the indicator function of the set E (1 if $\mathbf{z} \in E$, 0 otherwise).
- $\lfloor z \rfloor$ is the highest integer strictly less than z.
- $\mathbf{z} \sim p(\mathbf{z})$: $\mathbf{z}$ is distributed according to $p(\mathbf{z})$.
- $\mathbf{z}|\mathbf{y} \sim p(\mathbf{z})$: the conditional distribution of $\mathbf{z}$ given $\mathbf{y}$ is $p(\mathbf{z})$.

Probability distribution	$\mathcal{F}$	$f_{\mathcal{F}}(\cdot)$
Inverse Gamma	$\mathcal{IG}(\alpha, \beta)$	$\frac{\beta^\alpha}{\Gamma(\alpha)} z^{-\alpha-1} \exp(-\beta/z)\mathbb{I}_{[0,+\infty)}(z)$
Gamma	$\mathcal{G}a(\alpha, \beta)$	$\frac{\beta^\alpha}{\Gamma(\alpha)} z^{\alpha-1} \exp(-\beta z)\mathbb{I}_{[0,+\infty)}(z)$
Gaussian	$\mathcal{N}(\mathbf{m}, \Sigma)$	$\lvert 2\pi\mathbf{\Sigma}\rvert^{-1/2} \exp(-\frac{1}{2}(\mathbf{z}-\mathbf{m})^t\mathbf{\Sigma}^{-1}(\mathbf{z}-\mathbf{m}))$
Uniform	$\mathcal{U}_A$	$[\int_A d\mathbf{z}]^{-1}\mathbb{I}_A(\mathbf{z})$

APPENDIX B
PROOF OF THEOREM 1

Before proving Theorem 1, we need the two following Lemmas:

Lemma 1: We denote $\mathbf{P}_k^*$ as the matrix $\mathbf{P}_k$ for which $\delta^2 \to +\infty$. This is the projection matrix onto the space orthogonal to $\text{span}\{[\mathbf{D}(\boldsymbol{\omega}_k)]_{1:N,j}; j = 1, \cdots, 2k\}$. Let $\mathbf{v} \in \mathbb{R}^N$ then $\mathbf{v}^t\mathbf{P}_k^*\mathbf{v} = 0$ if and only if $\mathbf{v}$ belongs to the space spanned by the columns of $\mathbf{D}(\boldsymbol{\omega}_k)$, $\boldsymbol{\omega}_k \in \mathbf{\Omega}_k$.

Then, noting that $\mathbf{y}^t\mathbf{P}_k\mathbf{y} = 1/(1+\delta^2)\mathbf{y}^t\mathbf{y} + (\delta^2/(1+\delta^2))\mathbf{y}^t\mathbf{P}_k^*\mathbf{y}$, we obtain the following results.

Corollary 2: If the observed data $\mathbf{y}$ are really noisy, i.e., $\mathbf{y} \notin \text{span}\{[\mathbf{D}(\boldsymbol{\omega}_k)]_{1:N,j}; j = 1, \cdots, 2k\}$ for any $(k, \boldsymbol{\omega}_k) \in \mathbf{\Omega}$, then there exists $\varepsilon > 0$ such that for all $k \leq k_{\max}$ ($k_{\max} = \lfloor (N-1)/2 \rfloor$), $\delta^2 \in \mathbb{R}^+$ and $\boldsymbol{\omega}_k \in \mathbf{\Omega}_k$

$$\mathbf{y}^t\mathbf{P}_k\mathbf{y} \geq \varepsilon > 0. \tag{25}$$

Lemma 2: For all $k \leq k_{\max}$, $\delta^2 \in \mathbb{R}^+$, and $\boldsymbol{\omega}_k \in \mathbf{\Omega}_k$

$$\mathbf{y}^t\mathbf{P}_k\mathbf{y} \leq \mathbf{y}^t\mathbf{y}. \tag{26}$$

Proof of Theorem 1: Let $\mathcal{K}(k_1, \boldsymbol{\omega}_{k_1}; k_2, d\boldsymbol{\omega}_{k_2})$ denote the transition kernel of the Markov chain, i.e., for fixed $(k_1, \boldsymbol{\omega}_{k_1})$

$$\Pr((k_2, \boldsymbol{\omega}_{k_2}) \in \{j\} \times A | (k_1, \boldsymbol{\omega}_{k_1})) = \int_A \mathcal{K}(k_1, \boldsymbol{\omega}_{k_1}; j, d\boldsymbol{\omega}_j)$$

where $A \in \mathcal{B}(\mathbf{\Omega}_j)$. By construction $\mathcal{K}(k_1, \boldsymbol{\omega}_{k_1}; k_2, d\boldsymbol{\omega}_{k_2})$ admits $p(k, \boldsymbol{\omega}_k|\mathbf{y})$ as an invariant probability distribution. We now prove the $p(\cdot|\mathbf{y})$-irreducibility of the Markov chain. To prove this result, we first establish $\phi(\cdot)$-irreducibility of the Markov chain, where $\phi(\cdot)$ is another probability measure on $\mathbf{\Omega}$.

For each $k_1 = 1, \cdots, k_{\max}$ and $\boldsymbol{\omega}_{k_1} \in \mathbf{\Omega}_{k_1}$, we have

$$\mathcal{K}(k_1, \boldsymbol{\omega}_{k_1}; k_2, d\boldsymbol{\omega}_{k_2}) \geq \min\{1, r_{\text{death}}\} d_{k_1} \frac{\delta_{S_{\boldsymbol{\omega}_{k_1}}}(d\boldsymbol{\omega}_{k_2})}{k_1} \tag{27}$$

for any $(k_1, \boldsymbol{\omega}_{k_1}), (k_2, \boldsymbol{\omega}_{k_2}) \in \mathbf{\Omega}$, where $1/k_1$ is the probability of choosing one of the sinusoids to suppress it, and $S_{\boldsymbol{\omega}_{k_1}} \triangleq \{\boldsymbol{\omega}' \in \mathbf{\Omega}_{k_1-1} / \exists l \in \{1, \cdots, k_1\}$ such that $\boldsymbol{\omega}' = \boldsymbol{\omega}_{k_1,-l}\}$. Then, from (20) and for all $k_1 = 1, \cdots, k_{\max}$ we obtain

$$\begin{aligned} r_{\text{death}}^{-1} &= \left(\frac{\gamma_0 + \mathbf{y}^t\mathbf{P}_{k_1-1}\mathbf{y}}{\gamma_0 + \mathbf{y}^t\mathbf{P}_{k_1}\mathbf{y}}\right)^{(N+\upsilon_0)/2} \frac{1}{(1+\delta^2)k_1} \\ &\leq \left(\frac{\gamma_0 + \mathbf{y}^t\mathbf{y}}{\varepsilon}\right)^{(N+\upsilon_0)/2} \frac{1}{(1+\delta^2)k_1} < M < +\infty \end{aligned} \tag{28}$$

where we have used Lemmas 1 and 2 for the existence of ε and M.

Thus, a sufficiently high M exists such that

$$\mathcal{K}(k_1, \boldsymbol{\omega}_{k_1}; k_1 - 1, d\boldsymbol{\omega}_{k_1-1}) \geq \frac{d_{k_1}}{M} \frac{\delta_{S_{\boldsymbol{\omega}_{k_1}}}(d\boldsymbol{\omega}_{k_1-1})}{k_1} \tag{29}$$

and there exists $\alpha > 0$ such that for all $1 \leq k_1 \leq k_{\max}$ and $\boldsymbol{\omega}_{k_1} \in \mathbf{\Omega}_{k_1}$

$$\begin{aligned} &\mathcal{K}(k_1, \boldsymbol{\omega}_{k_1}; k_1 - 1, d\boldsymbol{\omega}_{k_1-1}) \\ &\quad \geq \frac{\alpha}{k_1} \delta_{S_{\boldsymbol{\omega}_{k_1}}}(d\boldsymbol{\omega}_{k_1-1}) \quad \text{with } \alpha > 0. \end{aligned} \tag{30}$$

Therefore, for all k_1, we can reach the empty configuration with a nonzero probability. Let $\phi(\cdot)$ be the probability distribution defined as $\phi(k, d\boldsymbol{\omega}_k) \triangleq \mathbb{I}_{\{0\}}(k)$. As $p(k, \boldsymbol{\omega}_k|\mathbf{y})$ is an invariant distribution of $\mathcal{K}(\cdot;\cdot)$, and the Markov chain is ϕ-irreducible, then from [19, Prop. 4.2.2] and [25, Th. 1*, p. 1758], the Markov chain is $p(k, \boldsymbol{\omega}_k|\mathbf{y})$-irreducible. Aperiodicity is straightforward. Indeed, there is a nonzero probability of choosing the update move in the empty configuration leading to

$$\mathcal{K}(0, \boldsymbol{\omega}_0;\ 0, d\boldsymbol{\omega}_0) \geq u_0 > 0. \tag{31}$$

(In this case, i.e., $k = 0$, we keep for notational convenience the same notation for the transition kernel even if $\boldsymbol{\omega}_0$ does not exist.) The Markov chain thus admits $p(k, \boldsymbol{\omega}_k|\mathbf{y})$ as equilibrium distribution [25, Th. 1*, p. 1758].

We now prove that the Markov chain is uniformly ergodic using the fact that $\boldsymbol{\Omega}$ is included in a compact set, i.e., we show that $\boldsymbol{\Omega}$ is a small set. From (30) and (31), we deduce for $k_1 = 1, \cdots, k_{\max}$ that when we iterate $k_{\max}$ times the kernel $\mathcal{K}(\cdot;\cdot)$, then the resulting transition kernel denoted $\mathcal{K}^{(k_{\max})}$ satisfies

$$\begin{aligned}
&\mathcal{K}^{(k_{\max})}(k_1, \boldsymbol{\omega}_{k_1};\ 0, d\boldsymbol{\omega}_0) \\
&\quad = \int_{\boldsymbol{\Omega}} \mathcal{K}^{(k_1)}(k_1, \boldsymbol{\omega}_{k_1};\ l, d\boldsymbol{\omega}) \mathcal{K}^{(k_{\max}-k_1)}(l, \boldsymbol{\omega};\ 0, d\boldsymbol{\omega}_0) \\
&\quad \geq \int_{\boldsymbol{\Omega}_0} \mathcal{K}^{(k_1)}(k_1, \boldsymbol{\omega}_{k_1};\ l, d\boldsymbol{\omega}) \mathcal{K}^{(k_{\max}-k_1)}(l, \boldsymbol{\omega};\ 0, d\boldsymbol{\omega}_0) \\
&\quad = \mathcal{K}^{(k_1)}(k_1, \boldsymbol{\omega}_{k_1};\ 0, d\boldsymbol{\omega}_0) \mathcal{K}^{(k_{\max}-k_1)}(0, \boldsymbol{\omega}_0;\ 0, d\boldsymbol{\omega}_0) \\
&\quad \geq \alpha^{k_1} u_0^{k_{\max}-k_1} \phi(0, d\boldsymbol{\omega}_0).
\end{aligned} \tag{32}$$

The last inequality is also valid for $k_1 = 0$. It allows us to write for any $(k_1, \boldsymbol{\omega}_{k_1}) \in \boldsymbol{\Omega}$

$$\mathcal{K}^{(k_{\max})}(k_1, \boldsymbol{\omega}_{k_1};\ k_2, d\boldsymbol{\omega}_{k_2}) \geq \eta \phi(k_2, d\boldsymbol{\omega}_{k_2}) \tag{33}$$

where $\eta \triangleq \min_{k \in \{0, \cdots, k_{\max}\}} \alpha^k u_0^{k_{\max}-k}$. Thus, from [19, Th. 5.4] and [25, Prop. 2], there exists $C_0 > 0$ such that for all i

$$||p^{(i)}(k, \boldsymbol{\omega}_k) - p(k, \boldsymbol{\omega}_k|\mathbf{y})||_{TV} \leq C_0 (1 - \eta)^{\lfloor i/k_{\max} \rfloor}. \tag{34}$$

■

REFERENCES

[1] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 716–723, 1974.

[2] C. Andrieu and A. Doucet, "Efficient stochastic maximum *a posteriori* estimation for harmonic signals," in *Proc. EUSIPCO*, Island of Rhodes, Greece, Sept. 1998.

[3] ———, "Joint Bayesian detection and estimation of harmonics signals via reversible jump MCMC," Tech. Rep. CUED-F-TR318, Dept. Eng., Univ. Cambridge, Cambridge, U.K., 1998.

[4] P. Barone and R. Ragona, "Bayesian estimation of parameters of a damped sinusoidal model by a Markov chain Monte Carlo method," *IEEE Trans. Signal Processing*, vol. 45, pp. 1806–1814, July 1997.

[5] J. M. Bernardo and A. F. M. Smith, *Bayesian Theory*. New York: Wiley, 1994, Series in Applied Probability and Statistics.

[6] J. Besag, P. J. Green, D. Hidgon, and K. Mengersen, "Bayesian computation and stochastic systems," *Stat. Sci.*, vol. 10, pp. 3–66, 1995.

[7] W. B. Bishop and P. M. Djurić, "Model order selection of damped sinusoids in noise by predictive densities," *IEEE Trans. Signal Processing* vol. 44, pp. 611–619, Mar. 1996.

[8] G. L. Bretthorst, *Bayesian Spectrum Analysis and Parameter Estimation.* New York: Springer-Verlag, 1988, Lecture Notes in Statistics, vol. 48.

[9] ———, "Bayesian analysis—I: Parameter estimation using quadrature NMR models," *J. Magn. Reson.*, vol. 88, pp. 533–551, 1990.

[10] ———, "Bayesian analysis—II: Signal detection and model selection," *J. Magn. Reson.*, vol. 88, pp. 552–570, 1990.

[11] L. Devroye, *Non-Uniform Random Variate Generation.* New York: Springer, 1986.

[12] P. M. Djurić and H. Li, "Bayesian spectrum estimation of harmonic signals," *IEEE Signal Processing Lett.*, vol. 2, pp. 213–215, Nov. 1995.

[13] P. M. Djurić, "A model selection rule for sinusoids in white Gaussian noise," *IEEE Trans. Signal Processing*, vol. 44, pp. 1744–1751, July 1996.

[14] L. Dou and R. J. W. Hodgson, "Bayesian inference and Gibbs sampling in spectral analysis and parameter estimation—I," *Inv. Probl.*, pp. 1069–1085, 1995.

[15] ———, "Bayesian inference and Gibbs sampling in spectral analysis and parameter estimation—II," *Inv. Probl.*, pp. 121–137, 1996.

[16] W. R. Gilks and G. O. Roberts, "Strategies for improving MCMC," in *Markov Chain Monte Carlo in Practice.* London, U.K.: Chapman & Hall, 1996, pp. 89–110.

[17] P. J. Green, "Reversible jump MCMC computation and Bayesian model determination," *Biometrika*, vol. 82, pp. 711–732, 1995.

[18] J. S. Liu, W. H. Wong, and A. Kong, "Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes," *Biometrika*, vol. 81, no. 1, pp. 27–40, 1994.

[19] S. P. Meyn and R. Tweedie, *Markov Chains and Stochastic Stability.* London, U.K.: Springer-Verlag, 1993.

[20] V. Nagesha and S. M. Kay, "Spectral analysis based on the canonical autoregressive decomposition," *IEEE Trans. Signal Processing*, vol. 44, pp. 1719–1733, July 1996.

[21] S. Richardson and P. J. Green, "On Bayesian analysis of mixtures with unknown number of components," *J. R. Stat. Soc. B*, vol. 59, no. 4, pp. 731–792, 1997.

[22] D. C. Rife and R. R. Boorstyn, "Multiple-tone parameter estimation from discrete-time observations," *Bell Syst. Tech. J.*, vol. 55, pp. 1389–1410, 1976.

[23] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, pp. 465–471, 1978.

[24] P. Stoica, R. L. Moses, B. Friedlander, and T. Soderstrom, "Maximum likelihood estimation of the parameters of multiple sinusoids from noisy measurements," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 378–392, 1989.

[25] L. Tierney, "Markov chains for exploring posterior distributions," *Ann. Stat.*, pp. 1701–1762, 1994, with discussion.

[26] A. Zellner, "On assessing prior distributions and Bayesian regression analysis with g-prior distributions," in *Bayesian Inference and Decision Techniques*, P. Goel and A. Zellner, Eds. New York: Elsevier, 1986, pp. 233–243.

Christophe Andrieu was born in France in 1968. He received the M.S. degree from the Institut National des Télécommunications, Paris, France, in 1993 and the D.E.A. and Ph.D. degrees from the University of Paris XV in 1994 and 1998, respectively.

Since 1998, he has been a Research Associate with the Signal Processing Group, Cambridge University, Cambridge, U.K. His research interests include spectral analysis, source separation, Markov chain Monte Carlo methods, and stochastic optimization.

Arnaud Doucet was born in Melle, France, on November 2, 1970. He received the M.S. degree from the Institut National des Télécommunications, Paris, France, in 1993 and the Ph.D. degree from the University of Paris XI, Orsay, France, in 1997.

In 1998, he was a Visiting Researcher with the Signal Processing Group, Cambridge University, Cambridge, U.K. Since 1999, he has been a Research Associate with the same group. His research interests include Markov chain Monte Carlo methods, sequential Monte Carlo methods, Bayesian estimation, and hidden Markov models.