EFFICIENT STOCHASTIC MAXIMUM A POSTERIORI ESTIMATION FOR HARMONIC SIGNALS

*Christophe Andrieu - Arnaud Doucet
Signal Processing Group, Department of Engineering
University of Cambridge, Trumpington St., CB2 1PZ Cambridge, UK.
e-mail: ca226@eng.cam.ac.uk - ad2@eng.cam.ac.uk

ABSTRACT

In this paper, we address the problems of ML (Maximum Likelihood) parameter estimation and model order selection for harmonic signals using classical criteria. Solving these problems requires the maximization of complex multimodal functions. These optimization problems are shown as being equivalent to the estimation of joint and marginal maximum *a posteriori* (MAP) estimates under given Bayesian models. Efficient stochastic algorithms based on non-homogeneous Markov chain Monte Carlo methods are presented to solve these problems and their convergence is established. Computer simulations demonstrate the efficiency of these algorithms.

1 INTRODUCTION

The harmonic retrieval problem is a fundamental problem in signal processing. Under the assumption of an additive white Gaussian noise, it has already often be addressed in the literature. When the number of harmonics is known, the Maximum Likelihood (ML) estimate is the most popular estimate [5, 8, 10]. Unfortunately, obtaining this estimate involves the optimization of a typically complex multimodal function and this problem admits no analytical solution. Numerous methods have been proposed to solve these problems. They are mainly deterministic and thus very sensitive to initialization. As a consequence, they are not reliable as soon as the signal to noise ratio is weak [5, 8, 10]. Standard stochastic optimization algorithms have also already been applied but they do not use the statistical structure of the model and are not very efficient in practice [6, 7]. When the number of harmonics is unknown, the detection problem, *i.e.* the selection of the model order, is often addressed using classical criteria such as AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion) or MDL (Minimum Description Length). These criteria require the evaluation of the ML estimate for each possible model order. The aim of this article is to present efficient stochastic algorithms to solve these optimization problems.

* Alphabetical order

The problem of harmonic retrieval is first addressed in a Bayesian framework. We show that obtaining the ML estimate and estimating the model order using classical information criteria is equivalent to obtaining some marginal and joint MAP (Maximum a Posteriori) estimates for given Bayesian models. Then we develop efficient stochastic algorithms to obtain these estimates. When the model order is fixed, stochastic optimization to obtain the ML estimate is performed using a stochastic annealing version of an efficient MCMC (Markov chain Monte Carlo) sampler, see [9] for an introduction to MCMC algorithms. To estimate the model order based on classical information criteria, we develop a stochastic annealing version of a reversible jump MCMC sampler [4]. Sufficient conditions to ensure convergence towards the set of global maxima of the functions to optimize are established using general state-space Markov chain theory. The efficiency of these algorithms is demonstrated via computer simulations.

2 ESTIMATION OBJECTIVES AND BAYESIAN MODELS

2.1 Model

Let $\mathbf{y} \triangleq (y_0, y_1, \ldots, y_{T-1})^{\mathrm{T}}$ be an observed vector of T real data samples. The elements of $\mathbf{y}$ may be represented by different models $\mathcal{M}_k$ corresponding either to samples of noise only or to the superimposition of k sinusoids corrupted by noise:

$$\begin{array}{ll} \mathcal{M}_0: & y_t = n_t^0 \qquad\qquad\qquad\qquad\qquad\quad k=0 \\ \mathcal{M}_k: & y_t = \sum_{j=1}^{k} a_{j,k} \cos\left[\omega_{j,k} t + \phi_{j,k}\right] + n_t^k \quad k \geq 1 \end{array}$$

where $\omega_{j_1,k} \neq \omega_{j_2,k}$ for $j_1 \neq j_2$, $a_{j,k}$, $\omega_{j,k}$, ϕ_j being respectively the amplitude, the radial frequency and the phase of the j^{th} sinusoid for the model with k sinusoids. The noise sequence $\mathbf{n}_k \triangleq \left(n_0^k, \ldots, n_{T-1}^k\right)^{\mathrm{T}}$ is assumed zero-mean white Gaussian of variance σ_k^2. In a vector-matrix form, we have

$$\mathbf{y} = \mathbf{D}\left(\boldsymbol{\omega}_k\right) \mathbf{a}_k + \mathbf{n}_k \tag{1}$$

We denote $\mathbf{D}\left(\boldsymbol{\omega}_k\right)$ the $T \times 2k$ matrix defined by

$$\mathbf{D}\left(\boldsymbol{\omega}_k\right) \triangleq \left(\begin{array}{ccccccc} \mathbf{d}_{c_{1,k}} & \mathbf{d}_{s_{1,k}} & \mathbf{d}_{c_{2,k}} & \cdots & \mathbf{d}_{c_{k,k}} & \mathbf{d}_{s_{k,k}} \end{array} \right)$$

$$\mathbf{d}_{c_{j,k}} \triangleq \left(\begin{array}{cccc} 1 & \cos[\omega_{j,k}] & \cdots & \cos[\omega_{j,k}(T-1)] \end{array} \right)^{\mathrm{T}}$$
$$\mathbf{d}_{s_{j,k}} \triangleq \left(\begin{array}{cccc} 0 & \sin[\omega_{j,k}] & \cdots & \sin[\omega_{j,k}(T-1)] \end{array} \right)^{\mathrm{T}}$$
$$\mathbf{a}_k \triangleq \left(\begin{array}{cccccc} a_{c_{1,k}} & a_{s_{1,k}} & \cdots & \cdots & a_{c_{k,k}} & a_{s_{k,k}} \end{array} \right)^{\mathrm{T}}$$
$$\boldsymbol{\omega}_k \triangleq (\omega_{1,k}, \ldots, \omega_{k,k})^{\mathrm{T}}$$
with $a_{c_j} = a_j \cos[\phi_j]$ and $a_{s_j} = -a_j \sin[\phi_j]$.

2.2 Estimation objectives

2.2.1 Fixed model order

When the model order k is known, we want to obtain the ML estimate $\widehat{\boldsymbol{\omega}_k}$ of the frequencies $\boldsymbol{\omega}_k$. This is equivalent to minimize the following function [8]:

$$\widehat{\boldsymbol{\omega}_k} = \arg\min_{\boldsymbol{\omega}_k \in \Omega_k} \mathbf{y}^{\mathrm{T}} \mathbf{P}(\boldsymbol{\omega}_k) \mathbf{y} \quad (2)$$

where $\Omega_k \triangleq \left\{ \boldsymbol{\omega}_k \in (0,\pi)^k \right\}$ and

$$\mathbf{P}(\boldsymbol{\omega}_k) \triangleq \mathbf{I}_T - \mathbf{D}(\boldsymbol{\omega}_k) \left[\mathbf{D}^{\mathrm{T}}(\boldsymbol{\omega}_k) \mathbf{D}(\boldsymbol{\omega}_k) \right]^{-1} \mathbf{D}^{\mathrm{T}}(\boldsymbol{\omega}_k) \quad (3)$$

2.2.2 Unknown model order

When the model order k is unknown, we want to select it according to a classical criterion such as AIC, BIC, MDL or the approximated Bayes factor DMAP introduced in [2]. All these criteria are of the following form:

$$\frac{T}{2} \ln\left(\mathbf{y}^{\mathrm{T}} \mathbf{P}(\boldsymbol{\omega}_k) \mathbf{y} \right) + ck \quad (4)$$

The associated values of c are given in the following table:

Criterion	AIC	BIC	MDL	DMAP
Value of c	3	$\frac{1}{2}\ln(T)$	$\frac{3}{2}\ln(T)$	$\frac{5}{2}\ln(T)$

The model order is then selected by minimizing (4) over $k = 0, \ldots, k_{\max}$ and $\boldsymbol{\omega}_k \in \Omega_k$, $k_{\max}$ being the maximum possible model order.

We show now that obtaining these estimates is equivalent to obtaining MAP estimates for given Bayesian models.

2.3 Bayesian models

We underline here that the following Bayesian models are defined in an *adhoc* way so that the so-built posterior distributions admit marginal and joint MAP estimates matching the ML estimate and classical model order selection criteria.

2.3.1 Fixed model order

The parameters $(\boldsymbol{\omega}_k, \mathbf{a}_k, \sigma_k^2)$ are unknown. To define a Bayesian model, it is necessary to set a prior distribution on $(\boldsymbol{\omega}_k, \mathbf{a}_k, \sigma_k^2)$. We set the following improper prior distribution:

$$p(\boldsymbol{\omega}_k, \mathbf{a}_k, \sigma_k^2) \propto \left| \mathbf{D}^{\mathrm{T}}(\boldsymbol{\omega}_k) \mathbf{D}(\boldsymbol{\omega}_k) \right|^{\frac{1}{2}} \mathbb{I}_{\Omega_k}(\boldsymbol{\omega}_k) / \sigma_k^2 \quad (5)$$

where $\Omega_k \triangleq \left\{ \boldsymbol{\omega}_k \in (0,\pi)^k \right\}$. $\propto$ means "proportional to" and $\mathbb{I}_A(\cdot)$ is the indicator function of the set A. Under such a prior distribution, one obtains

$$p(\boldsymbol{\omega}_k | \mathbf{y}) \propto \left[\mathbf{y}^{\mathrm{T}} \mathbf{P}(\boldsymbol{\omega}_k) \mathbf{y} \right]^{-\frac{T-2k}{2}} \mathbb{I}_{\Omega_k}(\boldsymbol{\omega}_k) \quad (6)$$

Thus under this Bayesian model the marginal MAP estimate of the frequencies $\boldsymbol{\omega}_k$ and the ML estimate are equal.

2.3.2 Unknown model order

The parameters $(k, \boldsymbol{\omega}_k, \mathbf{a}_k, \sigma_k^2)$ are unknown. We set a prior distribution on $(k, \boldsymbol{\omega}_k, \mathbf{a}_k, \sigma_k^2)$. $p(k, \boldsymbol{\omega}_k, \mathbf{a}_k, \sigma_k^2)$ is assumed proportional to

$$\frac{1}{\sigma_k^2 \left| 2\pi\sigma_k^2 \boldsymbol{\Sigma}_k \right|^{1/2}} \exp\left[-\frac{\mathbf{a}_k^{\mathrm{T}} \boldsymbol{\Sigma}_k^{-1} \mathbf{a}_k}{2\sigma_k^2} \right] \frac{\mathbb{I}_{\Omega_k}(\boldsymbol{\omega}_k)}{\pi^k} \quad (7)$$

where $\boldsymbol{\Sigma}_k^{-1} = \delta^{-2} \mathbf{D}^{\mathrm{T}}(\boldsymbol{\omega}_k) \mathbf{D}(\boldsymbol{\omega}_k)$. By integrating out the nuisance parameters $(\mathbf{a}_k, \sigma_k^2)$, we obtain

$$p(k, \boldsymbol{\omega}_k) \propto \left[\mathbf{y}^{\mathrm{T}} \mathbf{P}(\boldsymbol{\omega}_k) \mathbf{y} \right]^{-\frac{T}{2}} \left[\left(\delta^2 + 1 \right) \pi \right]^{-k} \quad (8)$$

If we select δ^2 such that $\left[\left(\delta^2 + 1 \right) \pi \right]^{-k} = \exp(-ck)$ then obtaining the joint MAP estimate $(k, \boldsymbol{\omega}_k)$ of this posterior distribution is equivalent to the minimization of the criterion (4).

We now propose some efficient algorithms to obtain the marginal and joint MAP estimates.

3 STOCHASTIC OPTIMIZATION FOR ML ESTIMATION

3.1 Algorithm

Let $\{\gamma_i; i \in \mathbb{N}\}$ be a given positive increasing sequence of numbers, then our optimization method simulates a non homogeneous Markov chain $\left\{ \boldsymbol{\omega}_k^{(i)}; i \in \mathbb{N} \right\}$ using the following algorithm.

Stochastic optimization for ML estimation

1. Initialisation: randomly set $\boldsymbol{\omega}_k^{(0)}$ on Ω_k and $i = 1$.
2. Iteration i, $i \geq 1$.
   - Sample $\boldsymbol{\omega}_k^*$ according to a reversible uniformally geometrically ergodic kernel $\mathcal{K}\left(\boldsymbol{\omega}_k^{(i-1)}; d\boldsymbol{\omega}_k^* \right)$ admitting $p(\boldsymbol{\omega}_k | \mathbf{y})$ as invariant distribution.
   - Evaluate $\alpha_i\left(\boldsymbol{\omega}_k^{(i-1)}, \boldsymbol{\omega}_k^* \right)$ given by
   $$\min\left\{ 1, \left(\frac{p(\boldsymbol{\omega}_k^* | \mathbf{y})}{p\left(\boldsymbol{\omega}_k^{(i-1)} \middle| \mathbf{y} \right)} \right)^{\gamma_i - 1} \right\} \quad (9)$$
   - Simulate $u \sim \mathcal{U}_{[0,1]}$. If $u < \alpha\left(\boldsymbol{\omega}_k^{(i-1)}, \boldsymbol{\omega}_k^* \right)$ then $\boldsymbol{\omega}_k^{(i)} = \boldsymbol{\omega}_k^*$ otherwise $\boldsymbol{\omega}_k^{(i)} = \boldsymbol{\omega}_k^{(i-1)}$.
3. Set $i \leftarrow i + 1$ and go to step 2.

These different steps are detailed in the following subsection.

3.2 Implementation issues

To implement this algorithm, one must be able to evaluate $\alpha_i\left(\boldsymbol{\omega}_k^{(i-1)}, \boldsymbol{\omega}_k^* \right)$. It can be easily done using (6). It is necessary to define a reversible uniformally geometrically ergodic kernel $\mathcal{K}\left(\boldsymbol{\omega}_k^{(i)}; d\boldsymbol{\omega}_k^* \right)$ admitting $p(\boldsymbol{\omega}_k | \mathbf{y})$ as invariant distribution [9]. There are a huge number of possibilities. We propose to use the transition kernel described in ([3], 2.3.1), that is we sample the frequencies one-at-a-time using a mixture of MH steps of invariant distribution $p\left(\omega_{j,k} | \mathbf{y}, \boldsymbol{\omega}_{-j,k}^* \right)$ [9]. It is easily seen that the so-built transition kernel is uniformally geometrically ergodic and reversible.

3.3 Convergence issues

We have obtained the following convergence result.

Theorem 1 *Under weak assumptions [1], it exists* $C, \alpha > 0$ *such that for* $\gamma_i = C \ln(i + \alpha)$ *then the probability distribution of* $\boldsymbol{\omega}_k^{(i)}$ *denoted* $p^{(i)}(\boldsymbol{\omega}_k)$ *satisfies*

$$\lim_{i \to +\infty} \left\| p^{(i)}(\boldsymbol{\omega}_k) - \overline{p}^{\gamma_i}(\boldsymbol{\omega}_k | \mathbf{y}) \right\|_{TV} = 0$$

($\|\cdot\|_{TV}$ *being the total variation norm [9]) where* $\overline{p}^{\gamma_i}(\boldsymbol{\omega}_k | \mathbf{y}) \propto [p(\boldsymbol{\omega}_k | \mathbf{y})]^{\gamma_i}$ *is a probability distribution that concentrates on the set of global maxima of* $p(\boldsymbol{\omega}_k | \mathbf{y})$ *as* $i \to +\infty$.

4 STOCHASTIC OPTIMIZATION FOR MODEL ORDER SELECTION

4.1 Algorithm

Let $\{\gamma_i; i \in \mathbb{N}\}$ be a given positive increasing sequence of numbers, then our method simulates a non homogeneous Markov chain $\left\{ \left(k^{(i)}, \boldsymbol{\omega}_{k^{(i)}}^{(i)} \right); i \in \mathbb{N} \right\}$ using the following algorithm.

Stochastic Optimization for Model Order Selection

1. Initialisation : set randomly $\left(k^{(0)}, \boldsymbol{\omega}_{k^{(0)}}^{(0)} \right)$ on $\{0, \ldots, k_{\max}\} \times \Omega_{k^{(0)}}$ and $i = 1$.
2. Iteration i
   - Sample k^* and $\boldsymbol{\omega}^*$ using a reversible uniformally geometrically ergodic kernel $\mathcal{K}\left(\boldsymbol{\omega}_k^{(i-1)}, k^{(i-1)}; d\boldsymbol{\omega}_{k^*}^*, k^* \right)$ admitting $p(\boldsymbol{\omega}_k, k | \mathbf{y})$ as invariant distribution.
   - Evaluate $\alpha_i\left(\left(k^{(i-1)}, \boldsymbol{\omega}_{k^{(i-1)}}^{(i-1)} \right), (k^*, \boldsymbol{\omega}_{k^*}^*) \right)$ given by
   $$\min\left\{ 1, \left(\frac{p(\boldsymbol{\omega}_{k^*}^*, k^* | \mathbf{y})}{p\left(\boldsymbol{\omega}_{k^{(i-1)}}^{(i-1)}, k^{(i-1)} \middle| \mathbf{y} \right)} \right)^{\gamma_i - 1} \right\} \quad (10)$$
   - Sample $u \sim \mathcal{U}_{[0,1]}$. If $u < \alpha_i\left(\boldsymbol{\omega}^{(i-1)}, k^{(i-1)}; \boldsymbol{\omega}^*, k^* \right)$ then $\left(k^{(i)}, \boldsymbol{\omega}_{k^{(i)}}^{(i)} \right) = (k^*, \boldsymbol{\omega}_{k^*}^*)$ otherwise $\left(k^{(i)}, \boldsymbol{\omega}_{k^{(i)}}^{(i)} \right) = \left(k^{(i-1)}, \boldsymbol{\omega}_{k^{(i-1)}}^{(i-1)} \right)$.
3. Set $i \leftarrow i + 1$ and go to step 2.

These different steps are detailed in the following subsections.

4.2 Implementation issues

To implement this algorithm, one must be able to evaluate $\alpha_i\left(\left(k^{(i-1)}, \boldsymbol{\omega}_{k^{(i-1)}}^{(i-1)} \right), (k^*, \boldsymbol{\omega}_{k^*}^*) \right)$. It is done using (8). We now propose a reversible uniformally geometrically ergodic kernel $\mathcal{K}\left(\left(k^{(i-1)}, \boldsymbol{\omega}_{k^{(i-1)}}^{(i-1)} \right); k^*, d\boldsymbol{\omega}_{k^*}^* \right)$ admitting $p(\boldsymbol{\omega}_k, k | \mathbf{y})$ as invariant distribution. Note that to be ergodic this transition kernel must involve moves between subspaces of different dimensions as $\dim(\boldsymbol{\omega}_k) \neq \dim(\boldsymbol{\omega}_{k'})$ for $k \neq k'$. This is not possible using classical MCMC methods and we need to use a reversible jump MCMC method, a general methodology that has been recently introduced by Green [4]. We propose here to use a mixture that consists of the transition kernel used previously and a pair of reversible birth/death moves. Assume that there are $k^{(i-1)}$ sinusoids, then with probability $b_{k^{(i-1)}}$ we perform the following move.

Birth move

- Select a new frequency ω' uniformly in $(0, \pi)$ and set $\boldsymbol{\omega}'_{k^{(i-1)}+1} = \left(\boldsymbol{\omega}_{k^{(i-1)}}^{(i-1)}, \omega' \right)$.
- Evaluate $\alpha_{birth} = \min\{r_{birth}, 1\}$ with
$$r_{birth} = \left(\frac{\mathbf{y}^{\mathrm{T}} \mathbf{P}\left(\boldsymbol{\omega}_{k^{(i-1)}}^{(i-1)} \right) \mathbf{y}}{\mathbf{y}^{\mathrm{T}} \mathbf{P}\left(\boldsymbol{\omega}'_{k^{(i-1)}+1} \right) \mathbf{y}} \right)^{\frac{T}{2}} \frac{\left(1 + \delta^2\right)^{-1}}{\left(k^{(i-1)} + 1\right)}$$
- Sample $u \sim \mathcal{U}_{[0,1]}$. If $u \leq \alpha_{birth}$ then $(\boldsymbol{\omega}_{k^*}^*, k^*) = \left(\boldsymbol{\omega}'_{k^{(i-1)}+1}, k^{(i-1)} + 1 \right)$ otherwise $(\boldsymbol{\omega}_{k^*}^*, k^*) = \left(\boldsymbol{\omega}_k^{(i-1)}, k^{(i-1)} \right)$.

Now assume that there are $k^{(i-1)} \geq 1$ sinusoids, then with probability $d_{k^{(i-1)}}$ we perform the following reversible move, the details can be found in [1].

Death move

- Select uniformly among the $k^{(i-1)}$ existing frequencies the one to kill, say ω' and set $\boldsymbol{\omega}'_{k^{(i-1)}-1} = \boldsymbol{\omega}_{k^{(i-1)}}^{(i-1)} \setminus \{\omega'\}$.
- Evaluate $\alpha_{death} = \min\{r_{death}, 1\}$ with
$$r_{death} = \left(\frac{\mathbf{y}^{\mathrm{T}} \mathbf{P}\left(\boldsymbol{\omega}'_{k^{(i-1)}-1} \right) \mathbf{y}}{\mathbf{y}^{\mathrm{T}} \mathbf{P}\left(\boldsymbol{\omega}_{k^{(i-1)}}^{(i-1)} \right) \mathbf{y}} \right)^{\frac{T}{2}} k^{(i-1)} \left(1 + \delta^2\right)$$
- Sample $u \sim \mathcal{U}_{[0,1]}$. If $u \leq \alpha_{death}$ then $(\boldsymbol{\omega}_{k^*}^*, k^*) = \left(\boldsymbol{\omega}'_{k^{(i-1)}-1}, k^{(i-1)} - 1 \right)$ otherwise $(\boldsymbol{\omega}_{k^*}^*, k^*) = \left(\boldsymbol{\omega}_k^{(i-1)}, k^{(i-1)} \right)$.

4.3 Convergence issues

We underline that the proposed algorithm performs global optimization on an union of subspaces of different dimension and naturally focus on those that are likely according to the criterion. We are not aware of similar results in the literature. The following convergence result has been established for this algorithm.

Theorem 2 *Under weak assumptions [1], it exists* $C, \alpha > 0$ *such that for* $\gamma_i = C \ln(i + \alpha)$ *then the probability distribution of* $\left(k^{(i)}, \boldsymbol{\omega}_{k^{(i)}}^{(i)} \right)$ *denoted* $p^{(i)}(\boldsymbol{\omega}_k, k)$ *satisfies*

$$\lim_{i \to +\infty} \left\| p^{(i)}(\boldsymbol{\omega}_k, k) - \overline{p}^{\gamma_i}(\boldsymbol{\omega}_k, k | \mathbf{y}) \right\|_{TV} = 0$$

where $\overline{p}^{\gamma_i}(\boldsymbol{\omega}_k, k | \mathbf{y}) \propto [p(\boldsymbol{\omega}_k, k | \mathbf{y})]^{\gamma_i}$ *is a probability distribution that concentrates on the set of global maxima of* $p(\boldsymbol{\omega}_k, k | \mathbf{y})$ *as* $i \to +\infty$.

5 SIMULATIONS

The following parameters have been selected for the sinusoids: $T = 64$, $k = 3$. We define $E_i \triangleq a_{c_i}^2 + a_{s_i}^2$. $E_1 = E_3 = 20$, $E_2 = 6.32$,$-\arctan(a_{s_1}/a_{c_1}) = 0$, $-\arctan(a_{s_2}/a_{c_2}) = \pi/4$, $-\arctan(a_{s_3}/a_{c_3}) = \pi/3$,$\omega_1/2\pi = 0.2$, $\omega_2/2\pi = 0.2 + 1/T$ and$\omega_3/2\pi = 0.2 + 2/T$. The SNR is defined as $10\log_{10} E_1/(2\sigma^2)$. Theoretically, the algorithms require a so-called logarithmic cooling schedule γ_i and an infinite number of iterations to converge. This sequence goes to $+\infty$ to slowly to be used practically. We run here the algorithms for 5000 iterations and select a linear growing cooling schedule $\gamma_i = A + Bi$ where $\gamma_0 = 1$ and $\gamma_{5000} = 10^2$.

To test the ML procedure, we simulated 100 realisations of a noisy signal with a SNR ranging from $-1dB$ to $3dB$. In Tab. 1, we present the average mean values $m(f_1)$ and standard deviations $\sigma(f_i)$ of the obtained estimates.

SNR	$m(f_1)/\sigma(f_1)$	$m(f_2)/\sigma(f_2)$	$m(f_3)/\sigma(f_3)$
-1	0.193/0.027	0.216/0.009	0.247/0.046
0	0.194/0.023	0.215/0.010	0.243/0.038
1	0.195/0.020	0.214/0.006	0.236/0.022
2	0.199/0.002	0.216/0.005	0.235/0.021
3	0.200/0.001	0.215/0.003	0.233/0.011

Table 1

In all simulations, we found that the likelihood value of the true parameters is weaker than the one of the obtained estimate. If it does not prove that our algorithm converges towards a global maximum, it shows that it is very efficient in practice.

To test the model order selection procedure, we simulated 100 realisations of a noisy signal with a SNR ranging from $0dB$ to $3dB$. The results are displayed in Tab. 2.

Criterion	*SNR*	$k \leq 1$	$k = 2$	$k = 3$	$k \geq 4$
MDL	$0db$	0	39	43	18
	$1db$	0	24	53	23
	$2db$	0	13	59	28
	$3db$	0	6	63	31
DMAP	$0db$	1	92	7	0
	$1db$	0	72	28	0
	$2db$	0	80	20	0
	$3db$	0	51	49	0

Table 2

The results for other criteria are presented in [1].

We underline that in both cases these algorithms were intialized randomly. Contrary to classical deterministic methods, they do not appear to be sensitive to initialization and are very reliable, the cost to pay is that they are much more computationally demanding.

6 CONCLUSION

In this paper, efficient stochastic optimization algorithms have been proposed to perform ML estimation and model order selection for harmonic signals. These algorithms are based on non homogeneous versions of Markov chain Monte Carlo methods. Sufficient conditions to ensure convergence of these algorithms towards the set of global maxima have been established. Their practical efficiency has been demonstrated via computer simulations. These algorithms outperform classical methods as they are not sensitive to initialization and give reliable estimates even in difficult cases where the other methods fail.

References

[1] C. Andrieu and A. Doucet, "Stochastic algorithms for optimisation of harmonic signals," technical report, Univ. Cambridge, Dept. Eng., 1998. Available on request.

[2] P.M. Djurić, "A model selection rule for sinusoids in white Gaussian noise," *IEEE Trans. Sig. Proc.*, vol. 44, no. 7, pp. 1744-1751, 1996.

[3] A. Doucet and C. Andrieu, "Robust Bayesian spectral analysis using MCMC," in *Proc. EUSIPCO'98*, Island of Rhodes.

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

[5] 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.

[6] K. Sharman, "Maximum likelihood estimation by simulated annealing," in *Proc. IEEE ICASSP*, pp. 2241-2244, 1988.

[7] K. Sharman and G. McClurkin, "Genetic algorithms for maximum likelihood parameter estimation," in *Proc. IEEE ICASSP*, pp. 2716-2719, 1989.

[8] P. Stoica, R.L. Moses, B. Friedlander and T. Soderstrom, "Maximum likelihood estimation of the parameters of multiple sinusoids from noisy measurements," *IEEE Trans. Acou. Speech Sig. Proc.*, vol. 37, pp. 378-392, 1989.

[9] L. Tierney, "Markov chains for exploring posterior distributions," *Ann. Stat.*, pp. 1701-1728, 1994.

[10] I. Ziskind and M. Wax, "Maximum likelihood localisation of multiple sources by alternating projection," *IEEE Trans. Acou. Speech Sig. Proc.*, vol. 36, pp. 1553-1560, 1988.