BAYESIAN ESTIMATION OF THE PARAMETERS OF A POLYNOMIAL PHASE SIGNAL USING MCMC METHODS

Céline Theys, Michelle Vieira and André Ferrari

I3S, CNRS/UNSA, 41 Bd. Napoléon III, 06041 Nice, France
theys@unice.fr - vieira@unice.fr - ferrari@unice.fr

ABSTRACT

The aim of this paper is Bayesian estimation of the parameters of a polynomial phase signal. This problem, encountered in radar systems for example, is usually solved using a time-frequency analysis or phase-only algorithms, see [4] for a detailed introduction. A Bayesian approach using Markov chain Monte Carlo (MCMC) methods for estimating a posteriori densities of the polynomial parameters is proposed. This approach has the main advantage : it gives a direct estimation of all polynomial coefficients, contrary to recently developed algorithms, [4], [2].

1. PROBLEM STATEMENT

The signal under study is a noisy polynomial phase signal s_n as:

$$s_n = A\exp(j\sum_{i=0}^{M} a_i n^i) + e_n \qquad (1)$$

where e_n is a complex gaussian i.i.d. noise of known variance σ^2 and zero mean.

The aim of this paper is to estimate the parameter $\mathbf{a} = \{a_0, a_1, \ldots, a_M\}$. The Bayesian solution is to evaluate the joint posterior probability density of $\mathbf{a}$ conditional on the data $\mathbf{s} = \{s_0, s_1, \ldots, s_{N-1}\}$ and the prior information I, abbreviated as $p(\mathbf{a}|\mathbf{s}, I)$.

The knowledge of this probability density will then allow us to sample the marginal posterior probability density of a_i, $p(a_i|\mathbf{s}, \mathbf{a} - \{a_i\}, I)$ (We note $\mathbf{a} - \{a_i\}$ the set $\mathbf{a}$ without the element a_i). We construct chain of samples whose the untractable target density is $p(a_i|\mathbf{s}, \mathbf{a} - \{a_i\}, I)$ using a stochastic algorithm, a Markov Chain Monte Carlo (MCMC).

Finally, a Minimum Mean Square Error (MMSE) Bayesian estimator is applied (see [3]).

2. DETERMINATION OF P(a|s, I)

By applying Bayes' theorem, the joint posterior probability density of all of parameters, $p(A, \mathbf{a}|\mathbf{s}, I)$, is :

$$p(A, \mathbf{a}|\mathbf{s}, I) = \frac{p(A, \mathbf{a}|I)p(\mathbf{s}|A, \mathbf{a}, I)}{p(\mathbf{s}|I)} \qquad (2)$$

Three probabilities are required :

- A priori probability density of the parameters given a priori information I, $p(A, \mathbf{a}|I)$.
- Direct probability density or likehood function, $p(\mathbf{s}|A, \mathbf{a}, I)$.
- Probability of the data given I, a normalization constant here, $p(\mathbf{s}|I)$.

Then,

$$p(A, \mathbf{a}|\mathbf{s}, I) \propto p(A, \mathbf{a}|I)p(\mathbf{s}|A, \mathbf{a}, I) \qquad (3)$$

Amplitude A of the signal is referred to as a *nuisance parameter*. To remove this parameter, the joint probability density is integrated on all possible values of A :

$$p(\mathbf{a}|\mathbf{s}, I) = \int_{-\infty}^{+\infty} p(A, \mathbf{a}|\mathbf{s}, I)dA \qquad (4)$$

Calculation of the $p(\mathbf{a}|\mathbf{s}, I)$ is at least a three step problem : assign the direct then the prior probability density and finally remove the dependence on the amplitude by integration.

2.1. Assignment of direct probability

From the knowledge of noise probability, likehood function is :

$$p(\mathbf{s}|\mathbf{a}, A, I) = \prod_{i=0}^{N-1} p(e_i)$$

$$= \frac{1}{\pi^N \sigma^{2N}} \exp[-\frac{1}{\sigma^2}\sum_{k=0}^{N-1} |s_k - A\exp j\sum_{i=0}^{M} a_i k^i|^2] \qquad (5)$$

It is worth noticing that the likehood function depends on σ^2, but as the noise variance is assumed to be known, we do not make it appear.

2.2. Assignment of prior probability

Non-informative prior probability densities are chosen to express ignorance about the value of the parameter vector in absence of data.

All parameters are assumed to be uniformly distributed, $p(a_0|I) = k_1, \ldots, p(a_M|I) = k_M$ and $p(A|I) = k_{M+1}$ where k_i is constant, $i = 1 \cdots M+1$.

Eq. (3) becomes :

$$p(A, \mathbf{a}|\mathbf{s}, I) \propto p(\mathbf{s}|A, \mathbf{a}, I) \tag{6}$$

2.3. Elimination of nuisance parameter

Once all terms in the posterior probability density function have been assigned, amplitude of the signal is removed by integration. Using the following standard identity :

$$\int_{-\infty}^{+\infty} \exp(-ax^2 - bx - c)dx = \sqrt{\frac{\pi}{a}} \exp(\frac{b^2}{4c} - c) \tag{7}$$

Eq. (4) becomes :

$$p(\mathbf{a}|\mathbf{s}, I) \propto \exp\left[-\frac{1}{\sigma^2}\left(\sum_{k=0}^{N-1}|s_k|^2 - \frac{1}{N}\times\right.\right.$$
$$\left.\left.\left[\sum_{k=0}^{N-1}\Re(s_k)\cos(\sum_{i=0}^{M}a_i k^i) + \Im(s_k)\sin(\sum_{i=0}^{M}a_i k^i)\right]^2\right)\right] \tag{8}$$

where $\Re(s_k)$ and $\Im(s_k)$ are respectively real and imaginary part of the data.

3. MCMC METHODS

The objective is to produce samples of marginal posterior probability density of each parameter a_i, $p(a_i|\mathbf{a} - \{a_i\}, \mathbf{s}, \sigma, I)$ from the joint posterior probability density (8).

$$p(a_i|\mathbf{a} - \{a_i\}, \mathbf{s}, I) = \frac{p(\mathbf{a}|\mathbf{s}, I)}{p(\mathbf{a} - \{a_i\}|\mathbf{s}, I)} \tag{9}$$

The problem is that $p(\mathbf{a} - \{a_i\}|\mathbf{s}, I)$ cannot be obtained by integration over a_i and conventional numerical integration methods are not appropriate. Then, we turn to MCMC methods, straightforward to implement.

3.1. Random Walk Metropolis-Hasting method

In recent years, an increasingly amount of attention has been devoted to MCMC methods, because of the large spread of applications. The usual approach to Markov chain theory on a continuous state space is to start with a transition kernel $P(x, A)$ for $x \in \mathcal{R}^d$ and $A \in \mathcal{B}$, where $\mathcal{B}$ is the Borel σ-field on $\mathcal{R}^d$. The transition kernel is a conditional distribution function that represents the probability of moving from x to a point in the set A, $\mathcal{P}(x, \mathcal{R}^d) = 1$ and $\mathcal{P}(x, \{x\})$ is not necessarily 0. Under certain conditions, it can be shown that the n^{th} iterate converges to the invariant distribution.

General idea of the MCMC methods is : the invariant density is known (perhaps up to a constant multiple), it is $p(.)$, the target density for which samples are desired , but the transition kernel is unknown. To generate samples from $p(.)$, the methods find and utilize a transition kernel $P(x, dy)$ whose n^{th} iterate converge to $p(.)$ for large n. The process is started at an arbitrary x and iterated a large number of times. One of the usefulness MCMC method is the Metropolis-Hastings (M-H) algorithm, the transition kernel for the M-H chain is defined by:

$$P_{MH}(x, dy) = q(x, y)\alpha(x, y)dy + \left[1 - \int_{\mathcal{R}} q(x, y)\alpha(x, y)dy\right]\delta_x(dy), \tag{10}$$

where $\alpha(x, y)$ is the probability of move from x to y and is given by:

$$\alpha(x, y) = \begin{cases} \min[\frac{p(y)q(y,x)}{p(x)q(x,y)}, 1] & \text{if } p(x)q(x, y) > 0 \\ 1 & \text{otherwise} \end{cases} \tag{11}$$

Expression of $\alpha(x, y)$ ensures reversibility of P_{MH}. The second term of (10) denotes the possibly non-zero probability that the process remains at x. $q(x, y)$ is the candidate generating density and is usually selected from a family of distributions that requires the specification of such tuning as the location and scale.

An important family of candidate generating densities is given by $q(x, y) = q_1(x - y)$ where $q_1(.)$ is a multivariate density. The candidate y is thus drawn according to the process $y = x + z$, where z is called the increment random variable and follows q_1. This *random walk M-H* algorithm is relevant in our case since it does not require the precise location of the target density, eq. (8). See [1] for a simple exposition of the M-H algorithm.

As we deal with joint densities of size set proportional to the polynomial order, M, a one "variable-at-a-time" algorithm combining $M+1$ updates is proposed. The practical significance of this principle is important since it allows us to take draws in succession from each of the kernels, instead of having to run each of the kernels to convergence for every value of the conditioning variable and it is usually far easier to find several conditional kernels that converge to their respective conditional densities than to find one kernel that converges to the joint.

Suppose that there exists a conditional transition kernel $P_i(a_i, dy_i|\mathbf{a}-\{a_i\})$ with the property that, for a fixed value of $\mathbf{a}-\{a_i\}$, $p^*(.|\mathbf{a}-\{a_i\})$ is its invariant distribution with density $p(.|\mathbf{a}-\{a_i\})$, i.e.:

$$p^*(dy_i|\mathbf{a}-\{a_i\}) = \int P_i(a_i, dy_i|\mathbf{a}-\{a_i\})p(.|\mathbf{a}-\{a_i\})da_i. \quad (12)$$

If we suppose that $P_0(a_0, dy_0|\mathbf{a}-\{a_0\})$ produces y_0 given $\mathbf{a}-\{a_0\}$ and $P_i(a_i, dy_i|y_{0\to i-1}, \mathbf{a}-\{a_0,\ldots,a_{i-1}\})$ produces y_i given $y_0,\ldots,y_{i-1}, \mathbf{a}-\{a_0,\ldots,a_{i-1}\}$ then the kernel formed by multiplying the conditional kernels,

$$\int_{M+1}\cdots\int P_0(a_0, dy_0|\mathbf{a}-\{a_0\})\cdots \times P_M(a_M, dy_M|y_{0\to M-1})p(\mathbf{a})da_0\cdots da_M \quad (13)$$

has $p^*(dy_0,\cdots,dy_M)$ as its invariant distribution.

Proof :

$$\int_{M+1}\cdots\int P_0(a_0, dy_0|\mathbf{a}-\{a_0\})\cdots P_M(a_M, dy_M|y_{0\to M-1})$$
$$\times\, p(\mathbf{a})da_0\cdots da_M$$
$$= \int_{M}\cdots\int P_1(a_1, dy_1|\mathbf{a}-\{a_0,a_1\})\cdots P_M(a_M, dy_M|y_{0\to M-1})$$
$$\times\, p(\mathbf{a}-\{a_0\})da_1\cdots da_M$$
$$\times \int P_0(a_0, dy_0|\mathbf{a}-\{a_0\})p(a_0|\mathbf{a}-\{a_0\})da_0$$
$$\text{as } p^*(y_0|\mathbf{a}-\{a_0\}) = \frac{p(\mathbf{a}-\{a_0\}|y_0)p^*(y_0)}{p(\mathbf{a}-\{a_0\})}$$
$$= p^*(y_0)\int_{M}\cdots\int P_1(a_1, dy_1|\mathbf{a}-\{a_0,a_1\})\cdots$$
$$\times\, P_M(a_M, dy_M|y_{0\to M-1})p((\mathbf{a}-\{a_0\}|y_0)da_1\cdots da_M$$
$$= p^*(y_0)p^*(y_1|y_0)\int_{M-1}\cdots\int P_2(a_2, dy_2|\mathbf{a}-\{a_0,a_1,a_2\})\cdots$$
$$\times\, P_M(a_M, dy_M|y_{0\to M-1})p(\mathbf{a}-\{a_0,a_1\}|y_0,y_1)da_2\cdots da_M$$
$$\vdots$$
$$= p^*(y_0)p^*(y_1|y_0)\ldots p^*(y_M|y_{0\to M-1})$$
$$= p^*(y_0,\ldots,y_M) \quad \square$$

3.2. Implementation

The one "variable-at-a-time" *random walk* M-H method has been implemented following algorithm :

find the initial values: $\mathbf{a}^{(0)}$
for $j = 0...n-1$ **do:**
  for $k = 0...M$
    generate y_k **from** $q_k(y_k - a_k^{(j)})$
    generate u_k **from** $\mathcal{U}(1,0)$

$$\alpha_k(a_k^{(j)}, y_k) = \min\{1, \frac{p(y_k|a_0^{(j+1)},\ldots,a_{k-1}^{(j+1)},a_{k+1}^{(j)},\ldots,a_M^{(j)})}{p(a_k^{(j)}|a_0^{(j+1)},\ldots,a_{k-1}^{(j+1)},a_{k+1}^{(j)},\ldots,a_M^{(j)})}\}$$

    if $u_k \le \alpha_k(a_k^{(j)}, y_k)$ **then**
      $a_k^{(j+1)} = y_k$
    else
      $a_k^{(j+1)} = a_k^{(j)}$
    endif
  end
end

where $q_k(y_k - a_k^{(j)})$ is the candidate-generating density of the k^{th} phase parameter and has been chosen as an independent uniform univariate distribution on the interval $[-\delta_k, \delta_k]$.

Implementation gives rise to important remarks. The initial values of the Markov chain must be chosen in such a way that the probability of move $\alpha(\mathbf{a}, y)$ should be defined, i.e., $p(\mathbf{a}^{(0)}) > 0$ leading to $p(\mathbf{a}^{(j)}) > 0, \forall j \in [0, ..., n-1]$

The choice of δ_k affects the behavior of chain in at least two dimensions : the acceptance rate (i.e. the percentage of times a move to a new point is made) and the region of the sample space that is covered by a chain. If δ_k is too large, the acceptance rate is low and if δ_k is too small, large number of iterations must be necessary to traverse the support of the k^{th} a posteriori density. For the *random walk* M-H algorithm, a high acceptance rate can signify that the random walk moves too slowly on the considered support. In the contrary, if the acceptance rate is small, the random walk explores quickly the considered support.

Gelman and al. (1994) enjoin an acceptance rate equal to 0.5 for models of dimension 1 and 2, the δ_k can be deduced experimentally.

4. SIMULATION RESULTS

In order to estimate performances of the proposed method, computer simulations using polynomial phase signal of N=100 samples have been drawn for M=2 and M=3.

With an acceptance rate around 0.5, $\delta_0 = 0.025$, $\delta_1 = 5.10^{-4}$, $\delta_2 = 5.10^{-6}$ for the 2^{nd} order signal and $\delta_0 = 0.025$, $\delta_1 = 25.10^{-5}$, $\delta_2 = 5.10^{-6}$, $\delta_3 = 10^{-7}$ for the third order one have been found.

In sampling process, the first 1000 draws have been ignored and we collect the next 4000 (n=5000) to approximate the posterior distribution of $p(a_i|\mathbf{a}-\{a_i\}, \mathbf{s}, \sigma, I)$. The posterior mean, standard deviation are reported for the second case in the array below :

parameter	mean	standard deviation
a_0 = -0.01	6.19e-3	2.38e-2
a_1 = 0.04	3.7e-2	2.05e-3
a_2 = 0.0005	5.8e-4	5.37e-4
a_3 = -0.000005	-5.51e-6	3.7e-7

According to [5] and supported by obtained results, relative error and support of posterior densities are always the largest for $i = 0$, and decrease monotonically with i. For illustrating this remark, a posteriori distribution histograms for the parameters of polynomial phase signal for M=2 are given (see figure 1, 2, 3) for $a_0 = -0.01$, $a_1 = -0.03$ and $a_2 = 0.0007$.

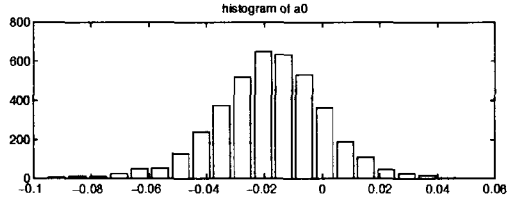

figure 1 : Estimation of $p(a_0|\mathbf{a} - \{a_0\}, \mathbf{s}, I)$

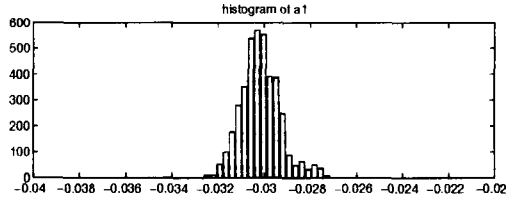

figure 2 : Estimation of $p(a_1|\mathbf{a} - \{a_1\}, \mathbf{s}, I)$

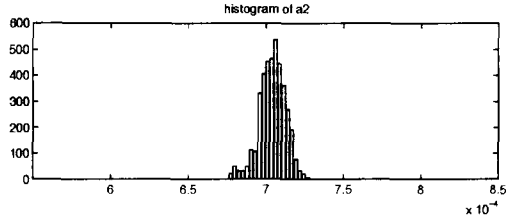

figure 3 : Estimation of $p(a_2|\mathbf{a} - \{a_2\}, \mathbf{s}, I)$

In general, according to [3], Bayesian approach can be applied only when parameters is random but, in practice, it is often used for deterministic parameter estimation, since non-informative prior probability density does add any information to the problem (see 2.2). Mean of the posterior density gives an optimal Bayesian estimator, called MMSE estimator, which minimize the Mean Square Error (MSE) for random parameter. We define, thus, $\hat{a}_i = \frac{1}{4000} \sum_{j=1000}^{5000} a_i^{(j)}$.

In order to evaluate the performance of the proposed approach, computer simulation using the second order signal are drawn. The MSE of $\hat{a}_2$, $\hat{a}_1$ and $\hat{a}_0$ have been estimated on 100 realizations and different SNR (see figure 4), and the Cramer-Rao lower bound (CRLB), derived in [5], are given.

It can be noticed that, from a SNR of 5 dB, performances are very close to the CRLB.

5. CONCLUSION

A Bayesian estimation of the parameters of noisy polynomial phase signal has been proposed. This approach gives at least three main advantages : it requires only the a priori density form of the process allowing others noise probability densities than the Gaussian one, it works directly on the noisy samples, contrary to phase-only algorithms and it gives a whole estimation of the polynomial coefficients.

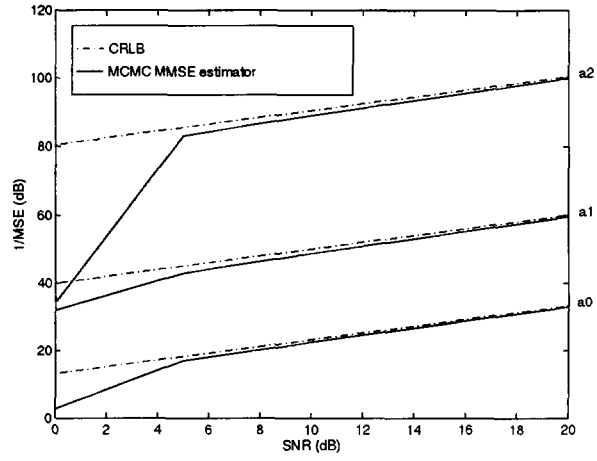

figure 4 : 1/MSE for $\hat{a}_0$, $\hat{a}_1$ et $\hat{a}_2$

6. REFERENCES

[1] S. Chib and E. Greenberg. Understanding the Metropolis-Hastings algorithm. *The American Statistician*, 49:327–335, 1995.

[2] A. Ferrari, C. Theys, and G. Alengrin. Polynomial Phase Signal Analysis using Stationary Moments. *EURASIP Signal Processing*, 1995. To be published.

[3] S. M. Kay. *Fundamentals of Statistical Signal Processing : Estimation Theory.* Signal Processing Series. Prentice Hall, 1993.

[4] S. Peleg and B. Friedlander. The discrete polynomial-phase transform. *IEEE Transactions on Signal Processing*, 43(8):1901–1912, August 1995.

[5] S. Peleg and B. Porat. The Cramer-Rao Lower Bound for Signals with Constant Amplitude and Polynomial Phase. *IEEE Transactions on Signal Processing*, 39(3):749–752, March 1991.