



HAL
open science

Estimation of the Amplitude and the Frequency of Nonstationary Short-time Signals

Meryem Jabloun, Nadine Martin, François Léonard, Michelle Vieira

► **To cite this version:**

Meryem Jabloun, Nadine Martin, François Léonard, Michelle Vieira. Estimation of the Amplitude and the Frequency of Nonstationary Short-time Signals. *Signal Processing*, 2008, 88 (7), pp.1636-1655. hal-00200115

HAL Id: hal-00200115

<https://hal.science/hal-00200115>

Submitted on 20 Dec 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Estimation of the Instantaneous Amplitude and Frequency of Non-stationary Short-time Signals

Meryem JABLOUN, Nadine MARTIN, Francois LEONARD* and Michelle VIEIRA

LIS-CNRS/INPG, 961 rue de la houille blanche, BP 46, 38402 Saint Martin

*d'Heres, France. firstname.lastname@lis.inpg.fr. *Institut de recherche*

d'Hydro-Quebec, 1800 Bd. Lionel Boulet, Varennes (QC) Canada, J3X 1S1.

leonard.francois@ireq.ca

Abstract

We consider the modeling of non-stationary discrete signals whose amplitude and frequency are assumed to be nonlinearly modulated over very short-time duration. We investigate the case where both instantaneous amplitude and frequency can be approximated by orthonormal polynomials.

Previous works dealing with polynomial approximations refer to orthonormal bases built from a discretization of continuous-time orthonormal polynomials. As this leads to a loss of the orthonormal property, we propose to use discrete orthonormal polynomial bases: the discrete orthonormal Legendre polynomials and a discrete base we have derived using Gram-Schmidt procedure. We show that in the context of short-time signals the use of these discrete bases leads to a significant improvement in the estimation accuracy.

We manage the model parameter estimation by applying two approaches. The first is maximization of the likelihood function. This function being highly nonlinear,

we propose to apply a stochastic optimization technique based on the simulated annealing algorithm. The problem can also be considered as a Bayesian estimation which leads us to apply another stochastic technique based on Monte Carlo Markov Chains. We propose to use a Metropolis Hastings algorithm. Both approaches need an algorithm parameter tuning that we discuss according our application context. Monte Carlo simulations show that the results obtained are close to the Cramer-Rao bounds we have derived. We show that the first approach is less biased than the second one. We also compared our results with the Higher Ambiguity Function-based method. The methods proposed outperform this method at low signal to noise ratios in terms of estimation accuracy and robustness. Both proposed approaches are of a great utility when scenarios in which signals having a small sample size are non-stationary at low signal to noise ratios. They provide accurate system descriptions which are achieved with only a reduced number of basis functions.

Key words: Polynomial phase signal, time-frequency analysis, parameter estimation, simulated annealing, Metropolis Hastings algorithm.

1 Introduction

Estimation of the amplitude and frequency of non-stationary signals is a fundamental task, which has applications in a wide range of fields such as speech, music, seismic, radar, sonar, biomedicine and mechanics [1–3]. We are particularly interested in the signals where the frequency and amplitude modulations are fast. Real examples of such signals can be of a long time duration [4] or of a short-time duration (few samples) [5,6]. In [5,6], the cumulative annual mass balance data for several glaciers in the Alps over fifty years (from 1949 to 2000) are presented. Their fluctuations are considered as an evidence of the climate change in the Alps. These data are non-stationary signals with

non-linear instantaneous amplitude and frequency and only fifties samples are available.

The instantaneous amplitude and frequency of the signals are frequently presented as time-varying functions [7–15] where polynomial function models have been assigned to the signal phase. In [12], Francos and Porat’s algorithm combined a time-frequency distribution of a minimum-cross-entropy with the Higher Ambiguity Function (HAF) [7] to achieve the model parameter estimation. The HAF technique, which is a suboptimal compared to the maximum likelihood procedure, transforms a signal of a M^{th} - order polynomial phase written as

$$\Phi[n] = \sum_{m=0}^M \varphi_m n^m \quad (1)$$

into a single harmonic at a frequency proportional to the M^{th} order coefficient φ_M and to a constant lag τ . This comes from the fact that $(\Phi[n] - \Phi[n - \tau])$ is a $(M - 1)^{th}$ order polynomial. When an estimation $\hat{\varphi}_M$ is computed, the HAF technique is repeated for the signal multiplied by $e^{-j\hat{\varphi}_M n^M}$ which becomes a signal of a $(M - 1)^{th}$ - order polynomial phase, its higher-order coefficient is φ_{M-1} . This step is restarted at lower orders.

Despite its low computational complexity, the robustness of the HAF technique is considerably limited by the lag choice and the propagation error phenomenon [15] which occurs in the presence of a higher-order polynomial modeling and/or a strong noise. Consequently, the Francos and Porat’s algorithm [12] fails at low Signal to Noise Ratios (SNRs) and high model orders. Moreover the resolution of the time-frequency distribution worsens and makes this algorithm unusable in the case of very short-time signals.

In this paper, we propose to process very-short-time signals with nonlinear amplitude and frequency modulations. Examples of real cases of short-time

signals are presented in [5,6]. Since in this context the data samples are few (less than 65 samples), each of the following assumptions is very important for decreasing the estimation variance. First, the model origin and the initial phase are referenced at the center of the time window. Then, contrary to [7–15], the instantaneous frequency instead of the phase is approximated by a discrete polynomial function. We compare different polynomial bases and we show that a discrete orthonormal base induces a significant improvement in the parameter estimates since it achieves amplitude parameter decoupling.

We then discuss and compare two methods for estimating the model parameters. The first is based on maximization of the likelihood function in order to preserve optimality. As this function is highly nonlinear, we use a stochastic optimization technique based on Simulated Annealing (SA) algorithm. In the second approach, we extend a Bayesian method [16] to short-time signals, the amplitude and frequency of which are modeled by nonlinear polynomials. A posterior probability distribution on the model parameters is defined by assuming them to be random variables with uninformative prior distributions. To overcome the common problem due to the evaluation of high dimensional integrals, we use a Markov Chain Monte Carlo method with a Metropolis Hastings algorithm (MH-MCMC). Both proposed approaches are of a great utility in the context of short-time non-stationary signals at low SNRs [5,6]. They provide accurate system descriptions which are achieved with only a reduced number of basis functions.

The approaches proposed offer great interest when addressing scenarios in which long-time signals are highly non-stationary at low SNRs. The basic idea consists in segmenting a long signal on short-time windows, each window being approximated with a low-order polynomial model of the instantaneous amplitude and frequency. The estimation of the local-model parameters is then

achieved using the approaches described in the present paper, for example by maximizing the likelihood function using the simulated annealing technique. We already proposed segmentation and merging procedures of several local models in [4,17,18], so higher orders are not necessary for modeling entire signal modulations. This method seems to be quite flexible for modeling a wide range of modulations, whatever the approximation order and the data size. This method is developed in [4] and applied to a canary song.

In the present paper, we focus on short-time signals, the model of which is given in Section 2. Section 3 describes the two proposed approaches to estimating the model parameters. In Section 4, we give the appropriate Cramer-Rao Bounds (CRB) and we discuss the discrete polynomial base choices in Section 5. Section 6 is devoted to Monte Carlo simulations for assessing the algorithm performance in terms of Minimum Squares Errors (MSE), a comparison with the HAF-based method is also given. Lastly, the conclusion and work progress are given in Section 7.

2 Polynomial amplitude and frequency Model

The discrete signal to be dealt with is a single component having a short-time duration and is given by

$$y[n] = s[n] + e[n] \quad \text{for } \frac{-N}{2} \leq n \leq \frac{N}{2}, \quad (2)$$

$$s[n] = A[n]e^{j\Phi[n]}. \quad (3)$$

where $y[n]$ represents the noisy signal, $s[n]$ is the deterministic signal and $e[n]$ is a white complex Gaussian noise with zero mean and unknown variance σ_e^2 . $A[n]$ and $\Phi[n]$ are the instantaneous amplitude (IA) and the instantaneous

phase respectively. j is the complex number verifying $j^2 = -1$. The time reference is set to the center of the window and the sample size $(N + 1)$ is low, about 30. In the following, N is assumed to be even to simplify the algorithm implementation.

The signal phase $\Phi[n]$ is obtained by numerical integration of Instantaneous Frequency (IF) $F[n]$

$$\Phi[n] = \varphi_0 + 2\pi \left(\sum_{k=-\frac{N}{2}}^n F[k] - \sum_{k=-\frac{N}{2}}^0 F[k] \right). \quad (4)$$

The initial phase $\varphi_0 = \Phi[0]$ is referenced to the time-window center to minimize the estimation error [19]; φ_0 is not set to zero contrary to all literature examples [7–15]. We assume $\varphi_0 \in] - \pi, \pi]$.

In order to avoid the ambiguity definition [20] on the signal model of (3), we assume a non-discontinuous phase and a positive-definite IA

$$\text{for } n = -\frac{N}{2}, \dots, \frac{N}{2}, \quad A[n] > 0. \quad (5)$$

The IF verifies

$$\text{for } n = -\frac{N}{2}, \dots, \frac{N}{2}, \quad 0 < F[n] < \frac{Fs}{2}, \quad (6)$$

where Fs is the sampling frequency.

A positive definite amplitude resolves the ambiguity on the phase definition. This ambiguity occurs if the amplitude is zero at certain instant of time. For those time instants, the phase may incur a discontinuity, and induces an instantaneous frequency definition with infinite jumps. The non-discontinuity of the phase ensures an instantaneous frequency definition without infinite jumps.

In [7–15] the polynomial phase signal models are motivated by Weierstrass'

theorem. The amplitude $A[n]$ is constant in [7,10,15], random in [13] and linear time functions in [8,11] and the IF estimate is obtained by derivation of a phase polynomial model in a continuous time. Consequently, an error on the m^{th} phase coefficient produces an m -fold error on the $(m - 1)^{th}$ IF coefficient and hence the IF estimate can be drastically affected in the context of very short-time signals. Therefore we choose to model IF instead of the phase and we approximate both the IA and IF by polynomial functions as follows

$$A[n] = \sum_{m=0}^p a_m g_m[n], \quad (7)$$

$$F[n] = \sum_{m=0}^q f_m g_m[n].$$

p and q are the approximation orders of IA and IF respectively and $g_m[n]$ is a polynomial of order m . a_m and f_m are the amplitude and frequency parameters respectively. The choice of the polynomial base $(g_m[n])_{m=0, \dots, \max(p,q)}$ is discussed in Section 5.

To sum up, a vector of $p + q + 3$ parameters needs to be estimated

$$\begin{aligned} \boldsymbol{\theta} &= [a_0, \dots, a_p, \varphi_0, f_0, \dots, f_q]^T \\ &= [\theta_0, \theta_1, \dots, \theta_{p+q+2}]^T, \end{aligned} \quad (8)$$

where T denotes the transpose symbol. In the particular case of short-time signals, we assume p and q are limited to low orders

$$0 \leq (p, q) \leq 3. \quad (9)$$

Therefore, 9 parameters at most are to be estimated. The signal length must be larger than $p + q + 3$ and should be about at least one or two periods of the signal to ensure the uniqueness of the solution of the polynomial model. In the next section, the approaches that we propose to estimate (8) are discussed.

3 Proposed Estimation Approaches

In order to estimate the model parameter vector $\boldsymbol{\theta}$ (8) of a short-time signal, we propose two different methods. The first one is a deterministic approach, the likelihood function is maximized by means of a stochastic technique based on the SA algorithm. In the second method a Bayesian approach is adopted and a MH-MCMC algorithm, random walk with a one-variable-at-a-time technique, is used to perform the Bayesian computation. These two approaches will be compared and the best one in terms of variance and bias will be employed in future work for estimating long-time signals by merging different short-time segments [4,18] and for estimating multicomponent short-time signals [21].

3.1 Maximum Likelihood Approach

3.1.1 Maximum likelihood estimation

In the case of very short-time signals, the number of model parameters is assumed low (9) and a maximum-likelihood estimation can therefore be performed in a reasonable computing time thanks to substantial advances in numerical analysis. In the presence of Gaussian noise ($e[n]$ (2)), the maximum likelihood procedure is equivalent to minimization of the Least Squares (LS) function $\ell_{LS}(\boldsymbol{\theta})$ given by

$$\ell_{LS}(\boldsymbol{\theta}) = \sum_{n=-\frac{N}{2}}^{\frac{N}{2}} |y[n] - s[n]|^2, \quad (10)$$

where the signal model $s[n]$ is evaluated for each $\boldsymbol{\theta}$ by substituting (4) and (7) into (3). This produces a multidimensional and nonlinear functional with

respect to $\boldsymbol{\theta}$ (8) to be minimized in \mathbb{R}^{p+q+3}

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta} \in \mathbb{R}^{p+q+3}} \ell_{LS}(\boldsymbol{\theta}). \quad (11)$$

Parameter estimation by direct minimization of $\ell_{LS}(\boldsymbol{\theta})$ is extremely difficult and classical optimization techniques such as gradient descent, Gauss-Newton and EM algorithm do not ensure convergence to the global minimum in the presence of many local minima. To overcome this problem, there is a variety of meta-heuristic approaches for escaping local minima. We propose to use the SA algorithm since it is adapted to continuous optimization problems ($\boldsymbol{\theta} \in \mathbb{R}^{p+q+3}$). The proposed SA algorithm is detailed in the next paragraph.

3.1.2 Simulated-Annealing-based algorithm

The SA algorithm based on Metropolis algorithm was initially proposed by Kirckpatrick et al. [22]. It is suitable for large-scale optimization problems, especially when a desired global *extremum* is hidden among many local extrema [23]. It has an analogy with thermodynamics in the way that metal cools and anneals [24,25]. The main steps of the SA algorithm include the initialization of the model parameters denoted $\boldsymbol{\theta}^{(0)}$, an iterative loop controlled by a temperature \mathcal{T} in order to generate new candidates that minimize $\ell_{LS}(\boldsymbol{\theta})$ and finally the verification of a stop criterion of the algorithm run. In this paragraph we discuss about the tuning of these steps according to our context.

Generally, initialization $\boldsymbol{\theta}^{(0)} = [a_0^{(0)}, \dots, a_p^{(0)}, \varphi_0^{(0)}, f_0^{(0)}, \dots, f_q^{(0)}]^T$ can be randomly chosen. However, good starting values of the parameters are helpful in

reducing "wanders" among local minima. Therefore, we initialize the parameter vector using the Fourier Transform of the discrete noisy signal $y[n]$, given by $\mathcal{F}_y(f) = \sum_{n=-\frac{N}{2}}^{\frac{N}{2}} y[n] e^{-2\pi jfn}$. The expression of the initialization of f_0 , a_0 and φ_0 are given as follows

$$\begin{aligned} f_0^{(0)} &= \arg \max_f |\mathcal{F}_y(f)|, \\ a_0^{(0)} &= |\mathcal{F}_y(f_0^{(0)})|, \\ \varphi_0^{(0)} &= \arg \mathcal{F}_y(f_0^{(0)}). \end{aligned} \tag{12}$$

The remaining parameters $\{a_1^{(0)}, \dots, a_p^{(0)}, f_1^{(0)}, \dots, f_q^{(0)}\}$ are initialized with zeros. We could also determine a good initialization of the amplitude and frequency parameters by using the spectrogram of $y[n]$. Since the signal is a single component the spectrogram presented one energy ridge which is extracted and decomposed on the chosen polynomial base to obtain parameter initialization [4]. The effect on the SA algorithm performance is the same as initialization using (12).

In the iteration loop, the generation of new candidates is a procedure controlled by the temperature \mathcal{T} for taking a random step from $\boldsymbol{\theta}$ to $\boldsymbol{\theta} + \Delta\boldsymbol{\theta}$, where $\Delta\boldsymbol{\theta} = [\Delta\theta_0, \dots, \Delta\theta_i, \dots, \Delta\theta_{p+q+2}]^T$. Several schemes [23] for choosing $\Delta\boldsymbol{\theta}$ are possible but we use a Gaussian random move with a variance σ_i^2 for each parameter θ_i of (8)

$$\Delta\theta_i \sim \mathcal{N}(0, \sigma_i^2), \text{ for } i = 0, \dots, p + q + 2 \tag{13}$$

and we define the monitoring vector by

$$\boldsymbol{\sigma}^2 = [\sigma_0^2, \sigma_1^2, \dots, \sigma_i^2, \dots, \sigma_{p+q+2}^2]^T. \tag{14}$$

According to this Gaussian move (13), the new generated candidate has more probability to belong to the neighboring of the current candidate than to the outside. Hence this move is a good compromise between selective and uninformative laws such as uniform, Jeffreys law ...

The change in the energy system is then evaluated with respect to the temperature \mathcal{T} as follows

$$\xi_{\boldsymbol{\theta}} = \exp\left(\frac{\ell_{LS}(\boldsymbol{\theta}) - \ell_{LS}(\boldsymbol{\theta} + \boldsymbol{\Delta}\boldsymbol{\theta})}{\mathcal{T}}\right). \quad (15)$$

The generated candidate is accepted using the Metropolis law: we generate a random variable ξ from a uniform distribution law \mathcal{U} on $[0, 1]$. If $\xi \leq \xi_{\boldsymbol{\theta}}$, the generated candidate is accepted.

A successful search for an acceptable candidate is determined by the choice of the temperature \mathcal{T} and the monitoring vector $\boldsymbol{\sigma}^2$. Both \mathcal{T} and $\boldsymbol{\sigma}^2$ are gradually reduced to allow better solution explorations. Indeed, at high \mathcal{T} and large $\boldsymbol{\sigma}^2$, a parameter research into a wide range is allowed and both increases and decreases of $\ell_{LS}(\boldsymbol{\theta})$ can be accepted to escape local minima. At low \mathcal{T} and small $\boldsymbol{\sigma}^2$ the research is more focused in order to provide better local explorations when progressing towards a solution. Several schemes to reduce both \mathcal{T} and $\boldsymbol{\sigma}^2$ are possible, we propose an original one. We generate a random variable u from a Bernoulli distribution law \mathcal{B} with parameter ρ , if $u = 1$, $\boldsymbol{\sigma}^2 = (1 - \epsilon_{\sigma}) \boldsymbol{\sigma}^2$ and $\mathcal{T} = (1 - \epsilon_{\mathcal{T}})\mathcal{T}$. ϵ_{σ} , $\epsilon_{\mathcal{T}}$ and ρ are determined by experiment.

An other strategic point to remember when monitoring the algorithm is to re-initialize both \mathcal{T} and $\boldsymbol{\sigma}^2$ after T iterations [23,26], T is fixed and chosen by the users. This avoids staying in local minima when the stop criterion is not verified.

As stop criterion for the SA-algorithm, we propose to supervise the power of the noise $\sum_{n=-\frac{N}{2}}^{\frac{N}{2}} |e[n]|^2$ where $e[n] = y[n] - s[n]$ (2). From knowledges about

the noise distributions, $\frac{2}{\sigma_e^2} \sum_{n=-\frac{N}{2}}^{\frac{N}{2}} |e[n]|^2$ is a χ^2 variable with $2(N+1)$ degrees of freedom. This distribution law tends toward a Gaussian since $2(N+1)$ is large. The expectation and variance of the variable $\sum_{n=-\frac{N}{2}}^{\frac{N}{2}} |e[n]|^2$ are $E_N = \sigma_e^2(N+1)$ and $\Sigma_N = \sigma_e^4(N+1)$ respectively. Hence, we define a stop criterion for α -significance level by

$$\sum_{n=-\frac{N}{2}}^{\frac{N}{2}} |y[n] - s[n]|^2 \in [E_N - z_\alpha \sqrt{\Sigma_N}, E_N + z_\alpha \sqrt{\Sigma_N}], \quad (16)$$

where z_α is the α -quantile of a unitary Gaussian distribution. For example $z_\alpha = 1.96$ corresponds to $\alpha = 5\%$.

Table 1 shows the main steps of the SA-based algorithm. $\mathcal{T}^{(0)}$ denotes the initialization of \mathcal{T} determined by experiment whereas $\sigma^2^{(0)}$ is the initialization of σ^2 and must be larger than the constraints defined in Section 5 which are functions of the polynomial base.

Many papers [27,28] have investigated the SA convergence in the discrete optimization problems. However few studies are addressed in the case of continuous parameters [26] which is our case since $\theta \in \mathbb{R}^{p+q+3}$ (11).

From the theoretical point of view, some assumptions described in [26] guarantee the theoretical convergence in probability of the SA algorithm to reach the global optimum value with a given accuracy:

- For example, over a number of iterations the tendency to sample better candidates (with a better objective function values) is stronger than the acceptance of worse candidates.
- Also, the objective function can be approximated by a strictly convex quadratic function in a neighborhood of the global optimum.
- To help the SA convergence, the choice of the way of reducing the tempera-

Table 1

Simulated Annealing based algorithm

- a- Initialization $\boldsymbol{\theta} = \boldsymbol{\theta}^{(0)}$,
 - Evaluate $s[n]$ for $\boldsymbol{\theta}$ using (4), (7) and (3).
 - Set $e[n] = y[n] - s[n]$.
 - b- Repeat until $e[n]$ verifies the whiteness criterion (16),
 - b.1- Set $\boldsymbol{\sigma}^2 = \boldsymbol{\sigma}^{2(0)}$ and $\mathcal{T} = \mathcal{T}^{(0)}$.
 - b.2- Iterations from $t = 1$ to T
 - Generate $\Delta\boldsymbol{\theta}$ as given in (13).
 - Evaluate $\xi_{\boldsymbol{\theta}}$ using (15). Draw $\xi \sim \mathcal{U}[0, 1]$.
 - If $\xi < \xi_{\boldsymbol{\theta}}$, set $\boldsymbol{\theta}$ to $\boldsymbol{\theta} + \Delta\boldsymbol{\theta}$, otherwise $\boldsymbol{\theta}$ is not modified.
 - Sample $u \sim \mathcal{B}(\rho)$. If $u = 1$, set $\boldsymbol{\sigma}^2 = (1 - \epsilon_{\sigma}) \boldsymbol{\sigma}^2$ and $\mathcal{T} = (1 - \epsilon_{\mathcal{T}})\mathcal{T}$.
 - Go to b.2-.
 - b.3- Evaluate $s[n]$ for $\boldsymbol{\theta}$ using (4), (7) and (3). Set $e[n] = y[n] - s[n]$. Go to b-.
-

ture \mathcal{T} and monitoring vector $\boldsymbol{\sigma}^2$ must be done according to the variability of the objective function.

The assessment of the theoretical convergence of the proposed SA algorithm is out of the scope of the present paper. However from the practical point of view the SA algorithm proposed, which generates at each iteration a Gaussian random candidate over a sphere and accept or reject it according to the Metropolis acceptance function, has been experimentally tested in [29] with quite encouraging results. The tests have been done using simulated AM/FM signals the model parameters of which are randomly generated. The choice of the reducing way of both the temperature \mathcal{T} and the monitoring vector $\boldsymbol{\sigma}^2$ seems quite adapted to the variability of the objective function and the stop criterion (16)

helps to avoid staying at local optimum.

3.2 Bayesian Approach

In this paragraph we present the Bayesian model we adopt in the context of short-time non-stationary signals. The Metropolis Hastings algorithm is also detailed.

3.2.1 The Bayesian Model

Authors in [16,30] employed MCMC methods to estimate the parameters of a polynomial phase signal with constant amplitude. In [31], estimation of the constant frequencies of harmonic signals was studied but, to the best of our knowledge, estimation of both amplitude and frequency parameters of short-time signals with nonlinear polynomial IA and IF has never been addressed using a Bayesian approach based on the MH-MCMC method.

In this section, we adopt a Bayesian approach where unknown parameters $\boldsymbol{\theta}$ and the noise variance σ_e^2 are regarded as random variables. An appropriate prior distribution is assigned to each parameter. Then an expression of the posterior distribution is obtained, up to a normalizing constant.

Let $\mathbf{y} = [y[-\frac{N}{2}], \dots, y[\frac{N}{2}]]^T$ be the observation time-vector and I_p the prior information. According to Bayes' theorem, the joint posterior distribution of all unknown parameters conditional on \mathbf{y} and I_p is given by

$$p(\boldsymbol{\theta}, \sigma_e^2 | \mathbf{y}, I_p) = \frac{p(\mathbf{y} | \boldsymbol{\theta}, \sigma_e^2, I_p) p(\boldsymbol{\theta}, \sigma_e^2 | I_p)}{p(\mathbf{y} | I_p)}, \quad (17)$$

where $p(\mathbf{y} | \boldsymbol{\theta}, \sigma_e^2, I_p)$ is the likelihood function, $p(\boldsymbol{\theta}, \sigma_e^2 | I_p)$ is the parameter prior distribution and $p(\mathbf{y} | I_p)$ is the probability of the data given I_p , a nor-

malization constant here. Using (3), (4), (7) and the knowledge of noise probability, the likelihood writes

$$p(\mathbf{y}|\boldsymbol{\theta}, \sigma_e^2, I_p) = (\pi\sigma_e^2)^{-(N+1)} \exp \left[-\frac{1}{\sigma_e^2} \sum_{n=-\frac{N}{2}}^{\frac{N}{2}} |y[n] - s[n]|^2 \right]. \quad (18)$$

As prior independence of the parameters is supposed, the parameter prior distribution is written

$$p(\boldsymbol{\theta}, \sigma_e^2 | I_p) = p(\sigma_e^2 | I_p) \prod_{i=0}^{p+q+2} p(\theta_i | I_p), \quad (19)$$

where θ_i is given in (8). We use uninformative prior distributions to express ignorance about the parameter vector in the absence of data: the classical Jeffrey distribution for σ_e^2 [31]

$$p(\sigma_e^2 | I_p) \propto \frac{1}{\sigma_e^2}, \quad (20)$$

and a uniform distribution with a wide definition domain (40) for θ_i

$$p(\theta_i | I_p) \propto \text{constant} \quad \text{for } i = 0, \dots, p + q + 2. \quad (21)$$

Hence using (18),(19),(20) and (21), we obtain the following expression for the posterior distribution, up to a normalizing constant

$$p(\boldsymbol{\theta}, \sigma_e^2 | \mathbf{y}, I_p) \propto \frac{\pi^{-(N+1)}}{\sigma_e^{2(N+2)}} \exp \left[-\frac{1}{\sigma_e^2} \sum_{n=-\frac{N}{2}}^{\frac{N}{2}} |y[n] - s[n]|^2 \right]. \quad (22)$$

The integration of σ_e^2 yields

$$p(\boldsymbol{\theta} | \mathbf{y}, I_p) \propto \left[\sum_{n=-\frac{N}{2}}^{\frac{N}{2}} |y[n] - s[n]|^2 \right]^{-(N+1)}. \quad (23)$$

This posterior distribution is highly nonlinear and a closed form can not be obtained by conventional numerical integration methods. We therefore use a MH-MCMC algorithm to sample from this distribution. The Minimum Mean-square Error Bayesian Estimators (MMSE) [32] are subsequently used for all

model parameters.

3.2.2 Metropolis Hastings MCMC method

The idea of MCMC methods [33] is to construct a Markov chain, the invariant distribution of which is the target distribution from which samples are desired (23). To generate samples, the methods find and use a transition kernel whose t^{th} iterate converges to the target distribution for large t . This transition kernel represents a conditional probability function of moving from a current state of the Markov chain to another state. Under certain regularity conditions [34], the Markov chain sample path mimics a random sample from the target distribution. In [35], a tutorial review of some of the most MCMC methods is provided and some implementation issues are detailed.

One of the usefulness MCMC methods is the Metropolis Hastings (MH) algorithm which is more general than the Gibbs sampler. As we deal with joint distributions (23) of $p + q + 3$ parameters, a one variable-at-a-time MH algorithm combining $p + q + 3$ updates is proposed. This simplifies the search of the suitable joint kernel transition and allows us to take draws in succession from each of the kernels, it is easier to find several conditional kernels that converge to their respective distributions than to find one kernel that converges to the joint [36].

The MH-MCMC algorithm begins with a candidate generating density which depends upon the current state of the chain $\theta_i^{(t-1)}$, we denote it by $p(\theta_i|\theta_i^{(t-1)})$. We use subscript (i) for the i^{th} parameter and superscript (t) for iteration t . In order to ensure the reversibility of the kernel transition [35,36] the new

candidate θ_i is accepted with the probability $\alpha_i(\theta_i, \theta_i^{(t-1)})$ defined by

$$\alpha_i(\theta_i, \theta_i^{(t-1)}) = \min \{1, \mathcal{R}_i^{(t)}\}, \quad (24)$$

with

$$\mathcal{R}_i^{(t)} = \frac{p(\theta_i \mid \theta_{0,\dots,i-1}^{(t)}, \theta_{i+1,\dots,p+q+2}^{(t-1)}, \mathbf{y}, I_p) p(\theta_i^{(t-1)} \mid \theta_i)}{p(\theta_i^{(t-1)} \mid \theta_{0,\dots,i-1}^{(t)}, \theta_{i+1,\dots,p+q+2}^{(t-1)}, \mathbf{y}, I_p) p(\theta_i \mid \theta_i^{(t-1)})} \quad (25)$$

and $\theta_{0,\dots,i-1}^{(t)} = (\theta_0^{(t)}, \theta_1^{(t)}, \dots, \theta_{i-1}^{(t)})$. In our context to simplify (25), we use a generating distribution that is symmetric and easily simulated $p(\theta_i \mid \theta_i^{(t-1)}) = p(\theta_i^{(t-1)} \mid \theta_i)$ written as

$$\theta_i \mid \theta_i^{(t-1)} \sim \mathcal{N}(\theta_i^{(t-1)}, \sigma_{RW,i}^2). \quad (26)$$

This normal distribution describes a random walk around the current parameter value $\theta_i^{(t-1)}$ with variance $\sigma_{RW,i}^2$. The random-walk MH algorithm is relevant in our case since it does not require the precise location of the target distribution.

The main steps of the proposed MH-MCMC algorithm are described in Table 2. The initialization of the parameter vector is provided by a Fourier Transform of $y[n]$ as in the SA algorithm (12). The MH-MCMC algorithm with one-variable-at-time technique converges to the target distribution after T_0 iterations. Therefore the first T_0 draws are ignored and we collect the next $(T - T_0)$ draws to approximate the posterior distribution, T being the total iteration number and is large. T , T_0 and $\sigma_{RW,i}^2$ (26) can be deduced experimentally and the acceptance rate (24) may be 0.45 [36]. Finally, the MMSE estimators of the model parameters are calculated using samples $\boldsymbol{\theta}^{(t)} = [\theta_0^{(t)}, \dots, \theta_{p+q+2}^{(t)}]^T$ obtained from iteration $t = T_0 + 1, \dots, T$.

In Table 3, we sum up the principle elements of the approaches proposed for

Table 2

A one variable-at-time MH algorithm with a random walk

- 1- Parameter initialization, $\boldsymbol{\theta}^{(0)} = [\theta_0^{(0)}, \theta_1^{(0)}, \dots, \theta_{p+q+2}^{(0)}]^\top$ (12)
 - 2- Iteration from $t = 1$ to T
 - 2.1- Iteration from $i = 0$ to $p + q + 2$
 - 2.1.1- Generate a candidate $\theta_i \sim \mathcal{N}(\theta_i^{(t-1)}, \sigma_{RW,i}^2)$ (26).
 - 2.1.2- Evaluate $\alpha_i(\theta_i, \theta_i^{(t-1)})$ (24).
 - 2.1.3- Draw $u \sim \mathcal{U}([0, 1])$.
If $u \leq \alpha_i(\theta_i, \theta_i^{(t-1)})$, update $\theta_i^{(t)} = \theta_i$, otherwise $\theta_i^{(t)} = \theta_i^{(t-1)}$.
 - 2.2- $\boldsymbol{\theta}^{(t)} = [\theta_0^{(t)}, \dots, \theta_{p+q+2}^{(t)}]^\top$, $t = t + 1$, go to step 2.
 - 3- $\boldsymbol{\theta} = \frac{1}{T-T_0} \sum_{t=T_0+1}^T \boldsymbol{\theta}^{(t)}$.
-

estimating the model parameters (8) of a short-time signal with a nonlinear polynomial IA and IF. Concerning the first approach, our contribution consists in using the maximum likelihood function maximized using the SA algorithm. We give an initialization of the parameters and of the monitoring vector $\boldsymbol{\sigma}^2$ (see Section 5), we define a stop criterion of the SA algorithm and finally we propose an original way in reducing the temperature and the $\boldsymbol{\sigma}^2$. In the Bayesian approach, we extend the existent approaches [16,30,31] to signals whose IA and IF are both nonlinear polynomials. Both approaches that we proposed have the main advantages: they give a direct estimation of all polynomial coefficients, contrary to the existent algorithms [7,15] which iteratively estimate them. In the next section, we give the CRB for assessing the performance of the proposed algorithms.

Table 3

The principle differences between the two proposed algorithms.

Proposed approach	deterministic: LS + SA	Bayesian: MH-MCMC
Objective function	$\ell_{LS}(\boldsymbol{\theta})$ (10)	$p(\boldsymbol{\theta} \mathbf{y}, I_p)$ (23)
Initialization	$\boldsymbol{\theta}^{(0)}$ (12)	$\boldsymbol{\theta}^{(0)}$ (12)
Candidate generation: random walk	σ^2 (14) reduced	$\sigma_{RW,i}^2$ (26) given
Acceptance/rejection	eq.(15) + Metropolis law	eq.(24) + Metropolis law
Parameters of control	$\mathcal{J}, \epsilon_{\mathcal{J}}, \epsilon_{\sigma}, \rho$ and T	T_0 and T
Estimators	Maximum likelihood	MMSE

4 Cramer-Rao Bounds

In [8], CRBs of phase and amplitude parameters are derived for polynomial phase signals, whatever the polynomial function bases. We adapt the formula given in [8] for the frequency and amplitude parameters of a single-component signal modeled by (7). We distinguish two types of CRBs : parameter CRB which are functions of SNR and IA/IF CRBs which are time-varying functions at given SNRs. The former is interesting to evaluate the robustness of the proposed algorithms at low SNRs whereas the latter is to check the algorithm robustness through time.

4.1 Parameter CRBs

To give the explicit Fisher Information Matrix (*FIM*) expression, we introduce the following notations. First, the primitives of the polynomial functions $(g_m[n])_{0 \leq m \leq q}$ (7) are denoted $\eta_m[n]$ and are written for $n = -\frac{N}{2}, \dots, \frac{N}{2}$ and $m = 0, \dots, q$

$$\eta_m[n] = 2\pi(\sum_{k=-\frac{N}{2}}^n g_m[k] - \sum_{k=-\frac{N}{2}}^0 g_m[k]). \quad (27)$$

We also define $\eta_{-1}[n] = 1$. Then, the time vectors \mathbf{s} , Φ , \mathbf{g}_m and $\boldsymbol{\eta}_m$ of the noise-free signal $s[n]$, the signal phase $\Phi[n]$ and the polynomial functions $g_m[n]$ and $\eta_m[n]$ are defined respectively by

$$\begin{aligned} \mathbf{s} &= [s[-\frac{N}{2}], \dots, s[\frac{N}{2}]]^T, \\ \Phi &= [\Phi[-\frac{N}{2}], \dots, \Phi[\frac{N}{2}]]^T, \\ \mathbf{g}_m &= [g_m[-\frac{N}{2}], \dots, g_m[\frac{N}{2}]]^T, \\ \boldsymbol{\eta}_m &= [\eta_m[-\frac{N}{2}], \dots, \eta_m[\frac{N}{2}]]^T. \end{aligned} \quad (28)$$

We then calculate the *FIM* of $\boldsymbol{\theta}$ using (7), (27) and the formula given in [8]

$$FIM(\boldsymbol{\theta}) = \frac{2}{\sigma_e^2} Re \left\{ \begin{bmatrix} \mathbf{A}^\dagger \mathbf{A} & \mathbf{A}^\dagger \Phi \\ \Phi^\dagger \mathbf{A} & \Phi^\dagger \Phi \end{bmatrix} \right\}, \quad (29)$$

where σ_e^2 is the noise variance (2), \dagger is the transpose conjugate symbol and *Re* is the real part symbol. We introduce the following matrices

$$\mathbf{A} = [\mathbf{g}_0 \cdot e^{j\Phi}, \mathbf{g}_1 \cdot e^{j\Phi}, \dots, \mathbf{g}_p \cdot e^{j\Phi}], \quad (30)$$

$$\Phi = j [\boldsymbol{\eta}_{-1} \cdot \mathbf{s}, \boldsymbol{\eta}_0 \cdot \mathbf{s}, \dots, \boldsymbol{\eta}_q \cdot \mathbf{s}], \quad (31)$$

where (\cdot) denotes the element by element multiplication of vectors. The CRBs for $\boldsymbol{\theta}$ are the diagonal elements of the *FIM* matrix inverse. From (29), the *FIM* depends on the polynomial base, the signal and the phase waveforms only. In the case of a single-component signal, $\mathbf{A}^\dagger \boldsymbol{\Phi}$ is purely imaginary and the *FIM* is a block diagonal matrix. Thus amplitude parameters are decoupled from frequency ones. Moreover, an orthonormal base makes $\mathbf{A}^\dagger \mathbf{A}$ a diagonal matrix and consequently, the amplitude parameters are also decoupled and the estimation is improved. In this case the *FIM* expression is written

$$FIM(\boldsymbol{\theta}) = \frac{2}{\sigma_e^2} \begin{pmatrix} \mathbf{I}_p & 0 \\ 0 & \boldsymbol{\Phi}^\dagger \boldsymbol{\Phi} \end{pmatrix}, \quad (32)$$

where \mathbf{I}_p is the identity matrix of $(p+1) \times (p+1)$ size and $\boldsymbol{\Phi}^\dagger \boldsymbol{\Phi}$ is a matrix of $(q+2) \times (q+2)$ size.

4.2 IA/IF CRBs

Since we are also interested on the evaluation of the algorithm performance for the estimation of the IA and IF waveforms, we give the CRBs of both of them by adapting formula of [8] to the signal model of (7)

$$CRB(A[n]) = \frac{\sigma_e^2}{2} \mathbf{g}^\dagger (\mathbf{A}^\dagger \mathbf{A})^{-1} \mathbf{g}, \quad (33)$$

$$CRB(F[n]) = \frac{\sigma_e^2}{2} \mathbf{h}^\dagger (\boldsymbol{\Phi}^\dagger \boldsymbol{\Phi})^{-1} \mathbf{h}, \quad (34)$$

where $\mathbf{g} = [g_0[n], g_1[n], \dots, g_p[n]]^\top$ and $\mathbf{h} = [0, g_0[n], g_1[n], \dots, g_q[n]]^\top$. For an orthonormal polynomial base, eq. (33) is reduced to the following expressions

$$CRB(A[n]) = \frac{\sigma_e^2}{2} \sum_{m=0}^p g_m[n]^2, \text{ for } \frac{-N}{2} \leq n \leq \frac{N}{2} \quad (35)$$

since $\mathbf{A}^\dagger \mathbf{A} = \mathbf{I}_p$. These CRBs (33), (34) and (35) are time-varying functions and consequently sensitive to SNR variations due to the non stationarity of the signal amplitude. We shall see in Section 6.1 that the local decrease of the SNR has a strong effect on the IF and IA estimates.

In the next section, we give the discrete orthonormal polynomial bases that make $\mathbf{A}^\dagger \mathbf{A}$ a diagonal matrix.

5 Discrete orthonormal polynomial-bases

An orthonormal polynomial base is useful for providing uncoupled parameters (32), thereby improving the estimation. Previous works [7] refer to orthonormal bases built from a discretization of continuous-time orthonormal polynomials (Legendre, Tchebychev, Hermite ...). However the orthogonal property is lost when discretizing these functions. Consequently, the smaller the data size ($N + 1$), the more significant the parameter coupling. Therefore, we propose two discrete orthonormal polynomial bases: the discrete orthonormal Legendre polynomials and a discrete base we have derived using Gram' Schmidt procedure.

The discrete Legendre polynomial base [37], denoted $g_m^{DL}[i]$, is given for $0 \leq i \leq N + 1$ by

$$g_m^{DL}[i] = \frac{1}{C_m^N} \sum_{l=0}^m (-1)^l \binom{l}{m} \binom{l}{m+l} \frac{i^{\{l\}}}{(N+1)^{\{l\}}}, \quad (36)$$

where $i^{\{l\}} = i(i-1)\dots(i-l+1)$ is the backward factorial function of order l , $C_m^N = \left(\frac{(N+2+m)^{\{m+1\}}}{(2m+1)(N+1)^{\{m\}}}\right)^{\frac{1}{2}}$ is the normalization coefficient and $\binom{l}{m} = \frac{m!}{l!(m-l)!}$ is the binomial coefficient. In [38] a fast algorithm is proposed to evaluate the polynomial coefficients of (36) since they need computation.

At low approximation orders, we can also derive a discrete orthonormal polynomial base using the Gram' Schmidt procedure. So, let

$$g_m[n] = b_{m,0}n^0 + b_{m,1}n^1 + \dots + b_{m,m-1}n^{m-1} + b_{m,m}n^m$$

be a m^{th} order polynomial defined over $[-\frac{N}{2}, \frac{N}{2}]$. The polynomial coefficients are calculated using the Gram-Schmidt procedure

$$\langle \mathbf{g}_m, \mathbf{g}_l \rangle = \sum_{n=-\frac{N}{2}}^{\frac{N}{2}} g_m[n] g_l[n] = \delta_{m,l}, \quad (37)$$

where \langle, \rangle is the scalar product, $\delta_{m,l}$ is the Kronecker symbol and \mathbf{g}_m defined in (28). For example, the second-order coefficients ($m = 2$) are as follows

$$\begin{aligned} b_{2,0} &= \frac{-\sqrt{5N(N+2)}}{2\sqrt{(N+3)(N+1)(N-1)}}, & b_{1,0} &= 0, & b_{0,0} &= \frac{1}{\sqrt{N+1}}, \\ b_{2,1} &= 0, & b_{1,1} &= \frac{2\sqrt{3}}{\sqrt{N(N+1)(N+2)}}, & & \\ b_{2,2} &= \frac{6\sqrt{5}}{\sqrt{(N+3)(N+2)(N+1)N(N-1)}}. \end{aligned} \quad (38)$$

After some calculation steps, we find that at low approximation orders the discrete Legendre polynomials and the derived base are linearly dependent

$$g_m[n] = (-1)^m g_m^{DL}\left[n + \frac{N}{2}\right]. \quad (39)$$

Therefore, the amplitude and frequency parameters calculated using the derived base (38) are equal to those obtained using the discrete Legendre base (36) up to a multiplicative constant $(-1)^m$. In the following, we only consider the base given by (38).

Using (5), (6) and (38), the constraints on $A[n]$ and $F[n]$ are then transformed into constraints on the model parameters, which give a good initialization of the monitoring vector $\boldsymbol{\sigma}^2$ (14) of the SA algorithm. For example, for a second-order polynomial modeling ($p=q=2$) the definition domains of the frequency parameters are

$$\begin{aligned} |f_0| &\leq \frac{Fs}{2|b_{0,0}|} \left(1 + \frac{4|b_{2,0}|}{N^2|b_{2,2}|}\right) = \sigma_4^2, \\ |f_1| &\leq \frac{Fs}{2|b_{1,1}|} \left(\frac{1}{N} + \frac{4|b_{2,1}|}{N^2|b_{2,2}|}\right) = \sigma_5^2, \\ |f_2| &\leq \frac{2Fs}{N^2|b_{2,2}|} = \sigma_6^2, \end{aligned} \tag{40}$$

where F_s , σ_i^2 and $b_{m,i}$ are given by (6), (14) and (38) respectively. The same expressions are obtained for the amplitude parameters by substituting $\max_n |s[n]|$ for $\frac{Fs}{2}$ and a_i for f_i into (40). This gives initializations of σ_0^2 , σ_1^2 and σ_2^2 . The initial phase verifies $|\varphi_0| \leq \pi = \sigma_3^2$.

In the next section, we give some simulation results for assessing the performance of the proposed approaches.

6 Simulation Results

To study the robustness of the proposed algorithms, we pay attention to the following points. First, we study the effect of the SNR variation through time on the estimation accuracy. Second, we show that the algorithm performance differs according to the polynomial base choice. We then compare the approaches proposed with the HAF-based method. Finally, we discuss the effects of the model order limitation on the estimation results. Two single component signals with quadratic polynomial IA/IF are considered. These signals denoted by s_1 and s_2 , their parameter vectors $\boldsymbol{\theta}_{s_1}$ and $\boldsymbol{\theta}_{s_2}$, defined in the discrete or-

thonormal base (38), are as follows

$$\boldsymbol{\theta}_{s_1} = [a_0=16, a_1=-3, a_2=2, \varphi_0=0.4, f_0=2, f_1=-0.2, f_2=-0.5]^T, \quad (41)$$

$$\boldsymbol{\theta}_{s_2} = [a_0=16, a_1=10.4, a_2=4.5, \varphi_0=0.4, f_0=1.5, f_1=-0.2, f_2=-0.5]^T. \quad (42)$$

Here p and q equal two, the sample number is 33 and the sampling frequency F_s is equal to $1Hz$. Figure 1 shows the signals s_1 and s_2 having different modulation rates.

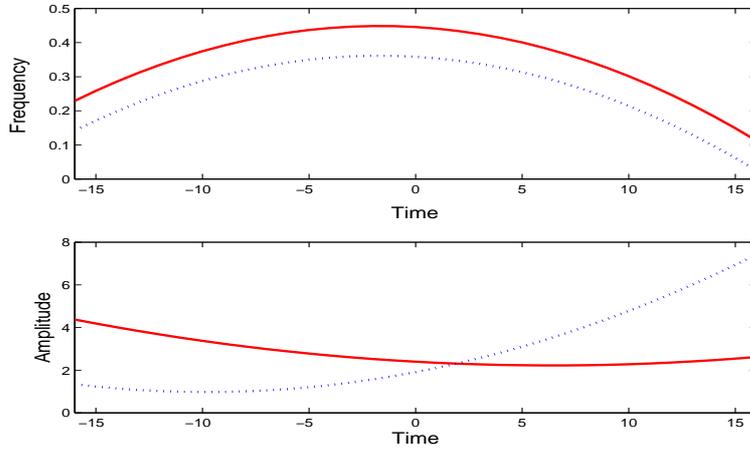


Fig. 1. IA and IF of signals: (—) s_1 and (...) s_2 with two different frequency modulation rates.

We introduce two different SNR definitions interesting for the algorithm performance study. First, the mean SNR is defined as the ratio of the energy of a constant amplitude signal (equal to that of the time-varying amplitude signal) to the noise variance. The mean SNR is different from the operative SNR (instantaneous SNR) which is a time-varying function due the IA changes. We define the operative SNR as the ratio of the energy of a constant amplitude signal (equal to that of the time-varying signal) on a sliding window of a small

sample size (about $\frac{N}{4}$) to the noise variance. Figure 2 displays the operative SNR of s_1 and s_2 versus the mean SNR=5 dB. In the following Monte Carlo noise simulations are performed over 200 runs at each mean SNR and for each signal.

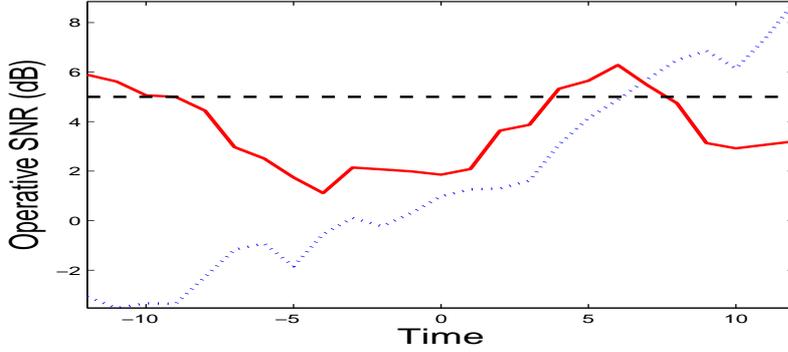


Fig. 2. Operative SNR of (—) s_1 and (...) s_2 versus (—) the mean SNR=5 dB. Operative SNR is calculated with a sliding window of 8 samples, the total sample number being 33.

6.1 Time-varying SNR effects

The aim in this section is to study the effect of the time-varying SNR on the estimation accuracy in the case of short-time signals. We consider the signals s_1 (41) and s_2 (42) and we compare the estimation results obtained using the SA-based algorithm. We choose by experiment $\epsilon_\sigma = \epsilon_\gamma = 0.03$, $T = 50000$ and $\rho = 0.01$ (see Table 1). The polynomial base considered is the discrete orthogonal (38).

The MSE values for the estimated IAs and IFs of signals s_1 and s_2 and the corresponding time-varying CRBs (33) are depicted in Fig. 3 at a mean SNR

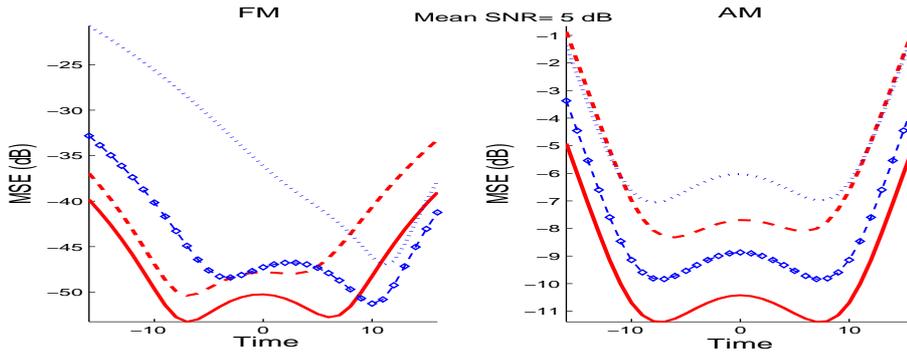


Fig. 3. Operative SNR effect on the estimation efficiency of the IF (left side) and IA (right side) at a mean SNR=5 dB: MSE (—) in the case of s_1 versus the appropriate CRB (---) and MSE (...) in the case of s_2 versus the appropriate CRB (- -◇).

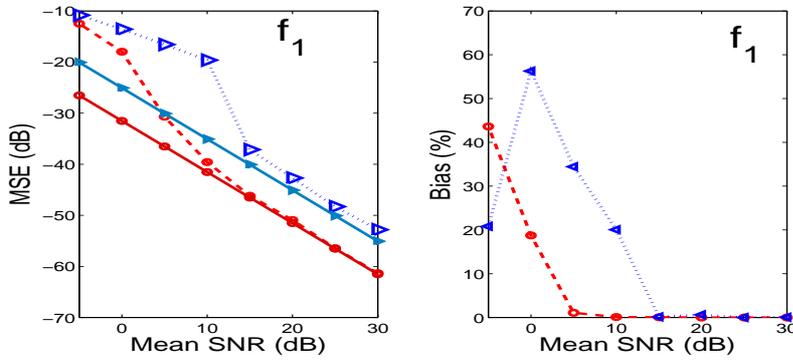


Fig. 4. Operative SNR effect on parameter estimation: (left side) the MSE (—o) of f_1 estimate of signal s_1 versus the appropriate CRB (---o) and the MSE (...△) of f_1 estimate of signal s_2 versus the appropriate CRB (---△). (right side) the bias (—o) of f_1 estimate of signal s_1 and the bias (...△) of f_1 estimate of signal s_2 .

equal to 5 dB. By comparing the different curves, we observe that the difference between the IF MSE and the corresponding CRB is larger for s_2 than for s_1 . Figure 4 depicts the MSE and bias of frequency parameter f_1 versus mean SNRs varying from -5 dB to 30 dB in cases of s_1 and s_2 . The parameter CRBs (32) evaluated for f_1 of signal s_2 and f_1 of signal s_1 are also plotted. We note that the difference between MSE of parameter f_1 of signal s_2 and the corresponding CRB is higher than that in the case of signal s_1 . Moreover the

threshold effect in the case of s_1 appears at mean SNR below 8 dB whereas, in the case of s_2 it appears at mean SNR lower than 15 dB . This difference in the estimation accuracy in cases of s_1 and s_2 is essentially due to the significant variation in the operative SNR. Indeed, as shown in Fig. 2, the operative SNR for the signal s_2 is much lower than 5 dB especially for the left side of the time window; which consequently induces estimation accuracy differences according to time as shown in Fig. 3. The more the significant variation of the operative SNR, the more the estimation accuracy decreasing.

6.2 Base influence on the parameter estimation

In this section, three different bases are compared: the proposed orthonormal base (38), the canonical base ($g_m[n] = n^m$) and the discretized Legendre base. The latter is obtained from the classical Legendre polynomials which are defined on continuous time (on $[-1, 1]$), then translated and sampled to obtain expressions on $[-\frac{N}{2}, \frac{N}{2}]$ [7]. This discretized base is different from the discrete Legendre base (36). The canonic base and the discretized Legendre one are not orthogonal. The SA-based algorithm is used to achieve the maximum likelihood estimation of the model parameters of signal s_1 (41) and the results obtained are compared to the derived CRBs (32). We note that it is difficult to theoretically demonstrate the strong influence of the polynomial base on the algorithm running, therefore we perform Monte Carlo simulation studies.

Figure 5 displays the MSE and the bias of the estimated parameters a_0 , f_2 and θ_0 of the signal s_1 versus the CRBs (29). The bias (in percent) is normalized by the known parameter values. Actually, the amplitude parameters are

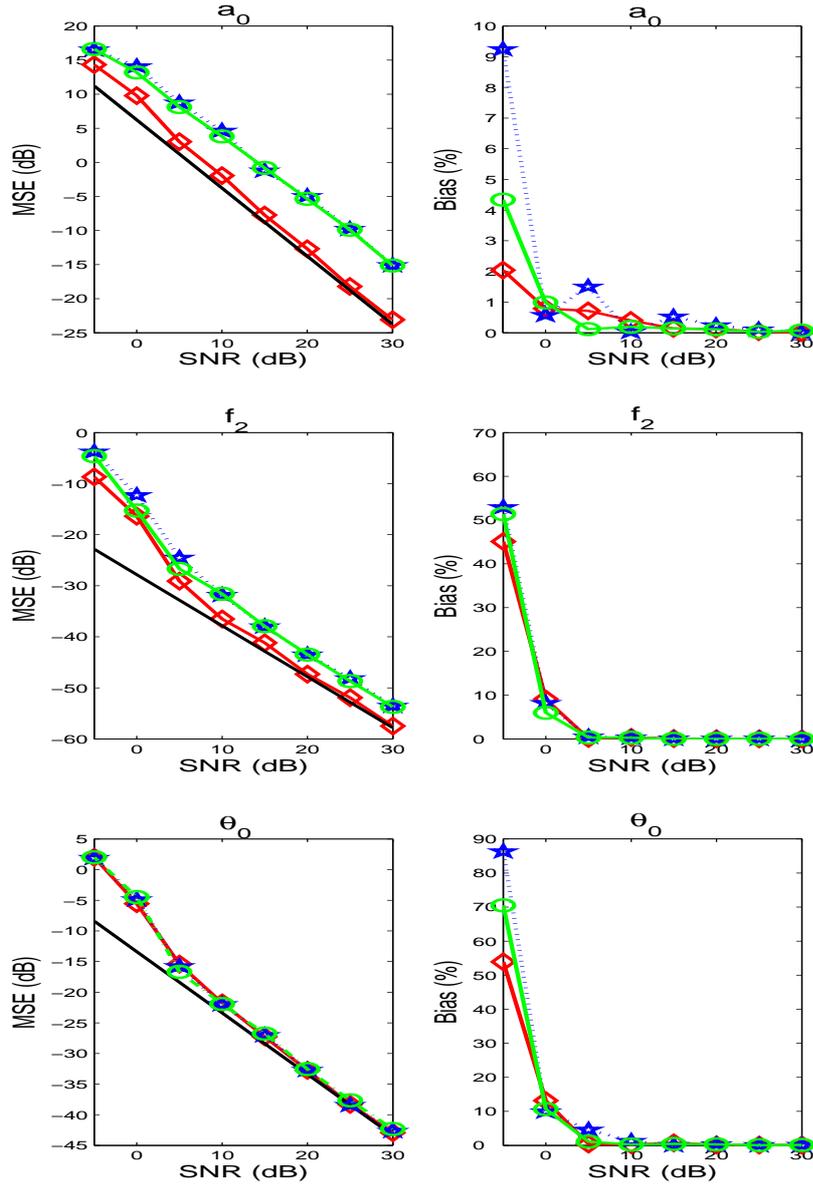


Fig. 5. Influence of the base choice on the estimated parameters: (—) CRB, (\diamond) discrete orthonormal base, (\circ) discretized Legendre base and (\star) canonical base: the MSEs and biases of s_1 parameters (41) versus the mean SNRs. The estimation is achieved by using the SA algorithm.

coupled in the case of the discretized Legendre and canonical bases albeit they are not coupled in the orthonormal proposed base (38), as shown by (32). As may be seen from Fig. 5, the orthonormal property enhances the efficiency of the parameter estimation by significantly reducing the MSE. Indeed, we gain

more than 10 dB on the MSE of the amplitude parameters and more than 5 dB on the frequency ones. The discrete orthonormal base outperforms the other bases in terms of variances and bias. From Fig. 5, we also note that the SA algorithm leads to a biased estimation of all the parameters, especially at low SNRs. This is mainly due to the small sample size. As the CRB is the lowest bound for the variance of an unbiased estimator, estimation errors are not directly comparable to the CRB but the latter is used as a reference in evaluating the performance of the algorithm. Moreover the likelihood function is a nonlinear function of the frequency parameters, which explains the threshold that appears at SNR below 5 dB . This threshold effect is well-known in the literature [19].

6.3 Algorithm Comparison

In order to compare the performance of the algorithms presented in Section 3 and the existing HAF-based method in terms of estimation accuracy, we consider the short-time signal s_1 (41), the IF and IA of which are quadratic polynomials. However, parameter φ_0 is assumed to be known in the HAF algorithm. Indeed, the HAF technique is not able to correctly estimate the initial phase, and the estimation error of the phase is stretched over the amplitude parameter estimation. All algorithms are coded using MATLAB and simulations are performed on 'bi-Pentium IV¹, 3.2 GHz, 4 Go of RAM'. Each processing of one realization takes an average of 120s, 320s and less than 1s using the SA algorithm, the MH-MCMC method and the HAF-based method, respectively.

¹ 25593.00 Bogomips.

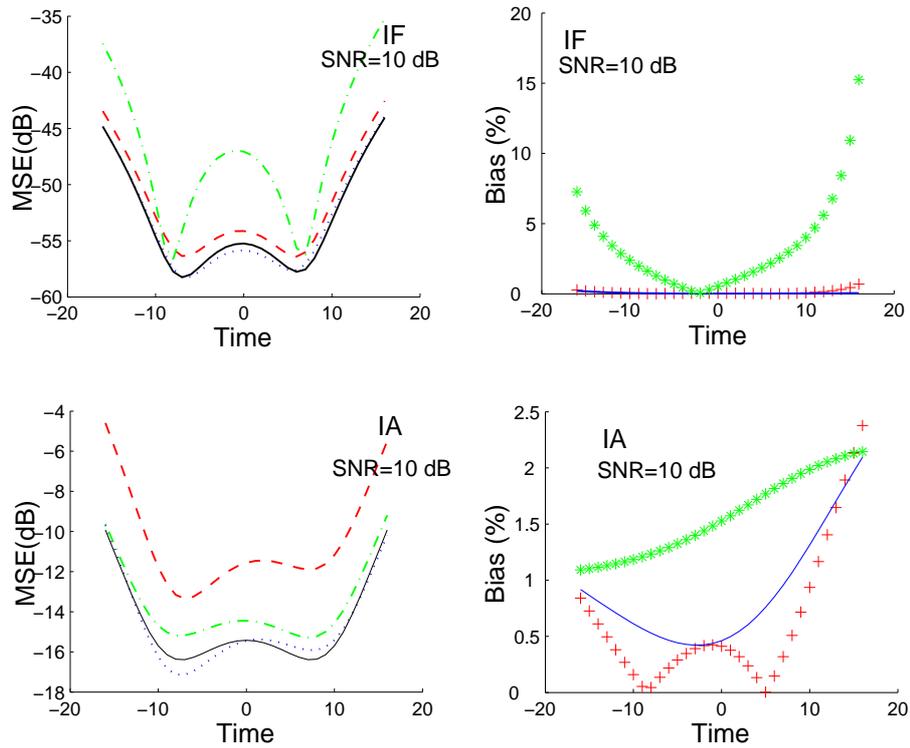


Fig. 6. Algorithm performances at mean $SNR = 10$ dB: (left column) The MSE of the IF and IA using (...) MH-MCMC, (---) SA and (-.-) HAF. (right column) The bias of the IF and IA: (—) MH-MCMC, (+++) SA and (***) for the HAF.

Figures 6 and 7 show the MSE and the bias of the IF and IA estimates using these three techniques at 10 dB and 0 dB. As expected, the HAF technique fails at low SNR. For this technique, the MSE and bias obtained are larger than those of the MH-MCMC or the SA technique. Each of the approaches proposed outperforms the HAF in term of estimation accuracy at low SNR. The MH-MCMC estimators have the smallest MSE but are very biased. However Fig. 8 shows the Markov Chain, generated using the proposed MH-MCMC algorithm, is sufficiently mixed, the acceptance rate is 0.4432. The SA-based method is a good compromise between small MSE, bias and CPU time.

In Table 4, the empirical means, bias and Standard Deviation (STD) of es-

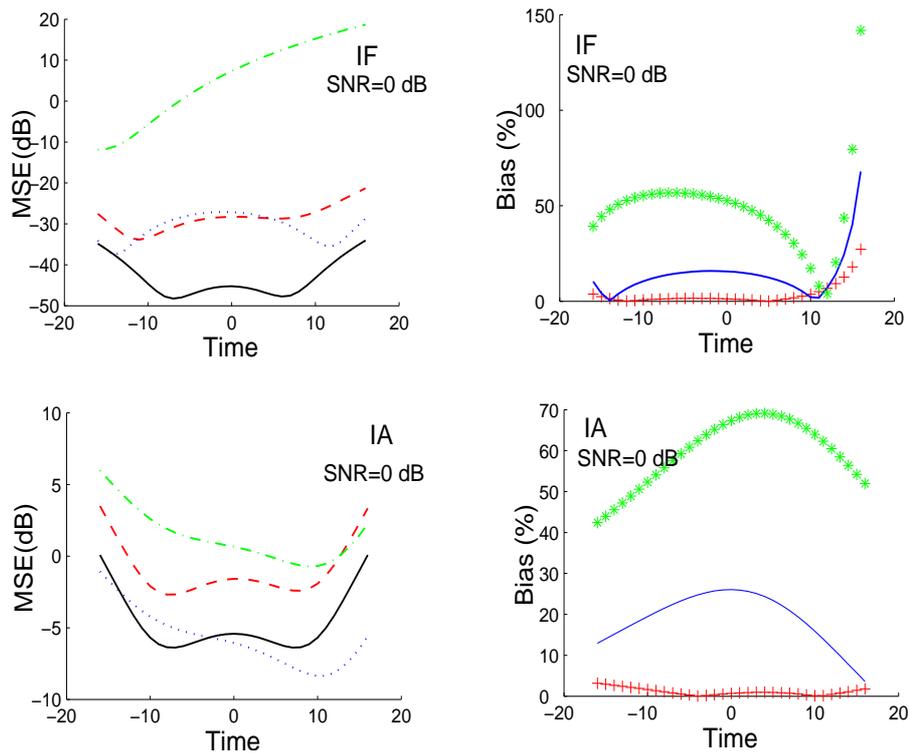


Fig. 7. Algorithm performances at mean $SNR = 0$ dB: (left column) The MSE of the IF and the IA using (...) MH-MCMC, (---) SA and (-.-) HAF. (right column) The bias of the IF and IA: (—) MH-MCMC, (+++) SA and (***) HAF.

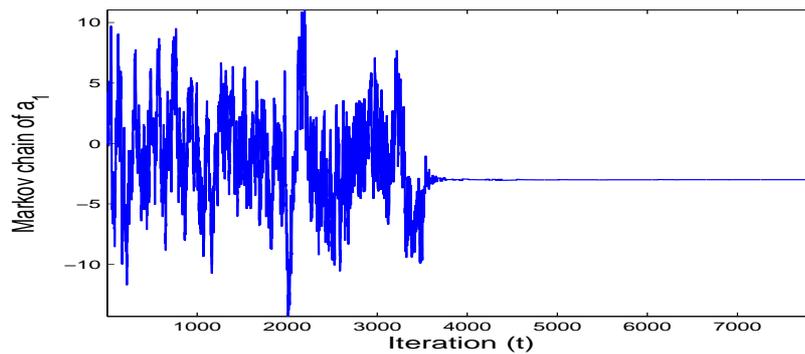


Fig. 8. The Markov Chain generated for the parameter a_1 at mean $SNR=30$ dB using the proposed MH-MCMC algorithm. The algorithm converges to the distribution target after $T_0 = 5000$ iterations. $T = 8000$ is the total iteration number.

estimated parameters f_0 , f_2 and a_2 are reported at a mean SNR equal to 5 dB. Using the SA, the estimates are significantly closer to the true values than those obtained using the MH-MCMC method or the HAF.

Table 5 compares the mean SNR evaluated from the residue obtained with the Table 4

Algorithm comparison: mean, bias and STD of f_0 , f_2 and a_2 estimates at SNR=5 dB.

		f_0	f_2	a_2
True value		2	-0.5	2
Simulated	Mean	2.0015	-0.4995	2.0290
	Bias	0.0015	0.0005	0.0290
	STD	0.0251	0.0351	1.7572
MH-MCMC	Mean	2.0034	-0.4915	1.7150
	Bias	0.0034	0.0085	0.2850
	STD	0.0199	0.0319	1.1865
HAF	Mean	0.9433	-0.5621	1.8326
	Bias	1.0567	0.0621	0.1674
	STD	7.6622	1.6436	1.4775

three algorithms, to the true mean SNR (10 dB and 0 dB). The HAF-based method totally fails at SNR equal to 0 dB. As a result, we conclude that the SA-based method gives the best approximation, which means that it performs better at finding the global *extremum* of the likelihood function in the case of

Table 5

Algorithm comparison: Estimated SNR evaluated from residue obtained with the three algorithms for the true SNR values 10 dB and 0 dB.

True value		mean SNR (dB)	
		10	0
SA	Estimated SNR	9.8421	-0.2383
MH-MCMC	Estimated SNR	9.8048	-1.8771
HAF	Estimated SNR	9.7519	-7.1053

few sample numbers and low SNR.

The approaches proposed give a better accuracy compared to the HAF and to the spectrogram when estimating non-stationary signals with both non-linear IA and IF and having a very short-time duration. The MSE and bias of the parameter estimates are larger using the HAF than using the approaches proposed. The HAF fails at low SNR and/or at high-order polynomial modeling (2nd order polynomial IA and IF with a high frequency modulation rate), this is mainly due to the error propagation phenomenon. The spectrogram is a non-parametric method which has a very bad resolution because of the small sample number. Moreover the IF and IA cannot be reconstructed using the spectrogram at the beginning and ending of the time window.

Three particularities of the approaches proposed contribute considerably to reducing the estimation errors. The first is that the polynomial model is more adapted to approximate non-linear AM and FM signals than that used in the existing techniques. Indeed, we model the IF instead of the phase and the

initial phase is referenced to the center of the time segment. The second particularity is that the discrete orthonormal polynomial base helps improving the estimation accuracy by decoupling the amplitude parameters. Finally, to achieve the parameter estimation, we use techniques which are robust in the presence of noise: the MCMC method using the Metropolis Hastings algorithm and the maximization of the likelihood using the simulated annealing technique. The model parameter are simultaneously estimated thus avoiding the error propagation phenomenon known with the HAF method [15].

6.4 *Effects of low-order polynomial modeling*

In this section, we aim to study the effects of the low-order polynomial modeling on signals whose phase is a high-order polynomial and on signals whose phase is non polynomial. Indeed the estimation error of the IA and IF is due not only to the errors in the estimation parameters caused by noise but also to this low-order approximation. The SA-based algorithm is employed to provide parameter estimation and the orthonormal base (38) is used. A total of 200 runs of Monte Carlo noise simulations are performed at each mean SNR.

We consider two different signals:

- a synthetic long-time signal whose IF and IA are both fifth-order polynomials. This signal, denoted s_3 , is highly nonlinear with a phase of order six. The total sample size is 256.
- a sinusoidal IA/IF signal, denoted s_4 , whose frequency and amplitude are

given by

$$F[n] = 0.25 + 0.1\cos\left(\frac{2\pi n}{256}\right) \quad \text{for } 0 \leq n \leq 256. \quad (43)$$

$$A[n] = 5 + 5\sin\left(\frac{\pi n}{256}\right)$$

In both cases of s_3 and s_4 , we model the IA and IF locally by second-order polynomials. Then the SA algorithm is locally applied to short-time segments chosen from the entire signal. Figure 9(a) displays the estimation of two disjoint segments with lengths of 37 samples (index (1)) and of 33 samples (index (2)) in the case of signal s_3 . Figure 10(a) shows the local segments considered in the case of signal s_4 , whose length is 51 samples (index (1)) and 71 samples (index (2)) respectively. The CRBs are computed using (33) and (34) for the mean SNR evaluated on the chosen segments.

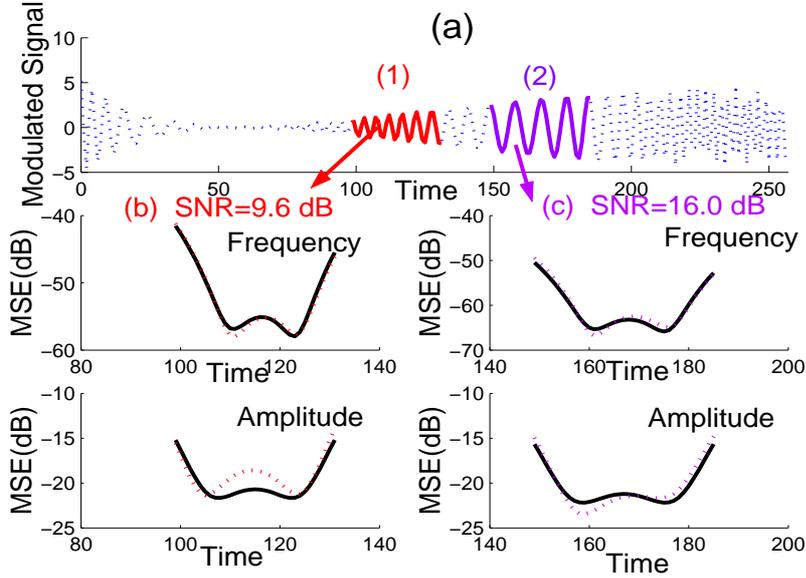


Fig. 9. (a): A non stationary signal with a fifth order IA/IF polynomial modulation (...) and the two segments considered (—). Fig. (b) and (c): the CRB (—) and the MSE (...) of the IF and IA of each segment are plotted.

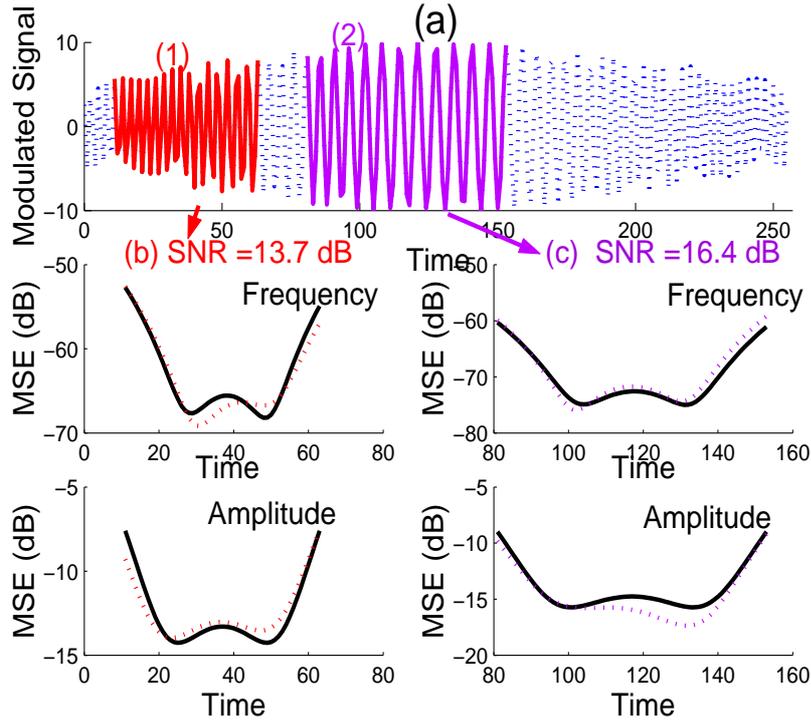


Fig. 10. (a) A nonstationary signal with sinusoidal modulations (...) and the two segments considered (—). Figs.(b) and (c): CRB (—) and the MSE (...) of the IF and IA of each segment.

Figures 9(b) and (c) display the MSE of the local IA and IF estimates in the case of signal s_3 whereas Figs. 10(b) and (c) display the MSE in the case of signal s_4 . These figures show a good estimation accuracy of the local IA and IF, the MSEs are close to the derived CRBs. We can see differences in the performance from one segment to another due to significant changes in the operative SNR and due to amplitude variations. Therefore the estimation of each segment depends strongly on the SNR variations. Reconstructed curves of IA and IF of the high-order polynomial signal s_3 are plotted versus the original ones in Fig. 11 (left side), the estimated curves are close to the original ones. Normalized biases are depicted in Fig. 11 (right side).

From these examples, we illustrate the great interest of the approaches pro-

posed when addressing the estimation of highly non stationary signals with long-time duration. Indeed, a low-order (second-order) polynomial model suffices to locally model the non-stationarity on short-time segments [4]. The results obtained on each segment using the SA algorithm proposed are close to the CRB and the estimation accuracy is high.

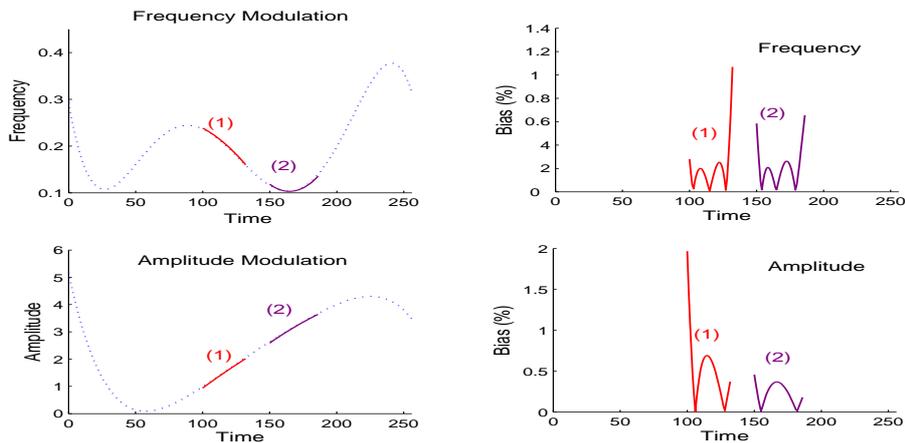


Fig. 11. Approximation of the IA and IF, which are 5^{th} -order polynomials, by low-order polynomials: (left column) original curves of the IA and IF (...) and the estimated curves (—). (right column) bias of the IA and IF estimates.

7 Conclusion

In this paper, we have investigated the estimation of short-time signals with nonlinear amplitude and frequency modulations. We have proposed and compared two different approaches to process single-component signals whose both the IA and IF could be modeled by low-order polynomial models. In the first approach, a Maximum-Likelihood procedure, which has the advantage of preserving the optimality, is maximized. To perform the maximization efficiently, we had to evaluate a highly nonlinear equation, for which we used

a SA algorithm. The second approach is based on a Bayesian approach. A MH-MCMC method, one variable at a time, was used together with a random walk technique to sample from the posterior distribution. Monte Carlo simulations showed that the SA offers a good compromise between small variances and low bias compared to the MH-MCMC method. These two methods outperform the well-known HAF technique.

In the context of short-time signals, we also showed the importance of the orthogonal property in the choice of polynomial model by comparing different discrete polynomial bases. The discrete orthonormal base proposed is the best in terms of the estimation accuracy, since the orthogonality property in discrete time reduces the parameter coupling.

We also studied the influence of low-order polynomial modeling on the IA and IF estimation. We considered two cases: a signal whose true phase is a polynomial of high order, and a signal whose phase is non-polynomial. For each, we considered only short-time segments, where the IA and IF are locally approximated by second-order polynomials. The MSE obtained using the proposed approaches are close to the CRBs. This approach is employed to estimate long-time signals with high non-stationarity at low SNR [4]. The long-time signal is segmented in short-time windows. A low-order polynomial model is assigned for each window, as described in the present paper. We have already presented the first results of segmentation [4,17,18], where a procedure for segmentation and merging of different local models is described. Thus, we do not need higher order to model the entire signal phase. Error propagation phenomena will be avoided. The great advantage of this method is its flexibility for modeling a wide range of highly non-stationary signals whatever the modulation order and signal length. In [4], we present an example of a real canary song where variations in the signal are so fast that the proposed local

short-time modeling is necessary.

We would also like to point out possible extensions of this work. Given the small sample number and the biased estimators, it will be interesting to study other statistical bounds such as Barankin bounds [39]. We can also focus on the case of multicomponent signals with more challenging problems such as crossing or closely spaced frequencies in colored noise. The first results of short-time multicomponent signals embedded in Gaussian noise are given in [21] and [40]. We also aim to make comparisons with other existent techniques [41,42] which do not suppose polynomial modeling of the IA and IF.

References

- [1] P. Maragos, J. Kaiser, T. Quatieri, Energy separation in signal modulations with application to speech analysis, *IEEE Trans. on Signal Processing* 41, No. 10 (1993) 3024–3051.
- [2] T. Quatieri, R. Dunn, R. McAulay, T. Hanna, Underwater signal enhancement using a sine-wave representation, in: *IEEE Oceans'92 Proceedings, 1992*, pp. 449–454.
- [3] F. Leonard, J. Lantaigne, S. Lalonde, Y. Turcotte, Vibration behavior of a cracked cantilever beam, *Mechanical Systems and Signal Processing* 15 (2001) 529–548.
- [4] M. Jabloun, N. Martin, M. Vieira, F. Leonard, A new flexible approach to estimate highly non-stationary signals, Accepted in *IEEE Trans. on Signal Processing* 2007.
- [5] C. Vincent, G. Kappenberger, F. Valla, A. Bauder, M. Funk, E. L. Meur, Ice ablation as evidence of climate change in the Alps over 20th century, *Journal*

of Geophysical Research 109 (2004) D10104.

- [6] E. T. M. Gay, C. Chauveau, Bilan de masse du glacier de sarennes, 55 et 56 ieme saisons 2002-2003-2004, Grenoble - France, 2005.
- [7] S. Peleg, B. Porat, B. Friedlander, The achievable accuracy in estimating the instantaneous phase and frequency of a constant amplitude, *IEEE Trans. on Signal Processing* 71 (1993) 2216–2224.
- [8] B. Friedlander, J. Francos, Estimation of amplitude and phase parameters of multicomponent signals, *IEEE Trans. on Signal Processing* 43 (1995) 917–926.
- [9] S. Saha, S. Kay, Maximum likelihood parameter estimation of superimposed chirps using Monte Carlo importance sampling, *IEEE Trans. on Signal Processing* 50 (1995) 224–230.
- [10] S. Barbarossa, A. Scaglione, G. Giannakis, Product High-order Ambiguity Function for multicomponent polynomial-phase signal modeling, *IEEE Trans. on Signal Processing* 49 (1998) 691–708.
- [11] G. Zhou, G. Giannakis, A. Swami, On polynomial phase signals with time-varying amplitudes, *IEEE Trans. on Signal Processing* 44, No. 4 (1996) 848–861.
- [12] A. Francos, M. Porat, Analysis and synthesis of multicomponent signals using positive Time-Frequency distributions, *IEEE Trans. on Signal Processing* 47 (1999) 493–504.
- [13] M. Ikram, G. Zhou, Estimation of multicomponent polynomial phase signals of mixed orders, *Signal Processing* 81 (2001) 2293–2308.
- [14] P. O’Shea, A fast algorithm for estimating the parameters of a quadratic FM signal, *IEEE Trans. on Signal Processing* 52 - No. 2 (2004) 385–393.

- [15] C. Ioana, C. Cornu, A. Quinquis, Polynomial phase signal processing via warped higher-order ambiguity function, in: Proc. of EUSIPCO, Vienna, Austria, 2004, pp. 1159–1162.
- [16] C. Theys, M. Vieira, A. Ferrari, Bayesian estimation of the parameters of polynomial phase signals, in: ICASSP, Seattle, USA, 1997, pp. 3553–3356.
- [17] M. Jabloun, M. Vieira, F. Leonard, N. Martin, Local orthonormal decomposition for both instantaneous amplitude and frequency of highly non-stationary discrete signals, in: IMA: The Institute of Mathematics and Its Applications, Cirencester, United Kingdom, 2004, pp. 107–110.
- [18] M. Jabloun, M. Vieira, F. Leonard, N. Martin, A AM/FM single component signal reconstruction using a non-sequential time segmentation and polynomial modeling., in: NSIP: IEEE workshop on Nonlinear Signal and Image Processing, Sapporo, Japan, ISBN 0-7803-9064-4, 2005.
- [19] D. Rife, R. Boorstyn, Single-tone parameter estimation from discrete-time observations, IEEE Trans. on information theory IT-20 (1974) 591–598.
- [20] L. Cohen, P. Loughlin, D. Vakman, On an ambiguity in the definition of the amplitude and phase of a signal, Signal Processing 79 (1999) 301–307.
- [21] M. Jabloun, N. Martin, M. Vieira, F. Leonard, Multicomponent signal: Local analysis and estimation, in: EUSIPCO, Antalya - Turkey, 2005.
- [22] S. Kirkpatrick, C. Gelatt, J. Vecchi, Optimization by simulated annealing, Science 220-4598 (1983) 670–679.
- [23] W. Press, S. Teukolsky, W. Vetterling, B. Flannery, Numerical recipes in fortran 77, The Art of Scientific Computing, Cambridge University Press (Second Edition).
- [24] P. V. Laarhoven, E. Aarts, Simulated annealing: Theory and applications, Kluwer Academic Publishers, ISBN: 9-027-72513-6 (1987).

- [25] M. Vieira, F. Leonard, M. Jabloun, N. Martin, Short time single polynomial phase using Legendre functions, in: Proc. of EUSIPCO, Vienna, Austria, 2004, pp. 793–796.
- [26] M. Locatelli, Simulated annealing algorithm for continuous global optimization: convergence conditions, *Journal of Optimization theory and applications* 104(1) (2000) 121–154.
- [27] L. Ingber, Simulated annealing: practice versus theory, *Mathematical and Computer Modelling* 18 (1993) 29–57.
- [28] L. Herault, Rescaled simulated annealing- accelerating convergence of simulated annealing by rescaling the states energies, *Journal of Heuristics* 6 (2000) 215–252.
- [29] M. Jabloun, Phd thesis : Modelisation de signaux fortement non stationnaires a phase et a amplitude locales polynomiales, directed by N. martin, Institut National Polytechnique de Grenoble, GIPSA lab, 38402 St Martin d’Heres France (defended the 10th July 2007).
- [30] Y. Lin, X. Wang, Y. Peng, Parameter estimation of chirp signals using the Metropolis-Adjusted-Langevin’s algorithm, in: ICSP, 2004, pp. 160–163.
- [31] C. Andrieu, A. Doucet, Joint bayesian model selection and estimation of noisy sinusoids via reversible jump MCMC, *IEEE Trans. on Signal Processing* 47 (1999) 2667–2676.
- [32] S. M. Kay, *Fundamentals of statistical signal processing: estimation theory*, Signal Processing Series, Prentice Hall (1993).
- [33] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, E. Teller, Equation of state calculations by fast computing machines, *Journal of Chemical Physics* 21 (1953) 1087–1092.

- [34] G. O. Roberts, A. F. M. Smith, Simple conditions for the convergence of the Gibbs sampler and Metropolis Hastings algorithms, *Stochastic Processes and applications* 49 (1994) 207–216.
- [35] S. P. Brooks, Markov chain Monte Carlo and its application, *The Statistician* 47 (1998) 69–100.
- [36] S. Chib, E. Greenberg, Understanding the Metropolis Hastings algorithm, *American Statistician* 49, No. 4 (1995) 327–335.
- [37] N. Morrison, Introduction to sequential smoothing and prediction, McGraw-Hill Book Company, New York, Vol 1 (Second Edition 1969).
- [38] M. Aburdene, On the computation of discrete legendre polynomial coefficients, in: *Multidimensional systems and signal processing*, Kluwer Academic Publishers, Boston. Manufactured in The Netherlands, 1993, pp. 181–186.
- [39] P. Ciblat, M. Ghogho, Ziv-Zakai bound for harmonic retrieval in multiplicative and additive Gaussian noise, in: *Proc. IEEE workshop Statistical Signal Processing*, Bordeaux, France, 2005.
- [40] M. Jabloun, N. Martin, M. Vieira, F. Leonard, Maximum likelihood parameter estimation of short-time multicomponent signals with nonlinear AM/FM modulation, in: *IEEE workshop on Statistical Signal Processing*, Bordeaux, France, 2005.
- [41] A. Boudraa, J. Cexus, F. Salzenstein, L. Guillon, IF estimation using empirical mode decomposition and nonlinear teager energy operator, in: *ICASSP*, Montreal, Canada, 2004, pp. 45–48.
- [42] C. Dubois, M. Davy, J. Idier, Tracking of time-frequency components using particle filtering, in: *ICASSP*, Philadelphia, US, 2005, pp. IV.9–IV.12.