View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Open Archive Toulouse Archive Ouverte

# Classification of Chirp Signals Using Hierarchical Bayesian Learning and MCMC Methods

Manuel Davy, Christian Doncarli, and Jean-Yves Tourneret

***Abstract*—This paper addresses the problem of classifying chirp signals using hierarchical Bayesian learning together with Markov chain Monte Carlo (MCMC) methods. Bayesian learning consists of estimating the distribution of the observed data conditional on each class from a set of training samples. Unfortunately, this estimation requires to evaluate intractable multidimensional integrals. This paper studies an original implementation of hierarchical Bayesian learning that estimates the class conditional probability densities using MCMC methods. The performance of this implementation is first studied via an academic example for which the class conditional densities are known. The problem of classifying chirp signals is then addressed by using a similar hierarchical Bayesian learning implementation based on a Metropolis-within-Gibbs algorithm.**

***Index Terms*—Classification of chirps, hierarchical Bayesian learning, MCMC methods, supervised classification.**

## I. Introduction

BAYES classification theory has received much attention in the pattern recognition literature (for an overview, see [1], [2], and references therein). This popularity may be justified by the fact that the Bayesian classifier is optimal in terms of an appropriate overall risk or the average probability of error. The Bayes classifier assumes the knowledge of the prior probabilities and the conditional probability densities of each class. Unfortunately, these quantities are unknown in many practical applications and have to be estimated. The estimation of the prior probabilities generally presents no serious difficulty [1, p. 44]. However, estimation of the class-conditional probability density functions (pdfs) raises troublesome problems.

An usual approach for estimating the class-conditional probability densities is the so-called hierarchical Bayesian learning (HBL), which has received much attention in the literature [3]–[7]. Hierarchical modeling assumes that the training samples have distributions depending on different but related parameters. These parameters are modeled as random variables or vectors whose distribution is parametrized by so-called hyperparameters. Our knowledge about these hyperparameters is contained in a known prior distribution. Unfortunately, the class-conditional densities resulting from HBL cannot be generally expressed in closed-form because of intractable multidimensional integrals. Similarly, closed-form expressions for the predictive distribution of the observations, conditional on the training set of each class, are generally difficult to obtain. Such problems can be solved by approximating the integrals using stochastic sampling or deterministic numerical integration. However, stochastic sampling usually yields better approximations, as explained in [5] and [8]. This paper proposes to approximate the integrals required for HBL by using Markov chain Monte Carlo (MCMC) methods. These methods have recently been considered with increasing interest in many signal processing applications (see [9] and references therein). MCMC methods consist of generating samples by running an ergodic Markov chain (MC), which converges to a desired target distribution. These samples can then be used to approximate multidimensional integrals using the ergodic theorem.

Bayesian learning implementations using MCMC methods have been used in many applications. In [10], many integration techniques, including MCMC methods, have been proposed for the computation of Bayes factors, with a particular interest on five scientific applications in genetics, sports, ecology, sociology, and psychology. A general framework for the classification and discrimination of normal-mixture models using Bayesian learning and MCMC methods has been studied in [11]. The main contribution of this paper is to apply the methodology of Lavine and West [11] to the problem (recently introduced in [12]–[15]) of classifying multicomponent linear (sinusoids) and quadratic (chirps) frequency-modulated signals. This paper assumes that the number of components in each signal is known, which can be justified in many applications including target identification and knock detection (presented in Section IV). However, for applications where this number of components is unknown, a reversible jump MCMC algorithm similar to the one developed in [7] can be derived (see [16]).

Section II briefly recalls HBL principle for supervised classification theory and studies an original HBL implementation based on MCMC methods. Section III studies an interesting example providing closed-form expressions for the posterior distributions and the class-conditional probability densities. The performance of the proposed HBL implementation is evaluated by comparing these closed-form expressions with MCMC approximations. The application of HBL to the classification of chirp signals is investigated in Section IV. Conclusions are reported in Section V.

Manuscript received February 16, 2001; revised October 9, 2001. The associate editor coordinating the review of this paper and approving it for publication was Dr. Petar M. Djurić.

M. Davy is with the Engineering Department, University of Cambridge, Cambridge, U.K. (e-mail: md283@eng.cam.ac.uk).

C. Doncarli is with the Institut de Recherche en Communication et Cybernétique de Nantes (IRCCyN), Nantes, France (e-mail: Christian.Doncarli@irccyn.ec-nantes.fr).

J.-Y. Tourneret is with ENSEEIHT/TESA, Toulouse Cedex, France (e-mail: jean-yves.tourneret@tesa.prd.fr).

Publisher Item Identifier S 1053-587X(02)00558-5.

1053–587X/02$17.00 © 2002 IEEE

Authorized licensed use limited to: INP TOULOUSE. Downloaded on July 19, 2009 at 12:41 from IEEE Xplore. Restrictions apply.

## II. Hierarchical Bayesian Learning

An available set $\mathbf{X}$ of independent so-called training samples is separated in $c$ subsets corresponding to the $c$ classes $\omega_i$, $i = \{1, \ldots, c\}$. Denote $\mathbf{X} = \{\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_c\}$, where $\mathcal{X}_i = \{x_{1,i}, \ldots, x_{n_i,i}\}$ contains the training samples $x_{j,i}$, $j = \{1, \ldots, n_i\}$ corresponding to the class $\omega_i$. Choosing a zero-one loss function (which minimizes the average error probability), the Bayes classification rule can be written as

$$\text{Decide } x \in \omega_i \quad \text{if} \quad P(\omega_i|x, \mathbf{X}) \geq P(\omega_k|x, \mathbf{X}), \quad \forall k \neq i \tag{1}$$

where $k = \{1, \ldots, c\}$, and $P(\omega_i|x, \mathbf{X})$ is the posterior probability of the class $\omega_i$. Assuming that the classes have the same prior probability and that the subset $\mathcal{X}_i$ provides no information about the class $\omega_k$, $k \neq i$ (separability condition) [1, p. 50], the Bayes classification rule reduces to

$$\text{Decide } x \in \omega_i \quad \text{if} \quad p(x|\mathcal{X}_i, \omega_i) \geq p(x|\mathcal{X}_k, \omega_k), \quad \forall k \neq i \tag{2}$$

where $k = \{1, \ldots, c\}$, and $p(x|\mathcal{X}_i, \omega_i)$ is the probability density of class $\omega_i$ whose training samples are summarized in $\mathcal{X}_i$. This equation shows that the Bayesian classifier only requires computation of the pdfs $p(x|\mathcal{X}_i, \omega_i)$. The separability condition ensures that each class can be treated independently. Consequently, by forgetting the class index for brevity, we have to solve the following problem for each class: Estimate an unknown distribution $p(x|\mathcal{X})$ by using a set $\mathcal{X} = \{x_1, \ldots, x_n\}$ of $n$ samples drawn independently according to this unknown distribution.

### A. Hierarchical Modeling

Hierarchical models are used when the samples $x_j$, inside a given class, have pdfs depending on different (but related) unknown parameters $\theta_j$, which are denoted $p(x_j|\theta_j)$. Inside the class $\omega_i$, the parameters $\theta_j$ are distributed according to a pdf depending on an unknown hyperparameter $\Phi$ characterizing the class, which is denoted $p(\theta_j|\Phi)$. Finally, the prior distribution of $\Phi$ is assumed to be known and is denoted $p(\Phi)$. Such a hierarchical modeling has received much attention for the analysis of signals defined by random parameters [3]–[6]. However, this modeling is also interesting for the analysis of signals with deterministic parameters since it introduces an additional level in prior modeling, increasing the robustness of the learning process [5, p. 112], [7]. The desired predictive distribution of an observation $x$ denoted $p(x|\mathcal{X})$ is then defined by

$$p(x|\mathcal{X}) = \iint p(x|\theta, \Phi)p(\theta, \Phi|\mathcal{X})\, d\theta\, d\Phi. \tag{3}$$

Since $p(x|\theta, \Phi) = p(x|\theta)$, (3) reduces to

$$\boxed{p(x|\mathcal{X}) = \iint p(x|\theta)p(\theta, \Phi|\mathcal{X})\, d\theta\, d\Phi} \tag{4}$$

where $p(\theta, \Phi|\mathcal{X}) = p(\theta|\Phi)p(\Phi|\mathcal{X})$ (using the hierarchical structure), and

$$p(\Phi|\mathcal{X}) = \frac{p(\Phi)\prod_{j=1}^{n} p(x_j|\Phi)}{\int p(\Phi)\prod_{j=1}^{n} p(x_j|\Phi)\, d\Phi} \tag{5}$$

i.e.,

$$\boxed{p(\Phi|\mathcal{X}) \propto p(\Phi)\prod_{j=1}^{n} p(x_j|\Phi)} \tag{6}$$

where $\propto$ means *proportional to*. Using the hierarchical structure, $p(x_j|\Phi)$ can be computed as follows:

$$\boxed{p(x_j|\Phi) = \int p(x_j|\theta_j)p(\theta_j|\Phi)\, d\theta_j.} \tag{7}$$

Equations (4), (6), and (7) are the key HBL equations. Note that when $p(\theta|\Phi)$ is the Dirac delta function, i.e., when $p(\theta|\Phi) = \delta(\theta - \Phi)$, HBL reduces to general Bayesian learning, which has also received much attention in the literature (for instance, see [1, p. 51]).

### B. HBL Implementation Using MCMC Methods

Unfortunately, closed-form expressions for the integrals in (4) and (7) are rarely available. This section studies an HBL implementation based on MCMC methods. Equation (4) shows that the predictive distribution of the observation $x$ can be written as follows:

$$p(x|\mathcal{X}) = \int p(x|\theta)\left\{\int p(\theta, \Phi|\mathcal{X})\, d\Phi\right\} d\theta. \tag{8}$$

In other words, $p(x|\mathcal{X})$ can be interpreted as the mean of $p(x|\theta)$, where $\theta$ is distributed according to the marginal distribution

$$\int p(\theta, \Phi|\mathcal{X})\, d\Phi = p(\theta|\mathcal{X}). \tag{9}$$

The proposed Bayesian learning implementation first generates samples $(\tilde{\theta}_l, \tilde{\Phi}_l)$ distributed according to $p(\theta, \Phi|\mathcal{X})$ using the following two-step procedure:

1) Generation of $L$ samples $\tilde{\Phi}_l$ distributed according to

$$p(\Phi|\mathcal{X}) = \int p(\theta_1, \theta_2, \ldots, \theta_n, \Phi|\mathcal{X})\, d\theta_1\, d\theta_2 \cdots d\theta_n. \tag{10}$$

This is achieved by using the Gibbs Sampler to draw $L$ vectors $(\tilde{\theta}_{1l}, \tilde{\theta}_{2l}, \ldots, \tilde{\theta}_{nl}, \tilde{\Phi}_l)$, $l = \{1, \ldots, L\}$, according to the joint pdf $p(\theta_1, \theta_2, \ldots, \theta_n, \Phi|\mathcal{X})$;

2) For each sample $\tilde{\Phi}_l$, generation of one sample $\tilde{\theta}_l$ distributed according to $p(\theta|\tilde{\Phi}_l)$.

Authorized licensed use limited to: INP TOULOUSE. Downloaded on July 19, 2009 at 12:41 from IEEE Xplore. Restrictions apply.

After generating $L$ vectors $(\tilde{\theta}_l, \tilde{\Phi}_l)$, $l = \{1, \ldots, L\}$ that are distributed according to $p(\theta, \Phi|\mathcal{X})$, the predictive distribution of the observation $x$ conditional on the training set $\mathcal{X}$ is estimated as follows:

$$\hat{p}(x|\mathcal{X}) = \frac{1}{L} \sum_{l=1}^{L} p\left(x|\tilde{\theta}_l\right). \tag{11}$$

## III. Academic Example

This section studies an interesting example, where closed-form expressions of $p(\Phi|\mathcal{X})$ and $p(x|\mathcal{X})$ are available. Consequently, approximations of these distributions obtained using MCMC simulations can be compared with the corresponding theoretical expressions. Consider a class $\omega$ whose training set is denoted $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$. Each training sample is an $N$-dimensional vector $\boldsymbol{x}_j = (\boldsymbol{x}_j[0], \ldots, \boldsymbol{x}_j[N-1])^T$, which is defined as[1]

$$\forall i = \{0, \ldots, N-1\}, \qquad \boldsymbol{x}_j[i] = m_j + \boldsymbol{\epsilon}_j[i] \tag{12}$$

where $p(m_j) \sim \mathcal{N}(\mu, \sigma_m^2)$, and $(\boldsymbol{\epsilon}_j[i])_{i=\{0\cdots N-1\}}$ is a sequence of i.i.d. zero mean Gaussian variables with variance $\sigma_\epsilon^2$. The variances $\sigma_m^2$ and $\sigma_\epsilon^2$ are assumed to be known. The variables $m_j$ and $m_k$, as well as the sequences $(\boldsymbol{\epsilon}_j[i])_{i=\{0\cdots N-1\}}$ and $(\boldsymbol{\epsilon}_k[i])_{i=\{0\cdots N-1\}}$, are assumed independent for $j \neq k$. Moreover, $m_j$ and $\boldsymbol{\epsilon}_j[i]$ are also independent variables. Note that two training vectors $\boldsymbol{x}_j$ and $\boldsymbol{x}_k$ have different distributions since $m_j \neq m_k$ for $j \neq k$. Consequently, the general Bayesian learning principle [1, p. 51] cannot be used to solve this problem. This section shows how hierarchical Bayesian learning allows the estimation of $p(\boldsymbol{x}|\mathcal{X})$, which is necessary for the classification of $\boldsymbol{x}$. In this example, the model parameters are $\theta_j = m_j$, $j = \{1, \ldots, n\}$ (where $n$ is the number of training samples), and the hyperparameter is $\Phi = \mu$.

### A. HBL Closed-Form Expressions

Standard computations lead to

$$p(\boldsymbol{x}_j|\mu) \propto \exp\left\{ -\frac{\left[N(\overline{v}_j - \overline{x}_j^2)\left(\frac{\sigma_m}{\sigma_\epsilon}\right)^2 + \overline{v}_j + \mu^2 - 2\mu\overline{x}_j\right]}{2\left(\sigma_m^2 + \frac{1}{N}\sigma_\epsilon^2\right)} \right\} \tag{13}$$

where

$$\overline{x}_j = \frac{1}{N}\sum_{i=0}^{N-1} \boldsymbol{x}_j[i] \quad \text{and} \quad \overline{v}_j = \frac{1}{N}\sum_{i=0}^{N-1} \boldsymbol{x}_j^2[i]. \tag{14}$$

Note that $p(\boldsymbol{x}_j|\mu)$ is proportional to the given closed-form expression (13) up to a factor that does not depend on $\boldsymbol{x}_j$ and $\mu$. By choosing a Gaussian prior for $\mu$ i.e., $p(\mu) \sim \mathcal{N}(\mu_0, \sigma_0^2)$, $p(\mu|\mathcal{X})$ can be expressed in closed-form [see (6)]

$$p(\mu|\mathcal{X}) \sim \mathcal{N}(\mu_n, \sigma_n^2) \tag{15}$$

where

$$\mu_n = \frac{n\sigma_0^2}{n\sigma_0^2 + \sigma_m^2 + \frac{1}{N}\sigma_\epsilon^2} \cdot \frac{1}{n}\sum_{j=1}^{n} \overline{x}_j + \frac{\sigma_m^2 + \frac{1}{N}\sigma_\epsilon^2}{n\sigma_0^2 + \sigma_m^2 + \frac{1}{N}\sigma_\epsilon^2}\mu_0$$

$$\sigma_n^2 = \frac{\sigma_0^2\left(\sigma_m^2 + \frac{1}{N}\sigma_\epsilon^2\right)}{n\sigma_0^2 + \sigma_m^2 + \frac{1}{N}\sigma_\epsilon^2}. \tag{16}$$

Straightforward computations show that the predictive distribution of an observation $\boldsymbol{x}$ denoted $p(\boldsymbol{x}|\mathcal{X})$ is the pdf of a multivariate Gaussian distribution given by

$$p(\boldsymbol{x}|\mathcal{X}) \propto \exp\left\{ -\frac{1}{2}\left[\frac{\overline{v} - \overline{x}^2}{\frac{1}{N}\sigma_\epsilon^2} + \frac{(\overline{x} - \mu_n)^2}{\sigma_n^2 + \sigma_m^2 + \frac{1}{N}\sigma_\epsilon^2}\right]\right\} \tag{17}$$

where

$$\overline{x} = \frac{1}{N}\sum_{i=0}^{N-1} \boldsymbol{x}[i] \quad \text{and} \quad \overline{v} = \frac{1}{N}\sum_{i=0}^{N-1} \boldsymbol{x}^2[i]. \tag{18}$$

The corresponding mean vector $\boldsymbol{\mu} = E[\boldsymbol{x}]$ and covariance matrix $\boldsymbol{\Sigma} = E[(\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})^T]$ can be computed as

$$\boldsymbol{\mu} = \mu_n(1, \ldots, 1)^T$$

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{\sigma_\epsilon^2}\boldsymbol{I}_N + \left(\frac{1}{N^2\left(\sigma_n^2 + \sigma_m^2 + \frac{\sigma_\epsilon^2}{N}\right)} - \frac{1}{N\sigma_\epsilon^2}\right)\mathbf{1}_N \tag{19}$$

where $\boldsymbol{I}_N$ is the $N \times N$ identity matrix, and $\mathbf{1}_N$ is the $N \times N$ matrix of ones.

### B. HBL Implementation Using MCMC Methods

The *learning step* of the HBL implementation can be summarized into two parts:

*First Part—Generation of $L$ Samples $\tilde{\mu}_l$ Distributed According to $p(\mu|\mathcal{X})$:* Vectors $(\tilde{m}_{1l}, \tilde{m}_{2l}, \ldots, \tilde{m}_{nl}, \tilde{\mu}_l)$, $l = \{1, \ldots, L\}$, which are distributed according to the joint pdf $p(m_1, m_2, \ldots, m_n, \mu|\mathcal{X})$, are generated using the Gibbs sampler (GS). The GS algorithm is summarized as follows (see Appendix A for more details):

- Generation of $L$ vectors $\tilde{m}_l = (\tilde{m}_{1l}, \ldots, \tilde{m}_{nl})$ distributed according to $p(m_1, \ldots, m_n|\tilde{\mu}_{l-1}, \mathcal{X})$. Using the independence assumption regarding the training samples $\boldsymbol{x}_i$, $n$ samples $\tilde{m}_{il}$ are generated independently according to $p(m_i|\tilde{\mu}_{l-1}, \boldsymbol{x}_i) \propto p(\boldsymbol{x}_i|m_i)p(m_i|\tilde{\mu}_{l-1})$.
- Generation of samples $\tilde{\mu}_l$ distributed according to $p(\mu|\tilde{m}_{1l}, \ldots, \tilde{m}_{nl}) \propto p(\mu)\prod_{j=1}^{n} p(\tilde{m}_{jl}|\mu)$.

*Second Part—Generation of Samples $\tilde{m}_l$ Distributed According to $p(m|\tilde{\mu}_l)$:* The *test step* of the HBL implementation consists of estimating the predictive distributions of test samples from each class (different from the training samples) as

$$\hat{p}(\boldsymbol{x}|\mathcal{X}) = \frac{1}{L}\sum_{l=1}^{L} p(\boldsymbol{x}|\tilde{m}_l). \tag{20}$$

[1] Throughout this paper, we adopt the following: Vectors are denoted using bold fonts, whereas scalars are denoted using normal fonts. The $i$th component of a vector $\boldsymbol{v}$ is denoted $\boldsymbol{v}[i]$.

Authorized licensed use limited to: INP TOULOUSE. Downloaded on July 19, 2009 at 12:41 from IEEE Xplore. Restrictions apply.

TABLE I
ACADEMIC EXAMPLE: SIMULATION PARAMETERS

| | $p(m_j) = \mathcal{N}(\mu, \sigma_m^2)$ | | $p(\epsilon_j) = \mathcal{N}(m_\epsilon, \sigma_\epsilon^2)$ | | $N$ |
|---|---|---|---|---|---|
| | $\mu$ | $\sigma_m$ | $m_\epsilon$ | $\sigma_\epsilon$ | |
| Class 1 | 0 | 0.5 | 0 | 0.1 | 10 |
| Class 2 | 0.5 | 0.5 | 0 | 0.1 | 10 |

TABLE II
ACADEMIC EXAMPLE: SIMULATION RESULTS FOR THE MARKOV CHAIN GENERATION

| nb. of training samples $n_1 = n_2 \rightarrow$ | | 10 | 20 | 50 | 100 | 200 | 500 |
|---|---|---|---|---|---|---|---|
| simulation parameters | $\widetilde{\mu}_l$ Markov chain length $L$ | 5,000 | | | | | |
| | Burn-in length $B$ | 100 | | | | | |
| | prior pdf $p(\mu)$ | $\mathcal{N}(0, 4)$ | | | | | |
| Class $\omega_1$ | mean of $(\widetilde{\mu}_l)$: $\widehat{\mu}_n$ | 0.197 | 0.086 | -0.017 | -0.006 | -0.026 | 0.020 |
| | exact value of $\mu_n$ | 0.189 | 0.082 | -0.014 | -0.005 | -0.026 | 0.021 |
| | variance of $(\widetilde{\mu}_l)$: $\widehat{\sigma}_n^2$ | 0.0233 | 0.0122 | 0.0048 | 0.0026 | 0.0013 | 0.0005 |
| | exact value of $\sigma_n^2$ | 0.0249 | 0.0125 | 0.0050 | 0.0025 | 0.0012 | 0.0005 |
| Class $\omega_2$ | mean of $(\widetilde{\mu}_l)$: $\widehat{\mu}_n$ | 0.681 | 0.578 | 0.486 | 0.494 | 0.470 | 0.527 |
| | exact value of $\mu_n$ | 0.686 | 0.581 | 0.485 | 0.494 | 0.473 | 0.521 |
| | variance of $(\widetilde{\mu}_l)$: $\widehat{\sigma}_n^2$ | 0.0242 | 0.0129 | 0.0059 | 0.0026 | 0.0012 | 0.0005 |
| | exact value of $\sigma_n^2$ | 0.0249 | 0.0125 | 0.0050 | 0.0025 | 0.0012 | 0.0005 |
| Computation time (arbitrary unit) | | 10 | 20 | 50 | 100 | 200 | 500 |

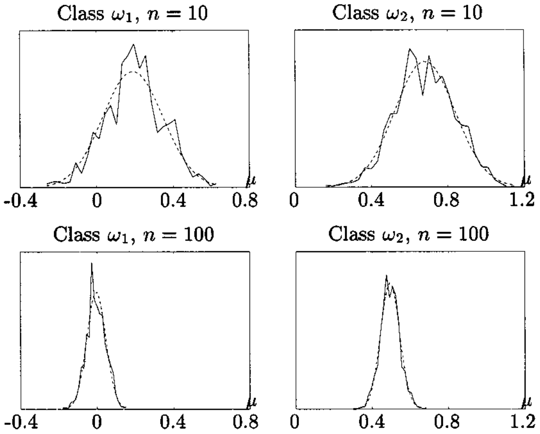


Fig. 1. Academic example: Histograms of the MC samples $\tilde{\mu}_l$ generated using HBL for different numbers of training samples (solid lines). Exact distribution $p(\mu|\mathcal{X})$ for the corresponding training samples (dashed lines).

These test samples are then classified using the Bayes rule defined in (2), after replacing the predictive densities by their estimates.

### C. Simulation Results

Simulation results are presented for a two-class problem whose parameters are displayed in Table I. Note that the two classes only differ by their mean vectors. The results of the *learning step* are presented in Table II. In this implementation, the number of training samples varies from $n_1 = n_2 = 10$ to $n_1 = n_2 = 500$. The mean and variance of the Markov Chain samples are clearly in excellent agreement with the exact values deduced from the closed-form expressions. Fig. 1 shows the histograms of the MC elements $\tilde{\mu}_l$ compared with the Gaussian distribution $\mathcal{N}(\mu_n, \sigma_n^2)$ (see [1, p. 54] for similar plots). As can be seen, the proposed HBL implementation based on MCMC methods shows good performance for learning $p(\mu|\mathcal{X})$.

During the *test step*, 1000 test samples (i.e., 500 test samples from each class that are different from the training samples) have been generated according to (12). These test samples are then classified using the previous HBL implementation.

TABLE III
ACADEMIC EXAMPLE: ERROR RATES ESTIMATED WITH 500 TEST SAMPLES IN EACH CLASS

| nb. of training samples | $n_1 = n_2 \rightarrow$ | 10 | 20 | 50 | 100 | 200 | 500 |
|---|---|---|---|---|---|---|---|
| HBL | Error rates (%) | 31.4 | 31.6 | 31.4 | 31.2 | 31.2 | 31.0 |
| HBL MCMC implementation | Error rates (%) | 31.3 | 31.5 | 31.5 | 31.3 | 31.2 | 31.3 |
| | Computation time | 200 (arbitrary unit) | | | | | |

Table III shows the corresponding error rates, which can be compared with those obtained from the closed-form expression of $p(\boldsymbol{x}|\mathcal{X})$. The comparison shows that the proposed HBL implementation performs very similarly to the optimal closed-form decision rule, i.e., is highly reliable.

## IV. CLASSIFICATION OF CHIRP SIGNALS USING HBL IMPLEMENTATION

This section studies the classification of multicomponent chirp signals using HBL implementation. This example has been recently introduced in [12] and [15]. The test signals are generated using the following model:

$$\boldsymbol{x}[i] = \sum_{k=1}^{m} a_k D_k[i] + \boldsymbol{\epsilon}[i], \qquad i = \{0, \ldots, N-1\} \quad (21)$$

where $D_k[i] = \cos(2\pi[\phi_k + f_k i + s_k i^2])$ is the $k$th chirp component with amplitude $a_k$, initial frequency $f_k$, initial phase $\phi_k$, and slope $s_k$. The additive noise $\boldsymbol{\epsilon}[i]$ is a Gaussian zero-mean white noise of unknown variance $\sigma_\epsilon^2$. The number of chirp components $m$ is assumed to be known in each class. This modeling is interesting in many realistic applications including the following.

- **Target Identification**

  When there is a relative motion between the target and the receiver, the radar signal can be modeled as a chirp signal with quadratic phase $\phi(t) = a_0 + a_1 t + a_2 t^2$. The parameters $a_1$ and $a_2$ are either related to speed and acceleration or range and speed, depending on what the radar is intended for and on the kind of waveforms transmitted [17], [18, p. 56-65]. In the presence of a harmonic jammer, the received signal is the sum of a chirp and a sinusoid, which corresponds to $m = 2$. In this particular case, the proposed classification problem is interesting to determine the target nature (high versus low acceleration/speed) (see [19] for a similar problem).
- **Knock Detection**

  The knock detection problem has received much attention in the literature [20], [21]. Several studies have shown that the knock data consist of several resonances with decreasing resonance frequencies. More precisely, the signals presented in [20] and [21] are the sum of three chirps (which corresponds to $m = 3$) whose parameters depend on the motor rotation speed. The classification algorithm proposed in this paper is useful to classify knocking and nonknocking signals.

For the sake of simplicity, we introduce the following notations:

- *Signal:* $\boldsymbol{x} = [\boldsymbol{x}[0], \boldsymbol{x}[1], \ldots, \boldsymbol{x}[N-1]]^t$;
- *Amplitudes:* $\boldsymbol{a} = [a_1, a_2, \ldots, a_m]^t$;

Authorized licensed use limited to: INP TOULOUSE. Downloaded on July 19, 2009 at 12:41 from IEEE Xplore. Restrictions apply.

- *Initial phases:* $\boldsymbol{\phi} = [\phi_1, \phi_2, \ldots, \phi_m]^t$;
- *Initial frequencies:* $\boldsymbol{f} = [f_1, f_2, \ldots, f_m]^t$;
- *Slopes:* $\boldsymbol{s} = [s_1, s_2, \ldots, s_m]^t$;
- *Noise:* $\boldsymbol{\epsilon} = [\boldsymbol{\epsilon}[0], \boldsymbol{\epsilon}[1], \ldots, \boldsymbol{\epsilon}[N-1]]^t$.

Using these notations, (21) can be written in matrix form (following the same approach as in [7])

$$\boldsymbol{x} = \boldsymbol{D}\boldsymbol{a} + \boldsymbol{\epsilon} \tag{22}$$

where $\boldsymbol{D} = [\boldsymbol{D}_1, \ldots, \boldsymbol{D}_m]$, and $\boldsymbol{D}_k = [D_k[0], \ldots, D_k[N-1]]^t$. Consequently, the likelihood of $\mathbf{x}$ is expressed as

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = \frac{1}{(2\pi\sigma_\epsilon^2)^{N/2}} \exp\left\{-\frac{1}{2\sigma_\epsilon^2}(\mathbf{x} - \boldsymbol{D}\boldsymbol{a})^t(\mathbf{x} - \boldsymbol{D}\boldsymbol{a})\right\} \tag{23}$$

i.e., $p(\boldsymbol{x}|\boldsymbol{\theta}) \sim \mathcal{N}\left(\boldsymbol{D}a, \sigma_\epsilon^{2N} I_N\right)$, where $I_N$ is the $N \times N$ identity matrix, and $\boldsymbol{\theta} = [\boldsymbol{a}^t, \boldsymbol{\phi}^t, \boldsymbol{f}^t, \boldsymbol{s}^t, \sigma_\epsilon^2]^t$ is the unknown parameter vector. Using Bayes rule, the parameter posterior distribution can be computed as follows:

$$\begin{aligned} p(\boldsymbol{\theta}|\boldsymbol{x}) &= p(\boldsymbol{a}, \boldsymbol{\phi}, \boldsymbol{f}, \boldsymbol{s}, \sigma_\epsilon^2|\boldsymbol{x}) \\ &\propto p(\boldsymbol{x}|\boldsymbol{a}, \boldsymbol{\phi}, \boldsymbol{f}, \boldsymbol{s}, \sigma_\epsilon^2)p(\boldsymbol{a}, \boldsymbol{\phi}, \boldsymbol{f}, \boldsymbol{s}, \sigma_\epsilon^2) \end{aligned} \tag{24}$$

which requires specification of appropriate parameters priors.

### A. Parameter Priors

Following [7], the following hierarchical structure has been chosen for the parameter priors:

$$p(\boldsymbol{a}, \boldsymbol{\phi}, \boldsymbol{f}, \mathrm{s}, \sigma_\epsilon^2) = p(\boldsymbol{a}|\boldsymbol{f}, \sigma_\epsilon^2)p(\boldsymbol{s}|\boldsymbol{f})p(\boldsymbol{f})p(\boldsymbol{\phi})p(\sigma_\epsilon^2). \tag{25}$$

This natural hierarchical structure (see [7] and [22] for a similar choice) can be motivated as follows.

- The amplitude vector $\boldsymbol{a}$ and the additive noise level $\sigma_\epsilon^2$ are closely related since they both determine the amplitude of the observed signal as well as the SNR. Moreover, the chirp components are labeled by their initial frequency, which requires the knowledge of $\boldsymbol{f}$ (or at least, its ordering), such as to assign each chirp component its corresponding amplitude. Based on these comments, the amplitude prior of the form $p(\boldsymbol{a}|\boldsymbol{f}, \sigma_\epsilon^2)$ has been chosen.
- Similarly, the knowledge of $\boldsymbol{f}$ is required to assign each chirp component its corresponding slope, which explains that the slope prior depends on the frequencies, i.e., $p(\boldsymbol{s}|\boldsymbol{f})$.

Some comments are now appropriate before specifying the parameter priors chosen for the classification of chirp signals. The principle of HBL consists of learning the hyperparameter posterior distribution for each class from the training samples. When combined with the selected parameter prior distribution, this leads to the following *parameter posterior distribution*:

$$p(\theta|\mathcal{X}) = \int p(\theta|\Phi)p(\Phi|\mathcal{X})\, d\Phi. \tag{26}$$

This equation shows that implementing an effective learning procedure requires a choice of hyperparametered priors $p(\theta|\Phi)$ for every parameter that may be discriminant (i.e., which provides helpful information for the classification of the current observation signal). As an example, the initial phases in (21) are not discriminant. As a consequence, the corresponding prior distributions do not depend on hyperparameters. On the contrary, the chirp amplitudes, frequencies, and slopes are discriminant in most chirp classification problems. Consequently, the corresponding priors depend on hyperparameters. Moreover, among all possible prior distributions, it is interesting to select pdfs that can be either vague (in the case where the parameter values are quite different from one learning signal to another) or very informative (when the parameter values are similar from one learning signal to another). A typical example of such a prior is the Gaussian distribution that is parametrized by its mean and variance and can be vague or informative, depending on the value of its variance. Other examples of appropriate priors include the inverse Gamma distribution or the uniform distribution defined on an interval whose bounds are hyperparameters.

In the proposed HBL implementation, the following priors have been chosen.

- The initial phases are independent and uniformly distributed in $[0, 1]$:

$$p(\boldsymbol{\phi}) = \prod_{k=1}^{m} \mathbb{I}_{[0,1]}(\boldsymbol{\phi}[k]) = \mathbb{I}_{[0,1]^m}(\boldsymbol{\phi}) \tag{27}$$

  where $\mathbb{I}_A(u) = 1$ if $u \in A$ and $\mathbb{I}_A(u) = 0$ else. Since the phases are not discriminant in our classification problem, the phase priors have been chosen as uniformative uniform priors.
- The slope prior distribution is defined as

$$\begin{aligned} &p(\boldsymbol{s}|\boldsymbol{f}, \boldsymbol{s}_{\min}, \boldsymbol{s}_{\max}) \\ &= \prod_{k=1}^{m} \frac{1}{\boldsymbol{s}_{\max}[k] - \boldsymbol{s}_{\min}[k]} \mathbb{I}_{[\boldsymbol{s}_{\min}[k], \boldsymbol{s}_{\max}[k]]}(\boldsymbol{s}[\mathrm{asc}_{\boldsymbol{f}}(k)]) \end{aligned} \tag{28}$$

  where $\mathrm{asc}_{\boldsymbol{f}}(k)$ is a function that rearranges the indices $k$ such that the frequencies $\boldsymbol{f}[\mathrm{asc}_{\boldsymbol{f}}(k)]$ are sorted in ascending order. This sorting operation ensures that $\boldsymbol{s}_{\min}$ and $\boldsymbol{s}_{\max}$ correspond to the same chirp component (labeled by its initial frequency) from one training signal to another (for instance, $s_1$ is the slope associated to the lowest initial frequency $f_1$ and is uniformly distributed on $[\boldsymbol{s}_{\min}[1], \boldsymbol{s}_{\max}[1]]$).
- The initial frequency prior distribution is a truncated Gaussian distribution of mean $\boldsymbol{m}_f$ and variance-covariance matrix $\Sigma_f$

$$\begin{aligned} p(\boldsymbol{f}|\boldsymbol{m}_f, \boldsymbol{\Sigma}_f) &\propto \frac{1}{|\boldsymbol{\Sigma}_f|^{1/2}} \\ &\times \exp\left\{-\tfrac{1}{2}(\boldsymbol{f}' - \boldsymbol{m}_f)^t\boldsymbol{\Sigma}_f^{-1}(\boldsymbol{f}' - \boldsymbol{m}_f)\right\} \mathbb{I}_{[0,0.5]^m}(\boldsymbol{f}) \end{aligned} \tag{29}$$

  where $\boldsymbol{\Sigma}_f$ is a diagonal matrix whose diagonal elements are $\boldsymbol{\sigma}_f^2 = (\boldsymbol{\sigma}_f^2[1], \ldots, \boldsymbol{\sigma}_f^2[m])$, and $\boldsymbol{f}'$ is the reordered vector $\boldsymbol{f}'[k] = \boldsymbol{f}[\mathrm{asc}_{\boldsymbol{f}}(k)]$, $k = \{1, \ldots, m\}$.
- The amplitude prior distribution is normal, of mean $\boldsymbol{m}_a$, and of variance-covariance matrix $\sigma_\epsilon^2\boldsymbol{\Sigma}_a$

$$\begin{aligned} p(\boldsymbol{a}|\boldsymbol{f}, \boldsymbol{m}_a, \sigma_\epsilon^2, \boldsymbol{\Sigma}_a) &= \frac{1}{|2\pi\sigma_\epsilon^2\Sigma_a|^{1/2}} \\ &\times \exp\left[-\frac{1}{2\sigma_\epsilon^2}(\boldsymbol{a}' - \boldsymbol{m}_a)^t\Sigma_a^{-1}(\boldsymbol{a}' - \boldsymbol{m}_a)\right] \end{aligned} \tag{30}$$

Authorized licensed use limited to: INP TOULOUSE. Downloaded on July 19, 2009 at 12:41 from IEEE Xplore. Restrictions apply.

where $\mathbf{\Sigma}_a$ is a diagonal matrix, whose diagonal terms are $\boldsymbol{\sigma}_a^2 = (\boldsymbol{\sigma}_a^2[1], \ldots, \boldsymbol{\sigma}_a^2[m])$, and $\boldsymbol{a}'$ is the reordered vector $\boldsymbol{a}'[k] = \boldsymbol{a}[\mathrm{asc}_{\boldsymbol{f}}(k)]$, $k = \{1, \ldots, m\}$.

As can be seen, the frequency, slope, and amplitude priors are parametrized by hyperparameters that are related to the training samples via (10). This choice is natural since the frequencies, slopes, and amplitudes are discriminant for our classification problem.

- The prior distribution of the additive noise variance $\sigma_\epsilon^2$ is an inverse Gamma distribution with hyperparameters $\alpha_\epsilon/2$ and $\beta_\epsilon/2$ [denoted $p(\sigma_\epsilon^2|\alpha_\epsilon, \beta_\epsilon) \sim \mathcal{IG}(\alpha_\epsilon/2, \beta_\epsilon/2)$]

$$p(\sigma_\epsilon^2|\alpha_\epsilon, \beta_\epsilon) = \frac{(\beta_\epsilon/2)^{\alpha_\epsilon/2}}{\Gamma(\alpha_\epsilon/2)}(\sigma_\epsilon^2)^{-\alpha_\epsilon/2-1}e^{-\beta_\epsilon/2\sigma_\epsilon^2}\mathbb{I}_{\mathbb{R}^+}(\sigma_\epsilon^2) \tag{31}$$

where $\Gamma$ is the Euler Gamma function, and $\mathbb{I}_\Omega(u)$ is the indicator of the set $\Omega$ (1 if $u \in \Omega$, and 0 elsewhere). The inverse Gamma distribution is a conjugate prior, which has been chosen in order to obtain closed-form expressions of the posteriors in the first step of the Gibbs sampler. As explained in [23], the inverse Gamma prior is interesting since vague or specific prior information can be incorporated through suitable choice of $\alpha_\epsilon$ and $\beta_\epsilon$.

*Remark 1:* In estimation problems (see, e.g., [7]), the initial phases $\boldsymbol{\phi}$ in (21) can be included in the amplitudes, leading to the following equivalent model:

$$\boldsymbol{x}[i] = \sum_{k=1}^{m} a_k \cos(2\pi[f_k i + s_k i^2]) + b_k \sin(2\pi[f_k i + s_k i^2]) + \boldsymbol{\epsilon}[i] \qquad i = \{0, \ldots, N-1\} \tag{32}$$

in which both $\boldsymbol{a}$ and $\boldsymbol{b}$ are linear coefficients. This solution avoids consideration of the nonlinear initial phase coefficients. However, this model is not appropriate for our classification problem. Indeed, the chirp amplitude components have Gaussian priors, which implicitly assumes that the amplitudes are similar for every signal, whereas the initial phase can be very different from one signal to another. By using the model (32), the parameters $a_k$ and $b_k$ may vary significantly from one signal to another (depending on the value of $\phi_k$), even if the amplitude is similar for all signals. This precludes the selected Gaussian modeling and, as a consequence, the use of the model (32).

### B. Hyperparameter Priors

The previous Bayesian model for the classification of chirps is specified by the following hyperparameter vector:

$$\boldsymbol{\Phi} = [\boldsymbol{m}_f^t, \boldsymbol{\sigma}_f^{2^t}, \boldsymbol{m}_a^t, \boldsymbol{\sigma}_a^{2^t}, \boldsymbol{s}_{\min}^t, \boldsymbol{s}_{\max}^t, \alpha_\epsilon, \beta_\epsilon]^t. \tag{33}$$

Of course, the hyperparameter vector $\boldsymbol{\Phi}$ is unknown in practical applications. Following the hierarchical Bayesian model, the hyperparameters are assumed to be random, whose priors (displayed in Table IV) are summarized in $p(\boldsymbol{\Phi})$:

$$p(\boldsymbol{\Phi}) = p(\boldsymbol{m}_f)\ p(\boldsymbol{\sigma}_f^2)p(\boldsymbol{m}_a)p(\boldsymbol{\sigma}_a^2)p(\alpha_\epsilon)p(\beta_\epsilon) \\ p(\boldsymbol{s}_{\max}|\mathrm{s}_{\min})p(\boldsymbol{s}_{\min}) \tag{34}$$

TABLE IV
MULTICOMPONENT SIGNAL CLASSIFICATION: HYPERPARAMETER PRIORS

| Hyperparameter | Distribution | Distribution expression (proportional to) |
|---|---|---|
| $\boldsymbol{m}_f$ | Uniform | $\mathcal{U}([0, 0.5]^m)$ |
| $\boldsymbol{\sigma}_f^2$ | Jeffrey's | $\prod_{k=1}^{m} \frac{1}{\boldsymbol{\sigma}_f^2[k]}$ |
| $\boldsymbol{m}_a$ | Uniform | $\mathcal{U}([0, a_{\max}]^m)$, with $a_{max} = 10$ |
| $\boldsymbol{\sigma}_a^2$ | Jeffrey's | $\prod_{k=1}^{m} \frac{1}{\boldsymbol{\sigma}_a^2[k]}$ |
| $\boldsymbol{s}_{\min}$ | Uniform | $\mathcal{U}([s_{\inf}, s_{\sup}]^m)$ with $s_{\inf} = -0.5/2N, s_{\sup} = 0.5/2N$ |
| $\boldsymbol{s}_{\max} - \boldsymbol{s}_{\min}$ | Truncated Jeffrey's | $\prod_{k=1}^{m} (\boldsymbol{s}_{\max}[k] - \boldsymbol{s}_{\min}[k])^{-1}\ \mathbb{I}_{[\boldsymbol{s}_{\min}[k],\, s_{\sup}]}(\boldsymbol{s}_{\max}[k])$ |
| $\alpha_\epsilon$ | Uniform | $\mathcal{U}([0, \alpha_{\max}]$ with $\alpha_{max} = 200$ |
| $\beta_\epsilon$ | Gamma | $\beta_\epsilon^{\nu_0 - 1}e^{-\gamma_0\beta_\epsilon}$ with $\nu_0 = 2$, $\gamma_0 = 2$ |

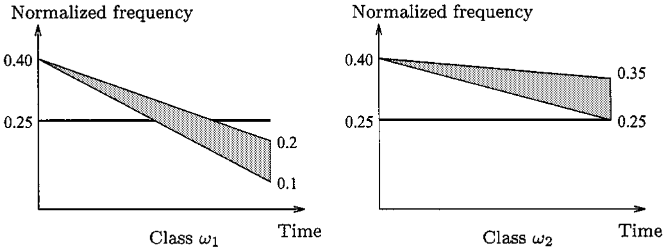


Fig. 2. Example 1: Instantaneous frequencies for the second chirp component (gray areas).

where

- $p(\boldsymbol{\sigma}_f^2)$ and $p(\boldsymbol{\sigma}_a^2)$ are the products of independent Jeffreys' priors [24, p. 44];
- $\boldsymbol{m}_f$, $\boldsymbol{m}_a$ and $\alpha_\epsilon$ have uniform priors on $[0, 0.5]^m$, $[0, a_{\max}]^m$, $[0, \alpha_{\max}]$;
- the components of $\boldsymbol{s}_{\min}$ and $\boldsymbol{s}_{\max}$ are independent and uniformly distributed on appropriate intervals (i.e., $[s_{\inf}, s_{\sup}]$ and $[\boldsymbol{s}_{\min}[k], s_{\sup}]$), ensuring $\boldsymbol{s}_{\max}[k] \geq \boldsymbol{s}_{\min}[k]$ for all $k = \{1, \ldots, m\}$;
- the prior of $\beta_\epsilon$ is a Gamma distribution with parameters $\nu_0$ and $\gamma_0$.

The hyperparameters $a_{\max}$, $s_{\inf}$, $s_{\sup}$, $\alpha_{\max}$, $\nu_0$, and $\gamma_0$ have been chosen in order to provide vague prior information. However, the learning procedure is robust to the values of these hyperparameters (as outlined in [7] in a similar context).

*Remark 2:* It is interesting to note that an alternative approach called *empirical Bayes analysis* [5, p. 307] could also be used for the classification of chirps. Empirical Bayes analysis consists of estimating the unknown hyperparameters from the observed data by using an appropriate estimation technique such as the moment method or the maximum likelihood method. The unknown hyperparameters are then replaced by their estimated values in the Bayesian model. However, empirical Bayes analysis may suffer from several problems. In particular, the derivation of the estimator is generally too complicated, and the approximated posterior obtained after replacing the hyperparameters by their estimates is only acceptable for large sample sizes (see [5, p. 309] for more details). These problems have motivated our choice of implementing HBL instead of empirical Bayes analysis.

### C. HBL Implementation Using MCMC Methods

As outlined before, the first part of the HBL implementation using MCMC methods consists of generating vectors $(\tilde{\boldsymbol{\theta}}_{1l}, \ldots, \tilde{\boldsymbol{\theta}}_{nl}, \tilde{\boldsymbol{\Phi}}_l)$ for $l = \{1, \ldots, L\}$ distributed according to $p(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n, \boldsymbol{\Phi}|\mathcal{X})$. For the classification of chirps, the

Authorized licensed use limited to: INP TOULOUSE. Downloaded on July 19, 2009 at 12:41 from IEEE Xplore. Restrictions apply.

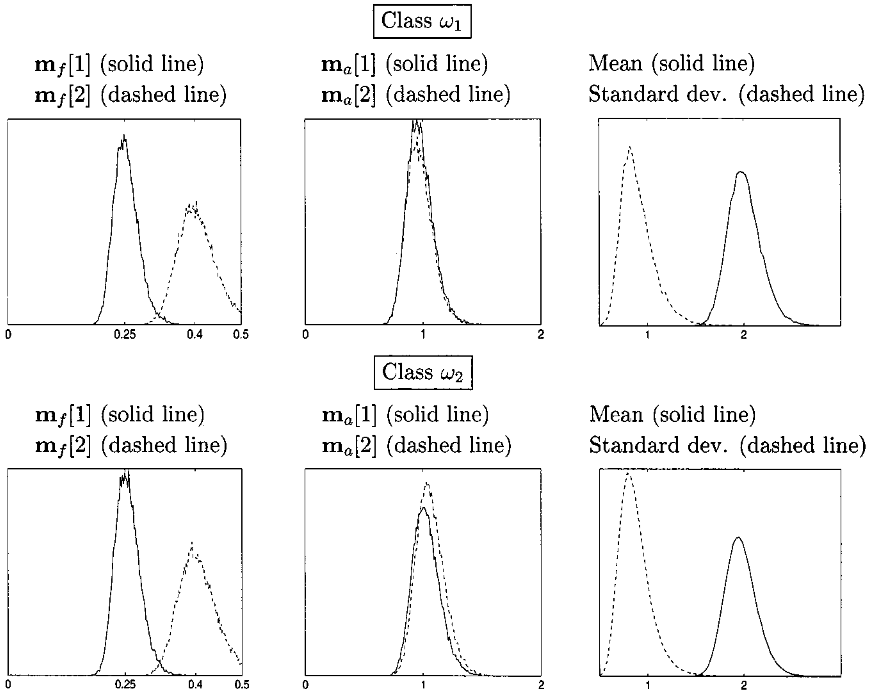


Fig. 3. Example 1: Histograms of MC samples distributed according to $p(\boldsymbol{m}_f(1)|\mathcal{X})$, $p(\boldsymbol{m}_f(2)|\mathcal{X})$, $p(\boldsymbol{m}_a(1)|\mathcal{X})$, $p(\boldsymbol{m}_a(2)|\mathcal{X})$, $p(\beta_c/(\alpha_c-2)|\mathcal{X})$ (mean of the inverse Gamma distribution of parameters $\alpha_c/2$ and $\beta_c/2$), and $p([\beta_c^2/((\alpha_c-2)^2(\alpha_c/2-2))]^{1/2}|\mathcal{X})$ (standard deviation) for classes $\omega_1$ and $\omega_2$.

generation is conducted by using the following one-variable-at-a-time MH algorithm (see Appendix B for more details):

- *First Step of the Sampler:* Generation of $L$ vectors $(\tilde{\boldsymbol{\theta}}_{1l}, \ldots, \tilde{\boldsymbol{\theta}}_{nl})$, $l = \{1, \ldots, L\}$, distributed according to $p(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n|\tilde{\boldsymbol{\Phi}}_{l-1}, \mathcal{X})$. Using the independence of training samples $\boldsymbol{x}_j$, this step can be decomposed into $n$ independent generations of $\tilde{\boldsymbol{\theta}}_{jl}$ distributed according to $p(\boldsymbol{\theta}_j|\tilde{\boldsymbol{\Phi}}_{l-1}, \boldsymbol{x}_j)$. Given a training signal $\boldsymbol{x}_j$ written in vector form, the selected priors lead to (the subscript $j$ is forgotten for simplicity)

$$
\begin{aligned}
p(\boldsymbol{a}, \boldsymbol{\phi}, \boldsymbol{f}, \boldsymbol{s}, \sigma_\epsilon^2|\boldsymbol{x}, \Phi)
&\propto (\sigma_\epsilon^2)^{-((N+\alpha_c)/2)-1} \exp\left\{-\frac{1}{2\sigma_\epsilon^2}(\boldsymbol{a}'-\boldsymbol{m})^t\boldsymbol{M}^{-1}(\boldsymbol{a}'-\boldsymbol{m})\right\}\\
&\times \frac{1}{|\sigma_\epsilon^2\boldsymbol{\Sigma}_a|^{1/2}}\\
&\times \exp\left\{-\frac{1}{2\sigma_\epsilon^2}(\boldsymbol{x}^t\boldsymbol{x}+\boldsymbol{m}_a^t\boldsymbol{\Sigma}_a{}^{-1}\boldsymbol{m}_a-\boldsymbol{m}^t\boldsymbol{M}^{-1}\boldsymbol{m})\right\}\\
&\times \frac{1}{|\boldsymbol{\Sigma}_f|^{1/2}}\\
&\times \exp\left\{-\frac{1}{2}(\boldsymbol{f}'-\boldsymbol{m}_f)^t\boldsymbol{\Sigma}_f^{-1}(\boldsymbol{f}'-\boldsymbol{m}_f)\right\}\\
&\times \prod_{k=1}^{m}\frac{1}{\boldsymbol{s}_{\max}[k]-\boldsymbol{s}_{\min}[k]}\mathbb{1}_{[\boldsymbol{s}_{\min}[k],\, \boldsymbol{s}_{\max}[k]]}(\boldsymbol{s}[\mathrm{asc}_{\boldsymbol{f}}(k)])\\
&\times e^{-\beta_c/2\sigma_\epsilon^2}\mathbb{1}_{[0,\,0.5]^m}(\boldsymbol{f})\mathbb{1}_{[0,\,1]^m}(\boldsymbol{\phi}) \qquad (35)
\end{aligned}
$$

where $\boldsymbol{M}^{-1} = \boldsymbol{D}^t\boldsymbol{D}+\boldsymbol{\Sigma}_a^{-1}$, and $\boldsymbol{m} = \boldsymbol{M}[\boldsymbol{D}^t\boldsymbol{x}+\boldsymbol{\Sigma}_a^{-1}\boldsymbol{m}_a]$. Following the same approach as in [7], the parameters $\boldsymbol{a}$ and $\sigma_\epsilon^2$ are integrated out analytically. As a consequence, the generation of $\tilde{\boldsymbol{\theta}}_l = (\tilde{\boldsymbol{a}}, \tilde{\boldsymbol{\phi}}, \tilde{\boldsymbol{f}}, \tilde{\boldsymbol{s}}, \tilde{\sigma}_\epsilon^2)$ is decomposed into the following three moves for numerical efficiency (see Appendix B for details):

1) Generation of $(\tilde{\boldsymbol{\phi}}, \tilde{\boldsymbol{f}}, \tilde{\boldsymbol{s}})$ distributed according to

$$
\begin{aligned}
p(\boldsymbol{\phi}, \boldsymbol{f}, \boldsymbol{s}|\boldsymbol{x}, \boldsymbol{\Phi})
&\propto |\boldsymbol{M}|^{1/2}\exp\left\{-\tfrac{1}{2}(\boldsymbol{f}'-\boldsymbol{m}_f)^t\boldsymbol{\Sigma}_f^{-1}(\boldsymbol{f}'-\boldsymbol{m}_f)\right\}\\
&\times \prod_{k=1}^{m}\frac{1}{\boldsymbol{s}_{\max}[k]-\boldsymbol{s}_{\min}[k]}\mathbb{1}_{[\boldsymbol{s}_{\min}[k],\, \boldsymbol{s}_{\max}[k]]}(\boldsymbol{s}[\mathrm{asc}_{\boldsymbol{f}}(k)])\\
&\times (\boldsymbol{x}^t\boldsymbol{x}+\boldsymbol{m}_a^t\boldsymbol{\Sigma}_a^{-1}\boldsymbol{m}_a-\boldsymbol{m}^t\boldsymbol{M}^{-1}\boldsymbol{m}+\beta_\epsilon)^{-(N+\alpha_c)/2}\\
&\times \mathbb{1}_{[0,\,0.5]^m}(\boldsymbol{f})\mathbb{1}_{[0,\,1]^m}(\phi) \qquad (36)
\end{aligned}
$$

using Metropolis Hastings steps (see Appendix B for details).

2) Given $(\tilde{\boldsymbol{\phi}}, \tilde{\boldsymbol{f}}, \tilde{\boldsymbol{s}})$, sample the noise variance $\tilde{\sigma}_\epsilon^2$ distributed according to the inverse Gamma distribution

$$
\begin{aligned}
p\left(\sigma_\epsilon^2|\boldsymbol{x}, \tilde{\boldsymbol{\phi}}, \tilde{\boldsymbol{f}}, \tilde{\boldsymbol{s}}, \boldsymbol{\Phi}\right)
&\sim \mathcal{IG}\left(\frac{N+\alpha_\epsilon}{2}, \frac{\boldsymbol{x}^t\boldsymbol{x}+\boldsymbol{m}_a^t\boldsymbol{\Sigma}_a{}^{-1}\boldsymbol{m}_a-\boldsymbol{m}^t\boldsymbol{M}^{-1}\boldsymbol{m}+\beta_\epsilon}{2}\right). \qquad (37)
\end{aligned}
$$

3) Given all the other parameters, sample the amplitude vector distributed according to

$$
p(\boldsymbol{a}|\boldsymbol{x}, \tilde{\boldsymbol{\phi}}, \tilde{\boldsymbol{f}}, \tilde{\boldsymbol{s}}, \tilde{\sigma}_\epsilon^2, \boldsymbol{\Phi}) \sim \mathcal{N}(\mathbf{m}, \tilde{\sigma}_\epsilon^2\,\boldsymbol{M}). \qquad (38)
$$

- *Second Step of the Sampler:* Using $\{\tilde{\boldsymbol{\theta}}_{1l}, \ldots, \tilde{\boldsymbol{\theta}}_{nl}\}$ obtained during the previous step, generation of one vector $\tilde{\boldsymbol{\Phi}}_l$ distributed according to $p(\boldsymbol{\Phi}|\tilde{\boldsymbol{\theta}}_{1l}, \ldots, \tilde{\boldsymbol{\theta}}_{nl}, \mathcal{X}) = p(\boldsymbol{\Phi}|\tilde{\boldsymbol{\theta}}_{1l}, \ldots, \tilde{\boldsymbol{\theta}}_{nl})$. This requires to compute the hyperparameter posterior distribution $p(\boldsymbol{\Phi}|\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n)$ (see Appendix C for details).

Authorized licensed use limited to: INP TOULOUSE. Downloaded on July 19, 2009 at 12:41 from IEEE Xplore. Restrictions apply.

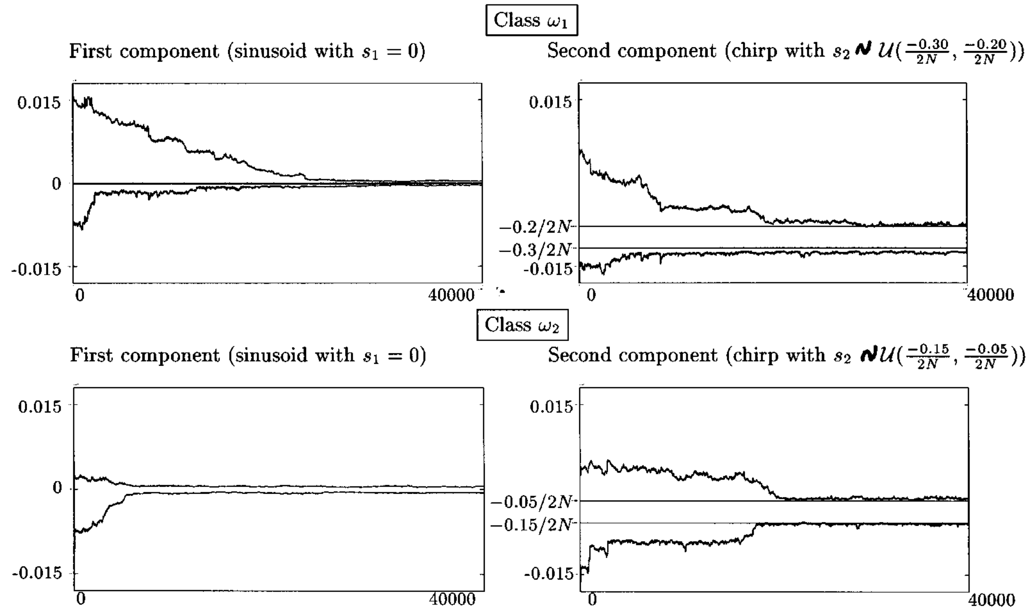


Fig. 4. Example 1: MC samples distributed according to $p(\boldsymbol{s}_{\min}[1]|\mathcal{X})$, $p(\boldsymbol{s}_{\max}[1]|\mathcal{X})$ (figures left), and $p(\boldsymbol{s}_{\min}[2]|\mathcal{X})$, $p(\boldsymbol{s}_{\max}[2]|\mathcal{X})$ (figures right) for class $\omega_1$ and class $\omega_2$.

The second part of the HBL implementation for the classification of chirps consists of generating one sample $\tilde{\boldsymbol{\theta}}_l$ distributed according to $p(\boldsymbol{\theta}|\tilde{\boldsymbol{\Phi}}_l)$ (where $\tilde{\boldsymbol{\Phi}}_l$ has been obtained during the first part of the algorithm). This leads to the sampling procedure summarized in Appendix B.

### D. Simulation Results

The performance of the proposed HBL implementation for the classification of chirps has been studied via several simulation results. This paper focuses on two examples:

1) *Example 1—Radar Target Identification:* We assume that the signal observed by the receiver is two-component: one tone, emitted by an harmonic jammer, and one chirp, corresponding to the signal reflected by the target. Moreover, an additive ambient noise corrupts the observations. A sensible model is then obtained as the sum of one chirp and one tone embedded in additive white Gaussian noise, namely

$$\boldsymbol{x}[i] = \sum_{k=1}^{2} a_k \cos(2\pi[\phi_k + f_k i + s_k i^2]) + \boldsymbol{\epsilon}[i] \qquad (39)$$

where $i = \{0, \ldots, N-1\}$, $\boldsymbol{\epsilon}[i] \sim \mathcal{N}(0, \sigma_\epsilon^2)$, and $N = 128$, $\sigma_\epsilon^2 = 2$ [the signal to noise ratio is $\text{SNR} = 10\log_{10}(a_1^2 + a_2^2)/\sigma_\epsilon^2 = 0$ dB]. The signal amplitudes and initial frequencies are $a_1 = a_2 = 1$, $f_1 = 0.25$, $f_2 = 0.4$. The initial phases $\phi_1$ and $\phi_2$ are independent and uniformly distributed in [0, 1]. The slope of the first component is $s_1 = 0$, which corresponds to a pure harmonic. The problem addressed in [12] consists of classifying a given signal $\boldsymbol{x}[i]$ into one of the two classes defined as follows:

1) Class $\omega_1$: $p(s_2) = \mathcal{U}(-0.30/2N, -0.20/2N)$;
2) Class $\omega_2$: $p(s_2) = \mathcal{U}(-0.15/2N, -0.05/2N)$;

where $\mathcal{U}(a, b)$ is the uniform distribution on $[a, b]$. These classes correspond to targets having the same speed but different accelerations (or the same range and different speeds).

TABLE V
EXAMPLE 1: ERROR RATES FOR THE CLASSIFICATION OF CHIRPS USING THE PROPOSED HBL IMPLEMENTATION AND NONPARAMETRIC TIME–FREQUENCY CLASSIFIERS (THE LEARNING SET IS COMPOSED OF $n_1 = n_2$ SIGNALS FOR EACH CLASS, WHEREAS THE TEST SET IS COMPOSED OF 20 000 SIGNALS WITH 10 000 FOR EACH CLASS)

| | Number of training samples $n_1 = n_2 =$ | | | |
|---|---|---|---|---|
| Classification method | 50 | 100 | 200 | 500 |
| Wigner distribution [25] | 22.30 % | 19.39 % | 12.53 % | 5.38 % |
| Ambiguity plane [26] | 4.56 % | 3.84 % | 2.68 % | 2.16 % |
| Time-Frequency [12] | 2.25 % | 1.91 % | 1.85 % | 1.64 % |
| HBL implementation [this paper] | 5.24 % | 2.76 % | 1.42 % | 0.85 % |

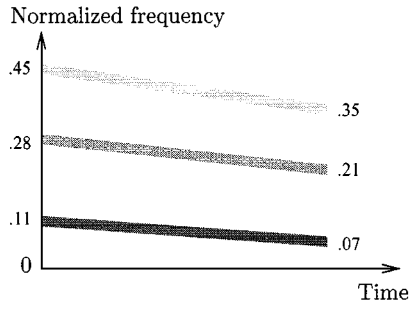


Fig. 5. Example 2: Instantaneous frequencies for the three chirp components (both classes).

The possible instantaneous frequencies of the chirp component are plotted in Fig. 2 for classes $\omega_1$ and $\omega_2$ (gray areas). Note that the two classes only differ by the slope of the second chirp component and that the jammer and target frequencies overlap. In order to simulate a realistic situation, we assume that the signal model parameters are unknown. However, the model given by (39) is known (it can be derived from a nonparametric analysis; see [12]).

The first simulation results illustrate the performance of the *learning step*. Fig. 3 displays histograms of Markov chain samples distributed according to the posterior distributions of hyperparameters $\boldsymbol{m}_f$, $\boldsymbol{m}_a$, $\beta_\epsilon/(\alpha_\epsilon - 2)$, and

Authorized licensed use limited to: INP TOULOUSE. Downloaded on July 19, 2009 at 12:41 from IEEE Xplore. Restrictions apply.

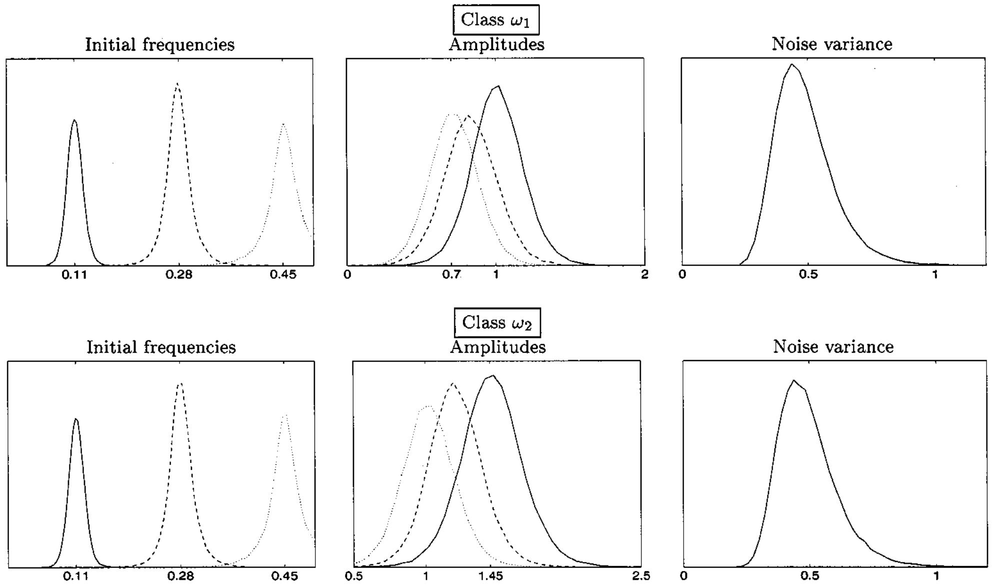


Fig. 6. Example 2: Histograms of the MC samples distributed according to $p(\boldsymbol{\theta}|\mathcal{X})$ for some components of $\boldsymbol{\theta}$. Left column: $p(f_1|\mathcal{X})$ (dotted line), $p(f_2|\mathcal{X})$ (dashed line), and $p(f_3|\mathcal{X})$ (solid line). Middle column: $p(a_1|\mathcal{X})$ (dotted line), $p(a_2|\mathcal{X})$ (dashed line), and p$(a_3|\mathcal{X})$ (solid line). Right column: $p(\sigma_\epsilon^2|\mathcal{X})$.

$\beta_\epsilon^2/((\alpha_\epsilon-2)^2(\alpha_\epsilon/2-2))]^{1/2}$. These histograms are clearly in good agreement with the true values of parameters $f_1$, $f_2$, $a_1$ and $a_2$. The histograms displaying $\beta_\epsilon/(\alpha_\epsilon-2)$ (mean of the inverse Gamma distribution of parameters $\alpha_\epsilon/2$ and $\beta_\epsilon/2$) and $[\beta_\epsilon^2/((\alpha_\epsilon-2)(\alpha_\epsilon/2-2))]^{1/2}$ (standard deviation of this inverse Gamma distribution) are also coherent with the true value of $\sigma_\epsilon^2$. MC samples drawn from the posterior distributions $p(\mathbf{s}_{\min}[1]|\mathcal{X})$, $p(\boldsymbol{s}_{\max}[1]|\mathcal{X})$, $p(\boldsymbol{s}_{\min}[2]|\mathcal{X})$, and $p(\mathbf{s}_{\max}[2]|\mathcal{X})$ of class $\omega_1$ are shown in the upper row of Fig. 4. Similar results associated with class $\omega_2$ are shown in the lower row of Fig. 4. These plots show that 1) $\mathbf{s}_{\min}[1]$ and $\boldsymbol{s}_{\max}[1]$ converge to 0 for classes $\omega_1$ and $\omega_2$, and 2) $[\mathbf{s}_{\min}[2], \boldsymbol{s}_{\max}[2]]$ converge to values close to $[-0.30/2N, -0.20/2N]$ for class $\omega_1$ and $[-0.15/2N, -0.05/2N]$ for class $\omega_2$. These results are in good agreement with the definition of $s_1$ and $s_2$ for both classes. Consequently, the learning step of the HBL MCMC implementation shows good performance for the classification of chirps (in particular, the minimum and maximum values of the slope $s_2$, which are the discriminant features, have been learned with good accuracy).

During the *test step*, 20 000 test vectors (i.e., 10 000 test vectors from each class different from the training vectors) are classified using the MCMC implementation of HBL. Table V displays the corresponding error rates, which are compared with those obtained with classifiers based on time-frequency representations [12], [25], [26]. For a large number of training vectors, the MCMC method outperforms nonparametric time–frequency classifiers and provides a reference to which suboptimal classifiers can be compared. This result can be explained as follows: 1) The Bayesian classifier is optimal when the prior probabilities and the class-conditional densities are known, and 2) the observation predictive distributions estimated with the proposed HBL implementation converge to the true predictive distributions (according to the ergodic theorem) when the number of MC samples increases. However, for a small training set, the time–frequency classifiers outperform the proposed Bayesian learning MCMC implementation. This is mainly due to the complexity of the implicit model used for the classification of chirps. Due to this complexity, the Bayesian classifier requires a large number of training samples to learn the class-conditional distributions. In contrast, the model used for the classification of chirps is well suited to time–frequency classifiers since energy in the time–frequency plane is well localized for each class.

*2) Example 2—Knock Detection:* This second example addresses the knock detection problem in car engines (see [20] and [21] for an overview and a time-frequency detection algorithm). The recorded vibration signal is modeled as the sum of three chirps embedded in additive white Gaussian noise

$$\boldsymbol{x}[i]=\sum_{k=1}^{3} a_k\cos(2\pi[\phi_k+f_k i+s_k i^2])+\boldsymbol{\epsilon}[i] \qquad (40)$$

where $i=\{0,\ldots,N-1\}, N=64, \boldsymbol{\epsilon}[i]\sim\mathcal{N}(0,\sigma_\epsilon^2)$, and $\sigma_\epsilon^2=0.5$. The instantaneous frequencies of each chirp component are the same in classes $\omega_1$ and $\omega_2$ and are defined as $f_1\sim\mathcal{N}(0.45, 10^{-4}), f_2\sim\mathcal{N}(0.28, 10^{-4}), f_3\sim\mathcal{N}(0.11, 10^{-4})$, $s_1=-0.1/2N, s_2=-0.07/2N$, and $s_3=-0.04/2N$ (see also Fig. 5). The initial phases $\phi_1, \phi_2$, and $\phi_3$ are independent and uniformly distributed in $[0, 1]$. In this example, the presence or absence of knock is characterized by different chirp amplitudes:

- Class $\omega_1$ (absence of knock): $p(a_1)=\mathcal{N}(0.7, 0.0025)$, $p(a_2)=\mathcal{N}(0.82, 0.0025), p(a_3)=\mathcal{N}(1, 0.0025)$;
- Class $\omega_2$ (presence of knock): $p(a_1)=\mathcal{N}(1, 0.0025)$, $p(a_2)=\mathcal{N}(1.2, 0.0025), p(a_3)=\mathcal{N}(1.45, 0.0025)$.

The problem addressed in this example is the classification of an observed vector $\boldsymbol{x}$ in one of the two classes $\omega_1$ and $\omega_2$. The first simulations illustrate the performance of the *learning step* with $n_1=n_2=200$. Fig. 6 displays histograms of MC samples distributed according to the posterior distributions of frequencies, amplitudes, and noise variance. As can be seen, parameters have been learned with good accuracy. During the *test step*, 20 000 test vectors (i.e., 10 000 test vectors from each class different from the learning vectors) are classified using the proposed HBL implementation. The estimated error rate is

Authorized licensed use limited to: INP TOULOUSE. Downloaded on July 19, 2009 at 12:41 from IEEE Xplore. Restrictions apply.

TABLE VI
EXAMPLE 2: ERROR RATES FOR THE CLASSIFICATION OF CHIRPS USING THE PROPOSED HBL IMPLEMENTATION AND NONPARAMETRIC TIME–FREQUENCY CLASSIFIERS (THE LEARNING SET IS COMPOSED OF $n_1 = n_2 = 200$ SIGNAL FOR EACH CLASS, WHEREAS THE TEST SET IS COMPOSED OF 20 000 SIGNALS WITH 10 000 FOR EACH CLASS)

| Classification method | | Probability of error |
|---|---|---|
| Wigner distribution | [25] | ≈ 50 % |
| Ambiguity plane | [26] | ≈ 50 % |
| Time-Frequency | [12] | 1.27 % |
| HBL implementation [this paper] | | 21.72 % |

compared in Table VI with those obtained with the previous time–frequency classifiers. As can be seen, the proposed HBL implementation outperforms the classifiers based on the Wigner distribution and the ambiguity plane for the knock detection problem. However, the optimized time–frequency classifier proposed in [12] and [13] performs considerably better for this problem. This result indicates that more than 200 learning samples are necessary to obtain a good approximation of the predictive distributions, i.e., to achieve the optimal Bayesian classifier performance.

Simulation results presented in this section have shown the good performance of the proposed MCMC-based HBL implementation for the classification of chirp signals. Two different scenarios have been studied, for which chirp slopes and chirp amplitudes are the respective discriminant parameters. This shows the robustness of the proposed classification procedure to the nature of discriminant parameters. However, it is important to note that the performance of the proposed HBL implementation strongly depends on the number of training samples.

## V. CONCLUSION

This paper showed that MCMC methods, which are widely used for Bayesian estimation, are also a suitable tool for supervised classification using hierarchical Bayesian learning. In the proposed HBL implementation, the training samples were used to build Markov chains whose target distributions were the parameter posterior distributions. The class-conditional probability densities were then obtained by Monte Carlo ergodic averaging of the MC elements. An academic example showed that the performance of the proposed HBL implementation is very close to the theoretical performance obtained with closed-form expressions of the class conditional distributions. Finally, the problem of classifying chirp signals by using HBL combined with MCMC methods was studied. The proposed classifier was shown to outperform conventional time–frequency classifiers, provided the number of training vectors is sufficient to learn the class-conditional distributions.

## APPENDIX A
## FIRST EXAMPLE: HBL ALGORITHM

```
% Initialization
set $l \leftarrow 1$
sample $\tilde{\mu}_1$ according to $p(\mu)$.
for $j = \{1, \ldots, n\}$, do
  sample $\tilde{m}_{j\,1} \sim \mathcal{N}(\tilde{\mu}_1, \sigma_m^2)$
% Generation of the Markov Chain elements
$\tilde{\mu}_l$
while $l < L$, do
  % First step of the Gibbs algorithm
  for each training sample $x_j$ in
  $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$, do
    sample $\tilde{m}_{jl}$ according to $p(m_j|\tilde{\mu}_{l-1}, \boldsymbol{x}_j) \propto$
        $p(\boldsymbol{x}_j|m_j)p(m_j|\tilde{\mu}_{l-1})$
  % Second step of the Gibbs algorithm
  sample $\tilde{\mu}_l$ according to $p(\mu|\tilde{m}_{1l}, \ldots, \tilde{m}_{nl}) \propto$
        $p(\mu)\prod_{j=1}^{n} p(\tilde{m}_{jl}|\mu)$
% Generation of the samples $\tilde{m}_l$
For $l = \{1, \ldots, L\}$, do
  sample $\tilde{m}_l \sim p(m|\tilde{\mu}_l) = \mathcal{N}(\tilde{\mu}_l, \sigma_m^2)$.
```

## APPENDIX B
## HBL ALGORITHM FOR THE CLASSIFICATION OF CHIRPS

This appendix derives the HBL algorithm for the classification of chirps. The main algorithm is first presented. Some subprocedures are then detailed.

```
Main Algorithm
% Hyperparameter initialization
for each chirp component $k$, repeat
  sample $\tilde{\boldsymbol{m}}_{f1}[k] \sim \mathcal{U}[0, 0.5]$
  sample $\tilde{\boldsymbol{\sigma}}_{f1}[k]$ according to the Jeffrey's
  prior
  sample $\tilde{\boldsymbol{m}}_{a1}[k] \sim \mathcal{U}[0, a_{\max}]$ with $a_{\max} = 10$
  sample $\tilde{\boldsymbol{\sigma}}_{a1}[k]$ according to the Jeffrey's
  prior
  sample $\tilde{\boldsymbol{s}}_{\min 1} \sim \mathcal{U}[s_{\inf}, s_{\sup}]$ and $\tilde{\boldsymbol{s}}_{\max 1} \sim$
  $\mathcal{U}[\tilde{\boldsymbol{s}}_{\min 1}, s_{\sup}]$
  set $\tilde{\alpha}_{\epsilon 1} \leftarrow 2$ and $\tilde{\beta}_{\epsilon 1} \leftarrow 2$
% Parameter initialization
for each training signal $x_j$ do
  for each chirp component do
    sample $\tilde{\boldsymbol{\phi}}_{j1}[k] \sim \mathcal{U}[0, 1]$
    sample $\tilde{\boldsymbol{s}}_{j1}[k] \sim \mathcal{U}[s_{\inf}, s_{\sup}]$
  for $k = \{1, \ldots, m\}$, set the frequencies
  $\tilde{\boldsymbol{f}}_{j1}[k]$ ... linearly in $[0, 0.5]$
set $l \leftarrow 1$
% Generation of the Markov chain $\tilde{\boldsymbol{\Phi}}_l$
while $l < L$ do
  % First step of the Gibbs Sampler
  % Generation of samples $\tilde{\boldsymbol{\theta}}_{jl}$
  for each training signal $\boldsymbol{x}_j$, do
    for each chirp component $k$ do
      sample $\tilde{\boldsymbol{\phi}}_{j\,l}[k], \tilde{f}_{j\,l}[k], \tilde{\boldsymbol{s}}_{j\,l}[k]$ according to
  the  pdf (36) using subprocedure 1
    sample $\tilde{\sigma}_{\epsilon jl}^2$ distributed according to
  (37)
    sample $\tilde{\boldsymbol{a}}_{jl}$ distributed according to
  (38)
  % Second step of the Gibbs algorithm
  % Generation of samples $\tilde{\Phi}_l$
  for each chirp component $k$ do
    sample $\tilde{\boldsymbol{m}}_{fl}[k]$ as in (41)
```

Authorized licensed use limited to: INP TOULOUSE. Downloaded on July 19, 2009 at 12:41 from IEEE Xplore. Restrictions apply.

```
    sample $\tilde{\boldsymbol{\sigma}}_{f\,l}[k]$ as in (42)
    sample $\tilde{\boldsymbol{s}}_{\min l}[k], \tilde{\mathrm{s}}_{\max l}[k]$ according to the
      pdf (43) using subprocedure 2
    sample $\tilde{\boldsymbol{m}}_{al}[k]$ as in (44)
    sample $\tilde{\boldsymbol{\sigma}}_{a\,l}[k]$ as in (45)
    sample $\tilde{\alpha}_{\epsilon l}$ according to the pdf (46)
      using subprocedure 3
    sample $\tilde{\beta}_{\epsilon\, l}$ as in eq. (47)
% Generation of samples $\tilde{\boldsymbol{\theta}}_l$
for $l = \{1, \ldots, L\}$, do
  sample $\tilde{\boldsymbol{\theta}}_l$ according to $p(\boldsymbol{\theta}|\tilde{\boldsymbol{\Phi}}_l)$, such as
  defined in Section IV-A.
```

The three subprocedures consist of a Metropolis–Hastings step using either a local random walk proposal distribution or a global proposal distribution. In each subprocedure, the choice of the proposal distribution depends on a parameter $\lambda$. Each subprocedure has the following structure [where the variable to be sampled is denoted $\xi$, the target pdf is $p(\xi)$, the global proposal probability density is denoted $q_1$, and the local random walk proposal probability density is denoted $q_2$].

```
Subprocedures
if $\lambda >$ Rand then
  % the global proposal probability den-
  sity is used
  sample $\xi^\star$ distributed according $q_1(\xi^\star|\tilde{\xi}_{l-1})$
  if $(p(\xi^\star)q_1(\tilde{\xi}_{l-1}|\xi^\star))/(p(\tilde{\xi}_{l-1})q_1(\xi^\star|\tilde{\xi}_{l-1})) >$ Rand
  then
    $\tilde{\xi}_l \leftarrow \xi^\star$
  else
    $\tilde{\xi}_l \leftarrow \tilde{\xi}_{l-1}$
else
  % the local proposal probability
  % density is used (randomn walk)
  sample $v$ according to $q_2$
  set $\xi^\star \leftarrow \tilde{\xi}_{l-1} + v$
  if $p(\xi^\star)/p(\tilde{\xi}_{l-1}) >$ Rand
  then
    $\tilde{\xi}_l \leftarrow \xi^\star$
  else
    $\tilde{\xi}_l \leftarrow \tilde{\xi}_{l-1}$
```

In the three subprocedures, $\lambda = 0.2$, and $q_2$ is the pdf of a zero-mean Gaussian vector with diagonal variance-covariance matrix $\boldsymbol{\Sigma}_{q_2}$. In subprocedure 1, the diagonal elements of $\boldsymbol{\Sigma}_{q_2}$ are $(0.002, 0.001, 0.005/2N)$, corresponding to $(\boldsymbol{\phi}[k], \boldsymbol{f}[k], \boldsymbol{s}[k])$. In subprocedure 2, the elements of $\boldsymbol{\Sigma}_{q_2}$ are $(0.001/2N, 0.001/2N)$, corresponding to $(\boldsymbol{s}_{\min}[k], \boldsymbol{s}_{\max}[k])$. Finally, $\boldsymbol{\Sigma}_{q_2} = 0.5$ in subprocedure 3. These values ensure an acceptance rate of about 25%. (Note that the convergence rate of the algorithm may depend on these parameters, as opposed to the Markov Chain invariant distribution.) The proposal probability densities $q_1$ are the following.

- *Subprocedure 1:* The phase is sampled uniformly in [0, 1], whereas the frequency and slope parameters are jointly sampled from a stepwise constant distribution, i.e., $(\phi^\star, f^\star, s^\star) \sim (\mathcal{U}[0;\ 1], q_1'(f, s))$, where $q_1'(f, s)$ is obtained by discretizing the function
$$q_1'(f, s) \propto \mathcal{S}_{\text{begin}}(f)\mathcal{S}_{\text{end}}(f + 2Ns)$$
where $\mathcal{S}_{\text{begin}}(f)$ denotes the frequency spectrum of the $N/4$ first points of the signal, and $\mathcal{S}_{\text{end}}(f)$ is the frequency spectrum of the $N/4$ last points of the signal. One has, e.g.,
$$\mathcal{S}_{\text{begin}}(f) = \frac{|[fft(\boldsymbol{x}_{\text{begin}})](f)|^2}{N/4}$$
and $\boldsymbol{x}_{\text{begin}} = [\boldsymbol{x}[0], \ldots, \boldsymbol{x}[N/4 - 1]]$.
- *Subprocedure 2:* The proposal probability density $q_1$ is the uniform distribution $\mathcal{U}([s_{b_1}, s_{b_2}] \times [s_{b_3}, s_{b_4}])$ with $b_1 = s_{\inf}$, $b_2 = \min(s_{\sup}, \boldsymbol{s}_{\min}[1], \ldots, \boldsymbol{s}_{\min}[m])$, $b_3 = \max(s^\star_{\min}, \boldsymbol{s}_{\max}[1], \ldots, \boldsymbol{s}_{\max}[m])$, $b_4 = s_{\sup}$;
- *Subprocedure 3:* $q_1 = \mathcal{U}[0, \alpha_{\max}]$.

## APPENDIX C
## HYPERPARAMETER POSTERIOR DISTRIBUTIONS FOR THE CLASSIFICATION OF CHIRPS

The hyperparameter priors displayed in Table IV lead to the following distributions.

- Frequency hyperparameters

For each component $k = \{1, \ldots m\}$
$$\begin{aligned} p(\boldsymbol{m}_f[k]|\boldsymbol{\theta}_{1:n}) &\propto 2\mathbb{I}_{[0,\,0.5]}(\boldsymbol{m}_f[k])\left[\frac{(\boldsymbol{m}_f[k] - \tau_f[k])^2}{\nu_f\xi_f[k]} + 1\right]^{-(\nu_f+1)/2} \\ &\propto \mathbb{I}_{[0,\,0.5]}(\boldsymbol{m}_f[k])\mathcal{T}_1(\nu_f, \boldsymbol{\tau}_f[k], \boldsymbol{\xi}_f[k]) \end{aligned} \tag{41}$$
where $\boldsymbol{\theta}_{1:n} = [\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n]$, and $\mathcal{T}_1$ is a one-dimensional Student-t distribution with parameters $\nu_f = n - 1$, $\nu_f\boldsymbol{\xi}_f[k] = \overline{f^2[k]} - \overline{f[k]}^2$, and $\boldsymbol{\tau}_f[k] = \overline{f[k]}$ (using the usual notations $\overline{f[k]} = (1/n)\sum_{i=1}^n \boldsymbol{f}_i[k]$, and $\overline{f^2[k]} = (1/n)\sum_{i=1}^n \boldsymbol{f}_i^2[k]$). Furthermore, the variance terms are distributed according to the following inverse Gamma distribution:
$$p(\boldsymbol{\sigma}_f^2[k]|\boldsymbol{m}_f[k], \boldsymbol{\theta}_{1:n}) \sim \mathcal{IG}\left(\frac{n}{2}, \frac{1}{2}\sum_{i=1}^n(\boldsymbol{f}_i[k] - \boldsymbol{m}_f[k])^2\right). \tag{42}$$

- Slope hyperparameters

For each component $k = \{1, \ldots m\}$, the posterior distribution of $\boldsymbol{s}_{\min}[k]$, $\boldsymbol{s}_{\max}[k]$ is a truncated inverse Gamma distribution
$$\begin{aligned} &p(\boldsymbol{s}_{\min}[k], \boldsymbol{s}_{\max}[k]|\boldsymbol{\theta}_{1:n}) \\ &\quad\propto [(n-1)!(\boldsymbol{s}_{\max}[k] - \boldsymbol{s}_{\min}[k])^{n+1}]^{-1} \\ &\qquad\times \mathbb{I}_{[s_{\inf},\, s_{\sup}]}(\boldsymbol{s}_{\min}[k])\mathbb{I}_{[\boldsymbol{s}_{\min}[k],\, s_{\sup}]}(\boldsymbol{s}_{\max}[k]). \end{aligned} \tag{43}$$

- Amplitude hyperparameters

For each component $k = \{1, \ldots m\}$
$$p(\boldsymbol{m}_a[k]|\boldsymbol{\theta}_{1:n}) \propto \mathbb{I}_{[0,\,a_{\max}]}(\boldsymbol{m}_a[k])\mathcal{T}_1(\nu_a, \boldsymbol{\tau}_a[k], \boldsymbol{\xi}_a[k]) \tag{44}$$
where $\nu_a = n - 1$, $\boldsymbol{\tau}_a[k] = (1/\mu_a)\sum_{i=1}^n(\boldsymbol{a}_i[k]/\sigma_{\epsilon_i}^2)$, $\mu_a = \sum_{i=1}^n(1/\sigma_{\epsilon_i}^2)$, and $\nu_a\boldsymbol{\xi}_a[k] = (1/\mu_a)\sum_{i=1}^n(\boldsymbol{a}_i^2[k]/\sigma_{\epsilon_i}^2) - (1/\mu_a^2)(\sum_{i=1}^n(\boldsymbol{a}_i[k]/\sigma_{\epsilon_i}^2))^2$. This is a one-dimensional Student-t distribution for the variable $\mathbf{m}_a[k]$, with parameters

Authorized licensed use limited to: INP TOULOUSE. Downloaded on July 19, 2009 at 12:41 from IEEE Xplore. Restrictions apply.

$\nu_a$, $\boldsymbol{\tau}_a[k]$, $\boldsymbol{\xi}_a[k]$, which is limited to the interval $[0, a_{\max}]$. Moreover

$$p(\boldsymbol{\sigma}_a^2[k]|\boldsymbol{m}_a[k], \boldsymbol{\theta}_{1:n}) \sim \mathcal{IG}\left(\frac{n}{2}, \frac{1}{2}\sum_{i=1}^{n}\frac{(\boldsymbol{a}_i[k]-\boldsymbol{m}_a[k])^2}{\sigma_{\epsilon_i}^2}\right). \tag{45}$$

- Additive noise hyperparameters

The hyperparameter $\alpha_\epsilon$ is distributed according to the (proper) distribution

$$p(\alpha_\epsilon|\boldsymbol{\theta}_{1:n}) \sim \mathbb{I}_{[0,\, \alpha_{\max}]}(\alpha_\epsilon)\frac{\Gamma\left(n\frac{\alpha_\epsilon}{2}+\nu_0\right)}{\Gamma(\alpha_\epsilon/2)^n}\left[\frac{P_\epsilon}{[S_\epsilon+2\gamma_0]^n}\right]^{\alpha_c/2} \tag{46}$$

where the sufficient statistics are $P_\epsilon = \prod_{i=1}^n (1/\sigma_{\epsilon_i}^2)$, and $S_\epsilon = \sum_{i=1}^n (1/\sigma_{\epsilon_i}^2$. The second hyperparameter $\beta_\epsilon$ is distributed according to a Gamma distribution

$$\begin{aligned} p(\beta_\epsilon|\alpha_\epsilon, \boldsymbol{\theta}_{1:n}) &\sim \mathcal{G}a\left(\frac{n\alpha_\epsilon}{2}+\nu_0, \frac{S_\epsilon}{2}+\gamma_0\right) \\ &\propto \beta_\epsilon^{(n\alpha_c/2)+\nu_0-1}e^{-(S_c/2+\gamma_0)\beta_c}\mathbb{I}_{\mathbb{R}}+(\beta_\epsilon). \end{aligned} \tag{47}$$

## Acknowledgment

The authors would like to thank A. Doucet, P. Djuric, and C. Andrieu for useful discussions and comments related to this work.

## References

[1] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.

[2] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*. New York: Academic, 1999.

[3] J. J. Rajan, P. J. W. Rayner, and S. J. Godsill, "A Bayesian approach to parameter estimation and interpolation of time-varying autoregressive processes using the Gibbs sampler," *Proc. Inst. Elect. Eng., Vis., Image Signal Process.*, vol. 144, no. 4, Aug. 1997.

[4] C. Guihenneuc-Jouyaux, S. Richardson, and V. Lasserre, "Convergence assessment in latent variable models: Application to the longitudinal modeling of a marker of HIV progression," in *Discretization and MCMC Convergence Assessment*, C. P. Robert, Ed. New York: Springer Verlag, 1998, pp. 147–159.

[5] C. P. Robert, *The Bayesian Choice—A Decision-Theoretic Motivation*. New York: Springer Verlag, 1994.

[6] J. M. Bernardo and A. F. M. Smith, *Bayesian Theory*. New York: Wiley, 1994.

[7] C. Andrieu and A. Doucet, "Joint Bayesian detection and estimation of noisy sinusoids via reversible jump mcmc," *IEEE Trans. Signal Processing*, vol. 47, pp. 2667–2676, Oct. 1999.

[8] W. J. Fitzgerald, "Markov chain Monte Carlo methods with applications to signal processing," *Signal Process.*, vol. 81, no. 1, pp. 3–18, Jan. 2001.

[9] P. M. Djuric, "Bayesian methods for signal processing," *IEEE Signal Processing Mag.*, vol. 15, pp. 26–28, May 1998.

[10] R. E. Kass and A. E. Raftery, "Bayes factors," *J. Amer. Statist. Assoc.*, vol. 90, pp. 773–795, 1995.

[11] M. Lavine and M. West, "A Bayesian method for classification and discrimination," *Canad. J. Stat.*, vol. 20, no. 4, pp. 451–461, 1992.

[12] M. Davy, C. Doncarli, and G. F. Boudreaux-Bartels, "Improved optimization of time-frequency based signal classifiers," *IEEE Signal Processing Lett.*, vol. 8, pp. 52–57, Feb. 2001.

[13] M. Davy, "Noyaux optimisés pour la classification dans le plan temps-fréquence," Ph.D. dissertation (in French), Univ. Nantes, Nantes, France, 2000.

[14] M. Davy and C. Doncarli, "Optimal kernels of time-frequency representations for signal classification," in *Proc. IEEE Int. Symp. TFTS*, Pittsburgh, PA, 1998, pp. 581–584.

[15] M. Davy, C. Doncarli, and J. Y. Tourneret, "Supervised classification using MCMC methods," in *Proc. IEEE ICASSP*, Istanbul, Turkey, June 2000, pp. 33–36.

[16] ——, "Hierarchical Bayesian learning using reversible jumps MCMC for chirps classification," Univ. Cambridge Eng. Dept., Cambridge, U.K., Tech. Rep. CUED/F-INFENG/TR.422, 2001.

[17] O. Besson, M. Ghogho, and A. Swami, "Parameter estimation for random amplitude chirp signals," *IEEE Trans. Signal Processing*, vol. 47, pp. 3208–3219, Dec. 1999.

[18] A. W. Rihaczek, *Principles of High-Resolution Radar*. New York: McGraw-Hill, 1969.

[19] B. W. Gillespie and L. E. Atlas, "Optimization of time and frequency resolution for radar transmitter identification," in *Proc. IEEE ICASSP*, Phoenix, AZ, May 1999, pp. 2127–2131.

[20] G. Matz and F. Hlawatsch, "Time-frequency subspace detectors and application to knock detection," *AEÜ Int. J. Electron. Commun.*, vol. 53, no. 6, pp. 1–6, Aug. 1999.

[21] ——, "Time-frequency methods for signal detection with application to the detection of knock in car engines," in *Proc. IEEE-SP Workshop Stat. Signal Array Process.*, Portland, OR, Sept. 1998, pp. 196–199.

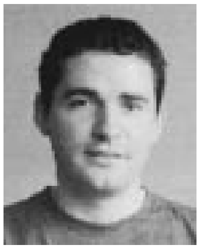
[22] P. J. Green, "Reversible jump MCMC computation and Bayesian model determination," *Biometrika*, vol. 82, pp. 711–732, 1995.

[23] S. J. Godsill and P. J. W. Rayner, "Statistical reconstruction and analysis of autoregressive signals in impulsive noise using the Gibbs sampler," *IEEE Trans. Speech, Audio Processing*, vol. 6, pp. 352–372, July 1999.

[24] A. Zellner, *An Introduction to Bayesian Inference in Econometrics*. New York: Wiley, 1971.

[25] S. Abeysekera and B. Boashash, "Methods of signal classification using the images produced by the Wigner distribution," *Pattern Recognit. Lett.*, vol. 12, pp. 717–729, Nov. 1991.

[26] B. W. Gillespie and L. E. Atlas, "Optimizing time-frequency kernels for classification," *IEEE Trans. Signal Processing*, vol. 49, pp. 485–496, Mar. 2001.



**Manuel Davy** was born in Caen, France, in 1972. He received the ingénieur degree in electrical engineering in 1996 from Ecole Centrale de Nantes, Nantes, France, and the Ph.D. degree from the University of Nantes in 2000.

He is currently an Associate Researcher at the Engineering Department, University of Cambridge, where his research activity is centered around parametric and nonparametric methods for audio processing.

**Christian Doncarli** was born in Marseille, France. He received the Ph.D. degree in automatic control in 1977.

He is First Class Professor of Signal Processing at Ecole Centrale de Nantes, Nantes, France. His current research interest is the area of nonstationary signal classification, using time–frequency/time-scale representations as well as parametric methods. His preferred application field is audio. He is in charge of the signal division of the National Group of Research on Information, Signal, Image, and Vision (ISIS).

**Jean-Yves Tourneret** was born in Besançon, France, in 1965. He received the ingénieur degree in electrical engineering in 1989 from Ecole Nationale Supérieure d'Electronique, d'Electrotechnique, d'Informatique et d'Hydraulique de Toulouse (ENSEEIHT), Toulouse, France, and the Ph.D. degree from the National Polytechnic Institute, Toulouse, in 1992.

He is currently an Associate Professor with ENSEEIHT, where his research activity is centered around estimation, detection, and classification of non-Gaussian and nonstationary processes.

Authorized licensed use limited to: INP TOULOUSE. Downloaded on July 19, 2009 at 12:41 from IEEE Xplore. Restrictions apply.