SIGNAL DETECTION AND ESTIMATION USING ATOMIC DECOMPOSITION AND INFORMATION-THEORETIC CRITERIA

Gustavo López-Risueño, Jesús Grajal, Omar A. Yeste-Ojeda

Departamento de Señales, Sistemas y Radiocomunicaciones
ETSIT, Universidad Politécnica de Madrid
Ciudad Universitaria s/n, 28040 Madrid, Spain
{risueno,jesus,omar}@gmr.ssr.upm.es

ABSTRACT

An algorithm to detect and estimate a linear mixture of signals corrupted by white Gaussian noise is presented. The number of signals is assumed to be unknown. The algorithm is based on a type of information theoretic criterion capable of adjusting the probability of false alarm, and uses atomic decomposition refined by the expectation maximization method to efficiently compute the maximum likelihood estimate of the signal parameters. Signals are modeled as chirplets. Algorithm consistency and efficiency are shown by simulation.

1. INTRODUCTION

The paper deals with the detection and estimation of the components of a linear mixture of signals collected by a single sensor. This problem arises in areas such as signal interception, also termed spectrum surveillance, where typically neither the number of signals nor their parameters are known. Our only assumption is that signals belong to an a-priori-known set or dictionary. Since the particular application is the interception of radar signals, the selected dictionary is made of chirplets, i.e. Gaussian-envelope functions with linear frequency modulation. Among other properties (e.g. optimum time-frequency concentration [1]), linear frequency modulation makes the chirplet dictionary suitable to model a wide variety of radar signals, specially those transmitted by low-probability-of-intercept radars [2].

The proposed algorithm relies upon the use of an information-theoretic criterion [3], whose structure is similar to other well-known ones, such as the Akaike information criterion (AIC) and the minimum description length criterion (MDL). Information criteria have been typical tools in model order selection problems; regarding signal detection and estimation, they have been profusely applied in areas such as array processing [3, 4]. In that area, most of the papers have dealt with the application of the MDL and the AIC, and the study of their asymptotic consistency (when increasing the number of measurements). On the one hand, these criteria present a limited aplicability since they fix the probability of false alarm ($\mathrm{P_{FA}}$) [5]. On the other hand, we model the signals by the parametric and deterministic chirplet model instead of the random stationary Gaussian model commonly used in array processing; therefore, provided that chirplets are time-limited, the asymptotic consistency becomes a secondary issue, since more measurements in time do not imply more information for time-limited signals.

This work was supported by CIDA and the National Board of Science and Technology (CICYT) under project TIC-2002-04569-C02-01

The proposed algorithm exhibits two important features. First, the information criterion penalty function includes a multiplicative constant to adjust the $\mathrm{P_{FA}}$. $\mathrm{P_{FA}}$ is defined as the probability of detecting one or more signals for noise only, due to its importance in real interception systems. Second, for a given model order, the signals are sequentially estimated by the use of Atomic Decomposition (AD) [6, 7, 8], and subsequently refined by the Expectation Maximization method (EM) [9]. AD plus EM provides an efficient manner to compute the maximum likelihood estimate (MLE), which is hard to compute in a direct way even for a low number of signals in the mixture (MLE is required for the cost function of the information criterion [3]).

AD, also known as Matching Pursuit [6, 7], is an adaptive approximation technique expanding the signal onto an overcomplete dictionary of elementary signals, called atoms (in this case, the chirplet dictionary). In general, AD provides a compact, physically-meaningful representation of signals, with better resolution than other adaptive techniques like the Best Basis Selection [6, 7], and with more efficiency than the Basis Pursuit [6]. The atoms estimated by the AD act as the starting point for the EM method. The joint use of the AD and EM algorithms was suggested earlier in [8].

The present work is an extension of the algorithm in [10], based only on AD. It also extends the work in [11], where EM and a "modified" AD were jointly applied to estimate the parameters of a known number of signals. The paper is organized as follows. Section 2 states the problem, addresses the application of the information criterion, and defines the MLE. AD, the EM method, and their mutual relationship are described in Section 3. The algorithm structure is described in Section 4, and its performance is shown in Section 5 through a number of simulations. Finally, the conclusions are drawn in Section 6.

2. THEORETICAL OVERVIEW

The problem of detecting and estimating the signals in a linear mixture can be stated by the following hypothesis test:

$$\begin{aligned} H_0 &: \mathbf{x} = \mathbf{r} \\ H_1 &: \mathbf{x} = \mathbf{s} + \mathbf{r}\,, \end{aligned} \tag{1}$$

where $\mathbf{r}$ is a complex, white, zero-mean, Gaussian noise of power σ^2 (vector notation is used, and the signal length is denoted by N). The signal $\mathbf{s}$ is the linear mixture of M signals,

$$\mathbf{s} = \sum_{k=1}^{M} b_k \, \mathbf{h}_{\gamma_k} \, . \tag{2}$$

The signals $\mathbf{h}_{\gamma_k}$ are unit-energy signals belonging to the a-priori-known dictionary, $D = \{\mathbf{h}_\gamma : \gamma \in \Gamma\}$. $\{\gamma_k\}_{k=1}^M$ are the corresponding indices (usually finite-dimension vectors). Coefficients $\{b_k\}_{k=1}^M$ are complex valued, and $|b_k|^2$ is the energy of each signal. A useful figure of merit is the energy-to-noise ratio (ENR), defined as $\mathrm{ENR}_k = |b_k|^2 / \sigma^2$ for each signal.

The objective is to determine the number of signals (M), and their parameters, namely: $\{\gamma_k\}_{k=1}^M$ and $\{b_k\}_{k=1}^M$. The noise power can be either known or unknown, so that an algorithm for each case is devised. Let $\widehat{M}$ be the number of signals estimate, $\mathrm{P_{FA}}$ is defined as

$$\mathrm{P_{FA}} = \mathrm{P}_{H_0}\left(\widehat{M} > 0\right) \, . \tag{3}$$

Likewise, the probability of correct detection ($\mathrm{P_{CD}}$), missing detections ($\mathrm{P_{MD}}$) and false detections ($\mathrm{P_{FD}}$) are given, respectively, by

$$\mathrm{P_{CD}} = \mathrm{P}_{H_1}\left(\widehat{M} = M\right) \, , \mathrm{P_{MD}} = \mathrm{P}_{H_1}\left(\widehat{M} < M\right) \, , \quad \mathrm{P_{FD}} = \mathrm{P}_{H_1}\left(\widehat{M} > M\right) \, . \tag{4}$$

We follow an information criterion to estimate $\widehat{M}$. In general, information criterion approaches estimate the model order (q) by minimizing a cost function of the form [3]

$$\mathrm{IC}(q) = -\log f(\mathbf{x}; \boldsymbol{\theta}^{(q)}) + p(q) \, , \tag{5}$$

where $f(\mathbf{x}; \widehat{\boldsymbol{\theta}}^{(q)})$ is the pdf corresponding to the family of qth-order distributions, and $\widehat{\boldsymbol{\theta}}^{(q)}$ is the maximum likelihood estimate (MLE) of the distribution parameters, i.e.,

$$\widehat{\boldsymbol{\theta}}^{(q)} = \arg\max f(\mathbf{x}; \boldsymbol{\theta}^{(q)}) \, . \tag{6}$$

$p(q)$ is a penalty function of the model order depending on the specific information-theoretic criterion considered. It is usually split into two factors

$$p(q) = g(N) \, m(q) \, , \tag{7}$$

where $g(N)$ is a function of the number of observations (i.e., the signal length N), and $m(q)$ is the number of free parameters associated to the family of qth-order distributions. For instance, $g(N) = 1$ for the AIC, and $g(N) = 0.5 \log N$ for the MDL.

For the problem at hand, the model order coincides with the number of signals in the mixture. Besides, $g(N)$ does not follow any predefined expression, but it is computed by simulation to meet a given $\mathrm{P_{FA}}$ value (Section 4). AIC or MDL will hold a greater (smaller) $\mathrm{P_{FA}}$ than the required value if their $g(N)$ values are smaller (greater) than the selected one. The computation of the cost function (5) requires the MLE of the distribution parameters. For the assumption of q signals, it becomes

$$\widehat{\boldsymbol{\theta}}^{(q)} = (\widehat{b}_1, \widehat{\gamma}_1^{\mathrm{T}}, \ldots, \widehat{b}_q, \widehat{\gamma}_q^{\mathrm{T}})^{\mathrm{T}} = \arg\min_{\boldsymbol{\theta}^{(q)}} \|\mathbf{x} - \sum_{k=1}^{q} b_k \, \mathbf{h}_{\gamma_k}\|^2 \, , \tag{8}$$

for known σ. If it is unknown, its MLE,

$$\widehat{\sigma}_q^2 = \frac{\|\mathbf{x} - \sum_{k=1}^{q} \widehat{b}_k \, \mathbf{h}_{\widehat{\gamma}_k}\|^2}{N} \, , \tag{9}$$

has to be added to the parameter vector $\widehat{\boldsymbol{\theta}}^{(q)}$.

2.1. The chirplet dictionary

The concrete form of Eqns. (5) and (8) depends on the utilized dictionary. Regarding the chirplet dictionary, each chirplet is defined by the four parameters vector $\gamma = [\alpha, \beta, T, f]^{\mathrm{T}}$ and has the following unit-energy waveform:

$$h_\gamma(n) = \left(\frac{\alpha}{\pi}\right)^{1/4} e^{-\frac{\alpha}{2}(n-T)^2} \cdot e^{j\left[2\pi f(n-T) + \pi\beta(n-T)^2\right]} \, . \tag{10}$$

T, f and β are the mean time, mean frequency, and chirp rate, respectively. α is inversely related to the duration (d) of the chirplet: $d = \sqrt{2/\alpha}$, in the framework of Time-Frequency Analysis [1]. Chirplets are assumed to be time- and band-limited, and the sampling rate is one. There are 6 free parameters per chirplet (the coefficient is complex-valued); thus, for order q, the number of free parameters becomes

$$m(q) = \begin{cases} 6\,M & \text{, for known } \sigma, \\ 6\,M + 1 & \text{, for unknown } \sigma. \end{cases} \tag{11}$$

3. ATOMIC DECOMPOSITION AND EM

For most dictionaries, including the chirplet one, the optimization of (8) is numerically inefficient for $q > 1$. The cost function is highly dimensional, with many local maxima, and constant-valued regions [10]. However, although still a hard problem, the optimization of (8) is computationally feasible for $q = 1$. Thus, the sequential estimation of the signals arises as an efficient way to compute (8) for $q > 1$. This is performed by AD in a natural way [6, 7, 8]. Let $\mathbf{x}$ be the signal under analysis, and $D = \{\mathbf{h}_\gamma : \gamma \in \Gamma\}$ be the dictionary of atoms, AD obtains the expansion

$$\mathbf{x} \approx \sum_{p} \widehat{b}_p \, \mathbf{h}_{\widehat{\gamma}_p} \tag{12}$$

by means of the equations [6, 7, 8]:

$$\widehat{\gamma}_p = \arg\max_{\gamma} \left|\mathbf{h}_\gamma^{\mathrm{H}} \, \mathbf{x}_{p-1}\right|^2 \, , \tag{13}$$

$$\widehat{b}_p = \mathbf{h}_\gamma^{\mathrm{H}} \, \mathbf{x}_{p-1} \, , \tag{14}$$

where $\mathbf{x}_p$ denotes the pth order residual that are defined as

$$\mathbf{x}_p = \mathbf{x}_{p-1} - \widehat{b}_p \, \mathbf{h}_{\widehat{\gamma}_p} \quad p = 1, \ldots \, , \tag{15}$$

$$\mathbf{x}_0 = \mathbf{x} \, . \tag{16}$$

AD approximates the MLE of (8) by splitting the q-signal estimation into q consecutive single-signal estimations; hence it is suboptimal. We suggest the use of the EM method in order to improve the MLE approximation. EM [9] is an iterative algorithm that computes the MLE using a "completed" data space of the observed data. Each iteration consists of an expectation step (E-step)

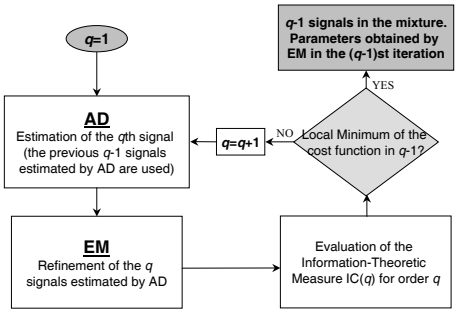

Fig. 1. The block diagram of the detection and estimation algorithm.

followed by a maximization step (M-step). By the space extension, the MLE computation typically becomes simpler than directly using the observed data. In addition, under fairly general conditions, EM converges to a stationary point of the likelihood function [9]. In our case, following the work by Feder and Weinstein [12], the complete data are obtained by decomposing the observed signal $\mathbf{x}$ into its components. Assuming a known number of signals in the mixture (q), and an initial estimate of them, the error between the signal under analysis ($\mathbf{x}$) and its estimate becomes

$$\mathbf{e} = \mathbf{x} - \sum_{k=1}^{q} \widehat{b}_k \; \mathbf{h}_{\widehat{\gamma}_k} \; . \tag{17}$$

Then, the complete data estimate (E-step) are formed as

$$\widehat{\mathbf{x}}_k = \widehat{b}_k \; \mathbf{h}_{\widehat{\gamma}_k} + \beta_k \mathbf{e} \, , \quad k = 1, \ldots, q \, , \tag{18}$$

with the condition $\sum_{k=1}^{q} \beta_k = 1$ (see [11] for details on the selection of weights β_k) . The M-step computes the MLE of each $\widehat{\mathbf{x}}_k$, i.e.,

$$\widehat{\gamma}_k = \arg \max_{\gamma} \left| \mathbf{h}_{\gamma}^{\mathrm{H}} \, \widehat{\mathbf{x}}_k \right|^2 \, , \quad \widehat{b}_k = \mathbf{h}_{\widehat{\gamma}}^{\mathrm{H}} \, \widehat{\mathbf{x}}_k \; . \tag{19}$$

This procedure iterates to refine the estimation. Interestingly, every maximization in the M-step can be seen as the estimation of an atom by the AD, so that the same optimization techniques can be used.

4. ALGORITHM DESCRIPTION

The algorithm flowchart is depicted in Fig. 1. For the qth iteration, the qth signal is estimated by AD. Using this and the $q-1$ previously estimated signals, the EM refinement is carried out, and the information criterion cost function (eqn. 5) is computed. The algorithm should seek for the global minimum. Nevertheless, for computational efficiency, it seeks for the first local minimum. This approximate strategy has resulted in a very good performance for the experiments evaluated (see the example in Section 5).

The AD algorithm is implemented by a global search algorithm plus a gradient-search refinement due to the complex structure of the objective function in (13) [10]. In this case, the optimization is carried out by a genetic algorithm followed by a Quasi-Newton search. The reader may find the algorithm description in [10] (the novelty here is the use of the chirplet duration d instead of the parameter α in the Quasi-Newton search for convergence reasons). For the optimizations of the M-step of the EM method, only the above-mentioned Quasi-Newton algorithm is utilized. Regarding the computational burden, the genetic algorithm and the Quasi-Newton search take 1.8 second and 0.1 second, respectively (Pentium IV 2.4 GHz with 1 GB RAM using MATLAB 6.1; 1024-sample signals).

It is noteworthy to emphasize the efficiency of our algorithm since only a new AD estimation (i.e. the most computationally demanding step) is required every time the order q increases. For the estimation by the AD of the qth signal, the previous $q-1$ signals estimated by the AD must be used. The use of the $q-1$ signals obtained by the EM refinement may lead to the divergence from the MLE. Regarding $\mathrm{P_{FA}}$, if only the first local minimum is sought, its definition turns out to be:

$$\mathrm{P_{FA}} = \mathrm{P}_{H_0} \left(\mathrm{IC}(1) < \mathrm{IC}(0) \right) = \mathrm{P}_{H_0} \left(\left| \widehat{b}_1 \right|^2 / \sigma^2 > Th \right) \, , \tag{20}$$

where the threshold Th is related to the information criterion by $Th = 6 \; g(N)$ for a known σ. If σ is unknown, the estimate $\widehat{\sigma}_0^2$ (eqn. 9) must be used; in addition, the threshold Th is written as $Th = N \left(1 - \exp\{-6g(N)/N\}\right)$, which becomes $Th = 6 \; g(N)$ for a small $g(N)/N$.

5. RESULTS

This section considers a very general multiple-signal example made up of 6 chirplets with different parameters. Chirplets 1 to 4 are time-shifted pulses with $\alpha = 10^{-3}$, $f = 1/14$, $\beta = 0$, and mean times 150, 300, 600 and 800, respectively. Chirplets 5 and 6 have the same mean time and frequency ($T = 500$, and $f = 0.25$). For the chirplet 5, $\alpha = 10^{-3}$ and $\beta = 3 \cdot 10^{-3}$; for the chirplet 6, $\alpha = 10^{-4}$ and $\beta = -10^{-3}$. Chirplets are uncorrelated except the chirplets 5 and 6 (correlation coefficient 16%), and the chirplets 3 and 6 (correlation coefficient 2.5%). Its time-frequency representation is shown in Fig. 2. The signal length is $N = 1024$ and the noise power is $\sigma^2 = 1$. In the experiments, different ENRs have been studied. Chirplets 1 to 5 have the same ENR, while the ENR of the chirplet 6 is 2 dB higher. The algorithm of Section 4 (using AD+EM) is analyzed for known and unknown noise power. It is also compared with a version using only AD (no EM). $\mathrm{P_{FA}} = 10^{-6}$ in all cases.

Figure 3 shows the performance in the estimation of the number of signals. When EM is used, the information criterion approach leads to a consistent estimate of the number of signals for moderate ENRs (Moderate ENRs in the sense that they are around 5 dB over the sensitivity of the matched filter detector, $\mathrm{ENR} = 14$ dB, for $\mathrm{P_D} = 90\%$ and $\mathrm{P_{FA}} = 10^{-6}$).

This is related to the fact that the EM method provides an estimate approaching the MLE when the ENR increases. The former statement relies on: 1) in the experiment, the chirplet parameter estimates attain the Cramer-Rao bound (CRB), and 2) the MLE has been found to be statistically efficient for the considered problem when ENR increases [11]. However, when EM is not used, estimation errors deviate from the CRB and even increase for a high ENR. This results in the over-estimation of the number of signals (Fig. 3, right). An illustration of the previous comments on the estimation performance is in Fig. 4. The root mean square error (RMSE) of the estimation of α and the corresponding CRB are plotted for chirplets 5 and 6. The other parameters exhibit the same behavior and are not displayed because of the limitation in space.

Focusing on the proposed algorithm (AD+EM), there are no differences between the cases with known and unknown noise power in terms of estimation performance. For detection, a slight degradation appears when the noise power is unknown, due to the need to estimate it. This is more clearly shown in Table 1, where the sensitivity for each chirplet in the mixture is shown. Sensitivity is defined as the minimum ENR for the chirplet to be detected by the algorithm with a probability of 90%. Losses range from 0.5 dB to 1 dB.

Regarding the use of other information criteria, we have carried out the same experiments using the MDL and the AIC instead of the proposed information criterion. Since $g(N)$ is smaller for these criteria ($g(N)$ is 6, 3.5 and 1, for the proposed approach, the MDL and the AIC, respectively), $\mathrm{P_{FA}}$ becomes greater: 0.3 for the MDL, and approximately 1 for the AIC. The proposed information criterion and the MDL estimate the true number chirplets for $\mathrm{ENR} \geq 25$ dB; however, the AIC is not able to consistently estimate the number of signals for $\mathrm{ENR} < 35$ dB (the ENR is referred to chirplets 1-5).

σ^2 \ Chirplet	1	2	3	4	5	6
known σ^2	17	17	17	17	16	20
unknown σ^2	17.5	17.5	17.5	17.6	17	20.5

Table 1. Sensitivity in terms of ENR (dB) for each chirplet in the mixture for the cases with known and unknown noise power (algorithm AD+EM).

6. CONCLUSIONS

This paper presents an algorithm to detect the number of signals in a linear mixture corrupted by white Gaussian noise, and to estimate their parameters. It is based on an information-theoretic criterion that, unlike others, permits the control of $\mathrm{P_{FA}}$. The algorithm uses the AD and EM algorithms to efficiently estimate the MLE. The consistency in estimating the number of signals as well as the efficiency of the signal estimates have been shown by simulation. The important role of the EM method has been pointed out as well. The formal generalization of the properties shown by the algorithm in the simulations would need, nevertheless, a further investigation. Signals have been modeled as chirplets, although the algorithm can be easily extended to other dictionaries.

7. REFERENCES

[1] L. Cohen, *Time-Frequency Analysis*, Prentice Hall, 1995.

[2] J. Tsui, *Digital Techniques for Wideband Receivers*, Artech House, 1995.

[3] E. Fishler, M. Grossmann, and H. Messer, "Detection of Signals by Information Theoretic Criteria: General Asymptotic performance analysis," *IEEE Trans. SP*, vol. 50, no. 5, 2002.

[4] H. L. van Trees, *Optimum Array Processing*, Wiley, 2002.

[5] R. E. Bethel and K. L. Bell, "A Correlated Signals MLE Approach to Joint Source Detection and Direction of Arrival Estimation," in *IEEE SSP*, Sept. 2003.

[6] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, 1998.

[7] S. G. Mallat and Z. Zhang, "Matching Pursuits with Time-Frequency Dictionaries," *IEEE Trans. on SP*, vol. 41, no. 12, 1993.

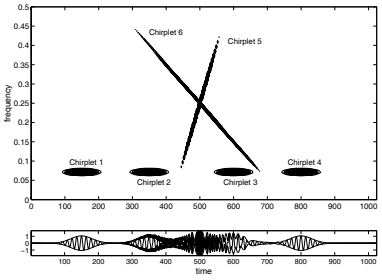

Fig. 2. Adaptive Spectrogram [6] of the multichirplet signal (noise free). Signal envelope and real part at the bottom.

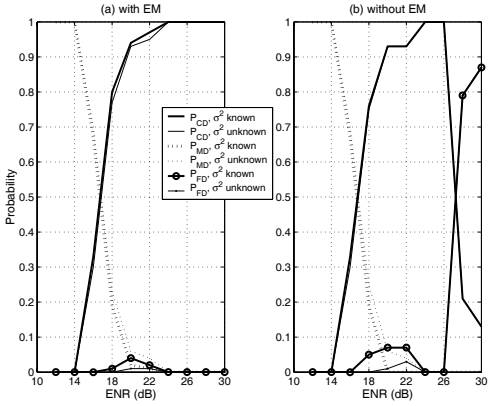

Fig. 3. $\mathrm{P_{CD}}$, $\mathrm{P_{MD}}$ and $\mathrm{P_{FD}}$ for the example in Fig. 2. ENR referred to chirplets 1 to 5 in abcisas

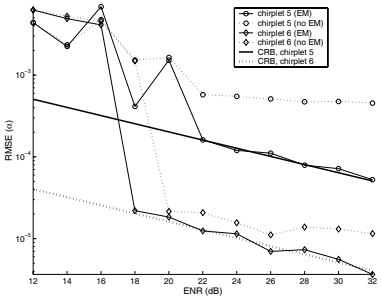

Fig. 4. Estimation vs. ENR of parameter α for chirplets 5 and 6, when EM is either used or not. ENR referred to chirplets 1 to 5 in abcisas.

[8] J. C. O'Neill, P. Flandrin, and W. C. Karl, "Sparse Representation with Chirplets via Maximum Likelihood Estimation," 2000, Available at http://mdsp.bu.edu/jeffo.

[9] G. J. McLachlan and K. Thrimyambakam, *The EM Algorithm and Extensions*, Wiley, 1997.

[10] G. López-Risueño, J. Grajal, and O. Yeste-Ojeda, "Atomic Decomposition-based Complex Radar Interception," *IEE Proc.-Radar, Sonar and Navigation*, August 2003.

[11] G. Lopez-Risueño, J. Grajal, and O. A. Yeste-Ojeda, "On the Use of EM and High Resolution Atomic Decomposition to Estimate a Linear Mixture of Chirplets," in *IEEE SSP*, Sept. 2003, pp. 257–260.

[12] M. Feder and E. Weinstein, "Parameter Estimation of Superimposed Signals Using the EM Algorithm," *IEEE Trans. on ASSP*, vol. 36, no. 4, 1988.