DECOMPOSITION OF A MIXTURE OF GAUSSIAN AR PROCESSES

Christophe Couvreur and Yoram Bresler

Coordinated Science Laboratory and
Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign
1308 West Main Street, Urbana, IL 61801

ABSTRACT

We consider the problem of detecting and classifying an unknown number of multiple simultaneous Gaussian autoregressive (AR) signals with unknown variances given a finite length observation of their sum and a dictionary of candidate AR models. We show that the problem reduces to the maximum likelihood (ML) estimation of the variances of the AR components for every subset from the dictionary. The "best" subset of AR components is then found by applying the minimum description length (MDL) principle. The ML estimates of the variances are obtained by combining the EM algorithm with the Rauch-Tung-Striebel optimal smoother. The performance of the algorithm is illustrated by numerical simulations. Possible improvements of the method are discussed.

1. INTRODUCTION

It is well known that autoregressive (AR) models can represent a wide variety of signals, and they have been extensively researched [1]. However, in certain applications multiple signals, each described by an AR model, may arise simultaneously. To deal with such situations we define a new model for simultaneous AR signals. A time series $\{y_t\}$ is defined as a mixture of Gaussian AR processes (denoted ΣAR) if $\{y_t\}$ is a sum of c components $\{x_{i,t}\}$ plus some white noise $\{n_t\}$, and the sequences $\{x_{i,t}\}$ are generated by uncorrelated Gaussian AR(p_i) processes. That is,

$$y_t = \sum_{i=1}^{c} x_{i,t} + n_t, \tag{1}$$

$$x_{i,t} = -\sum_{k=1}^{p_i} a_{i,k} x_{i,t-k} + e_{i,t}, \tag{2}$$

where $\{e_{i,t}\}$ are Gaussian mutually uncorrelated innovation sequences with respective variances ρ_i, and $\{n_t\}$ is a white Gaussian process with variance σ_n^2. Such a process is represented by the system of Fig. 1, where $A_i(z) = 1 + \sum_{k=1}^{p_i} a_{i,k} z^{-k}$. The sum of AR processes results in an ARMA(p, q) process with a particular structure, as it can be seen by writing the power spectral density of $\{y_t\}$ as

$$P_{yy}(\omega) = \sum_{i=1}^{c} \frac{\rho_i}{|A_i(e^{j\omega})|^2} + \sigma_n^2 = \sigma_u^2 \frac{|B(e^{j\omega})|^2}{|A(e^{j\omega})|^2}, \tag{3}$$

with $B(z) = 1 + \sum_{k=1}^{q} \beta_k z^{-k}$ and $A(z) = 1 + \sum_{k=1}^{p} \alpha_k z^{-k}$, $p = q = \sum_{i=1}^{c} p_i$.

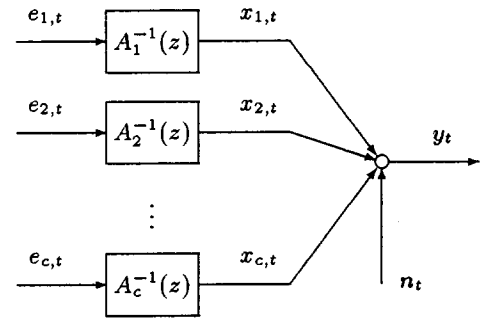

Figure 1: A mixture of Gaussian AR processes.

In this paper we consider the problem of detecting the AR components that are present in a ΣAR signal when a set of candidate AR models is available. That is, we have a dictionary of known component models $A_i(z)$, $i = 1, \ldots, M$, and we want to find the subset of c of these components that are present in the observed signal $\{y_t\}$. Typical examples of applications of such tests would be in environmental sound recognition [2] or in recognition and processing of speech in noisy background [3] where multiple signal components well represented by AR models may be observed simultaneously. The problem is, in a stochastic sense, equivalent to a vector quantization (VQ) problem where we could select multiple codewords simultaneously from a known codebook to best fit the signal.

2. PROBLEM FORMULATION

Let us assume that a dictionary Ω of M candidate AR models is available:

$$\Omega = \left\{ A_i(z) : A_i(z) = 1 + \sum_{k=1}^{p_i} a_i[k] z^{-k}, i = 1, \ldots, M \right\}.$$

This work was supported by a Rotary Foundation Ambassadorial Scholarship. Also supported in part by a National Science Foundation grant No. MIP 91-57377. Christophe Couvreur is a Research Assistant of the National Fund for Scientific Research Belgacom with the Faculté Polytechnique de Mons, Mons, Belgium.

0-7803-2431-5/95 $4.00

We want to find the subset $\xi \subset \Omega$ of c elements from Ω that are present in the signal y_t based on the observation of a realization $\mathbf{y} = [y_1 \ \cdots \ y_N]^T$ of $\{y_t\}$. The number of AR component c is not known a priori.

A naive attempt to solve the problem would be to choose the subset ξ that maximizes the generalized likelihood function

$$L(\mathbf{y};\xi) = \log \max_{\theta} p(\mathbf{y};\xi,\theta) \tag{4}$$

where the vector parameter $\theta = [\rho_1 \ \cdots \ \rho_c \ \sigma_n^2]^T$ represents the unknown variances of the c AR components in ξ and the unknown noise variance σ_n^2. This attempt will fail because the generalized log-likelihood $L(\mathbf{y};\xi)$ is a monotonically increasing (or at least nondecreasing) function of c. It will therefore reach its maximum at $c = M$. To avoid selecting this trivial and uninteresting solution, it is necessary to add a *penalty term* to (4). One form of penalized likelihood based on information theoretic considerations is the *minimum description length* criterion (MDL) of Rissanen [4]. Under some mild assumptions on the variances ρ_i and σ_n^2, the MDL solution ξ_{MDL} is given by [5]

$$\xi_{\mathrm{MDL}} = \arg \min_{\xi \subset \Omega} \left\{ -L(\mathbf{y};\xi) + \frac{1}{2}(c+1)\log N \right\}. \tag{5}$$

The optimum subset ξ_{MDL} is obtained by computing $L(\mathbf{y};\xi)$ for all possible subsets ξ and then selecting the the maximizer of (5). From (4), we can see that the computation of $L(\mathbf{y};\xi)$ is equivalent to the estimation of θ in the maximum likelihood (ML) sense. In the next section, we will derive the ML estimator of θ for a given ξ. That is, we will assume that c and the AR parameters are known and we will estimate the variances $\rho_1, \ldots, \rho_c, \sigma_n^2$.

3. ESTIMATION OF VARIANCES

3.1. State-Space Model

The ΣAR signal model defined by (1)–(2) can be reformulated as a state-space model. Let $\mathbf{F}_i$ be the $(p_i+1)\times(p_i+1)$ augmented top-row companion matrix

$$\mathbf{F}_i = \begin{bmatrix} -a_{i,1} & -a_{i,2} & \cdots & -a_{i,p_i} & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix},$$

and let $\mathbf{x}_{i,t}$ and $\mathbf{g}_i$ be the $(p_i+1)\times 1$ column vectors $\mathbf{x}_{i,t} = [x_{i,t} \ x_{i,t-1} \ \cdots \ x_{i,t-p_i}]^T$ and $\mathbf{g}_i = [1 \ 0 \ \cdots \ 0]^T$. The AR($p_i$) process $\{x_{i,t}\}$ of (2) is alternately defined by the vector difference equation

$$\mathbf{x}_{i,t} = \mathbf{F}_i \mathbf{x}_{i,t-1} + \mathbf{g}_i e_{i,t}. \tag{6}$$

We can collate the state equations (6) for $i = 1, \ldots, c$, and introduce an observation equation for $\{y_t\}$ to obtain the state-space model

$$\mathbf{x}_t = \mathbf{F}\mathbf{x}_{t-1} + \mathbf{G}\mathbf{e}_t, \tag{7}$$

$$y_t = \mathbf{h}^T \mathbf{x}_t + n_t, \tag{8}$$

where $\mathbf{x}_t = [\mathbf{x}_{1,t}^T \ \mathbf{x}_{2,t}^T \ \cdots \ \mathbf{x}_{c,t}^T]^T$, and $\mathbf{e}_t = [e_{1,t} \ e_{2,t} \ \cdots \ e_{c,t}]^T$. The $(p+c)\times(p+c)$ block diagonal matrix $\mathbf{F}$ is given by $\mathbf{F} = \mathrm{block\,diag}(\mathbf{F}_1, \ldots, \mathbf{F}_c)$. The $(p+c)\times c$ matrix $\mathbf{G}$ and the $(p+c)\times 1$ column vector $\mathbf{h}$ are given respectively by $\mathbf{G} = \mathrm{block\,diag}(\mathbf{g}_1, \ldots, \mathbf{g}_c)$, and $\mathbf{h} = [1 \ 0 \ \cdots \ 0]^T$. The covariance matrix of the input vector $\mathbf{e}_t$ is the $c \times c$ diagonal matrix $\mathbf{Q} = \mathrm{Cov}(\mathbf{e}_t) = \mathrm{diag}(\rho_1, \ldots, \rho_c)$.

3.2. EM Algorithm

The Expectation-Maximization (EM) algorithm presented first by Dempster *et al.* [6] is a general iterative method to compute ML estimates when the observed data can be regarded as "incomplete" and the incomplete data set can be related to some "complete" data set through a noninvertible transformation. Let the vector of observations $\mathbf{y}$ be the incomplete data set which has probability density $p(\mathbf{y};\theta)$. The ML estimate of θ is obtained by maximizing the log-likelihood function

$$\theta_{\mathrm{ML}} = \arg \max_{\theta} \log p(\mathbf{y};\theta). \tag{9}$$

A natural choice for the complete data set $\mathcal{X}$ is the set $\mathcal{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N, n_1, \ldots, n_N\}$. The non-invertible transformation relating the complete data set and the incomplete data set is simply the observation equation (8). The two stages of the EM algorithm can be stated as follows. Let $\theta^{(0)}$ denote an arbitrary initial guess for θ obtained, for example, by the method of moments method described in section 3.3, and let $\theta^{(n)}$ denote the current estimate of θ after k iterations of the algorithm. Then, the next iteration cycle is given by

E step: compute

$$U(\theta,\theta^{(k)}) = E_{\theta^{(k)}}\{\log p(\mathcal{X};\theta) \mid \mathbf{y}\}, \tag{10}$$

M step:

$$\theta^{(k+1)} = \arg \max_{\theta} U(\theta,\theta^{(k)}). \tag{11}$$

First, consider the evaluation of $U(\theta,\theta^{(k)})$ for the E-step. The log-likelihood function $\log p(\mathcal{X};\theta)$ is equal to

$$\begin{aligned} &\log p(\mathcal{X};\theta) \\ &\quad = \sum_{i=1}^{c} \log p(\mathbf{x}_{i,1}, \ldots, \mathbf{x}_{i,N}) + \log p(n_1, \ldots, n_N) \\ &\quad = \sum_{i=1}^{c} \log p(e_{i,1}, \ldots, e_{i,N} | \mathbf{x}_{i,1}) p(\mathbf{x}_{i,1}) \\ &\qquad\qquad + \log p(n_1, \ldots, n_N) \end{aligned}$$

Using the Gaussian assumption. we find, after some algebraic manipulations [5],

$$\begin{aligned} \log p(\mathcal{X};\theta) \ = \ & \sum_{i=1}^{c} \left[-\frac{1}{2}(N+p_i)\log \rho_i \right. \\ & \left. -\frac{1}{2\rho_i}\left(\mathbf{a}_i^T \left(\sum_{t=2}^{N} \mathbf{x}_{i,t}\mathbf{x}_{i,t}^T \right) \mathbf{a}_i + \mathrm{trace}\{\mathbf{\Gamma}_i^{-1}\mathbf{x}_{i,1}\mathbf{x}_{i,1}\} \right) \right] \\ & -\frac{N}{2}\log \sigma_n^2 - \frac{1}{2\sigma_n^2}\sum_{t=1}^{N}\left(y_t^2 - 2y_t\mathbf{h}^T\mathbf{x}_t + \mathbf{h}^T\mathbf{x}_t\mathbf{x}_t^T\mathbf{h} \right) \\ & + K, \end{aligned} \tag{12}$$

where K is a constant independent of θ, $\mathbf{a}_i$ is the vector of AR parameters $\mathbf{a}_i = [1\ a_{i,1}\ \cdots\ a_{i,p_i}]^T$, and $\mathbf{\Gamma}_i$ is the normalized covariance matrix $\mathbf{\Gamma}_i = \rho_i^{-1}\,\mathrm{Cov}(\mathbf{x}_{i,1})$. Since the elements of the vector $\mathbf{x}_{i,t} = [x_{i,t}\ \cdots\ x_{i,t-p_i}]^T$ are consecutive samples from a stationary AR process, $\mathbf{\Gamma}_i$ is a symmetric Toeplitz matrix whose elements are the first $p_i + 1$ autocorrelations of $\{x_{i,t}\}$ (see sec. 3.3 below). Alternately, the inverse $\mathbf{\Gamma}_i^{-1}$ can be obtained directly from the AR parameters by the Godberg-Semencul formula [1, p. 166]. We then define the smoothed conditional means

$$\hat{\mathbf{x}}_{t|N}^{(k)} = E_{\theta^{(k)}}\{\mathbf{x}_t|\mathbf{y}\}, \tag{13}$$

$$\hat{\mathbf{x}}_{i,t|N}^{(k)} = E_{\theta^{(k)}}\{\mathbf{x}_{i,t}|\mathbf{y}\}, \tag{14}$$

and the smoothed conditional covariance matrices

$$\mathbf{P}_{t|N}^{(k)} = \mathrm{Cov}_{\theta^{(k)}}(\mathbf{x}_t|\mathbf{y}), \tag{15}$$

$$\mathbf{P}_{i,t|N}^{(k)} = \mathrm{Cov}_{\theta^{(k)}}(\mathbf{x}_{i,t}|\mathbf{y}). \tag{16}$$

Note that $\hat{\mathbf{x}}_{i,t|N}^{(k)}$ can be extracted from $\hat{\mathbf{x}}_{t|N}^{(k)}$, and $\mathbf{P}_{i,t|N}^{(k)}$ is a block diagonal submatrix of $\mathbf{P}_{t|N}^{(k)}$. Combining (10) with (12) and (13)–(16), we obtain

$$\begin{aligned} U(\theta,\theta^{(k)}) = \sum_{i=1}^{c}\Bigg[& -\frac{1}{2}(N+p_i)\log\rho_i \\ & -\frac{1}{2\rho_i}\left(\mathbf{a}_i^T\left(\sum_{t=2}^{N}\left(\hat{\mathbf{x}}_{i,t|N}^{(k)}(\hat{\mathbf{x}}_{i,t|N}^{(k)})^T + \mathbf{P}_{i,t|N}^{(k)}\right)\right)\mathbf{a}_i \right. \\ & \left. +\mathrm{trace}\left\{\mathbf{\Gamma}_i^{-1}\left(\hat{\mathbf{x}}_{i,1|N}^{(k)}(\hat{\mathbf{x}}_{i,1|N}^{(k)})^T + \mathbf{P}_{i,1|N}^{(k)}\right)\right\}\right)\Bigg] \\ & -\frac{N}{2}\log\sigma_n^2 - \frac{1}{2\sigma_n^2}\sum_{t=1}^{N}\left(y_t^2 - 2y_t\mathbf{h}^T\hat{\mathbf{x}}_{t|N}^{(k)}\right. \\ & \left. +\mathbf{h}^T\left(\hat{\mathbf{x}}_{t|N}^{(k)}(\hat{\mathbf{x}}_{t|N}^{(k)})^T + \mathbf{P}_{t|N}^{(k)}\right)\mathbf{h}\right) + K. \end{aligned} \tag{17}$$

The smoothed conditional means (13) and smoothed conditional covariances matrices (15) can be computed by the Kalman filter in its Rauch-Tung-Striebel optimal smoother version [7]. Table 1, where the index $(\cdot)^{(k)}$ has been omitted for brevity, summarizes the smoothing algorithm. The values of $\rho_1,\ldots,\rho_c$, and σ_n^2 used in the algorithm are the current estimates at step k of the EM algorithm.

The maximization of (17) during the M-step is trivial. Setting the partial derivatives of $U(\theta,\theta^{(k)})$ with respect to $\rho_1,\ldots,\rho_c$, and σ_n^2 to zero yields the new estimate $\theta^{(k+1)}$.

We conclude this section by noting that Ziskand and Hertz [8] proposed a conceptually similar approach to the maximum likelihood estimation of the parameters of multiple AR signals in a frequency estimation problem.

3.3. Method of Moments

The EM algorithm presented above requires an initial guess $\theta^{(0)}$. This initial value can be obtained by the method of moments (MoM). The method of moments estimate θ_{MoM} is obtained by equating $(c+1)$ theoretical moments of the process, which are a function of θ, to $(c+1)$ sample moments computed from the data, and solving the resulting system of equations for θ. Since we are considering zero-mean Gaussian processes, the moments of interest are the covariances. Let $\{r_k\}$ be the covariance sequence of $\{y_t\}$ and let $\{r_{i,k}\}$ be the covariance sequence of $\{x_{i,t}\}$ normalized by ρ_i (i.e., the autocorrelation sequence): $r_k = E\{y_t y_{t-k}\}$, and $r_{i,k} = E\{x_{i,t}x_{i,t-k}\}/\rho_i$. The autocorrelation sequence $\{r_{i,k}\}$ of an AR process can be computed from the AR parameters $\{a_{i,1},\ldots,a_{i,p_i}\}$ of the process [1, pp. 44–45]. The covariance sequence of the white noise sequence $\{n_t\}$ is given by $r_{n,k} = \sigma_n^2\delta_k$, where δ_k is Kronecker's delta function. Because the processes $\{x_{i,t}\}$ and $\{n_t\}$ are mutually uncorrelated, their covariances add up, and we have

$$r_k = \rho_1 r_{1,k} + \ldots + \rho_c r_{c,k} + \sigma_n^2\delta_k. \tag{18}$$

The unbiased sample covariances $\tilde{r}_k$ are defined by

$$\tilde{r}_k = \tilde{r}_{-k} = \frac{1}{N-k}\sum_{t=1}^{N-k} y_{k+t}y_t. \tag{19}$$

The MoM estimate is obtained by replacing r_k by $\tilde{r}_k$ in (18) for $k = 1,\ldots,c+1$, and solving the resulting system of linear equations for $\rho_1,\ldots,\rho_c$, and σ_n^2. The MoM estimator for our problem can be shown to be unbiased and consistent [5].

Set $\hat{\mathbf{x}}_{1|0} = \mathbf{0}$,
Set $\mathbf{P}_{1|0} = \mathrm{block\ diag}(\rho_1\mathbf{\Gamma}_1,\ldots,\rho_c\mathbf{\Gamma}_c)$
For $t = 1$ to N do
  $\mathbf{K}_t = \mathbf{P}_{t|t-1}\mathbf{h}(\mathbf{h}^T\mathbf{P}_{t|t-1}\mathbf{h} + \sigma_n^2)^{-1}$
  $\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + \mathbf{K}_t(y_t - \mathbf{h}^T\hat{\mathbf{x}}_{t|t-1})$
  $\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{K}_t\mathbf{h}^T\mathbf{P}_{t|t-1}$
  $\hat{\mathbf{x}}_{t+1|t} = \mathbf{F}\hat{\mathbf{x}}_{t|t}$
  $\mathbf{P}_{t+1|t} = \mathbf{F}\mathbf{P}_{t|t}\mathbf{F}^T + \mathbf{G}\mathbf{Q}\mathbf{G}^T$
For $t = N$ to 1 do
  $\mathbf{A}_t = \mathbf{P}_{t|t}\mathbf{F}^T\mathbf{P}_{t+1|t}^{-1}$
  $\hat{\mathbf{x}}_{t|N} = \hat{\mathbf{x}}_{t|t} + \mathbf{A}_t(\hat{\mathbf{x}}_{t+1|N} - \mathbf{F}\hat{\mathbf{x}}_{t|t})$
  $\mathbf{P}_{t|N} = \mathbf{P}_{t|t} + \mathbf{A}_t(\mathbf{P}_{t+1|N} - \mathbf{P}_{t+1|t})\mathbf{A}_t^T$

Table 1: Rauch-Tung-Striebel optimal smoother.

4. PRELIMINARY RESULTS

To verify the proposed algorithm and test its performance, we conducted several Monte-Carlo experiments with simulated data. Here we present the results obtained for mixtures of two AR(2) components chosen from the dictionary of six AR models described in Table 2. They represent a combination of low-pass, high-pass, band-pass, and band-cut broadband processes.

In the first experiment, we generated 25 different realizations of length $N = 512$ of a ΣAR process for each of the $\binom{6}{2} = 15$ possible subsets of two components out of the dictionary. The innovation sequences for both AR processes had unit variance $\rho_1 = \rho_2 = 1$. The noise variance σ_n^2 was set equal to 0.01, which corresponds roughly

Table 2: AR(2) models parameter dictionary

i	1	2	3	4	5	6
$a_{i,1}$	1.5588	0.9000	0.0000	-0.9000	-1.5588	0.0000
$a_{i,2}$	0.8100	0.8100	0.8100	0.8100	0.8100	0.8100

to a SNR of 30 dB. We computed the value of the MDL criterion for all possible subsets of one, two, or three AR processes. There are thus $\binom{6}{1} + \binom{6}{2} + \binom{6}{3} = 41$ hypotheses for the subsets of AR components that are compared. A subset of two elements was selected 269 times out of $15 \times 25 = 375$. The two element subset selected was the correct subset 260 times. When a subset of three components was selected, it always contained the two correct AR models plus a spurious one. Subsets of one elements were never selected. The ΣAR processes most difficult to identify were the ones composed of the pair of AR processes $\{1,4\}, \{2,5\}, \{2,6\}$, or $\{4,6\}$ for which the error rate was close to 60%. The ΣAR processes composed of the pair of AR processes $\{1,5\}, \{2,4\}$, or $\{3,6\}$ were the most easily identified with an average error rate of 4%. It is possible to interpret this results in term of the spectral separation of the different AR models. Subset hypotheses that have "close" spectra are difficult to identify. Conversely, subset hypothesis that are well separated in the spectral domain are easier to identify [5].

The iterative EM algorithm with the Rauch-Tung-Striebel smoother run at each iteration is computationaly intensive. One way to reduce the computational load is to stop the EM algorithm after an arbitrary number of iterations $k_{\max}$, before complete convergence. If the initial value obtained by the method of moments is close enough to the maximizer, the variance estimate $\theta^{(k_{\max})}$ when the EM algorithm is stopped can be viewed as an approximate maximum likelihood estimate. Using this estimate in (5) will therefore give an approximate MDL criterion which can be used to select the desired subset. To verify this hypothesis, we conducted a second experiment to analyze the effect of the number of iterations of the EM algorithm on the classification error rate. The dictionary and parameters values were the same as in the first experiment. With only one iteration of the EM algorithm, a subset of two element was selected 256 times, and this subset was the correct subset 247 times. With no iteration at all, that is, by using directly the MoM estimate in (4), a subset of two elements was selected 247 times, and this subset was the correct subset 246 times. The ΣAR mixture leading to the highest and lowest error rates were the same as in the first experiment.

Other results with different dictionaries will be shown at the conference.

5. SUMMARY AND CONCLUDING REMARKS

In this paper we have considered the problem of detecting and classifying an unknown number of multiple simultaneous Gaussian AR signals with unknown variances given a finite length observation of their sum and a dictionary of candidate AR models. We have shown that the problem reduces to the estimation in the maximum likelihood sense of the variances of the AR components for every subset of AR models from the dictionary. The "best" subset is then found by applying the minimum description length principle. The ML estimator of the variances has been derived by combining the EM algorithm with the Rauch-Tung-Striebel optimal smoother. Another estimator based on the method of moments has also been proposed.

The selection of the optimal subset of components from the dictionary requires the evaluation of the MDL criterion for every possible combination of components. To avoid the combinatorial explosion associated with large dictionaries, it is necessary to resort to sub-optimal search strategies. This is similar to the issue of feature subset selection in pattern recognition [9], and the numerous algorithms that have been developed to address that problem can be applied. The search can be further accelerated by performing a first selection of possible subsets with only a few iterations of the EM algorithm. The results of this coarse search through the dictionary with an approximate MDL criterion can then be refined with a fully iterated EM algorithm to retain only the "best" subset.

In our formulation of the mixture decomposition problem, we have made the hypothesis that a dictionary of *fixed* parameters for the candidate AR models is available. The algorithms proposed here can be extended to treat the problem in a Bayesian framework. That is, the dictionary of fixed parameters can be replaced by a dictionary of *a priori distributions* for the parameters of the candidate AR models [5].

6. REFERENCES

[1] B. Porat, *Digital Processing of Random Signal, Theory and Methods,* Prentice-Hall, Inc., Englewood Cliffs, NJ, 1994.

[2] J. P. Woodard, "Modeling and Classification of Natural Sounds by Product Code Hidden Markov Models," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-40, pp. 1833–1835, July 1992.

[3] Y. Ephraim, "Statistical-Model-Based Speech Enhancement Systems," *Proc. IEEE,* vol. 80, pp. 1526–1555, Oct. 1992.

[4] J. Rissanen, "Modeling by Shortest Data Description," *Automatica,* vol. 14, pp. 465-471, 1978.

[5] C. Couvreur, "Estimation of Parameters and Classification of Mixtures of Autoregressive Processes," M.S. Thesis, University of Illinois at Urbana-Champaign, Jan. 1995.

[6] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the *EM* Algorithm," *Ann. Roy. Stat. Soc. (B),* vol. 39, pp. 1–38, Dec. 1977.

[7] P. S. Maybeck, *Stochastic Models, Estimation, and Control,* vol. 1, Academic Press, Orlando, FL, 1979.

[8] I. Ziskand and D. Hertz, "Multiple Frequencies and AR Parameters Estimation from One Bit Quantized Signal via the EM Algorithm," *IEEE Trans. Signal Processing,* vol. SP-41, no. 11, pp. 3202–3206, Nov. 1993.

[9] K. Fukunaga, *Introduction to Statistical Pattern Recognition,* 2d edition, Academic Press. San Diego, 1990.