PARAMETER ESTIMATION FOR AUTOREGRESSIVE GAUSSIAN-MIXTURE PROCESSES: THE EMAX ALGORITHM

Shawn M. Verbout *James M. Ooi* *Jeffrey T. Ludwig* *Alan V. Oppenheim**

Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

ABSTRACT

The problem of estimating parameters of discrete-time non-Gaussian autoregressive (AR) processes is addressed. The subclass of such processes considered is restricted to those whose driving noise samples are statistically independent and identically distributed according to a Gaussian-mixture probability density function (pdf). Because the likelihood function for this problem is typically unbounded in the vicinity of undesirable, degenerate parameter estimates, a global maximum likelihood approach is not appropriate. Hence, an alternative approach is taken whereby a finite local maximum of the likelihood surface is sought. This approach, which is termed the quasi-maximum likelihood (QML) approach, is used to obtain estimates of the AR parameters as well as the means, variances, and weighting coefficients that define the Gaussian-mixture pdf. A technique for generating solutions to the QML problem is derived using a generalized version of the expectation-maximization principle.

1. INTRODUCTION

Estimation of parameters of discrete-time non-Gaussian autoregressive (AR) processes has typically been approached using methods based on higher-order statistics (HOS) [5]. The approach developed in this paper is fundamentally different from the HOS approach in that it assumes a specific form for the probability density function (pdf) of the observed data, and is therefore entirely parametric. In particular, we consider processes that can be represented as the output of a linear time-invariant (LTI) AR system driven by noise samples that are statistically independent and identically distributed (i.i.d.) according to a Gaussian-mixture pdf (i.e., a pdf that is a weighted average of a finite number of Gaussian densities having arbitrary means and variances); we call such processes ARGMIX (AutoRegressive Gaussian-MIXture) processes. We seek estimates of the AR parameters jointly with the mixture parameters—i.e., the means, variances, and weighting coefficients—that define the Gaussian-mixture pdf. Joint maximum likelihood (ML) estimates have not been directly pursued in the past because the value of the likelihood function is infinite for certain known, degenerate parameter values. In general, these parameter values are not useful as estimates, even though, strictly speaking, they do maximize the likelihood function. However, as we shall see in this paper, strategies based on finding nondegenerate *local* maxima of the likelihood function yield solutions that are useful.

The Gaussian-mixture model is capable of closely approximating many densities, and has been considered by a number of researchers for this purpose (see, for example, [3], [6], [9]). Yet only a few researchers, most notably Sengupta and Kay [8] and Zhao et al. [10], have considered Gaussian-mixture models in conjunction with AR systems. Sengupta and Kay [8] address the problem of ML estimation of AR parameters for ARGMIX processes in which two Gaussian pdfs constitute the mixture, each with zero mean and known variance, but with unknown relative weighting. They use a conventional Newton-Raphson optimization algorithm that is initialized by the least-squares solution to find ML estimates for the AR parameters and the single weighting coefficient, and show that the performance of the ML estimate is superior to that of the standard forward-backward least-squares method. In a separate investigation, Zhao et al. [10] also consider ML estimation of the AR parameters of ARGMIX processes and derive a set of linear equations whose solution gives the ML estimate for the AR parameters when all the mixture parameters are known. When the mixture parameters are unknown, they combine these linear equations with an ad hoc clustering technique to produce an iterative algorithm for obtaining a joint estimate of both the AR parameters and the mixture parameters. They do not guarantee convergence of this algorithm or optimality of the estimate in any sense, but demonstrate empirically that the performance of their algorithm is superior to that of HOS-based methods in certain cases.

We use the expectation-maximization (EM) method to derive an iterative algorithm, called the EMAX algorithm (which stands for EM principle applied to ARGMIX processes), for jointly estimating the AR parameters and mixture parameters for ARGMIX processes. The EMAX algorithm finds local maxima of the likelihood function. We demonstrate that, when initialized appropriately, the estimates corresponding to these local maxima are desirable solutions, and hence that the likelihood function can still guide us to useful answers via its local maxima even though the ML estimation problem is degenerate.

2. PROBLEM FORMULATION

2.1. Notation

We adopt the convention of writing random variables in upper case and particular realizations of random variables in lower case. If X is a random variable, then we denote its pdf by $f_X(\cdot)$. If this density depends on a parameter θ then it is written $f_X(\cdot;\theta)$. Expectations associated with densities that depend on a parameter θ are analogously denoted by $E\{\cdot;\theta\}$. Vector-valued variables, both random and deterministic, are written in boldface. If $\mathbf{x}$ is an n-dimensional vector, then the ith element of $\mathbf{x}$ is denoted by $x_{(i)}$ for

*This research was supported in part by the Department of the Navy, Office of the Chief of Naval Research, contract number N00014-93-1-0686, in part by the U. S. Air Force Office of Scientific Research under Grant AFOSR-F49620-96-1-0072, and in part through collaborative participation in the Advanced Sensors Consortium sponsored by the U. S. Army Research Laboratory under Cooperative Agreement DAAL01-96-2-0001.

0-8186-7919-0/97 $10.00

$i = 1, \cdots, n$. Finally, we introduce the function definition $\mathcal{N}(\cdot; \mu, \sigma)$ to represent a Gaussian pdf with mean μ and standard deviation σ.

2.2. Statistical Signal Model

We consider a discrete-time scalar-valued random process $\{Y_t\}$ that satisfies the Kth-order autoregressive difference equation $Y_t = \sum_{k=1}^{K} a_k Y_{t-k} + V_t$, where $\{a_k\}_{k=1}^{K}$ are the real-valued AR coefficients of the process, and $\{V_t\}$ is a sequence (termed the driving process or driving noise) that consists of i.i.d. random variables having a Gaussian-mixture pdf defined by $f_V(v) = \sum_{i=1}^{M} \rho_i \ \mathcal{N}(v; \mu_i, \sigma_i)$, where the weighting coefficients $\{\rho_i\}_{i=1}^{M}$ satisfy $\rho_i \geq 0$ for $i = 1, 2, \cdots, M$ and $\sum_{i=1}^{M} \rho_i = 1$. Alternatively, we can express the tth sample of the driving process as $V_t = \sigma(\Phi_t) W_t + \mu(\Phi_t)$, where $\{W_t\}$ is a sequence of i.i.d., zero-mean, unit-variance Gaussian random variables, σ and μ are mappings defined by $\sigma(i) = \sigma_i$ and $\mu(i) = \mu_i$ for $i = 1, 2, \cdots, M$, $\{\Phi_t\}$ is a sequence of i.i.d. discrete-valued random variables distributed according to the probability law $\Pr(\Phi_t = i) = \rho_i$ for $i = 1, 2, \cdots, M$, and the processes $\{W_t\}$ and $\{\Phi_t\}$ are assumed statistically independent.

We assume that the order of the autoregression, K, and the number of constituent densities in the Gaussian-mixture pdf, M, are given, and that the parameters $\boldsymbol{\mu} = [\mu_1 \cdots \mu_M]^T$, $\boldsymbol{\sigma} = [\sigma_1 \cdots \sigma_M]^T$, $\boldsymbol{\rho} = [\rho_1 \cdots \rho_M]^T$, and $\mathbf{a} = [a_1 \cdots a_K]^T$ are unknown. Our observations for the random variables $Y_{-K}, \cdots, Y_{N-1}$ are the values $y_{-K}, \cdots, y_{N-1}$, respectively; from these observations, we wish to estimate the parameter vector $\boldsymbol{\Psi} = (\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\rho}, \mathbf{a})$. For notational convenience, we define the random vectors $\mathbf{Y} = [Y_0 \cdots Y_{N-1}]^T$ and $\mathbf{Y}_t = [Y_{t-1} \cdots Y_{t-K}]^T$ for $t = 0, 1, \cdots, N$, and denote the realizations of these vectors by $\mathbf{y}$ and $\mathbf{y}_t$, respectively.

2.3. Approach to Parameter Estimation

As mentioned earlier, we are not seeking a global ML estimate because in most cases degenerate estimates exist that have infinite likelihood. To see how such degenerate estimates can arise, one can easily verify that if we put, say, $\hat{a}_i = 0$ for $i = 1, \cdots, K$, $(\hat{\mu}_i, \hat{\sigma}_i, \hat{\rho}_i) = (0, 1, 1/M)$ for $i = 2, \cdots, M$, and $\hat{\mu}_1 = y_0$, and then let $\hat{\sigma}_1 \to 0$, then the likelihood function $f_{\mathbf{Y}_0, \mathbf{Y}}(\mathbf{y}_0, \mathbf{y}; \boldsymbol{\Psi}')$ will increase without bound. This assignment of parameter values corresponds to choosing the unknown AR system to be an identity system and one of the Gaussian densities in the mixture to be an impulse centered directly on one of the observations.

It is apparent that degenerate estimates are obtained only if one or more of the standard deviation estimates is chosen to be zero. We may be tempted to avoid this problem by restricting all of the standard deviation estimates to be greater than some prespecified positive threshold. However, if this minimum threshold is set too low, then meaningless estimates can arise when the largest likelihood value occurs on the boundary of the restricted parameter space near a singularity at which $\hat{\sigma}_i = 0$ for some i. Yet if the threshold is set too high, we risk excluding the best available estimate, since a component of the true Gaussian-mixture pdf may have a standard deviation smaller than the artificially set threshold.

One alternative to maximizing the likelihood function is to find the parameters that achieve the largest of the finite local maxima [3]. In general, no closed-form solution exists for this estimate, and a numerical method must typically be used. Because the likelihood surface usually has a large number of local maxima, classical optimization techniques cannot be guaranteed to find the largest local maximum. However, Titterington [9] has found that methods based on finding local maxima (not necessarily the largest finite local maximum) yield useful estimates. Accordingly, we take the approach of searching for local maxima of the likelihood function using the generalized EM algorithm.

More formally, if we let $\mathcal{P}$ denote the set of all possible values for the parameter vector $\boldsymbol{\Psi}$, then the estimate we seek for $\boldsymbol{\Psi}$ is any $\widehat{\boldsymbol{\Psi}}$ satisfying

$$\widehat{\boldsymbol{\Psi}} \in \overline{\arg\max}_{\boldsymbol{\Psi}' \in \mathcal{P}} \left\{ \log f_{\mathbf{Y}_0}(\mathbf{y}_0; \boldsymbol{\Psi}') + \log f_{\mathbf{Y}|\mathbf{Y}_0}(\mathbf{y}|\mathbf{y}_0; \boldsymbol{\Psi}') \right\}, \tag{1}$$

where the notation $\overline{\arg\max}_{x \in \mathcal{P}} \{g(x)\}$ stands for the set of all values in $\mathcal{P}$ achieving finite local maxima of g.

Since the estimate $\widehat{\boldsymbol{\Psi}}$ is defined in terms of the likelihood function, but is not obtained through a standard global maximization, we refer to this estimate as a quasi-maximum likelihood (QML) estimate. In the sequel, we shall assume that $N \gg K$, i.e., that the number of samples in the observed sequence is much greater than the number of AR parameters to be estimated. Under this assumption, we may, as is standard in the derivation of ML estimates for Gaussian AR processes, ignore the effect of the initial condition (represented by the first term of the log-likelihood function appearing on the right-hand side of (1)) and assume that a QML estimate is any $\widehat{\boldsymbol{\Psi}}$ satisfying

$$\widehat{\boldsymbol{\Psi}} \in \overline{\arg\max}_{\boldsymbol{\Psi}' \in \mathcal{P}} \left\{ \log f_{\mathbf{Y}|\mathbf{Y}_0}(\mathbf{y}|\mathbf{y}_0; \boldsymbol{\Psi}') \right\}. \tag{2}$$

3. SOLUTION VIA THE EM PRINCIPLE

The EM algorithm, which was first proposed by Dempster et al. [1], is an iterative technique for finding local maxima of likelihood functions. Although its convergence rate is slow, this algorithm converges reliably to a local maximum of the likelihood function under appropriate conditions, requires no derivatives of the likelihood function, and often yields equations that have an intuitively pleasing interpretation. The EM algorithm is best suited to problems in which there is a "complete" data specification $\mathbf{Z}$, from which the original observations $(\mathbf{Y}, \mathbf{Y}_0)$ can be derived, and such that the expectation $E\{\log f_{\mathbf{Z}}(\mathbf{Z}; \boldsymbol{\Psi}') \mid \mathbf{Y} = \mathbf{y}, \mathbf{Y}_0 = \mathbf{y}_0; \boldsymbol{\Psi}''\}$ can be easily computed for any two parameter vectors $\boldsymbol{\Psi}', \boldsymbol{\Psi}'' \in \mathcal{P}$. For our problem, we use the complete data specification $\mathbf{Z} = (\mathbf{Y}, \mathbf{Y}_0, \boldsymbol{\Phi})$, where $\boldsymbol{\Phi}$ is the vector of pdf-selection variables defined by $\boldsymbol{\Phi} = [\Phi_0 \cdots \Phi_{N-1}]^T$. With this choice of complete data, the EM algorithm as applied to our problem generates a sequence of estimates $\{\boldsymbol{\Psi}^{(s)}\}_{s=1}^{\infty}$ according to the recursive formula $\boldsymbol{\Psi}^{(s+1)} = \arg\max_{\boldsymbol{\Psi}' \in \mathcal{P}} U(\boldsymbol{\Psi}', \boldsymbol{\Psi}^{(s)})$, where

$$U(\boldsymbol{\Psi}', \boldsymbol{\Psi}^{(s)}) = E\{\log f_{\mathbf{Y}, \boldsymbol{\Phi}|\mathbf{Y}_0}(\mathbf{Y}, \boldsymbol{\Phi}|\mathbf{y}_0; \boldsymbol{\Psi}') \mid \mathbf{Y} = \mathbf{y}, \mathbf{Y}_0 = \mathbf{y}_0; \boldsymbol{\Psi}^{(s)}\}, \tag{3}$$

and some starting estimate $\boldsymbol{\Psi}^{(0)}$ must be chosen to initialize the recursion.

To derive the EMAX algorithm, we let $\boldsymbol{\Psi}' = (\boldsymbol{\mu}', \boldsymbol{\sigma}', \boldsymbol{\rho}', \mathbf{a}')$ and let

$$U_1(\boldsymbol{\rho}', \boldsymbol{\Psi}^{(s)}) = E\{\log f_{\boldsymbol{\Phi}|\mathbf{Y}_0}(\boldsymbol{\Phi}|\mathbf{y}_0; \boldsymbol{\rho}') \mid \mathbf{Y} = \mathbf{y}, \mathbf{Y}_0 = \mathbf{y}_0; \boldsymbol{\Psi}^{(s)}\} \tag{4}$$

$$U_2(\mathbf{a}', \boldsymbol{\mu}', \boldsymbol{\sigma}', \boldsymbol{\Psi}^{(s)}) = E\{\log f_{\mathbf{Y}|\boldsymbol{\Phi}, \mathbf{Y}_0}(\mathbf{y}|\boldsymbol{\Phi}, \mathbf{y}_0; \mathbf{a}', \boldsymbol{\mu}', \boldsymbol{\sigma}') \mid \mathbf{Y} = \mathbf{y}, \mathbf{Y}_0 = \mathbf{y}_0; \boldsymbol{\Psi}^{(s)}\}. \tag{5}$$

Then the EM recursion can be equivalently expressed as

$$\boldsymbol{\rho}^{(s+1)} = \arg\max_{\boldsymbol{\rho}'} U_1(\boldsymbol{\rho}', \boldsymbol{\Psi}^{(s)}) \tag{6}$$

$$(\mathbf{a}^{(s+1)}, \boldsymbol{\mu}^{(s+1)}, \boldsymbol{\sigma}^{(s+1)}) = \arg\max_{\mathbf{a}', \boldsymbol{\mu}', \boldsymbol{\sigma}'} U_2(\mathbf{a}', \boldsymbol{\mu}', \boldsymbol{\sigma}', \boldsymbol{\Psi}^{(s)}) \tag{7}$$

To find $\boldsymbol{\rho}^{(s+1)}$ so that (6) is satisfied, we first let $C_j(\boldsymbol{\Phi})$ be the number of times the symbol j appears in the vector $\boldsymbol{\Phi}$ and let

$$P_{t,j}(\boldsymbol{\Psi}') = \Pr\{\Phi_t = j \mid \mathbf{Y} = \mathbf{y}, \mathbf{Y}_0 = \mathbf{y}_0; \boldsymbol{\Psi}'\} \quad (8)$$

for all $\boldsymbol{\Psi}' \in \mathcal{P}$, for $t = 0, \cdots, N-1$ and $j = 1, \cdots, M$. Using these definitions, we can write

$$U_1(\boldsymbol{\rho}', \boldsymbol{\Psi}^{(s)}) = E\{ \log \prod_{j=1}^{M} \rho'^{C_j(\boldsymbol{\Phi})}_{(j)} \mid \mathbf{Y} = \mathbf{y}, \mathbf{Y}_0 = \mathbf{y}_0; \boldsymbol{\Psi}^{(s)}\} \quad (9)$$

$$= \sum_{j=1}^{M} \sum_{t=0}^{N-1} P_{t,j}(\boldsymbol{\Psi}^{(s)}) \log \rho'_{(j)}. \quad (10)$$

Then, using Jensen's inequality, the maximization in (6), which is over all $\boldsymbol{\rho}'$ such that $\rho'_{(j)} \geq 0$ and $\sum_{j=1}^{M} \rho'_{(j)} = 1$, has the solution

$$\rho^{(s+1)}_{(j)} = \frac{1}{N} \sum_{t=0}^{N-1} P_{t,j}(\boldsymbol{\Psi}^{(s)}). \quad (11)$$

To attempt the maximization in (7), we use the knowledge that the driving process is a sequence of i.i.d. Gaussian-mixture random variables to write the pdf for $\mathbf{Y}$ conditioned on $\boldsymbol{\Phi}$ and $\mathbf{Y}_0$ as

$$f_{\mathbf{Y}|\boldsymbol{\Phi},\mathbf{Y}_0}(\mathbf{y}|\boldsymbol{\Phi}, \mathbf{y}_0; \mathbf{a}', \boldsymbol{\mu}', \boldsymbol{\sigma}') = \prod_{t=0}^{N-1} \mathcal{N}(y_t - \mathbf{y}_t^T \mathbf{a}', \mu'_{(\Phi_t)}, \sigma'_{(\Phi_t)}) \quad (12)$$

Notice that the term $y_t - \mathbf{y}_t^T \mathbf{a}'$ represents the residual or prediction error obtained by using $\mathbf{a}'$ as the AR parameter vector. The function being maximized in (7) can then be written as

$$U_2(\mathbf{a}', \boldsymbol{\mu}', \boldsymbol{\sigma}', \boldsymbol{\Psi}^{(s)}) = - \frac{N}{2} \log 2\pi - \sum_{t=0}^{N-1} \sum_{j=1}^{M} P_{t,j}(\boldsymbol{\Psi}^{(s)}) \log \sigma'_{(j)} - \sum_{t=0}^{N-1} \sum_{j=1}^{M} P_{t,j}(\boldsymbol{\Psi}^{(s)}) \frac{(y_t - \mathbf{y}_t^T \mathbf{a}' - \mu'_{(j)})^2}{2\sigma'^2_{(j)}}. \quad (13)$$

Taking derivatives of this expression with respect to the quantities $\boldsymbol{\mu}'$, $\boldsymbol{\sigma}'$, and $\mathbf{a}'$ and setting the resulting expressions equal to zero yields three coupled nonlinear equations that define a stationary point of the right-hand side of (13). Because we are unable to solve these nonlinear equations analytically, it is difficult to find a global maximum. We instead use the method of coordinate ascent to numerically find a *local* maximum, resulting in a GEM algorithm rather than an EM algorithm. Coordinate ascent increases a multivariate function at each iteration by changing one variable at a time. If, at each iteration, the variable that is allowed to change is chosen to achieve the maximum of the function while the other variables are kept fixed, then coordinate ascent converges to a local maximum of the function [4]. Coordinate ascent is attractive because it is simple to maximize (13) separately over each variable. The coordinate-ascent algorithm is initialized by setting $\tilde{\mu}^{(0)}_{(j)} = \mu^{(s)}_{(j)}$ and $\tilde{\sigma}^{(0)}_{(j)} = \sigma^{(s)}_{(j)}$ for $j = 1, \cdots, M$, and setting $\tilde{\mathbf{a}}^{(0)} = \mathbf{a}^{(s)}$. It then proceeds iteratively by using the following formulas:

$$\tilde{\mu}^{(i+1)}_{(j)} = \arg\max_{\mu'_j} U_2(\tilde{\mathbf{a}}^{(i)}, \tilde{\mu}^{(i)}_{(1)}, .., \mu'_j, .., \tilde{\mu}^{(i)}_{(M)}, \tilde{\boldsymbol{\sigma}}^{(i)}, \boldsymbol{\Psi}^{(s)}) = \frac{\sum_{t=0}^{N-1} P_{t,j}(\boldsymbol{\Psi}^{(s)})(y_t - \mathbf{y}_t^T \tilde{\mathbf{a}}^{(i)})}{\sum_{t=0}^{N-1} P_{t,j}(\boldsymbol{\Psi}^{(s)})} \quad (14)$$

$$\tilde{\sigma}^{(i+1)}_{(j)} = \arg\max_{\sigma'_j} U_2(\tilde{\mathbf{a}}^{(i)}, \tilde{\mu}^{(i+1)}, \tilde{\sigma}^{(i)}_{(1)}, .., \sigma'_j, .., \tilde{\sigma}^{(i)}_{(M)}, \boldsymbol{\Psi}^{(s)}) = \sqrt{\frac{\sum_{t=0}^{N-1} P_{t,j}(\boldsymbol{\Psi}^{(s)})(y_t - \mathbf{y}_t^T \tilde{\mathbf{a}}^{(i)} - \tilde{\mu}^{(i+1)}_{(j)})^2}{\sum_{t=0}^{N-1} P_{t,j}(\boldsymbol{\Psi}^{(s)})}} \quad (15)$$

$$\tilde{\mathbf{a}}^{(i+1)} = \arg\max_{\mathbf{a}'} U_2(\mathbf{a}', \tilde{\boldsymbol{\mu}}^{(i+1)}, \tilde{\boldsymbol{\sigma}}^{(i+1)}, \boldsymbol{\Psi}^{(s)}) = \left[\sum_{t=0}^{N-1} \sum_{j=1}^{M} \frac{P_{t,j}(\boldsymbol{\Psi}^{(s)})}{(\tilde{\sigma}^{(i+1)}_{(j)})^2} \mathbf{y}_t \mathbf{y}_t^T \right]^{-1} \left[\sum_{t=0}^{N-1} \sum_{j=1}^{M} \frac{P_{t,j}(\boldsymbol{\Psi}^{(s)})}{(\tilde{\sigma}^{(i+1)}_{(j)})^2} (y_t - \tilde{\mu}^{(i+1)}_{(j)}) \mathbf{y}_t \right]. \quad (16)$$

If this recursion is iterated for $i = 0, \cdots, J-1$, then we define our parameter updates by $\mathbf{a}^{(s+1)} = \tilde{\mathbf{a}}^{(J)}$, $\boldsymbol{\mu}^{(s+1)} = \tilde{\boldsymbol{\mu}}^{(J)}$, $\boldsymbol{\sigma}^{(s+1)} = \tilde{\boldsymbol{\sigma}}^{(J)}$. For sufficiently large values of J, the updated parameters are, for practical purposes, local maxima of (13). Since the EMAX algorithm is a GEM algorithm that chooses the updated parameter estimates to be local maxima of (13), it converges to a stationary point. In summary, then, a single iteration of the EMAX algorithm consists of computing $\{P_{t,j}(\boldsymbol{\Psi}^{(s)})\}$, applying (11), and iterating (14)–(16) until convergence.

4. NUMERICAL EXAMPLE: AR PROCESS WITH LAPLACIAN DRIVING NOISE

In many applications, we would like to obtain ML estimates for the parameters of an AR system, but the ML problem is ill-posed because the marginal pdf characterizing the driving noise is unknown. In certain cases, however, it may be reasonable to assume that the true marginal pdf is accurately modeled by a Gaussian-mixture pdf. For such cases, if we process our observations with the EMAX algorithm, then we might expect the EMAX algorithm to find the mixture parameters that yield a good approximation to the true driving-noise pdf and simultaneously to produce good approximations to the ML estimates for the AR parameters. With the present example we demonstrate the validity of this approach to the ML estimation problem. In particular, we consider the parameter estimation problem for a fifth-order AR process whose AR coefficients are given by $(a_1, a_2, a_3, a_4, a_5) = (1.934, -2.048, 1.072, -0.340, 0.027)$. The driving noise for this process consists of i.i.d. samples distributed according to a Laplacian pdf defined by $f_V(v) = (1/2\beta) \exp\{-|v|/\beta\}$, where the scale parameter β (which is related to the standard deviation σ for this density by $\sigma = \sqrt{2}\beta$) was put at $\beta = 5$.

To find parameter estimates for this problem with the EMAX algorithm, we fixed the number of Gaussian densities in the mixture at $N = 3$ and constrained the means of these constituent densities to be zero. We used the following simple method for generating an initial parameter estimate: The vector $\mathbf{a}^{(0)}$ was computed using the forward-backward least-squares technique from traditional AR signal analysis. Each of the M elements of $\boldsymbol{\sigma}$ was randomly generated according to a uniform pdf having region of support $[0, \max_t\{v^{(0)}_{(t)}\} - \min_t\{v^{(0)}_{(t)}\}]$, where $v^{(0)}_{(t)}$ is the tth element of the residual sequence $\mathbf{v}^{(0)}$ produced by applying the filter $1 - \sum_{i=1}^{K} a^{(0)}_{(i)} z^{-i}$ to the sequence of observations. The elements of the weighting coefficient vector $\boldsymbol{\rho}^{(0)}$ were all set equal to $1/M$. Finally, because the means were assumed to be zero, $\boldsymbol{\mu}^{(s)}$ was set to zero for all $s \geq 0$.

	True Value	Sample Mean (LS)	Sample Mean (EMAX)	Sample MSE (LS) ×.001	Sample MSE (EMAX) ×.001
a_1	1.934	1.931	1.932	1.075	0.630
a_2	-2.048	-2.041	-2.045	5.157	2.878
a_3	1.072	1.065	1.069	8.400	4.777
a_4	-0.340	-0.336	-0.338	4.971	2.919
a_5	0.027	0.026	0.026	1.051	0.629

Table 1. Sample means and sample mean square error values for parameter estimates produced by the least squares (LS) method and the EMAX algorithm.

We performed a total of 500 trials. On each trial, a sequence of 1000 data points was generated and processed with the EMAX algorithm. The sample means and sample mean square error (MSE) values of the parameter estimates produced by the EMAX algorithm are presented in Table 1. To provide a convenient point of reference, we have also shown the sample means and sample MSE values of the AR parameter estimates given by the classical forward-backward least-squares method.

Observe from the table that, for each AR parameter, the ratio of the MSE of the least-squares estimate to that of the EMAX estimate ranges approximately from 1.7 to 1.8. The superior performance of the EMAX algorithm may be attributed to the ability of its assumed Gaussian-mixture pdf to closely approximate the Laplacian pdf, as is shown for a typical case in Figure 1(a). It is clear from this figure that the approximation is very good over the interval containing most of the driving-noise samples. However, since the number of Gaussian densities in the mixture is finite, an accurate model for the Laplacian density may be obtained only over a finite region of support. Eventually, the tails of the Gaussian-mixture pdf become bounded by a function of the form $k_1 \exp\{-k_2 v^2\}$ for appropriately chosen constants k_1 and k_2. Indeed, Figure 1(b) reveals this phenomenon with the aid of a log-magnitude scale.

As a further test of the EMAX algorithm, we compared its performance to that of the true ML estimator for this problem. It can be shown [2] that if the samples of the driving noise for an AR process are i.i.d. and Laplacian, then the ML estimate for the AR parameter vector $\mathbf{a}$ is given by the value of $\mathbf{a}'$ that minimizes the sum of absolute residuals $\sum_{t=0}^{N-1} |y_t - \mathbf{y}_t^T \mathbf{a}'|$. An algorithm for finding such a value for $\mathbf{a}'$ was proposed by Schlossmacher [7]; this algorithm is based on the method of iteratively reweighted least squares and is therefore easy to implement on a computer. The experimental results obtained using this algorithm are given in Table 2. A comparison of the sample MSE values in Table 2 with those given in Table 1 reveals that the ML estimator performs only slightly better than the EMAX algorithm, despite the fact that the ML estimator is able to exploit exact knowledge of the driving-noise pdf.

5. CONCLUSION

We have presented a general iterative technique known as the EMAX algorithm for estimating the parameters of a process that can be represented as the output of an autoregressive LTI system driven by a sequence of i.i.d. random

	True Value	Sample Mean (ML)	Sample MSE (ML) ×.001
a_1	1.934	1.933	0.577
a_2	-2.048	-2.045	2.725
a_3	1.072	1.070	4.289
a_4	-0.340	-0.339	2.423
a_5	0.027	0.027	0.540

Table 2. Sample means and sample mean square error values for parameter estimates produced the ML estimation algorithm of Schlossmacher.

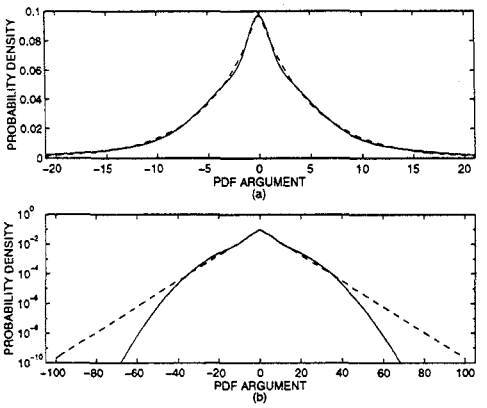

Figure 1. Laplacian pdf (dashed curve) for driving process of the example and a typical estimate of the pdf (solid curve) produced by the EMAX algorithm: (a) linear-magnitude scale (with horizontal axis spanning ±3 standard deviations), and (b) log-magnitude scale (with horizontal axis spanning ±15 standard deviations).

variables having a Gaussian-mixture pdf. The Gaussian-mixture assumption for the driving-noise pdf provides a convenient and general parametric framework for analyzing non-Gaussian AR signals. Although the likelihood function associated with ARGMIX processes is typically unbounded in the vicinity of undesirable, degenerate parameter values, we have seen in our numerical example that good estimates can still be obtained by searching for finite local maxima of the likelihood surface. The EMAX algorithm, by finding these local maxima, provides a useful way of exploiting the ARGMIX model to obtain high-quality estimates of signal parameters.

REFERENCES

[1] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society*, Series B, pp. 1-38, 1977.

[2] N. R. Draper and H. Smith, *Applied Regression Analysis*, John Wiley and Sons, New York, 1950.

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

[4] D. G. Luenberger, *Linear and Nonlinear Programming*, second edition, Addison Wesley, Reading, MA, 1984.

[5] C. L. Nikias, A. P. Petropulu, *Higher-Order Spectra Analysis: A Nonlinear Signal Processing Framework*, Prentice-Hall, Englewood Cliffs, NJ, 1993.

[6] R. A. Redner and H. F. Walker, "Mixture densities, maximum likelihood and the EM algorithm," *SIAM Review*, vol. 26, no. 2, April 1984.

[7] E. J. Schlossmacher, "An interactive technique for absolute deviations curve fitting," *Journal of the American Statistical Association*, vol. 68, no. 344, December 1973.

[8] D. Sengupta and S. Kay, "Efficient estimation of parameters for non-Gaussian autoregressive processes," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 6, June 1989.

[9] D. M. Titterington, A. F. M. Smith, and U. E. Makov, *Statistical Analysis of Finite Mixture Distributions*, John Wiley and Sons, Chichester, 1985.

[10] Y. Zhao, X. Zhuang, and S.-J. Ting, "Gaussian mixture density modeling of non-Gaussian source for autoregressive process," *IEEE Transactions on Signal Processing*, vol. 43, no. 4, April 1995.