A stochastic approximation type EM algorithm for the mixture problem

Gilles Celeux, Jean Diebolt

UNITÉ DE RECHERCHE
INRIA-ROCQUENCOURT

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
B.P. 105
78153 Le Chesnay Cedex
France
Tél.:(1) 39 63 55 11

Rapports de Recherche

N° 1383

Programme 5
Traitement du signal, Automatique et Productique

A STOCHASTIC APPROXIMATION TYPE EM ALGORITHM FOR THE MIXTURE PROBLEM

Gilles CELEUX
Jean DIEBOLT

Janvier 1991

A STOCHASTIC APPROXIMATION TYPE EM ALGORITHM FOR THE MIXTURE PROBLEM

UNE APPROXIMATION STOCHASTIQUE DE L'ALGORITHME EM POUR L'ESTIMATION DE MELANGES

Gilles CELEUX
INRIA, Rocquencourt

Jean DIEBOLT
CNRS, Paris 6

Abstract: The EM algorithm is a widely applicable approach for computing maximum likelihood estimates for incomplete data. We present a stochastic approximation type EM algorithm: SAEM. This algorithm is an adaptation of the stochastic EM algorithm (SEM) that we have previously developed. Like SEM, SAEM overcomes most of the well-known limitations of EM. Moreover, SAEM performs better for small samples. Furthermore, SAEM appears to be more tractable than SEM, since it provides almost sure (a.s.) convergence, while SEM provides convergence in distribution. Here, we focus on the mixture problem. We state a theorem which asserts that each SAEM sequence converges a.s. to a local maximizer of the likelihood function and we give the asymptotic rate of convergence. We close this paper with Monte-Carlo experiments which show that SAEM performs better than EM and SEM for small samples.

Keywords : maximum likelihood, mixture, simulated annealing, stochastic algorithm.

Résumé : L'algorithme EM est très répandu pour l'estimation par le maximum de vraisemblance de paramètres de modèles où les données sont incomplètes. Nous présentons une nouvelle version stochastique de l'algorithme EM. Cet algorithme, désigné algorithme SAEM, est une adaptation de l'algorithme stochastique SEM que nous avons précédemment développé. Comme ce dernier, l'algorithme SAEM répond aux limitations bien connues de l'algorithme EM ; mais de plus il se comporte mieux pour traiter de petits échantillons. Par ailleurs, il est plus simple à appréhender que l'algorithme SEM dans la mesure où il converge presque sûrement tandis que l'algorithme SEM converge en loi. Ici, on limite la présentation détaillée de l'algorithme SAEM au problème des mélanges de lois de probabilité. On établit un théorème qui assure que toute suite d'estimés par SAEM converge p.s. vers un maximum local de la fonction de vraisemblance et précise sa vitesse de convergence asymptotique. On conclut cet article par des simulations de Monte-Carlo qui montrent que SAEM se comporte mieux que EM et SEM pour de petits échantillons.

Mots-clés : maximum de vraisemblance, algorithme stochastique, mélange, recuit simulé.

1. Introduction

The EM algorithm is a widely applicable approach for computing maximum likelihood (m.l.) estimates for incomplete data. The first description and study of the EM algorithm dates back to Dempster, Laird and Rubin [5]. Since this date, some authors have examined with great attention the convergence properties of EM: e.g., Wu [12] and Redner and Walker [10]. EM has been successfully applied to a wide variety of problems. Moreover, Louis [7] and Meilijson [9] among others have investigated methods of acceleration of its convergence. A detailed account of EM focusing on the mixture problem is given in Section 2.

Despite several appealing features, the EM algorithm has severe limitations. Its limiting position is strongly dependent on its starting position. It has been known in some instances to converge to a saddle point of the likelihood function (l.f.) L. In some cases, its rate of convergence is unbearably slow. Moreover, in many important situations, certain key parameters of the model must be known before running the EM algorithm. If the value of one of these parameters is wrongly set, then EM fails in the sense that, generally, it does not even indicate that some assumptions on these parameters are not actually true. For instance, in the mixture context, an important but delicate parameter of the model is the number of components K of the mixture. Testing for the number of components in a mixture is an important but very difficult problem which has not been completely resolved (see, e.g., McLachlan and Basford [8], p. 21-29, and references therein).

In an attempt to overcome some of the above limitations of EM, we have defined and studied a stochastic version of EM, that we have called the SEM algorithm, in Celeux and Diebolt [2], [3] and [4]. In these studies, we paid special attention to the mixture case.

Our proposal for avoiding the problems of wrong (or slow) convergence of EM was to incorporate a stochastic step between the E and the M steps, which we called the S step, hence the name SEM.

The S step prevents the sequence $\{\phi^n\}$ from being immobilized near an unstable stationary point of the log-likelihood function. In section 3, we briefly describe the SEM algorithm and provide the detailed formulas for the mixture case.

But the usage of SEM has also some drawbacks. Firstly, since it converges weakly to a stationary probability distribution, it needs experience to handle its numerical results. Secondly, it tends to give unreliable results for small sample sizes, because in this case the random perturbations of the S step are too important.

An alternative to the SEM algorithm is the subject of the present paper. It consists in a stochastic approximation type algorithm derived from SEM, in which the variance of the random drawings is to decrease to zero as the number of iterations increases to infinity. We have called it the SAEM algorithm (Stochastic Approximation EM). SAEM has a median position between EM and SEM. A detailed description of SAEM will be found in Section 4.

In this paper, we focus on the mixture case. The main result of this paper is Theorem 1, which asserts that in the particular case of mixtures of densities from exponential families and under technical assumptions, the sequence generated by SAEM converges almost surely (a.s.) to a local maximizer of the l.f. This theorem will be stated and proved in Section 5. Finally, in Section 6, we will report numerical simulations where we compare the behaviour of EM, SEM and SAEM in the case of univariate Gaussian mixtures for small sample sizes.

The results of our Theorem 1 and of the above-mentioned simulations suggest that SAEM is likely to generally converge to the global maximizer of the l.f., whatever its initial position. This leads us to the conclusion that SAEM is a fruitful line of approach.

Since we will only consider the behaviour of SAEM in the mixture context, we give a short account of the mixture problem. The observed $\mathbf{R}^d$-valued sample $\mathbf{x} = (x_1,...,x_N)$ is assumed to be drawn from the mixture density

$$h(x) = \sum_{k=1}^{K} p_k \, h(x, a_k), \tag{1.1}$$

where the mixing weights p_k satisfy $0 < p_k < 1$ and sum to one and the densities $h(x, a_k)$ are distinct members from the same exponential family. Note that the component densities of most of the mixtures of interest are members of exponential families. The number of components K is assumed known. Note that the functions $h(x, a_k)$ are densities with respect to a measure which is either Lebesgue measure or counting measure on some finite or countably infinite subset of $\mathbf{R}^d$. This allows us to treat mixtures of discrete or continuous distributions in one exercise. The generic density h(x,a) has the form

$$h(x, a) = D(a)^{-1}\, n(x) \exp\{a^T b(x)\}, \qquad (1.2)$$

where a is a vector of dimension s, a^T denotes the transpose of a, $n : \mathbf{R}^d \to \mathbf{R}$ and $b : \mathbf{R}^d \to \mathbf{R}^s$ and $D(a) = \int n(x) \exp\{a^T b(x)\}\, d\mu$ is a normalizing factor, for an appropriate underlying measure μ on $\mathbf{R}^d$. Within this framework, the parameter $\phi = (p_1, \dots, p_{K-1}, a_1, \dots, a_K)$ lies in a subset of the (K-1 + sK)-dimensional Euclidian space **E**. The true parameter ϕ of the mixture is to be estimated.

2. The EM algorithm

The EM algorithm is designed to find iteratively m.l. estimates in the context of parametric models when the observed data can be viewed as incomplete. More precisely, the observed sample **x** is assumed to be drawn from a parametrized family of probability density functions $\mathbf{f}(\mathbf{x};\phi)$ defined on the sample space **X** equipped with some reference σ-finite measure. This observed sample **x** is assumed to be the image through a many-to-one mapping π: $\mathbf{Y} \to \mathbf{X}$ of a complete sample **y** drawn from a density $\mathbf{g}(\mathbf{y};\phi)$ defined on the sample space **Y**. In all relevant situations, the space **Y** can be split into the product **X** x **Z**, where **Z** denotes the space of the unobserved samples **z**, the mapping π is the projection on the first factor and $\mathbf{f}(\mathbf{x};\phi)$ and $\mathbf{g}(\mathbf{y};\phi)$ are related by

$$\mathbf{f}(\mathbf{x};\phi) = \int_{\mathbf{Z}} \mathbf{g}\{(\mathbf{x},\mathbf{z});\phi\}\, d\mathbf{z}. \qquad (2.1)$$

The EM algorithm is directed at finding the global maximizer, or at least a local maximizer, of the likelihood function (l.f.) $\mathbf{f}(\mathbf{x};\phi)$ by taking advantage of the fact that the m.l. estimate of the complete data $\mathbf{y} = (\mathbf{x},\mathbf{z})$ can usually be expressed in closed form. The EM method replaces the maximization of the unknown l.f. $\mathbf{g}(\mathbf{y};\phi)$ of the complete data by successive maximizations of the conditional expectation of log $\mathbf{g}(\mathbf{y};\phi)$ given **x** for the current fit ϕ^n of the parameter. Let $\mathbf{k}(\mathbf{z}|\mathbf{x};\phi) = \mathbf{g}\{(\mathbf{x},\mathbf{z});\phi\}/\mathbf{f}(\mathbf{x};\phi)$ denote the conditional density of **z** given **x** and $Q(\phi';\phi)$ denote the conditional expectation of log $\mathbf{g}(\mathbf{y};\phi')$ given **x** for the value ϕ of the parameter:

$$Q(\phi';\phi) = E\left\{\log \mathbf{g}(\mathbf{y};\phi')|\mathbf{x};\phi\right\} = \int_{\mathbf{Z}} \mathbf{k}(\mathbf{z}|\mathbf{x};\phi) \log \mathbf{g}\{(\mathbf{x},\mathbf{z});\phi'\}\, d\mathbf{z}. \qquad (2.2)$$

The expectation $Q(\phi';\phi)$ is assumed to be well defined for every ϕ and ϕ' in the parameter space.

Starting from an initial position ϕ^0, the standard iteration $\phi^n \to \phi^{n+1}$ of the EM algorithm consists in the steps E and M.

E step : Compute $Q(\phi;\phi^n)$.

M step : Choose ϕ^{n+1} to maximize $Q(\phi;\phi^n)$ in ϕ.

The basic property of EM is that, if we define $L(\phi) = \log \mathbf{f}(x;\phi)$, then $L\{T(\phi)\} \geq L(\phi)$ for each ϕ.

We now turn to a detailed description of EM for mixtures of densities all belonging to some exponential family. First, we describe the incomplete data structure of the problem. The complete sample $\mathbf{y} = (\mathbf{x},\mathbf{z}) = \{(x_i, z_i),\ i = 1, ..., N\}$, where the vector of indicator variables $z_i = (z_{ij}, j=1,...,K)$ is defined by $z_{ij} = 1$ or 0 according as x_i has been drawn from the density $h(x,a_j)$ or not. The random variables $z_1, ..., z_N$ are i.i.d. following a multinomial distribution consisting of one draw of K categories with probabilities $p_1,...,p_K$, respectively. Owing to independence, the conditional density $\mathbf{k}(\mathbf{z}|\mathbf{x};\phi)$ of $\mathbf{z}$ given $\mathbf{x}$ can be split into the product $\prod_i k(z_i|x_i;\phi)$ where

$$k(z_i|x_i;\phi) = \frac{p(z_i)\, h(x_i,a(z_i))}{\sum_{j=1}^{K} p_j\, h(x_i,a_j)}, \tag{2.3}$$

with $p(z_i) = p_j$ and $a(z_i) = a_j$ if $z_{ij} = 1$. We define the probability a posteriori that x_i has been drawn from the *j*th mixture component by

$$t_j(x_i) = k(z_i \mid x_i ; \phi) \text{ if } z_{ij} = 1. \tag{2.4}$$

Since the mixture component densities are from some exponential family, it follows from (2.3) and (2.4) that the $t_j(x_i)$'s are C^∞ functions of ϕ. In this particular context, (2.1) takes the form

$$\mathbf{f}(\mathbf{x};\phi) = \prod_{i=1}^{N} \sum_{j=1}^{K} p_j\, h(x_i,a_j) \tag{2.5}$$

Hence, the log-likelihood function $L(\phi) = \log \mathbf{f}(\mathbf{x} ; \phi)$ is a C^∞ function of ϕ. Similary, (2.2) takes the form

$$Q(\phi';\phi) = \sum_{i=1}^{N} \sum_{j=1}^{K} t_j(x_i)\, \{\log p'_j + \log h(x_i,a'_j)\}. \tag{2.6}$$

The *n*th iteration $\phi^{n+1} = T_N(\phi^n)$ of EM can be summarized as follows (see (5.3) p. 222 of Redner and Walker [10] or Titterington, Smith and Makov [11]).

E step : Compute the probabilities a posteriori $t_j^n(x_i)$, $i=1, ..., N$ and $j = 1, ..., K$, using (2.3) and (2.4), as

$$t_j^n(x_i) = \frac{p_j^n\, h(x_i,a_j^n)}{\sum_{r=1}^{K} p_r^n\, h(x_i,a_r^n)}. \tag{2.7}$$

M step: Compute

$$p_j^{n+1} = \sum_{i=1}^{N} t_j^n(x_i) /N, \quad j= 1, ..., K, \tag{2.8}$$

and

$$a_j^{n+1} = \frac{\sum_{i=1}^{N} t_j^n (x_i)\, b(x_i)}{\sum_{i=1}^{N} t_j^n (x_i)}, \quad j = 1, ..., K \tag{2.9}$$

Clearly, T_N is C^∞. In view of (2.9), the estimates a_j^n , $n \geq 1$, remain in the convex hull B of $b(x_1), ..., b(x_N)$. Thus, the estimates ϕ^n, $n \geq 1$, remain in the bounded convex subset

$$C = (0, 1)^{K-1} \text{ x } B^K \text{ of } \mathbf{E}. \tag{2.10}$$

The EM sequence $\{\phi^n\}$ can be viewed as generated by a discrete-time dynamical system $\phi^{n+1} = T_N (\phi^n)$, where T_N denotes the nonlinear operator of the EM algorithm. Here and in the sequel, the subscript N indicates dependence on the sample $\mathbf{x} = (x_1, ..., x_N)$. Furthermore, $L(\phi)$ can be thought of as a Lyapunov function of this dynamical system. In all pertinent applications, no working analytic expression of T_N is available. Hence, any study of the convergence properties of $\{\phi^n\}$ involves assuming hypotheses on T_N and $Q(\phi'; \phi)$.

Wu's [12] very careful theoretical study reveals that no general convergence result for the EM sequence can be derived, except under strict assumptions. In the case of mixtures of densities all belonging to some exponential family, Redner and Walker [10], Theorem 5.2 (p. 223), have proved the following local result: the sequence $\{\phi^n\}$ is ensured to converge to the consistent m.l. estimator ϕ_N of ϕ, provided that the initial position ϕ^0 is close enough to ϕ_N.

This result is of theoretical primary importance. Its local nature highlights the main limitations of EM, since it does not guarantee that the sequence $\{\phi^n\}$ converges to ϕ_N as $n \to \infty$, whatever its starting point. Numerous numerical experiments report that the limiting position of EM can actually depend greatly on its initial position (see, e.g., [2], [8] and [11]) and happens to be a saddle-point of the likelihood function.

Two additional mathematical results of Redner and Walker [10] will be used in the proof of our Theorem 1. Their Theorem 4.1 (p. 218) and Theorem 5.1 (p. 223) state that, in this particular mixture case with all densities belonging to the same exponential family, all the fixed points of T_N are stationary points of L and $L\{T_N(\phi)\} > L(\phi)$ whenever $T_N(\phi) \neq \phi$.

3. The SEM algorithm

The SEM algorithm has been designed to give an answer to the fundamental limitations of EM mentioned in Section 2. It incorporates a stochastic step (S step) between the E and the M steps. This S step is directed by the following **Random Imputation Principle** (RIP): generate a completed sample $\mathbf{y}^n = (\mathbf{x}, \mathbf{z}^n)$ by drawing it at random from the conditionnal density $\mathbf{k}(\mathbf{z} \mid \mathbf{x} ; \phi^n)$ given the observed sample $\mathbf{x}$ and for a current fit ϕ^n of the parameter.

Once the S step has been performed, the new estimate ϕ^{n+1} of the parameter is the m.l. estimate computed on the basis of the completed sample $\mathbf{y}^n = (\mathbf{x}, \mathbf{z}^n)$ and, in all relevant situations, an analytic expression of ϕ^{n+1} as a function of $\mathbf{y}^n$ can easily be derived in closed form. In this section, we outline the main properties of SEM in the mixture context. First, we describe the three steps of the standard SEM iteration $\phi^n \rightarrow \phi^{n+1}$, starting from the initial value ϕ^0.

E step : Compute the conditional density $\mathbf{k}(\mathbf{z}|\mathbf{x};\phi^n)$;

S step : According to the above mentioned RIP, draw at random a complete sample $\mathbf{y}^n = (\mathbf{x},\mathbf{z}^n)$ from $\mathbf{k}(\mathbf{z}|\mathbf{x};\phi^n)$;

M step : Define ϕ^{n+1} to be the solution of the m. l. equations based on the completed sample $\mathbf{y}^n = (\mathbf{x}, \mathbf{z}^n)$, i. e. solve $\partial\, \mathbf{g}\, (\mathbf{y}^n ; \phi) / \partial\, \phi = 0$. In all relevant situations, the M step can be expressed in closed form.

In the context of mixtures of densities from the exponential family (1.2), the nth iteration $\phi^n \rightarrow \phi^{n+1}$ of SEM can be summarized as follows.

E step: Compute $t_j^n(x_i)$, i = 1, ..., N and j = 1, ..., K, using (2.7).

S step: For i = 1, ..., N, draw independently the random indicator variable z_i^n from a multinomial distribution consisting of one draw of K categories with probabilities $t_1^n(x_i)$, ..., $t_K^n(x_i)$. If

$$N^{-1} \sum_{i=1}^{N} z_{ij}^n \geq c(N) \text{ for all } j = 1, ..., K, \tag{3.1}$$

then the M step described below can be accomplished. Here, c(N) is a threshold satisfying $0 < c(N) < 1$ and $c(N) \rightarrow 0$ as $N \rightarrow \infty$. The role of the condition (3.1) is to avoid the occurence of numerical singularities. For instance, for the normal mixture problem, c(N) must be chosen $\geq (d + 1) / N$ so that, with probability 1, the covariance matrices derived from (3.3) are nondegenerate. On the other hand, for a mixture of exponential distributions, $c(N) \geq 1 / N$. Now, if $N^{-1} \sum_i z_{ij}^n < c(N)$ for some j, $1 \leq j \leq K$, then we perform the S-step again as follows. We draw the new $z_{ij}^{n'}$ s from some *ad hoc* preassigned distribution on $\mathbf{Z}$ designed to ensure that condition (3.1) holds.

M step : Compute ϕ^{n+1} as follows:

$$p_j^{n+1} = \sum_{i=1}^{N} z_{ij}^n / N \quad \text{for } j = 1, ..., K \tag{3.2}$$

and

$$a_j^{n+1} = \frac{\sum_{i=1}^{N} z_{ij}^n \; b(x_i)}{\sum_{i=1}^{N} z_{ij}^n} \quad \text{for } j = 1, ..., K. \tag{3.3}$$

The formulas (3.2) and (3.3) have been derived from (2.8) and (2.9) by substituting z_{ij}^n for $t_j^n(x_i)$ according to the RIP. It results from (3.1) that the SEM estimates ϕ^n remain in the compact convex subset

$$H_N = [c(N), 1 - c(N)]^{K-1} \times B^K \tag{3.4}$$

of E. Recall that E denotes the Euclidian space where the parameters $\phi = (p_1, \ldots, p_{K-1}, a_1, \ldots, a_K)$ take their value, B denotes the convex hull of $b(x_1), \ldots, b(x_N)$ and b(x) is defined in (1.2). The process $\{\phi^n\}$ generated by the SEM iterations n = 1, 2, ... on the basis of a given sample $x_1, \ldots, x_N$ of size N is an homogeneous Markov chain for which ergodicity usually holds. In the mixture case, the ergodicity of $\{\phi^n\}$ has been established in Celeux and Diebolt [3]. When ergodicity holds, ϕ^n converges in distribution, as the iteration index $n \to \infty$, to the unique stationary distribution Λ_N, which is supported by a subset of H_N (see (3.4)). Since the sample $x_1, \ldots, x_N$ is fixed, ϕ^n cannot be expected to converge in a stronger way (e.g., in probability or wp 1).

Thus, in the strictest sense, the estimate provided by SEM is not pointwise and is nothing but the probability distribution Λ_N. A natural pointwise estimator of ϕ derived from Λ_N is the mean $\hat{\phi}_N$ of Λ_N. Note that there is no reason why $\hat{\phi}_N$ should be equal to the m.l. consistent estimator ϕ_N of the parameter ϕ (for the existence of a m.l. consistent estimator in the mixture case, see Redner and Walker [10]).

Under technical conditions on the EM operator T_N, the more restrictive of which is that T_N has only one stable fixed point, Celeux and Diebolt [3] have established that, if X denotes a r.v. drawn from the stationary distribution Λ_N, then $N^{1/2}(X - \phi_N)$ converges in distribution, as the sample size $N \to \infty$, to a Gaussian r.v. with mean 0 and regular variance matrice Γ which can be expressed in terms of the true mixture parameters. Moreover, Celeux and Diebolt [3] have proved that, under the same assumptions, $\hat{\phi}_N - \phi_N = O(N^{-1/2})$ and the variance matrix of Λ_N takes the form $\Gamma/N + o\,(1/N)$ as $N \to \infty$. As a consequence, it results that for almost all samples, the distribution $\Lambda_N = \Lambda(x_1, \ldots, x_N)$ converges to the Dirac distribution δ_ϕ in the weak topology as $N \to \infty$.

Empirical means of the ϕ^n's when n is large enough provide satisfactory approximations of $\hat{\phi}_N$. Numerous numerical experiments (see, e.g., [2]) show that $\hat{\phi}_N$ is close to ϕ_N for moderate or large sample size. These experiments also show that SEM overcomes most of the limitations of EM, provided that the sample size is not too small. In contrast, when the sample size is small, the magnitude of the random perturbations is too large. In such a situation, the sequence $\{\phi^n\}$ generated by SEM hits the boundary of H_N very often. Each time such an event happens, ϕ^n starts afresh from a preassigned distribution concentrated on H_N, as explained in the above description of the SEM algorithm in the mixture context. Thus, the stationary distribution Λ_N cannot take into account much information on the underlying EM operator T_N. In such a situation, the mean $\hat{\phi}_N$ of Λ_N is no more a reliable approximation of ϕ_N and the whole SEM methodology fails. This is the main reason why another version of the SEM algorithm has to be defined and studied. This new version is the SAEM algorithm.

In order to prepare Section 4, we close this section with an identity which expresses the SEM sequence as a randomly perturbated dynamical system. We have

$$\phi^{n+1} = T_N(\phi^n) + U_N(\phi^n, z^n), \tag{3.5}$$

where $U_N(\phi^n, z^n) = M_N(z^n) - T_N(\phi^n)$, with M_N denoting the operator which conducts to ϕ^{n+1} from z^n via the M step of SEM. The r. v. $U_N(\phi^n, z^n)$ is conditionally independent of $(\phi^0, \phi^1, \ldots, \phi^{n-1})$ given ϕ^n, its conditional expectation is small but not necessarily zero and its conditional variance matrix has order N^{-1} as N increases to ∞.

4. The SAEM algorithm

In this section, we describe the SAEM algorithm for the mixture problem. Our approach is to modify the SEM algorithm in order to replace convergence in distribution by almost sure convergence (which is much easier to handle) and to attenuate the erratic behaviour of SEM in case of small samples without sacrificing the stochastic nature of the algorithm.

This is accomplished by replacing $U_N(\phi^n, z^n)$ by $\gamma_n U_N(\phi^n, z^n)$, where $\gamma_0 = 1$ and $\{\gamma_n\}$ is a sequence of positive real numbers decreasing to zero at a sufficiently slow rate as $n \to \infty$. Thus, any SAEM sequence has the form

$$\phi^{n+1} = T_N(\phi^n) + \gamma_n U_N(\phi^n, z^n). \tag{4.1}$$

The standard SAEM iteration $\phi^n \to \phi^{n+1}$ proceeds as follows.

E step : Compute $t_j^n(x_i)$, $i = 1, \ldots, N$ and $j = 1, \ldots, K$, using (2.7).

S step : For $i = 1, \ldots, N$, draw independently the random indicator variables z_i^n from a multinomial distribution consisting of one draw of K categories with probabilities $t_1^n(x_i), \ldots, t_K^n(x_i)$. If (3.1) holds, then the M step described below is accomplished. If (3.1) does not hold, then we draw new z_{ij}'s from some preassigned distribution on **Z** designed to ensure that condition (3.1) holds.

M step : Compute ϕ^{n+1} as follows:

$$p_j^{n+1} = N^{-1} \sum_{i=1}^{N} [(1 - \gamma_n)\, t_j^n(x_i) + \gamma_n z_{ij}^n] \quad \text{for } j = 1, \ldots, K, \tag{4.2}$$

and

$$a_j^{n+1} = (1 - \gamma_n) \frac{\sum_{i=1}^{N} t_j^n(x_i)\, b(x_i)}{\sum_{i=1}^{N} t_j^n(x_i)} + \gamma_n \frac{\sum_{i=1}^{N} z_{ij}^n\, b(x_i)}{\sum_{i=1}^{N} z_{ij}^n} \quad \text{for } j = 1, \ldots, K. \tag{4.3}$$

Some comments are in order. Informally, the SAEM algorithm can be schematized by the equation SAEM = $(1 - \gamma)$EM + γSEM, i.e. we go from pure SEM at the beginning towards pure EM at the end.

Since a_j^{n+1} is a convex combination of two elements of B, it is still in B. As a consequence, we have from (2.7) and (3.1) that, for all i = 1, ..., N and j = 1, ..., K,

$$t_j^n(x_i) \geq \frac{c(N)\, m_0}{\sum\limits_{r=1}^{K} p_r^n\, h(x_i, a_r^n)}, \tag{4.4}$$

where $m_0 = \inf\{h(x, a) : x = x_1, \ldots, x_N \text{ and } a \in B\}$. Now, since the p_r^n's are ≤ 1, we have

$$t_j^n\,(x_i) \geq \frac{m_0}{KM_0}\, c(N), \tag{4.5}$$

where $M_0 = \sup\{h(x, a) : x = x_1, \ldots, x_N \text{ and } a \in B\}$. To guarantee that $m_0 > 0$ and $M_0 < \infty$, it is sufficient to assume that h(x, a) is positive and continuous, since $\{x_1, \ldots, x_N\}$ and B are compact. We assume that these conditions will be in force throughout the remainder of this paper. Now, from (4.5) and since $\Sigma_j\, p_j^n = 1$, it results that, for each j = 1, ..., K,

$$\begin{aligned} p_j^n &= 1 - \sum_{r \neq j} p_r^n \\ &\leq 1 - \frac{(K-1)\, m_0}{KM_0}\, c(N) \\ &\leq 1 - \frac{m_0}{KM_0}\, c(N). \end{aligned} \tag{4.6}$$

Hence, each SAEM estimate ϕ^n remains in the compact convex subset

$$G_N = [c'(N),\, 1 - c'(N)]^{K-1} \times B^K \tag{4.7}$$

of E, where $c'(N) = \frac{m_0}{KM_0}\, c(N)$. Moreover, from (4.2) and (4.3) it follows directly that $|\, U_N\, (\phi, z)\, |$ is uniformly bounded when $\phi \in G_N$ and z satisfies (3.1).

SAEM exhibits some striking similarities with simulated annealing. Simulated annealing is a general approach for solving approximately large combinatorial optimization problems when no additional information about the structure of the function to be optimized is used. The simulated annealing method shares with SAEM the basic property of not terminating when reaching the first local optimum they encounter. In both cases, this is possible since the transitions $\phi^n \rightarrow \phi^{n+1}$ corresponding to a decrease (resp. increase) of the function to be maximized (resp. minimized) can be accepted, in some limited fashion, with non-zero probability. Moreover, in both cases the probability of accepting such transitions decreases to zero as the algorithm proceeds. By contrast, it must be stressed that SAEM primarily takes care of searching a significant local maximum of the log-l.f. $L(\phi)$ defined on some continuous parameter space (see, e.g., Section 6, where we deal with Gaussian mixtures), whereas simulated annealing is concerned with the search for the optimum of functions defined on discrete spaces. Moreover, SAEM is a tailored algorithm designed to find a stable fixed point of the log-l.f. $L(\phi)$, by taking advantage of the basic properties of the EM algorithm (especially, that each EM iteration increases the l.f.). Finally, even if SAEM were used to deal with a discrete parameter space, it could be expected to be more efficient than simulated annealing since it is based on the discrete-time dynamical system EM which increases $L(\phi)$ at each iteration.

5. The a.s. convergence of SAEM in the mixture case

The algorithm SAEM has been described in detail in Section 4, see (4.2)-(4.3). In this section, we will establish Theorem 1 which asserts the a.s. convergence of SAEM. Throughout this section and the Appendix, we will denote T_N, G_N and U_N by T, G and U, respectively, for the sake of brevity. Before proceeding to state and prove Theorem 1, we give a result which precises some important properties of $T(\phi)$ and $L(\phi)$ in the mixture context. This is the object of the following proposition, where $<h, h'>$ denotes the standard inner product of the Euclidian space **E**. Recall that **E** is the (K - 1 + sK) - dimensional space where the parameters ϕ take value.

PROPOSITION 1. *Let* ϕ^* *be any fixed point of* T. *Then,*

(i). There exists a symmetric definite positive matrix $A = A(\phi^*)$ *depending on* ϕ^* *such that*

$$<ADT(\phi^*)\, h, h'> = <Ah, DT(\phi^*)\, h'> \text{ for all h and h' in } \mathbf{E}. \qquad (5.1)$$

(ii). The eigenvalues of $DT(\phi^*)$ *are positive real numbers.*

(iii). Any fixed point ϕ^* *of T is stable iff it is a proper maximizer of the log-likelihood function* L. *It is hyperbolic iff it is a saddle-point of* L. *It is unstable iff it is a proper local minimizer of* L.

Proof of Proposition 1. We first prove (i). The relation (3.19), p. 9, of Dempster, Laird and Rubin [5] asserts that, if the matrix $D^{20}\,Q(\phi^*\,;\,\phi^*)$ is regular, then

$$DT(\phi^*) = D^{20}\,H(\phi^*\,;\,\phi^*)\ \{D^{20}\,Q(\phi^*\,;\,\phi^*)\}^{-1}, \qquad (5.2)$$

where $Q(\phi'; \phi)$ has been defined in (2.2) and (2.6), $H(\phi'\,;\,\phi) = E\{\log \mathbf{k}\,(\mathbf{z} \mid \mathbf{x}\,;\,\phi') \mid \mathbf{x}\,;\,\phi\} = L(\phi) - Q(\phi'; \phi)$ (see (2.3)) and D^{20} denotes the operator of second derivatives with respect to the first variable. Suppose first that we have proved that the symmetric matrices $D^{20}\,H(\phi^*\,;\,\phi^*)$ and $D^{20}\,Q(\phi^*\,;\,\phi^*)$ are definite negative. Writing $B = -\,D^{20}\,H(\phi^*\,;\,\phi^*)$ and $C = -\,\{D^{20}\,Q(\phi^*\,;\,\phi^*)\}^{-1}$ and defining $A = B^{-1}$, it follows from (5.2) that

$$<ADT(\phi^*)h, h'> = <Ch, h'> = <BAh, Ch'> = <Ah, DT(\phi^*)h'> \text{ for all h, h' in } \mathbf{E},$$

which is (i). Now, (5.1) means that $DT(\phi^*)$ is symmetric with respect to the inner product $<h, h'>_A = <Ah, h'>$ of **E**, implying that $DT(\phi^*)$ is diagonalizable with real eigenvalues. Moreover, (5.1) implies that these eigenvalues are positive. So, it remains to check that $D^{20}\,Q(\phi^*\,;\,\phi^*)$ and $D^{20}\,H(\phi^*\,;\,\phi^*)$ are definite negative. We begin with $D^{20}\,Q(\phi^*\,;\,\phi^*)$. In view of (2.6), the matrix $D^{20}\,Q(\phi^*\,;\,\phi^*)$ has zero nondiagonal blocks, whereas its diagonal blocks are given by

$$(\partial^2 / \partial p_j^2)\, Q(.\,;\,\phi^*)_{|\,p_j = p_j^*} = -\sum_{i=1}^{N} \left\{t_j^*(x_i) \,/\, p_j^{*2}\right\},\ \ j = 1, ..., K \qquad (5.3)$$

and

$$(\partial^2 / \partial a_j^2)\, Q(.\,;\phi^*)|_{a_j = a_j^*} = -\sum_{i=1}^{N} t_j^*(x_i)\, (\partial^2 / \partial a_j^2) \log h(x_i, a_j)|_{a_j = a_j^*},\ j = 1, \ldots, K \qquad (5.4)$$

respectively, where $t_j^*(x_i)$ is given by (2.7) with $\phi = \phi^*$. Since all the $t_j^*(x_i)$'s are positive, the right-hand side of (5.3) is negative. In addition, since the h(x, a) 's are issued from some exponential family, the terms $(\partial^2 / \partial a_j^2) \log h(x_i, a_j)|_{a_j = a_j^*}$ in the right-hand side of (5.4) are the opposites of covariance matrices. Hence the right-hand side of (5.4) is definite negative. We now turn to $D^{20} H(\phi^* ; \phi^*)$. We have (see Appendix, p. 232, of Louis [7])

$$D^{20}H(\phi^*;\phi^*) = \sum_{i=1}^{N} \left\{ \bar{S}(x_i,\phi^*)\, \bar{S}^T(x_i,\phi^*) - \sum_{j=1}^{K} t_j^*(x_i)\, S(x_i,\phi_j^*)\, S^T(x_i,\phi_j^*) \right\}, \qquad (5.5)$$

where

$$S(x_i,\phi_j^*) = (\partial / \partial \phi_j) \log \{p_j\, h(x_i, a_j)\}|_{\phi_j = \phi_j^*},\ i = 1, \ldots, N \text{ and } j = 1, \ldots, K, \qquad (5.6)$$

and

$$\bar{S}(x_i, \phi^*) = \sum_{j=1}^{K} t_j^*(x_i)\, S(x_i,\phi_j^*),\ i = 1, \ldots, N \qquad (5.7)$$

with $\phi_j = (p_j, a_j)$ for $j = 1, \ldots, K - 1$ and $\phi_K = (1 - \sum_{j=1}^{K-1} p_j, a_K)$. Recall that S^T denotes the transpose of S. From Cauchy-Schwarz inequality and in view of (5.5)-(5.7), it follows that $D^{20}H(\phi^* ; \phi^*)$ is definite negative. Finally, (iii) follows from (ii), using

$$D^2 L(\phi^*) = \{I - DT(\phi^*)\}\, D^{20} Q(\phi^* ; \phi^*). \qquad (5.8)$$

The relation (5.8) is established in Dempster, Laird and Rubin [5] (p.10). Hence, the proof of Proposition 1 is complete.

As there exists no tractable analytic expression for $T(\phi)$ and $U(\phi, z)$, it is necessary to provide some technical assumptions which are simple enough to allow a complete resolution of the problem of the a.s. convergence of the algorithm SAEM. We now introduce the assumptions that we will need.

(H1) The set F of those fixed points of T which are contained in the compact subset G of E is finite.

(H2) For any $\phi^* \in F$, the matrix $D^2 L(\phi^*)$ is regular. (Recall that the fixed points of T are stationary points of L.)

(H3) There exists at least a stable fixed point of T in G.

(H4) There exists $\rho > 0$ such that $T(G)^\rho$ is contained in G.

Here, $T(G)^\rho = \{\phi \in E: d(\phi, T(G)) < \rho\}$, where $d(\phi,T(G)) = \inf_{\psi \in T(G)} \|\phi - \psi\|$.

(H5) For any $\phi \in G$, any hyperplane H of **E** such that $T(\phi) \in H$ and any half-space D of **E** spanned by H, the set of those points of the form $T(\phi) + U(\phi, \mathbf{z})$, $\mathbf{z} \in \mathbf{Z}$, which are in D is non-empty.

We are now in a position to state our Theorem 1.

THEOREM 1. *Suppose that the assumptions (H1)-(H5) hold. If* $\{\gamma_n\}$ *is a sequence of positive numbers decreasing to zero as* $n \to \infty$ *and such that* $\gamma_n \sim c\, n^{-\mu}$ *or* $\gamma_n \sim c\,(\log n)^{-\nu}$ *as* $n \to \infty$ *for some positive constant* c *and some* μ, $0 < \mu < 1$, *or* $\nu > 0$, *then the sequence* $\{\phi^n\}$ *generated by SAEM converges almost surely to a local maximizer of the log-likelihood function* $L(\phi)$, *whatever its starting point* ϕ^0.

Remarks

(i) Assumption (H3) is satisfied, e.g., if the consistent m.l. estimator ϕ_N of the true parameter ϕ of the mixture is in G. As G is defined by the very weak constraints $c'(N) \leq p_j \leq 1 - c'(N)$ for $j = 1, ..., K$ (see (4.4)-(4.7) above), Assumption (H3) appears to be a very mild condition.

(ii) Assumption (H4) seems more restrictive. Roughly speaking, we can consider that (H4) is essentially equivalent to the following assumption (but we have no proof of this equivalence): the vector field $\nabla L(\phi)$ (gradient of L), $\phi \in G$, satisifies

$$< \nabla L(\phi), N(\phi) > < 0 \text{ for all } \phi \in \partial G,$$

where ∂G denotes the boundary of G and $N(\phi)$ denotes the unit vector normal to ∂G pointing outward G. Since the number of components K of the mixture is assumed known, hence is correct, and if the consistent m.l. estimator ϕ_N is in G (see (i) above), this assumption gives a reasonable description of the behaviour of T near ∂G. Assumption (H4) is technically essential since it guarantees that for all n large enough, ϕ^n remains in G.

(iii) Assumption (H5) seems very technical in nature. It is closely related to Assumption (H4) for the following reason. The set S of points of C (defined in (2.10)) of the form $T(\phi) + U(\phi, \mathbf{z})$, $\mathbf{z} \in \mathbf{Z}$, is equal to the set of points of C accessible in one iteration of SEM, starting from ϕ. Owing to the multinomial nature of the random drawings $\mathbf{z}$ involved by the RIP in the S-step (see Section 3), the number of distinct elements of S is equal to K^N. Recall that $K \geq 2$ is the number of components of the mixture and that N is the sample size. Even for moderate sample sizes, this number is huge. For instance, if $K = 2$ and $N = 50$, K^N is larger than 10^{15}. Moreover, it appeared implicitly through simulations that the points of S are rather homogeneously distributed in C. Now, since by (H4) we have $d(T(\phi), \partial G) \geq \rho > 0$, points of S can be found all around $T(\phi)$. This gives a rough justification of our assumption that (H5) holds. Assumption (H5) is technically important in the proof of our Theorem 1 (stated above) for the following reason. The conditional expectation $E^{\mathbf{z}}\{U(\phi, \mathbf{z}) \mid \phi\}$ is generally different from 0 for $\phi \in G$. Assumption (H5) takes care of this drawback by proposing a reasonable and tractable condition in order to replace the centered nature of the perturbation term $U(\phi, z)$. Assumption (H5) will be used in the proof of Proposition A.1 in the Appendix.

(iv) The assumptions concerning the asymptotic behaviour of the sequence $\{\gamma_n\}$ in the statement of our Theorem 1 have been chosen in this concrete form for clarity's sake. The result stated in this manner is less general, but has the bonus of avoiding some rather obscure regularity conditions. The essential requirements about $\{\gamma_n\}$ that we need are that γ_n decreases to 0 as $n \to \infty$, $\lim_{n\to\infty} (\gamma_n / \gamma_{n+1}) = 1$ and $\sum_n \gamma_n = \infty$.

(v) For EM, the possibility of convergence to a saddle-type point of the l.f. is always present. On the contrary, Theorem 1 ensures that SAEM does not converge to such a point a.s. Moreover, inspection of the proof below shows that SAEM does not necessarily terminate in the first local maximum encountered, as does EM.

Proof of Theorem 1. Before proceeding, we mention some notation and conventions which will be adopted throughout this proof and in the Appendix. We will denote by F the finite set of the fixed points ϕ^* of T in G, see (H1). The notation $\| h \|$ will indicate the usual Euclidian norm on **E**, with $< h, k >$ denoting the corresponding inner product. For each $\phi^* \in F$, we define

$$A(\phi^*) = - D^{20} H(\phi^* ; \phi^*)^{-1}, \tag{5.9}$$

a symmetric definite positive matrix, where $D^{20} H(\phi^* ; \phi^*)$ has been defined in the proof of Proposition 1. When no risk of confusion can arise, we will denote $A(\phi^*)$ more compactly by A and we will write

$$< h, k >_A = < Ah, k > \text{ for all h and k in } \mathbf{E} \tag{5.10}$$

and

$$\| h \|_A = < h, h >_A \text{ for all h in } \mathbf{E}. \tag{5.11}$$

As **E** is finite-dimensional, there exists positive constants C_A and D_A such that

$$C_A \| h \| \leq \| h \|_A \leq D_A \| h \| \text{ for all h in } \mathbf{E}, \tag{5.12}$$

and we will define

$$C_0 = \min_{\phi^* \in F} C_A, \quad A = A(\phi^*), \tag{5.13}$$

which is positive since #F is finite. Moreover, for any C^2 function $f : \mathbf{E} \to \mathbf{R}$, we will denote the value of the quadratic form $D^2 f(\phi)$ at $h \in \mathbf{E}$ by $D^2 f(\phi)$ (h), i.e., in matrix notation,

$$D^2 f(\phi)\,(h) = h^T D^2 f(\phi)\, h \text{ for all h in } \mathbf{E}, \tag{5.14}$$

where h is written in vector form in the right-hand side of (5.14) and h^T indicates the transpose of h.

Furthermore, we will consider the following subsets of E. We will denote by $B(\phi, r)$ the open ball with center ϕ and radius r, i.e.

$$B(\phi\,;r) = \{\phi + h : \| h \| < r\} \tag{5.15}$$

and, similarly,

$$B_A(\phi\,;r) = \{\phi + h : \| h \|_A < r\}. \tag{5.16}$$

Given $\phi^* \in F$, we will define

$$V_n(\phi^*) = B(\phi^*\,;C_V\,\gamma_n^{1/2}) \cap G, \tag{5.17}$$

$$HV_n(\phi^*) = B_A(\phi^*\,;HC_V\,\gamma_n^{1/2}) \cap G \tag{5.18}$$

and

$$B_n(\phi^*) = B_A(\phi^*\,;K\,\gamma_n) \cap G \tag{5.19}$$

for all $n \geq 1$. Here, C_V, $H > 1$ and K denote positive constants to be suitably chosen in the course of the proof. When no risk of confusion can arise, the dependence of $V_n(\phi^*)$, $HV_n(\phi^*)$ and $B_n(\phi^*)$ on $\phi^* \in F$ will be suppressed in the notation. The subset of G defined by

$$A_n = G - \bigcup_{\phi^* \in F} V_n(\phi^*) \tag{5.20}$$

is a compact subset of G. Throughout the proof, the integer $n \geq 1$ will be assumed large enough to ensure that B_n is contained in V_n and V_n is contained in HV_n.

One of the key observations on which the proof relies is that there exists a constant SUP < ∞ such that

$$| U(\phi, \mathbf{z}) | \leq \text{SUP for all } \phi \in G \text{ and } \mathbf{z} \text{ satisfying (3.1).} \tag{5.21}$$

Finally, we will write

$$\| DL(\phi) \| = \sup_{\| h \| \leq 1} | DL(\phi)(h) |, \tag{5.22}$$

$$\| D^2L(\phi) \| = \sup_{\| h \|, \| k \| \leq 1} | D^2L(\phi)(h, k) | \tag{5.23}$$

and

$$\| DT(\phi) \| = \sup_{\| h \| \leq 1} \| DT(\phi)(h) \| \tag{5.24}$$

and note that the constants

$$b_1 = \sup_{\phi \in G} \| DL(\phi) \|, \tag{5.25}$$

$$\beta = \sup_{\phi \in G} \| D^2L(\phi) \|, \tag{5.26}$$

$$C_T = \sup_{\phi \in G} \| DT(\phi) \| \tag{5.27}$$

and

$$b_2 = \sup_{\phi \in G} \| D(L \circ T)(\phi) \| \leq b_1 C_T \tag{5.28}$$

are finite since the corresponding functions $DL(\phi)$, $D^2L(\phi)$, $DT(\phi)$ and $D(L \circ T)(\phi)$ are continuous on the compact subset G of **E**.

The proof is organized as follows. It is divided into 3 steps. In Step 1, we establish that after each entrance in A_n, the sequence $\{\phi_m\}$ will escape from A_m for some finite $m > n$. In Step 2, we essentially prove that, if $DT(\phi^*)$ has at least an eigenvalue larger than 1 (i.e., ϕ^* is unstable) and ϕ_n has entered $V_n(\phi^*)$ for some n, then with probability 1 the sequence $\{\phi_m\}$ cannot remain in $V_m(\phi^*)$ for all $m \geq n$, provided that the constants C_V and $H > 1$ have been suitably chosen. In Step 3, we conclude the proof of Theorem 1 by collecting the various results obtained in the first two steps. Step 2 is in turn divided into three parts : Points 1, 2 and 3. In Point 1, we establish that if ϕ^* is unstable and n is large enough and ϕ_n has entered $B_n(\phi^*)$, then ϕ_m exits from $B_m(\phi^*)$ for some finite $m > n$ with probability 1. In proving Point 1, we make use of a crucial but technical result that we have stated as Proposition A.1 and postponed in the Appendix in the interests of clarity. In Point 2, we prove that, if K has been chosen large enough, then a. s. ϕ_m exits from $HV_m(\phi^*)$ after a finite time. In Point 3, we prove that, given any $\phi^* \in F$ (stable or unstable), if ϕ_n has entered $V_n(\phi^*)$ at some time $n = t$ and has exited from $HV_n(\phi^*)$ at some time $n = n_* > t$, then

$$L(\phi^m) > \sup_{\phi \in V_{n_*}(\phi^*)} L(\phi) \quad \text{for all } m \geq n_*, \tag{5.29}$$

provided that $H > 1$ has been chosen large enough and that the entrance time t is assumed sufficiently large. Inequality (5.29) is crucial, since it implies that ϕ^m can never re-enter $V_m(\phi^*)$. The proof of Step 1 involves establishing Lemma 1, in which lower bounds for $\inf_{\phi \in A_n} (L \circ T - L)(\phi)$ and $L(\phi^{n+1}) - L(\phi^n)$ for $\phi^n \in A_n$ are derived. The proof of Point 3 in Step 2 relies on several technical inequalities stated in Lemma 2 and Lemma 3.

Step 1. We claim that C_V can be chosen so large as to ensure that, if ϕ_n has entered A_n for some n, then ϕ_m will exit from A_m for some finite $m > n$, a. s.

Proof of Step 1. We preface the proof with a lemma.

Lemma 1. (i)*There exists a positive constant* α *such that the inequality*

$$c_n = \inf_{\phi \in A_n} (L \circ T - L)(\phi) \geq \alpha\, C_V^2\, \gamma_n \tag{5.30}$$

holds for all $n \geq 1$ *sufficiently large.*

(ii) *The constant* C_V *in (5.17) can be chosen so large that*

$$L(\phi^{n+1}) - L(\phi^n) \geq (\alpha / 2)\, C_V^2\, \gamma_n \tag{5.31}$$

for $\phi^n \in A_n$ *and for all* $n \geq 1$ *sufficiently large.*

Proof of Lemma 1. Since L and T are C^∞, so is $L \circ T - L : \mathbf{E} \to \mathbf{R}$. From Proposition 1, it follows that $(L \circ T - L)(\phi)$ reaches its minimum value at $\phi = \phi^* \in F$. From Assumption (H2) and Proposition 1 it results that the symmetric matrix $D^2 (L \circ T - L)(\phi^*)$ is definite positive. This implies that

$$D^2 (L \circ T - L)(\phi^*)(h) \geq \lambda_{\min}(\phi^*) \| h \|^2 \text{ for all } h \in \mathbf{E}, \tag{5.32}$$

where the positive number $\lambda_{\min}(\phi^*)$ denotes the smallest eigenvalue of $D^2 (L \circ T - L)(\phi^*)$. Since #F is finite, $4\alpha = \min_{\phi^* \in F} \lambda_{\min}(\phi^*) > 0$. Thus,

$$D^2 (L \circ T - L)(\phi^*)(h) \geq 4\alpha \| h \|^2 \text{ for all } \phi^* \in F \text{ and } h \in \mathbf{E}. \tag{5.33}$$

Now, a quadratic Taylor expansion of $L \circ T - L$ about ϕ^* gives

$$(L \circ T - L)(\phi^* + h) = (1/2) D^2 (L \circ T - L)(\phi^*)(h) + O(\| h \|^3). \tag{5.34}$$

Substituting (5.33) into (5.34) we obtain that, if $\| h \| < \varepsilon$ for some sufficiently small $\varepsilon > 0$, then

$$(L \circ T - L)(\phi^* + h) \geq \alpha \| h \|^2. \tag{5.35}$$

Inequality (5.35) implies that

$$\inf \{(L \circ T - L)(\phi^* + h) : \phi^* \in F, C_V \gamma_n^{1/2} \leq \| h \| < \varepsilon\} \geq \alpha C_V^2 \gamma_n \tag{5.36}$$

whenever $n \geq 1$ is large enough to ensure that $C_V \gamma_n^{1/2} < \varepsilon$. Thus it remains to examine the infimum of $(L \circ T - L)(\phi)$ when $\phi \in A_n$ and $\phi \notin \bigcup_{\phi^* \in F} B(\phi^* ; \varepsilon)$. By the continuity of $L \circ T - L$ and the compacity of A_n, we have that $c_n > 0$. We will prove (5.30) by contradiction. If (5.30) did not hold, then we would have $c_n < \alpha C_V^2 \gamma_n$ for some subsequence $\{n'\}$ of $\{n\}$. For simplicity, we write $n' = n$. From the continuity of $L \circ T - L$ and the compacity of A_n, there would exist $g_n \in A_n$ satisfying

$$\eta_n = (L \circ T - L)(g_n) < \alpha C_V^2 \gamma_n. \tag{5.37}$$

By compacity, there would exist a subsequence $\{g_{\beta(n)}\}$ converging to some $g \in G$ as $n \to \infty$, implying $(L \circ T - L)(g) = 0$, hence $g \in F$. Put $\phi^* = g$ and suppose that n is so large that $g_{\beta(n)}$ has the form $\phi^* + h$, with $\| h \| < \varepsilon$. Then, using (5.35) and (5.37) we would have

$$\alpha \| \phi^* - g_{\beta(n)} \|^2 \leq (L \circ T - L)(g_{\beta(n)}) = \eta_{\beta(n)} < \alpha C_V^2 \gamma_{\beta(n)},$$

implying that $g_{\beta(n)} \in V_{\beta(n)}(\phi^*)$, thus contradicting $g_{\beta(n)} \in A_{\beta(n)}$. This proves (i). We now turn to (ii). By making use of a linear Taylor expansion we can derive that, if $\phi^n \in A_n$, then

$$L(\phi^{n+1}) - L(\phi^n) \geq \alpha\, C_V^2\, \gamma_n - b_1 \,\mathrm{SUP}\, \gamma_n, \tag{5.38}$$

in view of (5.21), (5.25) and (5.30). Thus, if C_V is chosen so large that $\alpha\, C_V^2 \geq 2b_1$ SUP, then (5.31) holds. This proves (ii).

We are now in a position to complete the proof of Step 1. The proof is by contradiction. If ϕ^m remained in A_m for all $m \geq n$, then we could deduce from (5.31) and $\sum \gamma_m = \infty$ that $L(\phi^m) \to \infty$ as $m \to \infty$, thus contradicting the boundedness of L on G. This completes the proof of Step 1.

Step 2. Suppose that $DT(\phi^*)$ has at least an eigenvalue larger than 1. We claim that, if $\phi_n \in V_n(\phi^*)$ for some $n \geq 1$, then with probability 1 ϕ_m will not remain in $V_m(\phi^*)$ for all $m \geq n$.

Proof of Step 2. The proof of Step 2 will be dissected in stating and proving Points 1, 2 and 3 below.

Point 1. We claim that, if $\phi_n \in B_n(\phi^*)$ for some $n \geq 1$ and provided that the constant K in (5.19) has been chosen large enough, then with probability 1 ϕ_m will not remain in $B_m(\phi^*)$ for all $m \geq n$.

Proof of Point 1. Since we have assumed that $DT(\phi^*)$ has an eigenvalue $\lambda > 1$, it follows that there exists $v \in \mathbf{E}$ satisfying

$$DT(\phi^*)\, v = \lambda v \text{ and } \| v \|_A = 1, \tag{5.39}$$

where A has been defined in (5.9) and $\| v \|_A = < Av, v >^{1/2}$ has been defined in (5.10)-(5.11). Define

$$q_n = < v, \phi^n - \phi^* >_A \text{ and } w_n = < v, U(\phi^n, z^n) >_A. \tag{5.40}$$

Then, from Cauchy-Schwarz inequality together with (5.39) and (5.40) and in connection with a quadratic Taylor expansion of T about ϕ^*, it follows that

$$\begin{aligned} q_{n+1} &= < v, \phi^{n+1} - \phi^* >_A = < v, T(\phi^n) - T(\phi^*) + \gamma_n U(\phi^n, z^n) >_A \text{ (since } T(\phi^*) = \phi^*) \\ &= < v, DT(\phi^*)\,(\phi^n - \phi^*) >_A + O(q_n^2) + \gamma_n w_n = < DT(\phi^*)\, v, \phi^n - \phi^* >_A + O(q_n^2) + \gamma_n w_n \\ &= \lambda < v, \phi^n - \phi^* >_A + O(q_n^2) + \gamma_n w_n, \end{aligned}$$

whence

$$q_{n+1} = \lambda q_n + O(q_n^2) + \gamma_n w_n. \tag{5.41}$$

For any integers n and $j \geq 1$ and any positive a, we introduce the event $E_{n,j} = E^+_{n,j} \cup E^-_{n,j}$, with

$$E^{+}_{n,j} = \{q_n \geq 0, w_n > a, \ldots, w_{n+j} > a\} \tag{5.42}$$

and

$$E^{-}_{n,j} = \{q_n \leq 0, w_n < -a, \ldots, w_{n+j} < -a\}. \tag{5.43}$$

Proposition A.1 in the Appendix implies that the positive number a can be chosen so that, given any $j \geq 1$, the events $E_{n,j}$, $n \geq 1$, occur infinitely often (i.o.) almost surely. Define $r \geq 1$ to be the first entrance-time of ϕ_n in B_n posterior to a given $t \geq 1$ assumed sufficiently large. Of course, r is assumed finite.

Applying Cauchy-Scharwz inequality to $q_n = <v, \phi^n - \phi^*>$, Point 1 will be proved if we can prove that with probability 1

$$| q_n | \geq K\gamma_n \text{ for some finite } n \geq r. \tag{5.44}$$

We will prove (5.44) by contradiction. To this end, assume that the event

$$\Omega_r = \{| q_n | < K\gamma_n, \text{ all } n \geq r\} \tag{5.45}$$

has positive probability. Then, by using Proposition A.1 as indicated above, the event $\Omega_{j,r} = \Omega_r \cap \{E_{n,j}$ occurs i.o.$\}$ has positive probability. Now, assume that $\omega \in \Omega_r$ and let n be any integer $\geq r$ such that $E_{n,j}$ occurs. If $q_n \geq 0$, then using (5.41) and (5.42) we obtain that

$$q_{n+1} \geq a\gamma_n - c_0 \gamma_n^2 \tag{5.46}$$

for some positive constant c_0 (independent of ω). Similarly, if $q_n \leq 0$ then we have $q_{n+1} \leq - (a\gamma_n - c_0 \gamma_n^2)$. Moreover, we may and will select n so large that

$$1 - (c_0 \gamma_n / a) \geq 1/2. \tag{5.47}$$

Observe that we can suppose without loss of generality that $q_n \geq 0$, since (5.39)-(5.45) are invariant under the transformation $q_n \rightarrow - q_n$. Comparing (5.46) and (5.47) and using (5.41), it follows that

$$q_{n+j} \geq (a/2) \sum_{\ell=0}^{j-1} \lambda^{j-\ell-1} \gamma_{n+\ell}, \tag{5.48}$$

so that $q_{n+j} \geq K\gamma_{n+j}$ provided that j has been chosen so large as to guarantee that $(a/2)\lambda^{j-1} \geq 2K$ (recall that $\lambda > 1$) and n is assumed so large that $\gamma_{n+j} \geq \gamma_n / 2$. (Recall that in Proposition A.1 mentioned above, the choice of a and the choice of j are independent.) This completes the proof of Point 1.

Point 2. We claim that, if K is large enough, then, if $\phi^n \in V_n$ but $\phi^n \notin B_n$, then ϕ^m will exit from HV_m in a finite time, whatever $H > 1$.

Proof of Point 2. Let $b > 0$ be such that $\lambda > 1 + b$ and K so large that

$$\lambda K - H^2 C_v^2 - \mathrm{SUP} > (1 + b)\, K. \tag{5.49}$$

Then, in view of (5.21) and (5.39)-(5.41) it can be deduced that

$$| q_{n+k} | \geq (1 + b)^k K \gamma_{n+k} \tag{5.50}$$

for all $k \geq 1$ such that $| q_{n+k-1} | < HC_v \gamma_{n+k-1}^{1/2}$. Since, given $n \geq 1$, $(1 + b)^k \gamma_{n+k} \to \infty$ as $k \to \infty$, the proof of Point 2 is complete.

Point 3. Let ϕ^* be any fixed point of T in G. Assume that at some time t, $\{\phi^n\}$ has entered V_t (ϕ^*). Denote by n^* the first exit-time of $\{\phi^n ; n > t\}$ from $HV_{n^*}(\phi^*)$. Assume that n^* is finite. We claim that the constant $H > 1$ above can be chosen so large as to ensure that

$$L(\phi^m) > \sup_{\phi \in V_{n^*}} L(\phi) \text{ for all } m \geq n^*. \tag{5.51}$$

Remark that this statement is true for all the fixed points. Point 3 will also appear to be crucial in Step 3.

Recall that t is an entrance-time of ϕ^n in V_n. Of course, t is assumed finite and sufficiently large.

Proof of Point 3. We let $k \geq t$ denote the last exit-time from V_n before n_*, i.e. ϕ^n is in HV_n but ϕ^n is not in V_n for all n, $k \leq n \leq n_* - 1$, and ϕ^{n^*} is not in HV_{n^*}. We first need to prove preliminary technical inequalities involving both $T(\phi)$ and $L(\phi)$ for ϕ in V_n or in HV_n. These preliminary results are established in Lemmas 2 and 3 below.

Lemma 2. *For any fixed point* $\phi^* \in F$ *and any positive constant* D, *there exists* n_0 *such that*

$$(L \circ T - L)(\phi^* + h) \geq (\alpha / 2)\, C_0^2 D^2 \gamma_n \textit{ for all h such that}$$

$$\| h \| < \varepsilon \textit{ and } \| h \|_A \geq D \gamma_n^{1/2} \textit{ and all } n \geq n_0, \tag{5.52}$$

where $\varepsilon > 0$ *has been defined in the proof of Lemma 1 and* C_0 *has been defined in (5.13).*

Proof of Lemma 2. The proof of Lemma 2 closely parallels the proof of (5.36) in the proof of Lemma 1 and will not be detailed here. Next, we establish the following inequalities.

Lemma 3. *Let* ϕ^* *be any fixed point of* T *in* G. *Assume that at some time* t $\{\phi^n\}$ *has entered* $V_t(\phi^*)$. *Denote by* n^* *the first exit-time of* $\{\phi^n; n > t\}$ *from* $HV_{n^*}(\phi^*)$. *Assume that* n^* *is finite. Denote by* k, $t < k \leq n^*$, *the last exit-time of* $\{\phi^n; n > t\}$ *from* $V_k(\phi^*)$ *before* n^*.

(i) *If* n *is large enough, then*

$$\sup_{\phi \in V_n} | L(\phi) - L(\phi^*) | \leq b_1 C_v^2 \gamma_n, \tag{5.53}$$

-

$$\sup_{\phi \in V_n} | (L \circ T)(\phi) - (L \circ T)(\phi^*) | \leq b_2 C_v^2 \gamma_n \tag{5.54}$$

and

$$\sup_{\phi \in V_n} | (L \circ T) (\phi) - L(\phi) | \leq (b_1 + b_2) C_v^2 \gamma_n, \tag{5.55}$$

where b_1 *and* b_2 *have been defined in (5.25) and (5.28), respectively.*

(ii) *If* k *is large enough, then*

$$| L(\phi^k) - L(\phi^*) | \leq (4b_1 + 2b_2) C_v^2 \gamma_k. \tag{5.56}$$

(iii) *If* k *is large enough and if we let* $n_* = k + p$, *then*

$$p \geq 1, \tag{5.57}$$

provided $H > 1$ *has been chosen large enough,*

$$\| T(\phi^{k+p-1}) - \phi^* \|_A \geq (HC_v / 2) \gamma_{k+p}^{1/2}, \tag{5.58}$$

and

$$\| \phi^{k+p-1} - \phi^* \| \geq (HC_v / 2C_T D_A) \gamma_{k+p}^{1/2}, \tag{5.59}$$

where C_T *and* D_A *have been defined in (5.27) and (5.12), respectively.*

Remark. If ϕ^* is an unstable fixed point, it has been proved in Point 2 that n^* was finite a.s.

Proof of Lemma 3. Inequality (5.53) (resp. (5.54)) follows from a linear Taylor expansion of $L(\phi)$ (resp. $(L \circ T)(\phi)$) about ϕ^*, and Inequality (5.55) follows by adding (5.53) and (5.54). Next, we turn to (5.56). Splitting $L(\phi^k) - L(\phi^*)$ into a sum of two terms and using (5.53) in connection with the fact that $\phi^{k-1} \in V_{k-1}$ as well as a quadratic expansion of $L(\phi)$ about $T(\phi^{k-1})$ yields, in view of (5.55)

$$\begin{aligned} | L(\phi^k) - L(\phi^*) | &\leq | L(\phi^k) - L(\phi^{k-1}) | + | L(\phi^{k-1}) - L(\phi^*) | \\ &\leq (2b_1 + b_2) C_v^2 \gamma_{k-1} + \mathrm{SUP} \sup_{\phi \in V_{k-1}} | DL\{T(\phi)\} | \gamma_{k-1} + O(\gamma_{k-1}^2). \end{aligned} \tag{5.60}$$

By a linear Taylor expansion of $DL\{T(\phi)\}$ about ϕ^* and $DL\{T(\phi^*)\} = 0$, we have

$$\sup_{\phi \in V_{k-1}} | DL\{T(\phi)\} | \leq \beta C_T C_v \gamma_{k-1}, \tag{5.61}$$

where β has been defined in (5.26). Substituting (5.61) into (5.60) gives, in view of $\lim_{k \to \infty} (\gamma_k / \gamma_{k-1}) = 1$,

$$| L(\phi^k) - L(\phi^*) | \leq (2b_1 + b_2) C_v^2 \gamma_k + o(\gamma_k),$$

which implies (5.56) whenever γ_k is small enough. This proves (ii).

We now turn to (iii). We first prove that $p \geq 1$, i.e. that $\phi^k \in HV_k$. As $\phi^{k-1} \in V_{k-1}$,

$$\| \phi^k - \phi^{k-1} \| \leq (C_T C_v + C_v) \gamma_{k-1}^{1/2} + \text{SUP}\, \gamma_k$$

from which it follows that, in view of (5.12),

$$\| \phi^k - \phi^{k-1} \|_A \leq D_A \{(C_T C_v + 2C_v) \gamma_{k-1}^{1/2} + \text{SUP}\, \gamma_k\}. \tag{5.62}$$

Thus, the right-hand side of (5.62) is smaller than $HC_v \gamma_k^{1/2}$, provided that $H > 1$ has been chosen large enough and k is assumed large enough. Hence (5.57). In order to prove (5.58) we first observe that, as $\phi_{k+p} \notin HV_{k+p}$, $p \geq 1$ and $\lim_{n\to\infty} (\gamma_n / \gamma_{n-1}) = 1$, we have

$$\begin{aligned} \| T(\phi^{k+p-1}) - \phi^* \|_A &\geq \| \phi^{k+p} - \phi^* \|_A - \gamma_{k+p-1}\, \text{SUP} \\ &\geq HC_v\, \gamma_{k+p}^{1/2} + O(\gamma_{k+p}) \\ &\geq (HC_v /2)\, \gamma_{k+p}^{1/2} \end{aligned}$$

for all k sufficiently large, which is (5.58). Finally, a linear Taylor expansion of $T(\phi)$ about ϕ^* gives in view of (5.12),

$$\| T(\phi^{k+p-1}) - \phi^* \|_A \leq C_T D_A \| \phi^{k+p-1} - \phi^* \|. \tag{5.63}$$

Substituting (5.63) into (5.58) yields (5.59). This completes the proof of Lemma 3.

We now return to the proof of Point 3. We first split

$$L(\phi^{k+p}) - L(\phi^*) = I + II + III, \tag{5.64}$$

with

$$I = L(\phi^{k+p}) - L(\phi^{k+p-1}), \tag{5.65}$$

$$II = L(\phi^{k+p-1}) - L(\phi^k) \tag{5.66}$$

and

$$III = L(\phi^k) - L(\phi^*), \tag{5.67}$$

and we seek suitable lower bounds for I, II and III. First, from (5.56) it results that

$$| III | \leq (4b_1 + 2b_2)\, C_v^2\, \gamma_k \tag{5.68}$$

for all k sufficiently large. Thus, in order to prove (5.51), in view of (5.53), it suffices to prove that,

$$I + II \geq (4b_1 + 2b_2)\, C_v^2 \gamma_k + 2 \sup_{\phi \in V_{k+p}} | L(\phi) - L(\phi^*) |,$$

where $\sup_{\phi \in V_{k+p}} | L(\phi) - L(\phi^*) | \leq b_1 C_v^2 \gamma_{k+p}$ by (5.53).

Since $\gamma_{k+p} < \gamma_k$, it suffices to prove that

$$I + II \geq (4b_1 + 2b_2) C_v^2 \gamma_k + 2b_1 C_v^2 \gamma_k,$$

i.e.

$$I + II \geq (6b_1 + 2b_2) C_v^2 \gamma_k. \qquad (5.69)$$

The idea of the proof of (5.69) is as follows. Since $\phi^{k+\ell} \in A_{k+\ell}$ for all $0 \leq \ell \leq p - 1$, it follows from (5.31) that $I > 0$ and $II > 0$ when k is large enough. If p is large, then II becomes large. In this case, we will prove that

$$II \geq (6b_1 + 2b_2) C_v^2 \gamma_k. \qquad (5.70)$$

If $p \geq 1$ is small, then we will prove that, if $H > 1$ has been chosen large enough, then

$$I \geq (6b_1 + 2b_2) C_v^2 \gamma_k. \qquad (5.71)$$

More precisely, let $p_0 \geq 1$ be a fixed integer to be chosen in the course of the proof. We first derive a suitable lower bound for I. Since $\phi^{k+p-1} \in HV_{k+p-1}$, $\| \phi^{k+p-1} \| < \varepsilon$ for k large enough, where $\varepsilon > 0$ has been defined in (5.35). Using (5.52) in Lemma 2 with $D = HC_v/(2C_T D_A)$ given by (5.59) in connection with an upper bound for $|DL\{T(\phi^{k+p-1})\}|$ similar to the right-hand side of (5.61) with $k + p - 1$ replacing $k - 1$, we obtain

$$I \geq H^2 M(\gamma_{k+p_0} / \gamma_k)\gamma_k + o(\gamma_k), \quad 1 \leq p \leq p_0,$$

for some positive constant M. As $\lim_{k \to \infty} (\gamma_{k+p_0} / \gamma_k) = 1$, k may be chosen large enough to ensure $\gamma_{k+p_0} / \gamma_k \geq 1/2$.

Hence, for k sufficiently large,

$$I \geq (H^2 M/4) \gamma_k \text{ for all p, } 1 \leq p \leq p_0. \qquad (5.72)$$

Consequently, substituting (5.72) into (5.71) we see that (5.71) holds if H is chosen large enough. We will now prove (5.70) and determine p_0 in the same exercise. The key observation is that, since $\phi^m \in A_m$, $k \leq m \leq k+p-1$,

$$II \geq (\alpha/2) C_v^2 \sum_{m=k}^{k+p-1} \gamma_m. \qquad (5.73)$$

On the other hand, since $p \geq 1$,

$$\sum_{k=1}^{k+p-1} m^{-\mu} \geq \int_k^{k+p} t^{-\mu}\, dt \geq (1 - \mu)^{-1} k^{1-\mu} [\{1 + (p/k)\}^{1-\mu} - 1]. \qquad (5.74)$$

Since $1 \leq p \leq p_0$, the right-hand side of (5.74) is equivalent to $pk^{-\mu}$ as $k \to \infty$. Hence, for k sufficiently large and $1 \leq p \leq p_0$, we have

$$II \geq (\alpha/4)\, C_v^2 \;\; p\gamma_k, \tag{5.75}$$

from which it results that

$$II \geq (\alpha/4)\, C_v^2 \;\; p_0\, \gamma_k \tag{5.76}$$

for all $p \geq p_0$. This implies that (5.70) holds if p_0 is chosen so large that $(\alpha/4)p_0 > 6b_1 + 2b_2$. The proof is similar if $\gamma_n \sim c(\log n)^{-\nu}$, where c and ν are positive constants. This completes the proof of Point 3. Hence, the proof of Step 2 is completed.

Step 3. We will now collect the facts obtained above in order to complete the proof of Theorem 1. First, observe that since # F is finite, we can choose n_0 so large as to ensure that

$$\bigcap_{\phi^* \in F} HV_n(\phi^*) = \varnothing \tag{5.77}$$

for all $n \geq n_0$.

Let $\phi^m \in V_m(\overset{*}{\phi}_0)$ for a $\overset{*}{\phi}_0 \in$ F and some finite $m \geq n_0$: from Step 1, such an m exists a.s. Either $\phi^n \in HV_n(\overset{*}{\phi}_0)$ for all $n \geq m$, or ϕ_n exits from $HV_n(\overset{*}{\phi}_0)$ at a finite time $n = n_1$. In the latter case, in view of (5.51), either ϕ^{n1} is in $V_{n1}(\overset{*}{\phi}_1)$ for a $\overset{*}{\phi}_1$ in F such that $L(\overset{*}{\phi}_1) > L(\overset{*}{\phi}_0)$ or $\phi^{n1} \in A_{n1}$. In the latter case, it results from Step 1 that $L(\phi^{n1}) < L(\phi^{n1+1}) < \ldots < L(\phi^n)$ until eventually $\phi^n \in V_n(\overset{*}{\phi}_2)$ for a $\overset{*}{\phi}_2$ in F such that $L(\overset{*}{\phi}_2) > L(\overset{*}{\phi}_1)$. From (5.51) and (5.77), it follows that $L(\phi^n) > L(\phi^{n1})$ for all $n > n_1$. We now prove by contradiction that there exists $\phi^* \in$ F and $n_2 \geq n_0$ such that ϕ_n is in $HV_n(\phi^*)$ for all $n \geq n_2$. This will imply the convergence of ϕ^n to ϕ^*. Otherwise, there would exist a sequence $\overset{*}{\psi}_1, \overset{*}{\psi}_2, \ldots$ in F such that $L(\overset{*}{\psi}_1) < L(\overset{*}{\psi}_2) < \ldots$, implying the existence of an infinite sequence of distinct points in F. But this is a contradiction since F is finite. It remains to prove that ϕ^* is a stable point. But, from Steps 1 and 2, ϕ^* cannot be unstable, since ϕ_n is necessarily in all the $HV_n(\phi^*)$'s for n large enough. This completes the proof of Theorem 1. Incidentally, we have proved that

$$\|\phi^n - \phi^*\|_A < HC_V\, \gamma_n^{1/2} \tag{5.78}$$

for n large enough. This provides an asymptotic upper bound for the rate of convergence.

6. Numerical comparaisons of EM, SEM and SAEM

In this section, we compare the practical behaviour of the algorithms EM, SEM and SAEM for estimating the parameters of a four-component univariate Gaussian mixture for small samples, on a basis of a Monte-Carlo experiment. We chose the following values of the parameters. The proportions of the mixture were $p_1 = p_2 = p_3 = p_4 = 0.25$; the means $m_1 = 2$, $m_2 = 5$, $m_3 = 9$ and $m_4 = 15$ and the variances $\sigma_1^2 = 0.0625$, $\sigma_2^2 = 0.25$, $\sigma_3^2 = 1$ and $\sigma_4^2 = 4$.

We generated 100 samples from this mixture: The first fifty ones with a size N = 100 and the other ones with a size N = 60. Using these 50 samples for each sample size, the numerical experiments were performed as follows. For each generated sample, we performed 200 iterations of EM, SEM and SAEM starting first with KINIT = 5 components, then with 4 components. The initial positions of the parameters was drawn at random in the following way: The initial centers of the mixture components were drawn as follows. We first draw uniformly KINIT points in $\{x_1, ..., x_n\}$, then we aggregated the x_i's around the nearest of these points, thus obtaining a partition of the sample into KINIT clusters. Finally, the initial parameters were computed on the basis of this generated partition using each cluster as if it were a sample drawn from one of the components of the mixture. Note that new initial positions were drawn before running EM, SEM and SAEM for each simulated sample; thus, 50 initial positions were drawn for experimenting the three algorithms over the 50 generated samples of size 100 with the initial value KINIT = 5 ; 50 more initial positions were drawn for KINIT = 4; then, 100 more initial positions were drawn in the same fashion for the 50 generated samples of size 60. Note also that we chose a comparatively large number of iterations for EM instead of a stopping rule because the EM iteration number at which a given degree of accuracy is first obtained can be very large (see Table 6.1, p.233, of Redner and Walker [10]).

Our procedure to obtain a reasonable value of the number K of components of the mixture was the following. Each time one of the mixing weights p_j^n became smaller than 2/N, we cancelled the corresponding *j* th component and started afresh, using the same initialization scheme on the basis of the remaining components for 200 new iterations. Observe that this procedure is just taking c(N) = 2/N in (3.1). The resulting K, that we denote KFINAL in Table 1 below, was the largest integer for which no cancellation occured during 200 consecutive iterations.

In order to derive in a simple way, from the SEM scheme, a reliable pointwise estimate of the mixture parameters, we used an hybrid algorithm. We ran 200 SEM iterations. Then we ran 10 EM iterations starting from the position which achieved the largest value of the likelihood function among these 200 SEM iterations.

The choice of the rate of convergence to 0 of the sequence $\{\gamma_n\}$ for SAEM is very important and delicate, as for simulated annealing. From our experience, it seems that the following cooling schedule turns out to give good results : $\gamma_n = \cos n\alpha$, for $0 \leq n \leq 20$, and $\gamma_n = c/\sqrt{n}$, for $21 \leq n \leq 200$, where $\cos(20\,\alpha) = c/\sqrt{20} = 0.3$. The cooling schedule for $n \leq 20$ made it possible to obtain a comparatively small number of overestimations of K, whereas the schedule for $n \geq 21$ fitted the assumptions on γ_n in Theorem 1. Note that the alternative rate $\gamma_n = c\,(\log n)^{-v}$ revealed to be much too slow. For conciseness we only report the SAEM results using this particular cooling schedule.

		EM				SEM				SAEM			
	Size	100		60		100		60		100		60	
	KINIT	5	4	5	4	5	4	5	4	5	4	5	4
KFINAL	5	50	-	50	-	1	-	0	-	7	-	5	-
	4	0	50	0	50	32	24	7	7	42	34	43	27
	3	0	0	0	0	17	26	43	43	1	16	2	23

Table 1 Estimation of the number K of components for two sample sizes and two initial values of K.

We summarize these numerical experiments in three tables. Table 1 displays the frequency of KFINAL for the three algorithms from two positions of KINIT (5 or 4 components). From this table, it appears that EM needs to know the exact number K of components for good performance, and that, when it is initiated with the good K, it does not provide degenerate solutions, even for small samples. Concerning SEM, it turns out that its ability to find out the right K as reported in [2] does not hold for small sample sizes: If the results can be regarded as correct for the sample size 100 (56 good values out of 100 trials), the performances are poor for the sample size 60 (14 good values out of 100 trials). From our numerical simulations, it is clear that the estimation of K works better for SAEM, especially for the small sample size N = 60.

	EM		SEM		SAEM	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
p_1	0.280	0.143	0.243	0.047	0.254	0.077
p_2	0.262	0.101	0.238	0.039	0.242	0.070
p_3	0.205	0.090	0.265	0.058	0.251	0.075
p_4	0.253	0.097	0.255	0.044	0.253	0.064
m_1	2.344	0.701	2.004	0.052	2.094	0.390
m_2	5.958	2.096	4.980	0.127	5.194	1.038
m_3	9.791	2.306	9.067	0.260	9.293	1.302
m_4	15.000	1.288	14.969	0.645	15.026	0.803
σ_1^2	0.714	1.073	0.061	0.018	0.194	0.583
σ_2^2	0.770	1.550	0.220	0.078	0.415	0.918
σ_3^2	1.095	1.480	1.084	0.668	1.162	1.356
σ_4^2	3.910	3.250	3.642	1.997	3.613	2.463

Table 2 Means and Standard Deviation of the estimates of the mixture parameters for sample size 100 using EM, SEM and SAEM when they provide the right value of K.

	EM		SEM		SAEM	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
p_1	0.316	0.138	0.246	0.055	0.253	0.567
p_2	0.257	0.115	0.259	0.060	0.257	0.060
p_3	0.205	0.103	0.253	0.067	0.249	0.069
p_4	0.222	0.107	0.242	0.063	0.241	0.578
m_1	2.454	0.757	2.037	0.219	2.042	0.237
m_2	6.222	1.9525	5.102	0.616	5.134	0.672
m_3	10.129	2.534	9.139	0.787	9.162	0.890
m_4	15.280	1.444	15.070	0.638	15.013	0.734
σ_1^2	0.797	1.225	0.102	0.326	0.124	0.375
σ_2^2	0.962	1.882	0.254	0.136	0.304	0.540
σ_3^2	0.893	0.963	0.877	0.597	0.909	0.662
σ_4^2	3.252	2.997	3.343	2.476	3.493	2.236

Table 3 Means and Standard Deviation of the estimates of the mixture parameters for sample size 60 using EM, SEM and SAEM when they provide the right value of K.

Tables 2 and 3 display the mean and the standard deviation of the estimates of the mixture parameters using EM, SEM and SAEM when they provide the right value of K for both sample sizes. Some comments are in order.

Since the samples were small, the likelihood functions were littered with many local maxima, and thus the EM algorithm provided solutions which greatly depend on its initial position. This is apparent from the mean values of the first two component parameters, which are far from the true values, and from the large standard deviation of these estimates.

The behaviour of SEM is completely different from EM's one: either SEM stabilizes on the right number of components, in which case it converges to the right solution, or it fails to find the true number of components, in which case we did not record its results. This is the reason why the values of the means of the estimates of the parameters and the standard deviations are small in Table 2 and 3.

As for SAEM, it provides the right value of K in much more cases than SEM. But, the values of the means of the estimates of the parameters using SAEM are not as close to the true values of the parameters as for SEM, although they are good, and the standard deviations are sensibly larger than for SEM. This is a consequence of the fact that in some cases, the SAEM sequence was captured by a wrong local maximum of the likelihood function, far from the correct one.

Nevertheless, a detailed investigation of the results showed that EM provided good values of K and of the estimates of the parameters 66 times out of 200 trials (N =100 and 60, KINIT = 5 and 4), SEM provide good values 70 times out of 200 and SAEM provided good values 137 times out of 200. Finally, it can be seen that SAEM performs better for small sample sizes.

7. Conclusion

In this paper, we have introduced a new stochastic version of the EM algorithm for the statistical analysis of finite mixtures that we called the SAEM (Stochastic Approximation EM) algorithm. SAEM has a median position between EM and the SEM algorithm, an other stochastic version of EM that we have previously studied.

Like SEM, SAEM makes use of a probabilistic teacher scheme, but in contrast with SEM the random perturbations decrease to 0 as the iteration index grows to infinity.

From an heuristic perspective, we have compared SAEM with Simulated Annealing since both make use of a sequence of positive numbers decreasing to 0 to attenuate the intensity of the Monte-Carlo drawings.

We have studied in detail the behaviour of SAEM for mixtures of densities from some exponential familly. We have established a theorem which asserts that, under mild assumptions, the sequence generated by SAEM converges almost surely to a local maximizer of the likelihood function.

We have reported Monte-Carlo simulations which show that SAEM, like SEM, overcomes the limitations of EM (dependence on initial position, possible convergence to a saddle point of the likelihood function, the number of the mixture components has to be known, ...) and that SAEM performs better than SEM for small samples. Moreover, SAEM appears to be more tractable than SEM since SAEM provides almost sure convergence, while SEM provides convergence in distribution. Therefore, we think that SAEM, with a slow rate of convergence of $\{\gamma_n\}$, could be implemented in currently available software.

REFERENCES

[1] M. Broniatowski, G. Celeux and J. Diebolt, Reconnaissance de mélange de densités par un algorithme d'apprentissage probabiliste. *Data Analysis and Informatics* **3** (1983), 359-374.

[2] G. Celeux and J. Diebolt, The SEM algorithm: a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem. *Computational Statistics Quaterly* **2**, Issue 1 (1985), 73-82.

[3] G. Celeux and J. Diebolt, Comportement asymptotique d'un algorithme d'apprentissage probabiliste pour les mélanges de lois de probabilité. *Rapport de recherche INRIA* 563 (1986).

[4] G. Celeux and J. Diebolt, A probabilistie teacher algorithm for iterative maximum likelikood estimation. *Classification and related methods of Data Analysis*. North Holland (1987), 617-623.

[5] A.P. Dempster, N.M. Laird and D.B. Rubin, Maximum likelikood from incomplete data via the EM algorithm (with discussion). *J. R. Statist. Soc. B* **39** (1977), 1-38.

[6] P. Hall and C.C. Heyde, *Martingale limit theory and its application.* Academic Press, New York 1980.

[7] T.A. Louis, Finding the observed information matrix when using the EM algorithm. *J. R. Statist. Soc. B* **44** (1982), 226-233.

[8] G.J. McLachlan and K.E. Basford, *Mixture models - Inference and applications to clustering.* Marcel Dekker New York 1988.

[9] I. Meilijson, A fast improvement to the EM algorithm on its own terms. *J. R. Statist. Soc. B* **51** (1989), 127-[12]8.

[10] R.A. Redner and H.F. Walker, Mixtures densities, maximum likelikood and the EM algorithm. *SIAM Rev.* **26** (1984), 195-249.

[11] D.M. Titterington, A.F.M. Smith and U.E. Makov, *Statistical analysis of finite mixture distribution.* Wiley New York 1985.

[12] C.F. Wu, On the convergence properties of the EM algorithm. *Ann. Statis.* **11** (1983), 95-103.

APPENDIX

Recall the definition of the event $E_{n,j}$, $n \geq 1, j \geq 1$, which has been given in (5.42)-(5.43). We will prove the following technical result.

Proposition A.1. *There exists* $a > 0$ *such that with probability* 1 *the event* $E_{n,j}$ *occurs for an infinite number of integers* n, *whatever the fixed integer* $j \geq 1$.

Proof of Proposition A.1. We first establish the following lemma.

Lemma A.2. *There exists* $a > 0$ *such that*

$$\min+ = \inf_{\phi \in G} P_z \{w(\phi, \mathbf{z}) > a\} > 0 \qquad (A.1)$$

and

$$\min - = \inf_{\phi \in G} P_z \{w(\phi, \mathbf{z}) < -a\} > 0, \qquad (A.2)$$

where P_z *denotes the conditional probability given* ϕ *and* $w(\phi,\mathbf{z}) = < U(\phi, \mathbf{z}), v >_A$ *has been defined in (5.40).*

Proof of Lemma A.2. We will first prove that the conditional probability distribution $B \longrightarrow P_z \{U(\phi, \mathbf{z}) \in B\}$, for B in the Borel σ-field of **E**, is weakly continuous in the argument $\phi \in G$. The components of $U(\phi, \mathbf{z})$ are given by

$$p_j (\phi, \mathbf{z}) = N^{-1} \sum_{i=1}^{N} \{z_{ij} - t_j (x_i)\} \text{ for } j = 1, ..., K \qquad (A.3)$$

and

$$a_j(\phi, \mathbf{z}) = \{\sum_{i=1}^{N} z_{ij}\}^{-1} \{\sum_{i=1}^{N} z_{ij} b(x_i)\} - \{\sum_{i=1}^{N} t_j (x_i)\}^{-1} \{\sum_{i=1}^{N} t_j (x_i) b(x_i)\} \text{ for } j = 1, ..., K, \qquad (A.4)$$

respectively. Denote by f a real-valued continuous bounded function defined on **E**. By the formula of transfert we have

$$\int_{\mathbf{E}} f\{U(\phi, \mathbf{z})\} P_z \{U(\phi, \mathbf{z}) \in dh\} = \sum_{\mathbf{z} \in \mathbf{Z}} f\{p_j (\phi, \mathbf{z}), a_j (\phi, \mathbf{z}), 1 \leq j \leq K\} \mathbf{k}(\mathbf{z} \mid \mathbf{x} ; \phi), \qquad (A.5)$$

where $\mathbf{k} (\mathbf{z} \mid \mathbf{x} ; \phi) = \prod_{i=1}^{N} k(z_i \mid x_i ; \phi)$ has been defined in (2.3) and has been shown to be continuous in $\phi \in G$. As the $t_j (x_i)$'s are continuous functions of the argument $\phi \in G$, (A.3)-(A.5) together imply the required weak continuity result. This result entails in turn the weak continuity of $w(\phi ; \mathbf{z})$ with respect to $\phi \in G$. We will now prove (A.1) and (A.2) by contradiction. Denote by $F(t, \phi)$ the distribution function of the real-valued r.v. $\mathbf{z} \longrightarrow w(\phi, \mathbf{z})$, i.e. $F(t, \phi) = P_z \{w(\phi, \mathbf{z}) \leq t\}$ for $t \in \mathbf{R}$. Suppose that $\inf_{\phi \in G} 1 - F(t, \phi) = 0$ for all $t > 0$. Since G is a compact subset of **E**, there exists for all $t > 0$ a sequence $\{\phi_m(t)\}$ in G converging to some $\phi(t)$ in G as $m \longrightarrow \infty$, such that

$$\lim_{m\to\infty} 1 - F(t, \phi_m(t)) = 0. \tag{A.6}$$

If $t + u$ with $u > 0$ is any continuity point of $F(., \phi(t))$ larger than t, we have from (A.6) that

$$\begin{aligned} 0 \leq 1 - F(t + u, \phi(t)) &= \lim_{m\to\infty} 1 - F(t + u, \phi_m(t)) \\ &\leq \limsup_{m\to\infty} 1 - F(t, \phi_m(t)) \\ &= 0. \end{aligned} \tag{A.7}$$

Pick $t_k = 1/k$, $k \geq 1$, and select u_k so that $0 < u_k < 1/k$, $k \geq 1$. By the compacity of G, there exists a subsequence of $\{\phi(t_k) ; k \geq 1\}$ which converges to some ϕ in G. For the sake of simplicity, we still denote this subsequence $\{\phi(t_k) ; k \geq 1\}$. Let $a > 0$ be any positive continuity point of $F(., \phi)$. We have from (A.7) that

$$\begin{aligned} 0 \leq 1 - F(a, \phi) &= \lim_{k\to\infty} 1 - F(a, \phi(t_k)) \\ &\leq \limsup_{k\to\infty} 1 - F(t_k + u_k, \phi(t_k)) \\ &= 0. \end{aligned} \tag{A.8}$$

As $a > 0$ can be chosen arbitrarily close to 0, (A.8) implies that $P_z \{ < U(\phi, z), v >_A > 0\} = 1 - F(0+, \phi) = 0$, which contradicts Assumption (H5). The proof is similar if it is assumed that $\inf_{\phi \in G} F(t, \phi) = 0$ for all $t < 0$. Thus, the proof of Lemma A.2 is complete.

We can now return to the proof of Proposition A.1. This proof relies on the following conditional version of the Borel-Cantelli Lemma, as given in the Corollary 2.3, p. 32, of Hall and Heyde [6]. Let $\{\mathcal{F}_n ; n \geq 0\}$ denote a nondecreasing sequence of σ-fields and $\{A_n ; n \geq 0\}$ denote a sequence of events such that A_n is $\mathcal{F}_n$-measurable for all n. If $\sum P(A_{n+1} \mid \mathcal{F}_n) = \infty$ a.s., then A_n occurs infinitely often a.s. Here, $\mathcal{F}_n = \sigma_{(n+1)(j+1)}$, where $\sigma_m = \sigma(z^0, ..., z^m)$ is the σ-field generated by the random drawings $z^0, ..., z^m$, and $A_n = E_{nj,j}$, where $j \geq 1$ is assumed fixed. We will prove that there exists a positive δ such that $P(E_{nj,j} \mid \mathcal{F}_{n-1}) > \delta$ a.s. for all $n \geq 1$. We first give lower bounds for $P(E^+_{nj,j} \mid \mathcal{F}_{n-1})$ and $P(E^-_{nj,j} \mid \mathcal{F}_{n-1})$, respectively. Recall that $q_n = < \phi^n - \phi^*, v >_A$ and $w_n = < U(\phi^n, z^n), v>$ have been defined in (5.40). As q_{nj} is $\mathcal{F}_{n-1}$-measurable, we have

$$\begin{aligned} P(E^+_{nj,j} \mid \mathcal{F}_{n-1}) &= \int I\{q_{nj} \geq 0\}\, I\{w_{nj} \geq a\} \ldots I\{w_{nj+j} \geq a\} dz^{nj} \ldots dz^{nj+j} \\ &= I\{q_{nj} \geq 0\} \int I\{w_{nj} \geq a\} \ldots I\{w_{nj+j} \geq a\} dz^{nj} \ldots dz^{nj+j} \\ &= I\{q_{nj} \geq 0\} \int I\{w_{nj} \geq a\}\, dz^{nj} \ldots \int I\{w_{nj+j} \geq a\}\, dz^{nj+j}, \end{aligned} \tag{A.9}$$

where $I\{A\}$ is the indicator function of the event A. Now, (A.1) in Lemma A.2 implies that $a > 0$ can be chosen so small that

$$\int I\{w_{nj+j} \geq a\}\, dz^{nj+j} = P_z \{w(\phi^{nj+j}, z) \geq a\} \geq \min + > 0. \tag{A.10}$$

Substituting (A.10) into (A.9) we obtain

$$P(E^{+}_{nj,j} \mid \mathcal{F}_{n-1}) \geq (\min+)\ I\{q_{nj} \geq 0\} \int I\{w_{nj} \geq a\}\ dz^{nj} \ldots \int I\{w_{nj+j-1} \geq a\}\ dz^{nj+j-1}. \quad (A.11)$$

Now, (A.10) is still true with j-1 replacing j. Proceeding recursively, we finally obtain

$$P(E^{+}_{nj,j} \mid \mathcal{F}_{n-1}) \geq I\{q_{nj} \geq 0\}\ (\min+)^{j+1} \text{ a.s.} \qquad (A.12)$$

Similarly,

$$P(E^{-}_{nj,j} \mid \mathcal{F}_{n-1}) \geq I\{q_{nj} \leq 0\}\ (\min-)^{j+1} \text{ a.s.} \qquad (A.13)$$

Adding (A.12) and (A.13) we obtain

$$\begin{aligned} P(E_{nj,j} \mid \mathcal{F}_{n-1}) &\geq [I\{q_{nj} \geq 0\} + I\{q_{nj} \leq 0\}] \inf(\min+, \min-)^{j+1} \text{ a.s.} \\ &\geq \inf(\min+, \min-)^{j+1} \text{ a.s.,} \end{aligned}$$

which provides the required result. Since if $E_{nj,j}$ occurs i.o. then the same is true for $E_{n,j}$, this completes the proof of Proposition A.1.

Imprimé en France
par
l'Institut National de Recherche en Informatique et en Automatique.

ISSN 0249-6399