Mathematica
Balkanica

New Series Vol. 23, 2009, Fasc. 3-4

EM Algorithm and Varible Neighborhood Search for Fitting Finite Mixture Model Parameters

A. Bessadok [1], *P. Hansen* [2] *and A. Rebaï* [3]

Finding maximum likelihood parameter values for Finite Mixture Model (FMM) is often done with the Expectation Maximization (EM) algorithm. However the choice of initial values can severely affect the time to attain convergence of the algorithm and its efficiency in finding global maxima. We alleviate this defect by embedding the EM algorithm within the variable Neighborhood Search (VNS) methaheurestic framework. Computational experiment in several problems in literature as well as some larger ones are reported.

Key Words: Expectation Maximization algorithm, Metaheuristic, Variable Neighborhood Search, Maximum Likelihood Estimation, Finite Gaussian Mixture Model and Global Optimization

1. Introduction Finite Mixture Models (FMM) are strong tools for modeling of a wide variety of random phenomena and to cluster data sets [2]. They provide an efficient platform for apprehending data with complex structure; (see [8]). Because of their usefulness as a very flexible method of modeling, FMM have proved to be of great interest over the years, both in theory and practice; see [63]. Many problems of biology, physics and social sciences are modelled using a finite mixture of distributions; (see [55]). Recently mixture model-based methods has became very popular in cluster analysis, (see [42]) for an application of FMM to micro array data. Modeling using mixture distributions consists in the determination of the estimation of model parameters. Maximum Likelihood Estimation (MLE) is considered to be the best method among many others, such as the method of moments used by Pearson [54] and by Cohen [12], and the graphical techniques deployed during the early and mid-1900's by Harding [31] and extended by Cassie [11] (see [21]). For mixture models problems, the likelihood equations are almost always nonlinear and can not be solved by analytic means. Consequently, one must resort to seeking an aproximate solution via some iterative procedure. Our main interest here, is in a special iterative method called by Dempster et al, [14] EM algorithm.

However, the EM algorithm has some weaknesses in practice. Indeed it does not automatically provide an estimate of the covariance matrix of the parameter estimates, it is sometimes very slow to converge and in some problems, the E- or M-steps may be analytically intractable, and may converge to local optima [[62], [43]]. Moreover, one should emphasize that in the FMM the likelihood function is usually not unimodal and the likelihood equation has multiple roots corresponding each to local maxima; hence the EM algorithm will be very sensitive to the choice of starting values. Wu [72] reported that in general, if the log-likelihood has several (local or global) maxima and stationary points, like in the case of FMM, convergence of the EM sequence to either type of points depends on the choice of starting point.

Many variants of the original EM algorithm have been proposed in this last decade by several authors to overcome this local maxima problem. Mclachlan [43] proposed the use of principle components to provide suitable starting values in FMM context. Wright and Kennedy [71] use interval analysis methods to locate multiple stationary points of a log likelihood within any designated region of the parameter space. Ueda and Nakano [66] presented a deterministic annealing EM algorithm (DAEM) method for the EM iterative process in order to be able to recover from a poor choice of starting value. By introducing the temperature parameters to modify the posteriori probability in the E-step, they provided the EM with the ability to avoid the local maxima in some specific cases but in the process make it slow so this approach may not be appropriate in general (see [45]). Moreover, the DAEM and other similar extensions of EM are useless with respect to the problem of inappropriate distribution of the components in data space when locally trapped. Recently, in order to circumvent the local optimum problem of the EM algorithm for parameter estimation in a FMM, Ueda et al. [67] proposed a split-and-merge EM (SMEM) algorithm in which they applied a split-and-merge operation to the usual EM algorithm. The basic idea of the SMEM algorithm is the folowing: after convergence of the usual EM algorithm, one first uses the split-and-merge operation to update the values of some parameters among all the parameters, then one performs the next round of the usual EM algorithm, and alternatively iterate the split-and-merge operation and the EM algorithm until some stopping criterion is met. Obviously, this not only benefits from the appealing simplicity of the usual EM algorithm, but also improves its global convergence capability. Vlassis and Likas [68] proposed a Greedy EM algorithm for learning a FMM to overcome the limitation of getting trapped in one of the many local maxima of the likelihood function when using the EM algorithm.

The choice of initial values is considered as a crucial point in the algorithm-based literature, as it can severely affect the time to convergence of the algorithm and its efficiency to pinpoint the global maxima (see [10]). Finch et al. [19] used a quasi-Newton method as an iterative algorithm and proposed, for two component Gaussian mixture, that only the mixing weight should be given an initial value and the rest of the parameters be automatically estimated based on this values. Karlis and Xekalki [36] presented a brief review of such method. In recent paper Biernacki, Celeux, and Govaert [5] , propose a method for getting the highest likelihood value in the framework of FMM: they identify a strategy which is based on random initialization of EM, characterized in three steps search/run/select in a fixed number of iterations. Biernacki [6] proposes a strategy to initialize the EM algorithm in FMM context by defining a starting value distribution on a mixture parameter space including all possible EM trajectories.

In this paper we develop a new method that combines a metaheuristic variable neighborhood search (VNS) algorithm with the EM algorithm to overcome as much as possible the local maxima problem.

The present work thus illustrates the association of these two algorithms to overcome the negative effect of the poor starting value choices; it is organized in six sections: (i) after this introduction (ii) we give a general presentation of the FMM and MLE in literature and formulation of the data to be treated, (iii) introduction and application of the EM algorithm for the FMM, (iv) presentation of the VNS algorithm associated with EM algorithm as a basic solution (v) some implementation issues, and simulation as an experimental results (vi) summary and discussion of the general procedure and results of the new approach in conclusions.

2. General model and data presentation

2.1 Finite Mixture Model Formulation

As a simple and general presentation of mixture distributions we suppose a random variable, X, takes values in a sample space, Ω, with probability distribution $P(x)$, where x is its realization. The probability distribution can be written in the form of

$$P(x) = \sum_{j=1}^{k} \pi_j P_j(x) \tag{1}$$

where $P_j, j = 1, 2, ..., k$ are the components distribution of the mixture, verifying a probability distribution properties. The $\pi_j, j = 1, 2, ..., k$ are called the mixing weights where $\pi_j > 0$ for $j = 1, 2, ..., k$ and $\sum_{j=1}^{k} \pi_j = 1$. Frequently, component distributions are assumed to have non parametric forms. In this case, they are parametrized by the elements of a set $\alpha = \alpha_1, \alpha_2, ..., \alpha_k$ where α_k is the unknown vector parameters from the k^{th} component of the mixture. The Eq(1) will be

$$P(x) = \sum_{j=1}^{k} \pi_j P_j(x|\alpha_j) \tag{2}$$

Thus, $\theta = \{\pi_1, \pi_2, ..., \pi_k, \alpha_1, \alpha_2, ..., \alpha_k\}$ is considered to be the complete collection of all FMM parameters. The mixture distribution then takes the form

$$P(x|\theta) = \sum_{j=1}^{k} \pi_j \tilde{P}_j(x|\alpha_j) \tag{3}$$

where, each of $\alpha_1, \alpha_2, ..., \alpha_k$ belongs to the same parameter space, denoted by Θ and $\tilde{P}_j(.|\alpha_j)$ is assumed to be a conditional distribution in each component. In this case $\pi = (\pi_1, \pi_2, ..., \pi_k)$ may be defined as a probability distribution over θ, where $\pi_j = Pr(\alpha = \alpha_j), j = 1, 2, ..., k$.

2.2 Maximum Likelihood Estimation formulation

Given $X = \{X_1, X_2, ..., X_i, ..., X_m\}$, considered as m independent observations from the mixture, their joint probability distribution is the product of each individual distribution. Therefore the likelihood function is given by

$$P(X|\theta) = \prod_{i=1}^{m} [\sum_{j=1}^{k} \pi_j \tilde{P}_j(x_i|\alpha_j)] \tag{4}$$

The maximum likelihood principle states that we should choose as an estimate of θ, a value of the observed data x which maximizes $P(x|\theta)$. That is,

$$\theta^* = \arg\max_{\theta} P(X|\theta). \tag{5}$$

θ^* is called the MLE of θ. In order to estimate θ, it is typical to introduce the log likelihood function defined as

$$L(\theta) = \ln P(X|\theta) \tag{6}$$

Since $\ln P(x|\theta)$ is a strictly increasing function, the value of θ which maximizes $P(x|\theta)$ also maximizes $L(\theta)$. MLE corresponds to a solution of the following likelihood equation:

$$\partial L/\partial \theta = 0. \tag{7}$$

Involving the log of the sum makes the maximization of L numerically difficult. $Eq(7)$ becomes non linear and closed form solution of this equation cannot be found. Therefore, some iterative methods should be applied.

The likelihood log function associated to this model comprises multiple local maxima. In such situations, the MLE must be sought numerically using non linear optimization algorithms and it may possible to compute iteratively the MLE by using iterative procedures. There are many general iterative procedures which are suitable for finding an approximate solution of likelihood equations. Rao [56] and Mendenhall [47] developed iterative procedures which successfully used to obtain approximate solutions of nonlinear equations satisfied by MLE. The main methods deployed are the Newton-Raphson maximization procedure or some variant such as Fisher's scoring method and quasi-Newton methods. Our main interest here, however, is in a special iterative method which is unrelated to the above ones and which has been applied to a wide variety of mixture problems over the last fifteen or so years called the EM algorithm.

3. EM algorithm

3.1 EM Derivation

In the comments folowing in Eq (7), estimating FMM parameters using MLE amounts to solving a non-linear system of equations. However, the intuitive idea behind EM algorithm shows that, if we know the component of the mixture from which an observed data point is generated then the problem would be simpler and we get a closed-form solution of $Eq(7)$. Since this information is not known when the data is observed, the sample of data is then termed as incomplete data. Define the complete data to be the fully categorized data and the problem will be reformulated as an incomplete or hidden problem. Let $C = (O, H)$ being a sample of the complete data where $O = \{o_1, o_2, .., o_i, .., o_m\}$ is the sample of m observed data and $H = \{h_1, h_2, ..., h_i, ..., h_m\}$ defined as a sample of m hidden data. Each $h_i = (h_{i1}, h_{i2}, ..., h_{ij}, ..., h_{ik})$ where $h_{ij} \in \{0, 1\}$ corresponding to the mixture component to which o_i belongs. Thus, the likelihood function for the complete data can be written as

$$P(C|\theta) = \prod_{i=1}^{m} \prod_{j=1}^{k} [\pi_j P_j(o_i|\alpha_j)]^{h_{ij}} \tag{8}$$

This function is considered to be as joint complete distribution where the marginal distribution of O will be

$$P(O|\theta) = \sum_h P(C|\theta) = \sum_z P(O|h,\theta)P(h|\theta) \tag{9}$$

The goal of EM in its basic idea is to find θ such that the likelihood $P(O|\theta)$ or equivalently $L(\theta)$ is maximized. EM algorithm is an iterative procedure for maximizing $L(\theta)$. Assume that after the n^{th} iteration the current estimate for θ is given by $\theta^{(n)}$. Since the objective is to maximize $L(\theta)$, we wish to compute an update estimate θ, such that

$$L(\theta) > L(\theta^{(n)}) \tag{10}$$

Equivalently, we want to find a new updated parameter, say $\theta^{(n+1)}$ such

$$\theta^{(n+1)} = \arg\max_\theta \{L(\theta) - L(\theta^{(n)})\} \tag{11}$$

then the difference is

$$\begin{aligned} L(\theta) - L(\theta^{(n)}) &= \log(\textstyle\sum_h P(O|h,\theta)P(h|\theta)) - \log(P(O|\theta^{(n)}) \\ &= \log(\frac{\sum_h P(O|h,\theta)P(h|\theta)}{P(O|\theta^{(n)})}) \end{aligned} \tag{12}$$

However, using Bayes rules we get

$$P(h|O,\theta^{(n)}) = \frac{P(O,h|\theta^{(n)})}{P(O|\theta^{(n)})} \tag{13}$$

the $Eq(12)$ could be written with

$$L(\theta) - L(\theta^{(n)}) = \log[\frac{\sum_h P(O|h,\theta)P(h|\theta).P(h|O,\theta^{)n)})}{P(O,h|\theta^{(n)})}] \tag{14}$$

Notice that this expression involves the logarithm of a sum. Since $P(h|O,\theta^{(n)})$ is a probability measure, we have that $P(h|O,\theta^{(n)}) \geq 0$ and $\sum_h P(h|O,\theta^{(n)}) = 1$. For instance, we may apply Jensen's inequality [1] to get

$$L(\theta) - L(\theta^{(n)}) \geq \sum_h P(h|O,\theta^{(n)}) \log(\frac{P(O|h,\theta)P(h|\theta)}{P(O,h|\theta^{(n)})}) \tag{15}$$

$$\triangleq \Delta(\theta|\theta^{(n)}) \tag{16}$$

[1] for constants $\lambda_i \geq 0$ with $\sum_i \lambda_i = 1$ it shown that $\log \sum_i \lambda_i x_i \geq \sum_i \lambda_i \log(x_i)$

equivalently, Eq(11) will be,

$$\theta^{(n+1)} = \arg\max_{\theta}\{\Delta(\theta|\theta^{(n)})\} \quad (17)$$

$$= \arg\max_{\theta}\sum_{h} P(h|O,\theta^{(n)})\log P(O,h|\theta) \quad (18)$$

$$= \arg\max_{\theta} E_{h|O,\theta^{(n)}}\log P(O,h|\theta) \quad (19)$$

$$= \arg\max_{\theta} E_{h|O,\theta^{(n)}}\ L(C|\theta) \quad (20)$$

$$= \arg\max_{\theta} Q(\theta;\theta^{(n)}) \quad (21)$$

In going from $Eq(17)$ to $Eq(18)$ we drop terms which are constant with respect to θ. Hence, the $Eq(20)$ shows well the two EM algorithm steps which are :

1. **E-step**: Compute the conditional expectation of the complete data log likelihood $L_c(\theta)$, given the observed data O, using the current iteration $\theta^{(n)}$ for θ.

2. **M-step**: Update the value of θ, say $\theta^{(n+1)}$, that maximizes the E-step

The **E-** and **M-** steps are continuously repeated until the difference $|\theta^{(n+1)}-\theta^{(n)}|$ or $|L(\theta^{(n+1)}) - L(\theta^{(n)})|$ changes by an arbitrarily small amount.

3.2 EM convergence

The crucial property of the EM algorithm proved by Dempster et al. (1977) is that the observed data log-likelihood $L(\theta)$ can never decrease during the EM sequence. This process continues until $L(\theta)$ converges. In fact we had in $Eq(16)$ that

$$\begin{aligned} L(\theta) &\geq L(\theta^{(n)}) + \Delta(\theta|\theta^{(n)}) \\ &\triangleq \ell(\theta|\theta^{(n)}) \end{aligned}$$

Additionally, it is staightforward to show that $\Delta(\theta^{(n)}|\theta^{(n)}) = 0$, hence,

$$\ell(\theta|\theta^{(n)}) = L(\theta^{(n)}) \quad (22)$$

thus, the $\ell(\theta|\theta^{(n)})$ function is bounded above by the $L(\theta)$ likelihood function. The following Figure 1, illustrates the EM procedure for two iterations.

Once our objective is to find $\theta^{(n+1)}$ that maximize $L(\theta)$, we can deduct from Figure1, that $\ell(\theta^{(n+1)}|\theta^{(n)}) \geq L(\theta^{(n)}|\theta^{(n)}) = L(\theta^{(n)})$. Therefore, at each iteration, $L(\theta)$ cannot decrease.

3.3 EM and Mixture Model

In MM context, the complete data likelihood in $Eq(8)$ is

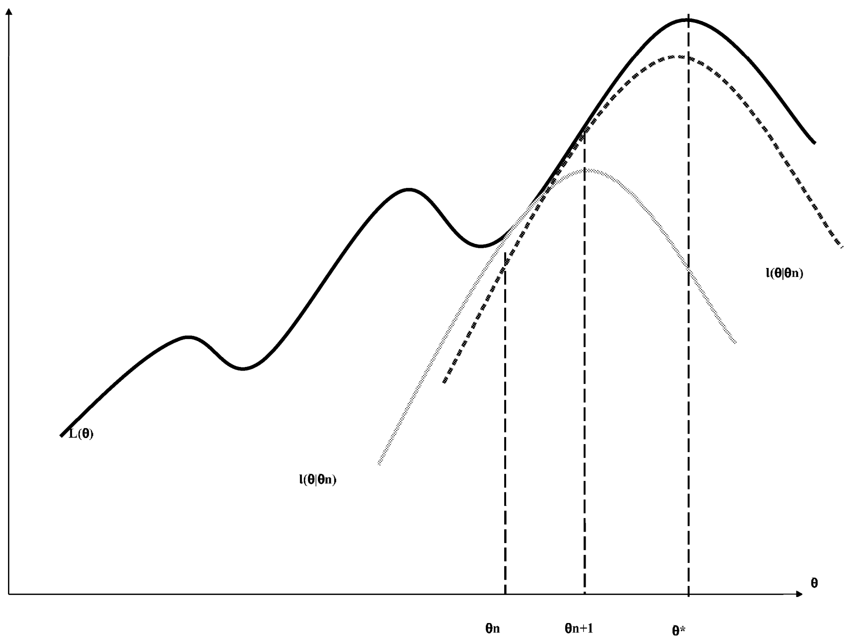


Figure 1: EM computes the function $\ell(\theta)$ using the current estimate $\theta^{(n)}$ and choose the update estimate $\theta^{(n)}$ as the maximum point of $\ell(\theta)$. In the next iteration at θ^{*} the same $\ell(\theta)$ will be generated causing the algorithm to end.

$$L(C|\theta) = \sum_{i=1}^{m}\sum_{j=1}^{k} h_{ij}\log\{\pi_j P_j(o_i|\alpha_j)\} \tag{23}$$

Since $L(C|\theta)$ is a linear function of the hidden indicator variables h_{ij}, the $E{-}step$ is reduced to the computation of the conditional expectation of h_{ij}, which, given an obseved data o_i, using the current estimate $\theta^{(n)}$ for θ, is computed as

$$\tau_{ij}^{(n)} = E[H_{ij}|o_i;\theta^{(n)}] = P(H_{ij}=1|o_i;\theta^{(n)}) \tag{24}$$

for each i,j, which is the current estimate of the posterior probability of the i^{th} observation generated from j component conditional on P_i and $\theta^{(n)}$ gived by

$$\tau_{ij}^{(n)} = \frac{\pi_j^{(n-1)}P_j(o_i;\alpha_j^{(n-1)})}{\sum_{j=1}^{k}\pi_j^{(n-1)}P_j(p_i;\alpha_j^{(n-1)})} \tag{25}$$

On the **M-step** of $(n+1)^{th}$ iteration, we update the value of θ, say $\theta^{(n+1)}$, that maximizes

$$Q(\theta;\theta^{(n)}) = \sum_{i=1}^{m}\sum_{j=1}^{k} \tau_{ij}^{(n)} \log\{\pi_j P_j(o_i|\alpha_j)\} \tag{26}$$

For MM, the estimation of mixing weight is done by differentiating $Q(\theta;\theta^{(n)}) - \lambda(\sum_{j=1}^{k} \pi_j - 1)$ with respect to π_j and setting derivative equal to zero, where λ is a lagrange multiplier, one has

$$\pi_j^{(n+1)} = \frac{1}{m}\sum_{i=1}^{m} \tau_{ij}^{(n)} \tag{27}$$

As for the updating of θ, it is obtained as an appropriate root of

$$\sum_{i=1}^{m}\sum_{j=1}^{k} \tau_{ij}^{(m)} \frac{\partial \log P_j(o_i|\alpha_j)}{\partial\theta} = 0 \tag{28}$$

In the FMM context as shown in example of Figure 1, the likelihood function will most probably have many local maxima, especially when the number of mixture-components is large (see [35]). However, performing of the EM algorithm in FMM context will provide a local maxima of the likelihood function of the observed data tend to converge to a local optima, and not necessarily the global one as results by Xu and Jordan [73]. Consequently, the EM algorithm will be very sensitive to the choice of the initial value $\theta^{(0)}$. Specifically, the effectiveness of the EM algorithm considerably depends on this first initialization. In the next section, we will present a new variant of the EM algorithm to alleviate the influence of the initial values on the performance of FMM estimation parameters.

4. Embedding EM in VNS

4.1 EM and Global optimization problem

When there are many local optima the EM algorithm can, and often does, lead to one of them instead of to the global maxima. In fact, as shown in section 3.2, in the M-step of the EM algorithm we are concerned with finding $\theta^{(n+1)}$ such that

$$\theta^{(n+1)} = \arg\max_{\theta} Q(\theta|\theta^{(n)}) \tag{29}$$

The $Q(\theta|\theta^{(n)})$ function is multimodal. However, we can formulate the problem of the form (29) as;

$$\text{(30)} \qquad \begin{array}{rc} \max & f(\theta) = Q(\theta|\theta^{(n)}) \\ \text{subject to} & \theta \in \Theta \end{array}$$

where $f(\theta)$ is the objective function to be maximized and Θ is the set of feasible solutions. A solution $\theta^* \in \Theta$ is optimal if $f(\theta^*) \geq f(\theta)$, $\forall\theta \in \Theta$.In other words, we can view this problem as a global optimization one.

Woodruff and Reiners [70] considered the modeling of such sophisticated topics commonly attends to $NP - hard$ problem. Such problems, paticulary when involve continious variables are often very difficult to solve. So one can either limit oneself to smal instances solvable golbal optomization technique, or limit oneself to heurestic optimization. the later is infact done by EM, but as we will show experementely below embedding it in the VNS format leads to a more efecien algorithm in terms of values of soliution obtained wile steel keepeng resolution time reasonable.

Combinatorial and global optimization algorithms are typically attracted in solving instances of problems that are believed to be hard in general. However, the use of available exact algorithms such branch-and-bound, cutting planes, decomposition, Lagrangian relaxation, column generation, and many others may not reach to solve very large instances. Moreover, Hansen and Mladenoviĉ [29], asserted that many practical instances of such problems of the form (30), arising in Operations Research and other fields, are too complex for an exact solution to be realized in conceivable time. Thus one has to endeavor to heuristics, which provide an approximate solution, or sometimes an optimal but without proof of its optimality but for the favor of a reasonable time realization. Local search is considered as one of the most used type of heuristic [28]. Local search algorithms move from an initial solution to another neighborhood solution in the space of candidate solutions by alternation of local changes, which improve at each time the objective function, until a solution deemed optimal is found or a time bound is elapsed. Although, an EM algorithm can be employed to perform local search for the problem of the form (30) [68].

4.2 VNS and Heuristics

Heuristic search procedures that aspire to find global optimal solutions to hard combinatorial optimization problems usually require some type of adjustment to overcome local optimality. In recent years, many authors extended this methodology and developed several metaheuristics algorithms for avoid being trapped in local optima with a poor value. The most relevant procedures in terms of their application to a wide variety of problems are: Tabu Search is by now a well-known metaheuristic for solving hard combinatorial optimization problems ([23], [24], [25]), Multi Start (MS) methods are devoted to the Monte Carlo random re-start in the context of nonlinear unconstrained optimization, where the method simply evaluates the objective function at randomly generated points [58], adaptive multi-start [9], simulated annealing [37], one of the most well known MS methods is the greedy adaptive search procedures (GRASP). The GRASP methodology was introduced by Feo and Resende [18] and many others have contributed to an abundant enhanced results for many combinatorial problems.

However, the performance apprehended and the sophistication of such heuristics makes it difficult to allocate with a accuracy the reasons for their effectiveness. Mladenoviȁ and Hansen [52] had examined a change of neighborhood in the search as a relatively unexplored reason and they proposed a new optimization technique called variable neighborhood search VNS.

VNS is a metaheuristic method that embeds a local search heuristic for solving combinatorial and global optimization problems. VNS performances systematically the idea of neighborhood change, both in ascendant to local maxima and in escape from the hills which contain them [30].

4.3 VNS algorithm

Local search algorithms are widely applied to numerous hard computational problems, including problems from computer science and in particularly artificial intelligence, mathematics, operations research, engineering, and bioinformatics. Moreover, they often are bulding blocks for more sophisticated heuristics. A basic scheme for local search can be presented as follows

---

Initialization. Select a neighborhood structure N, that will be used in the search; find an initial solution x;

Repeat the following until the stopping condition (i.e., finding a local optimum) is met:

(a) Find the best neighbor $x' \in N(x)$ of x;

(b) if x' is not better than x, stop. Otherwise, set $x = x'$ and return to (a);

Step of local search heuristic.

The stopping condition of this heuristic, using one neighborhood structure is staisfied as soon as a local optimum is reached. In our study the EM algorithm itself is considered as a local search structure. To improve upon the basic scheme so obtained, one can use a MS strategy that iterates for a number of times the local search from initial solution generated randomly until no further progress is made or an limit for computing time for the step is reach. However, for a considerable number of local optima, the best of those found by MS may be very far from the global optimum (Boese, Kahng and Muddu, [9]). Actually, the MS method concentrates in exploring many hills but without exploring properties of local optima so found.

VNS and contrary to other metaheuristics based on local search methods, does not follow a trajectory but explores increasingly distant neighborhoods of the current incumbent solution, and jumps from this solution to a new one if and only if an improvement has been made. Several questions about selection of neighborhood structures are in order [29]:

(i) What properties of the neighborhoods are mandatory for the resulting scheme to be able to find a globally optimal or near-optimal solution?

(ii) What properties of the neighborhoods will be helpful in finding a near-optimal solution?

(iii) Should neighborhoods be nested? Otherwise how should they be ordered?

(iv) What are desirable properties of the sizes of neighborhoods?

The basic VNS method described by Hansen and Mladenoviċ ([27], [28], [29]), combines deterministic and stochastic changes of neighborhood. Denote with $N_k (k = 1, 2, .., k_{max})$ a finite set of preselected neighborhood structures, and with $N_k(x)$ the set of solutions in the k^{th} neighborhood of x. Its steps are given as:

Initialization. Select the set of neighborhood structures N_k, $k = 1, 2, ...,$ k_{max}, that will be used in the search; find an initial solution x; choose a stopping condition;

Repeat the following until the stopping condition is met:

1. set $k \leftarrow 1$;
2. Repeat the following steps until $k = k_{max}$:
   (a) Shaking. Generate a point x' at random from the k^{th} neighborhood of x ($x' \in N_k(x)$);
   (b) Local search. Apply some local search method with x' as initial solution; denote with x'' the so obtained local optimum;

(c) Move or not. If this local optimum is better than the incumbent, move there ($x \leftarrow x''$),and continue the search with $N_1(k \leftarrow 1)$; otherwise, set $k \leftarrow k + 1$;

Steps of the basic VNS

The stopping condition criteria could be such as maximum fixed number of iterations, maximum CPU time allowed, or maximum number of iterations since the last increase in the Log likelihood function. One of the major challenges in the metaheuristic VNS is the selection of neighborhood structures properties and desirable properties of the sizes of neighborhoods in away to be able to find a global optimal or best optimal solution and to do so fairly in a reasonable time realization. In fact, To avoid being blocked in a hill, while there may be higher ones, Hansen and Mladenoviȁ [29] suggested that the union of the neighborhoods around any feasible solution θ should contain the whole feasible set:

$\Theta \subseteq N_1(x) \cup N_2(x) \cup ... \cup N_{kmax}(x), \forall x \in X.$

These sets may cover X without necessarily partitioning it, which is easier to implement, e.g. when using nested neighborhoods, i.e.,

$N_1(x) \subset N_2(x) \subset ..... \subset N_{kmax}(x), \forall x \in X.$

If these properties do not hold, one might still be able to explore X completely, by traversing small neighborhoods around parameters values on some trajectory, but it is no more guaranteed. For instance, we define a first neighborhood $N_1(x)$ as a subdivision of the interval data range and then iterating it k times to obtain neighborhoods $N_k(x)$ for $k = 2, ..., k_{max}$. They have the property that their sizes are increasing. Therefore if, as is often the case, one goes many times through the whole sequence of neighborhoods the first ones will be explored more thoroughly than the last ones.

4.4 EM and VNS

To let EM algorithm to be not totally depending on the first initialization is to reformulated it using the VNS method. We may consider EM as a local search in global optimization context and estimating the parameters model by maximizing the Log likelihood function subject to each parameters belong to the set of feasible solutions. We define the neighborhood structures as subintervals obtained from the data distribution range. The algorithm will be a combination of the EM and VNS (EMVNS). Therefore the basic EMVNS steps are:

Initialization. Choose an initial solution θ; select the set of neighborhood structure by defining the intervals range I_p for the means, covariances and mixing weights parameters and by choosing the maximum number of embedded intervals in I_p (I_{pk}, $k = 1, 2, ..k_{max}$); choose a stopping condition, that will be used in the perturbation phase; choose a stopping condition;

Repeat the following until the stopping condition is met:

1. set $k \leftarrow 1$;
2. Repeat the following steps until $k = k_{max}$:
   - (a) Perturbation. Generate a parameter θ' at random from the k^{th} neighborhood of θ ($\theta' \in I_{pk}(\theta)$);
   - (b) Application of the EM algorithm with θ' as initial solution; denote with θ'' the so obtained local optimum;
   - (c) Move or not. If this local optimum is better than the incumbent, move there ($\theta \leftarrow \theta''$),and continue the search with $I_{p1}(k \leftarrow 1)$; otherwise, set $k \leftarrow k + 1$;

Steps of the basic EMVNS

A general procedure of the EMVNS approach is presented in Figure 2.

5. Application to Finite Gaussian Mixture Model

In this section we apply our method to eight FGMM examples with different degrees of complexity in order to show the effectiveness of the EMVNS approach compared to MS method over these degrees of the problem complexity.

5.1 EM and FGMM

To establish FGMM parameters estimation, we explicitly derive the EM steps for Finite d-dimensional Gaussian Mixture Model. The mixed weight π_j is the unknown probability of occurrence of the j^{th} component in the mixture. Assume that each Gaussian component parameteres α_j has a vector mean μ_j and covariance matrix $\Sigma_j = \sigma^2\mathbf{I}$ where Σ_j is a positif definite symmetric matrix. The marginal Finite d-dimentional Gaussian Model distribution is given by

$$P(O) = \sum_{j=1}^{k} \frac{\pi_j}{(2\pi)^{d/2}|\Sigma_j|^{1/2}} \exp\{-\frac{1}{2}(o-\mu_j)^t\Sigma_j^{-1}(o-\mu_j)\} \tag{31}$$

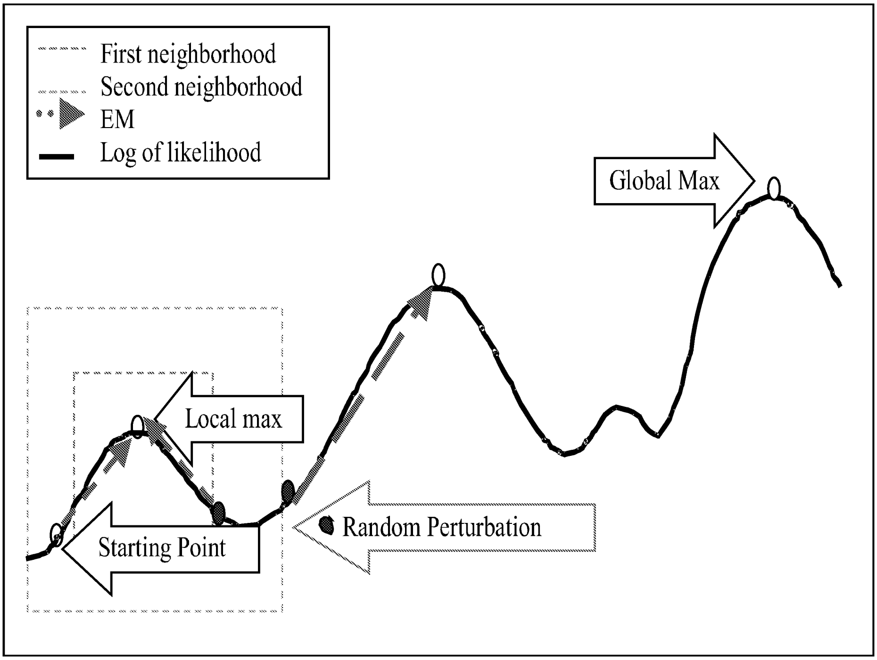


Figure 2: General procedure of EMVNS scheme

The parameters to be estimated are $\alpha_j = (\mu_j, \sigma_j)$, and π_j, $j = 1, 2, .., k$. Then, using the two steps for estimating the model parameters denoted by, $\theta = (\mu_j, \sigma_j, \pi_j; j = 1, 2..., k)$ it can shown that E-step:

$$\tau_{ij}^{(n)} = \frac{\sigma_j^{-d} exp\{-\|o - \mu_j(n)\|^2 / 2\sigma_j^2(n)\}}{\sum_{j=1}^{k} \sigma_j^{-d}(n) exp\{-\|o - \mu_j(n)\|^2 / 2\sigma_j^2(n)\}} \tag{32}$$

M-step:

$$\pi_j^{(n+1)} = \frac{1}{m} \sum_{i=1}^{m} \tau_{ij} \tag{33}$$

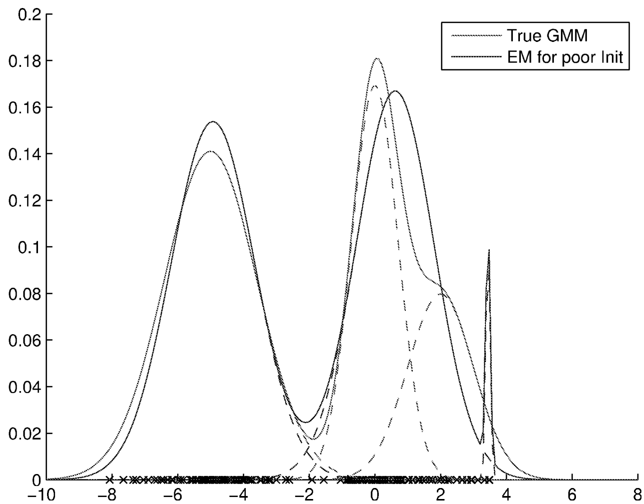
$$\mu_j^{(n+1)} = \frac{\sum_{i=1}^{m} \tau_{ij} o_i}{\sum_{i=1}^{m} \tau_{ij}} \tag{34}$$

$$\sigma_j^{(n+1)} = \frac{\sum_{i=1}^{m} \tau_{ij} \|o_i - \mu_j(n+1)\|^2}{\sum_{i=1}^{m} \tau_{ij}} \tag{35}$$

We denote by c-G-d-MM the c component of Gaussian distribution with d dimensional Mixture Model (e.g 6G2MM is FGMM with 6 components in two dimension space). The eight examples chosen are displayed in table .

5.2 Experimental Procedure

(a) EM is applied with poor initialization



(b) using EMVNS with the same poor initialization

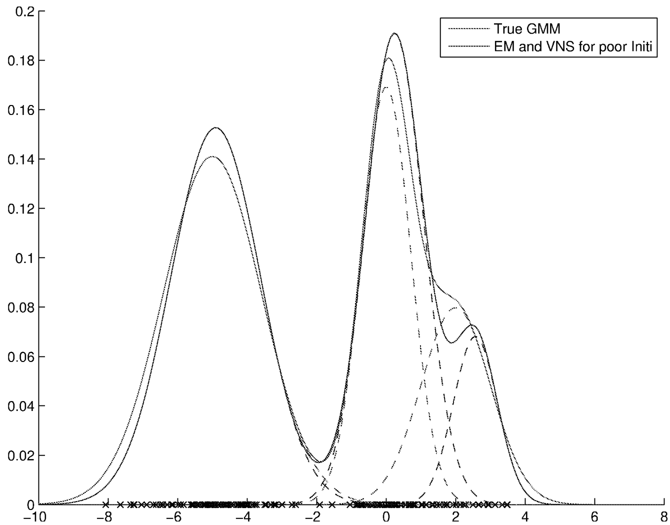


Figure 3: The use of EM can lead in general to a local maxima and not to the global one. However the implementation of EMVNS using the same poor initialization guarantee a better result.

Before detailing the experimental procedure, we describe the sensibility of EM using a "poor" initialization to get a local maxima. We use 150 observations generated from 3G1MM described in appendix A. Our algorithm is applied using the same 'poor' starting value and as shown in Figure 3 the EMVNS can easily improve the parameters model estimation.

Methods deployed In this paper, we limit our study to comparing the EMVNS with the MS method. Indeed, the MS method is the most used way of initiating EM and is considered as a reference method for almost any comparison method (see [5]). In the MS method, the means are generated from an interval within the data range, while the covariances are generated from an interval ranging from zero to the value of sample covariance, and the mixing weights are generated from a Dirichlet distribution. In the EMVNS approach, we choose I_p for the means as the data range interval, the covariance as an interval ranging from zero to the value of sample covariance, and an interval ranging from 0 to 0.9 for mixing weights. Note that EM was used in its standard form and without any acceleration scheme. To avoid the unbounded likelihoods problems, all components are chosen to have a common variance or equal determinants (see [45]).

Treated examples In order to give an acceptable credibility to the results provided by both methods, we choose a non-exhaustive variety of situations. In fact, as we will discuss later, the EM performance results are directly related to the attraction basin size as presented in Figure 4 and to the dimension of local and global maxima of the likelihood function (see [4]).

Thus, we considered the eight examples displayed in appendix with different degrees of complexity depicted in both component dimension and data distribution. From the Probability Density Function (PDF) of 8 FGMM examples shown in Figure 5, we can qualify the diversity of data distribution from well separated as in 8G2MM example, to poorly separated as in 2G2MM example. Figure 6 provides the information about the approximate local maxima number from very reduced as in 2G1MM to very considerable as in 10G2MM. The dendrogramms in Figure 6 illustrate that the number of local maxima values is dependent on the problem complexity characterized by the number of components and data distribution.

Criteria selected To analyze the performance of each method we define some criteria. For instance, our main objective is to reach the highest likelihood regardless of the time of realization even though the CPU time realization is less for EMVNS than for MS method. In fact, for the last 10G2MM example, one iteration with MS took 0.26s and only 0.14s with EMVNS. To be more realistic, we choose a "poor" and random starting parameter values for EMVNS method as shown in Appendex A.

Therefore, we choose a fixed number of iterations as a stopping condition for both methods. To perform the competition between the two methods we limit to 100 the number of iterations relatively to the maximum number of local maxima as presented in Figure 6. We fixed to 25 the maximum number of embedded intervals and for simplicity we choose the same incremental step for getting all I_{pk} as 30% of I_p (i,e. $I_{p1} = 30\% I_p, I_{p2} = 60\% I_P$, etc). We devote 10 trials twice for each example sample. For both methods, we record after each trial the highest Log-likelihood considered as the best associated global maxima. The second criterion which is the local maxima range, gives us an idea about the ability of EMVNS method to improve the search for getting the best local maxima by jumping from hill to hill. In the MS method, this second criterion informs us about the approximately wide range of the local maxima.The last criterion is the percentage of getting the associated global one; it characterizes the degree of complexity of the problem we treat. This percentage provided by EMVNS reflects the percentage of time being in this global maxima hill. Nevertheless, from MS method this percentage explained essentially the attraction basin size of the global maximum. For instance, if this last percentage is around 70% it means that attraction basin size of the global maximum is very large and that in 70% of the cases MS succeeded to get to this global one.

5.3 Results obtained

The results obtained for both poor and random starting values are the same. Thus, the initial values have no effect on the performence of the EMVNS algorithm. The results displayed in Appendix B, show that there are four types of problem complexity as shown in Figure 4. In fact, we can consider a very simple complexity problem characterized by a large basin attraction size of the global maximum or a reduced local maxima number as illustrated by 2G1MM, 2G2MM and 4G2MM examples. The second class complexity is represented by a considerable number of local maxima having relatively the same large attraction basin size as presented in 8G2MM example. The third complexity problem class is represented by a small local optima number with a small attraction basin size as described by 3G1MM. The forth types of problem complexity, considered as the most challenged one, have a relatively huge local maxima number where the global one have a very small attraction basin size as showing in 3G2MM, 6G2MM and 10G2MM. In first simple class complexity problems, both methods reached the same global maxima. Thus, we can depict from Figure 6 that the number of local maxima associated to 2G1MM example is very much reduced and despite of the attraction basin size of global maxima it is easy in this case for any simple method to gain in 100 iterations the global one. For the 2G2MM and 4G2MM illustrated from Biernacki et al [5] examples with a considerable number of local maxima as shown in Figure 6, in this case it's easy to get to the global maxima regardless of whether or not the model components are well or poorly separated. The 3G2MM example have relatively average local maxima number as showing in Figure 4 with a small attraction basin size; this can be depicted from the percentage of getting the maximum global by MS method as presented in Appendix B. In this situation, EMVNS succeeded far better than MS to get to the highest hill in the majority of trials see table 1. In 8G2MM example, where the components of the model are well separated, the MS method succeeded in only about 10% trials in getting the highest log likelihood compared to the EMVNS method. Due to the fixed number of iterations (i.e stopping criterion), this is the only case where MS can perform well against the EMVNS. For the 3G1MM considered as a simple example in one dimension space with a reduced local maxima number, the EMVNS achieved in the majority of trials to get to the highest Log likelihood. In fact, when the attraction basin size is large for a local maxima and a small for the global maximum (as explained by the quite significant percentage for getting the global maximum with MS), EMVNS succeeded to jump to other local maxima until getting to best hill; meanwhile, MS get trapped in this local maximum. In 6G2MM and 10G2MM examples, having a significant important number of local maxima, the competition between both methods is very rude. Indeed, the percentage of getting the global maxima in MS is very small as it did not exceeded 2%, consequently, the most of local maxima had a very small attraction basin. In this case, EMVNS had a more chance to get to the best result. We can resume all results in table 1with defferent degrees of complixty such:

1 Large basin attraction size of the global maximum or a reduced local maxima number.

2 Considerable number of local maxima having relatively the same large attraction basin size.

3 Small local optima number with a small attraction basin size.

4 Huge local maxima number where the global one have a very small attraction basin size.

Therefore, we can deduce from the local maxima range and from the comparatively great percentage of getting the global maxima with EMVNS method,

Table 1: Percentage of getting the global maximum by both methods in 20 replications

COMPLEXITY PROBLEM DEGREE	MS	EMVNS
1	100 %	100 %
2	100 %	90 %
3	60 %	90 %
4	30 %	70 %

that in almost all situations there is no difficulty for the EMVNS method to improve the accuracy of model parameters estimation and attain the best global maximum.

5.4 Discussions

As an iterative method, EM algorithm had viewing his sensitive to the initial values. In fact, Figure 3 gives the change in accuracy result when EM performs from poor initial value. From the experimental results, we can briefly resume all possible situations in four cases depending on complexity problem classes. First, when the attraction basin of the local maxima is relatively large supported by the percentage of getting the global maxima with MS method such as more than 15%. The second case is illustrated by a very important number of local maxima with relatively the same attraction basin size (i.e. low percentage of global maximum ranging from 5 to 15%); in this particular case MS outperforms EMVNS only in 10% of the trials. The last case is the most challenging one, because the number of local maxima is very important with a small attraction basin size; this corresponds to the case of a small percentage in getting the global maxima with MS method (less than 5%). This in fact is the most difficult one for which EMVNS succeeded in the best part of trials to reach the greatest global maximum. In fact, depending on the complexity of the problem, the large number of local maxima conducts to large number of parameters estimation solutions. As described earlier, when the number of local maxima is significantly reduced or when the basin attraction of the global maxima is visibly large, we can use an ordinary method as MS or any other simple method for getting best results to overcome initial values problem caused by EM algorithm. (However in practice, the data dimension is very large modelled by FMM and without guarantees to have a large basin attraction of the global maximum). In spite of this, the need of applying a robust method for complex problems is necessary. As described in section 4.2, contrarily to other methods based in local search, VNS provides a powerful and simple tool for getting best results compared to other competing methods. Moreover, the use of the appropriate structures in VNS leads not only to improve the maximization of the likelihood function but to obtain it with the best time realization.

6. Conclusions

The choice of initial values is considered as crucial point in the algorithm-based literature as it can severely affect the time realization of convergence of the algorithm and its efficiency to pinpoint the global maximum.

A novel EMVNS algorithm for estimating FMM parameters is proposed in this paper to overcome one of the main drawbacks of EM algorithm often getting trapped at local maxima. The VNS method largely deployed in many examples had shown his efficient in getting best improvement results which exploits systematically the idea of neighborhood change, both in ascendant to local maxima and in escape from the hills which contain them. The algorithm is computationally efficient and easily implemented.

The experimental results of employing FGMM for a variety of degrees of complexity of data dimension show that our algorithm can find excellent solutions with best time realization than MS method, especially in complicated situations.

The EMVNS algorithm use the VNS in his basic scheme and is focus on the estimation of FMM parameters supposing the number of FMM components are known before. therefore, developing EMVNS using several VNS extensions and finding more appropriate structures for resolving such problem appears to be desirable.

References

[1] M. Aitkin, D. Anderson and J. Hinde, Statistical Modelling of Data on Teaching Styles (with discussion). *J.Roy. Stat.Soc.*, **A 144**, 1981, 419-461.

[2] M. Aitkin and D. B. RUBIN, Estimation and Hypothesis Testing in Finite Mixture Distributions,*Journal of the Royal Statistical Society, Series B*, **47**, 1985, 67-75.

[3] K. E. Basford and G. J. McLachlan, The Mixture Method of Clustering Applied to Three-Way Data, *Journal of Classification*, **2**, 1985, 109-125.

[4] A. Berchtold , Optimisation of mixture models: Comparison of different strategies, *Computational Statistics*, **19**, 2004, 385-406.

[5] C. Biernacki, G. Celeux, G. Govaert, Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate

Gaussian mixture models, *Comput. Statist. Data Anal.*, **41**, 2003, 561-575.

[6] C . Biernacki, Initializing em using the properties of its trajectories in gaussian mixtures. *Statistics and Computing*, **14**(3), 2004, 267-279.

[7] C. G. Bhattacharya, A Simple Method for Resolution of a Distribution into its Gaussian Components, *Biometrics* , **23**, 1967. 115-135.

[8] K. Blekas and I. E. Lagaris, Split-Merge Incremental Learning (SMILE) of Mixture Models . *Inter. Conference on Artificial Neural Networks (ICANN)*, Porto 2007, *Lecture Notes on Artificial Neural Networks*, **4669**, 2007, 291-300.

[9] K. D. Boese, A.B. Kahng and S. Muddu, A New Adaptive Multi-Start Technique for Combinatorial Global Optimizations. *Oper. Res. Lett.*, **16**, 1994, 101-113.

[10]D. B hning , *Computer-Assisted Analysis of Mixtures and Applications: Meta-Analysis, Disease Mapping and Others.* , Chapman and Hall/CRC, New York, 1999.

[11] R. M. Cassie, Some Uses of Probability Paper for the Graphical Analysis of Polymodel Frequency Distributions, *Australian Journal of Marine and Freshwater Research*, **5**, 1954, 513-522.

[12] A.C. Cohen,Jr., Estimation in mixtures oftwo normal distributions, *Technometrics*, **9** , 1967, 15-28.

[13] N. E. DAY, Estimating the Components of a Mixture of two Normal Distributions, *Biometrika*, **56**, 1969, 463-474.

[14] 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*, **39** (1):1, November 1977.

[15] K-A. Do, G.J. McLachlan, R . Bean and S. Wen, Application of Gene Shaving and Mixture Models to Cluster Microarray Gene Expression Data, *Cancer Informatics* , **2**, 2007, 25-43.

[16] B. S. Everit and D. J. HAND, *Finite Mixture Distributions*, London: Chapman and Hall, 1981.

[17] Everit, B. S., Maximum Likelihood Estimation of the Parameters in a Mixture of two Univariate Normal Distributions: A Comparison of Different Algorithms, *Statistican*, **33**, 1984, 205-215.

[18] T. Feo and M. G. C. Resende, Greedy Randomized Adaptive Search Procedures. *Journal of Global Optimization*, **2**, 1995, 1-27.

[19] S. Finch, N. Mendell, H. Thode, Probabilistic measures of adequacy of a numerical search for a global maximum. *J. Amer. Statist. Assoc.*, **84**, 1989, 1020-1023.

[20] R. A. Fisher, The Case of Zero Survivors, (Appendix to Bliss, C.I. (1935)), *Annals of Applied Biology*, **22**, 1935, 164-165.

[21] E. B. Fowlkes,Some Methods for Studying Mixtures of two Normal (Lognormal) Distributions, *Journal of the American Statistical Association*, **74** , 1979, 561-575.

[22] I. G. Fryer and C. A. Robertson, A Comparison of Some Methods for Estimating Mixed Normal Distributions, *Biometrika* , **59**, 1972, 639-648.

[23] F. Glover, Heuristics for Integer Programming Using Surrogate Constraints, *Decision Sciences* , **8**, 1977, 156-166.

[24] F. Glover, Tabu Search: A tutorial, *Interfaces* , **20** , 1990, 74-94.

[25] F. Glover, *Multi-Start and strategic oscillation methods Principles to exploit adaptive memory Computing tools for modeling optimization and simulation*, Eds Laguna and Gonzalez-Velarde, Kluwer, 2000, 1-25.

[26] L. A. Goodman, Exploratory Latent Structure Analysis Using Both Identifiable and Unidentifiable Models, *Biometrika* , **61** ,1974, 215-231.

[27] P.Hansen and N. Mladenović, An Introduction to Variable Neighborhood Search. In: S. Voss et al. (eds.), *Metaheuristics, Advances and Trends in Local Search Paradigms for Optimization.* Dordrecht: Kluwer, 1998, 433-458.

[28] P.Hansen and N. Mladenovič, Variable neighborhood search: principles and applications. *European Journal of Operational Research*, **130**, 2001, 449-67.

[29] P.Hansen and N. Mladenovič, *Tutorial on variable neighborhood search.* Technical Report G-2003-46, Les Cahiers du GERAD, 2003.

[30] P.Hansen , N. Mladenov ič and J. Moreno Pérez, Variable neighborhood search. *European Journal of Operational Research,* **191**, 2008, 593-595.

[31] I. P. Harding, The Use of Probability Paper for the Graphical Analysis of Polymodel Frequency Distributions, *Journal of the Marine Biological Association (UK)*, **28**, 1948, 141-153.

[32] H. Hartley, Maximum likelihood estimates from incomplete data. *Biometrics*, **14**, 1958, 174-94.

[33] V. Hasselblad, Estimation of Parameters for a Mixture of Normal Distributions, *Technometrics*, **8**, 1966, 431-444.

[34] V. Hasselblad, Estimation of Finite Mixtures of Distributions from the Exponential Family, *Journal of the American Statistical Association*, **64**, 1969, 1459-1471.

[35] W. S. Jank, Ascent EM for fast and global model-based clustering: An

application to curve- clustering of online auctions, *Computational Statistics and Data Analysis*, **51**, 2006, 747-761.

[36] D. Karlis, E. Xekalaki, Choosing initial values for the EM algorithm for finite mixtures. *Comput. Statist. Data Anal.*, **41**, 2003, 577.

[37] S. Kirkpatrick, C.D. Gelatt and M. P. Vecchi, Optimization by Simulated Annealing. *Science* , **220**, 1983, 671-680.

[38] P. F. Lazarsfeld and N. W. Henry, , *Latent Structure Analysis*, Boston, Houghton-Mifflin , 1968.

[39] R.J.A. Little et D.B. Rubin, *Statistical Analysis with Missing Data*, second edition, New York: John Wiley and Sons, Inc. 2002,

[40] D.W. Hosmer,Jr. On MLE of the parameters of a mixture of two normal distributions when the sample size is small. *Communications in Statistics -Theory and Methods*, **1**, 1973, 217-227.

[41] A. G. McKendrick, Applications of Mathematics to Medical Problems, *Proceedings of the Erlinburgh Mathematical Society*, **44**, 1926, 98-130.

[42] G.J. McLachlan and T. Krishnam,*The EM algorithm and Extensions*. Wiley, New York, 1997.

[43] G.J. McLachlan, On the choice of initial values for the EM algorithm in)tting mixture models. *The Statistician*, **37**, 1988, 417-425.

[44] G. J. McLachlan and K. E. Basford, *Mixture Models: Inference and Applications to Clustering* . Marcel Dekker, New York, 1988.

[45] G. McLachlan and D. Peel, *Finite Mixture Models*, New-York: Wiley, 2000.

[46] G. J. McLachlan, Shu-Kay Ng and Richard Bean, Robust Cluster Analysis via Mixture Models. *AUSTRIAN JOURNAL OF STATISTICS*, **35**, Number 2 and 3, 2006, 157-3174.

[47] W. Mendenhaalnld , R. J. Hadere, Estimation of parameters of mixed exponentially distributed failure time distributions from censored life test data, *Biometrika*, **45**, 1958, 504-520.

[48] X. L. Meng and D. Rubin, Maximum likelihood estimation via the ECM algorithm:a general framework. *Biometrika*, **80**, 1993, 267-278.

[49] X. L. Meng and D. van Dyk, The EM algorithm - an old folk song sung to a fast new tune (with discussion), *Journal of the Royal Statistical Society B*, **59**, 1997, 511-567.

[50] X. L. Meng and S. Pedlow , Em: A bibliographic review with missing articles. *Proc. Statist. Comput.Sect. Am. Statist. Ass.*, **39**, 1992, 24-27.

[51] X. Meng, Missing data: dial M for ???, *Journal of the American Statistical Association.*, **95**, 1325-1330.

[52] N. Mladenoviċ, P. Hansen, 1997. Variable neighborhood search. *Computers and Operations Research*, **24**, 2000, 1097-1100.

[53] S. Newcomb, A Generalized Theory of the Combination of Observations So As To Obtain the Best Result, *American Journal of Mathematics*, **8**, 1886, 343-366.

[54] K. Pearson, Contributions to the Mathematical Theory of Evolution, *Philosophical Transactions, A,* **185**, 1894, 71-110.

[55] D. Peel and G. J. McLachlan, Robust mixture modeling using the t distribution. *Statistical Computing*, **10**, 2000, 335-344.

[56] C.R. RAO, The utilization of multiple measurements in problems of biological classification. *J. Royal Statist. Soc. Ser. B*, **10**, 1948, 159-193.

[57] R. A. Redner and H. F. Walker, Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review*, **26**, 1984, 195-239.

[58] A.H.G. Rinnooy Kan and G.T. Timmer, *Global Optimization Handbooks in operations research and management science,* Vol. 1, Ed. Rinnoy Kan and Todds, North Holland, 1989, 631-662.

[59] S. M. Stigler, Studies in the History of Probability and Statistics. XXXII Laplace, Fisher, and the discover of the concept of sufficiency, *Biometrika*, **60**, 1973, 439-445.

[60] W.Y. Tanand and W.C. Chang, Convolution approach to genetic analysis of quantitative characters of self-fertilized population *Biometrics*, **28** , 1972, 1073-1090.

[61] E. A. C. Thomas, Mathematical Models for the Clustered Firing of Single Cortical Neurons, *British Journal of Mathematical and Statistical Psychology,* **19**, 1966, 151-162.

[62] D. Titterington, A, Smith, U. Makov, *Statistical Analysis of Finite Distributions*, New York, NY, John Wiley, 1985.

[63] D. M. Titerington, Some Recent Research in the Analysis of Mixture Distributions, *Statistics*, **4** , 1990, 619-641.

[64] E. A. C. Thomas, Mathematical Models for the Clustered Firing of Single Cortical Neurons, *British Journal of Mathematical and Statistical Psychology*, **19**, 1966, 151-162.

[65] H. Teicher, Identifiability of Mixtures, *Annals of Mathematical Statistics*, **31**, 1961, 55-73.

[66] N. Ueda and R. Nakano, Deterministic annealing EM algorithm. *Neural Networks*, **11**, 1998, 271-282.

[67] N. Ueda, R. Nakano, Z. Ghahramani and G. E. Hinton, SMEM algorithm for mixture models, *Neural Computation*, **12**, 2000, 2109-2128.

[68] N. Vlassis, A. Likas, A greedy EM algorithm for Gaussian mixture learning, *Neural Process. Lett.*, **15**, 2002, 77-87.
[69] J. H. Wolfe, Pattern Clustering by Multivariate Mixture Analysis, *Multivariate Behavioral Research*, **5**, 1970, 329-350.
[70] D.L. Woodru7, T. Reiners, Experiments with, and on, algorithms for maximum likelihood clustering. *Comput. Statist. Data Analysis*, **47**, 2004, 237-253.
[71] K. Wright and W.J. Kennedy, An interval analysis approach to the EM algorithm. *Journal of Computational and Graphical Statistics*, **9**, 2000, 303-318.
[72] C.F.J. Wu, On the convergence properties of the EM algorithm. *Annals of Statistics*, **11**, 1983, 95-103.
[73] L. Xu and M.I. Jordan, On convergence properties of the em algorithm for gaussian mixtures. *Neural Computation*, **8**, 1996, 129151.

[1] *Institut Superieur de Gestion*
Université de GABÈS
Gabès 6072, TUNISIA
E-mail: a_bessadok@yahoo.com

[2] *GERAD and École des Hautes Études Commérciales*
Université de Montréal
3000, chemin de la Cote-sainte Catherine
Montréal, Qu., CANADA H3T 2A7
E-mail: pierre.hansen@gerad.ca

[3] *Department of Quantitative Methods*
Faculté des Sciences Économique et de Géstion Sfax Tunisia
BP 1088, Sfax 3018, TUNISIA
E-mail: abdelwaheb.rebai@fsegsfax.rnu.tn

Appendix A - Treated examples

MODEL	DATA	MEAN	COVARIANCE	MIXING WEIGHT π
2G1MM	REAL	$\mu = [15; 9.5]$	$\sigma = [2.1; 2.1]$	$[0.3; 0.7]$
	INITIAL	$\mu = [10.76; 16.62]$	$\sigma = [1.6; 0.01]$	$[0.98; 0.02]$
3G1MM	REAL	$\mu = [-5; 0; 2]$	$\sigma = [0.5; 0.5; 0.5]$	$[0.5; 0.3; 0.2]$
	INITIAL	$\mu = [0.73; -4.72; -5.86]$	$\sigma = [1.14; 0.23; 0.10]$	$[0.49; 0.44; 0.07]$
2G2MM	REAL	$\mu_1 = [0; 0]$ $\mu_2 = [0; 0]$	$\Sigma_1 = [3, 0; 0, 1/3]$ $\Sigma_2 = [1/3, 0; 0, 3]$	$[0.7, 0.3]$
	INITIAL	$\mu_1 = [0.14; -0.04]$ $\mu_2 = [-1.82; 1.94]$	$\Sigma_1 = [1, 0.3; 0.3, 1]$ $\Sigma_2 = [1, 0.5; 0.5, 1]$	$[0.96, 0.04]$
3G2MM	REAL	$\mu_1 = [0; 1]$ $\mu_2 = [0; -1]$ $\mu_3 = [-1; 2]$	$\Sigma_1 = \Sigma_2 = \Sigma 3$ $= [0.2, 0.1; 0.1, 0.2]$	$[0.3; 0.3; 0.4]$
	INITIAL	$\mu_1 = [-1.53; 1.73]$ $\mu_2 = [-0.08; 1.92]$ $\mu_3 = [0.52; 0.19]$	$\Sigma_1 = \Sigma_2 = \Sigma_3$ $= [1, 0.3; 0.3, 1]$	$[0.46; 0.01; 0.53]$
4G2MM	REAL	$\mu_1 = [0; -2] \mu_2 = [2, 0]$ $\mu_3 = [0; 2] \mu_4 = [-2, 0]$	$\Sigma_1 = \Sigma_3 = [3, 0; 0, 1/3]$ $\Sigma_2 = \sigma_4 = [1/3, 0; 0, 3]$	$[0.25, 0.25, 0.25, 0.25]$
	INITIAL	$\mu_1 = [-0.3; 1.1] \mu_2 = [1.9; 0]$ $\mu_3 = [-1.6; 1.2] \mu_4 = [0.6; 0.1]$	$\Sigma_1 = \Sigma_2 = \Sigma_3$ $= \Sigma_4 = [0.6, 0.9; 0.9, 0.4]$	$[0.41, 0.26, 0.18, 0.16]$
6G2MM	REAL	$\mu_1 = [0.75; -0.5]$ $\mu_2 = [0.5; 1]$ $\mu_3 = [0; 1.5]$ $\mu_4 = [-1; -0.5]$ $\mu_5 = [-1.5; 0]$ $\mu_6 = [1; -1.5]$	$\Sigma_1 = \Sigma_2 = \Sigma_3$ $\Sigma_4 = \Sigma_5 = \Sigma_6$ $= [0.05, 0; 0, 0.2]$	$[1/6; 1/6; 1/6; 1/6; 1/6; 1/6]$
	INITIAL	$\mu_1 = [-0.3; 1.1]$ $\mu_2 = [1.9; 0]$ $\mu_3 = [-1.6; 1.2]$ $\mu_4 = [0.6; 0.1]$ $\mu_5 = [0; 1.5]$ $\mu_6 = [1.1; -0.5]$	$\Sigma_1 = \Sigma_2 = \Sigma_3$ $\Sigma_4 = \Sigma_5 = \Sigma_6$ $= [0.6, 0.9; 0.9, 0.4]$	$[0.11; 0.20; 0.18; 0.16; 0.17; 0.18]$
8G2MM	REAL	$\mu_1 = [1.5; 0]$ $\mu_2 = [1; 1]$ $\mu_3 = [0; 1.5]$ $\mu_4 = [-1; 1]$ $\mu_5 = [-1.5; 0]$ $\mu_6 = [-1; -1]$ $\mu_7 = [0; -1.5]$ $\mu_8 = [1; -1]$	$\Sigma_1 = \Sigma_2 = \Sigma_3$ $\Sigma_4 = \Sigma_5 = \Sigma_6$ $\Sigma_7 = \Sigma_8$ $= [0.01, 0; 0, 0.1]$	$[1/8; 1/8; 1/8; 1/8; 1/8; 1/8; 1/8; 1/8]$
	INITIAL	$\mu_1 = [-0.3; 1.4]$ $\mu_2 = [1.9; -0.2]$ $\mu_3 = [-1.6; 1.2]$ $\mu_4 = [0.6; 0.1]$ $\mu_5 = [0.3; -0.7]$ $\mu_6 = [1.11; -0.5]$ $\mu_7 = [1.6; -0.5]$ $\mu_8 = [1.2; 1.3]$	$\Sigma_1 = \Sigma_2 = \Sigma_3$ $\Sigma_4 = \Sigma_5 = \Sigma_6$ $\Sigma_7 = \Sigma_8$ $= [0.06, 0.95; 0.95, 0.48]$	$[0.01; 0.23; 0.29; 0.12; 0.04; 0.06; 0.17; 0.08]$
10G2MM	REAL	$\mu_1 = [1.25; 0]$ $\mu_2 = [1; 1]$ $\mu_3 = [0; 1.5]$ $\mu_4 = [-1; 1]$ $\mu_5 = [-1.5; 0]$ $\mu_6 = [-1.5; -1]$ $\mu_7 = [0; -1.5]$ $\mu_8 = [1; -1]$ $\mu_9 = [0.5; -1.5] \mu_{10} = [1; -1.5]$	$\Sigma_1 = \Sigma_2 = \Sigma_3$ $\Sigma_4 = \Sigma_5 = \Sigma_6$ $\Sigma_7 = \Sigma_8 = \Sigma_9$ $= \Sigma_{10} = [0.01, 0; 0, 0.1]$	$[0.1; 0.1; 0.1; 0.15; 0.1; 0.15; 0.1; 0.1; 0.05; 0.05]$
	INITIAL	$\mu_1 = [-0.3; 1.4]$ $\mu_2 = [1.9; -0.2]$ $\mu_3 = [-1.6; 1.2]$ $\mu_4 = [0.6; 0.1]$ $\mu_5 = [0.3; -0.7]$ $\mu_6 = [1.11; -0.5]$ $\mu_7 = [1.6; -0.5]$ $\mu_8 = [1.2; 1.3]$ $\mu_9 = [1.1; -0.15] \mu_{10} = [1.5; -1.3]$	$\Sigma_1 = \Sigma_2 = \Sigma_3$ $\Sigma_4 = \Sigma_5 = \Sigma_6$ $\Sigma_7 = \Sigma_8 = \Sigma_9$ $= \Sigma_{10} = [0.06, 0.95; 0.95, 0.48]$	$[0.1; 0.1; 0.1; 0.15; 0.1; 0.15; 0.1; 0.1; 0.05; 0.05]$

Appendix B - Tables results

Using poor starting points

Model	Sample	TRIALS	Global Maxima		Local Maxima RANGE		% Global Maxima	
			MS	EMVNS	MS	EMVNS	MS	EMVNS
2G1MM	100	1	-235.8753	-235.8753	10.1791	8.4856	98	94
		2	-234.4101	-234.4101	13.2288	11.1921	96	95
		3	-239.2080	-239.2080	10.4160	10.4160	97	92
		4	-238.0905	-238.0905	14.4562	15.3911	99	94
		5	-235.1474	-235.1474	16.2545	13.2589	94	96
		6	-236.2357	-236.2357	11.6521	13.2587	91	93
		7	-233.9541	-231.9541	14.6523	14.3251	93	96
		8	-237.0245	-237.0245	16.2124	11.3547	98	94
		9	-238.1544	-238.1544	13.2574	11.2541	92	93
		10	-238.8563	-238.8563	11.6523	12.6523	96	91
3G1MM	150	1	-290.0475	**-286.6524**	1.9924	3.3951	17	17
		2	-307.6195	-307.6195	4.0851	3.4112	47	79
		3	-296.8935	**-296.8481**	2.0226	2.0680	19	70
		4	-318.4914	**-316.7653**	1.8197	3.5458	65	79
		5	-299.0124	-299.0124	12.2900	11.4197	42	88
		6	-306.7865	**-304.4735**	2.1404	5.0500	72	7
		7	-292.1598	-292.1598	5.6254	5.2654	76	75
		8	-303.1527	**-303.0161**	7.0874	7.1822	6	68
		9	-312.3613	-312.3613	3.1955	3.1675	74	91
		10	-301.1689	-301.1689	3.1113	3.1113	65	74
2G2MM	200	1	-641.2487	-641.2487	46.5814	22.3507	36	76
		2	-632.5786	-632.5786	45.3939	3.4112	27	79
		3	-600.5044	-600.5044	62.6585	48.2360	40	93
		4	-620.1937	-620.1937	36.1978	34.3310	31	82
		5	-627.4071	-627.4071	53.7468	42.9049	17	18
		6	-649.4448	-649.4448	57.4785	56.8515	29	86
		7	-611.8998	-611.8998	47.3719	32.4887	37	96
		8	-614.0520	-614.0520	24.3866	17.9981	3	80
		9	-601.6633	-601.6633	36.8120	34.5052	34	96
		10	-639.6007	-639.6007	51.6233	32.7748	12	73
3G2MM	200	1	-652.5217	**-648.1404**	47.2736	14.9758	6	83
		2	-646.0257	**-622.5508**	57.1086	23.4749	2	49
		3	-605.1128	**-584.6639**	65.7443	20.4489	35	33
		4	-655.8673	-655.8673	53.0580	3.7038	17	87
		5	**-610.5412**	-612.5774	71.2584	12.5263	12	20
		6	-630.8122	-630.8122	54.6093	6.9880	16	81
		7	-614.8514	-614.8514	61.9125	21.7185	14	68
		8	-643.9878	**-639.1087**	64.8320	6.9985	1	64
		9	-653.0959	**-652.4204**	67.4756	2.8292	1	34
		10	-646.1953	-646.1953	66.0131	4.2911	29	93

Model	Sample	TRIALS	Global Maxima		Local Maxima RANGE		% Global Maxima	
			MS	EMVNS	MS	EMVNS	MS	EMVNS
4G2MM	200	1	-770.1656	-770.1656	79.2570	57.8955	26	90
		2	-749.1386	-749.1386	89.8699	78.8091	24	74
		3	-741.2514	-741.2514	74.5417	65.2155	21	71
		4	-760.7066	-760.7066	68.8641	61.2430	32	55
		5	-776.4492	-776.4492	81.8411	59.6209	40	79
		6	-744.2546	-744.2546	80.3412	59.7544	23	81
		7	-751.1382	-751.1382	76.6545	66.2541	24	80
		8	-748.2112	-748.2112	81.6995	72.4121	23	73
		9	-752.2243	-752.2243	82.2546	77.5546	25	78
		10	-761.9663	-761.9663	72.8910	66.2096	37	66
6G2MM	500	1	-1.0041e+03	**-999.7815**	421.4680	349.8224	1	1
		2	-958.0554	-958.0554	450.9760	379.5808	2	64
		3	-989.4930	**-987.6576**	423.2859	357.9227	1	20
		4	-1.0137e+03	-1.0137e+03	383.7455	4328.2597	1	42
		5	-953.4334	**-944.4864**	468.9808	407.8856	1	54
		6	-989.7588	**-981.1017**	428.3935	354.6687	1	23
		7	-985.7730	**-969.1041**	439.6149	370.9343	1	5
		8	-1.0255e+03	-1.0255e+03	429.1275	330.0983	1	64
		9	-1.0173e+03	**-1.0083e+03**	394.4401	350.0143	1	77
		10	**-1.0300e+03**	-1.0314e+03	398.5337	315.0695	1	45
8G2MM	700	1	-1.0042e+03	-1.0042e+03	1.0154e+03	967.0880	8	55
		2	-1.0437e+03	-1.0437e+03	938.1917	910.6738	4	53
		3	-973.2614	-973.2614	1.0736e+03	974.4371	8	76
		4	-961.2759	-961.2759	1.0453e+03	989.5428	6	52
		5	-1.0055e+03	-1.0055e+03	1.0059e+03	967.4489	5	56
		6	-978.1461	-978.1461	1.0165e+03	982.0156	12	41
		7	**-973.6314**	-1.2444e+03	1.0098e+03	587.4345	12	27
		8	-1.0117e+03	-1.0117e+03	1.0023e+03	954.3250	8	18
		9	**-1.0513e+03**	-1.1150e+03	988.7539	917.9848	8	54
		10	-974.6313	-974.6313	1.0133e+03	923.3240	10	85
10G2MM	800	1	-1.1078e+03	**-1.1022e+03**	1.2953e+03	1.0856e+03	1	66
		2	-1.1656e+03	**-1.1623e+03**	1.2360e+03	1.0111e+03	1	10
		3	-1.1161e+03	**-1.1109e+03**	1.2805e+03	1.0489e+03	1	72
		4	-1.1512e+03	**-1.1392e+03**	1.2298e+03	1.0308e+03	1	29
		5	-1.1552e+03	-1.1552e+03	1.2319e+03	954.9397	1	42
		6	**-1.1330e+03**	-1.1681e+03	1.1882e+03	1.0101e+03	1	16
		7	-1.1360e+03	**-1.1341e+03**	1.2245e+03	1.0442e+03	1	55
		8	**-1.1137e+03**	-1.1850e+03	1.2506e+03	1.0221e+03	1	1
		9	-1.1642e+03	**-1.1560e+03**	1.2532e+03	1.0519e+03	1	48
		10	-1.1731e+03	**-1.1684e+03**	1.2141e+03	1.0132e+03	1	37

Model	Sample	TRIALS	Global Maxima		Local Maxima RANGE		% Global Maxima	
			MS	EMVNS	MS	EMVNS	MS	EMVNS
2G1MM	100	1	-234.8521	-234.8521	11.1269	9.5632	94	91
		2	-236.5411	-236.5411	16.2541	3.1695	95	98
		3	-239.6056	-239.6056	13.8891	1.9895e-13	96	100
		4	-236.2693	-236.2693	13.6254	7.3254	95	97
		5	-228.3542	-228.3542	4.5214e-13	2.9562e-13	100	100
		6	-237.1642	-237.1642	14.2671	10.1547	91	93
		7	-226.8521	-226.8521	12.0652	4.0516e-13	92	100
		8	-235.1245	-235.1245	16.2547	15.2541	96	92
		9	-234.0251	-234.0251	12.3658	10.3596	91	94
		10	-220.8229	-220.8229	17.3641	3.9790e-13	99	100
3G1MM	150	1	-290.1672	-290.1672	55.7924	0.2672	35	96
		2	-298.0510	**-296.8191**	2.7856	1.2320	74	25
		3	-288.9858	**-288.9070**	2.0077	1.5669	8	71
		4	-293.7079	-293.7079	1.0692	1.0692	64	99
		5	-281.0147	-281.0147	4.7916	1.3642e-12	72	100
		6	-289.9147	**-289.5405**	2.9777	2.0934	77	60
		7	-276.1808	-276.1808	12.4557	11.2750	70	98
		8	-307.4659	**-306.6627**	2.4843	3.2876	70	84
		9	-296.5679	**-295.2190**	4.7682	1.3489	81	71
		10	-304.4261	-304.4261	6.9570	6.2528e-13	74	100
2G2MM	200	1	-643.3488	-643.3488	48.9028	1.2506e-12	38	100
		2	-618.4387	-618.4387	39.2941	18.7931	16	98
		3	-629.2587	-629.2587	62.3592	44.0413	23	98
		4	-627.5673	-627.5673	37.3981	3.4106e-13	30	100
		5	-581.3516	-581.3516	12.2900	1.1369e-12	40	98
		6	-631.5586	-631.5586	43.3517	3.1243	29	81
		7	-601.5012	-601.5012	57.6534	39.2371	41	92
		8	-625.3583	-625.3583	41.2587	3.8471e-13	31	100
		9	-585.2847	-585.2847	22.12544	1.21459e-12	40	100
		5	-635.7149	-635.7149	38.8518	20.5429	1	61
3G2MM	200	1	-648.9412	**-649.4180**	76.1761	2.7151	1	15
		2	-659.7583	**-658.1108**	53.9242	6.8212e-13	2	100
		3	-635.7000	**-635.3556**	51.5436	43.8824	17	15
		4	-653.9328	**-652.8712**	59.6598	3.8281	8	11
		5	-620.8220	**-619.0844**	55.3170	1.7377	22	97
		6	-643.2921	-643.2921	66.4684	55.2612	2	72
		7	-642.6907	**-641.2610**	63.3844	1.4297	1	58
		8	-613.5389	-613.5389	51.6619	47.2346	1	33
		9	-619.8786	**-617.6769**	65.2008	56.4982	31	30
		10	**-611.3266**	-624.4769	77.3488	2.7285e-12	1	100

Model	Sample	TRIALS	Global Maxima		Local Maxima RANGE		% Global Maxima	
			MS	EMVNS	MS	EMVNS	MS	EMVNS
4G2MM	200	1	-752.1700	-752.1700	82.1010	15.1649	35	99
		2	-797.6400	-797.6400	67.9817	12.9877	4	79
		3	-742.1413	-742.1413	79.3266	68.4991	30	64
		4	-760.7066	-760.7066	68.8641	61.2430	32	55
		5	-758.0660	-758.0660	66.2370	34.3501	16	55
		6	-751.1227	-751.1227	87.0018	71.5188	24	99
		7	-772.2980	-772.2980	64.9957	0.4575	20	99
		8	-756.3823	-756.3823	79.1321	72.6345	14	57
		9	-740.3214	-740.3214	84.8219	58.9910	4	39
		10	-755.9385	-755.9385	87.8174	6.7298	21	92
6G2MM	500	1	-1.0039e+03	**-1.0002e+03**	417.7638	8.8476	1	8
		2	**-1.0264e+03**	-1.0278e+03	392.8282	247.8544	4	25
		3	-1.0264e+03	**-1.0056e+03**	392.8282	24.5343	4	93
		4	**-1.0072e+03**	-1.0101e+03	431.1072	333.6097	4	57
		5	-1.0072e+03	**-1.0068e+03**	431.1072	5.8514	1	5
		6	-1.0172e+03	**-992.2265**	378.7475	30.4581	1	42
		7	-1.0277e+03	**-1.0223e+03**	387.7640	273.8456	1	43
		8	-1.0183e+03	**-1.0179e+03**	393.6705	14.5584	3	8
		9	-1.0159e+03	-1.0159e+03	400.5788	19.9525	1	67
		10	-999.3922	**-977.3866**	450.8094	26.9714	1	87
8G2MM	700	1	-984.4266	-984.4266	1.0154e+03	76.1694	8	35
		2	-985.7864	-985.7864	1.0218e+03	913.5765	12	87
		3	-983.4621	-983.4621	1.0136e+03	53.6341	4	84
		4	-968.0664	-968.0664	1.0133e+03	967.1979	7	62
		5	**-993.8567**	-1.0562e+03	995.0358	906.2064	10	82
		6	-1.0237e+03	-1.0237e+03	996.4803	77.6997	7	99
		7	-1.0026e+03	-1.0026e+03	981.5858	81.8980	11	92
		8	-998.8686	-998.8686	1.0001e+03	66.4924	7	16
		9	-988.1429	-988.1429	975.6521	53.8557	8	99
		10	-1.0147e+03	-1.0147e+03	979.8081e+03	931.5988	10	64
10G2MM	800	1	-1.1385e+03	**-1.1366e+03**	1.2299e+03	22.2510	1	38
		2	-1.1551e+03	**-1.1478e+03**	1.2290e+03	38.0575	1	93
		3	-1.1093e+03	**-1.1068e+03**	1.2574e+03	1.1802e+03	1	44
		4	**-1.1694e+03**	-1.1732e+03	1.2309e+03	1.0507e+03	1	78
		5	-1.1483e+03	**-1.1472e+03**	1.2040e+03	33.5633	1	12
		6	-1.1082e+03	**-1.1030e+03**	1.3031e+03	10.8805	1	59
		7	-1.1360e+03	**-1.1341e+03**	1.2245e+03	1.0442e+03	1	55
		8	-1.1254e+03	**-1.1154e+03**	1.2621e+03	521.2814	1	12
		9	**-1.1621e+03**	-1.1698e+03	1.2486e+03	38.1673	1	10
		10	-1.5471e+03	**-1.5455e+03**	717.9538	684.8795	1	31

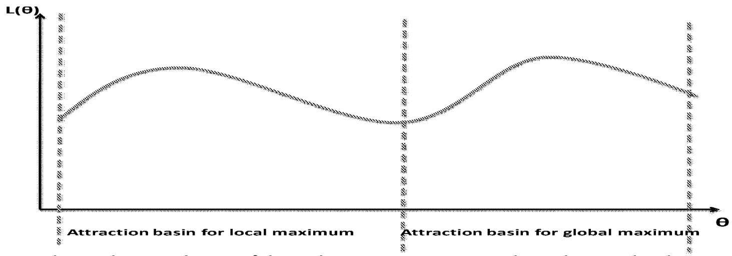
Using random starting points

Model	Sample	TRIALS	Global Maxima		Local Maxima RANGE		% Global Maxima	
			MS	EMVNS	MS	EMVNS	MS	EMVNS
2G1MM	100	1	-235.8753	-235.8753	10.1791	8.4856	98	94
		2	-236.5277	-236.5277	11.5142	9.6211	91	96
		3	-238.2114	-238.2114	12.8411	11.6244	94	96
		4	-234.1433	-234.1433	11.3266	10.6521	94	98
		5	-238.2510	-238.2510	13.5413	11.3210	91	93
		6	-235.0125	-235.0125	16.3214	10.1048	98	94
		7	-236.2301	-236.2301	14.0477	13.5109	91	96
		8	-237.4523	-237.4523	11.2036	9.3254	92	96
		9	-235.3580	-235.3580	13.0911	12.9154	92	91
		10	-233.2866	-233.2866	16.8205	10.2144	90	92
3G1MM	150	1	-288.5110	-288.5100	3.2262	1.4254	48	99
		2	-282.1450	**-278.7498**	5.1305	5.8677	4	47
		3	-294.5143	-294.5143	3.2514	2.6347	18	74
		4	-290.1672	-290.1672	4.7367	3.2314	16	81
		5	-295.4229	**-290.4305**	2.2856	6.3440	32	44
		6	-298.0510	**-296.8191**	2.7856	4.01764	70	1
		7	-275.8961	-275.8961	4.3314	3.8048	1	76
		8	-290.9747	**-289.2354**	5.3697	5.3159	58	83
		9	-317.6993	**-317.5027**	5.8401	5.3555	80	53
		10	-293.7200	**292.3509**	2.7478	3.8355	13	43
2G2MM	200	1	-641.2487	-641.2487	46.5814	22.3507	36	76
		2	-639.6214	-639.6214	41.2147	28.6211	24	75
		3	-614.6207	-614.6207	51.6503	46.2217	32	76
		4	-649.0452	-649.0452	41.9605	32.6072	38	90
		5	-601.2019	-601.2019	48.6263	42.2851	24	28
		6	-639.3088	-639.3088	48.3644	30.6249	32	92
		7	-620.3321	-620.3321	51.3017	38.6328	28	84
		8	-628.3622	-628.3622	50.9423	31.3301	38	94
		9	-638.0752	-638.0752	42.3285	31.0866	28	86
		10	-608.6429	-608.6429	45.6275	40.9004	32	76
3G2MM	200	1	-649.2168	-649.2168	59.0776	50.7962	2	73
		2	-623.7199	**-617.6543**	58.4495	62.3994	26	76
		3	-651.1126	**-643.0419**	54.814	46.1338	1	76
		4	-633.3075	-633.3075	59.0551	50.5428	13	81
		5	-642.5767	-642.5767	58.1452	50.2351	13	43
		6	-630.5347	**-625.4970**	54.6974	48.3538	3	54
		7	-640.8798	**-625.6880**	59.4298	71.9445	1	50
		8	-638.4954	-638.4954	53.5249	46.9954	14	61
		9	**-620.5147**	-623.3256	61.2547	51.2156	13	36
		10	-622.9948	**-607.1670**	62.7560	67.7480	3	67

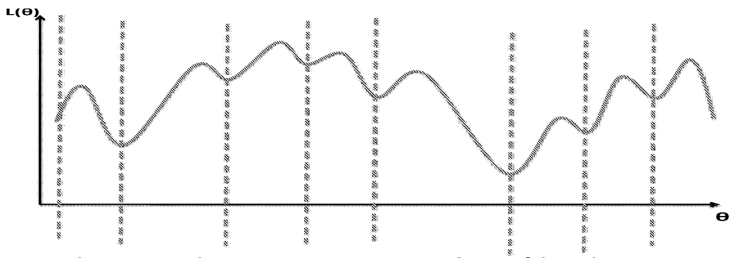
Model	Sample	TRIALS	Global Maxima		Local Maxima RANGE		% Global Maxima	
			MS	EMVNS	MS	EMVNS	MS	EMVNS
4G2MM	200	1	-743.1162	-743.1162	75.1187	73.6494	39	94
		2	-749.2199	-749.2199	86.3220	71.0821	28	94
		3	-756.3678	-756.3678	79.3128	65.2155	29	71
		4	-760.3703	-760.3703	91.1277	78.7588	38	92
		5	-763.3643	-763.3643	62.8596	71.2364	16	76
		6	-741.3652	-741.3652	79.6966	70.9122	24	75
		7	-752.9941	-752.9941	81.2247	71.3221	31	80
		8	-752.2411	-752.2411	74.3248	70.6912	30	76
		9	-748.2610	-748.2610	65.2544	69.6523	34	81
		10	-758.6311	-758.6311	80.9817	71.9122	31	91
6G2MM	500	1	**-1.0277e+03**	-1.0331e+03	366.3842	349.8224	1	40
		2	-977.9569	**-944.0497**	405.4793	367.8062	3	24
		3	-1.0408e+03	-1.0408e+03	399.9315	334.7967	1	21
		4	-975.4794	**-974.8529**	429.9173	306.8197	1	79
		5	-998.3713	**-995.8678**	423.5845	243.0707	1	41
		6	-989.1698	**-987.4594**	417.2788	232.7691	1	34
		7	-995.4418	**-990.2266**	434.6577	318.9361	1	21
		8	-1.0046e+03	-1.0046e+03	446.2635	376.4621	1	58
		9	**-1.0188e+03**	-1.0189e+03	429.0988	328.1266	1	83
		10	-1.0463e+03	**-1.0437e+03**	401.8871	267.3516	1	9
8G2MM	700	1	958.6579	958.6579	1.0557e+03	1.0242e+03	2	17
		2	-1.0081e+03	-1.0081e+03	9967.5115	905.1947	10	7
		3	-1.0472e+03	-1.0472e+03	954.5972	936.1669	5	28
		4	-1.0077e+03	-1.0077e+03	972.6653	901.8672	6	21
		5	-1.0040e+03	-1.0040e+03	986.8095	948.3994	11	72
		6	-1.0090e+03	-1.0090e+03	983.9204	953.1839	5	88
		7	-1.0084e+03	-1.0084e+03	994.5039	858.3365	6	27
		8	**-984.7680**	-1.0491e+03	1.06044	908.0154	7	45
		9	**-1.0171e+03**	-1.0941e+03	990.3082	834.4988	7	30
		10	-978.5214	-978.5214	985.9514	854.2145	11	64
10G2MM	800	1	-1.1093e+03	**-1.1068e+03**	1.2574e+03	1.1802e+03	1	44
		2	-1.1698e+03	**-1.1582e+03**	1.2375e+03	1.1437e+03	1	18
		3	-1.1093e+03	**-1.1068e+03**	1.2805e+03	1.1802e+03	1	44
		4	-1.1848e+03	**-1.1797e+03**	1.2298e+03	1.0362e+03	1	87
		5	-1.1469e+03	**-1.1447e+03**	1.2061e+03	1.0662e+03	1	1
		6	**-1.1682e+03**	-1.1826e+03	1.2315e+03	1.1525e+03	1	4
		7	**-1.1405e+03**	-1.1741e+03	1.2375e+03	1.1044e+03	1	18
		8	**-1.1202e+03**	-1.1226e+03	1.2888e+03	1.1900e+03	1	27
		9	-1.1308e+03	**-1.1303e+03**	1.2363e+03	1.1382e+03	1	26
		10	-1.0882e+03	**-1.0879e+03**	1.3139e+03	1.0151e+03	1	48

Model	Sample	TRIALS	Global Maxima		Local Maxima RANGE		% Global Maxima	
			MS	EMVNS	MS	EMVNS	MS	EMVNS
2G1MM	100	1	-235.8753	-235.8753	10.1791	1.6422e-13	98	100
		2	-236.5277	-236.5277	11.5142	8.2141	91	95
		3	-238.2114	-238.2114	12.8411	10.3249	94	91
		4	-234.1433	-234.1433	11.3266	1.5211e-13	94	100
		5	-238.2510	-238.2510	13.5413	2.3521	91	98
		6	-235.0125	-235.0125	16.3214	10.5211	98	96
		7	-236.2301	-236.2301	14.0477	1.4136e-13	91	100
		8	-237.4523	-237.4523	11.2036	8.2183	92	94
		9	-235.3580	-235.3580	13.0911	6.6271	92	96
		10	-233.2866	-233.2866	16.8205	11.9521	90	94
3G1MM	150	1	-288.5110	**-288.1290**	3.2262	1.8074	48	41
		2	-282.1450	**-279.2184**	5.1305	3.0781	4	18
		3	-294.5143	-294.5143	3.2514	1.6270	18	86
		4	-290.1672	-290.1672	4.7367	0.2672	16	88
		5	-295.4229	**-293.0061**	2.2856	2.4169	32	84
		6	-298.0510	**-296.8191**	2.7856	1.4551	70	35
		7	-275.8961	-275.8961	4.3314	0.2888	1	99
		8	-290.9747	**-289.2354**	5.3697	1.7393	58	52
		9	-317.6993	-317.6993	5.8401	5.1589	80	87
		10	-293.7200	**-291.5225**	2.7478	3.4538	13	71
2G2MM	200	1	-641.2487	-641.2487	46.5814	8.5470	36	84
		2	-639.6214	-639.6214	41.2147	10.9624	24	90
		3	-614.6207	-614.6207	51.6503	6.4063	32	86
		4	-649.0452	-649.0452	41.9605	44.9120	38	74
		5	-601.2019	-601.2019	48.6263	13.5209	24	92
		6	-639.3088	-639.3088	48.3644	36.0411	32	82
		7	-620.3321	-620.3321	51.3017	8.6157	28	92
		8	-628.3622	-628.3622	50.9423	40.6184	38	76
		9	-638.0752	-638.0752	42.3285	28.5713	28	74
		10	-608.6429	-608.6429	45.6275	36.2843	32	84
3G2MM	200	1	-649.2168	-649.2168	59.0776	5.7646	2	92
		2	-623.7199	-623.7199	58.4495	48.0424	26	16
		3	-651.1126	**-650.5690**	54.814	16.0921	1	47
		4	-633.3075	-633.3075	59.0551	32.7674	13	63
		5	-642.5767	-642.5767	58.1452	24.5217	13	68
		6	-630.5347	**-625.4970**	54.6974	8.2842	3	62
		7	-640.8798	**-634.9935**	59.4298	9.8502	1	97
		8	-638.4954	-638.4954	53.5249	37.5486	14	53
		9	**-620.5147**	-621.5142	61.2547	48.2547	13	41
		10	622.9948	**-621.5530**	62.7560	2.7384	3	33

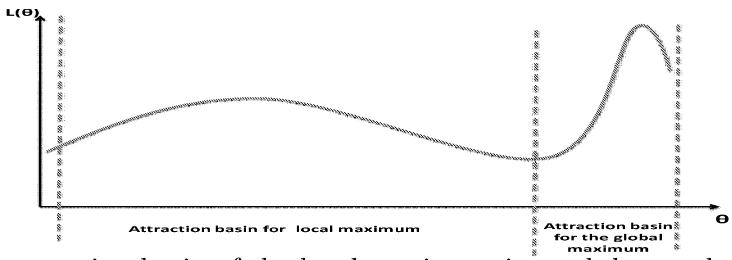
Model	Sample	TRIALS	Global Maxima		Local Maxima RANGE		% Global Maxima	
			MS	EMVNS	MS	EMVNS	MS	EMVNS
4G2MM	200	1	-743.1162	-743.1162	75.1187	50.8358	39	92
		2	-749.2199	-749.2199	86.3220	64.6878	28	92
		3	-756.3678	-756.3678	79.3128	68.4991	29	64
		4	-760.3703	-760.3703	91.1277	55.7623	38	81
		5	-763.3643	-763.3643	62.8596	58.2318	16	91
		6	-741.3652	-741.3652	79.6966	20.2514	24	91
		7	-752.9941	-752.9941	81.2247	34.6621	31	82
		8	-752.2411	-752.2411	74.3248	56.7752	30	74
		9	-748.2610	-748.2610	65.2544	12.5411	34	93
		10	-758.6311	-758.6311	80.9817	70.2286	31	87
6G2MM	500	1	-1.0277e+03	-1.0277e+03	403.7125	4.5753	1	98
		2	**-977.9569**	-981.0494	405.4793	4.3618	3	86
		3	-1.0408e+03	**-1.0400e+03**	399.9315	126.5601	1	5
		4	-975.4794	**-970.2479**	429.9173	7.1039	1	12
		5	-998.3713	**-995.6342**	423.5845	243.0707	1	41
		6	-989.1698	**-986.5160**	417.2788	12.1357	1	68
		7	-995.4418	**-994.0682**	434.6577	191.5254	1	31
		8	-1.0046e+03	**-995.0639**	446.2635	416.3952	1	31
		9	-1.0188e+03	**-1.0178e+03**	429.0988	13.4971	1	48
		10	-1.0463e+03	**-1.0451e+03**	401.8871	4.0569	1	85
8G2MM	700	1	-958.6579	-958.6579	1.0557e+03	75.3492	2	54
		2	-1.0081e+03	-1.0081e+03	967.5115	73.8963	10	86
		3	-1.0472e+03	-1.0472e+03	954.5972	125.2951	5	89
		4	-1.0077e+03	-1.0077e+03	972.6653	147.8416	6	95
		5	-1.0040e+03	-1.0040e+03	986.8095	201.4950	11	49
		6	-1.0090e+03	-1.0090e+03	983.9204	881.9820	5	75
		7	**-1.0084e+03**	-1.0732e+03	994.5039	10.3180	6	63
		8	-984.7680	-984.7680	948.6578	76.3896	7	61
		9	-1.0171e+03	-1.0171e+03	990.3082	744.9327	7	48
		10	-978.5214	-978.5214	985.9514	601.1251	11	64
10G2MM	800	1	-1.1093e+03	**-1.1065e+03**	1.2574e+03	11.4995	1	24
		2	-1.1698e+03	**-1.1685e+03**	1.2375e+03	6.0270	1	43
		3	-1.1093e+03	**-1.1076e+03**	1.2574e+03	166.1510	1	19
		4	-1.1848e+03	**-1.1711e+03**	1.2309e+03	18.0059	1	84
		5	-1.1469e+03	**-1.1466e+03**	1.2061e+03	54.1790	1	8
		6	**-1.1682e+03**	-1.1784e+03	1.2315e+03	971.8920	1	16
		7	-1.1405e+03	**-1.1329e+03**	1.2375e+03	31.4003	1	15
		8	**-1.1202e+03**	-1.1256e+03	1.2888e+03	30.4336	1	22
		9	-1.1308e+03	**-1.1074e+03**	1.2363e+03	33.1973	1	28
		10	**-1.0882e+03**	-1.0968e+03	1.3139e+03	5.9767	1	33



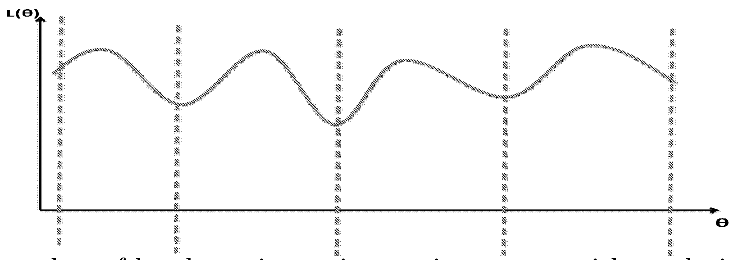
(a) A very reduced number of local maximum with relatively large attraction basin



(b) A general case with an important number of local maximum with a relatively small attraction basin

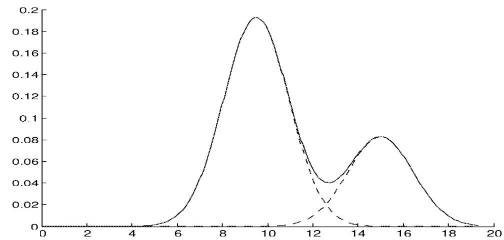


(c) The attraction basin of the local maximum is much larger than one of global maximum

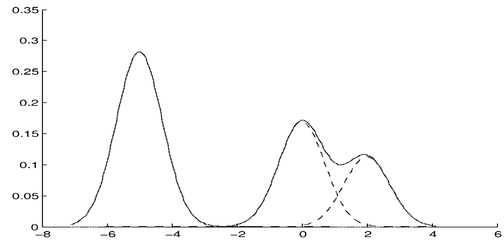


(d) The number of local maximum is very important with a relatively same large attraction basin

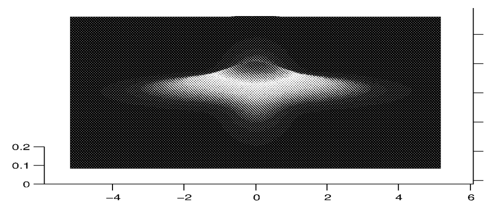
Figure 3: The complexity of FGMM examples dependent in both the number of local maximum and their attraction basin size.



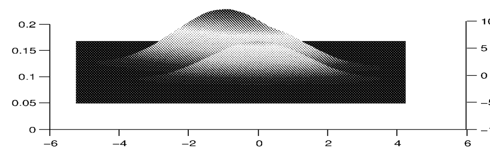
(a) A very simple FGMM



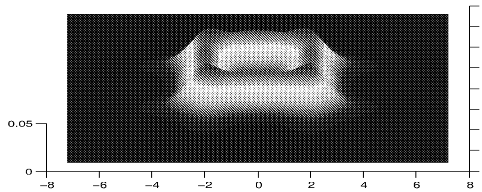
(b) FGMM in one dimension space



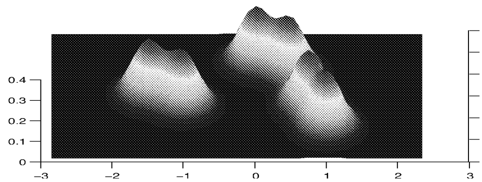
(c) The two model components are poorly separated



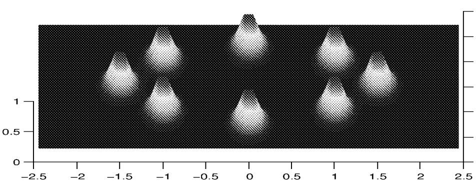
(d) The model components are relatively poorly separated



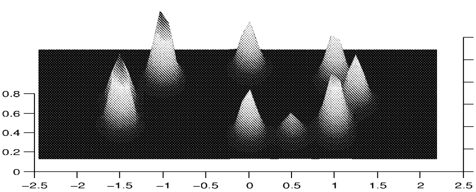
(e) The four model components are well separated



(f) The model components are quite poorly separated

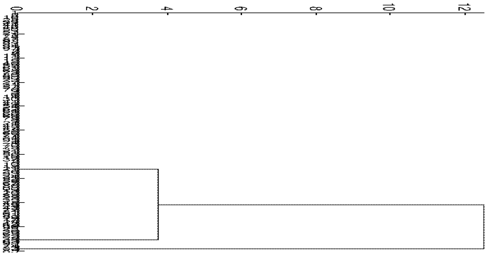


(g) The model components are very clearly separated

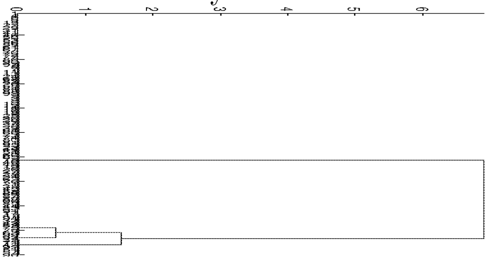


(h) The model components are quite separated

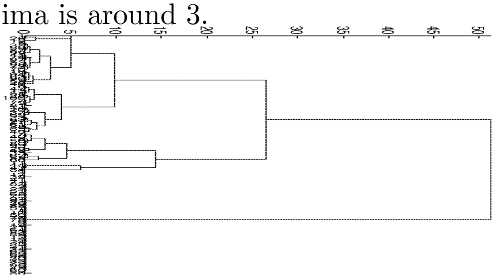
Figure 4: The PDF of eight FGMM examples are very diversified with different degrees of complexity summarized in both dimension and data distribution components.



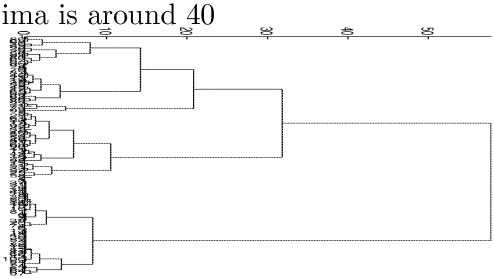
(a) The number of distinct local maxima is enormously reduced.



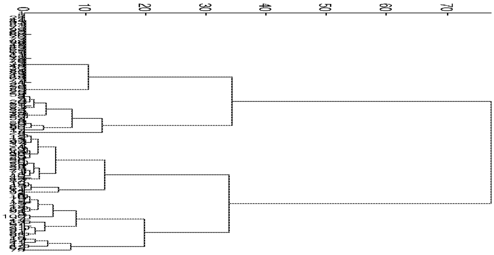
(b) The avarage number of local maxima is around 3.



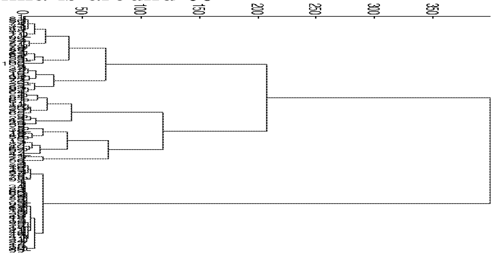
(c) The avarage number of local maxima is around 40



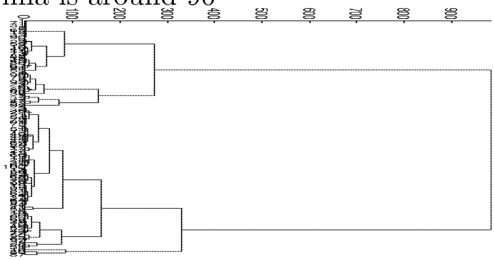
(d) The avarage number of local maxima is around 60



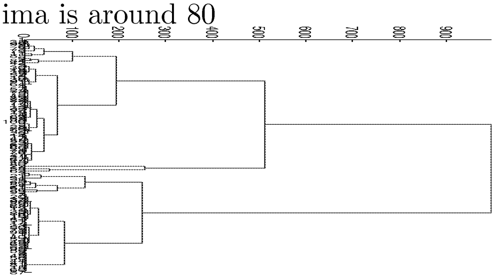
(e) The avarage number of local maxima is around 60



(f) The avarage number of local maxima is around 90



(g) The avarage number of local maxima is around 80



(h) The avarage number of local maxima is around 100

Figure 5: The dendrogramms are generated from 100 iterations of the EM algorithm using the MS method as initial value and the single linkage as criterion.