A SPECTRAL ANALYSIS APPROACH FOR GAUSSIAN MIXTURE ESTIMATION

Nicolas Paul, Michel Terre, Luc Fety

Electronique et Communications, CNAM, 292 rue Saint-Martin, 75141 Paris Cedex 3, France
Telephone: 33 1 40 27 25 67, Fax: 33 1 40 27 24 81, Email: nicolas.paul@cnam.fr

ABSTRACT
This paper deals with the estimation of one-dimensional Gaussian mixture. Given a set of observations of a K-component Gaussian mixture, we focus on the estimation of the component expectations. The number of components is supposed to be known. Our method is based on a spectral analysis of the estimated first characteristic function. We construct a Toeplitz matrix \( R_M \) with \( 2M - 1 \) estimated samples of the first characteristic function and show that the mixture component expectations can be derived from the eigenvector decomposition of \( R_M \). Simulations illustrate the performance of our algorithm on several configurations of a six-component Gaussian mixture. In the investigated scenarios the proposed method outperforms the Expectation-Maximization algorithm.

Index Terms—unsupervised learning, parameter estimation

1. INTRODUCTION
In this paper we deal with Gaussian mixture estimation. Given a set of one-dimensional observations originating from \( K \) possible Gaussian components, we focus on the estimation of the component expectations. The number of components is supposed to be known, and the component expectations are supposed to be all different.

One method consists in estimating a sampling of the observations probability density function (pdf), a mixture of \( K \) pdf, by associating a kernel to each observation and adding the contribution of all the kernels. A search of the pdf modes then leads to the component expectations. The drawback of such a method is that it requires the selection of extra-parameters (kernel design, sampling intervals). Furthermore, the final mode search algorithm might fail because of spurious local maxima in the estimated pdf.

An alternative method consists in using the Expectation-Maximization (EM) algorithm. Each EM iteration consists of two steps. The Expectation step estimates the probability for each observation to come from each mixture component. Then, during the Maximization step, these estimated probabilities are used to update the estimation of the mixture parameters. This procedure converges to a local maximum of the likelihood. The main drawback of the EM algorithm is the potential convergence to some local non-global maxima of the likelihood. Some solutions consist, for instance, in using smart initializations or stochastic optimization.

In this contribution we propose a new approach based on a spectral analysis of the first characteristic function (CF). We define a Toeplitz matrix \( R_M \) with \( 2M - 1 \) estimated samples of the CF and show that the mixture component expectations can be estimated from an eigenvector decomposition of \( R_M \). The proposed method is strongly inspired from the multiple signal classification algorithm MUSIC which aims at estimating the frequencies in a sum of sinusoids. The paper is organized as follows: In section 2 the observation model is presented and an analytical expression of the CF of a Gaussian mixture is given. In section 3 the matrix \( R_M \) is defined and some properties of \( R_M \) are described. Section 4 then presents the complete estimation algorithm. Section 5 illustrates the estimation performances on a six-component Gaussian mixture with different configurations. Conclusions are finally given in section 6, as well as perspectives for using the proposed method to estimate the number of components in a mixture.

2. GAUSSIAN MIXTURE

2.1. Probability density function (pdf)
Let \( \{p_k\}_{k \in \{1 \cdots K\}} \) be a set of \( K \) positive mixing weights that sum up to one. The multimodal pdf of the random observable variable \( Z \) is a finite mixture given by:

\[
 f_Z(z) = \sum_{k=1}^{K} p_k g(z, a_k, \sigma_k),
\]

where \( g(z, a_k, \sigma_k) \) is the Gaussian pdf given by: \( g(z, a_k, \sigma_k) = \frac{1}{\sqrt{2\pi\sigma_k}} \exp \left( -\frac{(z-a_k)^2}{2\sigma_k^2} \right) \) and \( a_k \) and \( \sigma_k \) are respectively the expectation and the standard deviation of component \( k \).

Given a set of \( N \) observed realizations \( \{z_n\}_{n \in \{1 \cdots N\}} \) of \( Z \) we focus on the estimation of the \( K \) component expectations \( \{a_k\}_{k \in \{1 \cdots K\}} \). Our proposal is mainly based on the estimated first characteristic function (CF) of the mixture.
2.2. First characteristic function (CF)

In general, the CF of a random variable \(X\) is defined by:
\[
\phi_X(t) = \mathbb{E}_X\{e^{itX}\}, \quad t \in \mathbb{R},
\]
(2)
where \(\mathbb{E}\{\cdot\}\) is the mathematical expectation with respect to the pdf of \(X\). For instance, the CF of a Gaussian random variable \(X\) with pdf \(g(x,a,\sigma)\) is given by [5]:
\[
\phi_X(t) = \int_{\mathbb{R}} e^{itx} g(x,a,\sigma) dx = e^{- \frac{a^2 t^2}{2}} e^{iat},
\]
(4)
Consequently the CF of the random variable \(Z\) with the pdf described in [1] is given by:
\[
\phi_Z(t) = \sum_{k=1}^{K} p_k \int_{\mathbb{R}} e^{itz} g(z, a_k, \sigma_k) dz
\]
\[
= \sum_{k=1}^{K} p_k e^{- \frac{a_k^2 t^2}{2}} e^{iat_k},
\]
(6)
Now let \(\phi_m\) be the sampled version of \(\phi_Z(t)\) with a sampling period \(T_c\). According to [6], we have:
\[
\phi_m = \frac{\Delta}{m} \phi_Z(mT_c), \quad m \in \mathbb{Z}
\]
\[
= \sum_{k=1}^{K} p_k a_{k,m} w_k^m,
\]
(8)
where \(w_k\) and \(a_{k,m}\) are defined by:
\[
w_k = e^{ia_k T_e}
\]
\[
a_{k,m} = e^{- \frac{2\pi (mT_c)^2}{T_e^2}}.
\]
(10)
In practical situation, \(\phi_m\) can be estimated from a set of \(N\) observations \(\{z_n\}_{n=1,\cdots,N}\) using:
\[
\hat{\phi}_m = \frac{1}{N} \sum_{n=1}^{N} e^{iz_n(mT_c)}
\]
(11)
In section 3 we will show how the \(\{w_k\}_{k=1,\cdots,K}\) defined in (9) can be estimated from the sampled CF. Once the \(\{w_k\}_{k=1,\cdots,K}\) are estimated, the \(\{a_k\}_{k=1,\cdots,K}\) can be obtained without ambiguity if the sampling period \(T_e\) is less than \(\frac{2\pi}{\max\{z_n\} - \min\{z_n\}}\). If we for instance choose:
\[
T_e = \frac{2\pi}{2\max\{z_n\} - \min\{z_n\}},
\]
(12)
then \(\frac{2\pi}{T_e} = 2(\max\{z_n\} - \min\{z_n\})\). Since \(\min\{z_n\} \leq a_k \leq \max\{z_n\}\) there is exactly one integer \(l_{w_k}\) such as:
\[
\frac{\angle(w_k)}{T_e} + l_{w_k} \frac{2\pi}{T_e} \in [\min\{z_n\}, \max\{z_n\}],
\]
and we have:
\[
a_k = \frac{\angle(w_k)}{T_e} + l_{w_k} \frac{2\pi}{T_e}, \quad k = 1, \cdots, K.
\]
(14)
Furthermore, if \(T_e\) verifies (12), and since the \(\{a_k\}_{k=1,\cdots,K}\) are supposed to be all different, then the \(\{w_k\}_{k=1,\cdots,K}\) are also all different. This will be used in section 3.

3. DEFINITION AND PROPERTIES OF \(R_M\)

Let \(R_M = (r_{jl})_{j,l=1,\cdots,M} \in \mathbb{C}^{M\times M}\) be the Toeplitz matrix with the following elements:
\[
r_{jl} = \phi_{l-j}, \quad j, l = 1, \cdots, M
\]
(15)
where \(\phi_m\) has been defined in (7). Note that \(\phi_m = \phi_{m}^{*}\), so \(R_M\) is a Hermitian matrix and one only has to compute \(M\) samples of the CF to build \(R_M\). Including (6) into (15):
\[
r_{jl} = \sum_{k=1}^{K} p_k \alpha_{k,l-j} w_{k}^{l-j}
\]
(16)
\[
= \sum_{k=1}^{K} p_k w_{k}^{l-j} + \sum_{k=1}^{K} p_k (\alpha_{k,l-j} - 1) w_{k}^{l-j}
\]
(17)
\[
= \sum_{k=1}^{K} w_{k}^{l-j} - \sum_{k=1}^{K} p_k (\alpha_{k,l-j} - 1) w_{k}^{l-j}.
\]
(18)
where in (18) we used that \(w_{k}^{l-j} = w_{k}^{l} w_{k}^{-j} = w_{k}^{l+1} w_{k}^{-j-1}\) since \(w_{k}^{-1} w_{k}^{-1} = 1\). A consequence of (18) is that \(R_M\) can be expressed as the sum of a "signal" matrix \(S_M\) and a "perturbation" matrix \(P_M\):
\[
R_M = S_M + P_M
\]
(19)
where the "signal" matrix \(S_M\) is given by:
\[
S_M = (w_1, \cdots, w_K) D (w_1^H, \cdots, w_K^H) \in \mathbb{C}^{M\times M}
\]
(20)
\[
w_k = (1, w_k^1, \cdots, w_k^{M-1})^H \in \mathbb{C}^M
\]
(21)
\[
D = \text{diag}(p_1, \cdots, p_K) \in \mathbb{R}^{K\times K}
\]
(22)
and the "perturbation" matrix \(P_M\) is given by:
\[
P_M = \sum_{k=1}^{K} p_k (\alpha_{k,l-j} - 1) w_{k}^{l-j} \in \mathbb{C}^{M\times M}
\]
(23)
The "signal" matrix \(S_M\) is a well-known matrix in the spectral analysis community. It is the auto-correlation matrix of a received sum of \(K\) sinusoids with angular frequencies \(a_k\) and
power $p_k$. High resolution algorithm such as MUSIC [4] estimate $S_M$ from some (potentially corrupted) signal samples then estimate the sinusoid frequencies from its eigenvector decomposition. Indeed, since the $w_k$ are all different (section 2.2), one can show that the rank of $S_M$ is equal to $K$ and that the signal vectors $w_k$ defined in (21) are orthogonal to any vector of the kernel of $S_M$. Consequently, if $\mathbf{V} \triangleq (v_{K+1}, \ldots, v_M) \in \mathbb{C}^{M \times M-K}$ contains $M-K$ orthogonal eigenvectors belonging to Ker($S_M$). Then estimate the sinusoid frequencies from its eigenvector decomposition. Indeed, since the “signal” matrix $S_M$ (24) is corrupted with the “perturbation” matrix $P_M$ (23). When all the component variances tend to zero (ideal case) the perturbation matrix $P_M$ tends to a null matrix: using (10) and (23) we have:

$$\lim_{\sigma_k \to 0} \alpha_{k,j} = 1, \quad k = 1, \ldots, K$$

and estimated ordered vector of component expectations:

$$\hat{c}_k = \frac{1}{M} \sum_{n=1}^{N} z_n$$

Table 1. Means, variances, weights of the simulated mixture

| Scenario | Mean | Variance | Weight |
|----------|------|----------|--------|
| 1        | $0$  | $\sigma^2$ | $0.2$ |
| 2        | $1$  | $\sigma^2$ | $0.1$ |
| 3        | $2$  | $\sigma^2$ | $0.2$ |
| 4        | $4$  | $\sigma^2$ | $0.2$ |

5. **SIMULATION**

In our simulations several types of a six-component Gaussian mixture have first been considered. The set of expectations is equal to $(0, 1, 2, 4, 5, 6)$, with a difference of one or two between two successive component expectations. Four cases have been studied: common variance and common weight (scenario 1), different variances and common weight (scenario 2), common variance and different weights (scenario 3) and different variances and different weights (scenario 4). A summary of the scenarios is given in Table 1. The parameter $\sigma$ in Table 1 enables to simulate different overlapping situation. The number $N$ of observations per simulation run is 200. The RA-based algorithm has been run as described in section 4 with $M = 2K$. This algorithm has been compared to the EM algorithm [2] with a uniform random start and a maximal number of 100 iterations. See [3] for a detailed description of the Gaussian mixture estimation with EM. A constrained version of the EM (EMc) which imposes a common variance and a common mixing weight has been used to prevent the convergence to components with an almost null variance. In all the scenario, EMc provides better estimates than the standard EM, even in the scenario where the component variances or the component weights are different. Therefore only the performances of EMc are presented here. To get rid of the permutation ambiguity, the estimation performance is evaluated as follows: If $a \in \mathbb{R}^K$ is the vector of the true component expectations and $\hat{a} \in \mathbb{R}^K$ is the vector of the estimated component expectations at simulation run $r$, the performance criterion $e_r$ is defined as the maximal absolute distance between the true and estimated ordered vector of component expectations:

$$e_r = \| \text{sort}(a) - \text{sort}(\hat{a}_r) \|_\infty.$$
Given a set of observations originating from a $K$-component univariate mixture, we focused on the estimation of the component expectations when the number $K$ of components is known. We proposed a method based on the eigenvector decomposition of a Toeplitz matrix $R_M$ built from some estimated samples of the first characteristic function. Simulations illustrated the superiority of the proposed method compared with the Expectation-Maximization algorithm on various configurations of a six-component Gaussian mixture. More theoretical investigations are now needed to study the influence of the perturbation matrix (25) on the performances.

Our current research also deals with the case of an unknown number of components. In figure 2 we plot the eigenvalues of $R_M$ with $M = 10$ in scenario 4 (6 mixture components) for $\sigma = 0.15$.

### 6. CONCLUSION

Given a set of observations originating from a $K$-component univariate mixture, we focused on the estimation of the component expectations when the number $K$ of components is