ON THE PARAMETERS ESTIMATION OF THE GENERALIZED GAUSSIAN MIXTURE MODEL

Ould Mohamed M. Mohamed, M.Jaïdane-Saïdane

Ecole Nationale d'Ingènieurs de Tunis, Signals and Systems Research Unit
Tunis, BP 37, Le Belvédère, 1002, Tunisie.
email: med.barrar@enit.rnu.tn, meriem.jaidane@wanadoo.tn

ABSTRACT

The parameters estimation of mixture distributions is an important task in statistical signal processing, Pattern recognition, blind equalization and other modern statistical tasks often call for mixture estimation. This paper aims to provide a realistic distribution based on Mixture of Generalized Gaussian distribution (MGG), which has the advantage to characterize the variability of shape parameter in each component in the mixture. We propose a formulation of the Expectation Maximization (EM) algorithm under Generalized Gaussian distribution. For this, two different methods are proposed to include the shape parameter estimation. In the first method a derivation of the Likelihood function is used to update the mixture parameters. In the second approach we propose an extension of the "classical" (EM) algorithm and to estimate the shape parameter in terms of Kurtosis. The Kullback-Leibler divergence (*KLD*) is used to compare, and evaluate these algorithms of MGG parameters estimation. An application of this technique is considered for modeling load distribution which exhibits an heterogeneity with a high variability of shape parameters [1].

1. INTRODUCTION

Modeling the distribution of the observed data with a parametric approach is an important tool of statistical signal processing. In applications where the data have many clusters, like image segmentation, a multi-component probabilistic model representation such as mixture modeling is required. Actually, in many real applications in the field of image processing, Audio, blind equalization, there are a need to better approximate the observed data and the use of mixture approach.
The mixture modeling technique has been widely used for the estimation of the probability density function and has found significant applications in various domains (see for example, Refs. [1, 2, 3]). However little work has been reported in the use of the MGG [4, 5]. To consider the shape variability we propose here the use of the generalized gaussian (GG) distribution as the basic distribution in the mixture. The GG distribution has been employed to model and detect Gaussian and non-Gaussian signals [2, 6], recently in speech modeling [7], and image and video coding [8].

[1] This work was supported by the 2005/2007 VRR (Projet de Valorisation des Résultats de Recherche du Ministère de la Recherche Scientifique de la Technologie et du Développement des Compétences) research project between the Tunisian Company of Electricity and Gas, and Signals and Systems Research Unit (U2S) of the ENIT.

A French version, not detailed, of this paper is accepted and presented at the International Conference JDS held in Bordeaux, France, May 25-29, 2009.

In this paper, we focus on the parameters estimation of the generalized gaussian mixture model. A particular interest is given to estimate the shape parameters. To this end two different methods which make use of the EM (Expectation Maximization) algorithm [9] to obtain an ML (Maximum Likelihood) has been developed. In the first method we extend the work in [4] which addresses only the case of 2 components by applying it more generally for estimating parameters considering any number of components in the mixture. In this method the shape parameters are updated by numerical optimization of the Likelihood function. In the second method we propose another efficient alternative to estimate the shape parameter of each component which is very easy and quick to be implemented. In this method the sharpness parameters are estimated as a function of a Higher Order Statistic (HOS), namely the Kurtosis.
An application of this technique in real load data obtained from the Tunisian power system is considered. For this purpose we use generalized gaussian mixture model to segregate the load data into two classes, related to the difference between daytime and evening load data and seasonal variation, which are assumed to follow a Generalized Gaussian (GG) distribution.
The rest of this paper is organized as follows: In section 2 we introduce the Mixture of Generalized Gaussian (MGG) approach. In Section 3, we present our approaches for the parameters estimation. In section 4, we discuss the performance of the MGG model applied at first to a theoretical example then on real load data. Technical details are provided in the Appendix.

2. THE MIXTURE OF GENERALIZED GAUSSIAN DISTRIBUTION

A mixture of generalized gaussian MGG is a parametric statistical model which assumes that the data originates from a weighted sum of several generalized Gaussian sources. More specifically, a MGG is defined as:

$$p(x \mid \theta) = \sum_{i=1}^{K} \omega_i p_i(x \mid m_i, \sigma_i, c_i) \tag{1}$$

Where

- $\theta = (\omega_i, m_i, \sigma_i, c_i)$, $i = 1, 2 \cdots K$
- K is the number of mixture density components.
- ω_i is the ith mixture weight and satisfies $\omega_i \geq 0$, $\sum_{i=1}^{K} \omega_i = 1$
- $p_i(x \mid m_i, \sigma_i, c_i)$ is an individual density of the Generalized Gaussian (GG) distribution which is characterized

by the following probability density function [10, 2]:

$$p_i(x) = \frac{c_i \gamma_i}{\Gamma(1/c_i)} e^{-\gamma_i^{c_i}[|(x-m_i)|]^{c_i}} \tag{2}$$

where γ_i is the ith scale parameter $\gamma_i = \frac{1}{\sigma_i}\frac{\Gamma(3/c_i)}{\Gamma(1/c_i)}^{1/2}$

- c_i is the ith shape parameter, m_i is the ith mean, σ_i is the standard deviation.

The shape parameter c_i is a measure of Kurtosis (flatness) and controls the deviation from the normality of the distribution.
$\Gamma(x) = \int_0^\infty \tau^{x-1} e^{-\tau} d\tau$ is the Gamma function.
By varying the shape parameter, it is possible to characterize a large class of distribution including gaussian, sub-gaussian (more peaked, than Gaussian, heavier tail) and super-gaussian (flatter, more uniform). It is noticed that if $c = 2$ the GG coincides with the gaussian model and if $c = 1$, it represents Laplace distribution.

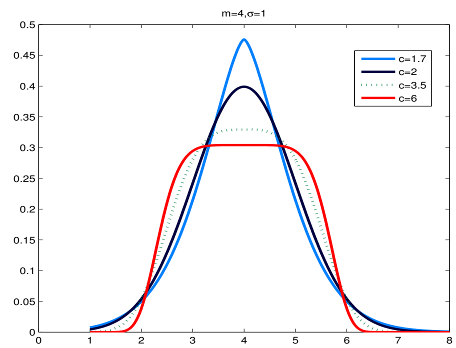

Figure 1: The probability density function of the Generalized Gaussian Mixture distribution for $K = 1, m_1 = 4, \sigma_1 = 1$. The MGG allows us to include the case of the gaussian mixture model ($c = 2$).

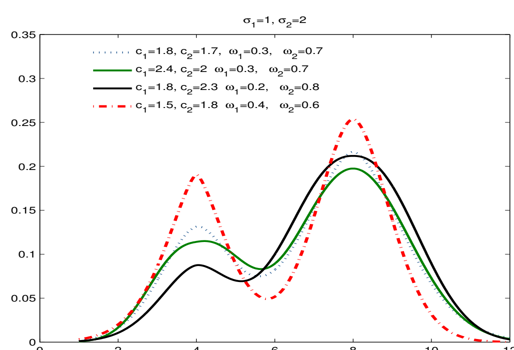

Figure 2: The probability density function of the Generalized Gaussian Mixture distribution for $K = 2$, and different shape, weight parameters, $\sigma_1 = \sigma_2 = 2$, $m_1 = 4$, $m_2 = 8$.

Fig.1 and Fig.2 show examples of pdf for MGG distribution for $K = 1$ and $K = 2$. Thanks to the shape parameter the MGG distribution is more flexible and can approximate a large class of statistical distributions. This distribution requires thus the estimation of 4*K parameters, $\theta(\omega_i, m_i, \sigma_i, c_i)$, $i = 1, 2 \cdots K$. Particularly we focus, in this paper, on the shape parameter estimation. This is discussed in detail in the following sections

3. REFORMULATION OF THE EM ALGORITHM IN THE CASE OF MGG MODEL TO ESTIMATE THE SHAPE PARAMETERS

The parameters estimation of the mixture of generalized gaussian (MGG) is more complex than in the case of mixture of gaussian. Difficulty lies in the estimate of the shape parameters c_i $i = 1, \cdots K$.

We propose in this study, to estimate the MGG parameters, with an extension of the "Expectation-Maximization"(EM) algorithm [11] which allows to maximize the complete log of Likelihood function based on equation 1, given by [12].

$$L(X \mid \theta) = \sum_{i=1}^{K} \sum_{j=1}^{N-1} h_{i,j} \ \ln \left[\omega_i p_i(x_j \mid m_i, \sigma_i, c_i)\right] \tag{3}$$

where
N the sample size
$h_{i,j} = p(i \mid x_j)$ ($i = 1, \cdots, K$ and $j \in [0, N-1]$) represents the conditional expectation of p_i given the observation x_j, means the posterior probability that x_j belongs to the ith component.
In the case of Generalized Gaussian distribution, if we substitute equation 2 into 3 and after some manipulation we obtain the following form of $L(X \mid \theta)$:

$$L(X \mid \theta) = \sum_{i=1}^{K}\sum_{j=1}^{N-1} h_{i,j} \ln(\omega_i) + \sum_{i=1}^{K}\sum_{j=1}^{N-1} \left[h_{i,j}\left(\ln c_i - \ln 2 - \ln \gamma_i - \ln\ \Gamma\left(\frac{1}{c_i}\right) - \gamma_i^{(c_i)}|x_j - m_i|\right)\right] \tag{4}$$

The notation which has been used throughout this paper is as follows: subscripts i refer to the mixture element, subscripts j to particular elements of a data vector superscripts (n) to the iteration of the algorithm.
The log likelihood $L(X \mid \theta)$ is optimized iteratively via the EM algorithm which has been extensively applied in the case of gaussian mixture.

The steps of the "classical" EM algorithm [11] in the case of generalized gaussian distribution can be summarized as followed:

- **Initialization** Initialize the model parameter θ,
- **Expectation step (E-step)**
  The expectation step is represented by the computation of the conditional expectation probability $h_{i,j}$:

$$h_{i,j}^{(n+1)} = \frac{\omega_i p(x_j|m_i^{(n)}, \sigma_i^{(n)}, c_i^{(n)})}{\sum_{r=1}^{K} \omega_r^{(n)} p(x_j|m_r^{(n)}, \sigma_r^{(n)}, c_r^{(n)})} \tag{5}$$

  In this step the computation of the $L(X \mid \theta)$ based on $\theta^{(n)}$ is made.
- **Maximization step (M-Step)** Allows numerical maximization of the log-likelihood function (equation 4) given $h_{i,j}^n$ and θ^n

$$\widehat{\theta}^{(n+1)} = arg\ max_\theta L(X \mid \theta^{(n)}) \tag{6}$$

$\widehat{\theta}^{(n+1)} = (\hat{\omega}_i^{(n+1)}, \hat{m}_i^{(n+1)}, \hat{\sigma}_i^{(n+1)}, \hat{c}_i^{(n+1)})$
In the case of mixture of simple gaussians the parameters $(\omega_i, m_i, \sigma_i)$ are estimated, in the iteration $(n+1)$, with a set of iterative equations [9, 12]

$$\hat{\omega}_i^{(n+1)} = \frac{1}{N}\sum_{j=1}^{N} h_{i,j}^{(n)} \tag{7}$$

$$\hat{m_i}^{(n+1)} = \frac{\sum_{j=1}^{N} h_{i,j}^{(n)} x_j}{\sum_{j=1}^{N} h_{i,j}^{(n)}} \quad (8)$$

$$\hat{\sigma_i^2}^{(n+1)} = \frac{\sum_{j=1}^{N} h_{i,j}^{(n)} (x_j - \hat{m}_i^{(n)})^2}{\sum_{j=1}^{N} h_{i,j}^{(n)}} \quad (9)$$

However *the problem is how to estimate the added shape parameter c_i in the mixture of GG*.
To deal with the problem we propose, in the following sections, two different methods.

3.1 Numerical optimization of the log likelihood function to estimate the shape parameters

We propose in this section a generalization of the method proposed in [4] which addresses only the case of 2 components by setting the derivatives of the log-likelihood function (equation 4) to zero with respect to m_i,γ_i, c_i respectively:

$$\frac{dL(X \mid \theta)}{dm_i} = 0 \ , \ \frac{dL(X \mid \theta)}{dc_i} = 0 \ , \frac{dL(X \mid \theta)}{d\gamma_i} = 0 \quad (10)$$

Accordingly, we obtain for $i = 1, \cdots, K$ the following nonlinear equation related to the shape parameters (The equations related to the estimation of mean and scale parameters are presented in Appendix 6):

$$\varphi(c_i) = \sum_{j=1}^{N-1} h_{i,j} \left[\frac{1}{c_i} + \frac{1}{c_i}^2 \Psi\left(\frac{1}{c_i}\right) - \left(\frac{|x_j - m_i^{(n)}|}{\gamma_i^{(n)}}\right)^{c_i} ln \left(\frac{|x_j - m_i^{(n)}|}{\gamma_i^{(n)}}\right)\right] = 0 \quad (11)$$

Where
$\Psi(\bullet)$ is the digamma function ($\Psi(x) = \Gamma'(x)/\Gamma(x)$). (for simplicity the subscript (n+1) referring to the iterations are omitted in this expression for c_i). To solve this equation numerical optimization method based on Newton Raphson procedure is applied. This procedure involve the following updating equation:

$$c_i^{(r+1)} = c_i^{(r)} - \frac{\varphi(c_i^{(r)})}{\varphi'(c_i^{(r)})} \quad (12)$$

$\varphi(c_i^{(r)})$ is given by equation (11), the calculation of the terms $\varphi'(c_i^{(r)})$ can be obtained in the reference [4]. In what follows, We denote this method by **NOP** (reference to Numerical OPtimisation).

This approach of shape parameter estimation is very complex, because the system to resolve is strongly nonlinear. This method also reveals an important sensibility on initial conditions and very important time of calculation

In the following section we propose a new method which extends the iterative process proposed in [12] to include shape parameter. This allows to avoid heavy computation in the previous method.

3.2 Use of HOS method in M-step to estimate the shape parameters

To include the shape parameter estimation in the mixture problem we propose, in the second method, to use the relation between c_i and the Kurtosis (κ_i) which describes sharpness variability. Hence the analytical relationship between c and κ has been introduced (see [13] for details):

$$\kappa_i = \frac{E(x - m_i)^4}{(E(x - m_i)^2)^2} = \frac{\Gamma(5/c_i)\Gamma(1/c_i)}{(\Gamma(3/c_i))^2} \quad (13)$$

where $\Gamma(x)$ is the Gamma function.

In the first step we estimate, in $(n+1)$ iteration, the kurtosis $\kappa_i^{(n+1)}$ in the same way (with the same weights) as m_i^{n+1} and $\sigma_i^{(n+1)}$ (equation 8-9) :

$$\hat{\kappa_i}^{(n+1)} = \frac{\sum_{j=1}^{N} h_{i,j}^{(n)} (x_j - m_i^{(n)})^4}{(\sigma_i^{(n)})^4 \sum_{j=1}^{N} h_{i,j}^{(n)}} \quad (14)$$

In second step we use this estimation of $\hat{\kappa_i}^{(n+1)}$ in the following approximation of the inverse of the expression (13), given by Regazzoni [10], which is obtained by applying the least squared method (LSM) on a generic second-order monotonic analytical expression of (13) :

$$\hat{c_i}^{(n+1)} \approx \sqrt{\frac{5}{\hat{\kappa_i}^{(n+1)} - 1.865}} - 0.12 \quad (15)$$

This expression allow a good approximation of $c_i^{(n+1)}$ as a function of $\hat{\kappa_i}$ for the range of validity $\hat{\kappa_i}^{(n+1)} > 1.865$. This range of validity includes about all the kurtosis values measured in the case of our application in real load data, in which kurtosis variability was measured in the range [2.07, 3.70].

4. EXPERIMENTAL RESULTS

In this section, several simulation experiments are carried out to demonstrate the performance of the proposed algorithms. We focus in these experiments to the shape parameters estimation. Before applying the MGG model to our real load data we propose in the following subsection to assess the algorithms performance against known densities.

4.1 Numerical example : known densities

The proposed methods of MGG parameters estimation are at first tested and confirmed on theoretical example from simulated data from various distributions and compare the simulated data to the fitting distribution with MGG model evaluated with method **NOP** (presented in section 3.1) and method **HOS** (presented in section 3.2). Unfortunately we have not

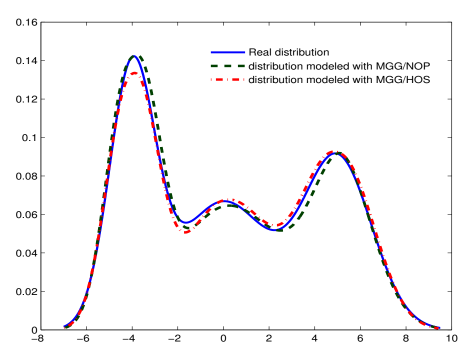

Figure 3: probability density (solid line) of mixture of gaussian random variable with $K = 3$, generated by simulating a mixture of gaussian with parameter values m=[-4, 0, 5], $\sigma^2 = [1, 2, 1.5]$ $N = 5000$ (sample size), and fitting by using the MGG model, parameters estimated by NOP method (dash line) and HOS method (dash-not line). The results of shape parameters estimation with HOS are $c_1 = 2.13$, $c_2 = 2.04$, $c_3 = 1.98$. in this case $KLD = 9.35.10^{-4}$. With the NOP method we obtain c_1 =2.07, c_2 =1.82, c_3 =2.06. in this case $KLD = 1.2.10^{-3}$

a generator of mixture of generalized gaussian. However we can use the generator of mixture of gaussian as particular case of MGG.

We present here the example of a "standard" mixture of gaussian for which the shape parameters are known ($c_i = 2$). For this, univariate multimodal Gaussian random number generator is used.

As example, the figure 3 represents the known densities (solid line) of mixture of gaussian random variable with $K = 3$ (number of components) generated with the parameters values $m = [-4, 0, 5]$, $\sigma^2 = [1, 2, 1.5]$, $\omega = [0.3, 0.3, 0.3]$ and sample size $N = 5000$. In the same figure we show our result of fitting this density by using the MGG model with the two algorithms of parameters estimation : method NOP (dash line) and HOS (dash-not line). We notice that in both cases the parameters of shape converged towards the real values $c = 2$: $\widehat{C} = [2.13, 2.04, 1.98]$ for HOS method and $\widehat{C} = [2.07, 1.82, 2.06]$ for NOP method.

These results show also that the two methods give close results for the estimation of the shape parameters. However, the NOP approach is heavier and requires more computational time. Furthermore, the measure of the Kullback-Leibler divergence [15] between two probability distributions in both cases, yielding $KLD = 9.35.10^{-4}$ for the HOS approach and $KLD = 1.2.10^{-3}$ for NOP approach shows that, globally, the HOS gives the better approximation of the mixture distribution. For its simplicity we retain in the following the HOS method.

4.2 Application in the case of real peak load data

[2] In this section, we present an application to real data. The MGG model is used to model the annual load distribution. Such information is greatly useful for estimating the operating cost of resource plans[14]. During the recent years, the

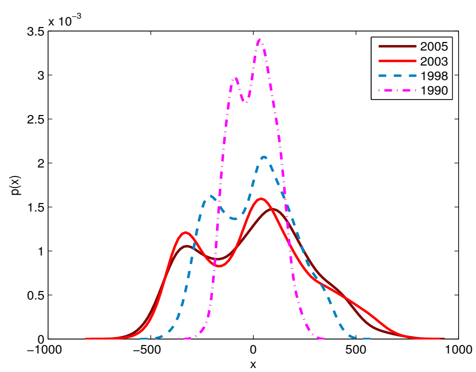

Figure 4: Qualitative modification of the annual hourly peak load (centered data) distribution (Example years 1990-1998-2003-2005).

load curve of tunisian power systems has undergone a dramatic change. This qualitative change is observable through the histogram of the dispersal of the daily peak load. This one represents non-parametric estimation of the probability density according to the Kernel method [15] during the period of 1990-2005. The histogram (Fig.4) makes it possible to notice the following :

- the annual load distribution revealed a bimodal form,
- high shape variation of the annual load distribution.

[2] Detailed application of this model is proposed in [5]to study the load variability in summer and winter seasons. In [5], we use the MGG model to give parametric expression of Load Duration Curve (LDC) which is connected the cumulative probability distribution function.

This observation of the power distribution behavior motivates the application of the mixture of generalized gaussian distribution. Thus the proposed model is evaluated for the annual load distribution. To obtain the estimation of the associated parameters we have used the hourly data (365*24=8760 points) and we retained, in this application, the second method of **HOS**.

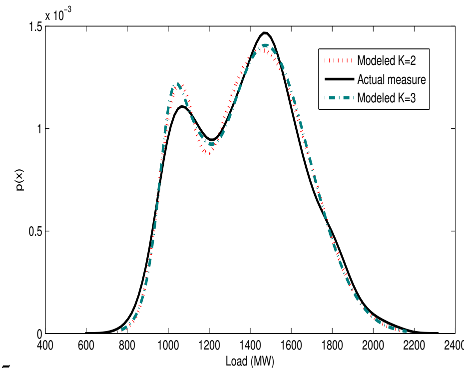

Figure 5: Measured annual load distribution (solid line) in comparison with modeled probability distribution MGG obtained with **HOS** method ($K = 2$ dash line) and ($K = 3$ dash-dot line) obtained from hourly data (8760 hours) 2005. The estimated shape parameters for $K = 2$ are : $c_1 = 1.97$, $c_2 = 2.10$

Fig.5 shows, for the example of 2005, the true distribution (solid line) and the modeled distribution for K =2 (dash line) and for K =3 (dash-dot line) obtained with **HOS** method (developed in section 3.2). Fig.6 illustrate an example of convergence of the **HOS** method for estimation of the MGG shape parameters in the case of mixture of two components.

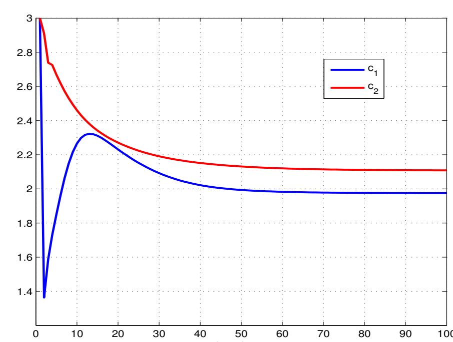

Figure 6: Convergence of the shape parameters for the example of the annual load distribution (2005), $K = 2$, with iterations number of 100. numerical results: $\widehat{c}_1 = 1.97$, $\widehat{c}_2 = 2.10$.

We can see from Fig.5 that the generalized gaussian mixture distribution makes a good representation for the load distribution. Also the results show that the estimated distributions for K=2 and K=3 are relatively very close. This is why, a choice of number of components $K = 2$ seems to be sufficient. In this case the load shape variation can be characterized by the two shape parameters c_1, c_2.

The results of estimation of c_1, c_2, in the recent years, are summarized in table 1.

We notice that the values of c1 and c2 vary from one year to another, the load classes have not necessarily the same form ($c_1 \neq c_2$), are not also necessarily gaussian ($c_1 \neq 2$, $c_2 \neq 2$). This results show that thanks to the shape parameter the MGG model is more flexible than the "standard" mixture

Table 1: The estimated values of shape parameters c_1 and c_2 from the hourly load data (2000-2006) using the HOS method.

	2000	2001	2002	2003	2004	2005	2006
$c1$	1.54	1.71	1.79	2.08	1.32	1.97	2.09
$c2$	2.42	2.65	2.39	2.18	2.40	2.10	2.26

of gaussian, and can approximate a large peak load behavior which varies according to the various load classes.

It is important to not that the problem related to the choice of initial conditions is more excessive in the first method where the EM algorithm convergence is not granted for a poor starting point . This constitutes a potential weakness of such methods. This point it is not treated in this work and remains to be investigated in other works. However, we note that in our application good initial parameters of load classes can be chosen if we consider the priori knowledge on the daytime and evening load classes separately.

5. CONCLUSION

The main motivation of this study is to provide estimators for the MGG parameters. For this, a new formulation of the EM algorithm is conducted to include the shape parameter estimation. Two different approaches for estimation of the MGG parameters are performed. The first method used numerical maximization of the log likelihood function of the mixture. In the second approach which is easy to be implemented make use of HOS in the maximization step to estimate shape parameter. Promising results have been obtained in the application of the MGG to encircle the load distribution variability. The results indicated that the HOS method has a lower computational time than the numerical maximization method.

In future work, further investigation will be conducted to deal with the problem of initial conditions sensitivity and to refine the choice of number of components in our application.

6. APPENDIX

M-Step based on the derivatives of the log-likelihood function

In this method the estimation is carried out by setting the derivatives of the log-likelihood function (equation 4) to zero. This allows to obtain the following equations:
For $m_i^{(n+1)}$:

$$\sum_{j=1}^{N-1} h_{i,j}^{(n)} \eta(x_j)|x_j - m_i^{(n+1)}|^{c_i^{(n)}-1} = 0 \qquad (16)$$

For $\gamma_i^{(n+1)}$:

$$\sum_{j=1}^{N-1} h_{i,j}^{(n)} [-\frac{1}{\gamma_i^{(n+1)}} + c_i^{(n)} (\gamma_i^{(n+1)})^{-c_i^{(n)}-1} |x_j - m_i^{(n)}|^{c_i^{(n)}}] = 0 \qquad (17)$$

where

$$\eta(x) = \begin{cases} 1 \text{ if } x_j - m_i^{(n+1)} < 0 \\ -1 \text{ if } x_j - m_i^{(n+1)} \geq 0 \end{cases} \qquad (18)$$

To calculate these parameters we need here to solve the non linear equations 16-17. For this numerical optimization method based on Newton Raphson procedure has been used.

REFERENCES

[1] Pekka Paalanen, Joni-Kristian Kamarainen, Jarmo Ilonen, Heikki Klviinen; "Feature representation and discrimination based on Gaussian mixture model probability densities Practices and algorithms," *Pattern Recognition* 39 (2006) pp. 1346-1358.

[2] C. Tzagkarakis, A. Mouchtaris, P. Tsakalides; "Musical Genre Classification VIA Generalized Gaussian and Alpha-Stable Modeling, ICASSP 2006 Proceedings. Vol.5, 14-19 May 2006 pp 217–220.

[3] Vimal Bhatia, Bernard Mulgrew, "Non-parametric likelihood based channel estimator for Gaussian mixture noise," *Signal Processing* Vol. 87 (2007) pp 2569-2586.

[4] Y. Bazia, L.Bruzzone, and F. Melgania "Image thresholding based on the EM algorithm and the generalized Gaussian distribution," *Pattern Recognition* vol.40 pp. 619–634 (2007).

[5] Mahmoud, M.O.M.; Jaidane-Saidane, M.; Hizaoui, N.; "The use of Mixture of Generalized Gaussian for trend analysis of the load duration curve: Summer and winter load variability in Tunisia," 43rd International UPEC 2008. 1-4 Sept Padova, Italy.

[6] Sharifi.K et al"Estimation of the shape parameter for generalized gaussian distribution in subband decomposition of Video," *IEEE Tans. on Circuits and Systems for Video Technology*, vol. 5, pp. 52–56, No.1 1995.

[7] K. Kokkinakis and A.K. Nandi. "Exponent Parameter Estimation for Generalized Gaussian Probability Density Functions with Application to Speech Modelling," Signal Processing, 85(9):1852-1858, 2005.

[8] Mohand Saïd Allili, Nizar Bouguila, "Finite Generalized Gaussian Mixture Modeling and Applications to Image and Video Foreground Segmentation," Conference on Computer and Robot Vision(CRV'07), IEEE 2007.

[9] J.A. Bilmes, "A gentle tutorial of the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models," Technical Report ICSI TR-97-021, UC Berkeley, 1997.

[10] A. Tesei, and C.S. Regazzoni,"HOS-based generalized noise pdfs models for signal detection optimisation," *IEEE Signal processing*, vol. 65, No.2 pp. 267–281, March. 1998.

[11] 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, 39:138, 1977.

[12] Figueiredo MAT, Jain AK. "Unsupervised learning of finite mixture models", *IEEE Trans Pattern Anal Mach Learn* 2002;24(3)pp.38196.

[13] A.Tesei, C.S.Regazzoni, "Use of fourth-order statistics in non-gaussian noise modelling for signal detection: the Generalized Gausian pdf in terms of kurtosis " in *proc. of EUSIPCO-96*, sep.1996.

[14] A.S Malik, B.J.Cory,, P.DC. Wijayatunga "Applications of probabilistic peak-shaving technique in Generation planning," *IEEE Trans on power systems*, Vol.14,No. 4, November 1999.

[15] David W. Scott, *Mulivariate Density Estimation*, Wiley-Interscience. 1992.