On Parameter Estimation of the Envelope Gaussian Mixture Model

Linyun Huang
Department of ECSE
Monash University, Australia
Linyun.Huang@monash.edu

Yi Hong
Department of ECSE
Monash University, Australia
Yi.Hong@monash.edu

Emanuele Viterbo
Department of ECSE
Monash University, Australia
Emanuele.Viterbo@monash.edu

Abstract—**In many communication systems, the Gaussian mixture model (GMM) is widely used to characterize non-Gaussian man-made and natural interference. The envelope distribution of such noise model is often expressed as a weighted sum of Rayleigh if in-phase and quadrature components of the noise are dependent. Instead, in this paper, a *simple* and *exact* closed form probability density function of the *envelope Gaussian mixture model* (i.e. the envelope of independent in-phase and quadrature components of complex non-Gaussian noise) is obtained. Furthermore, the problem of estimating of the *envelope Gaussian mixture* parameters is addressed. The proposed estimator of weights and variances is based upon the Expectation-Maximization (EM) algorithm.**

Index Terms—**Gaussian mixture noise, Envelope Gaussian Mixture, Maximum Likelihood, EM algorithm, parameter estimation.**

I. Introduction

In wireless communication systems, classical white Gaussian noise is often assumed to be a very accurate model. However, in other cases, such as the underwater [1] and the powerline communications [2], the noise may exhibit non-Gaussian behaviour and thus it is important to consider different versatile and robust noise/interference models. In 1977, Middleton [3] proposed the Middleton's Class A noise model, a Gaussian mixture density model with Poisson selection, to describe the electromagnetic (EM) interference from a variety of noise sources. Arzberger *et al.* [2] also suggested that the noise in powerline channels can also be modelled by the Gaussian mixture model, a parametric probability density function (pdf) expressed as summation of weighted Gaussian pdf's. The envelope distribution of this mixture density is often represented as weighted sum of Rayleigh under the assumption that the in-phase components and quadrature components of the noise are dependent [4]. In our work, we assume that both in-phase and quadrature components are independent and identically distributed (i.i.d.) random variables as motivated by [1], [2]. Hence, the envelope distribution of in-phase and quadrature noise components will not result in the Rayleigh mixture model, but give rise to the envelope Gaussian mixture model (EGMM).

Parameters of the envelope Gaussian mixture model can be estimated by maximum likelihood estimation (MLE). When closed form expression cannot be found for MLE, iterative methods, such as the Newton-based and the EM algorithm, are used. These algorithms iteratively maximize the log likelihood function. In an earlier work by Sari *et al.* [1], a conventional maximum likelihood estimator using the quasi-Newton method is used. In this paper, we consider the maximum likelihood estimator via the EM algorithm, a widely used method popularized by Dempster, Laird and Rubin in 1977 [5]. The EM algorithm has been used in many parameters estimation problems, especially in dealing with the curved exponential densities [6], such as the Middleton's Class A model [7], the Rayleigh mixture [8] and the Gaussian mixture density [9].

Maximum likelihood estimation via the EM algorithm is considered in this paper for the envelope Gaussian mixture model. The paper is organised as follows. In Section II, the model for non-Gaussian channel noise is described. In Section III, the envelope Gaussian mixture density function is derived. In Section IV, the maximum likelihood (ML) estimator via the EM algorithm is presented. The performance of the EM algorithm is compared to that of conventional ML estimator using the quasi-Newton method in Section V.

II. NON-GAUSSIAN NOISE MODEL

We first discuss the widely used two-term Gaussian mixture model for the in-phase and quadrature noise amplitudes Y and Z. Both in-phase and quadrature noise amplitudes are considered as independent random variables with the following two-term Gaussian mixture density function:

$$GM(a, \sigma_1^2, \sigma_2^2) = \frac{a}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{x^2}{2\sigma_1^2}} + \frac{(1-a)}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{x^2}{2\sigma_2^2}} \quad (1)$$

where $0 \leq a \leq 1$ is the mixing coefficient and σ_1^2, σ_2^2 are the variances of the two Gaussian components. An example where the above model can be used is when background noise is always present and impulsive noise events occur with probability $(1-a)$. The first Gaussian component can be seen as the nominal background noise with variance σ_1^2. The second component represents the combination of the background noise and the impulsive noise, when impulsive noise events occur. Since both background noise and impulsive noise are assumed to be Gaussian random variables, the sum of the two will also be a Gaussian random variable.

The two-term Gaussian mixture model was also considered as an approximation of Middleton's Class A noise model [10],[11] and has been used extensively in both modelling the

powerline noise [2] and the underwater communications noise [1]. In many practical cases, a small number K of Gaussian components (e.g., 2 or 3) are sufficient to accurately model the noise without overfitting. In general we have

$$GM(\underline{\pi}, \underline{\sigma}^2) = \sum_{k=1}^{K} \pi_k \cdot \frac{1}{\sqrt{2\pi\sigma_k^2}} e^{-\frac{y^2}{2\sigma_k^2}} \quad (2)$$

where $\underline{\pi} = \{\pi_k\}_{k=1}^{K}$, are mixing coefficients of each Gaussian density and $\sum_{k=1}^{K} \pi_k = 1$. The variances of the Gaussian pdf's are $\underline{\sigma}^2 = \{\sigma_k^2\}_{k=1}^{K}$, and we assume that $\sigma_k^2 > \sigma_{k-1}^2 > \sigma_{k-2}^2 > ... > \sigma_1^2$.

III. ENVELOPE GAUSSIAN MIXTURE

Let us now consider the noise envelope random variable X as a function of the in-phase and quadrature noise components Y and Z. We start with the case where $f_Y(y)$ and $f_Z(z)$ are two-term Gaussian mixtures as in (1) and Y and Z are i.i.d. random variables. Then X can be expressed as:

$$X = \sqrt{Y^2 + Z^2} \quad (3)$$

Given the joint pdf $f_{Y,Z}(y,z)$, the cumulative density function (cdf), $F_X(x)$, is defined as:

$$F_X(x) = \iint_{\sqrt{Y^2+Z^2} \leq x} f_{Y,Z}(y,z)dydz \quad (4)$$

We can then find $f_X(x)$ by differentiating $F_X(x)$ directly using differention rule due to Leibnitz [12], we have:

$$f_X(x) = \int_{-x}^{x} \frac{x}{\sqrt{x^2-z^2}} \Big(f_{Y,Z}(\sqrt{x^2-z^2}, z) $$
$$+ f_{Y,Z}(-\sqrt{x^2-z^2}, z) \Big) dz \quad (5)$$

Since Y and Z are assumed to be independent, the two terms $f_{Y,Z}(\sqrt{x^2-z^2}, z)$ and $f_{Y,Z}(-\sqrt{x^2-z^2}, z)$ will result in:

$$f_{Y,Z}(\sqrt{x^2-z^2}, z) = f_{Y,Z}(-\sqrt{x^2-z^2}, z)$$
$$= \frac{a^2}{2\pi\sigma_1^2} e^{-\frac{x^2-z^2+z^2}{2\sigma_1^2}} + \frac{a(1-a)}{2\pi\sigma_1\sigma_2} e^{-\frac{x^2-z^2}{2\sigma_1^2}-\frac{z^2}{2\sigma_2^2}}$$
$$+ \frac{a(1-a)}{2\pi\sigma_1\sigma_2} e^{-\frac{x^2-z^2}{2\sigma_2^2}-\frac{z^2}{2\sigma_1^2}} + \frac{(1-a)^2}{2\pi\sigma_2^2} e^{-\frac{x^2-z^2+z^2}{2\sigma_2^2}} \quad (6)$$

Substituting (6) into (5), yields:

$$f_X(x) = \int_{-x}^{x} \frac{2x}{\sqrt{x^2-z^2}} \Bigg[\frac{a^2}{2\pi\sigma_1^2} e^{-\frac{x^2}{2\sigma_1^2}} + \frac{(1-a)^2}{2\pi\sigma_2^2} e^{-\frac{x^2}{2\sigma_2^2}}$$
$$+ \frac{a(1-a)}{2\pi\sigma_1\sigma_2} \Big(e^{-\frac{x^2-z^2}{2\sigma_1^2}-\frac{z^2}{2\sigma_2^2}} + e^{-\frac{x^2-z^2}{2\sigma_2^2}-\frac{z^2}{2\sigma_1^2}} \Big) \Bigg] dy \quad (7)$$

The integral (7) can be computed in closed form by letting $z = x\sin(\theta)$ and $dz = x\cos(\theta)d\theta$ to yield:

$$f_X(x) = \frac{a^2 x}{\sigma_1^2} e^{-\frac{x^2}{2\sigma_1^2}} + \frac{(1-a)^2 x}{\sigma_2^2} e^{-\frac{x^2}{2\sigma_2^2}}$$
$$+ \frac{2a(1-a)x}{\sigma_1\sigma_2} e^{\left(-\frac{1}{4}\left(\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}\right)x^2\right)} I_0\left(\frac{1}{4}\left|\frac{1}{\sigma_2^2}-\frac{1}{\sigma_1^2}\right|x^2\right) \quad (8)$$

where

$$I_0(\eta) = \frac{1}{\pi}\int_0^\pi e^{\eta\cos(\theta)}d\theta$$

is the modified Bessel function of the first kind of *zero*-th order. In (8), first two terms follow a Rayleigh distribution. The term with modified Bessel function distinguished the envelope distribution of two-term Gaussian mixture from the two-term Rayleigh mixture. The envelope of two-term Gaussian mixture is shown in Fig. 1 for different values of a and and the ratio of two variances $c = \sigma_2^2/\sigma_1^2 = 10$. One can observe that when the value of mixing coefficient $a = 0.7$, the tail of the pdf is approximately linearly decaying from the peak value (dashed line in Fig. 1). When the ratio of the two variances $\sigma_2^2/\sigma_1^2 \to 1$, similar to the case of a two-term Rayleigh mixture model, the envelope of two-term Gaussian mixture (8) turns into a single Rayleigh distribution. The effect of varying the ratio of σ_2^2/σ_1^2 is illustrated in Fig. 2, in which the dashed grey line is the Rayleigh pdf with variance $\sigma^2 = 1$. However, when $\sigma_2^2 \gg \sigma_1^2$, the envelope Gaussian mixture is significantly different from the Rayleigh mixture as shown in Fig. 3. The two-term Rayleigh mixture model often exhibits a more pronounced muiti-modal behaviour than the envelope of two-term Gaussian mixture.

Assuming that the in-phase and quadrature components are not identically distributed and follow different GMM with K and L components, i.e,

$$f_Y(y) = \sum_{i=1}^{K} [\pi_y]_i \cdot \frac{1}{\sqrt{2\pi[\sigma_y^2]_i}} e^{-\frac{y^2}{2[\sigma_y^2]_i}} \quad (9)$$

$$f_Z(z) = \sum_{k=1}^{L} [\pi_z]_k \cdot \frac{1}{\sqrt{2\pi[\sigma_z^2]_k}} e^{-\frac{y^2}{2[\sigma_z^2]_k}} \quad (10)$$

then a more general envelope Gaussian mixture distribution $f_X(x)$ can be written as:

$$f_X(x) = \sum_{k=1}^{L}\sum_{i=1}^{K} \Bigg\{ \frac{[\pi_y]_i \cdot [\pi_z]_k \cdot x}{[\sigma_y]_i[\sigma_z]_k} \cdot e^{\left(-\frac{1}{4}\left(\frac{1}{[\sigma_z^2]_k}+\frac{1}{[\sigma_y^2]_i}\right)x^2\right)}$$
$$\cdot I_0\left(\frac{1}{4}\left|\frac{1}{[\sigma_z^2]_k}-\frac{1}{[\sigma_y^2]_i}\right|x^2\right) \Bigg\} \quad (11)$$

However, in many practical cases, the in-phase and quadrature noise components can be described by the same distribution with the same parameters, and the envelope Gaussian mixture density simplifies to:

$$f_X(x) = \sum_{k=1}^{K}\sum_{i=1}^{K} \Bigg\{ \frac{\pi_i\pi_k x}{\sigma_i\sigma_k} \cdot e^{\left(-\frac{1}{4}\left(\frac{1}{\sigma_k^2}+\frac{1}{\sigma_i^2}\right)x^2\right)}$$
$$\cdot I_0\left(\frac{1}{4}\left|\frac{1}{\sigma_k^2}-\frac{1}{\sigma_i^2}\right|x^2\right) \Bigg\} \quad (12)$$

IV. MAXIMUM LIKELIHOOD ESTIMATION VIA THE EM ALGORITHM

In this section, we adopt the well-known two-step iterative method called the EM Algorithm that finds the maximum likelihood or maximum a posteriori estimates of parameters
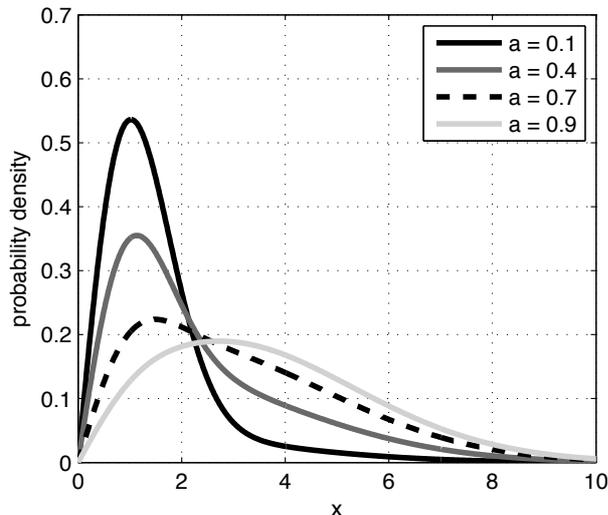
Fig. 1. Envelope of two-term Gaussian mixture pdf with mixing coefficient $0.1 \leq a \leq 0.9$ and $c = \sigma_2^2/\sigma_1^2 = 10$.
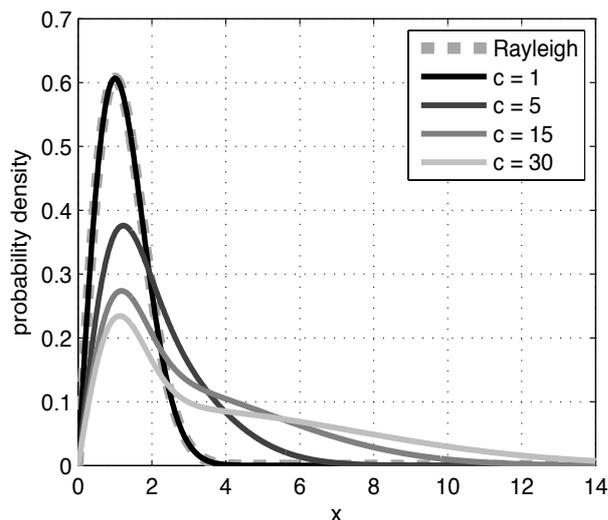
Fig. 2. Envelope of two-term Gaussian mixture pdf with mixing coefficient $a = 0.5$ and $c = \sigma_2^2/\sigma_1^2 = 1, 5, 15, 30$. The envelope Gaussian mixture converges to the Rayleigh distribution as $c \to 1$ (dashed grey curve).

in statistical models in which observations are treated as "incomplete data" [13]. Parameters of mixture densities, such as the Rayleigh mixture model, the Middleton's Class A model and the Gaussian mixture can be estimated by using the EM algorithm. Hence it is natural to predict that parameters of the envelope Gaussian mixture density can be estimated by using the same algorithm.1. The EM algorithm for the envelope of Gaussian mixture model is presented in Appendix A.

Note that in the previous section, Y and Z are used to represent the in-phase and quadrature noise amplitudes. Here, they will be used as the latent variables [13] in the EM algorithm. Given a data set $X = \{x_1, ..., x_N\}$, we assume that all data samples are i.i.d.. Let $p(x|\theta)$ be the pdf that is
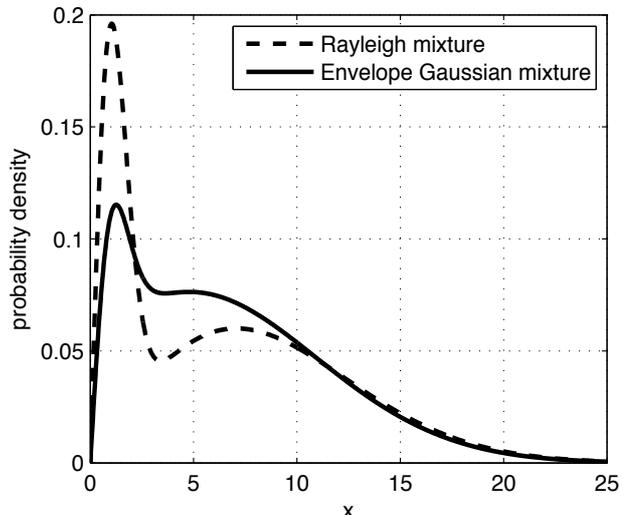
Fig. 3. Comparison of envelope of two-term Gaussian mixture pdf with two-term Rayleigh mixture both with mixing coefficient $a = 0.7$ and $c = \sigma_2^2/\sigma_1^2 = 50$.

governed by the set of parameters, θ, to be estimated. We have:

$$p(X|\theta) = \prod_{n=1}^{N} p(x_n|\theta)$$

$$\mathcal{L}(\theta|X) = \log p(X|\theta) = \sum_{n=1}^{N} \log p(x_n|\theta) \quad (13)$$

where $\mathcal{L}(\theta|X)$ is the log likelihood function. Our purpose is to find the parameters θ which maximize $\mathcal{L}(\theta|X)$ such that:

$$\hat{\theta}_{ML} = \arg \max_{\theta} \mathcal{L}(\theta|X) \quad (14)$$

Under the assumption that the envelope Gaussian mixture model is taken, $p(x_n|\theta)$ is replaced with (12) and the incomplete data log likelihood function is given by:

$$\mathcal{L}(\theta|X) = \sum_{n=1}^{N} \log \sum_{k=1}^{K} \sum_{i=1}^{K} \left\{ \frac{\pi_i \pi_k x_n}{\sigma_i \sigma_k} \cdot e^{\left(-\frac{1}{4} \left(\frac{1}{\sigma_k^2} + \frac{1}{\sigma_i^2} \right) x_n^2 \right)} \right.$$
$$\left. \cdot I_0 \left(\frac{1}{4} \left| \frac{1}{\sigma_k^2} - \frac{1}{\sigma_i^2} \right| x_n^2 \right) \right\} \quad (15)$$

In order to solve this equation, the EM algorithm is utilized by introducing the latent variables. However, unlike clustering problems in Gaussian mixture or Rayleigh mixture in which only one set of latent variables is introduced, here we employ two sets of latent variables $Y = \{y_i\}_{i=1}^{K}$ and $Z = \{z_k\}_{k=1}^{K}$ as binary indicator variables (i.e. $y_i \in \{0, 1\}$, $z_k \in \{0, 1\}$, $\sum_i y_i = 1$ and $\sum_k z_k = 1$). The value y_i indicates which Gaussian component in (9) generates the i-th in-phase noise sample and similarly z_k for the quadrature noise samples. It is important to note that since in-phase noise samples and quadrature noise samples are assumed to be independently generated, y_i and z_k are also independent. The product of these two latent variables, $y_i \cdot z_k$, is binary and forms a 2-dimensional indicator function (i.e. $y_i \cdot z_k \in \{0, 1\}$ and $\sum_i \sum_k y_i \cdot z_k = 1$).

We use the EM algorithm to estimate the parameters $\theta = \{\underline{\pi}, \underline{\sigma}^2\}$. The *expectation* step and *maximization* step are defined as follows:

- *E*-step: Compute $\mathcal{Q}(\theta|\theta^{(p)}) \triangleq \mathbb{E}[\mathcal{L}(\theta|X)|Y, Z, \theta^{(p)}]$
- *M*-step: Determine $\theta = \theta^{(p+1)}$ maximizing $\mathcal{Q}(\theta|\theta^{(p)})$

where $\theta^{(p)}$ is the estimation of θ at p-th iteration of the EM algorithm. We call X the 'incomplete data' and we assume that the complete data $\mathcal{S} = (X, Y, Z)$ includes the binary latent variables Y and Z. Then the joint density function $p(x, y, z)$ is:

$$p(x, y, z) = p(y, z)p(x|y, z) = p(y)p(z)p(x|y, z) \quad (16)$$

since Y and Z are independent. The proportion of the noise samples that are generated by k-th or i-th Gaussian component is π_k or π_i and therefore the joint distributions over Y and Z are specified in terms of the mixing coefficient π_i and π_k, such that $p(y_i = 1) = \pi_i$ and $p(z_k = 1) = \pi_k$. Both y_i and z_k are indicator variables, therefore we can write joint distribution in the following form:

$$p(y, z) = \prod_{i=1}^{K} \prod_{k=1}^{K} [\pi_i \pi_k]^{y_i \cdot z_k} \quad (17)$$

Similarly, the conditional distribution of X given particular values for Y and Z is the envelope Gaussian mixture component that is:

$$p(x|y, z) = \prod_{i=1}^{K} \prod_{k=1}^{K} [p(x|\sigma_i^2, \sigma_k^2)]^{y_i \cdot z_k} \quad (18)$$

Then we have:

$$
\begin{aligned}
p(x) &= \sum_{y} \sum_{z} p(y, z)p(x|y, z) \\
&= \prod_{i=1}^{K} \prod_{k=1}^{K} [\pi_i \pi_k \cdot p(x|\sigma_i^2, \sigma_k^2)]^{y_i \cdot z_k} \quad (19)
\end{aligned}
$$

This is an equivalent formulation of the mixture model involving two explicit latent variables. By dealing with the complete observation X, Y and Z, we can simplify the log likelihood function using (19). Replacing $p(x_n|\theta)$ in (15) with the equivalent formulation found in (19), we have:

$$
\begin{aligned}
\mathcal{L}(\theta|X) &= \sum_{n=1}^{N} \sum_{i=1}^{K} \sum_{k=1}^{K} y_i \cdot z_k \{ \log \pi_i + \log \pi_k \\
&\quad + \log p(x_n|\sigma_i^2, \sigma_k^2) \} \quad (20)
\end{aligned}
$$

Using (20), the $\mathcal{Q}$ function, becomes:

$$
\begin{aligned}
\mathcal{Q}(\theta|\theta^{(p)}) &\triangleq \mathbb{E}[\mathcal{L}(\theta|X)|Y, Z, \theta^{(p)}] \\
&= \sum_{n=1}^{N} \sum_{i=1}^{K} \sum_{k=1}^{K} \mathbb{E}[y_i^{(n)} \cdot z_k^{(n)}] \{ \log \pi_i + \log \pi_k \\
&\quad + \log p(x_n|\sigma_i^2, \sigma_k^2) \} \quad (21)
\end{aligned}
$$

where $\mathbb{E}[y_i^{(n)} \cdot z_k^{(n)}] = \gamma_{n,i,k}$ is the conditional probability of X given Y and Z, we also call it "soft" assignment (or

responsibility) which can be found by using Bayes' theorem:

$$
\begin{aligned}
\gamma_{n,i,k} &= \mathbb{E}[y_i^{(n)} \cdot z_k^{(n)}] = p(y_i = 1, z_k = 1|x_n) \\
&= \frac{p(y_i = 1)p(z_k = 1)p(x|y_i = 1, z_k = 1)}{p(x_n)} \\
&= \frac{\pi_i \pi_k p(x_n|\sigma_i^2, \sigma_k^2)}{\sum_{s=1}^{K} \sum_{t=1}^{K} \pi_s \pi_t p(x_n|\sigma_s^2, \sigma_t^2)} \quad (22)
\end{aligned}
$$

The *maximization* step of the EM algorithm finds the expression for π_i and π_k. We introduce the Lagrange multiplier λ with the constraint $\sum_i \pi_i = \sum_k \pi_k = 1$ and solve the following set of equations:

$$\frac{\partial}{\partial \pi_k} \left\{ \sum_{n=1}^{N} \sum_{i=1}^{K} \sum_{k=1}^{K} \gamma_{n,i,k} \log \pi_k + \lambda \left(\sum_k \pi_k - 1 \right) \right\} = 0 \quad (23)$$

for $k = 1, \ldots, K$. Summing left side over i or k gives $\lambda = -N$ which results in:

$$\pi_k = \frac{1}{N} \sum_{n=1}^{N} \sum_{i=1}^{K} \gamma_{n,i,k} \quad (24)$$

The variances σ_k^2 and σ_i^2 can be found by solving the set of K equations:

$$
\begin{aligned}
\frac{\partial}{\partial \sigma_k^2} \Bigg\{ &\sum_{n=1}^{N} \sum_{i=1}^{K} \sum_{k=1}^{K} \gamma_{n,i,k} \bigg[\log \frac{x_n}{\sigma_i \sigma_k} - \frac{1}{4} \left(\frac{1}{\sigma_k^2} + \frac{1}{\sigma_i^2} \right) x_n^2 \\
&+ \log I_0 \left(\frac{1}{4} \left| \frac{1}{\sigma_k^2} - \frac{1}{\sigma_i^2} \right| x_n^2 \right) \bigg] \Bigg\} = 0 \quad k = 1, \ldots, K (25)
\end{aligned}
$$

We get:

$$\sigma_k^2 = \frac{\sum_{n=1}^{N} \sum_{i=1}^{K} \gamma_{n,i,k} \frac{x_n}{2} (1 - \phi(\frac{1}{4} | \frac{1}{\sigma_k^2} - \frac{1}{\sigma_i^2} | x_n^2))}{\sum_{n=1}^{N} \sum_{i=1}^{K} \gamma_{n,i,k}} \quad (26)$$

where $\phi(\cdot) = \frac{I_1(\cdot)}{I_0(\cdot)}$ and $I_1(\cdot)$ is the modified Bessel function of first kind of *first* order. Equations (26) are non-linear equations and therefore can only be solved numerically.

V. SIMULATION RESULTS

A simulation of the proposed estimator via the EM algorithm is performed, in which we consider an envelope Gaussian mixture model with unknown variances and mixing coefficients, however we assume that the number of components is known. Random data samples are randomly generated from the distribution with known parameters. An alternative method in estimating parameters of the envelope Gaussian mixture apart from the EM algorithm is the quasi-Newton method (See [1] for detail). The performance of the EM algorithm will be compared with the performance of the quasi-Newton method with the BGFS step update for the envelope Gaussian mixture model [14]. The EM algorithm and the quasi-Newton method will terminate when the change in log likelihood function is less than stopping criterion, ϵ. The EM algorithm is more often used in parameter estimation of the curved exponential families, since the quasi-Newton method is more complicated to implement. On the other hand, the EM algorithm enjoys greater simplicity and stability but has

slower convergence speed. Here we focus on comparing their convergence speed and accuracy in terms of iteration numbers and Mean Square Error (MSE) between the estimated pdf and normalised data histograms.

Our first example takes 50000 data samples which are generated by the envelope of a two-term Gaussian mixture distribution with parameters $\pi_1 = 0.3$, $\pi_2 = 0.7$, $\sigma_1^2 = 1$ and $\sigma_2^2 = 10$. Fig. 4 illustrates the normalised data histogram with $N = 25$ bins of generated data samples, together with the estimated envelope Gaussian mixture pdf and true envelope Gaussian mixture pdf. The MSE between the envelope Gaussian mixture pdf and the normalised data histogram is defined as follows:

$$\mathbf{MSE} = \frac{1}{N}\sum_{i=1}^{N}(\hat{Y}_i - Y_i)^2 \qquad (27)$$

where Y_i is the frequency density of the i-th bin and $\hat{Y}_i$ is the estimated probability density taken at the midpoint of i-th bin. The MSE between the estimated pdf and histogram is 3.12779×10^{-4} and the parameters estimated are $\hat{\pi}_1 = 0.303502$, $\hat{\pi}_2 = 0.696498$, $\hat{\sigma}_1^2 = 1.04399$ and $\hat{\sigma}_2^2 = 9.946461$, which show that the ML estimator via the EM algorithm yields good performance. Example 2 shown in Fig. 5 also illustrates that parameters of the envelope Gaussian mixture model have been correctly estimated by the EM algorithm. In fact, for examples with distributions that are well separated (i.e. $\sigma_2^2 \gg \sigma_1^2$), both the EM algorithm and the quasi-Newton method perform well in terms of reproducing the true parameters. However, in some ill-conditioned cases (i.e. example 4, 5, 7 and 8) where too many components have been included in the model due to overfitting or to similar values of variances, both the EM algorithm and the quasi-Newton method perform poorly in terms of attaining the true values of parameters as shown in TABLE I. In addition, example 6 does not converge to the true value due to small sample size. Fortunately, both methods may achieve some local optima of the log likelihood functions in all cases, therefore the estimated envelope Gaussian mixture pdf's using both methods still exhibit excellent agreement with the histograms.

Despite that the EM algorithm performs well in terms of maximizing the log likelihood function, it has relatively slower convergence rate depending on the models and data size [14]. If we only compare the numbers of iterations to convergence, quasi-Newton would take fewer iterations, since the convergence speed of the quasi-Newton algorithm is super linear, whereas the EM algorithm converges linearly [14]. TABLE I shows that the computation iterations for estimation of the parameters of the envelope of two-term Gaussian mixture model with stopping criterion, $\epsilon = 10^{-6}$. It is clear in the TABLE I that the quasi-Newton method converges much faster than the EM algorithm. The quasi-Newton method, on average, took less than $1/8$ in iteration numbers of the EM algorithm. For example 4, 6 and 7, number of iteration for the EM algorithm can go over 1000. Moreover, the computation time per iteration of the EM algorithm is
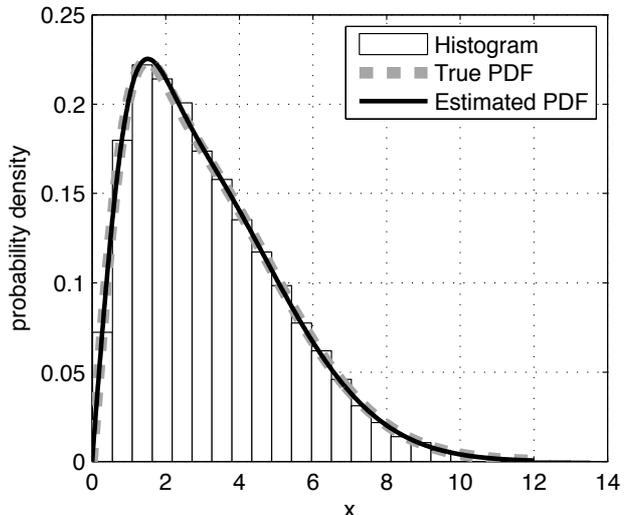
Fig. 4. Comparison of Normalised data histogram, true envelope Gaussian mixture pdf and estimated envelope Gaussian mixture pdf.
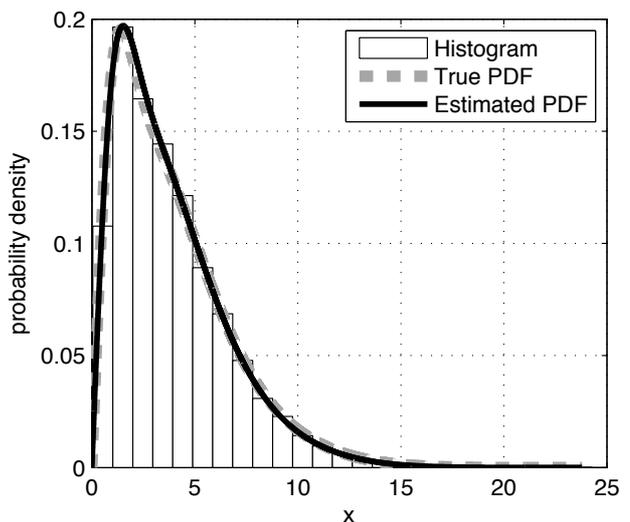
Fig. 5. Normalised data histogram generated with mixing coefficient $\pi_1 = 0.3$, $\pi_2 = 0.4$, $\pi_3 = 0.3$, $\sigma_1^2 = 1$, $\sigma_2^2 = 9$ and $\sigma_3^2 = 25$. Stopping criterion is set to be 10^{-6}. Number of iteration taken for the EM algorithm to converge is 1491. Parameters estimated are $\hat{\pi}_1 = 0.301052$, $\hat{\pi}_2 = 0.406479$, $\hat{\pi}_3 = 0.292469$, $\hat{\sigma}_1^2 = 1.109467$, $\hat{\sigma}_2^2 = 8.970086$ and $\hat{\sigma}_3^2 = 24.172147$.

much longer than that of the quasi-Newton method, since the variances, σ_k^2 and σ_i^2 in (26) have to be found numerically in each EM iteration. For this reason, we conclude that, in terms of iteration numbers, the quasi-Newton method should be preferred to the EM algorithm in estimation parameters of the envelope Gaussian mixture density functions. However, reader should take note that the EM algorithm is still an attractive method in this estimation problem due to greater simplicity (i.e. automatic satisfaction of probability constraints and monotonic convergence without the need to set a step size). On the other hand, implementation of the quasi-Newton method is indeed complicated. As Jamshidian and Jennrich

TABLE I
COMPARISON OF THE PERFORMANCE OF THE EM ALGORITHM WITH THE PERFORMANCE OF THE QUASI-NEWTON METHOD FOR THE ENVELOPE GAUSSIAN MIXTURE MODEL

#	True Parameters						Estimated parameters via EM						Estimated parameters via QN						EM	QN
	π_1	π_2	π_3	σ_1^2	σ_2^2	σ_3^2	π_1	π_2	π_3	σ_1^2	σ_2^2	σ_3^2	π_1	π_2	π_3	σ_1^2	σ_2^2	σ_3^2	Iter	Iter
1	0.3	0.7	-	1	10	-	0.29992	0.70008	-	1.01055	10.00381	-	0.29998	0.70002	-	1.01863	10.00449	-	168	20
2	0.95	0.05	-	1	10	-	0.94084	0.05916	-	0.97562	9.37184	-	0.94085	0.05915	-	0.97564	9.37315	-	145	22
3	0.5	0.5	-	1	10	-	0.50771	0.49229	-	1.00778	10.04186	-	0.50775	0.49225	-	1.00779	10.04249	-	121	19
4	0.5	0.5	-	1	1.5	-	0.40913	0.59087	-	1.01156	1.40103	-	0.365771	0.634229	-	1.10931	1.53166	-	1000 <	21
5	0.95	0.05	-	1	1.5	-	0.25011	0.74989	-	0.83423	1.09010	-	0.36550	0.63450	-	0.95190	1.18866	-	103	56
6	0.3	0.4	0.3	1	9	25	0.30305	0.42639	0.27056	1.03800	9.49385	26.098378	0.30265	0.42510	0.27225	1.03661	9.45800	26.04114	1000 <	63
7	0.4	0.4	0.2	1	2	25	0.40117	0.40481	0.19402	0.94892	2.11022	26.19052	0.27070	0.53347	0.19583	0.79183	1.87935	26.03970	1000 <	74
8	0.2	0.4	0.4	1	9	10	0.18559	0.33575	0.47866	0.80414	9.45651	9.45735	0.18555	0.65777	0.15668	0.80420	9.45740	9.45147	196	56

[15] have pointed out, to choose between the EM algorithm and the quasi-Newton method is more or less a personal choice.

VI. CONCLUSION

In this paper, we have derived the general expression for the envelope Gaussian mixture model. Such model is different from the Rayleigh mixture and may be applied to describe the envelope of powerline noise and the envelope of underwater communication noise. We proposed the EM algorithm to estimate the parameters of the envelope Gaussian mixture. Finally, we discussed the convergence speed and accuracy of the EM algorithm by simulations. This was compared with the quasi-Newton algorithm.

ACKNOWLEDGEMENT

REFERENCES

[1] F. Sari, N. Sari, and L. Mili, "Modelling of sea clutter with Gaussian mixtures and estimation of the clutter parameter," *Signal Processing and Communications Applications Conf. (SPCAC 2004)*, Proc. vol.12, pp. 53–56, Istanbul, Turkey.

[2] M. Arzberger, K. Dostert, T. Waldeck, and M. Zimmermann, "Fundamental properties of the low voltage power distribution grid," in *Proc. 1st Int. Symp. Power-Line Communications and its Applications (ISPLC 1997)*, Mar. 1997, pp. 45–50, Essen, Germany.

[3] D. Middleton, "Statistical-physical models of urban radio-noise environments-Part 1: Foundations," *IEEE Trans. Electromagn. Compat.*, vol. EMC-14, no. 1, pp. 38–56, May 1972.

[4] H. C. Ferreira, L. Lampe, J. Newbury, T. G. Swart, *Power Line Communications: Theory and Applications for Narrowband and Broadband Communications over Power Lines*, John Wiley and Sons Ltd, 2011.

[5] A. P. Dempster, N. M. Laird, and D. B. Rdin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc., Ser. B*, vol. 39, pp. 1–38, 1977.

[6] C.F. J. Wu, "On the convergence properties of the EM algorithm," *The Annals of Statistics*, vol. 11, no. 1, pp. 95–103, 1983.

[7] S. M. Zabin and H. V. Poor, "Efficient estimation of Class A noise parameters via the EM algorithm," *IEEE Trans. Inf. Theory*, vol. 37, no. 1, pp. 60–72, Jan. 1991.

[8] J. C. Seabra, F. Ciompi, O. Pujol, J. Mauri, P. Radeva, and J. Sanches, "Rayleigh mixture model for plaque characterization in intravascular ultrasound," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 5, pp. 1314–1324, May 2011.

[9] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford Univ. Press, 1995.

[10] D. Middleton, "Statistical-physical models of electromagnetic interference," *IEEE Trans. Electromagn. Compat.*, vol. EMC-19, no. 3, pp. 106–127, Aug. 1977.

[11] D. Middleton, "Canonical and quasi-canonical probability models of Class A interference," *IEEE Trans. Electromagn. Compat.*, vol. EMC-25, no. 2, pp. 76–106, May 1983.

[12] A. Papoulis, *Probability, Random Variables and Stochastic Processes*. New York: McGraw-Bill, 1965.

[13] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, Aug. 2006.

[14] M. Watanabe and K. Yamaguchi, *The EM Algorithm and Related Statistical Models*, Evanston, IL: Routledge, 2004, ch. 7.

[15] J. M. Jamshidian and R. I. Jennrich, "Acceleration of the EM Algorithm by using Quasi-Newton Methods," *J. R. Statistics Soc.*, vol. B59, pp. 569–587, 1997.

APPENDIX

A. EM algorithm for the Envelope Gaussian Mixture Model

1. Initialize the mixing coefficients π_i and π_k, together with variances σ_i^2 and σ_k^2 with random values.

2. **E step**: Given the data sample x_n, evaluate the responsibility using the current parameters π and σ^2 for all i and k from 1 to K

$$
\begin{aligned}
\gamma_{n,i,k} &= \mathbb{E}[y_i^{(n)} \cdot z_k^{(n)}] = p(y_i = 1, z_k = 1 | x_n) \\
&= \frac{\pi_i \pi_k p(x_n | \sigma_i^2, \sigma_k^2)}{\sum_{s=1}^K \sum_{t=1}^K \pi_s \pi_t p(x_n | \sigma_s^2, \sigma_t^2)}
\end{aligned} \tag{28}
$$

3. **M step**: Update the parameters using the current responsibility

$$
\pi_k = \frac{1}{N} \sum_{n=1}^N \sum_{i=1}^K \gamma_{n,i,k} \tag{29}
$$

$$
\sigma_k^2 = \frac{\sum_{n=1}^N \sum_{i=1}^K \gamma_{n,i,k} \frac{x_n}{2} \left(1 - \phi\left(\frac{1}{4}\left|\frac{1}{\sigma_k^2} - \frac{1}{\sigma_i^2}\right| x_n^2\right)\right)}{\sum_{n=1}^N \sum_{i=1}^K \gamma_{n,i,k}} \tag{30}
$$

for all k from 1 to K. $\phi(\cdot) = \frac{I_1(\cdot)}{I_0(\cdot)}$ and $I_1(\cdot)$ is the modified Bessel function of first kind of *first* order. Equation (30) are non-linear equations and therefore can only be solved numerically.

4. Evaluate the likelihood function

$$
\mathcal{L}(\theta | X) = \sum_{n=1}^N \log \left\{ \sum_{i=1}^K \sum_{k=1}^K \pi_k \pi_i \cdot p(x_n | \sigma_k^2, \sigma_i^2) \right\} \tag{31}
$$

and check the convergence of either the parameters or the likelihood function. If the convergence criterion is not satisfied, return to Step 2 (**E step**).