MA 690 Project 2

Project Statement: This project will focus on **Rayleigh Mixture Models (RMMs)**, which are mixture models in which the Rayleigh distribution is used (as opposed to the Gaussian distribution which we covered in class). Recently, there has been a surge of research which utilizes RMMs, including problems related to medical imaging, speech enhancement algorithms, and synthetic aperture radar imaging. Recall that the Rayleigh distribution is the distribution defined by the probability density function

$$f(x \mid \sigma) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right)$$

for $x \geq 0$ and $\sigma > 0$. Provide solutions for the following problems:

1. (**MATLAB**) Write code which will generate samples from an RMM with an *arbitrary* number of components. In order to check that you have generated samples correctly, generate a figure with a normalized histogram of the samples and the pdf of the prescribed mixture. You may use the MATLAB function `raylrnd` to generate the samples for a Rayleigh distribution and `raylpdf` for the probability density function of a Rayleigh distribution. **Hint:** If the vector which contains your probabilities is p and the vector which contains the parameters for each mixture component is called *params*, then the code `pdf = @(input) arrayfun(@(x) dot(p, raylpdf(x, params)), input)` is a vectorized version of the pdf of the given mixture model. Furthermore, you may find the code `latent_variable = sum(rand(1E5, 1) > cumsum(p), 2) + 1` useful.

2. (**By Hand**) Write down the quantity which one should maximize in order to perform the Expectation Maximization algorithm for an RMM. Then find a recursive formula for the parameters of each individual mixture. **Hint:** It suffices to find the formula for a single component of the mixture.

3. (**MATLAB**) Generate samples from a Rayleigh mixture model with some prescribed parameters (using your code from Part 1) and perform the expectation maximization algorithm in order to recover those parameters. The difficulty here lies in making the code work for an arbitrary number of mixtures. **Some Hints**:

   (a) You will have a for loop over the number of iterations of the EM algorithm you want to perform.

   (b) Inside that loop, you will have three additional loops all of which loop over the number of mixtures in your mixture model: one to compute $w_i^{(n)}$; one to compute γ_i, N_i, and your estimated parameters σ_i; and one to compute the updated probabilities of the mixture.

4. (**MATLAB**) For the last part of the project, download *project2_data.mat* from the CANVAS site. Write a script that will take the data from this file and perform the EM algorithm on it assuming that the data comes from a Rayleigh mixture model. Your code should output the following

   (a) The parameters of the RMM assuming that there is one mixture component,

   (b) The parameters of the RMM assuming that there are two mixture components,

   (c) The parameters of the RMM assuming that there are three mixture components,

   (d) The parameters of the RMM assuming that there are four mixture components,

   (e) The parameters of the RMM assuming that there are five mixture components,

   (f) The parameters of the RMM assuming that there are six mixture components.

   Furthermore, your code should display a figure with the normalized histogram of the data from the file along with the pdfs of each of the RMMs calculated above. Add a legend to the plot so that I know which line corresponds to which mixture model. **Note:** Your graphs will be dependent on the initial choices for the parameters.