AIP Assignment5

Khursheed, Ayush

April 2018

1 Question 3

1.1 Part A) Description of the application

1. Paper presents a methodology for analyzing polyphonic musical passages comprised by notes that exhibit a harmonically fixed spectral profile (such as piano notes).

2. Taking advantage of this unique note structure we can model the audio content of the musical passage by a linear basis transform and use non-negative matrix decomposition methods to estimate the spectral profile and the temporal information of every note. This approach results in a very simple and compact system that is not knowledge-based, but rather learns notes by observation.

3. Paper specifically addresses the problem of Polyphonic music transcription. Paper show that it is possible to efficiently perform polyphonic music transcription using non-negative matrix factorization of musical spectra.

4. Paper suggests that knowledge based approaches are highly complex. As compared to knowledge based approaches there approach a lighter approach and is data driven.

5. Result shown in paper is reasonably encouraging.

1.2 Part B) The main objective function being optimized in the paper

1. Objective Function : Paper uses following two objective functions.
   $C = \| X - W.H \|_F$
   OR
   $D = \| X \otimes ln\frac{X}{W.H} - X + W.H \|_F$
   where,
   X is a M by N non-negative matrix. The goal of NMF is to approximate it as a product of two non-negative matrices $W \in R^{>=0, M \times R}$ and $H \in R^{>=0, R \times N}$ and $R <= M$

$\otimes$ is the Hadamard product (an element-wise multiplication of the matrices),

2. second eqution is somewhat similar to the Kullback-Liebler divergence,

1.3 Part C) Some result from the paper

1. Let X contains musical spectra. Then the elements of W and H will respectively contain the spectrum and the temporal information of the notes in the analyzed passage.

2. Result are shown for following two cases:

   (a) **Isolated notes:** For the first example they consider the few notes(4 to be exact) of the piano. There are no major overlapping note regions.
   They used first cost function for the analysis and they found that with appropriate value of R the rows of H correspond to the temporal activity of the four notes, whereas the columns of W contain their respective spectra.

   (b) **Coinciding notes:** A simple example of polyphonic transcription.Two notes (B_4^b and G_3) sounding at the same time is considered.
   They perform NMF using second objective cost function.
   They see that the two simultaneous notes B_4^b and G_3 were consolidated as one component. The reason for this unexpected result is that a transcription technique is based on the system's accumulated experience from the presented input and not on predefined knowledge. Due to this all unique events are understood to be a new component. **It is important to understand that this method will not extract notes, but rather unique events.** We do not provide enough knowledge to purposely extract notes; but rather the algorithm has to examine the data to discover what appear to be unique events which should correspond to notes. The reason why in this case we end up with two notes being identified as one component is because they always occur at the same time. As far as the data goes the pair of B_4^b and G_3 is a unique event and it should be one component not two. The way to alleviate this problem is to present enough data so that all notes are exposed as either isolated events, or as parts of different polyphonic groups so as to highlight their individuality.