Non-Negative Matrix Factorization as Dictionary Learning for Audio Separation

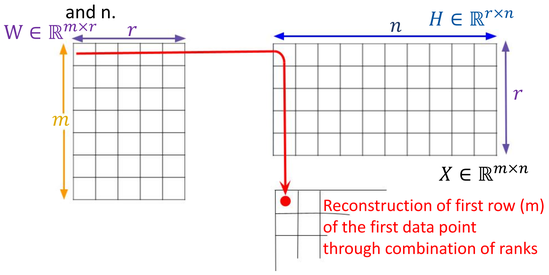
Evan Curtin

UNIVERSITY OF DELAWARE

Introduction

- Different instruments overlap in terms of their frequency even though they sound distinct. Machine learning is an approach to design software to automatically be excellent at identifying patterns. In music, these patterns, rhythmically or tonally (e.g., chord progression), could be simple or complex.

- Nonnegative matrix factorization (NMF) is an algorithm for analyzing nonnegative data. A matrix X of the data is approximated with two low-rank non-negative matrices $W \in \mathbb{R}^{m\times r}$ and $H \in \mathbb{R}^{r\times n}$ where ($X \in \mathbb{R}^{m\times n} \approx WH$). The rank, r, is the number of W's columns and the number of H's rows much less than m and n.



- W is the dictionary or the set of reappearing waveforms like notes or instruments. H is the activations or the timing of those waveforms.

- NMF models vary in the choice of the objective function that assesses the quality of an approximation by evaluating some distance, the error, between W H and X differs.

- The most widely used class of objective functions are component-wise and based on the β-divergences defined as follows: for x, y ∈ R+, $D_\beta(x,y)$

$$= \begin{cases} IS(x,y) for\ \beta = 0, Itakura - Saito\ (IS)\ divergence \\ KL(x,y)\ for\ \beta = 1, Kullback - Leibler\ (KL)\ divergence \\ \frac{1}{2}\|x-y\|_F^2\ for\ \beta = 2, Frobenius\ norm \\ \frac{1}{\beta(\beta-1)}\left(x^\beta + (\beta-1)y^\beta - \beta x y^{\beta-1}\right) for\ \beta \neq 0,1,2 \end{cases}$$

- DR NMF has multiple objective functions where the solution is one with low error for all the objectives, always the largest objective function value.

Purpose & Research Question

- As rank increases, how is the timing and error affected?

- Which NMF has the lowest error?

- **The purpose of this project is to take an audio file then use DR-NMF, training on 80% of the audio file, then recreating the last 20% of the song. Then being able to separate the notes of the music**.

Subjects & Methods

- I used the code of the algorithm described in the paper 'Distributionally Robust and Multi-Objective Nonnegative Matrix Factorization' by N. Gillis, L. T. K. Hien, V. Leplat and V. Y. F. Tan, to generate a graph of the error after 100 iterations from each NMF as the rank of the NMF increased of 'Mary had a little lamb'.

- I plotted the activation timings of the DR-NMF for some of the different ranks.

- Using 80% of the 'Ode to Joy', I used DR-NMF to generate a dictionary. Then I applied matching pursuit to reconstruct the final 20% of the song.

Analysis & Results

A. For IS and KL based NMF minimize their own error very well but at the cost of leaving the other error very high. The DR-NMF minimizes both somewhat equally. [Figure A]

B. The spectrogram of 'Mary had little lamb' shows that there are seven notes played in total as shown in [Figure B] with a temporal resolution of 32ms and a frequency resolution of 31.25Hz.

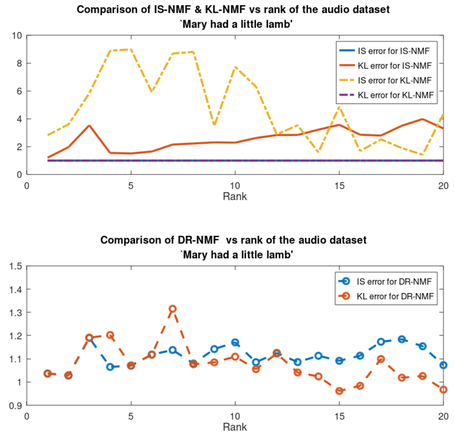
C. The timings of the DR-NMF for ranks from 1-5, 10, 15, and 20 are generated and displayed in [Figure C]. As the rank increases past the number of unique notes and general background noise, the DR-NMF finds more overlapping.

D. The recording of 'Ode to Joy'. is down sampled to fs = 16000Hz yielding T = 842196 samples plotted in [Figure D].

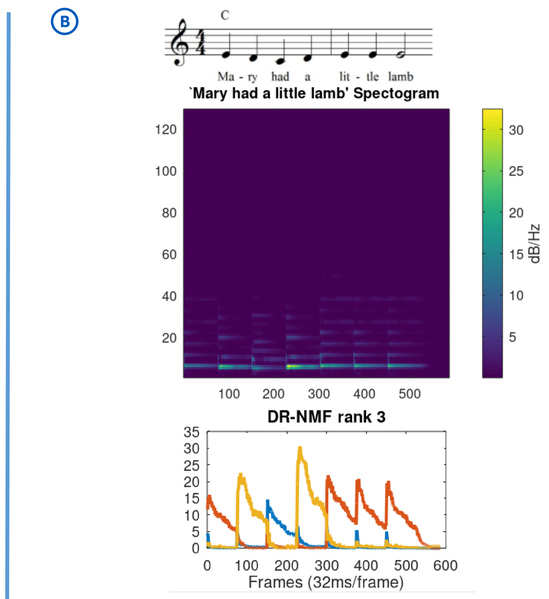
E. The Fast Fourier transform (FFT) of the input x is computed using a Hamming window of size F = 512. The waveform is displayed in log scale in [Figure E].

F. About 80% of the song is used to generate the dictionary from DR-NMF with a rank of seven then matching pursuit is used to reconstruct the last 20% shown in [Figure F]. This reconstruction shows that DR-NMF is great for dictionary learning.

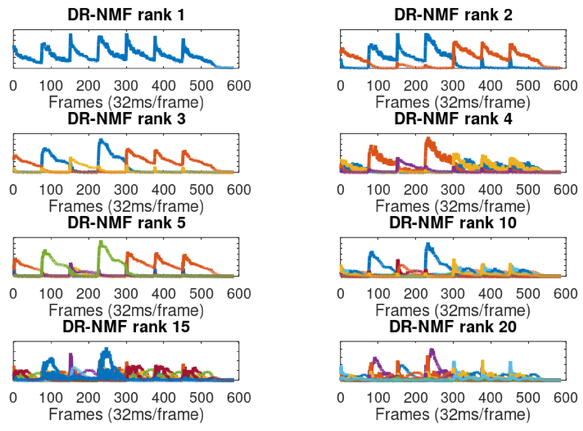
A



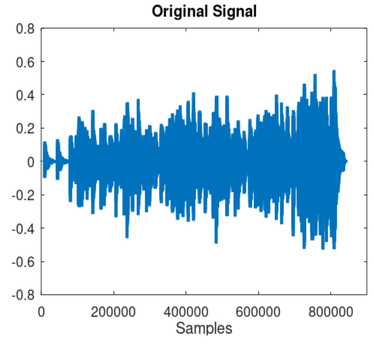
B



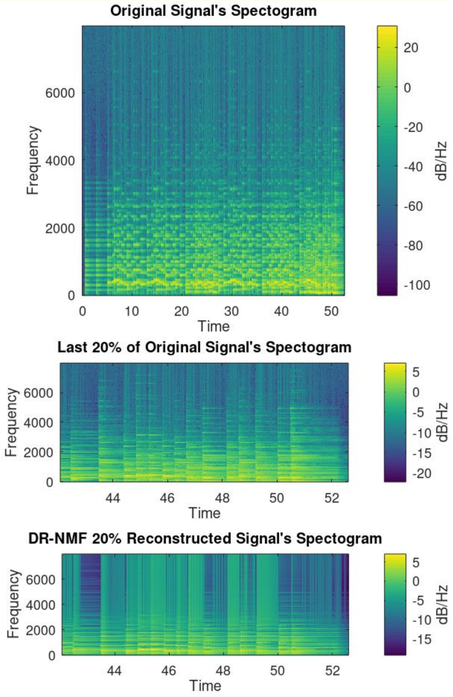
C



D



E



Conclusions

- The resulting reconstructed signal looks very similar to the original signal with the general patterns matching. This showed that the DR-NMF created a dictionary of the waveforms that are almost exact to the originals.

Directions for Future Research

- I will be using the last reconstructed signal and converting it back into time for more comparisons to the original. On top of a way for more in-depth error comparison for higher ranks reconstruction, better reconstructed algorithms, and testing with adding noise to the waveform.

- The goal is to be able to build a more complex dictionary to have more complex audio and separating it into either instruments, voice, and other sounds.

References

N. Gillis, L. T. K. Hien, V. Leplat and V. Y. F. Tan, "*Distributionally Robust and Multi-Objective Nonnegative Matrix Factorization*", IEEE Trans. on Pattern Analysis and Machine Intelligence, 2021.

FreeSheetPianoMusic. (2013). *Beethoven - "Ode to Joy" from Symphony No.9 Easy Piano Version*. *YouTube*. https://www.youtube.com/watch?v=AbXiKLA58SE.