MATRICES IN PHYSICS DATA PROCESSING

[1]TANG WEN KAI ADRIAN, [2]NG WEI SHEAN, [3]LIEW HOW HUI

[1,2,3]Department of Mathematical and Actuarial Sciences, Lee Kong Chian Faculty of Engineering & Science, UniversitiTunku Abdul Rahman, Bandar Sungai Long, Malaysia
E-mail: [1]adriantwk97@1utar.my,[2]ngws@utar.edu.my,[3]liewhh@utar.edu.my

Abstract - Source separation plays an important role in audio processing. Hence, we would like to incorporate matrix factorisation feature extraction properties which are well-known in image processing into audio source separation. Therefore, a comparison between Nonnegative Matrix Factorisation, Convolutive Nonnegative Matrix Factorisation and Interpolative Decomposition in performing source separation and evaluate the performance using the Signal-to-Noise ratio. In the end, we found out that Convolutive Nonnegative Matrix Factorisation with Itakura-Saito divergence has better performance in separating the musical instruments as compared to other methods.

Keywords - Audio Processing, Nonnegative Matrix Factorisation, Convolutive Nonnegative Matrix Factorisation, Interpolative Decomposition

I. INTRODUCTION

With the improvement of knowledge, different fields like physics, mathematics, data analysis and others have incorporated machine learning to solve their problems. Machine learning is important as it is efficient, used in decision making and increases accuracy. An example of machine learning used in physics is N-body simulation [23] which is used in cosmology and astrophysics to compare the simulation result with experimental observations. Besides that, boosted decision trees [5] are used in particle physics for particle identification. Recently, many researchers use the term physics-informed machine learning (PIML) [39] which merges physic-based modelling with machine learning. Examples of PIML are underground reservoir pressure management [10], additive manufacturing [37], improving RANS predicted Reynolds Stresses [24] and others.

Signal processing is a sub-field of electrical engineering. Signal processing involves representing data in a signal form which we are not able to see directly. Such data are images, audio, video and vibrations. The most popular signal processing problems are audio compression [27], speech recognition [36], economic modelling [25], weather forecast [11] and Oceanography [32].

This project studies the method to perform source separation. Source separation techniques can recover one or more original audio signals from the mixed audio signal. Source separation can be represented by the equation below:

$$y(t) = \sum_{j=1}^{N} x_j(t),$$

where $y(t)$ is the mixed audio signal and $x_j(t)$ is the number of sources that exist in the mixed audio signal. There are two approaches to solve source separation problems: multi-channel and single-channel source separation [13]. The main difference between these two approaches is that multi-channel can get information regarding the position of the audio source and the single-channel approach are not able to give such information.

In the past few years, researchers have proposed different methods to solve the source separation problems in both multi-channel and single-channel approaches. The proposed methods by other researchers are deep learning approaches such as Conv-TasNet [38], Wave-U-Net [12] and Multi-scale Multi-band DenseNets[28] and Primal-dual algorithm for Non-negative Matrix Factorisation with Kullback-Leibler divergence [15]. In this paper, we focus on matrix factorisation techniques with single-channel approaches.

II. MATRIX FACTORISATION

In this project, we use three matrix factorisation methods to solve the source separation problem which are Nonnegative Matrix Factorisation (NMF), Convolutive Nonnegative Matrix Factorisation (CNMF) and Interpolative Decomposition (ID).

2.1 Nonnegative Matrix Factorisation (NMF)

Let $A \in \mathbb{R}^{n\times m}$ be nonnegative matrix, then factorise A into nonnegative matrix $W \in \mathbb{R}^{n\times r}$ and nonnegative matrix $H \in \mathbb{R}^{r\times m}$ with rank, r, $0 \le r \le \min(n,m)$ such that:

$$V \approx WH,$$

where W denotes the basis matrix and H denotes the coefficient matrix. There are other variants of NMF that can be obtained in the review paper [40]. As for the algorithm, we are using [14]which is from the scikit-learn algorithm.

2.2 Convolutive Nonnegative Matrix Factorisation (CNMF)

Let $A \in \mathbb{R}^{n\times m}$ benonnegative matrix. CNMF is factorise A into nonnegative matrices $W_1, W_2, \dots, W_L$ with size $n \times r$ and nonnegative matrix H of size $r \times m$ with rank, r, $0 \le r \le \min(n,m)$ such that:

$$X \approx \sum_{t=0}^{L-1} W_t \overrightarrow{H}^{t},$$

where W_t denotes the basis matrix and $\overrightarrow{H}^{t}$ denotes the coefficient matrix shift to the right side by t step, i.e, $\overrightarrow{H}^{0} = H$. We use [29] to calculate the CNMF factorisation.

2.3 Interpolative Decomposition (ID)

Let $A \in \mathbb{R}^{n \times m}$ be an $n \times m$ nonnegative matrix, such that:

$$A = CZ, \quad (1)$$

where the columns of $C \in \mathbb{R}^{n \times r}$ is a subset of the columns of A and $H \in \mathbb{R}^{r \times m}$ such that:

- some size-r subset of the columns of Z form the $r \times r$ identity matrix, and
- all the absolute values in Z not greater than 2.
- rank, r, $0 \leq r \leq \min(n, m)$.

In this paper, we use the random version of the algorithm in [31]. This algorithm can be run only if $n < m$, as for the case where $n > m$, we need to input the transpose matrix of A, A^T into the algorithm and transpose C and Z when the calculation is done.

III. RELATED WORK

Researchers have performed research on matrix factorisation methods used in source separation. More popular matrix factorisation methods are Non-negative Matrix Factorisation (NMF) and these methods have different variants [40]. [8] have proposed NMF using Euclidean distance (Euc) and KullbackLeibler (KL) divergence [22] to calculate the low-rank approximated matrix WH. Then, [7] foundthat EUC and KL divergence does not separate the music audio well, thus the authors proposed Itakura-Saito (IS) divergence [2] to calculate the low-rank approximation matrix WH. This is because of the scale invariance property that the IS divergence processes can allocate the same weight on both low-energy and high-energy components preventing the elimination of low-energy components.

After solving the music source separation in the real number domain, [17] extended the NMF algorithm to solve the complex domain using the combination of NMF and Sparse Coding [30] at the same time. This method can extract recurrent patterns and by fulfilling some conditions multiplicative update rules can update the W and H iterative. Besides that, [26] proposed the positive semidefinite tensor factorisation (PSDTF) which was the natural extensionof NMF with IS divergence as the cost function. These methods can approximate the signals in the time domain but the computational cost is more expensive than the NMF.

The methods mentioned previously are only suitable for solving single-channel data. Therefore, researchers extended the application to multi-channel data where the researchers are interested in determining the source position. The idea of multi-channel processing is taken from [21] as humans and animals have two ears and can exploit the directions and locations of the audio signal. [33] represented the multi-channel data by taking each channel of the spectrogram and stack up into a single matrix. Therefore, [3] used convolution mixing and proposed two types of multi-channel NMF models which are the Expectation-maximisation (EM) algorithm [6] and multiplicative update (MU) algorithm but the computational cost is expensive.

Next, [20] proposed Hermitian positive semidefiniteness of a matrix which is equivalent to non-negativity, thus it can be used as an extension of NMF in solving multi-channel cases in complex domains with EM algorithm as the cost function. The authors also found that the multiplicative updates rules for multi-channel extensions of NMF are similar to single-channel NMF. In the previous work, [18] and [19] only manage to separate two sources and it is difficult to separate when the number of sources is more than two. In the recent paper, [20] managed to separate three sources by using the bottom-up clustering method and a condition where redundant spatial properties are allowed. The authors also mentioned that the multi-channel algorithm has the same effect as compared to the single-channel algorithms in terms of iterations needed for convergence.

The disadvantage of using NMF with a multiplicative update rule is the slow convergence rate when dealing with high-dimensional data. [15] sees this problem and proposed a first-order primal-dual algorithm for NMF with KL divergence using the Chambolle-Pock algorithm as the base. This result shows that the proposed methods converge faster than traditional multiplicative update rules [8] and alternating direction methods of multipliers [9].

The application of deep learning in source separation has been studied. [38] proposed Conv-TasNet to perform source separation in the time domain. As the authors mentioned some drawbacks when using the time-frequency domain. Besides this, [12] also applies their proposed method Wave-U-Net in the time domain. This method is adapted from U-Net [1] in the time-frequency domain but Wave-U-Net performs better based on their experiment. Finally, [28] proposed the multi-scale multi-band densenets which were modified from [16]. Originally densenets are used in image processing and performed well, hence the authors modified the methods and applied them in musical source separation.

IV. RESULTS AND DISCUSSION

In this section, comparisons in terms of convergence rate, computational time, and performance in musical source separation between different matrix factorisation algorithms are discussed.

4.1 Data Set

In this project, we use two musical instruments to test the algorithms. The first musical instrument is the conga which is a drum variant in Cuba. The second musical instrument is the guitar. Then, mix these two instruments manually. In this paper, we refer to conga as a drum. The data set is known as a simple audio mixture in this paper.

After that, we proceed to separate complex musical instruments and we use the song "Beat it" by Michael Jackson found on the open-source MIDI file website. Originally, there are more than two musical instruments in it, hence we use MIDI software to reduce the instrument to only a drum and guitar. This data set is known as a complex audio mixture in this paper.

4.2 Experiment Setup

First, we use librosa[4] to import our audio file and perform a Short-Time Fourier transform (STFT) to get the signal represented in the spectrogram which is in the frequency-time domain. The parameter of STFT is set to the default value. Then, we perform musical source separation on the spectrogram computed using STFT by applying different matrix factorisation algorithms and get its separated individual music audio. For example, we choose the rank to be 6 then we get 6 individual components. Then from the 6 individual audio, we group the audio into drum and guitar audio and add the signal together to get the separated drum and guitar audio. Then, we determine the performance of the algorithms by plotting the separated audio over the original audio, Signal-to-noise ratio (SNR) and listening to the separated audio by ears.

4.3 Convergence Rate

We compare the convergence rate of NMF and CNMF usingEuc, KL divergence andIS Divergence numerically. The mathematical proof can be found in [8]. Then, we use equation (2) to measure errors between the original and the audio separation.

$$\text{Error} = \frac{(X_O - X_{rd})^2}{(X_o)^2}, \qquad (2)$$

where X_o denote the original signal spectrogram and X_{rd} denote the reconstructed spectrogram that undergoes low-rank approximation.

Figure 1 shows that the NMF with different update rules formulas converges to a value, but NMF-ED and NMF-KL continue to converge after 100 iterations hencestopping conditions are needed. On the other hand, NMF-IS stops on its own when the lowest value of the loss function is obtained. From Figure 3,we also observe that CNMF-ED and CNMF-KL convergence patterns are similarto a straight line, but the CNMF-IS convergence value is larger as compared to other methodsshown in Figure 2. Then, we plot the graph consisting of the CNMF convergence rate.

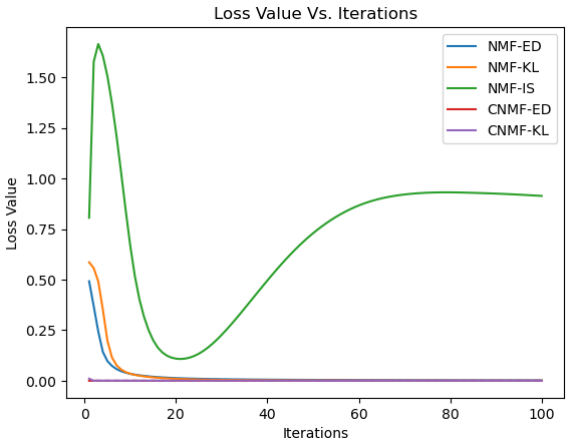


Fig. 1: Convergence rate of each algorithm except for CNMF.

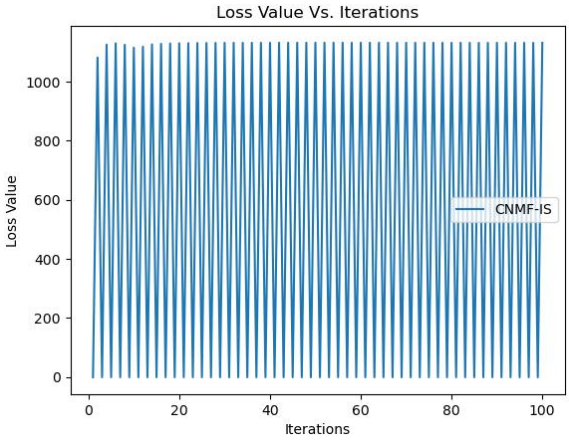


Fig. 2: Converge rate for CNMF-IS.

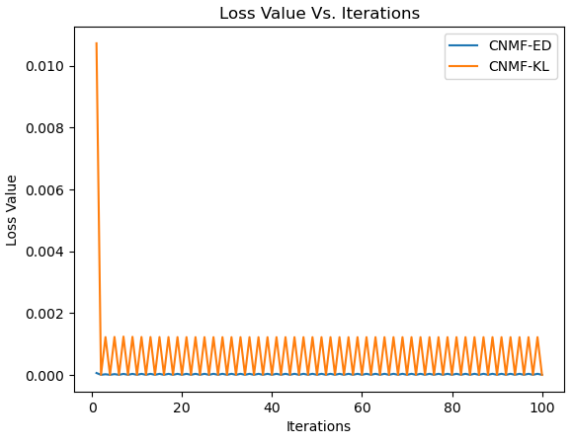


Fig.3: Convergence rate for CNMF-ED and CNMF-KL.

From Figures 2 and 3, we observe that the convergence of Convolutive Non-negative Matrix Factorisation is an oscillation between two numbers. The number at the bottom always remains the same but the number at the top changes until it stops at a certain loss value.

4.4 Computational Time

Computational time is an important aspect of an algorithm as society needs a fast efficient algorithm to solve problems. Therefore, we plot a graph to

study the computational time when we increase the rank. As the rank increases the dimension needs more time to calculate the factorised matrices. For this result, we choose 5, 10, 15, 20, 25, and 30 as the rank for the factorisation.

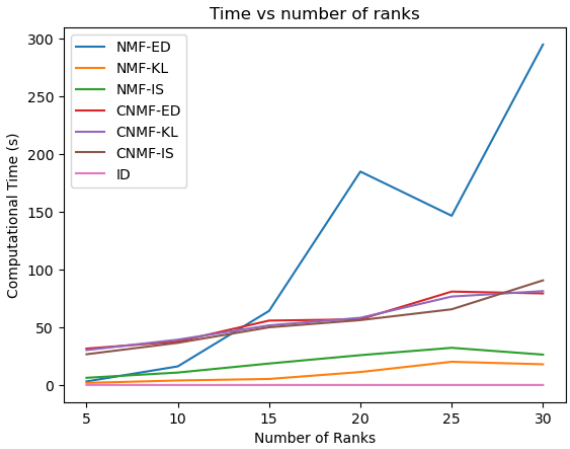


Fig. 4: Computational time of each rank.

From Figure 4, we observe that NMF-ED needs more time to compute as the rank increases. On the other hand, all three CNMF have similar computational time as rank increases. From all the methods, ID gives almost parallel lines to the x-axis. In terms of speed, ID has a better performance when the matrix dimension is getting larger.

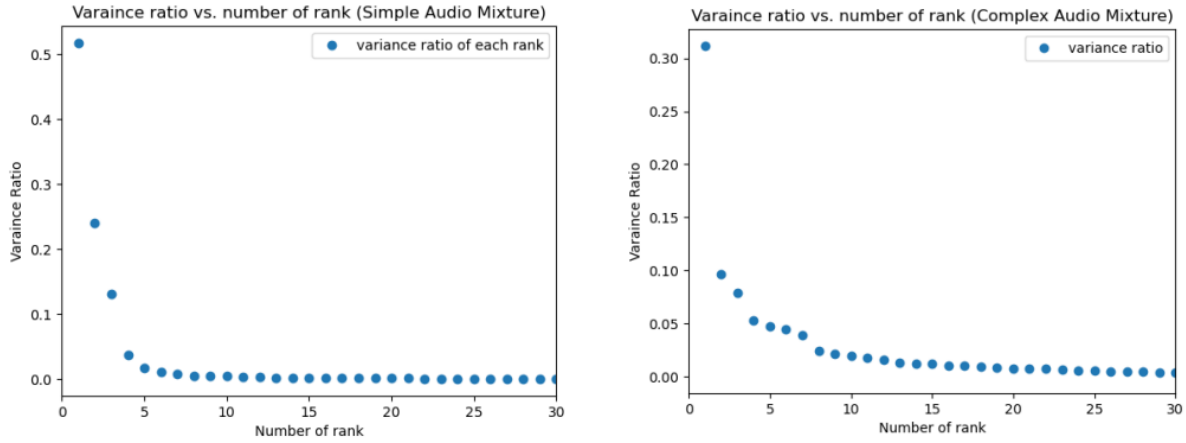
Another information, we observe that the NMF-ED at rank 25 has a lower computational time as compared to rank 20 and rank 30. This is because the solution of NMF is sensitive 9 to the initial value of W and H matrices [40]. Therefore, the scikit learn NMF algorithm initialises the initial W and H matrices and this may cause the values to sometimes be near or far from the converge value. Therefore,[34] has written a review paper on different types of initialisation methods other than random initialisation methods.

4.5 Performance in Musical Source Separation

In this section, we look into the performance of NMF, CNMF and ID in the simple audio mixture and complex audio mixture. The methods used to compare the performance in musical source separation are NMF-ED, NMF-KL, NMF-IS, CNMF-ED, CNMF-KL, CNMF-IS and ID.

The simple audio mixture consists of 2 individual musical instrument audios extracted from different music, hence overlapping of frequency in the spectrogram is less. On the other hand, a complex audio mixture consists of 2 individual musical instrument audios extracted from the same music, hence the overlapping of frequency in the spectrogram is high.

Then, we use Principal Component Analysis (PCA) to determine the suitable rank for the matrix factorisation methods and plotted the graph.



(a): Simple audio mixture (b): Complex audio mixture

Fig 5: Determine the rank using Principal Component Analysis (PCA).

From Figure 5(a), we take 7 as the rank to solve the source separation for the simple audio mixture. For the case in Figure 5(b), initially, we choose the rank to be 7 but the separated drum and guitar audio still have some background noise. Therefore, we increase the rank to 19 so that we can isolate the noise as an individual component.

After setting the rank, we apply matrix factorisation on the spectrogram to obtain the separated drum and guitar audio. Then we measure the performance using the Signal-to-noise ratio (SNR), graph and ears together. The formula for SNR is shown below:

$$\text{SNR} = 10 \log_{10} \frac{P_{signal}}{P_{noise}}, \qquad (3)$$

where P denotes the power of each signal SNR compared to the desired signal to background noise and higher SNR values denote that the separated audio is good. This means that the coverage of the separated audio region covers most of the original audio region. However, the disadvantage of equation

(3) is the inability to detect minor background noise in the separated audio and compute a high SNR value. Therefore, we cannot depend solely on SNR value to measure how good the method is in separating the source.

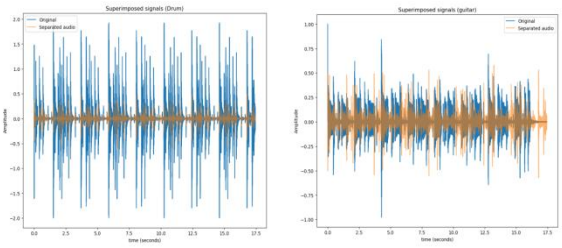
Method	SNR (Drum)	SNR (Guitar)
NMF-ED	3.1666	1.1135
NMF-KL	4.8019	2.7875
NMF-IS	2.6893	2.7873
CNMF-ED	12.6782	1.5997
CNMF-KL	9.9045	2.4574
CNMF-IS	2.1662	1.8315
ID	-3.1725	-0.8856

Table 1:SNR value for each method (Simple Audio Mixture).

Both Table 1 and Table 2, show that ID does not perform well in separating the source. When compared internally among the simple and complex mixture audio, ID performs well when applied in simple mixture audio as shown in Figure 6 and Figure 7. Another piece of information to note is the ID formula for Figure 6 and Figure 7 are equation (1). The columns of C in equation (1) are chosen randomly from the columns ofthe original matricesA_{simple} and $A_{complex}$.

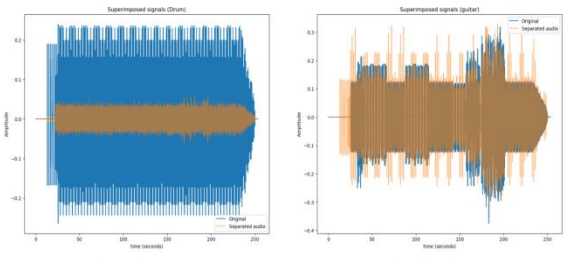
Method	SNR (Drum)	SNR (Guitar)
NMF-ED	6.0972	6.2619
NMF-KL	9.9319	5.1886
NMF-IS	6.0778	2.2496
CNMF-ED	5.3363	4.3901
CNMF-KL	7.3514	3.7712
CNMF-IS	2.9770	3.5599
ID	1.5640	0.1763

Table 2:SNR value for each method (Complex Audio Mixture).



(a): Drum (b): Guitar

Fig 6: Graph of original signal overlap with the separated signal using ID with the simple audio mixture.

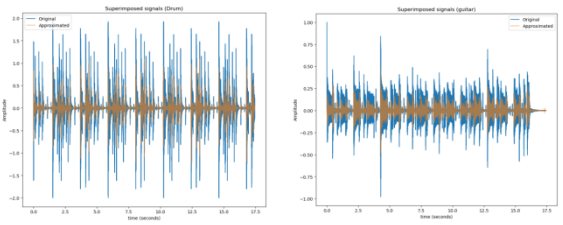


(a): Drum (b): Guitar

Fig 7: Graph of original signal overlap with the separated signal using ID with the complex audio mixture.

There is a certain column combination chosen from the original matrix that can reduce the error between the original and dimensional reduction matrix. Therefore, we need to run the ID algorithm multiple times to get better C and Z. Another finding is dealing with the complex audio mixture, there are certain column combinations chosen from the columns of $A_{complex}$ that made the matrix become either an ill-conditioned matrix or a singular matrix in which the algorithm returns an error message. This can be overcome by running the algorithm a few times until the error message does not pop up.

ID factorisation is not a part-based representation like NMF and CNMF. ID depends on the random combination chosen from matrix A. A good example is from Figure 6 where the separated region follows the pattern of the original region. On the other hand, Figure 7 shows that the guitar separation region is larger than the original region.This shows that the separation is unsuccessful. The result can be improved by increasing the rank, but it is time-consuming as we need to categorise each of the components manually.



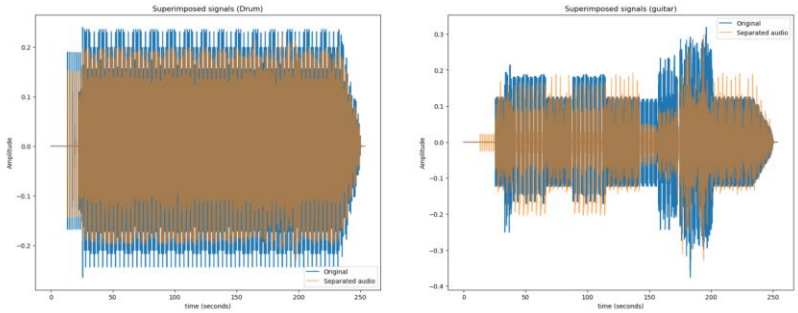
(a): Drum (b): Guitar

Fig 8: Graph of original signal overlap with the separated signal using CNMF-IS with the simple audio mixture.

Next, we move our discussion to CNMF and NMF. Every method of CNMF and NMF has a similar graph in Figure 8, but the difference is the coverage of the separated region over the original region. [35] shows that NMF-ED and NMF-KL work well with simple beat music audio, then we extend it using NMF-IS and CNMF with multiplicative update rules derived using ED, KL and IS in our music data. Then, we listen to every drum and guitar separated audio. We found that CNMF-KL and CNMF-IS have better audio source separation compared to other methods. Even though Table 1 shows that CNMF-ED has a higher SNR value in Drum audio. However, we can hear some background noise which is not part of the original audio. Therefore, in the case of simple mixtures, audio CNMF-KL and CNMF-IS perform better in audio separation for the simple audio mixtures.

Finally, we look into applying CNMF and NMF in a complex audio mixture. By looking at the separated guitar audio in Figure 9, we can see there are some separated signal regions outside of the original region. This is due to certain drumbeats being unable to separate from the guitar audio. This scenario happens on all the methods in this project,

but CNMF-KL, CNMF-IS, NMF-KL and NMF-IS can reduce this background sound as compared to CNMF-ED and CNMF-IS. Besides this, there is a cymbal sound in the original audio and some methods are not able to reproduce the sound in the separated audio. Methods like NMF-ED and CNMF-ED are not able to reproduce this sound, in addition, NMF-KL and CNMF-KL also not able to reproduce the sound at rank 13 but when the rank increase to 19 the cymbal sound is reproduced.



(a): Drum (b): Guitar

Figure 9: Graph of original signal overlap with the separated signal using CNMF-IS with the complex audio mixture.

Different from these two methods, NMF-IS and CNMF-IS can reproduce the cymbal sound even at a lower rank. This is because of the scale invariance property [7] that IS possesses able to low-energy components such as the cymbal sound. Between the two methods, NMF-IS perform as the sound is much louder and clearer when listening. To sum up, using Itakura-Saito Divergence to derive the multiplicative update rulesis suitable for solving music audio problems in either NMF or CNMF.

V. CONCLUSION

We compare the performance of NMF, CNMF and ID in music instrument source separation. When dealing with simple audio, a user in GitHub shows us the performance of NMF-ED and NMF-KL and we extend it using NMF-IS, CNMF and ID. From this experiment, all the methods except for NMF-ED, CNMF-ED and ID can separate the drum and guitar sources. It is smooth when listeningwith our ears. Next in complex audio mixtures, CNMF-IS is more suitable in musical audio separation as there is some music sound unable to detect when using either KL or ED as the update rules in NMF or CNMF. Some possible future works incorporate machine learning to help in determining each of the component categories automatically to speed up the source separation task. When apply to a fast pace musical instrument audio, it is hard for our ears to listen to which music it belongs to. A modification on ID is necessarily needed to apply in audio processing or ID is also possible to apply in image processing. As ID has a lower computational time when the rank increases.

ACKNOWLEDGMENTS

The research was supported by UniversitiTunku Abdul Rahman (UTAR) through UniverisitiTunku Abdul Rahman Research Fund IPSR/RMC/UTARRF/2021-C1/N03.

REFERENCES

[1] A. Jansson, E. Humphrey, N. Montecchio, R. Bittner, A. Kumar, and T. Weyde, “Singing voice separation with deep U-Net convolutional networks,” October 2017. https://openaccess.city.ac.uk/id/eprint/19289/.

[2] A. Lefevre, F. Bach, and C. Fevotte, “Online algorithms for nonnegative matrix factorization with the Itakura-Saito divergence,” *2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, October 2011, doi: 10.1109/aspaa.2011.6082314.

[3] A. Ozerov and C. Fevotte, “Multichannel Nonnegative Matrix Factorization in Convolutive Mixtures for Audio Source Separation,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 3, pp. 550 – 563, March 2010, doi: 10.1109/tasl.2009.2031510.

[4] B. McFee et al., "librosa/librosa: 0.9.2,"*Zenodo*, June 27, 2022. https://zenodo.org/record/6759664#.Y1n2S7ZBxPY.

[5] B. P. Roe, H.-J. Yang, J. Zhu, Y. Liu, I. Stancu, and G. McGregor, “Boosted decision trees as an alternative to artificial neural networks for particle identification,”*Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 543, no. 2 – 3, pp. 577 – 584, May 2005, doi: 10.1016/j.nima.2004.12.018.

[6] C. B. Do and S. Batzoglou, “What is the expectation maximization algorithm?,” *Nature Biotechnology*, vol. 26, no. 8, pp. 897 – 899, August 2008, doi: 10.1038/nbt1406.

[7] C. Févotte, N. Bertin, and J. -L. Durrieu, “Nonnegative Matrix Factorization with the Itakura-Saito Divergence: With Application to Music Analysis,” *Neural Computation*, vol. 21, no. 3, pp. 793 – 830, March 2009, doi: 10.1162/neco.2008.04-08-771.

[8] D. Lee, H. Seung, “Algorithms for non-negative matrix factorization,” *Advances in neural information processing systems*, vol. 13, 2000. Available: http://web.cs.ucla.edu/~yzsun/classes/2014Spring_CS7280/Papers/Clustering/NNF_lee01algorithms.pdf.

[9] D. L. Sun and C. Fevotte, “Alternating direction method of multipliers for non-negative matrix factorization with the beta-divergence,” *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, doi: 10.1109/icassp.2014.6854796.

[10] D. R. Harp, D. O’Malley, B. Yan, and R. Pawar, “On the feasibility of using physics-informed machine learning for underground reservoir pressure management,” *Expert Systems with Applications*, vol. 178, p. 115006, September 2021, doi: 10.1016/j.eswa.2021.115006.

[11] D. Schvartzman and C. D. Curtis, “Signal Processing and Radar Characteristics (SPARC) Simulator: A Flexible Dual-Polarization Weather-Radar Signal Simulation Framework

Based on Preexisting Radar-Variable Data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 1, pp. 135 – 150, January 2019, doi: 10.1109/jstars.2018.2885614.

[12] D. Stoller, S. Ewert, and S. Dixon, "Wave-U-Net: A Multi-Scale Neural Network for End-to-End Audio Source Separation," *19th International Society for Music Information Retrieval Conference (ISMIR 2018)*, June 2018, doi: 10.48550/ARXIV.1806.03185.

[13] E. Cano, D. FitzGerald, A. Liutkus, M. D. Plumbley, and F. R-. Stöter, "Musical Source Separation: An Introduction," *IEEE Signal Processing Magazine*, vol. 36, no. 1, pp. 31 – 40, January 2019, doi: 10.1109/MSP.2018.2874719.

[14] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, no. 85, pp. 2825 – 2830, 2011, [Online]. Available: https://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html.

[15] F. Yanez and F. Bach, "Primal-dual algorithms for non-negative matrix factorization with the Kullback -Leibler divergence,"*2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2017, doi: 10.1109/icassp.2017.7952558.

[16] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," *arXiv:1608.06993 [cs]*, January 2018, [Online]. Available: https://arxiv.org/abs/1608.06993v5

[17] H. Kameoka, Nobutaka Ono, Kunio Kashino, and Shigeki Sagayama, "Complex NMF: A new sparse representation for acoustic signals," *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, April 2009, doi: 10.1109/icassp.2009.4960364.

[18] H. Sawada, H. Kameoka, S. Araki, and N. Ueda, "New formulations and efficient algorithms for multichannel NMF," *2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, October 2011, doi: 10.1109/aspaa.2011.6082275.

[19] H. Sawada, H. Kameoka, S. Araki, and N. Ueda, "Efficient algorithms for multichannel extensions of Itakura-Saito nonnegative matrix factorization," *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2012, doi: 10.1109/icassp.2012.6287867.

[20] H. Sawada, H. Kameoka, S. Araki, and N. Ueda, "Multichannel Extensions of Non-Negative Matrix Factorization With Complex-Valued Data," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 5, pp. 971 – 982, May 2013, doi: 10.1109/tasl.2013.2239990.

[21] J. Blauert, "Spatial Hearing," 1996, doi: 10.7551/mitpress/6391.001.0001.

[22] J. M. Joyce, "Kullback-Leibler Divergence," *International Encyclopedia of Statistical Science*, pp. 720 – 722, 2011, doi: 10.1007/978-3-642-04898-2_327.

[23] J.S. Bagla, "Cosmological N-body simulation: Techniques, Scope and Status," *arxiv.org*, Nov. 2004, doi: 10.48550/arXiv.astro-ph/0411043.

[24] J. Wang, J. Wu, and H. Xiao, "A Physics-Informed Machine Learning Approach of Improving RANS Predicted Reynolds Stresses," *55th AIAA Aerospace Sciences Meeting*, January 2017, doi: 10.2514/6.2017-1712.

[25] J. Zhang, D. Li, Y. Hao, and Z. Tan, "A hybrid model using signal processing technology, econometric models and neural network for carbon spot price forecasting," *Journal of Cleaner Production*, vol. 204, pp. 958 – 964, December 2018, doi: 10.1016/j.jclepro.2018.09.071.

[26] K. Yoshii, R. Tomioka, D. Mochihashi, and M. Goto, "Beyond NMF: Time-Domain Audio Source Separation without Phase Reconstruction," *Semantic Scholar*, 2013. https://www.semanticscholar.org/paper/Beyond-NMF%3A-Time-Domain-Audio-Source-Separation-Yoshii-Tomioka/5fc6318684468678341f5e4fcc1c7317538c9ff3.

[27] N. S. Jayant and E. Y. Chen, "Audio Compression: Technology and Applications," *AT&T Technical Journal*, vol. 74, no. 2, pp. 23 – 34, March 1995, doi: 10.1002/j.1538-7305.1995.tb00398.x.

[28] N. Takahashi and Y. Mitsufuji, "Multi-Scale multi-band densenets for audio source separation," *2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, October 2017, doi: 10.1109/waspaa.2017.8169987.

[29] P. López-Serrano, C. Dittmar, Y. Özer, and M. Müller, "NMF TOOLBOX: MUSIC PROCESSING APPLICATIONS OF NONNEGATIVE MATRIX FACTORIZATION," in *Proceedings of the 22nd International Conference on Digital Audio Effects (DAFx-19)*, Birmingham, UK, 2019. [Online]. Available: http://www.dafx.de/paper-archive/2019/DAFx2019_paper_1.pdf.

[30] P. O. Hoyer, "Non-negative sparse coding," *Proceedings of the 12th IEEE Workshop on Neural Networks for Signal Processing*, 2002, doi: 10.1109/nnsp.2002.1030067.

[31] R. Advani and S. O'Hagan, "Efficient Algorithms for Constructing an Interpolative Decomposition," *arXiv:2105.07076 [cs, math]*, June 2022. [Online]. Available: https://arxiv.org/abs/2105.07076v2.

[32] R. J. Vaccaro, "The past, present, and the future of underwater acoustic signal processing," *IEEE Signal Processing Magazine*, vol. 15, no. 4, pp. 21 – 51, July 1998, doi: 10.1109/79.689583.

[33] R. M. Parry and I. Essa, "Estimating the Spatial Position of Spectral Components in Audio," *Independent Component Analysis and Blind Signal Separation*, pp. 666 – 673, 2006, doi: 10.1007/11679363_83.

[34] S. F. Hafshejani and Z. Moaberfard, "Initialization for Nonnegative Matrix Factorization: a Comprehensive Review," *arXiv:2109.03874[cs, math]*, September 2021, [Online]. Available: https://arxiv.org/abs/2109.03874.

[35] tky823, "audio source separation,"*GitHub*, 2021. https://github.com/tky823/audio_source_separaton.

[36] T. N. Sainath et al., "Multichannel Signal Processing With Deep Neural Networks for Automatic Speech Recognition," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 5, pp. 965 – 979, May 2017, doi: 10.1109/taslp.2017.2672401.

[37] Y. Du, T. Mukherjee, and T. DebRoy, "Physics-informed machine learning and mechanistic modeling of additive manufacturing to reduce defects," \textit{Applied Materials Today}, vol. 24, p. 101123, Sep. 2021, doi: 10.1016/j.apmt.2021.101123.

[38] Y. Luo and N. Mesgarani, "Conv-TasNet: Surpassing Ideal Time-Frequency Magnitude Masking for Speech Separation,"*IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 8, pp. 1256 – 1266, August 2019, doi: 10.1109/taslp.2019.2915167.

[39] Y. Xu, S. Kohtz, J. Boakye, P. Gardoni, and P. Wang, "Physics-informed machine learning for reliability and systems safety applications: State of the art and challenges," *Reliability Engineering & System Safety*, vol. 230, p. 108900, February 2022, doi: 10.1016/j.ress.2022.108900.

[40] Y. -X. Wang and Y. -J. Zhang, "Nonnegative Matrix Factorization: A Comprehensive Review," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 6, pp. 1336 – 1353, June 2013, doi: https://doi.org/10.1109/TKDE.2012.51.

★★★