INTERNATIONAL AUDIO LABORATORIES ERLANGEN
A joint institution of Fraunhofer IIS and Universität Erlangen-Nürnberg

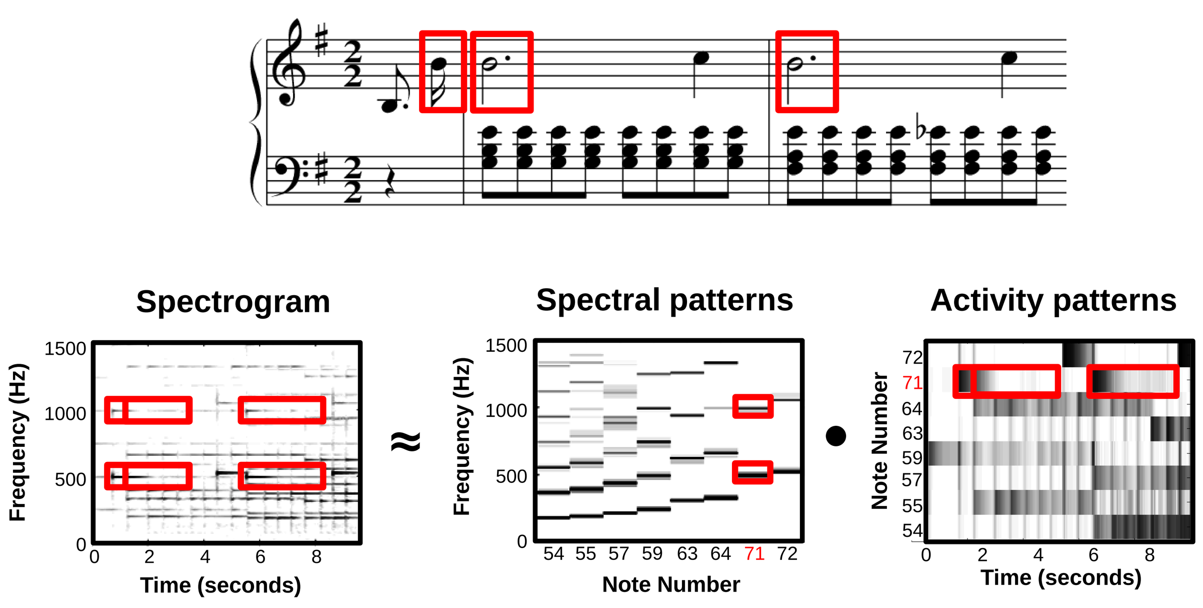
AUDIO LABS

INVESTIGATING NONNEGATIVE AUTOENCODERS FOR EFFICIENT AUDIO DECOMPOSITION

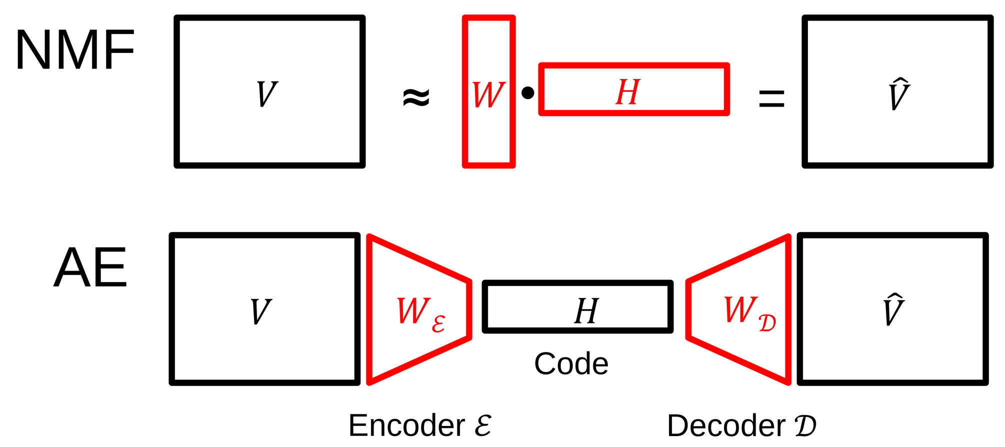
Yigitcan Özer, Jonathan Hansen, Tim Zunner, Meinard Müller

Introduction

- Task: Score-Informed Audio Decomposition
- Given the magnitude spectrogram of a music recording, encode
  - Spectral patterns, *templates*
  - Occurences in time, *activations*



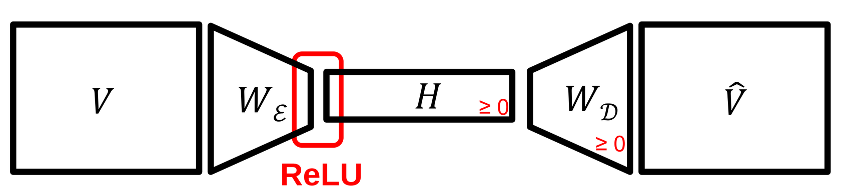
Nonnegative Autoencoders (NAE)



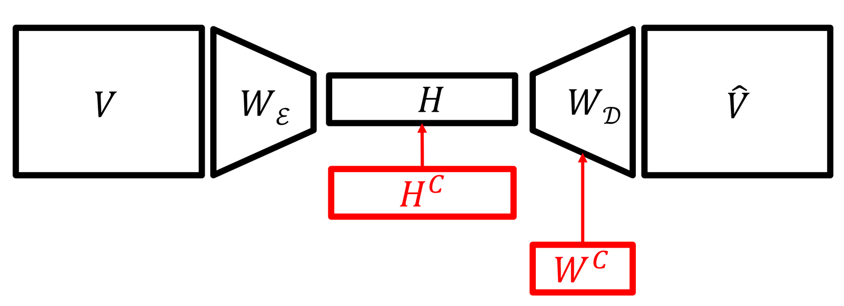
1. Layer: $H = W_\mathcal{E}\, V$
2. Layer: $\hat{V} = W_\mathcal{D}\, H$

NMF: Learn H and W
AE: Learn $W_\mathcal{E}$ and $W_\mathcal{D}$

- Nonnegativity constraints
  - **Code** $H \rightarrow$ ReLU
  - **Decoder** $\mathcal{D} \rightarrow$ Projected gradient



- Musical constraints
  - Activity constraints → Structured Dropout
  - Template constraints → Binary masking



NMF and NAE

- Multiplicative Nonnegative Matrix Factorization (NMF Mult.)

$$H \leftarrow H \odot (W^\top V) \oslash (W^\top W H + \varepsilon)$$
$$W \leftarrow W \odot (V H^\top) \oslash (W H H^\top + \varepsilon)$$

- This work: **Multiplicative NAE** (NAE Mult.)

$$W_\mathcal{E} \leftarrow W_\mathcal{E} \odot \Big(\big(((W_\mathcal{D}^\top V) \odot H^C)V^\top\big) \oslash \big(((W_\mathcal{D}^\top W_\mathcal{D} H') \odot H^C)V^\top + \varepsilon\big)\Big)$$
$$W_\mathcal{D} \leftarrow W_\mathcal{D} \odot \big((V H'^\top) \oslash (W_\mathcal{D} H' H'^\top + \varepsilon)\big)$$

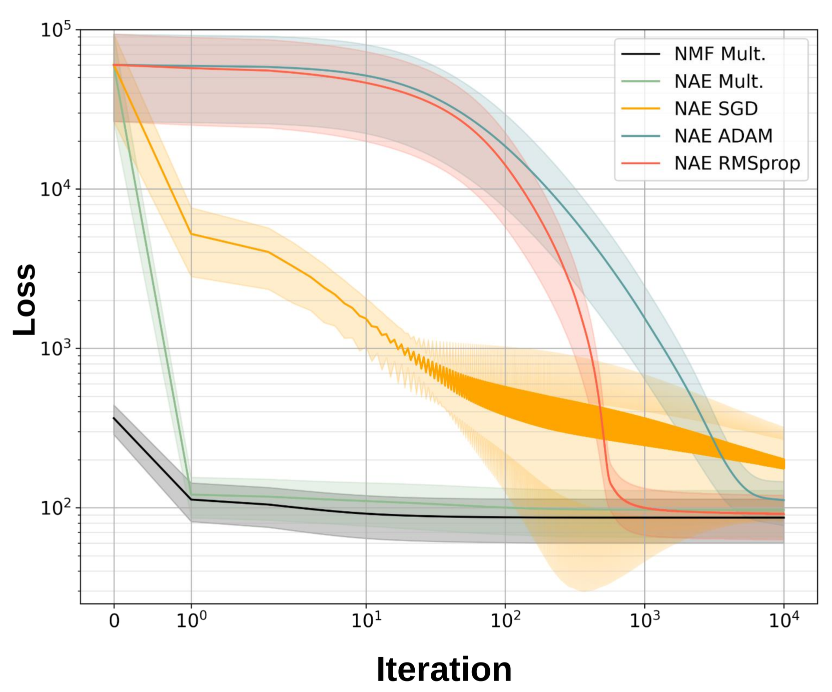
- Additive NAE
  - NAE SGD
  - NAE ADAM
  - NAE RMSprop

Results

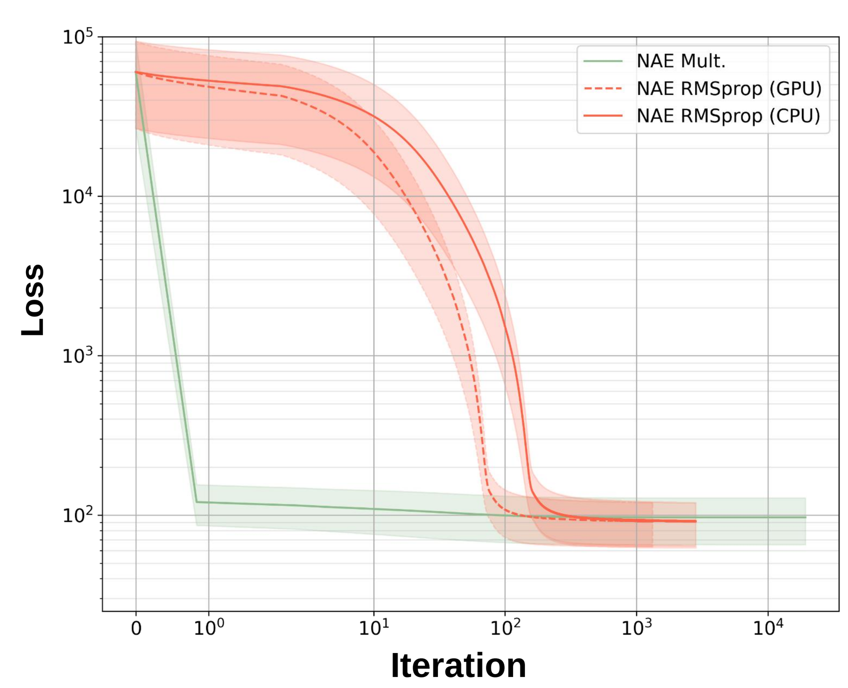
Approximation Error between V and $\hat{V}$

File ID \ Model	NMF Mult.	NAE Mult.	NAE SGD	NAE ADAM	NAE RMSprop
Chopin_Op028-04_SMD	46.4	49.0	62.4	57.6	48.1
Chopin_Op028-15_SMD	48.5	53.2	67.3	66.2	50.9
Chopin_Op066_SMD	79.5	87.2	139.0	101.1	85.7
Beethoven_Op031No2-01_SMD	90.7	99.2	105.1	104.4	94.9
Chopin_Op028-01_SMD	94.8	103.6	299.9	122.8	97.4
Bach_BWV875-01_SMD	97.5	107.3	219.9	129.2	104.3
Beethoven_Op111-01_EA	103.7	129.4	328.5	148.4	113.0
Chopin_Op064No1_EA	131.9	145.9	383.6	161.6	137.2

Average Approximation Error



Runtime Comparison



RMSprop outperforms other NAE variants in terms of the approximation quality and efficiency.

https://resources.mpi-inf.mpg.de/MIR/ICASSP2012-ScoreInformedNMF/

References

[1] Lee and Seung: Algorithms for Non-Negative Matrix Factorization. Proc. NIPS, 2000.

[2] Ewert and Müller: Using Score-Informed Constraints for NMF-based Source Separation. ICASSP, 2012.

[3] Smaragdis and Venkataramani: A Neural Network Alternative to Non-negative Audio Models. ICASSP, 2017.

[4] Ewert and Sandler: Structured Dropout for Weak Label and Multi-instance Learning and its Application for Score-Informed Source Separation. ICASSP, 2017.

Acknowledgements

This work was supported by the German Research Foundation (DFG MU 2686/10-2). The International Audio Laboratories Erlangen are a joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) and Fraunhofer Institut für Integrierte Schaltungen IIS.

FAU Friedrich-Alexander-Universität Erlangen-Nürnberg

Fraunhofer IIS