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



Nonnegative Autoencoders with **Applications to Music Audio Decomposing**

Meinard Müller

International Audio Laboratories Erlangen meinard.mueller@audiolabs-erlangen.de

Guest Lecture Sound and Music Computing (CS4347/CS5647) National University of Singapore

16.10.2023



Friedrich-Alexander-Universität Erlangen-Nürnberg



Meinard Müller

- Mathematics (Diplom/Master, 1997)
 Computer Science (PhD, 2001)
 Information Retrieval (Habilitation, 2007)
- Senior Researcher (2007-2012)
- Professor Semantic Audio Processing (since 2012)
- Former President of the International Society for Music Information Retrieval (MIR)
- IEEE Fellow for contributions to Music Signal Processing













Nonnegative Autoencoders with Applications to Music Audio Decomposing



Meinard Müller: Research Group

Semantic Audio Processing

- Michael Krause
- Yigitcan Özer
- Simon Schwär
- Johannes Zeitler
- Peter Meier (external)
- **Christof Weiß**
- Sebastian Rosenzweig
- Frank Zalkow
- **Christian Dittmar**
- Stefan Balke

. . .

- Jonathan Driedger
- **Thomas Prätzlich**















International Audio Laboratories Erlangen





- Fraunhofer Institute for Integrated Circuits IIS
- Largest Fraunhofer institute with
 ≈ 1000 members
- Applied research for sensor, audio, and media technology



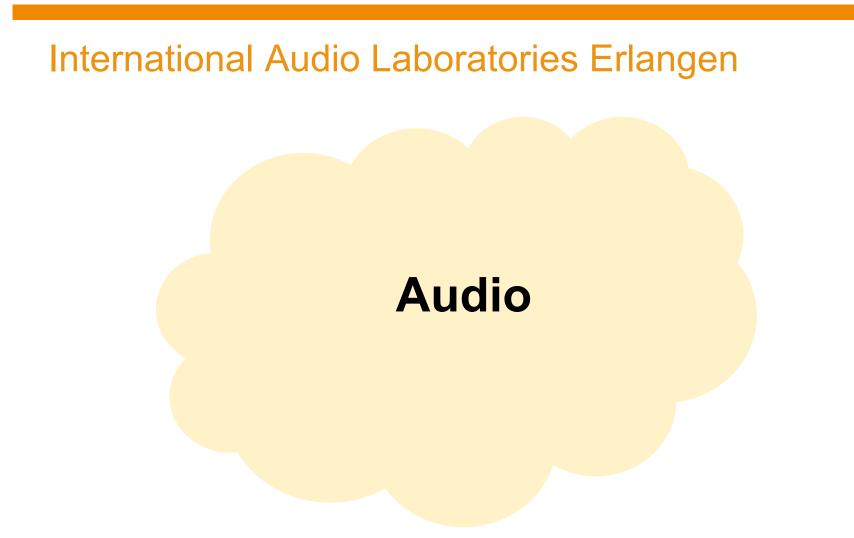






- Friedrich-Alexander Universität Erlangen-Nürnberg (FAU)
- One of Germany's largest universities with ≈ 40,000 students
- Strong Technical Faculty

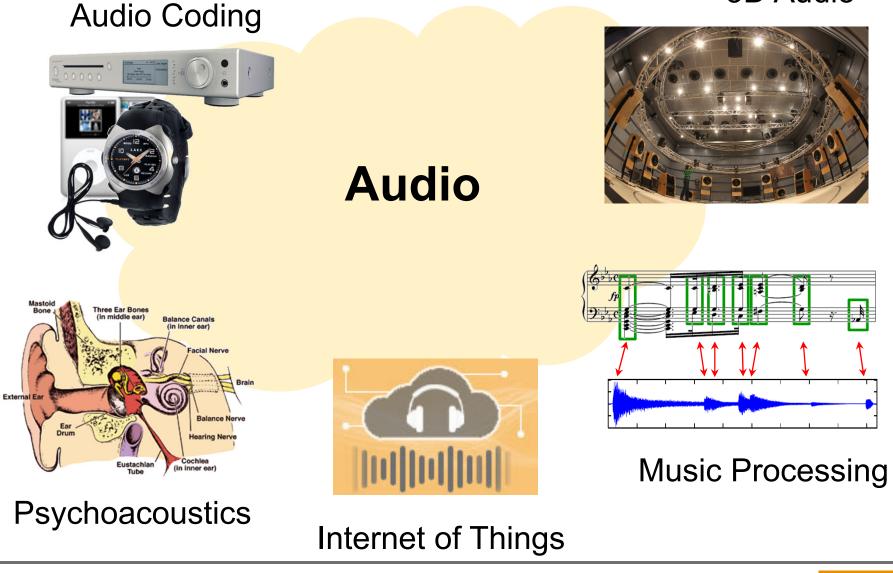






International Audio Laboratories Erlangen

3D Audio





AudioLabs – FAU

- Prof. Dr. Jürgen Herre Audio Coding
- Prof. Dr. Bernd Edler Audio Signal Analysis
- Prof. Dr. Meinard Müller Semantic Audio Processing
- Prof. Dr. Emanuël Habets
 Spatial Audio Signal Processing
- Prof. Dr. Nils Peters Audio Signal Processing
- Dr. Stefan Turowski Coordinator AudioLabs-FAU







© AudioLabs, 2023

Meinard Müller







Nonnegative Autoencoders with Applications to Music Audio Decomposing

Source Separation

- Decomposition of audio stream into different sound sources
- Central task in digital signal processing
- "Cocktail party effect"





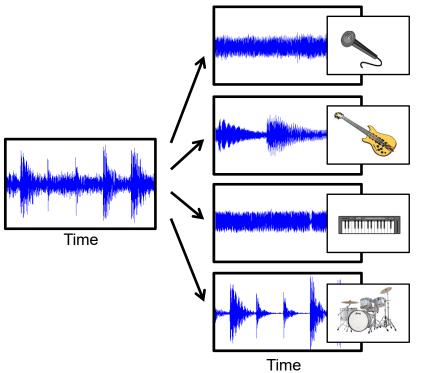
Source Separation

- Decomposition of audio stream into different sound sources
- Central task in digital signal processing
- "Cocktail party effect"
- Several input signals
- Sources are assumed to be statistically independent



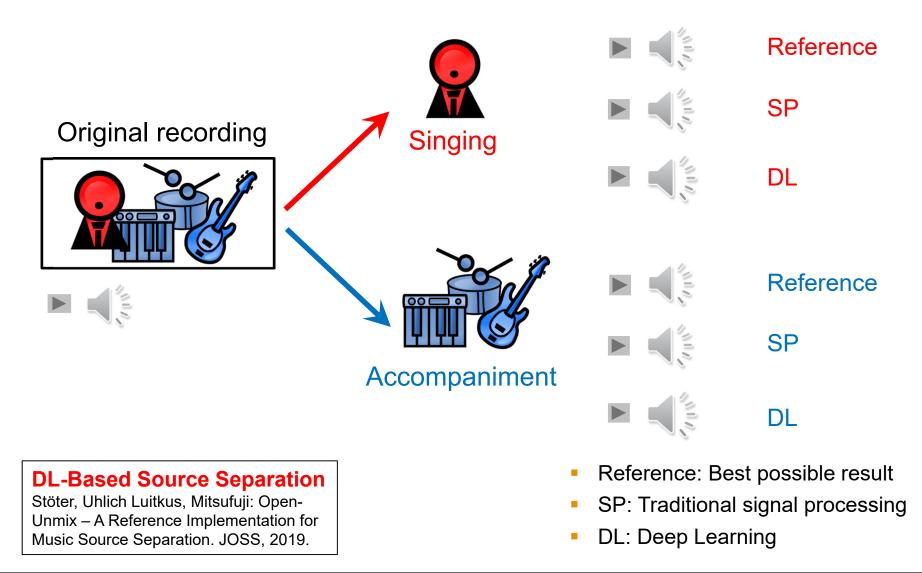
Source Separation (Music)

- Main melody, accompaniment, drum track
- Instrumental voices
- Individual note events
- Only mono or stereo
- Sources are often highly dependent





Source Separation (Singing Voice)





Score-Informed Source Separation

Exploit musical score to support decomposition process

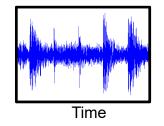
Musical Information



Ewert, Pardo, Müller, Plumbley: Score-Informed Source Separation for Musical Audio Recordings. IEEE SPM 31(3), 2014.

Audio Signal

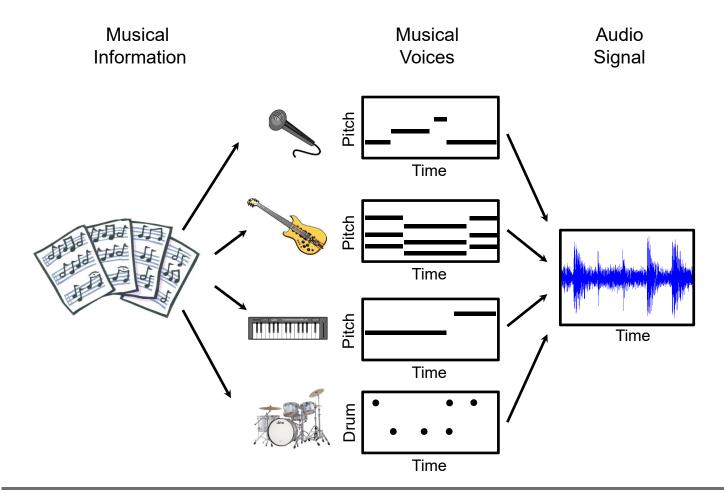






Score-Informed Source Separation

Exploit musical score to support decomposition process



Prior Knowledge Ewert, Pardo, Müller, Plumbley: Score-Informed Source Separation for Musical Audio Recordings.

IEEE SPM 31(3), 2014.

Nonnegative Autoencoders with Applications to Music Audio Decomposing

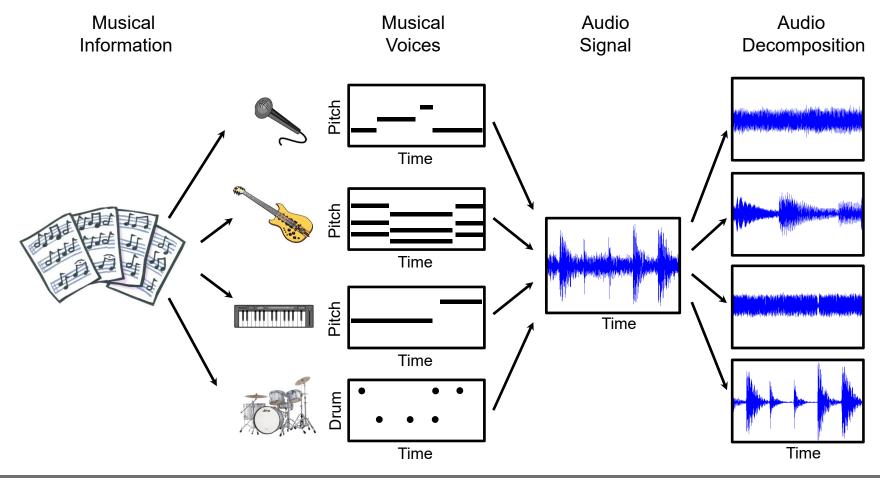


Score-Informed Source Separation

Exploit musical score to support decomposition process

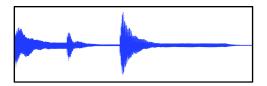
Prior Knowledge Ewert, Pardo, Müller, Plumbley:

Score-Informed Source Separation for Musical Audio Recordings. IEEE SPM 31(3), 2014.

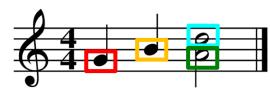


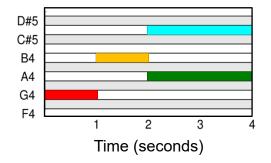


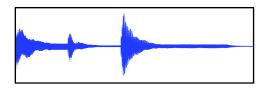




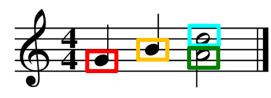


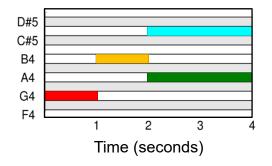


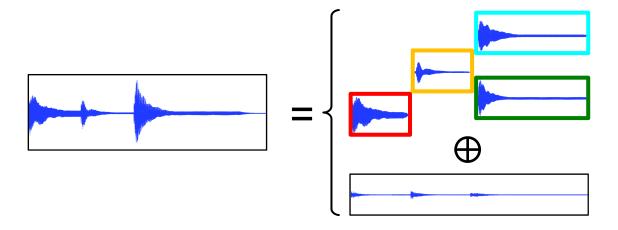




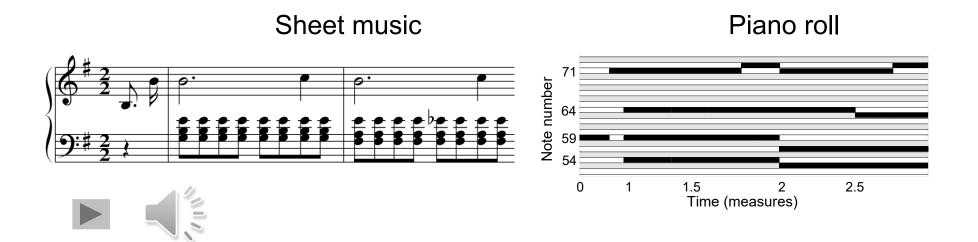




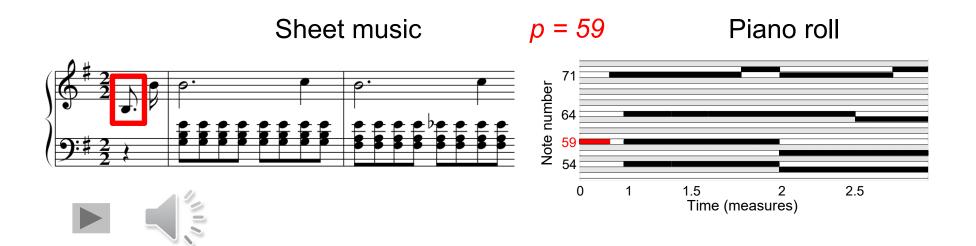




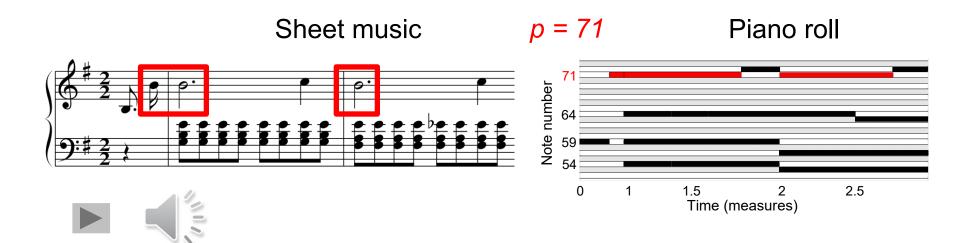




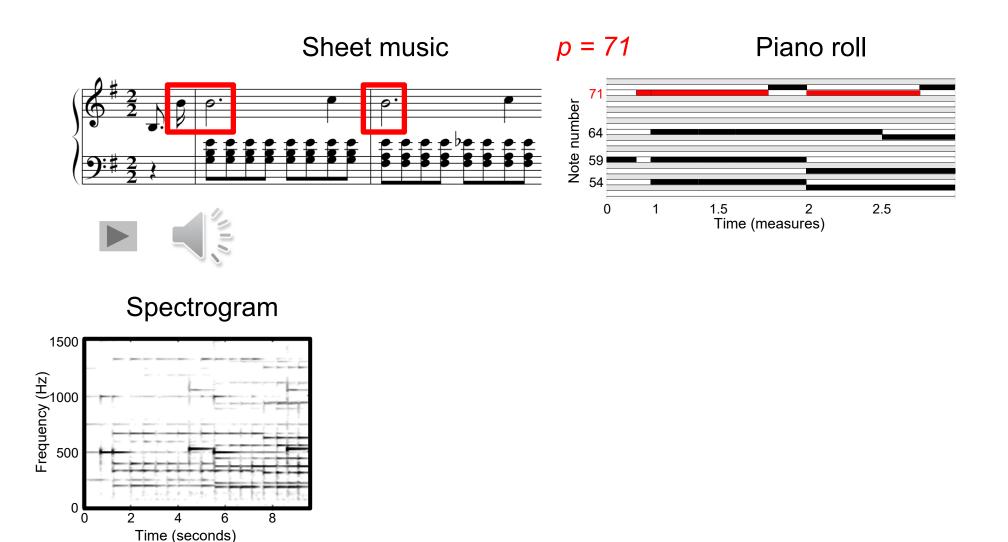




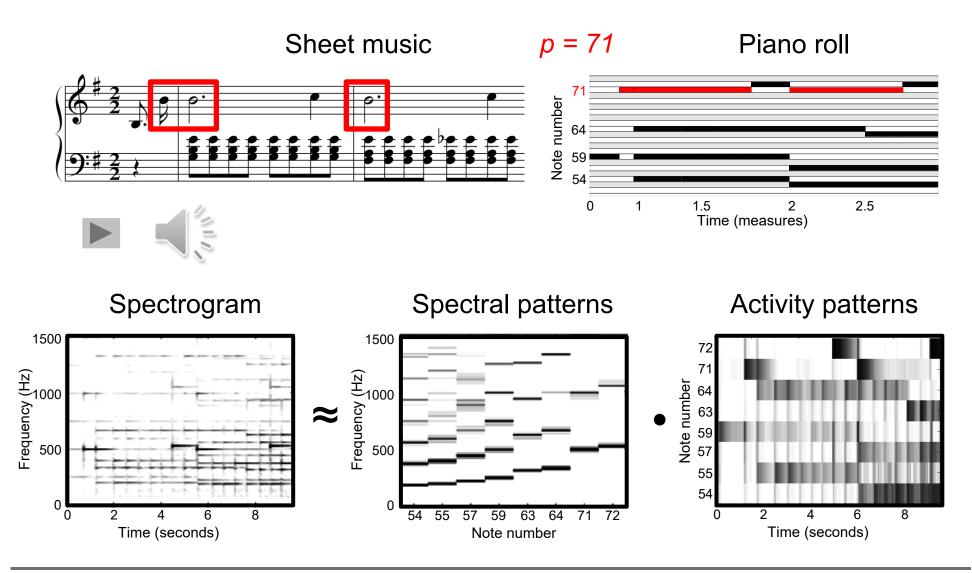






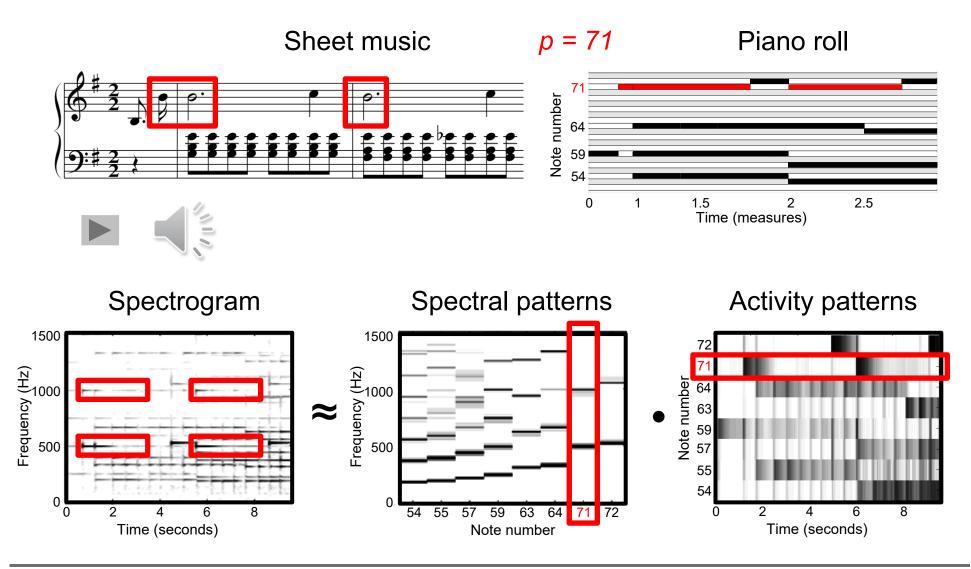






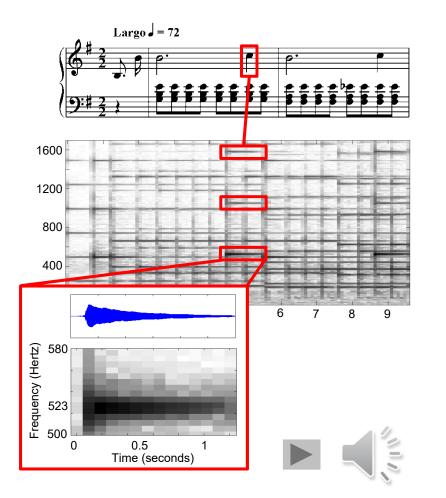
Nonnegative Autoencoders with Applications to Music Audio Decomposing

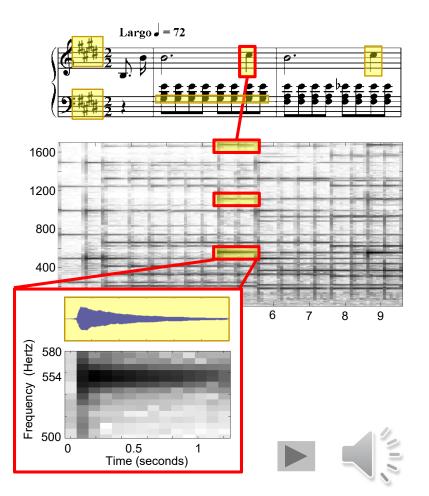




Nonnegative Autoencoders with Applications to Music Audio Decomposing



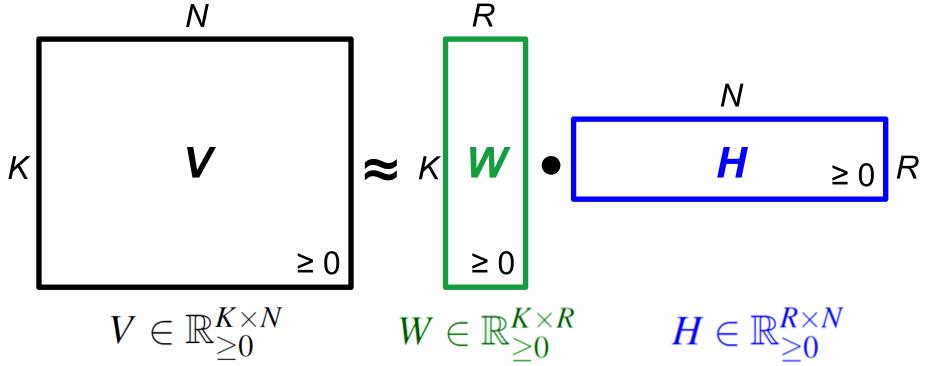




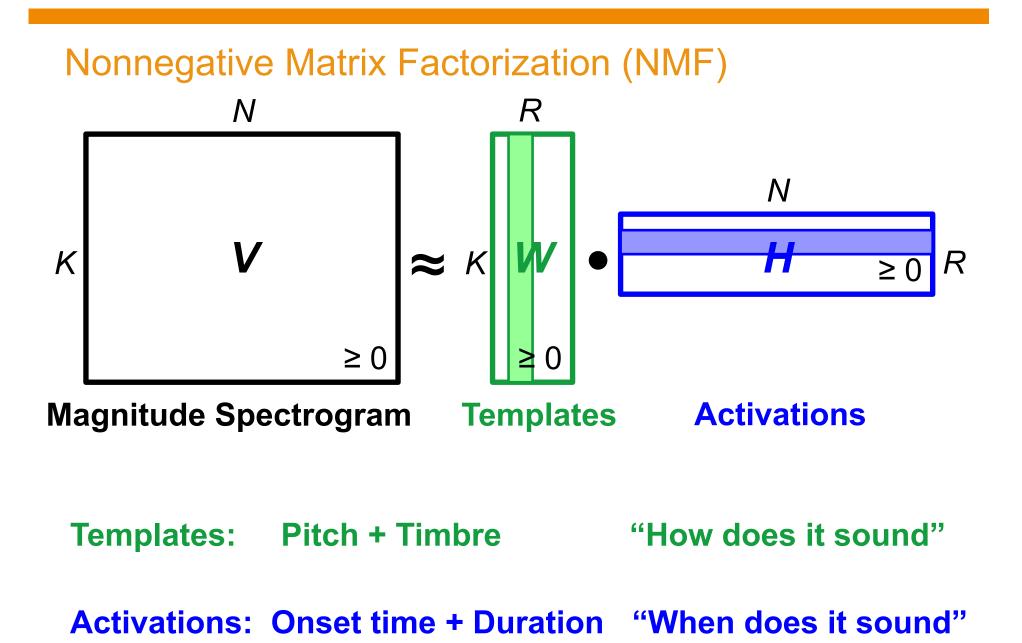
Nonnegative Autoencoders with Applications to Music Audio Decomposing



Nonnegative Matrix Factorization (NMF)

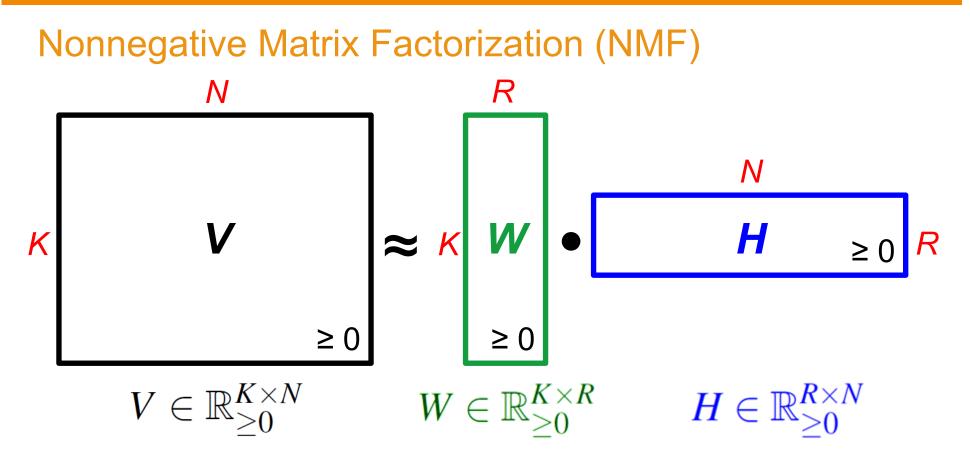






Nonnegative Autoencoders with Applications to Music Audio Decomposing





Dimensionality reduction

- K, N typically much larger than R (maximal rank)
- Example: N = 1000, K = 500, R = 20 K x N = 500,000, K x R = 10,000, R x N = 20,000

Nonnegative Autoencoders with Applications to Music Audio Decomposing



Nonnegative Matrix Factorization (NMF) Ν RΝ Η K ≥ 0 $V \in \mathbb{R}_{>0}^{K \times N}$ $W \in \mathbb{R}_{>0}^{K \times R}$ $H \in \mathbb{R}^{R \times N}_{>0}$

Nonnegativity:

- Prevents mutual cancellation of template vectors
- Encourages semantically meaningful decomposition

Nonnegative Autoencoders with Applications to Music Audio Decomposing



Optimization problem:

Given $V \in \mathbb{R}_{\geq 0}^{K imes N}$ and rank parameter R minimize
$\ V - WH\ ^2$
with respect to $\ W \in \mathbb{R}_{\geq 0}^{K imes R}$ and $\ H \in \mathbb{R}_{\geq 0}^{R imes N}$.

Optimization not easy:

- Nonnegativity constraints
- Nonconvexity when jointly optimizing W and H

Strategy: Iteratively optimize W and H via gradient descent



Computation of gradient with respect to H (fixed W)

$$D := RN$$
$$\varphi^{W} : \mathbb{R}^{D} \to \mathbb{R}$$
$$\varphi^{W}(H) := \|V - WH\|^{2}$$

Variables

 $H \in \mathbb{R}^{R imes N}$ $H_{
ho
u}$ $ho \in [1:R]$ $ho \in [1:N]$



Computation of gradient with respect to H (fixed W)

$$D := RN$$

$$\varphi^{W} : \mathbb{R}^{D} \to \mathbb{R}$$

$$\frac{\partial \varphi^{W}}{\partial H_{\rho v}} = \frac{\partial \left(\sum_{k=1}^{K} \sum_{n=1}^{N} \left(V_{kn} - \sum_{r=1}^{R} W_{kr} H_{rn} \right)^{2} \right)}{\partial H_{\rho v}}$$

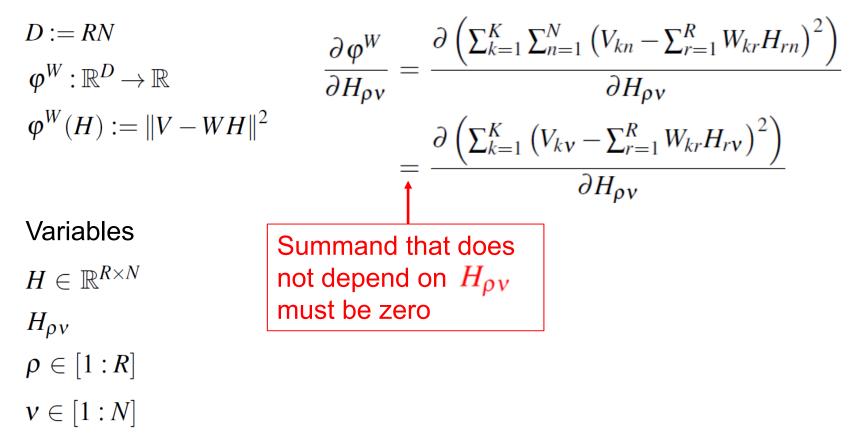
$$\varphi^{W}(H) := \|V - WH\|^{2}$$

Variables

 $H \in \mathbb{R}^{R imes N}$ $H_{
ho
u}$ $ho \in [1:R]$ $ho \in [1:N]$



Computation of gradient with respect to H (fixed W)





Computation of gradient with respect to H (fixed W)

$$D := RN$$

$$\varphi^{W} : \mathbb{R}^{D} \to \mathbb{R}$$

$$\frac{\partial \varphi^{W}}{\partial H_{\rho v}} = \frac{\partial \left(\sum_{k=1}^{K} \sum_{n=1}^{N} \left(V_{kn} - \sum_{r=1}^{R} W_{kr} H_{rn} \right)^{2} \right)}{\partial H_{\rho v}}$$

$$\varphi^{W}(H) := \|V - WH\|^{2}$$

$$= \frac{\partial \left(\sum_{k=1}^{K} \left(V_{kv} - \sum_{r=1}^{R} W_{kr} H_{rv} \right)^{2} \right)}{\partial H_{\rho v}}$$

$$Variables$$

$$H \in \mathbb{R}^{R \times N}$$

$$H_{\rho v}$$

$$\rho \in [1:R]$$

$$v \in [1:N]$$

$$\frac{\partial \varphi^{W}}{\partial H_{\rho v}} = \frac{\partial \left(\sum_{k=1}^{K} \left(V_{kv} - \sum_{r=1}^{R} W_{kr} H_{rv} \right)^{2} \right)}{\partial H_{\rho v}}$$



Computation of gradient with respect to H (fixed W)

 $\frac{\partial \varphi^{W}}{\partial H_{\rho \nu}} = \frac{\partial \left(\sum_{k=1}^{K} \sum_{n=1}^{N} \left(V_{kn} - \sum_{r=1}^{R} W_{kr} H_{rn} \right)^{2} \right)}{\partial H_{\rho \nu}}$ D := RN $\varphi^W: \mathbb{R}^D \to \mathbb{R}$ $\boldsymbol{\varphi}^{W}(H) := \|V - WH\|^2$ $=\frac{\partial\left(\sum_{k=1}^{K}\left(V_{k\nu}-\sum_{r=1}^{R}W_{kr}H_{r\nu}\right)^{2}\right)}{\partial H_{\rho\nu}}$ $= \sum_{k=1}^{K} 2 \left(V_{k\nu} - \sum_{r=1}^{R} W_{kr} H_{r\nu} \right) \cdot \left(-W_{k\rho} \right)$ Variables $H \in \mathbb{R}^{R \times N}$ $= 2\left(\sum_{r=1}^{R}\sum_{k=1}^{K}W_{k\rho}W_{kr}H_{r\nu} - \sum_{k=1}^{K}W_{k\rho}V_{k\nu}\right)$ $H_{\rho\nu}$ $\rho \in [1:R]$ Rearrange $\mathbf{v} \in [1:N]$ summands



Computation of gradient with respect to H (fixed W)

 $\frac{\partial \varphi^{W}}{\partial H_{\rho v}} = \frac{\partial \left(\sum_{k=1}^{K} \sum_{n=1}^{N} \left(V_{kn} - \sum_{r=1}^{R} W_{kr} H_{rn} \right)^{2} \right)}{\partial H_{\rho v}}$ D := RN $\varphi^W: \mathbb{R}^D \to \mathbb{R}$ $\boldsymbol{\varphi}^{W}(H) := \|V - WH\|^2$ $=\frac{\partial\left(\sum_{k=1}^{K}\left(V_{k\nu}-\sum_{r=1}^{R}W_{kr}H_{r\nu}\right)^{2}\right)}{\partial H_{\rho\nu}}$ $= \sum_{k=1}^{K} 2 \left(V_{k\nu} - \sum_{r=1}^{R} W_{kr} H_{r\nu} \right) \cdot \left(-W_{k\rho} \right)$ Variables $H \in \mathbb{R}^{R \times N}$ $= 2\left(\sum_{r=1}^{R}\sum_{k=1}^{K}W_{k\rho}W_{kr}H_{r\nu} - \sum_{k=1}^{K}W_{k\rho}V_{k\nu}\right)$ $H_{\rho\nu}$ $\rho \in [1:R]$ $= 2 \left(\sum_{r=1}^{R} \left(\sum_{k=1}^{K} W_{\rho k}^{\top} W_{kr} \right) H_{r\nu} - \sum_{k=1}^{K} W_{\rho k}^{\top} V_{k\nu} \right)$ $\mathbf{v} \in [1:N]$ Introduce transposed W^{+}



Computation of gradient with respect to H (fixed W)

 $\frac{\partial \varphi^{W}}{\partial H_{\rho v}} = \frac{\partial \left(\sum_{k=1}^{K} \sum_{n=1}^{N} \left(V_{kn} - \sum_{r=1}^{R} W_{kr} H_{rn} \right)^{2} \right)}{\partial H_{\rho v}}$ D := RN $\varphi^W: \mathbb{R}^D \to \mathbb{R}$ $\boldsymbol{\varphi}^{W}(H) := \|V - WH\|^{2}$ $=\frac{\partial\left(\sum_{k=1}^{K}\left(V_{k\nu}-\sum_{r=1}^{R}W_{kr}H_{r\nu}\right)^{2}\right)}{\partial H_{\rho\nu}}$ $= \sum_{k=1}^{K} 2 \left(V_{k\nu} - \sum_{r=1}^{R} W_{kr} H_{r\nu} \right) \cdot \left(-W_{k\rho} \right)$ Variables $H \in \mathbb{R}^{R \times N}$ $= 2\left(\sum_{r=1}^{R}\sum_{k=1}^{K}W_{k\rho}W_{kr}H_{r\nu} - \sum_{k=1}^{K}W_{k\rho}V_{k\nu}\right)$ $H_{\rho\nu}$ $\rho \in [1:R]$ $= 2 \left(\sum_{r=1}^{R} \left(\sum_{k=1}^{K} W_{\rho k}^{\top} W_{kr} \right) H_{r \nu} - \sum_{k=1}^{K} W_{\rho k}^{\top} V_{k \nu} \right)$ $\mathbf{v} \in [1:N]$ $= 2((W^{\top}WH)_{\rho\nu} - (W^{\top}V)_{\rho\nu}).$

36



NMF Optimization Gradient descent

Initialization $H^{(0)} \in \mathbb{R}^{R \times N}$ Iteration for $\ell = 0, 1, 2, ...$

$$H_{rn}^{(\ell+1)} = H_{rn}^{(\ell)} - \gamma_{rn}^{(\ell)} \cdot \left(\left(W^{\top} W H^{(\ell)} \right)_{rn} - \left(W^{\top} V \right)_{rn} \right)$$

with suitable learning rate $\gamma_{rn}^{(\ell)} \ge 0$



NMF Optimization Gradient descent

Initialization $H^{(0)} \in \mathbb{R}^{R \times N}$ Iteration for $\ell = 0, 1, 2, ...$

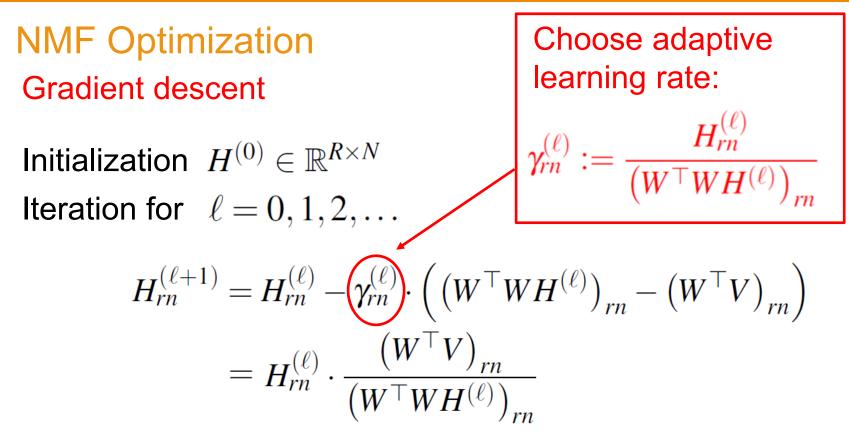
$$H_{rn}^{(\ell+1)} = H_{rn}^{(\ell)} - \gamma_{rn}^{(\ell)} \cdot \left(\left(W^{\top} W H^{(\ell)} \right)_{rn} - \left(W^{\top} V \right)_{rn} \right)$$

with suitable learning rate $\gamma_{rn}^{(\ell)} \ge 0$

Issues:

- How to do the initialization?
- How to choose the learning rate?
- How to ensure nonnegativity?

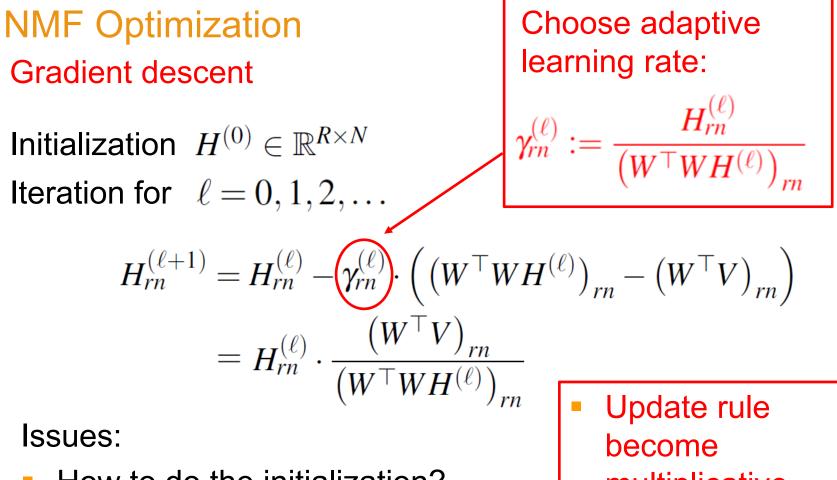




Issues:

- How to do the initialization?
- How to choose the learning rate?
- How to ensure nonnegativity?





- How to do the initialization?
- How to choose the learning rate?
- How to ensure nonnegativity?

- multiplicative
- Nonnegative values stay nonnegative



NMF Optimization

NMF Algorithm

Lee, Seung: Algorithms for Non-Negative Matrix Factorization. Proc. NIPS, 2000.

Algorithm: NMF ($V \approx WH$)

- Input:Nonnegative matrix V of size $K \times N$
Rank parameter $R \in \mathbb{N}$
Threshold ε used as stop criterionOutput:Nonnegative template matrix W of size $K \times R$
 - Nonnegative activation matrix *H* of size $R \times N$

Procedure: Define nonnegative matrices $W^{(0)}$ and $H^{(0)}$ by some random or informed initialization. Furthermore set $\ell = 0$. Apply the following update rules (written in matrix notation):

$$(1) \quad H^{(\ell+1)} = H^{(\ell)} \odot \left(((W^{(\ell)})^\top V) \oslash ((W^{(\ell)})^\top W^{(\ell)} H^{(\ell)}) \right)$$

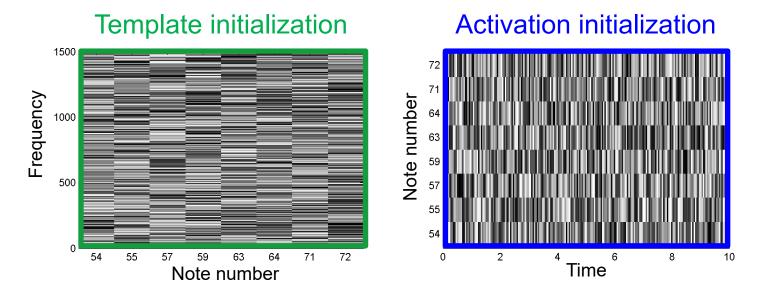
(2)
$$W^{(\ell+1)} = W^{(\ell)} \odot \left((V(H^{(\ell+1)})^{\top}) \oslash (W^{(\ell)}H^{(\ell+1)}(H^{(\ell+1)})^{\top}) \right)$$

(3) Increase ℓ by one.

Repeat the steps (1) to (3) until $||H^{(\ell)} - H^{(\ell-1)}|| \le \varepsilon$ and $||W^{(\ell)} - W^{(\ell-1)}|| \le \varepsilon$ (or until some other stop criterion is fulfilled). Finally, set $H = H^{(\ell)}$ and $W = W^{(\ell)}$.



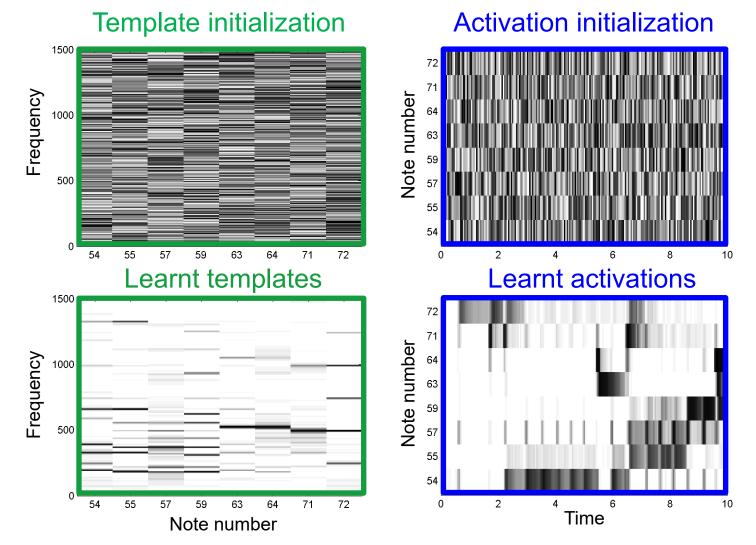
NMF-based Spectrogram Decomposition



Random initialization



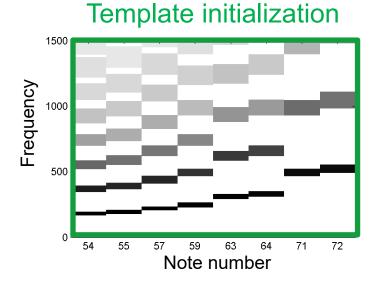
NMF-based Spectrogram Decomposition



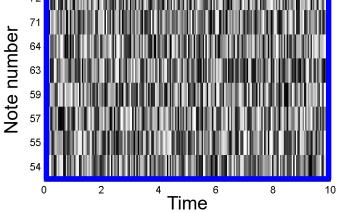
Random initialization \rightarrow No semantic meaning

Nonnegative Autoencoders with Applications to Music Audio Decomposing





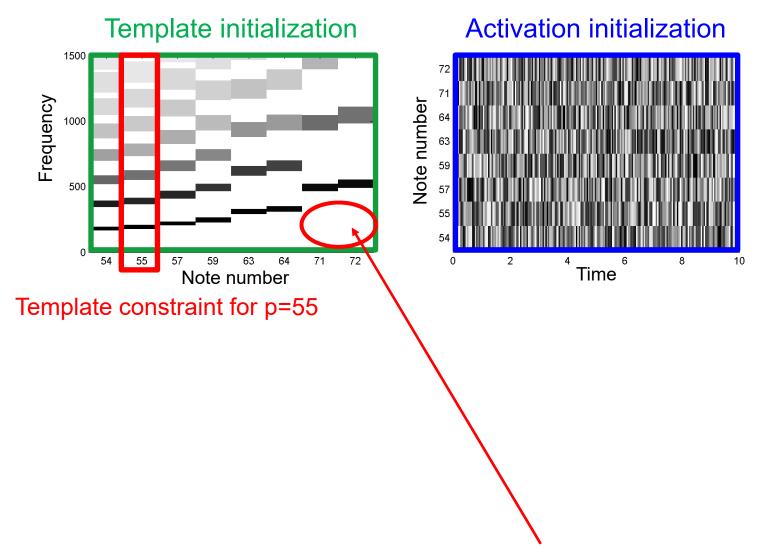
Activation initialization



Enforce harmonic structure with zero-valued entries

Nonnegative Autoencoders with Applications to Music Audio Decomposing

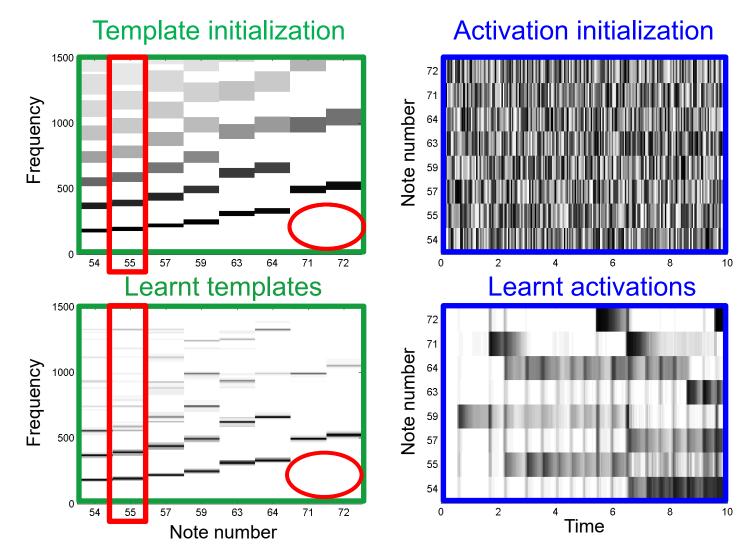




Enforce harmonic structure with zero-valued entries

Nonnegative Autoencoders with Applications to Music Audio Decomposing

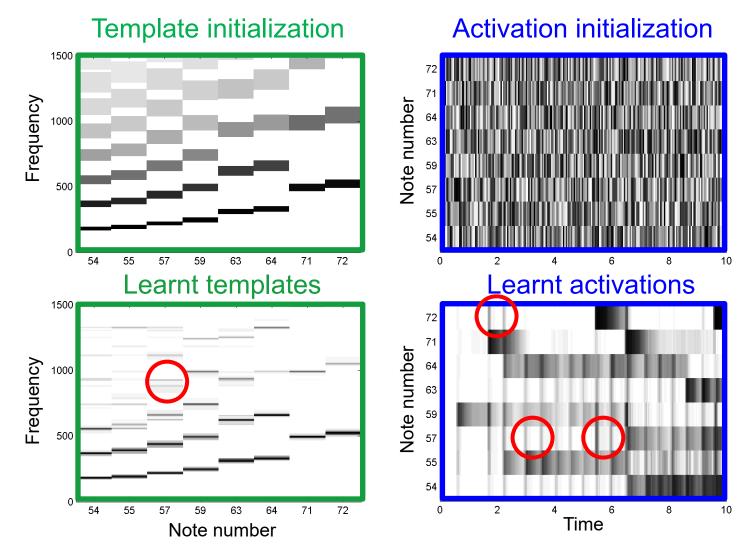




Zero-valued entries remain zero-valued entries!

Nonnegative Autoencoders with Applications to Music Audio Decomposing

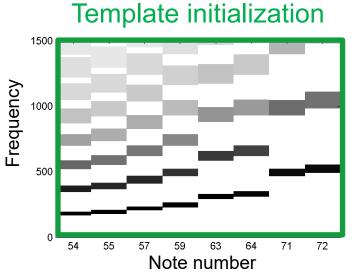




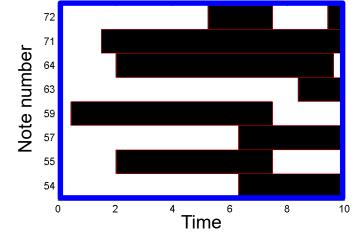
Pitch templates misused to represent onsets

Nonnegative Autoencoders with Applications to Music Audio Decomposing

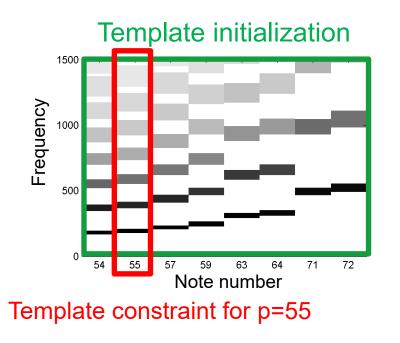




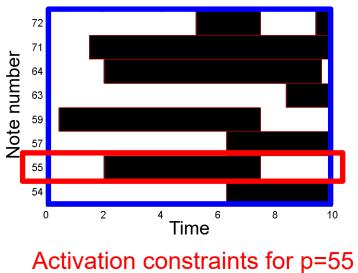
Activation initialization



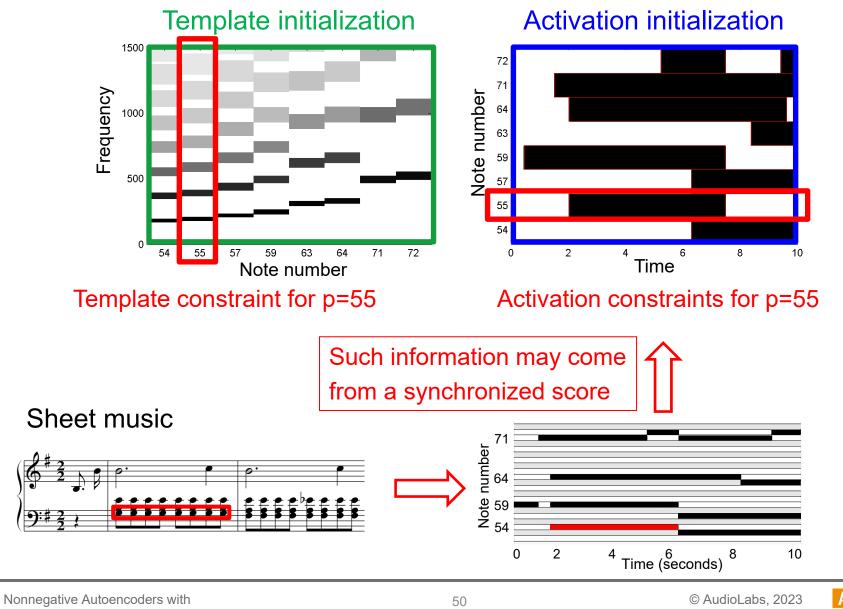




Activation initialization

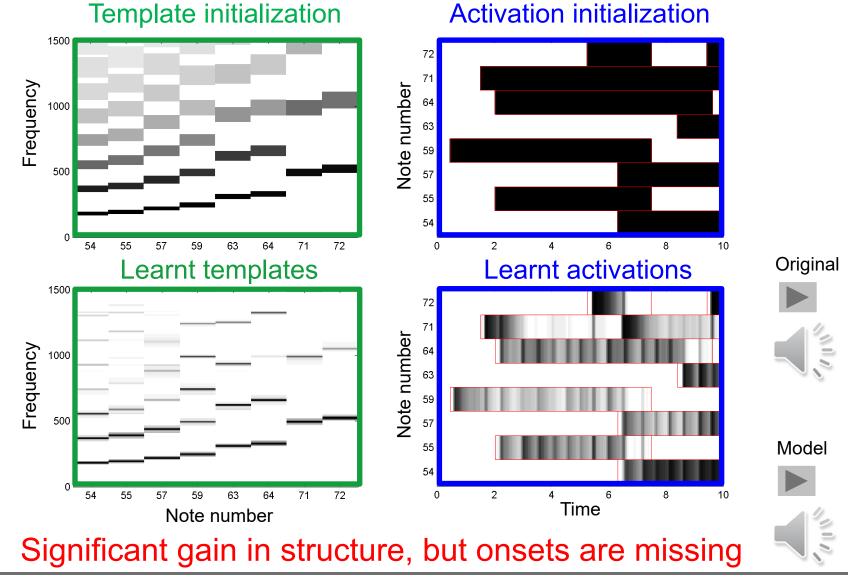






Applications to Music Audio Decomposing

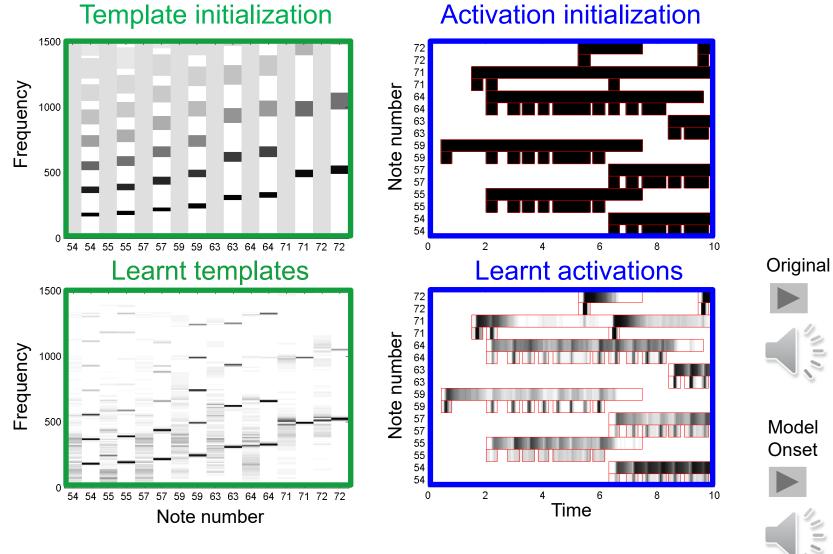




Nonnegative Autoencoders with Applications to Music Audio Decomposing



Constrained NMF: Onset Templates



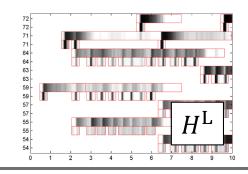


Application: Separating left and right hands for piano



1. Split activation matrix



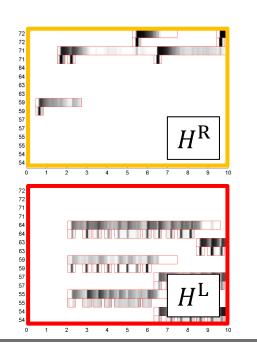




Application: Separating left and right hands for piano

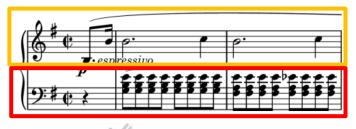


1. Split activation matrix



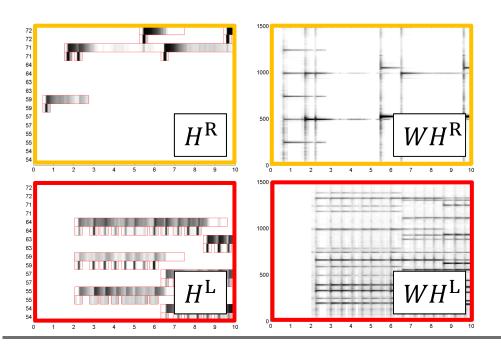


Application: Separating left and right hands for piano





2. Model spectrogram for left/right

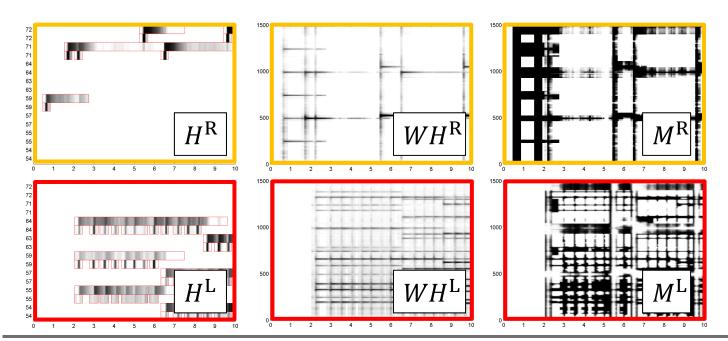




Application: Separating left and right hands for piano

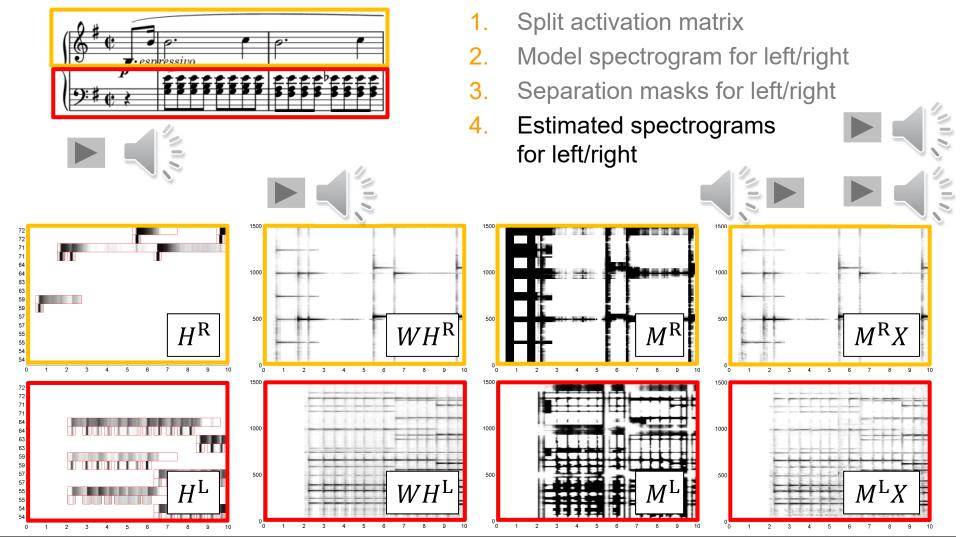


- 1. Split activation matrix
- 2. Model spectrogram for left/right
- 3. Separation masks for left/right





Application: Separating left and right hands for piano





Application: Separating left and right hands for piano

Chopin, Waltz Op. 64, No. 1



Original



Score-Informed Constraints

Ewert, Müller: Using Score-Informed Constraints for NMF-based Source Separation. Proc. ICASSP, 2012.

Further results available at http://www.mpi-inf.mpg.de/resources/MIR/ICASSP2012-ScoreInformedNMF/



Application: Separating left and right hands for piano

Chopin, Waltz Op. 64, No. 1





Score-Informed Constraints

Ewert, Müller: Using Score-Informed Constraints for NMF-based Source Separation. Proc. ICASSP, 2012.

Further results available at http://www.mpi-inf.mpg.de/resources/MIR/ICASSP2012-ScoreInformedNMF/

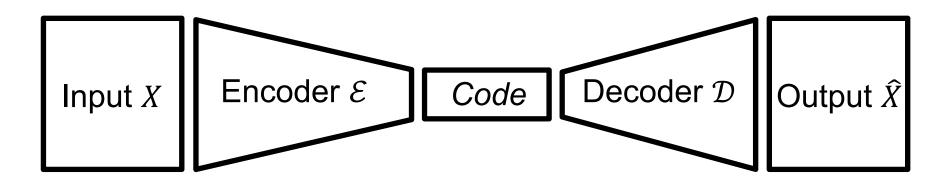


Conclusions (NMF)

- NMF used for spectrogram decomposition
- Multiplicative update rules make it easy to constrain NMF model via zero initialization
- Exploiting score information to guide separation process (requires score—audio synchronization)
- Application: Separation of arbitrary note groups from given audio recording



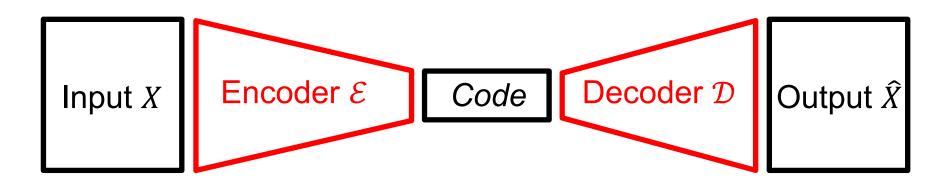
Autoencoder



- Specific type of neural network
- Encoder: Compress input X into a low-dimensional code
- Decoder: Reconstruct output \hat{X} from code



Autoencoder



- Specific type of neural network
- Encoder: Compress input X into a low-dimensional code
- Decoder: Reconstruct output \hat{X} from code
- Goal: Learn parameters for encoder and decoder such that output is close to input with respect to some loss function:

$$\mathcal{L}(X, \hat{X}) \approx 0$$



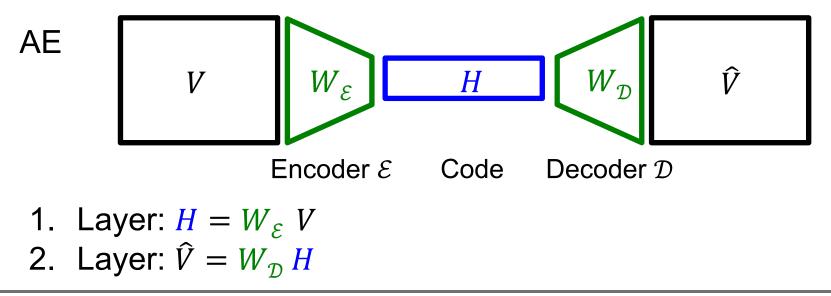
NMF and Autoencoder (AE)Nonnegative Autoencoder
Smaragdis, Venkataramani: A Neural
Network Alternative to Non-Negative
Audio Models, Proc. ICASSP 2017.NMF
$$V$$
 \thickapprox W H $=$ \hat{V}

 $V \approx WH$ implies $W^+V \approx H$ with pseudoinverse W^+

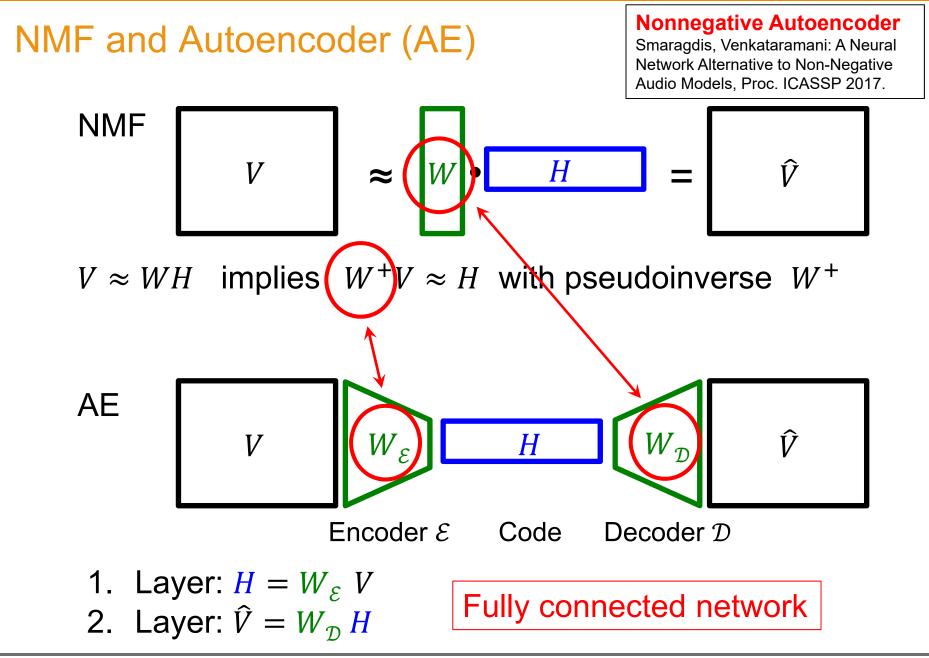


NMF and Autoencoder (AE)Nonnegative Autoencoder
Smaragdis, Venkataramani: A Neural
Network Alternative to Non-Negative
Audio Models, Proc. ICASSP 2017.NMF
$$V$$
 \thickapprox W H $=$ \hat{V}

 $V \approx WH$ implies $W^+V \approx H$ with pseudoinverse W^+



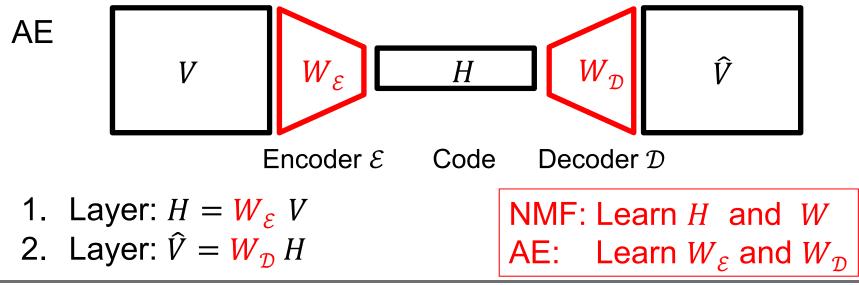




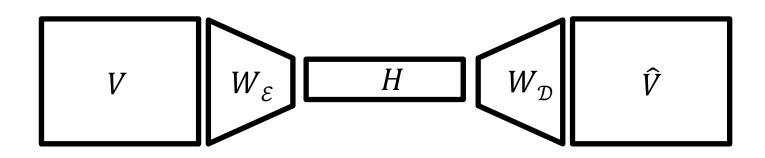


NMF and Autoencoder (AE)Nonnegative Autoencoder
Smaragdis, Venkataramani: A Neural
Network Alternative to Non-Negative
Audio Models, Proc. ICASSP 2017.NMF
$$V$$
 \thickapprox W H $=$ \hat{V}

 $V \approx WH$ implies $W^+V \approx H$ with pseudoinverse W^+





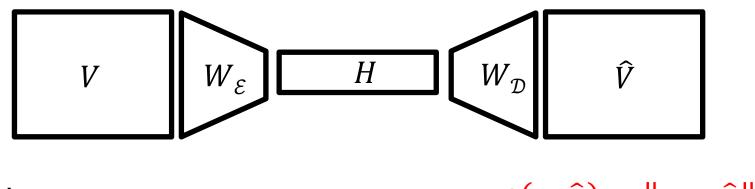


1. Layer: $H = W_{\mathcal{E}} V$ 2. Layer: $\hat{V} = W_{\mathcal{D}} H$

- How can one adjust the AE to simulate NMF?
- How can one achieve nonnegativity?
- How can one incorporate musical knowledge?

Nonnegative Autoencoders with Applications to Music Audio Decomposing





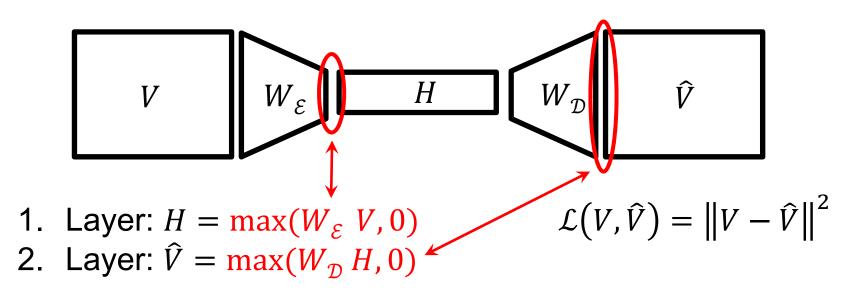
1. Layer:
$$H = W_{\mathcal{E}} V$$

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

 $\mathcal{L}(V,\widehat{V}) = \left\|V - \widehat{V}\right\|^2$

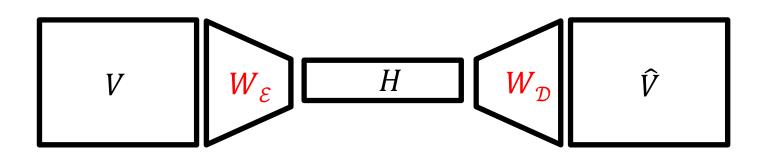
Loss function: same as in NMF





- Loss function: same as in NMF
- Activation function (ReLU) makes H and \hat{V} nonnegative



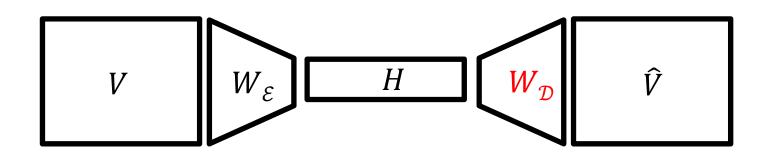


1. Layer: $H = \max(W_{\mathcal{E}} V, 0)$ $\mathcal{L}(V, \hat{V}) = \|V - \hat{V}\|^2$ 2. Layer: $\hat{V} = \max(W_{\mathcal{D}} H, 0)$ $W_{\mathcal{D}} \leftarrow \max\left(W_{\mathcal{D}} - \gamma \frac{\partial \mathcal{L}}{\partial W_{\mathcal{D}}}, 0\right)$

- Loss function: same as in NMF
- Activation function (ReLU) makes H and \hat{V} nonnegative
- Projected gradient descent can be used to keep $W_{\mathcal{D}}$ (and $W_{\mathcal{E}}$) nonnegative



Musical Constraints

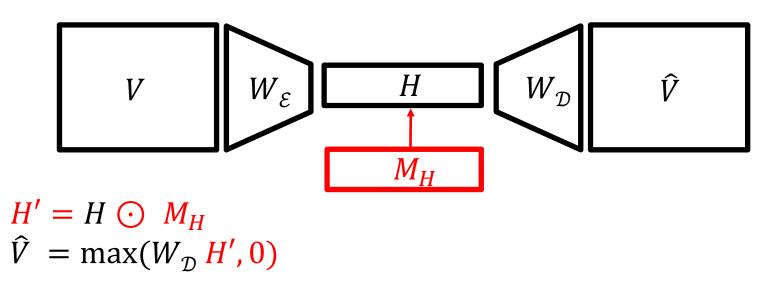


- $H = \max(W_{\mathcal{E}} \ V, 0)$ $\hat{V} = \max(W_{\mathcal{D}} \ H, 0)$
- Template constraints: Project certain entries in $W_{\mathcal{D}}$ to zero values (using projected gradient decent)



Musical Constraints

Ewert, Sandler: Structured Dropout for Weak Label and Multi-Instance Learning and Its Application to Score-Informed Source Separation. Proc. ICASSP, 2017.



- Template constraints: Project certain entries in $W_{\mathcal{D}}$ to zero values (using projected gradient decent)
- Activation constraints: Use structured dropout by applying pointwise multiplication with binary mask M_H



NAE with Multiplicative Update Rules

- Multiplicative update rules in NMF:
 - Preserve nonnegativity
 - Lead to fast convergence
- Question: Can one introduce multiplicative update rules to train network weights for NAE?
- Use in additive gradient descent

$$W^{(\ell+1)} = W^{(\ell)} - \gamma \cdot \frac{\partial \mathcal{L}}{\partial W}$$

a suitable (adaptive) learning rate γ .



NAE with Multiplicative Update Rules

Encoder:

$$H = W_{\mathcal{E}}V$$

Structured Dropout:

 $H' = H \odot M_H$

- Decoder: $\hat{V} = W_{\mathcal{D}}H'$

NMF vs. NAE

Özer, Hansen, Zunner, Müller: Investigating Nonnegative Autoencoders for Efficient Audio Decomposition. Proc. EUSIPCO, 2022.



NAE with Multiplicative Update Rules

Encoder:

Decoder:

$$H = W_{\mathcal{E}}V$$

$$W_{\mathcal{E},rk}^{(\ell+1)} = W_{\mathcal{E},rk}^{(\ell)} \cdot \frac{\left(\left(\left(W_{\mathcal{D}}^{\top}V\right) \odot M_{H}\right)V^{\top}\right)_{rk}}{\left(\left(\left(\left(W_{\mathcal{D}}^{\top}W_{\mathcal{D}}H'^{(\ell)}\right) \odot M_{H}\right)V^{\top}\right)_{rk}\right)_{rk}}$$

• Structured Dropout: $H' = H \odot M_H$

 $\hat{V} = W_{\mathcal{D}}H'$

$$W_{\mathcal{D},kr}^{(\ell+1)} = W_{\mathcal{D},kr}^{(\ell)} \cdot \frac{\left(VH^{\prime\top}\right)_{kr}}{\left(W_{\mathcal{D}}^{(\ell)}H^{\prime}H^{\prime\top}\right)_{kr}}$$

Similar idea and computation as for NMF.

NMF vs. NAE

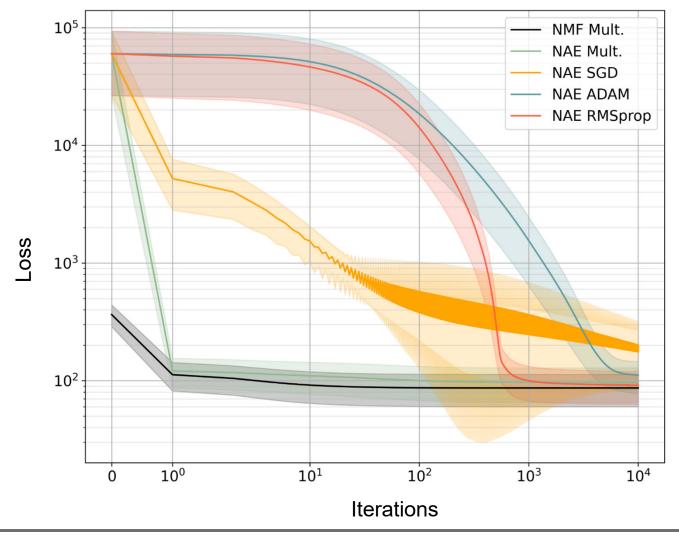
Özer, Hansen, Zunner, Müller: Investigating Nonnegative Autoencoders for Efficient Audio Decomposition. Proc. EUSIPCO, 2022.



Approximation Loss

NMF vs. NAE

Özer, Hansen, Zunner, Müller: Investigating Nonnegative Autoencoders for Efficient Audio Decomposition. Proc. EUSIPCO, 2022.

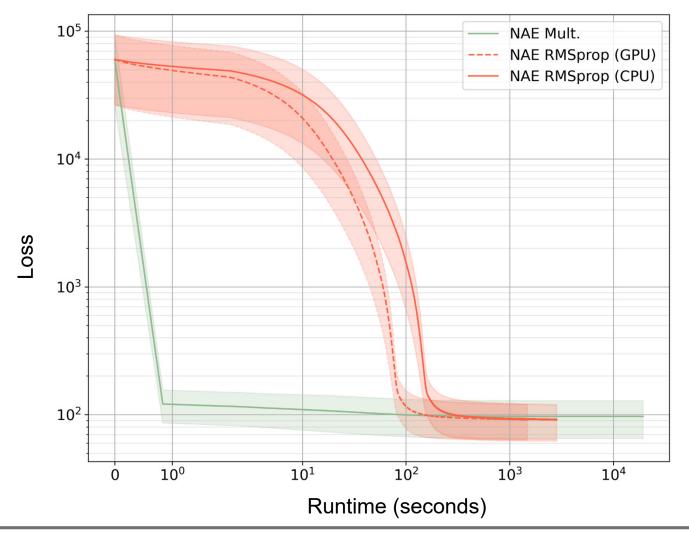




Approximation Loss

NMF vs. NAE

Özer, Hansen, Zunner, Müller: Investigating Nonnegative Autoencoders for Efficient Audio Decomposition. Proc. EUSIPCO, 2022.





Conclusions (NAE)

- Simulation of NMF:
 - Decoder corresponds to NMF templates
 - Encoder learns a kind of pseudo-inverse
 - Code corresponds to NMF activations
- Nonnegativity can be achieved via
 - activation function (ReLU)
 - projected gradient descent
 - multiplicative update rules
- Musical knowledge can be integrated via
 - removing network weights (template constraints)
 - structured dropout (activation constraints)



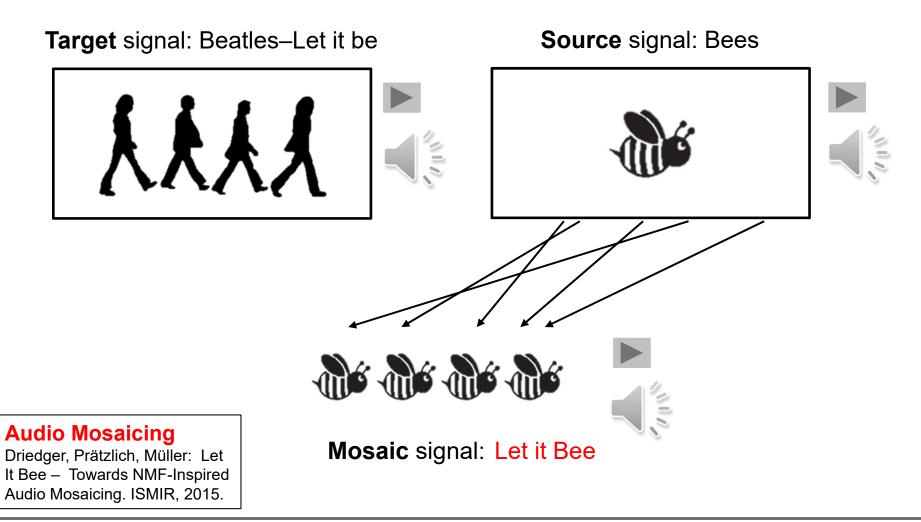
Outlook

- More complex networks
 - Deeper networks (more layers)
 - Different layer types (CNN, RNN, ...) and activation functions
 - Modification of loss function and regularization terms
- Understanding encoder decoder relationship
 - Nonnegativity
 - Pseudo-inverse
- Update rules
 - Constraints and convergence issues
 - Adaptive learning rates and projected gradient descent



Score-Informed Audio Decomposition

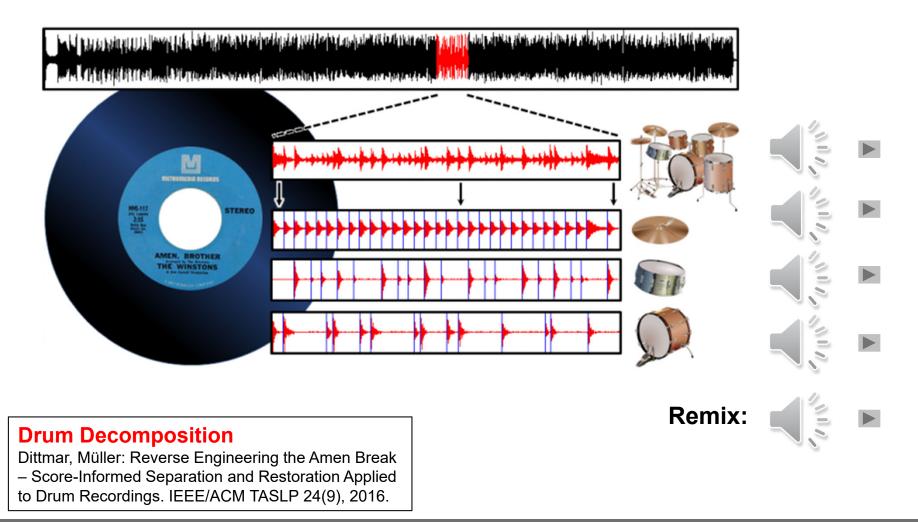
Audio mosaicing (style transfer)





Score-Informed Audio Decomposition

Informed Drum-Sound Decomposition





Score-Informed Audio Decomposition

Major challenge: Reconstructed sound events often have artifacts

Approaches:

- Resynthesize certain sound components
- Differentiable Digital Signal Processing (DDSP) combines classical DSP and deep learning
- Generative adversarial networks may help to reduce the artifacts

DDSP

Engel et al.: DDSP: Differentiable Digital Signal Processing. ICLR, 2020.



- Yigitcan Özer
- PhD student in engineering
- Pianist





- Yigitcan Özer
- PhD student in engineering
- Pianist



Only Piano!

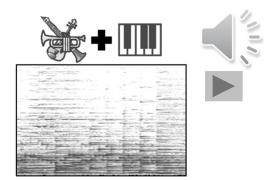


Where is the orchestra?

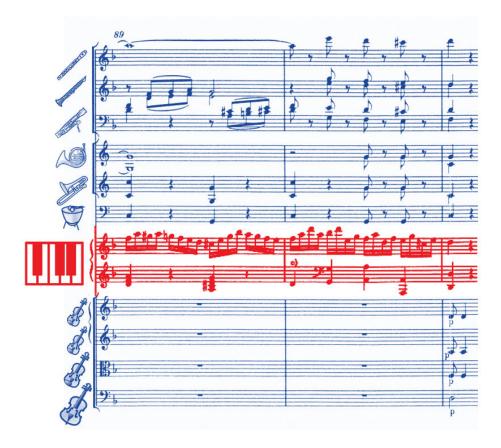


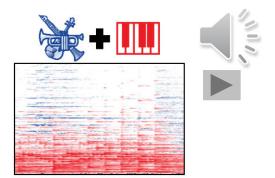




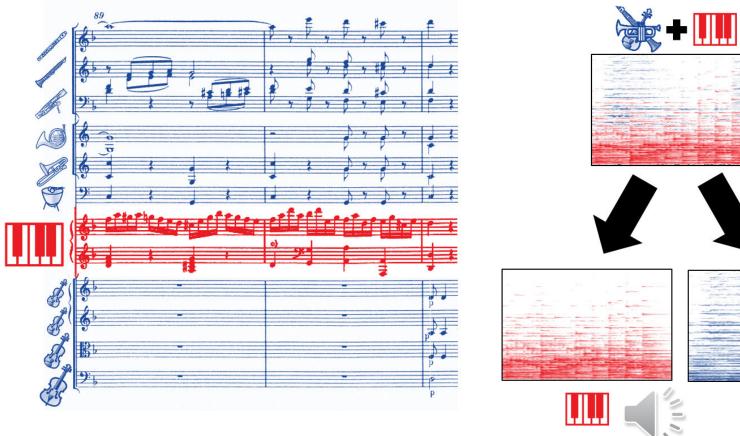


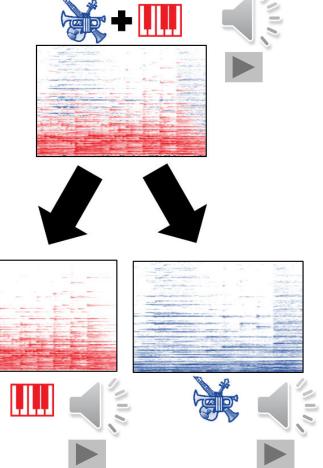




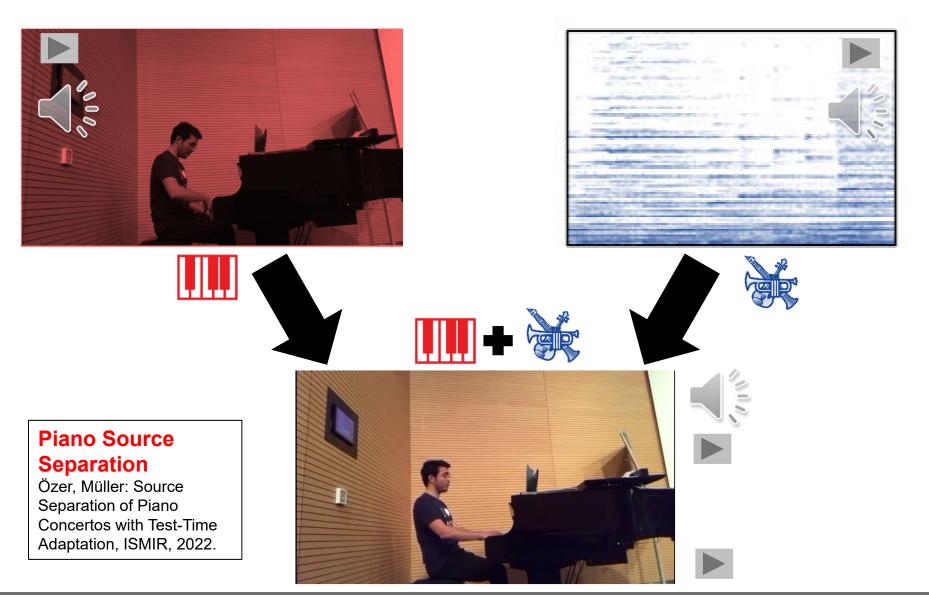






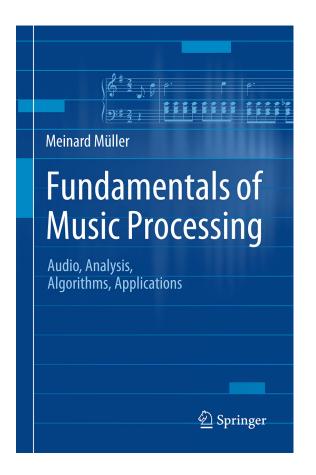








Fundamentals of Music Processing (FMP)

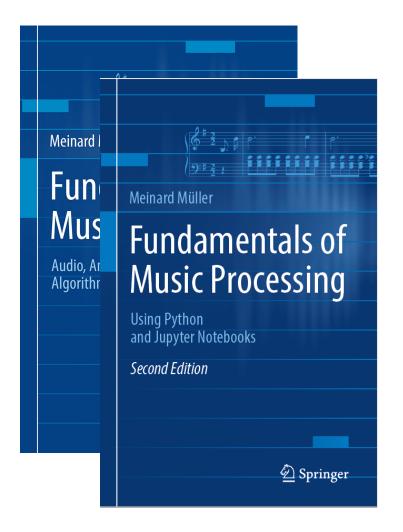


Meinard Müller Fundamentals of Music Processing Audio, Analysis, Algorithms, Applications Springer, 2015

Accompanying website: www.music-processing.de



Fundamentals of Music Processing (FMP)



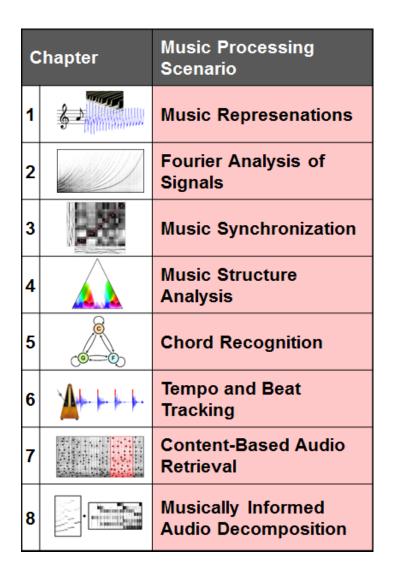
Meinard Müller Fundamentals of Music Processing Audio, Analysis, Algorithms, Applications Springer, 2015

Accompanying website: www.music-processing.de

2nd edition Meinard Müller Fundamentals of Music Processing Using Python and Jupyter Notebooks Springer, 2021



Fundamentals of Music Processing (FMP)



Meinard Müller Fundamentals of Music Processing Audio, Analysis, Algorithms, Applications Springer, 2015

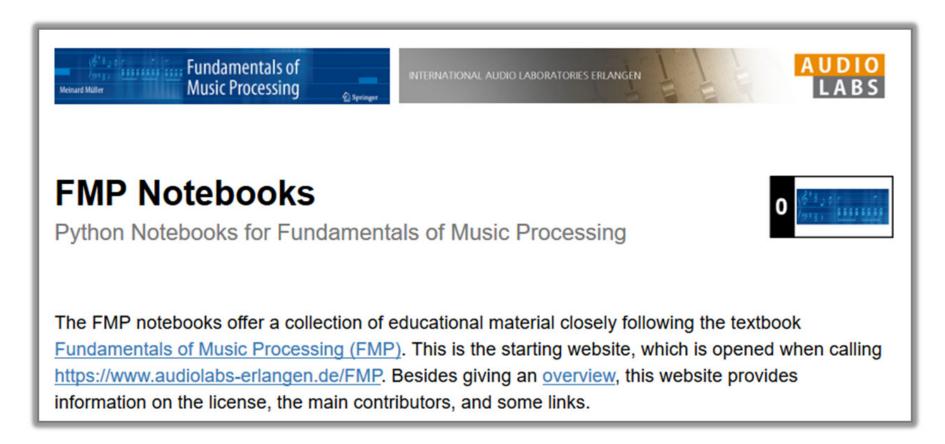
Accompanying website: www.music-processing.de

2nd edition Meinard Müller Fundamentals of Music Processing Using Python and Jupyter Notebooks Springer, 2021

91



FMP Notebooks: Education & Research



https://www.audiolabs-erlangen.de/FMP



Resources (Group Meinard Müller)

FMP Notebooks:

https://www.audiolabs-erlangen.de/FMP

libfmp:

https://github.com/meinardmueller/libfmp

synctoolbox:

https://github.com/meinardmueller/synctoolbox

libtsm:

https://github.com/meinardmueller/libtsm

Preparation Course Python (PCP) Notebooks:

https://www.audiolabs-erlangen.de/resources/MIR/PCP/PCP.html

https://github.com/meinardmueller/PCP



Resources

librosa:

https://librosa.org/

• madmom:

https://github.com/CPJKU/madmom

Essentia Python tutorial:

https://essentia.upf.edu/essentia_python_tutorial.html

mirdata:

https://github.com/mir-dataset-loaders/mirdata

• open-unmix:

https://github.com/sigsep/open-unmix-pytorch

• Open Source Tools & Data for Music Source Separation:

https://source-separation.github.io/tutorial/landing.html



ESSENTIA







References (FMP Textbook & Notebooks)

- Meinard Müller: Fundamentals of Music Processing Using Python and Jupyter Notebooks. 2nd Edition, Springer, 2021. https://www.springer.com/gp/book/9783030698072
- Meinard Müller and Frank Zalkow: libfmp: A Python Package for Fundamentals of Music Processing. Journal of Open Source Software (JOSS), 6(63): 1–5, 2021.
 https://joss.theoj.org/papers/10.21105/joss.03326
- Meinard Müller: An Educational Guide Through the FMP Notebooks for Teaching and Learning Fundamentals of Music Processing. Signals, 2(2): 245–285, 2021. <u>https://www.mdpi.com/2624-6120/2/2/18</u>
- Meinard Müller and Frank Zalkow: FMP Notebooks: Educational Material for Teaching and Learning Fundamentals of Music Processing. Proc. International Society for Music Information Retrieval Conference (ISMIR): 573–580, 2019. <u>https://zenodo.org/record/3527872#.YOhEQOgzaUk</u>
- Meinard Müller, Brian McFee, and Katherine Kinnaird: Interactive Learning of Signal Processing Through Music: Making Fourier Analysis Concrete for Students. IEEE Signal Processing Magazine, 38(3): 73–84, 2021.
 https://ieeexplore.ieee.org/document/9418542



References (NMF, NAE)

- Daniel Lee and Sebastian Seung: Algorithms for Non-Negative Matrix Factorization. Proc. NIPS, 2000.
- Sebastian Ewert and Meinard Müller: Using Score-Informed Constraints for NMF-Based Source Separation. Proc. ICASSP, 2012.
- Paris Smaragdis and Shrikant Venkataramani: A Neural Network Alternative to Non-Negative Audio Models. Proc. ICASSP, 2017.
- Sebastian Ewert and Mark B. Sandler: Structured Dropout for Weak Label and Multi-Instance Learning and Its Application to Score-Informed Source Separation. Proc. ICASSP, 2017.
- Yigitcan Özer, Jonathan Hansen, Tim Zunner, and Meinard Müller: Investigating Nonnegative Autoencoders for Efficient Audio Decomposition. Proc. EUSIPCO, 2022.

