Temporal models with low-rank spectrograms

Cédric Févotte
Institut de Recherche en Informatique de Toulouse (IRIT)

IEEE MLSP
Aalborg, Sep. 2018

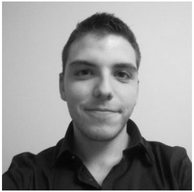
Outline

Collaborators

Low-rank time-frequency synthesis

Matthieu Kowalski
(Paris-Saclay University)

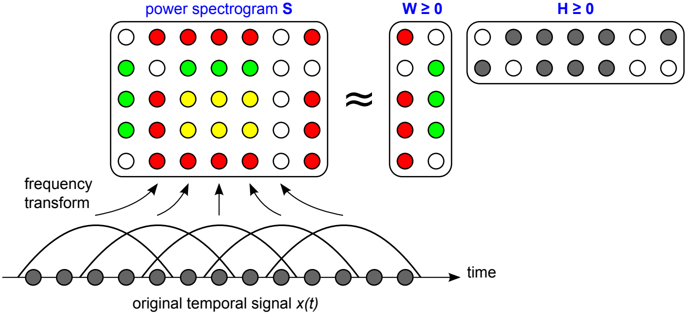
NMF with transform learning



Dylan Fagot Herwig Wendt
(CNRS, IRIT, Toulouse)

NMF for audio spectral unmixing

(Smaragdis and Brown, 2003)



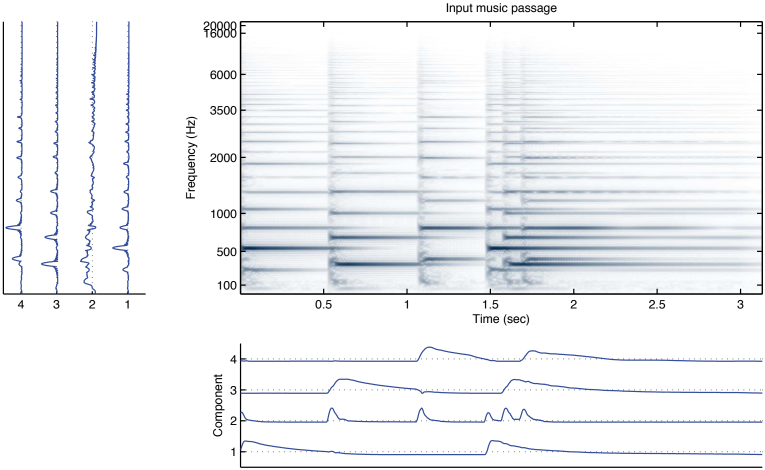
- y_{fn}: short-time Fourier transform (STFT) of temporal signal $x(t)$.
- $s_{fn} = |y_{fn}|^2$: power spectrogram.
- NMF extracts recurring spectral patterns from the data by solving

$$\min_{\mathbf{W},\mathbf{H}\geq 0} D(\mathbf{S}|\mathbf{WH}) = \sum_{fn} d(s_{fn}|[\mathbf{WH}]_{fn})$$

- Successful applications in audio source separation and music transcription.

NMF for audio spectral unmixing

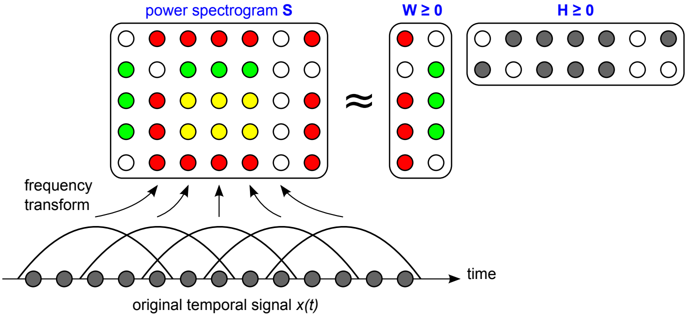
(Smaragdis and Brown, 2003)



reproduced from (Smaragdis, 2013)

NMF for audio spectral unmixing

(Smaragdis and Brown, 2003)



- What is the right time-frequency transform **S** ?
- What is the right measure of fit $D(\mathbf{S}|\mathbf{WH})$?
- NMF approximates the spectrogram by a sum of rank-one spectrograms. How to reconstruct temporal components ? What about phase ?

Itakura-Saito NMF

(Févotte, Bertin, and Durrieu, 2009)

- Gaussian low-rank variance model of the complex-valued STFT:

$$y_{fn} \sim N_c(0, [\mathbf{WH}]_{fn})$$

Related work by (Benaroya et al., 2003; Abdallah and Plumbley, 2004; Parry and Essa, 2007)

Itakura-Saito NMF
(Févotte, Bertin, and Durrieu, 2009)

- Gaussian low-rank variance model of the complex-valued STFT:

$$y_{fn} \sim N_c(0, [\mathbf{WH}]_{fn})$$

- Log-likelihood equivalent to Itakura-Saito (IS) divergence:

$$-\log p(\mathbf{Y}|\mathbf{WH}) = D_{\mathrm{IS}}(|\mathbf{Y}|^2|\mathbf{WH}) + \mathrm{cst}$$

Related work by (Benaroya et al., 2003; Abdallah and Plumbley, 2004; Parry and Essa, 2007)

Itakura-Saito NMF

(Févotte, Bertin, and Durrieu, 2009)

- Gaussian low-rank variance model of the complex-valued STFT:

$$y_{fn} \sim N_c(0, [\mathbf{WH}]_{fn})$$

- Log-likelihood equivalent to Itakura-Saito (IS) divergence:

$$-\log p(\mathbf{Y}|\mathbf{WH}) = D_{\mathrm{IS}}(|\mathbf{Y}|^2|\mathbf{WH}) + \mathrm{cst}$$

- Zero-mean assumption: $\mathrm{E}[x(t)] = 0$ implies $\mathrm{E}[y_{fn}] = 0$ by linearity.

Related work by (Benaroya et al., 2003; Abdallah and Plumbley, 2004; Parry and Essa, 2007)

Itakura-Saito NMF

(Févotte, Bertin, and Durrieu, 2009)

- Gaussian low-rank variance model of the complex-valued STFT:

$$y_{fn} \sim N_c(0, [\mathbf{WH}]_{fn})$$

- Log-likelihood equivalent to Itakura-Saito (IS) divergence:

$$-\log p(\mathbf{Y}|\mathbf{WH}) = D_{\text{IS}}(|\mathbf{Y}|^2|\mathbf{WH}) + \text{cst}$$

- Zero-mean assumption: $\mathrm{E}\left[x(t)\right] = 0$ implies $\mathrm{E}\left[y_{fn}\right] = 0$ by linearity.
- Underlies a Gaussian composite model (GCM):

$$y_{fn} = \sum_k z_{kfn},$$

$$z_{kfn} \sim N_c(0, w_{fk} h_{kn})$$

Related work by (Benaroya et al., 2003; Abdallah and Plumbley, 2004; Parry and Essa, 2007)

Itakura-Saito NMF

(Févotte, Bertin, and Durrieu, 2009)

- Gaussian low-rank variance model of the complex-valued STFT:

$$y_{fn} \sim N_c(0, [\mathbf{WH}]_{fn})$$

- Log-likelihood equivalent to Itakura-Saito (IS) divergence:

$$-\log p(\mathbf{Y}|\mathbf{WH}) = D_{\text{IS}}(|\mathbf{Y}|^2|\mathbf{WH}) + \text{cst}$$

- Zero-mean assumption: $\text{E}[x(t)] = 0$ implies $\text{E}[y_{fn}] = 0$ by linearity.
- Underlies a Gaussian composite model (GCM):

$$y_{fn} = \sum_k z_{kfn},$$

$$z_{kfn} \sim N_c(0, w_{fk} h_{kn})$$

- Latent STFT components can be estimated a posteriori by Wiener filter:

$$\hat{z}_{kfn} = \text{E}[z_{kfn}|\mathbf{Y}, \mathbf{W}, \mathbf{H}] = \frac{w_{fk} h_{kn}}{[\mathbf{WH}]_{fn}} y_{fn}$$

Related work by (Benaroya et al., 2003; Abdallah and Plumbley, 2004; Parry and Essa, 2007)

Itakura-Saito NMF

(Févotte, Bertin, and Durrieu, 2009)

- Gaussian low-rank variance model of the complex-valued STFT:

$$y_{fn} \sim N_c(0, [\mathbf{WH}]_{fn})$$

- Log-likelihood equivalent to Itakura-Saito (IS) divergence:

$$-\log p(\mathbf{Y}|\mathbf{WH}) = D_{\text{IS}}(|\mathbf{Y}|^2|\mathbf{WH}) + \text{cst}$$

- Zero-mean assumption: $\mathrm{E}[x(t)] = 0$ implies $\mathrm{E}[y_{fn}] = 0$ by linearity.
- Underlies a Gaussian composite model (GCM):

$$y_{fn} = \sum_k z_{kfn},$$

$$z_{kfn} \sim N_c(0, w_{fk} h_{kn})$$

- Latent STFT components can be estimated a posteriori by Wiener filter:

$$\hat{z}_{kfn} = \mathrm{E}[z_{kfn}|\mathbf{Y}, \mathbf{W}, \mathbf{H}] = \frac{w_{fk} h_{kn}}{[\mathbf{WH}]_{fn}} y_{fn}$$

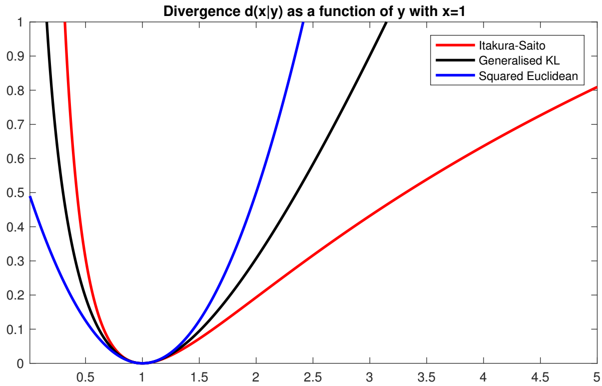
- Inverse-STFT of $\{\hat{z}_{kfn}\}_{fn}$ produces temporal components such that:

$$x(t) = \sum_k \hat{c}_k(t)$$

Related work by (Benaroya et al., 2003; Abdallah and Plumbley, 2004; Parry and Essa, 2007)

The Itakura-Saito divergence

(Itakura and Saito, 1968; Gray et al., 1980)



IS divergence: $d_{\mathrm{IS}}(x|y) = \frac{x}{y} - \log \frac{x}{y} - 1$

- Nonconvex in y.
- Scale-invariance: $d_{\mathrm{IS}}(\lambda\, x|\lambda\, y) = d_{\mathrm{IS}}(x|y)$
  Very relevant for spectral data with high dynamic range.
  In comparison, $d_{\mathrm{Euc}}(\lambda\, x|\lambda\, y) = \lambda^2\, d_{\mathrm{Euc}}(x|y)$, $d_{\mathrm{KL}}(\lambda\, x|\lambda\, y) = \lambda\, d_{\mathrm{KL}}(x|y)$.

Optimisation for IS-NMF

(Févotte and Idier, 2011)

Objective

$$\min_{\mathbf{W},\mathbf{H}\geq 0} D(\mathbf{S}|\mathbf{WH}) = \sum_{fn} \left[\frac{s_{fn}}{[\mathbf{WH}]_{fn}} - \log \frac{s_{fn}}{[\mathbf{WH}]_{fn}} - 1 \right]$$

$$= \sum_{fn} \left[\frac{s_{fn}}{[\mathbf{WH}]_{fn}} + \log [\mathbf{WH}]_{fn} \right] + \text{cst}$$

State of the art

- Block-coordinate descent $(\mathbf{W}, \mathbf{H})$ with majorisation-minimisation (MM)
- Updates of $\mathbf{W}$ and $\mathbf{H}$ equivalent by transposition of $\mathbf{S}$
- MM leads to multiplicative updates (linear complexity per iteration)

$$h_{kn} \leftarrow h_{kn} \frac{\sum_f w_{fk} s_{fn} [\mathbf{WH}]_{fn}^{-2}}{\sum_f w_{fk} [\mathbf{WH}]_{fn}^{-1}}$$

- Nonconvex problem (because of bilinearity and the divergence), initialisation matters.

Piano toy example

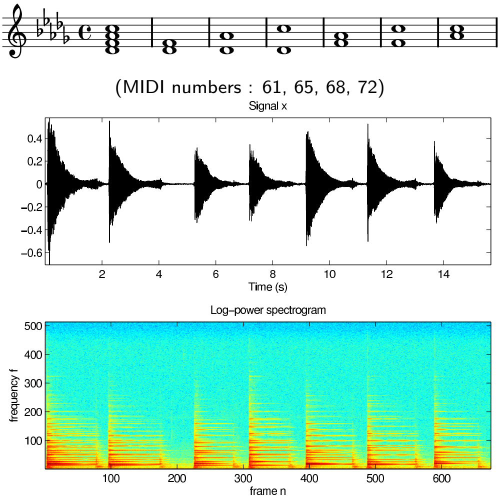
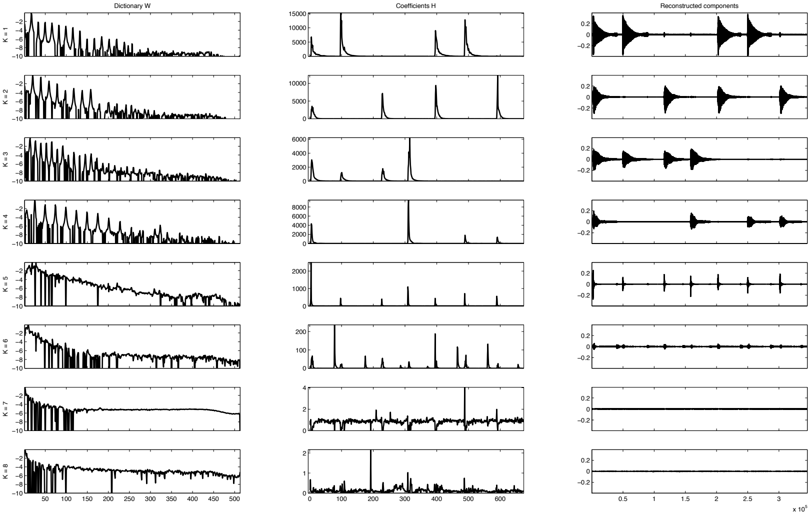


Figure: Three representations of data.

demo available at `https://www.irit.fr/~Cedric.Fevotte/extras/machine_audition/`

Piano toy example

IS-NMF on power spectrogram with $K = 8$

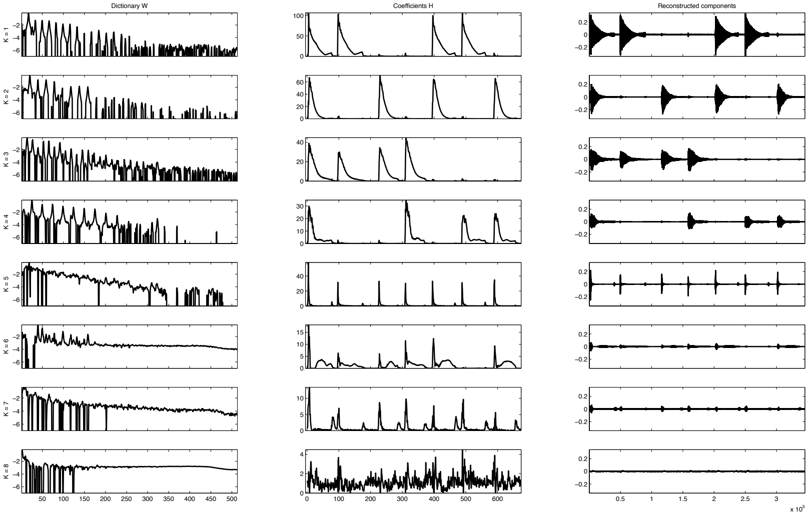


Pitch estimates: 65.0 68.0 61.0 72.0 0 0 0 0

(True values: 61, 65, 68, 72)

Piano toy example

KL-NMF on magnitude spectrogram with $K = 8$

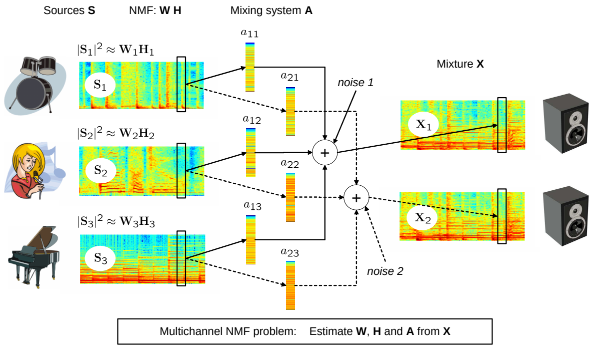


Pitch estimates: 65.2 68.2 61.0 72.2 0 56.2 0 0

(True values: 61, 65, 68, 72)

Follow-up on IS-NMF

- penalised versions promoting sparsity or dynamics (Lefèvre, Bach, and Févotte, 2011a; Févotte, 2011; Févotte, Le Roux, and Hershey, 2013)
- model order selection (Tan and Févotte, 2013)
- online/incremental variants (Dessein et al., 2010; Lefèvre et al., 2011b)
- Bayesian approaches (Hoffman et al., 2010; Dikmen and Févotte, 2011; Turner and Sahani, 2014)
- full-covariance models (Liutkus et al., 2011; Yoshii et al., 2013)
- improved phase models (Badeau, 2011; Magron, Badeau, and David, 2017)
- multichannel variants (Ozerov and Févotte, 2010; Sawada et al., 2013; Kounades-Bastian et al., 2016; Leglaive et al., 2016)



Outline

Analysis vs synthesis

- IS-NMF is a generative model of the STFT but not of the raw signal itself.
- Low-rank time-frequency synthesis (LRTFS) fills in this ultimate gap.
- STFT is an analysis transform

$$y_{fn} = \sum_t x(t)\phi^*_{fn}(t)$$

- LRTFS is a synthesis model

$$x(t) = \sum_{fn} \alpha_{fn}\, \phi_{fn}(t) + e(t)$$

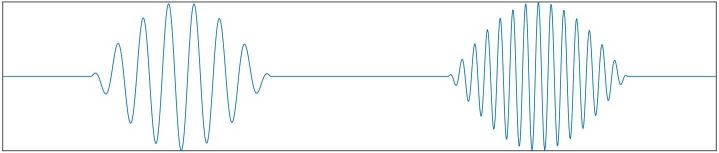


Figure: Two Gabor atoms $\phi_{fn}(t)$

Low-rank time frequency synthesis (LRTFS)

(Févotte and Kowalski, 2014)

- Gaussian low-rank variance model of the synthesis coefficients:

$$x(t) = \sum_{fn} \alpha_{fn}\, \phi_{fn}(t) + e(t),$$
$$\alpha_{fn} \sim N_c(0, [\mathbf{WH}]_{fn}),$$
$$e(t) \sim N_c(0, \lambda).$$

- LRTFS is a generative model of raw signal $x(t)$.
- Like in IS-NMF, latent composite structure of the synthesis coefficients:

$$\alpha_{fn} = \sum_{k} z_{kfn},$$
$$z_{kfn} \sim N_c(0, w_{fk}\, h_{kn}).$$

- Given $\{\hat{z}_{kfn}\}_{fn}$, temporal components can be reconstructed as

$$\hat{c}_k(t) = \sum_{fn} \hat{z}_{kfn}\, \phi_{fn}(t).$$

Relation to sparse Bayesian learning (SBL)

- Generative signal model in vector/matrix form:

$$\mathbf{x} = \mathbf{\Phi}\boldsymbol{\alpha} + \mathbf{e}.$$

  - $\mathbf{x}$, $\mathbf{e}$: vectors of signal and residual time samples (size T),
  - $\boldsymbol{\alpha}$: vector of synthesis coefficients α_{fn} (size FN),
  - $\mathbf{\Phi}$: time-frequency dictionary (size $T \times FN$).

- Synthesis coefficients model in vector/matrix form:

$$p(\boldsymbol{\alpha}|\mathbf{v}) = N_c(\boldsymbol{\alpha}|\mathbf{0}, \operatorname{diag}(\mathbf{v})).$$

  - $\mathbf{v}$: vector of variance coefficients $v_{fn} = [\mathbf{WH}]_{fn}$ (size FN).

- Similar to sparse Bayesian learning (Tipping, 2001; Wipf and Rao, 2004) except that the variance parameters are tied together by the low-rank structure $\mathbf{WH}$.

Maximum joint likelihood estimation in LRTFS

- Optimise

$$
\begin{aligned}
C(\boldsymbol{\alpha}, \mathbf{W}, \mathbf{H}) &\stackrel{\text{def}}{=} -\log p(\mathbf{x}, \boldsymbol{\alpha} | \mathbf{W}, \mathbf{H}, \lambda) \\
&= \frac{1}{\lambda} \|\mathbf{x} - \boldsymbol{\Phi}\boldsymbol{\alpha}\|_2^2 + \sum_{fn} \left[\frac{|\alpha_{fn}|^2}{[\mathbf{WH}]_{fn}} + \log\, [\mathbf{WH}]_{fn} \right] + \text{cst}
\end{aligned}
$$

- Block coordinate descent $(\boldsymbol{\alpha}, \mathbf{W}, \mathbf{H})$

Maximum joint likelihood estimation in LRTFS

- Optimise

$$C(\boldsymbol{\alpha}, \mathbf{W}, \mathbf{H}) \stackrel{\text{def}}{=} -\log p(\mathbf{x}, \boldsymbol{\alpha} | \mathbf{W}, \mathbf{H}, \lambda)$$
$$= \frac{1}{\lambda}\|\mathbf{x} - \boldsymbol{\Phi}\boldsymbol{\alpha}\|_2^2 + \sum_{fn}\left[\frac{|\alpha_{fn}|^2}{[\mathbf{WH}]_{fn}} + \log\,[\mathbf{WH}]_{fn}\right] + \text{cst}$$

- Block coordinate descent $(\boldsymbol{\alpha}, \mathbf{W}, \mathbf{H})$

Optimisation of $\boldsymbol{\alpha}$

$$\min_{\boldsymbol{\alpha} \in \mathbb{C}^M} \frac{1}{\lambda}\|\mathbf{x} - \boldsymbol{\Phi}\boldsymbol{\alpha}\|_2^2 + \sum_{fn} \frac{|\alpha_{fn}|^2}{[\mathbf{WH}]_{fn}}$$

Ridge regression with complex-valued FISTA (Chaâri et al., 2011; Florescu et al., 2014)

Maximum joint likelihood estimation in LRTFS

- Optimise

$$C(\boldsymbol{\alpha}, \mathbf{W}, \mathbf{H}) \stackrel{\text{def}}{=} -\log p(\mathbf{x}, \boldsymbol{\alpha} | \mathbf{W}, \mathbf{H}, \lambda)$$
$$= \frac{1}{\lambda} \|\mathbf{x} - \boldsymbol{\Phi}\boldsymbol{\alpha}\|_2^2 + \sum_{fn} \left[\frac{|\alpha_{fn}|^2}{[\mathbf{WH}]_{fn}} + \log [\mathbf{WH}]_{fn} \right] + \text{cst}$$

- Block coordinate descent $(\boldsymbol{\alpha}, \mathbf{W}, \mathbf{H})$

Optimisation of W, H

$$\min_{\mathbf{W},\mathbf{H} \geq 0} \sum_{fn} d_{\text{IS}}(|\alpha_{fn}|^2 | [\mathbf{WH}]_{fn})$$

IS-NMF with majorisation-minimisation
(Févotte and Idier, 2011)

Maximum joint likelihood estimation in LRTFS

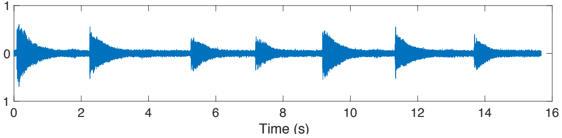
Algorithm 1: Block coordinate descent for LRTFS

Set $L \geq \|\mathbf{\Phi}\|_2^2$ (gradient inverse step size)
Set $\boldsymbol{\alpha}^{(0)} = \mathbf{\Phi}^{\mathsf{H}}\mathbf{x}$ (STFT)
repeat
  % Update **W** and **H** with NMF
  $\{\mathbf{W}^{(i+1)}, \mathbf{H}^{(i+1)}\} = \arg\min_{\mathbf{W},\mathbf{H}\geq 0} \sum_{fn} d_{\mathsf{IS}}(|\alpha_{fn}^{(i)}|^2 | [\mathbf{WH}]_{fn})$
  % Update $\boldsymbol{\alpha}$ with FISTA
  repeat
    % Gradient descent
    $\mathbf{z}^{(i)} = \boldsymbol{\alpha}^{(i)} + \frac{1}{L}\mathbf{\Phi}^{\mathsf{H}}(\mathbf{x} - \mathbf{\Phi}\boldsymbol{\alpha}^{(i)})$
    % Shrink
    $\alpha_{fn}^{(i+1)} = \frac{[\mathbf{WH}]_{fn}^{(i+1)}}{[\mathbf{WH}]_{fn}^{(i+1)} + \lambda/L} z_{fn}^{(i)}$
    % Accelerate with momentum
  until *convergence*;
until *convergence*;

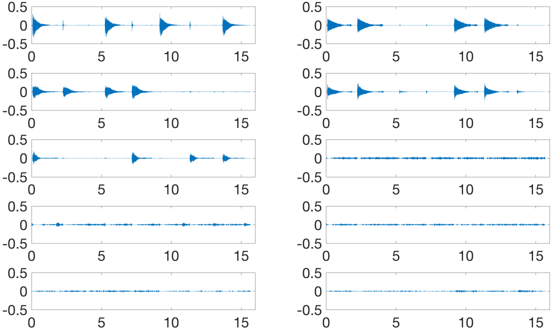
Complexity is one NMF per update of synthesis coefficients
Efficient multiplication by $\mathbf{\Phi}$ or $\mathbf{\Phi}^{\mathsf{H}}$ using the LTFAT time-frequency toolbox

Noisy piano example

(a) noisy signal



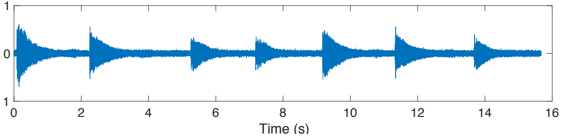
(b) IS-NMF decomposition



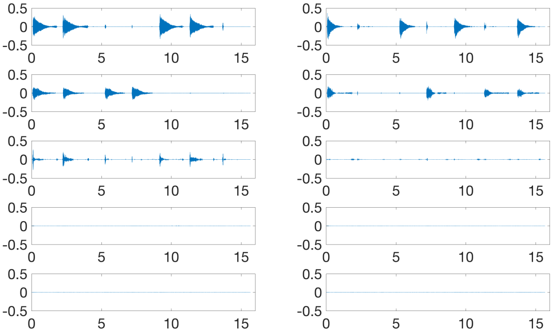
audio: $c_1(t)$ $c_2(t)$ $c_3(t)$ $c_4(t)$ $c_5(t)$ $c_6(t)$ $c_7(t)$ $c_8(t)$ $c_9(t)$ $c_{10}(t)$

Noisy piano example

(a) noisy signal



(b) LRTFS decomposition



audio: $c_1(t)$ $c_2(t)$ $c_3(t)$ $c_4(t)$ $c_5(t)$ $c_6(t)$ $c_7(t)$ $c_8(t)$ $c_9(t)$ $c_{10}(t)$

Remarks about LRTFS

Real-valued signals (Févotte and Kowalski, 2018)

- $x(t)$ previously assumed complex-valued for simplicity.
- In practice, $x(t)$ is real-valued and $\mathbf{\Phi}$, $\boldsymbol{\alpha}$ have Hermitian symmetry:

$$x(t) = \sum_{f=1}^{F/2} \sum_{n=1}^{N} 2\Re[\alpha_{fn}\phi_{fn}(t)] + e(t)$$

- More difficult to address but leads to essentially the same algorithm.

Multi-layer representations (Févotte and Kowalski, 2014, 2015)

- LRTFS allows for multi-resolution hybrid representations:

$$\mathbf{x} = \mathbf{\Phi}_a \boldsymbol{\alpha}_a + \mathbf{\Phi}_b \boldsymbol{\alpha}_b + \mathbf{e}.$$

  - $\mathbf{\Phi}_a$ and $\mathbf{\Phi}_b$ are time-frequency dictionaries with possibly different resolutions,
  - $\boldsymbol{\alpha}_a$ and $\boldsymbol{\alpha}_b$ have their own structure, either low-rank or sparse.
- Not possible with standard NMF !

Remarks about LRTFS

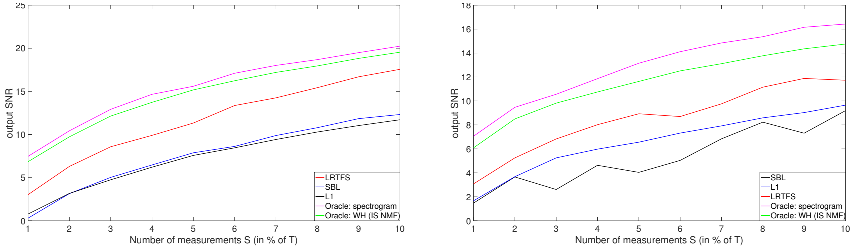
Compressive sensing (Févotte and Kowalski, 2018)

- Signal $\mathbf{x}$ of size T sensed through linear operator $\mathbf{A}$ of size $S \times T$:

$$\begin{aligned} \mathbf{b} &= \mathbf{A}\mathbf{x} + \mathbf{e} \\ &= \mathbf{A}\boldsymbol{\Phi}\boldsymbol{\alpha} + \mathbf{e} \end{aligned}$$

- Thanks to LRTFS, low-rankness can be used instead of sparsity.
- Estimate $\boldsymbol{\alpha}$ from $\mathbf{b}$ under $\alpha_{fn} \sim N_c(0, [\mathbf{WH}]_{fn})$, similar algorithm.

toy piano　　　　song excerpt*



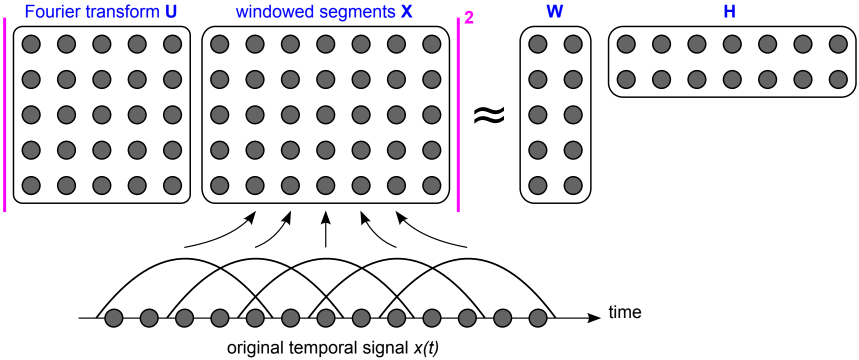
Recovery with **LRTFS**, **SBL**, **LASSO** (optimal hyperparameter) and two oracles

* *Mamavatu* by S. Raman (acoustic guitar, percussion, drums)

Outline

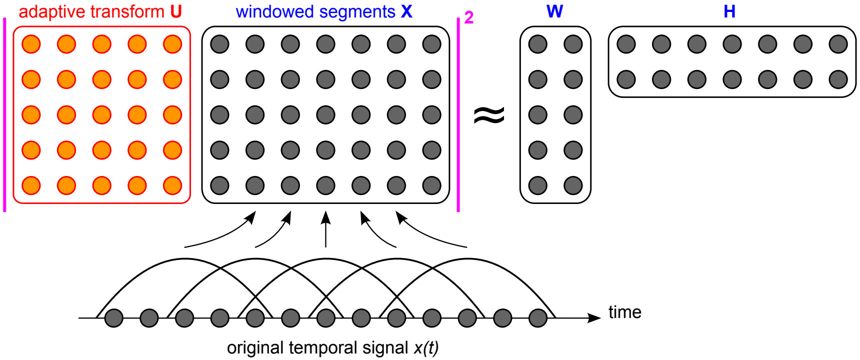
NMF with transform learning

(Fagot, Wendt, and Févotte, 2018)



NMF with transform learning

(Fagot, Wendt, and Févotte, 2018)



NMF with transform learning

(Fagot, Wendt, and Févotte, 2018)

- Power spectrogram $\mathbf{S}$ can be written as

$$\mathbf{S} = |\mathbf{U}_{\mathrm{FT}}\mathbf{X}|^2$$

  - $\mathbf{X}$ of size $F \times N$ contains adjacent and windowed segments of $x(t)$.
  - $\mathbf{U}_{\mathrm{FT}}$ of size $F \times F$ is the orthogonal Fourier matrix.

NMF with transform learning

(Fagot, Wendt, and Févotte, 2018)

- Power spectrogram $\mathbf{S}$ can be written as

$$\mathbf{S} = |\mathbf{U}_{\text{FT}}\mathbf{X}|^2$$

  - $\mathbf{X}$ of size $F \times N$ contains adjacent and windowed segments of $x(t)$.
  - $\mathbf{U}_{\text{FT}}$ of size $F \times F$ is the orthogonal Fourier matrix.
- Traditional NMF:

$$\min_{\mathbf{W},\mathbf{H}} D(|\mathbf{U}_{\text{FT}}\mathbf{X}|^2|\mathbf{WH}) \quad \text{s.t.} \quad \mathbf{W}, \mathbf{H} \geq 0$$

NMF with transform learning

(Fagot, Wendt, and Févotte, 2018)

- Power spectrogram $\mathbf{S}$ can be written as

$$\mathbf{S} = |\mathbf{U}_{\mathrm{FT}}\mathbf{X}|^2$$

  - $\mathbf{X}$ of size $F \times N$ contains adjacent and windowed segments of $x(t)$.
  - $\mathbf{U}_{\mathrm{FT}}$ of size $F \times F$ is the orthogonal Fourier matrix.
- Traditional NMF:

$$\min_{\mathbf{W},\mathbf{H}} D(|\mathbf{U}_{\mathrm{FT}}\mathbf{X}|^2|\mathbf{W}\mathbf{H}) \quad \text{s.t.} \quad \mathbf{W}, \mathbf{H} \geq 0$$

- NMF with transform learning (TL-NMF):

$$\min_{\mathbf{W},\mathbf{H},\mathbf{U}} D(|\mathbf{U}\mathbf{X}|^2|\mathbf{W}\mathbf{H}) \quad \text{s.t.} \quad \mathbf{W}, \mathbf{H} \geq 0, \mathbf{U} \text{ orthogonal}$$

- Can be interpreted as a one-layer factorising network.

NMF with transform learning

(Fagot, Wendt, and Févotte, 2018)

- Power spectrogram $\mathbf{S}$ can be written as

$$\mathbf{S} = |\mathbf{U}_{\mathrm{FT}}\mathbf{X}|^2$$

  - $\mathbf{X}$ of size $F \times N$ contains adjacent and windowed segments of $x(t)$.
  - $\mathbf{U}_{\mathrm{FT}}$ of size $F \times F$ is the orthogonal Fourier matrix.
- Traditional NMF:

$$\min_{\mathbf{W},\mathbf{H}} D(|\mathbf{U}_{\mathrm{FT}}\mathbf{X}|^2|\mathbf{WH}) \quad \text{s.t.} \quad \mathbf{W},\mathbf{H} \geq 0$$

- NMF with transform learning (TL-NMF):

$$\min_{\mathbf{W},\mathbf{H},\mathbf{U}} D(|\mathbf{UX}|^2|\mathbf{WH}) \quad \text{s.t.} \quad \mathbf{W},\mathbf{H} \geq 0, \mathbf{U} \text{ orthogonal}$$

- Can be interpreted as a one-layer factorising network.
- Inspired by the sparsifying transform of (Ravishankar and Bresler, 2013):

$$\min_{\mathbf{U}} \|\mathbf{UX}\|_1 \quad \text{s.t.} \quad \mathbf{U} \text{ square invertible}$$

NMF with transform learning

(Fagot, Wendt, and Févotte, 2018)

Objective

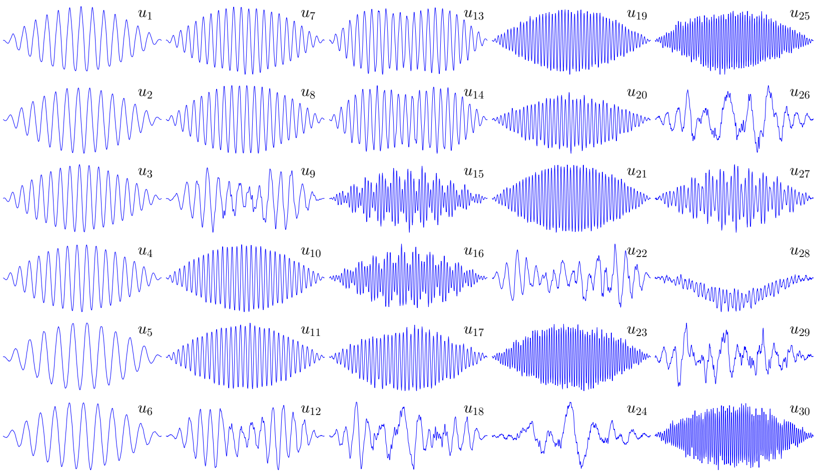
$$\min_{\mathbf{W},\mathbf{H},\mathbf{U}} D_{\text{IS}}(|\mathbf{U}\mathbf{X}|^2|\mathbf{W}\mathbf{H}) + \lambda\|\mathbf{H}\|_1 \quad \text{s.t.} \quad \begin{cases} \mathbf{W},\mathbf{H} \geq 0 \\ \|\mathbf{w}_k\|_1 = 1 \\ \mathbf{U}\mathbf{U}^T = \mathbf{I} \end{cases}$$

- For simplicity, $\mathbf{U}$ real-valued orthogonal matrix.
- Practical importance of imposing sparsity on $\mathbf{H}$.

Block coordinate descent $(\mathbf{U},\mathbf{W},\mathbf{H})$

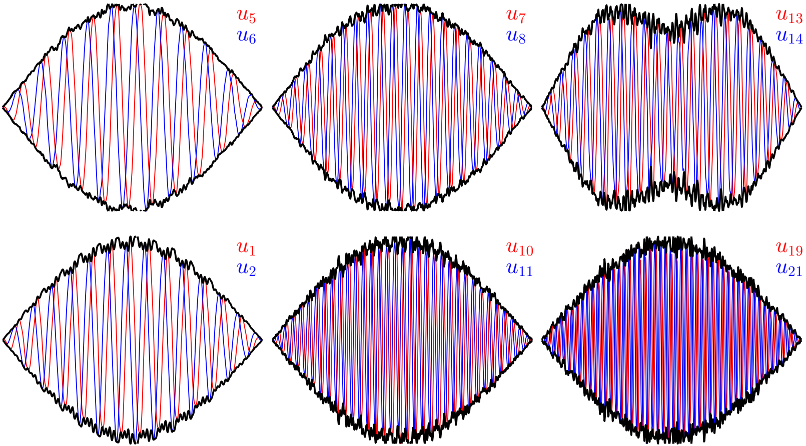
- Update of $(\mathbf{W},\mathbf{H})$ with majorisation-minimisation.
- Update of $\mathbf{U}$:
  - Projected gradient descent with line-search (Manton, 2002)
  - Jacobi algorithm ($\mathbf{U}$ decomposed as a product of Givens rotations) (Wendt, Fagot, and Févotte, 2018)

Learnt atoms from the piano toy example



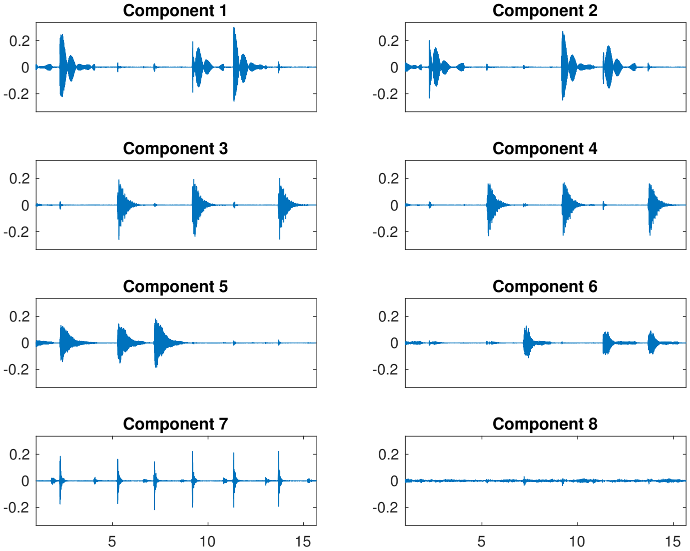
30 most active atoms learnt with TL-NMF
(random initialisation)

Learnt atoms from the piano toy example



Some atoms form pairs in phase quadrature

Temporal decomposition with TL-NMF



Audio: $c_1(t)$ $c_2(t)$ $c_3(t)$ $c_4(t)$ $c_5(t)$ $c_6(t)$ $c_7(t)$ $c_8(t)$

Summary

- IS-NMF is a generative model of the STFT.
- LRTFS is a generative model of the signal itself, with low-rank variance structure of the synthesis coefficients.
- TL-NMF learns a short-time transform together with the factorisation.
- Both LRTFS and TL-NMF take the raw signal as input.

Summary

- IS-NMF is a generative model of the STFT.
- LRTFS is a generative model of the signal itself, with low-rank variance structure of the synthesis coefficients.
- TL-NMF learns a short-time transform together with the factorisation.
- Both LRTFS and TL-NMF take the raw signal as input.

Available **PhD & postdoc** positions
in machine learning & signal processing
within ERC project FACTORY
`http://projectfactory.irit.fr/`

References I

S. A. Abdallah and M. D. Plumbley. Polyphonic transcription by nonnegative sparse coding of power spectra. In *Proc. International Symposium Music Information Retrieval Conference (ISMIR)*, pages 318–325, Barcelona, Spain, Oct. 2004.

R. Badeau. Gaussian modeling of mixtures of non-stationary signals in the time-frequency domain (HR-NMF). In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2011. doi: 10.1109/ASPAA.2011.6082264.

L. Benaroya, R. Gribonval, and F. Bimbot. Non negative sparse representation for Wiener based source separation with a single sensor. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 613–616, Hong Kong, 2003.

L. Chaâri, J.-C. Pesquet, A. Benazza-Benyahia, and P. Ciuciu. A wavelet-based regularized reconstruction algorithm for SENSE parallel MRI with applications to neuroimaging. *Medical Image Analysis*, 15(2):185–201, 2011.

L. Daudet and B. Torrésani. Hybrid representations for audiophonic signal encoding. *Signal Processing*, 82(11):1595 – 1617, 2002. ISSN 0165-1684. doi: http://dx.doi.org/10.1016/S0165-1684(02)00304-3.

A. Dessein, A. Cont, and G. Lemaitre. Real-time polyphonic music transcription with non-negative matrix factorization and beta-divergence. In *Proc. International Society for Music Information Retrieval Conference (ISMIR)*, 2010.

O. Dikmen and C. Févotte. Nonnegative dictionary learning in the exponential noise model for adaptive music signal representation. In *Advances in Neural Information Processing Systems (NIPS)*, pages 2267–2275, Granada, Spain, Dec. 2011. URL https://www.irit.fr/~Cedric.Fevotte/publications/proceedings/nips11.pdf.

References II

D. Fagot, H. Wendt, and C. Févotte. Nonnegative matrix factorization with transform learning. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2018. URL https://www.irit.fr/~Cedric.Fevotte/publications/proceedings/icassp18.pdf.

C. Févotte. Majorization-minimization algorithm for smooth Itakura-Saito nonnegative matrix factorization. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, May 2011. URL https://www.irit.fr/~Cedric.Fevotte/publications/proceedings/icassp11a.pdf.

C. Févotte and J. Idier. Algorithms for nonnegative matrix factorization with the beta-divergence. *Neural Computation*, 23(9):2421–2456, Sep. 2011. doi: 10.1162/NECO_a_00168. URL https://www.irit.fr/~Cedric.Fevotte/publications/journals/neco11.pdf.

C. Févotte and M. Kowalski. Low-rank time-frequency synthesis. In *Advances in Neural Information Processing Systems (NIPS)*, Dec. 2014. URL https://www.irit.fr/~Cedric.Fevotte/publications/proceedings/nips14.pdf.

C. Févotte and M. Kowalski. Hybrid sparse and low-rank time-frequency signal decomposition. In *Proc. European Signal Processing Conference (EUSIPCO)*, Nice, France, Sep. 2015. URL https://www.irit.fr/~Cedric.Fevotte/publications/proceedings/eusipco15.pdf.

C. Févotte and M. Kowalski. Estimation with low-rank time-frequency synthesis models. *IEEE Transactions on Signal Processing*, 66(15):4121–4132, Aug. 2018. doi: https://doi.org/10.1109/TSP.2018.2844159. URL https://arxiv.org/pdf/1804.09497.

C. Févotte, N. Bertin, and J.-L. Durrieu. Nonnegative matrix factorization with the Itakura-Saito divergence. With application to music analysis. *Neural Computation*, 21(3):793–830, Mar. 2009. doi: 10.1162/neco.2008.04-08-771. URL https://www.irit.fr/~Cedric.Fevotte/publications/journals/neco09_is-nmf.pdf.

References III

C. Févotte, J. Le Roux, and J. R. Hershey. Non-negative dynamical system with application to speech and audio. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Vancouver, Canada, May 2013. URL https://www.irit.fr/~Cedric.Fevotte/publications/proceedings/icassp13a.pdf.

A. Florescu, E. Chouzenoux, J.-C. Pesquet, P. Ciuciu, and S. Ciochina. A majorize-minimize memory gradient method for complex-valued inverse problems. *Signal Processing*, 103:285–295, 2014.

R. M. Gray, A. Buzo, A. H. Gray, and Y. Matsuyama. Distortion measures for speech processing. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 28(4):367–376, Aug. 1980.

M. Hoffman, D. Blei, and P. Cook. Bayesian nonparametric matrix factorization for recorded music. In *Proc. 27th International Conference on Machine Learning (ICML)*, Haifa, Israel, 2010.

F. Itakura and S. Saito. Analysis synthesis telephony based on the maximum likelihood method. In *Proc 6^{th} International Congress on Acoustics*, pages C–17 – C–20, Tokyo, Japan, Aug. 1968.

D. Kounades-Bastian, L. Girin, X. Alameda-Pineda, S. Gannot, and R. Horaud. A variational EM algorithm for the separation of time-varying convolutive audio mixtures. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 24(8):1408–1423, Aug. 2016. ISSN 2329-9290. doi: 10.1109/TASLP.2016.2554286.

A. Lefèvre, F. Bach, and C. Févotte. Itakura-Saito nonnegative matrix factorization with group sparsity. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, May 2011a. URL https://www.irit.fr/~Cedric.Fevotte/publications/proceedings/icassp11c.pdf.

A. Lefèvre, F. Bach, and C. Févotte. Online algorithms for nonnegative matrix factorization with the Itakura-Saito divergence. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, Mohonk, NY, Oct. 2011b. URL https://www.irit.fr/~Cedric.Fevotte/publications/proceedings/waspaa11.pdf.

References IV

S. Leglaive, R. Badeau, and G. Richard. Multichannel audio source separation with probabilistic reverberation priors. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 24(12): 2453–2465, Dec. 2016. ISSN 2329-9290. doi: 10.1109/TASLP.2016.2614140.

A. Liutkus, R. Badeau, and G. Richard. Gaussian processes for underdetermined source separation. *IEEE Transactions on Signal Processing*, 59(7):3155–3167, July 2011. doi: 10.1109/TSP.2011.2119315.

P. Magron, R. Badeau, and B. David. Phase-dependent anisotropic Gaussian model for audio source separation. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017. doi: 10.1109/ICASSP.2017.7952212.

J. H. Manton. Optimization algorithms exploiting unitary constraints. *IEEE Transactions on Signal Processing*, 50(3):635–650, Mar. 2002.

A. Ozerov and C. Févotte. Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation. *IEEE Transactions on Audio, Speech and Language Processing*, 18(3): 550–563, Mar. 2010. doi: 10.1109/TASL.2009.2031510. URL https://www.irit.fr/~Cedric.Fevotte/publications/journals/ieee_asl_multinmf.pdf.

R. M. Parry and I. Essa. Phase-aware non-negative spectrogram factorization. In *Proc. International Conference on Independent Component Analysis and Signal Separation (ICA)*, pages 536–543, London, UK, Sep. 2007.

S. Ravishankar and Y. Bresler. Learning sparsifying transforms. *IEEE Transactions on Signal Processing*, 61(5):1072–1086, Mar. 2013. ISSN 1053-587X. doi: 10.1109/TSP.2012.2226449.

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*, 21(5):971–982, May 2013. ISSN 1558-7916. doi: 10.1109/TASL.2013.2239990.

P. Smaragdis. About this non-negative business. WASPAA keynote slides, 2013. URL http://web.engr.illinois.edu/~paris/pubs/smaragdis-waspaa2013keynote.pdf.

References V

P. Smaragdis and J. C. Brown. Non-negative matrix factorization for polyphonic music transcription. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, Oct. 2003.

V. Y. F. Tan and C. Févotte. Automatic relevance determination in nonnegative matrix factorization with the beta-divergence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(7): 1592 – 1605, July 2013. URL `https://www.irit.fr/~Cedric.Fevotte/publications/journals/pami13_ardnmf.pdf`.

M. E. Tipping. Sparse Bayesian learning and the relevance vector machine. *Journal of Machine Learning Research*, 1:211–244, 2001.

R. E. Turner and M. Sahani. Time-frequency analysis as probabilistic inference. *IEEE Transactions on Signal Processing*, 62(23):6171–6183, Dec 2014. doi: 10.1109/TSP.2014.2362100.

H. Wendt, D. Fagot, and C. Févotte. Jacobi algorithm for nonnegative matrix factorization with transform learning. In *Proc. European Signal Processing Conference (EUSIPCO)*, Sep. 2018. URL `https://www.irit.fr/~Cedric.Fevotte/publications/proceedings/eusipco2018.pdf`.

D. P. Wipf and B. D. Rao. Sparse bayesian learning for basis selection. *IEEE Transactions on Signal Processing*, 52(8):2153–2164, Aug. 2004.

K. Yoshii, R. Tomioka, D. Mochihashi, and M. Goto. Infinite positive semidefinite tensor factorization for source separation of mixture signals. In *Proc. International Conference on Machine Learning (ICML)*, 2013.