Itakura-Saito NMF: un modèle probabiliste à facteurs latents pour la transformée de Fourier court-terme

Cédric Févotte

Laboratoire Lagrange, Nice

Journée GdR ISIS "Signaux complexes"
Marseille, juin 2013

Outline

Generalities about NMF

- Concept of NMF
- Majorization-minimization algorithms

Itakura-Saito NMF

- A statistical model of the STFT
- Piano decomposition example

Nonnegative dynamical system for speech & audio

- Statistical model
- Speech enhancement

Nonnegative matrix factorization (NMF)

Given a *nonnegative* matrix $\mathbf{V}$ of dimensions $F \times N$, NMF is the problem of finding a factorization

$$\mathbf{V} \approx \mathbf{W}\mathbf{H}$$

where $\mathbf{W}$ and $\mathbf{H}$ are *nonnegative* matrices of dimensions $F \times K$ and $K \times N$, respectively.

Nonnegative matrix factorization (NMF)

Given a *nonnegative* matrix $\mathbf{V}$ of dimensions $F \times N$, NMF is the problem of finding a factorization

$$\mathbf{V} \approx \mathbf{W}\mathbf{H}$$

where $\mathbf{W}$ and $\mathbf{H}$ are *nonnegative* matrices of dimensions $F \times K$ and $K \times N$, respectively.

Dimensions:

- If $\mathbf{W}$ tall ($K < F$), NMF produces a low-rank approximation.
- If $\mathbf{W}$ fat ($K > F$), NMF produces an overcomplete representation (e.g., sparse coding).

An unsupervised part-based representation

Along VQ, PCA or ICA, NMF provides an **unsupervised linear representation** of data

$\mathbf{v}_n$	$\approx$	$\mathbf{W}$	$\mathbf{h}_n$
data vector		"explanatory variables"	"regressors"
		"basis", "dictionary"	"expansion coefficients"
		"patterns"	"activation coefficients"

and $\mathbf{W}$ is learnt from the set of data vectors $\mathbf{V} = [\mathbf{v}_1 \ldots \mathbf{v}_N]$.

An unsupervised part-based representation

Along VQ, PCA or ICA, NMF provides an **unsupervised linear representation** of data

$\mathbf{v}_n$	$\approx$	$\mathbf{W}$	$\mathbf{h}_n$
data vector		“explanatory variables”	“regressors”
		“basis”, “dictionary”	“expansion coefficients”
		“patterns”	“activation coefficients”

and $\mathbf{W}$ is learnt from the set of data vectors $\mathbf{V} = [\mathbf{v}_1 \ldots \mathbf{v}_N]$.

- **nonneg. of W** ensures *interpretability* of the dictionary (features $\mathbf{w}_k$ and data $\mathbf{v}_n$ belong to same space).
- **nonneg. of H** tends to produce *part-based* representations because subtractive combinations are forbidden.

Early work by Paatero and Tapper (1994), landmark paper in *Nature* by Lee and Seung (1999).

NMF as a constrained minimization problem

Minimize a measure of fit between data **V** and model **WH**, subject to nonnegativity of **W** and **H**:

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

where $d(x|y)$ is a scalar cost function.

Regularization terms are often added to $D(\mathbf{V}|\mathbf{WH})$ to favor certain properties of **W** or **H** (sparsity, smoothness).

Divergences used in NMF

(selected references)

- Euclidean distance (Paatero and Tapper, 1994; Lee and Seung, 2001)
- Kullback-Leibler divergence (Lee and Seung, 1999; Finesso and Spreij, 2006)
- α-divergence (Cichocki et al., 2008)
- β-divergence (Cichocki et al., 2006; Févotte and Idier, 2011)
- Bregman divergences (Dhillon and Sra, 2005)

Divergences used in NMF

(selected references)

- Euclidean distance (Paatero and Tapper, 1994; Lee and Seung, 2001)
- Kullback-Leibler divergence (Lee and Seung, 1999; Finesso and Spreij, 2006)
- α-divergence (Cichocki et al., 2008)
- β-divergence (Cichocki et al., 2006; Févotte and Idier, 2011)
- Bregman divergences (Dhillon and Sra, 2005)

- Itakura-Saito divergence (Févotte et al., 2009)

Common NMF algorithm design

- Block-coordinate update of $\mathbf{H}$ given $\mathbf{W}^{(i-1)}$ and $\mathbf{W}$ given $\mathbf{H}^{(i)}$.

Common NMF algorithm design

- Block-coordinate update of $\mathbf{H}$ given $\mathbf{W}^{(i-1)}$ and $\mathbf{W}$ given $\mathbf{H}^{(i)}$.
- The updates of $\mathbf{W}$ and $\mathbf{H}$ are equivalent by transposition:

$$\mathbf{V} \approx \mathbf{W}\mathbf{H} \Leftrightarrow \mathbf{V}^T \approx \mathbf{H}^T\mathbf{W}^T$$

Common NMF algorithm design

- Block-coordinate update of $\mathbf{H}$ given $\mathbf{W}^{(i-1)}$ and $\mathbf{W}$ given $\mathbf{H}^{(i)}$.
- The updates of $\mathbf{W}$ and $\mathbf{H}$ are equivalent by transposition:

$$\mathbf{V} \approx \mathbf{W}\mathbf{H} \Leftrightarrow \mathbf{V}^T \approx \mathbf{H}^T\mathbf{W}^T$$

- The objective function is separable in the columns of $\mathbf{H}$ or the rows of $\mathbf{W}$:

$$D(\mathbf{V}|\mathbf{W}\mathbf{H}) = \sum_n D(\mathbf{v}_n|\mathbf{W}\mathbf{h}_n)$$

Common NMF algorithm design

- Block-coordinate update of **H** given $\mathbf{W}^{(i-1)}$ and **W** given $\mathbf{H}^{(i)}$.
- The updates of **W** and **H** are equivalent by transposition:

$$\mathbf{V} \approx \mathbf{W}\mathbf{H} \Leftrightarrow \mathbf{V}^T \approx \mathbf{H}^T\mathbf{W}^T$$

- The objective function is separable in the columns of **H** or the rows of **W**:

$$D(\mathbf{V}|\mathbf{W}\mathbf{H}) = \sum_n D(\mathbf{v}_n|\mathbf{W}\mathbf{h}_n)$$

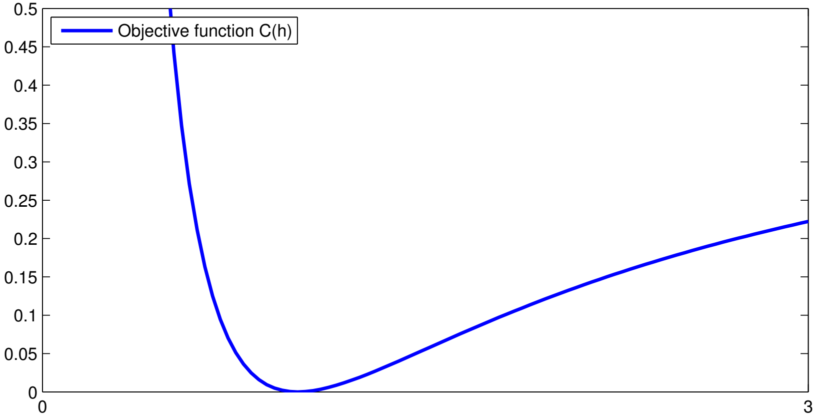
- In the end we are left with *nonnegative linear regression*

$$\min_{\mathbf{h}\geq\mathbf{0}} C(\mathbf{h}) \overset{\text{def}}{=} D(\mathbf{v}|\mathbf{W}\mathbf{h})$$

Numerous references in the image restoration literature (Richardson, 1972; Lucy, 1974; Daube-Witherspoon and Muehllehner, 1986)

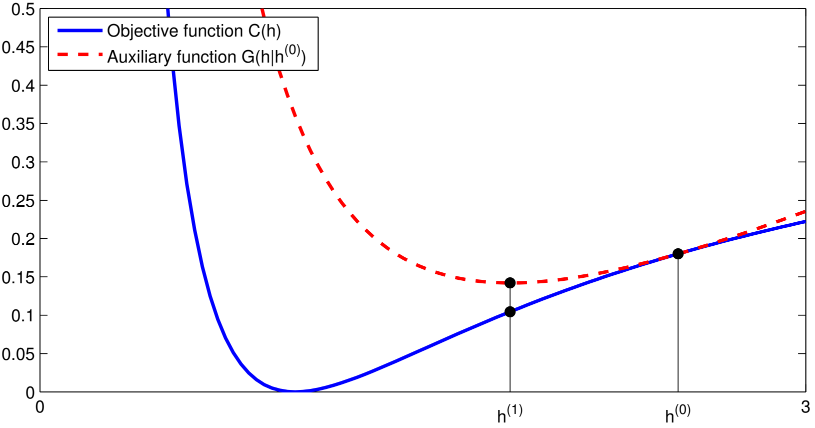
Majorization-minimization (MM)

Build $G(\mathbf{h}|\tilde{\mathbf{h}})$ such that $G(\mathbf{h}|\tilde{\mathbf{h}}) \geq C(\mathbf{h})$ and $G(\tilde{\mathbf{h}}|\tilde{\mathbf{h}}) = C(\tilde{\mathbf{h}})$.
Optimize (iteratively) $G(\mathbf{h}|\tilde{\mathbf{h}})$ instead of $C(\mathbf{h})$.



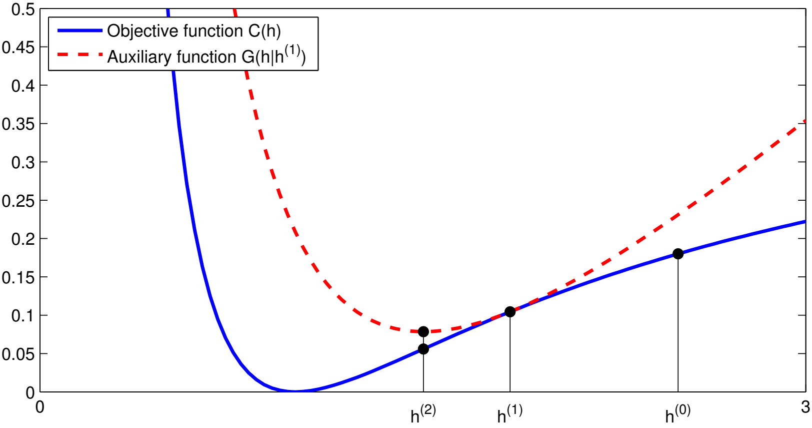
Majorization-minimization (MM)

Build $G(\mathbf{h}|\tilde{\mathbf{h}})$ such that $G(\mathbf{h}|\tilde{\mathbf{h}}) \geq C(\mathbf{h})$ and $G(\tilde{\mathbf{h}}|\tilde{\mathbf{h}}) = C(\tilde{\mathbf{h}})$.
Optimize (iteratively) $G(\mathbf{h}|\tilde{\mathbf{h}})$ instead of $C(\mathbf{h})$.



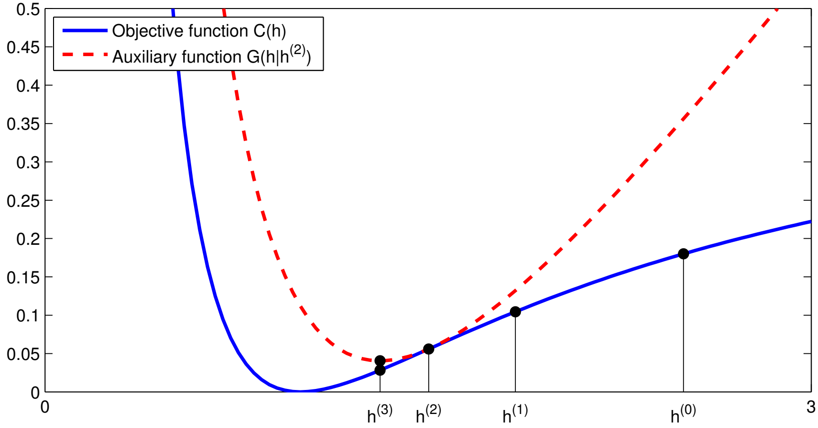
Majorization-minimization (MM)

Build $G(\mathbf{h}|\tilde{\mathbf{h}})$ such that $G(\mathbf{h}|\tilde{\mathbf{h}}) \geq C(\mathbf{h})$ and $G(\tilde{\mathbf{h}}|\tilde{\mathbf{h}}) = C(\tilde{\mathbf{h}})$.
Optimize (iteratively) $G(\mathbf{h}|\tilde{\mathbf{h}})$ instead of $C(\mathbf{h})$.



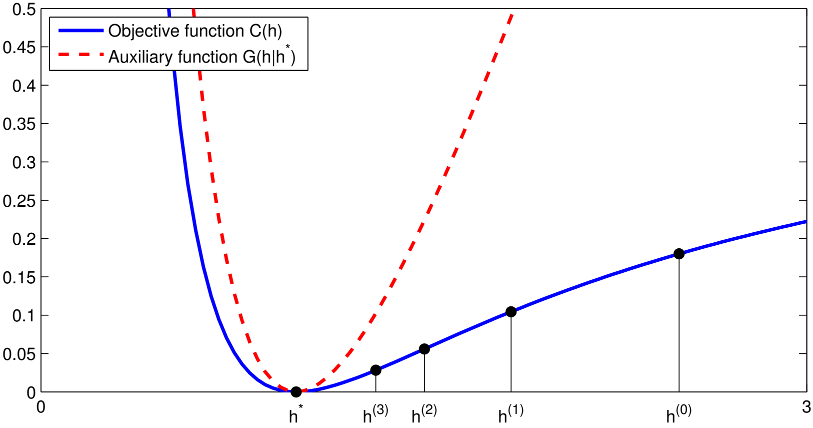
Majorization-minimization (MM)

Build $G(\mathbf{h}|\tilde{\mathbf{h}})$ such that $G(\mathbf{h}|\tilde{\mathbf{h}}) \geq C(\mathbf{h})$ and $G(\tilde{\mathbf{h}}|\tilde{\mathbf{h}}) = C(\tilde{\mathbf{h}})$.
Optimize (iteratively) $G(\mathbf{h}|\tilde{\mathbf{h}})$ instead of $C(\mathbf{h})$.

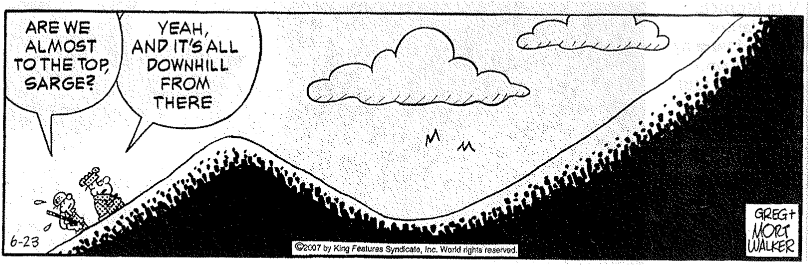


Majorization-minimization (MM)

Build $G(\mathbf{h}|\tilde{\mathbf{h}})$ such that $G(\mathbf{h}|\tilde{\mathbf{h}}) \geq C(\mathbf{h})$ and $G(\tilde{\mathbf{h}}|\tilde{\mathbf{h}}) = C(\tilde{\mathbf{h}})$.
Optimize (iteratively) $G(\mathbf{h}|\tilde{\mathbf{h}})$ instead of $C(\mathbf{h})$.



Local convergence

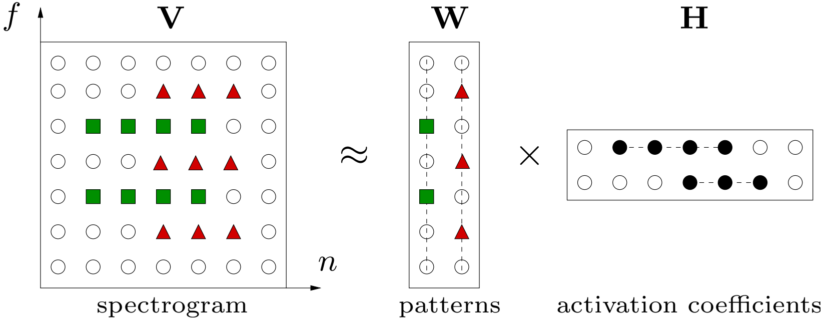

ARE WE ALMOST TO THE TOP, SARGE?
YEAH, AND IT'S ALL DOWNHILL FROM THERE
6-23
GREG+ MORT WALKER

©2007 by King Features Syndicate, Inc. World rights reserved.

- If $d(x|y)$ is convex w.r.t to y, $D(\mathbf{V}|\mathbf{WH})$ convex w.r.t either $\mathbf{W}$ or $\mathbf{H}$ but not both.
- Not even true if $d(x|y)$ not convex w.r.t y.

Application to music signal processing

(Smaragdis and Brown, 2003)



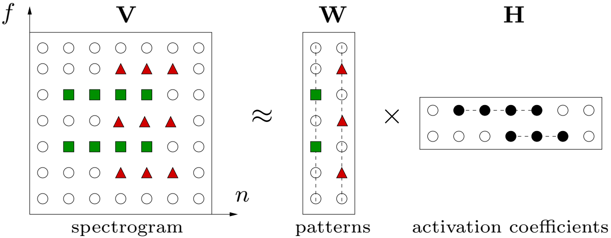
Outline

Generalities about NMF
  Concept of NMF
  Majorization-minimization algorithms

Itakura-Saito NMF
  A statistical model of the STFT
  Piano decomposition example

Nonnegative dynamical system for speech & audio
  Statistical model
  Speech enhancement

Model choices



- Magnitude or power spectrogram ?
- Which measure of fit should be used for the factorization ?
- NMF approximates the spectrogram by a sum of rank-one spectrograms. How can we invert these ? What about phase ?

Itakura-Saito NMF: a generative approach

(Févotte, Bertin, and Durrieu, 2009)

Let $\mathbf{X} = \{x_{fn}\}$ be the (complex-valued) STFT of the signal. Assume

$$x_{fn} = \sum_{k=1}^{K} c_{k,fn}$$

$$c_{k,fn} \sim \mathcal{N}_c(0, w_{fk} h_{kn})$$

and the components $c_{1,fn}, \ldots, c_{K,fn}$ are independent given $\mathbf{W}$ and $\mathbf{H}$.

Itakura-Saito NMF: a generative approach

(Févotte, Bertin, and Durrieu, 2009)

Let $\mathbf{X} = \{x_{fn}\}$ be the (complex-valued) STFT of the signal.
Assume

$$x_{fn} = \sum_{k=1}^{K} c_{k,fn}$$

$$c_{k,fn} \sim \mathcal{N}_c(0, w_{fk} h_{kn})$$

and the components $c_{1,fn}, \ldots, c_{K,fn}$ are independent given $\mathbf{W}$ and $\mathbf{H}$. Then

$$-\log p(\mathbf{X}|\mathbf{W},\mathbf{H}) = D_{IS}(|\mathbf{X}|^2|\mathbf{WH}) \quad + cst.$$

Additivity assumed in the STFT domain. Phase is preserved in the model, though in a noninformative way (uniform distribution).

Related work by Benaroya et al. (2003); Parry and Essa (2007)

Itakura-Saito NMF: a generative approach

(Févotte, Bertin, and Durrieu, 2009)

Main message: Itakura-Saito NMF of the power spectrogram corresponds to maximum likelihood estimation in a well-defined generative composite model of the STFT.

This in particular gives a statistically grounded way of reconstructing the components:

$$\hat{c}_{k,fn} = \mathrm{E}\{c_{k,fn} | \mathbf{X}, \mathbf{W}, \mathbf{H}\} = \underbrace{\frac{w_{fk} h_{kn}}{\sum_j w_{fj} h_{jn}}}_{\text{time-freq. mask}} \; x_{fn}$$

Lossless decomposition: $x_{fn} = \sum_k \hat{c}_{k,fn}$

Itakura-Saito NMF: a generative approach

(Févotte, Bertin, and Durrieu, 2009)

Alternatively, IS-NMF can be interpreted as maximum likelihood in multiplicative noise:

$$v_{fn} = |x_{fn}|^2 = [\mathbf{WH}]_{fn} \cdot \epsilon_{fn}$$

where ϵ_{fn} is Gamma multiplicative noise with mean value 1.

Related work by Abdallah and Plumbley (2004).

Noteworthy properties of the IS divergence

- The IS divergence is scale-invariant:

  $$d_{IS}(\lambda\, x|\lambda\, y) = d_{IS}(x|y)$$

  Implies higher accuracy in the representation of data with large dynamic range, such as audio spectra. In contrast,

  $$\begin{aligned} d_{EUC}(\lambda\, x|\lambda\, y) &= \lambda^2\, d_{EUC}(x|y) \\ d_{KL}(\lambda\, x|\lambda\, y) &= \lambda\, d_{KL}(x|y) \end{aligned}$$

- The IS divergence in nonconvex (inflexion at $y = 2x$); was found to lead to more local minima in practice.

Other statistical factor models of the spectrogram

Latent factor models for count data inspired from text analysis:

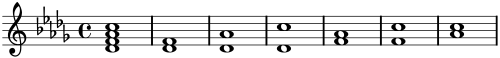
- Poisson models (Le Roux et al., 2007; Cemgil, 2009), similar to *GaP* (Canny, 2004)
- Multinomial models (Shashanka et al., 2008; Smaragdis et al., 2009), similar to *PLSI* (Hofmann, 1999) or *LDA* (Blei et al., 2003; Buntine and Jakulin, 2006)

Not generative models:

- Data $|x_{fn}|$ is modeled as integer.
- Additivity is assumed at the magnitude level

$$|x_{fn}| = \sum_k |c_{k,fn}|.$$

Small-scale example



(MIDI numbers : 61, 65, 68, 72)

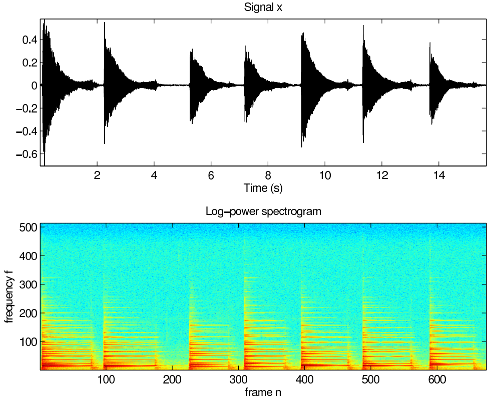
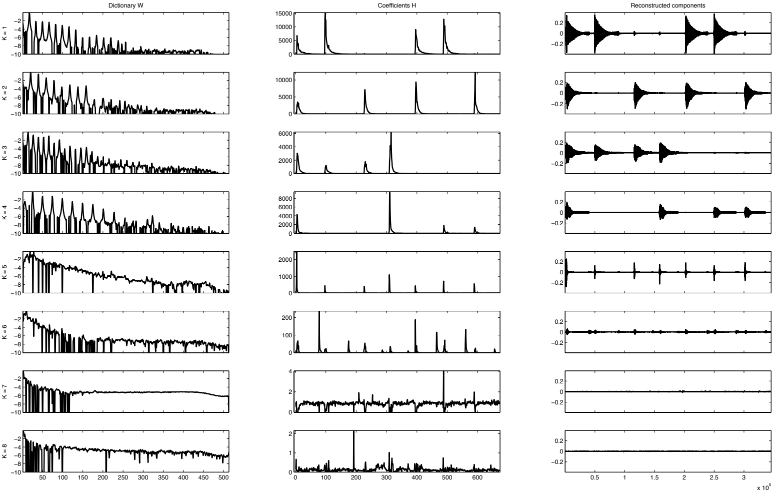


Figure: Three representations of data.

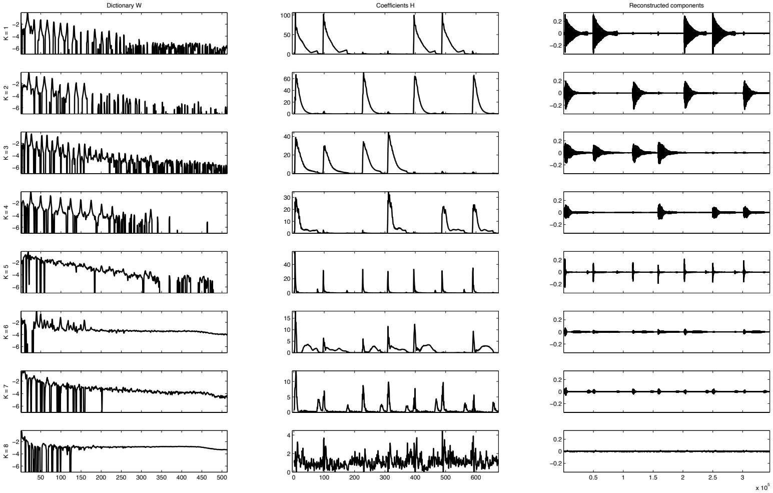
IS-NMF on power spectrogram with $K = 8$



Pitch estimates: 65.0 68.0 61.0 72.0 0 0 0 0

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

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)

Outline

Generalities about NMF
  Concept of NMF
  Majorization-minimization algorithms

Itakura-Saito NMF
  A statistical model of the STFT
  Piano decomposition example

Nonnegative dynamical system for speech & audio
  Statistical model
  Speech enhancement

Motivation

- Model the temporal structure in audio signals.
  - more accurate estimation of **H**, and **W**,
  - reduced identifiability ambiguities,
  - perceptually more pleasant component reconstruction.
- Many existing models for nonnegative data or for sequences, but hard to gather desirable properties.

Dynamical models for real-valued data:	**Static models for nonnegative data**:
Linear dynamical system (LDS) Hidden Markov model (HMM)	Nonnegative matrix factorization (NMF)

- **Goal**: bring advantages of LDS, HMM and NMF together in a simple and elegant framework.

Linear dynamical system (LDS)

The classic Gaussian model for real-valued data

$$\begin{aligned} \mathbf{h}_n &= \mathbf{A}\mathbf{h}_{n-1} \;+\; \boldsymbol{\xi}_n \quad \text{(state dynamics)} \\ \mathbf{v}_n &= \mathbf{W}\mathbf{h}_n \quad\;\; +\; \boldsymbol{\epsilon}_n \quad \text{(observation model)} \end{aligned}$$

- Continuous Markov chain with **real-valued** variables and parameters.
- **Additive Gaussian** innovations with zero mean value.

Linear dynamical system (LDS)

The classic Gaussian model for real-valued data

$$\begin{aligned} \mathbf{h}_n &= \mathbf{A}\mathbf{h}_{n-1} \;+\; \xi_n \quad \text{(state dynamics)} \\ \mathbf{v}_n &= \mathbf{W}\mathbf{h}_n \quad\;\; +\; \epsilon_n \quad \text{(observation model)} \end{aligned}$$

- Continuous Markov chain with **real-valued** variables and parameters.
- **Additive Gaussian** innovations with zero mean value.

$$\begin{aligned} \mathrm{E}\left[\mathbf{h}_n \middle| \mathbf{A}\mathbf{h}_{n-1}\right] &= \mathbf{A}\mathbf{h}_{n-1} \\ \mathrm{E}\left[\mathbf{v}_n \middle| \mathbf{W}\mathbf{h}_n\right] &= \mathbf{W}\mathbf{h}_n \end{aligned}$$

Nonnegative dynamical system (NDS)

(Févotte, Le Roux, and Hershey, 2013)

$$\begin{aligned} \mathbf{h}_n &= \mathbf{A}\mathbf{h}_{n-1} \circ \xi_n && \text{(state dynamics)} \\ \mathbf{v}_n &= \mathbf{W}\mathbf{h}_n \circ \epsilon_n && \text{(observation model)} \end{aligned}$$

- Continuous Markov chain with **nonnegative** variables and parameters.
- **Multiplicative Gamma** innovations with mean value 1.

$$\mathrm{E}\left[\mathbf{h}_n \middle| \mathbf{A}\mathbf{h}_{n-1}\right] = \mathbf{A}\mathbf{h}_{n-1}$$
$$\mathrm{E}\left[\mathbf{v}_n \middle| \mathbf{W}\mathbf{h}_n\right] = \mathbf{W}\mathbf{h}_n$$

Nonnegative dynamical system (NDS)

(Févotte, Le Roux, and Hershey, 2013)

$$\begin{aligned} \mathbf{h}_n &= \mathbf{A}\mathbf{h}_{n-1} \circ \xi_n \quad \text{(state dynamics)} \\ \mathbf{v}_n &= \mathbf{W}\mathbf{h}_n \quad \circ \epsilon_n \quad \text{(observation model)} \end{aligned}$$

- The observation model underlies an Itakura-Saito (IS) pseudo-likelihood:

$$-\log p(\mathbf{V}|\mathbf{WH}) = D_{IS}(\mathbf{V}|\mathbf{WH}) + cst$$

- When $\mathbf{A} = \mathbf{I}_K$, we obtain smooth IS-NMF, i.e., $\mathrm{E}\left[h_{kn}|h_{k(n-1)}\right] = h_{k(n-1)}$. (Févotte, 2011)

Parameter estimation

MAP approach:

$$\min_{\mathbf{W},\mathbf{A},\mathbf{H}\geq\mathbf{0}} C(\mathbf{W},\mathbf{H},\mathbf{A}) = \underbrace{-\log p(\mathbf{V}|\mathbf{WH})}_{\textit{fit}} \underbrace{-\log p(\mathbf{H}|\mathbf{A})}_{\textit{dynamics}}$$

Parameter estimation

MAP approach:

$$\min_{\mathbf{W},\mathbf{A},\mathbf{H}\geq\mathbf{0}} C(\mathbf{W},\mathbf{H},\mathbf{A}) = \underbrace{-\log p(\mathbf{V}|\mathbf{W}\mathbf{H})}_{\textit{fit}}\underbrace{-\log p(\mathbf{H}|\mathbf{A})}_{\textit{dynamics}}$$

Optimization:

- Block-coordinate descent algorithm that updates **W**, **A** and **H** in turn.
- Adjacent columns of **H** are coupled in the optimization; we used a left-to-right block-coordinate descent:

$$\underbrace{\mathbf{h}_1^{(i)} \rightarrow \ldots \rightarrow \mathbf{h}_{n-1}^{(i)}}_{\text{already updated}} \rightarrow \mathbf{h}_n \rightarrow \underbrace{\mathbf{h}_{n+1}^{(i-1)} \rightarrow \ldots \rightarrow \mathbf{h}_{n+1}^{(i-1)}}_{\text{not yet updated}}$$

- Updates obtained by majorization-minimization (MM).

Training a speech model

- Data
  - Speech from TIMIT, 16 kHz. exemple 1 exemple 2 exemple 3
  - 1000 files per gender ($\approx$ 50 minutes per gender).
  - Speaker-independent, gender-dependent training.

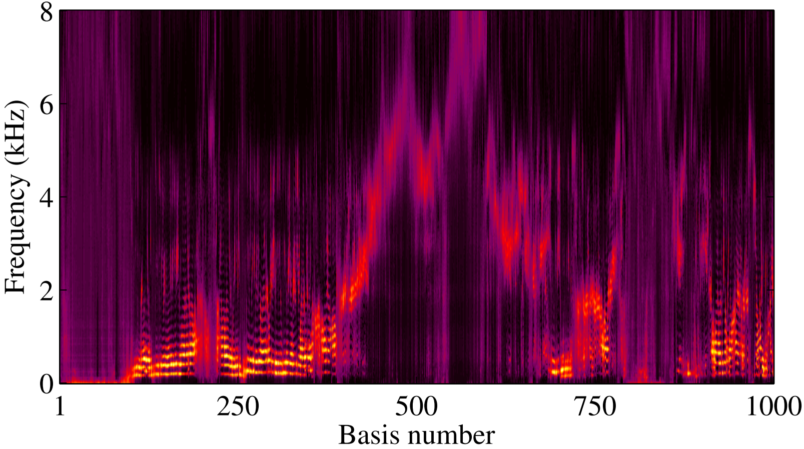
Training a speech model

- Data
    - Speech from TIMIT, 16 kHz. exemple 1 exemple 2 exemple 3
    - 1000 files per gender ($\approx$ 50 minutes per gender).
    - Speaker-independent, gender-dependent training.
- Power spectrogram computation
    - Sine window with 50% overlap.
    - Frame length: 32 ms to 60 ms.

Training a speech model

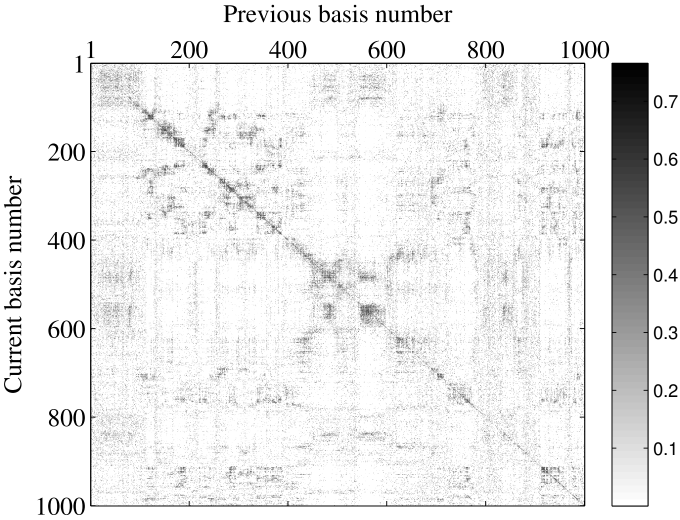
- Data
  - Speech from TIMIT, 16 kHz. exemple 1 exemple 2 exemple 3
  - 1000 files per gender ($\approx$ 50 minutes per gender).
  - Speaker-independent, gender-dependent training.
- Power spectrogram computation
  - Sine window with 50% overlap.
  - Frame length: 32 ms to 60 ms.
- NDS specifications
  - $K = 1000$
  - Observation: $\mathbf{v}_t = \mathbf{W}\mathbf{h}_t \circ \epsilon_t$ with exponential innovation ($\Leftrightarrow$ Gaussian modeling of the complex-valued STFT)
  - Dynamics: $\mathbf{h}_t = \mathbf{A}\mathbf{h}_t \circ \xi_t$ with very sparse innovation
    $\Rightarrow$ favors very few active coefficients in every frame
    $\Rightarrow$ dictionary elements look like phonems
    $\Rightarrow$ "relaxed" HMM model

Trained dictionary **W** (female)



(spectral patterns sorted greedily by similarity)

Trained transition matrix **A** (female)



Speech enhancement

$$\underbrace{x}_{\text{noisy signal}} = \underbrace{s}_{\text{clean speech}} + \underbrace{n}_{\text{noise}}$$

Speech enhancement

$$\underbrace{x}_{\text{noisy signal}} = \underbrace{s}_{\text{clean speech}} + \underbrace{n}_{\text{noise}}$$

1. Learn $\mathbf{W}^{\text{train}}$, $\mathbf{A}^{\text{train}}$ from clean speech (training data).

Speech enhancement

$$\underbrace{x}_{\text{noisy signal}} = \underbrace{s}_{\text{clean speech}} + \underbrace{n}_{\text{noise}}$$

1. Learn $\mathbf{W}^{\text{train}}$, $\mathbf{A}^{\text{train}}$ from clean speech (training data).
2. Compute the power spectrogram $\mathbf{V} = |\mathbf{X}|^2$ of noisy signal.

Speech enhancement

$$\underbrace{x}_{\text{noisy signal}} = \underbrace{s}_{\text{clean speech}} + \underbrace{n}_{\text{noise}}$$

1. Learn $\mathbf{W}^{\text{train}}$, $\mathbf{A}^{\text{train}}$ from clean speech (training data).
2. Compute the power spectrogram $\mathbf{V} = |\mathbf{X}|^2$ of noisy signal.
3. Produce the decomposition

$$\mathbf{V} = \mathbf{W}^{\text{train}}\mathbf{H} + \mathbf{W}^{\text{noise}}\mathbf{H}^{\text{noise}}$$

where

- $\mathrm{E}\left[\mathbf{h}_n | \mathbf{A}^{\text{train}}\mathbf{h}_{n-1}\right] = \mathbf{A}^{\text{train}}\mathbf{h}_{n-1}$ a priori.
- $\mathbf{W}^{\text{noise}}\mathbf{H}^{\text{noise}}$ forms a “garbage” NMF model of the noise.

Speech enhancement

$$\underbrace{x}_{\text{noisy signal}} = \underbrace{s}_{\text{clean speech}} + \underbrace{n}_{\text{noise}}$$

1. Learn $\mathbf{W}^{\text{train}}$, $\mathbf{A}^{\text{train}}$ from clean speech (training data).
2. Compute the power spectrogram $\mathbf{V} = |\mathbf{X}|^2$ of noisy signal.
3. Produce the decomposition

$$\mathbf{V} = \mathbf{W}^{\text{train}}\mathbf{H} + \mathbf{W}^{\text{noise}}\mathbf{H}^{\text{noise}}$$

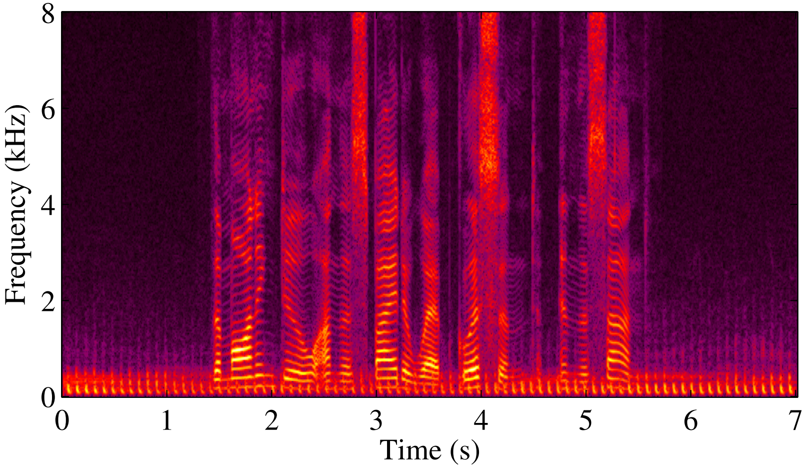
   where
   - $\mathrm{E}\left[\mathbf{h}_n | \mathbf{A}^{\text{train}}\mathbf{h}_{n-1}\right] = \mathbf{A}^{\text{train}}\mathbf{h}_{n-1}$ a priori.
   - $\mathbf{W}^{\text{noise}}\mathbf{H}^{\text{noise}}$ forms a "garbage" NMF model of the noise.
4. Produce the source estimate STFT by Wiener filtering

$$\hat{s}_{fn} = \frac{[\mathbf{W}^{\text{train}}\mathbf{H}]_{fn}}{[\mathbf{W}^{\text{train}}\mathbf{H} + \mathbf{W}^{\text{noise}}\mathbf{H}^{\text{noise}}]_{fn}} x_{fn}$$

Speech enhancement results

Helicopter

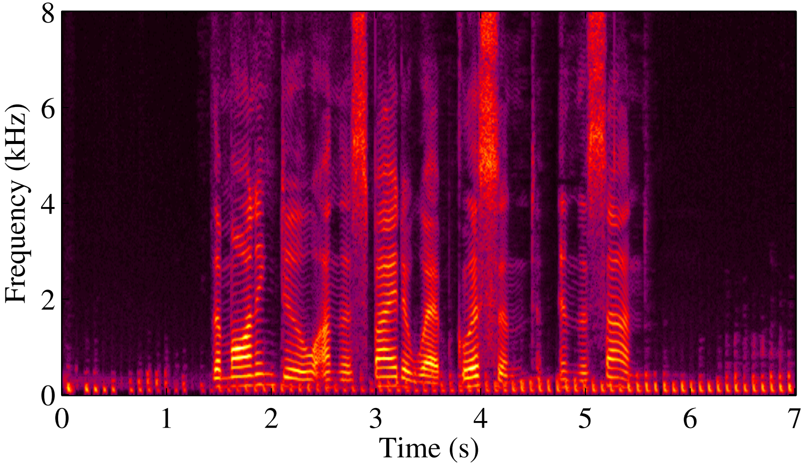
noisy signal



Speech enhancement results

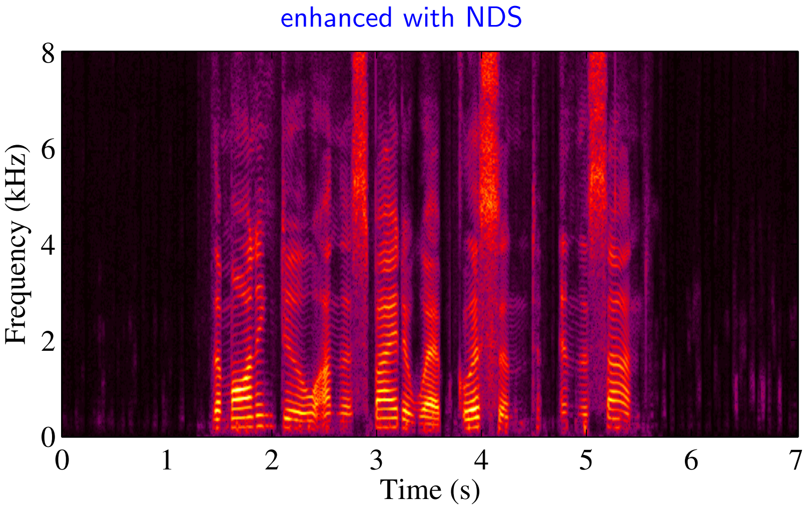
Helicopter

enhanced with OMLSA



Speech enhancement results

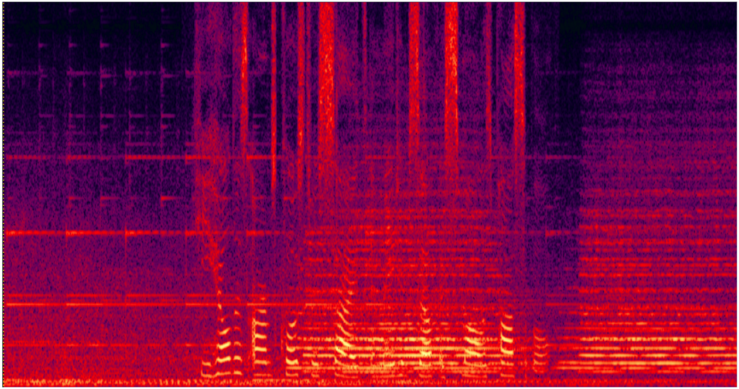
Helicopter



Speech enhancement results

Train

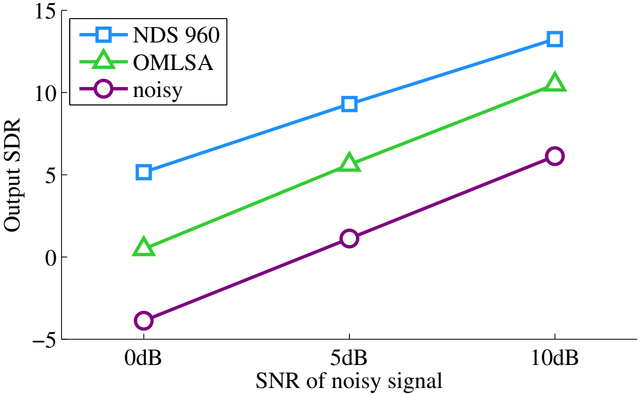
noisy signal



enhanced with OMLSA enhanced with NDS

Speech enhancement results

- 150 mixtures: 10 speech files *x* 15 texture sounds; 3 SNRs.
- Compared with state-of-the-art OMLSA (Cohen, 2002, 2003).
- 2 “garbage” spectral patterns to model the noise.



Conclusions

- Itakura-Saito NMF of the power spectrogram is underlain by a statistical model of superimposed Gaussian components.
- This model is relevant to the representation of audio signals.
- Algorithms can be designed in a principled way in the majorization-minimization setting.
- Regularized variants, e.g., NDS.

Conclusions

- Itakura-Saito NMF of the power spectrogram is underlain by a statistical model of superimposed Gaussian components.
- This model is relevant to the representation of audio signals.
- Algorithms can be designed in a principled way in the majorization-minimization setting.
- Regularized variants, e.g., NDS.

- Possible extension to multichannel data for audio source separation.
- The latent statistical model opens doors to fully Bayesian approaches that integrates over $\mathbf{W}$ and/or $\mathbf{H}$ (Févotte and Cemgil, 2009; Hoffman et al., 2010; Févotte et al., 2011; Dikmen and Févotte, 2011)

References I

S. A. Abdallah and M. D. Plumbley. Polyphonic transcription by nonnegative sparse coding of power spectra. In *Proc. 5th International Symposium Music Information Retrieval (ISMIR)*, pages 318–325, Barcelona, Spain, Oct. 2004.

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.

D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, Jan. 2003.

W. L. Buntine and A. Jakulin. Discrete component analysis. In *Lecture Notes in Computer Science*, volume 3940, pages 1–33. Springer, 2006. URL `http://www.springerlink.com/content/d53027666542q3v7/`.

J. F. Canny. GaP: A factor model for discrete data. In *Proceedings of the 27th ACM international Conference on Research and Development of Information Retrieval (SIGIR)*, pages 122–129, 2004.

A. T. Cemgil. Bayesian inference for nonnegative matrix factorisation models. *Computational Intelligence and Neuroscience*, 2009(Article ID 785152):17 pages, 2009. doi:10.1155/2009/785152.

References II

A. Cichocki, R. Zdunek, and S. Amari. Csiszar's divergences for non-negative matrix factorization: Family of new algorithms. In *Proc. 6th International Conference on Independent Component Analysis and Blind Signal Separation (ICA)*, pages 32–39, Charleston SC, USA, Mar. 2006.

A. Cichocki, H. Lee, Y.-D. Kim, and S. Choi. Non-negative matrix factorization with α-divergence. *Pattern Recognition Letters*, 29(9):1433–1440, July 2008.

I. Cohen. Optimal speech enhancement under signal presence uncertainty using log-spectral amplitude estimator. *IEEE Signal Processing Letters*, 9(4):113–116, 2002.

I. Cohen. Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging. *IEEE Transactions on Speech and Audio Processing*, 11(5):466–475, 2003.

M. Daube-Witherspoon and G. Muehllehner. An iterative image space reconstruction algorthm suitable for volume ECT. *IEEE Transactions on Medical Imaging*, 5(5):61 – 66, 1986. doi: 10.1109/TMI.1986.4307748.

I. S. Dhillon and S. Sra. Generalized nonnegative matrix approximations with Bregman divergences. *Advances in Neural Information Processing Systems (NIPS)*, 19, 2005.

References III

O. Dikmen and C. Févotte. Nonnegative dictionary learning in the exponential noise model for adaptive music signal representation. In J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K. Weinberger, editors, *Advances in Neural Information Processing Systems 24 (NIPS)*, pages 2267–2275, Granada, Spain, Dec. 2011. MIT Press. URL `http://www.unice.fr/cfevotte/publications/proceedings/nips11.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 `http://www.unice.fr/cfevotte/publications/proceedings/icassp11a.pdf`.

C. Févotte and A. T. Cemgil. Nonnegative matrix factorisations as probabilistic inference in composite models. In *Proc. 17th European Signal Processing Conference (EUSIPCO)*, pages 1913–1917, Glasgow, Scotland, Aug. 2009. URL `http://www.unice.fr/cfevotte/publications/proceedings/eusipco09a.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 `http://www.unice.fr/cfevotte/publications/journals/neco11.pdf`.

References IV

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 `http://www.unice.fr/cfevotte/publications/journals/neco09_is-nmf.pdf`.

C. Févotte, O. Cappé, and A. T. Cemgil. Efficient Markov chain Monte Carlo inference in composite models with space alternating data augmentation. In *Proc. IEEE Workshop on Statistical Signal Processing (SSP)*, pages 221 – 224, Nice, France, June 2011. URL `http://www.unice.fr/cfevotte/publications/proceedings/ssp11.pdf`.

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 `http://www.unice.fr/cfevotte/publications/proceedings/icassp13a.pdf`.

L. Finesso and P. Spreij. Nonnegative matrix factorization and I-divergence alternating minimization. *Linear Algebra and its Applications*, 416:270–287, 2006.

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.

References V

T. Hofmann. Probabilistic latent semantic indexing. In *Proc. 22nd International Conference on Research and Development in Information Retrieval (SIGIR)*, 1999. URL `http://www.cs.brown.edu/~th/papers/Hofmann-SIGIR99.pdf`.

J. Le Roux, H. Kameoka, N. Ono, A. de Cheveigné, and S. Sagayama. Single and multiple F0 contour estimation through parametric spectrogram modeling of speech in noisy environments. *IEEE Transactions on Audio, Speech and Language Processing*, 15(4):1135–1145, May 2007.

D. D. Lee and H. S. Seung. Learning the parts of objects with nonnegative matrix factorization. *Nature*, 401:788–791, 1999.

D. D. Lee and H. S. Seung. Algorithms for non-negative matrix factorization. In *Advances in Neural and Information Processing Systems 13*, pages 556–562, 2001.

L. B. Lucy. An iterative technique for the rectification of observed distributions. *Astronomical Journal*, 79:745–754, 1974. doi: 10.1086/111605.

P. Paatero and U. Tapper. Positive matrix factorization : A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, 5: 111–126, 1994.

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.

References VI

W. H. Richardson. Bayesian-based iterative method of image restoration. *Journal of the Optical Society of America*, 62:55–59, 1972.

M. Shashanka, B. Raj, and P. Smaragdis. Probabilistic latent variable models as nonnegative factorizations. *Computational Intelligence and Neuroscience*, 2008 (Article ID 947438):8 pages, 2008. doi:10.1155/2008/947438.

P. Smaragdis and J. C. Brown. Non-negative matrix factorization for polyphonic music transcription. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'03)*, Oct. 2003.

P. Smaragdis, M. Shashanka, and B. Raj. A sparse non-parametric approach for single channel separation of known sounds. In Y. Bengio, D. Schuurmans, J. Lafferty, C. K. I. Williams, and A. Culotta, editors, *Advances in Neural Information Processing Systems 22*, pages 1705–1713. MIT Press, 2009.