Some applications of non-negative matrix factorization and of information geometry in audio signal processing

Arnaud Dessein (Arshia Cont, Gérard Assayag, Guillaume Lemaitre)

Institute for Research and Coordination of Acoustics and Music, Paris, France
Japanese-French Laboratory for Informatics, Tokyo, Japan

October 22nd 2010
Outline

1. Introduction
2. Non-negative matrix factorization
3. Information geometry
4. Conclusion
Outline

1. Introduction
   - Presentation of the IRCAM
   - Research at the IRCAM
   - Motivations towards NMF and IG

2. Non-negative matrix factorization

3. Information geometry

4. Conclusion
What is the IRCAM?

- Status:
  - Institut de Recherche et Coordination Acoustique/Musique.
  - Associated with the Centre Georges Pompidou in Paris.

Figure: Centre Georges Pompidou (Renzo Piano and Richard Rogers).
What is the IRCAM?

- **Status:**
  - Institut de Recherche et Coordination Acoustique/Musique.
  - Associated with the Centre Georges Pompidou in Paris.
  - General head: Frank Madlener.
  - Scientific director: Hugues Vinet.

**Figure:** Frank Madlener and Hugues Vinet.
What is the IRCAM?

- **Status:**
  - Institut de Recherche et Coordination Acoustique/Musique.
  - Associated with the Centre Georges Pompidou in Paris.
  - General head: Frank Madlener.
  - Scientific director: Hugues Vinet.

- **History:**
  - 1970: the President Georges Pompidou asked Pierre Boulez to found an institution for musical research.

**Figure:** Georges Pompidou and Pierre Boulez.
What is the IRCAM?

- **Status:**
  - Institut de Recherche et Coordination Acoustique/Musique.
  - Associated with the Centre Georges Pompidou in Paris.
  - General head: Frank Madlener.
  - Scientific director: Hugues Vinet.

- **History:**
  - 1970: the President Georges Pompidou asked Pierre Boulez to found an institution for musical research.
  - 1973: the part underneath Place Igor Stravinsky was finished.

**Figure:** Stravinsky fountain (Tinguely and Niki de Saint Phalle) and scale model of the IRCAM.
What is the IRCAM?

- Status:
  - Institut de Recherche et Coordination Acoustique/Musique.
  - Associated with the Centre Georges Pompidou in Paris.
  - General head: Frank Madlener.
  - Scientific director: Hugues Vinet.

- History:
  - 1970: the President Georges Pompidou asked Pierre Boulez to found an institution for musical research.
  - 1973: the part underneath Place Igor Stravinsky was finished.
  - 1977: the center opened.

**Figure:** IRCAM (Renzo Piano and Richard Rogers).
What is the IRCAM?

- **Status:**
  - Institut de Recherche et Coordination Acoustique/Musique.
  - Associated with the Centre Georges Pompidou in Paris.
  - General head: Frank Madlener.
  - Scientific director: Hugues Vinet.

- **History:**
  - 1970: the President Georges Pompidou asked Pierre Boulez to found an institution for musical research.
  - 1973: the part underneath Place Igor Stravinsky was finished.
  - 1977: the center opened.

- **Figures:**
  - 150 people: artists, scientists, technicians, administrative staff.
  - 11,000,000 euros of annual budget.
  - 8 research teams.

*Figure: IRCAM (Renzo Piano and Richard Rogers).*
What do we do?

- Research teams:
  - Instrumental Acoustics.
  - Acoustic and Cognitive Spaces.
  - Perception and Sound Design.
  - Sound Analysis-Synthesis.
  - Musical Representations.
  - Analysis of musical practices.
  - Real-Time Musical Interactions.
  - Online Services.

“Working transversally for music research.”
Researchers and musicians working together on multidisciplinary projects centered around music and exploration of sounds.
What do we do?

- Research teams:
  - Instrumental Acoustics.
  - Acoustic and Cognitive Spaces.
  - Perception and Sound Design.
  - Sound Analysis-Synthesis.
  - Musical Representations.
  - Analysis of musical practices.
  - Real-Time Musical Interactions.
  - Online Services.

- Research thematics:
  - Sound synthesis and processing.

“Creating new sounds as an extension of acoustic instruments.”
Writing of sound, digital signal processing for sound transformation and synthesis, physical modeling, virtual instrument design.
What do we do?

- **Research teams:**
  - Instrumental Acoustics.
  - Acoustic and Cognitive Spaces.
  - Perception and Sound Design.
  - Sound Analysis-Synthesis.
  - Musical Representations.
  - Analysis of musical practices.
  - Real-Time Musical Interactions.
  - Online Services.

- **Research thematics:**
  - Sound synthesis and processing.
  - Live interaction.

"Interacting with the computer during performances."
Writing of time, performance capture/analysis (audio, gesture), synchronization (event detection, score following, shape recognition), multimedia/multimodality (dance, image, text).
What do we do?

- Research teams:
  - Instrumental Acoustics.
  - Acoustic and Cognitive Spaces.
  - Perception and Sound Design.
  - Sound Analysis-Synthesis.
  - Musical Representations.
  - Analysis of musical practices.
  - Real-Time Musical Interactions.
  - Online Services.

- Research thematics:
  - Sound synthesis and processing.
  - Live interaction.
  - Computer-aided composition.

“Using the computer as a reflexive support of composition.”
Writing of music, formalizing/producing/managing complex musical structures, assisting composition/orchestration.
What do we do?

- **Research teams:**
  - Instrumental Acoustics.
  - Acoustic and Cognitive Spaces.
  - Perception and Sound Design.
  - Sound Analysis-Synthesis.
  - Musical Representations.
  - Analysis of musical practices.
  - Real-Time Musical Interactions.
  - Online Services.

- **Research thematics:**
  - Sound synthesis and processing.
  - Live interaction.
  - Computer-aided composition.
  - Sound spatialization.

"Composing space as a dimension of musical expression."

Writing of space, simulation of static/mobile sources and of acoustic spaces, perception and cognition of space.
What do we need?

<table>
<thead>
<tr>
<th>Representation</th>
<th>Type</th>
<th>Data rate</th>
<th>Low information quantity</th>
<th>High information quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Semantic</td>
<td>Semantic</td>
<td>&lt; 0.1 Hz</td>
<td>Implicit knowledge</td>
<td>Explicit representations</td>
</tr>
<tr>
<td>Symbolic</td>
<td>Symbolic</td>
<td>0.1-25 Hz</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>Control</td>
<td>10Hz-1kHz</td>
<td>Information Generation: Synthesis</td>
<td>Information Reduction: Analysis</td>
</tr>
<tr>
<td>Signal</td>
<td>Signal</td>
<td>10-100 kHz</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical</td>
<td>Physical</td>
<td>10-100 kHz</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure:** Levels of representation of audio, waveform and spectrogram representations.
What do we need?

<table>
<thead>
<tr>
<th>Representation</th>
<th>Type</th>
<th>Data rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Semantic</td>
<td>&lt; 0.1 Hz</td>
<td></td>
</tr>
<tr>
<td>Symbolic</td>
<td>0.1-2.5 Hz</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>10 Hz-1 kHz</td>
<td></td>
</tr>
<tr>
<td>Signal</td>
<td>10-100 kHz</td>
<td></td>
</tr>
<tr>
<td>Physical</td>
<td>10-100 kHz</td>
<td></td>
</tr>
</tbody>
</table>

Figure: Levels of representation of audio, waveform and spectrogram representations.

- Fill in the gap between signal and symbolic representations.
What do we need?

- Fill in the gap between signal and symbolic representations.
- Devise computational tools for complex real-time settings.

**Figure:** Levels of representation of audio, waveform, and spectrogram representations.
What do we need?

- Fill in the gap between signal and symbolic representations.
- Devise computational tools for complex real-time settings.
- Two approaches:
  - NMF: current trend, structural a priori, reductionist.
  - IG: new trend, no structural a priori, holistic.

**Figure:** Levels of representation of audio, waveform and spectrogram representations.
Introduction

Non-negative matrix factorization

Background

Proposed system for real-time recognition of multiple sources

Sparsity and non-negative decomposition

Beta-divergence and non-negative decomposition

Results

Discussion

Information geometry

Conclusion
What is NMF?

Standard NMF model [Lee & Seung, 1999].

Let \( \mathbf{V} \in \mathbb{R}^{n \times m}_+ \) and \( r < \min(n, m) \), find \( \mathbf{W} \in \mathbb{R}^{n \times r}_+ \) and \( \mathbf{H} \in \mathbb{R}^{r \times m}_+ \) such that:

\[
\mathbf{V} \approx \mathbf{WH}
\]
What is NMF?

Standard NMF model [Lee & Seung, 1999].

Let $V \in \mathbb{R}_+^{n \times m}$ and $r < \min(n, m)$, find $W \in \mathbb{R}_+^{n \times r}$ and $H \in \mathbb{R}_+^{r \times m}$ such that:

$$V \approx WH$$

- **Interpretation:**
  $$v_j \approx Wh_j = \sum_i h_{ij}w_i$$
  - $w_i$: *basis vectors*.
  - $h_{ij}$: *decomposition coefficients*.
Introduction
Non-negative matrix factorization
Information geometry
Conclusion

How to solve NMF?

Standard NMF problem.

Minimize $C(W, H) = \frac{1}{2} \| V - WH \|_F^2$

subject to $W \in \mathbb{R}_{+}^{n \times r}, H \in \mathbb{R}_{+}^{r \times m}$

- Standard algorithms [Berry et al., 2007, Cichocki et al., 2009]:
How to solve NMF?

Standard NMF problem.

\[
\text{Minimize } \mathcal{C}(W, H) = \frac{1}{2} \| V - WH \|_F^2 \\
\text{subject to } W \in \mathbb{R}^{n \times r}_+, H \in \mathbb{R}^{r \times m}_+ 
\]

- Standard algorithms [Berry et al., 2007, Cichocki et al., 2009]
  - Alternate non-negative least-squares:
    \[
    H \leftarrow \text{arg min}_{H \in \mathbb{R}^{r \times m}_+} \| V - WH \|_F^2 \quad W \leftarrow \text{arg min}_{W \in \mathbb{R}^{n \times r}_+} \| V - WH \|_F^2 
    \]
How to solve NMF?

Standard NMF problem.

\[
\text{Minimize} \quad C(W, H) = \frac{1}{2} \| V - WH \|_F^2 \\
\text{subject to} \quad W \in \mathbb{R}^{n \times r}_+, H \in \mathbb{R}^{r \times m}_+
\]

- Standard algorithms [Berry et al., 2007, Cichocki et al., 2009]:
  - Alternate non-negative least-squares:
    \[
    H \leftarrow \text{arg min}_{H \in \mathbb{R}^{r \times m}_+} \| V - WH \|_F^2 \quad \text{subject to} \quad W \in \mathbb{R}^{n \times r}_+ \\
    W \leftarrow \text{arg min}_{W \in \mathbb{R}^{n \times r}_+} \| V - WH \|_F^2
    \]
  - Additive updates:
    \[
    h_{ij} \leftarrow h_{ij} - \mu_{ij} \frac{\partial C(W, H)}{\partial h_{ij}} \\
    w_{ij} \leftarrow w_{ij} - \eta_{ij} \frac{\partial C(W, H)}{\partial w_{ij}}
    \]
How to solve NMF?

**Standard NMF problem.**

\[
\text{Minimize } \quad C(W, H) = \frac{1}{2} \| V - WH \|_F^2
\]
subject to \( W \in \mathbb{R}^{n \times r}_{+}, H \in \mathbb{R}^{r \times m}_{+} \)

- Standard algorithms [Berry et al., 2007, Cichocki et al., 2009]:
  - Alternate non-negative least-squares:
    \[
    H \leftarrow \arg\min_{H \in \mathbb{R}^{r \times m}_{+}} \| V - WH \|_F^2, \quad W \leftarrow \arg\min_{W \in \mathbb{R}^{n \times r}_{+}} \| V - WH \|_F^2
    \]
  - Additive updates:
    \[
    h_{ij} \leftarrow h_{ij} - \mu_{ij} \frac{\partial C(W, H)}{\partial h_{ij}}
    \]
    \[
    w_{ij} \leftarrow w_{ij} - \eta_{ij} \frac{\partial C(W, H)}{\partial w_{ij}}
    \]
  - Multiplicative updates:
    \[
    H \leftarrow H \odot \frac{W^T V}{W^T WH}
    \]
    \[
    W \leftarrow W \odot \frac{VH^T}{WHH^T}
    \]
How to use NMF for sound analysis?

Model reminder.

\[ V \approx WH \]
\[ v_j \approx Wh_j = \sum_i h_{ij}w_i \]

- Usual setting:
  - \( V \): time-frequency representation.
  - \( v_j \): successive frames.
  - \( w_i \): spectral models.
  - \( h_{ij} \): activation coefficients.
How to use NMF for sound analysis?

Model reminder.

\[ V \approx WH \]

\[ v_j \approx Wh_j = \sum_i h_{ij}w_i \]

- Usual setting:
  - \( V \): time-frequency representation.
  - \( v_j \): successive frames.
  - \( w_i \): spectral models.
  - \( h_{ij} \): activation coefficients.

- Examples of application: source separation [Cichocki et al., 2009], but also polyphonic music transcription [Smaragdis & Brown, 2003, Abdallah & Plumbley, 2004, Virtanen & Klapuri, 2006, Raczyński et al., 2007, Bertin et al., 2010, Vincent et al., 2010].
How to use NMF for sound analysis?

Model reminder.

\[ V \approx WH \]
\[ v_j \approx Wh_j = \sum_i h_{ij}w_i \]

- Usual setting:
  - \( V \): time-frequency representation.
  - \( v_j \): successive frames.
  - \( w_i \): spectral models.
  - \( h_{ij} \): activation coefficients.

- Examples of application: source separation [Cichocki et al., 2009], but also polyphonic music transcription [Smaragdis & Brown, 2003, Abdallah & Plumbley, 2004, Virtanen & Klapuri, 2006, Raczyński et al., 2007, Bertin et al., 2010, Vincent et al., 2010].

- Limits for real-time settings.
How to adapt NMF to real-time settings?

Towards non-negative decomposition:
1. Learn source templates $w_i$ before decomposition.
2. Stack this templates in a dictionary $W$ kept fixed during decomposition.
3. Project the incoming audio stream onto the dictionary $W$ in real-time.
How to adapt NMF to real-time settings?

Towards *non-negative decomposition*:  
1. Learn source templates $w_i$ before decomposition.  
2. Stack this templates in a dictionary $W$ kept fixed during decomposition.  
3. Project the incoming audio stream onto the dictionary $W$ in real-time.

Applications:  
- Speech analysis [Sha & Saul, 2005].  
- Score following [Cont, 2006].  
- Multi-f0 and multi-instrument recognition [Cont et al., 2007].  
- Sight-reading evaluation [Cheng et al., 2008].  
- Polyphonic music transcription [Niedermayer, 2008].
General architecture

Figure: Schema of the general architecture of the system.
Goal: learn a dictionary $W$ source templates.

**Figure:** Schema of templates learning.
Template learning

- **Goal:** learn a dictionary $W$ source templates.
- **Method:** apply standard NMF to each sound sample $k$ with a factorization rank $r = 1$.

**Figure:** Schema of templates learning.
Template learning

- **Goal:** learn a dictionary $\mathbf{W}$ source templates.
- **Method:** apply standard NMF to each sound sample $k$ with a factorization rank $r = 1$.
- **Example:** the sources are the 88 notes of the piano.

**Figure:** Templates learned for the piano.

**Figure:** Schema of templates learning.

Template learning (off-line)

- Sound source database
- Short-time sound representation
- $\mathbf{V}^{(k)}$
- Non-negative matrix factorization
  $\mathbf{V}^{(k)} \approx \mathbf{W}^{(k)} \mathbf{H}^{(k)}$
- $\mathbf{W}^{(k)}$
- Source templates
Goal: obtain in real-time the activations of the sources present in the auditory scene.

Figure: Schema of audio stream decomposition.
Audio stream decomposition

- **Goal:** obtain in real-time the activations of the sources present in the auditory scene.
- **Method:** employ non-negative decomposition to project the audio stream onto $\mathbf{W}$.

**Figure:** Schema of audio stream decomposition.
Audio stream decomposition

- **Goal**: obtain in real-time the activations of the sources present in the auditory scene.
- **Method**: employ non-negative decomposition to project the audio stream onto $W$.
- **Example**: chromatic scale on the piano.

**Figure**: Activations for a chromatic scale.

**Figure**: Schema of audio stream decomposition.
Audio stream decomposition

- **Goal**: obtain in real-time the activations of the sources present in the auditory scene.
- **Method**: employ non-negative decomposition to project the audio stream onto $\mathbf{W}$.
- **Example**: chromatic scale on the piano.

**Figure**: Activations for a chromatic scale.

- Two approaches investigated to control the decomposition: sparsity, beta-divergence.
What is sparsity?

**Definition.**

A vector $\mathbf{x}$ is *sparse* when its energy is concentrated in a few coefficients.

- Sparsity measures [Karvanen & Cichocki, 2003]:

  - Sparsity measures [Karvanen & Cichocki, 2003]:
What is sparsity?

Definition.

A vector $x$ is *sparse* when its energy is concentrated in a few coefficients.

- Sparsity measures [Karvanen & Cichocki, 2003]:

$$sp(x) = \frac{\|x\|_0}{n} = \frac{\text{card} \{ i : x_i \neq 0 \}}{n}$$
Definition.

A vector $\mathbf{x}$ is *sparse* when its energy is concentrated in a few coefficients.

- Sparsity measures [Karvanen & Cichocki, 2003]:

$$\text{sp}(\mathbf{x}) = \frac{\|\mathbf{x}\|_0}{n} = \frac{\text{card} \{ i : x_i \neq 0 \}}{n}$$

$$\text{sp}(\mathbf{x}) = \frac{\|\mathbf{x}\|_{0,\varepsilon}}{n} = \frac{\text{card} \{ i : |x_i| \geq \varepsilon \}}{n} \text{ with } \varepsilon > 0$$
What is sparsity?

**Definition.**

A vector \( \mathbf{x} \) is *sparse* when its energy is concentrated in a few coefficients.

- Sparsity measures [Karvonen & Cichocki, 2003]:
  
  \[
  \text{sp}(\mathbf{x}) = \frac{\|\mathbf{x}\|_0}{n} = \frac{\text{card}\{i : x_i \neq 0\}}{n}
  \]

  \[
  \text{sp}(\mathbf{x}) = \frac{\|\mathbf{x}\|_{0, \varepsilon}}{n} = \frac{\text{card}\{i : |x_i| \geq \varepsilon\}}{n} \quad \text{with } \varepsilon > 0
  \]

  \[
  \text{sp}(\mathbf{x}) = \frac{\sum_i \tanh \left( |a x_i|^b \right)}{n} \quad \text{with } a > 0 \text{ et } b \geq 1
  \]
What is sparsity?

**Definition.**

A vector $\mathbf{x}$ is *sparse* when its energy is concentrated in a few coefficients.

- Sparsity measures [Karvanen & Cichocki, 2003]:

  $\text{sp}(\mathbf{x}) = \frac{\|\mathbf{x}\|_0}{n} = \frac{\text{card } \{i : x_i \neq 0\}}{n}$

  $\text{sp}(\mathbf{x}) = \frac{\|\mathbf{x}\|_{0, \varepsilon}}{n} = \frac{\text{card } \{i : |x_i| \geq \varepsilon\}}{n}$ with $\varepsilon > 0$

  $\text{sp}(\mathbf{x}) = \frac{\sum_i \tanh(|ax_i|^b)}{n}$ with $a > 0$ et $b \geq 1$

  $\text{sp}(\mathbf{x}) = \frac{\|\mathbf{x}\|_p}{n} = \frac{\sum_i |x_i|^p}{n}$ with $0 < p \leq 1$
How to obtain sparse NMF?

- Penalty and multiplicative updates
How to obtain sparse NMF?

- Penalty and multiplicative updates [Eggert & Körner, 2004, Virtanen, 2007].
- Penalty and alternate non-negative least-squares [Albright et al., 2006].

\[ \text{sp}(x) = \sqrt{n} - \frac{\|x\|_1}{\|x\|_2} \sqrt{n} - 1 \]
How to obtain sparse NMF?

- Penalty and multiplicative updates [Eggert & Körner, 2004, Virtanen, 2007].
- Penalty and alternate non-negative least-squares [Albright et al., 2006].
- Penalty and projected gradient [Hoyer, 2002].
How to obtain sparse NMF?

- Penalty and multiplicative updates [Eggert & Körner, 2004, Virtanen, 2007].
- Penalty and alternate non-negative least-squares [Albright et al., 2006].
- Penalty and projected gradient [Hoyer, 2002].
- Constraint and projected gradient [Hoyer, 2004].

\[ sp(x) = \frac{\sqrt{n} - \|x\|_1/\|x\|_2}{\sqrt{n} - 1} \]

**Figure:** Projection onto the s-sparsity cone.
How to obtain sparse NMF?

- Penalty and multiplicative updates [Eggert & Körner, 2004, Virtanen, 2007].
- Penalty and alternate non-negative least-squares [Albright et al., 2006].
- Penalty and projected gradient [Hoyer, 2002].
- Constraint and projected gradient [Hoyer, 2004].

\[
sp(x) = \frac{\sqrt{n} - \|x\|_1/\|x\|_2}{\sqrt{n} - 1}
\]

- Constraint and second-order cone programming [Heiler & Schnörr, 2005, Heiler & Schnörr, 2006].

Figure: Optimization between the \( s_{\text{min}} \)- and \( s_{\text{max}} \)-sparsity cones.
How to obtain sparse NMF?

- Penalty and multiplicative updates [Eggert & Körner, 2004, Virtanen, 2007].
- Penalty and alternate non-negative least-squares [Albright et al., 2006].
- Penalty and projected gradient [Hoyer, 2002].
- Constraint and projected gradient [Hoyer, 2004].

\[
sp(x) = \sqrt{n - \|x\|_1/\|x\|_2} \frac{\sqrt{n} - 1}{\sqrt{n} - 1}
\]

- Constraint and second-order cone programming [Heiler & Schnörr, 2005, Heiler & Schnörr, 2006].
- Penalty and convex quadratic programming [Zdunek & Cichocki, 2008].
Proposed approach based on convex quadratic programming

Problem.

Minimize \[ \frac{1}{2} \| v - W h \|_2^2 + \lambda_1 \| h \|_1 + \frac{\lambda_2}{2} \| h \|_2^2 \]

subject to \( h \in \mathbb{R}_+^r, \ s_{\min} \leq \text{sp}(h) \leq s_{\max} \)
Proposed approach based on convex quadratic programming

Problem.

Minimize \[ \frac{1}{2} \| \mathbf{v} - \mathbf{W} \mathbf{h} \|_2^2 + \lambda_1 \| \mathbf{h} \|_1 + \frac{\lambda_2}{2} \| \mathbf{h} \|_2^2 \]
subject to \[ \mathbf{h} \in \mathbb{R}_{++}^r, \ s_{min} \leq \text{sp}(\mathbf{h}) \leq s_{max} \]

- Sparsity parameter: \( \lambda_1 \geq 0 \).
- Regularization parameter: \( \lambda_2 \geq 0 \).
- Constraint parameters: \( 0 \leq s_{min} < s_{max} \leq 1 \).
Proposed approach based on convex quadratic programming

Problem.

Minimize \( \frac{1}{2} \|v - Wh\|_2^2 + \lambda_1 \|h\|_1 + \frac{\lambda_2}{2} \|h\|_2^2 \)

subject to \( h \in \mathbb{R}^r_+ \), \( s_{\text{min}} \leq \text{sp}(h) \leq s_{\text{max}} \)

- Sparsity parameter: \( \lambda_1 \geq 0 \).
- Regularization parameter: \( \lambda_2 \geq 0 \).
- Constraint parameters: \( 0 \leq s_{\text{min}} < s_{\text{max}} \leq 1 \).
- Algorithm:
  - Update \( h \) with a sequence of convex quadratic programs.
  - Approximation of the sparsity cones with tangent planes.

Figure: Approximation of the sparsity cones.
Illustrative example

**Figure:** Activations for a chromatic scale, $\lambda_1 = 0$. 
Illustrative example

**Figure:** Activations for a chromatic scale, $\lambda_1 = 1$. 
Illustrative example

**Figure**: Activations for a chromatic scale, $\lambda_1 = 5$. 
Figure: Activations for a chromatic scale, $\lambda_1 = 10$. 
Figure: Activations for a chromatic scale, $\lambda_1 = 50$. 

[Diagram showing activations for a chromatic scale]
Illustrative example

Figure: Activations for a chromatic scale, $\lambda_1 = 100$. 
What is the beta-divergence?

Definition [Eguchi & Kano, 2001].

Let $\beta \in \mathbb{R}$ and $x, y \in \mathbb{R}^{++}$, the $\beta$-divergence from $x$ to $y$ is defined by:

$$d_{\beta}(x|y) = \frac{1}{\beta(\beta - 1)}(x^\beta + (\beta - 1)y^\beta - \beta xy^{\beta - 1})$$
**What is the beta-divergence?**

**Definition [Eguchi & Kano, 2001].**

Let $\beta \in \mathbb{R}$ and $x, y \in \mathbb{R}^+$, the $\beta$-divergence from $x$ to $y$ is defined by:

$$d_\beta(x|y) = \frac{1}{\beta(\beta - 1)}(x^\beta + (\beta - 1)y^\beta - \beta xy^{\beta-1})$$

- **Particular cases:**
  - **Itakura-Saito divergence:** $d_{\beta=0}(x|y) = \frac{x}{y} - \log \frac{x}{y} - 1$.
  - **Kullback-Leibler divergence:** $d_{\beta=1}(x|y) = x \log \frac{x}{y} + y - x$.
  - **Euclidean distance:** $d_{\beta=2}(x|y) = \frac{1}{2}(x - y)^2$. 

arnaud.dessein@ircam.fr October 22nd 2010
What is the beta-divergence?

**Definition [Eguchi & Kano, 2001].**

Let $\beta \in \mathbb{R}$ and $x, y \in \mathbb{R}^{++}$, the $\beta$-divergence from $x$ to $y$ is defined by:

$$d_{\beta}(x|y) = \frac{1}{\beta(\beta - 1)}(x^\beta + (\beta - 1)y^\beta - \beta xy^{\beta - 1})$$

- **Particular cases:**
  - Itakura-Saito divergence: $d_{\beta=0}(x|y) = \frac{x}{y} - \log \frac{x}{y} - 1$.
  - Kullback-Leibler divergence: $d_{\beta=1}(x|y) = x \log \frac{x}{y} + y - x$.
  - Euclidean distance: $d_{\beta=2}(x|y) = \frac{1}{2}(x - y)^2$.
  - Generalized distance: $d_{\beta}(x|y) \geq 0$ and $d_{\beta}(x|y) = 0$ iff $x = y$. 

[arnaud.dessein@ircam.fr October 22nd 2010 RIKEN BSI Seminar]
What is the beta-divergence?

Definition [Eguchi & Kano, 2001].

Let $\beta \in \mathbb{R}$ and $x, y \in \mathbb{R}^{++}$, the $\beta$-divergence from $x$ to $y$ is defined by:

$$d_\beta(x|y) = \frac{1}{\beta(\beta - 1)} (x^\beta + (\beta - 1)y^\beta - \beta xy^{\beta-1})$$

- Particular cases:
  - Itakura-Saito divergence: $d_{\beta=0}(x|y) = \frac{x}{y} - \log \frac{x}{y} - 1$.
  - Kullback-Leibler divergence: $d_{\beta=1}(x|y) = x \log \frac{x}{y} + y - x$.
  - Euclidean distance: $d_{\beta=2}(x|y) = \frac{1}{2} (x - y)^2$.
- Generalized distance: $d_\beta(x|y) \geq 0$ and $d_\beta(x|y) = 0$ iff $x = y$.
- Scaling property: $d_\beta(\lambda x|\lambda y) = \lambda^\beta d_\beta(x|y)$. 

arnaud.dessein@ircam.fr  
October 22nd 2010  
RIKEN BSI Seminar
How to use the beta-divergence in NMF?

NMF problem with the beta-divergence.

\[
\text{Minimize} \quad D_\beta(V|WH) = \sum_{i,j} d_\beta(v_{ij} | [WH]_{ij}) \\
\text{subject to} \quad W \in \mathbb{R}^{n \times r}_{++}, \ H \in \mathbb{R}^{r \times m}_{++}
\]
How to use the beta-divergence in NMF?

NMF problem with the beta-divergence.

Minimize \( D_\beta(V \mid WH) = \sum_{i,j} d_\beta(v_{ij} \mid [WH]_{ij}) \)

subject to \( W \in \mathbb{R}^{n \times r}_{++}, H \in \mathbb{R}^{r \times m}_{++} \)

- Algorithms [Cichocki et al., 2009]:
  - Multiplicative updates:

\[
H \leftarrow H \otimes \frac{W^T((WH)^{\beta-2} \otimes V)}{W^T(WH)^{\beta-1}}
\]

\[
W \leftarrow W \otimes \frac{(WH)^{\beta-2} \otimes V}{(WH)^{\beta-1}H^T}
\]

- Employed to interpolate between \( d_E \) and \( d_{KL} \) [Kompass, 2007].
- Employed in audio [O'Grady & Pearlmutter, 2008, Bertin et al., 2009, Bertin et al., 2010, Vincent et al., 2010].
- Employed in audio for the particular case \( d_{IS} \) [Févotte et al., 2009].
How to use the beta-divergence in NMF?

NMF problem with the beta-divergence.

Minimize $\mathcal{D}_\beta(V \mid WH) = \sum_{i,j} d_\beta(v_{ij} \mid [WH]_{ij})$

subject to $W \in \mathbb{R}^{n \times r}_{++}, H \in \mathbb{R}^{r \times m}_{++}$

- Algorithms [Cichocki et al., 2009]:
  - Multiplicative updates:
    $$H \leftarrow H \otimes \frac{W^T ((WH)^{\beta-2} \otimes V)}{W^T (WH)^{\beta-1}}$$
    $$W \leftarrow W \otimes \frac{(WH)^{\beta-2} \otimes V H^T}{(WH)^{\beta-1} H^T}$$

- Employed to interpolate between $d_E$ and $d_{KL}$ [Kompass, 2007].
How to use the beta-divergence in NMF?

NMF problem with the beta-divergence.

Minimize $D_\beta(V | WH) = \sum_{i,j} d_\beta(v_{ij} | [WH]_{ij})$

subject to $W \in \mathbb{R}^{n \times r}_{++}, H \in \mathbb{R}^{r \times m}_{++}$

- Algorithms [Cichocki et al., 2009]:
  - Multiplicative updates:
    $$H \leftarrow H \otimes \frac{W^T (WH)^{\beta-2} \otimes V}{W^T (WH)^{\beta-1}} \\ W \leftarrow W \otimes \frac{(WH)^{\beta-2} \otimes V H^T}{(WH)^{\beta-1} H^T}$$

- Employed to interpolate between $d_E$ and $d_{KL}$ [Kompass, 2007].

- Employed in audio [O'Grady & Pearlmutter, 2008, Bertin et al., 2009, Bertin et al., 2010, Vincent et al., 2010].
How to use the beta-divergence in NMF?

NMF problem with the beta-divergence.

Minimize \( D_\beta(V|WH) = \sum_{i,j} d_\beta(v_{ij} | [WH]_{ij}) \)

subject to \( W \in \mathbb{R}^{n \times r}_{++}, H \in \mathbb{R}^{r \times m}_{++} \)

- Algorithms [Cichocki et al., 2009]:
  - Multiplicative updates:
    \[
    H \leftarrow H \otimes \frac{W^T((WH)^{\beta-2} \otimes V)}{W^T(WH)^{\beta-1}} \quad W \leftarrow W \otimes \frac{(WH)^{\beta-2} \otimes V}{(WH)^{\beta-1}H^T}
    \]

- Employed to interpolate between \( d_E \) and \( d_{KL} \) [Kompass, 2007].

- Employed in audio [O'Grady & Pearlmutter, 2008, Bertin et al., 2009, Bertin et al., 2010, Vincent et al., 2010].

- Employed in audio for the particular case \( d_{IS} \) [Févotte et al., 2009].
Proposed approach based on multiplicative updates

Problem.

Minimize \( D_\beta(v|Wh) \) subject to \( h \in \mathbb{R}^r_{++} \)
Problem.

Minimize $D_\beta(v|Wh)$ subject to $h \in \mathbb{R}^r_+$

- Decomposition parameter: $\beta \in \mathbb{R}$. 
Proposed approach based on multiplicative updates

Problem.

Minimize \( D_\beta(v | Wh) \) subject to \( h \in \mathbb{R}^r_+ \)

- Decomposition parameter: \( \beta \in \mathbb{R} \).
- Algorithm:
  1. Initialize \( h \) with positive values.
  2. Update \( h \) until convergence:

\[
  h \leftarrow h \odot \frac{W^T((Wh)\cdot\beta^{-2} \otimes v)}{W^T(Wh)\cdot\beta^{-1}}
\]
Proposed approach based on multiplicative updates

Problem.

\[
\text{Minimize } D_\beta(v \vert Wh) \text{ subject to } h \in \mathbb{R}^r_{++}
\]

- Decomposition parameter: \( \beta \in \mathbb{R} \).
- Algorithm:
  1. Initialize \( h \) with positive values.
  2. Update \( h \) until convergence:

\[
h \leftarrow h \odot \frac{W^T((Wh)^{\beta-2} \odot v)}{W^T(Wh)^{\beta-1}}
\]

- Updates tailored to real-time:

\[
h \leftarrow h \odot \frac{(W^T \odot (ev^T))(Wh)^{\beta-2}}{W^T(Wh)^{\beta-1}}
\]
Illustrative example

\textbf{Figure:} Activations for a chromatic scale, $\beta = 3$. 

\begin{center}
\includegraphics[width=0.8\textwidth]{activation_plot.png}
\end{center}
Illustrative example

**Figure:** Activations for a chromatic scale, $\beta = 2$. 
Illustrative example

Figure: Activations for a chromatic scale, $\beta = 1.5$. 
Illustrative example

**Figure:** Activations for a chromatic scale, $\beta = 1.$
**Illustrative example**

*Figure:* Activations for a chromatic scale, $\beta = 0.5$.  

---

arnaud.dessein@ircam.fr

October 22nd 2010

RIKEN BSI Seminar
Illustrative example

**Figure**: Activations for a chromatic scale, $\beta = 0$. 

氨基 acid residues for a chromatic scale, $\beta = 0$. 

arnaud.dessein@ircam.fr October 22nd 2010
Drum transcription

Figure: Decomposition of a drum loop.
Environmental sound detection in complex scenes

Figure: Decomposition of a complex environmental scene.

(a) MU.  
(b) PGC.  
(c) SCQP.
Polyphonic music transcription

- Subjective evaluation: music synthesized with real piano samples (demo).
Polyphonic music transcription

- Subjective evaluation: music synthesized with real piano samples (demo).
- Objective evaluation: recorded piano music.

<table>
<thead>
<tr>
<th>Alg.</th>
<th>$P_1$</th>
<th>$R_1$</th>
<th>$F_1$</th>
<th>$A_1$</th>
<th>$M_1$</th>
<th>$P_2$</th>
<th>$R_2$</th>
<th>$F_2$</th>
<th>$A_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCQP</td>
<td>64.1</td>
<td>56.2</td>
<td>59.9</td>
<td>42.7</td>
<td>53.7</td>
<td>25.4</td>
<td>22.3</td>
<td>23.7</td>
<td>13.5</td>
</tr>
<tr>
<td>BMU</td>
<td>75.5</td>
<td>67.1</td>
<td>71.1</td>
<td>55.1</td>
<td>56.7</td>
<td>30.0</td>
<td>26.6</td>
<td>28.2</td>
<td>16.4</td>
</tr>
<tr>
<td>MU</td>
<td>57.9</td>
<td>58.2</td>
<td>58.1</td>
<td>40.9</td>
<td>53.9</td>
<td>21.4</td>
<td>21.6</td>
<td>21.5</td>
<td>12.0</td>
</tr>
<tr>
<td>Hoyer</td>
<td>57.2</td>
<td>56.3</td>
<td>56.8</td>
<td>39.6</td>
<td>54.1</td>
<td>21.0</td>
<td>20.7</td>
<td>20.8</td>
<td>11.6</td>
</tr>
<tr>
<td>Vincent</td>
<td>58.1</td>
<td>73.7</td>
<td>65.0</td>
<td>48.1</td>
<td>57.7</td>
<td>20.7</td>
<td>26.3</td>
<td>23.2</td>
<td>13.1</td>
</tr>
<tr>
<td>Yeh</td>
<td>33.0</td>
<td>58.8</td>
<td>42.3</td>
<td>26.8</td>
<td>55.1</td>
<td>11.6</td>
<td>20.7</td>
<td>14.9</td>
<td>8.0</td>
</tr>
</tbody>
</table>

Table: Results of the evaluation for recorded piano music.
Polyphonic music transcription

- Subjective evaluation: music synthesized with real piano samples (demo).
- Objective evaluation: recorded piano music.

<table>
<thead>
<tr>
<th>Alg.</th>
<th>$P_1$</th>
<th>$R_1$</th>
<th>$F_1$</th>
<th>$A_1$</th>
<th>$M_1$</th>
<th>$P_2$</th>
<th>$R_2$</th>
<th>$F_2$</th>
<th>$A_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCQP</td>
<td>64.1</td>
<td>56.2</td>
<td>59.9</td>
<td>42.7</td>
<td>53.7</td>
<td>25.4</td>
<td>22.3</td>
<td>23.7</td>
<td>13.5</td>
</tr>
<tr>
<td>BMU</td>
<td>75.5</td>
<td>67.1</td>
<td>71.1</td>
<td>55.1</td>
<td>56.7</td>
<td>30.0</td>
<td>26.6</td>
<td>28.2</td>
<td>16.4</td>
</tr>
<tr>
<td>MU</td>
<td>57.9</td>
<td>58.2</td>
<td>58.1</td>
<td>40.9</td>
<td>53.9</td>
<td>21.4</td>
<td>21.6</td>
<td>21.5</td>
<td>12.0</td>
</tr>
<tr>
<td>Hoyer</td>
<td>57.2</td>
<td>56.3</td>
<td>56.8</td>
<td>39.6</td>
<td>54.1</td>
<td>21.0</td>
<td>20.7</td>
<td>20.8</td>
<td>11.6</td>
</tr>
<tr>
<td>Vincent</td>
<td>58.1</td>
<td>73.7</td>
<td>65.0</td>
<td>48.1</td>
<td>57.7</td>
<td>20.7</td>
<td>26.3</td>
<td>23.2</td>
<td>13.1</td>
</tr>
<tr>
<td>Yeh</td>
<td>33.0</td>
<td>58.8</td>
<td>42.3</td>
<td>26.8</td>
<td>55.1</td>
<td>11.6</td>
<td>20.7</td>
<td>14.9</td>
<td>8.0</td>
</tr>
</tbody>
</table>

**Table:** Results of the evaluation for recorded piano music.

- International evaluation: 2nd rank at MIREX 2010 (against off-line systems) for note-level transcription of polyphonic music (not only piano).
What we (don’t) have

Summary and perspectives.

- Representations.
- Audio decomposition.
- Template learning.
- Temporality of events.
What we (don’t) have

Summary and perspectives.

- Representations.
- Audio decomposition.
- Template learning.
- Temporality of events.

- Many possibilities.
- Complex representations for $\mathbf{V}$ and $\mathbf{W}$.
- Tensors for multichannel information [Cichocki et al., 2009].
Summary and perspectives.

- Representations.
- Audio decomposition.
- Template learning.
- Temporality of events.

- Euclidean geometry and sparsity.
- More general geometries and beta-divergence.
- Bayesian interpretation [Févotte et al., 2009, Bertin et al., 2010].
What we (don’t) have

Summary and perspectives.

- Representations.
- Audio decomposition.
- Template learning.
- Temporality of events.

- Rank-one standard NMF.
- Extensions of standard NMF.
- Harmonicity constraints [Bertin et al., 2010, Vincent et al., 2010].
What we (don’t) have

Summary and perspectives.

- Representations.
- Audio decomposition.
- Template learning.
- Temporality of events.

- Extended NMF model [Smaragdis, 2004].
- State representation.
- Non-negative HMM [Mysore, 2010].
Outline

1. Introduction
2. Non-negative matrix factorization
3. Information geometry
   - Background
   - Proposed system for real-time mining of audio data streams
   - Results
   - Other applications
   - Discussion
4. Conclusion
What is IG?

**Statistical differentiable manifold.**

Under certain assumptions, a statistical model forms a differentiable manifold:

\[ S = \{ p_\xi = p(x; \xi) : \xi = [\xi^1, \ldots, \xi^n] \in \Xi \} \]
What is IG?

Statistical differentiable manifold.

Under certain assumptions, a statistical model forms a differentiable manifold: 
\[ S = \{ p_\xi = p(x; \xi) : \xi = [\xi^1, \ldots, \xi^n] \in \Xi \} \]

- Example: 
  \[ p(x; \xi) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(x - \mu)^2}{2\sigma^2} \right\} \] with \( \xi = [\mu, \sigma^2] \).
What is IG?

Statistical differentiable manifold.

Under certain assumptions, a statistical model forms a differentiable manifold: 
\[ S = \{ p_{\xi} = p(x; \xi) : \xi = [\xi^1, \ldots, \xi^n] \in \Xi \}. \]

- Example: 
  \[ p(x; \xi) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(x - \mu)^2}{2\sigma^2} \right\} \text{ with } \xi = [\mu, \sigma^2]. \]

Fisher information metric [Rao, 1945, Chentsov, 1982].

Under certain assumptions, the Fisher information matrix defines the unique Riemannian metric \( g \) on \( S \):
\[ g_{ij}(\xi) = \int \partial_i \log p(x; \xi) \cdot \partial_j \log p(x; \xi) \cdot p(x; \xi) \cdot dx. \]
What is IG?

Statistical differentiable manifold.

Under certain assumptions, a statistical model forms a differentiable manifold: 
\[ S = \{ p_\xi = p(x; \xi) : \xi = [\xi^1, \ldots, \xi^n] \in \Xi \} \].

- Example: 
  \[ p(x; \xi) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(x - \mu)^2}{2\sigma^2} \right\} \text{ with } \xi = [\mu, \sigma^2] \].

Fisher information metric [Rao, 1945, Chentsov, 1982].

Under certain assumptions, the Fisher information matrix defines the unique Riemannian metric \( g \) on \( S \): 
\[ g_{ij}(\xi) = \int \partial_i \log p(x; \xi) \cdot \partial_j \log p(x; \xi) \cdot p(x; \xi) \cdot dx \].


Under certain assumptions, there is a family of dual affine connections \( \{ \nabla^{(\alpha)}, \nabla^{(-\alpha)} \}_{\alpha \in \mathbb{R}} \) on \( (S, g) \) called \( \alpha \)-connections.
How to use IG from a computational viewpoint?

Exponential family.

\[ p_\theta(x) = \exp(\theta^T T(x) - F(\theta) + C(x)) \]

where \( F \) is a strictly convex function.

- \( \theta \): natural parameters.
- \( F(\theta) \): log-normalizer.
- \( C(x) \): carrier measure.
- \( T(x) \): sufficient statistic.
How to use IG from a computational viewpoint?

\[ p_{\theta}(x) = \exp \left( \theta^T T(x) - F(\theta) + C(x) \right) \]
where \( F \) is a strictly convex function.

**Figure:** A taxonomy of exponential families.

\[ B_{F}(\theta_1, \theta_2) = F(\theta_1) - F(\theta_2) - (\theta_1 - \theta_2)^T \nabla F(\theta_2) \]
where \( F \) is a strictly convex differentiable function.
### Exponential family.

\[ p_\theta(x) = \exp(\theta^T T(x) - F(\theta) + C(x)) \]  

where \( F \) is a strictly convex function.

### Bregman divergence.

\[ B_F(\theta_1, \theta_2) = F(\theta_1) - F(\theta_2) - (\theta_1 - \theta_2)^T \nabla F(\theta_2) \]  

where \( F \) is a strictly convex differentiable function.

- Links with dually flat spaces, Legendre transform and expectation parameters [Amari & Nagaoka, 2000, Banerjee et al., 2005]:

\[ D_{KL}(p \parallel q) = B_F(\theta_q \parallel \theta_p) = B_G(\eta_p \parallel \eta_q) \]
How to use IG from a computational viewpoint?

Exponential family.

\[ p_\theta(x) = \exp (\theta^T T(x) - F(\theta) + C(x)) \] where \( F \) is a strictly convex function.

Bregman divergence.

\[ B_F(\theta_1, \theta_2) = F(\theta_1) - F(\theta_2) - (\theta_1 - \theta_2)^T \nabla F(\theta_2) \] where \( F \) is a strictly convex differentiable function.

- Links with dually flat spaces, Legendre transform and expectation parameters [Amari & Nagaoka, 2000, Banerjee et al., 2005]:
  \[ D_{KL}(p \parallel q) = B_F(\theta_q \parallel \theta_p) = B_G(\eta_p \parallel \eta_q) \]

- Generic algorithms that handle many generalized distances (demo) [Banerjee et al., 2005, Cayton, 2008, Cayton, 2009, Nielsen & Nock, 2009, Nielsen et al., 2009, Garcia et al., 2009]:
How to use IG from a computational viewpoint?

**Exponential family.**

\[ p_\theta(x) = \exp(\theta^T T(x) - F(\theta) + C(x)) \]

where \( F \) is a strictly convex function.

**Bregman divergence.**

\[ B_F(\theta_1, \theta_2) = F(\theta_1) - F(\theta_2) - (\theta_1 - \theta_2)^T \nabla F(\theta_2) \]

where \( F \) is a strictly convex differentiable function.

- Links with dually flat spaces, Legendre transform and expectation parameters [Amari & Nagaoka, 2000, Banerjee et al., 2005]:
  \[ D_{KL}(p \parallel q) = B_F(\theta_q \parallel \theta_p) = B_G(\eta_p \parallel \eta_q) \]

- Generic algorithms that handle many generalized distances (demo) [Banerjee et al., 2005, Cayton, 2008, Cayton, 2009, Nielsen & Nock, 2009, Nielsen et al., 2009, Garcia et al., 2009]:
  - Centroids and hard clustering (\( k \)-means).
How to use IG from a computational viewpoint?

**Exponential family.**

\[ p_\theta(x) = \exp(\theta^T T(x) - F(\theta) + C(x)) \text{ where } F \text{ is a strictly convex function.} \]

**Bregman divergence.**

\[ B_F(\theta_1, \theta_2) = F(\theta_1) - F(\theta_2) - (\theta_1 - \theta_2)^T \nabla F(\theta_2) \text{ where } F \text{ is a strictly convex differentiable function.} \]

- Links with dually flat spaces, Legendre transform and expectation parameters [Amari & Nagaoka, 2000, Banerjee et al., 2005]:
  \[ D_{KL}(p \parallel q) = B_F(\theta_q \parallel \theta_p) = B_G(\eta_p \parallel \eta_q) \]

- Generic algorithms that handle many generalized distances (demo) [Banerjee et al., 2005, Cayton, 2008, Cayton, 2009, Nielsen & Nock, 2009, Nielsen et al., 2009, Garcia et al., 2009]:
  - Centroids and hard clustering (k-means).
  - Parameter estimation and soft clustering (expectation-maximization).
How to use IG from a computational viewpoint?

**Exponential family.**

\[ p_\theta(x) = \exp(\theta^T T(x) - F(\theta) + C(x)) \]  

where \( F \) is a strictly convex function.

**Bregman divergence.**

\[ B_F(\theta_1, \theta_2) = F(\theta_1) - F(\theta_2) - (\theta_1 - \theta_2)^T \nabla F(\theta_2) \]  

where \( F \) is a strictly convex differentiable function.

- Links with dually flat spaces, Legendre transform and expectation parameters [Amari & Nagaoka, 2000, Banerjee et al., 2005]:

  \[ \mathcal{D}_{KL}(p \parallel q) = B_F(\theta_q \parallel \theta_p) = B_G(\eta_p \parallel \eta_q) \]

- Generic algorithms that handle many generalized distances (demo) [Banerjee et al., 2005, Cayton, 2008, Cayton, 2009, Nielsen & Nock, 2009, Nielsen et al., 2009, Garcia et al., 2009]:
  - Centroids and hard clustering (\( k \)-means).
  - Parameter estimation and soft clustering (expectation-maximization).
  - Ball trees and search queries (nearest-neighbors search, range search).
How to design real-time systems for audio based on IG?

Scheme:
1. Represent the incoming audio stream with short-time sound descriptors \( d_j \).
2. Model these descriptors as probability distributions \( p_{\theta_j} \) from a given exponential family.
3. Use the framework of computational information geometry on these distributions.
How to design real-time systems for audio based on IG?

Scheme:
1. Represent the incoming audio stream with short-time sound descriptors \( d_j \).
2. Model these descriptors as probability distributions \( p_{\theta_j} \) from a given exponential family.
3. Use the framework of computational information geometry on these distributions.

In particular, it allows to define the notion of similarity in an information setup through divergences.
How to design real-time systems for audio based on IG?

Scheme:

1. Represent the incoming audio stream with short-time sound descriptors \(d_j\).
2. Model these descriptors as probability distributions \(p_{\theta j}\) from a given exponential family.
3. Use the framework of computational information geometry on these distributions.

In particular, it allows to define the notion of similarity in an information setup through divergences.

Potential applications:

- Audio content analysis.
- Segmentation of audio streams.
- Automatic structure discovery of audio signals.
- Sound processing and synthesis.
General architecture

Figure: Schema of the general architecture of the system.
Sound descriptors modeling

- Computation of a sound descriptor $d_j$:
  - Fourier or constant-Q transforms for information on the spectral content.
  - Mel-frequency cepstral coefficients for information on the timbre.
  - Many other possibilities.

Figure: Sound descriptors modeling.
Sound descriptors modeling

- Computation of a sound descriptor $d_j$:
  - Fourier or constant-Q transforms for information on the spectral content.
  - Mel-frequency cepstral coefficients for information on the timbre.
  - Many other possibilities.

- Modeling with a probability distribution $p_{\theta_j}$ from an exponential family:
  - Categorical distributions.
  - Many other possibilities.

Figure: Sound descriptors modeling.
Temporal information modeling

- Model formation: from signal to symbol.
  - Assumption of quasi-stationary audio chunks.
  - Change detection adapted from CuSum [Basseville & Nikiforov, 1993].

Figure: Model formation at time $t$. 
Temporal information modeling

- Model formation: from signal to symbol.
  - Assumption of quasi-stationary audio chunks.
  - Change detection adapted from CuSum [Basseville & Nikiforov, 1993].

![Figure: Model formation at time \( t \).](image1)

- Factor oracle: from symbol to syntax (and from genetics to music!).
  - Forward transitions: original sequence factors.
  - Backward links: suffix relations, common context.

![Figure: Factor oracle of the word abbbaab.](image2)
Audio segmentation

**Figure:** Segmentation of the *1st Piano Sonate, 1st Movement, 1st Theme*, Beethoven.
Music similarity analysis

Figure: Similarity analysis of the 1st Piano Sonate, 3rd Movement, Beethoven.
Musical structure discovery

Figure: Structure discovery of the 1st Piano Sonate, 3rd Movement, Beethoven.
Figure: Query by similarity of the 1st Theme over the entire 1st Piano Sonate, 1st Movement, Beethoven.
Figure: Audio recombination of African drums by concatenative synthesis of congas.
Computer-assisted improvisation

Figure: Computer-assisted improvisation, Fabrizio Cassol and Philippe Leclerc.
Summary and perspectives.

- Representations.
- Descriptors modeling.
- Temporal modeling.
- Temporality of events.
What we (don’t) have

Summary and perspectives.

- Representations.
- Descriptors modeling.
- Temporal modeling.
- Temporality of events.

- Many possibilities.
- Combinations of descriptors.
- Complex representations.
What we (don’t) have

Summary and perspectives.

- Representations.
- Descriptors modeling.
- Temporal modeling.
- Temporality of events.

- Exponential families and Bregman divergences.
- Mixture models of a given exponential family.
- Other geometries, divergences, metrics.
What we (don’t) have

Summary and perspectives.

- Representations.
- Descriptors modeling.
- Temporal modeling.
- Temporality of events.

- On-line segmentation and factor oracle.
- On-line clustering and equivalence between symbols.
- Overlap between symbols and other temporal models.
What we (don’t) have

Summary and perspectives.
- Representations.
- Descriptors modeling.
- Temporal modeling.
- Temporality of events.
- Assumption of quasi-stationarity.
- Non-stationarity modeling.
- Time series.
Outline

1. Introduction
2. Non-negative matrix factorization
3. Information geometry
4. Conclusion
The story so far

Motivations:

- Fill in the gap between signal and symbolic representations.
- Devise computational tools for complex real-time settings.
The story so far

- **Motivations:**
  - Fill in the gap between signal and symbolic representations.
  - Devise computational tools for complex real-time settings.

- **Resources on NMF:** [http://imtr.ircam.fr/imtr/Realtime_Transcription](http://imtr.ircam.fr/imtr/Realtime_Transcription)
The story so far

- **Motivations:**
  - Fill in the gap between signal and symbolic representations.
  - Devise computational tools for complex real-time settings.

- **Resources on NMF:** [http://imtr.ircam.fr/imtr/Realtime_Transcription](http://imtr.ircam.fr/imtr/Realtime_Transcription)

- **Resources on IG:** [http://imtr.ircam.fr/imtr/Music_Information_Geometry](http://imtr.ircam.fr/imtr/Music_Information_Geometry)

- National research group: IRCAM, Ecole Polytechnique, Thales, etc.
- Brillouin seminar: [http://www.informationgeometry.org/Seminar/seminarBrillouin.html](http://www.informationgeometry.org/Seminar/seminarBrillouin.html)
- IGAIA 2012.
The story so far

- **Motivations:**
  - Fill in the gap between signal and symbolic representations.
  - Devise computational tools for complex real-time settings.

- **Resources on NMF:** [http://imtr.ircam.fr/imtr/Realtime_Transcription](http://imtr.ircam.fr/imtr/Realtime_Transcription)

- **Resources on IG:** [http://imtr.ircam.fr/imtr/Music_Information_Geometry](http://imtr.ircam.fr/imtr/Music_Information_Geometry)

- **National research group:** IRCAM, Ecole Polytechnique, Thales, etc.
- **Brillouin seminar:** [http://www.informationgeometry.org/Seminar/seminarBrillouin.html](http://www.informationgeometry.org/Seminar/seminarBrillouin.html)
- **IGAIA 2012.**

Thank you very much for your attention! Questions?

arnaud.dessein@ircam.fr

October 22nd 2010
Polyphonic music transcription by non-negative sparse coding of power spectra.
In Proc. of the 5th Int. Conf. on Music Information Retrieval (pp. 318–325). Barcelona, Spain.

Algorithms, initializations, and convergence for the non negative matrix factorization.
Technical report, NC State University.

Methods of information geometry, volume 191 of Translations of Mathematical Monographs.
American Mathematical Society.

Clustering with Bregman divergences.

Detection of abrupt changes: Theory and application.

Algorithms and applications for approximate nonnegative matrix factorization.
Computational Statistics & Data Analysis, 52(1), 155–173.

Enforcing harmonicity and smoothness in Bayesian non-negative matrix factorization applied to polyphonic music transcription.
IEEE Transactions on Audio, Speech and Language Processing, 18(3), 538–549.
Bibliography II

A tempering approach for Itakura-Saito non-negative matrix factorization. with application to music transcription.
In Proc. of the IEEE 34th Int. Conf. on Acoustics, Speech and Signal Processing (pp. 1545–1548). Taipei, Taiwan.

Fast nearest neighbor retrieval for Bregman divergences.

Efficient Bregman range search.

Nonnegative matrix factorization for real time musical analysis and sight-reading evaluation.
In Proc. of the IEEE 33rd Int. Conf. on Acoustics, Speech and Signal Processing (pp. 2017–2020). Las Vegas, NV, USA.

Chentsov, N. N. (1982).


Realtime multiple pitch observation using sparse non-negative constraints.
In Proc. of the 7th Int. Conf. on Music Information Retrieval Victoria, Canada.
Realtime multiple-pitch and multiple-instrument recognition for music signals using sparse non-negative constraints.
In Proc. of the 10th Int. Conf. on Digital Audio Effects Bordeaux, France.

Sparse coding and NMF.
In Proc. of the IEEE Int. Joint Conf. on Neural Networks (pp. 2529–2533). Budapest, Hungary.

*Robustifying maximum likelihood estimation.*
Technical report, Institute of Statistical Mathematics, Tokyo, Japan.

Nonnegative matrix factorization with the Itakura-Saito divergence: with application to music analysis.

Levels of details for Gaussian mixture models.

Learning non-negative sparse image codes by convex programming.
In Proc. of the IEEE 10th Int. Conf. on Computer Vision Beijing, China.

Learning sparse representations by non-negative matrix factorization and sequential cone programming.
Hoyer, P. O. (2002).
Non-negative sparse coding.
In Proc. of the IEEE 12th Workshop on Neural Networks for Signal Processing (pp. 557–565). Martigny, Switzerland.

Non-negative matrix factorization with sparseness constraints.

Measuring sparseness of noisy signals.

A generalized divergence measure for nonnegative matrix factorization.

Learning the parts of objects by non-negative matrix factorization.

Mysore, G. J. (2010).
PhD thesis, Stanford University, Palo Alto, CA, USA.

Non-negative matrix division for the automatic transcription of polyphonic music.
In Proc. of the 9th Int. Conf. on Music Information Retrieval (pp. 544–549). Philadelphia, PA, USA.
Bibliography V

Sided and symmetrized Bregman centroids.

Tailored Bregman ball trees for effective nearest neighbors.
In *Proceedings of the 25th European Workshop on Computational Geometry (EuroCG)* (pp. 29–32). Brussels, Belgium.

Discovering speech phones using convolutive non-negative matrix factorisation with a sparseness constraint.
*Neurocomputing*, 72, 88–101.

Harmonic nonnegative matrix approximation for multipitch analysis of musical sounds.
In *Proc. of ASJ Autumn Meeting* (pp. 827–830).

Information and accuracy attainable in the estimation of statistical parameters.

Real-time pitch determination of one or more voices by nonnegative matrix factorization.
Bibliography VI

Non-negative matrix factor deconvolution; extraction of multiple sound sources from monophonic inputs.

Non-negative matrix factorization for polyphonic music transcription.
In Proc. of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (pp. 177–180). New Paltz, NY, USA.

Adaptive harmonic spectral decomposition for multiple pitch estimation.
IEEE Transactions on Audio, Speech and Language Processing, 18(3).

Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria.

Analysis of polyphonic audio using source-filter model and non-negative matrix factorization.

Nonnegative matrix factorization with quadratic programming.
Neurocomputing, 71(10–12), 2309–2320.