Music Information Geometry

Arnaud Dessein\textsuperscript{1,2} and Arshia Cont\textsuperscript{1}

\textsuperscript{1}Institute for Research and Coordination of Acoustics and Music, Paris, France
\textsuperscript{2}Japanese-French Laboratory for Informatics, Tokyo, Japan

IRMA, Strasbourg, April 7th 2011
Outline

1. Introduction
2. Information geometry
3. Proposed framework
4. Obtained results
5. Conclusion
1. Introduction
   - A bit of history about science and music
   - Motivations towards information geometry

2. Information geometry

3. Proposed framework

4. Obtained results

5. Conclusion
Where do we come from?

- Pythagoras (∼ 570–495 BC): relation between string length and produced sound, Pythagorean tuning.

  *There is geometry in the humming of the strings, there is music in the spacing of the spheres.*
Where do we come from?

- Pythagoras (~ 570–495 BC): relation between string length and produced sound, Pythagorean tuning.
  
  *There is geometry in the humming of the strings, there is music in the spacing of the spheres.*

- Helmholtz (1821–1894): Helmholtz resonator, harmonics and frequency spectrum of sounds.
Where do we come from?

- Pythagoras (~ 570–495 BC): relation between string length and produced sound, Pythagorean tuning.
  
  *There is geometry in the humming of the strings, there is music in the spacing of the spheres.*

- Helmholtz (1821–1894): Helmholtz resonator, harmonics and frequency spectrum of sounds.

- But also indirectly Fourier, Shannon, etc.
What do we need?

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

- Develop a comprehensive framework that allows to quantify, process and represent the information contained in audio signals.

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

- Develop a comprehensive framework that allows to quantify, process and represent the information contained in audio signals.
- Fill in the gap between signal and symbolic representations.

**Figure:** Levels of representation of audio, waveform and spectrogram representations.
What is information geometry?

**Statistical differentiable manifold.**

Under certain assumptions, a parametric statistical model $S = \{ p_\xi : \xi \in \Xi \}$ of probability distributions defined on $\mathcal{X}$ forms a differentiable manifold.
What is information geometry?

Statistical differentiable manifold.

Under certain assumptions, a parametric statistical model $S = \{p_\xi: \xi \in \Xi\}$ of probability distributions defined on $\mathcal{X}$ forms a differentiable manifold.

- Example: $p_\xi(x) = \frac{1}{\sqrt{2\pi}\sigma^2} \exp \left\{ -\frac{(x - \mu)^2}{2\sigma^2} \right\}$ for all $x \in \mathcal{X} = \mathbb{R}$, with $\xi = [\mu, \sigma^2] \in \Xi = \mathbb{R} \times \mathbb{R}_{++}$. 
What is information geometry?

Statistical differentiable manifold.
Under certain assumptions, a parametric statistical model $\mathcal{S} = \{p_\xi : \xi \in \Xi\}$ of probability distributions defined on $\mathcal{X}$ forms a differentiable manifold.

- Example: $p_\xi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(x - \mu)^2}{2\sigma^2} \right\}$ for all $x \in \mathcal{X} = \mathbb{R}$, with $\xi = [\mu, \sigma^2] \in \Xi = \mathbb{R} \times \mathbb{R}^+$.  

Fisher information metric [Rao, 1945, Chentsov, 1982].
Under certain assumptions, the Fisher information matrix defines the unique Riemannian metric $g$ on $\mathcal{S}$: $g_{ij}(\xi) = \int_{x \in \mathcal{X}} \partial_i \log p_\xi(x) \cdot \partial_j \log p_\xi(x) \cdot p_\xi(x) \cdot dx$. 

What is information geometry?

Statistical differentiable manifold.

Under certain assumptions, a parametric statistical model $\mathcal{S} = \{ p_\xi : \xi \in \Xi \}$ of probability distributions defined on $\mathcal{X}$ forms a differentiable manifold.

- Example: $p_\xi(x) = \frac{1}{\sqrt{2\pi}\sigma^2} \exp \left\{ -\frac{(x - \mu)^2}{2\sigma^2} \right\}$ for all $x \in \mathcal{X} = \mathbb{R}$, with $\xi = [\mu, \sigma^2] \in \Xi = \mathbb{R} \times \mathbb{R}^+$.  

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

Under certain assumptions, the Fisher information matrix defines the unique Riemannian metric $g$ on $\mathcal{S}$: $g_{ij}(\xi) = \int_{x \in \mathcal{X}} \partial_i \log p_\xi(x) \cdot \partial_j \log p_\xi(x) \cdot p_\xi(x) \cdot dx$.


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

**Exponential family.**

\[ p_\theta(x) = \exp \left( \theta^T T(x) - F(\theta) + C(x) \right) \text{ for all } x \in \mathcal{X}. \]

- \( \theta \): natural parameters, vector belonging to a convex open set \( \Theta \).
- \( F \): log-normalizer, real-valued, strictly convex smooth function on \( \Theta \).
- \( C \): carrier measure, real-valued function on \( \mathcal{X} \).
- \( T \): sufficient statistic, vector-valued function on \( \mathcal{X} \).
**Exponential family.**

\[ p_\theta(x) = \exp \left( \theta^T T(x) - F(\theta) + C(x) \right) \text{ for all } x \in \mathcal{X}. \]

**Figure:** A taxonomy of exponential families.
How to use information geometry from a computational viewpoint?

**Exponential family.**

\[ p_\theta(x) = \exp(\theta^T T(x) - F(\theta) + C(x)) \text{ for all } x \in \mathcal{X}. \]

- We consider a statistical manifold \( S = \{p_\theta : \theta \in \Theta\} \) equipped with \( g \) and the dual exponential and mixture connections \( \nabla^{(1)} \) and \( \nabla^{(-1)} \).
How to use information geometry from a computational viewpoint?

**Exponential family.**

\[ p_\theta(x) = \exp(\theta^T T(x) - F(\theta) + C(x)) \text{ for all } x \in \mathcal{X}. \]

- We consider a statistical manifold \( \mathcal{S} = \{ p_\theta : \theta \in \Theta \} \) equipped with \( g \) and the dual exponential and mixture connections \( \nabla^{(1)} \) and \( \nabla^{(-1)} \).
- \( (\mathcal{S}, g, \nabla^{(1)}, \nabla^{(-1)}) \) possesses two dual affine coordinate systems, natural parameters \( \theta \) and expectation parameters \( \eta = \nabla F(\theta) \).
How to use information geometry from a computational viewpoint?

**Exponential family.**

\[
p_\theta(x) = \exp(\theta^T T(x) - F(\theta) + C(x)) \text{ for all } x \in \mathcal{X}.
\]

- We consider a statistical manifold \( S = \{p_\theta : \theta \in \Theta\} \) equipped with \( g \) and the dual exponential and mixture connections \( \nabla^{(1)} \) and \( \nabla^{(-1)} \).
- \((S, g, \nabla^{(1)}, \nabla^{(-1)})\) possesses two dual affine coordinate systems, natural parameters \( \theta \) and expectation parameters \( \eta = \nabla F(\theta) \).
- Dually flat geometry, Hessian structure \((g = \nabla^2 F)\), generated by the potential \( F \) together with its conjugate potential \( F^* \) defined by the Legendre-Fenchel transform: \( F^*(\eta) = \sup_{\theta \in \Theta} \theta^T \eta - F(\theta) \), which verifies \( \nabla F^* = (\nabla F)^{-1} \) so that \( \theta = \nabla F^*(\eta) \).
Exponential family.

$p_\theta(x) = \exp(\theta^T T(x) - F(\theta) + C(x))$ for all $x \in \mathcal{X}$.

- We consider a statistical manifold $S = \{p_\theta : \theta \in \Theta\}$ equipped with $g$ and the dual exponential and mixture connections $\nabla^{(1)}$ and $\nabla^{(-1)}$.
- $(S, g, \nabla^{(1)}, \nabla^{(-1)})$ possesses two dual affine coordinate systems, natural parameters $\theta$ and expectation parameters $\eta = \nabla F(\theta)$.
- Dually flat geometry, Hessian structure $(g = \nabla^2 F)$, generated by the potential $F$ together with its conjugate potential $F^*$ defined by the Legendre-Fenchel transform: $F^*(\eta) = \sup_{\theta \in \Theta} \theta^T \eta - F(\theta)$, which verifies $\nabla F^* = (\nabla F)^{-1}$ so that $\theta = \nabla F^*(\eta)$.
- Generalizes the self-dual Euclidean geometry, with notably two canonically associated Bregman divergences $B_F$ and $B_{F^*}$ instead of the self-dual Euclidean distance, but also dual geodesics, a generalized Pythagorean theorem and dual projections.
How to use information geometry from a computational viewpoint?

**Exponential family.**

\[ p_\theta(x) = \exp(\theta^T T(x) - F(\theta) + C(x)) \text{ for all } x \in \mathcal{X}. \]

**Bregman divergence.**

\[ \mathcal{B}_F(\theta, \theta') = F(\theta) - F(\theta') - (\theta - \theta')^T \nabla F(\theta'). \]
### Information geometry

#### Proposed framework

#### Obtained results

#### Conclusion

---

How to use information geometry from a computational viewpoint?

---

#### Exponential family.

$$p_\theta(x) = \exp \left( \theta^T T(x) - F(\theta) + C(x) \right) \text{ for all } x \in \mathcal{X}.$$  

#### Bregman divergence.

$$B_F(\theta, \theta') = F(\theta) - F(\theta') - (\theta - \theta')^T \nabla F(\theta').$$

- Canonical divergences of dually flat spaces, “bijection” with exponential families [Amari & Nagaoka, 2000, Banerjee et al., 2005]:
  $$D_{KL}(p_\xi \parallel p_{\xi'}) = B_F(\theta' \parallel \theta) = B_{F^*}(\eta \parallel \eta').$$
How to use information geometry from a computational viewpoint?

### Exponential family.

\[
p_\theta(x) = \exp(\theta^T T(x) - F(\theta) + C(x)) \quad \text{for all } x \in \mathcal{X}.
\]

### Bregman divergence.

\[
\mathcal{B}_F(\theta, \theta') = F(\theta) - F(\theta') - (\theta - \theta')^T \nabla F(\theta').
\]

- Canonical divergences of dually flat spaces, “bijection” with exponential families [Amari & Nagaoka, 2000, Banerjee et al., 2005]:
  \[
  D_{KL}(p_\xi \parallel p_{\xi'}) = \mathcal{B}_F(\theta' \parallel \theta) = \mathcal{B}_{F^*}(\eta \parallel \eta').
  \]
- No symmetry nor triangular inequality in general, but an information-theoretic interpretation.
How to use information geometry from a computational viewpoint?

**Exponential family.**

\[ p_\theta(x) = \exp(\theta^T T(x) - F(\theta) + C(x)) \text{ for all } x \in \mathcal{X}. \]

**Bregman divergence.**

\[ \mathcal{B}_F(\theta, \theta') = F(\theta) - F(\theta') - (\theta - \theta')^T \nabla F(\theta'). \]

- Canonical divergences of dually flat spaces, “bijection” with exponential families [Amari & Nagaoka, 2000, Banerjee et al., 2005]:
  \[ D_{KL}(p_\xi \parallel p_{\xi'}) = \mathcal{B}_F(\theta' \parallel \theta) = \mathcal{B}_{F^*}(\eta \parallel \eta'). \]

- No symmetry nor triangular inequality in general, but an information-theoretic interpretation.

  - Centroid computation and hard clustering (k-means).
  - Parameter estimation and soft clustering (expectation-maximization).
  - Proximity queries in ball trees (nearest-neighbors and range search).
Outline

1. Introduction
2. Information geometry
3. Proposed framework
   - General architecture
   - Sound descriptors modeling
   - Temporal modeling
4. Obtained results
5. Conclusion
How to design an audio system based on information geometry?

**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.

*Figure:* Schema of the general architecture of the system.
How to design an audio system based on information geometry?

- **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.**

**Figure:** Schema of the general architecture of the system.
How to design an audio system based on information geometry?

**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.
- Important need for temporal modeling.

*Figure:* Schema of the general architecture of the system.
How to design an audio system based on information geometry?

- 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.

- Important need for temporal modeling.

- Potential applications [Cont et al., 2011]:
  - Audio content analysis.
  - Segmentation of audio streams.
  - Automatic structure discovery of audio signals.
  - Sound processing and synthesis.

**Figure:** Schema of the general architecture of the system.
How to model sounds?

- 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.

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.
How to model sounds?

- 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.
How to take time into account?

- 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 \( t \).](image)
How to take time into account?

- **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.](image)

- **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 abbbbaab.](image)
Outline

1 Introduction
2 Information geometry
3 Proposed framework
4 Obtained results
   - Audio segmentation
   - Music similarity analysis
   - Musical structure discovery
   - Query by similarity
   - Audio recombination by concatenative synthesis
   - Computer-assisted improvisation
5 Conclusion
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.
Audio recombination by concatenative synthesis

Figure: Audio recombination of African drums by concatenative synthesis of congas.
Computer-assisted improvisation

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

1 Introduction
2 Information geometry
3 Proposed framework
4 Obtained results
5 Conclusion
What we (don’t) have

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.
Summary and perspectives.

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

- Resources on IG: 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
- IGAIA 2012.
What we (don’t) have

Summary and perspectives.

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

- Resources on IG: 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
- IGAIA 2012.
- Thank you very much for your attention! Questions?
- This work was supported by a doctoral fellowship from the UPMC (EDITE) and by a grant from the JST-CNRS ICT (Improving the VR Experience).

Clustering with Bregman divergences.


Fast nearest neighbor retrieval for Bregman divergences.

Efficient Bregman range search.

Chentsov, N. N. (1982).
On the information geometry of audio streams with applications to similarity computing.
IEEE Transactions on Audio, Speech and Language Processing, 19.
To appear.

Levels of details for Gaussian mixture models.

Sided and symmetrized Bregman centroids.

Tailored Bregman ball trees for effective nearest neighbors.

Information and accuracy attainable in the estimation of statistical parameters.