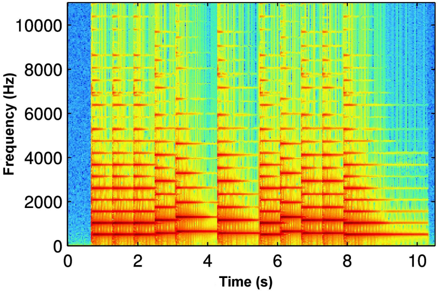
High resolution spectral analysis and nonnegative decompositions applied to music signal processing

Roland Badeau
Associate Professor, Telecom ParisTech
Visiting researcher at C4DM
roland.badeau@eecs.qmul.ac.uk

Music representations

Musical score

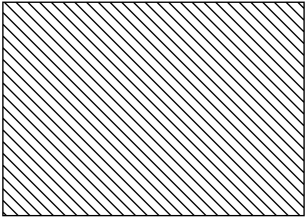
Music representations



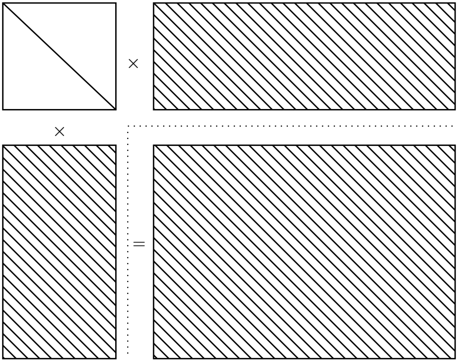
Spectrogram of "Au clair de la lune"

Musical score

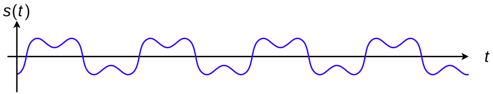
Low-rank approximations



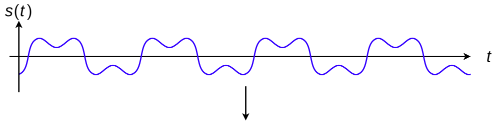
Low-rank approximations

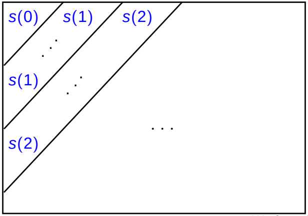


High Resolution (HR) spectral analysis

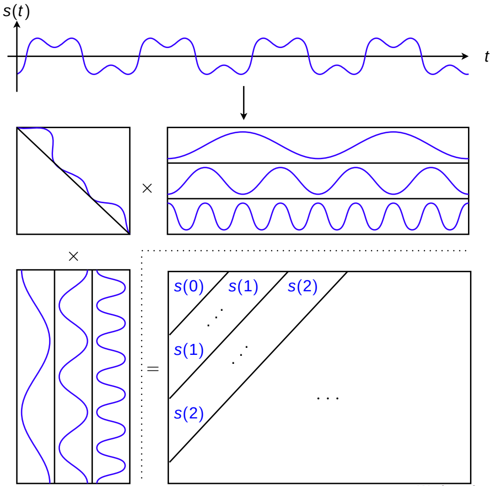


High Resolution (HR) spectral analysis



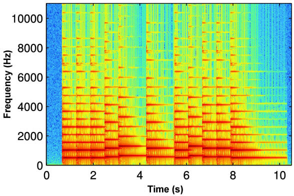


High Resolution (HR) spectral analysis



Nonnegative Matrix Factorization (NMF)

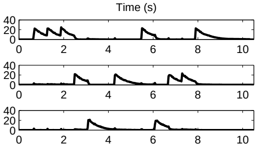
Musical score



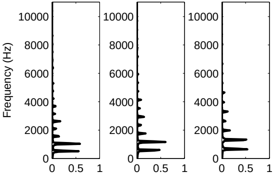
Spectrogram

Nonnegative Matrix Factorization (NMF)

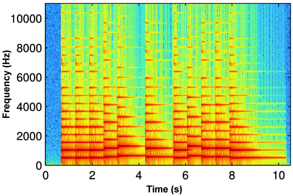
Musical score



Temporal activations



Spectral atoms



Spectrogram

Applications

- Applications of High Resolution methods
  - Spectral analysis (modal analysis, spectroscopy)
  - Array processing (beamforming, direction of arrival (DOA) estimation)
  - Digital communications (channel identification)
- Applications of NMF
  - Image analysis (face recognition)
  - Text mining, spectroscopy, finance, etc.
- Applications to audio signal processing
  - Source separation, audio coding
  - Pitch and tempo estimation, automatic transcription

Applications

- Applications of High Resolution methods
  - Spectral analysis (modal analysis, spectroscopy)
  - Array processing (beamforming, direction of arrival (DOA) estimation)
  - Digital communications (channel identification)
- Applications of NMF
  - Image analysis (face recognition)
  - Text mining, spectroscopy, finance, etc.
- Applications to audio signal processing
  - Source separation, audio coding
  - Pitch and tempo estimation, automatic transcription

Applications

- Applications of High Resolution methods
  - Spectral analysis (modal analysis, spectroscopy)
  - Array processing (beamforming, direction of arrival (DOA) estimation)
  - Digital communications (channel identification)
- Applications of NMF
  - Image analysis (face recognition)
  - Text mining, spectroscopy, finance, etc.
- Applications to audio signal processing
  - Source separation, audio coding
  - Pitch and tempo estimation, automatic transcription

Part I

High Resolution spectral analysis

Exponential Sinusoidal Model (ESM)

- Real-valued model: $s(t) = \sum_{k=1}^{r} a_k\, e^{-\delta_k t} \cos(2\pi\nu_k t + \phi_k)$
  - $a_k \in \mathbb{R}_+^*$ and $\phi_k \in]-\pi, \pi]$ are the **amplitude** and **phase**
  - $\delta_k \in \mathbb{R}$ and $\nu_k \in]-\frac{1}{2}, \frac{1}{2}]$ are the **damping factor** and **frequency**
- Complex-valued model: $s(t) = \sum_{k=1}^{r} \alpha_k\, {z_k}^t$
  - $\alpha_k = a_k\, e^{i\phi_k} \in \mathbb{C}^*$ is a **complex amplitude**
  - $z_k = e^{-\delta_k + i2\pi\nu_k} \in \mathbb{C}^*$ is a **complex pole**
- Noisy model: $x(t) = s(t) + b(t)$ ($b(t)$ is a white Gaussian noise)
- Model estimation
  - Data vector: $\mathbf{s}(t) = [s(t), \ldots, s(t+n-1)]^T$ with $n > r$
  - Fourier analysis: spectral resolution of the order of $\frac{1}{n}$
  - Subspace analysis: **high spectral resolution**

Exponential Sinusoidal Model (ESM)

- Real-valued model: $s(t) = \sum_{k=1}^{r} a_k\, e^{-\delta_k t} \cos(2\pi\nu_k t + \phi_k)$
  - $a_k \in \mathbb{R}_+^*$ and $\phi_k \in]-\pi, \pi]$ are the **amplitude** and **phase**
  - $\delta_k \in \mathbb{R}$ and $\nu_k \in]-\frac{1}{2}, \frac{1}{2}]$ are the **damping factor** and **frequency**
- Complex-valued model: $s(t) = \sum_{k=1}^{r} \alpha_k\, {z_k}^t$
  - $\alpha_k = a_k\, e^{i\phi_k} \in \mathbb{C}^*$ is a **complex amplitude**
  - $z_k = e^{-\delta_k + i2\pi\nu_k} \in \mathbb{C}^*$ is a **complex pole**
- Noisy model: $x(t) = s(t) + b(t)$ ($b(t)$ is a white Gaussian noise)
- Model estimation
  - Data vector: $\mathbf{s}(t) = [s(t), \ldots, s(t+n-1)]^T$ with $n > r$
  - Fourier analysis: spectral resolution of the order of $\frac{1}{n}$
  - Subspace analysis: **high spectral resolution**

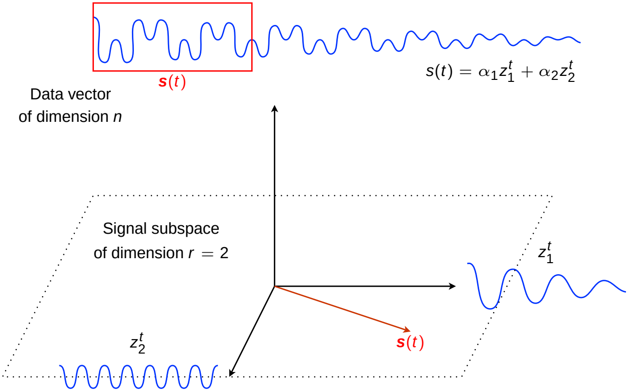
Exponential Sinusoidal Model (ESM)

- Real-valued model: $s(t) = \sum_{k=1}^{r} a_k\, e^{-\delta_k t} \cos(2\pi\nu_k t + \phi_k)$
  - $a_k \in \mathbb{R}_+^*$ and $\phi_k \in]-\pi, \pi]$ are the **amplitude** and **phase**
  - $\delta_k \in \mathbb{R}$ and $\nu_k \in]-\frac{1}{2}, \frac{1}{2}]$ are the **damping factor** and **frequency**
- Complex-valued model: $s(t) = \sum_{k=1}^{r} \alpha_k\, {z_k}^t$
  - $\alpha_k = a_k\, e^{i\phi_k} \in \mathbb{C}^*$ is a **complex amplitude**
  - $z_k = e^{-\delta_k + i2\pi\nu_k} \in \mathbb{C}^*$ is a **complex pole**
- Noisy model: $x(t) = s(t) + b(t)$ ($b(t)$ is a white Gaussian noise)
- Model estimation
  - Data vector: $\mathbf{s}(t) = [s(t), \ldots, s(t+n-1)]^T$ with $n > r$
  - Fourier analysis: spectral resolution of the order of $\frac{1}{n}$
  - Subspace analysis: **high spectral resolution**

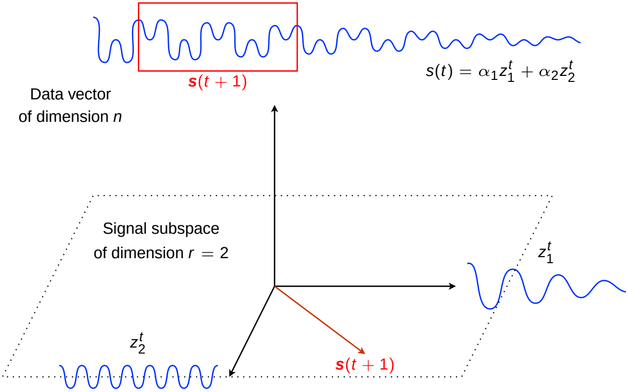
Exponential Sinusoidal Model (ESM)

- Real-valued model: $s(t) = \sum_{k=1}^{r} a_k \, e^{-\delta_k t} \, \cos(2\pi\nu_k t + \phi_k)$
  - $a_k \in \mathbb{R}_+^*$ and $\phi_k \in]-\pi, \pi]$ are the **amplitude** and **phase**
  - $\delta_k \in \mathbb{R}$ and $\nu_k \in]-\frac{1}{2}, \frac{1}{2}]$ are the **damping factor** and **frequency**
- Complex-valued model: $s(t) = \sum_{k=1}^{r} \alpha_k \, {z_k}^t$
  - $\alpha_k = a_k \, e^{i\phi_k} \in \mathbb{C}^*$ is a **complex amplitude**
  - $z_k = e^{-\delta_k + i2\pi\nu_k} \in \mathbb{C}^*$ is a **complex pole**
- Noisy model: $x(t) = s(t) + b(t)$ ($b(t)$ is a white Gaussian noise)
- Model estimation
  - Data vector: $\boldsymbol{s}(t) = [s(t), \ldots, s(t+n-1)]^T$ with $n > r$
  - Fourier analysis: spectral resolution of the order of $\frac{1}{n}$
  - Subspace analysis: **high spectral resolution**

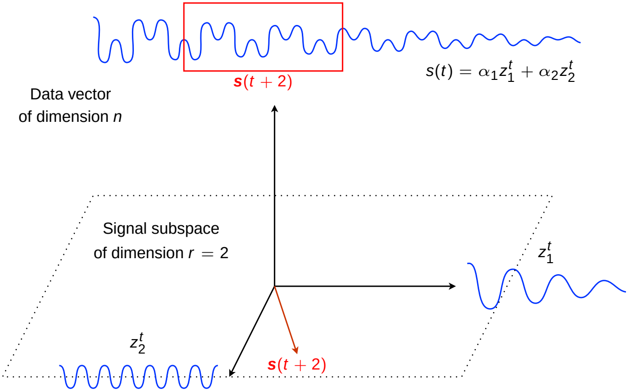
Subspace analysis



Subspace analysis



Subspace analysis



Model estimation

- Choose a window $(\gamma_\tau)_{\tau\in\mathbb{N}}$ (exponential, rectangular, hybrid)
- Compute a "correlation" matrix

$$\boldsymbol{C}_{xx}(t) = \sum_{\tau=0}^{t} \gamma_\tau \, \boldsymbol{x}(t-\tau)\,\boldsymbol{x}(t-\tau)^H$$

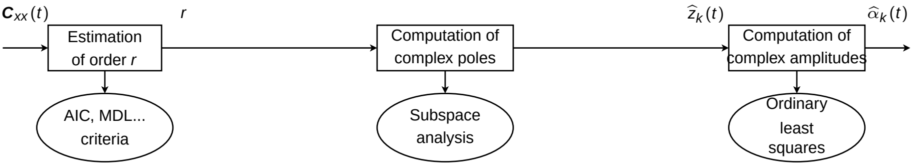
- Estimate the model parameters

Model estimation

- Choose a window $(\gamma_\tau)_{\tau\in\mathbb{N}}$ (exponential, rectangular, hybrid)
- Compute a "correlation" matrix

$$\boldsymbol{C}_{xx}(t) = \sum_{\tau=0}^{t} \gamma_\tau \, \boldsymbol{x}(t-\tau)\, \boldsymbol{x}(t-\tau)^H$$

- Estimate the model parameters

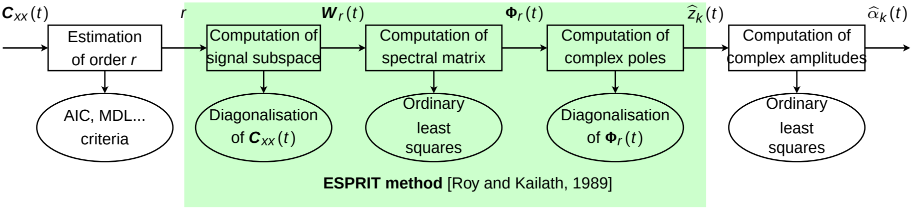


Model estimation

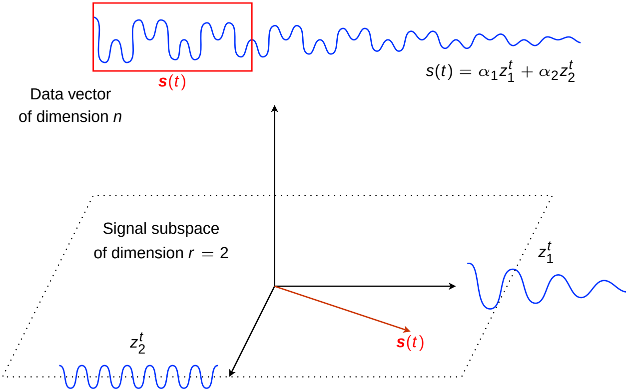
- Choose a window $(\gamma_\tau)_{\tau\in\mathbb{N}}$ (exponential, rectangular, hybrid)
- Compute a "correlation" matrix

$$\boldsymbol{C}_{xx}(t) = \sum_{\tau=0}^{t} \gamma_\tau \, \boldsymbol{x}(t-\tau)\,\boldsymbol{x}(t-\tau)^H$$

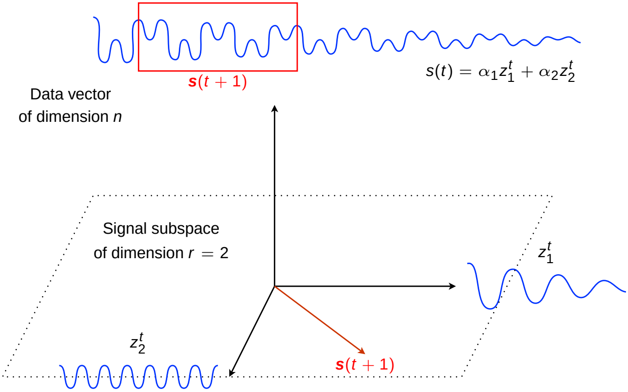
- Estimate the model parameters



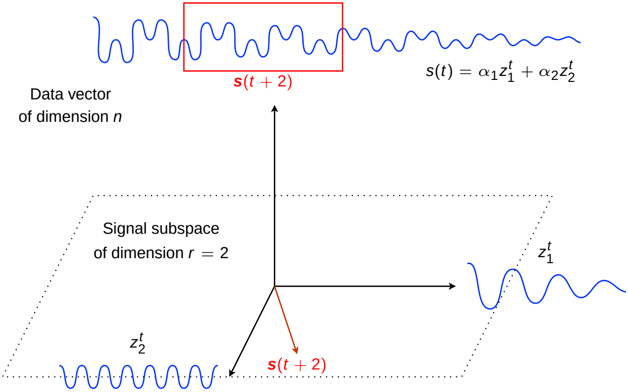
Subspace analysis



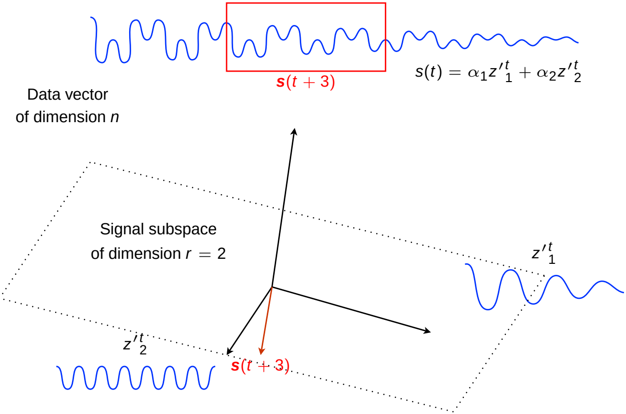
Subspace analysis



Subspace analysis



Subspace analysis

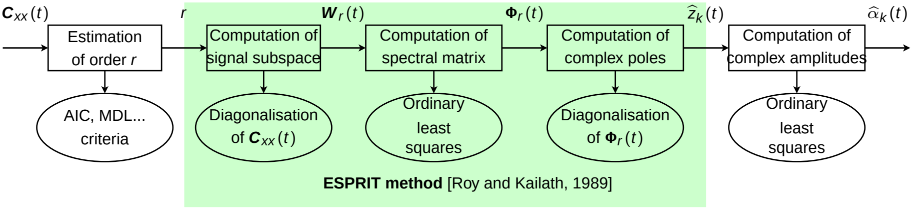


Model estimation

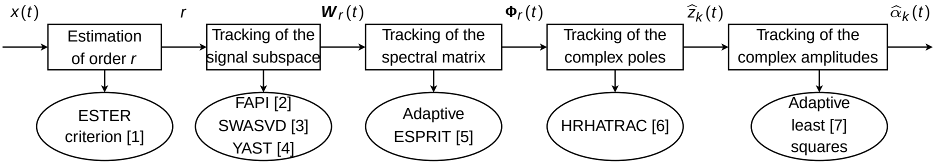
- Choose a window $(\gamma_\tau)_{\tau \in \mathbb{N}}$ (exponential, rectangular, hybrid)
- Compute a "correlation" matrix

$$\boldsymbol{C}_{xx}(t) = \sum_{\tau=0}^{t} \gamma_\tau \, \boldsymbol{x}(t-\tau)\, \boldsymbol{x}(t-\tau)^H$$

- Estimate the model parameters



Time-frequency analysis



[1] Roland Badeau, Bertrand David, and Gaël Richard. "A new perturbation analysis for signal enumeration in rotational invariance techniques". *IEEE Transactions on Signal Processing*, 54(2): 450–458, February 2006.

[2] Roland Badeau, Bertrand David, and Gaël Richard. "Fast Approximated Power Iteration Subspace Tracking". *IEEE Transactions on Signal Processing*, 53(8): 2931-2941, August 2005.

[3] Roland Badeau, Gaël Richard, and Bertrand David. "Sliding window adaptive SVD algorithms". *IEEE Transactions on Signal Processing*, 52(1): 1-10, January 2004.

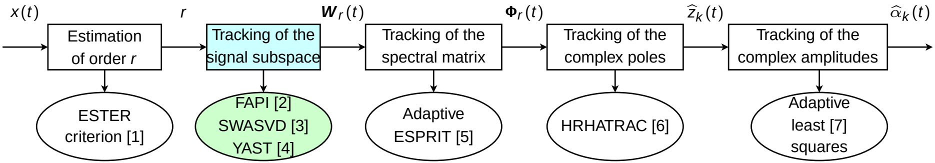
[4] Roland Badeau, Gaël Richard, and Bertrand David. "Fast and stable YAST algorithm for principal and minor subspace tracking". *IEEE Transactions on Signal Processing*, 56(8): 3437-3446, August 2008.

[5] Roland Badeau, Gaël Richard, and Bertrand David. "Fast adaptive ESPRIT algorithm". In *Proc. of IEEE Workshop on Statistical Signal Processing (SSP)*, Bordeaux, France, July 2005.

[6] Bertrand David, Roland Badeau, and Gaël Richard. "HRHATRAC Algorithm for Spectral Line Tracking of Musical Signals". In *Proc. of IEEE ICASSP*, volume 3, pages 45-48, Toulouse, France, May 2006.

[7] Bertrand David and Roland Badeau. "Fast sequential LS estimation for sinusoidal modeling and decomposition of audio signals". In *Proc. of IEEE WASPAA*, pages 211-214, New Paltz, New York, USA, October 2007.

Time-frequency analysis



[1] Roland Badeau, Bertrand David, and Gaël Richard. "A new perturbation analysis for signal enumeration in rotational invariance techniques". *IEEE Transactions on Signal Processing*, 54(2): 450–458, February 2006.

[2] Roland Badeau, Bertrand David, and Gaël Richard. "Fast Approximated Power Iteration Subspace Tracking". *IEEE Transactions on Signal Processing*, 53(8): 2931-2941, August 2005.

[3] Roland Badeau, Gaël Richard, and Bertrand David. "Sliding window adaptive SVD algorithms". *IEEE Transactions on Signal Processing*, 52(1): 1-10, January 2004.

[4] Roland Badeau, Gaël Richard, and Bertrand David. "Fast and stable YAST algorithm for principal and minor subspace tracking". *IEEE Transactions on Signal Processing*, 56(8): 3437-3446, August 2008.

[5] Roland Badeau, Gaël Richard, and Bertrand David. "Fast adaptive ESPRIT algorithm". In *Proc. of IEEE Workshop on Statistical Signal Processing (SSP)*, Bordeaux, France, July 2005.

[6] Bertrand David, Roland Badeau, and Gaël Richard. "HRHATRAC Algorithm for Spectral Line Tracking of Musical Signals". In *Proc. of IEEE ICASSP*, volume 3, pages 45-48, Toulouse, France, May 2006.

[7] Bertrand David and Roland Badeau. "Fast sequential LS estimation for sinusoidal modeling and decomposition of audio signals". In *Proc. of IEEE WASPAA*, pages 211-214, New Paltz, New York, USA, October 2007.

Power iteration method

- Power iteration method (recursive computation of $\boldsymbol{W}_r(t)$)
  1) $\boldsymbol{C}_{xy}(t) = \boldsymbol{C}_{xx}\boldsymbol{W}_r(t-1)$ (compression of $\boldsymbol{C}_{xx}$)
  2) $\boldsymbol{W}_r(t)\boldsymbol{R}(t) = \boldsymbol{C}_{xy}(t)$ (orthonormalisation of $\boldsymbol{C}_{xy}(t)$)
     - $\mathrm{Span}(\boldsymbol{W}_r(t))$ exponentially converges to the signal subspace
     - If 2) is an orthogonal-triangular (QR) decomposition, $\boldsymbol{W}_r(t)$ converges to the r principal eigenvectors of $\boldsymbol{C}_{xx}$
- Signal subspace tracking if $\boldsymbol{C}_{xx}(t)$ is time-varying
- Fast algorithm [Strobach, 1996] (complexity of nr^2 instead of n^2r)

Power iteration method

- Power iteration method (recursive computation of $\boldsymbol{W}_r(t)$)
  1) $\boldsymbol{C}_{xy}(t) = \boldsymbol{C}_{xx}\boldsymbol{W}_r(t-1)$ (compression of $\boldsymbol{C}_{xx}$)
  2) $\boldsymbol{W}_r(t)\boldsymbol{R}(t) = \boldsymbol{C}_{xy}(t)$ (orthonormalisation of $\boldsymbol{C}_{xy}(t)$)
     - $\mathrm{Span}(\boldsymbol{W}_r(t))$ exponentially converges to the signal subspace
     - If 2) is an orthogonal-triangular (QR) decomposition, $\boldsymbol{W}_r(t)$ converges to the r principal eigenvectors of $\boldsymbol{C}_{xx}$
- Signal subspace tracking if $\boldsymbol{C}_{xx}(t)$ is time-varying
- Fast algorithm [Strobach, 1996] (complexity of nr^2 instead of n^2r)

Power iteration method

- Power iteration method (recursive computation of $\boldsymbol{W}_r(t)$)
  1) $\boldsymbol{C}_{xy}(t) = \boldsymbol{C}_{xx}\boldsymbol{W}_r(t-1)$ (compression of $\boldsymbol{C}_{xx}$)
  2) $\boldsymbol{W}_r(t)\boldsymbol{R}(t) = \boldsymbol{C}_{xy}(t)$ (orthonormalisation of $\boldsymbol{C}_{xy}(t)$)
     - $\mathrm{Span}(\boldsymbol{W}_r(t))$ exponentially converges to the signal subspace
     - If 2) is an orthogonal-triangular (QR) decomposition, $\boldsymbol{W}_r(t)$ converges to the r principal eigenvectors of $\boldsymbol{C}_{xx}$
- Signal subspace tracking if $\boldsymbol{C}_{xx}(t)$ is time-varying
- Fast algorithm [Strobach, 1996] (complexity of nr^2 instead of n^2r)

Power iteration method

- Power iteration method (recursive computation of $\boldsymbol{W}_r(t)$)
  1) $\boldsymbol{C}_{xy}(t) = \boldsymbol{C}_{xx}\boldsymbol{W}_r(t-1)$ (compression of $\boldsymbol{C}_{xx}$)
  2) $\boldsymbol{W}_r(t)\boldsymbol{R}(t) = \boldsymbol{C}_{xy}(t)$ (orthonormalisation of $\boldsymbol{C}_{xy}(t)$)
     - $\mathrm{Span}(\boldsymbol{W}_r(t))$ exponentially converges to the signal subspace
     - If 2) is an orthogonal-triangular (QR) decomposition, $\boldsymbol{W}_r(t)$ converges to the r principal eigenvectors of $\boldsymbol{C}_{xx}$
- Signal subspace tracking if $\boldsymbol{C}_{xx}(t)$ is time-varying
- Fast algorithm [Strobach, 1996] (complexity of nr^2 instead of n^2r)

Power iteration method

- Power iteration method (recursive computation of $\boldsymbol{W}_r(t)$)
  1) $\boldsymbol{C}_{xy}(t) = \boldsymbol{C}_{xx}\boldsymbol{W}_r(t-1)$ (compression of $\boldsymbol{C}_{xx}$)
  2) $\boldsymbol{W}_r(t)\boldsymbol{R}(t) = \boldsymbol{C}_{xy}(t)$ (orthonormalisation of $\boldsymbol{C}_{xy}(t)$)
     - $\mathrm{Span}(\boldsymbol{W}_r(t))$ exponentially converges to the signal subspace
     - If 2) is an orthogonal-triangular (QR) decomposition, $\boldsymbol{W}_r(t)$ converges to the r principal eigenvectors of $\boldsymbol{C}_{xx}$
- Signal subspace tracking if $\boldsymbol{C}_{xx}(t)$ is time-varying
- Fast algorithm [Strobach, 1996] (complexity of nr^2 instead of n^2r)

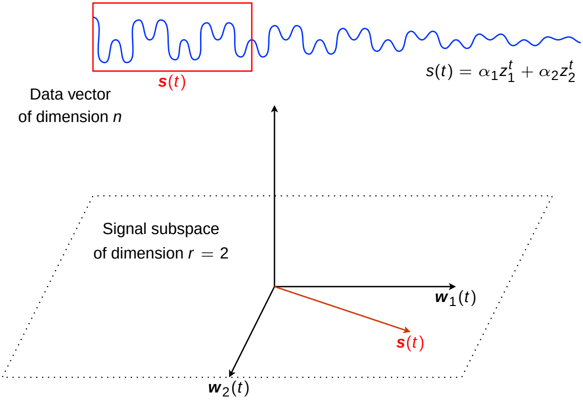
Power iteration method

- Power iteration method (recursive computation of $\boldsymbol{W}_r(t)$)
  1) $\boldsymbol{C}_{xy}(t) = \boldsymbol{C}_{xx}\boldsymbol{W}_r(t-1)$ (compression of $\boldsymbol{C}_{xx}$)
  2) $\boldsymbol{W}_r(t)\boldsymbol{R}(t) = \boldsymbol{C}_{xy}(t)$ (orthonormalisation of $\boldsymbol{C}_{xy}(t)$)
     - $\mathrm{Span}(\boldsymbol{W}_r(t))$ exponentially converges to the signal subspace
     - If 2) is an orthogonal-triangular (QR) decomposition, $\boldsymbol{W}_r(t)$ converges to the r principal eigenvectors of $\boldsymbol{C}_{xx}$
- Signal subspace tracking if $\boldsymbol{C}_{xx}(t)$ is time-varying
- Fast algorithm [Strobach, 1996] (complexity of nr^2 instead of n^2r)

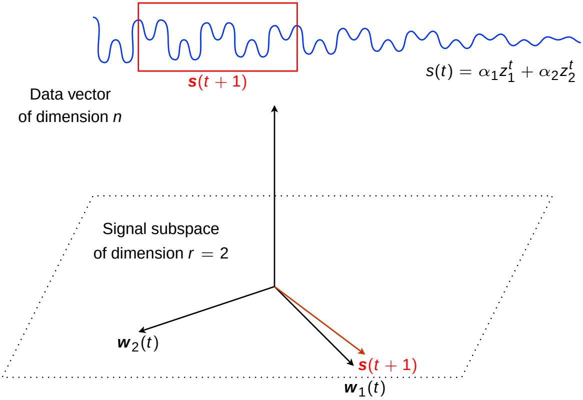
Power iteration method

- Power iteration method (recursive computation of $\boldsymbol{W}_r(t)$)
  1) $\boldsymbol{C}_{xy}(t) = \boldsymbol{C}_{xx}\boldsymbol{W}_r(t-1)$ (compression of $\boldsymbol{C}_{xx}$)
  2) $\boldsymbol{W}_r(t)\boldsymbol{R}(t) = \boldsymbol{C}_{xy}(t)$ (orthonormalisation of $\boldsymbol{C}_{xy}(t)$)
     - $\mathrm{Span}(\boldsymbol{W}_r(t))$ exponentially converges to the signal subspace
     - If 2) is an orthogonal-triangular (QR) decomposition, $\boldsymbol{W}_r(t)$ converges to the r principal eigenvectors of $\boldsymbol{C}_{xx}$
- Signal subspace tracking if $\boldsymbol{C}_{xx}(t)$ is time-varying
- Fast algorithm [Strobach, 1996] (complexity of nr^2 instead of n^2r)

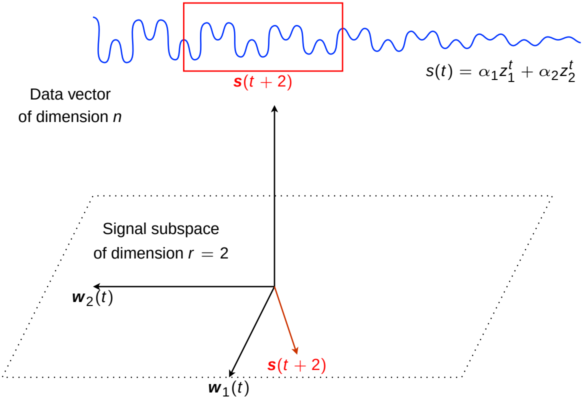
Subspace tracking



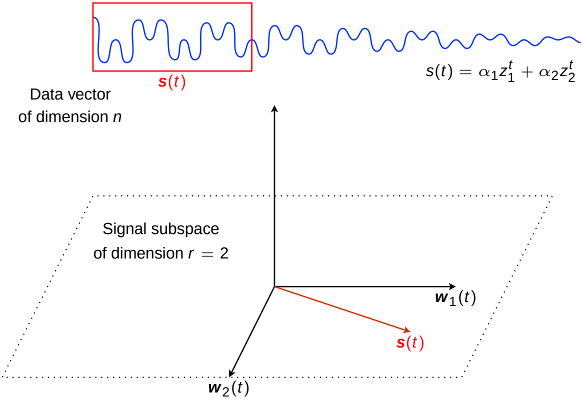
Subspace tracking



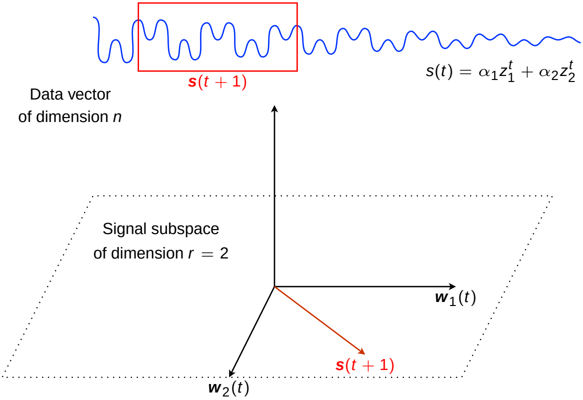
Subspace tracking



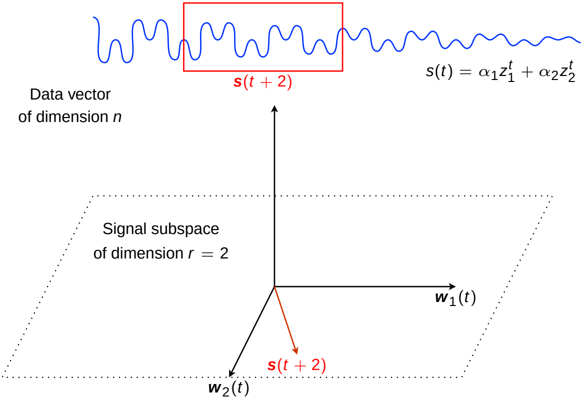
Subspace tracking



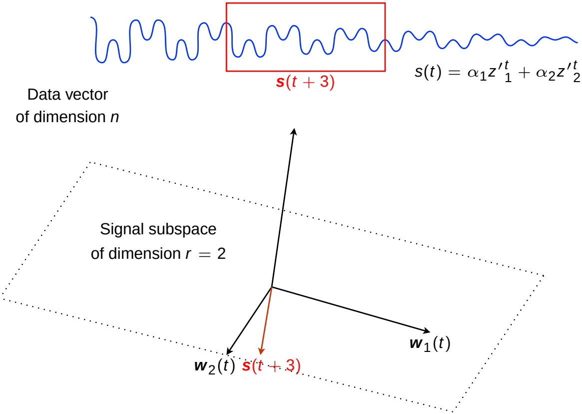
Subspace tracking



Subspace tracking



Subspace tracking



Natural power method

- Natural power method
  1) $\mathbf{C}_{xy}(t) = \mathbf{C}_{xx}(t)\mathbf{W}_r(t-1)$ (compression of $\mathbf{C}_{xx}$)
  2) $\mathbf{W}_r(t) = \mathbf{C}_{xy}(t)\left(\mathbf{C}_{xy}(t)^H\mathbf{C}_{xy}(t)\right)^{-\frac{1}{2}}$ (orthonormalisation of $\mathbf{C}_{xy}(t)$)
- FAPI algorithm [1] (complexity of $3nr$ instead of nr^2)
  - reaches the complexity lower bound ($3nr$)
  - converges faster than PAST and its variants
  - guarantees the orthonormality of $\mathbf{W}_r(t)$ and the numerical stability

[1] Roland Badeau, Bertrand David, and Gaël Richard. "Fast Approximated Power Iteration Subspace Tracking". *IEEE Transactions on Signal Processing*, 53(8): 2931-2941, August 2005.

Natural power method

- Natural power method
  1) $\mathbf{C}_{xy}(t) = \mathbf{C}_{xx}(t)\mathbf{W}_r(t-1)$ (compression of $\mathbf{C}_{xx}$)
  2) $\mathbf{W}_r(t) = \mathbf{C}_{xy}(t)\left(\mathbf{C}_{xy}(t)^H\mathbf{C}_{xy}(t)\right)^{-\frac{1}{2}}$ (orthonormalisation of $\mathbf{C}_{xy}(t)$)
- FAPI algorithm [1] (complexity of $3nr$ instead of nr^2)
  - reaches the complexity lower bound ($3nr$)
  - converges faster than PAST and its variants
  - guarantees the orthonormality of $\mathbf{W}_r(t)$ and the numerical stability

[1] Roland Badeau, Bertrand David, and Gaël Richard. "Fast Approximated Power Iteration Subspace Tracking". *IEEE Transactions on Signal Processing*, 53(8): 2931-2941, August 2005.

Natural power method

- Natural power method
  1) $\mathbf{C}_{xy}(t) = \mathbf{C}_{xx}(t)\mathbf{W}_r(t-1)$ (compression of $\mathbf{C}_{xx}$)
  2) $\mathbf{W}_r(t) = \mathbf{C}_{xy}(t)\left(\mathbf{C}_{xy}(t)^H\mathbf{C}_{xy}(t)\right)^{-\frac{1}{2}}$ (orthonormalisation of $\mathbf{C}_{xy}(t)$)
- FAPI algorithm [1] (complexity of $3nr$ instead of nr^2)
  - reaches the complexity lower bound ($3nr$)
  - converges faster than PAST and its variants
  - guarantees the orthonormality of $W_r(t)$ and the numerical stability

[1] Roland Badeau, Bertrand David, and Gaël Richard. "Fast Approximated Power Iteration Subspace Tracking". *IEEE Transactions on Signal Processing*, 53(8): 2931-2941, August 2005.

Natural power method

- Natural power method
  1) $\boldsymbol{C}_{xy}(t) = \boldsymbol{C}_{xx}(t)\boldsymbol{W}_r(t-1)$ (compression of $\boldsymbol{C}_{xx}$)
  2) $\boldsymbol{W}_r(t) = \boldsymbol{C}_{xy}(t)\left(\boldsymbol{C}_{xy}(t)^H\boldsymbol{C}_{xy}(t)\right)^{-\frac{1}{2}}$ (orthonormalisation of $\boldsymbol{C}_{xy}(t)$)
- FAPI algorithm [1] (complexity of $3nr$ instead of nr^2)
  - reaches the complexity lower bound ($3nr$)
  - converges faster than PAST and its variants
  - guarantees the orthonormality of $\boldsymbol{W}_r(t)$ and the numerical stability

[1] Roland Badeau, Bertrand David, and Gaël Richard. "Fast Approximated Power Iteration Subspace Tracking". *IEEE Transactions on Signal Processing*, 53(8): 2931-2941, August 2005.

Natural power method

- Natural power method
  1) $\boldsymbol{C}_{xy}(t) = \boldsymbol{C}_{xx}(t)\boldsymbol{W}_r(t-1)$ (compression of $\boldsymbol{C}_{xx}$)
  2) $\boldsymbol{W}_r(t) = \boldsymbol{C}_{xy}(t)\left(\boldsymbol{C}_{xy}(t)^H\boldsymbol{C}_{xy}(t)\right)^{-\frac{1}{2}}$ (orthonormalisation of $\boldsymbol{C}_{xy}(t)$)
- FAPI algorithm [1] (complexity of $3nr$ instead of nr^2)
  - reaches the complexity lower bound ($3nr$)
  - converges faster than PAST and its variants
  - guarantees the orthonormality of $\boldsymbol{W}_r(t)$ and the numerical stability

[1] Roland Badeau, Bertrand David, and Gaël Richard. "Fast Approximated Power Iteration Subspace Tracking". *IEEE Transactions on Signal Processing*, 53(8): 2931-2941, August 2005.

Natural power method

- Natural power method
  1) $\mathbf{C}_{xy}(t) = \mathbf{C}_{xx}(t)\mathbf{W}_r(t-1)$ (compression of $\mathbf{C}_{xx}$)
  2) $\mathbf{W}_r(t) = \mathbf{C}_{xy}(t)\left(\mathbf{C}_{xy}(t)^H\mathbf{C}_{xy}(t)\right)^{-\frac{1}{2}}$ (orthonormalisation of $\mathbf{C}_{xy}(t)$)
- FAPI algorithm [1] (complexity of $3nr$ instead of nr^2)
  - reaches the complexity lower bound ($3nr$)
  - converges faster than PAST and its variants
  - guarantees the orthonormality of $\mathbf{W}_r(t)$ and the numerical stability

[1] Roland Badeau, Bertrand David, and Gaël Richard. "Fast Approximated Power Iteration Subspace Tracking". *IEEE Transactions on Signal Processing*, 53(8): 2931-2941, August 2005.

Natural power method

- Natural power method
  1) $\mathbf{C}_{xy}(t) = \mathbf{C}_{xx}(t)\mathbf{W}_r(t-1)$ (compression of $\mathbf{C}_{xx}$)
  2) $\mathbf{W}_r(t) = \mathbf{C}_{xy}(t)\left(\mathbf{C}_{xy}(t)^H\mathbf{C}_{xy}(t)\right)^{-\frac{1}{2}}$ (orthonormalisation of $\mathbf{C}_{xy}(t)$)
- FAPI algorithm [1] (complexity of $3nr$ instead of nr^2)
  - reaches the complexity lower bound ($3nr$)
  - converges faster than PAST and its variants
  - guarantees the orthonormality of $\mathbf{W}_r(t)$ and the numerical stability

[1] Roland Badeau, Bertrand David, and Gaël Richard. "Fast Approximated Power Iteration Subspace Tracking". *IEEE Transactions on Signal Processing*, 53(8): 2931-2941, August 2005.

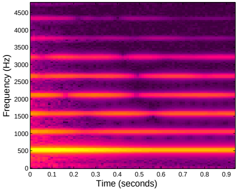
Applications of High Resolution analysis

- Analysis / Synthesis
  - High resolution time-frequency representation
  - Analysis of the sympathetic string modes in a concert harp
  - Audio coding
- Automatic transcription
  - Pitch estimation of piano notes
  - Musical tempo estimation
- Other applications
  - Channel estimation in digital communications
  - Adaptive multilinear SVD for structured tensors

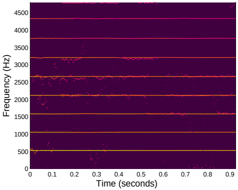
Applications of High Resolution analysis

- Analysis / Synthesis
  - High resolution time-frequency representation
  - Analysis of the sympathetic string modes in a concert harp
  - Audio coding
- Automatic transcription
  - Pitch estimation of piano notes
  - Musical tempo estimation
- Other applications
  - Channel estimation in digital communications
  - Adaptive multilinear SVD for structured tensors

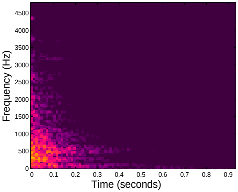
Decomposition of a piano sound



Spectrogram



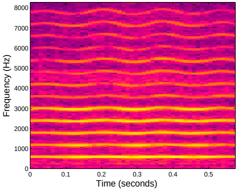
HR-ogram



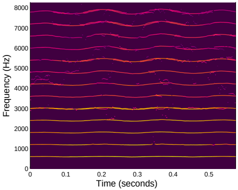
Residual

[1] Roland Badeau and Bertrand David. "Adaptive subspace methods for high resolution analysis of music signals". In Acoustics'08, Paris, France, July 2008.
[2] Bertrand David, Gaël Richard, and Roland Badeau. "An EDS modelling tool for tracking and modifying musical signals". In Proc. of Stockholm Music Acoustics Conference (SMAC), volume 2, pages 715-718, Stockholm, Sweden, August 2003.
[3] Roland Badeau, Rémy Boyer, and Bertrand David. "EDS parametric modeling and tracking of audio signals". In Proc. of the 5th International Conference on Digital Audio Effects (DAFx), pages 139-144, Hamburg, Germany, September 2002.

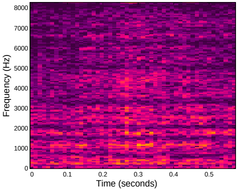
Decomposition of a violin sound



Spectrogram



HR-ogram



Residual

[1] Roland Badeau and Bertrand David. "Adaptive subspace methods for high resolution analysis of music signals". In Acoustics'08, Paris, France, July 2008.
[2] Bertrand David, Gaël Richard, and Roland Badeau. "An EDS modelling tool for tracking and modifying musical signals". In Proc. of Stockholm Music Acoustics Conference (SMAC), volume 2, pages 715-718, Stockholm, Sweden, August 2003.
[3] Roland Badeau, Rémy Boyer, and Bertrand David. "EDS parametric modeling and tracking of audio signals". In Proc. of the 5th International Conference on Digital Audio Effects (DAFx), pages 139-144, Hamburg, Germany, September 2002.

Sinusoids and noise separation

- Principle: projection onto the signal or the noise subspace [1,2]

Instrument	Original	Sinusoids	Noise
Piano			
Guitar			
Violin			
Flute			
Saxophone			
Bell			

[1] Roland Badeau, Rémy Boyer, and Bertrand David. "EDS parametric modeling and tracking of audio signals". In Proc. of the 5th International Conference on Digital Audio Effects (DAFx), pp. 139-144, Hamburg, Germany, September 2002.
[2] Bertrand David, Gaël Richard, and Roland Badeau. "An EDS modelling tool for tracking and modifying musical signals". In Proc. of Stockholm Music Acoustics Conference (SMAC), volume 2, pp. 715-718, Stockholm, Sweden, August 2003.

Drum separation and beat estimation

- Drum source separation [1]
  - Original (Aerosmith):
  - Separated drums:
  - Remix - more drums:
  - Remix - less drums:
- Beat tracking [2]
  - Pink Floyd:
  - Brad Mehldau:

[1] Olivier Gillet and Gaël Richard. Transcription and separation of drum signals from polyphonic music. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(3): 529-540, March 2008.
[2] Miguel Alonso Arevalo, Roland Badeau, Bertrand David, and Gaël Richard. "Musical tempo estimation using noise subspace projections". In *Proc. of IEEE WASPAA*, pages 95-98, New Paltz, New York, USA, October 2003.

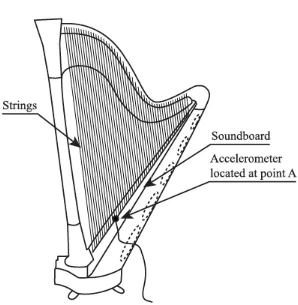
Drum separation and beat estimation

- Drum source separation [1]
  - Original (Aerosmith):
  - Separated drums:
  - Remix - more drums:
  - Remix - less drums:
- Beat tracking [2]
  - Pink Floyd:
  - Brad Mehldau:

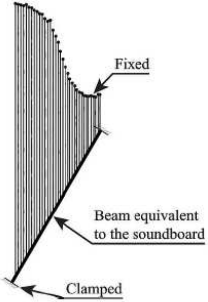
[1] Olivier Gillet and Gaël Richard. Transcription and separation of drum signals from polyphonic music. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(3): 529-540, March 2008.
[2] Miguel Alonso Arevalo, Roland Badeau, Bertrand David, and Gaël Richard. "Musical tempo estimation using noise subspace projections". In *Proc. of IEEE WASPAA*, pages 95-98, New Paltz, New York, USA, October 2003.

Sympathetic string modes in a concert harp

Modelling sympathetic string modes in a concert harp [1]



Experimental protocol

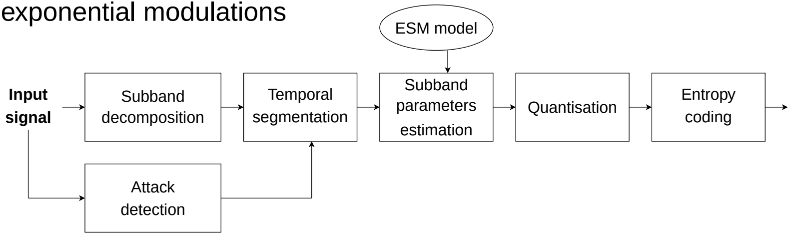


Physical model

[1] Jean-Loïc Le Carrou, François Gautier, and Roland Badeau. "Sympathetic string modes in the concert harp". *Acta Acustica united with Acustica*, 95(4): 744-752, July/August 2009.

Audio coding

- Parametric coder based on the ESM model [1]
  → exponential modulations



- Joint scalar quantisation with entropy constraint [2,3]

[1] Olivier Derrien, Gaël Richard, and Roland Badeau. "Damped sinusoids and subspace based approach for lossy audio coding". In *Acoustics'08*, Paris, France, July 2008.
[2] Olivier Derrien, Roland Badeau, and Gaël Richard. "Entropy-constrained quantization of exponentially damped sinusoids parameters". In *Proc. of IEEE ICASSP*, Prague, Czech Republic, May 2011.
[3] Olivier Derrien, Roland Badeau, and Gaël Richard. "Calculation of an entropy-constrained quantizer for exponentially damped sinudoids parameters". Technical report, Laboratoire de Mécanique et d'Acoustique, Marseille, France, June 2010.

Audio coding

Original sound:

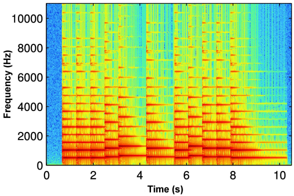
MDCT		ESM	
9 bits/spl		8.9 bits/spl	
8 bits/spl			
7 bits/spl		6.8 bits/spl	
6 bits/spl		6.4 bits/spl	
5 bits/spl		4.7 bits/spl	
4 bits/spl		4.4 bits/spl	
3 bits/spl		3.2 bits/spl	
2 bits/spl		2.1 bits/spl	

Part II

Nonnegative decompositions

Nonnegative Matrix Factorization (NMF)

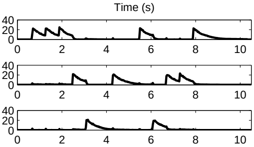
Musical score



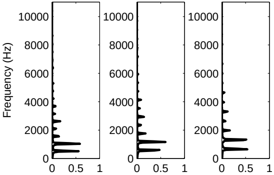
Spectrogram ***V***

Nonnegative Matrix Factorization (NMF)

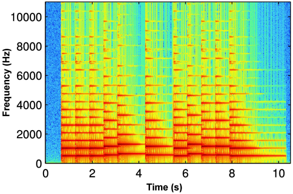
Musical score



Temporal activations ***H***



Spectral atoms ***W***



Spectrogram ***V***

β-divergence and multiplicative rules

- Minimisation of the criterion $D(\boldsymbol{V}|\boldsymbol{WH}) = \sum_{n=1}^{N}\sum_{f=1}^{F} d\left(v_{fn}\,|\hat{v}_{fn}\right)$
- β-**divergence** [Eguchi and Kano, 2001]:

$$d_\beta(a|b) = \frac{1}{\beta(\beta-1)}\left(a^\beta + (\beta-1)b^\beta - \beta a b^{\beta-1}\right)$$

  - $\beta = 2$ corresponds to **Euclidean distance** (EUC),
  - $\beta = 1$ corresponds to **Kullback-Leibler divergence** (KL),
  - $\beta = 0$ corresponds to **Itakura-Saito divergence** (IS),
  - $d_\beta(a|b)$ is convex w.r.t b if and only if $\beta \in [1,2]$,
- Multiplicative update rules [Kompass, 2007]:
  - $$\begin{cases} \boldsymbol{W} & \leftarrow & \boldsymbol{W} \otimes \frac{(\boldsymbol{V} \otimes (\boldsymbol{WH})^{\beta-2})\boldsymbol{H}^T}{(\boldsymbol{WH})^{\beta-1}\boldsymbol{H}^T} \\ \boldsymbol{H} & \leftarrow & \boldsymbol{H} \otimes \frac{\boldsymbol{W}^T(\boldsymbol{V} \otimes (\boldsymbol{WH})^{\beta-2})}{\boldsymbol{W}^T(\boldsymbol{WH})^{\beta-1}} \end{cases}$$
  - $D(\boldsymbol{V}|\boldsymbol{WH})$ is non-increasing if and only if $\beta \in [1,2]$.

β-divergence and multiplicative rules

- Minimisation of the criterion $D(\boldsymbol{V}|\boldsymbol{WH}) = \sum_{n=1}^{N}\sum_{f=1}^{F} d\left(v_{fn}\,|\hat{v}_{fn}\right)$
- β**-divergence** [Eguchi and Kano, 2001]:

$$d_\beta(a|b) = \frac{1}{\beta(\beta-1)}\left(a^\beta + (\beta-1)b^\beta - \beta a b^{\beta-1}\right)$$

  - $\beta = 2$ corresponds to **Euclidean distance** (EUC),
  - $\beta = 1$ corresponds to **Kullback-Leibler divergence** (KL),
  - $\beta = 0$ corresponds to **Itakura-Saito divergence** (IS),
  - $d_\beta(a|b)$ is convex w.r.t b if and only if $\beta \in [1, 2]$,
- Multiplicative update rules [Kompass, 2007]:
  - $$\begin{cases} \boldsymbol{W} & \leftarrow & \boldsymbol{W} \otimes \frac{(\boldsymbol{V}\otimes(\boldsymbol{WH})^{\beta-2})\boldsymbol{H}^T}{(\boldsymbol{WH})^{\beta-1}\boldsymbol{H}^T} \\ \boldsymbol{H} & \leftarrow & \boldsymbol{H} \otimes \frac{\boldsymbol{W}^T(\boldsymbol{V}\otimes(\boldsymbol{WH})^{\beta-2})}{\boldsymbol{W}^T(\boldsymbol{WH})^{\beta-1}} \end{cases}$$
  - $D(\boldsymbol{V}|\boldsymbol{WH})$ is non-increasing if and only if $\beta \in [1, 2]$.

β-divergence and multiplicative rules

- Minimisation of the criterion $D(\boldsymbol{V}|\boldsymbol{WH}) = \sum_{n=1}^{N} \sum_{f=1}^{F} d\left(v_{fn}|\hat{v}_{fn}\right)$
- β**-divergence** [Eguchi and Kano, 2001]:

$$d_\beta(a|b) = \frac{1}{\beta(\beta-1)}\left(a^\beta + (\beta-1)b^\beta - \beta a b^{\beta-1}\right)$$

  - $\beta = 2$ corresponds to **Euclidean distance** (EUC),
  - $\beta = 1$ corresponds to **Kullback-Leibler divergence** (KL),
  - $\beta = 0$ corresponds to **Itakura-Saito divergence** (IS),
  - $d_\beta(a|b)$ is convex w.r.t b if and only if $\beta \in [1, 2]$,
- Multiplicative update rules [Kompass, 2007]:
  - $$\begin{cases} \boldsymbol{W} & \leftarrow & \boldsymbol{W} \otimes \frac{(\boldsymbol{V} \otimes (\boldsymbol{WH})^{\beta-2})\boldsymbol{H}^T}{(\boldsymbol{WH})^{\beta-1}\boldsymbol{H}^T} \\ \boldsymbol{H} & \leftarrow & \boldsymbol{H} \otimes \frac{\boldsymbol{W}^T(\boldsymbol{V} \otimes (\boldsymbol{WH})^{\beta-2})}{\boldsymbol{W}^T(\boldsymbol{WH})^{\beta-1}} \end{cases}$$
  - $D(\boldsymbol{V}|\boldsymbol{WH})$ is non-increasing if and only if $\beta \in [1, 2]$.

Stability of multiplicative update rules

- Introduction of an exponentiation step η into NMF multiplicative update rules designed for minimizing the β-divergence [1]:

$$\boldsymbol{W} \leftarrow \boldsymbol{W} \otimes \left(\frac{(\boldsymbol{V} \otimes (\boldsymbol{WH})^{\beta-2}) \boldsymbol{H}^T}{(\boldsymbol{WH})^{\beta-1} \boldsymbol{H}^T} \right)^{\eta} \quad (1)$$

$$\boldsymbol{H} \leftarrow \boldsymbol{H} \otimes \left(\frac{\boldsymbol{W}^T (\boldsymbol{V} \otimes (\boldsymbol{WH})^{\beta-2})}{\boldsymbol{W}^T (\boldsymbol{WH})^{\beta-1}} \right)^{\eta} \quad (2)$$

- Monotonic decrease of the criterion if $\beta \in [1, 2]$ and $\eta \in]0, 1]$
- Exponential or asymptotic stability of both rules (1) and (2) if $\eta \in]0, \eta^*[$, where $\forall \beta \in \mathbb{R}$, $\eta^\star \in]0, 2]$ and if $\beta \in [1, 2]$, $\eta^\star = 2$
- Instability if $\eta \notin [0, 2]$ $\forall \beta \in \mathbb{R}$
- Step η permits to control the convergence rate

[1] Roland Badeau, Nancy Bertin, and Emmanuel Vincent. "Stability analysis of multiplicative update algorithms and application to nonnegative matrix factorization". *IEEE Transactions on Neural Networks*, vol.21, no. 12, pp. 1869-1881, December 2010.

Stability of multiplicative update rules

- Introduction of an exponentiation step η into NMF multiplicative update rules designed for minimizing the β-divergence [1]:

$$\boldsymbol{W} \leftarrow \boldsymbol{W} \otimes \left(\frac{(\boldsymbol{V} \otimes (\boldsymbol{W}\boldsymbol{H})^{\beta-2})\boldsymbol{H}^T}{(\boldsymbol{W}\boldsymbol{H})^{\beta-1}\boldsymbol{H}^T}\right)^{\eta} \quad (1)$$

$$\boldsymbol{H} \leftarrow \boldsymbol{H} \otimes \left(\frac{\boldsymbol{W}^T(\boldsymbol{V} \otimes (\boldsymbol{W}\boldsymbol{H})^{\beta-2})}{\boldsymbol{W}^T(\boldsymbol{W}\boldsymbol{H})^{\beta-1}}\right)^{\eta} \quad (2)$$

- Monotonic decrease of the criterion if $\beta \in [1, 2]$ and $\eta \in]0, 1]$
- Exponential or asymptotic stability of both rules (1) and (2) if $\eta \in]0, \eta^*[$, where $\forall \beta \in \mathbb{R}$, $\eta^\star \in]0, 2]$ and if $\beta \in [1, 2]$, $\eta^\star = 2$
- Instability if $\eta \notin [0, 2]$ $\forall \beta \in \mathbb{R}$
- Step η permits to control the convergence rate

[1] Roland Badeau, Nancy Bertin, and Emmanuel Vincent. "Stability analysis of multiplicative update algorithms and application to nonnegative matrix factorization". *IEEE Transactions on Neural Networks*, vol.21, no. 12, pp. 1869-1881, December 2010.

Stability of multiplicative update rules

- Introduction of an exponentiation step η into NMF multiplicative update rules designed for minimizing the β-divergence [1]:

$$\boldsymbol{W} \leftarrow \boldsymbol{W} \otimes \left(\frac{(\boldsymbol{V} \otimes (\boldsymbol{W}\boldsymbol{H})^{\beta-2})\boldsymbol{H}^T}{(\boldsymbol{W}\boldsymbol{H})^{\beta-1}\boldsymbol{H}^T}\right)^{\eta} \quad (1)$$

$$\boldsymbol{H} \leftarrow \boldsymbol{H} \otimes \left(\frac{\boldsymbol{W}^T(\boldsymbol{V} \otimes (\boldsymbol{W}\boldsymbol{H})^{\beta-2})}{\boldsymbol{W}^T(\boldsymbol{W}\boldsymbol{H})^{\beta-1}}\right)^{\eta} \quad (2)$$

- Monotonic decrease of the criterion if $\beta \in [1, 2]$ and $\eta \in]0, 1]$
- Exponential or asymptotic stability of both rules (1) and (2) if $\eta \in]0, \eta^*[$, where $\forall \beta \in \mathbb{R}$, $\eta^\star \in]0, 2]$ and if $\beta \in [1, 2]$, $\eta^\star = 2$
- Instability if $\eta \notin [0, 2]$ $\forall \beta \in \mathbb{R}$
- Step η permits to control the convergence rate

[1] Roland Badeau, Nancy Bertin, and Emmanuel Vincent. "Stability analysis of multiplicative update algorithms and application to nonnegative matrix factorization". *IEEE Transactions on Neural Networks*, vol.21, no. 12, pp. 1869-1881, December 2010.

Stability of multiplicative update rules

- Introduction of an exponentiation step η into NMF multiplicative update rules designed for minimizing the β-divergence [1]:

$$\boldsymbol{W} \leftarrow \boldsymbol{W} \otimes \left(\frac{(\boldsymbol{V} \otimes (\boldsymbol{W}\boldsymbol{H})^{\beta-2})\boldsymbol{H}^T}{(\boldsymbol{W}\boldsymbol{H})^{\beta-1}\boldsymbol{H}^T} \right)^{\eta} \quad (1)$$

$$\boldsymbol{H} \leftarrow \boldsymbol{H} \otimes \left(\frac{\boldsymbol{W}^T(\boldsymbol{V} \otimes (\boldsymbol{W}\boldsymbol{H})^{\beta-2})}{\boldsymbol{W}^T(\boldsymbol{W}\boldsymbol{H})^{\beta-1}} \right)^{\eta} \quad (2)$$

- Monotonic decrease of the criterion if $\beta \in [1, 2]$ and $\eta \in]0, 1]$
- Exponential or asymptotic stability of both rules (1) and (2) if $\eta \in]0, \eta^*[$, where $\forall \beta \in \mathbb{R}$, $\eta^\star \in]0, 2]$ and if $\beta \in [1, 2]$, $\eta^\star = 2$
- Instability if $\eta \notin [0, 2]$ $\forall \beta \in \mathbb{R}$
- Step η permits to control the convergence rate

[1] Roland Badeau, Nancy Bertin, and Emmanuel Vincent. "Stability analysis of multiplicative update algorithms and application to nonnegative matrix factorization". *IEEE Transactions on Neural Networks*, vol.21, no. 12, pp. 1869-1881, December 2010.

Stability of multiplicative update rules

- Introduction of an exponentiation step η into NMF multiplicative update rules designed for minimizing the β-divergence [1]:

$$\boldsymbol{W} \leftarrow \boldsymbol{W} \otimes \left(\frac{(\boldsymbol{V} \otimes (\boldsymbol{W}\boldsymbol{H})^{\beta-2})\boldsymbol{H}^T}{(\boldsymbol{W}\boldsymbol{H})^{\beta-1}\boldsymbol{H}^T} \right)^{\eta} \quad (1)$$

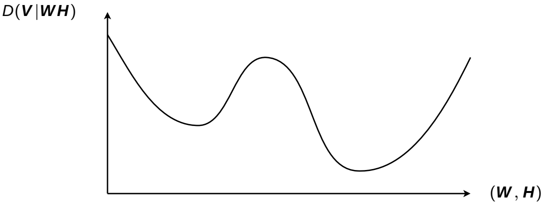
$$\boldsymbol{H} \leftarrow \boldsymbol{H} \otimes \left(\frac{\boldsymbol{W}^T(\boldsymbol{V} \otimes (\boldsymbol{W}\boldsymbol{H})^{\beta-2})}{\boldsymbol{W}^T(\boldsymbol{W}\boldsymbol{H})^{\beta-1}} \right)^{\eta} \quad (2)$$

- Monotonic decrease of the criterion if $\beta \in [1,2]$ and $\eta \in]0,1]$
- Exponential or asymptotic stability of both rules (1) and (2) if $\eta \in]0, \eta^*[$, where $\forall \beta \in \mathbb{R}$, $\eta^\star \in]0,2]$ and if $\beta \in [1,2]$, $\eta^\star = 2$
- Instability if $\eta \notin [0,2]$ $\forall \beta \in \mathbb{R}$
- Step η permits to control the convergence rate

[1] Roland Badeau, Nancy Bertin, and Emmanuel Vincent. "Stability analysis of multiplicative update algorithms and application to nonnegative matrix factorization". *IEEE Transactions on Neural Networks*, vol.21, no. 12, pp. 1869-1881, December 2010.

Avoiding local minima

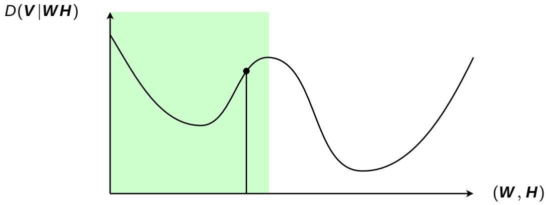
- The three divergences EUC, KL, and IS have local minima [1]



[1] Nancy Bertin and Roland Badeau. "Initialization, distances and local minima in audio applications of the nonnegative matrix factorization". In *Acoustics'08*, Paris, France, July 2008.

Avoiding local minima

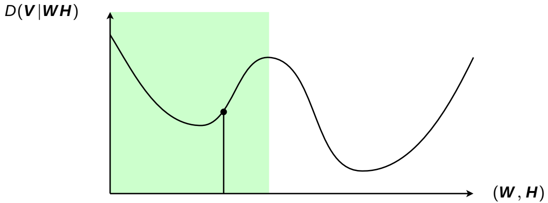
- The three divergences EUC, KL, and IS have local minima [1]



[1] Nancy Bertin and Roland Badeau. "Initialization, distances and local minima in audio applications of the nonnegative matrix factorization". In *Acoustics'08*, Paris, France, July 2008.

Avoiding local minima

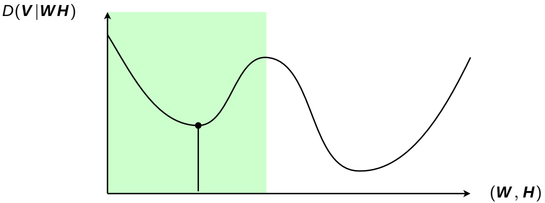
- The three divergences EUC, KL, and IS have local minima [1]



[1] Nancy Bertin and Roland Badeau. "Initialization, distances and local minima in audio applications of the nonnegative matrix factorization". In *Acoustics'08*, Paris, France, July 2008.

Avoiding local minima

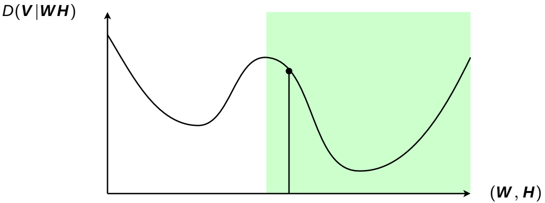
- The three divergences EUC, KL, and IS have local minima [1]



[1] Nancy Bertin and Roland Badeau. "Initialization, distances and local minima in audio applications of the nonnegative matrix factorization". In *Acoustics'08*, Paris, France, July 2008.

Avoiding local minima

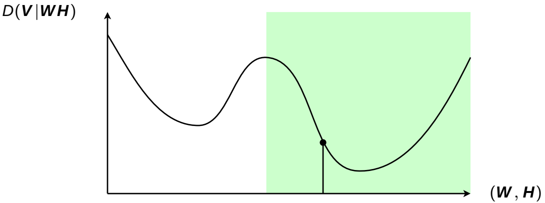
- The three divergences EUC, KL, and IS have local minima [1]



[1] Nancy Bertin and Roland Badeau. "Initialization, distances and local minima in audio applications of the nonnegative matrix factorization". In *Acoustics'08*, Paris, France, July 2008.

Avoiding local minima

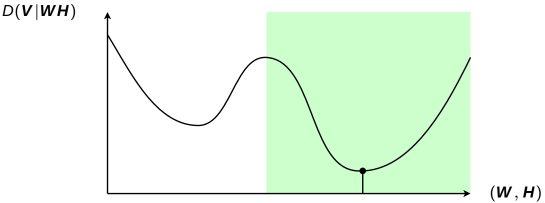
- The three divergences EUC, KL, and IS have local minima [1]



[1] Nancy Bertin and Roland Badeau. "Initialization, distances and local minima in audio applications of the nonnegative matrix factorization". In *Acoustics'08*, Paris, France, July 2008.

Avoiding local minima

- The three divergences EUC, KL, and IS have local minima [1]



[1] Nancy Bertin and Roland Badeau. "Initialization, distances and local minima in audio applications of the nonnegative matrix factorization". In *Acoustics'08*, Paris, France, July 2008.

Avoiding local minima

- Strategies for initialising the algorithm [1]
  - Failure of algorithms from automatic classification
- "Simulated cooling" algorithm for IS-NMF [2]
  - Parameter β becomes a function of the iteration index p:
  - The best transcription is not obtained by finding the lowest minimum of the criterion, but rather by constraining the decomposition

[1] Nancy Bertin and Roland Badeau. "Initialization, distances and local minima in audio applications of the nonnegative matrix factorization". In *Acoustics'08*, Paris, France, July 2008.
[2] Nancy Bertin, Cédric Févotte, and Roland Badeau. "A tempering approach for Itakura-Saito nonnegative matrix factorization. With application to music transcription". In *Proc. of IEEE ICASSP*, pages 1545–1548, Taipei, Taiwan, April 2009.

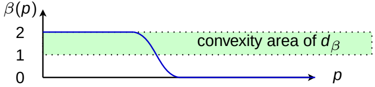
Avoiding local minima

- Strategies for initialising the algorithm [1]
  - Failure of algorithms from automatic classification
- "Simulated cooling" algorithm for IS-NMF [2]
  - Parameter β becomes a function of the iteration index p:
  - The best transcription is not obtained by finding the lowest minimum of the criterion, but rather by constraining the decomposition

[1] Nancy Bertin and Roland Badeau. "Initialization, distances and local minima in audio applications of the nonnegative matrix factorization". In *Acoustics'08*, Paris, France, July 2008.
[2] Nancy Bertin, Cédric Févotte, and Roland Badeau. "A tempering approach for Itakura-Saito nonnegative matrix factorization. With application to music transcription". In *Proc. of IEEE ICASSP*, pages 1545–1548, Taipei, Taiwan, April 2009.

Avoiding local minima

- Strategies for initialising the algorithm [1]
  - Failure of algorithms from automatic classification
- "Simulated cooling" algorithm for IS-NMF [2]
  - Parameter β becomes a function of the iteration index p:

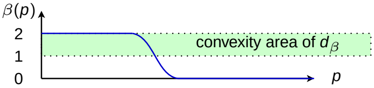


  - The best transcription is not obtained by finding the lowest minimum of the criterion, but rather by constraining the decomposition

[1] Nancy Bertin and Roland Badeau. "Initialization, distances and local minima in audio applications of the nonnegative matrix factorization". In *Acoustics'08*, Paris, France, July 2008.
[2] Nancy Bertin, Cédric Févotte, and Roland Badeau. "A tempering approach for Itakura-Saito nonnegative matrix factorization. With application to music transcription". In *Proc. of IEEE ICASSP*, pages 1545–1548, Taipei, Taiwan, April 2009.

Avoiding local minima

- Strategies for initialising the algorithm [1]
  - Failure of algorithms from automatic classification
- "Simulated cooling" algorithm for IS-NMF [2]
  - Parameter β becomes a function of the iteration index p:

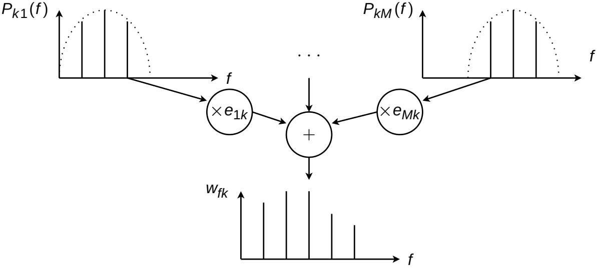


  - The best transcription is not obtained by finding the lowest minimum of the criterion, but rather by constraining the decomposition

[1] Nancy Bertin and Roland Badeau. "Initialization, distances and local minima in audio applications of the nonnegative matrix factorization". In *Acoustics'08*, Paris, France, July 2008.
[2] Nancy Bertin, Cédric Févotte, and Roland Badeau. "A tempering approach for Itakura-Saito nonnegative matrix factorization. With application to music transcription". In *Proc. of IEEE ICASSP*, pages 1545–1548, Taipei, Taiwan, April 2009.

Harmonicity and spectral smoothness

- Model [1]: $\hat{v}_{fn} = \sum_{k=1}^{K} w_{fk}(\theta) h_{kn}$ where $w_{fk}(\theta) = \sum_{m=1}^{M} e_{mk} P_{km}(f)$

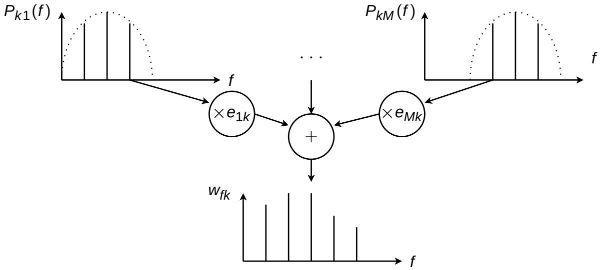


- $P_{km}(f)$ is a predefined harmonic spectral pattern
- e_{mk} and h_{kn} are estimated by means of a multiplicative algorithm

[1] Emmanuel Vincent, Nancy Bertin, and Roland Badeau. "Adaptive harmonic spectral decomposition for multiple pitch estimation". *IEEE Transactions on Audio, Speech, and Language Processing*, 18(3): 528- 537, March 2010.

Harmonicity and spectral smoothness

- Model [1]: $\hat{v}_{fn} = \sum_{k=1}^{K} w_{fk}(\theta) h_{kn}$ where $w_{fk}(\theta) = \sum_{m=1}^{M} e_{mk} P_{km}(f)$

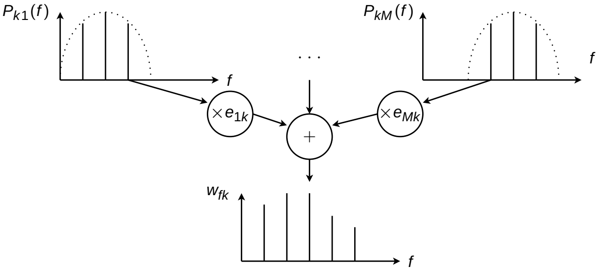


- $P_{km}(f)$ is a predefined harmonic spectral pattern
- e_{mk} and h_{kn} are estimated by means of a multiplicative algorithm

[1] Emmanuel Vincent, Nancy Bertin, and Roland Badeau. "Adaptive harmonic spectral decomposition for multiple pitch estimation". *IEEE Transactions on Audio, Speech, and Language Processing*, 18(3): 528- 537, March 2010.

Harmonicity and spectral smoothness

- Model [1]: $\hat{v}_{fn} = \sum_{k=1}^{K} w_{fk}(\theta) h_{kn}$ where $w_{fk}(\theta) = \sum_{m=1}^{M} e_{mk} P_{km}(f)$



- $P_{km}(f)$ is a predefined harmonic spectral pattern
- e_{mk} and h_{kn} are estimated by means of a multiplicative algorithm

[1] Emmanuel Vincent, Nancy Bertin, and Roland Badeau. "Adaptive harmonic spectral decomposition for multiple pitch estimation". *IEEE Transactions on Audio, Speech, and Language Processing*, 18(3): 528- 537, March 2010.

Temporal smoothness

- MAP estimator: $C(\boldsymbol{\Theta}) = L(\boldsymbol{\Theta}) + \log(p(\boldsymbol{\Theta}))$ où $\boldsymbol{\Theta} = \{e_{mk}, h_{kn}\}$
- Markov chain structured a priori distribution:

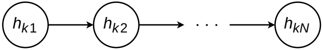
$$p(\boldsymbol{H}) = \prod_{k=1}^{K} p(h_{k1}) \prod_{n=2}^{N} p(h_{kn}|h_{k(n-1)})$$

  where $p(h_{kn}|h_{k(n-1)})$ follows an inverse-Gamma distribution
- SAGE algorithm [1] and multiplicative update rules [2] with $\eta = 0.5$

[1] Nancy Bertin, Roland Badeau, and Emmanuel Vincent. "Enforcing Harmonicity and Smoothness in Bayesian Nonnegative Matrix Factorization Applied to Polyphonic Music Transcription". *IEEE Transactions on Audio, Speech, and Language Processing*, 18(3): 538-549, March 2010.
[2] Nancy Bertin, Roland Badeau, and Emmanuel Vincent. "Fast Bayesian NMF algorithms enforcing harmonicity and temporal continuity in polyphonic music transcription". In *Proc. of IEEE WASPAA*, pages 29-32, New York, USA, October 2009.

Temporal smoothness

- MAP estimator: $C(\boldsymbol{\Theta}) = L(\boldsymbol{\Theta}) + \log(p(\boldsymbol{\Theta}))$ où $\boldsymbol{\Theta} = \{e_{mk}, h_{kn}\}$
- Markov chain structured a priori distribution:



$$p(\boldsymbol{H}) = \prod_{k=1}^{K} p(h_{k1}) \prod_{n=2}^{N} p(h_{kn}|h_{k(n-1)})$$

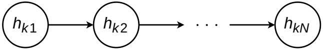
where $p(h_{kn}|h_{k(n-1)})$ follows an inverse-Gamma distribution

- SAGE algorithm [1] and multiplicative update rules [2] with $\eta = 0.5$

[1] Nancy Bertin, Roland Badeau, and Emmanuel Vincent. "Enforcing Harmonicity and Smoothness in Bayesian Nonnegative Matrix Factorization Applied to Polyphonic Music Transcription". *IEEE Transactions on Audio, Speech, and Language Processing*, 18(3): 538-549, March 2010.
[2] Nancy Bertin, Roland Badeau, and Emmanuel Vincent. "Fast Bayesian NMF algorithms enforcing harmonicity and temporal continuity in polyphonic music transcription". In *Proc. of IEEE WASPAA*, pages 29-32, New York, USA, October 2009.

Temporal smoothness

- MAP estimator: $C(\boldsymbol{\Theta}) = L(\boldsymbol{\Theta}) + \log(p(\boldsymbol{\Theta}))$ où $\boldsymbol{\Theta} = \{e_{mk}, h_{kn}\}$
- Markov chain structured a priori distribution:



$$p(\boldsymbol{H}) = \prod_{k=1}^{K} p(h_{k1}) \prod_{n=2}^{N} p(h_{kn}|h_{k(n-1)})$$

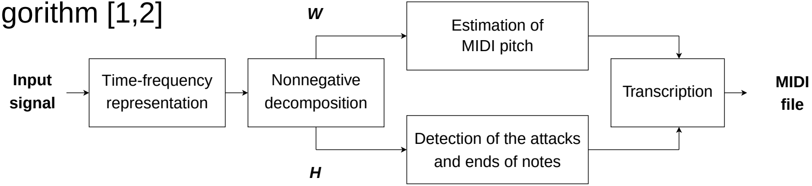
where $p(h_{kn}|h_{k(n-1)})$ follows an inverse-Gamma distribution

- SAGE algorithm [1] and multiplicative update rules [2] with $\eta = 0.5$

[1] Nancy Bertin, Roland Badeau, and Emmanuel Vincent. "Enforcing Harmonicity and Smoothness in Bayesian Nonnegative Matrix Factorization Applied to Polyphonic Music Transcription". *IEEE Transactions on Audio, Speech, and Language Processing*, 18(3): 538-549, March 2010.
[2] Nancy Bertin, Roland Badeau, and Emmanuel Vincent. "Fast Bayesian NMF algorithms enforcing harmonicity and temporal continuity in polyphonic music transcription". In *Proc. of IEEE WASPAA*, pages 29-32, New York, USA, October 2009.

NMF-based automatic transcription

- Algorithm [1,2]



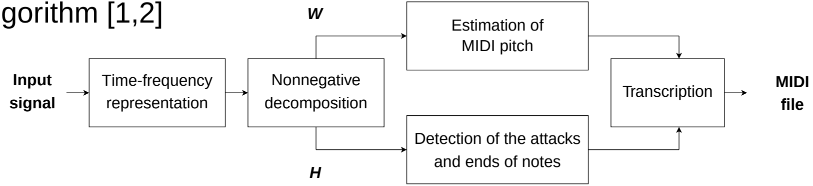
- Demo
  - Original signal (Liszt):
  - Transcribed signal:

[1] Nancy Bertin, Roland Badeau, and Gaël Richard. "Blind signal decompositions for automatic transcription of polyphonic music: NMF and K-SVD on the benchmark". In *Proc. of IEEE ICASSP*, volume 1, pages 65-68, Honolulu, Hawaii, USA, April 2007.

[2] Nancy Bertin, Roland Badeau, and Emmanuel Vincent. "Enforcing Harmonicity and Smoothness in Bayesian Nonnegative Matrix Factorization Applied to Polyphonic Music Transcription". *IEEE Transactions on Audio, Speech, and Language Processing*, 18(3): 538-549, March 2010.

NMF-based automatic transcription

- Algorithm [1,2]

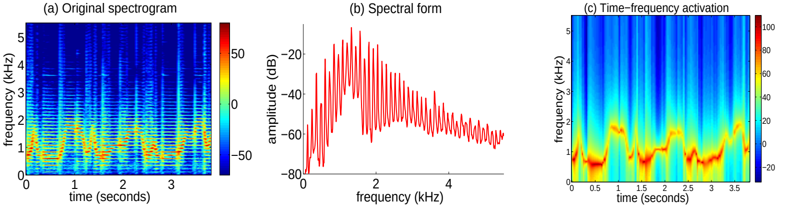


- Demo
  - Original signal (Liszt):
  - Transcribed signal:

[1] Nancy Bertin, Roland Badeau, and Gaël Richard. "Blind signal decompositions for automatic transcription of polyphonic music: NMF and K-SVD on the benchmark". In *Proc. of IEEE ICASSP*, volume 1, pages 65-68, Honolulu, Hawaii, USA, April 2007.
[2] Nancy Bertin, Roland Badeau, and Emmanuel Vincent. "Enforcing Harmonicity and Smoothness in Bayesian Nonnegative Matrix Factorization Applied to Polyphonic Music Transcription". *IEEE Transactions on Audio, Speech, and Language Processing*, 18(3): 538-549, March 2010.

Time-frequency activations

Model [1]: $\hat{v}_{fn} = \sum_{k=1}^{K} w_{fk} h_{kn}(f)$ where $h_{kn}(f) = \sigma_{kn}^2 \frac{\left|1+\sum_{q=1}^{Q} b_{kn}^{(q)} e^{-i2\pi\nu_f q}\right|^2}{\left|1+\sum_{p=1}^{P} a_{kn}^{(p)} e^{-i2\pi\nu_f p}\right|^2}$

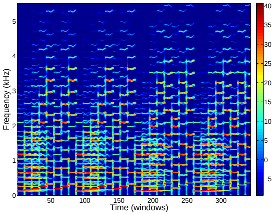


Jew's harp signal decomposed with an ARMA of order (1,1)

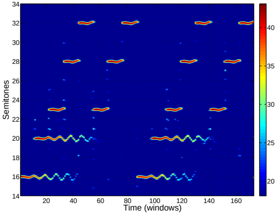
[1] Romain Hennequin, Roland Badeau, and Bertrand David. "NMF with time-frequency activations to model non-stationary audio events". *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 4, pp. 744-753, May 2011.

Fundamental frequencies variations

Model [1]: $\hat{v}_{fn} = \sum_{k=1}^{K} w_{fk}(\nu_{kn}^0)\, h_{kn}$ where $w_{fk}(\nu_{kn}^0) = \sum_{h=1}^{n_h(\nu_{kn}^0)} a_h\, g(\nu_f - h\,\nu_{kn}^0)$



Original spectrogram



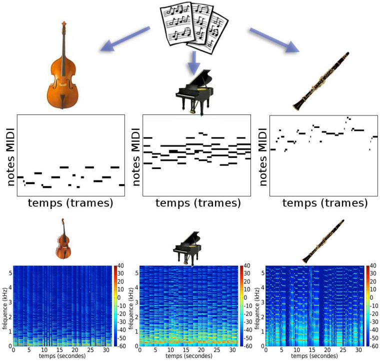
Temporal activations

Decomposition of an excerpt of J.S. Bach's first prelude

[1] Romain Hennequin, Roland Badeau, and Bertrand David. "Time-dependent parametric and harmonic templates in nonnegative matrix factorization". In *Proc. of DAFx*, Graz, Austria, September 2010.

Score-based informed source separation

- Algorithm [1]

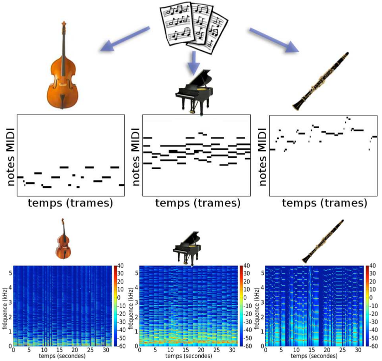


- Round Midnight (Thelonious Monk):

[1] Romain Hennequin, Bertrand David, and Roland Badeau. "Score informed audio source separation using a parametric model of non-negative spectrogram". In *Proc. of IEEE ICASSP*, Prague, Czech Republic, May 2011.

Score-based informed source separation

- Algorithm [1]



- Round Midnight (Thelonious Monk):

[1] Romain Hennequin, Bertrand David, and Roland Badeau. "Score informed audio source separation using a parametric model of non-negative spectrogram". In *Proc. of IEEE ICASSP*, Prague, Czech Republic, May 2011.

Conclusions

- Nonstationary signal modelling
  - Adaptive high resolution methods
  - Nonnegative decompositions enforcing harmonicity and smoothness
- Applications to audio and music signals
  - Source separation, audio coding,
  - Pitch and tempo estimation, automatic transcription
- Outlooks
  - Is it possible to merge HR methods and NMF in some way?
  - .. to be continued in an upcoming seminar (March 6)

Conclusions

- Nonstationary signal modelling
  - Adaptive high resolution methods
  - Nonnegative decompositions enforcing harmonicity and smoothness
- Applications to audio and music signals
  - Source separation, audio coding,
  - Pitch and tempo estimation, automatic transcription
- Outlooks
  - Is it possible to merge HR methods and NMF in some way?
  - .. to be continued in an upcoming seminar (March 6)

Conclusions

- Nonstationary signal modelling
  - Adaptive high resolution methods
  - Nonnegative decompositions enforcing harmonicity and smoothness
- Applications to audio and music signals
  - Source separation, audio coding,
  - Pitch and tempo estimation, automatic transcription
- Outlooks
  - Is it possible to merge HR methods and NMF in some way?
  - .. to be continued in an upcoming seminar (March 6)

Conclusions

- Nonstationary signal modelling
  - Adaptive high resolution methods
  - Nonnegative decompositions enforcing harmonicity and smoothness
- Applications to audio and music signals
  - Source separation, audio coding,
  - Pitch and tempo estimation, automatic transcription
- Outlooks
  - Is it possible to merge HR methods and NMF in some way?
  - .. to be continued in an upcoming seminar (March 6)