Introduction

→ Context and motivations:
  → Real-time system for polyphonic music transcription.
  → Non-negative matrix factorization and β-divergence.
  → Front-end for musical interactions in live performances.

→ Contributions:
  → Non-negative decomposition scheme tailored to real-time.
  → Intuition in understanding the relevancy of the β-divergence.
  → Comparative evaluation with off-line algorithms at the state-of-the-art.

Standard non-negative matrix factorization

→ Problem formulation and multiplicative updates:
  → Given \( \mathbf{V} \in \mathbb{R}_{+}^{m \times n} \) and \( r < \min(n, m) \), \( \mathbf{V} \) is modeled as follows:
    \[
    \mathbf{V} \approx \mathbf{W} \mathbf{H} \quad \text{with} \quad \mathbf{W} \in \mathbb{R}_{+}^{m \times r} \quad \text{and} \quad \mathbf{H} \in \mathbb{R}_{+}^{r \times n} \hspace{1cm} (1)
    \]

  → In the standard formulation, \( \mathbf{W} \) and \( \mathbf{H} \) are found by minimizing:
    \[
    \frac{1}{2} \| \mathbf{V} - \mathbf{W} \mathbf{H} \|_{F}^{2} \quad \text{subject to} \quad \mathbf{W}_{i,j} \geq 0 \quad \forall i, j
    \]
    \[
    \frac{1}{2} \sum_{j} \| \mathbf{v}_{j} - \mathbf{w}_{j} \mathbf{h}_{j} \|_{2}^{2} \hspace{1cm} (2)
    \]

→ A popular scheme alternates between two multiplicative updates:
    \[
    \mathbf{H} \leftarrow \mathbf{H} \odot \frac{\mathbf{W} \mathbf{V}}{\mathbf{W} \mathbf{H} \mathbf{H}^{T}} \quad \mathbf{W} \leftarrow \mathbf{W} \odot \frac{\mathbf{V} \mathbf{H}^{T}}{\mathbf{W} \mathbf{H} \mathbf{H}^{T}} \hspace{1cm} (3)
    \]

→ Applications in sound recognition:
  → \( \mathbf{V} \) is a time-frequency representation of the sound to analyze.
  → The basis vectors \( \mathbf{w} \) contain spectral templates, while the decomposition coefficients \( \mathbf{h} \) represent their successive activations.
  → Numerous off-line applications to polyphonic music transcription.
  → Some on-line applications by employing non-negative decomposition.

Non-negative decomposition with the beta-divergence

→ Beta-divergence:
  → For \( \beta \in \mathbb{R} \) and \( x, y \in \mathbb{R}_{+} \), the β-divergence from \( x \) to \( y \) is defined by:
    \[
    d_{\beta}(x|y) = \frac{1}{\beta - 1} \left( x^{\beta} + (\beta - 1) y^{\beta} - \beta xy^{\beta - 1} \right) \hspace{1cm} (4)
    \]

  → The scaling property is relevant to polyphonic music transcription:
    \[
    d_{\beta}(\lambda x|\lambda y) = \lambda^{\beta} d_{\beta}(x|y) \quad \text{for any} \lambda \in \mathbb{R}_{+} \hspace{1cm} (5)
    \]

→ \( \beta \) can help to reduce octave and harmonic errors since it controls the trade-off between the fundamental, the first, and the higher partials.

→ Problem formulation and multiplicative update:
  → The incoming signal \( \mathbf{V} \in \mathbb{R}_{+}^{m \times n} \) is decomposed onto a dictionary of spectral templates \( \mathbf{W} \in \mathbb{R}_{+}^{m \times r} \) which is kept fixed:
    \[
    \mathbf{v} \approx \mathbf{w}_{h} \quad \text{with} \quad \mathbf{h} \in \mathbb{R}_{+}^{r} \hspace{1cm} (6)
    \]

  → Using the β-divergence as a cost function, \( \mathbf{h} \) is found by minimizing:
    \[
    D_{\beta}(\mathbf{v}|\mathbf{w}_{h}) = \sum_{j} d_{\beta}(\mathbf{v}_{j}|\mathbf{w}_{h_{j}}) \hspace{1cm} (7)
    \]

→ A multiplicative update tailored to real-time is applied iteratively:
    \[
    \mathbf{h} \leftarrow \mathbf{h} \odot \frac{(\mathbf{W} \odot (\mathbf{v} \mathbf{W}^{T})^{\beta - 2} (\mathbf{W} \mathbf{h}^{T})^{\beta - 1}}{\mathbf{W} \mathbf{H}^{T}} \hspace{1cm} (8)
    \]

Conclusion

→ The proposed system can outperform off-line approaches.

→ Perspectives:
  → Employ multi-channel information.
  → Improve the template learning scheme.
  → Consider adaptive and non-stationary templates.

→ Additional resources:
  → http://imr.ircam.fr/imr/Arnaud_Dessein
  → http://imr.ircam.fr/imr/Realtime_Transcription

Table 1: Comparative results for frame and note-level transcription.