Multiplicative Update Rules for Nonnegative Matrix Factorization With Co-occurrence Constraints

Steven K. Tjoa and K. J. Ray Liu

Signals and Information Group, Department of Electrical and Computer Engineering
University of Maryland – College Park, MD, USA

Introduction

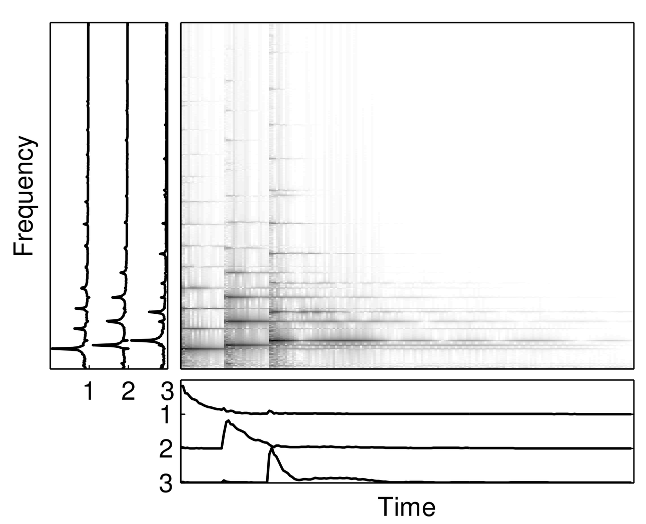
Nonnegative Matrix Factorizaton (NMF) has become a popular tool for decomposing spectrograms of musical signals.

Goal: Given nonnegative $\mathbf{X}$, find two nonnegative matrices:

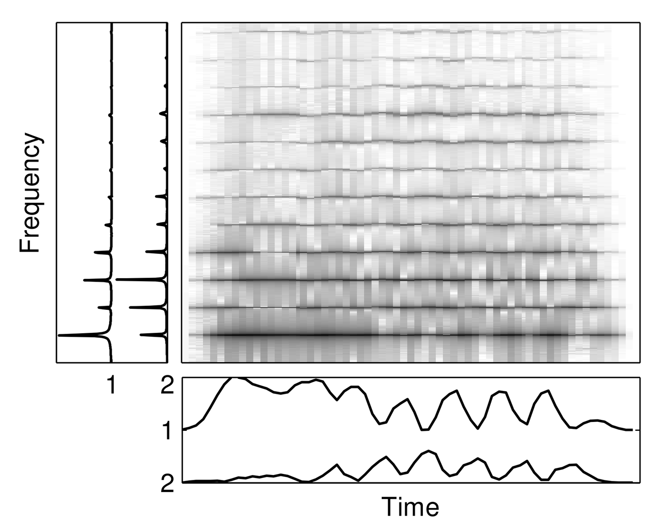
1. dictionary: $\mathbf{A} \in \mathbb{R}_+^{M \times K}$, where $\mathbf{A} = [\mathbf{a}_1\ \mathbf{a}_2\ \ldots\ \mathbf{a}_K]$
2. gain matrix: $\mathbf{S} \in \mathbb{R}_+^{K \times N}$, where $\mathbf{S} = [\mathbf{s}_1\ \mathbf{s}_2\ \ldots\ \mathbf{s}_N]^T$

that minimize some distance between $\mathbf{X}$ and $\mathbf{AS}$.

Piano: three notes



Violin: one note



Common Multiplicative Update Rules

Euclidean distance: $d_{EUC}(x, y) = |x - y|^2$

- $\min_{\mathbf{A},\mathbf{S}} d_{EUC}(\mathbf{X}, \mathbf{AS})$:

$$\mathbf{A} \leftarrow \mathbf{A} \cdot \frac{\mathbf{XS}^T}{\mathbf{ASS}^T} \qquad \mathbf{S} \leftarrow \mathbf{S} \cdot \frac{\mathbf{A}^T\mathbf{X}}{\mathbf{A}^T\mathbf{AS}}$$

Kullback-Leibler divergence: $d_{KL}(x, y) = x \log \frac{x}{y} - x + y$

- $\min_{\mathbf{A},\mathbf{S}} d_{KL}(\mathbf{X}, \mathbf{AS})$:

$$\mathbf{A} \leftarrow \mathbf{A} \cdot \frac{\frac{\mathbf{X}}{\mathbf{AS}}\mathbf{S}^T}{\mathbf{1S}^T} \qquad \mathbf{S} \leftarrow \mathbf{S} \cdot \frac{\mathbf{A}^T\frac{\mathbf{X}}{\mathbf{AS}}}{\mathbf{A}^T\mathbf{1}}$$

where $\mathbf{1}$ is a matrix of ones.

Itakura-Saito divergence: $d_{IS}(x, y) = \frac{x}{y} - \log \frac{x}{y} - 1$

- $\min_{\mathbf{A},\mathbf{S}} d_{IS}(\mathbf{X}, \mathbf{AS})$:

$$\mathbf{A} \leftarrow \mathbf{A} \cdot \frac{\frac{\mathbf{X}}{(\mathbf{AS})^2}\mathbf{S}^T}{\frac{1}{\mathbf{AS}}\mathbf{S}^T} \qquad \mathbf{S} \leftarrow \mathbf{S} \cdot \frac{\mathbf{A}^T\frac{\mathbf{X}}{(\mathbf{AS})^2}}{\mathbf{A}^T\frac{1}{\mathbf{AS}}}$$

Motivation

Problem: Some musical "objects" require more than one atom for accurate representation.

Solution: Impose **co-occurrence constaints** that force certain atoms to be highly correlated or uncorrelated.

Problem Formulation

New minimization problem:

$$\min_{\mathbf{A},\mathbf{S}} d(\mathbf{X}, \mathbf{AS}) + \lambda d(\mathbf{Q}, \mathbf{SS}^T)$$

- $\lambda > 0$ is a regularization parameter,
- $\mathbf{SS}^T$ is a temporal co-occurrence matrix,
- $\mathbf{Q} \in \mathbb{R}_+^{K \times K}$ is a predefined nonnegative symmetric matrix.

How is $\mathbf{Q}$ chosen?

- $q_{ii} = 1$ for all i to normalize each row of $\mathbf{S}$,
- $q_{ij} \approx 1$ for all pairs of atoms i and j that we want to be dependent,
- $q_{ij} \approx 0$ for all other pairs of atoms.

Proposed Multiplicative Update Rules

To minimize $d(\mathbf{Q}, \mathbf{A}^T\mathbf{A})$ or $d(\mathbf{Q}, \mathbf{SS}^T)$,

$$\text{EUC}: \mathbf{A} \leftarrow \mathbf{A} \cdot \frac{\mathbf{AQ} + \varepsilon}{\mathbf{AA}^T\mathbf{A} + \varepsilon} \quad \text{or} \quad \mathbf{S} \leftarrow \mathbf{S} \cdot \frac{\mathbf{QS} + \varepsilon}{\mathbf{SS}^T\mathbf{S} + \varepsilon}$$

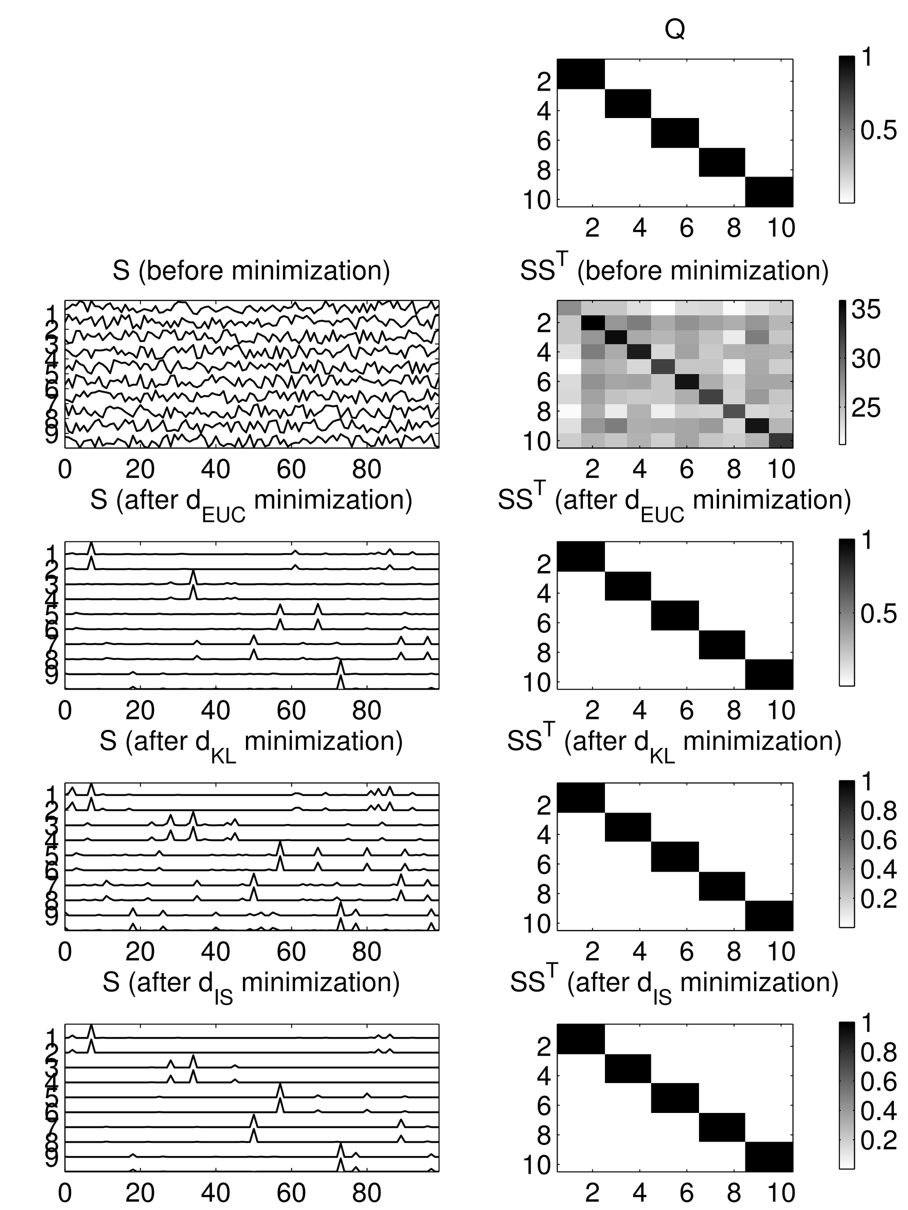
$$\text{KL}: \mathbf{A} \leftarrow \mathbf{A} \cdot \frac{\mathbf{A}\frac{\mathbf{Q}}{\mathbf{A}^T\mathbf{A}} + \varepsilon}{\mathbf{A1} + \varepsilon} \quad \text{or} \quad \mathbf{S} \leftarrow \mathbf{S} \cdot \frac{\frac{\mathbf{Q}}{\mathbf{SS}^T}\mathbf{S} + \varepsilon}{\mathbf{1S} + \varepsilon}$$

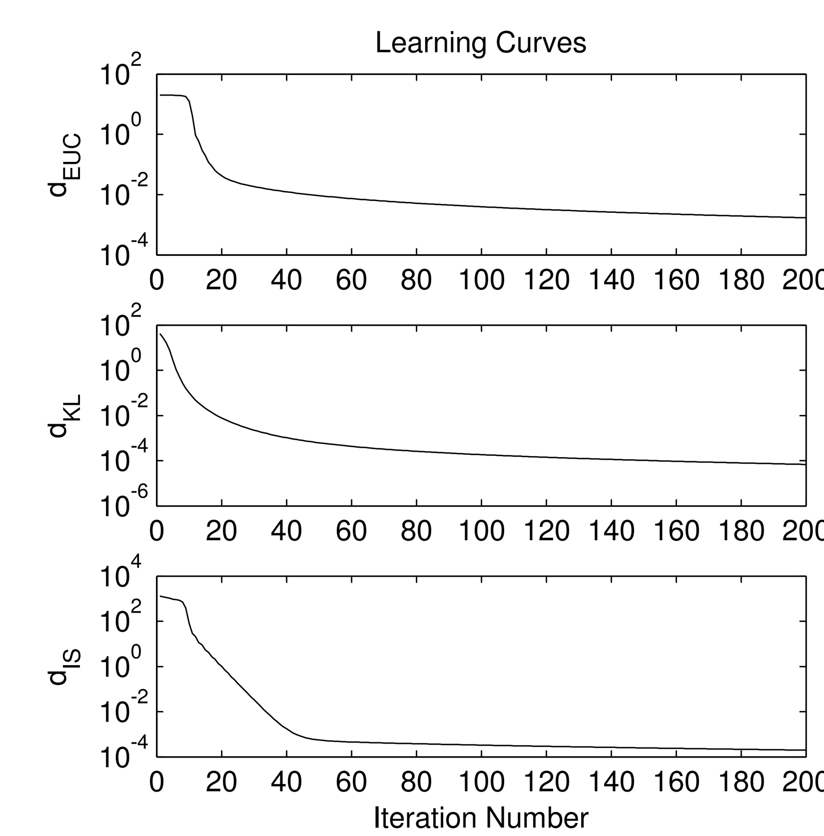
$$\text{IS}: \mathbf{A} \leftarrow \mathbf{A} \cdot \frac{\mathbf{A}\frac{\mathbf{Q}}{(\mathbf{A}^T\mathbf{A})^2} + \varepsilon}{\mathbf{A}\frac{\mathbf{1}}{\mathbf{AA}^T} + \varepsilon} \quad \text{or} \quad \mathbf{S} \leftarrow \mathbf{S} \cdot \frac{\frac{\mathbf{Q}}{(\mathbf{SS}^T)^2}\mathbf{S} + \varepsilon}{\frac{\mathbf{1}}{\mathbf{SS}^T}\mathbf{S} + \varepsilon}$$

To incorporate into initial NMF problem, the update rule changes to (e.g., for Euclidean distance):

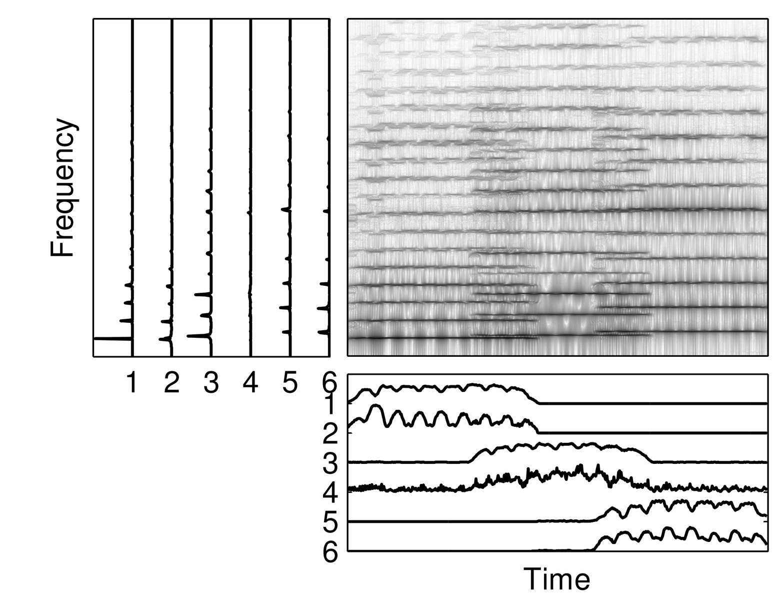
$$\mathbf{S} \leftarrow \mathbf{S} \cdot \frac{\mathbf{A}^T\mathbf{X} + \lambda\mathbf{QS} + \varepsilon}{\mathbf{A}^T\mathbf{AS} + \lambda\mathbf{SS}^T\mathbf{S} + \varepsilon}$$

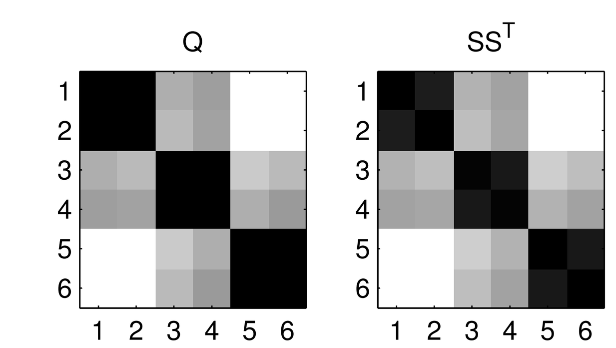
Experiment: Convergence



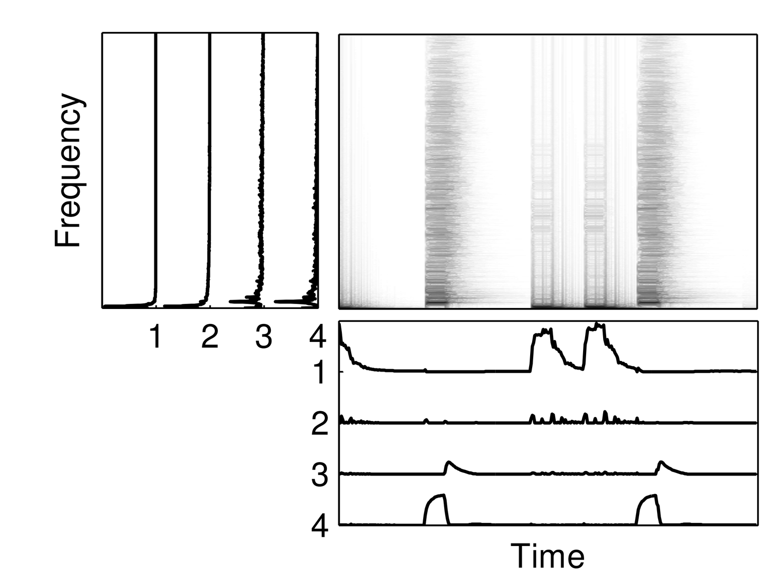


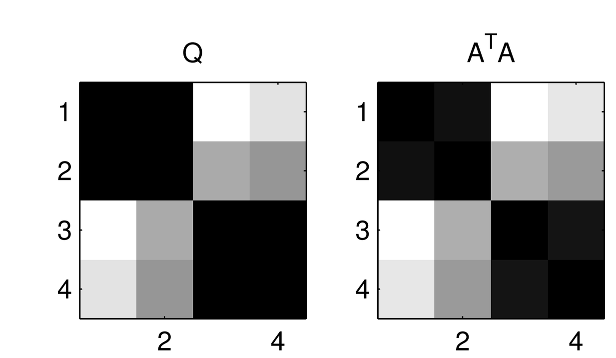
Experiment: Violin





Experiment: Drums





Conclusions

- Co-occurrence constraints successfully enforce dependence upon multiple atoms that describe the same musical object.
- Either temporal or spectral co-occurrence constraints can be enforced successfully.