



Practical work on Non-negative matrix factorization

Roland Badeau

from the subjects written by U. Simsekli, B. David and P. Magron



Contexte académique } **sans modifications**

Voir page 3

Master Sciences et Technologies - Parcours ATIAM - Module TSM



The files necessary for this practical work are available on the ATIAM Moodle. You can load a sound file in Matlab by typing the command `[x,Fs] = wavread('file.wav')`. To listen to it, you can type `soundsc(x,Fs)`. The template of a Python notebook `NMF_TP.ipynb` is also provided.

1 Non-Negative Matrix Factorization with β -divergence

In this practical work, we will deal with non-negative matrix factorization (NMF) with the β -divergence. The problem that we aim to solve is defined as follows :

$$(\mathbf{W}^*, \mathbf{H}^*) = \operatorname{argmin}_{\mathbf{W} \geq 0, \mathbf{H} \geq 0} \sum_{i=1}^I \sum_{j=1}^J d_{\beta}(x_{ij} \| \hat{x}_{ij}), \quad (1)$$

where x_{ij} is an element of $X \in \mathbb{R}_+^{I \times J}$, that is the *non-negative* data matrix, and $\mathbf{W} \in \mathbb{R}_+^{I \times K}$ and $\mathbf{H} \in \mathbb{R}_+^{K \times J}$ are the unknown *non-negative* factor matrices. We also define $\hat{x}_{ij} = \sum_k w_{ik} h_{kj}$. The cost function that we are minimizing is called the β -divergence, which is defined as follows :

$$d_{\beta}(x \| \hat{x}) = \frac{x^{\beta}}{\beta(\beta-1)} - \frac{x \hat{x}^{\beta-1}}{\beta-1} + \frac{\hat{x}^{\beta}}{\beta}. \quad (2)$$

When $\beta = 2$, we obtain the squared Euclidean distance (EUC), when $\beta = 1$, we obtain the Kullback-Leibler (KL) divergence, when $\beta = 0$ we obtain the Itakura-Saito (IS) divergence.

One of the most popular algorithms for NMF is called the multiplicative update rules (MUR). The MUR algorithm has the following update rules :

$$\mathbf{W} \leftarrow \mathbf{W} \circ \frac{(X \circ \hat{X}^{\beta-2}) \mathbf{H}^{\top}}{\hat{X}^{\beta-1} \mathbf{H}^{\top}} \quad (3)$$

$$\mathbf{H} \leftarrow \mathbf{H} \circ \frac{\mathbf{W}^{\top} (X \circ \hat{X}^{\beta-2})}{\mathbf{W}^{\top} \hat{X}^{\beta-1}}, \quad (4)$$

where \circ denotes element-wise multiplication and $/$ and \div denote element-wise division.

By following the technique that we used in the lecture, derive the MUR algorithm by yourselves.

2 Variations on a simple example

The aim here is to program the NMF as described above. For that, we will rely on the script `test_nmf.m` which manages the display part.

2.1 Construction of the simple example

We propose to synthesize the NMF of a simple musical example : a C, followed by an E, followed by a C/E chord, which we will note C-E-C/E.

This example is obtained by constructing the matrix \mathbf{W} of the spectral basis and the matrix \mathbf{H} of the activations in a synthetic way. No sound is calculated here. For this we can for example :

- use the sampling frequency $F_e = 8000$ Hz, the DFT order : $N_{\text{fft}} = 512$.
- calculate the fundamental frequencies F_0 , of C and E located under A 440.

- for each note r , create a (non-negative) harmonic spectrum of the form $\mathbf{W}_r(f) = \sum_{k=1}^K a_k w(f - kF_{0,r})$ where w is a usual window (e.g. Hann of width 80 Hz) and a_k describes the amplitude of the spectral lines (e.g. decreasing, of the form $a_k = \exp(-k/K)$). K is fixed using the Shannon criterion. Deduce the spectral matrix \mathbf{W}_s .
- for the activations, we will use a function of the type $h_r(t) = \sum_p b_{r,p} h(t - p\Delta T)$ where $h(t) = e^{-t/\tau}$ for $t \in [0, \Delta T]$ with ΔT of the order of 0.5 seconds and $h(t) = 0$ elsewhere. b_p is 0 or 1. Deduce the matrix \mathbf{H}_s . Attention, to put ourselves in realistic conditions, we will suppose that the representation is obtained with frames of length $N = N_{\text{fft}}$ and an overlap of 75%, i.e. a sampling rate of $F_e/(N_{\text{fft}}/4)$ for the temporal activations. τ will be selected nearby $\Delta T/3$.
- then construct the time-frequency matrix $\mathbf{X}_s = \mathbf{W}_s \mathbf{H}_s$.

2.2 Standard NMF

- Use the multiplicative update rules to factor the matrix \mathbf{X}_s into a product of non-negative matrices. The matrices \mathbf{W} and \mathbf{H} will be initialized with random values (respecting however the orders of magnitude and the normalization).
- Draw the graph of the cost function as a function of the iteration. What do you notice?

2.3 Variations

Test your algorithm by varying the data and parameters :

- by repeating with several draws of initialization (what differences do you notice on the results ?)
- with addition of noise (be careful to respect the non-negativity, use `rand`)
- with different values of β (0, 1 and 2 typically)
- with more notes and indeterminations like C/E, C/G, C/E/G
- by making synthetic notes (waveforms) as sum of sinusoids and by calculating the STFT of the obtained temporal signal.

3 Automatic transcription using semi-supervised NMF

The isolated notes of a piano are provided, as well as a piece of music played with the same instrument. The semi-supervised NMF consists in defining the spectral matrix \mathbf{W} using the isolated notes (provided in the form of `xxx.wav` where `xxx` is the MIDI number, 21-108 for the tessitura of the piano), by simply calculating an estimate of the associated power spectral density. The only remaining update is therefore that of \mathbf{H} . We will work on signals sub-sampled at 22 kHz.

The provided script `build_signal.m` allows, with the help of a simple syntax, to build a piece from notes defined by their MIDI number, their attack time, their duration and the associated velocity.

1. Build the basis \mathbf{W} corresponding to the 88 notes of the piano.
2. Synthesize a test signal corresponding to the C3-C4 scale.
3. Program the semi-supervised NMF using the results of the previous part and apply it to the test signal. Observe the temporal activations obtained.
4. Program an attack detection function to be applied to the activations in order to estimate the times when the notes are played.



Contexte académique } **sans modifications**

Par le téléchargement ou la consultation de ce document, l'utilisateur accepte la licence d'utilisation qui y est attachée, telle que détaillée dans les dispositions suivantes, et s'engage à la respecter intégralement.

La licence confère à l'utilisateur un droit d'usage sur le document consulté ou téléchargé, totalement ou en partie, dans les conditions définies ci-après, et à l'exclusion de toute utilisation commerciale.

Le droit d'usage défini par la licence autorise un usage dans un cadre académique, par un utilisateur donnant des cours dans un établissement d'enseignement secondaire ou supérieur et à l'exclusion expresse des formations commerciales et notamment de formation continue. Ce droit comprend :

- le droit de reproduire tout ou partie du document sur support informatique ou papier,
- le droit de diffuser tout ou partie du document à destination des élèves ou étudiants.

Aucune modification du document dans son contenu, sa forme ou sa présentation n'est autorisée.

Les mentions relatives à la source du document et/ou à son auteur doivent être conservées dans leur intégralité.

Le droit d'usage défini par la licence est personnel et non exclusif. Tout autre usage que ceux prévus par la licence est soumis à autorisation préalable et expresse de l'auteur : sitepedago@telecom-paristech.fr

