MASTER ATIAM (Année 2022-2023)
Examen de Traitement du Signal Musical (UE TSM)
Partie *Applications et ouvertures*

Geoffroy Peeters, Roland Badeau

Mardi 3 janvier 2023

RENDRE DEUX COPIES: l'une pour Geoffroy Peeters (exercice 1)
et l'autre pour Roland Badeau (exercices 2 et 3)

Les documents ne sont pas autorisés.

1 Cours "Estimation de structure" et "Estimation multi-pitch"

Question 1
a) En quoi le cepstre réel peut-il être rapproché de l'auto-corrélation ? Montrez-le mathématiquement.
b) En quoi le cepstre réel permet-il de séparer la contribution du pitch et de l'enveloppe spectrale ? Montrez-le mathématiquement.
Question 2
a) Qu'est-ce que la méthode du produit spectral pour estimer la fréquence fondamentale ? Donnez son expression mathématique.
b) Quelle est sa principale hypothèse/limitation ?
Question 3
a) Expliquer l'algorithme itératif d'estimation de fréquences fondamentales dit de Klapuri. Qu'appelle-t-on le "lissage spectral" ?
b) Donner et expliquer un exemple de l'utilisation de la méthode NMF pour l'estimation de hauteurs fondamentales multiples.
Question 4
a) En utilisant d'une part la formule donnant la largeur fréquentielle à -3dB du lobe principal d'une fenêtre d'analyse, et d'autre part celle transformant les pitchs MIDI en fréquences en Hz, établissez l'inégalité permettant de déterminer la longueur temporelle de la fenêtre en fonction de celles des pitchs adjacents.
b) Pourquoi la représentation de type chroma est-elle sensible au timbre des instruments ?

2 Cours "NMF" : *Harmonicity and spectral smoothness*

When applying NMF to music, the spectral templates in the $F \times K$ matrix $\mathbf{W}$ are expected to represent individual notes, thus to be formed of harmonic partials with smooth spectral envelope. One way to enforce both harmonicity and spectral smoothness is to model each spectral template w_{fk} as a nonnegative linear combination of M narrowband, harmonic spectral patterns $g_{fk,m}$:

$$w_{fk} = \sum_{m=1}^{M} g_{fk,m} e_{mk} \tag{1}$$

where all spectral patterns $\{g_{fk,m}\}_m$ share the same fundamental frequency $\nu_k^0 > 0$, have smooth spectral envelopes and different spectral centroids, so as to form a filterbank-like decomposition of the whole spectrum, and $\{e_{mk}\}_m$ are the nonnegative coefficients of this decomposition. In this way, it is guaranteed that w_{fk} is a harmonic spectrum of fundamental frequency ν_k^0, with a smooth spectral envelope. If the signal of interest is a music signal, then the order K and the fundamental frequencies ν_k^0 can be preset according to the semitone scale.

When decomposing an observed music spectrogram $\mathbf{V}$ of size $F \times N$, since all spectral patterns $g_{fk,m}$ in the $F \times K$ matrix $\mathbf{G}_m$ are preset, we just have to estimate the coefficients e_{mk} in the $M \times K$ matrix $\mathbf{E}$ and the activations h_{kn} in the $K \times N$ matrix $\mathbf{H}$, so as to get the factorization $\mathbf{V} \approx \widehat{\mathbf{V}} = \mathbf{WH}$. To do so, we consider the following optimization problem:

$$(\mathbf{E}^\star, \mathbf{H}^\star) = \operatorname*{arg\,min}_{\mathbf{E}\geq 0, \mathbf{H}\geq 0} \mathcal{C}(\mathbf{V} \mid \widehat{\mathbf{V}}) \text{ with } \mathcal{C}(\mathbf{V} \mid \widehat{\mathbf{V}}) = \sum_{f=1}^{F}\sum_{n=1}^{N} d_\beta(v_{fn}; \hat{v}_{fn}), \tag{2}$$

where $\hat{v}_{fn} = \sum_{k=1}^{K} w_{fk} h_{kn}$ and $d_\beta(\cdot;\cdot)$ is the β-divergence, defined as follows for $\beta \in \mathbb{R}_+ \backslash \{0, 1\}$:

$$d_\beta(v; \hat{v}) = \frac{1}{\beta(\beta-1)}\Big(v^\beta + (\beta-1)\hat{v}^\beta - \beta v \hat{v}^{\beta-1}\Big). \tag{3}$$

Question 1 We remind that the update rule of $\mathbf{H}$ is derived by decomposing the gradient of the cost function $\mathcal{C}(\mathbf{V} \mid \widehat{\mathbf{V}})$ as the difference of two nonnegative terms: $\frac{\partial}{\partial h_{kn}}\mathcal{C}(\mathbf{V} \mid \widehat{\mathbf{V}}) = p_{k,n} - q_{k,n}$ where $p_{k,n} \geq 0$ and $q_{k,n} \geq 0$. Then matrix $\mathbf{H}$ can be updated as $h_{kn} \leftarrow h_{kn}\left(\frac{q_{k,n}}{p_{k,n}}\right)^\eta$, where $\eta > 0$ is a stepsize like in a gradient descent. Derive the multiplicative update rules for:
a) matrix $\mathbf{H}$, as a function of $\mathbf{V}$, $\widehat{\mathbf{V}}$ and $\mathbf{W}$,
b) the m-th row of matrix $\mathbf{E}$, as a function of $\mathbf{V}$, $\widehat{\mathbf{V}}$, $\mathbf{H}$ and $\mathbf{G}_m$.
Question 2 Provide short answers for the following questions. Let us assume that we obtain the optimal factors, $\mathbf{E}^\star$ and $\mathbf{H}^\star$. How can we use them for:
a) audio source separation?
b) automatic music transcription?

3 Cours "Méthodes à haute résolution"

Les méthodes à haute résolution vues en cours s'appuient sur le modèle d'observation $x(t) = s(t) + b(t)$, où $s(t) = \sum_{k=0}^{K-1} \alpha_k z_k^t$ (avec $\alpha_k \in \mathbb{C}^*$, $z_k \in \mathbb{C}^*$ et tous les pôles z_k sont distincts deux à deux), et où $b(t)$ est un bruit blanc. En pratique, cette hypothèse de bruit blanc est inadaptée aux signaux audiofréquences, qui sont généralement caractérisés par une forte dynamique spectrale. En toute rigueur, il faudrait donc modéliser $b(t)$ comme un processus stationnaire dont la densité spectrale de puissance n'est pas constante. Nous proposons ici de modéliser $b(t)$ comme un processus à moyenne ajustée (MA) d'ordre M: $b(t)$ est de moyenne nulle et sa fonction d'autocovariance $r_{bb}(m)$ est nulle quand $|m| > M$.
Question 1 On définit le polynôme $P(z) = \prod_{k=0}^{K-1}(z - z_k) = \sum_{\tau=0}^{K} p_\tau\, z^{K-\tau}$ (où $p_0 = 1$) et les vecteurs colonnes $\boldsymbol{p} = [1, p_1, \ldots, p_K]^T$ et $\boldsymbol{x}(t) = [x(t), x(t-1), \ldots, x(t-K)]^T$ (où le symbole ^{T} désigne le transposé). Vérifier que $\forall t$,

$$\boldsymbol{p}^T \boldsymbol{x}(t) = b(t) + \sum_{\tau=1}^{K} p_\tau\, b(t-\tau). \tag{4}$$

Question 2 En multipliant l'équation (4) à droite par $\boldsymbol{x}(t-K-M-1)^H$ (où le symbole ^{H} désigne le transposé conjugué), démontrer l'égalité $\boldsymbol{p}^T \boldsymbol{R} = 0$, où la matrice de corrélation $\boldsymbol{R}$ est définie par $\boldsymbol{R} = \mathbb{E}\left[\boldsymbol{x}(t)\, \boldsymbol{x}(t-K-M-1)^H\right]$.
Question 3 Proposer une méthode d'estimation des z_k reposant sur les résultats des questions précédentes.