Optimal spectral transportation with application to music transcription

Ghassan Najjar, Ruben Illouz

Ecole Polytechnique

Feburary 8, 2022

Overview

1. Problem presentation
2. Solution
3. Results

The unmixing problem

- Input: vector v that represents the frequency spectrum of a signal s ($v_i = FT(s)(f_i)$) $v \in \mathbb{R}^M$ normalized to 1.
- Dictionary : matrix W where each column is a template spectrum, $W \in \mathbb{R}^{M \times K}$ (dictionary of K templates normalized to 1)
- Goal : find a vector $h \in \mathbb{R}^K$ where $Wh \approx v$
- But which meaning to give to "$\approx$" ?

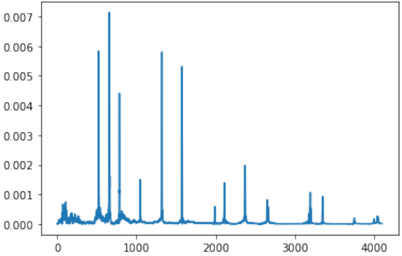


Figure: **Spectrum**

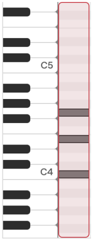


Figure: **Played notes**

First meaning : approximation with the meaning of KL

Approximation with the meaning of Kullback-Leibler

$$\min_{\mathbf{H}\geq 0} D_{\mathrm{KL}}(\mathbf{v} \mid \mathbf{W}\mathbf{h}) \quad \text{s.t} \quad \|\mathbf{h}\|_1 = 1,$$

avec $D_{\mathrm{KL}}(\mathbf{v} \mid \hat{\mathbf{v}}) = \sum_i v_i \log\left(v_i/\hat{v}_i\right)$

- Disproportionate recognition of small errors

Second Meaning: approximation with the meaning of the OT

Wasserstein distance

Let C a cost matrix.

$$J(\mathbf{T} \mid \mathbf{v}, \hat{\mathbf{v}}, \mathbf{C}) \stackrel{\text{def}}{=} \sum_{ij} c_{ij} t_{ij}$$

$$D_{\mathbf{C}}(\mathbf{V} \mid \mathbf{Wh}) \stackrel{\text{def}}{=} \min_{\mathbf{T}} J(\mathbf{T} \mid \mathbf{v}, \hat{\mathbf{v}}, \mathbf{C})$$

sous contrainte $\forall i, j = 1, \ldots, N, \sum_{j=1}^{M} t_{ij} = v_i, \sum_{i=1}^{M} t_{ij} = \hat{v}_j$

Approximation with the meaning of Wasserstein

$$\min_{\mathbf{h} \geq 0} D_{\mathbf{C}}(\mathbf{V} \mid \mathbf{Wh}) \quad \text{s.t} \quad \|\mathbf{h}\|_1 = 1,$$

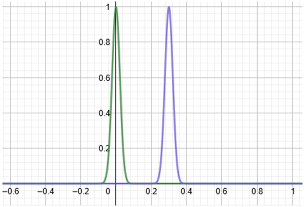
Second Meaning : Approximation with the meaning of l'OT

Approximation au sens de Wasserstein

$$\min_{\mathbf{h} \geq 0} D_{\mathbf{C}}(\mathbf{V} \mid \mathbf{Wh}) \quad \text{s.t} \quad \|\mathbf{h}\|_1 = 1,$$

Par exemple avec $C_{i,j} = (f_i - f_j)^2$

- No penalty for small mistakes



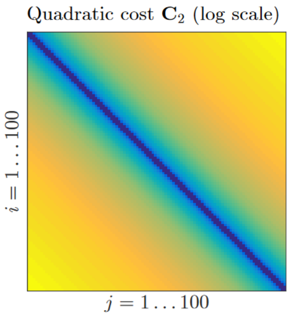
- Disadvantage: requires to have in W all fundamental+harmonic configurations

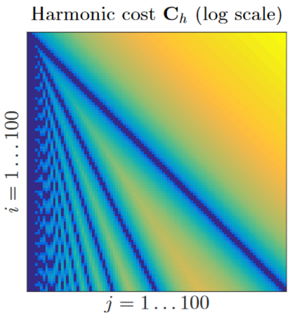
Construction of C

Approximation au sens de Wasserstein

$$C_{ij} = \min_{q=1,\dots,q_{\text{max}}} (f_i - q f_j)^2 + \epsilon \delta_{q \neq 1}$$

- Harmonic-invariant cost
- We can take diracs for the columns of W





Solution 1

Optimal transport problem

$$\min_{\mathbf{h}\geq 0, \mathbf{T}\geq 0} \langle \mathbf{T}, \mathbf{C} \rangle = \sum_{ij} t_{ij} c_{ij} \quad \text{s.t.} \quad \mathbf{T}\mathbf{1}_M = \mathbf{v}, \quad \mathbf{T}^\top \mathbf{1}_M = \mathbf{W}\mathbf{h},$$

Optimisation of the problelm

$$W \in \mathbb{R}^{M\times K} \text{ et } K << M$$

W is composed of K diracs of frequencies $(f_i)_{i\in S}$

$$\min_{\mathbf{h}\geq 0, \widetilde{\mathbf{T}}\geq 0} \langle \widetilde{\mathbf{T}}, \widetilde{\mathbf{C}} \rangle \quad \text{s.t.} \quad \widetilde{\mathbf{T}}\mathbf{1}_K = \mathbf{v}, \quad \widetilde{\mathbf{T}}^\top \mathbf{1}_M = \mathbf{h}.$$

Solution 1

Reformulation of the problem

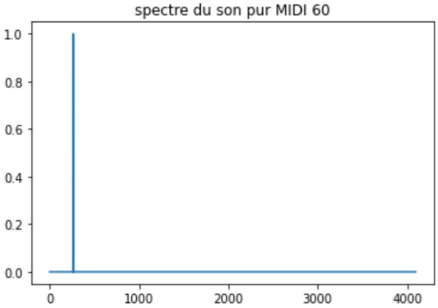
$$\min_{\tilde{t}_i \geq 0} \sum_k \tilde{t}_{ik} \tilde{c}_{ik} \quad \text{s.t.} \quad \sum_k \tilde{t}_{ik} = v_i.$$

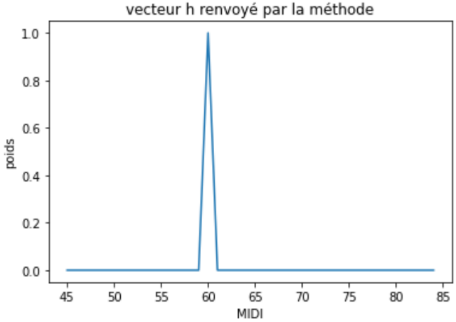
Solution 1

Let's define $k_i^\star = \arg\min_k \{\tilde{c}_{ik}\}$ by defining $\mathbf{L} = (\delta_{(i,k_i^\star)}(i,j))_{(i,j)}$, we get the solu tion h given by:

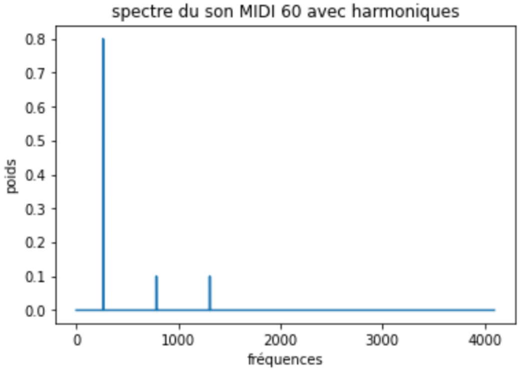
$$\hat{\mathbf{h}} = \mathbf{L}^\top \mathbf{v}$$

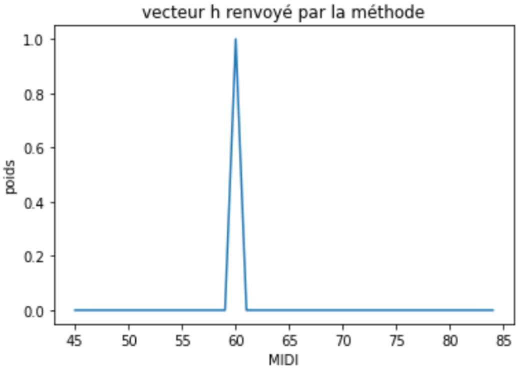
Implementation of solution 1: case of pure sound





Implementation of solution 1: presence of harmonics





Solution 2: Entropic regularisation

Optimal transport problem with entropy regularisation

$$\min_{\mathbf{h}\geq 0, \widetilde{\mathbf{T}}\geq 0} \langle \widetilde{\mathbf{T}}, \widetilde{\mathbf{C}} \rangle + \lambda_e \Omega_e(\widetilde{\mathbf{T}}) \quad \text{s.t.} \quad \widetilde{\mathbf{T}}\mathbf{1}_K = \mathbf{v}, \quad \widetilde{\mathbf{T}}^\top \mathbf{1}_M = \mathbf{h}$$

avec $\Omega_e(\widetilde{\mathbf{T}}) = \sum_{ik} \tilde{t}_{ik} \log(\tilde{t}_{ik})$

- This regulation promotes the transport of v_i to several destinations

Expression of the solution

$$\hat{\mathbf{h}} = \mathbf{L}_e^\top \mathbf{v}$$

avec $\mathbf{L_e} = (\exp(-\tilde{c}_{ik}/\lambda_e) / \sum_p \exp(-\tilde{c}_{ip}/\lambda_e))_{i,k}$

Solution 3 : group regularisation

Optimal transport problem with group regularisation

$$\min_{\mathbf{h}\geq 0, \widetilde{\mathbf{T}}\geq 0} \langle \widetilde{\mathbf{T}}, \widetilde{\mathbf{C}} \rangle + \lambda_g \Omega_g(\widetilde{\mathbf{T}}) \quad \text{s.t.} \quad \widetilde{\mathbf{T}}\mathbf{1}_K = \mathbf{v}, \quad \widetilde{\mathbf{T}}^\top \mathbf{1}_M = \mathbf{h}$$

avec $\Omega_g(\widetilde{\mathbf{T}}) = \sum_k \sqrt{\left\|\widetilde{\mathbf{t}}_k\right\|_1}$

- Cette régularisation favorise la régularisation group-sparse (force certaines colonnes de $\widetilde{\mathbf{T}}$ à être nulles).

Solution 3 : group regularisation

Expression of the solution

There is no explicit solution for this problem, but one can reason iteratively until convergence occurs, following the following algorithm:

Require: Sample $\mathbf{v}$, transportation cost matrix $\widetilde{\mathbf{C}}$, hyper-parameter λ_g.
Initialise $\widetilde{\mathbf{R}}^{(0)}$ with zeros, set $iter = 0$
repeat
    $iter = iter + 1$
    Compute $\widetilde{\mathbf{C}}^{(iter)} = \widetilde{\mathbf{C}} + \widetilde{\mathbf{R}}^{(\overset{iter-1}{\cancel{iter}})}$ with $\widetilde{\mathbf{R}}^{(iter)}$ computed with Eq. ($\boxed{8}$)
    Compute $\mathbf{L}_g^{(iter)}$ with Algorithm $\boxed{1}$ applied to $\widetilde{\mathbf{C}}^{(iter)}$
    Compute $\widetilde{\mathbf{T}}^{(iter)} = \operatorname{diag}(\mathbf{v})\mathbf{L}_g^{(iter)}$ and $\mathbf{h}^{(iter)} = \mathbf{L}_g^{(iter)\top}\mathbf{v}$
until convergence

- The equation (8) is $\widetilde{r}_{ik}^{(iter)} = \frac{1}{2}\left\|\widetilde{\mathbf{t}}_k^{(iter)}\right\|_1^{-\frac{1}{2}}$.

Comparison of the methods

- Choosing a chord on the piano: C5/E5/G5

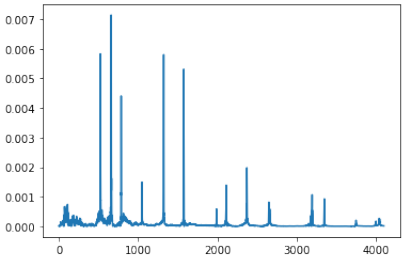


Figure: **Spectrum of the chord**

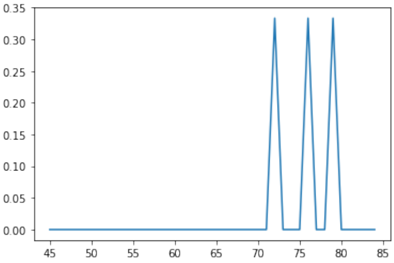


Figure: **Midis of the chord**

Comparison of the methods

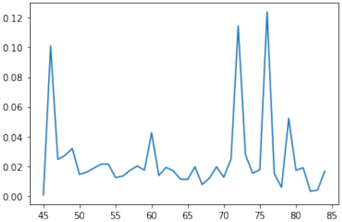


Figure: Méthode OST

l_1 error: **1.42**
Time (ms): 1.25

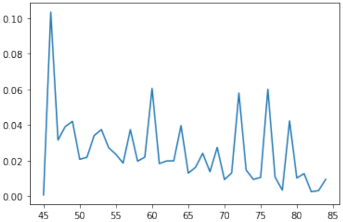


Figure: Méthode OST_e

l_1 error: 1.67
Time (ms): **0.67**

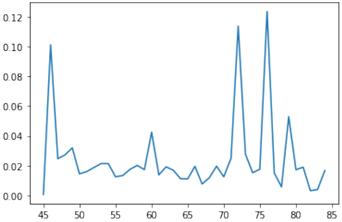


Figure: Méthode OST_g

l_1 error: **1.42**
Time (ms): 2 070

Conclusion

- The OST method is more efficient and faster than KL: the use of optimal transport is justified
- OSTE and OSTG are not efficient on this simple example of a chord
- In reality Deep Learning is more efficient than any OST method or KL one: but training data is needed