école
normale
supérieure
paris—saclay

Review of the paper : ”Optimal spectral transportation with application to music transcription”

Project for the course ”Computational Optimal Transport”

Lucas Haubert

Master MVA

ENS Paris-Saclay

January 19, 2024

The problem : Robustness of Music Transcription

Definition : Music transcription aims to ”translate” a raw musical signal according to its composition (notes)

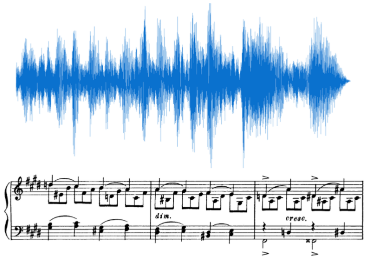

Figure: Music transcription on Prelude in C# minor, Op.3 No.2, S. Rachmaninoff

The idea : Approximate an input spectrogram V by a product $\hat{V} = WH$, where W is a dictionary of representation of notes, and H is the optimal allocation of notes. $V = \left(v_1, \ldots, v_n\right)$.

A relevant baseline : PLCA

Naive frequency-wise comparison : Probabilistic Latent Component Analysis (PLCA)

- The process :
  - Normalize columns v_n's of V (discrete probability distribution) ; same requirements for W and H
  - Comparison of V and $\hat{V}$ (or v_n and $\hat{v}_n$) ; $\min D_{KL}\left(V|\hat{V}\right)$ s.t. $\forall n, \left\|h_n\right\|_1 = 1$ over $H \geq 0$
  - Use of KL divergence : $D_{KL}\left(v_n|\hat{v}_n\right) = \sum_{i=1}^{M} v_i \log\left(v_i/\hat{v}_i\right)$ and $D_{KL}\left(V|\hat{V}\right) = \sum_{n=1}^{N} D_{KL}\left(v_n|\hat{v}_n\right)$
- Analysis :
  - Pros of KL : Tool for comparing distributions $=>$ notion of """"distance"""" (not a metric)
  - BUT : Separability of KL $=>$ frequency-wise comparison between v_n and $\hat{v}_n$
  - Lack of robustness to small displacements in the frequency support (disproportional changes of KL)
  - Use of other methods, such as OT-based approaches, to fix it (see content of [1])

OT-based approaches

Approach of the studied article [1] : "Optimal spectral transportation with application to music transcription"

- The idea :
  - Columns v_n's are energy distributions of intensities $v_{n1}, \ldots, v_{nM}$ at sampling frequencies $f_1, \ldots, f_M$
  - Use of a transportation matrix in $\{T \in \mathbb{R}^{M \times M} | \forall i, j \in [\![0, M]\!], \sum_{j=1}^{M} t_{i,j} = v_{ni}, \sum_{j=1}^{M} t_{i,j} = \hat{v}_{nj}\}$
  - Use of a cost matrix C to quantify the distance between uncorrelated frequencies
  - Define the "OT divergence" : $D_C(V|\hat{V}) = \sum_n D_C(v_n|\hat{v}_n) = \sum_n \min_T \sum_{i,j} c_{i,j} t_{i,j}$
  - Solve $\min D_C(V|\hat{V})$ s.t. $\forall n, \|h_n\|_1 = 1$ ($H \geq 0$)
- Analysis :
  - $D_C(\cdot|\cdot)$ is an OT-based measure that can handle robustness issues with C
  - Define C_h with $c_{i,j} = \min_{q=1,\ldots,q_{max}} (f_i - qf_j)^2 + \epsilon\delta_{q\neq1}$ for inharmonicities and variation of timbre
  - Define W as a set of Dirac vectors placed at the fundamental frequencies $\nu_1, \ldots, \nu_K$ of the notes to identify
  - This method can be extended to regularized ones by means of additional assumptions

Optimization and regularization

Improvement of classical OST : Analysis and regularization

- Analysis :
  - Naive OT unmixing : $\min_{h_n \geq 0, T \geq 0} < T, C >$ s.t. $T \in \Theta$: LP (computationally heavy)
  - With W set of Diracs and $K < M$, sparse transportation matrix $\tilde{T}$ related to fundamentals $\nu_1, ..., \nu_K$ and $\hat{h}_n = L^T v_n$ where L is sparse $=>$ computationally efficient algorithm
  - Define $\tilde{c}_{i,k} = \min_q (f_i - q\nu_k)^2 + \epsilon \delta_{q \neq 1} => \tilde{T}$ deals with the fundamentals directly
  - Opportunity to keep this method or to settle regularization(s)
- Regularizations :
  - Entropic regularization : $\min_{h_n \geq 0, T \geq 0} < \tilde{T}, \tilde{C} > + \lambda_e \Omega_e(\tilde{T})$ s.t $T \in \Theta$; $\Omega_e(\tilde{T}) = \sum_{i,k} \tilde{t}_{ik} \log(\tilde{t}_{ik})$
    Distribute energies a bit more in order to get a smoother estimate of transport $\tilde{T}$; closed-form solution
  - Group regularization : Add $\lambda_g \Omega_g(\tilde{T}) = \lambda_g \sum_k \sqrt{[\![\tilde{t}_k]\!]_1}$
    Use the group structure of $\tilde{T}$; solution obtained by iterations of classic OST with additive term on $\tilde{C}$
  - Entropic + Group regularization : Use both methods at the same time

Experiment 1 : Synthetic data

Synthetic harmonic templates : Setup and results

- Setup :
  - Generate a synthetic dictionary of template (12 here)
  - extract two templates to generate a new one, and apply a small shift on fundamental frequencies
  - Compute the l_1 error between h_n and $\hat{h}_n$ for each method

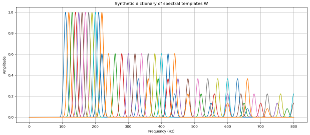

Figure: 12 simulated harmonic templates

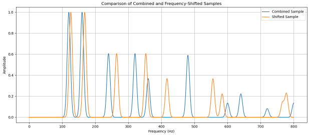

Figure: Small shift of fundamentals

- Results :
  - Group regularization approach is the best (due to sparse context of these synthesis templates)
  - Same conclusion for variation of timbre

Experiment 2 : Real-world data

Piano recordings from dataset MAPS : Setup and results

- Setup :
  - Produce a spectrogram V by means of Hann windows and overlap
  - Real-world music instrument can produce time varying signals, due to their physical composition
  - Comparison tools based on MIDI system

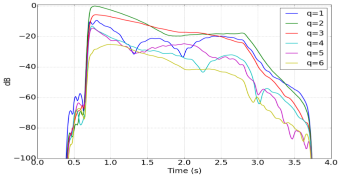

Figure: Harmonics of a single piano note along time

- Results :
  - Entropic regularization performs better in this situation (due to a better distribution of transportation)

Conclusion and opening

Overview of the study of article [1] : ”Optimal spectral transportation with application to music transcription”

- Problem and solutions :
  - Naive measure of fit : $\min D_{KL}(V|\hat{V}) =>$ frequency-wise comparison $=>$ lack of robustness
  - PLCA have been widely used to perform MT, but is no longer state-of-the-art
  - OT-based approaches : Leverage OT techniques to proceed a global comparison
  - Entropic regularization good on real-world musical data ; Group regularization efficient on sparse profiles
- Possible openings / extensions :
  - Quadratic regularization $=>$ maintains a sparse transport plan
  - Unbalanced optimal transport $=>$ better deal with outliers and noise
  - Deep Learning architectures such as LSTMs or Transformers to process the audio tracks

Thank You for Listening

Ready for Q&A session

Reference

N. Courty R. Flamary, C. Févotte and V. Emiya.
Optimal spectral transportation with application to music transcription, 2016.