CSCI-B659 Machine Perception and Audition (Fall 2016)

A SCORE-INFORMED SHIFT-INVARIANT EXTENSION OF COMPLEX MATRIX FACTORIZATION FOR IMPROVING THE SEPARATION OF OVERLAPPED PARTIALS IN MUSIC RECORDINGS

Francisco J. Rodriguez-Serrano, Sebastian Ewert, Pedro Vera Candeas , Mark Sandler

Presented By: Roshan Kathawate

Overview

- Introduction
- Non-Negative Matrix Factorization and Variants
- Score-Informed Shift-Invariant CMF
- Experiments
- Conclusion

Introduction

- The decomposition of a given music recording into its constituent parts, a task also known as source separation, is one of the central topics in music information retrieval and processing. Possible applications range from stereo-to-surround up-mixing, remixing tools for DJs or producers to instrument-wise equalizing or karaoke systems.

- In this paper we show that score information provides a source of prior knowledge rich enough to stabilize the CMF parameter estimation, without sacrificing its expressive power.

- We present a shift-invariant extension to CMF bringing the vibrato-modeling capabilities of NMF to CMF.

Basic framework for Source Separation

1. STFT
2. NMF
3. FILTER
4. ISTFT

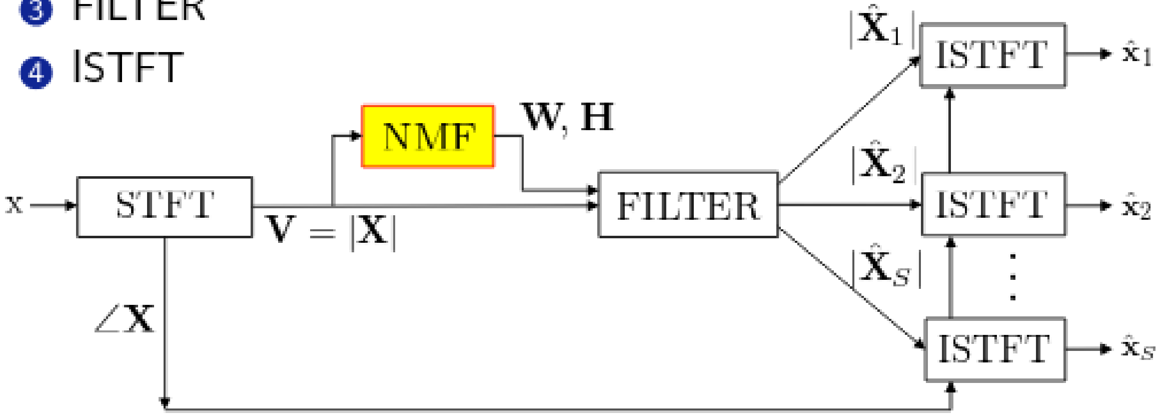


Image source: https://ccrma.stanford.edu/~njb/teaching/sstutorial/part2.pdf

Non Negative Matrix Factorization and Variants

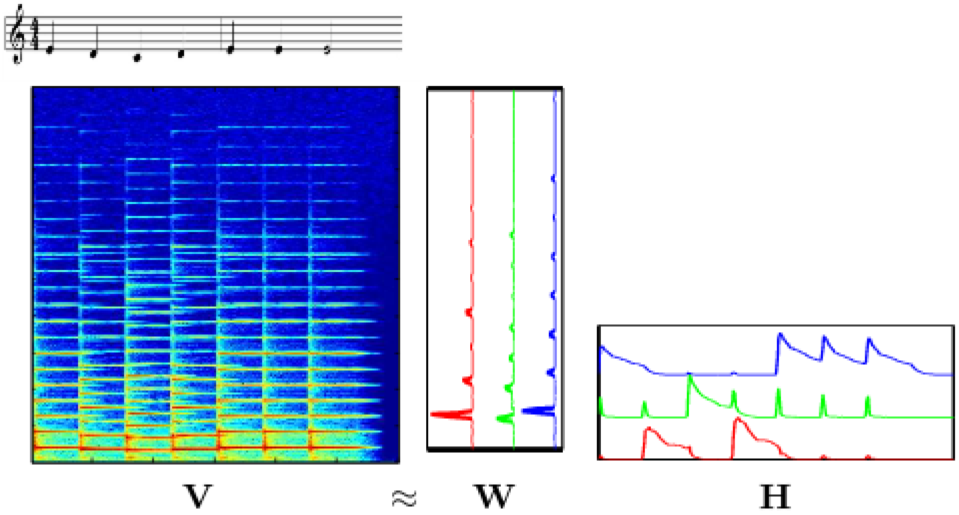


Figure : NMF of *Mary Had a Little Lamb* with $K = 3$ play stop

- The columns of W are often referred to as template vectors and provides information about the spectral energy distribution of a sound source
- The rows of H as the corresponding activations and encode when and how intense a source is active.

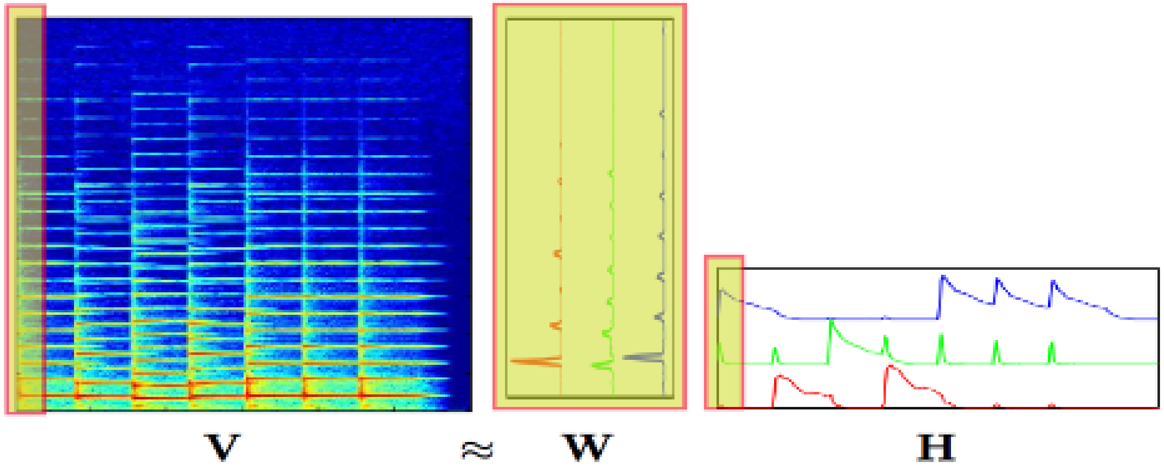
Cont...

Image source: https://ccrma.stanford.edu/~njb/teaching/sstutorial/part2.pdf

Non Negative Matrix Factorization and Variants

- Interpretation

Columns of $\mathbf{V} \approx$ as a weighted sum (mixture) of basis vectors



$$\begin{bmatrix} | & | & & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \ldots & \mathbf{v}_T \\ | & | & & | \end{bmatrix} \approx \begin{bmatrix} \sum_{j=1}^{K} \mathbf{H}_{j1}\,\mathbf{w}_j & \sum_{j=1}^{K} \mathbf{H}_{j2}\,\mathbf{w}_j & \ldots & \sum_{j=1}^{K} \mathbf{H}_{jT}\,\mathbf{w}_j \end{bmatrix}$$

Image source: https://ccrma.stanford.edu/~njb/teaching/sstutorial/part2.pdf

Shift-Invariant NMF

- By allowing the templates we learn using NMF to be shifted along the frequency axis, we obtain a slightly extended version of NMF often referred to as shift-invariant NMF.
- we choose a number of possible shifts S > 0 we want to consider, and approximate V component-wise by
  V (m, n) = $\sum_{k=1}^{K} \sum_{s=1}^{S} W(m - s, k) H_s(k, n),$
    - where each Hs contains the activations for shift s.
- If a log-distributed frequency scale is used in V , the shift corresponds to slight changes of the fundamental frequency for a template and thus enables accounting for vibrato or tuning differences without requiring an excessive amount of template vectors for a given musical pitch

Score-based constraints with NMF

- To incorporate score information into the NMF factorization process, various approaches have been proposed
  - Most of them employ parametric signal models, where the templates (and sometimes the activations) are described using a few meaningful parameters which can be associated with and constrained using score information but limits expressivity of the model
- Given a musical pitch, it is possible to roughly estimate the fundamental frequency and the harmonics and one can set entries in a template to zero that are not in a neighborhood of these frequencies and thus can be expected to be zero.
  - However, it is yet unclear if the score information might even allow for increasing the modelling detail even further.
  - It is one goal of this paper to find out if it is rich enough to support a CMF-based signal model which has many more parameters than an NMF model.

NMF Vs CMF

- CMF introduces introduction of a phase matrix for each source.
- The complex time-frequency representation S is approximated using CMF as

$$S \approx \sum_{k=1}^{K} W(m,k)H(k,n)e^{i\phi_k(m,n)}.$$

- As we can see, compared to the K(M + N) parameters used in NMF, CMF uses KM N additional phase parameters which is typically several times higher.
- CMF models the complex spectrogram, which in contrast to its magnitude is truly additive w.r.t. the individual sources, enabling CMF to account for phase cancellations between overlapping sources, which can potentially increase the separation performance.

Score-Informed Shift-Invariant CMF

- We integrate both the idea of shift-invariance and the idea of score-based constraints into complex matrix factorization.

a data fidelity term in the form of the square of the Frobenius norm between the given time-frequency representation S and our model

A regularization for the phase. The phase in frame n should not deviate much from the phase in frame n − 1 after being advanced using the fundamental frequency for a certain pitch.

$$\begin{aligned}
&\sum_{m,n} \left| S(m,n) - \sum_{q,p,s} \overline{W}_q(m-s,p) H_{q,s}(p,n) e^{i\phi_q(m,n)} \right|^2 \\
&+\sigma \sum_{p,q,m,n,r} M_q(m,p) A_q(p,n) \left| e^{i\phi_q(m,n)} - e^{i\phi_q(m,n-1)} e^{i2\pi h r f_{q,p}(n)/F} \right|^2 \\
&s.t.\ \ \overline{W} \geq 0,\ \ H \geq 0,\ \ H_{q,s}(p,n) > 0 \Leftrightarrow A_q(p,n) = 1, \\
&\overline{W}_q(m,p) > 0 \Leftrightarrow M_q(m,p) = 1
\end{aligned}$$

Score-Informed Shift-Invariant CMF

- Additional constraints for score-informed shift-invariant CMF
  - make the shift of templates ex

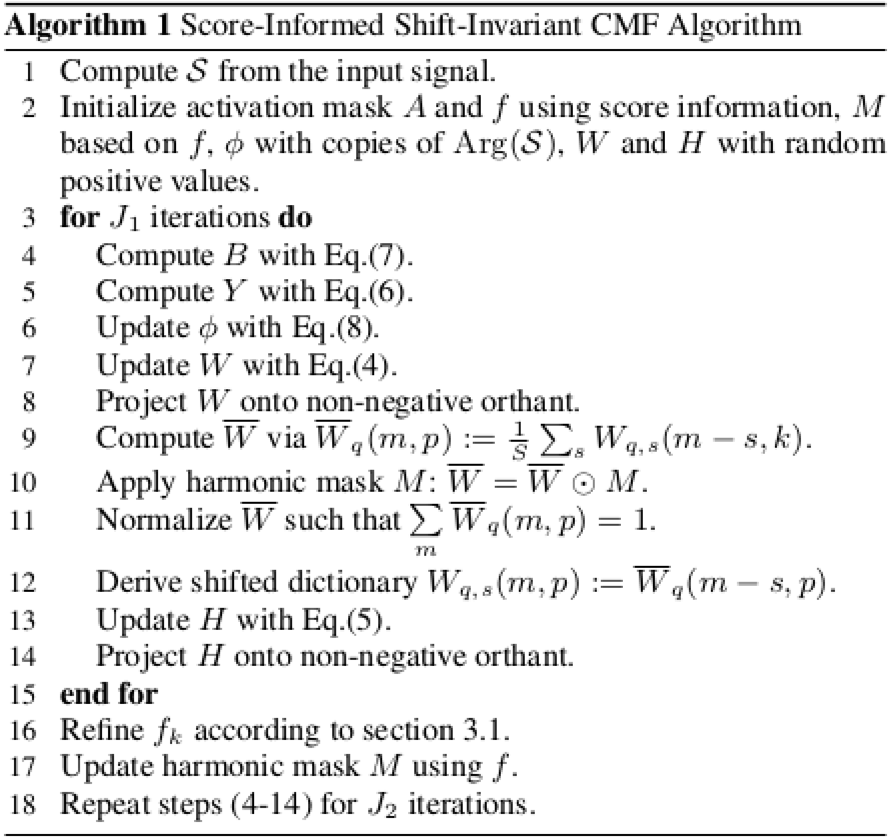
    W ,

    i.e. W q,s (m, p) = W q (m − s

$$\begin{aligned}
&\sum_{m,n}\left|S(m,n)-\sum_{q,p,s}W_{q,s}(m,p)H_{q,s}(p,n)e^{i\phi_q(m,n)}\right|^2\\
&+\sigma\sum_{p,q,m,n,r}M_q(m,p)A_q(p,n)\left|e^{i\phi_q(m,n)}-e^{i\phi_q(m,n-1)}e^{i2\pi hrf_{q,p}(n)/F}\right|^2\\
&s.t.\;\; W\geq 0,\;\; H\geq 0,\;\; H_{q,s}(p,n)>0\Leftrightarrow A_q(p,n)=1,\\
&W_{q,s}(m,p)>0\Leftrightarrow M_q(m-s,p)=1,\\
&W_{q,s}(m,p)=W_{q,0}(m-s,p)
\end{aligned}$$

  - we can now more easily define the iterative update rules for our score-informed shift-invariant CMF.

Score-Informed Shift-Invariant CMF



Algorithm 1 Score-Informed Shift-Invariant CMF Algorithm

```
1  Compute $\mathcal{S}$ from the input signal.
2  Initialize activation mask $A$ and $f$ using score information, $M$
   based on $f$, $\phi$ with copies of $\mathrm{Arg}(\mathcal{S})$, $W$ and $H$ with random
   positive values.
3  for $J_1$ iterations do
4      Compute $B$ with Eq.(7).
5      Compute $Y$ with Eq.(6).
6      Update $\phi$ with Eq.(8).
7      Update $W$ with Eq.(4).
8      Project $W$ onto non-negative orthant.
9      Compute $\overline{W}$ via $\overline{W}_q(m,p) := \frac{1}{S}\sum_s W_{q,s}(m-s,k)$.
10     Apply harmonic mask $M$: $\overline{W} = \overline{W} \odot M$.
11     Normalize $\overline{W}$ such that $\sum_m \overline{W}_q(m,p) = 1$.
12     Derive shifted dictionary $W_{q,s}(m,p) := \overline{W}_q(m-s,p)$.
13     Update $H$ with Eq.(5).
14     Project $H$ onto non-negative orthant.
15 end for
16 Refine $f_k$ according to section 3.1.
17 Update harmonic mask $M$ using $f$.
18 Repeat steps (4-14) for $J_2$ iterations.
```

Cont...

Score-Informed Shift-Invariant CMF

- During the first step, we initialize the fundamental frequencies f we need for the phase regularization using only rough estimates, i.e. we set f q,p (n) to the standard MIDI frequency corresponding to pitch p.
- Based on this f , we then derive the harmonic mask M which encodes the location of harmonics for each pitch.(first estimate of all parameters)
- Next, based on the resulting initial model and the score information, we track the fundamental frequency of each note specified by the score.
- we use the refined f to update M and re-estimate the remaining model parameters
- No need for manually provided fundamental frequency estimates.

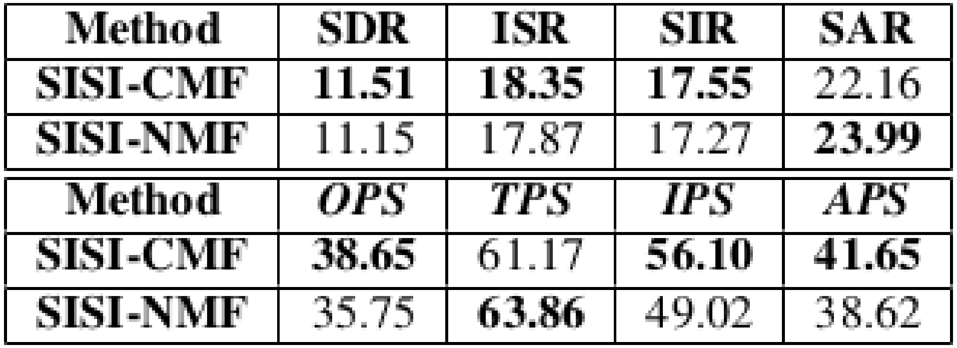
Score-Informed Shift-Invariant CMF

- Refinement of Fundamental Frequencies
  - after the initial step using the rough estimates, we refine the fundamental frequencies based on the initial model we have obtained so far based on a simple procedure
- For each fundamental frequency candidate, we simply compute a weighted sum of entries W q,S max (m, p) for all frequency bins m that correspond to that fundamental frequency or one of its harmonics.
- The candidate with the highest sum is used as the value for f q,p (n)
- After that, we can use the new f to update our harmonic mask M .
  - In particular, since during the first step f was only a rough estimate, we use rather wide regions of ones around possible positions of partials. During the second step, f is more accurate and we can use less entries of ones, which sharpens the mask and its constraints.

Experiments

- Dataset consisting of 10 four-part chorales by J.S. Bach which are given as multitrack recordings of real recordings, each approximately 30 seconds long.
- Each music excerpt consists of an instrumental duet among these instruments: violin, clarinet, tenor saxophone and bassoon.
- Mixture of individual tracks from each chorale to create 60 duets
- Used proposed method as well as a score-informed NMF-based method similar to [19] to decompose these mixes into the two instruments.
- Both methods use S = 5 shifts and one spectral template per pitch and instrument.

Result



Method	SDR	ISR	SIR	SAR
SISI-CMF	**11.51**	**18.35**	**17.55**	22.16
SISI-NMF	11.15	17.87	17.27	**23.99**

Method	*OPS*	*TPS*	*IPS*	*APS*
SISI-CMF	**38.65**	61.17	**56.10**	**41.65**
SISI-NMF	35.75	**63.86**	49.02	38.62

- The two methods are conceptually very similar and if any we expect a gain in separation quality only to be related to overlapped partials – and those contribute only to some degree to the overall signal energy.
- CMF variant aims at modeling the overlapped partials in such way that the interference between instruments could be reduced.
- A negative effect of the increase of the number of parameters to be estimated in CMF compared to NMF might be indicated by the SAR and TPS, where we see better values for NMF.
- overall, the results indicate that the score information is rich enough to contain the vast number of parameters in CMF and use the resulting freedom to improve the separation results.

Conclusion

- A novel score-informed shift-invariant extension of complex matrix factorization.
- Results indicate that incorporating and estimating phase information indeed leads to improved results in score-informed source separation.
- The score information provides enough prior knowledge to control and guide the CMF parameter estimation process, despite the vast number of parameters it contains.