Being low-rank in the time-frequency plane

Valentin Emiya[1] Ronan Hamon[1] Caroline Chaux[2]

[1]Aix Marseille Univ, Université de Toulon, CNRS, LIS, Marseille, France {firstname.lastname}@lis-lab.fr

[2]Aix Marseille Univ, CNRS, Centrale Marseille, I2M, Marseille, France {firstname.lastname}@univ-amu.fr

Open question

Is a **low-rank** prior well adapted to **complex time-frequency matrices** obtained through short-time Fourier transform (STFT)?

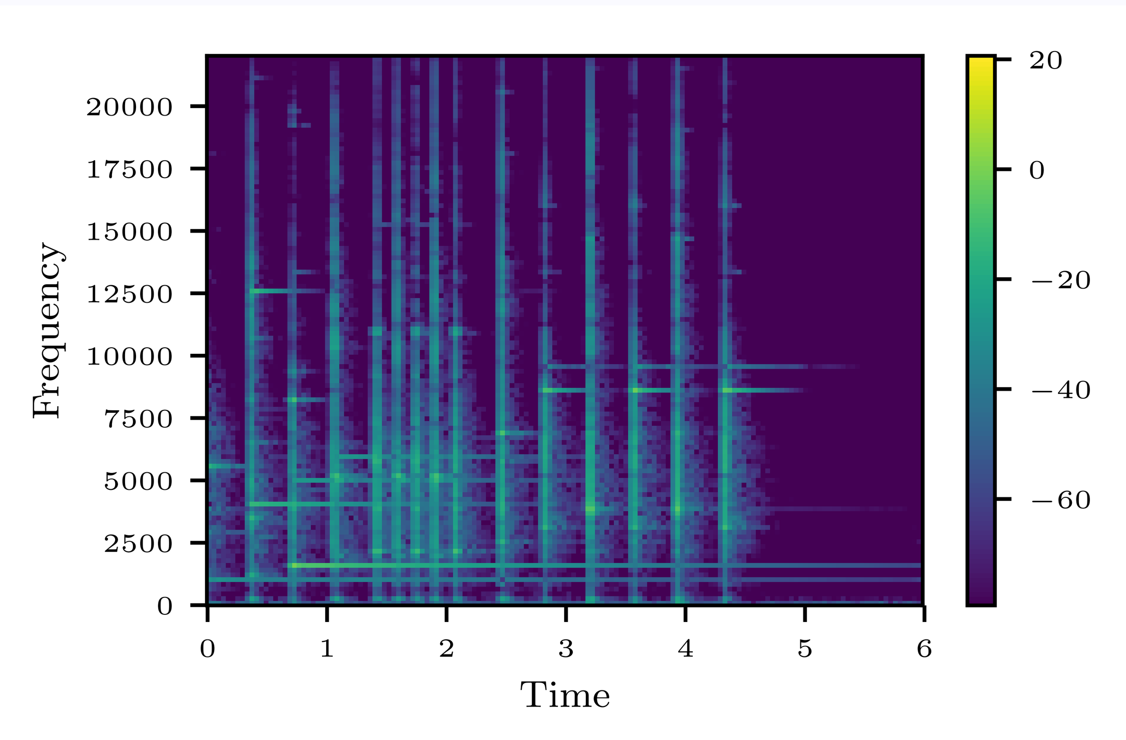
Introduction

Low-rankness prior

- Prior widely used on matrix variable : dictionary learning, image processing, ...
- Good results in audio signal processing when applied on spectrograms through non-negative matrix factorization (NMF) (source separation, inpainting).

Low rank prior for complex time-frequency matrices

- From a general viewpoint, how the intuitions of the low-rankness of the spectrograms can be extended to complex-valued time-frequency matrices, and how to validate them or not ?
- What is a rank-one matrix, or more generally a rank-r matrix, in the time-frequency plane? Can the set of rank-r time-frequency matrices be fully characterized?
- Do time-frequency matrices of real-world sounds have good low-rank approximations? Which kind of elementary patterns are obtained?



Spectrogram of the *Glockenspiel*, composed of about 50 spectral peaks distributed on 15 occurrences of 8 notes. Does the approximate rank of the complex-valued STFT matrix equal 8, 15, 50, or another value?

Notations and definitions

- $[\![L]\!] = \{0, \dots, L-1\}$: set of the first L integers;
- $(\mathbf{s}[m])_{m \in [\![L]\!]} \in \mathbb{C}^L$: complex-valued vectors of length L;
- $(\mathbf{h}[m])_{m \in [\![L]\!]} \in \mathbb{C}^L$: the window;
- L-periodic extension of signals considered.

- STFT defined on K discrete frequencies $\{\nu_k\}_{k \in [\![K]\!]}$ with $\nu_k = \frac{k}{K}$ and N time steps $\{t_n\}_{m \in [\![N]\!]}$, with $t_n = nh$, where h is an arbitrary hop size.
- $\mathcal{T}_n$: translation by t_n
- $\mathcal{M}_k$: modulation by ν_k

[Fourier matrix]
The Fourier matrix $\mathbf{E} \in \mathbb{C}^{L \times L}$ is defined by

$$\mathbf{E} = \left(e^{-2i\pi\frac{kt}{L}}\right)_{k \in [\![L]\!], t \in [\![L]\!]}$$

[DFT and IDFT]
The discrete Fourier transform (DFT) of $\mathbf{u} \in \mathbb{C}^L$ on L discrete frequencies is $\hat{\mathbf{u}} = \mathrm{DFT}(\mathbf{u}) = \mathbf{E}\mathbf{u}$. The adjoint of $\mathbf{E}$ being $\mathbf{E}^*$, the inverse discrete Fourier transform (IDFT) of $\mathbf{u} \in \mathbb{C}^L$ is
$\check{\mathbf{u}} = \mathrm{IDFT}(\mathbf{u}) = \mathbf{E}^{-1}\mathbf{u} = \frac{1}{L}\mathbf{E}^*\mathbf{u}$.

STFT conventions

[$(K \times N)$-STFT, band-pass convention]
In the so-called band-pass convention, the $(K \times N)$-STFT of $\mathbf{s} \in \mathbb{C}^L$ is defined on discrete frequency ν_k, $k \in [\![K]\!]$ and discrete time t_n, $n \in [\![N]\!]$ by

$$\mathbf{S}_{\mathrm{BP}}^{(K\times N)}[k,n] = \langle \mathcal{T}_n \mathcal{M}_k \mathbf{h}, \mathbf{s} \rangle = \sum_m \mathbf{s}[t_n + m]\, \mathbf{h}[m]\, e^{-2i\pi\nu_k m}.$$

[$(K \times N)$-STFT, low-pass convention]
In the so-called low-pass convention, the $(K \times N)$-STFT of $\mathbf{s} \in \mathbb{C}^L$ is defined on discrete frequency ν_k, $k \in [\![K]\!]$ and discrete time t_n, $n \in [\![N]\!]$ by

$$\mathbf{S}_{\mathrm{LP}}^{(K\times N)}[k,n] = \langle \mathcal{M}_k \mathcal{T}_n \mathbf{h}, \mathbf{s} \rangle = \sum_m \mathbf{s}[m]\, \mathbf{h}[m - t_n]\, e^{-2i\pi\nu_k m}.$$

[Relation between conventions]
$\forall k \in [\![K]\!]$, $n \in \mathbb{Z}$, $\mathbf{S}_{\mathrm{LP}}(k,n) = \mathbf{S}_{\mathrm{BP}}(k,n) \times e^{-2i\pi\nu_k m_n}$

[Maximum redundancy $K = N = L$]
$(L \times L)$-STFT of $\mathbf{s} \in \mathbb{C}^L$ in both conventions, denoted respectively by $\mathbf{S}_{\mathrm{BP}} = \mathbf{S}_{\mathrm{BP}}^{(L\times L)}$ and $\mathbf{S}_{\mathrm{LP}} = \mathbf{S}_{\mathrm{LP}}^{(L\times L)}$, are rewritten

$$\forall k, n, \mathbf{S}_{\mathrm{BP}}[k,n] = \sum_m \mathbf{s}[n+m]\, \mathbf{h}[m]\, e^{-2i\pi\frac{km}{L}}$$

$$\forall k, n, \mathbf{S}_{\mathrm{LP}}[k,n] = \sum_m \mathbf{s}[m]\, \mathbf{h}[m-n]\, e^{-2i\pi\frac{km}{L}}.$$

[$\mathbf{S}_{\mathrm{BP}}^{(K\times N)}$ vs $\mathbf{S}_{\mathrm{BP}}$ and $\mathbf{S}_{\mathrm{LP}}^{(K\times N)}$ vs $\mathbf{S}_{\mathrm{LP}}$]
Let $K, N \in \mathbb{N}$ be such that $K|L$ and $N|L$. Then for any $k \in [\![K]\!]$, $n \in [\![N]\!]$, we have

$$\mathbf{S}_{\mathrm{BP}}^{(K\times N)}[k,n] = \mathbf{S}_{\mathrm{BP}}\left[\frac{kL}{K}, \frac{nL}{K}\right]$$

and

$$\mathbf{S}_{\mathrm{LP}}^{(K\times N)}[k,n] = \mathbf{S}_{\mathrm{LP}}\left[\frac{kL}{K}, \frac{nL}{K}\right].$$

Characterization of rank-r STFT matrices

Based on the following factorization,

Factorization of STFT matrices

For any signal $\mathbf{s} \in \mathbb{C}^L$ and window $\mathbf{h} \in \mathbb{C}^L$, we have

$$\mathbf{S}_{\mathrm{BP}} = \mathbf{E}\,\mathrm{diag}(\mathbf{h})\,\mathbf{E}^{-1}\,\mathrm{diag}(\hat{\mathbf{s}})\,\mathbf{E}$$

and

$$\mathbf{S}_{\mathrm{LP}} = \mathbf{E}\,\mathrm{diag}(\mathbf{s})\,\mathbf{E}^{-1}\,\mathrm{diag}(\hat{\mathbf{h}})\,\mathbf{E}$$

the main result is the caracterization:

Rank-r STFT matrices

If $\mathbf{h} \in \mathbb{C}^L$ is a window that does not vanish, i.e., $\forall k \in [\![L]\!], \mathbf{h}[k] \neq 0$, then $\mathrm{rank}(\mathbf{S}_{\mathrm{BP}}) = \|\hat{\mathbf{s}}\|_0$.

$\Rightarrow$ *The set of rank-r STFT matrices in the band-pass convention is composed of the signals that are a sum of r pure complex exponentials at Fourier frequencies.*

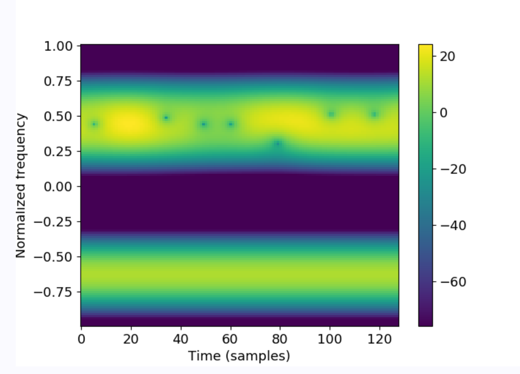
If $\mathbf{h} \in \mathbb{C}^L$ is a window such that $\hat{\mathbf{h}}$ does not vanish, i.e., $\forall k \in [\![L]\!], \hat{\mathbf{h}}[k] \neq 0$, then $\mathrm{rank}(\mathbf{S}_{\mathrm{LP}}) = \|\mathbf{s}\|_0$.

$\Rightarrow$ *The set of rank-r STFT matrices in the low-pass convention is composed of the signals that are a sum r diracs at integer times.*

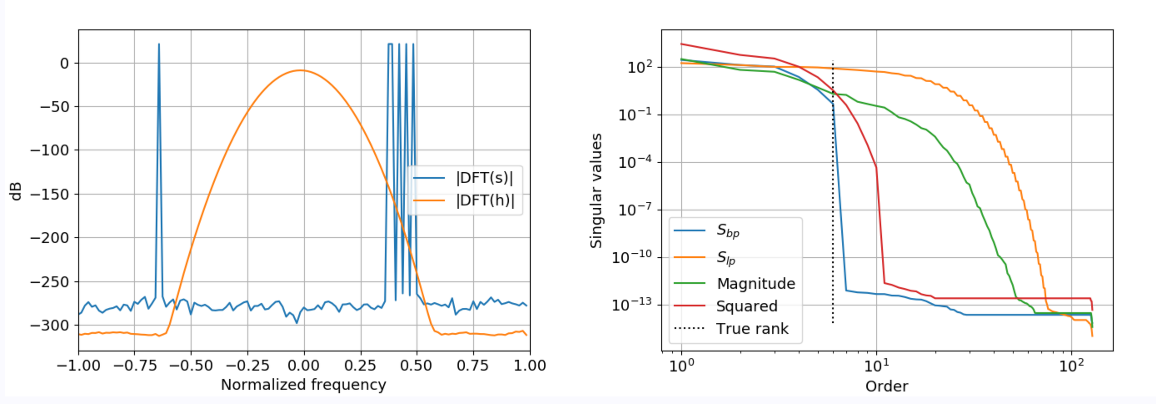
Illustrations

Analysis of low-rank STFT matrices

Context : Signal with length $L = 128$ composed of a sum of $N_c = 6$ complex sinusoids at exact Fourier frequencies (5 closed frequencies and 1 isolated frequency).



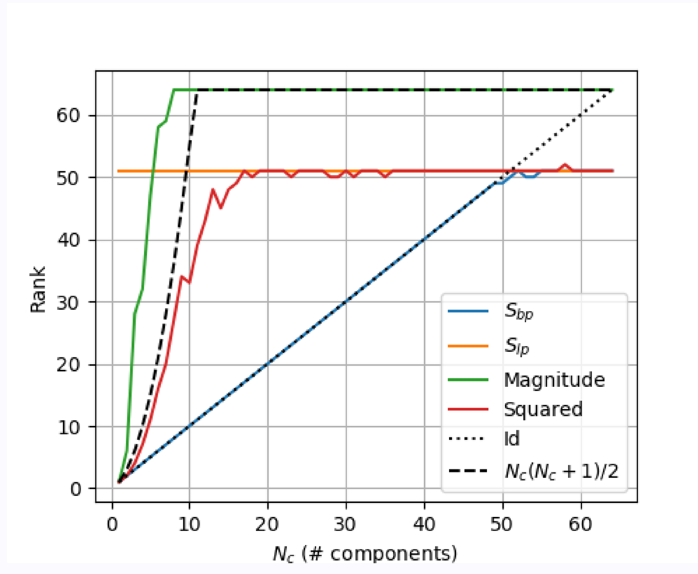
Results : $\mathrm{rg}\,\mathbf{S}_{\mathrm{BP}} = Nc$ while $\mathrm{rg}\,\mathbf{S}_{\mathrm{LP}}$ is higher.



Analysis with a Gaussian window: DFT of the signal and of the window (left) and singular values of STFT matrices, magnitude and energy spectrograms (right).

Context: rank of complex and magnitude TF matrices vs. number of components N_c (frequencies drawn randomly at exact Fourier frequencies), signal length $L = 64$.

Results: $\mathrm{rg}\,\mathbf{S}_{\mathrm{BP}} = N_c$ while $\mathrm{rg}\,\mathbf{S}_{\mathrm{LP}}$ is higher.
Trend in $\frac{N_c(N_c+1)}{2}$ followed by $\mathrm{rg}\,|\mathbf{S}|^2$ (upper bound) and rg STFT matrix < related spectrograms.



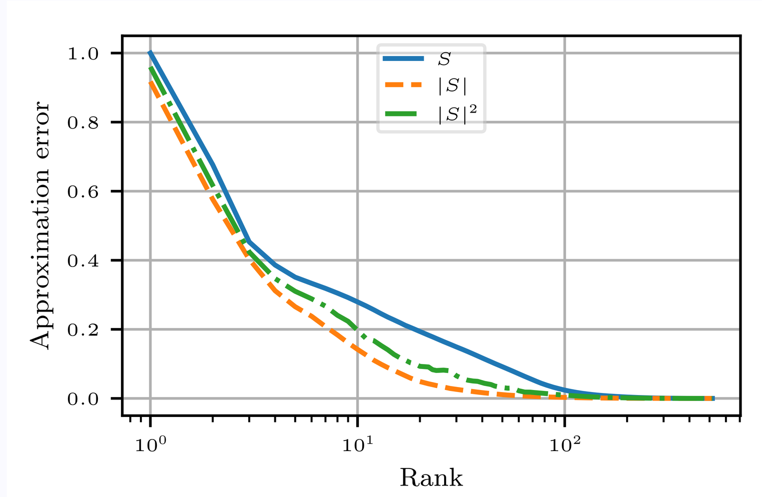
Rank of several types of time-frequency matrices vs. number of sinusoids in the signal.

Low-rank STFT approximation for real audio signals

Objective: find the best rank-r approximation $\widetilde{\mathbf{X}} \in \mathbb{C}^{K\times N}$ of a matrix $\mathbf{X} \in \mathbb{C}^{K\times N}$

$$\widetilde{\mathbf{X}} = \underset{\mathbf{Y} \in \mathbb{C}^{K\times N}, \mathrm{rg}(\mathbf{Y}) \leq r}{\arg\min} \|\mathbf{X} - \mathbf{Y}\|_F^2.$$

Solved for $\mathbf{X} = \mathbf{S}_{\mathrm{BP}}$ and $\mathbf{X} = |\mathbf{S}_{\mathrm{BP}}|$ for different r.



Glockenspiel sound: normalized approximation error of STFT/magnitude spectrogram/energy spectrogram when considering a low-rank decomposition.

Results: better approximation for spectrograms than STFT. STFT still somehow low-rank approximable.

Conclusion

- We have characterized exactly the set of low-rank matrices in a general context:
  1. the set of low-rank matrices is very narrow;
  2. the STFT phase convention is critical
  3. the STFT of a mixture of sinusoids and dirac cannot be jointly described by a low-rank model
- Extension of this work to the case of finite signals: K. Usevich et al., *Characterization of finite signals with low-rank STFT*, SSP 2018.
- Future work: local time-frequency low-rank models.

ICASSP 2018 - April 2018, Calgary, Canada - This work was supported by the Agence Nationale de la Recherche under grant JCJC MAD (ANR-14-CE27-0002).