Characterization of finite signals with low-rank STFT

Konstantin Usevich, Valentin Emiya, David Brie, Caroline Chaux

CHARACTERIZATION OF FINITE SIGNALS WITH LOW-RANK STFT

Konstantin Usevich[a], Valentin Emiya[b], David Brie[a], Caroline Chaux[c]

[a] Université de Lorraine, CNRS, CRAN, Vandœuvre-lès-Nancy, France
[b] Aix Marseille Univ, Université de Toulon, CNRS, LIS, Marseille, France
[c] Aix Marseille Univ, CNRS, Centrale Marseille, I2M, Marseille, France

ABSTRACT

The goal of this paper is characterize finite-length signals that have a low-rank short-time Fourier transform. By using the connection with Hankel matrices, we give a comprehensive answer in case of maximal overlap, where the class of signals includes products of complex exponentials and polynomials. In the general case, other exotic examples may appear.

Index Terms— Short-time Fourier transform, Hankel matrices, low-rank approximation

1. INTRODUCTION

Matrices of data are often modeled with a low-rank assumption, which can be conveniently handled in optimization procedures [1]. Such approximations have proven to be very useful for spectrograms in the context of audio signal processing, where nonnegative matrix factorisation is typically performed [2, 3]. However, the phase information is lost when considering the spectrograms. This results in limitations in audio inverse problems like denoising [4], source separation [3, 5] and inpainting [6, 7], where the amplitudes and phases are generally estimated sequentially. In [8, 9] it was proposed to perform low-rank approximation directly on the complex-valued short-time Fourier transform (STFT), in the context of inpainting missing coefficients in STFT matrices.

However, such an approach raises the following question: what kind of signals have a low-rank STFT? An answer was given in [8] and leads to a restricted class of possible signals, due to the circularity of the proposed time-frequency transform – which comes down to analyzing a periodized version of the original finite signal. The goal of this paper is to provide an insight into the structure of signals with low-rank STFT, without using a circular transform. This is beneficial not only to avoid boundary effects, but also to obtain a larger class of low-rank STFT matrices, allowing exponential damping and polynomial modulations. We use the relation between STFT and Hankel/block-Hankel matrices, which allows us to give a complete characterisation for the case of maximal overlap, and to give nontrivial examples for non-maximal overlap.

1.1. Notation and the definition of the STFT

We follow mainly the notation in [9]. For a nonnegative integer number K we define by $[\![K]\!] = \{0, \dots, K-1\}$. The vectors are in boldface, e.g., $\boldsymbol{a}$, and their elements will be indexed as $\boldsymbol{a}(k)$; the matrices are denoted by capital boldface letters, e.g., $\boldsymbol{A}$ and their elements are accessed as $\boldsymbol{A}(k, l)$. Note that the vectors and matrices in this paper are indexed starting from zero, similarly to many programming languages.

Let $\boldsymbol{s} = [\boldsymbol{s}(m)]_{m \in \mathbb{Z}}$ be a signal, and $\boldsymbol{w} = [\boldsymbol{w}(m)]_{m \in \mathbb{Z}}$ be a window. Next, fix two positive natural numbers: **(i)** h — so-called hop; **(ii)** K — number of frequencies. The STFT is then defined for K frequencies $\nu_k = \frac{k}{K}$, $k \in [\![K]\!]$ at time instants $t_n = hn$, $n \in \mathbb{Z}$ as follows.

Definition 1.1. *For $k \in [\![K]\!]$ and $n \in \mathbb{Z}$ we define the short time Fourier transform (STFT) as*

$$\widehat{\boldsymbol{S}}(k, n) = \sum_{m \in \mathbb{Z}} \boldsymbol{s}(t_n + m)\boldsymbol{w}(m)e^{-2i\pi\nu_k m}.$$

In this paper, we use only band-pass convention for STFT; there also exists a low-pass convention [8]. Note that in [8], the notation $\boldsymbol{S}_{BP}$ was used. We use the notation $\widehat{\boldsymbol{S}}$ instead, because only the band-pass convention is considered.

2. FINITE SIGNALS AND MATRIX FORM

We consider only the first N time instants, and the window $\boldsymbol{w}$ having finite length $L \leq K$, so that $\boldsymbol{w}(m) = 0$ if $m < 0$ or $m \geq L$. Then the STFT becomes a $K \times N$ matrix $\widehat{\boldsymbol{S}}^{(K \times N)}$:

$$\widehat{\boldsymbol{S}}^{(K \times N)}(k, n) = \sum_{m=0}^{L-1} \boldsymbol{s}(t_n + m)\boldsymbol{w}(m)e^{-2i\pi\nu_k m}, \quad (1)$$

2.1. Matrix factorization

First, let us define the following $K \times L$ matrix $\boldsymbol{F}^{(K \times L)}$:

$$\boldsymbol{F}^{(K \times L)} := \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & e^{\frac{-2i\pi}{K}} & \cdots & e^{\frac{-2i\pi(L-1)}{K}} \\ 1 & e^{\frac{-2i\pi 2}{K}} & \cdots & e^{\frac{-2i\pi 2(L-1)}{K}} \\ \vdots & \vdots & & \vdots \\ 1 & e^{\frac{-2i\pi(K-1)}{K}} & \cdots & e^{\frac{-2i\pi(K-1)(L-1)}{K}} \end{bmatrix},$$

i.e., $\boldsymbol{F}^{(K\times L)}(k,m) = e^{\frac{-2i\pi km}{K}}$. The matrix $\boldsymbol{F}^{(K\times L)}$ contains the first L columns of the Fourier transform matrix. Since $L \leq K$, this matrix is full column rank.

Next, we define the matrix $\boldsymbol{S}^{L\times N} \in \mathbb{C}^{L\times N}$ containing the "patches" of the signal stacked next to each other:

$$\boldsymbol{S}^{L\times N} := \begin{bmatrix} s(t_0) & s(t_1) & \cdots & s(t_{N-1}) \\ s(t_0+1) & s(t_1+1) & \cdots & s(t_{N-1}+1) \\ \vdots & \vdots & & \vdots \\ s(t_0+L-1) & s(t_1+L-1) & \cdots & s(t_{N-1}+L-1) \end{bmatrix}.$$

Then the matrix $\widehat{\boldsymbol{S}}^{(K\times N)}$ admits the following factorisation:

$$\widehat{\boldsymbol{S}}^{(K\times N)} = \boldsymbol{F}^{(K\times L)}\operatorname{Diag}(\boldsymbol{w})\boldsymbol{S}^{L\times N}, \tag{2}$$

where $\operatorname{Diag}(\boldsymbol{w})$ denotes the diagonal matrix with the elements of $\boldsymbol{w}$ on the diagonal.

2.2. Rank of the STFT matrix

Since $\boldsymbol{F}^{(K\times L)}$ is full column rank, the rank of $\widehat{\boldsymbol{S}}^{(K\times N)}$ can be expressed in terms of the rank of a submatrix of $\boldsymbol{S}^{L\times N}$.

In order to have a complete description, we need to know the support of the window $\boldsymbol{w}$, which is defined as the set

$$\mathcal{L} = \{\ell_0, \ldots, \ell_{M-1}\} \subset [\![L]\!] \tag{3}$$

such that

$$\boldsymbol{w}(m) \neq 0 \iff m \in \mathcal{L}.$$

Proposition 2.1. *Let $\mathcal{L}$ be the support* (3) *of the window $\boldsymbol{w}$. Then the rank of the $\widehat{\boldsymbol{S}}^{(K\times N)}$ is equal to the rank of the matrix*

$$\boldsymbol{S}^{\mathcal{L}\times N} := \begin{bmatrix} s(\ell_0+t_0) & s(\ell_0+t_1) & \cdots & s(\ell_0+t_{N-1}) \\ s(\ell_1+t_0) & s(\ell_1+t_1) & \cdots & s(\ell_1+t_{N-1}) \\ \vdots & \vdots & & \vdots \\ s(\ell_{M-1}+t_0) & s(\ell_{M-1}+t_1) & \cdots & s(\ell_{M-1}+t_{N-1}) \end{bmatrix}. \tag{4}$$

Proof. Since $\boldsymbol{F}^{(K\times L)}$ is full column rank, we have that

$$\operatorname{rank}(\widehat{\boldsymbol{S}}^{(K\times N)}) = \operatorname{rank}(\operatorname{Diag}(\boldsymbol{w})\boldsymbol{S}^{L\times N}).$$

Since ℓ-th row of $\operatorname{Diag}(\boldsymbol{w})\boldsymbol{S}^{L\times N}$ is nonzero only if $\boldsymbol{w}(\ell) \neq 0$, we have that $\operatorname{rank}(\operatorname{Diag}(\boldsymbol{w})\boldsymbol{S}^{L\times N}) = \operatorname{rank}(\boldsymbol{S}^{\mathcal{L}\times N})$. $\square$

2.3. Low-rank approximations

In the previous subsection, we established that the rank of the STFT matrix can be computed in time domain. Before studying signals with low-rank STFT matrices, we show that low-rank approximations can be also computed in time domain.

Remark 2.2. *For any unitarily invariant norm $\|\cdot\|$,*

$$\|\widehat{\boldsymbol{S}}^{(K\times N)}\| = \|\operatorname{Diag}(\boldsymbol{w})\boldsymbol{S}^{L\times N}\|.$$

This follows due to the fact that not only $\boldsymbol{F}^{(K\times L)}$ has full column rank, but also because its columns are orthogonal.

This has a direct consequence for low-rank approximations and convex relaxations. First, consider a low-rank approximation problem as stated in [9]

$$\min_{\boldsymbol{Y}\in\mathbb{C}^{K\times N},\operatorname{rank}(\boldsymbol{Y})\leq R} \|\widehat{\boldsymbol{S}}^{(K\times N)} - \boldsymbol{Y}\|_F^2. \tag{5}$$

Hence, by Proposition 2.1, we have that problem (5) is equivalent to low-rank approximation in time domain

$$\min_{\boldsymbol{Z}\in\mathbb{C}^{K\times N},\operatorname{rank}(\boldsymbol{Z})\leq R} \|\boldsymbol{S}^{L\times N} - \boldsymbol{Z}\|_{F,\boldsymbol{w}}^2, \tag{6}$$

where $\|\cdot\|_{F,\boldsymbol{w}}$ denotes the weighted Frobenius norm:

$$\|\boldsymbol{X}\|_{F,\boldsymbol{w}}^2 := \operatorname{trace}(\boldsymbol{X}^{\mathsf{T}}\operatorname{Diag}(\boldsymbol{w})^2\boldsymbol{X}).$$

In other words, the low-rank approximation problem can be posed and solved in time domain. Note that the nuclear norm is also orthogonally invariant, hence the nuclear norm minimization [9] can be also treated in the time domain.

3. LOW-RANK HANKEL MATRICES

Before we proceed to classification of signals with low-rank STFT, we recall Hankel matrices and their algebraic theory.

Let $\boldsymbol{f} = (f(0), \ldots, f(T-1)) \in \mathbb{C}^T$ be a vector and L be a nonnegative integer, so that $1 \leq L \leq T$. Then the Hankel matrix $\mathscr{H}_L(\boldsymbol{f}) \in \mathbb{C}^{L\times(T-L+1)}$ is defined as follows

$$\mathscr{H}_L(\boldsymbol{f}) = \begin{bmatrix} f(0) & f(1) & \cdots & f(T-L) \\ f(1) & f(2) & \cdots & f(T-L+1) \\ \vdots & \vdots & & \vdots \\ f(L-1) & f(L) & \cdots & f(T-1) \end{bmatrix}. \tag{7}$$

Note that there is a slight abuse of notation: L has a different meaning from the one used in Section 2. Now we consider a class of signals that is central to this paper: sum of products of polynomial and exponential.

Let $(\nu_1, \lambda_1), \ldots, (\nu_s, \lambda_s) \in \mathbb{N}^* \times (\mathbb{C}\setminus\{0\})$, and $\nu_0, \nu_\infty \in \mathbb{N}$ be the numbers such that

$$\nu_0 + \nu_1 + \cdots + \nu_s + \nu_\infty = R. \tag{8}$$

Consider a signal $\boldsymbol{f} \in \mathbb{C}^T$ such that

$$f(t) = \underbrace{\sum_{j=1}^{\nu_0} a_j \delta_{j-1}(t) + \sum_{l=1}^{\nu_\infty} b_l \delta_{T-l}(t)}_{\text{transient terms (beginning and end)}} + \sum_{k=1}^{s} p_k(t)\lambda_k^t, \tag{9}$$

where **(i)** $p_k(t)$ is a polynomial of degree $\nu_k - 1$; **(ii)** $\delta_x(t)$ is a (Kronecker) delta-function:

$$\delta_x(t) = \begin{cases} 1, & x = t, \\ 0, & x \neq t; \end{cases}$$

(iii) and $a_{\nu_0} \neq 0$, $b_{\nu_\infty} \neq 0$. As shown in [10], low-rank Hankel matrices correspond exactly to signals of the form (9). Note that similar results were obtained and used in sampling theory, in the context of finite rate of innovation [11].

Theorem 3.1 (Special case of [10, Thm. 8.2]). *Let R be an integer such that $0 < R < \min(L, T - L + 1)$. Then the Hankel matrix $\mathrm{rank}(\mathscr{H}_L(\boldsymbol{f}))$ is of rank R if and only if the signal $\boldsymbol{f}$ has the form (9). Moreover, the representation (9) (called* canonical representation *in [10]), is unique.*

Example 1 (Cisoids). *First, a signal*

$$s(t) = c_1 e^{2i\pi\omega_1 t} + c_2 e^{2i\pi\omega_2 t} + c_3 e^{2i\pi\omega_3 t},$$

has rank 3 for any distinct $\{\omega_1, \omega_2, \omega_3\}$, as in [9, Fig. 2]. Note that frequencies do not have to lie on the Fourier grid. We can also consider damped exponentials (i.e., $e^{(2i\pi\omega - \alpha)t}$).

Example 2 (Polynomial modulation). *The function*

$$f(t) = (at^2 + bt + c)\cos(2\pi\omega t + \phi)$$

corresponds to a rank-6 Hankel matrix. Signals with polynomial modulations are useful for modelling of music records [12] and in other contexts [13].

Example 3 (Transients). *Finally, let us give an example for a transient. The signal in Fig. 1 has rank-4 Hankel matrix,*
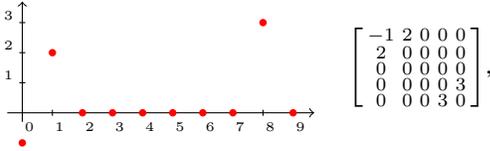

Fig. 1. A "transient" signal, and the Hankel matrix $\mathscr{H}_4(\boldsymbol{s})$.

since $\nu_0 = \nu_\infty = 2$. Adding the signal to any of previous examples increases the rank by 4.

Note that transient terms in (9) are allowed only in the beginning and/or at the end of the signal, thus there should be at least L zero-valued consecutive samples in between the transient terms.

4. CLASSES OF LOW-RANK STFT MATRICES

In this section we discuss the structure of (4), which will help us to describe the class of signals with low-rank STFT. The descriptions rely on the algebraic theory of Hankel matrices.

4.1. Some simplifications for most common windows

Note that for most windows used in practice, the support of the window (the set $\mathcal{L}$ in (3)) is the whole interval $[\![L]\!]$, maybe without the endpoints, i.e. one of the following three options:

$$\mathcal{L} = [\![L]\!], \text{ or } \mathcal{L} = \{1, \ldots, L - 1\}, \text{ or } \mathcal{L} = \{1, \ldots, L - 2\},$$

where the last case is more typical if a symmetric window is used. These cases can be represented in a unified form. Let

$$\Delta = \ell_0 \in \{0, 1\}, \ M \in \{L, L - 1\}, \text{ and } \mathcal{L} = [\![M]\!].$$

Then the matrix (4) has the form

$$
\begin{bmatrix}
s(\Delta) & s(\Delta+h) & \cdots & s(\Delta+h(N-1)) \\
s(\Delta+1) & s(\Delta+h+1) & \cdots & s(\Delta+h(N-1)+1) \\
\vdots & \vdots & & \vdots \\
s(\Delta+M-1) & s(\Delta+h+M-1) & \cdots & s(\Delta+h(N-1)+M-1)
\end{bmatrix}. \quad (10)
$$

4.2. Maximum overlap: a complete classification

In the case of the maximum overlap, $h = 1$. In this case, the matrix (10) is just a Hankel matrix. The following corollary characterizes all signals having low-rank STFT.

Corollary 4.1. *Let $\boldsymbol{s} \in \mathbb{C}^T$ be a signal, and $\widehat{\boldsymbol{S}}^{(K \times N)}$ be its STFT with the hop $h = 1$, and window parameters (Δ and M) as in Section 4.1. Then we have that (i) $\mathrm{rank}(\widehat{\boldsymbol{S}}^{(K \times N)}) \leq \min(M, N)$ for any signal; (ii) $\mathrm{rank}(\widehat{\boldsymbol{S}}^{(K \times N)}) = R < \min(M, N)$ if and only if the subvector $\boldsymbol{s}(\Delta : \Delta + N + M - 2)$ has the representation (9) with R given in (8).*

Remark 4.2. *Note that only a subvector of $\boldsymbol{s}$ is considered, so that the values $s(t)$ for $0 \leq t < \Delta$ or $\Delta + N + M - 1 \leq t < T$ can take arbitrary values. In other words, we may have more transient values in the beginning and in the end of the signal.*

Example 4. *Take a specific example of a signal given by*

$$s(t) = a_0 \delta_{10}(t) + (t^2 + t + 1)e^{(\ln(0.95) + 2i\pi 0.1)t}, \quad (11)$$

for which $\nu_0 = 11$ and $\nu_1 = 3$. Its real part is plotted in Fig. 2. Next, we compute the STFT of the signal with hop
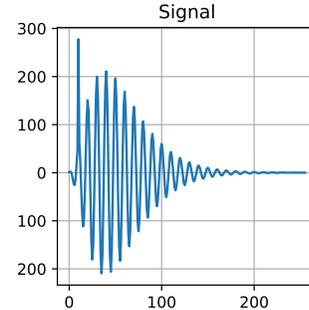

Fig. 2. The real part of the signal (11).

size $h = 1$, and its SVD, which are plotted in Fig. 3 (the singular vectors being represented by the modulus only). The numerical rank is $14 = \nu_0 + \nu_1$, as predicted.

The first two SVD components, with left singular vectors well-localized in frequency, mainly explain the modulated exponential component, while SVD components 2 to 13, with right singular vectors well-localized in time, explain the Dirac component. However, the shape of left singular vectors 2 to 13 around frequency 0.1 shows that there is no perfect assignation of SVD components to one of the terms in (11).
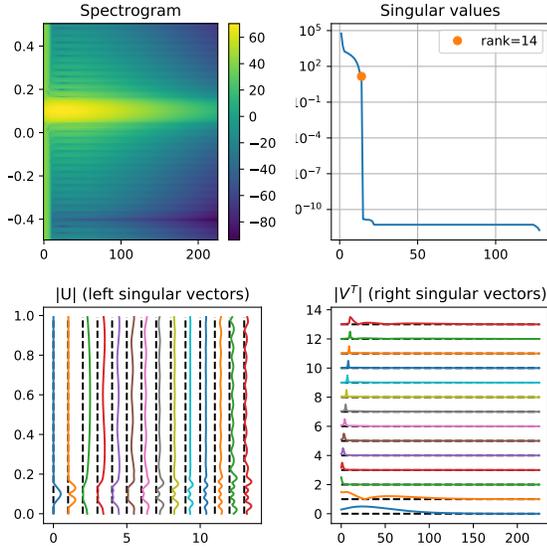
Fig. 3. SVD of the STFT of (11), maximum overlap ($h = 1$).

4.3. Case of general overlap

The matrix (10) is a submatrix of the Hankel matrix $\mathscr{H}_L(\boldsymbol{s})$, obtained by taking every h-th column. Therefore, any signal of the form (9) will have an STFT matrix with rank $\leq R$, as illustrated in Example 5.

Example 5. *We continue Example 4, but now we take $h = 8$. The computed STFT and its SVD are shown in Fig. 4. Note that the rank of the matrix was reduced to 5, because there are only two STFT columns where the transient appears.*
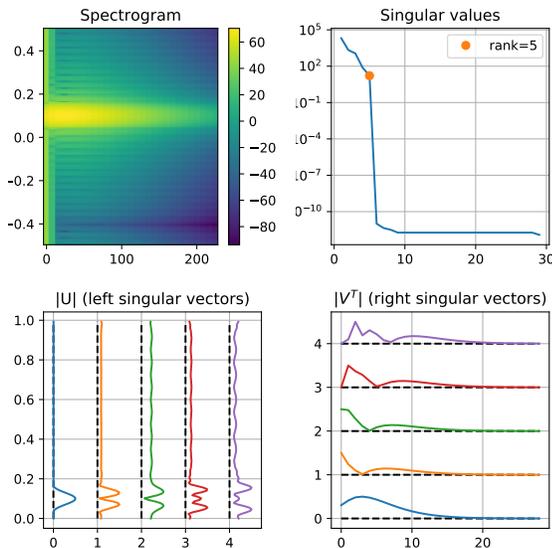

Fig. 4. SVD of the STFT of (11), general overlap ($h = 8$)

We now exhibit examples of signals that are not of the form (9) but have a low-rank STFT matrices for $h > 1$.

Example 6. *Suppose we consider an STFT with the hop $h > 1$. Take an arbitratry FIR filter $\boldsymbol{g} \in \mathbb{C}^h$ and consider a spike train[1], with the interval between spikes equal to h:*

$$f(t) = c_0\delta_0(t) + c_1\delta_h(t) + \cdots + c_{N-1}\delta_{h(N-1)}(t),$$

where $c_k = ak + b$. Then for the convolution $\boldsymbol{s} = \boldsymbol{f} * \boldsymbol{g}$, the rank of the matrix (10) is 2. An example for $h = 3$, $M = 6$, $c_k = k + 1$, together with the matrix (10), is shown in Fig. 5.
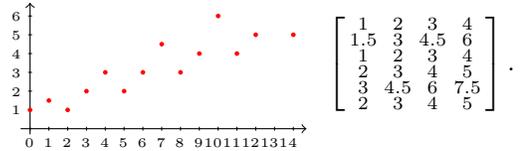

Fig. 5. Convolution with $\boldsymbol{g} = [1, 1.5, 1]$;

Remark 4.3. *The same could be said about arbitrary sequence c_k of the form (9), for example, $c_k = \cos(2\pi\omega k)$.*

Example 7. *For $h > 1$, and a piecewise polynomial of degree $d < h - 1$ function, where the breakpoints are equispaced (with hop h). Then the matrix (10) has rank at most $(d+1)N$.*

Again, take an example of $h = 3$. Then for the following signal, the corresponding matrix (10) has rank 4.
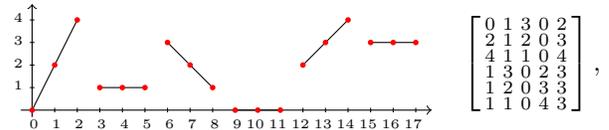
These examples are due to the fact that the matrix (10) is, in fact, a truncated block-Hankel matrix. Although there is also an algebraic theory of block-Hankel matrices [14], it is more difficult to have a simple description (as in (9)).

5. CONCLUSIONS

We have shown that the finite signals with low-rank STFT include polynomially modulated complex exponentials with possibly some transient terms (9); moreover, only signals (9) have a low-rank STFT in the case of maximal overlap. For the general overlap, the rank may become smaller for signals (9) and less dependent on transient terms. However, there are much more signals with a low-rank STFT in the general case.

We leave for future research the practical implications of these results. Other questions could be considered, for example, why and when the components of the signals separate into different components of the SVD[2] in Examples 4 and 5.

---

[1]Note that $\boldsymbol{f}$ does not play the same role as in Section 3.

[2]Note that in view of Section 2.3, for a rectangular window, the SVD of an STFT matrix can be obtained from the SVD of the Hankel matrix. Separability of signals in that case was studied in the context of time series analysis [15] and harmonic retrieval [13].

6. REFERENCES

[1] B. Mishra, G. Meyer, F. Bach, and R. Sepulchre, "Low-rank optimization with trace norm penalty," *SIAM J. Optim.*, vol. 23, no. 4, pp. 2124–2149, 2013.

[2] P. Smaragdis and J. C. Brown, "Non-negative matrix factorization for polyphonic music transcription," in *2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (IEEE Cat. No.03TH8684)*, Oct 2003, pp. 177–180.

[3] A. Ozerov and C. Févotte, "Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 3, pp. 550–563, March 2010.

[4] P. Mowlaee and R. Saeidi, "Iterative closed-loop phase-aware single-channel speech enhancement," *Signal Processing Letters, IEEE*, vol. 20, no. 12, pp. 1235–1239, Dec 2013.

[5] J. Le Roux and E. Vincent, "Consistent Wiener filtering for audio source separation," *Signal Processing Letters, IEEE*, vol. 20, no. 3, pp. 217–220, Mar. 2013.

[6] P. Smaragdis, B. Raj, and M. Shashanka, "Missing data imputation for spectral audio signals," in *Proc. of MLSP*, Grenoble, France, Sept. 2009.

[7] J. Le Roux, H. Kameoka, N. Ono, A. de Cheveigné, and S. Sagayama, "Computational auditory induction as a missing-data model-fitting problem with Bregman divergence," *Speech Communication*, vol. In Press, 2010.

[8] V. Emiya, R. Hamon, and C. Chaux, "Being low-rank in the time-frequency plane," in *Proc. of ICASSP*, Calgary, Canada, Apr. 2018.

[9] R. Hamon, C. Chaux, and V. Emiya, "De la faiblesse de rang en temps-fréquence," in *GRETSI*, Juan-Les-Pins, France, Sept. 2017.

[10] G. Heinig and K. Rost, *Algebraic methods for Toeplitz-like matrices and operators*, Birkhäuser, Boston, 1984.

[11] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Transactions on Signal Processing*, vol. 50, no. 6, pp. 1417–1428, March 2002.

[12] R. Badeau, *Méthodes à haute résolution pour l'estimation et le suivi de sinusoïdes modulées. Application aux signaux de musique.*, Ph.D. thesis, Ecole Nationale Supérieure des Télécommunications, Paris, 2005.

[13] R. Badeau, G. Richard, and B. David, "Performance of ESPRIT for estimating mixtures of complex exponentials modulated by polynomials," *IEEE Transactions on Signal Processing*, vol. 56, no. 2, pp. 492–504, Feb 2008.

[14] G. Heinig and P. Jankowski, "Kernel structure of block Hankel and Toeplitz matrices and partial realization," *Linear Algebra and its Applications*, vol. 175, pp. 1–30, 1992.

[15] N. Golyandina, V. Nekrutkin, and A. Zhigljavsky, *Analysis of Time Series Structure: SSA and Related Techniques*, Chapman&Hall/CRC, 2001.