Estimation with Low-Rank Time-Frequency Synthesis Models

Cédric Févotte, Matthieu Kowalski

Estimation with Low-Rank Time-Frequency Synthesis Models

Cédric Févotte, *Senior Member, IEEE,* and Matthieu Kowalski

***Abstract*—Many state-of-the art signal decomposition techniques rely on a low-rank factorization of a time-frequency (t-f) transform. In particular, nonnegative matrix factorization (NMF) of the spectrogram has been considered in many audio applications. This is an *analysis* approach in the sense that the factorization is applied to the squared magnitude of the analysis coefficients returned by the t-f transform. In this paper we instead propose a *synthesis* approach, where low-rankness is imposed to the synthesis coefficients of the data signal over a given t-f dictionary (such as a Gabor frame). As such we offer a novel modeling paradigm that bridges t-f synthesis modeling and traditional analysis-based NMF approaches. The proposed generative model allows in turn to design more sophisticated multi-layer representations that can efficiently capture diverse forms of structure. Additionally, the generative modeling allows to exploit t-f low-rankness for compressive sensing. We present efficient iterative shrinkage algorithms to perform estimation in the proposed models and illustrate the capabilities of the new modeling paradigm over audio signal processing examples.**

I. Introduction

MATRIX factorization methods currently enjoy a large popularity in machine learning and signal processing. In signal processing, the input data is usually a time-frequency (t-f) transform of some original time series $x(t)$. For example, in the audio setting, nonnegative matrix factorization (NMF) is commonly used to decompose magnitude or power spectrograms into elementary components [1]; the spectrogram $\mathbf{P}$ is approximately factorized into $\mathbf{WH}$, where $\mathbf{W}$ is the dictionary matrix collecting spectral patterns in its columns and $\mathbf{H}$ is the activation matrix. The approximate $\mathbf{WH}$ is generally of lower rank than $\mathbf{P}$, unless additional constraints are imposed on the factors. NMF is at the core of many state-of-the-art source separation systems, such as [2], [3].

The spectrogram $\mathbf{P}$ is usually obtained from the short-time Fourier transform $\mathbf{Y}$. The coefficients y_{fn} of $\mathbf{Y}$ are the inner products of $x(t)$ with t-f atoms $\phi_{fn}(t)$, where f and n index frequencies and time frames, respectively, and a common choice is $\mathbf{P} = |\mathbf{Y}|^2$. The STFT coefficients are so-called *analysis* coefficients and as such spectral decomposition by NMF can be viewed as a low-rank time-frequency *analysis* procedure. Leveraging on the potential of *synthesis* models as opposed to analytical ones (see, e.g., [4]–[7]), we propose to explore a dual view of the usual NMF approach and present a new paradigm that we name *low-rank time-frequency synthesis* (LRTFS). In this new paradigm, the signal is decomposed as

$$x(t) = \sum_{fn} \alpha_{fn} \phi_{fn}(t) + e(t) \quad (1)$$

where the synthesis coefficients $\{\alpha_{fn}\}$ are given a low-rank structure such that $|\alpha_{fn}|^2 \approx [\mathbf{WH}]_{fn}$. Formulation (1) provides a generative representation of the raw data $x(t)$ and extends the modeling capacities of standard NMF-based signal decomposition towards more advanced multilayer hybrid decompositions. Having a generative model of the raw data (instead of its transform) is also useful for some inverse problems such as compressive sampling, an application that will be considered in the paper.

The low-rankness of the synthesis coefficients $\{\alpha_{fn}\}$ is induced through a probabilistic model named *Gaussian Composite Model* (GCM) [8]. The GCM underlies Itakura-Saito NMF, a baseline method that will be recalled in Section II. Section III-A presents our new paradigm LRTFS in the general case of complex-valued signals. It also describes a alternate minimization algorithm for maximum joint likelihood estimation of the parameters. Section III-B shows how the methodology for complex signals can be adapted to real-valued signals. Section III describes how LRTFS can accommodate more advanced multilayer decompositions in which every layer can have its own t-f resolution or structure (e.g., a sparse instead of low-rank time-frequency structure). Section V describes a new approach to compressive sampling, that exploits latent low-rank time-frequency structure instead of sparsity, with superior results for the considered type of data. The article is illustrated throughout with experiments on audio signals (the presented methodology is however not limited to audio signals).

This article unifies and continues our work presented in the conference papers [9], [10]. In particular, it provides more detailed experiments and contributes the following novel methodological additions: the case of real-valued signals (which require to properly handle the Hermitian symmetry of their synthesis coefficients) is now rigorously treated in Section III-B, algorithm accelerations are presented in Sections III-A2, and the concept of compressive LRTFS presented in Section V is entirely novel.

II. A baseline: Itakura-Saito NMF and the Gaussian Composite Model (GCM)

NMF was originally designed in a deterministic setting [11]: a measure of fit between $\mathbf{P}$ and $\mathbf{WH}$ is minimized with respect to (w.r.t) $\mathbf{W}$ and $\mathbf{H}$. Choosing the "right" measure for a specific type of data and task is not straightforward. Furthermore, NMF-based spectral decompositions often arbitrarily discard phase information: only the magnitude of the complex-valued short-time Fourier transform (STFT) is considered. To remedy

C. Févotte is with IRIT, Université de Toulouse, CNRS, France.
M. Kowalski is with L2S, Univ Paris-Sud, Gif-sur-Yvette, France.

these limitations, a generative probabilistic latent factor model of the STFT, the GCM, was proposed in [8]. It is defined by

$$y_{fn} \sim N_c(0, [\mathbf{WH}]_{fn}), \tag{2}$$

where N_c refers to the circular complex-valued normal distribution.[1] As shown by Eq. (2), in the GCM the STFT is assumed centered and its variance has a low-rank structure. Many temporal waveforms (such as audio signals) can be assumed zero-mean and this remains true for they Fourier coefficients by linearity, hence the zero-mean assumption of the GCM. The low-rank variance structure on the other hand underlies a composite signal structure that makes the model relevant for decomposition task. Indeed, introducing the latent complex-valued components y_{kfn}, Eq. (2) is equivalent to

$$y_{fn} = \sum_k y_{kfn}, \tag{3}$$

$$y_{kfn} \sim N_c(0, w_{fk}h_{kn}). \tag{4}$$

The latent component $\mathbf{Y}_k$ with coefficients $\{y_{kfn}\}_{fn}$ reflects the contribution of the spectral pattern $\mathbf{w}_k$, the k^{th} column of $\mathbf{W}$, amplitude-modulated in time by the activation coefficients of the k^{th} row of $\mathbf{H}$.

Under these assumptions, the negative log-likelihood $-\log p(\mathbf{Y}|\mathbf{W},\mathbf{H})$ is equal, up to a constant, to the Itakura-Saito (IS) divergence $D_{\text{IS}}(\mathbf{P}|\mathbf{WH})$ between the power spectrogram $\mathbf{P} = |\mathbf{Y}|^2$ and $\mathbf{WH}$, where

$$D_{\text{IS}}(\mathbf{A}|\mathbf{B}) = \sum_{ij} \frac{a_{ij}}{b_{ij}} - \log \frac{a_{ij}}{b_{ij}} - 1 \tag{5}$$

for nonnegative matrices $\mathbf{A}$ and $\mathbf{B}$ [8].

The GCM is a step forward from traditional NMF approaches that fail to provide a valid generative model of the STFT itself – other approaches have only considered probabilistic models of the magnitude spectrogram under Poisson or multinomial assumptions, see [1] for a review. Still, the GCM is not yet a generative model of the raw signal $x(t)$ itself, but of its STFT. LRTFS fills in this ultimate gap.

III. Low-rank time-frequency synthesis (LRTFS)

In this section we first present LRTFS for complex-valued signals, closely following [9]. Then we rigorously address the case of real-valued signals represented as a complex-valued linear combination of complex-valued t-f atoms (such as Gabor atoms) with Hermitian symmetry. Finally, we discuss relevant connections with the state-of-the-art and illustrate the potential of LRTFS on an audio example.

A. Complex-valued signals

1) Model: Let $x(t)$ denote a complex-valued signal of length T and $\{\phi_{fn}(t)\}_{f=1..F,n=1..N}$ denote a dictionary of complex-valued t-f atoms of length T. LRTFS is defined as follows. For $t = 1,\dots,T$, $f = 1,\dots,F$, $n = 1,\dots,N$:

$$x(t) = \sum_{fn} \alpha_{fn}\phi_{fn}(t) + e(t), \tag{6}$$

$$\alpha_{fn} \sim N_c(0, [\mathbf{WH}]_{fn}), \tag{7}$$

$$e(t) \sim N_c(0, \lambda), \tag{8}$$

where $\{\alpha_{fn}\}$ are the complex-valued synthesis coefficients, $\mathbf{W}$ and $\mathbf{H}$ are nonnegative matrices of sizes $F\times K$ and $K\times N$, respectively, and $e(t)$ is an additive complex-valued residual term with Gaussian distribution $N_c(0,\lambda)$. The synthesis coefficients $\{\alpha_{fn}\}$ are furthermore assumed independent given $\mathbf{W}$ and $\mathbf{H}$. The synthesis coefficients are dual of the analysis coefficients, defined by $y_{fn} = \sum_t x(t)\phi_{fn}^H(t)$, where $\cdot^H$ denotes conjugate transpose. IS-NMF assumes that the *analysis* coefficients follow a GCM, see Eq. (2). In contrast, LRTFS assumes that the *synthesis* coefficients follow a GCM, as given by Eq. (7). As announced, LRTFS provides a generative model of the raw data $x(t)$, where IS-NMF only provided a generative model of the transformed data $\mathbf{Y}$.

Let us denote by $\mathbf{x}$ and $\mathbf{e}$ the column vectors of size T with coefficients $x(t)$ and $e(t)$, respectively. Given an arbitrary mapping from $(f,n) \in \{1,\dots,F\}\times\{1,\dots,N\}$ to $m \in \{1,\dots,M\}$, where $M = FN$, we denote by $\boldsymbol{\alpha}$ the column vector of dimension M with coefficients $\{\alpha_{fn}\}_{fn}$. Similarly, we denote by $\boldsymbol{\Phi}$ the matrix of size $T\times M$ with columns $\{\boldsymbol{\phi}_{fn}\}_{fn}$, where $\boldsymbol{\phi}_{fn}$ is the column vector of size T with coefficients $\phi_{fn}(t)$. Finally, let us denote by $\mathbf{v}$ the column vector of dimension M with coefficients $v_{fn} = [\mathbf{WH}]_{fn}$. In the following we will sometimes abuse notations by indexing the coefficients of $\boldsymbol{\alpha}$ or $\mathbf{v}$ by either m or (f,n). It should be understood that m and (f,n) are in one-to-one correspondence and the notation should be clear from the context. In particular we will write $\mathbf{v} = \text{vect}[\mathbf{WH}]$. Equipped with these notations, we may write Eq. (6) and (7) as

$$\mathbf{x} = \boldsymbol{\Phi}\boldsymbol{\alpha} + \mathbf{e}, \tag{9}$$

$$\boldsymbol{\alpha} \sim N_c(\mathbf{0}, \text{diag}(\mathbf{v})), \tag{10}$$

$$\mathbf{e} \sim N_c(\mathbf{0}, \lambda\mathbf{I}_T). \tag{11}$$

Ignoring the low-rank structure of $\mathbf{v}$, Eqs. (9)-(11) resemble sparse Bayesian learning (SBL), as introduced in [12], [13], where it is shown that marginal likelihood estimation of the variance induces sparse solutions of $\mathbf{v}$ (and as a consequence, of $\boldsymbol{\alpha}$). The essential difference between our model and SBL is that the coefficients are no longer unstructured in LRTFS. Indeed, in SBL, each coefficient α_m has a free variance parameter v_m. This property is fundamental to the sparsity-inducing effect of SBL [12]. In contrast, in LRTFS, the variances are now tied together and such that $v_m = v_{fn} = [\mathbf{WH}]_{fn}$.

2) Maximum joint likelihood estimation : We now address the estimation of $\mathbf{W}$, $\mathbf{H}$ and $\boldsymbol{\alpha}$ and possibly λ in LRTFS. We consider maximum joint likelihood estimation (MJLE), also referred to as type-I maximum likelihood estimation in [13].

[1] A random variable x has distribution $N_c(x|\mu,\lambda) = (\pi\lambda)^{-1}\exp -(|x-\mu|^2/\lambda)$ if and only if its real and imaginary parts are independent and with distribution $N(\Re[\mu],\lambda/2)$ and $N(\Im[\mu],\lambda/2)$, respectively.

MJLE relies on the minimization of the following objective function:

$$C_{\text{JL}}(\boldsymbol{\alpha}, \mathbf{W}, \mathbf{H}, \lambda) \stackrel{\text{def}}{=} -\log p(\mathbf{x}, \boldsymbol{\alpha}|\mathbf{W}, \mathbf{H}, \lambda) \tag{12}$$

$$= -\log p(\mathbf{x}|\boldsymbol{\alpha}, \lambda) - \log p(\boldsymbol{\alpha}|\mathbf{W}, \mathbf{H}) \tag{13}$$

$$= \frac{1}{\lambda}\|\mathbf{x} - \boldsymbol{\Phi}\boldsymbol{\alpha}\|_2^2 + \sum_{fn}\left[\frac{|\alpha_{fn}|^2}{[\mathbf{WH}]_{fn}} + \log[\mathbf{WH}]_{fn}\right] + cst \tag{14}$$

$$= \frac{1}{\lambda}\|\mathbf{x} - \boldsymbol{\Phi}\boldsymbol{\alpha}\|_2^2 + D_{\text{IS}}(|\boldsymbol{\alpha}|^2|\mathbf{v}) + \log(|\boldsymbol{\alpha}|^2) + cst \tag{15}$$

where $cst = M \log \pi$ and we recall that $\mathbf{v} = \text{vect}[\mathbf{WH}]$. The first term in Eq. (15) measures the discrepancy between the raw signal and its approximation. The second term ensures that the synthesis coefficients are approximately low-rank.

Another possible estimation procedure for LRTFS is maximum marginal likelihood estimation (MMLE), also referred to as type-II maximum likelihood estimation in [13]. It relies on the minimization of $-\log p(\mathbf{x}|\mathbf{W}, \mathbf{H}, \lambda)$, i.e., involves the marginalization of $\boldsymbol{\alpha}$ from the joint likelihood, following the principle of SBL. We considered MMLE for LRTFS in [9] and presented a valid EM algorithm. However our implementation does not scale with the dimensions involved in signal processing, as it requires solving linear systems of size $\min\{T, M\}$, with T large and smaller than M in typical signal processing settings. Large-scale algorithms for MMLE are left as future work.

3) Alternate minimization algorithm for MJLE : We now describe an alternate minimization algorithm that returns stationary points of $C_{\text{JL}}(\boldsymbol{\theta})$, where $\boldsymbol{\theta} = \{\boldsymbol{\alpha}, \mathbf{W}, \mathbf{H}, \lambda\}$. The optimization of $\boldsymbol{\alpha}$ given the other parameters reduces to

$$\min_{\boldsymbol{\alpha}} \frac{1}{\lambda}\|\mathbf{x} - \boldsymbol{\Phi}\boldsymbol{\alpha}\|_2^2 + \sum_{fn}\frac{|\alpha_{fn}|^2}{[\mathbf{WH}]_{fn}} \tag{16}$$

which defines a convex ridge regression problem. The problem has the closed-form solution

$$\hat{\boldsymbol{\alpha}} = \left[\boldsymbol{\Phi}^H\boldsymbol{\Phi} + \lambda\text{diag}(\mathbf{v})^{-1}\right]^{-1}\boldsymbol{\Phi}^H\mathbf{x} \tag{17}$$

which involves the inversion of a $M \times M$ matrix. The inversion can be reduced to dimension T thanks to the Woodbury identity, but this is still large in signal processing applications. As such, a numeral optimization procedure is to be preferred and several options are available, such as conjugate gradient descent, expectation-minimization (EM), forward-backward optimization or majorization-minimization. The latter three are closely related and lead in the present case to an iterative shrinkage algorithm (ISA) [14], [15]. We used in our implementation a complex-valued version of ISA, similar to the complex-valued cases treated in [16], [17], and using the acceleration described in [18]. This leads to a simple and parameter-free implementation with satisfactory speed of convergence. This in particular results in a faster algorithm than the original EM algorithm presented in our initial contribution [9]. The resulting updates are given in Algorithm 1. The value of the inverse step-size L should be set to the maximum eigenvalue of $\boldsymbol{\Phi}^H\boldsymbol{\Phi}$, i.e., the squared spectral norm of $\boldsymbol{\Phi}$. If this value is not available in closed form or difficult to compute, a larger value $L \geq \|\boldsymbol{\Phi}\|_2^2$ is also permissible but will result in smaller step sizes. In Algorithm 1, the operations $\mathbf{A} \circ \mathbf{B}$, $\mathbf{A}^{\circ p}$ and $\frac{\mathbf{A}}{\mathbf{B}}$ denote entry-wise multiplication, exponentiation and division, respectively.

The optimization of $\mathbf{W}$ and $\mathbf{H}$ given $\boldsymbol{\alpha}$ reduces to

$$\min_{\mathbf{W},\mathbf{H}\geq 0} \sum_{fn} D_{\text{IS}}(|\alpha_{fn}|^2|[\mathbf{WH}]_{fn}) \tag{18}$$

which defines a IS-NMF problem with input matrix $\mathbf{S} = [|\alpha_{fn}|^2]_{fn}$. This a non-convex problem that is generally approached with alternating updates of $\mathbf{W}$ and $\mathbf{H}$ and majorization-minimization (MM) [19]. This results in the multiplicative updates given in Algorithm 1.

Finally, the optimization of λ given $\boldsymbol{\alpha}$ is trivially given by $\hat{\lambda} = \|\mathbf{x} - \boldsymbol{\Phi}\boldsymbol{\alpha}\|/T$. However, the MJLE setting is known to be inefficient for the estimation of both the variance parameters of $\boldsymbol{\alpha}$ and of $\mathbf{e}$, with either $\boldsymbol{\Phi}\hat{\boldsymbol{\alpha}}$ or $\hat{\mathbf{e}}$ capturing most of the signal variance. As such, though the estimation of λ is possible in principle, we will consider λ to be a fixed hyper-parameter in the following.

The objective function C_{JL} being non-convex and because we are using an alternate minimization algorithm, the output of Algorithm 1 depends on the initialization. In all simulations we initialized the synthesis coefficients $\boldsymbol{\alpha}$ with the synthesis coefficients $\boldsymbol{\Phi}^H\mathbf{x}$. The matrices $\mathbf{W}$ and $\mathbf{H}$ are initialized using the absolute values of the complex SVD of the synthesis coefficients [20]. Finally, a tempering strategy with warm restart is used to speed up convergence for small target values of λ. The hyper-parameter λ is set to an arbitrarily large value in the first iterations and is then gradually decreased to its target value, as proposed in [21].

4) Reconstruction of the latent components: LRTFS assumes the synthesis coefficients α_{fn} follow a GCM, see Eq. (7). As such, α_{fn} may be written as a sum of Gaussian latent components, such that $\alpha_{fn} = \sum_k \alpha_{kfn}$, with $\alpha_{kfn} \sim N_c(0, w_{fk}h_{kn})$. Denoting by $\boldsymbol{\alpha}_k$ the column vector of dimension M with coefficients $\{\alpha_{kfn}\}_{fn}$, Eq. (9) may be written as

$$\mathbf{x} = \sum_k \boldsymbol{\Phi}\boldsymbol{\alpha}_k + \mathbf{e} = \sum_k \mathbf{c}_k + \mathbf{e}\ , \tag{19}$$

where $\mathbf{c}_k = \boldsymbol{\Phi}\boldsymbol{\alpha}_k$. The component $\mathbf{c}_k$ is the "temporal expression" of spectral pattern $\mathbf{w}_k$, the k^{th} column of $\mathbf{W}$. Given estimates of $\boldsymbol{\alpha}$, $\mathbf{W}$ and $\mathbf{H}$, the components may be reconstructed in various way. A natural choice is $\hat{\mathbf{c}}_k^{\text{MMSE}} = \boldsymbol{\Phi}\hat{\boldsymbol{\alpha}}_k^{\text{MMSE}}$ with

$$\hat{\boldsymbol{\alpha}}_k^{\text{MMSE}} \stackrel{\text{def}}{=} \mathbb{E}[\boldsymbol{\alpha}_k|\mathbf{x}, \hat{\boldsymbol{\theta}}] = \mathbb{E}[\boldsymbol{\alpha}_k|\hat{\boldsymbol{\alpha}}, \hat{\mathbf{W}}, \hat{\mathbf{H}}] \tag{20}$$

and whose coefficients are given by

$$\hat{\alpha}_{kfn}^{\text{MMSE}} = \frac{\hat{w}_{fk}\hat{h}_{kn}}{[\hat{\mathbf{W}}\hat{\mathbf{H}}]_{fn}}\hat{\alpha}_{fn}. \tag{21}$$

Using this estimate, the latent components are reconstructed by applying a t-f dependent "Wiener mask" to the synthesis coefficients. This procedure and the expression of $\hat{\alpha}_{fkn}^{\text{MMSE}}$ is

Algorithm 1: Alternate minimization for LRTFS

Set $L = \|\mathbf{\Phi}\|_2^2$ (or a larger value)
Compute the synthesis coefficients $\mathbf{y} = \mathbf{\Phi}^H \mathbf{x}$ (with matrix form $\mathbf{Y}$)
Set $\boldsymbol{\alpha}^{(0)} = \mathbf{y}$
Initialize $\mathbf{W}^{(0)}$ and $\mathbf{H}^{(0)}$ with the absolute values of the complex SVD of $\mathbf{Y}$
Set $i = 0$
repeat
  %% Update $\mathbf{W}$ and $\mathbf{H}$ with MM
  Compute spectrogram $\mathbf{S}^{(i)} = \left[|\alpha_{fn}^{(i)}|^2\right]_{fn}$
  Initialize inner loop: $\mathbf{W} = \mathbf{W}^{(i)}$, $\mathbf{H} = \mathbf{H}^{(i)}$
  repeat
    $\mathbf{W} \leftarrow \mathbf{W} \circ \frac{[\mathbf{S}^{(i)} \circ (\mathbf{WH})^{\circ -2}]\mathbf{H}^T}{[(\mathbf{WH})^{\circ -1}]\mathbf{H}^T}$
    $\mathbf{H} \leftarrow \mathbf{H} \circ \frac{\mathbf{W}^T[\mathbf{S}^{(i)} \circ (\mathbf{WH})^{\circ -2}]}{\mathbf{W}^T[(\mathbf{WH})^{\circ -1}]}$
  until *convergence*;
  Leave inner loop: $\mathbf{W}^{(i+1)} = \mathbf{W}$, $\mathbf{H}^{(i+1)} = \mathbf{H}$
  Set $\mathbf{v}^{(i+1)} = \text{vect}[\mathbf{W}^{(i+1)}\mathbf{H}^{(i+1)}]$
  %% Update $\boldsymbol{\alpha}$ with accelerated ISA
  Initialize inner loop: $\mathbf{a}^{(0)} = \mathbf{z}^{(0)} = \boldsymbol{\alpha}^{(i)}$
  Set $j = 0$
  repeat
    % Descend
    $\mathbf{z}^{(j+1/2)} = \mathbf{a}^{(j)} + \frac{1}{L}\mathbf{\Phi}^H(\mathbf{x} - \mathbf{\Phi}\mathbf{a}^{(j)})$
    % Shrink
    $\mathbf{z}^{(j+1)} = \frac{\mathbf{v}^{(i+1)}}{\mathbf{v}^{(i+1)} + \lambda/L} \circ \mathbf{z}^{(j+1/2)}$
    % Accelerate
    $\mathbf{a}^{(j+1)} = \mathbf{z}^{(j+1)} + \frac{j+1}{j+5}(\mathbf{z}^{(j+1)} - \mathbf{z}^{(j)})$
    $j \leftarrow j + 1$
  until *convergence*;
  Leave inner loop: $\boldsymbol{\alpha}^{(i+1)} = \mathbf{z}^{(j+1)}$
until *convergence*;

analog to the standard Wiener estimate of the latent components in IS-NMF applied to $|\mathbf{Y}|^2$ [8], given by

$$\hat{y}_{kfn}^{\text{MMSE}} = \frac{\hat{w}_{fk}\hat{h}_{kn}}{[\hat{\mathbf{W}}\hat{\mathbf{H}}]_{fn}} y_{fn}. \tag{22}$$

The estimate $\hat{\boldsymbol{\alpha}}$ is used as an intermediate variable in the expression of $\hat{\alpha}_{kfn}^{\text{MMSE}}$ given by Eq. (21). Another possible estimate, which marginalizes $\boldsymbol{\alpha}$, is

$$\begin{aligned}\hat{\boldsymbol{\alpha}}_k &= \mathbb{E}[\boldsymbol{\alpha}_k|\mathbf{x}, \hat{\mathbf{W}}, \hat{\mathbf{H}}, \hat{\lambda}] &(23)\\ &= \text{diag}(\mathbf{v}_k)\mathbf{\Phi}^H[\mathbf{\Phi}\text{diag}(\mathbf{v})\mathbf{\Phi}^H + \lambda \mathbf{I}_T]^{-1}\mathbf{x}, &(24)\end{aligned}$$

where $\mathbf{v}_k$ is the vector of dimension M with coefficients $\{w_{fk}h_{kn}\}_{fn}$. The input of the estimator is now the raw data $\mathbf{x}$ which may be more sensible. It however requires the resolution of a large-scale linear system of dimension $\min\{T, M\}$, which again is hardly feasible in usual signal processing scenarios.

B. Real-valued signals

1) Model: In many signal processing settings the data is a real-valued signal $x(t)$ expressed as a linear combination of complex-valued t-f atoms with Hermitian symmetry. More specifically, the dictionary and synthesis coefficients are such that $\boldsymbol{\phi}_{fn} = \boldsymbol{\phi}_{(F-f)n}^*$ and $\alpha_{fn} = \alpha_{(F-f)n}^*$ for $f = 1, \ldots, F/2$ (assuming F to be even-valued for simplicity), where $\cdot^*$ denotes conjugation. Under this particular structure, we have

$$\sum_{f=1}^{F}\sum_{n=1}^{N} \alpha_{fn}\phi_{fn}(t) = \sum_{f=1}^{F/2}\sum_{n=1}^{N} 2\Re[\alpha_{fn}\phi_{fn}(t)] \tag{25}$$

and we define real-valued LRTFS (rLFTS) as follows. For $t = 1, \ldots, T$, $f = 1, \ldots, F/2$, $n = 1, \ldots, N$:

$$\begin{aligned}x(t) &= \sum_{f=1}^{F/2}\sum_{n=1}^{N} 2\Re[\alpha_{fn}\phi_{fn}(t)] + e(t), &(26)\\ \alpha_{fn} &\sim N_c(0, [\mathbf{WH}]_{fn}), &(27)\\ e(t) &\sim N(0, \lambda). &(28)\end{aligned}$$

Note how F now runs from 1 to $F/2$ instead of 1 to F. The synthesis coefficients α_{fn} remain complex-valued and the residual $e(t)$ becomes real-valued. $\mathbf{W}$ and $\mathbf{H}$ are nonnegative matrices of sizes $F/2 \times K$ and $K \times N$, respectively.

Let us now denote by $\underline{\boldsymbol{\alpha}}$ and $\mathbf{v}$ the vectors of dimension $M/2$ with coefficients α_{fn} and $v_{fn} = [\mathbf{WH}]_{fn}$, respectively, and by $\underline{\mathbf{\Phi}}$ the matrix of dimension $T \times M/2$ with columns $\boldsymbol{\phi}_{fn}$, for $f = 1, \ldots, F/2$ and $n = 1, \ldots, N$. With these notations we have $\boldsymbol{\alpha} = [\underline{\boldsymbol{\alpha}}^T, \underline{\boldsymbol{\alpha}}^H]^T$, $\mathbf{\Phi} = [\underline{\mathbf{\Phi}}, \underline{\mathbf{\Phi}}^*]$ and $\mathbf{\Phi}\boldsymbol{\alpha} = 2\Re[\underline{\mathbf{\Phi}}\underline{\boldsymbol{\alpha}}]$. Consequently, we may write Eq. (26)-(28) as

$$\begin{aligned}\mathbf{x} &= 2\Re[\underline{\mathbf{\Phi}}\underline{\boldsymbol{\alpha}}] + \mathbf{e}, &(29)\\ \underline{\boldsymbol{\alpha}} &\sim N_c(\mathbf{0}, \text{diag}(\mathbf{v})), &(30)\\ \mathbf{e} &\sim N(\mathbf{0}, \lambda \mathbf{I}_T). &(31)\end{aligned}$$

2) Estimation: The MJLE objective function for rLRTFS writes

$$\begin{aligned}C_{\text{JL}}^{\Re}(\underline{\boldsymbol{\alpha}}, \mathbf{W}, \mathbf{H}, \lambda) &\stackrel{\text{def}}{=} -\log p(\mathbf{x}, \underline{\boldsymbol{\alpha}}|\mathbf{W}, \mathbf{H}, \lambda) &(32)\\ &= \frac{1}{2\lambda}\|\mathbf{x} - 2\Re[\underline{\mathbf{\Phi}}\underline{\boldsymbol{\alpha}}]\|_2^2 &(33)\\ &\quad + \sum_{f=1}^{F/2}\sum_{n=1}^{N}\left[\frac{|\alpha_{fn}|^2}{[\mathbf{WH}]_{fn}} + \log[\mathbf{WH}]_{fn}\right] + cst &(34)\\ &= \frac{1}{2\lambda}\|\mathbf{x} - 2\Re[\underline{\mathbf{\Phi}}\underline{\boldsymbol{\alpha}}]\|_2^2 + D_{\text{IS}}(|\underline{\boldsymbol{\alpha}}|^2|\mathbf{v}) + \log(|\underline{\boldsymbol{\alpha}}|^2) + cst &(35)\end{aligned}$$

where $cst = \frac{M}{2}\log 2\pi\lambda$. Using an alternate minimization setting like in Section III-A3, the updates of $\mathbf{W}$ and $\mathbf{H}$ are virtually unchanged. They amount to IS-NMF of the matrix form of the synthesis spectrogram $|\underline{\boldsymbol{\alpha}}|^2$ (of size $F/2 \times N$). The update of λ is easily given by $\hat{\lambda} = \|\mathbf{x} - 2\Re[\underline{\mathbf{\Phi}}\underline{\boldsymbol{\alpha}}]\|_2^2/T$, but here again we prefer to treat λ as an hyper-parameter. The update of $\underline{\boldsymbol{\alpha}}$ involves the following minimization problem:

$$\min_{\underline{\boldsymbol{\alpha}} \in \mathbb{C}^{F/2}} F(\underline{\boldsymbol{\alpha}}) \stackrel{\text{def}}{=} \frac{1}{2\lambda}\|\mathbf{x} - \Re[\underline{\mathbf{\Phi}}\underline{\boldsymbol{\alpha}}]\|_2^2 + \sum_{f=1}^{F/2}\sum_{n=1}^{N}\frac{|\alpha_{fn}|^2}{[\mathbf{WH}]_{fn}}. \tag{36}$$

The problem defined by Eq. (36) has a closed-form solution, with a however less simpler expression than Eq. (17). The solution is still computationally demanding and the following numerical procedure is preferable.

Theorem 1 (Iterative shrinking algorithm for rLRTFS)**.** *Let $L = \|\mathbf{\Phi}\|_2^2$ (with $\mathbf{\Phi} = [\underline{\mathbf{\Phi}}, \underline{\mathbf{\Phi}}^*]$) and $\underline{\boldsymbol{\alpha}}^{(0)}$ be an initial estimate. The following sequence of updates converge to the global solution of problem* (36)*:*

$$\underline{\boldsymbol{\alpha}}^{(j+1/2)} = \underline{\boldsymbol{\alpha}}^{(j)} + \frac{1}{L}\underline{\mathbf{\Phi}}^H(\mathbf{x} - 2\Re[\underline{\mathbf{\Phi}}\underline{\boldsymbol{\alpha}}^{(j)}]), \tag{37}$$

$$\underline{\boldsymbol{\alpha}}^{(j)} = \frac{\mathbf{v}}{\mathbf{v} + \lambda/L} \circ \underline{\boldsymbol{\alpha}}^{(j+1/2)}. \tag{38}$$

Proof. The proof consists in reformulating Eq. (36) as a quadratic optimization problem over the real and imaginary parts of $\underline{\boldsymbol{\alpha}}$ and applying ISA. Let $\mathbf{A} = 2[\Re[\underline{\mathbf{\Phi}}], -\Im[\underline{\mathbf{\Phi}}]]$, $\mathbf{b} = [\Re[\underline{\boldsymbol{\alpha}}]^T\Im[\underline{\boldsymbol{\alpha}}]^T]^T$ and $\mathbf{c} = \frac{1}{2}[\mathbf{v}^T\mathbf{v}^T]^T$. Then we may write

$$F(\underline{\boldsymbol{\alpha}}) = F(\mathbf{b}) = \frac{1}{2\lambda}\|\mathbf{x} - \mathbf{A}\mathbf{b}\|_2^2 + \frac{1}{2}\sum_{m=1}^{M}\frac{b_m^2}{c_m} \tag{39}$$

Denoting $L_{\mathbf{A}} = \|\mathbf{A}\|_2^2$, the ISA update for problem (39) writes [14], [15]

$$\mathbf{b}^{(j+1/2)} = \mathbf{b}^{(j)} + \frac{1}{L_{\mathbf{A}}}\mathbf{A}^T(\mathbf{x} - \mathbf{A}\mathbf{b}^{(j)}), \tag{40}$$

$$\mathbf{b}^{(j)} = \frac{\mathbf{c}}{\mathbf{c} + \lambda/L_{\mathbf{A}}} \circ \mathbf{b}^{(j+1/2)}. \tag{41}$$

Using the identities $\mathbf{A}\mathbf{b} = 2\Re[\underline{\mathbf{\Phi}}\underline{\boldsymbol{\alpha}}]$ and

$$\mathbf{A}^T\mathbf{e} = 2\begin{bmatrix}\Re[\underline{\mathbf{\Phi}}^H\mathbf{e}]\\ \Im[\underline{\mathbf{\Phi}}^H\mathbf{e}]\end{bmatrix}, \tag{42}$$

Eqs. (40)-(41) can be rearranged in complex form as

$$\underline{\boldsymbol{\alpha}}^{(j+1/2)} = \underline{\boldsymbol{\alpha}}^{(j)} + \frac{2}{L_{\mathbf{A}}}\underline{\mathbf{\Phi}}^H(\mathbf{x} - 2\Re[\underline{\mathbf{\Phi}}\underline{\boldsymbol{\alpha}}^{(j)}]) \tag{43}$$

$$\underline{\boldsymbol{\alpha}}^{(j)} = \frac{\mathbf{v}}{\mathbf{v} + 2\lambda/L_{\mathbf{A}}} \circ \underline{\boldsymbol{\alpha}}^{(j+1/2)}. \tag{44}$$

To complete the proof we only need to show that $L = 2L_{\mathbf{A}}$. Let $\boldsymbol{\alpha} = [\underline{\boldsymbol{\alpha}}_1^T, \underline{\boldsymbol{\alpha}}_2^T]^T$ be an eigenvector of $\mathbf{\Phi}^H\mathbf{\Phi}$ with maximum eigenvalue $L = \|\mathbf{\Phi}\|_2^2$. By definition we have

$$\begin{bmatrix}\underline{\mathbf{\Phi}}^H\underline{\mathbf{\Phi}} & \underline{\mathbf{\Phi}}^H\underline{\mathbf{\Phi}}^*\\ \underline{\mathbf{\Phi}}^T\underline{\mathbf{\Phi}} & \underline{\mathbf{\Phi}}^T\underline{\mathbf{\Phi}}^*\end{bmatrix}\begin{bmatrix}\underline{\boldsymbol{\alpha}}_1\\ \underline{\boldsymbol{\alpha}}_1\end{bmatrix} = L\begin{bmatrix}\underline{\boldsymbol{\alpha}}_1\\ \underline{\boldsymbol{\alpha}}_1\end{bmatrix}. \tag{45}$$

By taking the conjugate of Eq. (45), we easily show that $[\underline{\boldsymbol{\alpha}}_2^H, \underline{\boldsymbol{\alpha}}_1^H]^T$ is also an eigenvector with eigenvalue L. It follows that $[\underline{\boldsymbol{\alpha}}_1 + \underline{\boldsymbol{\alpha}}_2^H, \underline{\boldsymbol{\alpha}}_2 + \underline{\boldsymbol{\alpha}}_1^H]^T$ is also an eigenvector, which happens to have a Hermitian structure. We may thus impose $\underline{\boldsymbol{\alpha}}_2 = \underline{\boldsymbol{\alpha}}_1^*$ and as such $\boldsymbol{\alpha} = [\underline{\boldsymbol{\alpha}}^T, \underline{\boldsymbol{\alpha}}^H]^T$. Then we have the following series of equivalences:

$$\mathbf{\Phi}^H\mathbf{\Phi}\boldsymbol{\alpha} = L\boldsymbol{\alpha} \iff \underline{\mathbf{\Phi}}^H\underline{\mathbf{\Phi}}\underline{\boldsymbol{\alpha}} + \underline{\mathbf{\Phi}}^H\underline{\mathbf{\Phi}}^*\underline{\boldsymbol{\alpha}}^* = L\underline{\boldsymbol{\alpha}} \tag{46}$$

$$\iff 2\underline{\mathbf{\Phi}}^H\Re[\underline{\mathbf{\Phi}}\underline{\boldsymbol{\alpha}}] = L\underline{\boldsymbol{\alpha}} \tag{47}$$

$$\iff \frac{1}{2}\mathbf{A}^T\mathbf{A}\mathbf{b} = L\mathbf{b}. \tag{48}$$

As such, the spectra of $\underline{\mathbf{\Phi}}^H\underline{\mathbf{\Phi}}$ and $\mathbf{A}^T\mathbf{A}$ coincide up to a factor 2 and we have $L = 2L_{\mathbf{A}}$, which concludes the proof. ∎

Algorithm 2: Alternate minimization for rLRTFS

Set $L = \|\mathbf{\Phi}\|_2^2$ (or a larger value)
Compute the synthesis coefficients $\mathbf{y} = \underline{\mathbf{\Phi}}^H\mathbf{x}$ (with matrix form $\mathbf{Y}$)
Set $\underline{\boldsymbol{\alpha}}^{(0)} = \mathbf{y}$
Initialize $\mathbf{W}^{(0)}$ and $\mathbf{H}^{(0)}$ with the absolute values of the complex SVD of $\mathbf{Y}$
Set $i = 0$
repeat
  %% Update $\mathbf{W}$ and $\mathbf{H}$ with MM
  Compute spectrogram
  $\mathbf{S}^{(i)} = \left[|\underline{\alpha}_{fn}^{(i)}|^2\right]_{f=1,\dots,F/2,n=1,\dots,N}$
  Initialize inner loop: $\mathbf{W} = \mathbf{W}^{(i)}$, $\mathbf{H} = \mathbf{H}^{(i)}$
  repeat
    $\mathbf{W} \leftarrow \mathbf{W} \circ \frac{[\mathbf{S}^{(i)} \circ (\mathbf{W}\mathbf{H})^{\circ -2}]\mathbf{H}^T}{[(\mathbf{W}\mathbf{H})^{\circ -1}]\mathbf{H}^T}$
    $\mathbf{H} \leftarrow \mathbf{H} \circ \frac{\mathbf{W}^T[\mathbf{S}^{(i)} \circ (\mathbf{W}\mathbf{H})^{\circ -2}]}{\mathbf{W}^T[(\mathbf{W}\mathbf{H})^{\circ -1}]}$
  until *convergence*;
  Leave inner loop: $\mathbf{W}^{(i+1)} = \mathbf{W}$, $\mathbf{H}^{(i+1)} = \mathbf{H}$
  Set $\mathbf{v}^{(i+1)} = \text{vect}[\mathbf{W}^{(i+1)}\mathbf{H}^{(i+1)}]$
  %% Update $\boldsymbol{\alpha}$ with accelerated ISA
  Initialize inner loop: $\mathbf{a}^{(0)} = \mathbf{z}^{(0)} = \underline{\boldsymbol{\alpha}}^{(i)}$
  Set $j = 0$
  repeat
    % Descend
    $\mathbf{z}^{(j+1/2)} = \mathbf{a}^{(j)} + \frac{1}{L}\underline{\mathbf{\Phi}}^H(\mathbf{x} - 2\Re[\underline{\mathbf{\Phi}}\mathbf{a}^{(j)}])$
    % Shrink
    $\mathbf{z}^{(j+1)} = \frac{\mathbf{v}^{(i+1)}}{\mathbf{v}^{(i+1)} + \lambda/L} \circ \mathbf{z}^{(j+1/2)}$
    % Accelerate
    $\mathbf{a}^{(j+1)} = \mathbf{z}^{(j+1)} + \frac{j+1}{j+5}(\mathbf{z}^{(j+1)} - \mathbf{z}^{(j)})$
    $j \leftarrow j + 1$
  until *convergence*;
  Leave inner loop: $\underline{\boldsymbol{\alpha}}^{(i+1)} = \mathbf{z}^{(j+1)}$
until *convergence*;

3) Comments about implementation: Eq. (37) and (38) can be accelerated like before and this results in the general procedure summarized in Algorithm 2. As compared to Algorithm 1, $\boldsymbol{\alpha}$ is essentially replaced by $\underline{\boldsymbol{\alpha}}$, of size half, and the expression of $\mathbf{z}^{(i+1)}$ is changed with Eq. (37). Although the same notations are used for convenience, $\mathbf{Y}$, $\mathbf{S}^{(i)}$ and $\mathbf{W}$ become matrices with $F/2$ rows.

Eq. (37) can be read as follows. The operation $2\Re[\underline{\mathbf{\Phi}}\underline{\boldsymbol{\alpha}}]$ consists of reconstructing an approximation $\hat{\mathbf{x}}$ of $\mathbf{x}$ based on the current synthesis coefficients $\underline{\boldsymbol{\alpha}}$. The operation $\underline{\mathbf{\Phi}}^H\mathbf{e}$ then consists in computing the analysis coefficients (restricted to "positive" frequencies, i.e., $f = 1, \dots, F/2$) of the current residual $\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}}$. When $\mathbf{\Phi}$ is a tight Gabor frame, these operations can be efficiently performed with dedicated time-frequency libraries. For example, these operations can efficiently be performed in Matlab using The Large Time-Frequency Analysis Toolbox (LTFAT) [22].[2] When the Gabor

[2]The specific commands being of the like $\underline{\boldsymbol{\alpha}} =$ `dgtreal`($\mathbf{x}$, 'parameters') and $\mathbf{x} =$ `idgtreal`($\underline{\boldsymbol{\alpha}}$, 'parameters'), where `dgt` stands for discrete Gabor transform.

frame is tight, i.e., $\mathbf{\Phi}\mathbf{\Phi}^H\mathbf{x} = \mathbf{x}$, $\mathbf{\Phi}$ has a unit spectral norm and we may set $L = 1$. A Matlab implementation of Algorithm 2 is available online.[3]

Finally, given estimates of $\underline{\boldsymbol{\alpha}}$, $\mathbf{W}$ and $\mathbf{H}$, latent component coefficients $\underline{\hat{\boldsymbol{\alpha}}}_k$ may be reconstructed like in Eq. (21), and then $\hat{\mathbf{c}}_k = 2\Re[\mathbf{\Phi}\underline{\hat{\boldsymbol{\alpha}}}_k]$.

C. Related work

The closest to our work are probably the recent papers by Kameoka [23], [24] which addresses temporal models of the form $\mathbf{x} = \sum_k \mathbf{c}_k$, like Eq. (19), where the spectrograms of the latent components are approximately rank-one. In essence (and slightly simplifying) these papers address optimization problems of the form

$$\min_{\mathbf{c}_k,\mathbf{W},\mathbf{H}} \sum_{fkn} D([|\mathbf{\Phi}^H\mathbf{c}_k|^2]_{fn}|w_{fk}h_{kn}) \quad \text{s.t.} \quad \mathbf{x} = \sum_k \mathbf{c}_k \quad (49)$$

where $|\mathbf{\Phi}^H\mathbf{c}_k|^2$ stands as the spectrogram of $\mathbf{c}_k$, abusively indexed by f and n for clarity, and $D(\cdot|\cdot)$ is a divergence between nonnegative matrices (either the generalized Kullback-Leibler divergence or the quadratic cost in [23], [24]). Though very elegant in our opinion, the approaches of [23], [24] are still analysis-based and do not yet provide a fully generative synthesis-based model like LRTFS.

Another related trend of work are the approaches of [25]–[27] which essentially model $x(t)$ as a sum of variance-structured Gaussian processes. In [25], [26] $x(t)$ is modeled as a short-time stationary process. Each temporal frame n of $\mathbf{x}$ is assumed to follow a Gaussian process with covariance $\sum_k \mathbf{\Sigma}_{kn}$. Based on the Whittle approximation, $\mathbf{\Sigma}_{kn}$ is assumed diagonal in the Fourier basis and its eigenspectrum is approximated by $h_{kn}\mathbf{w}_k$. Using our notations, the model in [27] sets $N = T$ and assumes $x(t) = \sum_f \sqrt{[\mathbf{WH}]_{ft}}\,\Re[s_f(t)e^{-j2\pi ft/F}]$ where $s_f(t)$ is a complex-valued Gaussian autoregressive sequence. The papers [25]–[27] describe NMF-related generative probabilistic models of $x(t)$, like LRTFS. These approaches are rooted in time series analysis while LRTFS offers a different perspective and model, rooted in the sparse approximation literature. In particular, it can be used with any time-frequency dictionary $\mathbf{\Phi}$.

D. Example

We illustrate the performance of LRTFS compared to standard IS-NMF using the piano example used in [8]. The sequence has a simple structure: four notes are played together at once in the first measure and are then played by pairs in all possible combinations in the subsequent measures. The duration is 15.6 s and the sampling rate 22050 Hz. In noise-free conditions and with appropriate initialization, standard IS-NMF is able to recover the four notes, the transient part produced by the hammer and the sound produced by the sustain pedal when it is released [8]. We here consider a noisy example using additive white Gaussian noise with 20 dB input Signal to Noise Ratio (SNR). A tight Gabor dictionary (with

[3]https://www.irit.fr/~Cedric.Fevotte/extras/tsp2018/. Future references to online material refer to this same url.

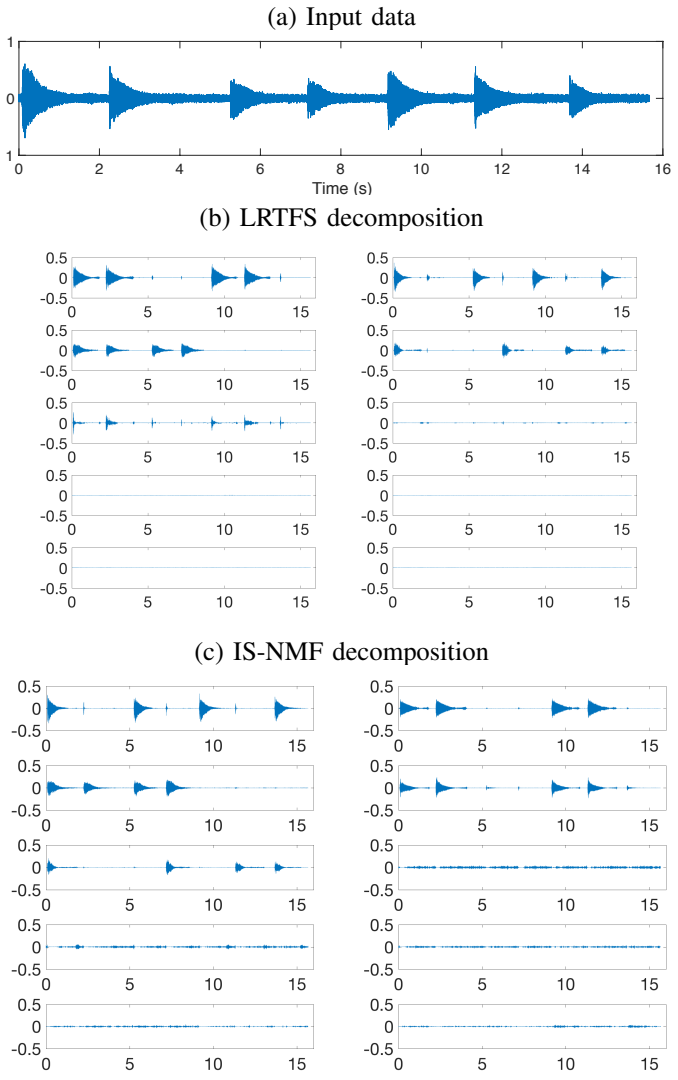

Fig. 1. Decomposition of a piano sequence consisting of four notes. The temporal components are displayed by decreasing energy (from left to right and top to bottom).

Hermitian symmetry) built on a Hann window of 1024 samples (46 ms) with 50% overlap is used for $\mathbf{\Phi}$. IS-NMF is applied to the analysis spectrogram $|\underline{\mathbf{\Phi}}\mathbf{x}|^2$. The number of latent components is arbitrarily set to $K = 10$ for both IS-NMF and rLTFS and the two methods are run from the same initialization (based on the SVD of $\mathbf{Y}$, see Alg. 2). Iteration of the main and inner loops is stopped when the relative error between two successive parameter iterates falls under 10^{-5}. rLTFS is run with six different values of λ logarithmically equally spaced between 10^{-1} and 10^{-6}. We show results corresponding to the value of λ that maximizes the output SNR given by

$$10\log\frac{\|\hat{\mathbf{x}} - \mathbf{x}\|_2^2}{\|\mathbf{x}\|_2^2}. \quad (50)$$

The results are reported in Fig. 1. LRTFS is able to recover the four notes in the first four components, while the fifth component recovers the transient components produced by the hammer and the sustain pedal. The remaining five components

are inaudible because of the denoising performed by LRTFS. In this setting, IS-NMF fails to recover this transient part and splits the first note into two components. The input noise is spread over the five remaining components as expected. Audio files are available online.

IV. Multi-layer LRTFS

Besides the advantage of modeling the raw signal itself, and not its STFT, another major strength of LRTFS is that it offers the possibility of multi-layer modeling. This means we may envisage models of the form

$$\mathbf{x} = \mathbf{x}_\mathrm{a} + \mathbf{x}_\mathrm{b} + \mathbf{e} = \mathbf{\Phi}_\mathrm{a}\boldsymbol{\alpha}_\mathrm{a} + \mathbf{\Phi}_\mathrm{b}\boldsymbol{\alpha}_\mathrm{b} + \mathbf{e} \tag{51}$$

where $\mathbf{x}_\mathrm{a} = \mathbf{\Phi}_\mathrm{a}\boldsymbol{\alpha}_\mathrm{a}$ and $\mathbf{x}_\mathrm{b} = \mathbf{\Phi}_\mathrm{b}\boldsymbol{\alpha}_\mathrm{b}$ are referred to as *layers*. This setting covers a variety of situations. $\mathbf{\Phi}_\mathrm{a}$ and $\mathbf{\Phi}_\mathrm{b}$ may be equal with $\boldsymbol{\alpha}_\mathrm{a}$ and $\boldsymbol{\alpha}_\mathrm{b}$ having a different structure. For example, $\boldsymbol{\alpha}_\mathrm{a}$ may follow a GCM like before and $\boldsymbol{\alpha}_\mathrm{b}$ may be given a sparsity-inducing prior. In such a case, multi-layer LRTFS offers a synthesis perspective to sparse + low-rank spectrogram decompositions, such as those presented in [28]–[30] which propose variants of robust principal component analysis (RPCA) [31] for spectral unmixing. Even more interestingly, the time-frequency dictionaries $\mathbf{\Phi}_\mathrm{a}$ and $\mathbf{\Phi}_\mathrm{b}$ may be chosen with different t-f resolutions. This yields so-called hybrid or morphological decompositions [32], [33], in which each layer may capture specific resolution-dependent structures. A typical audio example is transient + tonal decomposition: transient components are by nature adequately represented by a t-f dictionary with short time resolution while tonal components (such as the sustained parts of musical notes) are better represented by a t-f dictionary with larger time resolution (and as a consequence, finer frequency resolution). A variety of priors can be considered for $\boldsymbol{\alpha}_\mathrm{a}$ and $\boldsymbol{\alpha}_\mathrm{b}$, such as frequency grouping for the transient synthesis coefficients and temporal grouping for the tonal synthesis coefficients [34].

A. Sparse and low-rank time-frequency synthesis

We consider a special case of multi-layer LRTFS that illustrates the potential of the synthesis approach. We present the methodology in the complex case for simplicity, but the results can readily be adapted to the real case following the procedure described in Section III.

1) Model: Let $\mathbf{\Phi}_\mathrm{a}$ and $\mathbf{\Phi}_\mathrm{b}$ be time-frequency dictionaries consisting of atoms $\boldsymbol{\phi}^\mathrm{a}_{fn}(t)$ and $\boldsymbol{\phi}^\mathrm{b}_{fn}(t)$ with common dimension T and t-f pavings of size $F_\mathrm{a}\times N_\mathrm{a}$ and $F_\mathrm{b}\times N_\mathrm{b}$, respectively. We consider the following model, for $t = 1, \dots, T$:

$$x(t) = \sum_{f=1}^{F_\mathrm{a}}\sum_{n=1}^{N_\mathrm{a}} \alpha^\mathrm{a}_{fn}\phi^\mathrm{a}_{fn}(t) + \sum_{f=1}^{F_\mathrm{b}}\sum_{n=1}^{N_\mathrm{b}} \alpha^\mathrm{b}_{fn}\phi^\mathrm{b}_{fn}(t) + e(t) \tag{52}$$

$$\alpha^\mathrm{a}_{fn} \sim N_c(0, [\mathbf{WH}]_{fn}), f = 1, \dots, F_\mathrm{a}, n = 1, \dots, N_\mathrm{a} \tag{53}$$

$$\alpha^\mathrm{b}_{fn} \sim N_c(0, v^\mathrm{b}_{fn}), f = 1, \dots, F_\mathrm{b}, n = 1, \dots, N_\mathrm{b} \tag{54}$$

$$e(t) \sim N_c(0, \lambda) \tag{55}$$

where $\{\alpha^\mathrm{a}_{fn}\}$ and $\{\alpha^\mathrm{b}_{fn}\}$ are the complex-valued synthesis coefficients, $\mathbf{W}$ and $\mathbf{H}$ are nonnegative matrices of sizes $F_\mathrm{a}\times K$ and $K \times N_\mathrm{a}$, respectively, $\{v^\mathrm{b}_{fn}\}$ are nonnegative variance parameters and $e(t)$ is an additive complex-valued residual term. Eq. (52) is nothing but the scalar form of Eq. (51). Eq. (53) defines a GCM, while Eq. (53) defines the sparse-inducing prior that is used in SBL. Like before, we note by $\mathbf{v}^\mathrm{a}$ and $\mathbf{v}^\mathrm{b}$ the column vectors with coefficients $[\mathbf{WH}]_{fn}$ and v^b_{fn}, respectively. Both $\mathbf{v}^\mathrm{a}$ and $\mathbf{v}^\mathrm{b}$ are parameters of a hierarchical variance model. Notice however how $\mathbf{v}^\mathrm{b}$ is a *free* parameter, while $\mathbf{v}^\mathrm{a}$ is structured through $\mathbf{W}$ and $\mathbf{H}$. Overall, Eq. (52)-(55), define a multi-layer LRTFS model with latent low-rank t-f structure for layer $\mathbf{x}_\mathrm{a}$ and latent sparse t-f structure for layer $\mathbf{x}_\mathrm{b}$.

B. Estimation

The negative log-likelihood of the data and parameters in model (52)-(55) is given by

$$\begin{aligned}-\log p(\mathbf{x}, \boldsymbol{\alpha}_\mathrm{a}, \boldsymbol{\alpha}_\mathrm{b}|\mathbf{W}, \mathbf{H}, \mathbf{v}^\mathrm{b}) &= \frac{1}{\lambda}\|\mathbf{x} - \mathbf{\Phi}_\mathrm{a}\boldsymbol{\alpha}_\mathrm{a} - \mathbf{\Phi}_\mathrm{b}\boldsymbol{\alpha}_\mathrm{b}\|_2^2 \\ &+ D_\mathrm{IS}(|\boldsymbol{\alpha}_\mathrm{a}|^2|\mathbf{v}_\mathrm{a}) + \log(|\boldsymbol{\alpha}_\mathrm{a}|^2) \\ &+ D_\mathrm{IS}(|\boldsymbol{\alpha}_\mathrm{b}|^2|\mathbf{v}_\mathrm{b}) + \log(|\boldsymbol{\alpha}_\mathrm{b}|^2) + cst\end{aligned} \tag{56}$$

where $cst = (F_\mathrm{a}N_\mathrm{a} + F_\mathrm{b}N_\mathrm{b})\log\pi$. Unfortunately, and similarly to the difficulty of estimating λ raised in Section III-A2, MJLE fails to evenly distribute the signal variance onto the two layers, and one of the two layers takes it all in practice. Such a problem can be mitigated using MMLE instead of MJLE, but again, MMLE is too costly in our setting. To solve this issue we introduce an extra hyper-parameter μ that balances the contributions of each layer and propose to optimize the following objective

$$C_\mathrm{SLR}(\boldsymbol{\theta}) \overset{\text{def}}{=} \frac{1}{\lambda}\|\mathbf{x} - \mathbf{\Phi}_\mathrm{a}\boldsymbol{\alpha}_\mathrm{a} - \mathbf{\Phi}_\mathrm{b}\boldsymbol{\alpha}_\mathrm{b}\|_2^2 \tag{57}$$

$$\begin{aligned}&+ \mu\left[D_\mathrm{IS}(|\boldsymbol{\alpha}_\mathrm{a}|^2|\mathbf{v}_\mathrm{a}) + \log(|\boldsymbol{\alpha}_\mathrm{a}|^2)\right] \\ &+ (1-\mu)\left[D_\mathrm{IS}(|\boldsymbol{\alpha}_\mathrm{b}|^2|\mathbf{v}_\mathrm{b}) + \log(|\boldsymbol{\alpha}_\mathrm{b}|^2)\right] + cst,\end{aligned} \tag{58}$$

where $0 \le \mu \le 1$, $\boldsymbol{\theta} = \{\boldsymbol{\alpha}_\mathrm{a}, \boldsymbol{\alpha}_\mathrm{b}, \mathbf{W}, \mathbf{H}, \mathbf{v}_\mathrm{b}\}$ is the set of latent variables and parameters and SLR stands for "sparse + low-rank".

We may again find a stationary point of $C_\mathrm{SLR}(\boldsymbol{\theta})$ by alternate minimization. The update of $\mathbf{v}_\mathrm{b}$ is trivially given by $\mathbf{v}_\mathrm{b} = |\boldsymbol{\alpha}_\mathrm{b}|^2$. The update of $\mathbf{W}$ and $\mathbf{H}$ amounts to finding an IS-NMF of the synthesis spectrogram $|\alpha^\mathrm{a}_{fn}|^2$ like in Algorithm 1. The synthesis coefficients $\boldsymbol{\alpha}_\mathrm{a}$ and $\boldsymbol{\alpha}_\mathrm{b}$ may be updated jointly via ridge regression over the joint dictionary $[\mathbf{\Phi}_\mathrm{a}, \mathbf{\Phi}_\mathrm{b}]$. This leads to the following updates

$$\hat{\mathbf{e}}^{(j)} = \mathbf{x} - \mathbf{\Phi}_\mathrm{a}\boldsymbol{\alpha}_\mathrm{a}^{(j)} - \mathbf{\Phi}_\mathrm{b}\boldsymbol{\alpha}_\mathrm{b}^{(j)} \tag{59}$$

$$\boldsymbol{\alpha}_\mathrm{a}^{(j+1/2)} = \boldsymbol{\alpha}_\mathrm{a}^{(j)} + \frac{1}{L}\mathbf{\Phi}_\mathrm{a}^H\hat{\mathbf{e}}^{(j)} \tag{60}$$

$$\boldsymbol{\alpha}_\mathrm{b}^{(j+1/2)} = \boldsymbol{\alpha}_\mathrm{b}^{(j)} + \frac{1}{L}\mathbf{\Phi}_\mathrm{b}^H\hat{\mathbf{e}}^{(j)} \tag{61}$$

$$\boldsymbol{\alpha}_\mathrm{a}^{(j+1)} = \frac{\mathbf{v}_\mathrm{a}}{\mathbf{v}_\mathrm{a} + \lambda/L} \circ \boldsymbol{\alpha}_\mathrm{a}^{(j+1/2)} \tag{62}$$

$$\boldsymbol{\alpha}_\mathrm{b}^{(j+1)} = \frac{\mathbf{v}_\mathrm{b}}{\mathbf{v}_\mathrm{b} + \lambda/L} \circ \boldsymbol{\alpha}_\mathrm{b}^{(j+1/2)} \tag{63}$$

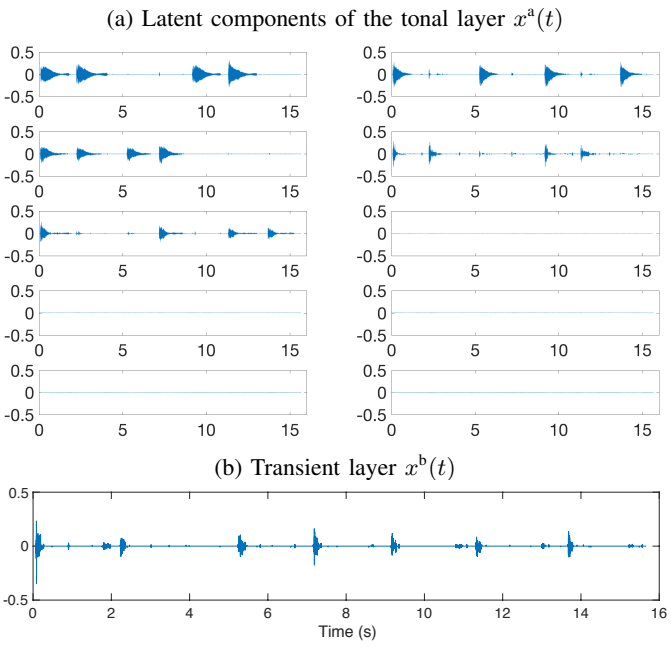

Fig. 2. Two-layer decomposition of the piano sequence displayed in Fig. 1.

where the inverse-step size should satisfy $L \geq \|[\mathbf{\Phi}_{\mathrm{a}}, \mathbf{\Phi}_{\mathrm{b}}]\|_2^2$. A convenient choice is $L = \|\mathbf{\Phi}_a\|_2^2 + \|\mathbf{\Phi}_{\mathrm{b}}\|_2^2$. Eq. (59) computes the current residual, Eqs. (60) and (61) produce a step in the descent direction and Eqs. (62) and (63) shrink the resulting iterates.

C. Example

We use exactly the same data and setting as in Section III-D but we now add a sparse layer $\mathbf{\Phi}_{\mathrm{b}}\boldsymbol{\alpha}_{\mathrm{b}}$ to the LRTFS layer. $\mathbf{\Phi}_{\mathrm{b}}$ is set to be a tight Gabor dictionary built on a Hann window of 128 samples (6 ms) with 50% overlap. $\mathbf{\Phi}_{\mathrm{a}}$ is set as in Section III-D. The parameter μ was experimentally fixed to $\mu = 0.05$, and λ was again chosen among logarithmically spaced vales. Fig. 2 displays the 10 latent components characterizing the tonal layer and the transient layer. The components of the tonal layer are similar to those obtained from the single-layer LRTFS decomposition of Fig. III-D. The fourth component captures part of the hammer attacks (especially from the first, most energetic note) with the shortest resolution components relegated to the transient layer $x^{\mathrm{b}}(t)$ as expected. Audio files are available online.

V. Compressive LRTFS

A striking advantage of LRTFS is that it may be used as a source model in inverse problems. A popular instance is compressive sensing (CS) in which a source signal $x(t)$ must be recovered from $S << T$ random projections. Traditionally, CS exploits the sparsity of the synthesis coefficients of $x(t)$ onto a suitable dictionary. In this section we show that sparsity can efficiently be replaced with low-rankness.

A. Model

Let us denote by $\mathbf{x} \in \mathbb{C}^T$ the vector source signal. The source is assumed to be sensed through the given linear operator $\mathbf{A} \in \mathbb{C}^{S \times T}$, with $S < T$, with output $\mathbf{b} \in \mathbb{C}^S$. We assume the following observation model:

$$\mathbf{b} = \mathbf{A}\mathbf{x} + \mathbf{e} \tag{64}$$

$$= \mathbf{A}\mathbf{\Phi}\boldsymbol{\alpha} + \mathbf{e} \tag{65}$$

where $\mathbf{\Phi} \in \mathbb{C}^{T \times M}$ is a given dictionary and $\boldsymbol{\alpha}$ are the synthesis coefficients of $\mathbf{x}$ and $\mathbf{e}$ is a residual term that accounts for noise or model errors. Where traditional CS assumes some form of sparsity for $\boldsymbol{\alpha}$, we assume the synthesis coefficients to have the LRTFS low-rank structure described by Eq. (7). Like in traditional CS settings, we assume $\mathbf{A}$ to be a random matrix. Finally, we assume $\mathbf{e}$ to follow a complex Gaussian distribution like in Eq. (8).

B. Estimation

MJLE amounts to minimizing the following objective function:

$$C_{\mathrm{CS}}(\boldsymbol{\alpha}, \mathbf{W}, \mathbf{H}) = -\log p(\mathbf{b}, \boldsymbol{\alpha} | \mathbf{W}, \mathbf{H}, \lambda) \tag{66}$$

$$= \frac{1}{\lambda}\|\mathbf{b} - \mathbf{A}\mathbf{\Phi}\boldsymbol{\alpha}\|_2^2 + D_{\mathrm{IS}}(|\boldsymbol{\alpha}|^2 | \mathbf{W}\mathbf{H}) + \log(|\boldsymbol{\alpha}|^2) + cst$$

where $cst = M \log \pi$. The problem of optimizing $C_{\mathrm{CS}}(\boldsymbol{\alpha}, \mathbf{W}, \mathbf{H})$ is equivalent to the one of optimizing $C_{\mathrm{JL}}(\boldsymbol{\alpha}, \mathbf{W}, \mathbf{H})$ given by Eq. (15). In the complex case, the methodology developed in Section III-A3 can be readily applied by replacing $\mathbf{\Phi}$ with $\mathbf{M} = \mathbf{A}\mathbf{\Phi}$. The spectral norm of $\mathbf{M}$ may be difficult to derive or compute and we may set $L = \|\mathbf{A}\|_2^2\|\mathbf{\Phi}\|_2^2$ thanks to the inequality

$$\|\mathbf{A}\mathbf{\Phi}\|_2^2 \leq \|\mathbf{A}\|_2^2\|\mathbf{\Phi}\|_2^2. \tag{67}$$

In the real case, i.e, when $\mathbf{x} \in \mathbb{R}^T$, the methodology developed in Section III-B may again be applied by assuming $\mathbf{A} \in \mathbb{R}^{S \times T}$ and replacing $\underline{\mathbf{\Phi}}$ with $\mathbf{M} = \mathbf{A}\underline{\mathbf{\Phi}}$. Posterior to estimation, an estimate of the original source is given by $\hat{\mathbf{x}} = \mathbf{\Phi}\hat{\boldsymbol{\alpha}}$.

Note that we have addressed compressive sampling of real or complex-valued signals by exploiting a latent NMF-type t-f structure, which is different from compressive sampling of non-negative signals, a topic addressed for example in [35].

C. Example

We evaluate the recovery accuracy of the piano sequence used in Sections III-D and IV-C using a number of measurements S varying increasingly from $T/100$ to $T/10$. For this experiment, the length of the sequence remains 15.6 s but the sampling rate has been fixed at 11025 Hz because of the memory and computational complexities. The Gabor parameters have been adjusted accordingly with a Hann window of length 512 samples (46 ms) with 50% overlap.

We compare CS recovery methods based on LRTFS, SBL and ℓ_1 regularization, using a common alternating minimization setting (only the shrinkage or thresholding operators are changed). Note that we here consider type-I SBL (equivalent to MJLE) and not type-II (which again does not scale with the dimensions of our problem). The algorithms are initialized with $\boldsymbol{\alpha} = \mathbf{0}_{M \times 1}$. The first IS-NMF step of the LRTFS estimation was initialized with the absolute value of the

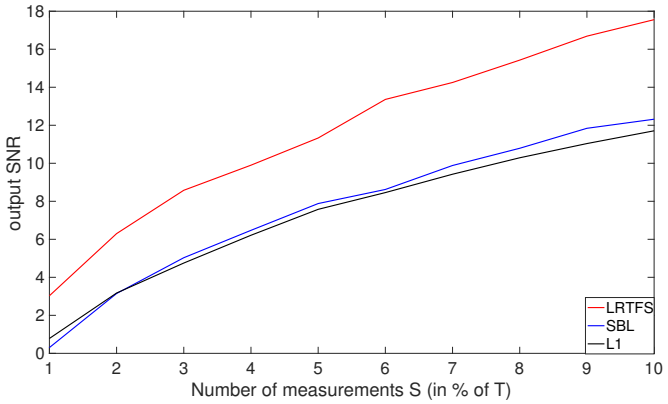

Fig. 3. Recovery of a compressively sensed piano sequence using LRTFS, SBL and ℓ_1 regularization.

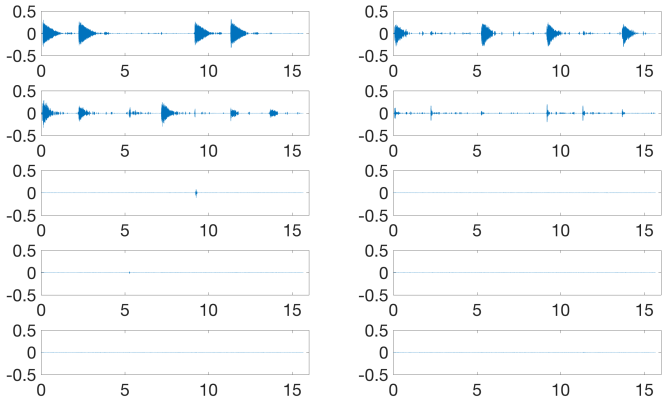

Fig. 4. Latent components of the compressively sensed piano sequence recovered by LRTFS. The temporal components are displayed by decreasing energy (from left to right and top to bottom).

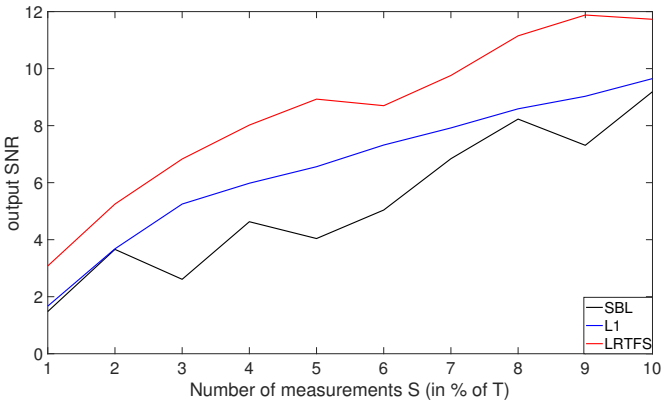

Fig. 5. Recovery of the compressively sensed Mamavatu sequence using LRTFS, SBL and ℓ_1 regularization.

complex-SVD as explained in Sec. III-A3.LRTFS was applied with $K = 10$ and the hyper-parameter λ was incrementally decreased from 10^3 to 10^{-2}.

Estimation accuracy was measured by means of output SNR. The results are displayed in Fig. 3 and show that LRTFS-based recovery improves accuracy by several dB as compared to sparsity-based methods. This means that for such audio signals, there is a significant gain in exploiting low-rankness instead of sparsity for CS. Such a recovery approach is made possible thanks to the generative design of LRTFS. Fig. 4 displays the estimated components $\hat{\mathbf{c}}_k$ returned by LRFTS. It is interesting to note that only 4 components are meaningful. The first two notes are well recovered, like the experiment of Section III-D, see Fig. 1-(b), while the two other notes are mixed in the third component. The fourth component still captures some transient information. We also run experiments for various values of the rank K in the case $S = 0.05\ T$. The recovery results appeared very robust to this parameter. For $K \in \{5, 8, 10, 15, 20, 30\}$ the largest difference in the output SNRs was less than 0.5 dB.

Finally, we run the same CS experiment using the first 12 s of the song *Mamavatu* from Susheela Raman. The excerpt contains acoustic guitar and drums. Output SNRs are displayed on Fig. 5. Again, LRTFS recovery outperforms ℓ_1 and SBL by several dB which confirms the potential of the proposed model for audio inverse problems.

VI. Conclusion

We have presented a new modeling paradigm that bridges t-f synthesis modeling and traditional analysis-based approaches. The proposed generative model allows in turn to design more sophisticated multi-layer representations that can efficiently capture diverse forms of structure. Additionally, the generative modeling allows to exploit NMF-like structure for compressive sensing which, to the best of our knowledge, is entirely new. Maximum joint likelihood estimation in the proposed models can be efficiently addressed using state-of-the art iterative shrinkage and NMF algorithms. They can be efficiently implemented thanks to dedicated time-frequency analysis/synthesis packages. In this paper, we also addressed the modeling and decomposition of real signals in a rigorous way, which was missing from our preliminary contributions and appeared more tricky than initially expected.

The MLJE objective function (15) induced by the proposed generative modeling suggests more general problems of the form

$$C(\boldsymbol{\alpha}, \mathbf{W}, \mathbf{H}, \lambda) = \frac{1}{\lambda}\|\mathbf{x} - \boldsymbol{\Phi}\boldsymbol{\alpha}\|_2^2 + D(|\boldsymbol{\alpha}|^p|\mathbf{v}) \qquad (68)$$

where $\mathbf{v} = \text{vec}[\mathbf{WH}]$, $D(\cdot|\cdot)$ is an arbitrary divergence between nonnegative numbers and p is an arbitrary exponent. $D = D_{\text{IS}}$ and $p = 2$ follow naturally from the GCM assumptions but other choices could be more suitable for other families of signals or images. Such problems do not seem to have been addressed yet in the literature and offer stimulating optimization problems. The exact reconstruction case $\lambda = 0$ is also very interesting in itself.

Another challenging line of research is the design of workable large-scale optimization algorithms for maximum marginal likelihood estimation. As known from [13], such an estimator would be robust to the joint estimation of λ and $\mathbf{v}$, something in which MJLE fails in practice.

References

[1] P. Smaragdis, C. Févotte, G. Mysore, N. Mohammadiha, and M. Hoffman, "Static and dynamic source separation using nonnegative

factorizations: A unified view," *IEEE Signal Processing Magazine*, vol. 31, no. 3, pp. 66–75, May 2014. [Online]. Available: https://www.irit.fr/~Cedric.Fevotte/publications/journals/spm2014.pdf

[2] T. Virtanen, "Monaural sound source separation by non-negative matrix factorization with temporal continuity and sparseness criteria," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15, no. 3, pp. 1066–1074, Mar. 2007.

[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, Mar. 2010. [Online]. Available: https://www.irit.fr/~Cedric.Fevotte/publications/journals/ieee_asl_multinmf.pdf

[4] M. Elad, P. Milanfar, and R. Rubinstein, "Analysis versus synthesis in signal priors," *Inverse problems*, vol. 23, no. 3, p. 947, 2007.

[5] P. Balazs, M. Doerfler, M. Kowalski, and B. Torrésani, "Adapted and adaptive linear time-frequency representations: a synthesis point of view," *IEEE Signal Processing Magazine*, vol. 30, no. 6, pp. 20–31, 2013.

[6] P. Sprechmann, R. Litman, T. B. Yakar, A. M. Bronstein, and G. Sapiro, "Supervised sparse analysis and synthesis operators," in *Advances in Neural Information Processing Systems*, 2013, pp. 908–916.

[7] S. Nam, M. E. Davies, M. Elad, and R. Gribonval, "The cosparse analysis model and algorithms," *Applied and Computational Harmonic Analysis*, vol. 34, no. 1, pp. 30–56, 2013.

[8] C. Févotte, N. Bertin, and J.-L. Durrieu, "Nonnegative matrix factorization with the Itakura-Saito divergence. With application to music analysis," *Neural Computation*, vol. 21, no. 3, pp. 793–830, Mar. 2009. [Online]. Available: https://www.irit.fr/~Cedric.Fevotte/publications/journals/neco09_is-nmf.pdf

[9] C. Févotte and M. Kowalski, "Low-rank time-frequency synthesis," in *Advances in Neural Information Processing Systems (NIPS)*, Dec. 2014. [Online]. Available: https://www.irit.fr/~Cedric.Fevotte/publications/proceedings/nips14.pdf

[10] ——, "Hybrid sparse and low-rank time-frequency signal decomposition," in *Proc. European Signal Processing Conference (EUSIPCO)*, Nice, France, Sep. 2015. [Online]. Available: https://www.irit.fr/~Cedric.Fevotte/publications/proceedings/eusipco15.pdf

[11] D. D. Lee and H. S. Seung, "Learning the parts of objects with nonnegative matrix factorization," *Nature*, vol. 401, pp. 788–791, 1999.

[12] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, vol. 1, pp. 211–244, 2001.

[13] D. P. Wipf and B. D. Rao, "Sparse bayesian learning for basis selection," *IEEE Transactions on Signal Processing*, vol. 52, no. 8, pp. 2153–2164, Aug. 2004.

[14] M. Figueiredo and R. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Transactions on Image Processing*, vol. 12, no. 8, pp. 906–916, 2003.

[15] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[16] L. Chaâri, J.-C. Pesquet, A. Benazza-Benyahia, and P. Ciuciu, "A wavelet-based regularized reconstruction algorithm for SENSE parallel MRI with applications to neuroimaging," *Medical Image Analysis*, vol. 15, no. 2, pp. 185–201, 2011.

[17] A. Florescu, E. Chouzenoux, J.-C. Pesquet, P. Ciuciu, and S. Ciochina, "A majorize-minimize memory gradient method for complex-valued inverse problems," *Signal Processing*, vol. 103, pp. 285–295, 2014.

[18] A. Chambolle and C. Dossal, "On the convergence of the iterates of the fast iterative shrinkage/thresholding algorithm," *Journal of Optimization Theory and Applications*, vol. 166, no. 3, pp. 968–982, 2015.

[19] C. Févotte and J. Idier, "Algorithms for nonnegative matrix factorization with the beta-divergence," *Neural Computation*, vol. 23, no. 9, pp. 2421–2456, Sep. 2011. [Online]. Available: https://www.irit.fr/~Cedric.Fevotte/publications/journals/neco11.pdf

[20] J. Becker, M. Menzel, and C. Rohlfing, "Complex SVD initialization for NMF source separation on audio spectrograms," *Proc. Deutsche Jahrestagung für Akustik (DAGA)*, 2015.

[21] E. T. Hale, W. Yin, and Y. Zhang, "Fixed-point continuation for ℓ_1-minimization: Methodology and convergence," *SIAM Journal on Optimization*, vol. 19, no. 3, pp. 1107–1130, 2008.

[22] Z. Průša, P. L. Sondergaard, N. Holighaus, C. Wiesmeyr, and P. Balazs, "The large time-frequency analysis toolbox 2.0," in *Sound, Music, and Motion, Lecture Notes in Computer Science*. Springer, 2014, pp. 419–442.

[23] H. Kameoka, "Multi-resolution signal decomposition with time-domain spectrogram factorization," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2015.

[24] ——, "Complex NMF with the generalized Kullback-Leibler divergence," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017.

[25] A. Liutkus, R. Badeau, and G. Richard, "Gaussian processes for underdetermined source separation," *IEEE Transactions on Signal Processing*, vol. 59, no. 7, pp. 3155–3167, July 2011.

[26] K. Yoshii, R. Tomioka, D. Mochihashi, and M. Goto, "Infinite positive semidefinite tensor factorization for source separation of mixture signals," in *Proc. International Conference on Machine Learning (ICML)*, 2013, pp. 576–584.

[27] R. E. Turner and M. Sahani, "Time-frequency analysis as probabilistic inference," *IEEE Transactions on Signal Processing*, vol. 62, no. 23, pp. 6171–6183, Dec 2014.

[28] P.-S. Huang, S. D. Chen, P. Smaragdis, and M. Hasegawa-Johnson, "Singing-voice separation from monaural recordings using robust principal component analysis," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2012, pp. 57–60.

[29] Z. Chen and D. P. W. Ellis, "Speech enhancement by sparse, low-rank, and dictionary spectrogram decomposition," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2013.

[30] C. Sun, Q. Zhu, and M. Wan, "A novel speech enhancement method based on constrained low-rank and sparse matrix decomposition," *Speech Communication*, vol. 60, pp. 44–55, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167639314000168

[31] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of ACM*, vol. 58, no. 1, pp. 1–37, 2009.

[32] L. Daudet and B. Torrésani, "Hybrid representations for audiophonic signal encoding," *Signal Processing*, vol. 82, no. 11, pp. 1595 – 1617, 2002.

[33] J.-L. Starck, Y. Moudden, J. Bobin, M. Elad, and D. Donoho, "Morphological component analysis," in *Optics & Photonics 2005*. International Society for Optics and Photonics, 2005, pp. 59 140Q–59 140Q.

[34] M. Kowalski, "Sparse regression using mixed norms," *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 303–324, 2009.

[35] P. D. O'Grady and S. T. Rickard, "Compressive sampling of non-negative signals," in *Proc. IEEE Workshop on Machine Learning for Signal Processing (MLSP)*, 2008.

Cédric Févotte is a CNRS senior researcher at Institut de Recherche en Informatique de Toulouse (IRIT). Previously, he has been a CNRS researcher at Laboratoire Lagrange (Nice, 2013-2016) & Télécom ParisTech (2007-2013), a research engineer at Mist-Technologies (the startup that became Audionamix, 2006-2007) and a postdoc at University of Cambridge (2003-2006). He holds MEng and PhD degrees in EECS from École Centrale de Nantes. His research interests concern statistical signal processing and machine learning, for inverse problems and source separation. He is a member of the IEEE Machine Learning for Signal Processing technical committee and an associate editor for the IEEE Transactions on Signal Processing. In 2014, he was the co-recipient of an IEEE Signal Processing Society Best Paper Award for his work on audio source separation using multichannel nonnegative matrix factorization. He is the principal investigator of the European Research Council project FACTORY (New paradigms for latent factor estimation, 2016-2021).

PLACE
PHOTO
HERE

Matthieu Kowalski received the engineering degree in computer science from the Université de Technologie de Compiègne in 2005, and the master degree in Mathematics Vision and Learning from the Ecole Normale Supérieur, Cachan, the same year. He received the PhD degree in applied mathematics from the University of Provence in 2008. His thesis was axed on sparse time-frequency decompositions. He is now an associate professor at the University of Paris-Sud, in the L2S Lab, and his research focuses on Inverse Problems and structured sparse approximations. He is an elected member of the SPARS Steering committee since 2013.