



**HAL**  
open science

# Hybrid sparse and low-rank time-frequency signal decomposition

Cédric Févotte, Matthieu Kowalski

► **To cite this version:**

Cédric Févotte, Matthieu Kowalski. Hybrid sparse and low-rank time-frequency signal decomposition. 23rd European Signal Processing Conference (EUSIPCO 2015), Aug 2015, Nice, France. hal-01199622

**HAL Id: hal-01199622**

**<https://hal.science/hal-01199622>**

Submitted on 15 Sep 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# HYBRID SPARSE AND LOW-RANK TIME-FREQUENCY SIGNAL DECOMPOSITION

Cédric Févotte\*, Matthieu Kowalski†‡

\* Laboratoire Lagrange (CNRS, OCA & Université Nice Sophia Antipolis), Nice, France

† Laboratoire des Signaux et Systèmes (CNRS, CentraleSupélec, Univ Paris-Sud), Gif-sur-Yvette, France

‡ Parietal project-team (INRIA & CEA), Saclay, France

## ABSTRACT

We propose a new hybrid (or morphological) generative model that decomposes a signal into two (and possibly more) layers. Each layer is a linear combination of localised atoms from a time-frequency dictionary. One layer has a low-rank time-frequency structure while the other as a sparse structure. The time-frequency resolutions of the dictionaries describing each layer may be different. Our contribution builds on the recently introduced Low-Rank Time-Frequency Synthesis (LRTFS) model and proposes an iterative algorithm similar to the popular iterative shrinkage/thresholding algorithm. We illustrate the capacities of the proposed model and estimation procedure on a tonal + transient audio decomposition example.

**Index Terms**— Low-rank time-frequency synthesis, sparse component analysis, hybrid/morphological decompositions, non-negative matrix factorisation.

## 1. INTRODUCTION

Sparse time-frequency signal representations (e.g., [1, 2]) on the one hand and low-rank time-frequency spectrogram factorizations on the other one (e.g., [3]) have attracted significant attention in recent years. Furthermore, so-called hybrid or morphological decompositions have been a popular subtopic of sparse representations [4, 5]. In the latter, the signal is expressed as a sum of elementary atoms from various dictionaries with different characteristics, such as dictionaries of Gabor atoms with different time resolutions.

This work describes a new framework that merges sparse and low-rank time-frequency signal representations for hybrid decomposition. In our framework, the signal is expressed as the sum of two layers (and potentially more); the first layer is a linear combination of time-frequency atoms tied by a low-rank variance structure while the other layer is a linear combination of time-frequency atoms characterised by structured sparsity. This is a relevant modelling for musical signals for example, for which the low-rank layer may appropriately capture the tonal structures while the sparse layer may capture transient sounds or even outliers.

The paper builds on our previous work [6], where a generative low-rank time-frequency signal synthesis model is described. We here further demonstrate the flexibility of the previously proposed model by showing how it can readily be adapted to hybrid decompositions settings with mixed designs.

The paper is organised as follows. Section 2 describes the generative hybrid/morphological model in a statistical setting. Section 3 describes an iterative algorithm for maximum a posteriori estimation in the proposed model. Section 4 illustrates our approach on a tonal + transient decomposition of an audio signal. Section 5 concludes.

## 2. MODEL

### 2.1. Observation model

#### 2.1.1. Notations

Let us denote by  $x(t)$  the observed signal in the time domain, with  $t = 1, \dots, T$ . Our two-layers model writes

$$x(t) = x^{\text{lr}}(t) + x^{\text{sp}}(t) + e(t), \quad (1)$$

where  $x^{\text{lr}}(t)$  and  $x^{\text{sp}}(t)$  refer respectively to the “low-rank” and “sparse” layers, and  $e(t)$  denotes a residual term, accounting for noise and/or modelling errors. Each of the layers is constructed as a linear combination of atoms from two time-frequency dictionaries, which may or may not be identical, such that

$$x^{\text{lr}}(t) = \sum_{f_n} \alpha_{f_n} \phi_{f_n}(t), \quad (2)$$

$$x^{\text{sp}}(t) = \sum_{q_l} \beta_{q_l} \psi_{q_l}(t), \quad (3)$$

where the variables  $\alpha_{f_n}$  and  $\beta_{q_l}$  denote synthesis coefficients and the variables  $\phi_{f_n}(t)$  and  $\psi_{q_l}(t)$  denote dictionary atoms. The variables  $f = 1, \dots, F$  and  $q = 1, \dots, Q$  index frequencies, while the variables  $n = 1, \dots, N$  and  $l = 1, \dots, L$  index time frames. The time-frequency tilings of dimensions  $F \times N$  and  $Q \times L$  characterising each dictionary are typically different. Their numbers of elements are denoted  $M = FN$  and  $P = QL$ .

Authorship based on alphabetical order to reflect an equal contribution.

In the following, we will sometimes more conveniently use the vectorial form of Eq. (1), such that

$$\mathbf{x} = \mathbf{x}^{\text{lr}} + \mathbf{x}^{\text{sp}} + \mathbf{e} = \mathbf{\Phi}\boldsymbol{\alpha} + \mathbf{\Psi}\boldsymbol{\beta} + \mathbf{e}, \quad (4)$$

where  $\mathbf{x}^{\text{lr}}$ ,  $\mathbf{x}^{\text{sp}}$  and  $\mathbf{e}$  are vectors of size  $T$ , with elements  $x^{\text{lr}}(t)$ ,  $x^{\text{sp}}(t)$  and  $e(t)$ , respectively. Given arbitrary mappings from  $(f, n)$  to  $m = 1, \dots, M$  and from  $(q, l)$  to  $p = 1, \dots, P$ ,  $\boldsymbol{\alpha}$  and  $\boldsymbol{\beta}$  are vectors of size  $M$  and  $P$  with elements  $\alpha_{fn}$  and  $\beta_{ql}$ , respectively. Similarly, the matrices  $\mathbf{\Phi}$  and  $\mathbf{\Psi}$ , of size  $T \times N$  and  $T \times P$ , collect the set of time-frequency atoms defining each dictionary.

### 2.1.2. Complex vs real-valued modelling

For generality but also simplicity of exposure, we will consider in this paper the complex-valued case, i.e., all the variables in Eq. (1)–(3) are assumed to be complex-valued. This is obviously relevant in cases where the data is naturally complex-valued, but can be interesting as well in audio cases where the real and imaginary parts of  $x$  can be the left and right channels of a stereophonic recording. However, in most cases  $x(t)$  will be real-valued and it may be more natural to assume all the variables in Eq. (1)–(3) to be real-valued. In such a case, the derivations presented for the complex case readily adapt to the real case.

A more intricate setting occur with some dictionaries, such as Gabor dictionaries (closely connected to the short-time Fourier transform), that are naturally defined in the complex-valued domain though they are commonly used for the representation of real-valued signals. In that case, about half of the dictionary atoms (corresponding to “negative frequencies”) are the complex conjugates of the other half (corresponding to “positive frequencies”). It is thus natural to assume that the synthesis coefficients satisfy the same Hermitian symmetry, automatically leading to a real-valued observation  $x(t)$  (assuming furthermore that the residual  $e(t)$  is real-valued). This setting is slightly more complicated to address but in the end only amounts to estimating synthesis coefficients corresponding to positive frequencies and deduce the other half deterministically by conjugation [7]. Another and closely connected perspective to this setting consists of analysing the complex-valued analytic transform of the real-valued signal and revert to the complex-valued signal setting where the “negative frequencies” synthesis coefficients are constrained to be zero.

## 2.2. Statistical assumptions

### 2.2.1. Low-rank layer modelling

The synthesis coefficients of the “low-rank” layer are assumed to be distributed as

$$p(\alpha_{fn} | [\mathbf{WH}]_{fn}) = N_c(\alpha_{fn} | 0, [\mathbf{WH}]_{fn}), \quad (5)$$

where  $N_c(x | \mu, \lambda)$  is the probabilistic density function of the circular complex Gaussian distribution with mean  $\mu$  and variance  $\lambda$ ,<sup>1</sup> and  $\mathbf{W}$  and  $\mathbf{H}$  are non-negative matrices of dimensions  $F \times K$  and  $K \times N$ .  $\mathbf{W}$  and  $\mathbf{H}$  are free parameters to be estimated and  $K$  is assumed lower than both  $F$  and  $N$ , so that  $\mathbf{WH}$  can be viewed as a low-rank approximation of the “power spectrogram” matrix with coefficients  $|\alpha_{fn}|^2$ . The model (5) was proposed in [6]. It is a “synthesis” version of the statistical model underlying non-negative matrix factorisation with the Itakura-Saito divergence [8].

The synthesis coefficients  $\{\alpha_{fn}\}_{fn}$  are further assumed mutually independent conditionally upon  $\mathbf{WH}$ .

### 2.2.2. Sparse layer modelling

The synthesis coefficients of the “sparse” layer are assumed to follow a distribution that reflects a sparse structure. In the simplest case where the synthesis coefficients are assumed mutually independent a priori, we may for example assume a Laplace distribution of the form

$$p(\beta_{ql} | \lambda_s) = \lambda_s^{-1} \exp[-\lambda_s^{-1} |\beta_{ql}|]. \quad (6)$$

However, it is often more realistic or desirable to assume that the synthesis coefficients bear some structure a priori (so-called structured sparsity) in the form of specific dependences (e.g., group-sparsity). In such cases, the prior for  $\boldsymbol{\beta}$  takes a more complicated expression. We here assume that it can nevertheless be written as

$$p(\boldsymbol{\beta} | \lambda_s) = C \exp[-\lambda_s^{-1} S(\boldsymbol{\beta})], \quad (7)$$

where  $S(\boldsymbol{\beta})$  reflects the structured sparsity assumptions and  $C$  is a normalising constant that may depend on  $\lambda_s$ . A large class of such sparsity-inducing priors is presented in [9]. For example, in the experiments we will consider a group-sparsity prior similar to the group-Lasso.

### 2.2.3. Residual term modelling

The sequence  $e(t)$  is assumed identically and independently distributed with distribution  $N_c(0, \lambda_e)$ .

### 2.2.4. Relation to prior works

As mentioned in the introduction, our work builds on hybrid decomposition frameworks proposed in the sparse literature, e.g., [4, 5], and on other works on audio decomposition via the factorisation of spectrograms [3]. In the latter area, so-called “robust factorisations” have been proposed, e.g., [10–12], where the spectrogram is decomposed into a low-rank part plus a sparse term, that accounts for outliers.

<sup>1</sup>The complex random variable  $x$  has distribution  $N_c(\mu, \lambda)$  if and only if its real and imaginary parts are real-valued independent random variables with Gaussian distributions  $N_c(\text{Re}[\mu], \lambda/2)$  and  $N_c(\text{Im}[\mu], \lambda/2)$ .

For example, in the speech/music separation of [10], the low-rank term accounts for the music part while the sparse component accounts for the more variable speech part. Our contribution bears some resemblance with such works. However fundamental differences are as follows. Firstly, our model is generative of the time-domain signal itself (i.e., of  $x(t)$ ) and not of a pre-computed spectrogram, and secondly, different time-frequency dictionaries can possibly be used to model each of the two terms.

### 3. ESTIMATION

#### 3.1. Maximum joint likelihood estimation

We address estimation in light of maximum joint likelihood and as such, we consider the optimisation of

$$\begin{aligned} C_{\text{JL}}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{W}, \mathbf{H}) &= -\log p(\mathbf{x}, \boldsymbol{\alpha}, \boldsymbol{\beta} | \mathbf{W}, \mathbf{H}, \lambda_e, \lambda_s) \\ &= -\log p(\mathbf{x} | \boldsymbol{\alpha}, \boldsymbol{\beta}, \lambda_e) - \log p(\boldsymbol{\alpha} | \mathbf{W}\mathbf{H}) - \log p(\boldsymbol{\beta} | \lambda_s), \end{aligned}$$

subject to non-negativity of the matrices  $\mathbf{W}$  and  $\mathbf{H}$ . The scale parameters  $\lambda_e, \lambda_s$  are here treated as fixed hyper-parameters. After rearrangements of these hyper-parameters, the minimisation of the joint likelihood is equivalent to the minimisation of

$$\begin{aligned} C(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{W}, \mathbf{H}) &= \frac{1}{2} \|\mathbf{x} - \boldsymbol{\Phi}\boldsymbol{\alpha} - \boldsymbol{\Psi}\boldsymbol{\beta}\|_2^2 + \sigma(1 - \gamma)S(\boldsymbol{\beta}) \\ &\quad + \sigma\gamma \sum_{fn} \frac{|\alpha_{fn}|^2}{[\mathbf{W}\mathbf{H}]_{fn}} + \sigma\gamma \sum_{fn} \log[\mathbf{W}\mathbf{H}]_{fn} \quad (8) \\ &= \frac{1}{2} \|\mathbf{x} - \boldsymbol{\Phi}\boldsymbol{\alpha} - \boldsymbol{\Psi}\boldsymbol{\beta}\|_2^2 + \sigma(1 - \gamma)S(\boldsymbol{\beta}) \\ &\quad + \sigma\gamma \sum_{fn} D_{\text{IS}}(|\alpha_{fn}|^2 | [\mathbf{W}\mathbf{H}]_{fn}) + \sigma\gamma \sum_{fn} \log |\alpha_{fn}|^2 + M \quad (9) \end{aligned}$$

where  $D_{\text{IS}}(x|y) = x/y - \log(x/y) - 1$  is the Itakura-Saito divergence between non-negative numbers, and  $0 \leq \gamma \leq 1$ ,  $\sigma \geq 0$  are the two new hyper-parameters. The objective function describes a least-squares problem with sparse penalisation on  $\boldsymbol{\beta}$  and low-rank penalisation on  $\boldsymbol{\alpha}$  by means of the Itakura-Saito divergence. The two equivalent formulations (8) and (9) exhibit two kinds of structure. With  $\mathbf{W}, \mathbf{H}$  (and  $\boldsymbol{\beta}$ ) fixed, the formulation of (8) describes a least-squares regression problem with weighted  $\ell_2$ -regularisation on  $\boldsymbol{\alpha}$ . Alternatively, with  $\boldsymbol{\alpha}$  (and  $\boldsymbol{\beta}$ ) fixed, the formulation of (9) reduces to an Itakura-Saito NMF problem.

#### 3.2. Algorithm

Various approaches can be considered to solve (8) or (9). We here adopt the Expectation-Minimization (EM) approach presented in [6, 13], which leads to the classical forward-backward algorithm [14] in the convex case. The resulting procedure is an iterative shrinkage algorithm described in

Alg. 1 that updates the synthesis coefficients for the two layers and the matrices  $\mathbf{W}, \mathbf{H}$  in turn. In short, the algorithm performs a gradient descent on the synthesis coefficients (jointly for the two layers) and then shrinks or thresholds the updated coefficients according on their prior low-rank or sparse prior. The matrices  $\mathbf{W}$  and  $\mathbf{H}$  are updated after each update of the synthesis coefficients, by solving a regular IS-NMF problem.

---

#### Algorithm 1: Iterative sparse & low-rank signal decomposition

---

Initialisation:  $L \geq \|\boldsymbol{\Phi}\|^2 \|\boldsymbol{\Psi}\|^2$ ,  $\boldsymbol{\alpha} \in \mathbb{C}^M$ ,  $\boldsymbol{\beta} \in \mathbb{C}^P$ ,  
 $\mathbf{W} \in \mathbb{R}_+^{F \times K}$ ,  $\mathbf{H} \in \mathbb{R}_+^{K \times N}$ ,  $\sigma > 0$ ,  $\gamma > 0$

**repeat**

```

% Gradient descent
 $\mathbf{r}^{(i-1)} = \boldsymbol{\Phi}\boldsymbol{\alpha}^{(i-1)} + \boldsymbol{\Psi}\boldsymbol{\beta}^{(i-1)}$ ;
 $\mathbf{z}_{\text{lr}}^{(i)} = \boldsymbol{\alpha}^{(i-1)} + \frac{1}{L}\boldsymbol{\Phi}^H(\mathbf{x} - \mathbf{r}^{(i-1)})$ ;
 $\mathbf{z}_{\text{sp}}^{(i)} = \boldsymbol{\beta}^{(i-1)} + \frac{1}{L}\boldsymbol{\Psi}^H(\mathbf{x} - \mathbf{r}^{(i-1)})$ ;
% Shrinkage for low-rank layer
 $\forall(f, n), \alpha_{fn}^{(i)} = \frac{[\mathbf{W}^{(i-1)}\mathbf{H}^{(i-1)}]_{fn}}{[\mathbf{W}^{(i-1)}\mathbf{H}^{(i-1)}]_{fn} + \sigma\gamma/L} z_{fn}^{(i)}$ ;
% Low-rank parameters update
 $\{\mathbf{W}^{(i)}, \mathbf{H}^{(i)}\} =$ 
arg min  $\sum_{fn} D_{\text{IS}}(|\alpha_{fn}^{(i)}|^2 | [\mathbf{W}\mathbf{H}]_{fn})$ ;
 $\mathbf{W}, \mathbf{H} \geq 0$ 
% Shrinkage for sparse layer
 $\boldsymbol{\beta}^{(i)} = \text{prox}_{\frac{\sigma(1-\gamma)}{L}S}(\mathbf{z}_{\text{sp}}^{(i)})$ 

```

**until convergence;**

---

In Alg. 1, we denote by  $\|\cdot\|$  the spectral norm, and by  $\text{prox}_{\frac{\gamma}{L}S}(\mathbf{z}) = \arg \min_{\boldsymbol{\beta}} \frac{1}{2} \|\mathbf{z} - \boldsymbol{\beta}\|_2^2 + \frac{\gamma}{L}S(\boldsymbol{\beta})$  the proximity operator associated with the function  $\frac{\gamma}{L}S$ , see, e.g., [14]. As mentioned in Section 2.2.2, one can choose any sparse regulariser on  $\boldsymbol{\beta}$ . In practice, the proximal step associated with this type of regulariser reduces to a shrinkage or thresholding step. One can also choose the expected shrinkage/proximal operator directly, instead of designing the regulariser [15]. For example, if the soft-thresholding operator is associated to the  $\ell_1$  regulariser, one can choose instead the empirical-Wiener/non-negative garrotte shrinkage operator. Indeed, the soft-thresholding operator is known to induce a bias in the estimation of large coefficients. This bias can be limited by designing nice shrinkage/thresholding operator (see, e.g., [16, 17] for a review of such operators).

#### 3.3. Estimation of the “rank-1” components

As explained in [6, 8], the low-rank model described by Eq. (5) underlies latent components with “rank-1” variance. Indeed, the model is equivalent to  $\alpha_{fn} = \sum_k \alpha_{k,fn}$  with  $\alpha_{k,fn} \sim N_c(0, w_{fk}h_{kn})$ . Given an estimate of  $\alpha_{fn}$ , the latent “rank-1” components can be estimated by Wiener fil-

tering, such that

$$\hat{\alpha}_{k,fn} = \frac{w_{fk}h_{kn}}{\sum_k w_{fk}h_{kn}} \hat{\alpha}_{fn}. \quad (10)$$

The set of synthesis coefficients  $\{\alpha_{k,fn}\}_{fn}$  form the time-frequency expression of spectral pattern  $\mathbf{w}_k$ , the  $k^{th}$  column of  $\mathbf{W}$ . Its temporal expression is given by  $x_k^{\text{lr}}(t) = \sum_{fn} \alpha_{k,fn} \phi_{fn}(t)$ . Note that, with these notations, our signal model can be further written as  $x(t) = \sum_k x_k^{\text{lr}}(t) + x^{\text{sp}}(t)$ .

#### 4. EXPERIMENTS

In order to illustrate the interest of the proposed model, we propose to perform a tonal + transient decomposition of a 6-seconds musical signal sampled at 20 kHz. Tonal + transient decompositions have enjoyed a large interest in audio signal processing, for example for audio coding [4, 18] or audio restoration [19]. See also [20] for a review.

The data is an excerpt from the song *Mamavatu* from Susheela Raman. It is composed of a guitar melody accompanied by percussions. As such, we may expect to capture the guitar in the low-rank tonal layer, and the percussive part in the sparse transient layer, similar to [21]. The experimental parameters are specified as follows:

- the Gabor dictionary for the low-rank tonal layer is built from a 51-ms Hann window with 50% overlap,
- the Gabor dictionary for the sparse transient layer is built from a 6-ms Hann window with 50% overlap,
- the number of components in the low-rank tonal layer is set to  $K = 3$ ,
- the shrinkage/thresholding operator used for the sparse transient layer is the group non-negative garrotte (aka, group empirical-Wiener) shrinkage operator, which can provide sharp transient estimates:

$$\alpha_{fn} \leftarrow \alpha_{fn} \left( 1 - \frac{\gamma^2}{\sum_f |\alpha_{fn}|^2} \right)^+, \quad (11)$$

- the hyper-parameters are chosen as  $\gamma = 5 \cdot 10^{-3}$  and  $\sigma = 1 \cdot 10^{-3}$ , which led to a satisfactory tonal + transient decomposition in practice, with an almost perfect reconstruction.

In the following, we compare the tonal + transient decomposition obtained by the proposed low-rank (LR) + sparse model with the decomposition obtained by the two-layers low-rank model proposed in our previous work [6]. We display on Fig. 1 the time-frequency synthesis coefficients estimated by the two models. One can see that, thanks to the chosen group-sparse structure, the transients estimated by the proposed model are well localised in time, and that all the frequencies are active in each group. When one listens to the estimated layer, the transients given by the two-layers low-rank model are far less satisfactory.

We also give for comparison the three “rank-1” component obtained for the tonal layer by the proposed approach, by the two-layers low-rank model [6] and by classic IS-NMF [8] used on the analysis coefficients computed with the Gabor dictionary chosen for the tonal layer on Figure 2. On the first and the second components, one can recognise the “stochastic” nature of the percussions. This figure shows that the LR + sparse decomposition helps the estimation of the tonal layer: the third component corresponds to the guitar and is much cleaner with the low-rank + sparse model. Indeed, with the two-layers LR model and the IS-NMF, the guitar is split between at least two rank-1 components.

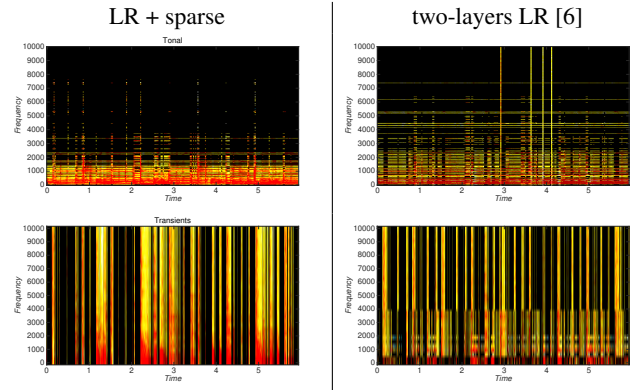


Fig. 1. Power spectrograms  $\{|\alpha_{fn}|^2\}_{fn}$  and  $\{|\beta_{ql}|^2\}_{ql}$  of the tonal (top) and transient (bottom) layers.

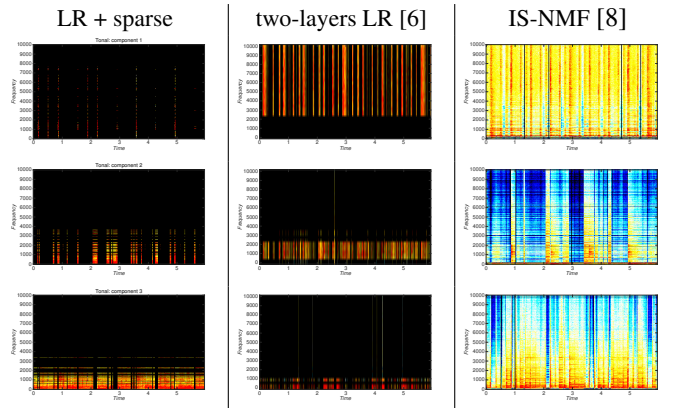


Fig. 2. Power spectrograms  $\{|\alpha_{k,fn}|^2\}_{fn}$  of the “rank-1” components obtained by three methods.

#### 5. CONCLUSION

We have presented a new hybrid/morphological model that allows to mix low-rank and sparse time-frequency designs, in a generative framework that can accommodate multiple time-frequency resolutions. We proposed an iterative estimation procedure similar to forward-backward. The benefit

of our new modelling paradigm is demonstrated on a classic transient + tonal audio decomposition example, in which sharper transients can be estimated thanks to the structured sparse prior.

## REFERENCES

- [1] S. S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic decomposition by basis pursuit,” *SIAM journal on scientific computing*, vol. 20, no. 1, pp. 33–61, 1998.
- [2] P. Balazs, M. Dörfler, M. Kowalski, and B. Torrèsani, “Adapted and adaptive linear time-frequency representations: a synthesis point of view,” *IEEE Signal Processing Magazine*, vol. 2013, pp. 20–31, 2013.
- [3] 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.
- [4] L. Daudet and B. Torrèsani, “Hybrid representations for audiophonic signal encoding,” *Signal Processing*, vol. 82, no. 11, pp. 1595 – 1617, 2002.
- [5] J.-L. Starck, F. Murtagh, and J. Fadili, *Sparse image and signal processing: wavelets, curvelets, morphological diversity*, Cambridge University Press, 2010.
- [6] C. Févotte and M. Kowalski, “Low-rank time-frequency synthesis,” in *Advances in Neural Information Processing Systems (NIPS)*, 2014.
- [7] P. J. Wolfe, S. J. Godsill, and W.-J. Ng, “Bayesian variable selection and regularization for time–frequency surface estimation,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 66, no. 3, pp. 575–589, 2004.
- [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.
- [9] A. Lee, F. Caron, A. Doucet, and C. Holmes, “A hierarchical bayesian framework for constructing sparsity-inducing priors,” Tech. Rep., arXiv, 2010.
- [10] 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.
- [11] 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.
- [12] 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.
- [13] 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.
- [14] P. L. Combettes and J.-C. Pesquet, “Proximal splitting methods in signal processing,” in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H. H. Bauschke, R. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, Eds., Springer Optimization and Its Applications, pp. 185–212. Springer New York, 2011.
- [15] M. Kowalski, “Thresholding rules and iterative shrinkage/thresholding algorithm: A convergence study,” in *International Conference on Image Processing (ICIP) 2014*, 2014.
- [16] A. Antoniadis, “Wavelet methods in statistics: Some recent developments and their applications,” *Statistics Surveys*, vol. 1, pp. 16–55, 2007.
- [17] Y. She, “Thresholding-based iterative selection procedures for model selection and shrinkage,” *Electronic Journal of Statistics*, vol. 3, 2009.
- [18] S. Molla and B. Torrèsani, “A hybrid scheme for encoding audio signal using hidden markov models of waveforms,” *Applied and Computational Harmonic Analysis*, vol. 18, no. 2, pp. 137–166, 2005.
- [19] C. Févotte, B. Torrèsani, L. Daudet, and S. J. Godsill, “Sparse linear regression with structured priors and application to denoising of musical audio,” *IEEE Transactions on Audio, Speech and Language Processing*, vol. 16, no. 1, pp. 174–185, Jan. 2008.
- [20] L. Daudet, “A review on techniques for the extraction of transients in musical signals,” in *Computer Music Modeling and Retrieval*, pp. 219–232. Springer, 2006.
- [21] M. Kowalski and B. Torrèsani, “Identification de signaux parcimonieux dans un dictionnaire hybride,” in *GRETSI 2007*, 2007, p. 1249.