



HAL
open science

Spectral Estimation for Multivariate Locally Time-Warped Signals

Adrien Meynard

► **To cite this version:**

Adrien Meynard. Spectral Estimation for Multivariate Locally Time-Warped Signals. IEEE Signal Processing Letters, 2020, 27, pp.311 - 315. 10.1109/LSP.2020.2970344 . hal-02455218

HAL Id: hal-02455218

<https://hal.science/hal-02455218>

Submitted on 25 Jan 2020

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.

Spectral Estimation for Multivariate Locally Time-Warped Signals

Adrien Meynard

Abstract—Spectral estimation generally aims at determining from a single realization of a given signal, the distribution of its power as a function of frequency. In this paper, we focus on multivariate, locally time-warped signals. We show that the spectral estimation problem can also be interpreted as a doubly nonstationary blind source separation (BSS) problem, where both the mixing matrix and the original sources contribute to nonstationarity. We then introduce a BSS algorithm for joint spectral estimation of nonstationary sources. The performance of the proposed approach is evaluated on numerical simulations, and compared with other nonstationary BSS algorithms.

Index Terms—Nonstationarity, spectral analysis, blind source separation, time warping, wavelet transform.

I. INTRODUCTION

IN time series analysis or statistical signal processing, *spectral estimation* is generally understood as the problem of estimating the frequency content of a signal. Many spectral estimation techniques have been proposed in this context, that can handle estimation from a single signal realization [1]. A familiar example is the so-called Welch estimator. In a more general nonstationary context, *spectral estimation* can also be interpreted as the problem of estimating and diagonalizing the covariance operator of the signal. Both notions coincide in the stationary case, where the covariance operator is a convolution operator, which is diagonalized by a suitable version of the Fourier transform. To depart from the stationary situation, several classes of locally stationary random processes have been proposed and studied in the literature [2], [3], [4], [5], each of them being characterized by a specific form of covariance operator. In this paper, we focus on a specific, physically relevant model of nonstationarity: the *locally deformed signals*. In a few words, those signals are stationary signals transformed by some stationarity-breaking operator. In particular, we focus on *time warping*, which is a specific type of deformation introduced in [6]. Such a model, combined with amplitude modulation, can account for physical phenomena as diverse as Doppler effect, speed variations of an engine, animal vocalization [7], or speech. The covariance operator of locally deformed signals only depends on the spectrum of the underlying stationary signal and the deformation operator. In this case, spectral analysis can be regarded as the joint estimation of these two quantities. We introduced in [8] the JEFAS algorithm (*Joint Estimation of Frequency Amplitude and Spectrum*), which performs this estimation.

In this paper, we extend the results of [8] to the multivariate case where the amplitude modulation is matrix-valued and

mixes decorrelated locally time-warped signals. We formulate this problem as a joint *blind source separation (BSS)* and *nonstationary spectral analysis* problem, which we call *doubly nonstationary BSS*: sources and mixing matrix are nonstationary.

Most BSS algorithms assume stationarity of the sources, and often also address instantaneous linear mixtures, where the mixing matrix is constant over time (see [9] for a thorough review of the field). For such problems, SOBI [10] is a reference approach that exploits second-order statistics and performs approximate joint diagonalization of covariance matrices for the estimation. Several adaptations to the BSS of nonstationary signals mixed by a constant matrix have been proposed. Many of them are based on quadratic time-frequency analysis [11], [12] and rely on a selection of specific points of the time-frequency domain where a single source is active. In this paper we will use a specific one of these termed QTF-BSS, for comparison. BSS algorithms relying on local approximations of the observations by piecewise stationary signals have also been studied [13]. These methods include the piecewise SOBI estimation on non-overlapping subintervals, called p-SOBI in the article. Models of non-instantaneous mixtures have also been considered. In [14], the authors review the general case of time-varying convolutive mixing (including instantaneous time-varying mixing). The proposed method assumes the existence of a sparse representation of the sources, which is not relevant for the situation considered here.

We tackle here the *doubly nonstationary BSS* problem, in the framework of Gaussian models. Based on prior works on locally time-warped signals [15], [8], we use an approximation of the wavelet transforms of the observations, and formulate an approximate maximum likelihood approach for the joint estimation of the mixing matrix and the time warping functions. The corresponding BSS algorithm consists in the alternate estimation of these two quantities, the latter being estimated via the above mentioned JEFAS method. The resulting algorithm is termed JEFAS-BSS. Its performance is evaluated on a synthetic example.

II. MODEL OF NONSTATIONARITY

A. A class of nonstationary signals

In the following, we consider nonstationary signals generated by deformation of stationary signals. Various classes of deformation operators have been studied before: time warping, frequency modulation [15], or amplitude modulation [8]. As mentioned above, we focus here on time warping combined with amplitude modulation, which allows addressing physically relevant situations.

More precisely, let x denote a stationary signal, modeled as a realization of a stationary random process X with power spectrum \mathcal{S}_X . Acting on x with a time warping operator \mathcal{D}_γ yields a signal y given by

$$y(t) = \mathcal{D}_\gamma x(t) \triangleq \sqrt{\gamma'(t)} x(\gamma(t)), \quad (1)$$

where $\gamma \in C^2$ is a strictly increasing smooth function. If γ is not an affine function, y is nonstationary. Time warping is combined with an amplitude modulation operator \mathcal{A}_α , the observed signal is then of the form

$$z(t) = \mathcal{A}_\alpha y(t) \triangleq \alpha(t) y(t) \quad \text{where } \alpha \in C^1. \quad (2)$$

In [8], we showed that the wavelet transform is a natural tool to analyze such signals. Given a wavelet function ψ , let \mathcal{W}_x denote the wavelet transform of a signal x , defined as

$$\mathcal{W}_x(s, \tau) = \int_{\mathbb{R}} x(t) q^{-\frac{s}{2}} \overline{\psi} \left(\frac{t - \tau}{q^s} \right) dt \quad \text{with } q > 1. \quad (3)$$

When the deformation functions α and γ' are slowly varying compared to variations of x , the respective wavelet transforms \mathcal{W}_z and \mathcal{W}_x of z and x are approximately related by:

$$\mathcal{W}_z(s, \tau) \approx \alpha(\tau) \mathcal{W}_x(s + \log_q(\gamma'(\tau)), \gamma(\tau)). \quad (4)$$

This mapping $\mathcal{W}_x \mapsto \mathcal{W}_z$ can be interpreted in terms of local time-scale shifts of the wavelet coefficients: the time warping function γ' yields a translation along the scale axis, while the amplitude modulation function α modifies the magnitude. In the setting of random processes, we showed (Theorem 1 in [8]) that the approximation error in (4) is a zero-mean complex random field, whose variance is controlled by the decay properties of the wavelet ψ , and the slowness of variations of γ' and α .

B. Nonstationary instantaneous mixture

From now on, we assume that N_z observed signals $z_i(t)$ are available, each of which is an instantaneous linear mixture of N_y nonstationary signals, called sources, modeled as in equation (1). The number of mixtures is assumed to be at least equal to the number of sources (i.e. $N_z \geq N_y$). The under-determined case (i.e. $N_z < N_y$) is out of the scope of the present work. In addition, these sources are assumed to be uncorrelated.

More formally, let $\mathbf{y}(t) \in \mathbb{R}^{N_y}$, $\mathbf{z}(t) \in \mathbb{R}^{N_z}$ denote the column vectors containing respectively all the nonstationary sources and observations at time t . Then, the linear instantaneous mixture is written as follows:

$$\mathbf{z}(t) = \mathbf{A}(t) \mathbf{y}(t), \quad (5)$$

where $\mathbf{A}(t) \in \mathbb{R}^{N_z \times N_y}$ denotes the mixing matrix at time t . This model extends the amplitude modulation model (2), which corresponds to the special case where all matrices $\mathbf{A}(t)$ are diagonal. The aim of the spectral analysis of the multivariate locally deformed signal \mathbf{z} is to determine jointly:

- the mixing matrices $\mathbf{A}(t)$;
- the time warping functions γ_i ;
- the spectra of the stationary sources \mathcal{S}_{X_i} .

In other words, we aim at performing a *doubly nonstationary blind source separation*. Indeed, this problem consists of the estimation of the nonstationary mixing matrix $\mathbf{A}(t)$ and the spectral analysis of the nonstationary sources \mathbf{y} .

Let us consider a fixed time τ . For each observation z_i , we denote by $\mathbf{w}_{z_i, \tau} = \mathcal{W}_{z_i}(\mathbf{s}, \tau)$ the row vector containing the values of the wavelet transform for a vector of scales \mathbf{s} (of size denoted by M_s). These vectors are gathered into the $N_z \times M_s$ matrix $\mathbf{w}_{\mathbf{z}, \tau} = (\mathbf{w}_{z_1, \tau}^T \cdots \mathbf{w}_{z_{N_z}, \tau}^T)^T$. We use the same notation for the wavelet transform of the sources $\mathbf{w}_{\mathbf{y}, \tau}$. The matrix $\mathbf{A}(t)$ is assumed to vary slowly with respect to the oscillations of the signals. Then the wavelet transforms of \mathbf{y} and \mathbf{z} approximately satisfy a linear relationship of the type (5), i.e.

$$\mathbf{w}_{\mathbf{z}, \tau} \approx \mathbf{A}(\tau) \mathbf{w}_{\mathbf{y}, \tau}. \quad (6)$$

The following theorem provides a quantitative bound for the approximation error in equation (6).

Theorem 1. *With the above notations, let $\epsilon_\tau \in \mathbb{C}^{N_z \times M_s}$ be the approximation error in the wavelet domain, defined as:*

$$\epsilon_\tau = \mathbf{w}_{\mathbf{z}, \tau} - \mathbf{A}(\tau) \mathbf{w}_{\mathbf{y}, \tau}. \quad (7)$$

Let the underlying stationary sources X_i ($i = 1, \dots, N_y$) be second-order zero-mean stationary random processes of power σ_X^2 . Then the approximation error ϵ_τ is a second-order zero-mean, complex circular random matrix. Besides, the variance of the error matrix coefficients is bounded as follows:

$$\mathbb{E} \{ |\epsilon_\tau|^2 \} \leq \sigma_X^2 k_\psi^2 \mathbf{A}'^{\circ 2} \boldsymbol{\gamma}'_\infty (q^{3s})^T, \quad (8)$$

where $\mathbf{A}'^{\circ 2}$ is the entrywise square of the matrix \mathbf{A} , and $k_\psi = \int_{\mathbb{R}} |t\psi(t)| dt$, $\mathbf{A}'_\infty \in \mathbb{R}_+^{N_z \times N_y}$, and $\boldsymbol{\gamma}'_\infty \in \mathbb{R}_+^{N_y}$ are such that:

$$(\mathbf{A}'_\infty)_{ij} = \sup_t |\mathbf{A}'_{ij}(t)|, \quad (\boldsymbol{\gamma}'_\infty)_i = \sup_t |\gamma'_i(t)|.$$

A proof of Theorem 1 is provided in the Supplementary material. A Taylor expansion of \mathbf{A} centered on τ enables the construction of the bound (8). Aside from the terms controlling the error bound for (4), the error bound for (6) is also controlled by the variations of the mixing matrix coefficients.

III. ESTIMATION PROCEDURE

The estimation procedure relies on the approximation equations (4) and (6), which we assume to be valid, i.e. \mathbf{A} and $\boldsymbol{\gamma}'$ are supposed to be slowly varying enough. Assuming a Gaussian signal, the wavelet coefficients follow a complex Gaussian law, from which a likelihood can be evaluated. Time-varying parameters are therefore estimated on a discrete-time grid D by maximizing the likelihood. In this section, we describe the estimation procedure for a given $\tau \in D$. For simplicity, we set $\theta_{i, \tau} = \log_q(\gamma'_i(\tau))$ and $\boldsymbol{\theta}_\tau = (\theta_{1, \tau} \cdots \theta_{N_y, \tau})^T$.

A. Dimension reduction

Our approach requires the mixing matrix to be invertible (see section III-B). The first step is then a dimension reduction step so that \mathbf{A} is a square matrix. We first perform the singular value decomposition of consecutive segments of \mathbf{z} and set N_y to the number of significant singular values. We then select N_y observations out of the N_z available (this selection is

further discussed in section III-C). Let \mathbf{z}_l denote the retained observations, $\mathbf{A}_{l,\tau}$ the restriction of $\mathbf{A}(\tau)$ to the corresponding rows, and $\mathbf{B}_{l,\tau} = \mathbf{A}_{l,\tau}^{-1}$ the corresponding separation matrix.

B. Probabilistic setting

The underlying stationary signals X_i are henceforth assumed to be Gaussian. It follows that the wavelet transform of the i -th source Y_i is a complex circular Gaussian random matrix (see Proposition 3 in [15]). Then, $\mathbf{w}_{y_i,\tau} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma}_i(\theta_{i,\tau}))$, where

$$[\boldsymbol{\Sigma}_i(\theta_{i,\tau})]_{kk'} = q^{\frac{s_k + s_{k'}}{2}} \int_{\mathbb{R}} \mathcal{S}_{X_i}(q^{-\theta_{i,\tau}} \xi) \overline{\hat{\psi}}(q^{s_k} \xi) \hat{\psi}(q^{s_{k'}} \xi) d\xi.$$

Therefore, the sources independence hypothesis yields the following negative log-likelihood of $\mathbf{w}_{\mathbf{z},\tau}$:

$$\begin{aligned} \ell_\tau(\mathbf{B}_{l,\tau}, \boldsymbol{\theta}_\tau) = & -M_s \log |\det(\mathbf{B}_\tau)| + \frac{1}{2} \sum_{i=1}^N \log |\det \boldsymbol{\Sigma}_i(\theta_{i,\tau})| \\ & + \frac{1}{2} \sum_{i=1}^N [\mathbf{B}_{l,\tau} \mathbf{w}_{\mathbf{z},\tau}]_i \cdot \boldsymbol{\Sigma}_i(\theta_{i,\tau})^{-1} [\mathbf{B}_{l,\tau} \mathbf{w}_{\mathbf{z},\tau}]_i^H, \end{aligned}$$

where $[\mathbf{M}]_i$ denotes the i -th row of the matrix \mathbf{M} , and \mathbf{M}^H is its conjugate transpose. Maximum likelihood (ML) estimates, i.e. minimizers of $\ell_\tau(\mathbf{B}_{l,\tau}, \boldsymbol{\theta}_\tau)$, can be evaluated numerically. These estimates are respectively denoted by $\tilde{\mathbf{B}}_{l,\tau}$ and $\tilde{\boldsymbol{\theta}}_\tau$.

C. Estimation algorithm

The algorithm performs the same estimation strategy on all possible combinations of N_y observations. It consists of an alternating estimation of $\mathbf{B}_{l,\tau}$, $\boldsymbol{\theta}_\tau$, and the underlying spectra. Algorithm 1 (named JEFAS-BSS) synthesizes all the estimation steps, which are detailed below.

a) *Initialization*: Piecewise SOBI estimation (named p-SOBI) provides a baseline method, which is also used to determine the initial time-varying mixing matrix.

b) *Mixing matrix estimation*: We numerically solve the ML problem using Newton's method. Besides, because of the assumption of slow variations of the matrix coefficients, we estimate $\mathbf{B}_{l,\tau}$ with a time step Δ_τ greater than the sampling period of the observations. Then, the unmixing is performed approximating $\mathbf{B}_{l,\tau}$ by a constant matrix on the interval $I_\tau = [\tau - \frac{\Delta_\tau}{2}, \tau + \frac{\Delta_\tau}{2}]$. The estimated sources $\tilde{\mathbf{y}}$ are obtained via $\tilde{\mathbf{y}}(t) = \tilde{\mathbf{B}}_{l,\tau} \mathbf{z}_l(t)$ when $t \in I_\tau$. Besides, continuity of $\tilde{\mathbf{B}}_{l,\tau}$ coefficients is ensured by initializing the Newton's method with $\tilde{\mathbf{B}}_{l,\tau - \Delta_\tau}$. This prevents the descent from converging to another minimum of ℓ_τ caused by the BSS indeterminacies.

c) *Deformations and spectra estimations*: For each source, we obtain the joint estimation of $\{\theta_{i,\tau}\}_{\tau \in D}$ and \mathcal{S}_{X_i} via the JEFAS algorithm (which we detail in [8]) where the input wavelet transform \mathbf{w}_{y_i} of the source y_i is replaced with its estimate $\left\{ \left[\tilde{\mathbf{B}}_{l,\tau} \mathbf{w}_{\mathbf{z}_l,\tau} \right]_i \right\}_{\tau \in D}$.

d) *Stopping criterion*: The Source to Interference Ratio (SIR), introduced in [16], quantifies the interference in an estimated source originating from the other sources. The weaker the interference, the larger the SIR. As we do not have access to the ground truth sources, we measure the SIR between $\tilde{\mathbf{y}}^{(k-1)}$ and $\tilde{\mathbf{y}}^{(k)}$ (instead of \mathbf{y}), which gives

Algorithm 1 JEFAS-BSS

Dimension reduction: Estimate the number of sources N_y for the l -th combination of N_y observations **do**

Initialization: Evaluate $\tilde{\mathbf{B}}_{l,\tau}^{(0)}$ by means of p-SOBI. Derive the estimated sources $\tilde{\mathbf{y}}^{(0)}(\tau) = \tilde{\mathbf{B}}_{l,\tau}^{(0)} \mathbf{z}_l(\tau)$.

• $k \leftarrow 1$

while stopping criterion is false **and** $k \leq k_{max}$ **do**

• **Spectral estimation**: For $i \leftarrow 1, \dots, N_y$, estimate parameters $\tilde{\theta}_{i,\tau}^{(k)}$, $\forall \tau \in D$ and spectrum $\tilde{\mathcal{S}}_{X_i}^{(k)}$ applying JEFAS to $\tilde{\mathbf{y}}^{(k-1)}$.

• **BSS**: For $\tau \leftarrow 0, \Delta_\tau, \dots, T$, estimate $\tilde{\mathbf{B}}_{l,\tau}^{(k)}$: solve the ML problem replacing $\boldsymbol{\theta}_\tau$ and \mathcal{S}_{X_i} with their current estimations $\tilde{\theta}_\tau^{(k)}$ and $\tilde{\mathcal{S}}_{X_i}^{(k)}$. Derive the sources $\tilde{\mathbf{y}}^{(k)}$.

• $k \leftarrow k + 1$

end while

end for

• Estimate the averaged separation matrix $\tilde{\mathbf{B}}_\tau$. Derive $\tilde{\mathbf{y}}$.
• Apply JEFAS to $\tilde{\mathbf{y}}$ to get final estimates of \mathcal{S}_{X_i} and γ_i .

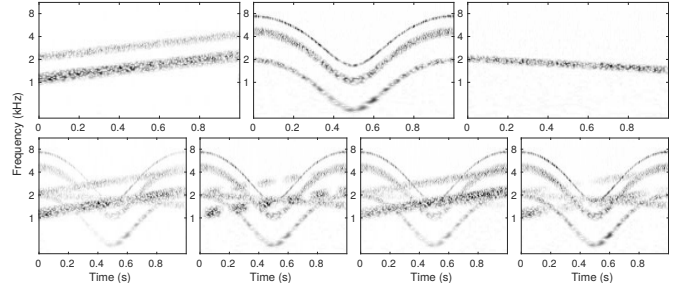


Fig. 1. Top: scalograms of the nonstationary sources. Bottom: scalograms of the observations.

an evaluation of the BSS update and is, therefore, a relevant convergence assessment. Practically, the convergence criterion consists in the comparison of the SIR averaged over sources with a given threshold, which is chosen sufficiently large.

e) *Algorithm estimations*: Finally, each row of the global separation matrix \mathbf{B}_τ is estimated $\binom{N_z-1}{N_y-1}$ times. The final estimation $\tilde{\mathbf{B}}_\tau$ is naturally obtained by averaging those different estimations. This enables us to get the final estimation $\tilde{\mathbf{y}}$ of the sources. Finally, JEFAS applied to $\tilde{\mathbf{y}}$ gives the final estimations of the time warping functions and spectra.

IV. RESULTS

We evaluate the performance of JEFAS-BSS on a synthetic example with $N_y = 3$ and $N_z = 4$. Observations are 1 second long, and sampled at $F_s = 44.1$ kHz. The power spectra of the three underlying Gaussian sources consist of non-overlapping Hann windows. The time-varying mixing matrix coefficients are sinusoidally varying over time (with different frequencies). The wavelet transforms of the locally time-warped sources, as well as those of the observations, are displayed in Fig. 1.

JEFAS-BSS is applied to the observations. It leads to 4 different estimates based on the 4 available combinations of 3 observations. Each of these estimations converges in 5 to 10 iterations. Each iteration takes about 60 seconds of

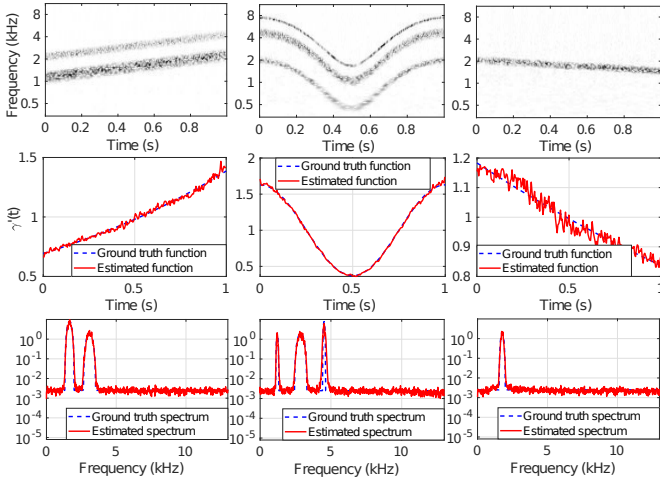


Fig. 2. Top: scalograms of the estimated sources. Middle: estimated time warping functions. Bottom: estimated spectra of the stationary sources.

CPU time on a computer running an Intel Xeon E5-2680 v4 processor. Notice that the computing time increases linearly with the observations length and dramatically increases with N_y , which limits the number of sources that we can handle. The scalograms of the estimated sources and the corresponding estimated time warping functions and spectra are displayed in Fig. 2. The estimated spectra are of good quality, even though one can notice a slight overestimation of the bandwidth and residual interference. For a complete study on the performance of JEFAS (i.e. spectra and time warping functions), we refer to [8], which in particular gives bounds on the variance of the time warping estimator and the bias of the spectrum estimator. In the following, we focus on the BSS performance.

To evaluate the quality of the BSS obtained via JEFAS-BSS, we compare it with other BSS algorithms (SOBI, p-SOBI, QTF-BSS). It is also compared with the strategy consisting in applying JEFAS-BSS on a single combination of N_y observations, called *Single JEFAS-BSS*. We apply those algorithms to 20 independent realizations of the above synthetic example. We evaluate the quality of the BSS algorithms thanks to the mean over 20 simulations of the SIR between the ground truth sources and their estimations. We give the results in Table I, together with the corresponding standard deviations (SD). Nonetheless, this quantity is a global indicator of BSS quality. To follow the temporal evolution of the BSS quality, we introduce the so-called normalized Amari index, initially constructed in [17] and called here $\rho(t)$. The Amari index takes values in $[0, 1]$ and actually measures the departure from $\hat{\mathbf{B}}_{l,\tau} \mathbf{A}_{l,\tau}$ to the identity matrix. The smaller the Amari index, the better the BSS. The mean over 20 simulations of $\rho(t)$ is displayed in Fig. 3 for each of the algorithms. Amari indices averaged with respect to time and simulations are also given in Table I, with the corresponding SD with respect to simulations.

As expected, the Amari indices and SIR clearly show that JEFAS-BSS outperforms all other algorithms. JEFAS-BSS takes advantage of the *a priori* knowledge of the nonstationarity class involved in the observations. p-SOBI allows improving the Amari index compared to SOBI, but the SIR is

TABLE I
MEAN OVER 20 REALIZATIONS OF THE SIR AND THE AVERAGED AMARI INDEX FOR DIFFERENT BSS STRATEGIES.

Algorithm	SIR (dB)		Averaged ρ (dB)	
	Mean	SD	Mean	SD
SOBI	5.93	0.60	-6.04	1.56
p-SOBI	4.54	1.62	-7.18	2.37
QTF-BSS	-0.13	3.05	-3.84	0.95
Single JEFAS-BSS	29.40	2.16	-15.43	3.08
JEFAS-BSS	29.96	2.61	-15.33	3.90

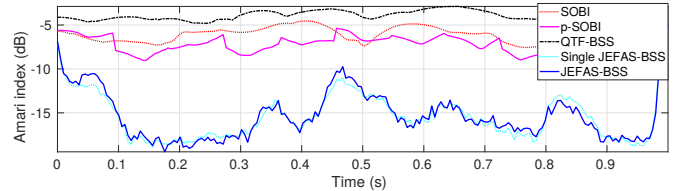


Fig. 3. Mean over 20 realizations of the temporal evolution of the Amari index (in dB) for various BSS algorithms.

impaired because of the imposed trade-off on the subintervals lengths: they must be long enough to ensure the relevance of SOBI, but must not be too long so that the mixing matrix can be approximately considered as piecewise constant. More surprisingly, QTF-BSS performance is of the same order as that of SOBI. This may be due to the fact that QTF-BSS and SOBI estimations rely on a constant mixing matrix model, then the time-varying mixing matrix deteriorates in the same way for both algorithms even though QTF-BSS is adapted to BSS of nonstationary sources while SOBI is not. Lastly, the results show that the strategy consisting in taking into account all the available combinations of N_y does not really improve the BSS quality in comparison with Single JEFAS-BSS, where a single combination of sources is considered. Nevertheless, JEFAS-BSS remains valuable thanks to its robustness in situations where one of the submatrices $\mathbf{A}_{l,\tau}$ is ill-conditioned. In contrast, Single JEFAS-BSS could diverge in such a case, as illustrated on an extra example available online¹.

V. CONCLUSION

We have introduced JEFAS-BSS, a spectral estimation algorithm for a class of multivariate nonstationary signals. We have shown that the problem can also be interpreted as a doubly nonstationary BSS problem. We evaluated JEFAS-BSS on a synthetic example on which it outperforms concurrent BSS methods, which were not designed to handle double nonstationarity. We believe the model is general enough to be applied to real situations, including audio and EEG signals. Such applications are ongoing works. Extension to other types of stationarity-breaking deformations [18] is also of interest.

The code and datasets used to produce the numerical results of this paper are available online¹. Another illustration on a synthetic mixture of real audio signals, where we evaluate the influence of the speed variation of the mixing matrix on JEFAS-BSS results, is given in Supplementary material.

¹MATLAB code at <https://github.com/AdMeynard/JEFAS>

REFERENCES

- [1] P. Stoica and R. Moses, *Spectral analysis of signals*. Upper Saddle River, N.J.: Pearson/Prentice Hall, 2005.
- [2] R. A. Silverman, "Locally stationary random processes," *IRE Transactions on Information Theory*, vol. 3, no. 3, pp. 182–187, Sep. 1957.
- [3] M. B. Priestley, *Spectral analysis and time series*, ser. Probability and mathematical statistics. Academic Press, 1982.
- [4] S. Mallat, G. Papanicolaou, and Z. Zhang, "Adaptive covariance estimation of locally stationary processes," *Ann. Statist.*, vol. 26, no. 1, pp. 1–47, Feb. 1998. [Online]. Available: <http://dx.doi.org/10.1214/aos/1030563977>
- [5] R. Dahlhaus, "Locally stationary processes," in *Time Series Analysis: Methods and Applications*, ser. Handbook of Statistics, T. S. Rao, S. S. Rao, and C. Rao, Eds. Elsevier, 2012, vol. 30, pp. 351 – 413.
- [6] M. Clerc and S. Mallat, "Estimating deformations of stationary processes," *Ann. Statist.*, vol. 31, no. 6, pp. 1772–1821, Dec. 2003. [Online]. Available: <http://dx.doi.org/10.1214/aos/1074290327>
- [7] D. Stowell, *Computational Analysis of Sound Scenes and Events*. Springer, 2018, ch. Computational Bioacoustic Scene Analysis, pp. 303–333.
- [8] A. Meynard and B. Torr sani, "Spectral Analysis for Nonstationary Audio," *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 26, no. 12, pp. 2371 – 2380, Dec. 2018. [Online]. Available: <https://hal.archives-ouvertes.fr/hal-01670187>
- [9] P. Comon and C. Jutten, *Handbook of Blind Source Separation, Independent Component Analysis and Applications*. Academic Press (Elsevier), Feb. 2010. [Online]. Available: <https://hal.archives-ouvertes.fr/hal-00460653>
- [10] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Transactions on Signal Processing*, vol. 45, no. 2, pp. 434–444, Feb. 1997.
- [11] A. Belouchrani, M. G. Amin, N. Thirion-Moreau, and Y. D. Zhang, "Source separation and localization using time-frequency distributions: An overview," *IEEE Signal Processing Magazine*, vol. 30, no. 6, pp. 97–107, Nov. 2013.
- [12] N. Thirion-Moreau and M. G. Amin, "Chapter 11 - quadratic time-frequency domain methods," in *Handbook of Blind Source Separation*, P. Comon and C. Jutten, Eds. Oxford: Academic Press, 2010, pp. 421 – 466.
- [13] D.-T. Pham and J.-F. Cardoso, "Blind separation of instantaneous mixtures of nonstationary sources," *IEEE Transactions on Signal Processing*, vol. 49, no. 9, pp. 1837–1848, Sep. 2001.
- [14] R. Kaftory and Y. Y. Zeevi, "Blind separation of time/position varying mixtures," *IEEE Transactions on Image Processing*, vol. 22, no. 1, pp. 104–118, Jan. 2013.
- [15] H. Omer and B. Torr sani, "Time-frequency and time-scale analysis of deformed stationary processes, with application to non-stationary sound modeling," *Applied and Computational Harmonic Analysis*, vol. 43, no. 1, pp. 1 – 22, 2017. [Online]. Available: <https://hal.archives-ouvertes.fr/hal-01094835>
- [16] E. Vincent, R. Gribonval, and C. F votte, "Performance measurement in blind audio source separation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 4, pp. 1462–1469, Jul. 2006.
- [17] E. Moreau and O. Macchi, "A one stage self-adaptive algorithm for source separation," in *Proceedings of ICASSP '94. IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. iii, Apr. 1994, pp. III/49–III/52.
- [18] A. Meynard and B. Torr sani, "Spectral estimation for non-stationary signal classes," in *Sampling Theory and Applications*, ser. Proceedings of SampTA17, Tallinn, Estonia, Jul. 2017. [Online]. Available: <https://hal.archives-ouvertes.fr/hal-01475534>

Spectral Estimation for Multivariate Locally Time Warped Signals: Supplementary Material

Adrien Meynard

I. PROOF OF THEOREM 1

The notations are the same as in the main body of the article.

Proof. The approximation error is given by:

$$\begin{aligned} \epsilon_\tau &= \mathbf{w}_{\mathbf{z},\tau} - \mathbf{A}(\tau)\mathbf{w}_{\mathbf{y},\tau} = \langle \mathbf{A}\mathbf{y}, \psi_{s\tau} \rangle - \langle \mathbf{A}(\tau)\mathbf{y}, \psi_{s\tau} \rangle \\ &= \langle (\mathbf{A} - \mathbf{A}(\tau))\mathbf{y}, \psi_{s\tau} \rangle. \end{aligned}$$

The Taylor-Lagrange formula states that for all the values of t , there exists a value $u_\tau(t)$ between t and τ such that:

$$\mathbf{A}(t) - \mathbf{A}(\tau) = \mathbf{A}'(u_\tau(t))(t - \tau).$$

Let $\mathbf{M}_\tau(t)$ denote the real matrix given by: $\mathbf{M}_\tau(t) = \mathbf{A}'(u_\tau(t))(t - \tau) \in \mathbb{R}^{N_z \times N_y}$. Then:

$$[\epsilon_\tau]_{im} = \sum_{j=1}^{N_y} \left\langle [\mathbf{M}_\tau]_{ij} \mathcal{D}_{\gamma_j} X_j, \psi_{s_m\tau} \right\rangle = \sum_{j=1}^{N_y} \left\langle X_j, \mathcal{D}_{\gamma_j^{-1}} [\mathbf{M}_\tau]_{ij} \psi_{s_m\tau} \right\rangle.$$

Let us denote $g_{ij,s\tau} = \mathcal{D}_{\gamma_j^{-1}} [\mathbf{M}_\tau]_{ij} \psi_{s\tau}$, then

$$\mathbb{E} \left\{ |[\epsilon_\tau]_{im}|^2 \right\} = \sum_{j=1}^{N_y} \sum_{k=1}^{N_y} \mathbb{E} \left\{ \left\langle X_j, g_{ik,s_m\tau} \right\rangle \overline{\left\langle X_k, g_{ij,s_m\tau} \right\rangle} \right\}.$$

Since the stationary sources X_j are two-by-two independent, only the terms where $k = j$ are non zero. Besides, by definition of the power spectrum, we have the following relation:

$$\sum_{j=1}^{N_y} \mathbb{E} \left\{ \left\langle X_j, g_{ij,s_m\tau} \right\rangle \overline{\left\langle X_j, g_{ij,s_m\tau} \right\rangle} \right\} = \sum_{j=1}^{N_y} \left\langle \mathcal{S}_{X_j}, |g_{ij,s_m\tau}|^2 \right\rangle.$$

Let us find an upper bound for $|g_{ij,s\tau}|$:

$$\begin{aligned} |\hat{g}_{ij,s_m\tau}(\xi)| &= \left| \int_{\mathbb{R}} [\mathbf{M}_\tau]_{ij}(t) \psi_{s_m\tau}(t) \sqrt{\gamma'_j(t)} e^{-2i\pi\gamma_j(t)\xi} dt \right| \\ &\leq \int_{\mathbb{R}} [\mathbf{A}'_\infty]_{ij} |t - \tau| |\psi_{s_m\tau}(t)| \|\gamma'_j\|_\infty^{\frac{1}{2}} dt = [\mathbf{A}'_\infty]_{ij} [\gamma'_\infty]_j^{\frac{1}{2}} q^{\frac{3s_m}{2}} k_\psi. \end{aligned}$$

Then, we obtain:

$$\mathbb{E} \left\{ |[\epsilon_\tau]_{im}|^2 \right\} \leq \sum_{j=1}^{N_y} q^{3s_m} \sigma_X^2 [\mathbf{A}'_\infty]_{ij}^2 [\gamma'_\infty]_j k_\psi^2 = q^{3s_m} \sigma_X^2 [\mathbf{A}'_\infty]_{i\cdot}^2 \gamma'_\infty k_\psi^2,$$

and finally, we get the main result of the theorem:

$$\mathbb{E} \left\{ |\epsilon_\tau|^2 \right\} \leq \sigma_X^2 k_\psi^2 \mathbf{A}'_\infty{}^{\circ 2} \gamma'_\infty (q^{3s})^T. \quad \square$$

II. ILLUSTRATION ON A SYNTHETIC MIXING OF AUDIO SIGNALS

We complement the results of the paper with a performance evaluation on a synthetic mixture of two audio signals in the determined case i.e. $N_z = N_y = 2$. We stress that a full quantitative assessment is not possible here, as the underlying spectra are not available. The code and datasets used to produce the numerical results of this example are available online¹.

The two nonstationary sources we have mixed are the sound produced by the wind blowing through a door and the sound of a singing woman. The mixing matrix coefficients are sinusoidally time-varying. The wavelet transforms of the

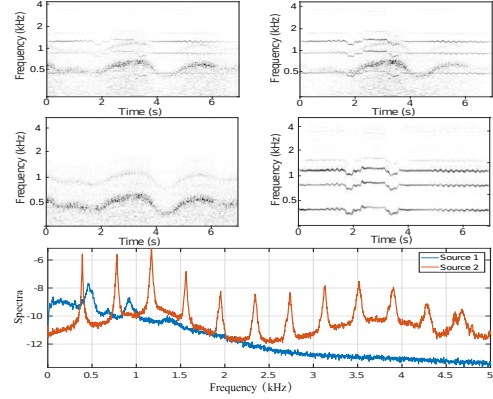


Fig. 4. Synthetic mixing of audio signals. Top: Scalograms of both observations. Middle: Scalograms of both estimated sources. Bottom: Estimated power spectra of both underlying stationary sources.

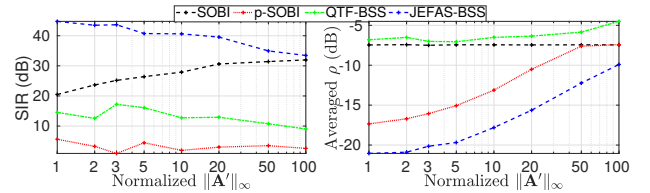


Fig. 5. Evolution of the SIR (left) and the Amari index (right) in function of the speed variation of the mixing matrix.

observations are displayed on the top of Fig. 4. The interest of this example comes from the fact that each observation is composed of two source signals whose wavelet transforms are superimposed on large time-scale zones. Therefore, it highlights JEFAS-BSS ability to separate broadband sources.

The scalograms of both estimated sources are displayed on the middle of Fig. 4, the one on the left corresponds to the estimated wind sound, while the one on the right corresponds to the voice sound. JEFAS-BSS also estimates the associated time warping functions and power spectra. Both estimated power spectra are superimposed on the bottom of Fig. 4. This figure shows the superposition of the spectral contents of the two sources, notably between 0 and 2 kHz (beyond 2 kHz, the values of the singing spectrum remain significant, and strongly dominate the values of the wind spectrum). The harmonic structure of the voice spectrum appears clearly.

Besides, we apply the same BSS algorithms on mixtures obtained from the same sources where the oscillation frequencies of the coefficients of the mixing matrix are evenly increased. We display on Fig. 5 the evolution of the performance indices as functions of the speed variation of the mixing matrix (quantified by $\|\mathbf{A}'\|_\infty$). This highlights the ability of JEFAS-BSS to outperform the other BSS algorithms when the variations of the mixing matrix are controlled, whereas JEFAS-BSS is not able to capture the variability of $\mathbf{A}(t)$ when the speed variation increases. Therefore, its SIR and Amari index converge to those of a stationary-based BSS algorithm, namely SOBI.

¹MATLAB code at <https://github.com/AdMeynard/JEFAS>