A robust algorithm for convolutive blind source separation in presence of noise

M. El Rhabi, H. Fenniri, A. Keziou, E. Moreau

A robust algorithm for convolutive Blind Source Separation in presence of noise

M. El Rhabi[a], H. Fenniri[b], A. Keziou[c,*], E. Moreau[d]

[a]*LAMAI, FSTG, Université Cadi Ayyad - Marrakech*
[b]*CReSTIC, Université de Reims Champagne-Ardenne*
[c]*Laboratoire de Mathématiques de Reims EA 4535, Université de Reims Champagne-Ardenne*
[d]*LSIS, UMR CNRS 7296, Université du Sud-Toulon-Var*

Abstract

We consider the Blind Source Separation (BSS) problem in the noisy context. We propose a new methodology in order to enhance separation performances in terms of efficiency and robustness. Our approach consists in denoising the observed signals through the minimization of their total variation, and then minimizing divergence separation criteria combined with the total variation of the estimated source signals. We show by the way that the method leads to some projection problems that are solved by means of projected gradient algorithms. The effeciency and robustness of the proposed algorithm using Hellinger divergence, are illustrated and compared with the classical mutual information approach, through numerical simulations.

Keywords: Blind source sparation, Noisy convolutive mixtures, Total variation, Divergences.

*Corresponding author
Email addresses: elrhabi@gmail.com (M. El Rhabi), hassan.fenniri@univ-reims.fr (H. Fenniri), amor.keziou@univ-reims.fr (A. Keziou), moreau@univ-tln.fr (E. Moreau)

1. Introduction

Blind Source Separation (BSS) is one of the most attractive research topics in signal processing nowadays. It has received attention because of numerous potential applications such as e.g. speech recognition systems [1], wireless digital communications [2] and biomedical signal processing [3, 4]. The main goal of BSS is to recover a set of source signals from unknown mixtures of them yielded by a set of sensors. These observed signals are assumed to be linear mixtures of the source signals. When the source signals are assumed to be statistically independent and at most one component is gaussian, the BSS principle consists in recovering this statistical property lost by the mixture process, see e.g. [5], [6], [7], [8] and the references therein. One can find numerous BSS algorithm that are based on the above principle. However, the considered mixture model is very often noise free or with a sufficiently high signal to noise ratio. In practice, the noise is an important problem that has to be considered carefully. Indeed, it changes seriously the structure of the estimation problem and makes it more difficult to tackle.

In this paper, we propose a separation process based on the minimization of measures of dependence under constraints that take into account the characteristics of the signals and the noise. Basically, it combines separating criteria using divergences between probability densities with regularization terms based on the Total-Variation (TV) of the signals. We use such sort of regularization because it allows a minimal assumption of the regularity of the signals, such that non-smooth signals with bounded variations. This results in two steps: a denoising step of the observed signals, and a second

step of simultaneous estimation and denoising of the source signals.

We will take as measures of dependence the class of D_φ-divergences between probability densities of the random signals. These criteria extend the mutual information (MI) approach (c.f. e.g. [9]) and share common features with the methods based on Renyi's distances (c.f. [6]) and β-divergences (c.f. e.g. [10], [11]). The choice of D_φ-divergences is justified by the following arguments. The divergences are convex functions which simplify the computation of their optimizers. Moreover, these criteria lead to efficient estimates as maximum likelihood (ML) method for less-noisy signals, and a suitable choice of divergences, such as Hellinger one, may improve the ML approach for noisy signals (c.f. e.g. [12], [13] and [14]). We expect that the use of Hellinger divergence combined with variational methods improves the separation performance for noisy mixed source signals.

The paper is organized as follows. In section 2, the convolutive BSS model is presented in a noisy framework. Section 3 deals with divergence criteria and their properties. In section 4, we describe our approach. Section 5 illustrates how to implement the proposed approach using both numerical and statistical techniques. In section 6, the method is used to separate noisy mixed images and signals.

2. The BSS model

Let us consider the linear convolutive BSS model in a noisy context. Here, the observed signals

$$\mathbf{x}^0(t) := (x_1^0(t), \ldots, x_M^0(t))^\tau \in \mathbb{R}^M,$$

where M denotes the number of the observed signals, is an unknown convolutive mixture of M source signals

$$\mathbf{s}^0(t) := (s_1^0(t), \ldots, s_M^0(t))^\tau \in \mathbb{R}^M,$$

trough the convolutive mixing system up to an (unknown) additive noise $\mathbf{n}(t) \in \mathbb{R}^M$:

$$\begin{aligned} \mathbf{x}^0(t) &:= A \star \mathbf{s}(t) + \mathbf{n}(t) \\ &:= \mathbf{x}(t) + \mathbf{n}(t), \end{aligned} \tag{1}$$

where "$\star$" denotes the convolutive product,

$$\mathbf{x}(t) := A \star \mathbf{s}(t)$$

is the noise-free mixed vector signals, A is the mixing operator and $\mathbf{s}(t)$ is the vector of source signals, to be estimated as precisely as possible. We assume that the mixing operator is invertible and that the components of the source vector are mutually statistically independent. The BSS, in the noiseless case, consists in finding a demixing system B operating on the observations as

$$\mathbf{y}(t) := B \star \mathbf{x}(t), \tag{2}$$

such that the components of the vector $\mathbf{y} \in \mathbb{R}^M$, called the output vector, will be statistically independent. It is known see e.g. [7], that this leads

to a consistent estimation of the source signals up to a permutation and a scalar filtering, provided that the source signals are statistically independent and that at most one source component is gaussian. This scalar filtering can eventually be reduced to a simple delay when sources are i.i.d. signals.

The estimated source signals obtained by a direct BSS, for the noisy case, can be written into the form

$$\begin{aligned} \mathbf{y}^0(t) &:= B \star \mathbf{x}^0(t) \\ &:= B \star A \star \mathbf{s}(t) + B \star \mathbf{n}(t) \\ &:= \mathbf{y}(t) + \mathbf{n}_1(t), \end{aligned} \tag{3}$$

where $\mathbf{y}(t) := B \star A \star \mathbf{s}(t)$ and $\mathbf{n}_1(t) := B \star \mathbf{n}(t)$. That is, the "noisy" estimated source $\mathbf{y}^0$ is the sum of $\mathbf{y}$, the "*ideal*" estimated source, and the noise $\mathbf{n}_1$. Ideally, we would like to retrieve $\mathbf{y}$ by denoising $\mathbf{y}^0$, but this is rather difficult since the noise $\mathbf{n}_1$ is unknown. Authors have considered the BSS in the noisy case. In [15], the authors propose a two-step approach by combining the fractional lower order statistic for the mixing estimation and minimum entropy criterion for noise-free source component estimation. The performance of this method depends on a non linear function properly chosen with relation to the source distribution and the characteristic of the noise. In [16], a whitening procedure is proposed to reduce the noise effect. Both methods are concerned by the linear instantaneous case. For linear convolutuve mixtures, in [17], the authors propose a robust algorithm against noise. They assume that noise could be decomposed in coherent and incoherent contributions. To increase robustness of their BSS algorithms against uncorrelated noise, bias removal techniques must be considered. In general, the

BSS framework for convolutive mixture was presented for the noiseless case. The robustness of the existing methods against the noise is often checked afterward. Our objective here is to propose a new strategy which takes into account the presence of the noise. The method is mainly based on a double action : (a) a denoising of the observed signal $\mathbf{x}^0$ before demixing; (b) a simultaneous BSS-denoising procedure which aims to get a noiseless estimation of source. Both of the actions lead to a regularized optimization problem. Notice that total variations methods are also used in image processing, see e.g. [18, 19, 20].

3. The divergences between probability densities

Let φ be a convex function from $[0, +\infty]$ into $[0, +\infty]$ such that $\varphi(1) = 0$. For any probabilities Q and P on $\mathbb{R}^M$ such that Q is absolutely continuous with respect to P, the D_φ-divergence between Q and P is defined through

$$D_\varphi(Q, P) := \int_{\mathbb{R}^M} \varphi\left(\frac{dQ}{dP}(\mathbf{t})\right)\ dP(\mathbf{t}), \tag{4}$$

in which $\frac{dQ}{dP}(\cdot)$ denotes the Radon-Nikodym derivative of Q with respect to P, and $\mathbf{t} := (t_1, \ldots, t_M)^\tau \in \mathbb{R}^M$. If Q is not absolutely continuous with respect to P, we set $D_\varphi(Q, P) = +\infty$. For any probability P, the function $Q \mapsto D_\varphi(Q, P)$ is convex and nonnegative. If $Q = P$, then $D_\varphi(Q, P) = 0$. Furthermore, if the function $x \mapsto \varphi(x)$ is strictly convex on a neighborhood of $x = 1$, we have the following fundamental property

$$D_\varphi(Q, P) = 0 \ \text{ iff } \ Q = P. \tag{5}$$

All the above properties are presented and proved in [21] and [22]. Note that if both Q and P have densities with respect to the Lebesgue measure on $\mathbb{R}^M$,

denoted $q(\cdot)$ and $p(\cdot)$ respectively, then the divergence (4) in this case can be written as

$$D_\varphi(Q,P) := \int_{\mathbb{R}^M} \varphi\left(\frac{q(\mathbf{t})}{p(\mathbf{t})}\right) p(\mathbf{t})\ d\mathbf{t}. \tag{6}$$

Widely used in information theory, the Kullback-Leibler divergence, noted KL-divergence, is associated to the real convex function $\varphi(x) = x\log x - x + 1$. It is defined through

$$KL(Q,P) := \int_{\mathbb{R}^M} \log\left(\frac{dQ}{dP}(\mathbf{t})\right)\ dQ(\mathbf{t}). \tag{7}$$

The modified Kullback-Leibler divergence, noted KL_m divergence, is associated to the convex function $\varphi(x) = -\log x + x - 1$, i.e.,

$$KL_m(Q,P) := \int_{\mathbb{R}^M} -\log\left(\frac{dQ}{dP}(\mathbf{t})\right)\ dP(\mathbf{t}). \tag{8}$$

It is also called mutual information. Other divergences, widely used in statistics, are the χ^2 and modified χ^2 (χ^2_m) divergences, associated respectively to the convex functions $\varphi(x) = \frac{1}{2}(x-1)^2$ and $\varphi(x) = \frac{1}{2}(x-1)^2/x$. The Hellinger ($H$) distance is also a D_φ-divergence. It is associated to the convex function $\varphi(x) = 2(\sqrt{x}-1)^2$, namely,

$$H(Q,P) := \int_{\mathbb{R}^M} 2\left(\sqrt{\frac{dQ(\mathbf{t})}{dP(\mathbf{t})}} - 1\right)^2\ dP(\mathbf{t}). \tag{9}$$

All the above examples of D_φ-divergences belong to the class of the so-called "power divergences" introduced by [23] and which are defined through the real convex functions

$$\varphi_\alpha : x \in \mathbb{R}^*_+ \mapsto \varphi_\alpha(x) := \frac{x^\alpha - \alpha x + \alpha - 1}{\alpha(\alpha-1)} \tag{10}$$

for all $\alpha \in \mathbb{R}\setminus\{0,1\}$,

$$\varphi_0(x) := -\log x + x - 1,$$

and

$$\varphi_1(x) := x\log x - x + 1.$$

The table 1 gives, according to the choice of φ_α, the associated D_{φ_α}-divergence.

Those divergences have been used in statistics, see e.g. [24], [25], [26], [12], [14] and [27] among others. It has been proved that the use of divergences in statistics extends the well known maximum likelihood approach, and leads to estimates with similar efficiency properties even better in some cases; the particular choice of Hellinger distance improves the maximum likelihood approach in terms of efficiency-robustness for noisy data, see e.g. [12], [14] and [27]. We expect that the use in BSS of divergences other than mutual information, such as Hellinger one, may lead to better results for noisy mixture signals.

4. The proposed approach

The aim here is to get a good approximation of $\mathbf{y}$ from (3) and to make their components statistically independent, in order to give a consistent estimation of the source signal vector $\mathbf{s}(t)$. Our approach proceeds in two steps. Step 1 : denoising the observed signal $\mathbf{x}^0$. Step 2 : a blind source separation combined with a denoising of the estimated source $\mathbf{y}^0$. Let us sketch the main features of each step.

4.1. Denoising the observed signal

Let $\mathbf{x}^0(t) := (x_1(t), \ldots, x_M(t))^\tau$, be the noisy observed random vector signal, which writes into the form

$$\mathbf{x}^0(t) := \mathbf{x}(t) + \mathbf{n}(t).$$

We would reconstruct the *ideal* observed signal

$$\mathbf{x}(t) := (x_1(t), ..., x_M(t))^\tau$$

from

$$\mathbf{x}^0(t) := (x_1^0(t), ..., x_M^0(t))^\tau$$

by means of the following variational problem

$$x_i = \underset{w_i \in \mathbb{X}}{\arg\min} \left\{ \frac{1}{2} E\left(|w_i - x_i^0|^2 \right) + \lambda E\left(\theta\left(|w'_i| \right) \right) \right\}, \tag{11}$$

for all $i \in \{1, ..., M\}$, where $\lambda > 0$ is a penalization parameter, $\theta(\cdot)$ is a well chosen function, $w'_i(t)$ the first derivative of $w_i(t)$ with respect to t, and $\mathbb{X}$ is an appropriate space. $E(\cdot)$ denotes the mathematical expectation. Notice that the first term of (11) is the fidelity term while the second one controls the variation of w_i. The Euler-Lagrange equation corresponding to this optimization problem writes formally

$$w_i - \lambda \left(\frac{\theta'(|w_i{}'|)}{|w_i{}'|} w'_i \right)' = x_i^0. \tag{12}$$

In practice, the function $\theta(\cdot)$ is chosen to encourage smoothing in regions where the variations of the signal are weak, that is $|\boldsymbol{w}'| \approx 0$, and to preserve discontinuities where $|\boldsymbol{w}'|$ is strong. The total variation case corresponds to the choice $\theta(x) = x$ which will be adopted here. One can make other

choices like $\theta(x) = \sqrt{1+x^2}$. The natural space for treating the continuous variational problem (11), when $\theta(x) = x$, is $\mathbb{X} := BV([0,T]$, the space of all real valued functions on the interval $[0,T]$ with bounded variation (see e.g. [28], [29]), where T is the observation time of the signals. The treatment of the continuous problem is a mathematical question which is beyond the scope of this paper. Here, we consider only the corresponding discrete problem which will be presented and solved in section 5.1 hereafter.

4.2. The simultaneous BSS-denoising procedure

The purpose of this second step is to reconstruct an estimated source signals $\mathbf{y}$ from the partially denoised observed signal $\mathbf{x}$. Following the arguments of the last section, our approach consists in minimizing, with respect to B, a statistical estimate of the following criterion

$$J(\mathbf{y}) := J_{sep}(\mathbf{y}) + J_{reg}(\mathbf{y}). \tag{13}$$

Here $\mathbf{y}(t) := B \star \mathbf{x}(t)$, $t \in [0,T]$, where $\mathbf{x} := (x_1, \ldots, x_M)^\tau$ is obtained from (11). The first term $J_{sep}(\cdot)$ is a measure of dependence between the random components of $\mathbf{y} := (y_1, \ldots, y_M)^\tau$, and $J_{reg}(\cdot)$ is a regularization term which takes the form

$$J_{reg}(\mathbf{y}) := \sum_{i=1}^{M} \left[\frac{\gamma}{2} E\left(|y_i - y_i^0|^2\right) + \mu E\left(\theta\left(|y_i'|\right)\right)\right], \tag{14}$$

where $\theta(\cdot)$ is defined as in section 4.1, y_i and y_i^0 are respectively the i^{th} component of $\mathbf{y}$ and $\mathbf{y}^0 := B \star \mathbf{x}^0$, for all $i \in \{1, ..., M\}$. $\mathbf{y}'$ is the first derivative of $\mathbf{y}$. Notice that the first term is the fidelity term while the second one controls the variation of $\mathbf{y}$. The reals $\gamma > 0$ and $\mu > 0$ are regularization

parameters. We use as measure of statistical dependence between the components $y_1, \ldots, y_M$ of the random vector $\mathbf{y}$, the D_φ-divergence between the product density $\prod_{i=1}^{M} f_{y_i}(\cdot)$ of the marginal densities f_{y_i} of the components y_i, $i \in \{1, \ldots, M\}$, and the joint density $f_{\mathbf{y}}(\cdot)$ of the random vector $\mathbf{y}$. Since we deal with convolutive mixtures, it is easy to show that the independence between two scalar random sources $y_1(n)$ and $y_2(n)$ (for all n) is not sufficient to separate the system. That is why additional constraints must be stated to ensure the mutual independence of the output signal components $y_i(t), i \in \{1, ..., M\}$. To make it easier to understand, let us consider now a bidimensional random vector $\mathbf{y}(n) := (y_1(n), y_2(n))^\tau$. In order to separate the signals, the independence of the components $y_1(n_1)$ and $y_2(n_2)$ is needed for all n_1 and n_2. In other words, the independence of $y_1(n)$ and $y_2(n-k)$, for all n and at all lags k, is necessary to ensure separation. As in [30, 31], we define the separating criterion J_{sep} by

$$J_{sep}(\mathbf{y}) := \sum_{\boldsymbol{q}} I_\varphi(\mathbf{y}^{\boldsymbol{q}}), \tag{15}$$

where

$$\boldsymbol{q} := (q_1 = 0, q_2, ..., q_M)^\tau \in \{0\} \times \mathbb{Z}^{M-1}$$

is an integer vector,

$$\begin{aligned} \mathbf{y}^{\boldsymbol{q}}(n) &:= (y_1^{q_1}(n), \ldots, y_M^{q_M}(n))^\tau \\ &:= (y_1(n - q_1), ..., y_M(n - q_M))^\tau, \end{aligned}$$

and $I_\varphi(\mathbf{y}^{\boldsymbol{q}})$ is the I_φ-divergence of the random vector $\mathbf{y}^{\boldsymbol{q}}(n)$, which we define to be the D_φ-divergence between the densities $\prod_{i=1}^{M} f_{y_i{}^{q_i}}(\cdot)$ and $f_{\mathbf{y}^{\boldsymbol{q}}}(\cdot)$, i.e.,

$$\begin{aligned} I_\varphi(\mathbf{y}^{\boldsymbol{q}}) &:= D_\varphi\left(\prod_{i=1}^{M} f_{y_i{}^{q_i}}, f_{\mathbf{y}^{\boldsymbol{q}}}\right) \\ &= \int_{\mathbb{R}^M} \varphi\left(\frac{\prod_{i=1}^{M} f_{y_i{}^{q_i}}(t_i)}{f_{\mathbf{y}^{\boldsymbol{q}}}(\mathbf{t})}\right) \; f_{\mathbf{y}^{\boldsymbol{q}}}(\mathbf{t}) \; d\mathbf{t}, \end{aligned} \tag{16}$$

where $\mathbf{t} := (t_1, \ldots, t_M)^\tau$. Notice that the particular choice $\varphi(x) := -\log x + x - 1$ leads to the mutual information of the random vector $\mathbf{y}^q(n)$. Notice also, from property (5), that $I_\varphi(\mathbf{y}^{\boldsymbol{q}}) \geq 0$, and

$$I_\varphi(\mathbf{y}^{\boldsymbol{q}}) = 0 \text{ iff } \text{ the components of } \mathbf{y}^{\boldsymbol{q}} \text{ are independent.} \tag{17}$$

Moreover, the above criterion (16) can be written as follows

$$I_\varphi(\mathbf{y}^{\boldsymbol{q}}) = E\left[\varphi\left(\frac{\prod_{i=1}^{M} f_{y_i^{q_i}}(y_i^{q_i})}{f_{\mathbf{y}^{\boldsymbol{q}}}(\mathbf{y}^{\boldsymbol{q}})}\right)\right]. \tag{18}$$

5. Discretization and statistical estimation

We now present how to make operational the above method described in section 4, by the use of numerical and statistical techniques. In the sequel, continuous scalar random signals are sampled at a period T_e. To each continuous scalar signal $u(t)$, we associate a vector $(u(1), ..., u(N))^\tau \in \mathbb{X} := \mathbb{R}^N$ defined by $u(k) := u(kT_e)$, for all $k = 1, \ldots, N$. This vector is denoted by

$$U := (u(1), ..., u(N))^\tau.$$

The vector space $\mathbb{X} := \mathbb{R}^N$ is equipped with the euclidian inner product

$$\langle U, V \rangle := \sum_{k=1}^{N} u(k)v(k),$$

for all $U, V \in \mathbb{X}$. The numerical first derivative of any $U \in \mathbb{X}$, denoted $U' := (u'(1), ..., u'(N))^\tau$, belongs to $\mathbb{X}$, and is defined by

$$u'(k) := \frac{u(k+1) - u(k)}{T_e}, \quad \text{for all } k = 1, \ldots, N-1, \tag{19}$$

and $u'(N) = 0$. Due to the discretization of the first derivatives and in order to avoid the *edge effect*, we define also the *backward derivative* of any $U \in \mathbb{X}$, denoted

$$U^* := (u^*(1), ..., u^*(N))^\tau,$$

by

$$u^*(1) := \frac{u(1)}{T_e}, \; u^*(N) := -\frac{u(N-1)}{T_e}$$

and

$$u^*(k) := \frac{u(k) - u(k-1)}{T_e} \text{ for all } k = 2, \ldots, N-1.$$

Here, the backward derivatives $(\cdot)^*$ is the discrete adjoint operator of $-(\cdot)'$. That is, for all $U, V \in \mathbb{X}$, we have

$$\langle U^*, V \rangle = -\langle U, V' \rangle. \tag{20}$$

All these definitions are extended in a natural way to elements of the cartesian product space $\mathbb{X}^M$. Elements of the vector space $\mathbb{X}$ will be denoted by capital symbols, while the elements of the vector space $\mathbb{X}^M$ are denoted by bold capital symbols, which can be considered as $M \times N$ dimension matrices.

5.1. The denoising of the discrete observed signal

In this section, we show how to estimate in practice $\mathbf{x}(t)$ from the noisy observation $\mathbf{x}^0(t)$. Recall that this estimation is obtained by solving the discrete version of the optimization problem (11) in which the mathematical

expectation is estimated trough the time mean. We start with the following proposition which gives a simple characterization of the solution (see [19] for a proof).

Proposition 1. *The discrete version of the problem (11) has a unique solution given by*

$$\mathbf{X} = \mathbf{X}^0 - \Pi_{\lambda G}\mathbf{X}^0, \tag{21}$$

where $\Pi_{\lambda G}$ *is the euclidian orthogonal projection operator on the convex set* λG *with* $G := \{\mathbf{V}^* \mid \mathbf{V} \in \mathbb{X}^M,\ |v_i(k)| \leq 1, \forall (i,k) \in \{1,\dots,M\}\times\{1,\dots,N\}\}$.

Thus, to compute $\mathbf{X}$ we are lead to compute the projection operator $\Pi_{\lambda G}$ on the convex set λG, i.e., the projection of the observed signal matrix $\mathbf{X}^0$ on the set λG:

$$\mathbf{X} = \underset{\mathbf{W}\in\lambda G}{\arg\min} \left\|\mathbf{W} - \mathbf{X}^0\right\|^2.$$

In other words, if D denotes the convex set given by

$$D := \{V \in \mathbb{X} \mid \forall k = 1,\dots N,\ |v(k)| \leq 1\},$$

it is equivalent to compute the projection operator Π_D, i.e., to solve the problem

$$\mathbf{X} = \underset{\mathbf{V}\in D^M}{\arg\min} \left(\left\|\lambda\mathbf{V}^* - \mathbf{X}^0\right\|^2\right). \tag{22}$$

If $\rho > 0$ denotes a real number, the optimality condition for the problem (22) could be written as follows

$$V_i = \Pi_D\left(V_i - 2\lambda\rho\left(\lambda V_i^* - X_i^0\right)'\right),\ \ \forall i = 1,...,M, \tag{23}$$

where $\Pi_D(\cdot)$ is the orthogonal projection on D. It is straightforward to see that

$$\forall V \in \mathbb{X}, \left(\Pi_D V\right)(k) = \begin{cases} \frac{v(k)}{|v(k)|}, & \text{if} \quad |v(k)| \geq 1, \\ v(k), & \text{else,} \end{cases} \tag{24}$$

for all $k = 1, ..., N$. In order to solve the problem (22), one can use a projected gradient method, that is

Data: $\mathbf{X}^0$ the observed signal
Result: $\mathbf{X}$ the denoised observed vector
Initialization : $\mathbf{X}^{(0)} = \mathbf{X}^0$. Given $\varepsilon > 0$, $\lambda > 0$ and $\rho > 0$ suitably choosen

do

• update $\mathbf{X}$:

for $i = 1, \ldots, M$

$$X_i^{(p+1)} = \Pi_D \left(X_i^{(p)} - 2\lambda\rho \left(\lambda \left(X_i^{(p)} \right)^* - X_i^0 \right)' \right)$$

until $\|\mathbf{X}^{(p+1)} - \mathbf{X}^{(p)}\| < \varepsilon$

$\mathbf{X} = \mathbf{X}^{(p+1)}$.

Algorithm 1: The denoising step.

Proposition 2. *The sequence $\mathbf{X}^{(p)}$ converges to $\mathbf{X}$, when $p \to \infty$, if λ and ρ satisfy the following condition*

$$\lambda^2 \rho < \frac{T_e^2}{4}. \tag{25}$$

Proof of proposition 2. If we analyze this algorithm, component by component, we have $\forall i \in \{1, ..., M\}$,

$$\begin{aligned} \left\| X_i^{(p)} - X_i \right\| &= \left\| X_i^{(p)} - X_i + 2\lambda^2\rho \left(\left(X_i^{(p)} - X_i \right)^* \right)' \right\| \\ &= \left\| \left(I + 2\lambda^2\rho\Delta \right) \left(X_i^{(p)} - X_i \right) \right\| \\ &\leq \left\| (I + 2\lambda^2\rho\Delta \right\| \left\| X_i^{(p)} - X_i \right\|, \end{aligned} \tag{26}$$

where Δ is the second order derivative operator defined by $\forall V \in \mathbb{X}, \Delta(V) := (V^*)' = V''$. Since the eigenvalues of Δ are negative, the sequence $\left\| \mathbf{X}^{(p)} - \mathbf{X} \right\|$

is necessary decreasing. Hence, it converges when

$$\lambda^2 \rho < \frac{1}{\|\Delta\|},$$

where $\|\Delta\|$ is defined by

$$\|\Delta\| := \sup_{V \in \mathbb{X}} \frac{\langle -\Delta V, V \rangle}{\|V\|^2}.$$

According to (20), we have, $\forall V \in \mathbb{X}$,

$$\begin{aligned} \langle -\Delta V, V \rangle &= \|V'\|^2 \\ &= \sum_{k=1}^{N-1} \left(\frac{v(k+1) - v(k)}{T_e} \right)^2 \\ &\leq \frac{4}{T_e^2} \|V\|^2, \end{aligned}$$

which ends the proof of proposition 2.

Afterwards, the filtered observed signal $\mathbf{X}$ will be considered as the denoised version of the observed signal $\mathbf{X}^0$. Now, we investigate the BSS step.

5.2. An algorithm for the simultanous BSS-denoising

In this section, we apply the gradient approach to separate convolutive mixtures based on the minimization of a statistical estimate of the criterion (13). Let us assume that the discrete separating system form of (2) is defined by

$$\mathbf{y}(n) := B \star \mathbf{x}(n) := \sum_{k=0}^{L} B_k \mathbf{x}(n-k), \ \forall n = 1, \ldots, N, \tag{27}$$

where $B := (B_0, B_1, \ldots, B_L)$ are finite impulse response (FIR) filters with maximum degree L. To estimate the $M \times M$ dimension matrices B_k, leading to estimated sources outputs, we expose hereafter how to estimate the

criterion $J(B) := J(\mathbf{y})$ (by abuse of notation) and how to compute the minimizer in $B := (B_0, B_1, \ldots, B_L)$ of the proposed estimate. Denote by $\widehat{J}(B)$ the estimate of the criterion $J(B)$. The criterion $J(B)$, under assumption (27), writes

$$J(B) := J_{sep}(B) + J_{reg}(B), \tag{28}$$

where

$$J_{sep}(B) := \sum_{\boldsymbol{q} \in \{0\} \times \{-2L, \ldots, 2L\}^{M-1}} I_{\varphi}(\mathbf{y}^{\boldsymbol{q}}) \tag{29}$$

and

$$J_{reg}(B) := \sum_{i=1}^{M} \left[\frac{\gamma}{2} E\left(|y_i - y_i^0|^2 \right) + \mu E\left(\theta\left(|y_i'| \right) \right) \right]. \tag{30}$$

From the formula (18), we propose to estimate the criterion (29) through

$$\widehat{J}_{sep}(B) := \sum_{\boldsymbol{q} \in \{0\} \times \{-2L, \ldots, 2L\}^{M-1}} \widehat{I}_{\varphi}(\mathbf{y}^{\boldsymbol{q}}), \tag{31}$$

where

$$\widehat{I}_{\varphi}\left(\mathbf{y}^{\boldsymbol{q}}\right) := \frac{1}{N} \sum_{n=1}^{N} \varphi\left(\frac{\prod_{i=1}^{M} \widehat{f}_{y_i^{q_i}}(y_i^{q_i}(n))}{\widehat{f}_{\mathbf{y}^{\boldsymbol{q}}}(\mathbf{y}^{\boldsymbol{q}}(n))} \right) \tag{32}$$

in which $\widehat{f}_{y_i^{q_i}}(\cdot)$ denotes the kernel estimate of the marginal density $f_{y_i^{q_i}}(\cdot)$, for all $i \in \{1, \ldots, M\}$, and $\widehat{f}_{\mathbf{y}^{\boldsymbol{q}}}(\cdot)$ is the kernel estimate of the joint density $f_{\mathbf{y}^{\boldsymbol{q}}}(\cdot)$, i.e.,

$$\widehat{f}_{y_i^{q_i}}(u_i) := \frac{1}{N h_i} \sum_{n=1}^{N} k\left(\frac{y_i^{q_i}(n) - u_i}{h_i} \right), \tag{33}$$

$\forall i \in \{1, \ldots, M\}$, and

$$\widehat{f}_{\mathbf{y}^{\boldsymbol{q}}}(\mathbf{u}) := \frac{1}{N h_1 \cdots h_M} \sum_{n=1}^{N} \prod_{i=1}^{M} k\left(\frac{y_i^{q_i}(n) - u_i}{h_i} \right), \tag{34}$$

in which $h_1, \ldots, h_M$ are the bandwidths and $k(\cdot)$ the kernel. Here we take $k(\cdot)$ to be the standard univariate gaussian density $x \in \mathbb{R} \mapsto k(x) := \frac{1}{\sqrt{2\pi}} \exp\{-x^2/2\}$, and we consider the "optimal" bandwidths

$$h_i := \widehat{\sigma}_i (3/4)^{-1/5} N^{-1/5},$$

where $\widehat{\sigma}_i$ is the empirical standard deviation of the scalar data $y_i^{q_i}(1), \ldots, y_i^{q_i}(N)$, for all $i \in \{1, \ldots, M\}$; see e.g. [32].

The second term $J_{reg}(\cdot)$ can be estimated as in section 5.1 by discretization and replacing the mathematical expectation $E(\cdot)$ by the time mean $\widehat{E}(\cdot)$, as follows

$$\widehat{J}_{reg}(B) := \sum_{i=1}^{M} \left(\frac{\gamma}{2} \widehat{E}\left(|y_i - y_i^0|^2 \right) + \mu \widehat{E}\left(\theta\left(|y_i'| \right) \right) \right). \tag{35}$$

Hence the problem leads to minimize in $B := (B_0, B_1, \ldots, B_L)$ the following estimate of the criterion $J(B)$

$$\widehat{J}(B) := \widehat{J}_{sep}(B) + \widehat{J}_{reg}(B), \tag{36}$$

where $\widehat{J}_{sep}(B)$ and $\widehat{J}_{reg}(B)$ are given by (31) and (35), respectively. The minimizer, denote it $\widehat{B}$, can be computed using gradient descent type algorithms. For this, we need the computation of the gradient, according to each $B_k, k = 0, 1 \ldots, L$, of the estimated criterion $\widehat{J}(B)$. The following proposition gives the explicit formula of the gradient.

Proposition 3. *Let us consider $\widehat{J}(B)$ defined by (36). Then,*

$$\begin{aligned}\frac{\partial \widehat{J}(B)}{\partial B_k} &= \sum_{\boldsymbol{q}\in\{0\}\times\{-2L,\ldots,2L\}^{M-1}} \frac{\partial \widehat{I}_\varphi(\mathbf{y}^{\boldsymbol{q}})}{\partial B_k} \\ &\quad +\gamma \widehat{E}\left[\left(\mathbf{y}-\mathbf{y}^0\right)(n)\left(\mathbf{x}-\mathbf{x}^0\right)(n-k)^\tau\right] \\ &\quad +\mu \widehat{E}\left[\left(\frac{\theta'\left(|\mathbf{y}'(n)|\right)\mathbf{y}'(n)}{|\mathbf{y}'(n)|}\right)^* \mathbf{x}(n-k)^\tau\right].\end{aligned}$$

Remark 1. *1) In the particular case of mutual information, i.e., when $\varphi(x) = -\log x$, we can write*

$$\begin{aligned}\widehat{I}_\varphi(\mathbf{y}^{\boldsymbol{q}}) &= \widehat{E}\left[\log \widehat{f}_{\mathbf{y}^{\boldsymbol{q}}}(\mathbf{y}^{\boldsymbol{q}}(n))\right] \\ &\quad -\sum_{i=1}^M \widehat{E}\left[\log \widehat{f}_{y_i^{q_i}}(y_i^{q_i}(n))\right],\end{aligned}$$

and then

$$\begin{aligned}\frac{\partial \widehat{I}_\varphi(\mathbf{y}^{\boldsymbol{q}})}{\partial B_k} &= \widehat{E}\left[\frac{(\partial/\partial B_k)\widehat{f}_{\mathbf{y}^{\boldsymbol{q}}}(\mathbf{y}^{\boldsymbol{q}}(n))}{\widehat{f}_{\mathbf{y}^{\boldsymbol{q}}}(\mathbf{y}^{\boldsymbol{q}}(n))}\right] \\ &\quad -\sum_{i=1}^M \widehat{E}\left[\frac{(\partial/\partial B_k)\widehat{f}_{y_i^{q_i}}(y_i^{q_i}(n))}{\widehat{f}_{y_i^{q_i}}(y_i^{q_i}(n))}\right],\end{aligned}$$

where, $\forall i = 1,\ldots,M$,

$$\begin{aligned}&\frac{\partial}{\partial B_k}\widehat{f}_{y_i^{q_i}}(y_i^{q_i}(n)) = \frac{1}{Nh_i^2}\sum_{j=1}^N \\ &k'\left(\frac{y_i^{q_i}(j)-y_i^{q_i}(n)}{h_i}\right)\left(\mathbf{x}^{q_i}(j-k)-\mathbf{x}^{q_i}(n-k)\right)^\tau\end{aligned}$$

and

$$\begin{aligned}&\frac{\partial}{\partial B_k}\widehat{f}_{\mathbf{y}^{\boldsymbol{q}}}(\mathbf{y}^{\boldsymbol{q}}(n)) = \\ &\frac{1}{Nh_1\cdots h_M}\sum_{j=1}^N \frac{\partial}{\partial B_k}\prod_{i=1}^M k\left(\frac{y_i^{q_i}(j)-y_i^{q_i}(n)}{h_i}\right).\end{aligned}$$

2) For any differentiable function φ, we have

$$\frac{\partial \widehat{I}_{\varphi}(\mathbf{y}^{\boldsymbol{q}})}{\partial B_k} = \widehat{E}\left[\varphi'\left(\frac{\prod_{i=1}^{M}\widehat{f}_{y_i^{q_i}}(y_i^{q_i}(n))}{\widehat{f}_{\mathbf{y}^{\boldsymbol{q}}}(\mathbf{y}^{\boldsymbol{q}}(n))}\right) \times \frac{\partial}{\partial B_k}\frac{\prod_{i=1}^{M}\widehat{f}_{y_i^{q_i}}(y_i^{q_i}(n))}{\widehat{f}_{\mathbf{y}^{\boldsymbol{q}}}(\mathbf{y}^{\boldsymbol{q}}(n))}\right].$$

We can then derive the following algorithm.

Data: $\mathbf{X}^0$ the observed signal
Result: $\mathbf{Y}$ the estimated source signal
Initialization : Compute $\mathbf{X} = \mathbf{X}^0 - \Pi_D \mathbf{X}^0$ from algorithm 1
Given $\varepsilon > 0$, $B_k^{(0)}$, $k \in \{0, 1, \ldots, L\}$, $\mathbf{Y}^{(0)} := B^{(0)} \star \mathbf{X}$ and $\tau > 0$
do

- update B_k

  for $k = 0, 1, \ldots, L$,

  $$B_k^{(p+1)} = B_k^{(p)} - \tau \frac{\partial \widehat{J}\left(B^{(p)}\right)}{\partial B_k^{(p)}}$$

- update $\mathbf{Y}$

  $$\mathbf{y}^{(p+1)}(n) = \sum_{k=0}^{L} B_k^{(p+1)} \mathbf{x}(n-k)$$

until $\|B_k^{(p+1)} - B_k^{(p)}\| < \varepsilon$

Algorithm 2: The separation algorithm.

6. Numerical results

In this section, we present simulation results for the proposed method. We dealt with three kinds of samples, namely observations obtained by a noisy instantaneous mixture of two images, a convolutive mixture of two i.i.d. uniform source signals, and a convolutive mixture of two source signals drawn

from the 4-ASK alphabet. The separation performance was evaluated in term of the signal-to-interference ratio (SIR) for each output y_i. Recall that the SIR is defined by

$$SIR_i := 10 \log_{10} \left(\frac{\widehat{E}\left(y_i^2\right)}{\widehat{E}\left(y_i^2|_{s_i=0}\right)} \right), \ i = 1, \ldots, M,$$

where $\widehat{E}\left(y_i^2|_{s_i=0}\right)$ stands for the time mean of the mixture y_i^2 from which the source signal s_i is cancelled.

The mixtures have been separated using mutual information (MI) method described in ([30, 31]), MI method where the data are pre-whitened (MI_Wh) and our proposed algorithm, denoted H_TV, using the separation criterion (36) described in sections 5.1 and 5.2, for the particular choice of Hellinger divergence (i.e., the D_{φ_α}-divergence with $\alpha = 0.5$) combined with the total variation. Since the criterion (36) is computationally expensive in the convolutive case, we implemented here a somehow stochastic version, where the sum is reduced to one term q_i, randomly chosen from the set $\{-2L, ..., 2L\}$, at each iteration. Concerning the penalization parameters, also called "hyperparameters", they are chosen (small enough) in an ad hoc manner. There exists methods to dynamically tuned these parameters in order to control the trade-off between conformance to data and conformance to the prior. We refer to [33] for a comparative study and a way to choose them objectively.

6.1. Example 1: Instantaneous case : noisy mixed images

In this example, we show the capability of the proposed algorithm using H_TV to successfully separate two noisy mixed images and compare its performance with the classical MI algorithm that does not take into account

the noise. The dimensions of the synthesized images which are used in this experiment are 256×256 pixels and the mixing matrix is

$$\mathbf{A}(\mathbf{z}) = \begin{bmatrix} 0.55 & 0.45 \\ 0.45 & 0.55 \end{bmatrix}.$$

The two original images, shown in figure 1, are mixed by the above mixing matrix, at which we add a gaussian noise. The signal to noise ratio (SNR) is defined as the power of the signal to the power of the noise and it is expressed in decibels. We take SNR $= -22.4$ dB. The mixed images so obtained are shown in figure 2. The gradient descent parameter for MI and H_TV is taken $\tau = 0.05$. For H_TV method, in the denoising step, see algorithm 1, we take $\rho = 0.25$, $\lambda = 0.01$ and $\varepsilon = 0.001$, and in the second step, see algorithm 2, we chose $\gamma = 0.0001$ and $\mu = 0.1$. To quantify the performance of the used algorithms, we consider the so-called peak signal to noise ratio (PSNR) defined by

$$PSNR_i := \sum_j \left(\frac{\max(u_{or}^j)^2 lc}{\|u_{app}^j - u_{or}^j\|^2} \right), \ i = 1, 2, \tag{37}$$

where $l \times c$ is the size of the image, j is the red (R), the green (G) or the blue (B) components of the image, u_{or} is the original image and u_{app} is the restored image.

The results, illustrated in figures 3 and 4 and presented in table 2, show that H_TV method provides a better visual quality of separation than the MI one.

6.2. Example 2

We compare the three methods, namely MI, MI_Wh and H_TV. We investigate the impact of the data set length, N, on the separation accuracy in

the first experiment, and the impact of the noise in the second experiment for a fixed signal sample length N. Indeed, in BSS approach, computational complexity increases with the data length. Thus, by optimizing the number of signal sample lengths, the computational complexity and operation time could be reduced. An additive white gaussian noise (AWGN) has been added to the observed signals. We take SNR $=-20$ dB, iteration number $= 2000$ and for every N, we consider 2 source signals drawn from the 4-ASK alphabet $\{-3, -1, 1, 3\}$. The study was carried out on set of 50 Monte Carlo runs of random 4-ASK source signals for different length N. The source signals were linearly mixed through the following RIF filters

$$\mathbf{A}(\mathbf{z}) = \begin{bmatrix} 1 + 0.2z^{-1} + 0.1z^{-2} & 0.5 + 0.3z^{-1} + 0.1z^{-2} \\ 0.5 + 0.3z^{-1} + 0.1z^{-2} & 1 + 0.2z^{-1} + 0.1z^{-2} \end{bmatrix}.$$

Figure 5 shows the averaged SIRs versus data length N for the three algorithms. We can see that the best separation accuracy for the three methods is achieved since $N = 5000$. For any considered signal sample length, we see that H_TV is the best. In the second experiment, we take $N = 5000$. The iteration number $= 5000$, and we let the SNR varying between -30 dB and 0 dB. For each SNR level the experiment is repeated 100 times with different realizations of i.i.d. uniform random sources mixed by the above RIFs. Figure 6 shows the averaged SIRs versus SNR for the three algorithms. For both experiments, the gradient descent parameter, for MI, MI_Wh and H_TV, is taken $\tau = 0.2$. For H_TV method, the hyperparameters are chosen to be $\rho = 0.25$, $\lambda = 0.01$, $\varepsilon = 0.001$, $\gamma = 0.0001$ and $\mu = 0.1$.

6.3. Example 3

Here, the source signals were linearly mixed through a randomly generated RIF filters of length 6 from uniform law on $[0, 1]$. The number of observations is taken equals 5000. The simulations are repeated 25 times with different realizations of i.i.d. uniform random sources. We consider two experiments. In the first one, we investigate, for the three criteria MI, MI_Wh and H_TV, the convergence of the corresponding descent gradient algorithms, in terms of the number of iterations. We add to the observed mixed sources an AWGN with $SNR = -20$ dB. The results are illustrated in figure 7. We can see that the rate of convergence is comparable for the three methods, but the separation is better for the H_TV one. In the second experiment, we investigate the impact of the noise on the separation performance. We consider the same source signals as above for a fixed length $N = 5000$, and we add an AWGN with different SNR values. The results are presented in Figure 8. This experiment shows that the H_TV approach outperform MI and MI_Wh. The parameters for all algorithms are taken to be the same as in example 2.

7. Conclusion

An efficient and robust algorithm for convolutive BSS in presence of noise is presented. This algorithm makes use of divergence separation criteria and variational methods to control the noise. It proceeds in two steps : a preprocessing which consists in reducing the noise on the data, and a second one minimizing a separation criterion based on divergences regularized through total variation. This regularization prevents from the error model and allows to denoise the estimate output. For the particular choice of the Hellinger

divergence, the simulation results show the high accuracy of this approach in terms of efficiency, and robustness against the noise, compared with the classical mutual information approach.

References

[1] S. Choi, C. A., Adaptive blind separation of speech signals: Cocktail party problem, in: Proc. Int. Conf. Speech Processing, 1997, pp. 617–622.

[2] D. M.R., E. B.L., Blind source separation with a time-varying mixing matrix, in: in Proc. of the Forty-First Asilomar Conference on Signal, System & Computer, Pacific Grove California, 2007, pp. 626–630.

[3] D. Ferreira, A. M. Sá, A. Cerqueira, J. Seixas, Ica-based method for quantifying eeg event-related desynchronization, Proceedings of the 8th International Conference on Independent Component Analysis and Signal Separation ICA'09 (2009) 403–410.

[4] M. Congedo, C. Gouy-Pailler, C. Jutten, On the blind source separation of human electroencephalogram by approximate joint diagonalization of second order statistics, Clinical Neurophysiology 119 (12) (2008) 2677 – 2686.

[5] P. Comon, Independent component analysis, a new concept ?, Signal Process. 36 (1994) 287–314.

[6] F. Vrins, D.-T. Pham, M. Verleysen, Is the general form of Renyi's

entropy a contrast for source separation?, Vol. 4666/2007 of Lecture Notes in Computer Science, Springer, Berlin, 2007.

[7] E. Moreau, J.-C. Pesquet, N. Thirion-Moreau, Convolutive blind signal separation based on asymmetrical contrast functions, IEEE Trans. Signal Process. 55 (1) (2007) 356–371.

[8] M. Castella, S. Rhioui, E. Moreau, J.-C. Pesquet, Quadratic higher order criteria for iterative blind separation of a MIMO convolutive mixture of sources, IEEE Trans. Signal Process. 55 (1) (2007) 218–232.

[9] D.-T. Pham, Mutual information approach to blind separation of stationary sources, IEEE Trans. Information Theory 48 (7).

[10] A. Cichocki, S.-i. Amari, Families of alpha- beta- and gamma-divergences: flexible and robust measures of similarities, Entropy 12 (6) (2010) 1532–1568.

[11] A. Basu, C. Park, H. Shioya, W.-t. Huang, Statistical Inference, The Minimum Distance Approach, Chapman & Hall, 2011.

[12] R. Beran, Minimum Hellinger distance estimates for parametric models, Ann. Statist. 5 (3) (1977) 445–463.

[13] B. G. Lindsay, Efficiency versus robustness: the case for minimum Hellinger distance and related methods, Ann. Statist. 22 (2) (1994) 1081–1114.

[14] R. Jiménez, Y. Shao, On robustness and efficiency of minimum divergence estimators, Test 10 (2) (2001) 241–248.

[15] M. Sahmoudi, H. Snoussi, M. G. Amin, Robust approach for blind source separation in non-gaussian noise environments, in: Proccedings of IS-CCSP, Marrakesh, Morocco, IEEE/EURASIP, March 2006.

[16] A. Belouchrani, A. Cichocki, Robust whitening procedure in blind source separation context, Electronics Letters 36, N. 24 (2000) 2050–2051.

[17] R. Aichner, H. Buchner, W. Kellermann, Convolutive blind source separation for noisy mixtures, In E. Hansler and G. Schmidt (eds.), Topics in Speech and Audio Processing in Adverse Environments, Springer-Verlag, Berlin/Heidelberg (2008) 469–524.

[18] L. Rudin, S. Oshers, E. Fatemi, Non linear total variation based noise removal algorithm, in: Annual internationnal conference N11,Los Alamos NM, Etats-Unis N 1-4, Vol. 60, 1992, pp. 259–268.

[19] A. Chambolle, An algorithm for total variation minimization and application, Journal of Mathematical Imaging and vision 20 issue 1-2 (2004) 89–97.

[20] A. Haddad, Y. Meyer, An improvement of rudin-osher-fatemi model, Appl. Comput. Harmon. Anal. 22(3) (2007) 319–334.

[21] I. Csiszár, Information-type measures of difference of probability distributions and indirect observations, Studia Scientiarum Mathematicarum Hungarica 2 (1967) 299–318.

[22] F. Liese, I. Vajda, Convex statistical distances, Vol. 95 of Teubner-Texte zur Mathematik [Teubner Texts in Mathematics], BSB B. G. Teubner Verlagsgesellschaft, Leipzig, 1987.

[23] N. Cressie, T. R. C. Read, Multinomial goodness-of-fit tests, J. Roy. Statist. Soc. Ser. B 46 (3) (1984) 440–464.

[24] A. Keziou, Dual representation of ϕ-divergences and applications, C. R. Math. Acad. Sci. Paris 336 (10) (2003) 857–862.

[25] M. Broniatowski, A. Keziou, Minimization of ϕ-divergences on sets of signed measures, Studia Sci. Math. Hungar. 43 (4) (2006) 403–442.

[26] M. Broniatowski, A. Keziou, Parametric estimation and tests through divergences and the duality technique, J. Multivariate Anal. 100 (1) (2009) 16–36.

[27] B. G. Lindsay, Efficiency versus robustness: the case for minimum Hellinger distance and related methods, Ann. Statist. 22 (2) (1994) 1081–1114.

[28] L. Evans, R. Gariepy, Measure theory and fine properties of function, Studies in Advanced Mathematics, CRC Press, Boca Raton, FL.

[29] A. Cohen, D. Wolfgang, , I. Daubechies, R. DeVore, Harmonic analysis of the space bv, Rev. Mat. Iberoamericana 19(1) (2003) 235–263.

[30] M. Babaie-Zadeh, C. Jutten, K. Nayebi, Convolutive mixtures by mutual information minimization, in: Proceedings of IWANN, Granada, Spain, 2001, pp. 834–842.

[31] M. E. Rhabi, G. Gelle, H. Fenniri, G. Delaunay, A penalized mutual information criterion for blind separation of convolutive mixtures, Signal Processing 84 (2004) 1979–1984.

[32] B. W. Silverman, Density estimation for statistics and data analysis, Monographs on Statistics and Applied Probability, Chapman & Hall, London, 1986.

[33] B. M. Graham, A. Adler, Objective selection of hyperparameter for eit, Physiol. Meas. 27 S5 (2006) 235–263.