CONTRAST FUNCTIONS FOR BLIND SEPARATION AND DECONVOLUTION OF SOURCES

Dinh Tuan Pham

Laboratoire de Modélisation et Calcul, CNRS, IMAG
BP. 53, 38041 Grenoble cedex, France

ABSTRACT

A general method to construct contrast functions for blind source separation is presented. It is based on a superadditive functional of class II applied to the distributions of the reconstructed sources. Examples of such functionals are given. Our approach permits exploiting the temporal dependence of the sources by using a functional on the joint distribution of the source process over a time interval. This yields many new examples and frees us from the constraint that the sources be non Gaussian. Contrasts functions based on cumulants requiring the orthogonality constraint is also covered. Finally, the case of convolutive mixtures is considered in relation with the problem of blind separation-deconvolution.

1. INTRODUCTION

The goal of blind source separation is to recover the sources from their unknown observed mixtures, typically of the form $\mathbf{X}(\cdot) = \mathbf{AS}(\cdot)$, where $\mathbf{S}(t)$ and $\mathbf{X}(t)$ represents the source and the observation vectors at time t, and $\mathbf{A}$ is a linear transformation, which can be either instantaneous or convolutive. We assume that there are a same number K of sources and of sensors and that an inverse transformation of $\mathbf{A}$ exists.

In the blind context, no particular knowledge on the source distribution is assumed, instead one will rely mainly on the assumption of independence between the sources to achieve their separation. Specifically, one tries to find a transformation $\mathbf{B}$ such that the components of $\mathbf{Y}(\cdot) = \mathbf{BX}(\cdot)$, which represent the reconstructed sources, are as independent as it is possible. A more general approach is to minimize a contrast function. Following Comon [1], we define a contrast function as a functional of the distribution of $\mathbf{Y}(\cdot)$ (and possibly also of $\mathbf{B}$) which attains its minimum when separation is achieved. Note that by relying solely on the independence of the sources, one can only achieve separation up to a permutation and a non mixing transformations on each source sequence: a scaling (and a translation) in the case of instantaneous mixtures or a convolution in the case of convolutive mixtures. Thus, *separation is always understood with the above indeterminacy attached*, not in the sense of having extracted exactly the sources. Following Comon [1], we call a contrast function discriminating if it attains its minimum *only when* separation is achieved.

This work provides a general method for constructing discriminating contrast functions. It is based on the concept of superadditive functionals introduced by Huber [2]. Subadditive functionals can also be used but they require *orthogonal constraint*. They yield cumulant based contrasts, which have been well studied and hence will be discussed only briefly. The main novelty of our work is that we consider also functionals of the joint distribution of the reconstructed sources at different time points and not only of their marginal distribution at a single time point (as in earlier works). Further, convolutive mixtures are treated as well. For such mixtures, the separation is much easier if one makes the assumption that the sources can be represented as the output of some filter applied to temporally independent processes. It is then actually easier to both separate and deconvolve the sources to obtain the underlying temporally independent processes. We call this the blind separation-deconvolution problem. It contains as a special case the classical deconvolution problem.

It should be noted that the construction of a contrast functions is only a first step toward a separation procedure. The contrasts introduced here are of theoretical nature because they depend on the distribution of the reconstructed sources, which is unknown. To obtain a usable contrast, this distribution or in fact certain functionals of it must be estimated from the data. This problem will be investigated in future works in specific settings since our general approach can have different implementations, adapted to the problem considered.

2. CONTRASTS FOR INSTANTANEOUS MIXTURES

Here the mixture model writes $\mathbf{X}(t) = \mathbf{AS}(t)$ where $\mathbf{A}$ is an unknown $K \times K$ mixing matrix and $\mathbf{X}(t)$ and $\mathbf{S}(t)$ are as

Thanks to the European Bliss Project for funding.

before. The separation is performed by making an inverse transformation $\mathbf{Y}(t) = \mathbf{B}\mathbf{X}(t)$ where $\mathbf{B}$ is a $K \times K$ separating matrix. In this setup, a contrast function is a functional of the distribution of $\mathbf{Y}(\cdot)$, which is minimized when $\mathbf{BA}$ equals the product of a permutation and a diagonal matrix.

2.1. Contrasts based on the marginal distribution

As the sources are reconstructed through an instantaneous transformation, one can expect to obtain contrast functions based only on the marginal distribution of the reconstructed source vector $\mathbf{Y}(t)$. By stationarity, this distribution does not depend on t, therefore we will drop this time index.

A natural contrast is obtained by considering the mutual information between the components $Y_1, \ldots, Y_K$ of $\mathbf{Y}$ [3]:

$$I(Y_1, \ldots, Y_K) = \mathrm{E} \log \frac{p_{Y_1,\ldots,Y_K}(Y_1, \ldots, Y_K)}{p_{Y_1}(Y_1) \cdots p_{Y_K}(Y_K)}$$

where E is the expectation operator and $p_{Y_1,\ldots,Y_K}$ and p_{Y_k} denote the joint and marginal densities of $Y_1, \ldots, Y_K$. This is a contrast since it is well known that it is non negative and vanishes if the random variables are independent. Note that

$$I(Y_1, \ldots, Y_K) = \sum_{k=1}^{K} h(Y_k) - h(Y_1, \ldots, Y_K) \qquad (1)$$

where for any random vector $\mathbf{Y}$ with density $p_{\mathbf{Y}}$, $h(\mathbf{Y}) = -\mathrm{E} \log p_{\mathbf{Y}}(\mathbf{Y})$ denotes its entropy and $h(Y_1, \ldots, Y_K)$ is the same as $h([Y_1 \ \cdots \ Y_K]^{\mathrm{T}})$ ($^{\mathrm{T}}$ denoting the transpose) and denotes the joint entropy of $Y_1, \ldots, Y_K$. Further, it can be shown that $h(\mathbf{Y}) = h(\mathbf{X}) + \log|\det \mathbf{B}|$ and thus the contrast (1) equals, up to an additive constant,

$$C(\mathbf{B}) = \sum_{k=1}^{K} h(Y_k) - \log|\det \mathbf{B}|. \qquad (2)$$

The advantage of (2) over (1) is that it involves only the distribution of a random variable, not the joint distribution of several random variables. This would avoid the problem of not having enough data to estimate the density distribution in high dimensional spaces (the "curse of dimensionality").

We now provide a *systematic* method to construct contrast functions, through the use of superadditive functionals of class II (of a distribution) introduced by Huber [2]. A functional Q of the distribution of a random variable Y, denoted by $Q(Y)$, is said to be of class II if it is (i) translation invariant, in the sense $Q(Y + a) = Q(Y)$ for any real number a, and (ii) scale equi-variant, in the sense that $Q(aY) = |a|Q(Y)$ for any real number a. Note that this definition, as appears in [2], doesn't require that Q be non negative, but is is clear that if Q satisfies it then so does $|Q|$, hence we can without loss of generality assume $Q \geq 0$. The functional Q is said to be superadditive if [2]:

$$Q^2(X + Y) \geq Q^2(X) + Q^2(Y) \qquad (3)$$

for any two independent random variables X and Y.

Proposition 2.1 *Let Q is a superadditive functional of class II, then*

$$\sum_{k=1}^{K} \log Q(Y_k) - \log|\det \mathbf{B}| \qquad (4)$$

is a contrast function for separating an instantaneous mixture of sources. This contrast is discriminating if $Q(S_k) > 0$ for every sources S_k and if the inequality (3) is strict (i.e not an equality) for any pair of non zero multiples of two distinct sources.

Examples

- 1: Let $Q(Y) = e^{h(Y)}$, the square root of the entropy power. Then Q is of class II and by a result of Blachman [4] (see also [3]), it is superadditive with inequality (3) being strict unless X and Y are Gaussian. Therefore, under the condition that there can be no more than one Gaussian source, the above functional yields a discriminating contrast, which can be easily seen to be the same as (2). Thus we get a proof that the mutual information contrast is discriminating (an earlier proof based on the Darmois Theorem has been given in Comon [1]).

- 2: Let $Q(Y) = J(Y)^{-1/2}$ where $J(Y) = \mathrm{E}\psi_Y^2(Y)$ (ψ_Y denoting minus the logarithmic derivative of the density of Y) is the Fisher information in the location estimation problem. Then Q is a functional of class II and, as proved in [2] and [4], is superadditive. The proof in [4] also shows that the inequality (3) is strict unless X and Y are Gaussian. Thus, the above functional also yields a discriminating contrast if there are no more than one Gaussian source.

- 3: For bounded random variables, the functional Q defined by $Q(Y) = R_X$, the range of X, is of class II and superadditive, since $R_{X+Y} = R_X + R_Y$ if X and Y are independent. The last equality also implies that the inequality (3) is strict unless R_X or R_Y equal 0. Thus $\sum_{k=1}^{K} \log R_{Y_k} - \log|\det \mathbf{B}|$ is a discriminating contrast, in the case of bounded non deterministic sources. It has in fact been introduced and proved to be so in [5].

- 4: Suppose that the sources are sub-Gaussian, that is they admit non positive fourth order cumulants. Clearly the same is true for any linear mixture of them. On the set of sub-Gaussian random variables, the class II functional Q defined by: $Q(Y) = [\mathrm{E}(Y - \mathrm{E}Y)^4]^{1/4}$ can be shown to be superadditive with inequality (3) being strict unless both X and Y have zero fourth order cumulant. This functional thus yields a discriminating contrast under the restriction that all sources are sub Gaussian with at most one can have zero fourth order cumulant.

2.2. Contrasts under orthogonality constraint

Many earlier works on blind source separation are based on higher (than 2) order cumulants. But cumulants are not superadditive but subadditive and the above method is not applicable. However, one can still construct cumulant based contrast functions *if the transformation matrix* $\mathbf{B}$ *is constrained to be orthogonal*. (In fact all cumulant based contrasts which we know of require this constraint.) The orthogonality constraint is justified by the fact that the data has been pre-whitened so that the observed vector $\mathbf{X}$ has covariance matrix the identity matrix. Therefore in order to preserve this property for the reconstructed source vector $\mathbf{Y}$, the matrix $\mathbf{B}$ must be orthogonal. Further, by absorbing the scales of the sources into $\mathbf{A}$ so that they have unit variance, their non correlation implies that $\mathbf{A}$ is also orthogonal.

A functional Q is called subadditive if [2]

$$Q^2(X+Y) \leq Q^2(X) + Q^2(Y)$$

for any pair of independent random variable X and Y. For generality we shall extend the above definition and call this functional α-subadditive ($\alpha \geq 0$) if one has instead

$$Q^\alpha(X+Y) \leq Q^\alpha(X) + Q^\alpha(Y).$$

It can be shown that α-subadditivity implies β-subadditivity for all $\beta \leq \alpha$, in particular it implies subadditivity if $\alpha \geq 2$.

Proposition 2.2 *Let Q is a α-subadditive functional of class II for some $\alpha \geq 2$ and assume that the mixing matrix $\mathbf{A}$ is orthogonal. Then both $-\sum_{k=1}^K Q^\alpha(Y_k)$ and $-\sum_{k=1}^K Q^{2\alpha}(Y_k)$ are contrast functions for separating an instantaneous mixture of sources under the restriction that the separating matrix $\mathbf{B}$ is orthogonal. These contrasts are discriminating if $\alpha > 2$ and the $Q(S_k)$ (S_k denoting the sources) can be zero for at most one index k.*

The above result shows that $-\sum_{k=1}^K Q^\beta(Y_k)$ is a contrast function for all β in the set $[2, \alpha] \cup [4, 2\alpha]$, which reduces to $[2, 2\alpha]$ if $\alpha \geq 4$. On the other hand, the functional Q defined by $Q(Y) = |\mathrm{cum}_r(Y)|^{1/r}$, where $\mathrm{cum}_r(Y)$ is the r-th cumulant of Y, can be shown to be r-subadditive. One can then deduce the results of Comon [1] and of Moreau and Macchi [6]. More general forms of cumulant based contrasts can also be found in [7, 8, 9].

2.3. Contrasts based on the joint distribution

When the sources possess temporal dependence (which is often the case), it is of interest to exploit this dependence by considering contrast functions depending on joint distribution of several consecutive values of the reconstructed sources. By stationarity, one only need to consider the distribution of the random vectors $[Y_k(1) \ \cdots \ Y_k(m)]^{\mathrm{T}}$, denoted by $Y_k(1:m)$ for short, where m is a given integer. (In the sequel, this kind of notation will be used for any sequence). Note that one can also consider the vector $[Y_k(q) \ Y_k(2q) \ \cdots \ Y_k(mq)]^{\mathrm{T}}$ instead of $Y_k(1:m)$, where q is an integer greater than 1. This can be useful when the observed process comes from the sampling of a continuous time process with a too fine sampling interval.

The results of section 2.1 can be easily generalized to functionals of the distribution of random *vectors* instead of random variables. As before, such a functional Q of is said to be of class II if for any random vector Y one has (i) $Q(Y) = Q(Y+a)$ for any real vector a (translation invariance) and (ii) $Q(aY) = |a|Q(Y)$ for any real number a (scale invariance). Again it is called superadditive if the inequality (3) holds for any pair of independent random vectors X and Y.

Proposition 2.3 *Let Q is a class II superadditive functional on m-dimensional distributions, then*

$$\sum_{k=1}^K \log Q[Y_k(1:m)] - \log|\det \mathbf{B}| \tag{5}$$

is a contrast function for separating an instantaneous mixture of sources. This contrast is discriminating if $Q[S_k(1:m)] > 0$ for all k (S_k denoting the k-th source) and the inequality (3) is strict for any pair of non zero multiples of $S_j(1:m)$ and $S_k(1:m)$ with distinct index j, k.

Examples

- 1: Let $Q(Y) = e^{h(Y)/m}$, the square root of the entropy power of the distribution of the m-vector Y. Then Q is a functional of class II and by the result of Blachman [4] (see also [3]), it is superadditive with inequality (3) being strict unless X and Y are Gaussian with proportional covariance matrices. Thus

  $$\frac{1}{m}\sum_{k=1}^K h[Y_k(1:m)] - \log|\det \mathbf{B}| \tag{6}$$

  is a discriminating contrast function under the condition that there can be no pair of Gaussian sources with proportional auto-covariances up to lag $m-1$. This contrast can be easily seen to be equivalent to the mutual information between $Y_1(1:m), \ldots, Y_K(1:m)$].

- 2: Let $Q(Y) = [\det \mathbf{J}(Y)]^{-1/(2m)}$ where m is the dimension of Y and $\mathbf{J}(Y) = \mathrm{E}[\psi_Y(Y)\psi_Y^{\mathrm{T}}(Y)]$ is the Fisher information matrix in the (vector) location estimation problem, ψ_Y here denoting minus the gradient of the logarithm of the density of Y. Then Q is a functional of class II and one can generalize the proof of Blachman [4] to show that it is superadditive with inequality (3) being strict unless X and Y are Gaussian random vector with proportional covariance matrices. Hence this functional yields a discriminating contrast under the same conditions as in example 1.

- 3: Let Q be the functional defined over bounded m-dimensional distributions by $Q(Y) = R_Y$ where R_Y is the m-th root of the volume of the support set of Y. By the Brunn-Minkowski inequality ([3], equation (16.98)), $R_{X+Y} \geq R_X + R_Y$ for independent random vectors X and Y. Thus the functional Q is superadditive of class II with the inequality (3) being strict unless R_X or R_Y equal 0 and hence the functional $\sum_{k=1}^K R_{Y_k} - \log\det \mathbf{B}$ is a discriminating contrast, in the case of bounded non deterministic sources.

The above examples are the analogues of examples 1 – 3 in section 2.1. But in the vector case, there are many other possibilities for constructing contrast functions.

Other examples

- 4: Let Q be the functional defined by

  $$Q[Y(1:m)] = e^{h[Y(m)|Y(1:m-1)]}$$

  where $h[Y(m)|Y(1:m-1)$ is the conditional entropy of $Y(m)$ given $Y(1)$, ..., $Y(m-1)$. Then, one can prove (using a result of Blachman [4]) the inequality (3) with X and Y replaced by $X(1:m)$, $Y(1:m)$, which is strict unless the above vectors are Gaussian with covariance matrix such that their inverses having last column proportional. Thus, by stationarity,

  $$\sum_{k=1}^K h[Y_k(m)|Y_k(1:m-1)] - \log \mathbf{B} \qquad (7)$$

  is a discriminating contrast under the same conditions as in example 1.

- 5: Let Q be the functional defined by $Q(Y) = [\mathbf{J}(Y)_{mm}]^{-1/2}$ where $\mathbf{J}(Y)$ is the Fisher information matrix as defined in example 2 above and $\mathbf{J}(Y)_{mm}$ its lower-right element. One can prove that Q is superadditive with inequality (3) being strict under the same condition as in example 4 above. Thus (7) with $\frac{1}{2}[\mathbf{J}(Y_k)_{mm}]^{-1/2}$ in place of $h[Y_k(m)|Y_k(1:m-1)]$ is a discriminating contrast under the same condition as in example 1.

It is of interest to note that the contrasts in the above examples do not require the sources to be non Gaussian to be discriminating, unlike the contrasts based on the marginal distribution. This points to the possibility of second order based procedures through the use of Gaussian functional. For any functional Q, we may defined a corresponding Gaussian functional Q_g by $Q_g(Y) = Q(\tilde{Y})$ where $\tilde{Y}$ is a Gaussian vector with the same covariance matrix as that of Y. Clearly if Q is a class II superadditive functional then so is Q_g, albeit strict inequality in (3) may not hold. In fact this is the case where Y is a scalar so that Q_g is of no interest. But in the vector case, one does get interesting functionals.

Examples of "Gausian" contrasts

- $1'$: let $Q(Y) = [\det \operatorname{cov}(Y)]^{1/(2m)}$ where cov$(\cdot)$ denotes the covariance matrix. This is the Gaussian analogue of the functionals in examples 1 and 2. The corresponding contrast is thus discriminating if there is no pair of sources which have proportional autocovariances up to lag $m-1$.

- $4'$: Let $Q(Y) = [\det \operatorname{cov}(Y)/\det \operatorname{cov}(Y)_-]^{1/2}$ where cov$(\cdot)$ denotes the covariance matrix and cov$(Y)_-$ is the matrix obtained by deleting the last row and column of cov(Y). This is the Gaussian analogue of the functionals in examples 4 and 5. The corresponding contrast is thus discriminating under the same condition as in example $1'$.

Filtering and combination of contrasts

In Proposition 2.3, the functional Q is applied to a vector of *consecutive* observations, but this is not necessary. We have noted that we can down sample the observed process. But there are other possibilities as well. In particular, we may consider a filter banks defined by the sequences $c_1(\cdot)$, ..., $c_m(\cdot)$ of their impulse responses. Then for a stationary process $Y(\cdot)$, we define

$$Q[Y(\cdot)] = \tilde{Q}([(c_1 \star Y)(1) \ \ \cdots \ \ (c_m \star Y)(1)]^{\mathrm{T}})$$

where $\star$ denotes the convolution, that is $(c_k \star Y)(t)$ is the output at time t of the filter with impulse response sequence $c_k(\cdot)$ applied to the process $Y(\cdot)$, and $\tilde{Q}$ is some class II superadditive functional over m-dimensional distributions. It is clear that the functional Q is of class II and superadditive. The conditions that the resulting contrast is discriminating are more complex but can be easily worked out case by case.

In examples 1, $1'$, 2, 3 above, $Q(Y)$ factorizes as $[\tilde{Q}(Y_1)\cdots\tilde{Q}(Y_m)]^{1/m}$ for some functional $\tilde{Q}$ over unidimensional distributions when the components $Y_1, \ldots, Y_m$ of Y are independent. This suggests considering the last expresion, which by itself defines a class II functional by itself and can be shown to be superadditive if $\tilde{Q}$ is, regardless whether the random variables $Y_1, \ldots, Y_m$ are independent or not. More generally, the functional Q defined by $Q(Y) = \tilde{Q}_1^{w_1}(Y)\cdots\tilde{Q}_p^{w_p}(Y)$ where the $\tilde{Q}_k$ are class II superadditive functionals operating on some subset (depending on k) of the components of the m-vector Y and $w_1, \ldots,$

w_p are positive numbers summing to 1, can be shown to be of class II superadditive.

The above consideration leads to the following contrast function:

$$\sum_{k=1}^{K}\sum_{n=1}^{m} w_n \log \tilde{Q}[(c_n \star Y_k)(1)] - \log|\det \mathbf{B}|$$

The condition for it to be discriminating can be easily worked out, on a case by case basis.

If we take the filters to be *narrow* band-pass filters at different frequency bands, then their outputs tend to be Gaussian, justifying the use of a Gaussian functional for $\tilde{Q}$. But in the uni-dimensional case, such a functional must be a multiple of the standard deviation. Hence we are led to the contrast function

$$\sum_{k=1}^{K}\frac{1}{2}\sum_{n=1}^{m} w_n \log \mathrm{var}[(c_n \star Y_k)(1)] - \log \mathbf{B} \qquad (8)$$

where var$(\cdot)$ denotes the variance. It can be shown that this contrast is discriminating if for each source S_k, $\mathrm{var}[(c_1 \star S_k)(1)]$, …, $\mathrm{var}[(c_m \star S_k)(1)]$ are all positive and that the vector having them as components is not proportional to any similarly defined vector corresponding to another source.

Let $\mathbf{f_Y}(n)$ be the matrix with general element the covariance between $(c_n \star Y_i)(1)$ and $(c_n \star Y_j)(1)$, which represents the smoothed spectral density matrix over the frequency band of the n-th filter. As $\mathbf{Y} = \mathbf{BX}$, $\mathbf{f_Y}(n) = \mathbf{B}\mathbf{f_X}(n)\mathbf{B}^{\mathrm{T}}$ where $\mathbf{f_X}$ is defined as $\mathbf{f_Y}$ but with the components of $\mathbf{Y}$ replaced by that of $\mathbf{X}$. Thus (8) can be rewritten, up to an additive constant as

$$\frac{1}{2}\sum_{n=1}^{m} w_n[\log\det\mathrm{diag}\,\mathbf{f_Y}(n) - \log\det\mathbf{f_Y}(n)]. \qquad (9)$$

The expression inside the above bracket [ ] can be interpreted as a measure of deviation from diagonality the matrix $\mathbf{f_Y}(n)$, since by the Hadamard inequality ([3], p. 233 or 502) $\det \mathrm{diag}\mathbf{M} \le \det \mathbf{M}$ for any positive definite matrix $\mathbf{M}$, with equality if and only if this matrix is diagonal. Thus, the contrast (9) is a joint diagonalization criterion.

The contrast (6) and (7) have been introduced in Pham [10]. Their Gaussian analogues and the contrast (8) or (9) have been introduced in Pham [11, 12]. Pham [13] also provides an efficient algorithm to solve the associated problem of joint approximate diagonalization of several matrices.

3. CONTRAST FOR CONVOLUTIVE MIXTURES

Consider now the case where the sources are mixed through a convolution

$$\mathbf{X}(t) = \sum_{l=-\infty}^{\infty} \mathbf{A}(l)\mathbf{S}(t-l) = (\mathbf{A}\star\mathbf{S})(t)$$

where $\mathbf{X}(\cdot)$ and $\mathbf{S}(\cdot)$ denote the observation and source process, respectively, $\mathbf{A}(\cdot)$ is a sequence of mixing matrices and $\star$ denotes the convolution. The separation then consists in applying an inverse convolution: $\mathbf{Y}(t) = (\mathbf{B}\star\mathbf{X})(t)$ to recover the sources.

The simplest (and most natural) way to construct an contrast for this problem is to consider the mutual information rate between the processes $Y_1(\cdot)$, …, $Y_K(\cdot)$:

$$I[Y_1(\cdot),\ldots,Y_1(\cdot)] = \sum_{k=1}^{K} h[Y_k(\cdot)] - h[\mathbf{Y}(\cdot)] \qquad (10)$$

where $h[Y(\cdot)]$ denotes the entropy rate of the (scalar or vector) process $Y(\cdot)$, defined as the limit of $h[Y(1:m)]/m$ as $m\to\infty$ [3]. This mutual information rate is clearly a contrast (but we haven't been able to prove that it is discriminating in all generalities).

There is a nice result relating the entropy rate of a filtered process to that of the original process [10]: if the process $\mathbf{Y}(\cdot)$ is related to $\mathbf{X}(\cdot)$ by $\mathbf{Y}(t) = (\mathbf{B}\star\mathbf{X})(t)$, then

$$h[\mathbf{Y}(\cdot)] = h[\mathbf{X}(\cdot)] + \int_0^{2\pi} \log\Big|\det\sum_{l=-\infty}^{\infty}\mathbf{B}(l)e^{il\lambda}\Big|\frac{d\lambda}{2\pi} \qquad (11)$$

Therefore the contrast (10) is equivalent to

$$C[\mathbf{B}(\cdot)] = \sum_{k=1}^{K} h[Y_k(\cdot)] - \int_0^{2\pi} \log\Big|\det\sum_{l=-\infty}^{\infty}\mathbf{B}(l)e^{il\lambda}\Big|\frac{d\lambda}{2\pi}. \qquad (12)$$

The advantage of (12) over (10) is that one is dispensed with the evaluation of the entropy rate of a vector process in (possibly) high dimension. Nevertheless, (12) is still mostly of theoretical interest since the entropy rate is defined through a limiting operation and hence is not easy to estimate.

There are however simple ways to obtain contrast for the convolutive mixture *if one restrict the sources to the (still general) class of linear processes.* Specifically, we will assume that the k-th source can be written as

$$S_k(t) = \sum_{l=-\infty}^{\infty} a_k(l)e_k(t-l) \qquad (13)$$

where $e_k(\cdot)$ is a sequence of independent identically distributed random variables and $a_k(\cdot)$ is a sequence of impulse responses of some (well behaved) filter. Thus, the observed process can be expressed as a convolutive mixtures of the temporally independent processes $e_1(\cdot)$, …, $e_K(\cdot)$. Then one may try to find an inverse convolution to extract the last processes. This may be called the blind separation-deconvolution problem since the sources are not only separated but also deconvolved as well. Note that in the convolutive mixture setup, the blind separation can only yield the sources up to a filtering, since replacing each of them by any filtered version does not affect their independence. Therefore the blind separation-decovolution problem is the same as the blind separation problem in which a particular filtered version, temporally independent, of each source is extracted. This would eliminate the indeterminacy with respect to filtering with the exception that one can still arbitrarily shift the time index of the reconstructed sources. Thus the word contrast in the following proposition should be understood as a function which is minimized when the reconstructed sources equal the real sources *up to a permutation, a scaling and a time shift*. The contrast is called discriminating if it is minimized only under such conditions.

Proposition 3.1 *Let Q is a superadditive functional of class II, then*

$$\sum_{k=1}^{K} \log Q[Y_k(1)] - \log \Big| \det \sum_{l=-\infty}^{\infty} \mathbf{B}(l) e^{il\lambda} \Big| \frac{d\lambda}{2\pi} \qquad (14)$$

is a contrast function for separating (and deconvoluting) a convolutive mixture of sources, under the assumption that the sources are linear processes. This contrast is discriminating if $Q(e_k) > 0$ for all k, $e_k(\cdot)$ denoting the temporally independent process in the representation (13) of the sources, and if the inequality (3) is strict (i.e not an equality) for any pair of non zero multiples of $e_j(t)$ and $e_k(t)$ with $j \neq k$ (the value of t is irrelevant since the processes are stationary).

The above result is very similar to the Proposition 2.1 for instantaneous mixtures. As in this case, one can also considers subadditive functionals instead, provided that the reconstruction sequence of matrices $\mathbf{B}(\cdot)$ satisfied a orthogonal constraint: $\sum_{l=-\infty}^{\infty} \mathbf{B}(l)\mathbf{B}(l)^{\mathrm{T}} =$ the identity matrix. Such constraint can be justified if the observed process has been pre-whitened so that they are temporally uncorrelated and uncorrelated among themselves. Then a similar result as Proposition 2.2 (with $\mathbf{A}$ replaced by $\mathbf{A} \star \operatorname{diag}(a_1, \ldots, a_K)$ and $Q(S_k)$ replaced by $Q(e_k)$) can be obtained. The proof is also similar to that of this Proposition and also to that in the papers [14] and [6]. Then one can construct cumulant based contrasts in the same way as in section 2.2. Such contrasts have actually appeared in [14] and some more general forms of cumulant based contrast can be found in [15].

4. REFERENCES

[1] P. Comon, "Independent components analysis, a new concept ?," *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.

[2] P. J. Huber, "Projection pursuit," *Ann. Statist.*, vol. 13, no. 2, pp. 435–475, 1985.

[3] Th. Cover and J. A. Thomas, *Elements of Information Theory*, New-York: Wiley, 1991.

[4] N. M. Blachman, "The convolution inequality for entropy powers," *IEEE Trans. Inform. Theory*, vol. 11, pp. 267–271, 1968.

[5] D.T. Pham, "Blind separation of instantenaous mixtrures of sources based on order statistics," *IEEE Trans. Signal Preocessing*, vol. 48, no. 2, pp. 363–375, 2000.

[6] E. Moreau and O. Macchi, "High order contrasts for self-adaptive source separation," *International Journal of Adaptive Control and Signal Processing*, vol. 10, no. 1, pp. 19–46, 1996.

[7] B. Stoll. and E. Moreau, "A generalized ica algorithm," *IEEE Signal Processing Letters*, vol. 7, no. 4, pp. 90–92, 2000.

[8] E. Moreau, , and N. Thirion-Moreau, "Non symmetrical contrasts for sources separation," *IEEE Trans. Signal Processing*, vol. 47, no. 8, pp. 2241–2252, 1999.

[9] E. Moreau, "A generalization of joint-diagonalization criteria for source separation," *IEEE Trans. Signal Processing*, vol. 10, 2001, To appear.

[10] D. T. Pham, "Mutual information approach to blind speparation for stationary sources," in *Proceeding of ICA'99 Conference*, J. F. Cardoso, C. Jutten, and Ph. Loubaton, Eds., Aussois, France, Jan. 1999, pp. 215–220.

[11] D. T. Pham, "Blind separation of instantaneous mixture of sources via the Gaussian mutual information criterion," in *Signal Processing X (Proceeding of EUSIPCO 2000 Conference)*, Moncef Gabbouj and Pauli Kuosmanen, Eds., Tampere, Finland, Sept. 2000, pp. 3–6.

[12] D. T. Pham, "Blind separation of instantaneous mixture of sources via the Gaussian mutual information criterion," *Signal Processing*, 2001, To appear.

[13] D. T. Pham, "Joint approximate diagonalization of positive definite matrices," *SIAM J. on Matrix Anal. and Appl.*, 2001, To appear.

[14] P. Comon, "Contrast for multichannel blind deconvolution," *IEEE Signal Processing Letters*, vol. 3, no. 7, pp. 209–211, 1996.

[15] E. Moreau and J.-C. Pesquet, "Generalized contrasts for multichannel blind deconvolution of linear systems," *IEEE Trans. Signal Processing Letters*, vol. 14, no. 6, pp. 182–183, 1997.