Blind Separation of Instantaneous Mixture of Sources based on order statistics

Dinh-Tuan Pham, Member IEEE

Abstract

In this paper we introduce a novel procedure for separating an instantaneous mixture of source based on the order statistics. The method is derived in a general context of independence component analysis, using a contrast function defined in term of the Kullback-Leibler divergence or of the mutual information. We introduce a discretized form of this contrast permitting its easy estimation through the order statistics. We show that the local contrast property is preserved and also derive a global contrast exploiting only the information of the support of the distribution (in the case this support is finite). Some simulations are given illustrating the good performance of the method.

Index Terms

Separation of sources. Contrast. Kullback-Leibler divergence. Entropy. Independence component analysis. Mutual information. Order statistic. Quantile.

I. INTRODUCTION

The problem of separation of sources has been the subject of rapid development in the signal processing literature recently (see for example [1], [4], [6], [7], [8], [10], [11] and the review [2] and the references therein). We consider here the simplest case where one observes K sequences $X_1(t)$, ..., $X_K(t)$, each being a linear combination of K independent sources $S_1(t)$, ..., $S_K(t)$. Thus $\mathbf{X}(t) = \mathbf{A}\mathbf{S}(t)$ where $\mathbf{X}(t)$ and $\mathbf{S}(t)$ denote the vectors with components $X_1(t)$, ..., $X_K(t)$ and $S_1(t)$, ..., $S_K(t)$, respectively, and $\mathbf{A}$ is some square matrix. The problem is to recover the sources from the observations, without any a priori knowledge on their probabilistic

D.T. Pham is with the Laboratory of of Modeling and Computation, IMAG - C.N.R.S., B.P. 53X, 38041 Grenoble cédex, France. E-mail: Dinh-Tuan.Pham@imag.fr

structure except that they are mutually independent (and therefore the separation is often referred to as blind). This problem is closely related to that of independent component analysis (ICA), introduced by Comon [5]. Given a probability distribution $P_{\mathbf{X}}$ of some a random vector $\mathbf{X}$, the ICA problem is to find a (square) transformation matrix $\mathbf{B}$ such that the components of the transformed vector $\mathbf{BX}$ are as independent (in some sense) as possible. Clearly if $\mathbf{X} = \mathbf{AS}$ with $\mathbf{S}$ having independent components, then $\mathbf{B} = \mathbf{A}^{-1}$ is a solution to the ICA problem. However, ICA is somewhat more general since it is not assumed that the observation vector $\mathbf{X}$ arises from a source separation model. (If this is not the case, ICA, of course, will not yield independent components but only a solution for which they as as independent as possible.)

In this paper we propose a method of separation of source based on the ICA concept. We consider the same contrast function as used in [9] to measure the lack of independence between the components of the transformed vector, so that its minimization would provide the separation. But unlike [9] which estimates this contrast directly through density estimation, we rewrite it in term of the quantile function leading to a discretized form, which can be easily estimated through the order statistics. We shall show that this form still retain the *local* contrast property. Further, in the case of distribution of finite support, we introduce an even simpler criterion exploiting only the information on this support, which we show to be a global contrast. We also provide a computational method for minimizing our criterion efficiently. Finally some simulations are performed showing the good behavior of the procedure.

For ease of reading, proofs of results will be relegated to an appendix. This part also contains some details on our computational algorithm and some explicit calculations in a simple case to illustrate the still-adequate behavior of the procedure when some of our conditions are not met. Finally, throughout the paper, the notations $^{\mathrm{T}}$ and tr will denote the transpose and the trace.

II. METHOD

We shall consider the problem in the general framework of ICA. Let $\mathbf{X}$ be random K-vector with a given probability distribution $P_{\mathbf{X}}$. The problem is to find a linear invertible transformation defined by a matrix $\mathbf{B}$ such that the components $(\mathbf{BX})_1, \ldots, (\mathbf{BX})_K$ of the transformed vector $\mathbf{BX}$ are the most independent possible. As in [9], we take as measure of independence the Kullback-Leibler entropy divergence between the joint probability distribution of the $(\mathbf{BX})_i$ and the one they would have if they are independent. Let $f_{(\mathbf{BX})_i}$ denote the density of $(\mathbf{BX})_i$

(assumed to exist) and $f_{\mathbf{BX}}$ be the joint density, this divergence is given by

$$\int \ln\Big[\frac{f_{\mathbf{BX}}(s_1,\ldots,s_K)}{\prod_{i=1}^K f_{(\mathbf{BX})_i}(s_i)}\Big] f_{\mathbf{BX}}(s_1,\ldots,s_K)ds_1\cdots ds_K = \sum_{i=1}^K H[(\mathbf{BX})_i] - H(\mathbf{BX})$$

where

$$H[(\mathbf{BX})_i] = -\int \ln[f_{(\mathbf{BX})_i}(s_i)] f_{\mathbf{BX}}(s_i)ds_i$$
$$H(\mathbf{BX}) = -\int \ln[f_{\mathbf{BX}}(s_1,\ldots,s_K)] f_{\mathbf{BX}}(s_1,\ldots,s_K)ds_i\cdots ds_K$$

are the Shannon marginal and joint entropy, respectively. Thus our independence measure is actually the mutual entropy between the components of $\mathbf{BX}$. By the same computation as in [9]: $H(\mathbf{BX}) = H(\mathbf{X}) + \ln|\det\mathbf{B}|$, $H(\mathbf{X})$ denoting the entropy of $\mathbf{X}$. Hence $H(\mathbf{BX})$ depends on $\mathbf{B}$ only through the term $\ln|\det\mathbf{B}|$ and the ICA problem is reduced to the minimization of

$$C(\mathbf{B}) = \sum_{i=1}^K H[(\mathbf{BX})_i] - \ln|\det\mathbf{B}|.$$

In the context of separation of sources, the function C is called a contrast (as introduced in [5]) meaning that if $\mathbf{X} = \mathbf{AS}$ with the matrix $\mathbf{A}$ invertible and the random vector $\mathbf{S}$ having independent components (the sources), then $C(\mathbf{B})$ is minimum if and only if $\mathbf{B} = \mathbf{A}^{-1}$ (up to the pre-multiplication by a permuted diagonal matrix).

In this work, we introduce a new way of computing the entropy based on the quantile function. We further replace the involved derivatives and integrals by divided differences and Riemann sums and we show that the resulting criterion remain a *local* contrast, under mild conditions. It involves only a finite number of values of the quantile function, which can be estimated easily from the order statistics. In the case where the probability distributions are of bounded support, we also provide a simple *global* contrast based only on information about the supports of the marginal distributions. These supports can be easily estimated by the extreme statistics[1].

Recall that the quantile function of a random variable Y is defined as

$$Q_Y(u) = \inf\{y \in \mathbb{R} : P(Y \leq y) \geq u\}.$$

We adopt this definition for $0 < u \leq 1$ but not for $u = 0$, since this would lead to $Q_Y(0) = -\infty$. We define instead $Q_Y(0) = \inf\{y \in \mathbb{R} : P(Y \leq y) > 0\}$. The function Q_Y is in fact the inverse function of the cumulative distribution function F_Y, defined as $F(y) = P(Y \leq y)$, in the case

[1]By extreme statistics, we mean the first (smallest) and last (largest) order statistics

where the later is strictly increasing. Simple computation yields that the derivative Q'_Y of Q_Y (if it exists) is related to the density f_Y of Y by

$$Q'_Y(u) = 1/f_Y[Q_Y(u)] \quad \text{or} \quad f_Y(y) = 1/Q'_Y[F_Y(y)]. \tag{2.1}$$

One then deduces that the entropy of the distribution of Y can also be written as

$$H(Y) = \int_{-\infty}^{\infty} \ln Q'_Y[F_Y(y)]\ f_Y(y)dy = \int_0^1 \ln Q'_Y(u)\ du$$

Thus our contrast function C can be written in term of the quantile functions $Q_{(\mathbf{BX})_i}$ of the components of $\mathbf{BX}$ as follows:

$$C(\mathbf{B}) = \sum_{i=1}^{K} \int_0^1 \ln Q'_{(\mathbf{BX})_i}(u)\ du - \ln|\det \mathbf{B}|.$$

In practice, the integral involved in the above criterion must be computed numerically. Therefore, we introduce a discretized form of this criterion as

$$C_d(\mathbf{B}) = \sum_{i=1}^{K}\sum_{l=2}^{L} \ln\Big[\frac{Q_{(\mathbf{BX})_i}(u_l) - Q_{(\mathbf{BX})_i}(u_{l-1})}{u_l - u_{l-1}}\Big]\frac{u_l - u_{l-1}}{u_L - u_1} - \ln|\det \mathbf{B}| \tag{2.2}$$

where $\{u_1,\ \ldots,\ u_L\}$ is a given set of strictly increasing numbers in $[0,1]$. Note that in the case the distribution has support the whole real line, taking $u_1 = 0$ would lead to $Q_{(\mathbf{BX})_i}(u_1) = -\infty$ and $u_L = 1$ would lead to $Q_{(\mathbf{BX})_i}(u_L) = \infty$. Therefore, u_1 and u_L should only be close, but not equal, to 0 and 1 in such case. Similar considerations apply when the distribution has support unbounded above or unbounded below. The difference between C and C_d is that when C is computed numerically, the number of discretizing points could be arbitrarily large, depending on the required accuracy, while C_d involves only a fixed and relatively small number of such points. Note that one can ignore any constant term on C_d and thus reduces it to

$$\sum_{i=1}^{K}\sum_{l=2}^{L} \ln[Q_{(\mathbf{BX})_i}(u_l) - Q_{(\mathbf{BX})_i}(u_{l-1})]\frac{u_l - u_{l-1}}{u_L - u_1} - \ln|\det \mathbf{B}| \tag{2.2$'$}$$

The method we propose consists in maximizing (2.2) or $(2.2')$, with the theoretical quantile $Q_{(\mathbf{BX})_i}$ being replaced by their empirical counterparts. Let $\mathbf{X}(1), \ldots, \mathbf{X}(n)$ be an observed record of independent random vectors with the same distribution as $\mathbf{X}$, from which one can construct a corresponding records $(\mathbf{BX})_i(1), \ldots, (\mathbf{BX})_i(n)$ of random variables with the same distribution as $(\mathbf{BX})_i$. The empirical quantile function of $(\mathbf{BX})_i$ at $j/(n+1)$, $(1 \le j \le n)$, is simply the j-th order statistic $(\mathbf{BX})_i(j:n)$, defined as the j-th smallest among the $(\mathbf{BX})_i(1),\ \ldots,\ (\mathbf{BX})_i(n)$. The empirical quantile at an arbitrary point $u \in [0,1]$ shall be taken as the one at the point of the

form $j/(n+1)$ which is closest to u. Actually, a better method to define the empirical quantile is to interpolate linearly from its values at $j/(n+1)$, but as the choices of the u_l in (2.2) is of minor importance, the above definition suffices for our purpose and has the virtue of simplicity. Thus we are led to the criterion

$$\hat{C}_d(\mathbf{B}) = \sum_{i=1}^{K}\sum_{l=2}^{L} \ln\Big[\frac{(\mathbf{BX})_i(n_l:n) - (\mathbf{BX})_i(n_{l-1}:n)}{n_l - n_{l-1}}\Big]\frac{n_l - n_{l-1}}{n_L - n_1} - \ln|\det\mathbf{B}|, \tag{2.3}$$

where $n_1, \ldots, n_L$ are integers in $\{1, \ldots, n\}$ such that $n_l/(n+1)$ is closest to u_l. We assume here that $n_l > n_{l-1}$, which is satisfied if $u_l - u_{l-1} > 1/(n+1)$. In practice, instead of choosing $u_1, \ldots, u_L$ and then finding the corresponding $n_1, \ldots, n_L$, one may simply choose the latter. However, the consideration of $u_1, \ldots, u_L$ allows us to work with the theoretical criterion C_d, without reference to the sample size n.

It is clear that the criteria C_d, $\hat{C}_d$ as well as C are invariant with respect to the pre-multiplication of their (matrix) argument by a permuted diagonal matrix. This simply reflects the well known fact that the separation of sources can be realized only up to a permutation and a change of scale [12]. Another interesting property of these criteria is their invariance with respect to linear transformations on the data. Indeed, they do not change when one subtracts from $\mathbf{X}$ or $\mathbf{X}(t)$ a constant vector. Further, pre-multiplying $\mathbf{X}$ or $\mathbf{X}(t)$ by a non singular matrix $\mathbf{T}$ and evaluating these criteria at $\mathbf{BT}^{-1}$ instead of $\mathbf{B}$, simply increases them by the constant $\ln|\det\mathbf{T}|$. Hence the above data transformation would not change the reconstructed sources (or independent components) resulting from the minimization of these criteria.

III. The contrast property

This section concerns the property of C_d being a contrast (C clearly is since the Kullback-Leibler divergence vanishes if and only if the distributions are the same). Actually we can only prove a local contrast property, in the sense that if $\mathbf{X} = \mathbf{AS}$ with $\mathbf{S}$ having independent components then $\mathbf{B} = \mathbf{A}^{-1}$ is a local minimum of C_d.

It is necessary to distinguish the cases where the distribution has infinite and has finite support. In fact it is necessary to consider the case of semi-infinite (bounded below, unbounded above or bounded above and unbounded below) support as well, but its treatment can be easily deduced from that of the above cases.

A. The case of distribution of infinite support

In this case, one must take $u_1 > 0$ and $u_L < 1$. We shall assume that the quantile function $Q_{(\mathbf{BX})_i}$ is continuously differentiable at u_1, ..., u_L, which is the same as assuming that $f_{(\mathbf{BX})_i}$ is strictly positive and continuously differentiable at $Q_{(\mathbf{BX})_i}(u_1)$, $\dots Q_{(\mathbf{BX})_i}(u_L)$. This is a very mild assumption.

We shall prove below that in the context of the separation of sources where $\mathbf{X} = \mathbf{AS}$ with $\mathbf{S}$ having independent components, the matrix $\mathbf{A}^{-1}$ is a stationary point of C_d. To show that $\mathbf{A}^{-1}$ is a local minimum, the standard method is to show that the Hessian of C_d at this point is positive definite. However this Hessian can at most be positive semi-definite since C_d is invariant with respect to scale change and is thus constant along a manifold consisting of matrices which can be deduced each from the other by pre-multiplication by a diagonal matrix. Hence we will show that, when the effect of scale change has been eliminated, the Hessian is indeed positive definite, under a mild condition involving the distribution of the source and the choices of the u_l. In practice, since the distribution of the sources is unknown, it will be very difficult to check this condition, but there are reasons to expect that it will be satisfied for a large class of distributions, if $\max\{u_1,\ u_l - u_{l-1},\ l = 2, \dots, L,\ 1 - u_L\}$ is small.

Instead of computing the gradient of C_d, it is easier to compute the relative gradient. This concept has been explained in [3], [2] and [9]. Briefly, the relative gradient is a matrix $C'_d(\mathbf{B})$ such that for any matrix ϵ tending to zero, $C_d(\mathbf{B} + \epsilon\mathbf{B}) - C_d(\mathbf{B})$ approximates to $\mathrm{tr}[\epsilon C'_d(\mathbf{B})^{\mathrm{T}}]$ with an error tending to 0 faster than ϵ. It is easily seen that the relative gradient $C'_d(\mathbf{B})$ is related to the ordinary gradient $\dot{C}_d(\mathbf{B})$ by $C'_d(\mathbf{B}) = \dot{C}_d(\mathbf{B})\mathbf{B}^{\mathrm{T}}$.

Proposition 1: The relative gradient of C_d is given by

$$C'_d(\mathbf{B}) = \sum_{l=2}^{L} \begin{bmatrix} \dfrac{\mathrm{E}_{\mathbf{X}^{\mathrm{T}}\mathbf{B}^T|(\mathbf{BX})_1}[Q_{(\mathbf{BX})_1}(u_l)] - \mathrm{E}_{\mathbf{BX}|(\mathbf{BX})_1}[Q_{(\mathbf{BX})_1}(u_{l-1})]}{Q_{(\mathbf{BX})_i}(u_l) - Q_{(\mathbf{BX})_i}(u_{l-1})} \\ \vdots \\ \dfrac{\mathrm{E}_{\mathbf{X}^{\mathrm{T}}\mathbf{B}^T|(\mathbf{BX})_K}[Q_{(\mathbf{BX})_K}(u_l)] - \mathrm{E}_{\mathbf{BX}|(\mathbf{BX})_K}[Q_{(\mathbf{BX})_K}(u_{l-1})]}{Q_{(\mathbf{BX})_i}(u_l) - Q_{(\mathbf{BX})_i}(u_{l-1})} \end{bmatrix} \frac{u_l - u_{l-1}}{u_L - u_1} - \mathbf{I}.$$

where $\mathrm{E}_{\mathbf{X}^{\mathrm{T}}\mathbf{B}^T|(\mathbf{BX})_i}(s)$ denotes the conditional expectation of $\mathbf{X}^{\mathrm{T}}\mathbf{B}^T$ given that $(\mathbf{BX})_i = s$.

One can see from the above result that the gradient indeed vanishes at $\mathbf{B} = \mathbf{A}^{-1}$, since for this $\mathbf{B}$, $\mathrm{E}_{\mathbf{X}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}|(\mathbf{BX})_i}(s)$ becomes a vector with i-th component being s and the other being constant (independent of s). Thus $\mathbf{A}^{-1}$ is a stationary point of C_d.

As a consequence of Proposition 3.1, one obtains the relative gradient of C by letting $L \to \infty$

and $\max\{u_1,\ u_2-u_1,\ \ldots,\ u_L-u_{L-1},\ 1-u_L\}\to 0$:

$$C'(\mathbf{B})=\begin{bmatrix}\int_0^1 \mathrm{E}'_{\mathbf{X}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}|(\mathbf{BX})_1}[Q_{(\mathbf{BX})_1}(u)]du\\ \vdots\\ \int_0^1 \mathrm{E}'_{\mathbf{X}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}|(\mathbf{BX})_K}[Q_{(\mathbf{BX})_K}(u)]du\end{bmatrix}-\mathbf{I},$$

$\mathrm{E}'_{\mathbf{X}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}|(\mathbf{BX})_i}(\cdot)$ denoting the derivative of the function $\mathrm{E}_{\mathbf{X}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}|(\mathbf{BX})_i}(\cdot)$. This matrix can be written differently. From (2.1)

$$\int_0^1 \mathrm{E}'_{\mathbf{X}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}|(\mathbf{BX})_i}[Q_{(\mathbf{BX})_i}(u)]du=\int_0^1 f_{(\mathbf{BX})_i}[Q_{(\mathbf{BX})_i}(u)]\ d\,\mathrm{E}_{\mathbf{X}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}|(\mathbf{BX})_i}[Q_{(\mathbf{BX})_i}(u)].$$

By integration by parts, the last right hand side equals

$$-\int_0^1 \mathrm{E}_{\mathbf{BX}|(\mathbf{BX})_i}[Q_{(\mathbf{BX})_i}(u)]\ d\,f_{(\mathbf{BX})_i}[Q_{(\mathbf{BX})_i}(u)]$$

which can be written, after a change of integration variable, as:

$$-\int_{-\infty}^{\infty} \mathrm{E}_{\mathbf{BX}|(\mathbf{BX})_i}(s)\ d\,f_{(\mathbf{BX})_i}(s)=\mathrm{E}\{\mathbf{BX}\psi_{(\mathbf{BX})_i}[(\mathbf{BX})_i]\},$$

where $\psi_{(\mathbf{BX})_i}$ is the derivative of $-\ln f_{(\mathbf{BX})_i}$. Hence, one is led to the same formula in [9]

$$C'(\mathbf{B})=\begin{bmatrix}\mathrm{E}\{\psi_{(\mathbf{BX})_1}[(\mathbf{BX})_1](\mathbf{BX})^{\mathrm{T}}\}\\ \vdots\\ \mathrm{E}\{\psi_{(\mathbf{BX})_K}[(\mathbf{BX})_K](\mathbf{BX})^{\mathrm{T}}\}\end{bmatrix}-\mathbf{I}.$$

Consider now the Hessian of C_d. As before we shall actually work with the relative Hessian. Further, to avoid cumbersome indexes, we prefer to work with the associated bilinear form, which can be defined as the one which approximates to the maps $(\varepsilon,\epsilon)\mapsto \mathrm{tr}[\epsilon\mathbf{B}\dot{C}_d(\mathbf{B}+\varepsilon\mathbf{B})^{\mathrm{T}}]-\mathrm{tr}[\epsilon C'_d(\mathbf{B})^{\mathrm{T}}]$ with an error tending to zero faster than $\|\varepsilon\|\|\epsilon\|$ as $(\varepsilon,\epsilon)\to 0$. (see [9] for more detail). We shall compute the relative Hessian only at the stationary point $\mathbf{A}^{-1}$. Although the computation can be done at an arbitrary point, the result is quite complex and is of lesser interest.

Proposition 2: The relative Hessian of C_d at $\mathbf{A}^{-1}$ is the matrix associated with the bilinear form

$$\epsilon,\varepsilon\mapsto\sum_{1\le i\ne j\le K}(\epsilon_{ij}\omega_i\sigma_j^2\varepsilon_{ij}+\epsilon_{ij}\varepsilon_{ji})$$

where ϵ_{ij}, ε_{ij} are the general elements of ϵ, ε, $\sigma_i^2=\mathrm{var}(S_i)$ and

$$\omega_i=\sum_{l=2}^{L}\frac{\psi_{S_i}[Q_{S_i}(u_l)]-\psi_{S_i}[Q_{S_i}(u_{l-1})]}{Q_{S_i}(u_l)-Q_{S_i}(u_{l-1})}\frac{u_l-u_{l-1}}{u_L-u_1},$$

ψ_{S_i} being the negative of the logarithmic derivative of f_{S_i}, the density of the i-th source.

It is important to note that the above bilinear form does not involve the diagonal elements of ϵ and ε. This arises from the fact that the criterion C_d is invariant with respect to a change of scale: $C_d(\mathbf{B} + \epsilon\mathbf{B}) = C_d(\mathbf{B})$ if ϵ is diagonal. Hence the above bilinear form must vanish whenever one of the matrices ϵ and ε is diagonal, entailing that it must not involve any diagonal term of ϵ or ε. Since one can always pre-multiply $\mathbf{B} + \epsilon\mathbf{B}$ by a diagonal matrix such that the result has the same diagonal as $\mathbf{B}$, one can restrict the matrices ϵ and ε to have zero diagonal. With this restriction, it is easily seen that the above bilinear form is positive definite if and only if $\omega_i\omega_j\sigma_i^2\sigma_j^2 > 1$ for all $i \neq j$. A sufficient condition for this is

$$\omega_i \geq 1/\sigma_i^2 \quad \text{with equality for at most one index } i \tag{3.1}$$

Thus for a given choice of $u_1, \ldots, u_L$, C_d is a local contrast if the densities of the sources satisfy the condition (3.1). To see what this condition means, note that ω_i can be viewed as the Riemann sum approximating the integral $\int_0^1 \psi'_{S_i}[Q_{S_i}(u)]du$, $'$ denoting the derivative operator. But as we will show, this integral is bounded below by $1/\sigma_i^2$, with the bound being attained if and only if $\psi'_{S_i} \equiv 1/\sigma_i^2$. Thus, excluding the case where $\psi'_{S_i} = 1/\sigma_i^2$ for more than one index i, that is the case where there are more than one Gaussian source, the condition (3.1) will be satisfied if ω_i is a sufficiently good approximation to $\int_0^1 \psi'_{S_i}[Q_{S_i}(u)]du$. The accuracy of this approximation will depend on the smoothness of ψ_{S_i} and the finesse of the subdivision $u_1, \ldots, u_L$. As most density encountered in practice are smooth, one can expect that for a given choice of $u_1, \ldots, u_L$ not too sparse in $(0, 1)$, the condition (3.1) will be satisfied for a large class of density of sources.

We now show that $\int_0^1 \psi'_{S_i}[Q_{S_i}(u)]du \geq 1/\sigma_i^2$. First note that the integral involved can be written as $\int \psi'_{S_i}(s)f_{S_i}(s)ds$ which, by integration by parts, equals $\int \psi_{S_i}^2(s)f_{S_i}(s)ds$, provided that $f'_{S_i} \to 0$ at infinity. Further, by the Schwartz inequality $\sigma_i^2 \int \psi_{S_i}^2(s)f_{S_i}(s)ds \geq [\int (s-m_i)\psi_{S_i}(s)f_{S_i}(s)ds]^2$ where $m_i = \int sf_{S_i}(s)ds$, with equality if and only if $\psi_{S_i}(s)$ is proportional to $s - m_i$. But $\int (s-m_i)\psi_{S_i}(s)f_{S_i}(s)ds = -\int (s-m_i)f'_{S_i}(s)ds$ which equals 1 by integration by parts, provided that provided that $sf_{S_i}(s)$ and $f_{S_i}(s)$ tend to zero as $s \to \pm\infty$. Thus $\sigma_i^2 \int \psi_{S_i}^2(s)f_{S_i}(s)ds \geq 1$ with equality if and only if $\psi_{S_i}(s) = (s - m_i)/\sigma_i^2$, which yields the announced result.

B. The case of distribution of finite support

In this case, one may take $u_1 = 0$ and $u_L = 1$. Actually, it is preferable to take them so since otherwise the criterion C_d may no longer possess the local contrast property. Indeed, although the Propositions 1 and 2 are still valid, the argument near the end of the above section may not be applicable. As it can be seen from the formula in Proposition 2, ω_i should be interpreted as the Riemann approximation of the integral of $\psi'_{S_i} \circ Q_{S_i}$ over the *open interval* $(0, 1)$ *excluding its boundary*. This integral would equal that of $\psi'_{S_i} f_{S_i}$ over the interior of the support of f_{S_i} *boundary excluded* and not over the whole line. Therefore, the calculations at the end of the previous section involving integration by parts are valid only if $f_{S_i}(s)$, $f'_{S_i}(s)$ and $sf_{S_i}(s)$ tend to zero as s tends (from the interior) to the boundaries of the support of the density. This condition is reasonable in the case this support is infinite, but is often not satisfied when it is finite. A simple illustrative example is the uniform distribution. The above condition is not satisfied as f_{S_i} is constant inside its support. The function ψ_{S_i} vanishes there and hence $\omega_i = 0$ whatever the choice of the u_i. The problem arises from the fact that f_{S_i} has jumps at the boundary of its support. To see more clearly the effect of these jumps, consider an approximating density $\tilde{f}_{S_i}$ having the same support and tending smoothly to 0 at its boundary. If the approximation is accurate enough, one again has $\int \tilde{\psi}'_{S_i}(s)\tilde{f}_{S_i}(s)ds > 1/\sigma_i^2$ where $\tilde{\psi}_{S_i} = -(\ln \tilde{f}_{S_i})'$. But $-\ln \tilde{f}_{S_i}$ must tend to infinity as s tend to the boundary points of the support of f_{S_i} and as a result $\tilde{\psi}_{S_i}$ and hence $\tilde{\psi}'_{S_i}$ must be very large in a neighborhood of these points. Thus the integral of $\tilde{\psi}'_{S_i}\tilde{f}_{S_i}$ over such neighborhood could have a significant contribution to the whole integral. By taking $u_1 > 0$ and $u_L < 1$, such contribution are more or less discarded in the definition of ω_i. This explains why (3.1) could fail to hold.

The above discussion show that the condition (3.1) may fail when $0 < u_1 < u_L < 1$, if the densities of sources have finite support with jumps at the boundaries. To avoid this, one should include the information on the support by taking $u_1 = 0$ and $u_L = 1$. We will show below that the resulting criterion C_d is then a local contrast. Note that Propositions 1 and 2 here are no longer applicable, since Q_{S_i} is *not* continuously differentiable at 0 and 1. In fact with $u_1 = 0$ and/or $u_L = 1$, the criterion C_d is not differentiable at $\mathbf{A}^{-1}$ so it makes no sense to compute its gradient and Hessian. However, it still admits $\mathbf{A}^{-1}$ as a minimum local. This result, as well as the non differentiability of the criterion, is a consequence of the following Proposition.

Proposition 3: Let $0 = u_1 < u_2 \cdots < u_{L-1} < u_L = 1$ and suppose that the densities f_{S_i} have finite support. Then

$$C_d(\mathbf{A}^{-1} + \epsilon\mathbf{A}^{-1}) - C_d(\mathbf{A}^{-1}) = \sum_{1\le i\neq j\le K} \Big[\frac{\max\{\epsilon_{ij}Q_{S_j}(0), \epsilon_{ij}Q_{S_j}(1)\} - \epsilon_{ij}\mathrm{E}S_j}{Q_{S_i}(1) - Q_{S_i}(u_{L-1})}(1-u_{L-1}) - \frac{\min\{\epsilon_{ij}Q_{S_j}(0), \epsilon_{ij}Q_{S_j}(1)\} - \epsilon_{ij}\mathrm{E}S_j}{Q_{S_i}(u_1) - Q_{S_i}(0)}u_2\Big] + o(\epsilon),$$

$o(\epsilon)$ denoting a term tending to 0 faster than ϵ as $\epsilon \to 0$

Due to the presence of the operators max and min, it is clear that the right hand side of the equality in the above Proposition cannot be approximated, up to a term $o(\epsilon)$, by a linear form in ϵ. This shows that the criterion C_d is not differentiable at $\mathbf{A}^{-1}$. On the other hand, since $Q_{S_j}(0) < \mathrm{E}S_j < Q_{S_j}(1)$ the expression in the bracket $[\ ]$ in this right hand side is non negative and can be zero only if $\epsilon_{ij} = 0$ for all $i \neq j$. This shows that $C_d(\mathbf{A}^{-1} + \epsilon\mathbf{A}^{-1}) < C_d(\mathbf{A}^{-1})$ for all sufficiently small non diagonal matrix ϵ. Thus $\mathbf{A}^{-1}$ is a local minimum of C_d.

IV. A GLOBAL CONTRAST BASED ON THE SUPPORT

In the case of distribution of finite support, it is of interest to simply take $L = 2$ and $u_1 = 0$, $u_2 = 1$. This results in a very simple criterion which exploits only the information on the support and moreover is a global contrast regardless of the distribution of the sources, as will be shortly shown.

Because of its importance, we reserve the notation C^* for this contrast. Its use is further justified in the case of nearly uniform sources. Indeed, in this case, the random variable $(\mathbf{BX})_i$ will be nearly uniform for $\mathbf{B}$ close to $\mathbf{A}^{-1}$. Therefore the function $Q'_{(\mathbf{BX})_i}$ is nearly constant. On the other hand, since $1 - 1/x \le \log x \le x - 1$, one has

$$1 - \frac{Q_{(\mathbf{BX})_i}(1) - Q_{(\mathbf{BX})_i}(0)}{Q'_{(\mathbf{BX})_i}(u)} \le \ln \frac{Q'_{(\mathbf{BX})_i}(u)}{Q_{(\mathbf{BX})_i}(1) - Q_{(\mathbf{BX})_i}(0)} \le \frac{Q'_{(\mathbf{BX})_i}(u)}{Q_{(\mathbf{BX})_i}(1) - Q_{(\mathbf{BX})_i}(0)}.$$

Hence by integration

$$1 - \int_0^1 \frac{Q_{(\mathbf{BX})_i}(1) - Q_{(\mathbf{BX})_i}(0)}{Q'_{(\mathbf{BX})_i}(u)}du \le \int_0^1 \ln Q'_{(\mathbf{BX})_i}(u)du - \ln[Q_{(\mathbf{BX})_i}(1) - Q_{(\mathbf{BX})_i}(0)] \le 0.$$

But the first left hand side can be written as

$$-\frac{1}{2}\int_0^1\int_0^1 \Big\{\Big[\frac{Q'_{(\mathbf{BX})_i}(v)}{Q'_{(\mathbf{BX})_i}(u)}\Big]^{1/2} - \Big[\frac{Q'_{(\mathbf{BX})_i}(u)}{Q'_{(\mathbf{BX})_i}(v)}\Big]^{1/2}\Big\}^2 dudv$$

which is small as $Q'_{(\mathbf{BX})_i}$ is nearly constant. Therefore $\int_0^1 \ln Q'_{(\mathbf{BX})_i}(u)du$ is well approximated to (and bounded above) by $\ln[Q_{(\mathbf{BX})_i}(1) - Q_{(\mathbf{BX})_i}(0)]$. With this approximation, $C(\mathbf{B})$ becomes:

$$\sum_{i=1}^{K} \ln[Q_{(\mathbf{BX})_i}(1) - Q_{(\mathbf{BX})_i}(0)] - \ln|\det \mathbf{B}|,$$

which is no other than $C^*(\mathbf{B})$.

The contrast C^* also possesses a nice geometrical interpretation. First, let us denote by R_Y the range of the random variable Y, that is $R_Y = Q_Y(1) - Q_Y(0)$. Then one may write $C^*(\mathbf{B})$ as

$$C^*(\mathbf{B}) = \sum_{i=1}^{K} \ln R_{(\mathbf{BX})_i} - \ln|\det \mathbf{B}| = \ln\Big[\Big(\prod_{i=1}^{K} R_{(\mathbf{BX})_i}\Big)\Big/|\det \mathbf{B}|\Big]. \tag{4.1}$$

The expression inside the last bracket [ ] can be interpreted as the volume of the set

$$\{\mathbf{x} \in \mathbb{R}^K : \ \mathbf{Bx} \in [Q_{(\mathbf{BX})_1}(0), Q_{(\mathbf{BX})_1}(1)] \times \cdots \times [Q_{(\mathbf{BX})_K}(0), Q_{(\mathbf{BX})_K}(1)]\}.$$

This set is a "hyper-parallepiped" with sides parallel to the columns of $\mathbf{B}^{-1}$. Further, by the definition of the quantile function, it is also the smallest among such parallepiped which contains the support of $P_{\mathbf{X}}$. Therefore minimizing C^* amounts to looking for a "hyper-parallepiped" with smallest volume enclosing the support of $P_{\mathbf{X}}$, the matrix argument realizing the minimum of C^* is then determined by the condition that its inverse have columns parallel to the sides of this minimum parallepiped. The fact that these columns can be determined only up to a multiplicative factor simply reflects the invariance of C^* with respect to the pre-multiplication of its argument by a diagonal matrix.

In practice, one would replace the criterion C^* by its empirical counterpart $\hat{C}^*$ obtained from (4.1) by replacing $R_{(\mathbf{BX})_i}$ by the difference between the maximum and the minimum of $(\mathbf{BX})_i(1)$, ..., $(\mathbf{BX})_i(n)$. The same argument as before shows that $\hat{C}^*(\mathbf{B})$ is the volume of the smallest "hyper-parallepiped" with sides parallel to the columns of $\mathbf{B}^{-1}$ and containing all the data points $\mathbf{X}(1)$, ..., $\mathbf{X}(n)$. Minimizing $\hat{C}^*$ thus amounts to looking for a "hyper-parallepiped" with smallest volume enclosing all the data points $X(1)$, ..., $\mathbf{X}(n)$, the matrix argument realizing the minimum of $\hat{C}^*$ is then determined by the condition that its inverse have columns parallel to the sides of this minimum "hyper-parallepiped".

The above method has some similarities with the "geometrical" method introduced in Puntonet *et al.* [11], in that it also exploits only the information on the support of the distribution. However,

the method of these authors proceeds through a direct approach which is to estimate the slope of the boundary of the distribution (a parallelogram in the case of 2 sources).

We conclude this section with the global contrast property of C^*.

Proposition 4: Suppose that $\mathbf{X} = \mathbf{AS}$ with $\mathbf{S}$ having independent components and finite support. Then C^*, as defined by (4.1), admits a minimum at $\mathbf{A}^{-1}$, unique up to pre-multiplication by a permutation and a diagonal matrix.

V. Minimization of the criterion

The application of our method requires the minimization of the criterion $\hat{C}_d$, defined by (2.3). However, classical gradient based methods are not applicable since this criterion is not everywhere differentiable (but C_d does, under the assumption of section 3.1).

We propose a simple method based on the relaxation principle. We minimize $\hat{C}_d(\mathbf{B})$ successively with respect to each of the non diagonal terms of the $\mathbf{B}$, keeping the other terms fixed. One needs not to consider the diagonal terms since the criterion is invariant with respect to scale change. However, to avoid possible explosion of the non diagonal terms in some row of $\mathbf{B}$ when in reality the diagonal term on this row should vanish, it is recommended to renormalize the rows of $\mathbf{B}$, either periodically or when they becomes too large. Practically, for each successive pair (i, j) of indexes ($i \neq j$), we minimize $\hat{C}_d(\hat{\mathbf{B}} + \beta \mathbf{E}_{ij})$ with respect to β, where $\mathbf{E}_{ij}$ is the matrix with 1 at the position (i, j) and 0 elsewhere and $\hat{\mathbf{B}}$ is the estimated matrix obtained at the proceeding minimization step. Let β^* be the value of β realizing the minimum, we replace $\hat{\mathbf{B}}$ by $\hat{\mathbf{B}} + \beta^* \mathbf{E}_{ij}$ and continue with another pair. Another possibility is to minimize $\hat{C}_d(\hat{\mathbf{B}} + \beta \mathbf{E}_{ij}\hat{\mathbf{B}})$ and then replace $\hat{\mathbf{B}}$ by $(\mathbf{I} + \beta^* \mathbf{E}_{ij})\hat{\mathbf{B}}$, β^* being the value of β realizing the minimum.

We are thus led to a one-dimensional problem which is to minimize $C_d(\hat{\mathbf{B}} + \beta \mathbf{E}_i)$ with respect to the real variable β, where $\hat{\mathbf{B}}$ et $\mathbf{E}_i$ are two fixed matrices with $\mathbf{E}_i$ having all rows, except the i-th, vanishing. Denote by $\hat{S}_1, \ldots, \hat{S}_K$ the components of $\hat{\mathbf{B}}\mathbf{X}$ and by T the only non vanishing (that is the i-th) component of $\mathbf{E}_i\mathbf{X}$, one has

$$\hat{C}_d(\hat{\mathbf{B}} + \beta \mathbf{E}_i) = \hat{C}_d(\hat{\mathbf{B}}) + \sum_{l=2}^{L} \ln \Big[\frac{(\hat{S}_i + \beta T)(n_l : n) - (\hat{S}_i + \beta T)(n_{l-1} : n)}{\hat{S}_i(n_l : n) - \hat{S}_i(n_{l-1} : n)} \Big] \frac{n_l - n_{l-1}}{n_L - n_1} - \ln |\det(\mathbf{I} + \beta \mathbf{E}_i \hat{\mathbf{B}}^{-1})|$$

where $\hat{S}_i(k : n)$ and $(\hat{S}_i + \beta T)(k : n)$ are the k-th order statistics of the samples $\hat{S}_i(1), \ldots, \hat{S}_i(n)$ and of the sample $S_i(1) + \beta T(1), \ldots, S_i(n) + \beta T(n)$, respectively. Since the matrix $\mathbf{E}_i \hat{\mathbf{B}}^{-1}$ has

all its rows, except the i-th, vanish, $\det(\mathbf{I}+\beta\mathbf{E}_i\hat{\mathbf{B}}^{-1})=1+c\beta$ where c is the i-th diagonal element of $\mathbf{E}_i\hat{\mathbf{B}}^{-1}$. Note that taking $\mathbf{E}_i=\mathbf{E}_{ij}\hat{\mathbf{B}}$ would lead to $c=0$ and also $T=\hat{S}_j$. This choice is in fact the most economical computationally.

The crucial point is that the order statistics $(\hat{S}_i+\beta T)(n_l:n)$, as functions of β, are piecewise-linear and continuous, with a finite number of slope changes. (Because of such changes, these functions are not everywhere differentiable and hence so is the criterion $\hat{C}_d$). To prove this point, note that $(\hat{S}_i+\beta T)(j:n)$, by definition, equals $\hat{S}_i(k_j)+\beta T_i(k_j)$, where $\{k_1,\ \ldots,\ k_n\}$ is a permutation of $\{1,\ldots,\ n\}$ such that

$$\hat{S}_i(k_1)+\beta T(k_1)\leq\cdots\leq\hat{S}_i(k_n)+\beta T(k_n).$$

The integers k_j would depend on β, however if there is no ties among the $\hat{S}_i(1)+\beta T(1)$, …, $\hat{S}_i(n)+\beta T(n)$, the above inequality will be strict and remain so as β change provided that the change is small enough. Only when the change exceeds a threshold for which a tie among the $\hat{S}_i(1)+\beta T(1)$, …, $\hat{S}_i(n)+\beta T(n)$ occurs then the k_j may change. Let β_1, …, β_N denote the values of β, in increasing order, for which such a tie occurs and put $\beta_0=-\infty$, $\beta_{L+1}=\infty$. Then for β in each of the open intervals (β_0,β_1), …, (β_N,β_{N+1}) and for each $j\in\{1,\ldots,n\}$, $(\hat{S}_i+\beta T)(j:n)$ is of the form $\hat{S}_i(k_j)+\beta T(k_j)$, for some k_j. Therefore, at such β, the function $\hat{C}_d(\hat{\mathbf{B}}+\beta\mathbf{E}_i)$ is differentiable with derivative

$$\sum_{l=2}^{L}\Big\{\frac{T(k_l)-T(k_{l-1})}{\hat{S}_i(k_l)-\hat{S}_i(k_{l-1})+\beta[T(k_l)-T(k_{l-1})]}-\frac{c}{1+c\beta}\Big\}\frac{n_l-n_{l-1}}{n_L-n_1}=$$
$$\frac{1}{1+c\beta}\sum_{l=2}^{L}\frac{\hat{T}_i(k_l)-\hat{T}_i(k_{l-1})}{\hat{S}_i(k_l)-\hat{S}_i(k_{l-1})+\beta[T_(k_l)-T(k_{l-1})]}\ \frac{n_l-n_{l-1}}{n_L-n_1}$$

where $\hat{T}_i=T-c\hat{S}_i$. The above right hand side times $(1+c\beta)^2$, which we denote by $f(\beta)$ for simplicity, is a decreasing function of β, as its derivative equals

$$f'(\beta)=-\sum_{l=2}^{L}\frac{[\hat{T}_i(k_l)-\hat{T}_i(k_{l-1})]^2}{\{\hat{S}_i(k_l)-\hat{S}_i(k_{l-1})+\beta[T(k_l)-T(k_{l-1})]\}^2}\ \frac{n_l-n_{l-1}}{n_L-n_1}.$$

Hence, it can vanish at most once. The second derivative of $\hat{C}_d(\hat{\mathbf{B}}+\beta\mathbf{E}_i)$, at a point where f vanish, is $f'(\beta)/(1+c\beta)^2<0$. This shows that the function $\hat{C}_d(\hat{\mathbf{B}}+\beta\mathbf{E}_i)$, in each of the open intervals (β_0,β_1), …, (β_N,β_{N+1}), is either monotonous or admits a maximum. Therefore, in the close interval, it either admits a minimum at one of the boundary points or converges to minus infinity at such a point. We will admit minus infinity as valid value of a criterion, smaller than

any other. Hence, to minimize $\hat{C}_d(\hat{\mathbf{B}} + \beta\mathbf{E}_i)$, one needs only to evaluate it at β_0, ..., β_{N+1} and look for the minimum among them.

From the computational point of view, the number of values of β for which ties occurs among the $\hat{S}_i(1) + \beta T(1)$, ..., $\hat{S}_i(n) + \beta T(n)$ can be quite large for large n. But fortunately, one needs not consider all of them. If one is interested only in the n_l-th order statistics, $l = 1, \ldots, L$, then one needs actually to consider only the values of β for which a tie among the $\hat{S}_i(1) + \beta T(1)$, ..., $\hat{S}_i(n) + \beta T(n)$ occurs *at the* $n_1, \ldots, n_L$ *place*, that is for which at least one of the order statistics $(\hat{S}_i + \beta T)(n_1 : n)$, ..., $(\hat{S}_i + \beta T)(n_L : n)$ is not uniquely defined. Such values are the only ones where the function $\beta \mapsto (\hat{S}_i + \beta T)(n_1 : n)$ can have a slope jump. This can be seen in our algorithm, given in the appendix, which searches all such slope jumps efficiently. Thus, the number of values of β to be examined can usually be drastically reduced. For example, in the case $L = 2$, $n_1 = 1$, $n_L = n$, considered in section 4 (which is in fact the most favorable case), one needs to consider only the values of β for which the maximum or the minimum of $\hat{S}_i(1) + \beta T(1)$, ..., $\hat{S}_i(n) + \beta T)(n)$ occur at more than one index. In our experiences, we found that there is only a few number of such values.

VI. Some simulation examples

We report here some of our simulation results to illustrate the performance of our procedure. We consider the separation of source problem with two sources and a mixing matrix $\mathbf{A} = \mathbf{I}$. The choice of this matrix is in fact unimportant since our method is invariant with respect to linear invertible transformation: pre-multiplying the data by some matrix would post-multiply the resulting estimated reconstruction matrix $\hat{\mathbf{B}}$ by its inverse. Hence $\hat{\mathbf{B}}\mathbf{A}$ depends only on the source sequence and not on the mixing matrix $\mathbf{A}$ and the same can be said about the reconstructed sources. Four different distributions of the source vector are considered. They are summarized as follows.

Case	Source 1	Source 2
a	$\mathcal{MN}(\frac{10}{\sqrt{116}}, \frac{16}{116})$	$\mathcal{N}(0, 1)$
b	$\mathcal{U}(-\sqrt{3}, \sqrt{3})$	$\mathcal{U}(-\sqrt{3}, \sqrt{3})$
c	$\mathcal{MN}(\frac{10}{\sqrt{116}}, \frac{16}{116})$	$\mathcal{U}(-\sqrt{3}, \sqrt{3})$
d	$\mathcal{MN}(\frac{10}{\sqrt{116}}, \frac{16}{116})$	$\mathcal{MN}(\frac{7.5}{\sqrt{72.25}}, \frac{16}{72.25})$

Fig. 1. about here

where $\mathcal{N}(0,1)$ denotes the standard normal distribution, $\mathcal{U}(-\sqrt{3},\sqrt{3})$ the uniform distribution over $[-\sqrt{3},\sqrt{3}]$ and $\mathcal{MN}(m,\sigma^2)$ the (symmetrically) mixed normal distribution with density being the average of the normal density with means $-m$ and m and variance σ^2. The above distributions are the same as those considered in [13] except that here they are normalized to have unit variance.

Fig. 2. about here

The results of application of our method are reported in figures 1 – 3. Since the reconstructed source vectors are related to the data vectors by a linear transformation, we choose to represent them by the same points, marked by the sign + on the figure: the observations on different channels are represented by the usual Cartesian coordinates of such a point while the reconstructed sources are represented by the coordinates of the same point with respect to a new system of coordinates. This new system is displayed as a parallelogram in the figure. The sides of this parallelogram are, of course, parallel to the new coordinates axes, and their placement is determined by the order statistics entering the criterion. For example, if only the extreme statistics are involved, as in the case of the contrast based on support considered in section 4, then this parallelogram is simply the one with smallest volume containing all the data points, as explained in section 4. In general we draw this parallelogram according to the n_1-th and n_L-th order statistics, so that there will be $n_l - 1$ data point on its left of it, $n - n_L$ data points on the its right, $n_1 - 1$ data points below its and $n - n_L$ data points above. (Note that a data point can be both on the left and above or both on the left and below, ...). Clearly the reconstruction of the sources is good if the sides of the parallelogram are nearly parallel to the x and y axes. A more precise measure of performance are the off diagonal terms b_{12} and b_{21} of the reconstruction matrix $\mathbf{B}$ obtained from the method. Since we have taken care to work with normalized sources (that is sources having unit variance), these terms are no other than the contamination coefficients

Fig. 3. about here

introduced in [9] and [10]. More precisely, b_{12}^2 and b_{21}^2 represent the fraction of power of the contamination of source 2 to source 1 and of sources 1 to source 2, respectively. The table 1 gives these contaminations coefficients in our experiments.

TABLE I

ABOUT HERE

It can be seen that our procedure performs quite well for the uniform distribution, even if the extreme statistics are not used. This is explained in the explicit calculation in the appendix. However, in the case of the normal or mixed normal distribution, which has infinite support, the use of the criterion based only on the extreme statistics leads to bad result. Using less extreme statistics (the 4-th and 197-th order statistics) can vastly improve performance, as can been seen in figure 2 and the second column in table 1. However, using two more order statistics does not seem to change significantly the performance, as suggested by figure 3 and the third column of the table. We have also tried to use even more order statistics, without obtaining any discernible improvement (it some case, it can be worse). This suggests that the simple procedure based on only two order statistics, not the extreme ones, but not too far from them either, could suffice.

APPENDIX

A. Some explicit calculations in a simple case

We give here some explicit calculations concerning the criterion C_d in the case of two uniform sources. We do not suppose $u_1 = 0$ or $u_L = 1$, to see the effect of choosing them in the interior of $[0, 1]$.

Since the scale factor has no effect, one can assume without loss of generality that the sources are uniform over $[-\frac{1}{2}, \frac{1}{2}]$. Further, by working with $\mathbf{BA}$ in place of $\mathbf{B}$, one can reduce the problem to the case $\mathbf{A}$ is the identity matrix. Finally, since the scale factor is of no importance, one can restrict oneself to the matrix argument of C_d is of the form $\mathbf{B} = \begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix}$. Then,

$$C_d(\mathbf{B}) = \Gamma(b_{12}) + \Gamma(b_{21}) - \ln|1 - b_{12}b_{21}|$$

where

$$\Gamma(b) = \sum_{l=2}^{L} \Big\{ \ln \Big[\frac{Q_{S_1+bS_2}(u_l) - Q_{S_1+bS_2}(u_{l-1})}{u_l - u_{l-1}} \Big] \frac{u_l - u_{l-1}}{u_L - u_1}$$

Let us compute $Q_{S_1+S_2}(u)$. The density of S_1+bS_2, in the case where $|b|\leq 1$, is

$$f_{S_1+bS_2}(x)=\begin{cases}1, & |x|\leq\frac{1-|b|}{2}\\ (\frac{1+|b|}{2}-x)/|b|, & \frac{1-|b|}{2}\leq|x|\leq\frac{1+|b|}{2}\\ 0, & |x|\geq\frac{1+|b|}{2}\end{cases}$$

and in the case where $|b|>1$, is

$$f_{S_1+bS_2}(x)=\begin{cases}1/|b|, & |x|\leq\frac{|b|-1}{2}\\ (\frac{|b|+1}{2}-x)/|b|, & \frac{|b|-1}{2}\leq|x|\leq\frac{|b|+1}{2}\\ 0, & |x|\geq\frac{|b|+1}{2}\end{cases}$$

Therefore, in the case where $|b|\leq 1$,

$$Q_{S_1+bS_2}(u)=\begin{cases}\sqrt{2|b|u}-\frac{1+|b|}{2}, & u\leq\frac{|b|}{2}\\ u-\frac{1}{2}, & \frac{|b|}{2}\leq u\leq 1-\frac{1}{2}|b|\\ \frac{1+|b|}{2}-\sqrt{2|b|(1-u)}, & u\geq 1-\frac{1}{2}|b|\end{cases}$$

and in the case where $|b|\geq 1$,

$$Q_{S_1+bS_2}(u)=\begin{cases}\sqrt{2|b|u}-\frac{1+|b|}{2}, & u\leq\frac{1}{2|b|}\\ (u-\frac{1}{2})|b|, & \frac{1}{2|b|}\leq u\leq 1-\frac{1}{2|b|}\\ \frac{1+|b|}{2}-\sqrt{2|b|(1-u)}, & u\geq 1-\frac{1}{2|b|}\end{cases}$$

One can write the above results differently: for $0\leq u\leq\frac{1}{2}$,

$$Q_{S_1+bS_2}(u)=\begin{cases}u-\frac{1}{2}, & |b|\leq 2u\\ \sqrt{2|b|u}-\frac{1+|b|}{2}, & 2u\leq|b|\leq 1/(2u)\\ (u-\frac{1}{2})|b|, & |b|\geq 1/(2u)\end{cases}$$

and for $\frac{1}{2}\leq u\leq 1$, $Q_{S_1+bS_2}(u)=-Q_{S_1+bS_2}(1-u)$ (since the distribution of S_1+bS_2 is symmetric around 0).

To simplify, we will restrict ourselves to the case where $u_l=1-u_{L+1-l}$ so that $Q_{S_1+bS_2}(u_l)=-Q_{S_1+bS_2}(u_{L+1-l})$. Therefore $\Gamma(b)$ equal 0 for $|b|\leq 2u_1$, equals

$$2\ln\Big(\frac{u_2-\sqrt{2|b|u_1}+|b|/2}{u_2-u_1}\Big)\frac{u_2-u_1}{u_L-u_1}$$

for $2u_1\leq|b|\leq 2u_2$ (in the case $L\geq 3$) and one can continue the formula for higher values of b, but it rapidly becomes too complex.

From the above results, one can see that the function $C_d(\mathbf{B})$ behaves, in a neighborhood of the identity matrix, as $-\ln|1-b_{12}b_{21}|$. The last function admits indeed a stationary point at $(b_{12},b_{21})=(0,0)$ but this is a saddle point and not a minimum. Further, this neighborhood

Fig. 4. about here

shrinks to the empty set as $[u_1, u_L]$ increases to $[0, 1]$. In the limiting case where $u_1 = 0$, $u_L = 1$, we have shown that the criterion admits the identity matrix as a local minimum, but only for $L = 2$, we have been able to show that it is a global minimum. We now examine what happens in this case when $u_1 = 1 - u_2 = u > 0$. We have

$$\Gamma(b) = \begin{cases} 0, & |b| \leq 2u \\ \ln \dfrac{1 - 2\sqrt{2|b|u} + |b|}{1 - 2u}, & 2u \leq |b| \leq 1/(2u) \\ \ln |b|, & |b| \geq 1/(2u) \end{cases}$$

The figure 4 plots the level curves of C_d in the case $L = 2$ with $u_1 = 1 - u_2 = u = .01$. This figure suggests that the criterion admits two minima at $(\pm\beta, \mp\beta)$. There would be two other minima at $(\pm 1/\beta^*, \mp 1/\beta^*)$, which correspond to the same matrix $\mathbf{B}$ (up to a permutation and a scale change), but they lie outside the figure. The exact value of β^* can be obtained by looking for the minimum of the function

$$\beta \mapsto 2 \ln[(1 - 2\sqrt{2u|\beta|} + |\beta|)/(1 - 2u)] - \ln(1 + \beta^2).$$

Equating the derivative of this function to 0, one gets that β^* is the solution of

$$(1 - \sqrt{2u/|\beta|}/(1 + |\beta| - 2\sqrt{2u|\beta|}) = |\beta|/(1 + \beta^2).$$

or $\sqrt{|\beta|} = \sqrt{2u}(1 - \beta^2) + |\beta|\sqrt{|\beta|}$. This yields $\beta^* = \pm.02084$ for $u = .01$. It is seen that β^* is quite close to $\pm 2u$. Thus, although the criterion C_d does not admit the identity matrix as the minimum (because we haven't chosen $u_1 = 0$, $u_2 = 1$), the minima (not unique) are not far. They are approximately at $\begin{bmatrix} 1 & \pm 2u \\ \mp 2u & 1 \end{bmatrix}$ in the case $u_1 = 1 - u_2 = u$. From this simple example, we conjecture that using the criterion C_d with $u_1 > 0$ and $u_L < 1$ in the case of distribution of finite support, may lead to biased estimator, but the bias is often quite small and can be tolerated.

B. Order statistics of linear combination of observations

Let $\{s(1), \ldots, s(n)\}$ and $\{t(1), \ldots, t(n)\}$ be two sets of real numbers, we would like to compute efficiently the p-th order statistic of $\{s(1) + \beta t(1), \ldots, s(n) + \beta t(n)\}$, for all real numbers β.

Our method is based on the following property: if there is a β for which

$$(p-k)\{[s(p)+\beta t(p)]-[s(k)+\beta t(k)]\}\geq 0, \qquad \forall k \tag{A.1}$$

then

$$\max_{k:(p-k)[t(p)-t(k)]>0}\frac{s(p)-s(k)}{t(k)-t(p)}\leq\beta\leq\min_{k:(p-k)[t(p)-t(k)]<0}\frac{s(p)-s(k)}{t(k)-t(p)} \tag{A.2}$$

and (A.1) also holds for all other β satisfying (A.2) and further for all such β, $s(p)+\beta t(p)$ is the required order statistic. (By convention, the maximum and minimum over an empty set in $-\infty$ and ∞, respectively.) To see this note that (A.1) is equivalent to $\beta(p-k)[t(k)-t(p)]\leq s(p)-s(k)$, hence it clearly implies (A.2) and if β satisfies (A.2) then it also satisfied (A.1) for any k such that $t(k)\neq t(p)$. But if $t(k)=t(p)$ then (A.1) is satisfied for all β since it is already satisfied by some β. This proves the announced result as (A.1) clearly implies that $s(p)+\beta t(p)$ is the required order statistic.

Observe is that the order statistics are unaffected by the simultaneous reordering of the $(s(k),t(k))$. Thus we begin by reordering them into $(s_0(1),t_0(1))$, ..., $(s_0(n),t_0(n))$ such that $(p-k)[t_0(p)-t_0(k)]\geq 0$ and that $t_0(k)=t_0(p)$ implies $(p-k)[s(p)-s(k)]\geq 0$. (This can be easily done by ordering the $(t(k),s(k))$ lexicographically). Then it can be seen that (A.1), with $s_0(\cdot)$, $t_0(\cdot)$ in place of $s(\cdot)$, $t(\cdot)$, is satisfied for β large enough. Indeed, it is clearly be satisfied for those k for which $t(p)=t(k)$ and for the other k, one has $(p-k)[t(p)-t(k)]>0$ and hence it suffices that

$$\beta\geq\beta_1=\max_{k:(p-k)[t_0(p)-t_0(k)]>0}\frac{s_0(p)-s_0(k)}{t_0(k)-t_0(p)} \tag{A.3}$$

for (A.2) and hence (A.1) to hold.

By the property just proved, $s_0(\cdot)+\beta t_0(\cdot)$ is the required order statistic for all $\beta\geq\beta_1$. Let q be the index k realizing the maximum in (A.3). Then $s_0(p)+\beta_1 t_0(p)=s_0(q)+\beta_1 t_0(q)$. Hence for $\beta=\beta_1$, (A.1) still holds for the new sequences obtained by permuting $(s_0(p),t_0(p))$ with $(s_0(q),t_0(q))$, that is for the sequence $s_1(k)=s_0(k)$ if $k\notin\{p,q\}$, $=s_0(p)$ if $k=q$, $=s_0(q)$ if $k=p$, and the sequence $t_1(k)$, defined similarly. Therefore, with now $s_1(\cdot)$, $t_1(\cdot)$ in place of $s(\cdot)$, $t(\cdot)$, (A.1) holds for all β satisfying (A.2). Note that the upper bound in (A.2) cannot be less β_1 since (A.1) holds for $\beta=\beta_1$. This bound actually equals β_1 as one has $\beta_1=[s_0(p)-s_0(q)]/[t_0(q)-t_0(p)]=[s_1(p)-s_1(q)]/[t_1(q)-t_1(p)]$ and $(p-q)[t_1(p)-t_1(q)]=(p-q)[t_0(q)-t_0(p)]<0$. Therefore $s_1(p)+\beta_1 t_1(p)$ is the required order statistic for all $\beta\in[\beta_1,\beta_2]$, where β_2 is defined as in (A.3) with $s_1(\cdot)$, $t_1(\cdot)$ in place of $s_0(\cdot)$, $t_0(\cdot)$.

The above construction clearly can be iterated. At the stage m, one has constructed sequences $s_m(\cdot)$ and $t_m(\cdot)$, then one just find the maximum in (A.3) with $s_m(\cdot)$ and $t_m(\cdot)$ in place of $s_0(\cdot)$, $t_0(\cdot)$. Let β_m be the maximum found and q be the index realizing this maximum, one then permute $(s_m(p),\ t_m(p))$ with $(s_m(q),\ t_m(q))$ to obtain the new sequences and so on. The algorithm terminates when the set $\{k : (p-k)[t_m(p) - t_m(k)] > 0\}$ is empty. To show that it must terminate, note that the number of elements in the above set cannot exceed the number of (unordered) pairs (j,k) for which $(j-k)[t_m(j) - t_m(k)] > 0$. But the permutation involved in our algorithm always decreases this number at each step. The algorithm therefore will terminate after at most $n(n-1)/2$ steps, but usually it terminates much earlier.

C. Proofs of results

To prove the Propositions 1 and 2, we shall make use of the Lemma A.2 of [13], which we recall here

Lemma 1: Assume that f_X is continuously differentiable. Let g be a continuously differentiable real function on $\mathbb{R}^K$ such that

$$\int_{\mathbb{R}^{k-1}} \Big(1 + \sum_{j\neq i} |s_j|\Big) \Big| \sup_{s_i} \frac{\partial}{\partial s_i} [g(\mathbf{B}^{-1}(\mathbf{s}) f_{\mathbf{BX}}(\mathbf{s})] \Big| \prod_{j\neq i} ds_j < \infty, \quad \mathbf{s} = (s_1 \quad \cdots \quad s_K)^{\mathrm{T}}$$

Then

$$\mathrm{E}_{g(\mathbf{X})|(\mathbf{BX}+\epsilon\mathbf{BX})_i}(s) f_{(\mathbf{BX}+\epsilon\mathbf{BX})_i}(s) - \mathrm{E}_{g(\mathbf{X})|(\mathbf{BX})_i}(s)] f_{(\mathbf{BX})_i}(s)$$
$$= -\epsilon_{i\cdot} [\mathrm{E}_{g(\mathbf{X})\mathbf{BX}|(\mathbf{BX})_i}(s) f_{(\mathbf{BX})_i}(s)]' + o(\epsilon)$$

where $\mathrm{E}_{g(\mathbf{X})|(\mathbf{BX})_i}(s)$ denotes the conditional expectation of $g(\mathbf{X})$ given $that(\mathbf{BX})_i = s$, $'$ is the derivative operator, $\epsilon_{i\cdot}$ is the i-th row of ϵ and $o(\epsilon)$ denotes a term tending to zero faster than ϵ.

Proof of Proposition 1

Taking $g \equiv 1$ in Lemma 1, one gets, after an integration

$$F_{(\mathbf{BX}+\epsilon\mathbf{BX})_i}(s) - F_{(\mathbf{BX})_i}(s) = -\epsilon_{i\cdot} \mathrm{E}_{\mathbf{BX}|(\mathbf{BX})_i}(s) f_{(\mathbf{BX})_i}(s) + o(\epsilon).$$

It follows that

$$Q_{(\mathbf{BX}+\epsilon\mathbf{BX})_i}(u) - Q_{(\mathbf{BX})_i}(u) = \epsilon_{i\cdot} \mathrm{E}_{\mathbf{BX}|(\mathbf{BX})_i}[Q_{(\mathbf{BX})_i}(u)] + o(\epsilon) \tag{A.4}$$

Therefore

$$C_d(\mathbf{B}+\epsilon\mathbf{B})-C_d(\mathbf{B}) = -\mathrm{tr}(\epsilon) + \sum_{i=1}^{K}\epsilon_{i\cdot}\sum_{l=2}^{L}\frac{\mathrm{E}_{\mathbf{BX}|(\mathbf{BX})_i}[Q_{(\mathbf{BX})_i}(u_l)]-\mathrm{E}_{\mathbf{BX}|(\mathbf{BX})_i}[Q_{(\mathbf{BX})_i}(u_{l-1})]}{Q_{(\mathbf{BX})_i}(u_l)-Q_{(\mathbf{BX})_i}(u_{l-1})}\frac{u_l-u_{l-1}}{u_L-u_1}+o(\epsilon).$$

From this, one obtains the result of the Proposition. ∎

Proof of Proposition 2

Recall that the bilinear form associated with the relative Hessian is the one which approximates to the maps $(\varepsilon,\epsilon)\mapsto \mathrm{tr}[\epsilon\mathbf{B}\dot{C}_d(\mathbf{B}+\varepsilon\mathbf{B})^{\mathrm{T}}]-\mathrm{tr}[\epsilon C_d'(\mathbf{B})^{\mathrm{T}}]$ with an error tending to zero faster than $\|\varepsilon\|\|\epsilon\|$ as $(\varepsilon,\epsilon)\to 0$, $\dot{C}_d$ denoting the ordinary gradient, which is related to the relative gradient by $\dot{C}_d(\mathbf{B})\mathbf{B}^{\mathrm{T}}=C_d'(\mathbf{B})$. For $\mathbf{B}=\mathbf{A}^{-1}$, $\mathbf{BX}=\mathbf{S}$ and $C_d'(\mathbf{B})=\mathbf{0}$. Hence, one is led to look for a bilinear approximation to

$$\sum_{i=1}^{k}\sum_{l=2}^{L}\epsilon_{i\cdot}\frac{\mathrm{E}_{\mathbf{S}|S_i+\varepsilon_{i\cdot}\mathbf{S}}[Q_{S_i+\varepsilon_{i\cdot}\mathbf{S}}(u_l)]-\mathrm{E}_{\mathbf{S}|S_i+\varepsilon_{i\cdot}\mathbf{S}}[Q_{S_i+\varepsilon_{i\cdot}\mathbf{S}}(u_{l-1})]}{Q_{S_i+\varepsilon_{i\cdot}\mathbf{S}}(u_l)-Q_{S_i+\varepsilon_{i\cdot}\mathbf{S}}(u_{l-1})}\frac{u_l-u_{l-1}}{u_L-u_1}-\mathrm{tr}[\epsilon(\mathbf{I}+\varepsilon)^{-1}] \quad \text{(A.5)}$$

as $(\varepsilon,\epsilon)\to\mathbf{0}$, S_i denoting the components of $\mathbf{S}$.

Applying again the Lemma 1, with $\mathbf{B}=\mathbf{A}$ and g being a component of $\mathbf{BX}$ or the constant 1, one gets

$$\mathrm{E}_{\mathbf{S}^{\mathrm{T}}|S_i+\varepsilon_{i\cdot}\mathbf{S}}(s)f_{S_i+\varepsilon_{i\cdot}\mathbf{S}}(s)=\mathrm{E}_{\mathbf{S}^{\mathrm{T}}|S_i}(s)f_{S_i}(s)-\varepsilon_{i\cdot}[\mathrm{E}_{\mathbf{S}\mathbf{S}^{\mathrm{T}}|S_i}(s)f_{S_i}(s)]'+o(\varepsilon)$$

$$f_{S_i+\varepsilon_{i\cdot}\mathbf{S}}(s)=f_{S_i}(s)-\varepsilon_{i\cdot}[\mathrm{E}_{\mathbf{S}|S_i}(s)f_{S_i}(s)]'+o(\varepsilon)$$

Therefore

$$\mathrm{E}_{\mathbf{S}^{\mathrm{T}}|S_i+\varepsilon_{i\cdot}\mathbf{S}}(s)-\mathrm{E}_{\mathbf{S}^{\mathrm{T}}|S_i}(s)= \varepsilon_{i\cdot}\{\mathrm{E}'_{\mathbf{S}|S_i}(s)\mathrm{E}_{\mathbf{S}^{\mathrm{T}}|S_i}(s)-\mathrm{E}'_{\mathbf{S}\mathbf{S}^{\mathrm{T}}|S_i}(s)+\psi_{S_i}(s)[\mathrm{E}_{\mathbf{S}\mathbf{S}^{\mathrm{T}}|S_i}(s)-\mathrm{E}_{\mathbf{S}|S_i}(s)]\mathrm{E}_{\mathbf{S}^{\mathrm{T}}|S_i}(s)\}+o(\varepsilon).$$

Since the components S_i of $\mathbf{S}$ are independent, the first term in the above right hand side reduces to a vector with j-th component given by

$$\varepsilon_{ij}\psi_{S_i}(s)\mathrm{var}(S_j)+o(\varepsilon) \quad \text{for} \quad j\neq i, \qquad -\varepsilon_{ii}s+o(\varepsilon) \quad \text{for} \quad j=i.$$

On the other hand, from the independence of the S_i again

$$\epsilon_{i\cdot}\frac{\mathrm{E}_{\mathbf{S}|S_i}[Q_{S_i+\varepsilon_{i\cdot}\mathbf{S}}(u_l)]-\mathrm{E}_{\mathbf{S}|S_i}[Q_{S_i+\varepsilon_{i\cdot}\mathbf{S}}(u_{l-1})]}{Q_{S_i+\varepsilon_{i\cdot}\mathbf{S}}(u_l)-Q_{S_i+\varepsilon_{i\cdot}\mathbf{S}}(u_{l-1})}=\epsilon_{ii}$$

Finally, $\mathrm{tr}[\epsilon(\mathbf{I}+\varepsilon)^{-1}] = \mathrm{tr}(\epsilon) - \mathrm{tr}(\epsilon\varepsilon) + o(\|\varepsilon\|\|\epsilon\|)$ as $(\epsilon,\varepsilon) \to \mathbf{0}$. By combining all the above results, one gets finally from (A.5) that, as $(\epsilon,\varepsilon) \to 0$,

$$\mathrm{tr}[\epsilon\mathbf{A}^{-1}\dot{C}_d(\mathbf{A}^{-1}+\varepsilon\mathbf{A}^{-1})] = \sum_{1\le i\ne j\le K} \epsilon_{ij}\omega_i\sigma_j^2\varepsilon_{ij} + \epsilon_{ij}\varepsilon_{ji} + o(\|\epsilon\|\|\varepsilon\|)$$

where σ_i^2 is the variance of S_i, ϵ_{ij}, ε_{ij} are the general elements of ϵ, ε and ω_i is as given is in Proposition 3.2. The result follows. ∎

Proof of Proposition 3

First observe that for $\mathbf{B} = \mathbf{A}^{-1}$, one has $\mathbf{BX} = \mathbf{S}$ and

$$Q_{(\mathbf{S}+\epsilon\mathbf{S})_i}(1) = \max\{(1+\epsilon_{ii})Q_{S_i}(0), (1+\epsilon_{ii})Q_{S_i}(1)\} + \sum_{j\ne i}\max\{\epsilon_{ij}Q_{S_j}(0), \epsilon_{ij}Q_{S_j}(1)\},$$

due to the independence of the components S_i of $\mathbf{S}$. Further, if $\epsilon_{ii} \ge -1$, the first term in the above right hand side reduces to $(1+\epsilon_{ii})Q_{S_i}(1)$ and hence in this case

$$Q_{(\mathbf{S}+\epsilon\mathbf{S})_i}(1) - Q_{S_i}(1) = \epsilon_{ii}Q_{\mathbf{S}_i}(1) + \sum_{j\ne i}\max\{\epsilon_{ij}Q_{S_j}(0), \epsilon_{ij}Q_{S_j}(1)\}$$

Similarly, one has

$$Q_{(\mathbf{S}+\epsilon\mathbf{S})_i}(0) - Q_{S_i}(0) = \epsilon_{ii}Q_{\mathbf{S}_i}(0) + \sum_{j\ne i}\min\{\epsilon_{ij}Q_{S_j}(0), \epsilon_{ij}Q_{S_j}(1)\},$$

provided that $\epsilon_{ii} > -1$.

On the other hand, for $0 < u < 1$, (A.4) yields, in the particular case where $\mathbf{B} = \mathbf{A}^{-1}$

$$Q_{(\mathbf{S}+\epsilon\mathbf{S})_i}(u) - Q_{S_i}(u) = \epsilon_{ii}Q_{S_i}(u) + \sum_{j\ne i}\epsilon_{ij}\mathrm{E}S_j + o(\epsilon)$$

(under the condition that $f_{(\mathbf{BX})_i}$ is continuous and does not vanishes at $Q_{S_i}(u)$).

Therefore, combining the above results, one gets the approximation to $C_d(\mathbf{A}^{-1}+\epsilon\mathbf{A}^{-1}) - C_d(\mathbf{A}^{-1})$, as given in the Proposition. ∎

Proof of Proposition 4

By assumption $\mathbf{X} = \mathbf{AS}$ with $\mathbf{S}$ having independent components $S_1, \ldots, S_K$. Thus, write $(\mathbf{BX})_i$ as $\sum_{j=1}^K \tilde{b}_{ij}S_j$ where $\tilde{b}_{ij}$ denotes the general term of the matrix $\tilde{\mathbf{B}} = \mathbf{BA}$, one gets, from the independence of the S_i, $R_{(\mathbf{BX})_i} = \sum_{j=1}^K |\tilde{b}_{ij}|R_{S_j}$. It follows that

$$C^*(\mathbf{B}) = \ln|\det\mathbf{A}| + \ln\Big[\Big(\prod_{i=1}^K\sum_{j=1}^K |\tilde{b}_{ij}|R_{S_j}\Big)/|\det\tilde{\mathbf{B}}|\Big].$$

But

$$\prod_{i=1}^{K}\sum_{j=1}^{K}|\tilde{b}_{ij}|R_{S_j} = R_{S_1}\cdots R_{S_K}\sum_{\pi}|\tilde{b}_{i\pi_j}| + \text{a sum of nonnegative termes}$$

where in the first sum, π runs through all permutations of the set $\{1,\ldots,K\}$. Therefore,

$$\prod_{i=1}^{K}\sum_{j=1}^{K}|\tilde{b}_{ij}|R_{S_j} \geq R_{S_1}\cdots R_{S_K}\det\tilde{\mathbf{B}}.$$

Further, one can verify that the above inequality reduces to an equality if and only if $\tilde{\mathbf{B}}$ is a product of a permutation and a diagonal matrix (the case where $\tilde{\mathbf{B}}$ contains a vanishing row being excluded). The results of the Proposition follows. ∎

REFERENCES

[1] Cardoso, J.P. "Iterative technique for blind source separation using only fourth order cumulants". In *Signal Processing VI: Theory and Application* (Proc. of EUSIPCO 92), J. Vandewalle, R. Boite, M. Moonen & A. Oosterlink eds, 1992, 739–742.

[2] Cardoso, J.-F. "Blind Signal Separation: Statistical principles", *Proc. IEEE*, **90**, 2009–2026. Aug. 1998.

[3] Cardoso, J. F., Laheld, B. "Equivariant adaptive source separation" *IEEE Trans. SP* **44**, 12, 3017–3030, 1996.

[4] Comon, P. "Blind Separation of sources: Problems statement". Signal Processing **24**, 1991, 11–20.

[5] Comon, P. "Independent component analysis, a new concept". Signal Processing **36**, 3, 1994, 287–314.

[6] Gaeta, M., Lacoume, J. L. "Source separation without a priori knowledge: the maximum likelihood approach". In Signal Processing V (Proc. of EUSIPCO 90), L.Tores, E. MasGrau and M. A. Lagunas eds, 1990, 621–624.

[7] Jutten, C. Herault, J. "Blind separation of sources, Part I: an adaptative algorithm based on neuromimetic structure". Signal Processing **24**, 1991, 1–10.

[8] Loubaton, Ph., Delfosse, N. "Séparation adaptive de sources indépendantes par une approch de déflation". In XIV Colloque sur le Traitement du Signal et des Images (GRETSI), Juan-Les-Pins, Sept. 1993, 325–328

[9] Pham, D. T. "Blind Separation of Instantaneous Mixture of Sources via an Independent Component Analysis." *IEEE Trans. SP* **44**, 11, 2768–2779, 1996.

[10] Pham, D. T., Garat, Ph. "Blind separation of a mixture of independent sources through a quasi-maximum likelihood approach". *IEEE Trans. SP* **45**, 7, 1712–1725, 1997.

[11] Puntonet, C. G., Prieto A., Jutten C, Rodriguez-Alvarez M and Ortega J. "Separation of sources: A geometrical based procedure for reconstsruction of n-valued signal. *Singal Processing* **46**, 3, 267–284, 1995.

[12] Tong, L., Inouye, Y., Liu, R. "Waveform preserving blind estimation of multiple independent sources." *IEEE Trans. on SP* **41**, 7, 1993, 2461–2470.

PLACE
PHOTO
HERE

Dinh-Tuan Pham was born in Hano, VietNam, on February 10, 1945. He graduated from the School of Applied Mathematics and Computer Sccience (ENSIMAG) of the Polytechnic Institute of Grenoble in 1968. He received the Ph. D. degree in Statistics in 1975 from the University of Grenoble. He was a Postdoctoral Fellow at Berkeley (Department of Statistics) in 1977-1978 and a Visisting Professor at Indiana University (Department of Mathematics) at Bloomingtion in 1879-1980. He is currently Director of Research at the French “Centre National de Recherche Scientifique”. His researches includes time series analysis, signal modelling, array processing and biomedical signal processing.