INDEPENDENCE/DECORRELATION MEASURES WITH APPLICATIONS TO OPTIMIZED ORTHONORMAL REPRESENTATIONS

Eric MOREAU [1] *and Jean-Christophe PESQUET* [2]

[1] MS-GESSY, ISITV, BP 56, F-83162 La Valette du Var Cedex, France
e-mail: moreau@isitv.univ-tln.fr
[2] Univ. Paris Sud, LSS/ESE, Plateau de Moulon, F-91192 Gif-sur-Yvette Cedex, France
e-mail: pesquet@lss.supelec.fr

ABSTRACT

In this work, extended forms of contrast functions are introduced to provide statistical measures of independence for orthogonal mixtures. We also define semicontrasts based on second-order statistics which, in some cases, may be sufficient to separate the mixed sources. The corresponding criteria are then used to obtain an optimized representation of a stochastic process in an orthonormal basis of wavelet packets or local cosines.

1. INTRODUCTION

We consider a linear orthogonal mixture of N real signals called sources. The vector of sources $\mathbf{a} = (a_1, \ldots, a_N)^T$ consists of N random variables. Furthermore, we assume that the sources do not necessarily have the same variance. The N outputs of the mixture are called the observations and the vector of observations is denoted by $\mathbf{x}$. In matrix and vector notations, the input/output relations of the mixing system are

$$\mathbf{x} = \mathbf{G}\mathbf{a} \tag{1}$$

where $\mathbf{G}$ is the orthogonal matrix characterizing the mixture. The sources are assumed to satisfy the following assumption:
A. The sources a_i, $i \in \{1, \ldots, N\}$, are statistically mutually independent.
Although not always explicitly mentioned, we will assume that the cumulants of some random variables exist whenever they are introduced. The general independence (resp. decorrelation) problem consists in determining a linear system operating on the observation vector $\mathbf{x}$, such that its N outputs y_i, $i \in \{1, \ldots, N\}$, are statistically independent (resp. decorrelated). This transformation reads

$$\mathbf{y} = \mathbf{H}\mathbf{x} \tag{2}$$

where $\mathbf{y} = (y_1, \ldots, y_N)^T$ is the output vector and $\mathbf{H}$ is an orthogonal matrix to be determined.
Making the components of $\mathbf{y}$ independent is closely related to the source separation problem because the former task is equivalent to restore the N independent sources. The independence is realized iff the global transform matrix

$$\mathbf{S} \triangleq \mathbf{H}\mathbf{G} \tag{3}$$

satisfies the so-called *independence property*: $\mathbf{S} = \mathbf{D}\mathbf{P}$ where $\mathbf{D} = \mathrm{diag}(\pm 1, \ldots, \pm 1)$ and $\mathbf{P}$ is a permutation matrix. Since one wishes to obtain a statistically independent vector $\mathbf{y}$, a measure of independence is needed. In the source separation context, such measures have been introduced by Comon [4] and are called *contrast functions* or *contrasts*. In this work, we focus on mixtures belonging to the set $\mathcal{U}$ of orthogonal matrices. The subset of $\mathcal{U}$ satisfying the independence property is denoted by $\mathcal{P}$. The set of the considered independent random source vectors is designated by $\mathcal{A}_i$ and the set of random vectors $\mathbf{y} = \mathbf{S}\mathbf{a}$ where $\mathbf{a} \in \mathcal{A}_i$ and $\mathbf{S} \in \mathcal{U}$ is denoted by $\mathcal{Y}_i$. We now recall the definition of a contrast [4]:

Definition 1 A contrast on $\mathcal{Y}_i$ is a multivariate mapping $\mathcal{I}$ from the set $\mathcal{Y}_i$ to $\mathbb{R}$ which satisfies the following three requirements:
R1. $\forall \mathbf{a} \in \mathcal{Y}_i$, $\forall \mathbf{S} \in \mathcal{P}$, $\mathcal{I}(\mathbf{S}\mathbf{a}) = \mathcal{I}(\mathbf{a})$;
R2. $\forall \mathbf{a} \in \mathcal{A}_i$, $\forall \mathbf{S} \in \mathcal{U}$, $\mathcal{I}(\mathbf{S}\mathbf{a}) \leq \mathcal{I}(\mathbf{a})$;
R3. $\forall \mathbf{a} \in \mathcal{A}_i$, $\forall \mathbf{S} \in \mathcal{U}$, $\mathcal{I}(\mathbf{S}\mathbf{a}) = \mathcal{I}(\mathbf{a}) \Leftrightarrow \mathbf{S} \in \mathcal{P}$.

2. STATISTICAL MEASURES

2.1. Measuring independence

A first contrast was proposed in [4], which is expressed as the sum of the squares of the auto-cumulants of the outputs of the separating system. In [5], it was shown that squaring the cumulants is not necessary to define a contrast.

It is interesting to note that there exists a common point between the two aforementioned contrasts. They are indeed sums of some convex functions, $(\cdot)^2$ or $(\cdot)$, of the absolute values of the q-th order cumulants. In this work, a generalization of the above measures of independence is proposed which is based on a convexity property.

Define

$$\mathcal{I}^f_{q_1,\dots,q_Q}(\mathbf{y}) \triangleq \sum_{i=1}^{N} \mathcal{J}^f_{q_1,\dots,q_Q}(y_i) \tag{4}$$

where $Q \in \mathbb{N}^*$, $(q_1, \dots, q_Q) \in (\mathbb{N}^* \setminus \{1\})^Q$,

$$\mathcal{J}^f_{q_1,\dots,q_Q}(y_i) \triangleq f\left(|\mathsf{Cum}_{q_1} y_i|, \dots, |\mathsf{Cum}_{q_Q} y_i|\right)$$

and $\forall k \in \{1, \dots, Q\}$, $\mathsf{Cum}_{q_k} y_i$ is the cumulant of order q_k of y_i. Here, f is a function defined on $(\mathbb{R}_+)^Q$, satisfying three properties:

(i) f is convex;

(ii) f is increasing w.r.t. any of its variables;

(iii) f has a unique minimum (which is at the origin).

In [8], the following result is proved:

Proposition 1 Under asumptions (i), (ii), (iii) and if (iv) there exists at most one $i \in \{1, \dots, N\}$ such that $\forall k_i \in \{1, \dots, Q\}$, $\mathsf{Cum}_{q_{k_i}} a_i = 0$;

(v) $\forall i \in \{1, \dots, Q\}$, $q_i \geq 3$;

the function $\mathcal{I}^f_{q_1,\dots,q_Q}$ is a contrast on $\mathcal{Y}_i$.

New forms of contrasts can be derived from the above proposition [8]. In particular, as also mentioned in [10], cumulants of different orders can be combined to improve the robustness of a contrast w.r.t. the variability of the statistics of the sources. Connections between some of these criteria and contrasts based on cross-cumulants have also been exhibited in [8].

2.2. Measuring decorrelation

Proposition 1 prescribes the use of high-order statistics to build contrast functions. Second-order moments can however be sufficient in some specific situations. The corresponding criteria are measures of decorrelation which will be called *semicontrasts* subsequently. First of all, we have to replace the assumption **A** by the weaker assumption:

A'. The sources a_i, $i \in \{1, \dots, N\}$, are statistically mutually uncorrelated.

Now the set of random vectors satisfying **A'** will be denoted by $\mathcal{A}_d$ and the set of vectors $\mathbf{y} = \mathbf{Sa}$ where $\mathbf{a} \in \mathcal{A}_d$ and $\mathbf{S} \in \mathcal{U}$ will be denoted by $\mathcal{Y}_d$. In fact, $\mathcal{Y}_d$ is the whole set of second-order random vectors as the linear decomposition of such a vector in uncorrelated components is always possible by performing its Karhunen-Loève transform.

Let us define a semicontrast:

Definition 2 A semicontrast on $\mathcal{Y}_d$ is a multivariate mapping $\mathcal{D}$ from the set $\mathcal{Y}_d$ to $\mathbb{R}$ which satisfies the following three requirements:

R1'. $\forall \mathbf{a} \in \mathcal{Y}_d$, $\forall \mathbf{S} \in \mathcal{P}$, $\mathcal{D}(\mathbf{Sa}) = \mathcal{D}(\mathbf{a})$;

R2'. $\forall \mathbf{a} \in \mathcal{A}_d$, $\forall \mathbf{S} \in \mathcal{U}$, $\mathcal{D}(\mathbf{Sa}) \leq \mathcal{D}(\mathbf{a})$;

R3'. $\forall \mathbf{a} \in \mathcal{A}_d$, $\forall \mathbf{S} \in \mathcal{U}$, $\mathcal{D}(\mathbf{Sa}) = \mathcal{D}(\mathbf{a}) \Leftrightarrow \mathbf{Sa} \in \mathcal{A}_d$.

We have the following result [8]:

Proposition 2 If φ is a strictly convex function from $(\mathbb{R}_+)^N$ to $\mathbb{R}$ which is invariant under any permutation of its variables, then a semicontrast on $\mathcal{Y}_d$ is given by

$$\mathcal{D}^\varphi(\mathbf{y}) \triangleq \varphi(\mathsf{Var}\,\mathbf{y}) \tag{5}$$

where $\mathsf{Var}\,\mathbf{y} \triangleq (\sigma^2_{y_1}, \dots, \sigma^2_{y_N})^T$ denotes the vector of component variances of $\mathbf{y}$.

An example of such a measure of decorrelation is

$$\varphi(\mathsf{Var}\,\mathbf{y}) = \sum_{i=1}^{N} f(\sigma^2_{y_i}) \tag{6}$$

where f is a strictly convex function from $\mathbb{R}_+$ to $\mathbb{R}$. This latter result was already obtained in [9] when $f(z) = z \log z$.

In general, decorrelation is a necessary but not sufficient condition for independence. However, a semicontrast is a contrast on $\mathcal{Y}_d$ when all the sources have different variances. Another interesting property can also be stated:

Proposition 3 If $\mathcal{I}$ is a contrast on $\mathcal{Y}_i \subset \mathcal{Y}_d$ and $\mathcal{D}$ is a semicontrast on $\mathcal{Y}_d$, then $\mathcal{I} + \mathcal{D}$ is a contrast on $\mathcal{Y}_i$.

Proposition 3 allows us to define contrast functions involving second and higher-order statistics and, thus, it provides a more general characterization of contrasts than Proposition 1.

3. APPLICATIONS TO "BEST BASIS" SEARCH

3.1. Connections between source separation and best basis search

In this part, we consider an application of measures of decorrelation/independence to the search of the "best" matched orthonormal representation of a discrete-time random signal $(y_n)_{n\in\mathbb{Z}}$, which is observed on a finite interval $\{1,\ldots,N\}$, $N\in\mathbb{N}^*\setminus\{1\}$. This problem can be viewed as a large dimension source separation problem as the size N of the vector $\mathbf{y}=(y_1,\ldots,y_N)^T$ of observations is generally large. In this context, it would not be feasible to search the optimal matrix $\mathbf{H}$ within the whole set of orthogonal matrices. To overcome this difficulty, we restrict our search to a specific subset of $\mathcal{U}$.

We can proceed in the following way to do this. First, $\forall j\in\{0,\ldots,J-1\}$ with $J\in\mathbb{N}^*$ and $J\le\log_2 N$, we define a set $\{\mathbf{H}^N_{j+1,m}, m\in\{0,\ldots,2^{j+1}-1\}\}$ of $N/2^{j+1}\times N$ matrices satisfying $\forall m\in\{0,\ldots,2^j-1\}$,

there exists an $N/2^j\times N/2^j$ orthogonal matrix $\mathbf{T}^N_{j,m}$ such that $\begin{bmatrix}\mathbf{H}^N_{j+1,2m}\\ \mathbf{H}^N_{j+1,2m+1}\end{bmatrix}=\mathbf{T}^N_{j,m}\mathbf{H}^N_{j,m}$

and $\mathbf{H}^N_{0,0}$ is an orthogonal matrix. Define

$$\mathcal{P}_J\triangleq\{I_{j_1,m_1},\ldots,I_{j_P,m_P}\}$$

where $P\in\{1,\ldots,2^J\}$ and $I_{j_p,m_p}=[2^{-j_p}m_p,2^{-j_p}(m_p+1))$ with $\forall p\in\{1,\ldots,P\}$, $j_p\in\{0,\ldots,J\}$, $m_p\in\{0,\ldots,2^{j_p}-1\}$ and

$$\begin{cases}\forall p\in\{1,\ldots,P-1\}\ , & 2^{-j_p}(m_p+1)=2^{-j_{p+1}}m_{p+1}\\ m_1=0\ , & 2^{-j_P}(m_P+1)=1\ .\end{cases}$$

Such a set $\mathcal{P}_J$ will be called a dyadic partition of $[0,1)$. Then, it can be checked that an $N\times N$ orthogonal matrix is obtained as

$$\mathbf{H}^N_{\mathcal{P}_J}\triangleq\mathbf{H}^N_{j_1,m_1}\oplus\cdots\oplus\mathbf{H}^N_{j_P,m_P}\qquad(7)$$

where $\oplus$ denotes the concatenation matrix operator defined, for all matrices $\mathbf{A}$ and $\mathbf{B}$ having the same number of columns, by $\mathbf{A}\oplus\mathbf{B}\triangleq\left[\mathbf{A}^T\mathbf{B}^T\right]^T$.
Subsequently, we focus on the specific subset of orthogonal matrices $\mathbf{H}^N_{\mathcal{P}_J}$ which is generated by considering all the possible dyadic partitions $\mathcal{P}_J$ of $[0,1)$ for a given choice of matrices $\mathbf{H}^N_{0,0}$ and $\mathbf{T}^N_{j,m}$ for $j\in\{0,\ldots,J-1\}$ and $m\in\{0,\ldots,2^j-1\}$.

If a dictionary of orthogonal transforms such as discrete-time forms of the decompositions onto wavelet packets or local cosines [2] is used, a fast tree search algorithm [2] may be applied to optimize a criterion of the form (6) or (4). The use of contrasts or semicontrasts thus provides a natural alternative to the statistical methods developed in [6, 1, 3] to optimize such time-frequency representations of a random signal.

3.2. Use of decorrelation measures

When the criterion in (6) is optimized, the resulting decomposition is suboptimal in terms of decorrelation but is less complex than the Karhunen-Loève transform. Furthermore, the method corresponds to a decomposition onto an orthonormal basis of sequences well-localized in time and frequency. This property is often desirable from a practical point of view but is not necessarily satisfied for the vectors of the Karhunen-Loève basis.

In the case of wavelet packet decompositions, asymptotic results can also be derived in order to gain insight into this kind of decompositions. Toward this end, we will focus on semicontrasts of the form:

$$\forall N\in\mathbb{N}^*\setminus\{1\}\ ,\ \mathcal{D}_N(\mathbf{y}^N)\triangleq\frac{1}{N}\sum_{i=1}^{N}f(\sigma^2_{y_i})\qquad(8)$$

where $\mathbf{y}^N=(y_1,\ldots,y_N)^T$ and f is a strictly convex function from $\mathbb{R}_+$ to $\mathbb{R}$, which does not depend on the original data length N. In particular, under some weak assumptions on boundary treatments, we have proved [8] the following result:

Proposition 4 Let $J\in\mathbb{N}^*$ and $(y_n)_{n\in\mathbb{Z}}$ be a second order stationary process to be analyzed. When $N\to\infty$, an orthonormal basis allowing to maximize any semicontrast of the form (8) within a given dictionary of possibly nonstationary wavelet packets is the equal subband analysis associated with the orthogonal matrix $\mathbf{H}^N_{J,0}\oplus\cdots\oplus\mathbf{H}^N_{J,2^J-1}$.

Recall that a nonstationary wavelet packet decompositions is implemented using different paraunitary filter banks at each resolution level.
The above proposition means that, when optimizing a semicontrast, arbitrary choices of wavelet packet bases may be interesting only for nonstationary signals.

3.3. Use of independence measures

In general, some of the variances of the components of the signal may be equal and contrast functions based on higher-order statistics must then be used to optimize the representation. Such a situation occurs in time-frequency multiplexing techniques [7] where the multiplexer is a paraunitary quadrature mirror synthesis filter bank whose structure is allowed to be adaptively varied in time. We have shown that, if the sources to be transmitted are iid, contrasts can be used by the receiver to blindly recover the structure of the multiplexer without any need for transmitting overhead information.

As an illustrative example, four PAM-4 signals have been multiplexed using the synthesis filter bank corresponding to a wavelet decomposition over 3 resolution levels. The values of a fourth-order cumulant based contrast have been evaluated at each possible node of the analysis filter banks used at the receiver. As depicted by the gray scale representation in Fig. 1, the wavelet basis is properly recovered. On the opposite, Fig. 2 shows that a standard averaged entropy criterion would lead us to conclude that the time-representation is the best one.

Figure 1: Contrast values computed from 40 realizations of a noisy multiplexed signal of length 256 ($j = 0$ at the top of the map).

Figure 2: Averaged entropy values.

4. REFERENCES

[1] J. B. Buckeit and D. L. Donoho, "Time-frequency tilings which best expose the non-Gaussian behavior of a stochastic process", *Proceedings of IEEE Symposium on Time-Frequency and Time-Scale Analysis*, Paris, France, June 18-21, 1996, pp. 1-4.

[2] R. R. Coifman and M. V. Wickerhauser "Entropy-based algorithms for best basis selection", *IEEE Trans. Inform. Theory*, Vol. 38, no. 2, pp. 713-718, Mar. 1992.

[3] R. R. Coifman and N. Saito, "The local Karhunen-Loève bases", *Proceedings of IEEE Symposium on Time-Frequency and Time-Scale Analysis*, Paris, France, June 18-21, 1996, pp. 129-132.

[4] P. Comon, "Independent Component Analysis, a New Concept?", *Signal Processing*, Vol. 36, pp 287-314, 1994.

[5] 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, Jan. 1996.

[6] H. Krim and J.-C. Pesquet, "On the statistics of best bases criteria", Wavelets and Statistics, Lecture Notes in Statistics, Springer Verlag, Ed. A. Antoniadis 1995, pp. 192-207.

[7] A.R. Lindsey, "Orthogonally multiplexed communication via wavelet packet bases", *PhD Dissertation*, Ohio University, Athens, Ohio, Aug. 1994.

[8] J.-C. Pesquet and E. Moreau, "Measures of independence for orthonormal mixtures", Technical report, LSS/Univ. Paris 11 (to be submitted to *IEEE Trans. on Information Theory*), Jul. 1996.

[9] S. Watanabe, "Karhunen-Loève expansion and factor analysis. Theoretical remarks and applications", *Trans. 4th Prague Conf. Informat. Theory, statist. Decision Functions Random Processes*, pp. 635-660, Publishing House of the Czechoslovak Academy of Sciences, 1965.

[10] D. Yellin and E. Weinstein, "Multichannel signal separation: method and analysis", *IEEE Trans. on Signal Processing*, vol. 41, No. 1, pp. 106-118, Jan. 1996.