L'Analyse en Composantes Indépendantes : un outil puissant pour le traitement de l'information

Christian JUTTEN[1], Rémi GRIBONVAL[2]

[1]INPG-LIS
4- avenue Félix Viallet, 38031 GRENOBLE Cedex
[2]IRISA
Campus de Beaulieu, 35042 RENNES Cedex
Christian.Jutten@inpg.fr, Remi.Gribonval@inria.fr

Résumé – L'analyse en composantes indépendantes (ACI) est une approche très générale qui a été développée pour résoudre le problème de séparation aveugle de sources (SAS) indépendantes. Au-delà de ce problème, l'ACI s'applique à l'étude de la représentation parcimonieuse de données liées à des phénomènes cognitifs complexes. Cet article est constitué de deux parties. La première est une synthèse sur les principes de la séparation de sources dans laquelle sont abordés la question de la séparabilité, les critères d'indépendance, et l'utilisation d'informations *a priori*. La seconde partie présente les approches de la SAS fondées sur des représentations parcimonieuses des signaux et montre comment elles permettent de traiter des problèmes sous-déterminés, et comment elles sont liées au codage parcimonieux.

Abstract – Independent Component analysis (ICA) is a very general approach, developed for solving the problem of Blind Source Separation (BSS). Beyond this problem, ICA can be used for studying sparse representations and coding in complex cognitive phenomena. This paper consists of two main parts. The first one is a survey on BSS principles, including the separability question, independence criteria and the use of priors on sources. The second part presents BSS approaches based on sparse representations of signals and points out how they can achieve BSS in underdetermined cases, and what are their relationships with sparse coding.

1 Introduction

L'analyse en composantes indépendantes (ACI) est une approche très générale qui a été initialement développée pour résoudre la problème de séparation aveugle de sources (SAS). Reposant sur l'hypothèse d'indépendance des sources, c'est une méthode essentiellement statistique. C'est aussi une méthode multi-dimensionnelle (multi-capteur), qui exploite la diversité spatiale, fréquentielle, etc. L'ACI est aussi une méthode de représentation de données complexes, à l'instar de l'Analyse en Composantes Principales, mais fortement associée à la notion de parcimonie. La première partie de ce papier sera consacrée à la SAS, la seconde aux représentations et codes parcimonieux.

2 Séparation Aveugle de Sources

Le problème de séparation de sources est un problème fondamental en traitement du signal. C'est finalement la généralisation du problème fondamental d'extraction d'un signal utile $s(t)$ dans une observation bruitée $x(t) = s(t) + n(t)$, qui est fondée sur l'exploitation d'informations *a priori* (spectres, distributions, modèle paramétrique) sur le signal ou sur le bruit. Cette généralisation prend tout son sens lorsque le signal utile et le bruit sont de même nature. Dans ce cas, on ne peut pas utiliser d'informations pour distinguer le signal utile de la perturbation. Les deux idées fondamentales de la séparation aveugle de sources consistent :

- à utiliser plusieurs capteurs (principe de diversité), qui fourniront chacun un mélange différent des sources.
- à supposer que les sources à extraire sont statistiquement indépendantes.

Si l'on observe N signaux $s_j(t)$ à l'aide de M capteurs, on obtient M mélanges $x_i(t)$:

$$\mathbf{x} = \mathcal{F}(\mathbf{s}) \tag{1}$$

où $\mathbf{x}(t)$ et $\mathbf{s}(t)$ sont les vecteurs des observations et des sources, respectivement.
Si le nombre de capteurs est supérieur ou égal au nombre de sources ($M \geq N$), et si la transformation $\mathcal{F}$ est inversible à gauche, estimer cet inverse $\mathcal{G}$ permet implicitement de séparer les sources. En effet,

$$\mathbf{y} = \mathcal{G}(\mathbf{x}) = \mathcal{G}[\mathcal{F}(\mathbf{s})] = \mathbf{s}. \tag{2}$$

Puisque la seule hypothèse sur les sources est l'indépendance statistique, $\mathcal{G}$ est estimée de sorte que les sources estimées $\mathbf{y}(t) = \mathcal{G}(\mathbf{x})(t)$ soient indépendantes. Cette hypothèse met en évidence le lien entre le *problème* de séparation aveugle de sources (SAS) et la *méthode* d'analyse en composantes indépendantes (ACI).

Dans cette partie, nous discuterons d'abord du problème de séparabilité. Ensuite, en partant de la divergence de Kullback-Leibler comme critère d'indépendance,

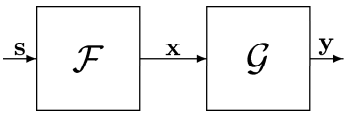


FIG. 1: *Les modèles de mélange et de séparation, dans le cas général.*

nous établirons les équations d'estimation, et montrerons les liens avec d'autres critères. Finalement, nous montrerons comment des informations *a priori*, même très faibles, peuvent être exploitées.

2.1 Séparabilité

2.1.1 La question fondamentale

En utilisant l'hypothèse d'indépendance, l'idée est d'estimer la transformation $\mathcal{G}$ de sorte que le vecteur de sources estimées $\mathbf{y}(t)$ soit à composantes indépendantes. La question essentielle est donc la suivante : est-ce que les sources $\mathbf{y}(t)$, à composantes indépendantes, sont obligatoirement les sources inconnues, $\mathbf{s}(t)$?
Autrement dit, existe-t-il des transformations $\mathcal{H} = \mathcal{G} \circ \mathcal{F}$ qui sont mélangeantes, c'est-à-dire à Jacobien non diagonal, et qui préservent l'indépendance ? La réponse est oui : l'indépendance statistique ne garantit pas la séparation ! Nous donnons ci-dessous un simple exemple, mais Darmois [27] propose une méthode simple de construction générale de telles transformations.

2.1.2 Un exemple simple

Soient $s_1(t) \in \mathbb{R}^+$, une variable aléatoire suivant une distribution de Rayleigh de densité de probablilité (ddp) $P_{s_1}(s_1(t)) = s_1(t)\exp(-s_1^2(t)/2)$, et $s_2(t) \in [0, 2\pi)$, une variable aléatoire uniforme, indépendante de $s_1(t)$. Considérons la transformation non linéaire :

$$\begin{aligned}[y_1(t), y_2(t)] &= \mathcal{H}(s_1(t), s_2(t)) \\ &= [s_1(t)cos(s_2(t)), s_1(t)sin(s_2(t))] \quad (3)\end{aligned}$$

qui est mélangeante car son Jacobien est non diagonal :

$$\mathbf{J} = \begin{pmatrix} cos(s_2(t)) & -s_1(t)sin(s_2(t)) \\ sin(s_2(t)) & s_1(t)cos(s_2(t)) \end{pmatrix}. \quad (4)$$

La ddp jointe de $y_1(t)$ et $y_2(t)$ s'écrit :

$$\begin{aligned}p_{y_1,y_2}(y_1(t), y_2(t)) &= \frac{p_{s_1,s_2}(s_1(t), s_2(t))}{|\mathbf{J}|} \\ &= \frac{1}{2\pi} exp(\frac{-y_1^2(t) - y_2^2(t)}{2}).\end{aligned}$$

Cette relation montre que les deux variables aléatoires $y_1(t)$ et $y_2(t)$ sont statistiquement indépendantes (au sens de (23)), quoique différentes des sources $s_1(t)$ et $s_2(t)$. D'autres exemples peuvent être trouvés dans la littérature (voir Lukacs [52]).

2.1.3 Cas des mélanges linéaires

Pour que l'ICA conduise à la SAS, on peut imposer des contraintes structurelles aux transformations $\mathcal{H}$, ce qui revient à restreindre l'espace de recherche. Puisque $\mathcal{H} = \mathcal{G} \circ \mathcal{F}$, ceci revient à imposer des contraintes cohérentes au mélange $\mathcal{F}$ et à la structure de séparation $\mathcal{G}$.
Dans le cas linéaire, Darmois [27] a établi (dans les années 50) le résultat suivant, connu sous le nom de théorème de Darmois-Skitovic. Soient deux variables aléatoires e_1 et e_2 :

$$e_1 = a_1 s_1 + \ldots + a_n s_n \quad (5)$$

$$e_2 = b_1 s_1 + \ldots + b_n s_n \quad (6)$$

où $s_1, \ldots, s_n$ sont des variables aléatoires indépendantes, l'indépendance de e_1 et e_2 implique que si $a_j b_j \neq 0$, la source s_j est gaussienne. Ceci montre le rôle particulier joué par les sources gaussiennes dans le cas de mélanges linéaires : pour pouvoir séparer les sources avec le seul critère d'indépendance, il faut supposer qu'au plus une source est gaussienne.

2.2 Mélanges de sources indépendantes

En pratique, l'estimation de l'inverse $\mathcal{G}$ du mélange nécessite une hypothèse (correcte) sur la nature du mélange $\mathcal{F}$ (les paramètres restant bien entendu inconnus) : on discutera dans ce paragraphe plusieurs modèles de mélanges.

2.2.1 Mélange linéaire instantané

Dans le cas le plus simple, la transformation $\mathcal{F}$ est linéaire et sans mémoire. On peut la modéliser par une matrice de mélange (inconnue) $\mathbf{A} = (a_{ij})$ à coefficients scalaires :

$$x_i(t) = \sum_j a_{ij} s_j(t),\ i = 1, \ldots, M. \quad (7)$$

Si on dispose d'un nombre de capteurs supérieur ou égal au nombre de sources ($M \geq N$), et que l'on suppose que la matrice $\mathbf{A}$ est de rang plein (ceci est vérifié si les mélanges sont linéairement indépendants - mélanges *différents*), on montre [23] que l'indépendance statistique du vecteur $\mathbf{y}$ permet d'estimer une matrice de séparation $\mathbf{B}$ telle que $\mathbf{BA} = \mathbf{P\Delta}$, où $\mathbf{P}$ et $\mathbf{\Delta}$ sont respectivement une matrice de permutation et une matrice diagonale, pourvu qu'une source au plus soit gaussienne. Le résultat met en évidence les indéterminations du problème : on ne peut retrouver exactement les sources qu'à un facteur d'échelle et une permutation près. Ceci s'explique facilement car ces indéterminations laissent les observations invariantes :

$$x_i(t) = \sum_j \left(\frac{a_{i\sigma(j)}}{\alpha_{\sigma(j)}}\right)(\alpha_{\sigma(j)} s_{\sigma(j)}(t)),\ i = 1, \ldots, M \quad (8)$$

où $\sigma(j)$ est une permutation sur $\{1, \ldots, N\}$.
En pratique, puisque N paramètres de gain sont indéterminés, il est possible de les fixer pour réduire les indéterminations : forcer N coefficients de la matrice $\mathbf{B}$ à 1, rechercher des sources estimées de puissance unité, etc.

2.2.2 Mélange linéaire convolutif

Si la transmission dans le canal fait intervenir des phénomènes de propagation, que l'on peut modéliser par des filtres linéaires, les mélanges s'écrivent

$$x_i(t) = \sum_j (a_{ij} * s_j)(t),\ i = 1, \ldots, M \tag{9}$$

soit $\mathbf{x}(t) = (\mathbf{A} * \mathbf{s})(t)$ où $\mathbf{A}(t) = (a_{ij}(t))$ est une matrice de filtres.
En adoptant une représentation à temps discret des signaux, et en passant à la transformée en z, les mélanges s'écrivent $\mathbf{x}(z) = \mathbf{A}(z)\mathbf{s}(z)$. Plusieurs auteurs [78, 71, 68] ont montré que des fonctions de contraste (découlant de l'indépendance statistique) conduisent à l'estimation d'une matrice $\mathbf{B}(z)$ telle que $\mathbf{B}(z)\mathbf{A}(z) = \mathbf{P}\mathbf{\Delta}(z)$. Dans ce cas, il existe une indétermination, plus sévère que dans le cas linéaire instantanée : chaque source ne peut être retrouvée qu'à un filtre près. Comme précédemment, il est facile de vérifier que ces indéterminations laissent les observations $x_i(t)$ invariantes.

La modélisation de mélanges convolutifs réalistes (par exemple, mélanges de sources audiophoniques dans une salle avec réverbération) nécessite des filtres avec un très grand nombre de paramètres, dont l'estimation sera fort coûteuse (temps de calcul, nombre d'échantillons). Il est possible de travailler sur un modèle fréquentiel [15, 26], déduit de (9) par transformée de Fourier discrète. On se ramène alors, à chaque fréquence discrète, à un mélange instantané à coefficients complexes, dont on peut facilement estimer un inverse aux indéterminations près. Ici, le problème de recontruction est cependant délicat : entre chaque paire de fréquences voisines, il faut compenser les permutations et égaliser les gains afin de reconstruire correctement les sources.

2.2.3 Mélange post-non-linéaire

Un mélange post-non-linéaire (PNL) est un mélange non linéaire particulier $\mathcal{F}$, constitué d'une partie linéaire (matrice de mélange $\mathbf{A}$ régulière, *i.e.*$M = N$) suivie de distorsions non linéaires f_i, supposées inversibles, agissant indépendamment sur chaque voie :

$$x_i(t) = f_i\Big(\sum_{j=1}^{N} a_{ij} s_j(t)\Big),\ i = 1, \ldots, N. \tag{10}$$

La figure 2 montre une représentation schématique de ce modèle, et sa structure de séparation, $\mathcal{G}$, adaptée au mélange. On peut montrer [70, 8] que ce mélange non linéaire particulier est séparable, c'est-à-dire que l'indépendance statistique permet d'identifier un inverse du mélange $\mathcal{F}$ avec les mêmes indéterminations qu'un mélanges linéaires, pourvu que la matrice $\mathbf{A}$ possède au moins deux éléments non nuls par ligne et par colonne, et qu'une source au plus soit gaussienne. Outre son intérêt théorique, ce modèle présente un réalisme certain, et il peut être utilisé pour modéliser des réseaux de capteurs [60], des liaisons satellite [65] et des systèmes biologiques [48].

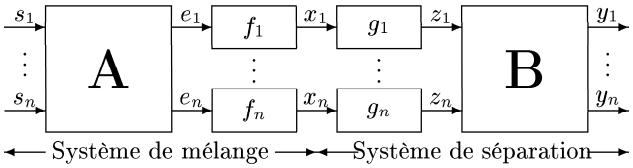


FIG. 2: *Mélange PNL et sa structure de séparation.*

2.2.4 Les mélanges multiplicatifs

Un exemple simple de mélange non linéaire séparable est le mélange multiplicatif de la forme :

$$x_i(t) = \prod_{j=1}^{N} s_j^{\alpha_j}(t),\ i = 1, \ldots, M \tag{11}$$

où $s_j(t)$ sont des sources indépendantes à valeurs positives. En prenant le logarithme, on obtient :

$$\ln x_i(t) = \sum_{j=1}^{N} \alpha_j \ln s_j(t),\ i = 1, \ldots, M \tag{12}$$

qui est un mélange linéaire de nouvelles variables aléatoires indépendantes (puisque la fonction ln est monotone) $\ln s_j(t)$. Ce type de mélanges modélise par exemple la dépendance entre la température et le champ magnétique de la tension Hall mesurée sur un capteur Hall à silicium :

$$V_H = kBT^{\alpha} \tag{13}$$

où α dépend du type de semiconducteur, car l'influence de la température est liée à la mobilité des porteurs majoritaires. Ainsi, en utilisant deux types de capteurs (N et P) on obtient :

$$V_{H_N}(t) = k_N B(t) T^{\alpha_N}(t) \tag{14}$$
$$V_{H_P}(t) = k_P B(t) T^{\alpha_P}(t). \tag{15}$$

Dans la suite, pour simplifier, on supprimera la variable t. Dans ce modèle, la température T est positive, mais le signe du champ magnétique B peut varier, on prendra donc le logarithme de la valeur absolue :

$$\ln \mid V_{H_N} \mid = \ln k_N + \ln \mid B \mid + \alpha_N \ln T \tag{16}$$
$$\ln \mid V_{H_P} \mid = \ln k_P + \ln \mid B \mid + \alpha_N \ln T. \tag{17}$$

En fait, ce modèle est trop simple, et peut être résolu facilement par simple décorrélation, car le champ B intervient avec le même exposant dans les deux équations. Le rapport des deux équations conduit alors à :

$$R = \frac{V_{H_N}}{V_{H_P}} = \frac{k_N}{k_P} T^{\alpha_N - \alpha_P} \tag{18}$$

qui ne dépend que de la température. Ensuite, R ne dépendant que de la température, pour estimer $\mid B(t) \mid$, il est suffisant de calculer le paramètre k tel que $V_{H_N} R^k$ soit décorrélé avec R. Le signe de B est très simple à estimer, puiqu'à chaque instant t, le signe de $k_N B(t)$ est égal au signe de $V_{H_N}(t)$.

2.2.5 Un ensemble de mélanges non linéaires séparables

L'extension du théorème de Darmois-Skitovic à des mélanges non linéaires a été étudiée dans le début des années 70's par Kagan *et al.* [45]. Ces résultats ont été récemment redécouvert dans le cadre de la séparation de sources par Eriksson and Koivunen [32]. L'idée de base est de considérer des mélanges particuliers $\mathcal{F}$ qui satisfont un *théorème d'addition* au sens de la théorie des équations fonctionnelles. Pour comprendre simplement cette idée, considérons les variables aléatoires :

$$x_1 = \frac{s_1 + s_2}{1 + s_1 s_2}$$
$$x_2 = \frac{s_1 - s_2}{1 - s_1 s_2}$$

où s_1 et s_2 sont deux variables aléatoires indépendantes. On remarque que, en posant $u_1 = \tan^{-1}(s_1)$ et $u_2 = \tan^{-1}(s_2)$, on arrive à :

$$x_1 = \tan(u_1 + u_2)$$
$$x_2 = \tan(u_1 - u_2).$$

En appliquant le même changement de variables à x_1 et x_2, on a :

$$v_1 = \tan^{-1}(x_1) = u_1 + u_2$$
$$v_2 = \tan^{-1}(x_2) = u_1 - u_2$$

qui est maintenant un mélange linéaire de deux variables indépendantes. Comme l'explique Kagan *et al.*, ce joli résultat est dû au fait que $\tan(a+b)$ (et $\tan(a-b)$) sont des fonctions de $\tan a$ et de $\tan b$:

$$\tan(a+b) = \frac{\tan a + \tan b}{1 + \tan a \tan b}. \tag{19}$$

De façon générale, cette propriété est valable si $\mathcal{F}$ satisfait un théorème d'addition :

$$f(a+b) = \mathcal{F}[f(a), f(b)] \tag{20}$$

Les propriétés requises pour $\mathcal{F}$ (dans le cas de deux variables, mais l'extension est évidente) sont les suivantes :

- $\mathcal{F}$ est continue par rapport aux 2 variables et possède des valeurs dans l'intervalle I;
- $\mathcal{F}$ est commutative, *i.e.* $\mathcal{F}(u,v) = \mathcal{F}(v,u)$;
- $\mathcal{F}$ est associative, *i.e.* $\mathcal{F}(\mathcal{F}(u,v),w) = \mathcal{F}(u,\mathcal{F}(v,w))$;
- Il existe un élément neutre e tel que $\forall u$, $\mathcal{F}(u,e) = \mathcal{F}(e,u) = u$;
- $\forall u$, Il existe un inverse u^{-1} tel que $\mathcal{F}(u,u^{-1}) = \mathcal{F}(u^{-1},u) = e$.

Autrement dit, en notant $u \circ v = \mathcal{F}(u,v)$, il faut que l'ensemble $(u, \circ)$ soit un groupe Abélien. Sous cette condition, il existe une fonction monotone et continue [4] $f : \mathbb{R} \to I$ telle que :

$$f(a+b) = f(a) \circ f(b). \tag{21}$$

Soit, en appliquant f^{-1}, on arrive à un mélange linéaire :

$$a + b = f^{-1}((f(a) \circ f(b))). \tag{22}$$

Pour avoir plus de détails et d'autres exemples de transformations qui satisfont ce théorème, on peut se reporter à [44, 32]. Ces résultats, quoique très intéressants d'un point de vue théorique, sont d'une portée pratique limitée, car ils supposent que les mélanges sont des fonctions non linéaires très particulières, et connues.

2.3 Critères d'indépendance

Après avoir discuté le problème de l'identifiabilité du mélange à l'aide du critère d'indépendance, nous devons examiner comment mettre en oeuvre l'indépendance statistique. Quoique le problème puisse être formulé de diverses manières : au sens du maximum de vraisemblance [63, 16], ou à l'aide de fonctions de contraste (voir au paragraphe 2.3.5), ou avec un critère quadratique [59, 3], nous nous concentrerons sur la divergence de Kullback-Leibler, qui donne un éclairage général à de nombreux algorithmes.

2.3.1 Divergence de Kullback-Leibler

Le vecteur aléatoire $\mathbf{y} = (y_j)$ a des composantes indépendantes y_j si pour tout $\mathbf{u}$:

$$p_{\mathbf{y}}(\mathbf{u}) = \prod_j p_{y_j}(u_j). \tag{23}$$

Cette relation, entre fonctions multivariables (qui plus est inconnues), est difficile à manipuler. Une mesure scalaire plus pratique de l'indépendance est la divergence (ce n'est pas une distance, parce qu'elle n'est pas commutative) de Kullback-Leibler qui mesure l'écart entre deux distributions p et q de la même variable $\mathbf{y}$:

$$KL(p \parallel q) = \int p(\mathbf{u}) \log \frac{p(\mathbf{u})}{q(\mathbf{u})} d\mathbf{u}. \tag{24}$$

On montre facilement que $KL(p \parallel q)$ est une quantité positive qui s'annule si et seulement si $p = q$. L'indépendance peut donc être mesurée par la divergence KL entre $p_{\mathbf{y}}$ et $\prod_j p_{y_j}$:

$$KL(p_{\mathbf{y}} \parallel \prod_j p_{y_j}) = \int p_{\mathbf{y}}(\mathbf{u}) \log \frac{p_{\mathbf{y}}(\mathbf{u})}{\prod_j p_{y_j}(u_j)} d\mathbf{u} \tag{25}$$

qui s'annule si et seulement si $p_{\mathbf{y}} = \prod_j p_{y_j}$. En théorie de l'information, cette relation coïncide avec l'information mutuelle (IM) $I(\mathbf{y})$ [25]. En notant $H(\mathbf{y})$ et $H(y_j)$ les entropies de Shannon jointes et marginales, respectivement :

$$H(\mathbf{y}) = -\int p_{\mathbf{y}}(\mathbf{u}) \log p_{\mathbf{y}}(\mathbf{u}) d\mathbf{u}$$
$$H(y_j) = -\int p_{y_j}(u_j) \log p_{y_j}(u_j) du_j$$

l'information mutuelle s'écrit :

$$I(\mathbf{y}) = KL(p_{\mathbf{y}} \parallel \prod_j p_{y_j}) = \sum_j H(y_j) - H(\mathbf{y}). \tag{26}$$

2.3.2 Equations d'estimation

Les propriétés de l'IM suggèrent une estimation de la structure de séparation par minimisation de l'IM. Quoique la relation (26) puisse être directement utilisée pour obtenir les équations d'estimation, on peut la simplifier en prenant en compte la structure de séparation.

Optimisation de l'IM pour des mélanges linéaires. Dans le cas de mélanges linéaires carrés, $\mathcal{F}$ est une matrice $\mathcal{F} = \mathbf{A}$ de taille $N \times N$, supposée inversible, et la transformation de séparation est aussi une matrice $\mathbf{B}$, de

taille $N \times N$. Puisque $\mathbf{y} = \mathbf{B}\mathbf{x}$ et $\mathbf{x} = \mathbf{A}\mathbf{s}$, la relation entre les ddp de $\mathbf{x}$ et $\mathbf{y}$ s'écrit $p_{\mathbf{y}}(\mathbf{u}) = p_{\mathbf{x}}(\mathbf{B}^{-1}\mathbf{u})/\mid \det \mathbf{B} \mid$ et l'IM devient :

$$I(\mathbf{y}) = \sum_j H(y_j) - H(\mathbf{x}) - \log \mid \det \mathbf{B} \mid . \qquad (27)$$

La minimisation de l'IM par rapport à la matrice de séparation $\mathbf{B}$ requiert la gradient de l'IM, qui s'écrit puisque $y_j = \sum_i b_{ji} x_i$,

$$\frac{\partial I(\mathbf{y})}{\partial \mathbf{B}} = E\Psi_{\mathbf{y}}\mathbf{x}^T - \mathbf{B}^{-T} \qquad (28)$$

où E est l'espérance mathématique et $\Psi_{\mathbf{y}} = [\Psi_{y_1} \dots \Psi_{y_N}]^T$ est le vecteur dont chaque composante Ψ_{y_j} est la fonction score de y_j définie par :

$$\Psi_{y_j}(y_j) = -\frac{\partial \log p_{y_j}(y_j)}{\partial y_j} = -\frac{p'_{y_j}}{p_{y_j}}(y_j). \qquad (29)$$

Après multiplication à droite par $\mathbf{B}^T$, on obtient l'équation d'estimation :

$$E\Psi_{\mathbf{y}}\mathbf{y}^T - \mathbf{I} = 0 \qquad (30)$$

où $\mathbf{I}$ est la matrice identité de taille $N \times N$.

L'expression (30) met en évidence l'importance des fonctions score qui dépendent des ddp et qui contiennent toute la connaissance statistique sur les sources estimées y_j. Hélas, dans le cadre aveugle, ces fonctions Ψ_{y_j} ainsi que les ddp p_{y_j} sont inconnues, et il faudra estimer ces fonctions pour concevoir des algorithmes efficaces. En effet, les fonctions score exactes fournissent les statistiques optimales dans l'équation (30). Si y_j est une variable gaussienne centrée réduite, sa fonction score est $\Psi_{y_j}(y_j) = y_j$, ce qui conduit aux équations d'estimation $Ey_iy_j = \delta_{ij}$, qui sont des équations de décorrélation (pour $i \neq j$). Au contraire, si y_j n'est pas gaussienne, sa fonction score est une fonction non linéaire et les équations d'estimation $E\psi_{y_i}(y_i)y_j = \delta_{ij}$ correspondent à l'annulation de moments statistiques d'ordre supérieur à 2.

Optimisation de l'IM pour des mélanges non linéaires. Dans le cas de mélanges non linéaires, on suppose que $\mathcal{F}$ et $\mathcal{G}$ sont des transformations non linéaires inversibles de dimension N. Puisque $\mathbf{y} = \mathcal{G}(\mathbf{x})$, la relation entre les ddp de $\mathbf{x}$ et $\mathbf{y}$ s'écrit :

$$p_{\mathbf{y}}(\mathbf{u}) = p_{\mathbf{x}}(\mathcal{G}^{-1}(\mathbf{u}))/ \mid \det \mathbf{J}_{\mathcal{G}}(\mathcal{G}^{-1}(\mathbf{u})) \mid \qquad (31)$$

et l'IM devient :

$$I(\mathbf{y}) = \sum_j H(y_j) - H(\mathbf{x}) - \log \mid \det \mathbf{J}_{\mathcal{G}}(\mathcal{G}^{-1}(\mathbf{u})) \mid . \qquad (32)$$

Dans le cas des mélanges PNL, on a $y_j = \sum_i b_{ji} g_i(x_i)$ et l'IM se simplifie :

$$I(\mathbf{y}) = \sum_j H(y_j) - H(\mathbf{x}) - \log \mid \Pi_i g'_i(x_i) \mid - \log \mid \det \mathbf{B} \mid . \qquad (33)$$

Ainsi, la minimisation de l'IM conduit à 2 jeux d'équations d'estimation, l'un pour la partie linéaire (similaire à (30)), l'autre pour la partie non linéaire de la structure de séparation, qui mettent encore en évidence le rôle important joué par les fonctions score [70].

Optimisation directe de l'IM. On peut optimiser directement $I(\mathbf{y}) = \sum_j H(y_j) - H(\mathbf{y})$ par rapport à $\mathbf{B}$:

$$\frac{\partial I(\mathbf{y})}{\partial \mathbf{B}} = E\Psi_{\mathbf{y}}\mathbf{x}^T - E\Phi_{\mathbf{y}}\mathbf{x}^T = E\boldsymbol{\beta}_{\mathbf{y}}\mathbf{x}^T \qquad (34)$$

où

- $\Psi_{\mathbf{y}} = [\Psi_{y_1} \dots \Psi_{y_N}]^T$ est le vecteur dont chaque composante Ψ_{y_j} est la fonction score (marginale) de y_j définie précédemment (29)
- $\Phi_{\mathbf{y}}(\mathbf{y}) = [\Phi_1(\mathbf{y}) \dots \Phi_N(\mathbf{y})]^T$ est le vecteur dont chaque composante $\Phi_j(\mathbf{y})$ est la fonction score jointe de $\mathbf{y}$ définie par :
$$\Phi_j(\mathbf{y}) = -\frac{\partial \log p_{\mathbf{y}}(\mathbf{y})}{\partial y_j} \qquad (35)$$
- $\boldsymbol{\beta}_{\mathbf{y}} \triangleq \Psi_{\mathbf{y}} - \Phi_{\mathbf{y}}$ est la différence entre les vecteurs des fonctions score marginales et jointes de $\mathbf{y}$, appelée DFS (différence de fonctions score).

Après multiplication à droite par $\mathbf{B}^T$, on obtient les équations d'estimation :

$$E(\Psi_{\mathbf{y}} - \Phi_{\mathbf{y}})\mathbf{y}^T = \quad E\boldsymbol{\beta}_{\mathbf{y}}\mathbf{y}^T = \mathbf{0} \qquad (36)$$

Optimisation de l'IM pour des mélanges convolutifs. Dans ce cas, $\mathcal{F}$ est une matrice inversible de filtres de taille $N \times N$. Par exemple, si $\mathbf{A}(z)$ contient des filtres d'ordre maximal égal à L, le mélange s'écrit :

$$\mathbf{x}(t) = (\mathbf{A} * \mathbf{s})(t) = \sum_{l=0}^{L} \mathbf{A}(l)\mathbf{s}(t-l). \qquad (37)$$

La structure de séparation est alors restreinte à une matrice de filtres $\mathbf{B}$ de taille $N \times N$:

$$\mathbf{y}(t) = (\mathbf{B} * \mathbf{x})(t) = \sum_{l=0}^{L'} \mathbf{B}(l)\mathbf{x}(t-l). \qquad (38)$$

Malheureusement, à cause des filtres, il n'y a plus de relation simple entre les ddp de $\mathbf{x}$ et $\mathbf{y}$, et il faut utiliser directement l'équation $I(\mathbf{y}) = \sum_j H(y_j) - H(\mathbf{y})$. En fait, on peut montrer [9] que, au premier ordre, on a :

$$I(\mathbf{y} + \boldsymbol{\Delta}) - I(\mathbf{y}) = E\boldsymbol{\Delta}^T\boldsymbol{\beta}_{\mathbf{y}} \qquad (39)$$

où $\boldsymbol{\Delta}$ est un 'petit' vecteur aléatoire. Cette équation montre que la DFS est le gradient stochastique de l'IM.

De plus, l'indépendance dans les mélanges convolutifs signifie indépendance de processus stochastiques, c'est-à-dire qu'il faut considérer l'indépendance entre les signaux avec tous les retards possibles. Pour des raisons de simplicité, considérons seulement les termes d'ordre l, et dérivons par rapport à $\mathbf{B}(l)$. Pour cela, posons $\hat{\mathbf{B}}(l) = \mathbf{B}(l) + \varepsilon$, où ε représente une 'petite' matrice. On a alors :

$$\hat{\mathbf{y}}(t) \triangleq (\mathbf{B} * \mathbf{x})(t) = \mathbf{y}(t) + \varepsilon\mathbf{x}(t-l). \qquad (40)$$

En combinant l'équation ci-dessus avec (39), et après quelques calculs, on trouve :

$$\frac{\partial I(\mathbf{y}(t))}{\partial \mathbf{B}(l)} = E\boldsymbol{\beta}_{\mathbf{y}}(\mathbf{y}(t))\,\mathbf{x}^T(t-l) \qquad (41)$$

Il faut encore calculer les gradients de l'IM pour des versions retardées des sorties $I(y_1(t), y_2(t-l))$ (dans le cas de 2 mélanges de 2 sources). On trouve des relations similaires à (41), détaillées dans [7]. Ce calcul peut aussi être étendu au cas de mélanges convolutifs avec des post non-linéarités [6].

2.3.3 Critères dans des structures contraintes

Des contraintes dans la structure de séparation peuvent réduire les indéterminations et même assurer la séparabilité. Elles peuvent aussi modifier voire simplifier le critère d'indépendance.

Séparation avec pré-blanchiment. De nombreuses méthodes [21, 57] utilisent une décomposition de la matrice de séparation $\mathbf{B}$:

$$\mathbf{B} = \mathbf{UW} \tag{42}$$

où $\mathbf{W}$ est une matrice de blanchiment (spatial) et $\mathbf{U}$ est une matrice orthogonale. Notons $\mathbf{z} = \mathbf{Wx}$ le vecteur aléatoire blanchi (et réduit) tel que $E\mathbf{zz}^T = I$. Puisque $\mathbf{U}$ est orthogonale, $\mathbf{y} = \mathbf{Uz}$ satisfait aussi $E\mathbf{yy}^T = I$ (ce qui fixe simplement l'indétermination d'échelle). Alors :

$$I(\mathbf{y}) = \sum_j H(y_j) - H(\mathbf{z}) - \log \mid \det \mathbf{U} \mid . \tag{43}$$

Après calcul de la matrice $\mathbf{W}$, l'entropie jointe $H(\mathbf{z})$ est constante. De plus, le déterminant de la matrice orthogonale $\mathbf{U}$ est égal à 1, l'IM se réduit donc à :

$$I(\mathbf{y}) = \sum_j H(y_j) + cst. \tag{44}$$

Minimiser l'IM est donc équivalent à minimiser la somme des entropies marginales (des variables normalisées). Or, il est bien connu que l'entropie est maximale (à variance unité) pour la distribution gaussienne. Autrement dit, sous contrainte de blanchiment, minimiser l'IM est équivalent à minimiser la gaussianité des sources estimées.

Infomax. Une autre approche, initialement proposée par Bell et Sejnowski [12], consiste à transformer chaque source estimée y_j en une source z_j à distribution uniforme dans l'intervalle $[0, 1]$. Il suffit d'écrire :

$$z_j = F_{y_j}(y_j),\ i = 1, \ldots, N \tag{45}$$

où F_{y_j} est la fonction de répartition de la variable aléatoire y_j. Puisque l'IM est invariante par une transformation diagonale inversible, on peut écrire :

$$I(\mathbf{z}) = I(\mathbf{y}) = \sum_j H(z_j) - H(\mathbf{z}). \tag{46}$$

Chaque variable z_j étant uniformément distribuée dans $[0, 1]$, l'IM devient :

$$I(\mathbf{z}) = I(\mathbf{y}) = cst - H(\mathbf{z}). \tag{47}$$

En conséquence, minimiser l'IM de la variable uniforme $\mathbf{z}$ est équivalent à maximiser son entropie jointe : l'algorithme associé à cette idée est appelé Infomax.

2.3.4 Estimation des fonctions score

Comme nous l'avons expliqué après le calcul du gradient de l'IM (2.3.2), l'information clé concernant les sources estimées est la fonction score. Si les fonctions score sont choisies *a priori*, les statistiques impliquées dans les équations ne sont pas optimales, et l'algorithme peut même diverger.

Cependant, pour estimer la matrice de séparation (dans le cas de mélanges linéaires), on peut noter qu'une estimation grossière des fonctions score est suffisante. En effet, la relation (30) conduit à l'ensemble d'équations :

$$E\psi_{y_i}(y_i)y_j = 0,\ i \neq j \tag{48}$$

car les termes diagonaux de (30) sont de simples équations de normalisation. Si les y_i sont des variables aléatoires centrées statistiquement independants (*i.e.* $Ey_j = 0$, $i = 1, \ldots, N$), il est évident que :

$$E\psi_{y_i}(y_i)y_j = E\psi_{y_i}(y_i)Ey_j = 0,\ i \neq j \tag{49}$$

même pour une estimation grossière de ψ_{y_i}. Quoiqu'une estimation précise améliore la vitesse de convergence, une estimation des fonctions score, même grossière, est préférable à un choix *a priori*, évitant notamment les risques de divergence de l'algorithme pour certaines distributions. Au contraire, dans le cas non linéaire, on montre [70] qu'une estimation précise des fonctions score est requise. Une comparaison fondée sur les estimations de 4-ème ordre obtenues avec un développement de Gram-Charlier (des ddp) et avec un critère des moindres carrés [70] met en évidence l'estimation médiocre du développement de Gram-Charlier, et explique les difficultés pour séparer les mélanges non linéaires difficiles [77].

Estimation indirecte. Les fonctions score s'exprimant en fonction de la ddp, une approche simple consiste à estimer la ddp, et à calculer l'estimation :

$$\hat{\psi}_{y_j}(y_j) = -\frac{\hat{p}'_{y_j}}{\hat{p}_{y_j}}(y_j). \tag{50}$$

Il existe de nombreuses méthodes d'estimation des densités en statistiques. Une première approche est fondée sur les développements de Gram-Charlier ou d'Edgeworth de la ddp p_{y_j} autour de la gaussienne [46]. Sans rentrer dans les détails, on peut dire que ces développement mettent en jeu explicitement des cumulants d'ordre supérieur (à 2). Les méthodes des noyaux, également très classiques [42], conduisent à :

$$\hat{p}_Y(u) = \frac{1}{T}\sum_{t=1}^{T} K_h(u - Y_t) \tag{51}$$

où les Y_t sont les échantillons de la variable aléatoire Y, et h est la largeur du noyau. Plus h est grand, plus l'estimation est lisse. Bien sûr, h dépend du nombre d'échantillons T. Expérimentalement, il semble qu'un choix optimal de h, fondé sur un validation croisée coûteuse, n'est pas nécessaire, même dans le cas non linéaire [70].

Estimation directe. Les fonctions score peuvent être estimées directement par une approche des moindres carrés. En effet, en notant $\hat{\psi}(\mathbf{w}, u)$ un modèle paramétrique de $\psi(u)$, et en utilisant la définition de la fonction score, le critère des moindres carrés :

$$J(\mathbf{w}) = \frac{1}{2}E(\hat{\psi}(\mathbf{w}, u) - \psi(u))^2 \tag{52}$$

conduit au gradient :

$$\frac{\partial J(\mathbf{w})}{\partial \mathbf{w}} = E[\hat{\psi}(\mathbf{w}, u)\frac{\partial \hat{\psi}}{\partial \mathbf{w}}(\mathbf{w}, u) + \frac{\partial^2 \hat{\psi}}{\partial u \partial \mathbf{w}}(\mathbf{w}, u)]. \tag{53}$$

Ce gradient peut être utilisé pour estimer un modèle paramétrique quelconque, par exemple une combinaison linéaire de fonctions non linéaires [64, 18], un réseau de neurones [70], où des modèles à base de splines ou de polynômes.

2.3.5 Fonctions de contraste

Les fonctions de contraste, initialement définies par Donoho [29] pour la déconvolution aveugle, ont été introduites par Comon [22, 23] dans le cadre de la séparation de sources.
Une fonction de contraste est une fonction réelle d'une distribution $\mathbf{y}$, qui doit être minimale si $\mathbf{y}$ est à composantes independantes. Pour des mélanges linéaires, la définition est la suivante :

Définition 1 *Une fonction $\phi[\mathbf{s}]$ de la distribution $\mathbf{s}$ est une fonction de contraste si, pour toute matrice $\mathbf{C}$ et tout vecteur aléatoire $\mathbf{s}$, $\phi[\mathbf{Cs}] \geq \phi[\mathbf{s}]$, avec égalité si et seulement si $\mathbf{C} = \mathbf{P\Delta}$, où $\mathbf{P}$ et $\mathbf{\Delta}$ sont des matrices de permutation et diagonale, respectivement.*

Cette définition a été étendue aux mélanges convolutifs [24] et non linéaires, en remplaçant $\mathbf{P\Delta}$ par des filtres triviaux $\mathbf{P\Delta}(z)$ où des transformations triviales $h_i(s_{\sigma(i)})$.
De façon évidente, les fonctions de contraste sont de bons critères pour concevoir des algorithmes de séparation de sources.
Deux questions importantes se posent sur les fonctions de contrastes : d'une part, chercher des classes générales de fonctions qui ont les propriétés de contraste [61], d'autre part trouver des fonctions de contraste aussi simples que possible, et ne présentant pas de minima locaux. A titre d'exemple, Comon [23] a montré que l'IM est une fonction de contraste. Plusieurs contrastes plus simples, à base de cumulants d'ordre 4, conduisent à des algorithmes simples et performants. Enfin, il existe des liens entre l'IM et ces contrastes : par exemple, l'approximation de l'IM par un développement de Gram-Charlier d'ordre 4 conduit à un contraste à base de cumulants d'ordre 4 [16].

2.4 Méthodes semi-aveugles

Dans les paragraphes précédents, nous avons considéré l'ACI comme une méthode générale pour résoudre le problème de séparation de sources à partir d'observations multiples (diversité spatiale) sans information sur les sources, à part leur indépendance statistique. Dans ce paragraphe, en restant dans le cas de mélanges linéaires, nous montrons que (i) des informations *a priori* même très faibles sur les sources peuvent conduire à des algorithmes plus simples et très performants, (ii) d'autres formes de diversité sont exploitables.

2.4.1 Sources non iid

Jusqu'à présent, les critères utilisés supposent implicitement les sources (temporellement) indépendantes et identiquement distribuées (iid) : cette hypothèse est utilisée pour écrire le maximum de vraisemblance (MV) [63]; de plus, on peut montrer les liens asymptotiques, entre MV et IM [16]. En fait, les méthodes sont robustes à l'hypothèse iid, mais n'exploitent aucune information temporelle sur les sources. Le coût à payer est que (i) les méthodes requièrent des statistiques d'ordre supérieur à deux, (ii) la séparation est impossible s'il y a plus d'une source gaussienne. L'utilisation d'informations temporelles, même très faibles, conduit à des méthodes de séparation à l'ordre 2, qui sont plus simples, et qui peuvent séparer des sources gaussiennes. Comme l'explique Cardoso [17], on peut relâcher l'hypothèse iid de deux manières.

Sources colorées. Relâcher le premier i revient à supposer les sources colorées. Dans ce cas, les matrices de variances-covariances $\Gamma_\tau = E\mathbf{y}(t)\mathbf{y}(t-\tau)^T$, pour différentes valeurs de τ sont différentes, pourvu que les sources aient des spectres différents. On peut alors séparer les sources à l'ordre 2, car la matrice qui diagonalise conjointement les Γ_τ est une matrice séparante. Cette idée conduit à des algorithmes très simples et très efficaces [72, 58, 13] puisqu'ils reposent sur la résolution d'un problème d'algèbre linéaire, et sont capables de séparer des sources gaussiennes.

Sources non stationnaires. Relâcher la nature *id* revient à supposer que les échantillons successifs sont distribués différemment. C'est le cas de sources non stationnaires. Dans ce cas, les matrices de variances-covariances $\Gamma_i = E_{[t_i,t_{i+1}[}\mathbf{y}(t)\mathbf{y}(t-\tau)^T$, calculées sur les fenêtres temporelles $[t_i, t_{i+1}[$, sont différentes, pourvu que les rapports des variances des sources varient différemment [56, 62]. Comme précédemment, cette idée aboutit à des méthodes très simples reposant sur des algorithmes de diagonalisation conjointe. Même algorithme donc, mais appliqué à des matrices de variances-covariances calculées différemment.

2.4.2 Autres formes de diversités

Le problème de séparation de sources peut être considéré d'un point de vue algébrique : il faut disposer d'au moins autant d'équations que de paramètres à déterminer. C'est le rôle des matrices de variances-covariances calculées pour différents retards ou sur différentes fenêtres. C'est aussi le rôle de la diversité spatiale : chaque capteur amène une nouvelle observation. D'autres formes de diversité peuvent être exploitées. En imagerie couleur ou multispectrale [51, 53], la diversité fréquentielle permet d'écrire autant d'équations différentes qu'il y a de bandes de fréquence. Pour séparer une image constituée d'une scène à laquelle est superposé un reflet dû à une vitre, Devlamink et Terrier [28] utilisent plusieurs images avec différents angles (diversité) de polarisation.

3 Représentations parcimonieuses

La parcimonie est une autre exemple d'information *a priori* très efficace pour déterminer la solution d'un problème de séparation de sources. Cette parcimonie peut être temporelle (signal non stationnaire qui s'éteint par

moment, ou signal discret [66]), fréquentielle (le signal n'occupe qu'une partie du spectre), ou combiner les deux [2]. Ces propriétés permettent notamment de résoudre les problèmes de SAS sous-déterminés (plus de sources que de capteurs, *i.e.* $M < N$), pour lesquels la séparation aveugle est impossible. En effet, s'il est possible dans le cas sous-déterminé d'identifier la matrice de mélange $\mathbf{A}$ [69], cela n'est pas suffisant pour retrouver les sources : l'ensemble des signaux sources satisfaisant

$$\mathbf{x} = \mathbf{As}$$

est en effet un sous-espace affine non réduit à un point. Une fois la matrice identifiée, il reste donc à choisir les "bonnes" sources parmi celles possibles. Par ailleurs, même si le mélange est déterminé ($M \geq N$), certaines séries temporelles $s_n(t)$ sont parfois très loin d'être i.i.d : c'est le cas des sources audio qui présentent des oscillations et des phénomènes entretenus caractéristiques.

Les approches de la SAS fondées sur les représentations parcimonieuses de signaux [74, 43, 79, 47, 14, 37, 38] permettent de traiter le cas sous-déterminé, et peuvent également constituer une alternative ou un complément intéressants des méthodes d'ACI lorsque le modèle de sources i.i.d. est peu réaliste. Ces approches partent du principe que, dans une représentation appropriée des signaux, les composantes des sources sont i.i.d et parcimonieuses, si bien que chaque composante des signaux de capteurs est due essentiellement à une seule source "active". Tout le problème devient alors d'affecter chaque composante observée à la bonne source. Ces méthodes mettent donc en jeu deux ingrédients : le premier consiste à choisir une représentation judicieuse des signaux (par exemple, temps-fréquence) [34, 54]; le second à classifier les coefficients de cette représentation pour effectuer la bonne affectation.

3.1 Changement de représentation

Les signaux (de capteurs ou de sources), qui peuvent être vus comme des vecteurs $y \in \mathbb{R}^T$, sont fournis sous la forme "brute" de leur succession d'échantillons $y(t), 1 \leq t \leq T$. On peut cependant les représenter sous d'autres formes, qu'il s'agisse de leur transformée de Fourier, de leur spectre à court terme ou de leur transformée en ondelettes, pour ne citer que quelques exemples [54]. Plus généralement, si l'on introduit la notion de *dictionnaire* $\mathcal{D} = \{g_k \in \mathbb{R}^T, 1 \leq k \leq K\}$ on peut considérer les représentations des signaux sous la forme

$$y = \sum_{k=1}^{K} a_k g_k. \tag{54}$$

Lorsque le dictionnaire est une base orthonormale de $\mathbb{R}^T$ (par exemple une base d'ondelettes orthogonales ou bien une base de cosinus locaux), les coefficients a_k sont donnés de façon unique et simple par les produits scalaires

$$a_k = \langle y, g_k \rangle = \sum_{t=1}^{T} y(t) g_k(t). \tag{55}$$

Par contre, lorsque le dictionnaire est une famille génératrice mais redondante, il existe une infinité de choix possibles pour ces coefficients.

Un certain nombre de dictionnaires redondants usuels constituent des *tight frames*, et l'on peut encore utiliser les produits scalaires comme coefficients :

$$y = \sum_{k=1}^{K} \langle y, g_k \rangle g_k. \tag{56}$$

Une telle représentation est dite linéaire.

Il existe cependant d'autres choix possibles de coefficients, qui sont *non-linéaires* et peuvent s'avérer plus judicieux. Ainsi, si l'on modélise les coefficients comme issus de variables aléatoires Laplaciennes indépendantes le jeux de coefficients le plus vraisemblable est

$$\{a_k^\star\}_{k=1}^K = \arg \min_{\{a_k\}_{k=1}^K} \sum_{k=1}^{K} |a_k| = \arg \min_{\{a_k\}_{k=1}^K} \|\{a_k\}_{k=1}^K\|_1. \tag{57}$$

Qu'elle soit linéaire ou non, le but du changement de représentation pour la SAS est faciliter la séparation de sources. On peut noter matriciellement $\mathbf{s} = \mathbf{SD}$ la représentation des sources inconnues

$$s_n(t) = \sum_{k=1}^{K} c_k^n g_k(t), \tag{58}$$

où $\mathbf{S}$ est une matrice $N \times K$ également inconnue des coefficients, et $\mathbf{D}$ est la matrice $K \times T$ du dictionnaire. C'est la structure plus "simple" de la matrice $\mathbf{S}$, comparée à la matrice de départ $\mathbf{s}$, qui va simplifier le problème.

3.2 Séparation dans le domaine transformé

Pour estimer les sources, il suffit manifestement d'en retrouver *une* représentation sous la forme d'*un* jeu de coefficients $\mathbf{S}$ tel que $\mathbf{s} = \mathbf{SD}$. Les méthodes de SAS fondées sur la parcimonie vont y parvenir en exploitant le fait que $\mathbf{S}$ contient peu de coefficients significativement non nuls.

En combinant la représentation $\mathbf{s} = \mathbf{SD}$ avec le modèle de mélange linéaire instantané $\mathbf{x} = \mathbf{As}$, on prédit que $\mathbf{x} = \mathbf{ASD}$, c'est-à-dire que la modélisation nous fournit *une* représentation $\mathbf{AS}$ des observation $\mathbf{x}$. Suivant la structure du dictionnaire $\mathbf{D}$, un certain nombre d'algorithmes sont disponibles pour calculer une (autre) représentation $\mathbf{x} = \mathbf{XD}$. Si l'on peut identifier $\mathbf{X} = \mathbf{AS}$, on obtient alors un nouveau problème de SAS dans le domaine tranformé.

Voyons comment l'hypothèse de parcimonie de $\mathbf{S}$ donne une structure particulière à ce nouveau problème, et comment elle facilite sa résolution. Nous verrons ensuite que sous certaines hypothèses (sur le dictionnaire, sur la parcimonie de $\mathbf{X}$ et de $\mathbf{S}$, sur l'algorithme utilisé), l'identification $\mathbf{X} = \mathbf{AS}$ est effectivement possible.

3.2.1 Structure de X

Par hypothèse, chaque ligne de $\mathbf{S}$ contient des coefficients $\{c_k^n\}_{k=1}^K$ qui sont parcimonieux, *au sens où* peu d'entre eux sont significativement différents de zéro. Autrement dit, la distribution des valeurs des coefficients exhibe un pic autour de zéro et des longues queues. Les sources étant supposées indépendantes, la probabilité de co-occurence de deux coefficients significatifs dans une même

colonne de $\mathbf{S} = [\mathbf{S}_1, \cdots, \mathbf{S}_K]$ est donc faible, et la plupart des colonnes contiennent au plus un coefficient non négligeable. Au final, en notant $n(k)$ la source qui induit un coefficient significatif sur la k-ème colonne, on a

$$\mathbf{S}_k \approx c_k^{n(k)}[0, \cdots, 1, \cdots, 0]^T.$$

Ecrivant l'égalité $\mathbf{X} = \mathbf{AS}$ colonne par colonne, on obtient

$$\mathbf{X}_k = \mathbf{AS}_k \approx c_k^{n(k)}\mathbf{A}_{n(k)}, \tag{59}$$

avec $\mathbf{X} = [\mathbf{X}_1, \ldots, \mathbf{X}_K]$ et $\mathbf{A} = [\mathbf{A}_1, \ldots, \mathbf{A}_N]$. C'est l'idée fondamentale des méthodes basées sur la parcimonie : grâce à la parcimonie de la représentation $\mathbf{S}$ des sources, les colonnes de la représentation calculée $\mathbf{X}$ sont (approximativement) proportionnelles à celles de la matrice de mélange.

3.2.2 Estimation des colonnes de A

Dans le cas d'un mélange déterminé carré ($M = N$), l'estimation de $\mathbf{A}$ est équivalente à celle de son inverse $\mathbf{B}$. On peut donc recourir aux méthodes décrites précédemment (minimisation de l'IM) en travaillant sur les données transformées $\mathbf{X}$. L'hypothèse de parcimonie de la représentation n'est alors pas cruciale, et l'intérêt du changement de représentation vient surtout du fait que chaque série de coefficients $\{c_k^n, 1 \leq k \leq K\}$ est plus proche du modèle i.i.d que les données brutes $\{s_n(t), 1 \leq t \leq T\}$.

Cependant, un autre type d'approche, spécifique à l'hypothèse de parcimonie, est possible et généralisable au cas des mélanges sous-déterminés. La propriété (59) suggère que le diagramme représentant les colonnes des données transformées $\{\mathbf{X}_k, 1 \leq k \leq K\}$ laisse apparaître des groupes de points accumulés le long des directions des colonnes de $\mathbf{A}$. On peut donc recourir à des algorithmes de classification (clustering) afin d'identifier ces directions pour estimer les colonnes $\mathbf{A}_n$.

Dans tous les cas, l'estimée est obtenue aux indéterminations près, c'est à qu'au mieux on retrouve $\widehat{\mathbf{A}} = \mathbf{ADP}$.

3.2.3 Estimation des sources

Pour les mélanges déterminés $M \geq N$, une fois estimée la matrice de mélange, on peut estimer les sources par séparation linéaire

$$\widehat{\mathbf{s}} = \widehat{\mathbf{A}}^{\dagger}\mathbf{x}$$

où $(\cdot)^{\dagger}$ dénote la (pseudo)inverse.

Dans le cas de mélanges sous-déterminés, les algorithmes de classification qui ont servi à estimer les colonnes de $\mathbf{A}$ permettent simultanément d' affecter chaque colonne $\mathbf{X}_k$ à la source d'indice $n(k)$ la plus vraisemblable. Une estimation aux moindres carrés fournit alors

$$c_k^{n(k)} = \arg\min_c \|\mathbf{X}_k - c\widehat{\mathbf{A}}_k\|^2 = \langle \mathbf{X}_k, \widehat{\mathbf{A}}_k\rangle / \|\widehat{\mathbf{A}}_k\|^2$$

et enfin

$$s_n = \sum_{k|\ n(k)=n} c_k^{n(k)} g_k.$$

3.3 Méthodes de représentation

Le choix du type de représentation des données est un élément essentiel des méthodes de séparation fondées sur la parcimonie. Il fait intervenir le choix du dictionnaire ainsi que d'un algorithme de décomposition.

3.3.1 Décompositions linéaires

La vaste majorité des représentations temps-fréquence ou temps-échelle répandues [54], sont associées à un dictionnaire qui constitue soit une base orthonormale, soit un *tight frame*, et la représentation standard des signaux est obtenue par analyse linéaire $\mathbf{S} := \mathbf{sD}^T$. De même on analyse linéairement les données observées, et l'on peut donc identifier $\mathbf{X} := \mathbf{xD}^T = \mathbf{AsD}^T = \mathbf{AS}$.

Pour une application donnée, on a le choix entre un certain nombre de représentations linéaires. Il peut donc être intéressant de choisir celle qui est la plus appropriée en évaluant ses performances sur une base de données [67, 49, 39, 75, 76]. Une alternative consiste à recourir à des algorithmes adaptatifs qui sélectionnent la "meilleure" représentation en fonction des observations $\mathbf{x}$.

3.3.2 Meilleure base orthonormale

Si l'on a le choix entre plusieurs bases orthonormales $\{\mathbf{D}_\lambda\}$ pour représenter les données $\mathbf{x}$, les algorithmes de sélection de meilleure base (Best Orthonormal Basis) proposent de retenir celle où la représentation conjointe des observations est la plus parcimonieuse, en utilisant un critère entropique [20, 38] $\lambda(\mathbf{x}) := \arg\min_\lambda C(\mathbf{x}, \mathbf{D}_\lambda)$. Au final on obtient la représentation $\mathbf{X} = \mathbf{xD}_{\lambda(\mathbf{x})}^T = \mathbf{AsD}_{\lambda(\mathbf{x})}^T$. La recherche d'une meilleures base est particulièrement intéressante, car algorithmiquement efficace, lorsque l'on dispose d'une "bibliothèque " de bases $\{\mathbf{D}_\lambda\}$ structurée de façon arborescente, comme les paquets d'ondelettes et les cosinus locaux [5].

3.3.3 Basis Pursuit, Matching Pursuit

Il arrive que les performances des algorithmes de meilleure base soient limitées par le fait que les composantes finalement obtenues sont nécessairement orthogonales. Les techniques de poursuite (Basis Pursuit [19] et Matching Pursuit [55]) permettent de s'affranchir de cette contrainte lorsqu'on en a besoin, au prix d'une plus grande complexité algorithmique. Le Basis Pursuit cherche la représentation la plus parcimonieuse au sens ℓ^1 (voir Eq. (57). Il s'appuie sur des techniques de programmation linéaire relativement gourmandes en temps de calcul, qui peuvent être adaptées et accélérées en tenant compte de la structure particulière du dictionnaire. Dans le cas de dictionnaires redondants arbitraires, le Matching Pursuit [55, 37, 38] est un algorithme itératif simple à mettre en oeuvre qui permet également de calculer une décomposition adaptative des observations. Des résultats récents montrent que le Basis Pursuit [30, 31, 40, 35], comme le Matching Pursuit [36, 73, 41], sont capables de retrouver la "bonne" représentation $\mathbf{X} = \mathbf{AS}$ à condition que le dictionnaire soit structuré (quasi-incohérent) et que les sources admettent une représentation suffisamment parcimonieuse.

Il est courant d'utiliser les techniques de poursuite avec des dictionnaires définis de manière analytique (Gabor, paquets d'ondelettes), mais leur intérêt réside aussi dans la possibilité d'employer un dictionnaire arbitraire, qui peut avoir été estimé ("appris") à partir d'un jeu de données. Les techniques d'ACI peuvent s'avérer très intéressantes

pour cet "apprentissage".

3.3.4 Le codage parcimonieux, un problème dual de la SAS

L'idée fondamentale du codage parcimonieux est de chercher quel dictionnaire de "formes d'ondes élémentaires" sert de briques de base à la construction de classes de signaux "naturels" [33, 11, 10, 50, 1]. Autrement dit, si l'on observe N signaux $s_n(t)$ modélisées par (58) avec un dictionnaire $\mathcal{D} = \{g_k(t), 1 \leq k \leq K\}$ inconnu et des coefficients c_n^k aléatoires, indépendants, et parcimonieux, le but est de retrouver le dictionnaire. A partir du jeu de N observations $\mathbf{s}$ et du modèle $\mathbf{s} = \mathbf{SD}$, on cherche donc à retrouver $\mathbf{D}$.

En passant au transposé, on obtient $\mathbf{s}^T = \mathbf{D}^T\mathbf{S}^T$. On reconnaît alors un problème de SAS où les observations sont $\mathbf{s}^T$, la matrice de mélange est $\mathbf{D}^T$ et les sources inconnues $\mathbf{S}^T$. Les méthodes d'ACI peuvent donc être utilisées, à la nuance près que le but est plus ici d'identifier la matrice de mélange $\mathbf{D}^T$ que les coefficients.

4 Conclusion

L'ACI est une méthode très puissante de traitement de l'information, dans la mesure où elle ne nécessite que très peu d'hypothèses *a priori* à part l'indépendance statistique des sources. En revanche, l'ACI estime un modèle (de l'inverse) du mélange, et la nature de ce dernier doit être connue. De plus, l'utilisation d'informations *a priori* faibles (qui préservent la puissance de la méthode) permet de simplifier les algorithmes.

Dans le cas de la SAS, des problèmes difficiles restent à approfondir, notamment la séparation dans des grands mélanges (grand nombre de sources et/ou de capteurs), dans des mélanges fortement bruités, convolutifs réalistes ou non linéaires et dans des mélanges sous-déterminés. Dans ce dernier cas, la restitution des sources est un problème mal posé, qui peut être régularisé en introduisant, en plus des contraintes structurelles, des hypothèses de parcimonie sur les signaux. L'ACI, en tant qu'analyse parcimonieuse de données complexes, doit s'enrichir vers des modèles non linéaires : en effet, comment des données non linéaires peuvent être représentées de façon pertinente par un seul modèle linéaire?

L'ACI peut être utilisée dans de nombreux domaines dès que l'on dispose d'observations multi-dimensionnelles : instrumentation, microélectronique, biomédical, sismique, chimie, astronomie. Il est temps de porter nos efforts en l'appliquant à des problèmes réels et difficiles. A nous de savoir expliquer cette approche et de contribuer à sa diffusion la plus large !

Références

[1] S.A. Abdallah and M.D. Plumbley. If edges are the independent components of natural images, what are the independent components of natural sounds? In *Proceedings of ICA2001)*, pages 534–539, San Diego, California, December 2001.

[2] F. Abrard and Y. Deville. From blind source seapration to blind source cancellation in the undetermined case: a new approach based on time-frequency analysis. In *Proceedings of ICA2001*, pages 734–739, San Diego (CA), December 2001.

[3] S. Achard, D.T. Pham, and C. Jutten. Quadratic dependence measure for nonlinear blind source separation. In *Proceedings of ICA2003*, pages 263–268, Nara (Japan), April 1-4 2003.

[4] J. Aczel. *Lectures on Functional Equations and Their Applications*. Academic Press, New-York, 1966.

[5] P. Auscher, G. Weiss, and M. V. Wickerhauser. *Wavelets – a tutorial in theory and applications*, chapter Local sine and cosine bases of Coifman and Meyer and the construction of smooth wavelets., pages 237–256. Academic Press, Boston, MA, 1992. Chui, C.K., Edt.

[6] M. Babaie-Zadeh, C. Jutten, and K. Nayebi. Separating convolutive mixtures by mutual information minimization. In *Proceedings of IWANN'2001*, pages 834–842, Granada, Spain, Juin 2001.

[7] M. Babaie-Zadeh, C. Jutten, and K. Nayebi. Separating convolutive post non-linear mixtures. In *Proceedings of ICA'2001)*, pages 138–143, San Diego (California, USA), 2001.

[8] M. Babaie-Zadeh, C. Jutten, and K. Nayebi. A geometric approach for separating post nonlinear mixtures. In *Proceedings of EUSIPCO 2002*, volume II, pages 11–14, Toulouse, France, 2002.

[9] M. Babaie-Zadeh, C. Jutten, and K. Nayebi. Differential of mutual information function. *IEEE Signal Processing Letters*, 2003. To appear.

[10] A.J. Bell and T.J. Sejnowski. Learning the higher-order structure of a natural sound. *Network*, 7(2), 1996.

[11] A.J. Bell and T.J. Sejnowski. The 'independent components' of natural scenes are edge filters. *Vision Research*, 37(23):3327–3338, 1997.

[12] T. Bell and T. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Comutation*, 7(6):1004–1034, 1995.

[13] A. Belouchrani, K. Abed Meraim, J. F. Cardoso, and E. Moulines. A blind source separation technique based on second order statistics. *IEEE Trans. on Signal Processing*, 45(2):434–444, 1997.

[14] P. Bofill and M. Zibulevsky. Underdetermined blind source separation using sparse representations. *Signal Processing*, 81:2353–2362, 2001.

[15] V. Capdevielle, Ch. Servière, and Lacoume J.-L. Blind separation of wide-band sources in the frequency domain. In *ICASSP*, pages 2080–2083, 1995.

[16] J.-F. Cardoso. Blind signal separation: statistical principles. *Proc. of the IEEE*, 9(10):2009–2025, 1998.

[17] J.-F. Cardoso. The three easy routes to independent component analysis: Contrasts and geometry. In *Proceedings of ICA2001*, pages 1–6, San Diego, CA, December 2001.

[18] N. Charkani and Y. Deville. Optimization of the asymptotic performance of time-domain convolutive source separation algorithms. In *Proceedings of ESANN 97*, pages 273–278, Bruges (Belgium), April 1997.

[19] S. Chen, D.L. Donoho, and M.A. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20(1):33–61, January 1999.

[20] R. Coifman and M.V. Wickerhauser. Entropy-based algorithms for best basis selection. *IEETIT*, 38(2):713–718, March 1992.

[21] P. Comon. Separation of sources using higher-order cumulants. In *SPIE Vol. 1152 Advanced Algorithms and Architectures for Signal Processing IV*, San Diego (CA), USA, August 8-10 1989.

[22] P. Comon. Independent component analysis. In J.-L. Lacoume, M. A. Lagunas, and C. L. Nikias, editors, *International Workshop on High Order Statistics*, pages 111–120, Chamrousse, France, July 1991.

[23] P. Comon. Independent component analysis, a new concept? *Signal Processing*, 36(3):287–314, 1994.

[24] P Comon. Contrasts for multichannel blind deconvolution. *IEEE Signal Processing Letters*, 3(7):209–211, 1996.

[25] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* Wiely Series in Telecommunications, 1991.

[26] A. Dapena and C. Servière. A simplified frequency-domain approach for blind separation of convolutive mixtures. In *Proceedings of ICA2001*, pages 569–574, San Diego (CA, USA), December 2001.

[27] G. Darmois. Analyse des liaisons de probabilité. In *Proc. of Int. Statistics Conferences 1947*, volume III A, page 231, Washington (D.C.), 1951.

[28] V. Devlamink and P. Terrier. Improving a separating reflection process using multispectral acquisitions. In *Physics in Signal and Image Processing 2003*, pages 1–4, Grenoble (France), January 2003.

[29] D. L. Donoho. On minimum entropy deconvolution. In *Proc. 2nd Applied Time Series Symp.*, Tulsa, 1980. reprinted in *Applied Time Series Analysis II*, Academic Press, New York, 1981, pp. 565-609.

[30] D.L. Donoho and Xiaoming Huo. Uncertainty principles and ideal atomic decompositions. *IEEE Trans. Inform. Theory*, 47(7):2845–2862, November 2001.

[31] M. Elad and A.M. Bruckstein. A generalized uncertainty principle and sparse representations in pairs of bases. *IEEE Trans. Inform. Theory*, 48(9):2558–2567, September 2002.

[32] J. Eriksson and V. Koivunen. Blind identifiability of a class of nonlinear instantaneous ICA models. In *Proceedings of EUSIPCO 2002*, Toulouse (France), September 2002.

[33] D.J. Field and B.A. Olshausen. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381:607–609, 1996.

[34] P. Flandrin. *Temps-Fréquence.* Hermes, Paris, France, 1993.

[35] J.-J. Fuchs. On sparse representations in arbitrary redundant bases. Technical report, IRISA, December 2002. submitted to IEEE Trans. Inf. Th.

[36] A.C. Gilbert, S. Muthukrishnan, and M.J. Strauss. Approximation of functions over redundant dictionaries using coherence. In *SIAM Symposium on Discrete Algorithms (SODA'03)*, 2003.

[37] R. Gribonval. Sparse decomposition of stereo signals with matching pursuit and application to blind separation of more than two sources from a stereo mixture. In *Proceedings of ICASSP'02*, Orlando, Florida, May 2002.

[38] R. Gribonval. Piecewise linear separation : a common framework for de-multiplexing, segmentation and blind source separation. In M.A. Unser, A. Aldroubi, and A.F. Laine, editors, *Wavelets X, Proc. SPIE*, San Diego, CA, August 2003.

[39] R. Gribonval, L. Benaroya, E. Vincent, and C. Févotte. Proposals for performance measurement in source separation. In *Proceedings of ICA2003*, pages 763–768, Nara, Japan, April 2003.

[40] R. Gribonval and M. Nielsen. Sparse decompositions in "incoherent" dictionaries. In *Proceedings ICIP'03*, Barcelona, Spain, sep 2003.

[41] R. Gribonval and P. Vandergheynst. Exponential convergence of Matching Pursuit in quasi-incoherent dictionaries. Technical report, IRISA, 2003. In preparation.

[42] W. Härdle. *Smoothing Techniques, with implementation in S.* Springer-Verlag, 1990.

[43] A. Jourjine, S. Rickard, and O. Yilmaz. Blind separation of disjoint orthogonal signals: Demixing n sources from 2 mixtures. In *Proceedings of ICASSP'00*, volume 5, pages 2985–2988, Istanbul, Turkey, June 2000.

[44] A. M. Kagan, Y. V. Linnik, and C. R. Rao. *Characterization Problems in Mathematics Statistics.* John Wiley & Sons, 1973.

[45] A. M. Kagan, Y. V. Linnik, and C. R. Rao. Extension of Darmois-Skitovich theorem to functions of random variables satisfying an addition theorem. *Communications in Statistics*, 1(5):471–474, 1973.

[46] M. Kendall and A. Stuart. *The Advanced Theory of Statistics, Distribution Theory*, volume 1. Griffin, 1977.

[47] P. Kisilev, M. Zibulevsky, Y. Y. Zeevi, and B. A. Pearlmutter. Multiresolution framework for blind source separation. Technical Report CCIT Report # 317, Technion University, June 2001.

[48] M.J. Korenberg and I.W. Hunter. The identification of nonlinear biological systems: LNL cascade models. *Biological Cybernetics*, 43(12):125–134, December 1995.

[49] R.H. Lambert. Difficulty measures and figures of merit for source separation. In *Proceedings of ICA'99)*, pages 133–138, Aussois, France, 1999.

[50] M.S. Lewicki. Efficient coding of natural sounds. *Nature Neurosci.*, 5(4):356–363, 2002.

[51] X Liu and J. Hérault. Colour image processing by a neural network model. In *Proceedings of INNC 90*, pages 3–6, Paris, France, July 1990.

[52] E. Lukacs. A characterization of the Gamma distribution. *Annals of Mathematical Statistics*, (26):319–324, 1955.

[53] G. M. Lennon, M.C. Mercier, L. Mouchot, and Hubert-Moy. Spectral unmixing of hyperspectral images with the independent component analysis and wavelet packets. In *Proceedings of International Geoscience And Remote Sensing Symposium*, Sydney (Australia), July 2001.

[54] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, San Diego, CA, 1998.

[55] S. Mallat and Z. Zhang. Matching pursuit with time-frequency dictionaries. *IEETS3*, 41(12):3397–3415, December 1993.

[56] K. Matsuoka, M. Ohya, and M. Kawamoto. A neural net for blind separation of nonstationary signals. *Neural Networks*, 8(3):411–419, 1995.

[57] A. Hyvärinen and E. Oja. A fast fixed point algorithm for independent component analysis. *Neural computation*, 9:1483–1492, 1997.

[58] L. Molgedey and H. G. Schuster. Separation of a mixture of independent signals using time delayed correlation. *Physical Review Letters*, 72:3634–3636, 1994.

[59] N. Murata. Properties of the empirical characteristic function and its applications to testing for independence. In *Proceedings ICA2001*, pages 19–24, San Diego, CA, December 2001.

[60] A. Parashiv-Ionescu, C. Jutten, and G. Bouvier. Source separation based processing for integrated hall sesnor arrays. *IEEE Sensors Journal*, 2(6):663–673, December 2002.

[61] D. T. Pham. Contrast functions for blind separation and deconvolution of sources. In *Proceedings of ICA2001*, pages 37–42, San Diego, CA, December 2001.

[62] D. T. Pham and J.-F. Cardoso. Blind separation of instantaneous mixtures of nonstationary sources. *IEEE Trans. on Signal Processing*, 49(9):1837–1848, 2001.

[63] D. T. Pham and Ph. Garat. Blind separation of mixture of independent sources through a quasi-maximum likelihood approach. *IEEE Trans. on S.P.*, 45(7):1712–1725, July 1997.

[64] D. T. Pham, Ph. Garat, and C. Jutten. Separation of a mixture of independent sources through a maximum likelihood approach. In *Proceedings of EUSIPCO 92*, pages 771–774, Brussels, Belgium, August 1992.

[65] S. Prakriya and D. Hatzinakos. Blind identification of LTI-ZMNL-LTI nonlinear channel models. *IEEE Trans. on Signal Processing*, 43(12):3007–3013, December 1995.

[66] C. Puntonet, A. Prieto, C. Jutten, M. Alvarez, and J. Ortega. Separation of sources: a geometry-based procedure for reconstruction of n-valued signals. *Signal Processing*, 46:267 284, 1995.

[67] D. Schobben, K. Torkkola, and P. Smaragdis. Evaluation of blind signal separation methods. In *Proceedings of ICA'99*, pages 261–266, Aussois, France, January 1999.

[68] C. Simon, Ph. Loubaton, and C. Jutten. Separation of a class of convolutive mixtures : a contrast function approach. *Signal Processing*, 81(4):838–888, 2001.

[69] A. Taleb and C. Jutten. On underdetermined source separation. In *Proceedings of ICASSP'99*, pages 2089–2092, Phoenix (AR, USA), May 1999.

[70] A. Taleb and C. Jutten. Source separation in post nonlinear mixtures. *IEEE Transactions on Signal Processing*, 47(10):2807 2820, 1999.

[71] H.L. Nguyen Thi and C. Jutten. Blind sources separation for convolutive mixtures. *Signal Processing*, 45:209–229, 1995.

[72] L. Tong, V. Soon, Y. Huang, and R. Liu. AMUSE: a new blind identification algorithm. In *Proceedings of ISCAS'90*, New Orleans (USA), 1990.

[73] J. Tropp. Greed is good : Algorithmic results for sparse approximation. Technical report, Texas Institute for Computational Engineering and Sciences, 2003. In preparation.

[74] M. Van Hulle. Clustering approach to square and non-square blind source separation. In *IEEE Workshop on Neural Networks for Signal Processing (NNSP99)*, pages 315–323, August 1999.

[75] E. Vincent, C. Févotte, L. Benaroya, and R. Gribonval. A tentative typology of audio source separation tasks. In *Proceedings of ICA2003*, pages 715–720, Nara, Japan, April 2003.

[76] E. Vincent, C. Févotte, R. Gribonval, et al. Comment évaluer les algorithmes de séparation de sources audio? In *Actes du GRETSI 2003*, ENST, Paris, France, sep 2003.

[77] H.-H. Yang, S.I. Amari, and A. Cichocki. Information-theoretic approach to blind separation of sources in non-linear mixtures. *Signal Processing*, pages 291–300, February 1998.

[78] D Yellin and E. Weinstein. Multichannel signal separation: methods and analysis. *IEEE Trans. Signal Processing*, pages 106–118, January 1996.

[79] M. Zibulevsky and B.A. Pearlmutter. Blind source separation by sparse decomposition in a signal dictionary. *Neural Computations*, 13(4):863–882, 2001.