

**Analyse en composantes indépendantes pour la séparation aveugle de sources**

C. Durieu<sup>1</sup> et M. Kieffer<sup>2</sup>

<sup>1</sup> Département EEA, ENS Cachan  
61 avenue du président Wilson  
F-94235 Cachan CEDEX, France

<sup>2</sup> L2S - CNRS - Supélec - Université de Paris-Sud  
F-91192 Gif-sur-Yvette CEDEX, France

# ANALYSE EN COMPOSANTES INDÉPENDANTES POUR LA SÉPARATION AVEUGLE DE SOURCES

Cécile DURIEU\* et Michel KIEFFER†

\*Département EEA – ENS Cachan, 61 avenue du Président Wilson, 94235 Cachan Cedex  
email : cecile.durieu@eea.ens-cachan.fr

†L2S – CNRS – Supélec – Université de Paris-Sud, 91192 Gif-sur-Yvette Cedex  
email : kieffer@lss.supelec.fr

**Résumé :** L'analyse en composantes indépendantes permet d'illustrer de nombreuses notions enseignées en traitement du signal tout en considérant une application ludique et attractive permettant de séparer un mélange de signaux sonores.

**Mots clés :** blanchiment, indépendance, kurtosis, méthode du point fixe, non corrélation, optimisation, séparation de sources, variables aléatoires gaussiennes.

## 1. MOTIVATIONS

Le *cocktail party problem* est le problème type que tente de résoudre l'analyse en composantes indépendantes (ICA). Lors d'une réception animée il s'agit d'isoler ce que dit chaque participant à partir d'enregistrements du brouhaha ambiant. Ce problème, dit de *séparation de sources*, se rencontre aujourd'hui dans de nombreuses applications. Ainsi il s'agit, par exemple, en électroencéphalographie de séparer les signaux utiles d'artéfacts dus au clignement des yeux, en astrophysique de classer des objets célestes à partir d'images obtenues dans différents domaines spectraux, et en télécommunications, dans un contexte d'accès multiple par code (CDMA), de séparer différents utilisateurs partageant les mêmes ressources fréquentielles et temporelles.

D'un point de vue pédagogique l'ICA permet d'illustrer des notions de traitement du signal souvent abstraites et difficiles à appréhender par les étudiants. La différence entre indépendance et non-corrélation de variables aléatoires, le caractère plus ou moins gaussien d'une distribution de probabilité, l'intérêt des moments ou encore du blanchiment sont illustrés par cette méthode. L'ICA conduit à résoudre un problème d'optimisation et peut donc aussi illustrer un cours d'optimisation.

Le principe de l'ICA est donné dans la section 2. Il nécessite la mesure de l'indépendance de variables aléatoires qui, dans la section 3, repose sur l'éloignement

d'une distribution de probabilité à une distribution gaussienne. Le blanchiment des signaux à traiter permet de simplifier la formulation du problème (section 4) qui est résolu dans la dernière section. Les points évoqués tout au long de l'article sont illustrés pour séparer un mélange instantané de signaux sonores. Cela rend l'application attractive et ludique, même si dans la pratique la séparation de tels signaux est plus complexe et fait encore l'objet de recherches.

## 2. OBJECTIF ET PRINCIPE DE L'ICA

L'objectif est d'estimer  $N$  signaux *sources*, supposés stationnaires, ergodiques et indépendants, à partir de la mesure de  $M$  signaux *observés* résultant d'un mélange des signaux sources. Le cas le plus simple est considéré ici : on suppose que le mélange est linéaire et instantané, que le bruit d'observation peut être négligé et que  $M = N$ . On note  $\mathbf{e}_t$  et  $\mathbf{s}_t$  le vecteur (colonne) des signaux sources et observés, respectivement, à l'instant  $t$  ( $t=1, \dots, T$ ). On a alors  $\mathbf{s}_t = \mathbf{A} \mathbf{e}_t$ , où  $\mathbf{A}$  est une matrice de *mélange* inconnue mais supposée constante et inversible. Dans le cas de signaux sonores le mélange n'est pas instantané et un modèle convolutif, permettant de prendre en compte les temps de propagation des signaux, est plus approprié.

À titre illustratif les signaux sources correspondant au mot "bonjour" et à un signal sinusoïdal sont considérés. La figure 1 présente l'évolution temporelle de ces si

gnaux (sous-fig. de gauche) ainsi que la répartition des échantillons (sous-fig. de droite). La répartition de chaque composante est également représentée (suivant les axes de la sous-fig. de droite). La figure 2 correspond aux signaux observés après mélange.

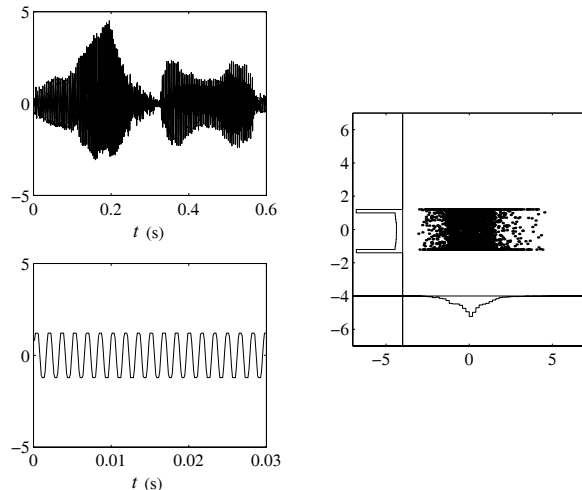


Figure 1 – Signaux sources

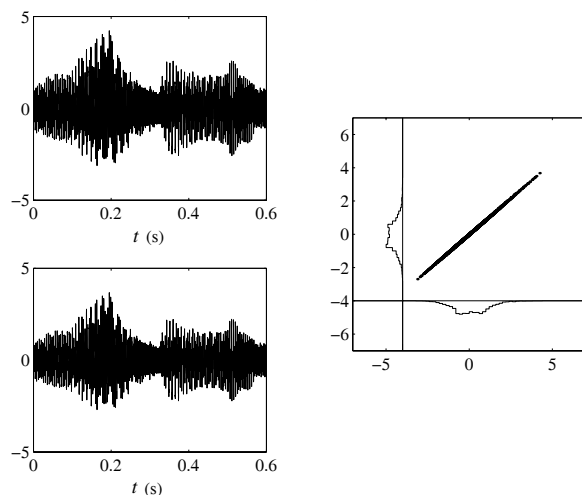


Figure 2 – Signaux observés

La séparation aveugle de sources consiste à rechercher, à partir de  $\mathbf{s}_t$ , une matrice de séparation  $\mathbf{W}$  telle que  $\mathbf{W}\mathbf{A}$  se réduise à une matrice diagonale, à une permutation près. Aucune information sur la structure temporelle des signaux n'est nécessaire. Plusieurs approches permettent de résoudre un tel problème. La méthode présentée dans cet article consiste à rechercher une transformation linéaire (non orthogonale dans le cas général) de l'espace des observations en un espace de représentation dans lequel les composantes sont aussi indépendantes que possible. Une mesure de l'indépendance statistique est alors nécessaire.

### 3. MESURE DE L'INDÉPENDANCE

Pour mesurer l'indépendance de variables aléatoires on pourrait mesurer l'écart entre la densité de probabilité conjointe et le produit des lois marginales. Cependant, compte tenu de la difficulté à estimer les densités de probabilité, cette mesure n'est pas retenue. La technique utilisée dans cet article et présentée dans [1] exploite une extension du théorème de la limite centrale : la somme de deux variables aléatoires indépendantes possède une distribution plus proche d'une gaussienne que la distribution de chacune des variables de la somme. Cette technique cherche à *dé-gaussianiser* les signaux observés et nécessite donc que les signaux sources ne soient pas gaussiens. Il reste alors à définir un critère pour caractériser l'éloignement à une distribution gaussienne.

Un critère couramment utilisé pour tester le caractère gaussien d'une variable aléatoire  $Y$  centrée repose sur le calcul de son cumulatif d'ordre 4 (*kurtosis*) défini par  $k(Y) = E[Y^4] - 3E[Y^2]^2$ , où  $E[Z]$  est la valeur moyenne de la variable aléatoire  $Z$ . Le kurtosis peut être positif ou négatif et il est nul pour une variable aléatoire gaussienne. Ainsi, la valeur absolue ou le carré du kurtosis peut être utilisé comme mesure de l'éloignement à une gaussienne. Le kurtosis dépendant du moment d'ordre 4, son estimation nécessite d'avoir plus d'échantillons que pour le calcul d'un moment d'ordre 2. L'estimateur du kurtosis est relativement sensible au bruit, ainsi en présence de peu d'échantillons et de bruit important d'autres mesures de distance reposant, par exemple, sur l'entropie sont utilisées [1]. En fonction de connaissances des étudiants ces mesures peuvent être introduites, cependant leur évaluation nécessite des approximations dont la justification n'est pas triviale. La mesure retenue dans cet article repose sur le kurtosis.

### 4. SIMPLIFICATION DU PROBLÈME

Sous sa formulation initiale la détermination de  $\mathbf{W}$  requiert l'estimation de  $N^2$  paramètres. En blanchissant les signaux observés le nombre de paramètres à déterminer peut, comme nous allons le voir, être réduit

d'un facteur deux environ.

La première étape consiste à centrer les données :  $\mathbf{s}_t$  est remplacé par  $\mathbf{s}_t - E[\mathbf{s}_t]$ . La seconde étape transforme, par une opération linéaire, le vecteur centré  $\mathbf{s}_t$  en un vecteur dont les composantes sont décorréélées deux à deux et de variance unité. L'opération de blanchiment s'effectue, par exemple, par décomposition en valeur propre de la matrice de covariance  $\Gamma = E[\mathbf{s}_t \mathbf{s}_t^T]$  de  $\mathbf{s}_t$ . Soient  $\Gamma = \mathbf{P} \mathbf{D} \mathbf{P}^T$ , où  $\mathbf{D} = \text{diag}(d_1, \dots, d_N)$  est une matrice diagonale contenant les valeurs propres  $d_i$  de  $\Gamma$  et  $\mathbf{P}$  est la matrice orthogonale des vecteurs propres associés. Alors  $\tilde{\mathbf{s}}_t = \mathbf{P} \mathbf{D}^{-1/2} \mathbf{P}^T \mathbf{s}_t$ , où  $\mathbf{D}^{-1/2} = \text{diag}(d_1^{-1/2}, \dots, d_N^{-1/2})$ , est tel que  $E[\tilde{\mathbf{s}}_t \tilde{\mathbf{s}}_t^T] = \mathbf{I}$ ,  $\mathbf{I}$  étant la matrice unité. Par conséquent les différentes composantes de  $\tilde{\mathbf{s}}_t$  sont décorréélées, on dit encore que  $\tilde{\mathbf{s}}_t$  est blanc. On a alors  $\tilde{\mathbf{s}}_t = \tilde{\mathbf{A}} \mathbf{e}_t$  avec  $\tilde{\mathbf{A}} = \mathbf{P} \mathbf{D}^{-1/2} \mathbf{P}^T \mathbf{A}$ .

Les signaux étant supposés stationnaires et ergodiques, la moyenne statistique  $E[Z_t]$  est évaluée à partir de la moyenne temporelle  $\sum_{t=1}^T Z_t / T$ .

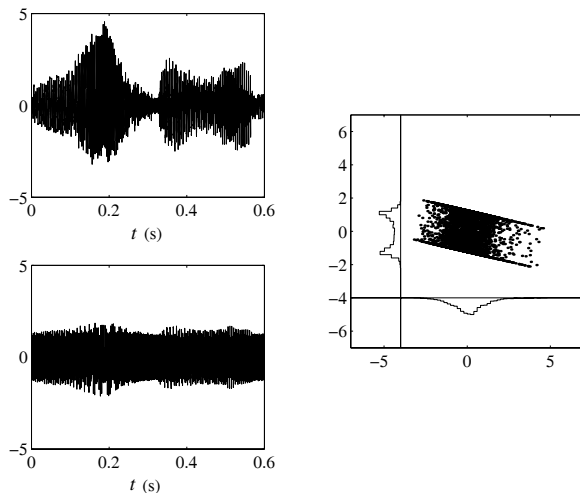


Figure 3 – Signaux blanchis

On constate que le blanchiment des données (cf. figure 3) ne suffit pas à rendre les signaux indépendants. À l'écoute les signaux avant et après blanchiment diffèrent très peu. Signalons que de nombreux traitements de données nécessitent au préalable une étape de

blanchiment.

Sans pour autant restreindre l'étude on suppose que les différentes composantes des signaux sources sont centrées et ont une variance unité. Ainsi, compte tenu des différentes hypothèses et notations, on a  $E[\tilde{\mathbf{s}}_t \tilde{\mathbf{s}}_t^T] = \tilde{\mathbf{A}}_t E[\mathbf{e}_t \mathbf{e}_t^T] \tilde{\mathbf{A}}_t^T = \tilde{\mathbf{A}} \tilde{\mathbf{A}}^T = \mathbf{I}$ . Le fait d'avoir blanchi les données se traduit par une matrice de mélange  $\tilde{\mathbf{A}}$  qui est orthogonale. Ceci se répercute sur la structure de la matrice de séparation  $\tilde{\mathbf{W}}$  qui doit être telle que  $\tilde{\mathbf{W}} \tilde{\mathbf{A}}$  est la matrice unité, à une permutation près. On a alors  $\tilde{\mathbf{W}} \tilde{\mathbf{W}}^T = \mathbf{I}$ . Ainsi, après blanchiment, le problème de recherche de la matrice de séparation se trouve simplifié car la matrice  $\tilde{\mathbf{W}}$  doit être orthogonale. Par la suite on suppose que l'étape de blanchiment des données a été appliquée et la notation tilt est omise.

## 5. MATRICE DE SÉPARATION

Comme nous allons le voir l'ICA consiste, après blanchiment des données, à résoudre un problème d'optimisation sous contraintes.

### 5.1. Recherche d'une seule composante

Dans un premier temps on s'intéresse à la recherche d'une seule composante indépendante ou encore d'un des vecteurs ligne  $\mathbf{w}^T$  de la matrice  $\mathbf{W}$ . Ce vecteur est recherché de manière à ce que le signal  $\mathbf{w}^T \mathbf{s}_t$  maximise le critère d'éloignement à une gaussienne  $J_g(\mathbf{w}^T \mathbf{s}_t)$  sous la contrainte  $\|\mathbf{w}\| = 1$  afin que  $\mathbf{W}$  soit orthogonale. Il s'agit donc de résoudre un problème d'optimisation sous contraintes.

Une méthode classique de résolution d'un tel problème consiste à introduire un multiplicateur de Lagrange  $\lambda$  et à maximiser la fonction  $L(\mathbf{w}, \lambda) = J_g(\mathbf{w}^T \mathbf{s}_t) + \lambda(\mathbf{w}^T \mathbf{w} - 1)$  par rapport à  $\mathbf{w}$  et à  $\lambda$  [2]. Il n'existe pas de solution analytique à ce problème d'optimisation et plusieurs méthodes itératives sont possibles pour rechercher la solution, par exemple la méthode du point fixe ou la méthode de Newton, qui converge plus rapidement. La méthode du point fixe permet de résoudre ce problème sans avoir à faire de

simplifications ; elle est détaillée ci-dessous.

Le vecteur  $\mathbf{w}$  doit annuler le gradient de  $L(\mathbf{w}, \lambda)$ . Il satisfait donc la condition  $\nabla J_g(\mathbf{w}^T \mathbf{s}_t) + 2\lambda \mathbf{w} = 0$  avec  $\|\mathbf{w}\|=1$ . Pour le critère reposant sur le kurtosis et défini par  $J_g(Y) = k^2(Y)$ , on a  $\nabla J_g(\mathbf{w}^T \mathbf{s}_t) = 2k(\mathbf{w}^T \mathbf{s}_t) \nabla k(\mathbf{w}^T \mathbf{s}_t)$ , avec  $\nabla k(\mathbf{w}^T \mathbf{s}_t) = 4\alpha - 12\beta$ , où  $\alpha = E[(\mathbf{w}^T \mathbf{s}_t)^3 \mathbf{s}_t]$  et  $\beta = E[(\mathbf{w}^T \mathbf{s}_t)^2] E[(\mathbf{w}^T \mathbf{s}_t) \mathbf{s}_t]$ . Le vecteur  $\mathbf{w}$  devant être normé et  $\mathbf{s}_t$  étant blanc, les expressions ci-dessus se simplifient. On démontre que le vecteur  $\mathbf{w}$  annulant le gradient est de la forme  $\mathbf{w} = \gamma(\alpha - 3\mathbf{w})$ , où  $\gamma$ , dépendant de  $\lambda$ , est choisi de manière à assurer  $\|\mathbf{w}\|=1$ . La méthode itérative du point fixe consiste partant d'un  $\mathbf{w}$  donné à le remplacer par  $\alpha - 3\mathbf{w}$  puis à le normer et ce jusqu'à convergence.

L'algorithme  $\mathcal{A}1$  de recherche de la première composante indépendante à partir de données blanchies est donné ci-dessous.

1. Poser  $n = 1$  et  $k = 0$ .
2. Initialiser aléatoirement  $\mathbf{w}(k)$  puis normer ce vecteur.
3. Calculer  $\mathbf{w}(k+1) = E[(\mathbf{w}(k)^T \mathbf{s}_t)^3 \mathbf{s}_t] - 3\mathbf{w}(k)$  puis normer ce vecteur.
4. Si  $|\mathbf{w}(k+1)^T \mathbf{w}(k)|$  est proche de 1 poser  $\mathbf{w}_n = \mathbf{w}(k+1)$  et arrêter  $\mathcal{A}1$ , sinon poser  $k = k+1$  et répéter les étapes 3 et 4.

Le résultat de l'algorithme dépend de l'initialisation car  $L(\mathbf{w}, \lambda)$  possède plusieurs maxima locaux.

## 5.2. Recherche des autres composantes

Il reste à déterminer les autres vecteurs lignes constituant la matrice orthogonale  $\mathbf{W}$ . L'algorithme  $\mathcal{A}2$  décrit ci-dessous et reposant sur une orthogonalisation de Gram-Schmidt peut être utilisé.

1. Poser  $n = n+1$  et  $k = 0$ .
2. Appliquer l'étape 2 de  $\mathcal{A}1$ .
3. Appliquer l'étape 3 de  $\mathcal{A}1$ .
4. Poser  $\mathbf{w} = \mathbf{w}(k+1)$  puis calculer  $\mathbf{w}(k+1) = \mathbf{w} - \sum_{p=1}^{n-1} \mathbf{w}^T \mathbf{w}_p \mathbf{w}_p$  et normer ce vecteur.

5. Si  $|\mathbf{w}(k+1)^T \mathbf{w}(k)|$  est proche de 1 poser  $\mathbf{w}_n = \mathbf{w}(k+1)$  et passer à l'étape 6 de  $\mathcal{A}2$ , sinon poser  $k = k+1$  et répéter les étapes 3 à 5 de  $\mathcal{A}2$ .

6. Si  $n < N$  répéter les étapes 1 à 5 de  $\mathcal{A}2$ , sinon poser  $\mathbf{W} = [\mathbf{w}_1 \dots \mathbf{w}_N]^T$  et arrêter  $\mathcal{A}2$ .

Les signaux sources estimés sont donnés par la relation  $\hat{\mathbf{e}}_t = \mathbf{W} \mathbf{s}_t$ . Le résultat dépend de l'ordre dans lequel les vecteurs  $\mathbf{w}_n$  sont introduits. Pour éviter ceci il est possible d'effectuer une orthogonalisation globale.

Après quelques itérations de l'algorithme les composantes de la matrice de séparation sont déterminées et les signaux sources sont estimés (cf. figure 4) avec une erreur très faible (d'un point de vue puissance de l'erreur et reconnaissance auditive).

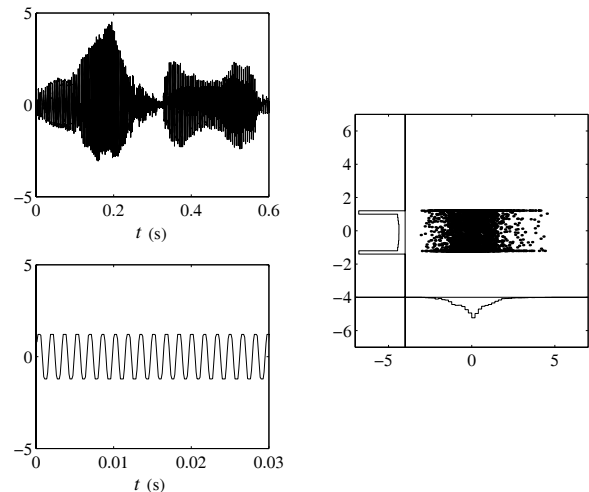


Figure 4 – Signaux estimés

Dans le cas de deux sources  $\mathbf{W}$  dépend d'un seul paramètre : un angle de rotation  $\theta$ . L'évolution en fonction de  $\theta$  du critère  $J_g(\mathbf{w}^T \mathbf{s}_t)$ , avec  $\mathbf{w} = [\cos \theta \sin \theta]^T$ , permet de déterminer graphiquement cet angle sans avoir à mettre en œuvre un algorithme d'optimisation sous contrainte.

## Bibliographie

- [1] A. Hyvärinen, J. Karhunen, E. Oja, *Independent Component Analysis*, John Wiley & Sons, New-York, NY, 2001 (<http://www.cis.hut.fi/~aapo/>).
- [2] R. Fletcher, *Practical Methods of Optimization*, 2<sup>nd</sup> ed., John Wiley & Sons, Inc., New York, 1987.