



SEPARATION DE MELANGES DE SIGNAUX *

P.COMON

THOMSON SINTRA, BP 53, 06801 Cagnes sur Mer

RESUME

Le problème consiste à estimer p processus stochastiques, à partir de l'observation d'une transformation linéaire inconnue de ces derniers. Les algorithmes développés ici utilisent uniquement l'hypothèse d'indépendance statistique des processus initiaux. Les processus recherchés sont appelés sources. Pour simplifier, on considère que l'observation est une transformée linéaire instantanée des sources, autrement dit définie par une matrice constante, A . De plus, on suppose que cette matrice est carrée et régulière, condition suffisante pour que le problème soit soluble à une matrice diagonale multiplicative près. Deux algorithmes sont comparés. On présente en parallèle l'algorithme original de Jutten et Herault proposé en 1985, dont on souligne les limitations, et un nouvel algorithme adaptatif plus performant. Ce dernier permet d'annuler les cumulants croisés de sortie.

SUMMARY

The problem consists of isolating p independent stochastic processes from the observation of a unknown linear transform of them. The algorithms designed in the sequel are solely based upon the assumption that original processes are statistically independent. Those unknown processes are called sources. For the sake of simplicity, the linear transform is assumed to be instantaneous, namely defined by a constant matrix, A . Moreover, A is assumed square and regular in order for the problem to be solvable up to a diagonal multiplicative matrix. Two algorithms are compared. We introduce both the original adaptive algorithm of Jutten and Herault proposed in 1985 whose limitations are emphasized, and a new one performing better. The latter algorithm aims at cancelling the output cross-cumulants.

* Supported by D.R.E.T., Paris.

1. Position du problème

L'observation est constituée de p processus aléatoires, et s'écrit de manière compacte:

$$\begin{aligned} e(t) &= A x(t); \\ x(t) &\triangleq [x_1(t), \dots, x_p(t)]^T \\ e(t) &\triangleq [e_1(t), \dots, e_p(t)]^T. \end{aligned} \quad (1)$$

Les inconnues à déterminer sont la transformation déterministe A , et le processus stationnaire à p variables, dit source, $x(t)$. Pour ce faire, il est nécessaire et suffisant de déterminer une matrice F telle que la sortie

$$s(t) \triangleq F e(t) \quad (2)$$

approche *au mieux* le processus source $x(t)$. Notons que si $s(t)$ est telle que toutes ses composantes $s_i(t)$ sont globalement (i.e. mutuellement dans le jargon de statistiques) indépendantes, alors il en est de même pour le processus $\Lambda P s(t)$, où Λ est une matrice diagonale régulière quelconque et P est une permutation quelconque. Nous constatons donc qu'il est impossible de recouvrer les amplitudes et signes des sources, ni l'ordre dans lequel elles se présentaient. Il faut donc comprendre par "au mieux" *conformément à une distance $d\{x(t), \Lambda P s(t)\}$* , où Λ et P sont laissées libres. Il est difficile de minimiser une telle distance, ou même de la définir correctement; nous choisissons plutôt de fixer p paramètres et de laisser subsister l'indétermination de signe et de permutation. La

transformation F contenant p^2 paramètres, il restera donc $p(p-1)$ inconnues. Nous présentons ci-après deux paramétrisations possibles.

1.1. Paramétrisation de Jutten-Herault (JH):

Dans leur approche [1] [2], Jutten et Herault imposent $\text{Diag}(F^{-1}) = I$, et cherchent C telle que

$$F = (I+C)^{-1} \text{ et } \text{Diag}(C) = 0. \quad (3)$$

Cette paramétrisation est telle qu'il existe théoriquement plusieurs solutions, équivalentes compte tenu de l'indétermination précitée:

$$C = AP.[\text{Diag}(AP)]^{-1} - I; \quad P \text{ permutation.}$$

1.2. Paramétrisation proposée

Dans notre approche [3], il semble plus naturel d'imposer

$$E\{s(t) s(t)^H\} = I. \quad (4)$$

L'indétermination qui subsiste sous cette contrainte est réduite à:

$$s(t) = \Delta P x(t), \quad \Delta \triangleq \text{Diag}(\pm 1). \quad (5)$$

Cette contrainte a l'avantage de rendre l'algorithme de séparation robuste vis-à-vis d'un mauvais conditionnement des amplitudes



Cette relation est due au fait que tous les cumulants croisés de $x(t)$ sont nuls, et que $e(t)$ est une transformée linéaire de $x(t)$. Dans le cas d'un traitement par paires en cascade, on peut exprimer explicitement cette relation conformément à la propriété 2.

Propriété 2

Pour $p=2$, posons $\sigma = (\gamma_{04} + \gamma_{40})/\gamma_{22}$ et $\rho = (\gamma_{13} - \gamma_{31})/\gamma_{22}$. Alors la relation suivante est toujours vérifiée:

$$\rho^2 - \sigma + 2 = 0 \quad (11)$$

Propriété 3

Γ_{31} et Γ_{13} vérifient $\Gamma_{ij}(\theta) = -\theta^4 \Gamma_{ij}(-1/\theta)$, et Γ_{22} vérifie $\Gamma_{22}(\theta) = \theta^4 \Gamma_{22}(-1/\theta)$. (12)

Corollaire 4

Les racines θ_c et $-1/\theta_c$ fournissent respectivement les solutions

$$\{Q, -Q\} \text{ et } \{\Delta P Q, -\Delta P Q\}, \quad (13)$$

où

$$\Delta = \begin{pmatrix} -10 & \\ 0 & 1 \end{pmatrix} \text{ et } P = \begin{pmatrix} 01 & \\ & 10 \end{pmatrix}.$$

D'après (12), le système (10) se réduit à deux équations indépendantes; (13) montre qu'il est nécessaire et suffisant de chercher les racines θ dans $]-1, 1[$; finalement (11) démontre que, seule, la dernière équation de (10) suffit à conclure puisque le polynôme de degré 4 dégénère en carré d'un polynôme de degré 2, dont presque sûrement une seule racine est dans $]-1, 1[$ d'après (12). Dans la pratique, à cause des erreurs commises dans l'estimation des cumulants, le polynôme $\Gamma_{22}(\theta)$ peut avoir 4 racines réelles, ou aucune. La meilleure approximation de la solution est obtenue en cherchant le minimum de $\Gamma_{22}(\theta)/\gamma_{22}$ dans $]-1, 1[$. On peut montrer que cette solution θ_0 est alors donnée par:

$$\theta_0 = \rho/2 - \text{sign}(\rho) \sqrt{\rho^2/4 + 1}. \quad (14)$$

On vérifie aussi expérimentalement que θ_0 est moins sensible aux erreurs d'estimation que la racine de $\Gamma_{13}(\theta)$ la plus proche.

2.5. Décorrélateur adaptatif

Formons la matrice $M(t)$:

$$M(t) = \begin{pmatrix} \beta(t) L(t-1) & \alpha(t) e(t) \\ 0 & 1/\alpha(t) \end{pmatrix}. \quad (15)$$

Supposons que l'on trouve une matrice de rotation Π , produit d'une séquence de p rotations de Givens, qui annule les p premières composantes de la dernière colonne. Alors nous avons

$$M(t) \Pi = \begin{pmatrix} L(t) & 0 \\ \tilde{e}^n & \tilde{n} \end{pmatrix}. \quad (16)$$

On peut vérifier que, dans ces conditions, nous avons:

$$L(t) L(t)^n = \beta(t)^2 L(t-1) L(t-1)^n + \alpha(t)^2 e(t) e(t)^n$$

et

$$L(t) \tilde{e}(t) = e(t).$$

Autrement dit, l'opération (16) permet à la fois (i) de remettre à jour le facteur de Cholesky $L(t)$ de la matrice de covariance de $e(t)$, sans la former explicitement, donc en évitant la perte de précision qui

serait entraînée par une quadrature des données (ii) d'obtenir directement la solution $\tilde{e}(t)$ du système $L(t) X = e(t)$ sans pour autant résoudre le système triangulaire par rétro-substitution. De plus, ce type d'algorithme peut être implanté en parallèle [6] en raison de la grande régularité des calculs et du flot de données.

Notons que l'estimateur choisi pour la matrice de corrélation fait intervenir une pondération de type "exponentiel". On doit avoir en régime permanent des valeurs constantes pour α et β , qui doivent satisfaire $\alpha^2 + \beta^2 = 1$ pour obtenir un estimateur non biaisé; α (ou β) fixe la longueur de la mémoire et est choisi a priori en fonction de la durée de stationnarité des phénomènes. En pratique, $L(0)$ sera initialisée par une matrice diagonale de petite norme, et on prendra pendant une phase transitoire d'adaptation $\alpha(t)^2 = 1/t$, par exemple pour $1 \leq t \leq \text{Ent}\{1/\alpha^2\}$ ($1/\alpha^2$ représente le nombre de moyennes équipondérées équivalent dans le cas Gaussien).

2.6. Algorithme adaptatif de rotation AROT

Les signaux sont supposé centrés. L'algorithme complet procède comme suit.

Pour $1 \leq i \leq p$ et pour $i < j \leq p$:

- 1) isoler la paire $E_1 \cong \tilde{e}_1(t)$ et $E_2 \cong \tilde{e}_j(t)$
- 2) Calculer les 3 moments d'ordre deux m_{20} , m_{11} , et m_{02} d'après:

$$m_{qr}(t) = \beta(t)^2 m_{qr}(t) + \alpha(t)^2 E_1^q E_2^r, \quad \{q,r\} \in \{0,1,2\}, \quad q+r=2.$$
- 3) Calculer les trois moments croisés d'ordre 4, m_{31} , m_{22} , et m_{13} :

$$m_{qr}(t) = \beta(t)^2 m_{qr}(t) + \alpha(t)^2 E_1^q E_2^r, \quad \{q,r\} \in \{0,1,2,3,4\}, \quad q+r=4.$$
- 4) En déduire les trois cumulants croisés d'ordre 4 d'après les relations:

$$\begin{aligned} \gamma_{31}(s_1, s_2) &= m_{31}(s_1, s_2) - 3 m_{20}(s_1, s_2) m_{11}(s_1, s_2) \\ \gamma_{22}(s_1, s_2) &= m_{22}(s_1, s_2) - m_{20}(s_1, s_2) m_{02}(s_1, s_2) - 2 m_{11}(s_1, s_2)^2 \\ \gamma_{13}(s_1, s_2) &= m_{13}(s_1, s_2) - 3 m_{02}(s_1, s_2) m_{11}(s_1, s_2). \end{aligned}$$
- 5) Calculer $\rho = (\gamma_{13} - \gamma_{31})/\gamma_{22}$, puis θ_0 conformément à (14).
- 6) Appliquer la transformation Q_{ij}^n au couple E_1, E_2 .

Dans cet algorithme, il peut paraître superflu de recalculer les moments d'ordre 2 en 2). En réalité, ce calcul est nécessaire pendant les périodes transitoires où le décorrélateur n'a pas encore fait converger (m_{20}, m_{11}, m_{02}) vers $(1, 0, 1)$.

3. Méthode JH: obtention et convergence

Une méthode de séparation a été proposée par Jutten [1] [2], et utilise un réseau de neurones linéaires monocouche totalement interconnecté, dont les poids sont contrôlés par un algorithme qui s'apparente à celui de l'itération stochastique. Nous donnons les justifications de cette procédure dans [3], et expliquons son fonctionnement qui était resté incompris jusqu'à présent. L'algorithme de séparation permet de mettre en oeuvre ce qui est baptisé l'analyse en composantes indépendantes [1]. L'itération est définie par

$$C(t+1) = C(t) + \mu \begin{pmatrix} 0 & s_1^3(t) \text{atan}[10 s_2(t)] \\ s_2^3(t) \text{atan}[10 s_1(t)] & 0 \end{pmatrix},$$

$$s(t+1) = [I + C(t+1)]^{-1} e(t).$$

En résumé, l'algorithme JH fonctionne car il fait appel à des fonctions fortement non linéaires, ici $\text{atan} x$ et x^3 , introduisant ainsi la contribution de moments d'ordre supérieur dans la fonctionnelle à annuler. Toutefois, il est nécessaire de restreindre cette combinaison aux moments pouvant effectivement s'annuler (m_{22} n'en fait par exemple pas partie puisque les fonctions $\text{atan} x$ et x^3 sont impaires).



4. Simulations

Les performances de *AROT* ont été analysées dans le cas de sources déterministes, aléatoires non Gaussiennes, puis Gaussiennes. On observe le même phénomène que dans la technique de Jutten et Herault, *JH*, à savoir que la séparation du mélange est possible même pour des signaux Gaussiens pourvu que la transformation *A* soit légèrement bruitée ou non stationnaire.

On présente ci-contre l'évolution des paramètres inconnus en fonction du nombre d'échantillons observés sur un exemple où nous avons $p=2$ sources aléatoires non Gaussiennes. La source $x_1(t)$ est trinaire et prend aléatoirement les valeurs:

$\{-\sqrt{2}, 0, +\sqrt{2}\}$ avec les probabilités $\{1/4, 1/2, 1/4\}$.

La source $x_2(t)$ est binaire et prend les valeurs:

$\{-\sqrt{p_+ / p_-}, \sqrt{p_- / p_+}\}$ avec les probabilités $\{p_-, p_+\}$;

dans notre exemple, $p_+=0.3$ et $p_-=0.7$. Les deux sources sont donc centrées et de variance unité. Afin de n'observer que le comportement de *AROT*, et non celui du décorrélateur, *A* a été choisie ici orthogonale et vaut

$$A = \begin{pmatrix} 0.32 & 0.95 \\ -0.95 & 0.32 \end{pmatrix}$$

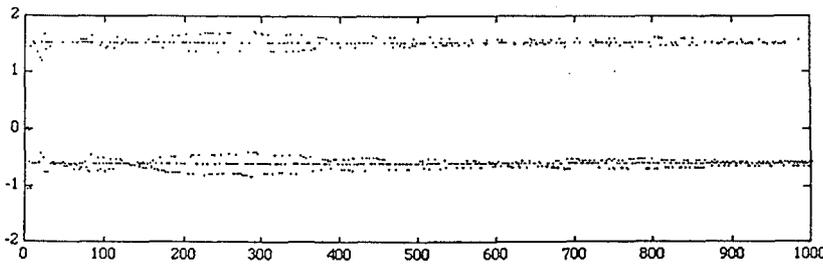
Les valeurs asymptotiques idéales sont $C_{12}=-1/3$, $C_{21}=1/3$, et $\theta=1/3$. Le comportement est en réalité similaire dans des cas plus généraux: la convergence est légèrement plus longue pour *AROT*, et peut être beaucoup plus longue -voire prohibitive- pour *JH*. Pour des signaux déterministes, la convergence est moins cahotique; pour des signaux plus Gaussiens, elle l'est plus.

5. Conclusions

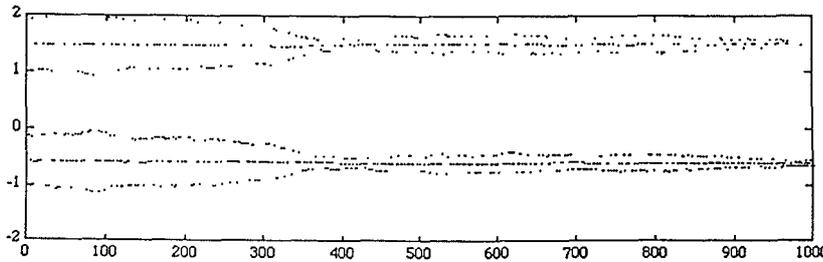
Notre algorithme *AROT* présente l'avantage d'être direct, donc de converger toujours (ceci n'est pas le cas pour l'algorithme *JH*). Il en résulte une meilleure fiabilité, une convergence plus rapide, et une insensibilité à la valeur initiale. Outre le fait qu'il est itératif, donc essentiellement plus lent, *JH* est pénalisé par l'impossibilité de contrôler séparément la longueur du pas de l'itération stochastique, et la longueur de sa mémoire. De plus, sa convergence dépend aussi de la valeur initiale de *C*; aucune valeur initiale (y compris $C=0$) n'est totalement satisfaisante. D'autre part, *AROT* est bien conditionné grâce à une normalisation préalable réalisée par un décorrélateur, dont peut aussi profiter *JH*. Des extensions peuvent être envisagées pour des transformations linéaires stationnaires non instantanées (déconvolution).

Références

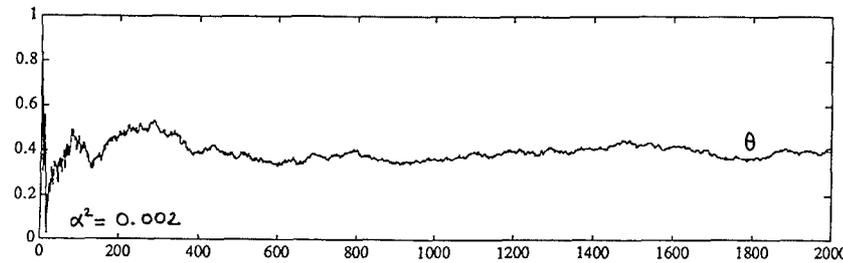
[1] C. JUTTEN, *Calcul Neuromimétique et Traitement du Signal, Analyse en Composantes Indépendantes*, Thèse d'Etat, USMG-INPG, Grenoble, 1987.
 [2] J. HERAULT et al, *Colloque GRETSI*, mai 1985, Nice, 1017-1022.
 [3] P.COMON, *NATO Workshop on Neuro Computing*, Les Arcs, feb 27-march 3, 1989, to be published by Springer Verlag, 1989.
 [5] P.COMON, *Workshop on High order Spectral Analysis*, Vail, Colorado, june 28-30, 1989.
 [6] J.G. McWHIRTER and T.J.SHEPHERD, *Proc. SPIE*, vol.975, Advanced Algorithms and Architectures, San Diego, 1988.



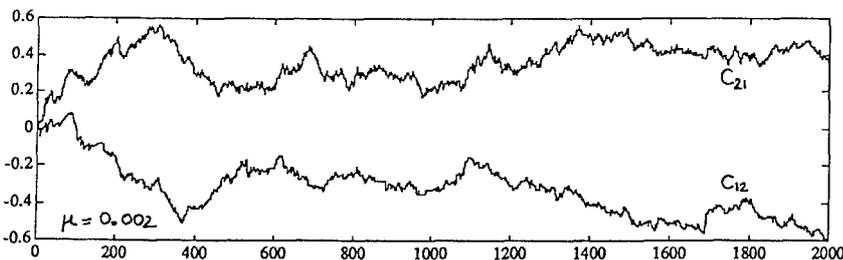
Algorithme *AROT*
Estimation de source, $s_1(t)$



Algorithme *JH*
Estimation de source, $s_1(t)$



Algorithme *AROT*
Estimation de la transformation linéaire



Algorithme *JH*
Estimation de la transformation linéaire