

Séparation de sources non stationnaires

Jean-François CARDOSO

ENST, département TSI
46, rue Barrault 75634 Paris Cedex 13
cardoso@tsi.enst.fr

Résumé — Nous considérons le problème de la séparation aveugle d'un mélange instantané de sources non stationnaires, éventuellement gaussiennes. La borne de Cramér-Rao asymptotique est calculée dans le cas où les distributions des sources sont connues ce qui permet de quantifier l'influence de la non stationnarité sur les performances optimales. Nous décrivons ensuite un algorithme adaptatif simple qui permet la séparation de sources gaussiennes en estimant simultanément la matrice de mélange et le profil de variance des signaux sources.

Abstract — We consider the blind separation of an instantaneous mixture of non-stationary source signals, possibly normally distributed. The asymptotic Cramér-Rao bound is exhibited in the case of known source distributions, showing in a quantitative manner how the non stationnarity governs the achievable performance. Finally a simple adaptive source separation algorithm is presented which is able to separate non stationary Gaussian sources.

1 Introduction

Le problème de la séparation aveugle de mélanges instantanés est généralement traité en exploitant la non gaussianité (éventuelle) des signaux sources [3]. Il existe cependant d'autres possibilités comme par exemple l'exploitation des différences spectrales [4, 1]. Dans cet article, nous considérons l'exploitation de la *non stationnarité* éventuelle des signaux sources. Dans ce cas, il est possible d'obtenir la séparation de signaux gaussiens même spectralement blancs.

Une première section décrit le modèle (très simple) considéré et donne les équations du maximum de vraisemblance. Une seconde section donne les résultats d'une analyse asymptotique dans le cas où les distributions (non stationnaires) des sources sont connues. Ce cas ne se présente jamais en pratique mais permet d'obtenir une borne aux performances en montrant comment un indice de non stationnarité contrôle la borne de Cramér-Rao. Dans une dernière section, nous montrons comment la solution du maximum de vraisemblance pour des distributions connues suggère une famille d'algorithmes adaptatifs capables d'obtenir la séparation de sources gaussiennes pour non stationnarité inconnue mais à variation suffisamment lente.

2 Modèle.

Nous considérons un modèle de mélange instantané où T échantillons d'un n -vecteur aléatoire $\mathbf{x}(t)$ sont représentés comme le mélange par une matrice A inversible $n \times n$ de

T échantillons d'un vecteur $\mathbf{s}(t)$ de 'signaux sources' :

$$\mathbf{x}(t) = A\mathbf{s}(t), \quad 1 \leq t \leq T. \quad (1)$$

La non-stationnarité de la séquence des signaux sources permet la séparation aveugle même lorsque ceux ci sont distribués selon une loi normale. Il est facile de s'en persuader dans des cas simples. Supposons par exemple l'existence de deux régime successifs: pendant une première période, la matrice de covariance des sources est une matrice diagonale Δ_1 puis, pendant une seconde période, c'est une matrice diagonale *différente* Δ_2 . Si ces deux périodes sont connues, on peut calculer sur chacune d'elles la matrice de covariance des observations, fournissant des estimées de $R_1 = A\Delta_1A^\dagger$ puis de $R_2 = A\Delta_2A^\dagger$. Cette structure détermine la matrice A puisque les colonnes de A sont les vecteurs propres de $R_1R_2^{-1}$. Ces derniers sont uniques (aux indéterminations habituelles près: permutation, signes et échelle). En notant $B = A^{-1}$, on remarque aussi que $BR_iB^\dagger = \Delta_i$: la matrice B diagonalise conjointement les deux matrices de corrélation. En fait, il est possible de reformuler la log-vraisemblance gaussienne sous la forme d'un critère de diagonalisation conjointe des covariances empiriques [6].

Dans cet article, nous considérons la solution du maximum de vraisemblance en supposant connues les distributions des sources. Le modèle est le suivant. Le processus $\{\mathbf{s}(t)\}$ n'est *pas* supposé stationnaire; nous considérons ici le cas le plus simple d'une distribution 't.i.d.' (sources Temporellement Indépendamment Distribuées), c'est-à-dire que $\mathbf{s}(t)$ est distribué indépendamment de $\mathbf{s}(t')$ si $t \neq t'$. De plus, nous supposons classiquement l'indépendance des composantes de $\mathbf{s}(t)$: en notant r_{ti} la densité de

la i -ième composante de $\mathbf{s}(t)$, la densité de l'échantillon $\mathbf{s}(1), \dots, \mathbf{s}(T)$ est le produit:

$$p(\mathbf{s}(1), \dots, \mathbf{s}(T)) = \prod_{t=1}^T \prod_{i=1}^n r_{ti}(s_i(t)) \quad (2)$$

Dans ce modèle t.i.d., le gradient relatif [2] de la log-vraisemblance est donné par

$$-\nabla \log p(\mathbf{x}(1), \dots, \mathbf{x}(T)|A) = \sum_{t=1}^T \phi_t(\mathbf{y}(t))\mathbf{y}(t)^\dagger - I \quad (3)$$

avec $\mathbf{y}(t) = A^{-1}\mathbf{x}(t)$ et $[\phi_t(\mathbf{y})]_i = -\frac{r'_{ti}(y_i)}{r_{ti}(y_i)}$ en généralisation directe du cas i.i.d.

L'algorithme du gradient relatif pour maximiser la vraisemblance d'un séparateur B dans le modèle gaussien non-stationnaire s'écrit donc, de nouveau en généralisation directe du cas stationnaire:

1. Initialisation : $B = I$ (par exemple).
2. Calcul du gradient relatif G de la log-vraisemblance

$$G(B) = \frac{1}{T} \sum_{t=1}^T \phi_t(\mathbf{y}(t))\mathbf{y}(t)^\dagger - I \quad (4)$$

3. Si $G(B)$ est 'petit', arrête; sinon mise à jour relative de la matrice de séparation :

$$B \leftarrow (I - \mu G(B))B \quad (5)$$

et retour à l'étape 2

Dans ce qui suit, nous serons particulièrement intéressés par le cas gaussien : le i -ème signal à l'instant t est tiré sous une loi gaussienne de moyenne nulle et de variance σ_{ti}^2 de sorte que $\mathbf{s}(t) \sim \mathcal{N}(0, \Delta_t)$ où

$$\Delta_t = \text{diag}(\sigma_{t1}^2, \dots, \sigma_{tn}^2) \quad (6)$$

et donc :

$$[\phi_t(\mathbf{y})]_i = \frac{y_i}{\sigma_{ti}^2} \quad \text{ou encore} \quad \phi_t(\mathbf{y}) = \Delta_t^{-1}\mathbf{y}. \quad (7)$$

Combinant (3) et (7), les points stationnaires de la vraisemblance sont, dans le cas gaussien, les solutions des équations d'estimation :

$$\frac{1}{T} \sum_{t=1}^T \frac{y_i(t)y_j(t)}{\sigma_{ti}^2} - \delta_{ij} = 0 \quad 1 \leq i, j \leq n. \quad (8)$$

Bien entendu, si les variances sont supposées constantes ($\sigma_{ti} = \sigma_i$), ces équations sont deviennent redondantes : le (i, j) -ème terme fournit la même condition que le (j, i) -ème terme et le modèle n'est pas identifiable. Par contre, dans le cas non-stationnaire, ces deux conditions sont *a priori* distinctes et le modèle doit être identifiable en général.

3 Borne de Cramér-Rao en non stationnaire

Dans cette section, nous donnons la variance asymptotique de l'estimateur du maximum de vraisemblance dans le cas t.i.d. en supposant connues les distributions des sources à chaque instant. Ceci permet de trouver la borne de Cramér-Rao et donne donc une idée de l'information de Fisher apportée par la non stationnarité (en supposant cette dernière connue).

Moyennes non-stationnaires. Les calculs asymptotiques du cas stationnaire tels qu'ils sont présentés par Pham dans [5] peuvent être étendus au cas non stationnaire t.i.d. Cependant un certain nombre de moments doivent recevoir une définition appropriée, en remplaçant les espérances mathématiques E par des limites de moyennes. Ceci sera noté par l'opérateur \bar{E} : si $\{X_t\}$ est une séquence de variables aléatoires pour laquelle la distribution de X_t dépend de t , nous notons

$$\bar{E}\{X\} \stackrel{\text{def}}{=} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T EX_t \quad (9)$$

en supposant que la somme converge effectivement. En particulier la puissance moyenne du i -ème signal est $\bar{E}\{s_i^2\} = \lim \frac{1}{T} \sum_t s_i(t)^2$.

Taux de réjection. La qualité d'une estimée \hat{A} en termes de séparation de sources peut être mesurée par le 'taux de réjection' associé. Si le vecteur des sources est estimé par $\hat{\mathbf{s}} = \hat{A}^{-1}\mathbf{x}$, alors la puissance moyenne de la j -ème source dans l'estimée de la i -ème source est donnée par $|\hat{A}^{-1}A]_{ij}|^2 \bar{E}\{s_j^2\}$ tandis que la puissance moyenne de la i -ème source est $|\hat{A}^{-1}A]_{ii}|^2 \bar{E}\{s_i^2\}$. Pour un estimateur régulier, la variance asymptotique de $[\hat{A}^{-1}A]_{ij}$ décroît comme T^{-1} pour $i \neq j$ tandis que $[\hat{A}^{-1}A]_{ii}$ converge vers une constante (égale à 1 si la distribution des sources est correctement spécifiée dans le modèle). Nous définissons

$$\rho_{ij} \stackrel{\text{def}}{=} \lim_{T \rightarrow \infty} TE \left(\frac{|\hat{A}^{-1}A]_{ij}|^2}{|\hat{A}^{-1}A]_{ii}|^2} \right) \frac{\bar{E}\{s_j^2\}}{\bar{E}\{s_i^2\}}. \quad (10)$$

pour exprimer le 'taux de réjection asymptotique' (de la j -ème source dans l'estimée de la i -ème source).

Indépendance des non stationnarités. Les expressions des taux de réjection asymptotiques ρ_{ij} peuvent être obtenues dans le cas non-stationnaire par la voie habituelle mais les résultats pour une paire de sources (i, j) ne se prêtent à une interprétation simple que lorsque la condition

$$\bar{E}\{\phi'_i(s_i)s_j^2\} = \bar{E}\{\phi'_i(s_i)\} \bar{E}\{s_j^2\} \quad (11)$$

est vérifiée. Cette condition s'interprète comme une sorte 'd'indépendance des non-stationnarités' en ce sens qu'elle exprime une 'décorrélation empirique' entre la séquence

des puissances de la j -ème source et la séquence $\{E\phi'_i(s_i(t))\}$ de la i -ème source.

Résultats asymptotiques Il s'avère que le comportement asymptotique dépend des quantités R_1, \dots, R_n définies par

$$R_i \stackrel{\text{def}}{=} \bar{E} \{ \phi'_i(s_i) \} \bar{E} \{ s_i^2 \}. \quad (12)$$

Sous l'hypothèse t.i.d. et sous celle d'indépendance des non-stationnarités introduite ci-dessus, la borne de Cramér-Rao s'obtient en calculant les taux de réjection asymptotiques de l'estimateur du maximum de vraisemblance. Le calcul donne

$$\rho_{ij} = \frac{R_j}{R_i R_j - 1}. \quad (13)$$

Ce résultat s'obtient sans supposer la normalité des signaux sources : il permet donc d'unifier la description des performances dans le cadre non-gaussien et dans le cadre non-stationnaire.

En effet, dans le cas *stationnaire non-gaussien*, la quantité R_i se réduit à $E\phi'_i(s_i)Es_i^2$: on retrouve les résultats de Pham [5] et R_i s'interprète comme une mesure de non-Gaussianité. Plus précisément, on a $R_i \geq 1$ dans le cas stationnaire avec égalité seulement si et seulement si s_i présente une distribution gaussienne. Les performances sont d'autant meilleures que les distributions des sources sont éloignées de la gaussienne et ceci est mesuré par l'écart de R_i à 1.

Dans le cas *non stationnaire gaussien*, le score est donné par (7) de sorte que $\phi'_i(s_i) = \sigma_{ti}^{-2}$. Par conséquent, la quantité R_i s'exprime comme

$$R_i = \bar{E}\sigma_{ti}^2 \bar{E}\sigma_{ti}^{-2} \geq 1. \quad (14)$$

Dans ce cas, l'égalité $R_i = 1$ est atteinte lorsque la séquence $\{\sigma_{ti}^2\}$ des variances est constante et c'est donc l'écart de R_i à la valeur 1 qui mesure (asymptotiquement) la capacité de la non stationnarité à permettre la séparation.

4 Une famille d'algorithmes

Dans la pratique, les distributions non stationnaires des sources ne sont pas connues : l'algorithme du maximum de vraisemblance résumé par les équations (4) et (5) ne peut être implanté tel quel.

Dans le cas stationnaire, il est certainement possible d'estimer les distributions de chaque source par exemple en adoptant la méthode de Pham [5] (qui estime non pas la distribution de chaque source mais la fonction score associée). Dans le cas non stationnaire, il n'est pas possible —sans hypothèses de régularité— d'estimer les distributions des sources.

Une solution possible est de supposer une certaine régularité dans l'évolution des distributions des sources : on fera l'hypothèse que les variances σ_{ti}^2 évoluent lentement en fonction du temps. Dès lors il est possible d'estimer les

variances instantanées, par exemple par un simple lissage temporel du carré y_{ti}^2 des sorties du séparateur. Si l'on adopte en outre un modèle gaussien, la connaissance des variances suffit à spécifier la distribution des sources (en supposant des sources de moyenne nulle).

L'approche proposée dans cet article consiste donc à modifier l'algorithme du gradient relatif (résumé par les équations (4) et (5)) en utilisant un score gaussien estimé, c'est-à-dire en prenant :

$$[\phi_t(\mathbf{y})]_i = \frac{y_i}{\hat{\sigma}_{ti}^2} \quad (15)$$

où, à chaque itération, les valeurs de $\hat{\sigma}_{ti}^2$ sont obtenues par filtrage passe-bas de $y_i(t)^2$.

Un exemple La figure 1 présente un exemple de trajectoire d'un tel algorithme où les variances de chaque source varient sur une trajectoire sinusoïdale comprise entre les valeurs 0.5 et 2. Ces variances σ_{ti}^2 sont estimées par la moyenne de $y_i(t)^2$ sur les échantillons entre $t-50$ et $t+50$.

Il faut noter que, dans cette approche, rien n'a été spécifié pour contrôler l'amplitude des sorties ou celle des composantes de la matrice de séparation: dans l'algorithme tel qu'il vient d'être décrit, si B est un point stationnaire alors ΛB en est aussi un si Λ est une matrice diagonale. Ceci est potentiellement dangereux (en termes de stabilité) dans le cas où l'on travaille à pas d'adaptation constant. Pour contrôler l'amplitude des sorties, une approche possible est de normaliser les fonctions scores par l'opération

$$\hat{\sigma}_{ti}^2 \rightarrow \frac{\hat{\sigma}_{ti}^2}{T^{-1} \sum_{t=1}^T \hat{\sigma}_{ti}^2} \quad 1 \leq i \leq n \quad (16)$$

ce qui a pour effet de produire des signaux sources estimés de variance moyenne égale à 1. C'est cette solution qui a été adoptée pour l'algorithme illustré par la figure 1.

5 Discussion

C'est la variation lente des variances des sources qui permet d'obtenir une estimation raisonnable de ces variances et donc de mettre en oeuvre l'algorithme présenté ci-dessus. Une situation opposée serait celle où à chaque instant les variances σ_{ti}^2 seraient tirés indépendamment au hasard. Dans ce cas, la distribution des sources devient —de fait— celle d'un mélange de gaussiennes qui est bien entendu une distribution *non* gaussienne. On voit que la notion de séquence non stationnaire de distributions gaussiennes est en fait une notion floue qui n'est bien cernée que dans la limite des variations lentes. Nous présenterons pendant la conférence d'autres résultats mettant en évidence la 'continuité' avec le cas non gaussien stationnaire.

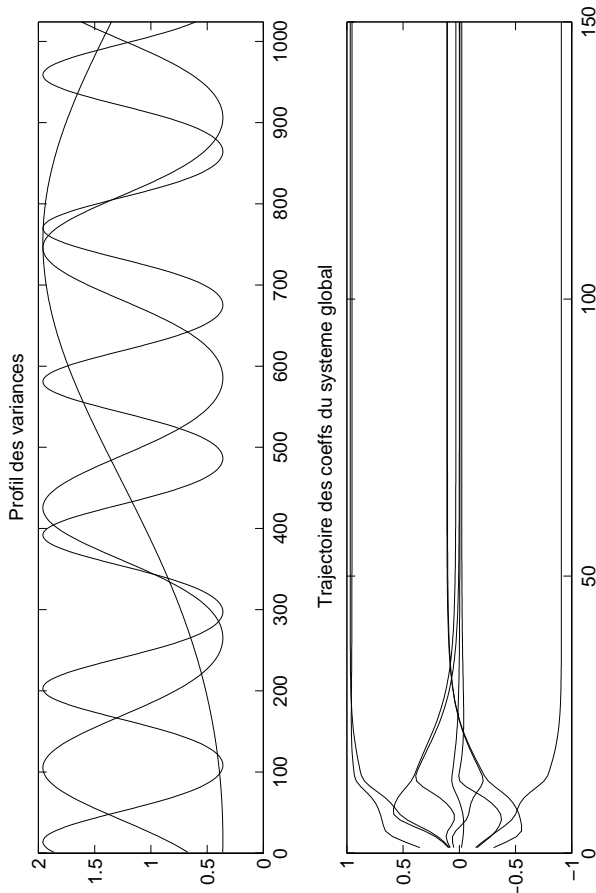


FIG. 1: Une trajectoire de l'algorithme du maximum de vraisemblance gaussien avec trois sources et trois capteurs. Les réalisations des sources sont obtenues en tirant $s_i(t)$ sous la loi $\mathcal{N}(0, \sigma_{i_i}^2)$ pour $1 \leq t \leq T = 1000$. La variation temporelle des variances de chaque source est choisie selon une loi déterministe: variation sinusoïdale représentée dans le premier panneau. Le second panneau montre l'évolution des 9 coefficients du système global BA . Trois de ces coefficients convergent vers une valeur proche de 1 tandis que 6 autres convergent vers des valeurs beaucoup plus faibles, mettant en évidence la convergence vers un séparateur en une trentaine d'itérations. Le pas d'adaptation choisi est $\mu = 0.5$, les variances σ_{ti}^2 des sources sont estimées par moyennage de $y_i(t)^2$ sur une fenêtre rectangulaire de 100 échantillons.

References

- [1] Adel Belouchrani, Karim Abed Meraim, Jean-François Cardoso, and Éric Moulines. A blind source separation technique based on second order statistics. *IEEE Trans. on Sig. Proc.*, 45(2):434–44, February 1997.
- [2] Jean-François Cardoso and Beate Laheld. Equivariant adaptive source separation. *IEEE Trans. on Sig. Proc.*, 44(12):3017–3030, December 1996.
- [3] P. Comon. Independent component analysis, a new concept ? *Signal Processing, Elsevier*, 36(3):287–314, April 1994. Special issue on Higher-Order Statistics.
- [4] Dinh-Tuan Pham and Philippe Garat. Séparation aveugle de sources temporellement corrélées. In *Proc. GRETSI*, pages 317–320, 1993.
- [5] Dinh-Tuan Pham and Philippe Garat. Blind separation of mixture of independent sources through a quasi-maximum likelihood approach. *IEEE Tr. SP*, 45(7):1712–1725, July 1997.
- [6] D.T. Pham. Joint approximate diagonalization of positive definite hermitian matrices. Technical report, LMC/IMAG, 1999.