

Séparation aveugle de sources temporellement corrélées par l'approche de maximum de vraisemblance

Shahram HOSSEINI¹, Christian JUTTEN¹, Dinh Tuan PHAM², Benoît STOLL¹

¹Laboratoire des Images et des Signaux
46 Avenue Félix Viallet, 38031 Grenoble, France

²Laboratoire de Modélisation et Calcul
BP 53X, 38041 Grenoble cedex 9, France
hosseini@lis.inpg.fr, Christian.Jutten@inpg.fr
Dinh-Tuan.Pham@imag.fr, stoll@lis.inpg.fr

Résumé – Nous utilisons une approche de maximum de vraisemblance pour la séparation aveugle des mélanges instantanés des sources temporellement corrélées sans aucune transformation préalable ni hypothèse *a priori* sur les densités de probabilité de sources. Un modèle de Markov du premier ordre est utilisé pour représenter la densité de probabilité jointe des échantillons successifs de chaque source. Ces densités sont estimées à partir des observations avec la méthode de noyaux.

Abstract – A quasi maximum likelihood approach is used for separating the instantaneous mixtures of temporally correlated, independent sources without any preliminary transformation nor a priori assumption about the probability distribution of the sources. A first order Markov model is used to represent the joint probability density of successive samples of each source. The joint probability density functions are estimated from the observations using a kernel method.

1 Introduction

Dans ce travail, l'approche de maximum de vraisemblance (MV) est utilisée pour la séparation aveugle de mélanges instantanés de sources. Dans le cadre général (sans bruit et avec même nombre de capteurs que de sources), ce problème peut être formulé de manière suivante. Ayant N échantillons d'un mélange instantané de K sources, $\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t)$, où $\mathbf{x}(t) = [x_1(t), x_2(t), \dots, x_K(t)]^T$ et $\mathbf{s}(t) = [s_1(t), s_2(t), \dots, s_K(t)]^T$ sont respectivement les vecteurs des observations et des sources et \mathbf{A} est une matrice inversible, on veut trouver une estimation de la matrice \mathbf{A} (ou son inverse, la matrice de séparation) à une permutation et un facteur d'échelle près.

L'une des approches pouvant être utilisées consiste à maximiser la fonction de vraisemblance des observations, (conditionnée sur la matrice \mathbf{A}). Cette approche a l'avantage de fournir un estimateur asymptotiquement efficace (variance minimum parmi les estimateurs non biaisés). Dans le cas des sources i.i.d., cette approche a été utilisée par Pham et Garat [1]. Ils montrent que la matrice de séparation peut être estimée au sens du maximum de vraisemblance par l'annulation de statistiques d'ordre supérieur de la forme $E[s_i\psi(s_j)]$ où s_i désigne la i -ème source et $\psi(s_j)$ est la fonction score de la j -ème source. Dans le même papier les auteurs proposent une autre méthode, dans le cas des sources temporellement corrélées qui consiste à calculer la transformée de Fourier discrète des sources et à appliquer l'approche de maximum de vraisemblance sur les résultats. Dans [2], les auteurs utilisent également la méthode MV mais ils modélisent les densités de proba-

bilité des sources par un développement de Gram-Charlier à l'ordre 4. Enfin, dans [3], la méthode MV a été utilisée pour la séparation des sources gaussiennes où la corrélation des sources est modélisée par un modèle autorégressif.

Dans ce travail, nous traitons le problème dans le cas des sources temporellement corrélées et notre but est de maximiser directement la fonction de vraisemblance sans aucune transformation préalable ni aucune hypothèse concernant les densités de probabilité des sources. En effet, ces densités seront estimées au cours de la procédure de maximisation avec une approche du type noyaux.

Le reste de ce papier est organisé de manière suivante. Dans la section 2, après l'exposé de problème, nous dérivons la fonction de vraisemblance à maximiser et nous présentons la méthode utilisée pour estimer les densités de probabilité à partir des estimations de sources, et un algorithme itératif pour maximiser la fonction dérivée. Dans la section 3, nos premiers résultats de simulation seront présentés. Enfin, la section 4 est consacrée à la conclusion et aux perspectives du travail.

2 Méthode

2.1 Exposé de problème

Ayant N échantillons d'un vecteur \mathbf{x} de dimension K , résultat d'une transformation linéaire $\mathbf{x} = \mathbf{A}\mathbf{s}$ où \mathbf{s} est un vecteur de composantes indépendantes et éventuellement corrélées dans le temps (les sources), et \mathbf{A} est une ma-

trice $K \times K$ et inversible, notre objectif est de trouver une matrice \mathbf{B} telle que les composantes du vecteur $\mathbf{y} = \mathbf{B}\mathbf{x}$ soient aussi indépendantes que possible (l'indépendance est acquise pour $\mathbf{B} = \mathbf{A}^{-1}$).

La méthode de Maximum de Vraisemblance (MV) consiste à maximiser la densité de probabilité jointe de tous les échantillons de toutes les composantes du vecteur \mathbf{x} (toutes les observations) par rapport à \mathbf{B} :

$$f(x_1(1), \dots, x_K(1), \dots, x_1(N), \dots, x_K(N)) \quad (1)$$

Étant donnée l'indépendance des sources, cette fonction est égale à:

$$\left(\frac{1}{|\det(\mathbf{B}^{-1})|}\right)^N \prod_{i=1}^K f_{s_i}(\mathbf{e}_i^T \mathbf{B}\mathbf{x}(1), \mathbf{e}_i^T \mathbf{B}\mathbf{x}(2), \dots, \mathbf{e}_i^T \mathbf{B}\mathbf{x}(N)) \quad (2)$$

où $f_{s_i}(\cdot)$ représente la densité jointe de N échantillons de la source s_i et \mathbf{e}_i est la colonne i de la matrice identité. Chaque terme $f_{s_i}(\cdot)$ peut être écrit comme le produit des densités conditionnelles. Même si l'approche peut être utilisée dans le cas général, pour simplifier la mise en œuvre, nous supposons pour $f_{s_i}(\cdot)$ un modèle de Markov du premier ordre, c'est à dire:

$$f_{s_i}(s_i(t)|s_i(1), \dots, s_i(t-1)) = f_{s_i}(s_i(t)|s_i(t-1)) \quad (3)$$

La fonction (2) se réduit alors à:

$$\left(\frac{1}{|\det(\mathbf{B}^{-1})|}\right)^N \prod_{i=1}^K [f_{s_i}(\mathbf{e}_i^T \mathbf{B}\mathbf{x}(1)) \prod_{t=2}^N f_{s_i}(\mathbf{e}_i^T \mathbf{B}\mathbf{x}(t)|\mathbf{e}_i^T \mathbf{B}\mathbf{x}(t-1))] \quad (4)$$

En prenant le logarithme de (4), on obtient la fonction de log-vraisemblance dont la maximisation nous permettra d'estimer la matrice de séparation \mathbf{B} . Les densités de probabilité des sources n'étant pas connues *a priori*, on doit les estimer à partir des observations. Même si on peut choisir l'estimation directe des densités conditionnelles, nous avons préféré procéder en estimant les densités jointes. Pour cela, après avoir calculé le logarithme de (4), en utilisant la formule de Bayes, nous remplaçons les densités conditionnelles par les densités jointes divisées par les densités marginales. La fonction de log-vraisemblance s'écrit alors:

$$L = N \log(|\det(\mathbf{B})|) + \sum_{i=1}^K [\log(f_{s_i}(\mathbf{e}_i^T \mathbf{B}\mathbf{x}(1))) + \sum_{t=2}^N \log\left(\frac{f_{s_i, s_i}(\mathbf{e}_i^T \mathbf{B}\mathbf{x}(t), \mathbf{e}_i^T \mathbf{B}\mathbf{x}(t-1))}{f_{s_i}(\mathbf{e}_i^T \mathbf{B}\mathbf{x}(t-1))}\right)] \quad (5)$$

Et après les simplifications, la fonction à maximiser L devient:

$$L = N \log(|\det(\mathbf{B})|) + \sum_{i=1}^K \left[\sum_{t=2}^N \log(f_{s_i, s_i}(\mathbf{e}_i^T \mathbf{B}\mathbf{x}(t), \mathbf{e}_i^T \mathbf{B}\mathbf{x}(t-1))) - \sum_{t=2}^{N-1} \log(f_{s_i}(\mathbf{e}_i^T \mathbf{B}\mathbf{x}(t))) \right] \quad (6)$$

En dérivant (6) par rapport à \mathbf{B} , on peut montrer que la matrice \mathbf{B} est la solution du système des équations sui-

vant¹:

$$\sum_{t=2}^N [\psi_{i,i}^{(1)}(\mathbf{e}_i^T \mathbf{B}\mathbf{x}(t), \mathbf{e}_i^T \mathbf{B}\mathbf{x}(t-1)) \mathbf{e}_j^T \mathbf{B}\mathbf{x}(t) + \psi_{i,i}^{(2)}(\mathbf{e}_i^T \mathbf{B}\mathbf{x}(t), \mathbf{e}_i^T \mathbf{B}\mathbf{x}(t-1)) \mathbf{e}_j^T \mathbf{B}\mathbf{x}(t-1)] - \sum_{t=2}^{N-1} [\psi_i(\mathbf{e}_i^T \mathbf{B}\mathbf{x}(t)) \mathbf{e}_j^T \mathbf{B}\mathbf{x}(t)] = 0 \quad i \neq j = 1, \dots, K \quad (7)$$

où $[\psi_{i,i}^{(1)}, \psi_{i,i}^{(2)}]^T = \nabla \log f_{s_i, s_i}(x, y)$ et $\psi_i(x) = \frac{d \log f_{s_i}(x)}{dx}$. Dans ces dernières expressions, $f_{s_i, s_i}(x, y)$ représente la densité jointe de deux échantillons successifs de la source s_i , et $f_{s_i}(x)$ est la densité marginale de la même source. Le système d'équations ci-dessus peut être résolu en utilisant, par exemple, l'algorithme itératif de Newton-Raphson. Cependant, dans ce papier, nous avons préféré maximiser directement la fonction de log-vraisemblance (6) avec un algorithme du gradient.

2.2 Estimation des densités de probabilité

Dans les relations ci-dessus, les densités de probabilité des sources sont supposées connues, ce qui n'est jamais le cas en pratique. Nous remplaçons donc ces densités par les densités de probabilité des sources estimées $\mathbf{y} = \mathbf{B}\mathbf{x}$. Évidemment, à la convergence, si $\mathbf{B} = \mathbf{A}^{-1}$ ces densités coïncident avec les densités de sources. Afin d'estimer ces densités, nous avons utilisé la méthode de noyaux. Pour l'estimation des densités jointes, étant donné que les échantillons successifs de chaque source sont corrélés, la formule de Fukunaga [4] avec les noyaux gaussiens a été utilisée. Cette méthode a l'avantage de s'adapter aux nuages des données non symétriques, en n'utilisant qu'un seul paramètre de lissage. La densité jointe de deux échantillons successifs de j -ème composante du vecteur \mathbf{y} est ainsi estimée par:

$$\hat{f}(y_j(t), y_j(t-1)) = \frac{(\det \mathbf{V}_j)^{-1/2}}{(N-1)h^2} \sum_{i=2}^N \frac{1}{2\pi} \exp\left(\frac{-1}{h^2}[y_j(t) - y_j(i), y_j(t-1) - y_j(i-1)]^T \mathbf{V}_j^{-1} [y_j(t) - y_j(i), y_j(t-1) - y_j(i-1)]\right) \quad (8)$$

Dans cette formule, \mathbf{V} représente la matrice de covariance de deux échantillons successifs de y_j , et h est le paramètre de lissage déterminant la largeur de chaque noyau gaussien et sa valeur optimale est: $h_{opt} = 0.96(N-1)^{-1/6}$.

Pour estimer les densités marginales, on peut procéder à l'intégration des densités jointes. Nous avons néanmoins préféré les estimer séparément, et avec la formule ci-dessous:

$$\hat{f}(y_j(t)) = \frac{1}{Nh} \sum_{i=1}^N \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-1}{2h^2}(y_j(t) - y_j(i))^2\right) \quad (9)$$

1. Il y a au total $K(K-1)$ équations. Les K autres équations nécessaires pour estimer les K^2 éléments de la matrice \mathbf{B} ne servent qu'à lever les indéterminations due au facteur d'échelle, et peuvent être choisies d'une façon plus ou moins arbitraire, selon la normalisation utilisée.

La valeur optimale de h dans cette dernière formule est: $h_{opt} = 1.06\sigma N^{-1/5}$ où σ représente l'écart-type de y_j .

2.3 Algorithme

L'estimation de la matrice \mathbf{B} s'effectue avec une approche itérative du type bloc. A chaque itération, en utilisant la valeur actuelle de la matrice \mathbf{B} , on estime les densités jointes et les densités marginales des signaux dé-mêlés avec les formules (8) et (9), on les remplace dans la fonction de log-vraisemblance (6) et on adapte la matrice \mathbf{B} pour maximiser cette fonction avec la méthode du gradient:

$$\mathbf{B} = \mathbf{B} + \mu \nabla_{\mathbf{B}} L \quad (10)$$

L'expérience montre que l'estimation des densités à partir des échantillons est très sensible à la normalisation des données. Ainsi, toutes les méthodes classiques utilisées pour lever les indéterminations ne sont pas équivalentes. Par exemple, si on contraint la matrice de séparation \mathbf{B} à avoir des "un" sur sa diagonale principale, le vecteur \mathbf{y} peut prendre des valeurs complètement aberrantes au début de l'optimisation, entraînant ainsi une estimation aberrante des densités qui peut aboutir à la divergence. En revanche, si on normalise les lignes de la matrice \mathbf{B} au début de chaque itération, ce problème ne se pose plus car le vecteur \mathbf{y} et par conséquent les estimations de densité prennent toujours les valeurs raisonnables. Avec ce constat, et pour lever les indéterminations d'échelle, de permutation et de signe, nous modifions la matrice \mathbf{B} au début de chaque itération de la manière suivante:

1. normer les lignes de \mathbf{B} à un.
2. permuter les lignes jusqu'à rendre maximale la somme des carrés des termes diagonaux de \mathbf{B} .
3. inverser si nécessaire les signes des termes d'une ligne, pour n'avoir sur la diagonale principale de \mathbf{B} que des termes positifs ou zéros.

Cette normalisation améliore considérablement la convergence de l'algorithme mais elle ne résout pas complètement le problème.

3 Résultats de simulation

Dans cette section, nous présentons nos premiers résultats de simulation. Il faut avouer que ces expériences ne suffisent pas pour une conclusion définitive. Cependant, elles peuvent confirmer le bon fonctionnement de la méthode proposée.

Notre expérience consiste en la séparation d'un mélange de deux sources indépendantes, chacune représentant une séquence autorégressive du premier ordre. L'une des sources est gaussienne et l'autre uniforme. Les expériences ont été effectuées pour différents nombre d'échantillons (N) et différents coefficients de corrélation de chaque source (ρ_1 et ρ_2). Pour chaque combinaison choisie, 10 expériences, correspondant à 10 différentes valeurs initiales du générateur de bruit sont faites et la moyenne

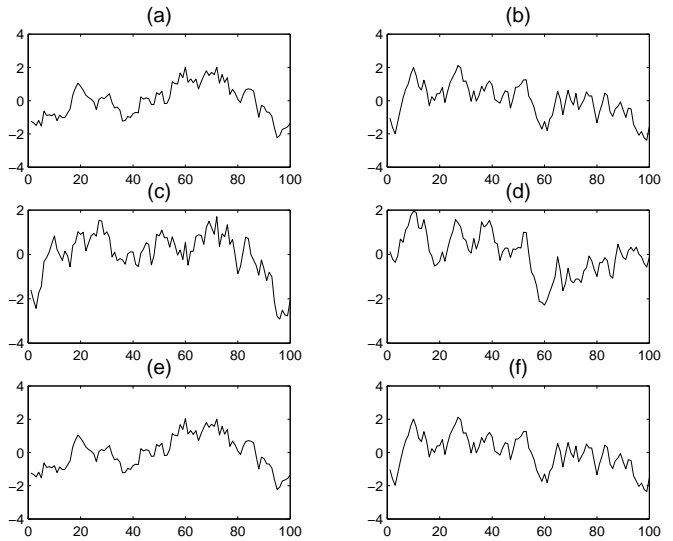


FIG. 1: (a) et (b) Sources, (c) et (d) mélanges, (e) et (f) signaux restitués pour $N = 100$ et $\rho_1 = \rho_2 = 0.9$.

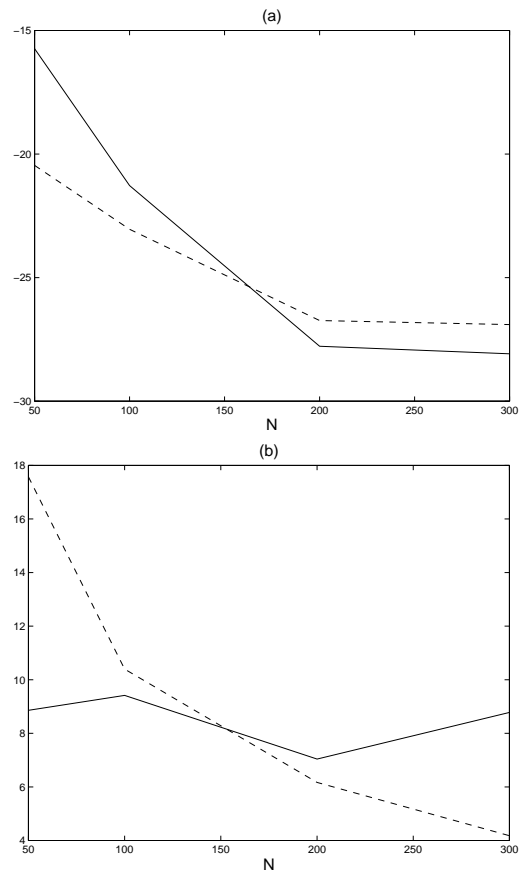


FIG. 2: (a) Moyenne et (b) écart-type de la diaphonie résiduelle pour la source gaussienne (lignes solides) et source uniforme (ligne coupée) par rapport au nombre d'échantillons (N).

et l'écart-type de ces 10 expériences sont présentés. Dans toutes les expériences, la matrice de mélange est: $\mathbf{A} = \begin{pmatrix} \sqrt{2}/2 & \sqrt{2}/2 \\ -\sqrt{2}/2 & \sqrt{2}/2 \end{pmatrix}$. Le résultat d'une expérience avec $N = 100$ et $\rho_1 = \rho_2 = 0.9$ est montré dans la figure 1.

On calcule les performances de l'algorithme en utilisant la diaphonie résiduelle sur les deux voies:

$$C_i = 10 \log_{10} E[(y_i - s_i)^2] \quad i = 1, 2 \quad (11)$$

où y_i et s_i ont une variance unité. La figure 2 montre la moyenne et l'écart-type de ces critères par rapport à N . On peut remarquer que la performance augmente en fonction de nombre des échantillons. Nos expériences montrent aussi qu'en général, la performance est meilleure pour les sources fortement corrélées. Il faut souligner que l'algorithme ne converge pas toujours. La divergence est en général due à la mauvaise estimation de densité au début de l'algorithme en raison de la distance considérable entre \mathbf{y} et \mathbf{s} . Dans le calcul de la moyenne et l'écart-type, seules les solutions convergentes ont été considérées.

4 Conclusion et perspectives

Dans ce papier, nous avons utilisé l'approche de maximum de vraisemblance pour la séparation aveugle des mélanges instantanés des sources temporellement corrélées sans aucune transformation préalable ni hypothèse *a priori* sur les densités de probabilité de sources. Un estimateur de noyaux gaussiens est utilisé pour estimer les densités des sources à partir des observations dans un algorithme itératif. Les premiers résultats confirment le bon fonctionnement de l'approche.

Plusieurs points restent pourtant à améliorer. D'abord, l'algorithme ne converge pas toujours. L'expérience montre que la convergence dépend principalement de la distance entre les signaux restitués ($\mathbf{B}\mathbf{x}$) et les sources (\mathbf{s}) au début de l'algorithme. La convergence est presque toujours acquise si on remplace l'estimateur de densités de $\mathbf{B}\mathbf{x}$ par un estimateur de densités de sources (Autrement dit, si on remplace \mathbf{y} par \mathbf{s} dans les relations (8 et 9)).

Ensuite, l'algorithme est assez lent en raison de coût de calcul élevé des densités de probabilité. Nous travaillons actuellement sur les algorithmes moins coûteux et plus performants de l'estimation de ces densités.

Enfin, l'algorithme du gradient simple utilisé dans ce papier est sensible au pas d'adaptation. Un algorithme du type gradient conjugué pourrait résoudre ce problème.

Les expériences présentées dans ce papier n'avaient pour but que de montrer le bon fonctionnement de la méthode. Plusieurs autres tests semblent nécessaires pour arriver à une conclusion convaincante. La comparaison de cette méthode avec les méthodes traditionnelles permet de vérifier si la modélisation de corrélation apporte une amélioration de performance ou non. On peut aussi tester l'algorithme

sur les corrélations générées par les filtres non linéaires. Nous soulignons que dans la formulation du problème, aucune hypothèse sur la nature des filtres temporels n'a été faite, excepté le modèle de Markov du premier ordre pour simplifier la mise en œuvre. Enfin, on peut envisager de tester l'algorithme avec les signaux réels.

Références

- [1] D. T. Pham et P. Garat. Blind separation of mixture of independent sources through a quasi-maximum likelihood approach. *IEEE Trans. on Signal Processing*, 45(7), July 1997.
- [2] M. Gaeta et J. L. Lacoume. Estimateur du maximum de vraisemblance étendus à la séparation de sources non gaussiennes. *Traitement du Signal*, 7(5): 419-437, 1990.
- [3] A. Zaïdi. *Séparation aveugle d'un mélange instantané de sources autorégressives gaussiennes par la méthode du maximum de vraisemblance*. Thèse de doctorat, soutenue le 14 décembre 2000.
- [4] B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, 1986.