

SÉPARATION ADAPTATIVE DE SOURCES DANS L'ESPACE SIGNAL.

Jean-François Cardoso (cardoso@sig.enst.fr)

Télécom Paris, Département Signal, 46 rue Barrault, 75634 Paris.

Résumé

Nous explorons la possibilité de coupler un algorithme adaptatif de séparation de sources avec un algorithme adaptatif de projection de projection sur l'espace signal afin de tirer parti d'un nombre de capteurs supérieur au nombre de sources. Nous considérons ici un algorithme de projection récemment publié qui met à jour un projecteur par multiplication à droite sans contrainte. En ce sens, cet algorithme est homogène à notre algorithme de séparation dont la mise à jour se fait par multiplication à gauche.

Abstract

To solve the source separation problem, it is possible to take advantage of 'extra' sensors to reduce the effect of noise by projecting the observations onto the 'signal subspace'. In this contribution, we investigate this possibility by coupling an adaptive source separation algorithm with an adaptive subspace tracker. Both stages are of the same nature: adaptive projection on the subspace is achieved by right multiplication of the separating matrix while the update related to source separation is obtained by left multiplication.

1 Introduction.

La séparation de sources consiste à retrouver n signaux statistiquement indépendants à partir de la seule observation de m mélanges instantanés de ces signaux. Cette observation étant éventuellement bruitée, le modèle des signaux capteurs $x(t)$ s'écrit

$$x(t) = As(t) + b(t) \quad (1)$$

où $s(t)$ est un vecteur de taille $n \times 1$ contenant n signaux (dits 'signaux sources') statistiquement indépendants. La matrice A (dite 'de mélange') est de taille $m \times n$ représente la propagation des sources aux m capteurs: elle est supposée fixe et de rang plein à l'exception de toute autre hypothèse. Le bruit additif $b(t)$ est supposé indépendant des signaux sources et spatialement blanc: $R_b \stackrel{\text{def}}{=} Eb(t)b^*(t) = \sigma I_m$. Tous les processus sont stationnaires et de moyenne nulle. Dans la suite nous considérons indifféremment des données réelles ou complexes, les algorithmes présentés ici prenant la même forme dans les deux cas.

Un algorithme adaptatif pour la séparation de sources consiste à définir une règle de mise à jour d'une matrice B_t de taille $n \times m$ dont la sortie $y(t) = B_t x(t)$ fournisse une estimation des signaux sources. La forme générale d'un tel algorithme est

$$\begin{aligned} y_t &= B_t x(t) \\ B_{t+1} &= B_t + \lambda_t f(B_t, x_t) \end{aligned}$$

où f doit être conçue pour faire converger B_t

vers une (pseudo-)inverse de A . La convergence est réglée par la séquence scalaire λ_t .

De nombreux algorithmes adaptatifs pour la séparation de sources ont été présentés récemment dans la littérature, voir par exemple [1, 2, 3]. Le problème est souvent traité dans un cadre simplifié : le bruit peut être négligé ($\sigma = 0$) et le nombre de sources présentes est égal au nombre de mélanges observés ($m = n$). Il est désirable de s'affranchir de ces deux limitations. Dans cet article, nous ne tentons pas d'obtenir un algorithme consistant en présence de bruit mais en nous plaçant précisément dans le cas où le nombre des capteurs est supérieur à celui des sources pour limiter les effets du bruit,

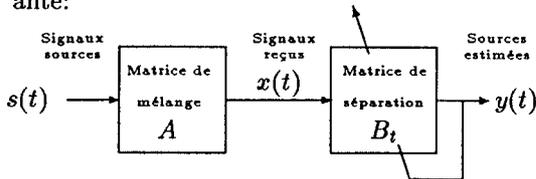
On peut profiter de capteurs surnuméraires pour effectuer une séparation en deux étapes. La première étape consiste à projeter les observations dans l'espace signal: cette opération ramène la dimension des observations à n tout en maximisant le rapport signal à bruit disponible. La seconde étape est une séparation de sources dans l'espace signal offrant l'avantage de travailler en dimension minimale et dans les meilleures conditions possibles de RSB. Dans cet article, nous considérons donc le couplage de deux algorithmes adaptatifs: un projecteur sur l'espace signal et un séparateur de sources.



2 Algorithmes adaptatif à mise à jour en série

2.1 Séparation adaptative

Nous avons présenté dans [2] un algorithme dont le schéma général est donné par la figure suivante:



qui fait apparaître un point important: la mise à jour du séparateur B_t ne dépend que des sorties $y(t)$ du séparateur et non de ses entrées $x(t)$. Plus précisément, nous préconisons la structure suivante d'adaptation:

$$B_{t+1} = \{I_n - \lambda_t G(y(t))\} B_t \quad (2)$$

où $G(\cdot)$ est une fonction qui associe à un vecteur y une matrice $G(y)$ de taille $n \times n$.

Il est facile de montrer qu'en l'absence de bruit ($x(t) = As(t)$), les performances en termes de séparation d'un tel algorithme ne dépendent pas de la matrice de mélange A . En effet, le but de l'algorithme consiste à adapter B de telle sorte que le système global $C_t \stackrel{\text{def}}{=} B_t A$, constitué de la matrice de mélange et du séparateur, converge vers la matrice identité (ou une permutation). En multipliant à droite la règle (2) par la matrice A , on obtient l'équation d'évolution du système global :

$$C_{t+1} = \{I_n - \lambda_t G(C_t s(t))\} C_t \quad (3)$$

où l'on a utilisé: $y(t) = B_t x(t) = B_t A s(t) = C_t s(t)$, valide à bruit nul. L'équation (3) met en évidence qu'en absence de bruit et pour toute fonction G , la trajectoire de C_t ne dépend que de la réalisation particulière des signaux sources. Autrement dit, la matrice de mélange A n'intervient ici que comme condition initiale. On en déduit que les performances de l'algorithme du point de vue de la séparation c'est-à-dire quantifiées en termes de la convergence ou des fluctuations du système global C_t , sont uniformes en la matrice de mélange. Cette propriété remarquable (et désirable) est acquise par la forme particulière de la mise à jour (2) où le système B_t est modifié par multiplication à gauche: en termes de 'circuit', ceci correspond à 'connecter' en série sur B_t le système $\{I_n - \lambda G\}$ (proche de l'identité si λ est faible).

Que les performances d'un algorithme série soient uniformes en A résulte uniquement du

fait que G ne soit fonction que de la sortie $y(t)$. Cependant, la qualité même de ces performances dépend ensuite d'un choix particulier de la fonction $G(y)$ et de son adéquation à la distribution des sources. Cette fonction peut dériver par exemple d'un potentiel comme la vraisemblance [4] et prend alors la forme

$$G(y) = I_n - g(y)y^* \quad (4)$$

ou encore d'une fonction de contraste orthogonale basée sur les cumulants comme celui proposé dans [5, 6] et l'on trouve [2]:

$$G(y) = yy^* - I_n + g(y)y^* - yg(y)^* \quad (5)$$

où, dans les deux cas, chaque composante du vecteur $g(y)$ est une fonction non-linéaire de la composante correspondante de y .

2.2 Points stationnaires.

Pour discuter du filtrage spatial du bruit, nous considérons d'abord à *bruit nul* les points stationnaires de l'algorithme (2) en distinguant d'une part les cas $m = n$ et $m > n$ et d'autre part les points stationnaires pour le séparateur B_t et pour le système global $C_t = B_t A$.

Considérons d'abord l'équation d'évolution (3) du système global C_t . Une matrice C_* de taille $n \times n$ est un point stationnaire seulement si la valeur moyenne du terme de mise à jour est nulle lorsque la matrice C est maintenue fixe à la valeur C_* . Le terme de mise à jour dans (3) étant $G(Cs)C$, la condition de stationnarité d'un système C_* inversible est donc simplement $EG(C_*s) = 0$. Puisque l'on souhaite que le système global converge vers l'identité, on voit alors que la fonction $G(\cdot)$ doit vérifier $EG(s) = 0$. C'est bien entendu le cas, sous les hypothèses convenables, pour les formes particulières de G données en (4) et (5). Nous faisons dans toute la suite l'hypothèse que $G(\cdot)$ vérifie effectivement $EG(s) = 0$.

Considérons maintenant les conditions de stationnarité pour le séparateur B_t lui-même. Pour les mêmes raisons, toute matrice B vérifiant $EG(Bx) = 0$ est un point stationnaire de (2). Dans le cas $m = n$ et en absence de bruit, la condition de stationnarité $EG(Bx) = 0$ est trivialement vérifiée pour $B = A^{-1}$ si $EG(s) = 0$. Dans le cas $m > n$, si $EG(s) = 0$, il suffit que $BA = I_n$ pour avoir $EG(Bx) = 0$. Par conséquent, si B_* est un point stationnaire, c'est aussi le cas de toute matrice $B = B_* + N$ où N est n'importe quelle matrice de taille $n \times m$ vérifiant $NA = 0$. Ces matrices forment un espace vectoriel de dimension $mn - n^2$ et à B_* .

est donc associé tout un continuum (de dimension $mn - n^2$) de points stationnaires. C'est une conséquence directe du fait que l'algorithme (2) adapte une matrice de taille $n \times m$ sur la base d'un critère défini par G de dimension n^2 . Ce fait est de peu d'importance dans le cas non-bruité puisque les matrices N sont celles dont les lignes sont dans l'espace bruit, c'est-à-dire orthogonales aux colonnes de A . Ainsi les séparateurs B_* et $B_* + N$ correspondent exactement au même système global et donc aux mêmes performances, *non en termes d'identification, mais non en termes de séparation*.

Dans le cas où le bruit additif ne peut être négligé, deux séparateurs B_1 et $B_2 = B_1 + N$ où $NA = 0$ correspondent au même système global $C = B_1A = B_2A$ mais donnent des estimées y_1 et y_2 des signaux sources différant par la quantité de bruit présent. L'approche la plus simple à ce problème consiste à gérer séparément le problème des capteurs surnuméraires et celui de la séparation proprement dite en projetant les données sur l'espace signal.

2.3 Projection adaptative sur l'espace signal

Nous appellerons 'projecteur sur l'espace signal', toute matrice P de dimension $n \times m$ dont les lignes sont orthonormées et engendrent l'espace signal.¹ Autrement dit, il faut vérifier:

$$PP^H = I_n \text{ et } P^H P A = A. \quad (6)$$

Une telle matrice n'est définie qu'à une matrice unitaire puisque si P vérifie (6), c'est aussi le cas de UP si U est unitaire de taille $n \times n$. Si une telle matrice de projection est connue, elle peut être utilisée pour un prétraitement spatial, en formant un nouveau processus $z(t)$:

$$z(t) \stackrel{\text{def}}{=} Px(t) = PA s(t) + Pb(t) \quad (7)$$

Le vecteur $z(t)$ obéit toujours à un modèle de mélange instantané mais la nouvelle matrice de mélange est désormais PA , carrée de dimension $n \times n$ et le bruit additif est $Pb(t)$ de matrice de covariance $\text{Cov}(Pb(t)) = PCov(b(t))P^H = P\sigma I_m P^H = \sigma I_n$: la blancheur spatiale du bruit est ainsi conservée en contraignant les lignes de P à l'orthonormalité. De plus, le bruit entachant $z(t)$ étant de covariance fixe égale à σI_n , le RSB dans $z(t)$ est maximisé en maximisant la contribution du signal. Sous la condition $PP^H = I_n$, ceci est obtenu sous la deuxième condition de (6),

¹Cette terminologie est abusive : c'est en fait $P^H P$ qui est *stricto sensu* un projecteur.

c'est-à-dire en alignant l'espace image de P^H sur celui de A .

Il est possible d'adapter un projecteur sur l'espace signal par un algorithme similaire à celui décrit plus haut pour la séparation de sources. Dans [7], Yang considère le critère :

$$c(P) = E|(I - P^H P)x|^2 \quad (8)$$

et montre qu'il n'admet pour uniques minima les matrices P satisfaisant les conditions (6). Un point remarquable est que cette propriété tient sans contraindre explicitement l'orthonormalité des colonnes de P . Il est facile d'obtenir le gradient de $c(P)$. Dans le cas réel, on trouve:

$$c(P + \delta P) = c(P) + 2E\text{Trace}\{H_P(x)P^H \delta P\} + o(\delta P) \quad (9)$$

où la matrice $H_P(x)$ est définie comme

$$H_P(x) = xx^* P + Pxx^* - 2xx^*. \quad (10)$$

On en déduit l'algorithme de gradient stochastique²:

$$P_{t+1} = P_t \{I_m - \lambda_t H_{P_t}(x_t)\}. \quad (11)$$

On trouvera dans [7] des exemples d'application à la localisation adaptative de sources. Pour notre part, nous combinons ci-dessous cet algorithme de projection avec l'algorithme de séparation de sources (2).

Il est frappant de constater la similarité entre les lois d'adaptation (2) et (11). Dans les deux cas, le système d'intérêt est adapté sans imposer explicitement de contraintes, et l'adaptation est obtenue par une loi prenant une forme simple lorsqu'elle est exprimée de façon multiplicative, ou encore 'en série'. Pour le séparateur B_t , cette façon de procéder confère à l'algorithme des performances uniformes; cette propriété, qui est liée à celle d'équivariance (voir [2] pour plus de détails) n'a toutefois pas d'équivalent en ce qui concerne le projecteur P_t .

On notera aussi que la mise à jour du projecteur se fait par prémultiplication tandis que celle du séparateur se fait par postmultiplication. Toutefois, le projecteur est mis à jour via la matrice $H_P(x)$ qui dépend explicitement de P : il n'est donc pas possible de combiner directement les deux algorithmes en un seul pour l'adaptation d'un système projecteur-séparateur $S_t = B_t P_t$ qui ferait disparaître l'individualité de chaque sous-système.

²Dans [7], l'auteur donne cette règle d'adaptation dans son équation (11) mais préfère élaborer plus avant. Nous nous en tenons ici à cette version simple. Voir aussi [8]



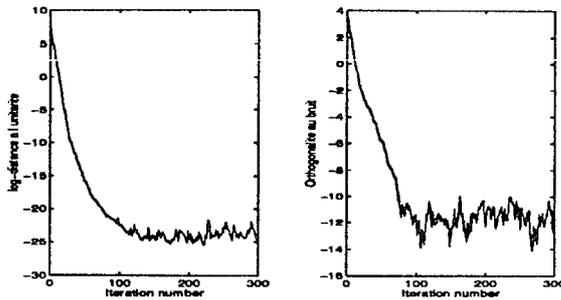
3 Simulations numériques

On utilise ici la version stabilisée de l'algorithme de séparation basé sur des non-linéarités cubiques pour $n = 2$ sources QAM4 et $m = 5$ capteurs.

La première figure illustre la convergence du projecteur P_t . Nous avons utilisé ici une version stabilisée de l'algorithme à pas fixe:

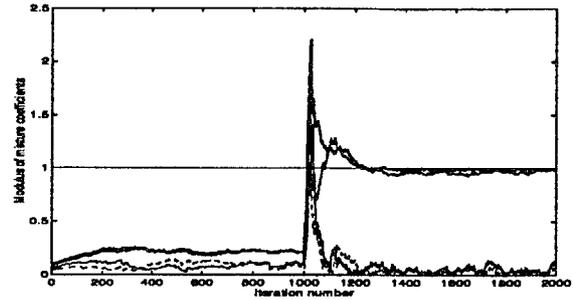
$$P_{t+1} = P_t \{ I_m - \lambda(1 + \lambda \|H_t\|_{\text{Fro}})^{-1} H_t \} \quad (12)$$

où $H_t = H_{P_t}(x_t)$. Les valeurs numériques sont $\sigma = 0.18$, $\lambda = 0.01$; P_0 et A sont tirées 'au hasard' avec les valeurs caractéristiques $\|A\|_{\text{Fro}} = 4.11$, $\text{cond}(A) = 2.2$.



Le panneau de gauche montre la trajectoire de $10 \log_{10} \|I_n - P_t P_t^H\|_{\text{Fro}}$, autrement dit illustre la vitesse de convergence de P_t vers une matrice orthonormée. Le panneau de droite montre la trajectoire de $10 \log_{10} \|P_t N\|_{\text{Fro}}$ où les colonnes de N forment une base orthonormée de l'espace bruit. Ceci illustre donc la convergence des lignes de P_t sur l'espace signal.

La seconde figure compare les performances du séparateur avec celles obtenues lorsqu'il est précédé du projecteur adaptatif. On montre l'évolution du module des coefficients de $B_t P_t A$ lorsque B_t est de taille $n \times n$ précédé d'un projecteur de taille $n \times m$. Le séparateur B_t est mis à jour par une version stabilisée de l'algorithme (2) où G est donné par (5) avec une non-linéarité cubique (voir [2] pour plus de détails). On souhaite voir converger les modules des coefficients du système global $B_t P_t A$ vers 0 ou 1. Dans cet exemple, on ne déclenche l'adaptation du projecteur que pour $t > t_0 = 1000$: avant t_0 , on observe une mauvaise convergence (vers un attracteur indésirable?) due à la présence du bruit; après t_0 , l'adaptation du projecteur débloque la situation: le rapport signal à bruit devient plus favorable et deux coefficients 'convergent' vers 0 et deux autres vers 1. Il s'agit ici cependant d'un exemple caricatural: pour obtenir un effet spectaculaire, nous avons volontairement initialisé A et P_0 de telle sorte que le RSB en sortie de $P_0 A$ soit mauvais.



Conclusion

En présence de moins de sources que de capteurs et de bruit additif, les algorithmes adaptatifs de séparation de sources peuvent être précédés par un projecteur sur le sous-espace signal. Ceci offre l'avantage de réduire la dimensionnalité des observations tout en maximisant le rapport signal à bruit. Il existe des algorithmes très simples pour ce faire, qui peuvent être utilisés en prétraitement pour tout séparateur adaptatif. Celui présenté ici est de structure homogène à notre technique adaptative de séparation de sources.

References

- [1] C. Jutten and J. Héroult, "Blind separation of sources: an adaptive algorithm based on neuromimetic architecture," *Signal Processing*, vol. 24, pp. 1-10, 1991.
- [2] B. Laheld and J.-F. Cardoso, "Adaptive source separation without prewhitening," in *Proc. EUSIPCO*, (Edinburgh), pp. 183-186, Sept. 1994.
- [3] E. Moreau and O. Macchi, "A one stage self-adaptive algorithm for source separation," in *Proc. ICASSP*, (Adelaide, Australia.), 1994.
- [4] D. T. Pham, P. Garrat, and C. Jutten, "Separation of a mixture of independent sources through a maximum likelihood approach," in *Proc. EUSIPCO*, pp. 771-774, 1992.
- [5] P. Comon, "Independent component analysis, a new concept?," *Signal Processing*, vol. 36, pp. 287-314, Apr. 1994.
- [6] M. Gaeta J.-L. Lacoume, "Source separation without a priori knowledge: the ML solution," in *Proc. EUSIPCO*, pp. 621-624, 1990.
- [7] B. Yang, "Projection approximation subspace tracking," *IEEE Tr. SP*, vol. 43, pp. 95-107, Jan. 1995.
- [8] J. Dehaene, M. Moonen, J. Vandewalle, "A continuous time approach to the analysis and design of algorithms for subspace tracking," *SVD and sig. proc. III*, pp. 259-266, Elsevier, 1995.