

Adaptation de SOBI à des données fréquentielles

Danielle NUZILLARD

Laboratoire d'Automatique et de Micro-électronique

Faculté des Sciences, Moulin de la Housse, BP 1039-51687 REIMS Cedex 2, France

danielle.nuzillard@univ-reims.fr

Résumé – Cet article traite de la séparation aveugle de sources orthogonales. La méthode présentée ici trouve tout son intérêt quand les signaux nécessitent des pré et post traitements dans le domaine fréquentiel, quand le contenu fréquentiel est faible devant le nombre d'échantillons temporels ou encore quand les données spectrales sont directement disponibles. Les matrices de covariance nécessaires sont établies à partir des spectres des signaux. La matrice séparante qui s'en déduit est appliquée sur les données spectrales et permet de restituer directement les spectres sources. Cette technique a été développée dans le cadre de SOBI. Elle est illustrée par une mise en œuvre dans une application d'analyse chimique par RMN.

Abstract – This article deals with source separation of orthogonal signals. The method presented herein is adapted to situations in which pre and post processing must be carried out in the frequency domain, when the frequency content is limited or when the signal acquisition takes place in this frequency domain. The separation itself is achieved by the SOBI algorithm but the required covariance matrices are evaluated directly from the spectra. The separation matrix is computed (rather than the mixing matrix) and applied to the spectral data. A real case of application is presented, using 2-D NMR spectra of mixtures of chemical compounds.

1. Introduction

Cette étude traite de la séparation aveugle de signaux sources orthogonaux et détectés en mélanges instantanés. La séparation est effectuée dans le domaine fréquentiel.

La séparation de sources dans le domaine fréquentiel se révèle attractive lorsque les données spectrales sont directement disponibles ou que les signaux nécessitent des pré et/ou post-traitements dans le domaine spectral. Dans ce cas, à partir des données temporelles, une première transformation de Fourier fournit les données spectrales. Celles-ci sont pré-traitées, pour corriger d'éventuels artefacts et/ou pour extraire les données pertinentes. Une transformation de Fourier inverse permet de revenir dans le domaine temporel pour appliquer un algorithme de séparation adéquat. Ensuite, le calcul des spectres sources nécessite une transformation de Fourier supplémentaire. Ces étapes successives de transformation de Fourier sont réduites si les données spectrales pré-traitées sont directement séparées.

Parmi les candidats à ces traitements, on peut citer les données pour lesquelles le contenu spectral est faible devant le nombre d'échantillons, mais aussi un cas bien particulier développé au paragraphe 2 pour lequel la position des raies spectrales fluctue au cours de l'enregistrement. D'autre part, la séparation dans le domaine fréquentiel nécessite moins de calculs lorsque le nombre de raies spectrales est inférieur au nombre d'échantillons temporels. C'est le cas pour les signaux expérimentaux. La transformation de Fourier est une opération linéaire, la matrice séparante peut donc s'appliquer indifféremment sur des données temporelles ou fréquentielles. Cette matrice est obtenue à partir des

matrices de covariance, elles-mêmes établies à partir des données spectrales. Ceci fera l'objet du troisième paragraphe. L'intérêt de la méthode est ensuite développé et son cadre d'application est précisé. Au cours du quatrième paragraphe, des résultats de séparation sur des données réelles issues de l'analyse chimique par résonance magnétique nucléaire (RMN) illustreront la méthode de séparation en fréquence.

2. Position du problème

La situation qui a motivé cette étude correspond à un cas bien particulier de l'acquisition des données. Les signaux sont produits par des échantillons soumis à une méthode physico-chimique d'analyse, la résonance magnétique nucléaire. Lorsqu'on étudie plusieurs mélanges de plusieurs substances, ils doivent être analysés séquentiellement. Les fréquences peuvent être globalement affectées par le changement d'échantillon, car elles sont obtenues par soustraction de la fréquence d'un oscillateur local dont la stabilité n'est pas infiniment parfaite. D'autre part, la concentration des composants n'étant pas la même d'un mélange à l'autre, elle peut induire des variations de fréquence dans les réponses d'un même constituant. Dans un cas comme dans l'autre, l'hypothèse des mélanges linéaires des signaux sources n'est plus respectée et conduit à un échec des méthodes de séparation de sources.

Il est tout de même possible de faire appel à la séparation en considérant que l'intégrale d'un signal dans le domaine des fréquences est proportionnelle à la concentration du constituant qui lui a donné naissance. Il convient alors d'intégrer les signaux observés sur un

domaine suffisamment large pour les englober individuellement pour tenir compte de la variation de leur fréquence. Un signal fréquentiel défini par plusieurs centaines ou milliers de points peut alors être réduit à un vecteur contenant quelques dizaines de valeurs d'intégrales seulement. Le problème de séparation se résume à la décomposition des vecteurs expérimentaux réduits en vecteurs sources. Chacun des vecteurs sources doit contenir des zéros pour les zones d'intégration où le constituant du mélange qui lui est associé ne fournit pas de réponse. Cette condition est l'analogie de la condition d'orthogonalité des signaux sources. On peut considérer les vecteurs expérimentaux réduits comme des signaux fréquentiels séparables par exemple par l'algorithme SOBI [1] moyennant une transformation de Fourier pour recréer des signaux temporellement corrélés. SOBI a servi de cadre pour développer cette méthode de séparation en fréquence. L'intérêt de cette approche réside dans la possibilité de calculer directement les coefficients d'intercorrélation des signaux en mélange à partir des données fréquentielles.

3. Matrices de covariance

3.1 Construction

A une collection d'amplitudes dans le domaine fréquentiel correspond une somme d'originaux temporels de forme sinusoïdale. Ces signaux sont bien temporellement cohérents. De plus, si les sources sont orthogonales en fréquence, elles sont orthogonales en temps puisque la transformation de Fourier est une opération qui conserve le produit scalaire. Les conditions d'utilisation de l'algorithme SOBI sont donc satisfaites.

A partir des raies spectrales, il est possible de reconstituer le signal temporel $A(l, \Delta t)$ correspondant par une transformation de Fourier inverse discrète, puis de calculer sa matrice de corrélation avec un autre signal et d'introduire des retards qui tiennent compte de la cohérence temporelle des signaux. Un signal temporel reconstitué s'écrit sous la forme suivante :

$$A(l, \Delta t) = \sum_k a_k \cdot e^{i2\pi \cdot \frac{kf_e}{T} \cdot l\Delta t} \quad (1)$$

$$\text{ou encore, } A(l, \Delta t) = \sum_k a_k \cdot w^{k \cdot l} \quad (2)$$

$$\text{avec } w = e^{i\frac{2\pi}{T}} \quad (3)$$

f_e est la fréquence d'échantillonnage, T est le nombre d'échantillons, $k \cdot f_e / T$ sont les fréquences des raies spectrales d'amplitude a_k , les $l \cdot \Delta t$ sont les instants d'échantillonnage.

Le coefficient d'intercorrélation entre deux signaux A et B a pour expression :

$$R_{AB}(\tau) = \frac{1}{L} \sum_l A(l, \Delta t) \cdot B^*(l, \Delta t - \tau), \quad (4)$$

$$\text{avec } \tau = n \cdot \Delta t, \quad (5)$$

alors :

$$R_{AB}(n) = \frac{1}{L} \sum_l \left(\sum_k a_k \cdot w^{k \cdot l} \right) \cdot \left(\sum_{k'} b_{k'}^* \cdot w^{-k' \cdot (l-n)} \right), \quad (6)$$

$$\text{d'où } R_{AB}(n) = \sum_k a_k \cdot b_k^* \cdot w^{k \cdot n}. \quad (7)$$

Les coefficients d'intercorrélation $R_{AB}(n)$ entre deux mélanges sont obtenus par transformation de Fourier inverse du produit terme à terme de deux vecteurs fréquentiels A et B^* . Les m vecteurs associés aux mélanges fournissent les $m \cdot (m+1) / 2$ coefficients d'intercorrélation qui constituent la partie triangulaire inférieure ou supérieure de la matrice de covariance $R_X(n)$. Celle-ci est complétée par symétrie. Le produit terme à terme de chaque paire de deux vecteurs fréquentiels A et B^* est mémorisé. Il entre dans le calcul des coefficients retardés qui eux-mêmes entrent dans la constitution des p matrices de covariance correspondantes. Puisque le nombre de valeurs de n (nombre de retards) et de k (nombre de raies spectrales) est faible, il n'y a pas lieu d'utiliser un algorithme de FFT pour calculer les $R_{AB}(n)$.

La suite de la démarche repose sur l'algorithme SOBI. La matrice $R_X(n=0)$ issue d'un retard nul est diagonalisée de façon à obtenir la matrice de blanchiment. Celle-ci est appliquée sur les données spectrales. La matrice séparante et la matrice de mélange sont obtenues à une constante et à une permutation près, par diagonalisation conjointe des autres matrices de covariance. L'application de la matrice séparante aux données initiales restitue les sources spectrales.

3.2 Avantages

La méthode décrite ci-dessus trouve tout son sens lorsque les données temporelles nécessitent des pré- et post-traitements dans le domaine fréquentiel où quand elles sont directement disponibles dans ce domaine. Un bilan est dressé dans le tableau 1, entre la séparation temporelle effectuée à l'aide de SOBI et cette méthode de séparation fréquentielle adaptée à ce type de problème. Les données temporelles sont composées de m vecteurs contenant chacun T échantillons à partir desquels sont recherchés n vecteurs sources orthogonaux. Les données pertinentes dans le domaine des fréquences ne contiennent plus que k échantillons. Ils sont utilisés pour calculer les coefficients d'intercorrélation des p matrices de covariance. La séparation en fréquence n'a lieu ensuite que sur k échantillons en fréquence au lieu de T échantillons en temps. De plus, deux transformations de Fourier sont évitées.

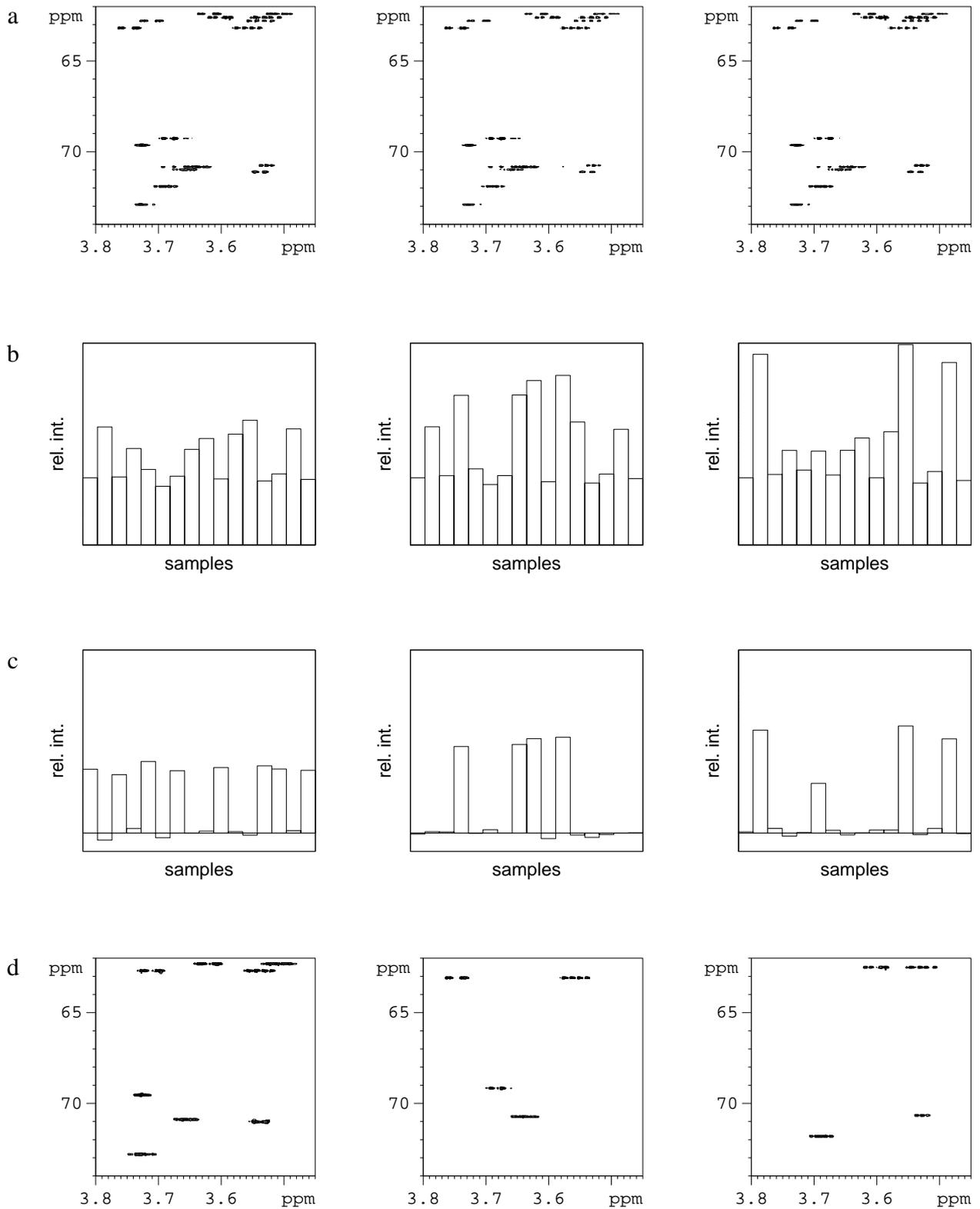


FIG. 1 : Spectres 2-D des mélanges (ligne a), données spectrales réduites avant séparation (ligne b), sources spectrales réduites (ligne c) et spectres des substances individualisées par séparation de sources (ligne d).

TAB. 1 : Comparaison des méthodes

	SOBI Séparation temporelle	Méthode adaptée Variante fréquentielle	Gain
Pré-traitement domaine spectral	- TF sur T points de m vecteurs, - Extraction de k points pertinents des m vecteurs, - TFI sur T points de m vecteurs.	- TF sur T points de m vecteurs, - Extraction de k points pertinents des m vecteurs.	- TFI sur T points de m vecteurs.
Séparation	- Calcul de $p+1$ matrices de covariance à partir de T points sur m vecteurs, - Diagonalisation conjointe de p matrices, - Séparation sur T points.	- Calcul de $p+1$ matrices de covariance à partir de k points sur m vecteurs, - Diagonalisation conjointe de p matrices, - Séparation sur k points.	- compte tenu de la symétrie des matrices, calcul des $m(m+1)/2$ coefficients d'intercorrélation pour chacune des p matrices de covariance, - ($T-k$) points sont inutiles.
Post-traitement domaine spectral	- TF sur T points de n vecteurs, - Réarrangement pour l'application.	- Réarrangement pour l'application.	- TF sur k points de n vecteurs.

4. Application à des données réelles

L'algorithme a été appliqué à des données réelles issues d'un spectromètre de résonance magnétique nucléaire (RMN). Les signaux accumulés subissent une transformation de Fourier. Ils sont ensuite disponibles sous la forme d'un spectre à une (1-D) ou deux dimensions (2-D). Chaque dimension est liée à la réponse d'un type d'atomes. Cette réponse dépend de la nature des liaisons entre les atomes d'une même molécule.

Trois mélanges de trois substances sont analysés. Ce sont le mannitol, le xylitol et le sorbitol. Chacun des trois spectres (2-D) de mélange (figure 1, ligne a) présente de larges parties où aucune information n'est significative. L'information pertinente est localisée. Elle est extraite des trois spectres selon le même patron. Ici, chaque spectre est découpé en seize zones. Sur la surface de chacune d'elle, les raies spectrales correspondantes sont intégrées. Ces intégrales constituent les nouvelles grandeurs utiles. Chacun des trois spectres ne comporte plus que 16 valeurs. L'information est donc réduite à 3 vecteurs ayant chacun 16 éléments. Ces vecteurs sont considérés comme des spectres 1-D réduits qu'il faut séparer pour identifier les composés en mélange (figure 1, ligne b).

Dans cet exemple, il est clair que les données fréquentielles ne sont pas séparables directement par l'algorithme SOBI puisqu'elles ne présentent pas le caractère de cohérence nécessaire.

Une première démarche imposait d'abord d'effectuer une transformation de Fourier inverse des spectres 1-D réduits pour séparer les données temporelles réduites par l'algorithme SOBI. Ensuite, les sources temporelles réduites subissaient une transformation de Fourier pour obtenir les sources fréquentielles réduites. Enfin, celles-ci étaient réintroduites éléments par éléments dans les zones d'extraction pour reconstituer les spectres 2-D séparés [2]. Chaque raie était normalisée en prenant en compte la hauteur maximale de la raie initiale.

La seconde démarche est l'application du nouvel algorithme adapté à la séparation en fréquence. Celui-ci est applicable directement sur les données fréquentielles réduites. Il fournit directement les spectres 1-D réduits séparés (figure 1, ligne c). Ceux-ci sont ensuite réintroduits dans les zones d'extraction des spectres 2-D comme dans le cas précédant (figure 1, ligne d).

Références

- [1] A. Belouchrani, K. Abed-Meraim, J-F. Cardoso, E. Moulines, « A blind Source Separation Technique Using Second-Order Statistics », *IEEE Trans. On Signal Processing*, vol. 45, no. 2, pp. 434-443, feb. 1997.
- [2] D. Nuzillard, S. Bourg, J.-M. Nuzillard, « Model-free Analysis of Mixtures by NMR », *Journal of Magnetic Resonance*, vol. 133, pp. 358-363, aug. 1998.