

Représentations parcimonieuses pour la séparation de sources avec un seul capteur

Laurent BENAROYA, Rémi GRIBONVAL, Frédéric BIMBOT

IRISA/INRIA, Campus de Beaulieu
35042 Rennes Cedex

laurent.benaroya@irisa.fr, remi.gribonval@inria.fr
frederic.bimbot@irisa.fr

Résumé – Nous étudions un paradigme pour la séparation de deux sources sonores avec un seul capteur. On compare expérimentalement trois méthodes de décomposition parcimonieuse pour un dictionnaire constitué de l’union de deux bases. On évalue la qualité de la séparation en fonction de la parcimonie de chaque source. De manière complémentaire, on étudie également l’apprentissage d’une base à partir d’une classe de signaux sonores définie par des exemples. On constate qu’à qualité de séparation sensiblement égale, notre méthode originale de seuillage alterné se distingue des méthodes FOCUSS et Basis Pursuit par des temps de calculs beaucoup plus faibles. Par ailleurs, il semble que l’apprentissage de bases par Analyse en Composantes Indépendante ou Principales soit insuffisant dans notre contexte.

Abstract – We investigate a paradigm for the separation of two sound sources with only one sensor. We experiment three sparse decomposition methods over a dictionary which is the union of two basis. We estimate the quality of separation as a function of the sparsity of the sources. Besides, we study methods that permit to learn a basis for a class of signals from a database of samples. We find that, with almost the same separation quality, our original method, an alternated thresholding method, requires noticeably less computing time. Furthermore, Independent and Principal Component Analysis for basis learning seem to be limited in our context.

1 Introduction

La séparation de sources consiste à estimer un ensemble de signaux dont on observe un ou plusieurs mélanges linéaires. Dans le cas où la matrice de mélange est carrée, l’Analyse en Composantes Indépendantes (ACI) fournit un algorithme d’estimation de cette matrice. Si celle-ci est inversible, on peut alors estimer linéairement les sources. Dans cet article, nous étudions un cas particulier, dans lequel la matrice de mélange est de taille $1 \times n$ où n est le nombre de sources (l’observation est mono-dimensionnelle). Nous proposons un algorithme d’estimation non linéaire des sources à partir de bases adaptées à celles-ci.

Dans un premier temps, nous étudions la séparation de deux sources pour chacune desquelles nous connaissons une base « adaptée ». Les sources sont estimées grâce à une décomposition parcimonieuse sur la réunion des bases adaptées. Dans un second temps, nous étudions la constitution d’une base adaptée à une classe de signaux sonores. Cette classe est caractérisée par des exemples sonores. Nous proposons d’estimer les bases par Analyse en Composantes Principales (ACP) ou Indépendantes (ACI). L’objectif est de trouver une base sur laquelle les exemples sonores sont « bien » approchés par un petit nombre de vecteurs. Deux méthodes rapides de décomposition parcimonieuse sont exposées dans la section 2. Ces méthodes sont étudiées expérimentalement dans la section 3, pour l’estimation de sources sur des bases adaptées connues. La section 4 concerne l’apprentissage d’une base pour une classe de signaux dont on connaît des exemples.

2 Décomposition parcimonieuse sur la réunion de deux bases

On dispose d’un signal $x = s_1 + s_2 \in \mathbb{R}^N$ de longueur N , que l’on suppose généré par la superposition de deux sources s_1 et s_2 . Sans modèle des sources, le problème de leur détermination à partir de x est mal posé. Cependant, si chaque source admet une représentation parcimonieuse $s_i = B_i \omega_i$ où B_i est la matrice d’une base \mathcal{B}_i et $\omega_i \in \mathbb{R}^N$ a peu de coefficients non nuls, une approche possible [6] [3] pour estimer les deux sources s_1 et s_2 consiste à décomposer parcimonieusement x sur le dictionnaire redondant $\mathcal{D} = \mathcal{B}_1 \cup \mathcal{B}_2$.

Il existe plusieurs méthodes (sous-optimales) pour décomposer un signal sur un système redondant. Le Basis Pursuit [2] est un algorithme performant, mais coûteux en temps de calcul. Nous proposons deux méthodes rapides de décomposition parcimonieuse par minimisation d’une norme ℓ_p ($p \leq 1$) des coefficients : la méthode *FOCUSS régularisée* [8] [9] et un algorithme original de *seuillage alterné* avec un seuil décroissant.

2.1 Algorithme FOCUSS régularisé

Dans le cas où les bases \mathcal{B}_1 et \mathcal{B}_2 sont quelconques, l’algorithme FOCUSS régularisé [8] [9] est une méthode itérative pour estimer la solution du problème suivant, où

l'on a noté $\omega = (\omega_1, \omega_2)$ et $A\omega = B_1\omega_1 + B_2\omega_2$ avec $p \leq 2$

$$\hat{\omega} = \arg \min_{\omega} \frac{1}{2} \|A\omega - x\|_2^2 + \lambda \|\omega\|_p \quad (1)$$

Nous avons utilisé l'algorithme avec $p = 0$. Nous utilisons un paramètre de régularisation $\lambda = \lambda_n$ décroissant en fonction des itérations (voir [9] pour plus de détails sur l'algorithme).

2.2 Seuillage alterné

Si les bases \mathcal{B}_1 et \mathcal{B}_2 sont *orthonormées*, un algorithme de *relaxation itérative par blocs* permet de résoudre le problème (1) avec $p = 1$ [7].

2.2.1 Algorithme

L'algorithme consiste en un seuillage alterné : si $\hat{\omega}_2^n$ est l'estimation de ω_2 à l'étape n , alors

$$\hat{\omega}_1^n = \arg \min_{\omega_1} \frac{1}{2} \|x - B_1\omega_1 - B_2\hat{\omega}_2^n\|_2^2 + \lambda(\|\omega_1\|_1 + \|\hat{\omega}_2^n\|_1) \quad (2)$$

est obtenu par seuillage doux [5] de $B_1^T(x - B_2\hat{\omega}_2^n)$ avec un seuil λ .

Puis, on ré-estime $\hat{\omega}_2^{n+1}$ par seuillage doux de $B_2^T(x - B_1\hat{\omega}_1^n)$.

2.2.2 Algorithme modifié

Nous proposons de remplacer le seuil constant λ dans l'équation (2) par un seuil décroissant λ_n . La décomposition fournie par l'algorithme de seuillage alterné (et par FOCUSS régularisé) dépend a priori de la suite (λ_n) . Si la série $\sum_n \lambda_n$ est convergente, alors l'algorithme converge. Nous conjecturons que la limite de $(\hat{\omega}_1, \hat{\omega}_2)$ est solution du problème :

$$(\hat{\omega}_1, \hat{\omega}_2) = \arg \min \|\omega_1\|_1 + \|\omega_2\|_1 \quad (3)$$

sous la contrainte : $x = B_1\omega_1 + B_2\omega_2$

3 Performances expérimentales

Dans cette section, on étudie l'influence de la parcimonie effective des sources à séparer, sur les performances de séparation, avec le Basis Pursuit (BP) et les algorithmes présentés dans la section précédente. Dans le cadre du BP, nous avons utilisé le logiciel *Atomizer* [1].

On se place dans le cas particulier où \mathcal{B}_1 est la base de Dirac et \mathcal{B}_2 est la base de Fourier réelle. On place autant de composantes non nulles dans chaque source. Ces composantes ont toutes une amplitude de 1 et sont de signe aléatoire.

On évalue les performances en terme de rapport signal à bruit (RSB) entre les sources s_i et leur estimation \hat{s}_i , mesuré en décibels pour chaque source, soit

$$RSB = -20 \log_{10} \frac{\|s_i - \hat{s}_i\|}{\|s_i\|} \quad (4)$$

Remarque : Comme $x = s_1 + s_2 \simeq \hat{s}_1 + \hat{s}_2$ et $\|s_1\| = \|s_2\|$, le RSB est presque identique pour les deux sources.

Afin de modéliser le bruit de quantification, nous ajoutons au mélange des sources un bruit uniforme de variance $\frac{(2^{-16})^2}{12}$ par échantillon. On considère alors qu'un RSB supérieur ou égal à 96 dB indique que la séparation est exacte. Noter que l'on peut avoir un RSB strictement supérieur à 96 dB, puisque le seuillage induit un débruitage. Ce n'est pas le cas pour le Basis Pursuit avec l'algorithme du point intérieur.

3.1 Conditions expérimentales

On a choisi $N = 64$. On effectue 100 tirages pour un nombre de composantes non nulles fixé. Le nombre total de composantes non nulles varie avec un pas de deux. Pour la méthode FOCUSS, on effectue 8 itérations et le paramètre de régularisation est $\lambda_n = (\frac{1}{100})^n$. Pour le seuillage alterné, on effectue 100 itérations et le seuil vaut $\lambda_n = O(n^{-4})$. Pour le Basis Pursuit, nous avons fixé à 10^{-3} les critères d'arrêt, qui donnent expérimentalement une bonne approximation [1].

3.2 Résultats

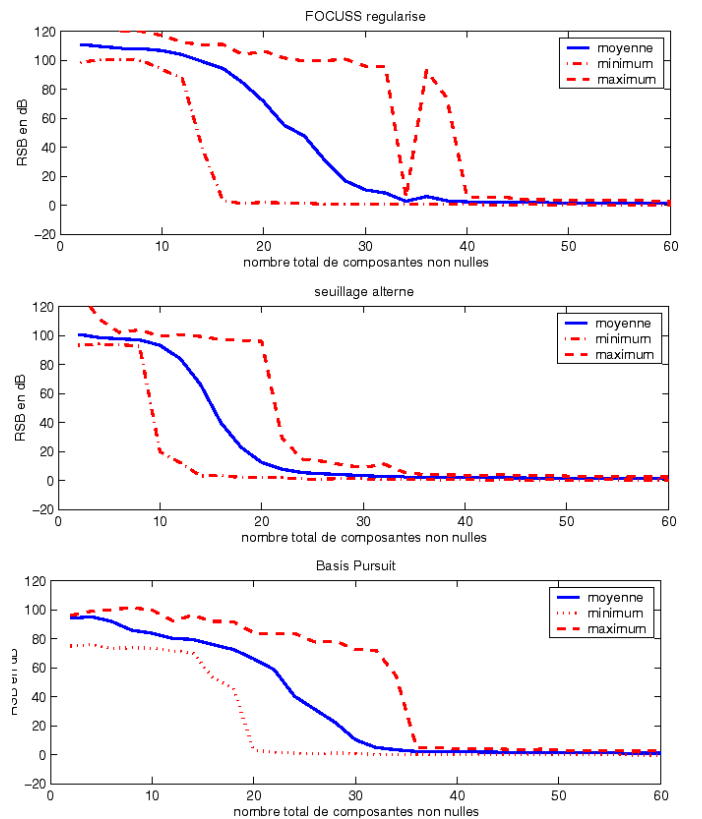


FIG. 1 – erreur relative de séparation par l'algorithme FOCUSS régularisé (en haut), par l'algorithme de seuillage alterné (au milieu) et le BP (en bas)

La Figure 1 indique le RSB entre les sources et leur estimation pour chaque algorithme. En abscisse, on trouve le nombre total de composantes non nulles effectivement présentes dans chaque source. En ordonnées se trouve le RSB en décibels. On a représenté trois courbes : la moyenne (courbe en trait plein) du RSB obtenu sur les 100 tirages à

nombre total de composantes fixe, le maximum (courbe en pointillés au dessus de la moyenne) et le minimum (courbe en pointillés au dessous de la moyenne). Sur les courbes correspondant à chacun des algorithmes, on distingue nettement trois zones :

- Pour un faible nombre de composantes non nulles, nous obtenons une très bonne séparation des deux sources. Le RSB est supérieur à 96 dB.
- Avec un peu plus de composantes non nulles, les performances des algorithmes décroissent en moyenne, avec une grande variabilité en fonction du tirage des sources. On peut penser que cette variabilité est due au fait qu’il existe plusieurs solutions de parcimonie comparable pour représenter le signal sur l’ensemble des deux bases.
- S’il y a beaucoup de composantes non nulles, le RSB est quasi nul : il devient impossible de séparer les sources.

3.3 Discussion

La première partie de la courbe correspond à une décomposition quasi exacte des sources, pour 100% des tirages. Cette partie se situe entre 2 et 10 composantes pour l’algorithme FOCUSS et entre 2 et 8 composantes pour le seuillage alterné. C’est la *zone de fonctionnement efficace* pour la séparation de sources. Cette zone est plus difficile à interpréter dans le cas du Basis Pursuit. Pour des sources réelles, si l’on peut identifier deux bases assez incohérentes, sur lesquelles elles ont une représentation suffisamment parcimonieuse, alors l’algorithme de décomposition « fonctionnera ». De fait, il existe un résultat théorique dans le cas de la minimisation de la norme ℓ_1 et de la quasi-norme ℓ_0 [3].

Théorème 1 (D.L. Donoho et X. Huo) *Si $x \in \mathbb{R}^N$ est la superposition de N_1 Diracs et de N_2 éléments de la base de Fourier réelle et si*

$$N_1 + N_2 \leq \frac{1}{2} \sqrt{N/2} \quad (5)$$

Alors la solution de l’équation (3) est unique et est aussi l’unique solution du problème de minimisation de la quasi-norme ℓ_0 .

Dans le cadre de notre expérience, le théorème s’applique donc si le nombre total de composantes est inférieur à $2\sqrt{2} \simeq 3$. On observe expérimentalement que la propriété de séparation exacte des sources s’applique bien au delà de cette borne. Par ailleurs, on peut envisager d’utiliser la méthode de seuillage alterné, qui nécessite des temps de calcul beaucoup plus faible, comme méthode d’initialisation pour l’accélération des algorithmes FOCUSS ou Basis Pursuit.

Le Tableau 1 récapitule les propriétés des trois algorithmes dans le cadre de la séparation de sources dans le cas des bases Dirac/Fourier. Nous étudions dans la section suivante le cas de bases plus complexes, adaptées à une classe de signaux.

TAB. 1 – Tableau Récapitulatif pour la séparation de source par décomposition parcimonieuse dans la réunion Base de Fourier/base de Dirac

	seuillage alterné	FOCUSS	Basis Pursuit
ℓ_p	ℓ_1	ℓ_0	ℓ_1
λ_n	$O(n^{-4})$	10^{-2n}	-
itérations	100	8	entre 4 et 10
temps cpu	20 min	2h30	>20h
zone efficace	2-8	2-12	2-14
RSB	>96 dB	>96 dB	≈ 80 dB

4 Adaptation d’une base à une classe de signaux sonores

4.1 Méthodes de construction

Une première méthode de construction classique est l’ACP (base de Karhunen-Loeve) [11]. La base est adaptée par minimisation de l’erreur moyenne d’approximation linéaire. Avec une méthode d’ACI [12] ou d’Analyse en composantes orthogonales le moins statistiquement dépendantes [4], on obtient des bases dans lesquelles les sources ont en théorie une représentation un peu plus parcimonieuse qu’avec l’ACP, ce qui devrait se traduire par une meilleure séparation des sources sur ces bases. Nous comparons les bases construites en terme de séparation de signaux par les différentes méthodes étudiées dans la section 3.

4.2 Expérimentation

Nous avons considéré 800 trames de longueur 128 d’un enregistrement de piano d’une part, et d’un enregistrement de violon, d’autre part. La fréquence d’échantillonnage était de 8kHz. Pour chaque source, nous avons construit deux bases (par ACP et ACI orthogonale), en utilisant la moitié des trames. Nous avons utilisé sur les trames restantes, la méthode de seuillage alterné pour évaluer les possibilités de séparation des instruments à partir des bases ainsi constituées. Par ailleurs, nous avons comparé les résultats à un filtrage linéaire de Wiener sur la base de Fourier:

On estime les matrices de covariances des deux sources, que l’on suppose diagonales dans le domaine de Fourier. On note $\mathcal{F}(x)$ la transformée de Fourier de x et $\sigma_1^2(f)$ (resp. $\sigma_2^2(f)$) la variance de la source s_1 (resp. s_2) à la fréquence f , calculée sur les données d’apprentissage. Les sources sont estimées par le filtre linéaire de Wiener suivant

$$\mathcal{F}(s_i)(f) = \frac{\sigma_i^2(f)}{\sigma_1^2(f) + \sigma_2^2(f)} \cdot \mathcal{F}(x)(f)$$

TAB. 2 – RSB en dB pour la séparation d’un mélange piano/violon en fonction de la méthode d’apprentissage

	piano	violon
ACP	3,3	2,7
ACI	5,0	4,2
Wiener	3,0	3,0

4.3 Discussion

Les résultats pour le filtre de Wiener ainsi que pour les bases apprises par ACP sont décevants, car on obtiendrait le même résultat en choisissant $\hat{s}_i = \frac{x}{2}$. L'ACI orthogonale donne un RSB légèrement supérieur. Ces résultats médiocres s'expliquent peut-être parce que les bases calculées sur chacun des instruments sont proches de celle de Fourier, et donc proches entre elles. En effet, on note $M = \max_{i,j} | \langle B_i^1, B_j^2 \rangle |$ le plus grand produit scalaire, en valeur absolue, entre les vecteurs de la base \mathcal{B}_1 et les vecteurs de la base \mathcal{B}_2 . M est une mesure de la cohérence des deux bases [3]. Pour l'ACP, on trouve $M = 0,98$ et pour l'ACI orthogonale $M = 0,93$. Dans le cas des bases Dirac/Fourier réelle, on a $M = 0,125$.

5 Conclusion

Nous avons mis en place une méthodologie en vue de la séparation de deux sources avec un seul capteur. Cette méthodologie comprend deux volets : la séparation de sources, connaissant une base adaptée pour chaque source; l'apprentissage de bases adaptées à partir d'un ensemble d'exemples sonores.

Les expériences sur trois méthodes de décomposition parcimonieuse suggèrent que la séparation de sources sur la réunion de deux bases adaptées est possible à condition que les sources admettent une représentation fortement parcimonieuse dans leur base adaptée et que les bases soient suffisamment incohérentes.

Pour la constitution de bases adaptées, nous avons comparé les performances de l'ACP et l'ACI en terme de séparation de source et les résultats sont décevants. Il semble que le calcul conjoint des deux bases pour les deux sources soit une piste susceptible d'améliorer la méthode. En terme de perspectives, nous suggérons l'utilisation d'algorithmes de type Expectation Maximisation (EM) [10] pour le calcul conjoint de sources séparées et des bases adaptées.

Références

- [1] S. Chen and D.L. Donoho *About Atomizer* Technical Report, Departement of Statistics, Stanford University, 1995
<http://www-stat.stanford.edu/~atomizer/>
- [2] S. Chen *Basis Pursuit* thesis, November 1995
- [3] D.L. Donoho and X. Huo *Uncertainty Principles and Ideal Atomic Decomposition* June 1999
- [4] N. Saito *Least statistically-dependent basis and its application to image modeling* Dans A. Laine, M. Unser et A. Aldroubi, rédacteurs, Wavelet Applications in Signal and Image Processing, tome 3458 de Proc. SPIE. San Diego CA., juillet 1998
- [5] D.L. Donoho *De-noising by soft-thresholding* 1992
- [6] M. Zibulevsky and B. A. Pearlmutter *Blind Source Separation by Sparse Decomposition* University of New Mexico Technical Report No. CS99-1
- [7] S.Sardy, A.G. Bruce and P. Tseng *Block Coordinate Relaxation Methods for nonparametric signal denoising with wavelet dictionaries*
- [8] Kjertsi Engan *Frame based signal representation and compression* Thesis, Departement of Electrical and Computer Engineering, Stavanger University College, Norway, 2000
- [9] B.D. Rao and K. Kreutz-Delgado *Basis Selection in the presence of Noise* Proceedings of the Thirty-second Asilomar Conference on Signals, Systems and Computers, Vol. 1, pp. 752-756, Pacific Grove, CA, Nov. 1-4 1998
- [10] O. Bermond and J.F. Cardoso *Approximate Likelihood for Noisy Mixtures* Proc. ICA '99, Aussois, France, pp. 325-330
- [11] J. Karhunen and J. Joutsensalo *Generalizations of principal component analysis, optimisation problems, and neural networks* Neural Networks, vol. 8, pp. 549-562, 1995
- [12] A. Hyvärinen and E. Oja. *Independent Component Analysis: Algorithms and Applications*. Neural Networks. 13(4-5):411-430, 2000