

Séparation de sources

Matthieu KOWALSKI

Table des matières

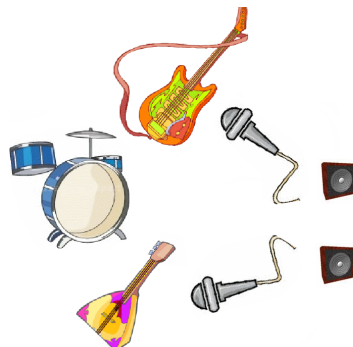
1	Introduction	1
2	Mélanges instantanés sur-déterminé	3
2.1	Mélanges non aveugles	3
2.2	Mélanges aveugles : ICA	4
2.2.1	ICA : formulation et pre-processing	4
2.2.2	ICA : principes généraux	5
2.2.3	Deux approches d'ICA	6
2.2.4	ICA dans un domaine transformé	8
3	Mélanges instantanés sous-déterminé	8
3.1	Parcimonie et séparation de sources	8
3.1.1	Exemples jouets	8
3.1.2	Influence de la parcimonie	9
3.2	Masquage temps-fréquence	10
3.2.1	Une seule source active	10
3.3	Approche type Basis Pursuit	12
4	Mélanges convolutifs non aveugle	13
4.1	Problème et 1ère approche	13
4.2	Approche variationnelle	13

1 Introduction

Table des matières

Problème et notations

Problème : On “observe” un mélange stéréo (bruité) de trois instruments. Comment retrouver ces trois instruments ?



Différences types de mélanges

Trois modèles principaux

1. Les mélanges instantanés
2. Les mélanges convolutifs
3. Cas particuliers des mélanges convolutifs : les mélanges anéchoïques

Trois situations

1. Mélange **sur-déterminé** : on a autant d'observation que de sources
2. Mélange **déterminé** : on a plus d'observation que de sources
3. Mélange **sous-déterminé** : on a moins d'observations que de sources

Mélanges instantanés

On observe M signaux issus d'un mélange de N signaux, avec $M < N$

Mélanges instantanés

$$X = AS$$

avec

- Les M observations $X \in \mathbb{R}^{M \times T}$
- La matrice de mélange $A \in \mathbb{R}^{M \times N}$
- Les sources $S \in \mathbb{R}^{N \times T}$

Mélanges instantanés

$$X = AS$$

$$\begin{pmatrix} x_1[1] & \dots & x_1[T] \\ \vdots & & \\ x_M[1] & \dots & x_M[T] \end{pmatrix} = \begin{pmatrix} a_{11} & \dots & \dots & a_{1N} \\ \vdots & & & \\ a_{M1} & \dots & \dots & a_{MN} \end{pmatrix} \begin{pmatrix} s_1[1] & \dots & s_1[T] \\ \vdots & & \\ s_N[1] & \dots & s_N[T] \end{pmatrix}$$

Mélanges convolutifs

Mélanges convolutifs

pour toutes les observations m , pour tous les instants t , on a

$$\mathbf{x}_m(t) = \sum_{n=1}^N a_{mn} \star s_n(t)$$

avec

- Les M observations $X \in \mathbb{R}^{M \times T}$
- Les $M \times N$ réponses impulsionnelles $a_{mn} \in \mathbb{R}^{T_a}$
- Les sources $S \in \mathbb{R}^{N \times T}$

Mélanges convolutifs

$$\begin{pmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_M \end{pmatrix} = \begin{pmatrix} \vec{a}_{11} & \dots & \dots & \vec{a}_{1N} \\ \vdots & & & \\ \vec{a}_{M1} & \dots & \dots & \vec{a}_{MN} \end{pmatrix} \star \begin{pmatrix} s_1[1] & \dots & s_1[T] \\ \vdots & & \\ s_N[1] & \dots & s_N[T] \end{pmatrix}$$

$$X = A \star S$$

Différentes classes de problèmes

Séparation non aveugle de sources

Hypothèses : On connaît le système de mélange A

But : Estimer les sources S à partir de X et A

- Séparation non aveugle
 - instantané
 - (sur)-déterminé
 - sous-déterminé
 - convolutif
 - (sur)-déterminé
 - sous-déterminé

Différentes classes de problèmes

Séparation aveugle de sources

But : Estimer les sources S et le système de mélange A à partir de X

- Séparation aveugle
 - instantané
 - (sur)-déterminé
 - sous-déterminé
 - convolutif
 - (sur)-déterminé
 - sous-déterminé

2 Mélanges instantanés sur-déterminé

2.1 Mélanges non aveugles

Problème

$$X = AS$$

avec

- Les M observations $X \in \mathbb{R}^{M \times T}$
- La matrice de mélange $A \in \mathbb{R}^{M \times N}$
- Les sources $S \in \mathbb{R}^{N \times T}$
- $M \geq N$

But : Estimer S à partir de A et X

Solution

Ici, la solution est simple et directe :

$$S = (A^T A)^{-1} A^T X$$

ce qui, dans le cas déterminé revient à

$$S = A^{-1} X$$

difficultés

si la matrice $A^T A$ est mal conditionnée, l'inversion peut-être difficile. On peut faire alors appel aux techniques de régularisation classiques.

2.2 Mélanges aveugles : ICA

2.2.1 ICA : formulation et pre-processing

Problème

$$X = AS$$

avec

- Les M observations $X \in \mathbb{R}^{M \times T}$
- La matrice de mélange $A \in \mathbb{R}^{M \times N}$
- Les sources $S \in \mathbb{R}^{N \times T}$
- $M \geq N$

But : Estimer S et A à partir de X

Autre formulation

Soit W une matrice de *demixage* tel que

$$S = WX$$

But : Estimer W à partir de X

Quelles hypothèses peut-on faire sur les sources ?

Analyse en composantes indépendantes

l'ICA cherche W telle que

$$WX = WAS = DPAS$$

avec

- D une matrice diagonale : *indétermination d'échelle*
- P une matrice de permutation : *indétermination d'ordre*
- Les M sources s_1, \dots, s_M sont *indépendantes*

Du sur-déterminé au déterminé

Hypothèse supplémentaire : $M = N$ (cas déterminé)

1ere étape de l'ICA dans le cas sur-déterminé

On se ramène à un problème déterminé par une SVD :

- On cherche les valeurs singulières et les vecteurs singuliers de X :

$$X = UYV^T$$

- On ne conserve que que les vecteurs singuliers correspondants aux M plus grandes valeurs singulières

Par la suite, on suppose $M = N$

Analyse en composantes indépendantes

Composantes Indépendantes ?

Définition : s_i et s_j sont indépendantes ssi

$$P(s_i, s_j) = P(s_i)P(s_j)$$

problème : on ne connaît pas les distributions ci-dessus.

Comment estimer l'indépendance ?

Indépendance et corrélation

Si les sources s_j sont gaussiennes, alors elles sont indépendantes ssi

$$E [SS^T] = I$$

Les séparer semble revenir à un simple blanchiment. . .

Cas de sources gaussiennes

Si les sources S sont gaussiennes et indépendantes (donc non corrélés). Si A est orthogonale, alors les observations

$$X = AS$$

sont toujours gaussiennes décorréelées.

la matrice A n'est pas identifiable dans le cas de sources gaussiennes

Hypothèse fondamentale de l'ICA

Au plus une source est gaussienne!

Blanchiment des observations

Si les sources S sont indépendantes, alors elles sont décorréelées

1ère étape de l'ICA : blanchir les observations

Blanchiment de X

à l'aide d'une ACP :

— Estimer la matrice de covariance des observations

$$\Sigma = XX^T$$

— La décomposer en valeurs propres :

$$D = P\Sigma P^T$$

— Choisir $B = D^{-1/2}P$ comme matrice de blanchiment :

$$\tilde{X} = BX = D^{-1/2}PX$$

2.2.2 ICA : principes généraux

ICA, reformulation

On observe

$$X = AS$$

tel que

— $M = N$

— au plus une source s_j est gaussienne

— les observations sont décorréelées, de variance 1 : $E [XX^T] = I$

— les observations sont centrées : $E[X] = 0$

But : estimer la matrice de démixage W , tel que

$$S = WX$$

à une indétermination d'échelle et de permutation près.

ICA : principe général

Soit J une mesure de l'indépendance de S . On cherche W tel que

$$\max_W J(WX) \quad \text{s.t. } WW^T = I \quad (1)$$

quelle mesure J choisir ?

ICA : indépendance et non gaussianité

Idée : le TCL dit qu'une somme de V.A. iid tend vers une distribution gaussienne.

Indépendance = non gaussianité

Mesures de non gaussianité :

— Kurtosis

$$k(y) = E[y^4] - 3$$

si y est gaussienne, alors $k(y) = 0$.

— Information mutuelle, directement reliée à la notion de neg-entropy.

ICA : negentropy

Entropy d'une source

$$H(Y) = - \sum_k P(Y = a_k) \log P(Y = a_k)$$

Neg-entropy

$$J(Y) = H(Y_{gauss}) - H(Y)$$

Approximation de la neg-entropy

$$J(Y) \simeq [E[G(Y)] - E[Y_{gauss}]]^2$$

où G est une fonction non quadratique. En pratique

$$G(u) = \frac{1}{a_1} \log \cosh(a_1 u) \quad G(u) = -e^{-u^2/2}$$

avec $1 \leq a_1 \leq 2$

2.2.3 Deux approches d'ICA

Fast ICA

Fast ICA propose de résoudre le problème suivant :

$$\max J(WX) = E[G(WX)] \text{ s.t. } WW^T = I$$

avec un algorithme de poursuite

Extraction d'une composante w_p

1. Initialisation de w_p (aléatoire)
2. itérer jusqu'à convergence :

(a)

$$\mathbf{w} = E[Xg'(\mathbf{w}^T X)^T] - E[g''(\mathbf{w}^T X)]\mathbf{w}$$

(b)

$$\mathbf{w} = \mathbf{w} / \|\mathbf{w}\|$$

si $g(u) = \log \cosh(u)$ on a $g'(u) = \tanh(u)$ et $g''(u) = 1 - \tanh(u)^2$

Fast ICA

Extraction de toutes les composantes :

- Extraire une composante w_{p+1}
- Faire une orthogonalisation de Gram-Schmidt avec les p premières composantes.

Fast ICA

Pour chaque composante $p \in \{1, \dots, N\}$

1. Initialisation de w_p (aléatoire)
2. itérer jusqu'à convergence :

(a)

$$\mathbf{w}_p = \mathbb{E} [Xg'(\mathbf{w}_p^T X)^T] - \mathbb{E} [g''(\mathbf{w}_p^T X)] \mathbf{w}_p$$

(b)

$$\mathbf{w}_p = \mathbf{w}_p - \sum_{k=1}^{p-1} \mathbf{w}_p^T \mathbf{w}_k \mathbf{w}_k$$

(c)

$$\mathbf{w}_p = \mathbf{w}_p / \|\mathbf{w}_p\|^2$$

ICA par maximum de vraisemblance

Modèle de mélange :

$$X = AS$$

On suppose que :

- $W = A^{-1}$
 - les sources s_n sont iid avec pour densité de probabilité p_s , supposée connue
- On a alors la vraisemblance du modèle

$$p(X|A) = p_s(WX) \det W$$

et la log-vraisemblance s'écrit :

$$L(W) = \log p(X|A) = \mathbb{E} \left[\sum_n p_s(w_n^T X) \right] + T \log |\det W|$$

ICA par maximum de vraisemblance

On estime alors W par maximum de vraisemblance :

$$W = \underset{W}{\operatorname{argmin}} \mathbb{E} \left[\sum_n -\log(p_s(w_n^T X)) \right] - T \log |\det W|$$

les p_s sont un paramètre du modèle et reflète l'a priori qu'on a sur les sources s

choix possible pour p_s :

- Laplace : $e^{-\lambda\|s\|} \rightarrow -\log(p_s(\mathbf{s})) = \lambda\|\mathbf{s}\|_1$
- $-\log(p_s(\mathbf{s})) = \lambda\|\mathbf{s}\|_0$
- $\log(p_s(\mathbf{s})) = \text{kurtosis}$
- $p_s(\mathbf{s}) = 1 - \tanh^2(\mathbf{s})$
- ...

2.2.4 ICA dans un domaine transformé

Limitation principale

On a formulé l'approche ICA pour

$$X = AS$$

En pratique, on cherche des sources S telles que leur distribution s'éloigne de la gaussienne par

- maximisation du kurtosis (FastICA)
- maximisation d'un critère de neg entropy (FastICA)
- minimisation ℓ_1, ℓ_0 (Maximum de vraisemblance)

Remarque

En audio, les sources S ont une distribution d'allure gaussienne dans le domaine temporel

Comment appliquer le principe d'ICA ?

ICA dans le domaine transformé

Soit Φ une transformée de type Gabor. Alors on a :

$$X = AS \Rightarrow X\Phi^* = AS\Phi^* \Rightarrow \Gamma = A\alpha$$

avec

- $\Gamma = X\Phi^*$ les coefficients STFT de X
- $\alpha = S\Phi^*$ les coefficients STFT de S

On applique le principe d'ICA dans le domaine transformée, sur les coefficients Γ .

Estimation des sources :

$$S = W\Gamma\Phi$$

3 Mélanges instantanés sous-déterminé

3.1 Parcimonie et séparation de sources

3.1.1 Exemples jouets

Un premier cas (très) simple - (1)

On considère un mélange stéréo particulier :

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = A \begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix}$$

Dans ce cas :

$$\begin{aligned} x_1[t] &= a_{11}s_1[t] + a_{12}s_2[t] + a_{13}s_3[t] \\ x_2[t] &= a_{21}s_1[t] + a_{22}s_2[t] + a_{23}s_3[t] \end{aligned}$$

Mais, si l'on se restreint à l'intervalle $[0, t_1]$

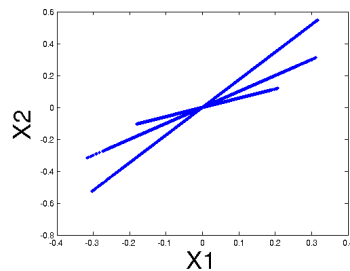
$$\begin{cases} x_1[t] = a_{11}s_1[t] \\ x_2[t] = a_{21}s_1[t] \end{cases} \implies x_2[t] = \frac{a_{21}}{a_{11}}x_1[t]$$

De même pour les autres intervalles de temps.

Un premier cas (très) simple - (2)

Comme pour tout $i \in 1, 2, 3$, $x_2[t] = \frac{a_{1i}}{a_{2i}} x_1[t]$, on peut en déduire A .

Graphiquement : on représente x_2 en fonction de x_1

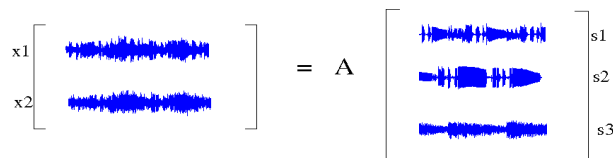


On peut alors estimer la matrice A et les sources S par un simple clustering :

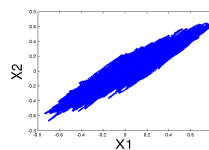
$$\hat{s}_i = \{x_i[t] : x_i[t] \in C_i\}$$

Mais en réalité...

Les mélanges sont du type



Si l'on dessine x_2 en fonction de x_1 on obtient :



clustering impossible!

La méthode ne fonctionne que si "une seule source est active pour un t donné"

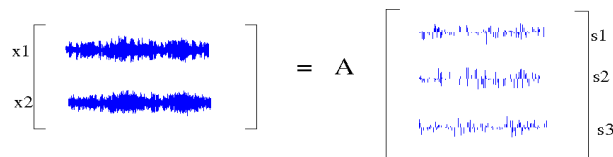
Autrement dit : les sources doivent être *parcimonieuses*

3.1.2 Influence de la parcimonie

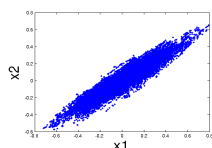
Quelle parcimonie ?

Hypothèse de séparabilité (1) : un seul signal actif à un instant t donné \Rightarrow les signaux doivent être parcimonieux.

Si la parcimonie n'est "que" sur les sources : cas particuliers où les sources sont parcimonieuses, mais actives **en même temps**



On dessine x_2 en fonction de x_1 :



clustering impossible!

Comment rendre les signaux parcimonieux à travers les canaux ?

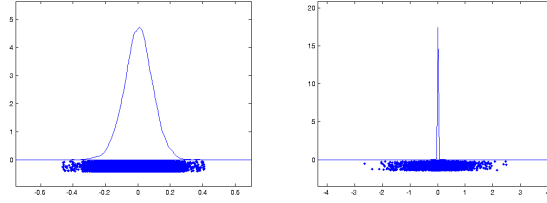


FIGURE : Gauche : pdf des échantillons. Droite : pdf des coefficients MDCT.

Résumons le problème

Pour l'instant

Le problème est

$$X = AS$$

Mais la séparation est impossible car l'hypothèse de séparabilité (1) n'est pas vérifiée.

But

Pour vérifier l'hypothèse (1), *il faut* trouver une représentation de S parcimonieuse.

ie : trouver une transformation Φ telle que $S = \underline{s}\Phi^T$ où \underline{s} est "parcimonieux"

Le problème devient alors

$$X = A \underline{s}\Phi^T$$

3.2 Masquage temps-fréquence

3.2.1 Une seule source active

Concentration de l'information

On représente les densités de probabilité des échantillons d'un signal et des coefficients MDCT de ce même signal

Une "bonne transformée" permet de concentrer l'information dans quelques coefficients

Exemples de transformée

- Fourier ;
- Gabor ou STFT ;
- MDCT ;
- Ondelettes ;
- *-lettres ;

Masquage temps-fréquence - (1)

Si l'on trouve une représentation parcimonieuse, on espère que dans ce nouveau repère l'hypothèse (1) est vérifiée.

Exemple avec une transformation dans une base orthonormée

Soit Φ une matrice dont les colonnes sont formées par les atomes d'une BON. On a $S = \underline{s}\Phi$

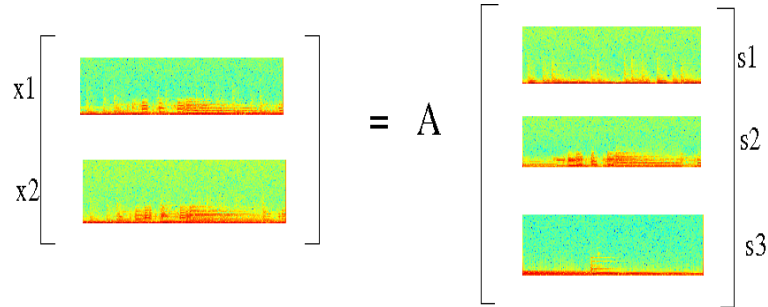
$$\begin{aligned}
 X = AS = A\underline{s}\Phi &\iff X\Phi^T = A\underline{s} \iff \underline{x} = A\underline{s} \\
 \iff \begin{pmatrix} \underline{x}_1[k_1, f_1] & \dots & \underline{x}_1[k_K, f_F] \\ \vdots \\ \underline{x}_M[k_1, f_1] & \dots & \underline{x}_M[k_K, f_F] \end{pmatrix} &= A \begin{pmatrix} \underline{s}_1[k_1, f_1] & \dots & \underline{s}_1[k_K, f_F] \\ \vdots \\ \underline{s}_N[k_1, f_1] & \dots & \underline{s}_N[k_K, f_F] \end{pmatrix}
 \end{aligned}$$

Si l'on choisit un repère non orthogonal (ex : de Gabor), on écrit en général l'équation du mélange directement dans ce repère :

$$\underline{x} = A\underline{s} \quad \text{où} \quad \underline{x} = X\Phi^T, \quad \Phi \text{ non orthogonale}$$

Masquage temps-fréquence - (2)

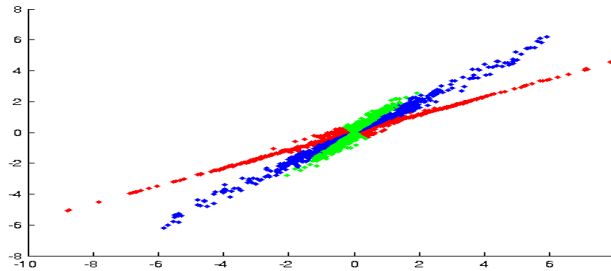
En image :



On espère que l'hypothèse “une seule source active par case temps fréquence” est vérifiée

Masquage temps-fréquence - (3)

On dessine \underline{x}_2 en fonction de \underline{x}_1



Les directions de la matrice de mélange apparaissent !

Le clustering et la séparation deviennent possible.

$$\hat{\underline{s}}_i = \{ \underline{x}_i[k, f] : \underline{x}_i[k, f] \in \mathcal{C}_i \}$$
$$\hat{\underline{s}}_i = \hat{\underline{s}}_i \Phi$$

Avantages et inconvénients du masquage temps-fréquence

Avantages

- Très simple à mettre en oeuvre ;
- Très rapide ;
- Permet l'estimation de la matrice de mélange A et de sources S .

Inconvénients

- Le clustering n'est pas évident ;
- Suppose **une seule** source active par case temps-fréquence, les estimations sont *très* (voire trop) parcimonieuses ;
- La “sur”-parcimonie limite les performances théoriques de la qualité des sources estimées.

Comment utiliser plusieurs sources actives par case temps-fréquence ?

3.3 Approche type Basis Pursuit

Une autre idée...

Plutôt que de supposer une seule source active et se “limiter” à l’équation

$$\underline{x}_1[k, f] = a_{1,i} \underline{s}_i[k, f] \implies \underline{s}_i[k, f] = \underline{x}_1[k, f]/a_{1,i}$$

On peut supposer qu’il y a exactement autant de sources actives qu’il y a d’observations.

Supposant la matrice de mélange A connue, cette hypothèse permet d’inverser localement A :

$$\begin{pmatrix} \underline{x}_1[k, f] \\ \underline{x}_2[k, f] \end{pmatrix} = \begin{pmatrix} a_{1i} & a_{1j} \\ a_{2i} & a_{2j} \end{pmatrix} \begin{pmatrix} \underline{s}_i[k, f] \\ \underline{s}_j[k, f] \end{pmatrix} \implies \begin{pmatrix} \hat{\underline{s}}_i[k, f] \\ \hat{\underline{s}}_j[k, f] \end{pmatrix} = A_{ij}^{-1} \begin{pmatrix} \underline{x}_1[k, f] \\ \underline{x}_2[k, f] \end{pmatrix}$$

Comment choisir ces sources ?

Basis Pursuit

L’hypothèse de parcimonie se traduit ici par M sources actives pour M observations.

Choisir autant de sources qu’il y a d’observations revient à choisir une base parmi les vecteurs $\{a_{:,1}, \dots, a_{:,M}\}$ dans laquelle représenter les observations $\{\underline{x}_1[k, f], \dots, \underline{x}_M[k, f]\}$.

On va donc choisir la “meilleure base” au sens où l’on cherche à minimiser une mesure de diversité E des sources $\underline{s}[k, f]$. Avec $E = \|\cdot\|_p^p$, $p < 2$:

$$\begin{aligned} & \min_{\underline{s}[k, f]} \|\underline{s}[k, f]\|_p^p \\ & \text{sous contrainte } \underline{\mathbf{x}}[k, f] = A\underline{\mathbf{s}}[k, f] \end{aligned}$$

Quand $p = 1$, on reconnaît le problème du *Basis Pursuit* (Chen et Donoho, 1998)

Résolution par FOCUSS

L’algorithme FOCUSS (Rao *et al*, 1999) permet de résoudre le problème précédent

$$\text{On a } \nabla_{\underline{s}_i[k, f]} \|\underline{s}_{k,f}\|_p^p = |p| \Pi(\underline{s}_{k,f}) \underline{s}_{k,f} \text{ avec } \Pi(\underline{s}_{k,f}) = \text{diag}(|s_{k,f,i}|^{p-2})$$

FOCUSS

En partant de $\underline{s}^{(0)}[k, f] = A^T(AA^T)^{-1}\underline{\mathbf{x}}[k, f]$

$$\underline{s}_{k,f}^{(r+1)} = \Pi^{-1}(\underline{s}_{k,f}^{(r)}) A^T \left(A \Pi^{-1}(\underline{s}_{k,f}^{(r)}) A^T \right)^{-1} \underline{\mathbf{x}}[k, f]$$

Avantages et inconvénients

Avantages

- Permet d’améliorer les estimations ;
- L’inversion locale permet une “remise à l’échelle” mieux adaptée ;
- Le choix de deux sources actives au lieu d’une seule permet théoriquement d’améliorer significativement les estimations.

Inconvénients

- Plus lent que le masquage simple (recherche des M “meilleures” sources actives au lieu d’une seule) ;
- Ne permet pas d’estimer A
- Il y a plus d’artefact dus aux sources voisines dans les estimations (mais compensé par le fait que l’estimation est meilleure...);

Et dans le cas où le mélange n’est plus instantané ?

4 Mélanges convolutifs non aveugle

4.1 Problème et 1ère approche

mélange convolutif : problème

On considère le cas où A est un “système de filtres” :

$$\begin{aligned} X &= A \star S \\ X &= \begin{pmatrix} \vec{a}_{11} & \dots & \dots & \vec{a}_{1N} \\ \vdots & & & \\ \vec{a}_{M1} & \dots & \dots & \vec{a}_{MN} \end{pmatrix} \star S \end{aligned}$$

où $X \in \mathbb{R}^{M \times T_x}$, $S \in \mathbb{R}^{N \times T_s}$ et $\vec{a}_{ij} \in \mathbb{R}^K$. On a $T_x = T_s + K - 1$.

Dans le cas “anéchoïque” A est du type $\vec{a}_{ij} = (0, \dots, 0, 1, 0, \dots, 0)$.

On considère toujours l’hypothèse de parcimonie, on choisi donc une matrice Φ telle que :

$$X = A \star (\underline{s}\Phi)$$

Mais on ne peut plus écrire

$$X\Phi^T = A \star \underline{s} \tag{2}$$

$$\underline{\mathbf{x}} = A \star \underline{\mathbf{s}} \tag{3}$$

Comment exploiter la parcimonie ?

Une première approximation

On considère la transformée de Fourier à court terme, qu’on applique à X et à $A \star S$. On peut alors considérer, $\forall k, f$:

$$\underline{\mathbf{x}}[k, f] \simeq A[f] \underline{\mathbf{s}}[k, f]$$

On peut alors résoudre :

$$\begin{aligned} \min_{\underline{\mathbf{s}}} \quad & \|\underline{\mathbf{s}}[t, f]\|_1 \\ \text{s.c.} \quad & \|\underline{\mathbf{x}}[t, f] - A[f] \underline{\mathbf{s}}[t, f]\| \leq \varepsilon \end{aligned}$$

Mais ça reste une approximation + Problème s’il y a du bruit...

4.2 Approche variationnelle

Vers une fonctionnelle convexe

Remarque : la parcimonie est “à travers les canaux”. On peut alors écrire le problème

$$\min_{\underline{\mathbf{s}}} \|\underline{\mathbf{s}}[t, f]\|_1 \quad \text{s.c.} \quad \|\underline{\mathbf{x}}[t, f] - A[f] \underline{\mathbf{s}}[t, f]\|_2^2 \leq \varepsilon$$

comme

$$\min_{\underline{\mathbf{s}}} \sum_{t,f} \sum_c |\underline{\mathbf{x}}[t, f] - A[f] \underline{\mathbf{s}}[t, f]|^2 + \lambda \sum_{t,f} \left(\sum_c |\underline{\mathbf{s}}_c[t, f]| \right)$$

Problèmes :

— λ doit être “assez” petit pour avoir la contrainte d’égalité $\underline{\mathbf{x}}[t, f] = A \underline{\mathbf{s}}[t, f]$.

— Si on “relache” cette contrainte, le risque est de trop pénaliser les hautes fréquences.

Solution : utiliser une norme mixte

$$\min_{\underline{s}} \sum_{t,f} \sum_c |\underline{x}[t, f] - A[f] \underline{s}[t, f]|^2 + \lambda \sum_{t,f} \left(\sum_c |\underline{s}_c[t, f]| \right)^2$$

Vers une fonctionnelle convexe

Solution : utiliser une norme mixte

$$\min_{\underline{s}} \sum_{t,f} \sum_c |\underline{x}[t, f] - A[f] \underline{s}[t, f]|^2 + \lambda \sum_{t,f} \left(\sum_c |\underline{s}_c[t, f]| \right)^2$$

Problème : le terme d’attache aux données traduit mal le bruit supposé gaussien et utilise toujours la même approximation.

Solution : écrire une attache aux données dans le domaine “temporel”

$$\min_{\underline{s}} \|X - A \star (\underline{s}\Phi)\|^2 + \lambda \|\underline{s}\|_{12}^2$$

Algorithmes

Pour résoudre le problème d’optimisation :

— Algorithmes de seuillage itératifs classiques : beaucoup trop lent ! (surtout avec la convolution)

— Utiliser les schéma de Nesterov (algo du 1er ordre optimal).

Les deux sont basés sur la notion d’opérateur de proximité :

definition : Proximity operator

Let $\varphi : \mathbb{C}^I \rightarrow \mathbb{C}$ be a lower semicontinuous, convex function. The proximity operator associated with φ denoted by $\text{prox}_\varphi : \mathbb{C}^I \rightarrow \mathbb{C}^I$ is given by

$$\text{prox}_\varphi(\mathbf{z}) = \frac{1}{2} \underset{\mathbf{u} \in \mathbb{C}^I}{\text{argmin}} \|\mathbf{z} - \mathbf{u}\|_2^2 + \varphi(\mathbf{u}). \quad (4)$$

Algorithmes

Prox du LASSO $\min_{\mathbf{u}} \|\mathbf{z} - \mathbf{u}\|_2^2 + \lambda \|\mathbf{u}\|_1$

$$\hat{u}_{t,f,c} = \text{sgn}(z_{t,f,c}) (|z_{t,f,c}| - \lambda)^+$$

Prox du E-LASSO $\min_{\mathbf{u}} \|\mathbf{z} - \mathbf{u}\|_2^2 + \lambda \|\mathbf{u}\|_{1,2}$

$$\hat{u}_{t,f,c} = \text{sgn}(z_{t,f,c}) \left(|z_{t,f,c}| - \frac{\lambda}{1 + \lambda L_{t,f}} \|\mathbf{z}_{t,f}\| \right)^+$$

Algorithmes

But : trouver $\text{argmin} \mathcal{L}(\mathbf{s}) + \lambda \mathcal{P}(\mathbf{s})$, avec \mathcal{L} L -Lipshitz différentiable

ISTA (en $O(1/k)$)

Initialization : $\tilde{\mathbf{s}}^{(0)} \in \mathbb{C}^{N \times B}$, $k = 1$.

Repeat :

$$\tilde{\mathbf{s}}^{(k)} = \text{prox}_{\frac{\lambda}{L} \mathcal{P}} \left(\tilde{\mathbf{s}}^{(k-1)} - \frac{\nabla \mathcal{L}(\tilde{\mathbf{s}}^{(k-1)})}{L} \right)$$

FISTA (en $O(1/k^2)$)

Initialization : $\tilde{\mathbf{s}}^{(0)} \in \mathbb{C}^{N \times B}$, $\mathbf{z}^{(0)} = \tilde{\mathbf{s}}^{(0)}$, $t^{(0)} = 1$, $k = 1$.

Repeat :

- $\tilde{\mathbf{s}}^{(k)} = \text{prox}_{\frac{\lambda}{L}\mathcal{P}} \left(\mathbf{z}^{(k-1)} - \frac{\nabla \mathcal{L}(\mathbf{z}^{(k-1)})}{L} \right)$
- $\tau^{(k)} = \frac{1 + \sqrt{1 + 4\tau^{(k-1)^2}}}{2}$
- $\mathbf{z}^{(k)} = \tilde{\mathbf{s}}^{(k)} + \frac{\tau^{(k-1)} - 1}{\tau^{(k)}} (\tilde{\mathbf{s}}^{(k)} - \tilde{\mathbf{s}}^{(k-1)})$
- $k = k + 1$

Results

TABLE : Average SDR in decibels as a function of RT_{60} and d over speech mixtures with $N = 4$ sources.

RT ₆₀	d	narrowband				wideband	
		DUET	ℓ_1 min.	Lasso	E-Lasso	Lasso	E-Lasso
anechoic	5 cm	5.4	4.5	6.4	6.7	5.7	6.5
	1 m	3.6	7.7	7.8	8.0	7.6	8.1
50 ms	5 cm	5.1	4.2	4.3	4.3	4.4	4.5
	1 m	3.1	6.3	6.4	6.4	7.0	7.4
250 ms	5 cm	2.4	1.8	2.1	2.1	5.9	5.0
	1 m	1.0	2.8	3.0	3.4	7.6	7.2

TABLE : Average SDR, SIR and SAR in decibels over speech mixtures with $N = 4$ sources, $\text{RT}_{60} = 250$ ms and $d = 1$ m.

method	narrowband				wideband	
	DUET	ℓ_1 min.	Lasso	E-Lasso	Lasso	E-Lasso
SDR	1.0	2.8	3.0	3.4	7.6	7.2
SIR	8.2	6.4	6.8	7.7	14.0	13.9
SAR	2.8	6.5	6.4	6.4	9.1	8.5

Results

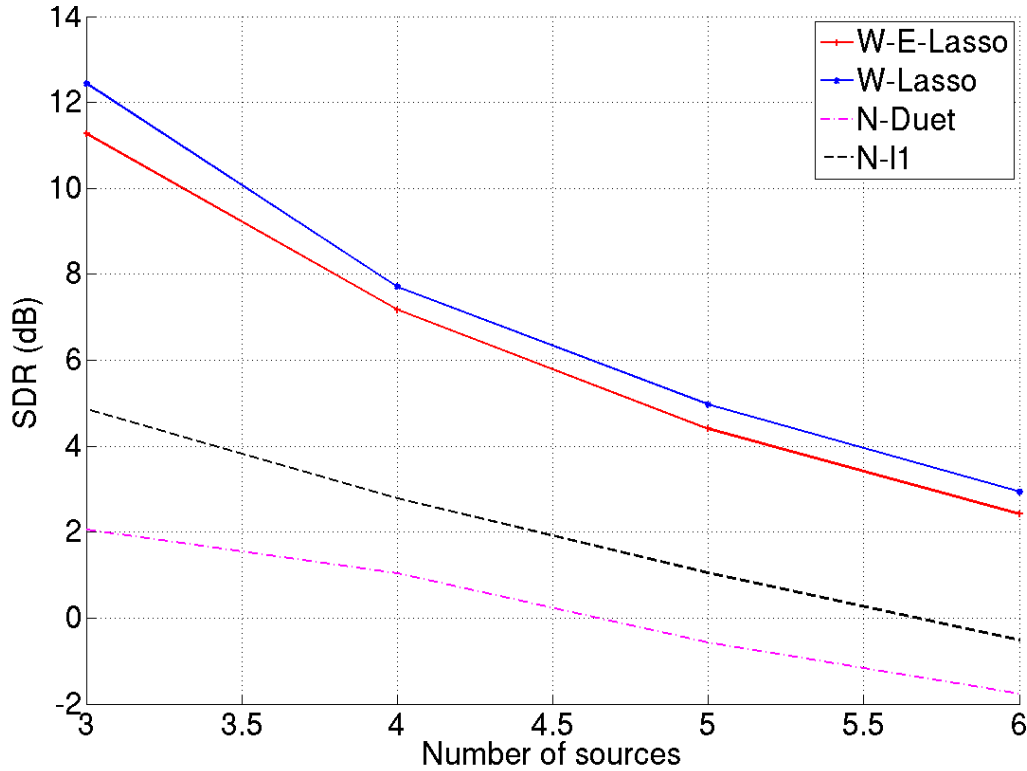


FIGURE : Variation of the average SDR as a function of N over speech mixtures with $\text{RT}_{60} = 250$ ms and $d = 1$ m.

Résultats

Results

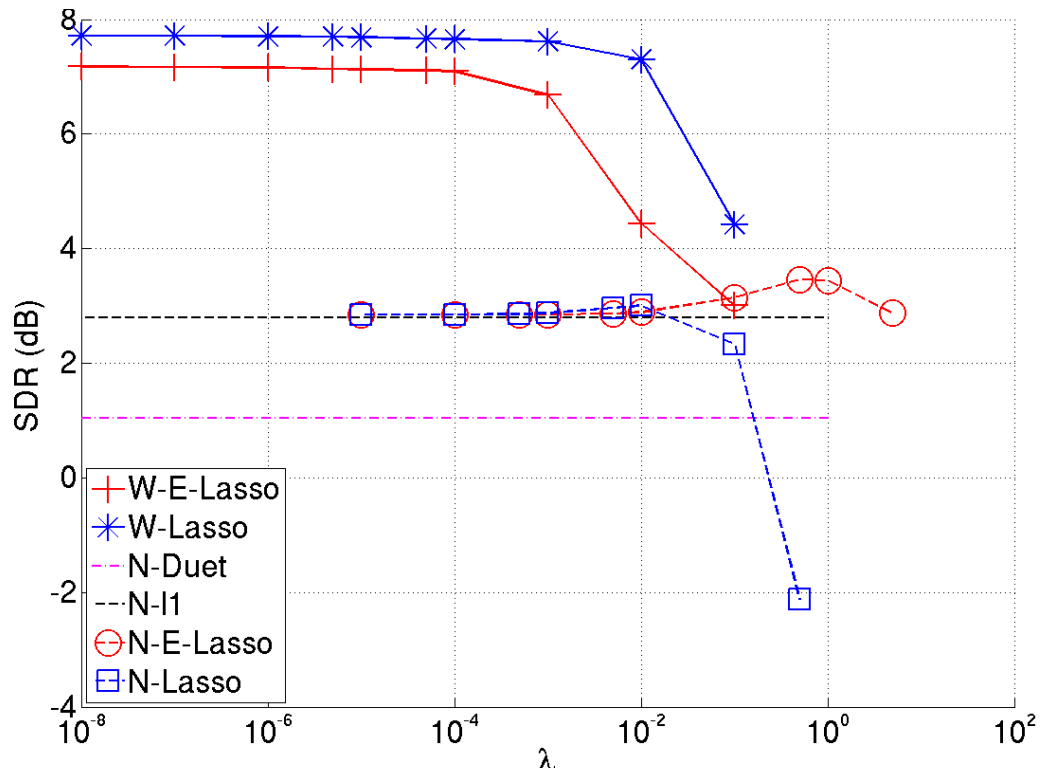


FIGURE : Variation of the average SDR as a function of λ over speech mixtures with $N = 4$ sources, $RT_{60} = 250$ ms and $d = 1$ m.

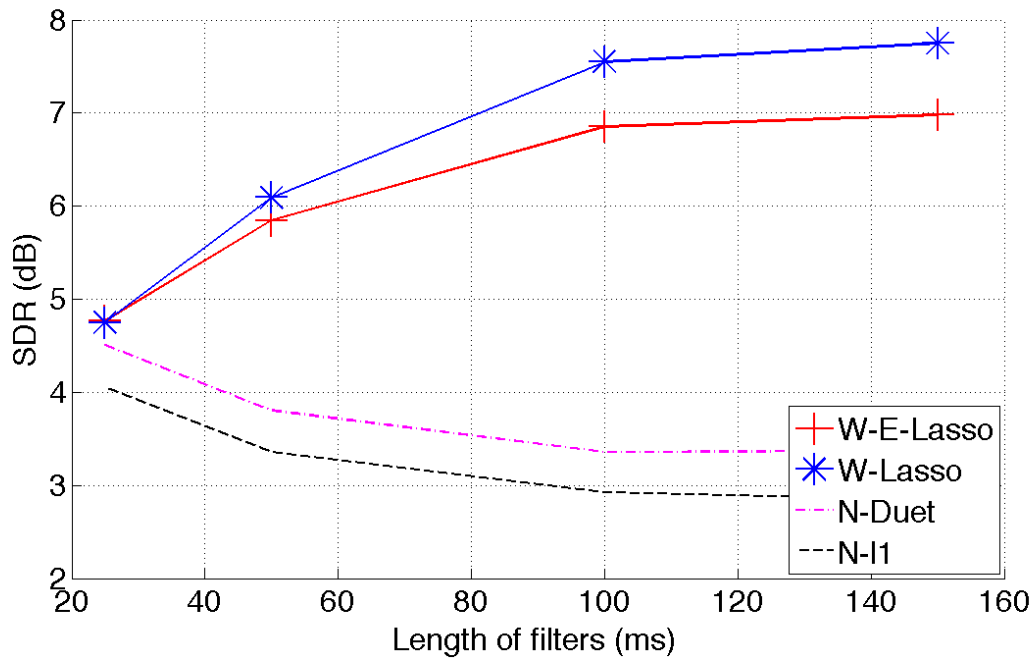


FIGURE : Variation of the average SDR as a function of the length of the filters over speech mixtures with $RT_{60} = 250$ ms and $d = 1$ m.

Résultats

Résultats

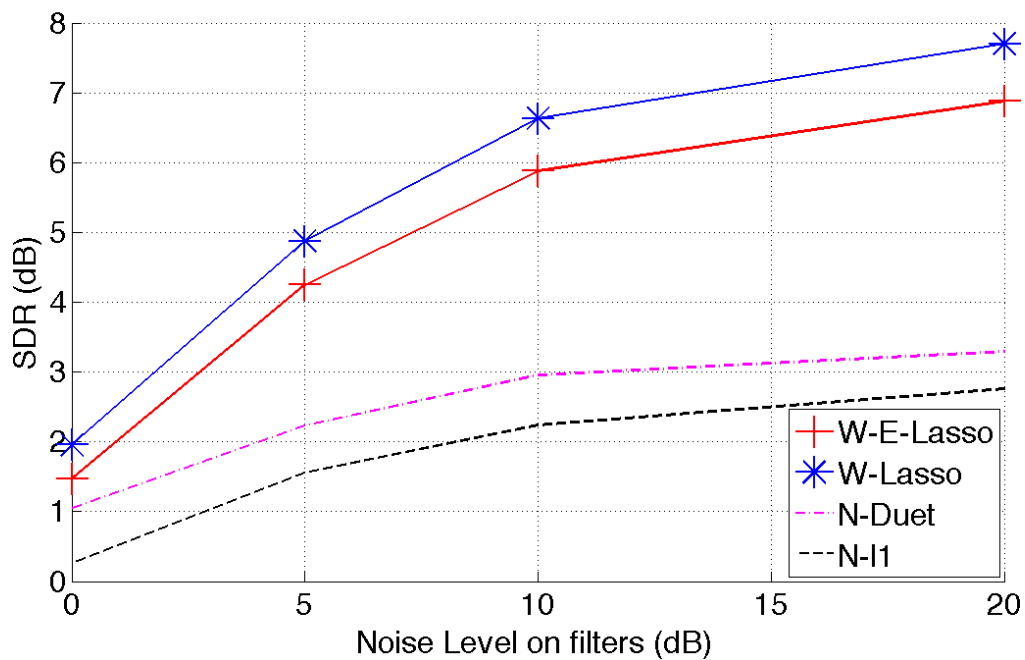


FIGURE : Variation of the average SDR as a function of the noise level over speech mixtures with $RT_{60} = 250$ ms and $d = 1$ m.

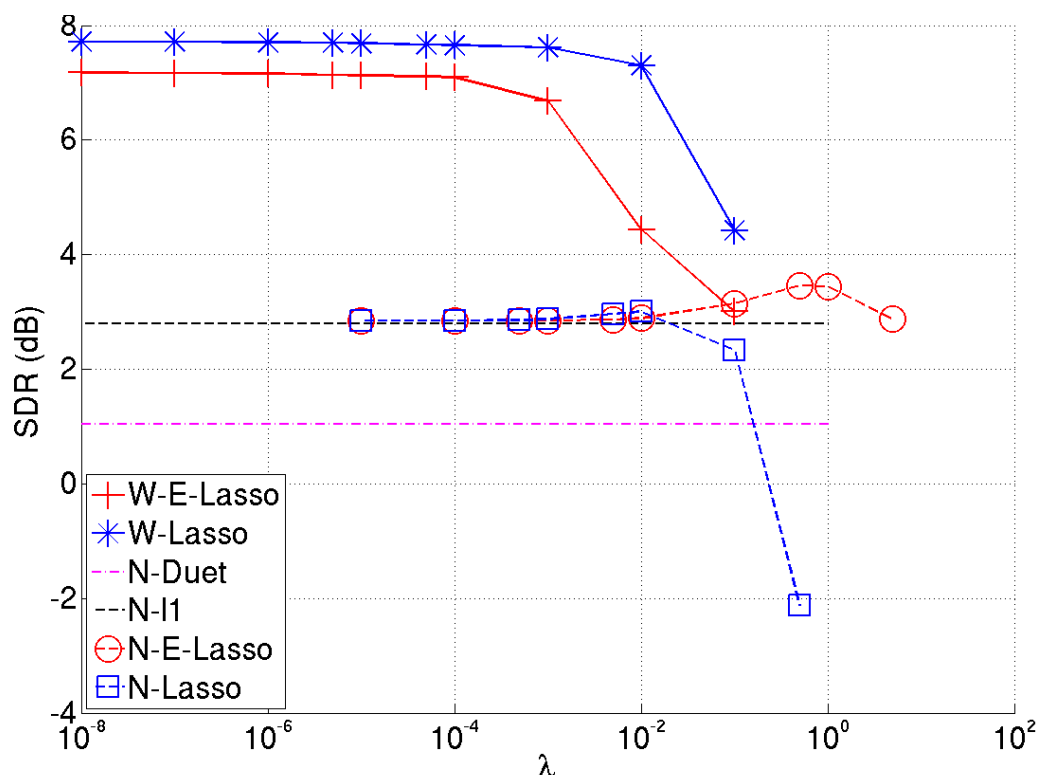


FIGURE : Variation of the average SDR as a function of λ over speech mixtures with $N = 4$ sources, $RT_{60} = 250$ ms and $d = 1$ m.

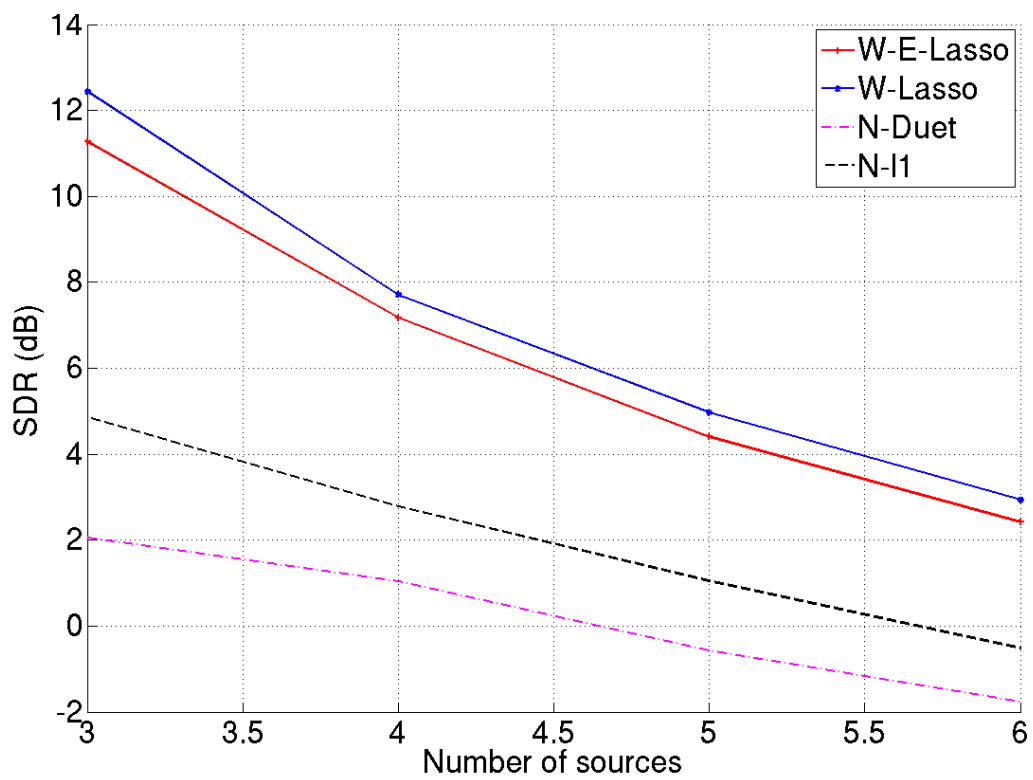


FIGURE : Variation of the average SDR as a function of N over speech mixtures with $RT_{60} = 250$ ms and $d = 1$ m.