Analyse par le spectre des valeurs singulières (Singular Spectrum Analysis) : introduction et exploration

F. Auger

IREENA

17 octobre 2014

C’est une technique d’analyse des signaux

- sans modèle *a priori*, dirigée par les données, non paramétrique
- qui fournit une décomposition d’un signal en composantes “significatives”
- basée sur la décomposition en éléments propres d’une “matrice de covariance”
- qui n’a qu’un seul paramètre de réglage
- qui conduit au concept de *séparabilité au sens du SSA*

Introduction

Applications

- analyse exploratoire
- débruitage
- prédiction, interpolation
- estimation de paramètres
- détection de défauts/ruptures
- estimation non paramétrique de densités de probabilité

Domaines d'application

- météorologie, climatologie, océanographie
- astronomie
- économie, finances
- médecine

- Deux “foyers” de création de la méthode :
  - occidental : Ghil *et al*, 1989-2012
  - pays de l’est : Caterpillar SSA, Golyandina, Zhigljavsky, Hassani *et al*, 2001-2013
- beaucoup de publications de congrès dans les domaines d’applications (géophysique, biomédical, économie . . .), peu (pas ?) de publications en théorie du signal

Principe et algorithme

1. Construction de la matrice de trajectoire

À partir d'un enregistrement de N points d'un signal réel, $x_1, x_2, x_3 \dots x_N$, on construit une matrice de trajectoire de L lignes et $K = N - L + 1$ colonnes

$$X = \begin{pmatrix} x_1 & x_2 & x_3 & \dots & x_K \\ x_2 & x_3 & x_4 & \dots & x_{K+1} \\ x_3 & x_4 & x_5 & \dots & x_{K+2} \\ \vdots & \vdots & \vdots & \vdots & \\ x_L & x_{L+1} & x_{L+2} & \dots & x_N \end{pmatrix}$$

X est une matrice de Hankel

Principe et algorithme

2. Décomposition en éléments propres

- On calcule les valeurs propres λ_i, $i \in [1; L]$ et les vecteurs propres U_i de $S = X\,X^T$, qui est un estimateur biaisé de la matrice de covariance de x_i.
- Les valeurs propres λ_i, triées dans l'ordre décroissant, forment *le spectre des valeurs singulières de* X.
- On calcule alors les vecteurs $V_i = X^T\,U_i$, vecteurs propres de $X^T\,X$, et $Y_i = U_i\,V_i^T$.
- On a alors

$$X = \sum_{i=1}^{d} Y_i \quad \text{avec} \quad d = \max\{i \text{ tel que } \lambda_i > 0\}$$

Principe et algorithme

3. Synthèse des composantes

- La matrice (L, K) est convertie en une composante y_i en *moyennant les termes anti-diagonaux*

$$y_{i,s} = \frac{1}{|A_s|} \sum_{(j,k)\in A_s} X_{i,jk}$$

avec $A_s = \{(j, k) \text{ tel que } j + k = s + 1, 1 \leq j \leq L, 1 \leq K \leq K\}$

- On a alors

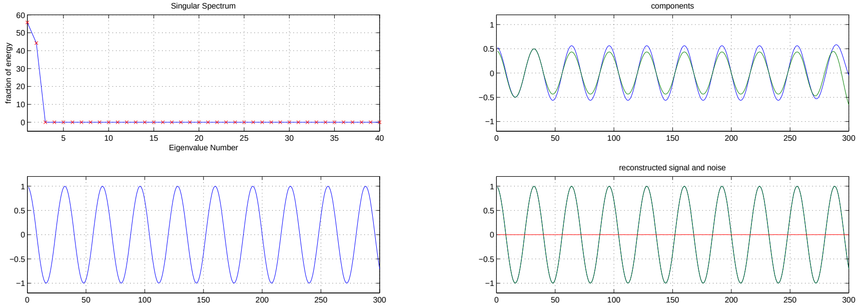
$$x_n = \sum_{i=1}^{d} y_{i,n} \quad \text{avec} \quad d = \max\{i \text{ tel que } \lambda_i > 0\}$$

Propriétés
linéarité

Si $x'_n = \alpha x_n$,
alors $X' = \alpha X$, $S' = \alpha^2 S$, $\lambda'_i = \alpha^2 \lambda_i$, $U'_i = U_i$, $V'_i = \alpha V_i$,
donc $Y'_i = \alpha Y_i$ et $y'_{i,n} = \alpha y_{i,n}$
x_n et $x'_n = \alpha x_n$ ont le même spectre des valeurs singulières normalisé

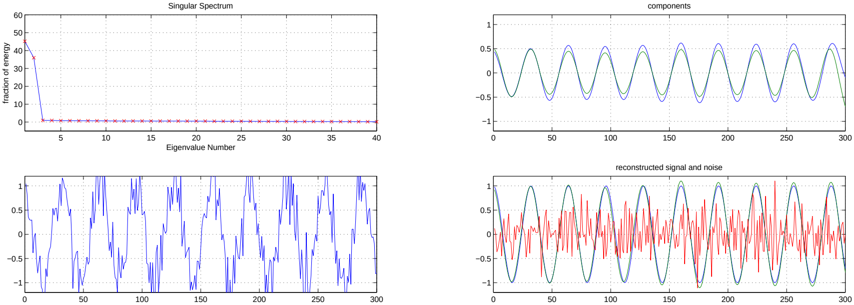
Résultats

Une sinusoïde pure $\lambda = 1/32$

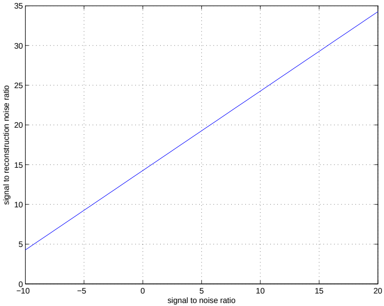


Résultats

Une sinusoïde bruitée (RSB=6 dB)

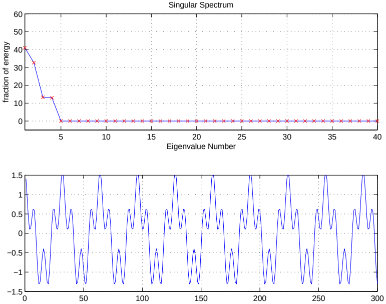


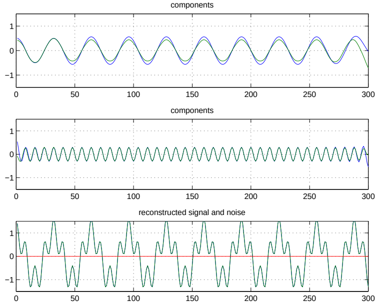
Résultats
Gain en débruitage



Résultats

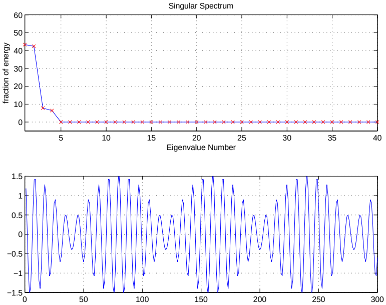
Deux sinusoïdes pures ($\lambda_1 = 1/32$, $\lambda_2 = 4/32$)

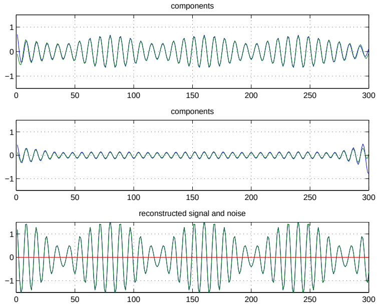




Résultats

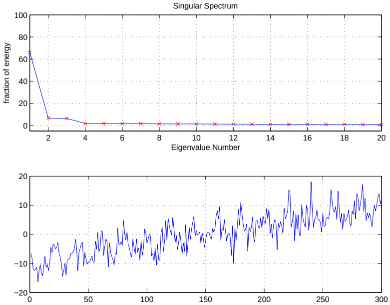
Deux sinusoïdes pures ($\lambda_1 = 3.6/32$, $\lambda_2 = 4/32$)

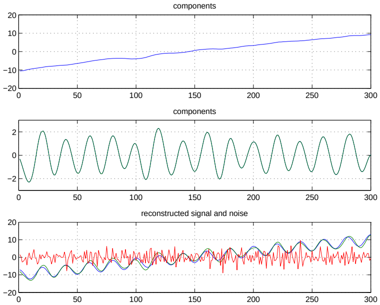




Résultats

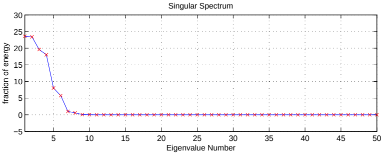
Une tendance linéaire avec une sinusoïde bruitée (RSB=6 dB)

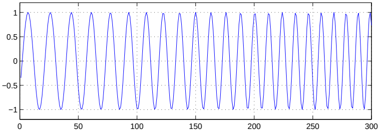


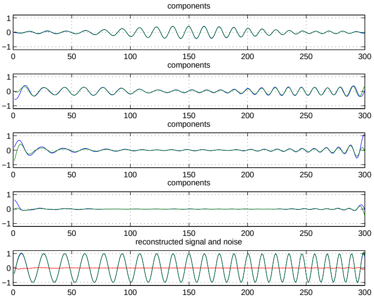


Résultats

modulation linéaire de fréquence de 0.05 à 0.1

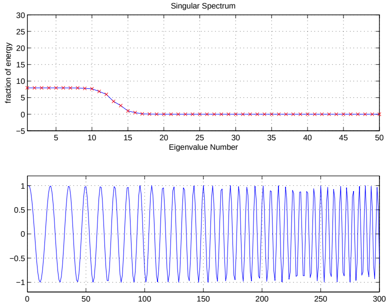


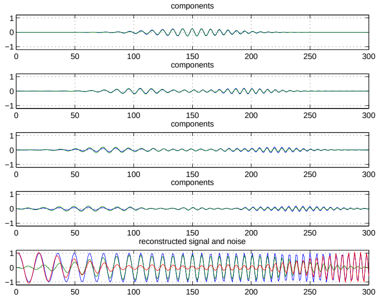




Résultats

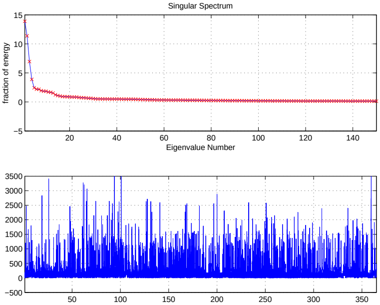
modulation linéaire de fréquence de 0.05 à 0.2

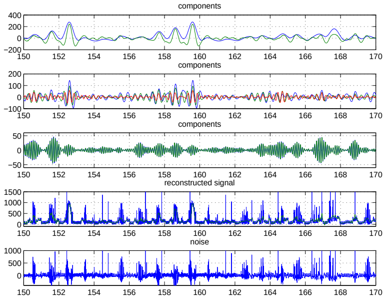


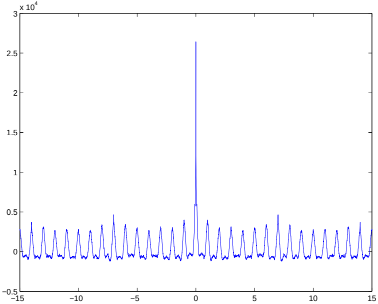


Résultats

consommation éléctrique d'une habitation







x 10^4
3
2.5
2
1.5
1
0.5
0
−0.5
−15
−10
−5
0
5
10
15

Conclusion

- Comment choisir L ? Il faut prendre L au moins égal à la plus grande période recherchée dans le signal.
- Comment savoir où sont les composantes recherchées ?
- Séparabilité ? Séparabilité asymptotique ? Quand y a t-il une ou deux composantes ?
- Perspectives ?
  - Séparabilité (résolution) de deux composantes sinusoïdales

$$\begin{aligned} x_n &= a_1 \cos(2\pi\lambda_1 n) + a_2 \cos(2\pi\lambda_2 n) \\ &= a_1 \left(\cos(2\pi\lambda_1 n) + \frac{a_2}{a_1} \cos(2\pi \frac{\lambda_2}{\lambda_1} \lambda_1 n) \right) \end{aligned}$$

  - Liens avec les méthodes de décomposition en sous espaces (Pisarenko, ESPRIT, MUSIC ...),
  - signaux complexes, **signaux non stationnaires** ?
  - coût algorithmique ? version récursive ?