

Statistiques fonctionnelles pour l'imagerie hyperspectrale

Laurent Delsol – Cécile Louchet

MAPMO

17e journée CASCIMODOT
6 décembre 2012

Introduction

Modélisation des spectres via les statistiques fonctionnelles

Application : segmentation d'images hyperspectrales

Introduction

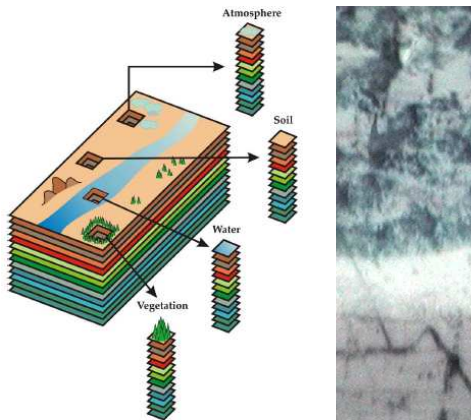
Modélisation des spectres via les statistiques fonctionnelles

Application : segmentation d'images hyperspectrales

Introduction à l'imagerie hyperspectrale

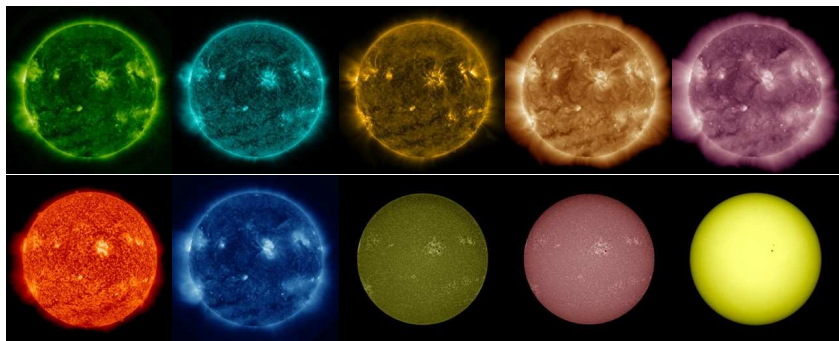
Image : fonction définie sur une grille de pixels.

Image hyperspectrale : en chaque pixel on a un spectre.



Bioucas-Dias et al., *Hyperspectral Unmixing Overview: Geometrical, Statistical, and Sparse Regression-Based Approaches*, IEEE Applied Earth Observations and remote sensing 2012

S. Cohen, E. Le Pennec, *Conditional Density Estimation by Penalized Likelihood Model Selection and Applications*, Tech. Report INRIA 2011



Mission NASA SDO (Solar Dynamics Observatory),
Atmospheric Imaging Assembly (AIA) (lumière visible, UV et
EUV), 10 spectres.

Autres jeux de données



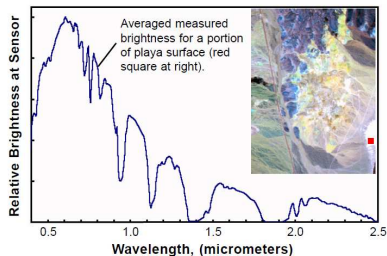
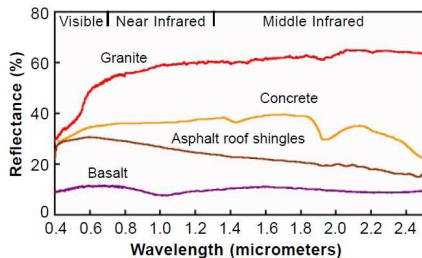
Lemon-slices



Watercolors

http://www1.cs.columbia.edu/CAVE/databases/multispectral/real_and_fake/

Réflectance / radiance



Images de réflectance

- ▶ c'est vraiment ce qui nous intéresse
- ▶ difficiles à acquérir

Images de radiance

- ▶ les conditions d'illumination compliquent l'interprétation
- ▶ mais faciles à acquérir.

Notations

Image hyperspectrale : $u \in \mathbb{R}^{\Omega \times \Lambda}$ où

- ▶ Ω est la grille de pixels;
- ▶ $\Lambda = (\lambda_i)_{i \in I}$ est un ensemble (fini) de fréquences (souvent $|\Lambda| = 31$ mais de plus en plus $|\Lambda| \sim 100$ ou 1000).
- ▶ $u(x) \in \mathbb{R}^\Lambda$ est un spectre.

On a besoin d'une notion de **distance** entre spectres :

$$d(u(x), u(y))$$

Distance entre spectres : pourquoi ?

- ▶ Définir une notion de qualité d'image reconstruite

$$PSNR(\hat{u}, u) = -10 \log \frac{\sum_x d^2(\hat{u}(x), u(x))}{\sum_x 255^2}.$$

- ▶ Définir la régularité d'une image hyperspectrale, par exemple

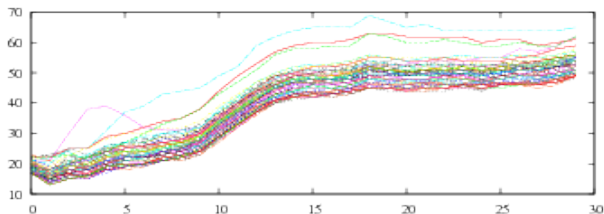
$$\|\nabla u\|_2^2 = \sum_{x \sim y} d^2(u(x), u(y))$$

ou

$$TV(u) = \sum_{x \sim y} d(u(x), u(y)).$$

- ▶ Étendre les méthodes classiques de TI aux images hyperspectrales (recalage, débruitage, segmentation, etc...).

Pseudo-distances (1/2)



► Distance euclidienne:

si $s, t \in \mathbb{R}^\Lambda$,
 $d(s, t) = \|s - t\|_2$.

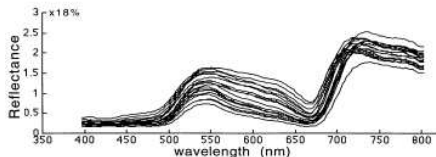
► Distance de Mahalanobis:

$d_\Sigma^2(s, t) = (s - t)^T \Sigma^{-1} (s - t)$, où
 $\Sigma \in \mathcal{S}^{++}(\mathbb{R}^\Lambda)$.

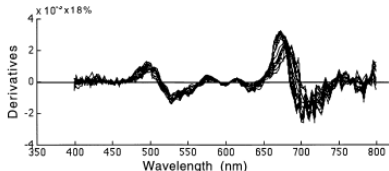
► Distance euclidienne sur les dérivées:

$d_{\text{der},q}(s, t) = \|s^{(q)} - t^{(q)}\|_2$ (en pratique $q = 1$ ou 2).

Pseudo-distances (2/2)



(a) Original spectra of manganese treated soybean plants



(b) Order-2 derivatives of (a) @ $\Delta\lambda=2.969\text{nm}$, filter size=7 points

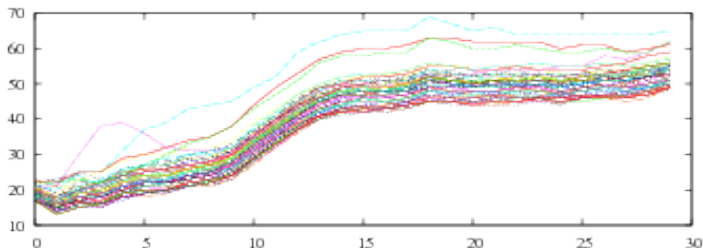
- ▶ **Distance euclidienne:**
invariant par permutation:
ne prend pas en compte la régularité des spectres.
Valable seulement en petite dimension.
- ▶ **Distance de Mahalanobis:**
 d_{Σ} est capable de prendre en compte la régularité des spectres, mais en pratique le calcul de Σ^{-1} est mal posé.
- ▶ **Distance sur les dérivées:** $d_{\text{der},q}$ prend en compte la régularité, simple, facilement généralisable en dimension infinie.

Introduction

Modélisation des spectres via les statistiques fonctionnelles

Application : segmentation d'images hyperspectrales

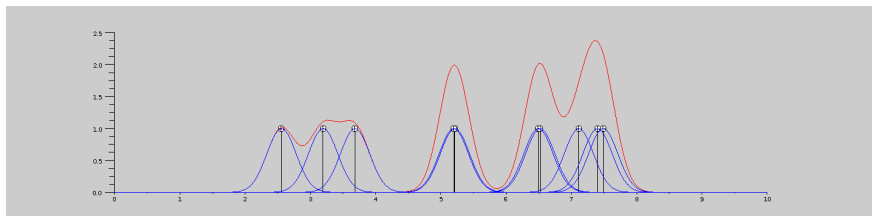
Modélisation des spectres



On veut une loi sur les spectres ...

- ▶ un loi qui respecte la régularité des courbes ;
- ▶ dans un premier temps on considère un spectre comme un vecteur de \mathbb{R}^Λ , avec $|\Lambda|$ grand...

Estimation de densité par noyau



On a $X_1, X_2, \dots, X_n \sim f$, et on veut estimer f .

En dimension 1

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)$$

En dimension d

$$\hat{f}(x) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{\|x - X_i\|}{h}\right)$$

- ▶ $K = \mathbb{I}_{[-\frac{1}{2}, \frac{1}{2}]}$ ou $K(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$ ou $K(t) = \frac{3}{4}(1 - t^2)_+ \dots$
- ▶ h petit \Rightarrow gros biais ; h grand \Rightarrow grosse variance.
 h doit dépendre de n et de la régularité de f .
- ▶ Choix de $\|\cdot\| \rightsquigarrow$ pseudo-distance d .

Estimation de densité pour $d \rightarrow \infty$

Le dénominateur h^d est relié au volume de $B^d(0, h)$, qui converge mal lorsque $d \rightarrow \infty \dots$ ($|B^d(0, h)|$ utilise la mesure de Lebesgue sur \mathbb{R}^d , non extensible à $d = \infty$).

\Rightarrow On remplace $|B^d(0, h)|$ par $\mathbb{E}K\left(\frac{d(X, x)}{h}\right)$.

Probabilité de petite boule

On note $\varphi_x(h) = \mathbb{P}(d(X, x) \leq h)$.

Proposition

Sous certaines hypothèses sur K et φ_x , il existe C et $C' > 0$ tels que

$$C\varphi_x(h) \leq \mathbb{E}K\left(\frac{d(X, x)}{h}\right) \leq C'\varphi_x(h)$$

Convergence de la densité estimée

On pose

$$\hat{f}(x) = \frac{1}{n Q(K, h)} \sum_{i=1}^n K\left(\frac{d(x, x_i)}{h}\right)$$

Théorème 1

Supposons

- ▶ f continue
- ▶ $h = (h_n)$ avec $h \rightarrow 0$
et $\frac{\log n}{n\varphi_x(h)} \rightarrow 0$
- ▶ autres hypothèses techniques.

Alors $\hat{f}(x) \rightarrow f(x)$ p.s.

Théorème 2

Si, de plus, f est Lipschitz d'ordre β , i.e.

$$\forall x, x', |f(x) - f(x')| \leq d(x, x')^\beta,$$

alors

$$\hat{f}(x) - f(x) = O(h^\beta) + O\left(\sqrt{\frac{\log n}{n\varphi_x(h)}}\right) \text{ p.s.}$$

Probabilités de petites boules

Processus fractals

Proposition : S'il existe $C > 0$ tel que

$$\varphi_x(h) \sim Ch^\tau \text{ lorsque } h \rightarrow 0,$$

alors, pour $h \sim \left(\frac{\log n}{n}\right)^{\frac{1}{2\beta+\tau}}$,

$$\hat{f}(x) - f(x) = O\left(\left(\frac{\log n}{n}\right)^{\frac{\beta}{2\beta+\tau}}\right)$$

presque sûrement.

Processus exponentiels

Proposition : S'il existe $C > 0$ tel que lorsque $h \rightarrow 0$

$$\varphi_x(h) \sim Ce^{-\frac{1}{h^{\tau_1}} \log(1/h)^{\tau_2}},$$

alors, pour $h \sim C(\log n)^{-u}$ (pour un $u > 0$), il existe $v > 0$ tel que

$$\hat{f}(x) - f(x) = O((\log n)^{-v})$$

presque sûrement.

Introduction

Modélisation des spectres via les statistiques fonctionnelles

Application : segmentation d'images hyperspectrales

Segmentation d'une image

D'une image v définie sur le domaine Ω , on veut trouver une partition de $\Omega = \sqcup_i \Omega_i$ qui reflète bien les régions de v .



Donnée v



Segmentation c en 80 classes

$c(x) =$ moyenne de v sur la région contenant x .

Segmentation = compromis entre 2 critères:

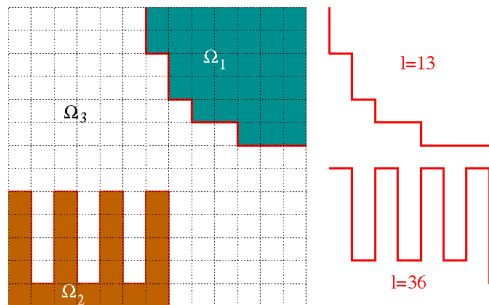
- ▶ v homogène sur chaque Ω_i ;
- ▶ Ω_i réguliers.

Modèle bayésien non-paramétrique

On se donne $v \in \mathbb{R}^{\Omega \times \Lambda}$, on veut trouver c .

Critère du maximum a posteriori (+ règle de Bayes):

$$\max_c P(C = c | V = v) \iff \max_c P(V = v | C = c) \cdot \underbrace{P(C = c)}$$



Loi a priori $P(C = c)$

On pénalise la longueur
des frontières des classes:

$$P(C = c) \propto e^{-\beta \sum_{x \sim y} \delta_{c(x) \neq c(y)}}.$$

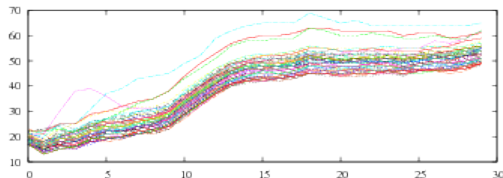
\iff Modèle de Potts

$\iff \min \mathcal{H}^1(\partial \Omega_i)$ dans
Mumford-Shah

Loi du bruit $P(V = v|C = c)$

- ▶ Si les pixels sont bruités indépendamment:

$$P(V = v|C = c) = \prod_{x \in \Omega} P(V(x) = v(x)|C(x) = c(x))$$



- ▶ Estimation non-paramétrique de densité:

$$P(v(x)|c(x)) = \frac{1}{Q(K, h)} \frac{\sum_{y|c(y)=c(x)} K\left(\frac{d(v(y), v(x))}{h}\right)}{\#\{y|c(y) = c(x)\}}$$

où par exemple

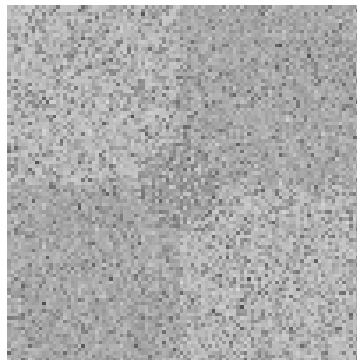
$$K(t) = \frac{1}{\sqrt{2\pi}} \exp(-t^2/2) \quad \text{et} \quad d = d_{\text{der}, q}$$

Maximisation de

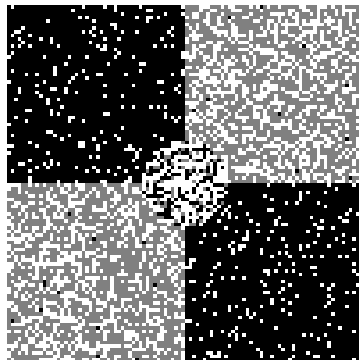
$$P(C = c | V = v) = \frac{1}{ZZ_h} e^{-\beta \sum_{x \sim y} \mathbf{1}_{c(x) \neq c(y)}} \\ \times \prod_{x \in \Omega} \frac{\sum_{y | c(y) = c(x)} K\left(\frac{d_{\text{der},q}(v(y), v(x))}{h}\right)}{\#\{y | c(y) = c(x)\}}$$

- ▶ problème non convexe
- ▶ *iterated conditional mode* : pour $x \sim \mathcal{U}_\Omega$, on choisit la classe qui maximise $P(c(x) | v(y), y \neq x)$, et on itère.
- ▶ initialisation:
sur-segmentation grossière ou *k-means* (longitudinal).
- ▶ choix du h ???

Résultats – données synthétiques



Données synthétiques

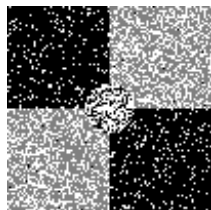


Initialisation

Initialisation par k -means (longitudinal) à 3 régions.

Algorithme : on prend $\beta = 0.5$, 8-connexité, distance euclidienne, et on attend la convergence complète.

Résultats – données synthétiques



iter. 0



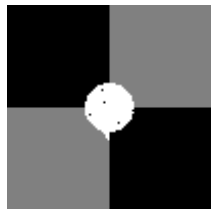
iter. 1



iter. 2



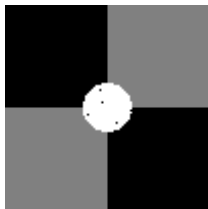
iter. 3



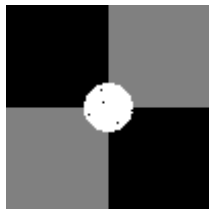
iter. 4



iter. 5

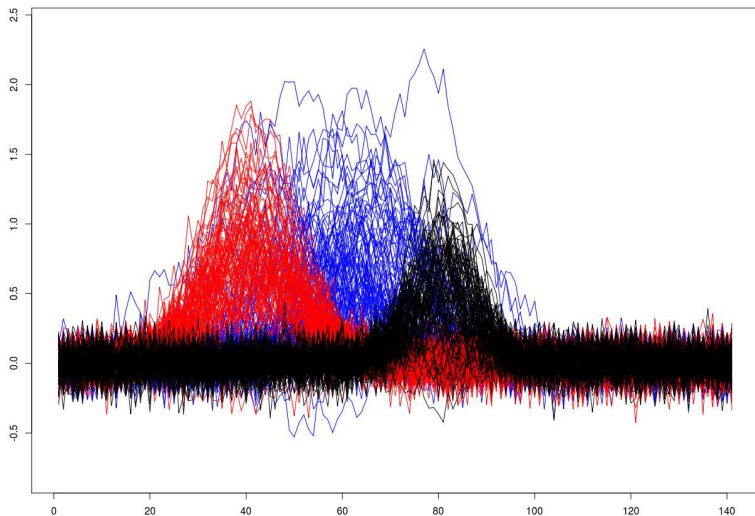


iter. 6



iter. 7

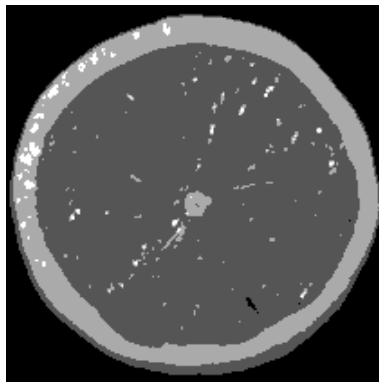
Résultats – données synthétiques



Résultats – Lemon-slices



données

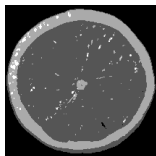


initialisation

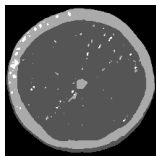
Initialisation par k -means (longitudinal) à 4 régions.

Algorithme : on prend $\beta = 0.6$, 8-connexité, distance euclidienne et on attend la convergence complète.

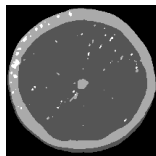
Résultats – Lemon-slices



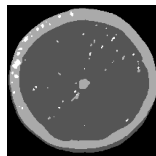
iter. 0



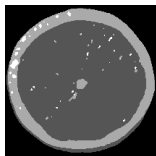
iter. 1



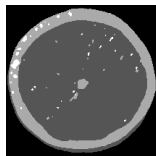
iter. 2



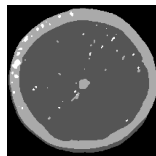
iter. 3



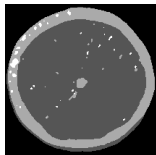
iter. 4



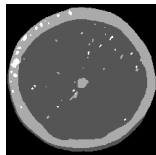
iter. 5



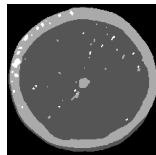
iter. 6



iter. 7

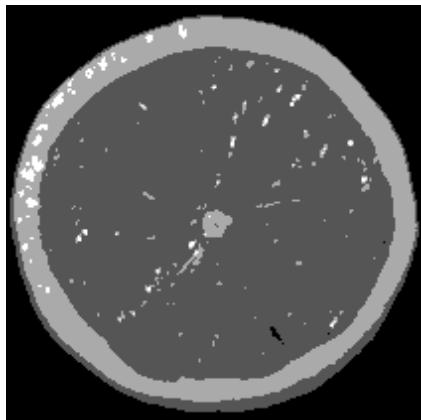


iter. 8

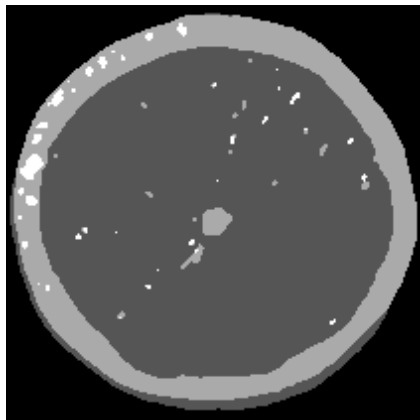


iter. 10

Résultats – Lemon-slices

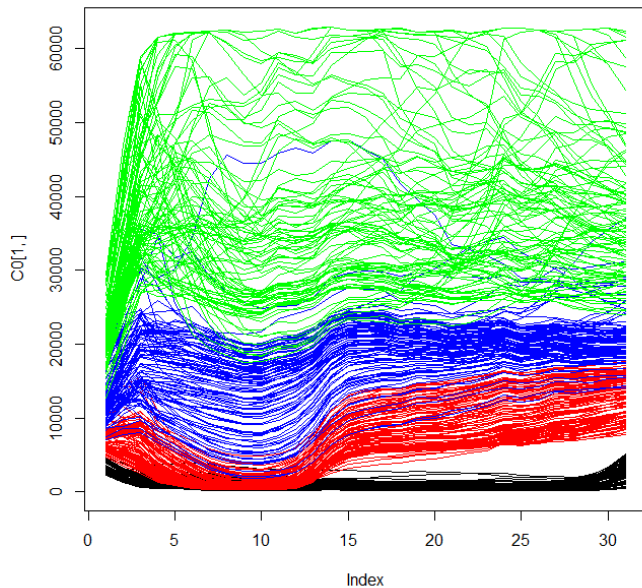


initialisation



instant final

Résultats – Lemon-slices



Conclusion

Conclusion

- ▶ Cadre fonctionnel pour les spectres;
- ▶ Méthodes adaptées et résultats théoriques pas si hors de portée;
- ▶ Généralisation possible à d'autres cadres que l'imagerie hyperspectrale.

Perspectives

- ▶ Modélisation plus fine des spectres (analyse de régularité, estimation de la probabilité de petite boule);
- ▶ Apprentissage d'une distance (sur spectres continus) à partir d'échantillons discrets;
- ▶ Stabilité de l'estimée non-paramétrique du bruit.
- ▶ Applications au débruitage, au démélange ...