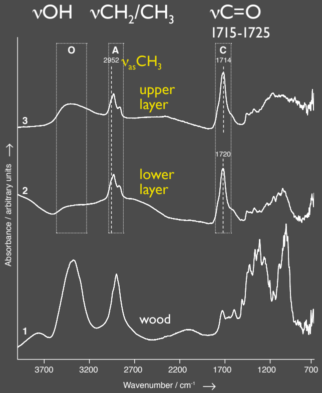
Segmentation d'images hyperspectrales par mélange de gaussiennes et sélection de modèles

E. Le Pennec
(SELECT - INRIA Saclay / Universités Paris Sud et Paris Diderot)
et
S. Cohen (IPANEMA - Soleil Saclay)
avec
G. Celeux et P. Massart (SELECT - INRIA Saclay / Université Paris Sud)
et C. Maugis (INSA Toulouse)

A. Stradivari (1644 - 1737)

Provigny (1716)

A. Giordan © Cité de la Musique



SOLEIL SYNCHROTRON

4 / 8 cm^{-1} resolution
64 / 128 scans
typ. 1 min/sp, 400sp

very simple process
no protein (amide I, amide II)
no gums, nor waxes
@SOLEIL: SMIS

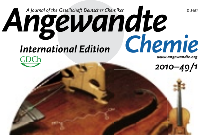
cité de la musique

SOLEIL SYNCHROTRON

CRCC

CENTRE DE RECHERCHE ET DE RESTAURATION DES MUSÉES DE FRANCE

Laboratoire de Dynamique, Interactions et Réactivité



J.-P. Echard, L. Bertrand, A. von Bohlen, A.-S. Le Hô, C. Paris, L. Bellot-Gurlet, B. Soulier, A. Lattuati-Derieux, S. Thao, L. Robinet, B. Lavédrine, and S. Vaiedelich. *Angew. Chem. Int. Ed.*, 49(1), 197-201, 2010.

Segmentation d'images hyperspectrales

- Données :
  - image de taille n comprise entre ~ 1000 et ~ 100000 pixels,
  - spectres $\mathcal{S}$ de ~ 1024 points,
  - résolution $\sim 4/8$ cm^{-1} (10 fois meilleure dans le visible),
  - possibilité de mesurer de très nombreux spectres par minute...
- Objectifs immédiats :
  - segmentation automatique de ces images,
  - sans intervention humaine,
  - aide à l'analyse des résultats.
- Objectifs lointains :
  - classification automatique,
  - interprétation...

Modélisation par un mélange de gaussiennes

- Modélisation stochastique des spectres $\mathcal{S}$:
  - existence de K classes de spectres,
  - proportion π_k pour chacune des classes ($\sum_{k=1}^{K} \pi_k = 1$),
  - loi gaussienne $\mathcal{N}(\mu_k, \Sigma_k)$ sur chacune des classes (hypothèse forte !)
- Densité :

$$\mathcal{S} \sim \sum_{k=1}^{K} \pi_k \mathcal{N}(\mu_k, \Sigma_k)(\mathcal{S}) d\mathcal{S}$$

- Objectif : estimer les paramètres K, π_k, μ_k, Σ_k à partir des données.
- Pourquoi ? : possibilité d'assigner ensuite une classe à une observation par maximum de vraisemblance

$$\widehat{k}(\mathcal{S}) = \operatorname{argmax} \pi_k \mathcal{N}(\mu_k, \Sigma_k)(\mathcal{S})$$

- Résultat en terme d'estimation de densité...

Modèle de mélange de gaussiennes

- Densités : $$\mathcal{S} \sim \sum_{k=1}^{K} \pi_k \mathcal{N}(\mu_k, \Sigma_k)(\mathcal{S}) d\mathcal{S}$$
- Modèle S_m :
  - choix d'un nombre de classe K,
  - choix d'une structure pour les moyennes μ_k et les covariances $\Sigma_k = L_k D_k A_k D_k'$
- Modèles $[\mu\, L\, D\, A]^K$: contraintes (valeurs connues, communes ou libres...) sur les moyennes μ_k, les volumes L_k, les bases de diagonalisation D_k et les valeur propres A_k.
- Modèle S_m : modèle paramétrique de dimension $(K-1) + \dim([\mu\, L\, D\, A]^K)$ dans un espace de dimension p.
- Estimation par maximum de vraisemblance des paramètres :
  - pour chaque classe, la moyenne μ_k et la covariance $\Sigma_k = L_k D_k A_k D_k'$
  - les proportions π_k du mélange.
- Technique classique avec algorithme (EM) efficace disponible.

Sélection de modèles

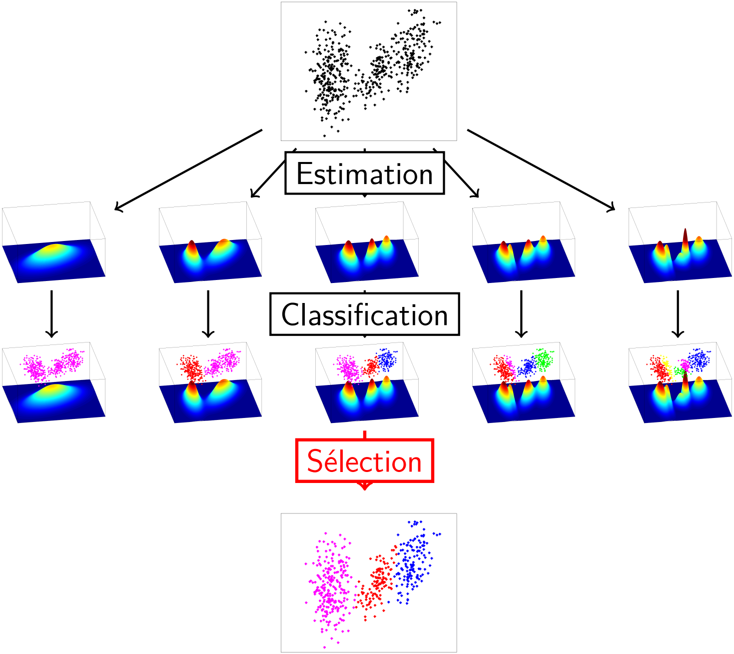
- Comment choisir le "modèle" S_m :
  - le nombre de classe K,
  - le modèle $[\mu\, L\, D\, A]^K$?
- Thème central du projet SELECT.
- Principe de sélection de modèles par pénalisation :
  - choix d'une collection de modèles $S_m = \{s_m\}$ avec $m \in \mathcal{M}$,
  - estimation par maximum de vraisemblance d'une densité $\hat{s}_m$ pour chaque modèle S_m,
  - sélection d'un modèle $\widehat{m}$ par

$$\widehat{m} = \mathrm{argmin} - \ln(\widehat{s}_m) + \mathrm{pen}(m).$$

  avec $\mathrm{pen}(m) = \kappa(\ln(n))\dim(S_m)$ (dimension intrinsèque de S_m),
- Résultats (Birgé, Massart, Celeux, Maugis, Michel...) :
  - pratique de classification non supervisée ($\neq$ segmentation),
  - théorique d'estimation du mélange : pour κ assez grand,

$$\mathbb{E}\left[d^2(s, \widehat{s}_{\widehat{m}})\right] \leq C \inf_{m \in \mathcal{M}} \left(\inf_{s_m \in S_m} KL(s, s_m) + \frac{\mathrm{pen}(m)}{n} \right) + \frac{C'}{n}.$$

Méthodologie



Segmentation et mélange de gaussiennes

- Objectif initial : segmentation $\neq$ classification non supervisée.
- Prise en compte de la position spatiale x du spectre à travers les proportions du mélange (Kolaczyk et al) :

$$S(x) \sim \sum_{k=1}^{K} \pi_k(x) \mathcal{N}(\mu_k, \Sigma_k)(S) dS.$$

- Modèle mélangeant paramétrique et “non-paramétrique”...
- Estimation à partir des données :
  - pour chaque classe, la moyenne μ_k et la covariance $\Sigma_k = L_k D_k A_k D'_k$,
  - de la fonction de mélange $\pi_k(x)$.
- $\pi_k(x)$ fonction : régularisation nécessaire.
- Principe de sélection de modèles...

Mélange de gaussiennes et partition hiérarchique

- Comment choisir le "modèle" S_m ? :
  - le nombre de classe K,
  - le modèle $[\mu\, L\, D\, A]^K$,
  - la structure des paramètres de mélange $\pi_k(x)$.
- Structure simple pour $\pi_k(x)$:
  - constant par morceau sur une partition "hiérarchique",
  - optimisation efficace possible,
  - performance d'approximation raisonnable.
- $\dim(S_m) = |\mathcal{P}|(K-1) + \dim([\mu\, L\, D\, A]^K)$.
- Pénalité $\mathrm{pen}(m) = \kappa \ln(n) \dim(S_m)$ suffisante pour
  - l'optimisation numérique (EM + programmation dynamique),
  - le contrôle théorique : pour κ assez grand,

$$\mathbb{E}\left[d^2(s, \widehat{s}_{\widehat{m}})\right] \leq C \inf_{m \in \mathcal{M}} \left(\inf_{s_m \in S_m} KL(s, s_m) + \frac{\mathrm{pen}(m)}{n} \right) + \frac{C'}{n}.$$

Theorem

- Assumption (H) : there is a non-increasing function $\tilde{\phi}_m(\delta, \beta_\phi)$ such that $\delta \mapsto \delta\phi_m(\delta)$ is non-decreasing on $(0, +\infty)$ and for every $\sigma \in \mathbb{R}^+$ and every $s_m \in S_m$

$$\frac{1}{\sigma}\int_0^\sigma \sqrt{H_{[\cdot],d^{\otimes n}}\left(\epsilon, S_m(s_m,\sigma)\right)}\, d\epsilon \leq \phi_m(\sigma).$$

- **Theorem (up to some technical conditions) :** Assume we observe (X_i, Y_i) with unknown law parametrized by s. Let $(S_m)_{m\in\mathcal{M}}$ a at most countable model collection.

  Assume that there is a family $(x_m)_{m\in\mathcal{M}}$ of non-negative number such that $\sum_{m\in\mathcal{M}} e^{-x_m} \leq \Sigma < +\infty$and, under assumption (H), let σ_m be the unique root of $\tilde{\phi}_m(\sigma) = \sqrt{n}\sigma$. and let $\widehat{s_m}$ be a ρ maximum likelihood minimizer in S_m

$$\sum_{i=1}^n -\ln(\widehat{s_m}(X_i, Y_i)) \leq \inf_{s_m\in S_m}\left(\sum_{i=1}^n -\ln(s_m(X_i, Y_i))\right) + \rho$$

  For any $C_1 > 1$, there are two absolute constants κ_0 and C_2 such as soon as for every model $m \in \mathcal{M}$

$$\mathrm{pen}(m) \geq \kappa\left(n\sigma_m^2 + x_m\right) \quad \text{with } \kappa > \kappa_0,$$

  the penalized likelihood estimate $\widehat{s_{\widehat{m}}}$ with $\widehat{m}$ defined by $\widehat{m} = \operatorname{argmin}_{m\in\mathcal{M}} \sum_{i=1}^n -\ln(\widehat{s_m}(X_i, Y_i)) + \mathrm{pen}(m)$ satisfies

$$\mathbb{E}\left[\mathfrak{d}^{2\otimes n}(s, \widehat{s_{\widehat{m}}})\right] \leq C_1 \inf_{S\in\mathcal{M}}\left(\inf_{s_m\in S_m} KL^{\otimes n}(s, s_m) + \frac{\mathrm{pen}(m)}{n}\right) + C_2\frac{\Sigma}{n} + \frac{\rho}{n}.$$

Kullback, Hellinger et extensions

- Inégalité oracle en sélection de modèles de la forme :

$$\mathbb{E}\left[d^2(s,\widehat{s}_{\widehat{m}})\right] \leq C\left(\inf_{m\in\mathcal{M}}\inf_{s_m\in S_m} KL(s,s_m) + \frac{\mathrm{pen}(m)}{n}\right) + \frac{C'}{n}.$$

- Densité : Hellinger $d^2(s,s')$ (ou affinité) (Kolaczyk, Barron, Bigot).
- Massart : raffinement avec $\mathfrak{d}^2(s,s') = 2KL(s,(s'+s)/2)$.
- Ici : observation de (X_i, S_i) où les X_i sont indépendants et S_i suit la loi $s(X_i,\cdot)$ (conditionnement au design...)
- Estimateur $\widehat{s}(x,\cdot)$
- Tensorisation de Kullback et $\mathfrak{d}^2(s,s')$

$$KL^{\otimes_n}(s,s') = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} KL\left(s(X_i,\cdot),s'(X_i,\cdot)\right)\right]$$

$$\mathfrak{d}^{2\otimes_n}(s,s') = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} \mathfrak{d}^2\left(s(X_i,\cdot),s'(X_i,\cdot)\right)\right]$$

- Distances raisonnables dans le cas fixed design et random design...

Inégalité oracle et distances

- Inégalité oracle de la forme

$$\mathbb{E}\left[\mathfrak{d}^{2\otimes_n}(s,\widehat{s}_{\widehat{m}})\right] \leq C \inf_{m\in\mathcal{M}}\left(\inf_{s_m\in S_m} KL^{\otimes_n}(s,s_m)+\frac{\mathrm{pen}(m)}{n}\right)+\frac{C'}{n}$$

sous une condition liant entropie à crochet des modèles et pénalité.

- On retrouve exactement le théorème classique si $s(X_i,\cdot)=s(\cdot)$.
- Bon scaling de $\mathfrak{d}^{2\otimes_n}(s,\widehat{s}_{\widehat{m}})$ et $KL^{\otimes_n}(s,s_m)$ avec n : ils restent du même ordre de grandeur.
- Problème dans Bigot et al avec Hellinger avec une uniforme pour X_i :

$$\frac{1}{n}d^2(s,\widehat{s}_{\widehat{m}}) \leq \frac{2}{n} \quad !$$

- Pas ce soucis avec Bhattacharyya-Renyi de Kolaczyk et Barron...

Pénalité et complexité

- Pénalité liée à la complexité du modèle et de la collection.
- Complexité du modèle S_m (entropie) :
  - $H_{[\cdot],d^{\otimes_n}}(\epsilon, S_m)$ entropie à crochet lié à la distance de Hellinger tensorisée $\left(d^{\otimes_n} = \sqrt{d^{2\otimes_n}} = \sqrt{\mathbb{E}\left[\frac{1}{n}\sum d^2(s(X_i,\cdot), s'(X_i,\cdot))\right]}\right)$.
  - Hypothèse (H) : pour tout modèle S_m, il existe une fonction décroissante $\tilde{\phi}_m(\delta)$ telle que $\delta \mapsto \delta\phi_m(\delta)$ soit croissante sur $(0, +\infty)$ et telle que pour tout $\sigma \in \mathbb{R}^+$ et tout $s_m \in S_m$

$$\frac{1}{\sigma}\int_0^\sigma \sqrt{H_{[\cdot],d^{\otimes_n}}(\epsilon, S_m(s_m, \sigma))}\, d\epsilon \leq \tilde{\phi}_m(\sigma),$$

  - Complexité mesurée par $\tilde{\phi}^2(\sigma_m)$ avec σ_m l'unique racine de $\tilde{\phi}_m(\sigma) = \sqrt{n}\sigma$
- Complexité de la collection (codage) :
  - complexité donnée par x_m satisfaisant Kraft $\sum_{m\in\mathcal{M}} e^{-x_m} \leq \Sigma < +\infty$
- Contrainte (classique) sur la pénalité

$$\mathrm{pen}(m) \geq \kappa\left(\tilde{\phi}^2(\sigma_m) + x_m\right) \quad \text{avec } \kappa > \kappa_0.$$

Retour vers les modèles de mélanges spatiaux

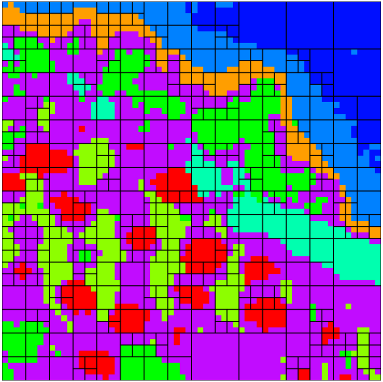
- Contrôle de $H_{[\cdot],d^{\otimes n}}(\epsilon, S_m(s_m,\sigma))$ pour les modèles de mélanges spatiaux (cf Maugis et Michel) :
  - contrôle d'un majorant de l'entropie : $H_{[\cdot],d^{\sup}}(\epsilon, S_m)$ où $d^{\sup} = \sqrt{d^{2\,\sup}} = \sqrt{\sup_x d^2(s(x,\cdot), s'(x,\cdot))}$,
  - résultat valide pour toutes les classes de mélanges ($[\mu\, L\, D\, A]^K$) et toutes les partitions :

$$H_{[\cdot],d^{\sup}}(\epsilon, S_m) \leq \dim(S_m)(C + \ln\frac{1}{\epsilon})$$

    avec C presque explicite (utilisation d'un lemme de Szarek sur l'entropie de $SO(n)$ sans constante explicite...).
  - implication : $\tilde{\phi}_m^2(\sigma_m) \leq \kappa' \ln(n) \dim(S_m)$.
- Codage de la collection avec $x_m \leq \kappa''|\mathcal{P}| \leq \frac{\kappa''}{K-1} \dim(S_m)$.
- Condition sur la pénalité :

$$\mathrm{pen}(m) \geq \left(\kappa' \ln(n) + \frac{\kappa''}{K-1}\right) \dim(S_m).$$

Le secret de Stradivarius



- Deux couches fines de vernis :
  - une première couche d'huile simple, similaire à celle des peintres, pénétrant légèrement le bois,
  - une seconde d'un mélange huile, résine de pin, pigments donnant cette couleur rouge caractéristique.
- Technique classique pour l'époque.
- Le secret de Stradivarius n'est pas dans le vernis !

Bibliographie

- Papier de départ présentant une solution complète :
  - A. Antoniadis, J. Bigot and R. von Sachs (2008) A multiscale approach for statistical characterization of functional images, Journal of Computational and Graphical Statistics, 18(1), pp. 216-237
  - Discussion paper qui lui correspond.
- Segmentation et sélection de modèles :
  - E.D. Kolaczyk, J. Ju and S. Gopal (2005). Multiscale, multigranular statistical image segmentation. Journal of the American Statistical Association, 100, 1358-1369.
  - E.D. Kolaczyk and R.D. Nowak (2004). Multiscale likelihood analysis and complexity penalized estimation. Annals of Statistics, 32, 500-527
- Sélection de modèles par MDL :
  - A.R. Barron, C. Huang, J. Q. Li and Xi Luo (2008). MDL Principle, Penalized Likelihood, and Statistical Risk. In Feschrift for Jorma Rissanen. Presented to Rissanen Nov. 2007.
- Sélection de modèles et entropie (à crochet) :
  - P. Massart (2003). Concentration inequalities and model selection : Ecole d'Eté de Probabilités de Saint-Flour XXXIII.
  - C. Maugis and B. Michel. (2009). A non asymptotic penalized criterion for Gaussian mixture model selection. Accepté à ESAIM : P&S (et l'erratum)
  - S. Cohen and E. Le Pennec (201 ?)
- Entropie de $SO(n)$:
  - J. Szarek (1998). Metric entropy of homogeneous spaces. Banach Center Pub., 43, 395–410