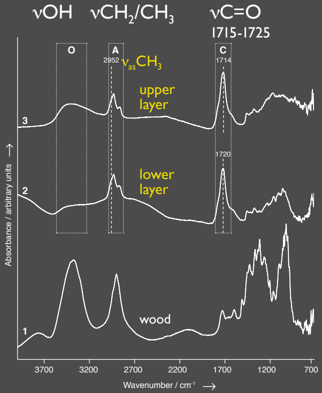
Hyperspectral image segmentation by Gaussian mixtures and model selection

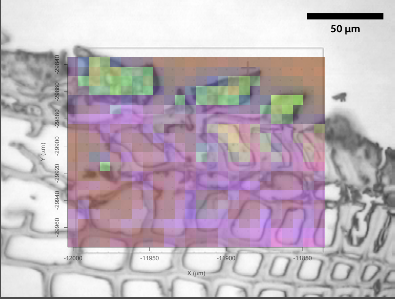
E. Le Pennec
(SELECT - INRIA Saclay / Université Paris Sud)
et
S. Cohen (IPANEMA - Soleil Saclay)
avec
G. Celeux et P. Massart (SELECT - INRIA Saclay / Université Paris Sud)
et C. Maugis (INSA Toulouse)

A. Stradivari (1644 - 1737)

Provigny (1716)

A. Giordan © Cité de la Musique



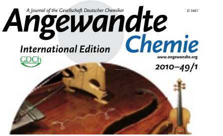


SOLEIL SYNCHROTRON

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.

Hyperspectral image segmentation

- Data :
  - image of size n between ~ 1000 and ~ 100000 pixels,
  - spectrum of $\mathcal{S}$ de ~ 1024 points,
  - resolution $\sim 4/8$ cm^{-1} (10 times better in the visible),
  - possibiliy to measure a lot of spectrums each minute...
- Immediate goals :
  - automatic segmentation,
  - without any human intervention,
  - provide help to analyse those results.
- Further goals :
  - automatic classification,
  - interpretation...

Gaussian mixture modeling

- Stochastic modeling of the spectrum $\mathcal{S}$:
  - existence of K classes of spectrum,
  - proportion π_k for each of these classes ($\sum_{k=1}^{K} \pi_k = 1$),
  - Gaussian law $\mathcal{N}(\mu_k, \Sigma_k)$ on each of these classes (strong assumption !)
- Density :

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

- Goal : estimate parameters K, π_k, μ_k, Σ_k from the data.
- Why ? : possibility to assign afterward a class to each observation by maximum likelihood :

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

- Theoretical results for density estimation...

Gaussian mixture model

- Densities : $$\mathcal{S} \sim \sum_{k=1}^{K} \pi_k \mathcal{N}(\mu_k, \Sigma_k)(\mathcal{S}) d\mathcal{S}$$
- Model S_m :
  - choice of a number of class K,
  - choice of astructure for the means μ_k and the covariances $\Sigma_k = L_k D_k A_k D'_k$
- Models $[\mu\, L\, D\, A]^K$: constraints (known values, common values or free) on the means μ_k, the volumes L_k, the diagonalization bases D_k and the eigenvalues A_k.
- Model S_m : parametric model of dimension $(K-1) + \dim([\mu\, L\, D\, A]^K)$ in a space of dimension p.
- Parameter estimation by maximum likelihood :
  - for each class, the mean μ_k and the covariance matrix $\Sigma_k = L_k D_k A_k D'_k$
  - the mixing proportions π_k.
- Classical technique with efficient algorithm (EM) available.

Model selection

- How to choose the "model" S_m :
  - the number of class K,
  - the model $[\mu\, L\, D\, A]^K$?
- Central theme of the SELECT project.
- Model selection by penalization :
  - choice of a model collection $S_m = \{s_m\}$ with $m \in \mathcal{M}$,
  - estimation by maximum likelihood of a density $\hat{s}_m$ for each model S_m,
  - selection of a model $\widehat{m}$ by

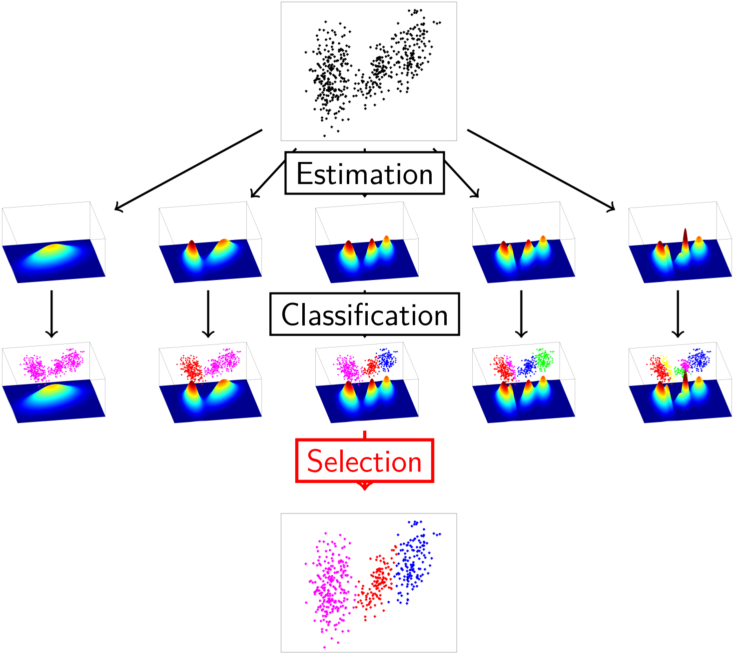
$$\widehat{m} = \mathrm{argmin} - \ln(\widehat{s}_m) + \mathrm{pen}(m).$$

  with $\mathrm{pen}(m) = \kappa(\ln(n))\,\dim(S_m)$ (intrinsic dimension of S_m),
- Results (Birgé, Massart, Celeux, Maugis, Michel...) :
  - theoretical (for mixture estimation) : for κ large enough,

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

  - practical : unsupervised classification ($\neq$ segmentation),

Methodology



Segmentation and Gaussian mixture

- Initial goal : segmentation $\neq$ unsupervised classification.
- Take into account the spatial position x of the spectrum trough the mixing proportion (Kolaczyk et al.) :

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

- Model mixing parametric and “non-parametric”
- Estimation from the data :
  - for each class, the mean μ_k and the covariance $\Sigma_k = L_k D_k A_k D_k'$,
  - of the mixing function $\pi_k(x)$.
- $\pi_k(x)$ function : regularization required.
- Model selection principle...

Gaussian mixture and hierarchical partition

- How to choose the “model” S_m ? :
  - the number of class K,
  - the model $[\mu\, L\, D\, A]^K$,
  - the structure of mixing function $\pi_k(x)$.
- Simple structure for $\pi_k(x)$:
  - piece-wise constant on a “hierarchical” partition,
  - efficient optimization possible,
  - good approximation performance.
- $\dim(S_m) = |\mathcal{P}|(K-1) + \dim([\mu\, L\, D\, A]^K)$.
- Penalty $\mathrm{pen}(m) = \kappa \ln(n) \dim(S_m)$ suitable for
  - the numerical optimization (EM + dynamic programming),
  - the theoretical control : for κ large enough,

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

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 and extensions

- Oracle inequality in model selection of type :

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

- Density : Hellinger $d^2(s,s')$ (or affinity) (Kolaczyk, Barron, Bigot).
- Massart : refinement with $\mathfrak{d}^2(s,s') = 2KL(s,(s'+s)/2)$.
- Here : observation of (X_i,S_i) with independent X_i and S_i of law $s(X_i,\cdot)$ (conditionning to the position...)
- Estimator $\widehat{s}(x,\cdot)$
- Tensorization of Kullback and $\mathfrak{d}^2(s,s')$

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

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

- Suitable distances for both fixed design and random design...

Oracle inequality and distances

- Oracle inequality of type

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

under a condition linking bracketing entropy of the models and penalty.
- Reduce to the classical theorem if $s(X_i,\cdot) = s(\cdot)$.
- Good scaling of $\mathfrak{d}^{2\otimes_n}(s,\widehat{s}_{\widehat{m}})$ et $KL^{\otimes_n}(s,s_m)$ with n : stay of the same order of magnitude.
- Issue in Bigot et al with Hellinger used with a uniform law for X_i :

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

- No issue with Bhattacharyya-Renyi of Kolaczyk and Barron...

Penality and complexity

- Penalty linked to the complexity of the model and of the collection.
- Complexity of the model S_m (entropy) :
  - $H_{[\cdot],d^{\otimes_n}}(\epsilon, S_m)$ bracketing entropy with the tensorized Hellinger distance $\left(d^{\otimes_n} = \sqrt{d^{2\otimes_n}} = \sqrt{\mathbb{E}\left[\frac{1}{n}\sum d^2(s(X_i,\cdot), s'(X_i,\cdot))\right]}\right)$.
  - Assumption (H) : for any model S_m, there is a non-increasing function $\tilde{\phi}_m(\delta)$ such that $\delta \mapsto \delta\phi_m(\delta)$ is non-decreasing on $(0, +\infty)$ and such that for any $\sigma \in \mathbb{R}^+$ and any $s_m \in S_m$

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

  - Complexity measured by $\tilde{\phi}^2(\sigma_m)$ with σ_m the unique root of $\tilde{\phi}_m(\sigma) = \sqrt{n}\sigma$
- Complexity of the collection (coding) :
  - complexity given by x_m satisfying Kraft $\sum_{m\in\mathcal{M}} e^{-x_m} \leq \Sigma < +\infty$
- (Classical) Constraint on the penalty :

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

Back to spatial mixture

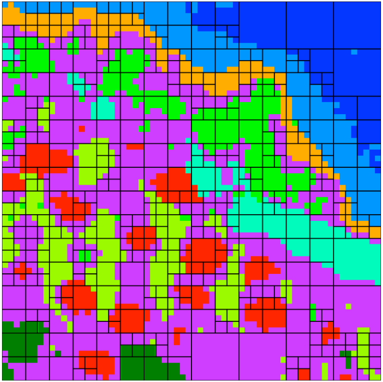
- Bound on $H_{[\cdot],d^{\otimes n}}(\epsilon, S_m(s_m,\sigma))$ for the spatial mixture models (cf Maugis et Michel) :
  - bound on a majoration of the entropy : $H_{[\cdot],d^{\sup}}(\epsilon, S_m)$ où $d^{\sup} = \sqrt{d^{2\,\sup}} = \sqrt{\sup_x d^2(s(x,\cdot), s'(x,\cdot))}$,
  - results for every mixture models ($[\mu\, L\, D\, A]^K$) and every parttions :

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

    with C almost explicit (rely on a lemma of Szarek on the entropy $SO(n)$ without an explicit constant...)
  - Implies : $\tilde{\phi}_m^2(\sigma_m) \leq \kappa' \ln(n) \dim(S_m)$.
- Collection coded with $x_m \leq \kappa''|\mathcal{P}| \leq \frac{\kappa''}{K-1}\dim(S_m)$.
- Constraint on the penalty :

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

Stradivarius secret



- Two fine varnish layers :
  - a first layer of simple oil, similar to the one used by painters, going slightly into the wood, légèrement le bois,
  - a second one with a mixture of oil, pine resin and pigments giving the characteristic red color.
- Classical technique for this period.
- Stradivarius secret is not in the varnish !

Bibliography

- My starting point describing a full solution :
  - 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
  - Corresponding discussion paper.
- Segmentation and model selection :
  - 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
- MDL based model selection :
  - 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.
- Model selection and (bracketing) entropy :
  - 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. ESAIM : P&S (plus erratum...)
  - S. Cohen and E. Le Pennec (201 ?)
- $SO(n)$ entropy :
  - J. Szarek (1998). Metric entropy of homogeneous spaces. Banach Center Pub., 43, 395–410