

Séparation de mélanges de données spectrales Raman sous contraintes de positivité et d'additivité

Sujet de stage 2011

Hérald Rabeson, Laurent Duval





IFP Energies nouvelles

**Établissement public de recherche,
d'innovation industrielle et de formation**

**Mission : développer des technologies performantes,
économiques, propres et durables,
pour relever les trois grands défis sociétaux du 21^e siècle :
changement climatique et impacts environnementaux,
diversification énergétique et gestion des ressources en eau**



**IFP Energies nouvelles apporte des solutions industrielles innovantes
dans ses domaines d'activité : énergie, transport, environnement**

**Centre de recherche appliquée, il assure le transfert entre
recherche fondamentale et développement industriel**



IFP Energies nouvelles

En bref

- 1 737 personnes*, dont 1 173 chercheurs (ingénieurs et techniciens), basés à Rueil-Malmaison et à Lyon
- 157 thésards et 27 postdoctorants
- Plus de 50 métiers représentés : du géologue au motoriste
- Un environnement technique (moyens d'essais, équipements) de très haut niveau

- Statut : établissement public à caractère industriel et commercial (EPIC)
- Financement : budget de l'État et ressources propres provenant de partenaires privés français et étrangers
- Budget 2009 : 307,8 M€ dont 249 M€ pour la R&D

En 2009 :

- 13 900 brevets vivants
- 305 articles publiés dans les revues scientifiques internationales

* effectif moyen équivalent temps plein



IFP Energies nouvelles

Positionnement stratégique

Le paysage énergétique

Croissance de la
demande et
prix du pétrole

Caractère par
nature limité des
énergies fossiles

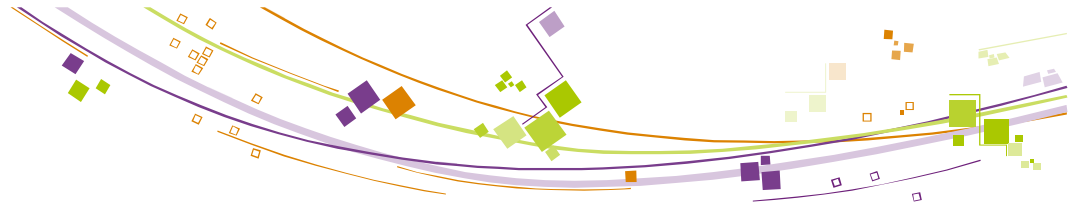
Changement
climatique

Difficile substitution
massive et rapide
des hydrocarbures
pour les transports et
la pétrochimie

Tension sur les
ressources
humaines

Préparer la transition énergétique

Concevoir les solutions permettant d'optimiser l'utilisation des énergies fossiles tout en développant de nouvelles technologies et sources d'énergies pour répondre aux besoins sociétaux dans les domaines de l'énergie, du transport et de l'environnement



Quelques activités TSI

- Diagnostic multi-varié de paramètres de combustion et détection d'événements pour les moteurs à combustion interne
 - Réduction de dimension
- Soustraction adaptative de multiples en sismique marine par une approche multi-échelle
 - Adaptation et décomposition en ondelettes 1D puis 2D directionnelles
- Caractérisation quantitative de supports de catalyseurs pour une meilleure compréhension entre morphologie du support et activité catalytique
 - Analyse d'images obtenues par tomographie électronique
- Analyse automatisée de signaux obtenus par différents types d'analyses (chromatographie, RMN, ..) pour la mise au point de nouveaux catalyseurs
 - Filtrage(bruit, ligne de base), quantification



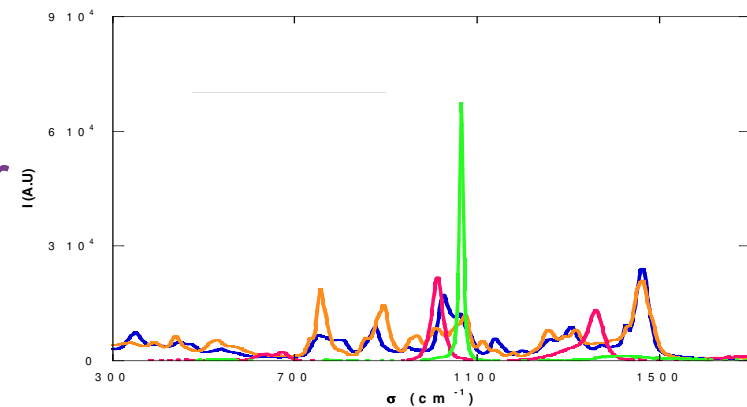
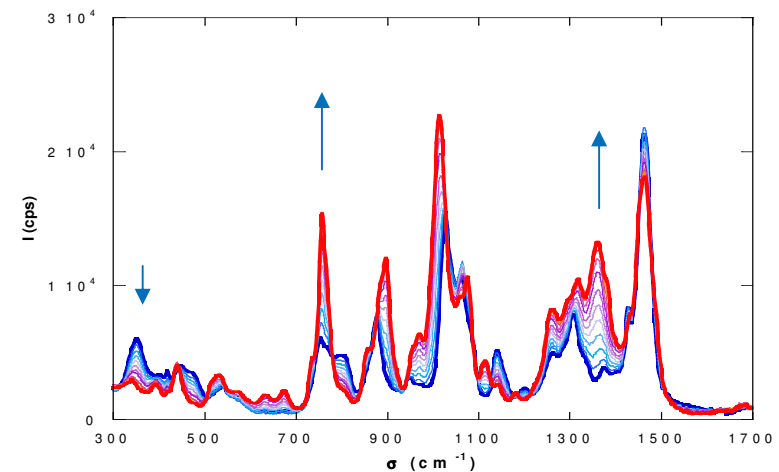
Analyse quantitative de mélanges de spectres Raman

■ Données:

- Séries de spectres, mélanges linéaires de composés dont les concentrations varient en fonction d'un paramètre donné (pH, temps, ...)

■ Objectif:

- Séparer les différentes contributions et estimer/quantifier les concentrations de chaque composé en fonction de l'évolution du paramètre



Exemples de spectres Raman: mélanges (en haut) et composés (en bas)
6 janvier 2011



Formulation de l'analyse quantitative de mélanges de spectres Raman

■ Notations:

- F le nombre de spectres d'une série, N le nombre de points par spectre, K le nombre de composés chimiques,
- X: matrice des spectres expérimentaux (F x N) ,
- A: matrice des profils de concentrations (F x K),
- S: matrice des spectres des composés chimiques 'purs' (K x N)

■ Objectif:

$$X = AS + E$$

où E est la contribution du bruit.

Estimer A (dans certains cas K et S) sachant que les éléments de A et S sont non-négatifs



Quelques Approches possibles

- Problème de factorisation et d'approximation

$$X \approx AS = \hat{X}$$

- Nombreux algorithmes envisageables
 - Moindres Carrés Alternés (type MCR-ALS)
 - Gradient conjugué ou projeté
 - Méthodes de Quasi-Newton
 - Analyse en Composantes Principales ou Indépendantes
- Prise en compte souvent non-naturelle de la positivité des éléments de A et S
- Factorisation en Matrices Non-négatives (NMF)
 - Algorithmes multiplicatifs



Minimisation d'une fonction coût appropriée

- Approche basée sur l'optimisation de la similarité (ou dissimilarité)

$$D(X|\hat{X}) = \sum_{f=1}^F \sum_{n=1}^N d(x_{fn}|\hat{x}_{fn})$$

- Distances et divergences

- distance euclidienne

- divergence de Kullback-Leibler $d_{KL}(x|y) = x \log\left(\frac{x}{y}\right) - x + y$

- divergence d'Itakura-Saito $d_{IS}(x|y) = \frac{x}{y} - \log\left(\frac{x}{y}\right) - 1$

- divergences généralisées (Csiszar, Kompass, β -divergence,...)

$$d_{\beta} = \frac{1}{\beta(\beta - 1)} (x^{\beta} + (\beta - 1)y^{\beta} - \beta xy^{\beta-1}) \quad \beta \notin \{0, 1\}$$

- Choix de la fonction coût lié aux données...



Exemple d'algorithmes standard de NMF

■ Lee & Seung, 1999

■ Règles de mise à jour multiplicatives

- Distance euclidienne

$$\mathbf{S} \leftarrow \mathbf{S} \otimes \frac{\mathbf{A}^T \mathbf{X}}{\mathbf{A}^T (\mathbf{A} \mathbf{S})} \quad \mathbf{A} \leftarrow \mathbf{A} \otimes \frac{\mathbf{X} \mathbf{S}^T}{(\mathbf{A} \mathbf{S}) \mathbf{S}^T}$$

- Divergence KL

$$\mathbf{S} \leftarrow \mathbf{S} \otimes \frac{\mathbf{A}^T (\mathbf{X} \oslash (\mathbf{A} \mathbf{S}))}{\mathbf{A}^T \mathbf{1}} \quad \mathbf{A} \leftarrow \mathbf{A} \otimes \frac{(\mathbf{X} \oslash (\mathbf{A} \mathbf{S})) \mathbf{S}^T}{\mathbf{1} \mathbf{S}^T}$$

⊗ désigne le produit terme à terme

— ⊘ désignent la division terme à terme



NMF avec contraintes (1/2)

■ Nombreuses Variantes Contraintes de la NMF

■ Formulation générale

$$\min_{\mathbf{A}, \mathbf{S}} D(\mathbf{X} | \mathbf{A}\mathbf{S}) + \lambda D_p(\mathbf{A}, \mathbf{S})$$

- D_p terme lié à une information *a priori*
- λ paramètre de régularisation, compromis entre l'importance de l'attache aux données et de l'*a priori* fourni

■ NMF Parcimonieuse

- Nombreux coefficients/lignes nuls dans \mathbf{S} . [Hoyer, 2004]
- Niveau de parcimonie contrôlé par un rapport de normes L^1 et L^2

■ NMF Orthogonale, NMF projective

$$\mathbf{A}^T \mathbf{A} = \mathbf{I} \quad \text{ou} \quad \mathbf{S}\mathbf{S}^T = \mathbf{I} \quad \mathbf{A}^T \mathbf{A} \approx \mathbf{I}$$



NMF avec contraintes (2/2)

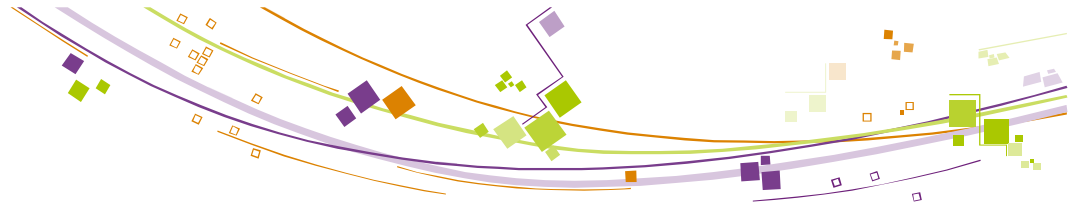
- Autres types d'information *a priori*

- Additivité

$$\sum_{(f,n) \in F_p \times N_p} a_{f,n} = c_0$$

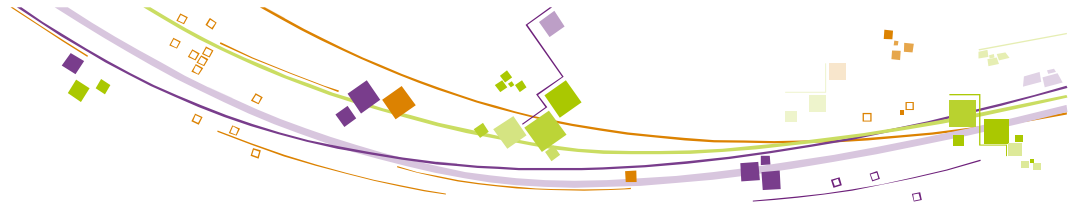
- Approche Statistique Bayésienne également possible

- Equivalence entre minimisation de la fonction cout et l'estimation de paramètres dans le cadre d'un modèle probabiliste des observations
- Lois de probabilités associés aux données, au bruit ainsi qu'à l'information *a priori* [N. Dobigeon, 2008]
 - Loi Gamma pour la non-négativité de S
 - Loi Inverse Gamma pour la variance du bruit
 - Loi de Dirichlet pour les éléments de A (non-négativité et additivité)



Bibliographie

- **R. Tauler, 2005**, Multivariate Curve Resolution: theory and applications
- **E. Kristien and W. Marquardt, 2008**, Fully automated indirect hard modeling of mixture spectra
- **N. Dobigeon, 2008**, Bayesian separation of spectral sources under non-negativity and full additivity constraints
- **P. Hoyer, 2004**, Non-negative Matrix Factorization with Sparseness Constraints
- **C. Gobinet, 2006**, Application de techniques de séparation de sources à la spectroscopie Raman et à la spectroscopie de fluorescence



Déroulement du stage

- Compléments d'information, données :

<http://www.laurent-duval.eu/lcd-2011-internship-bss.html>

- Conditions pratiques

- Stage de Master, début février/mars pour 5-6 mois, à Rueil-Malmaison
- Profil de candidat: spécialité Traitement du Signal (séparation de sources, régularisation, optimisation), pour la conception, le développement et la validation des algorithmes sur des données simulées puis application aux cas réels.
- Merci d'adresser votre candidature (CV et lettre de motivation) à Laurent Duval ou Hérald Rabeson

laurent.duval@ifpenergiesnouvelles.fr

01 47 52 61 02

herald.rabeson@ifpenergiesnouvelles.fr

01 47 52 62 84



Prolongation possible en thèse

- Sujet de thèse oct. 2011-2014

Représentations parcimonieuses pour la restauration et l'analyse quantitative de champs d'ondes en sismiques

- Encadrant : Laurent Duval, école doctorale : Université Paris-Est

Thèses précédentes :

- J. Gauthier, université de Paris-Est Marne-la-Vallée (dir. Jean-Christophe Pesquet), soutenance juin 2008

- Analyse de signaux et d'images par bancs de filtres. Applications aux géosciences

- Caroline Chaux, université de Paris-Est Marne-la-Vallée (dir. Jean-Christophe Pesquet), soutenance déc. 2006

- Analyse en ondelettes M-bandes en arbre dual ; application à la restauration d'images (prix 2008 de la meilleure thèse en signal et image du club EEA)