

# Stage Apprentissage non supervisé par approches spectrales (APO / Concace)

## Encadants

- E. Agullo, O. Coulaud (concace@Inria, Bordeaux)
- S. Mouysset, R. Guivarch, A. Buttari (APO, IRIT, Toulouse)

## Contexte

L'apprentissage non supervisé vise à créer une partition (un ensemble de classes) à partir des mesures de similarité (ou matrice de Gram) d'un ensemble de données. Les approches spectrales permettent d'une part, de réduire la dimension des données et de visualiser les données (*positionnement multidimensionnel (MDS)*) et d'autre part, de détecter automatiquement, de par la structure constante par morceaux des plus grands vecteurs propres extraits de la matrice de Gram, des structures blocs sur la matrice (*classification spectrale*).

Dans ce stage, on s'intéresse aux deux types d'approches spectrales suivantes :

- **positionnement multidimensionnel (MDS)** : de récents travaux ont montré qu'on peut construire une MDS réduite qui opère sur un sous-ensemble de données en utilisant un échantillon sans supposer que nous disposons d'un échantillon original et sans connaître le modèle. L'étude de ces vecteurs propres permettrait d'extraire des informations intéressantes à analyser de manière non supervisée.
- **classification spectrale** : une décomposition en sous-domaines a été définie pour découper la matrice de Gram et étudier la structure des vecteurs propres. Cependant le calcul des vecteurs propres reste coûteux et des améliorations peuvent être apportées en utilisant les techniques développées pour le MDS.

Les domaines d'application envisagés porteraient sur la biodiversité pour le MDS et la segmentation d'images pour la classification spectrale.

## Travail demandé :

- MDS :
  - analyse des vecteurs propres de grande dimension en utilisant des méthodes d'apprentissage non supervisé (kmeans, kppv...).
- Classification spectrale :
  - portage en C++ du code fortran
  - Utilisation de la bibliothèque [Fast Methods for Randomized numerical linear algebra \(FMR\)](https://gitlab.inria.fr/compose/legacystack/fmr) (<https://gitlab.inria.fr/compose/legacystack/fmr>)

## Profil du candidat :

L'étudiant devra avoir des connaissances en parallélisation et en calcul scientifique et quelques notions en apprentissage.

**Contact** : Emmanuel.Agullo@inria.fr, Olivier.Coulaud@inria.fr, Sandrine.Mouysset@irit.fr, Ronan.Guivarch@irit.fr, alfredo.buttari@enseeiht.fr