



HABILITATION A DIRIGER DES RECHERCHES

Date de la soutenance : **30 juin 2026**

Nom de famille et prénom de l'auteur : **Monsieur COHEN Jérémy**

Titre des travaux : « *Approximations de rang faible régularisées* »

Résumé



Contributions aux approximations de rang faible régularisées

Depuis le début de ma thèse en 2013, mes recherches se concentrent sur les modèles d'approximation de rang faible (LRA) comme outils permettant d'extraire des informations pertinentes à partir de matrices et de tenseurs, en utilisant des régularisations pour améliorer leur interprétabilité. Ces régularisations incluent souvent la non-négativité, qui est devenue ma spécialité. Mes travaux couvrent une diversité de sujets, allant de l'étude des propriétés des solutions des modèles rLRA à la proposition de nouvelles applications pour ces modèles. Une part significative de mes recherches porte sur la conception d'algorithmes permettant de calculer efficacement les solutions de rLRA dans divers contextes.

Défis dans les modèles rLRA

Malgré le nombre important de travaux existants sur les modèles rLRA, leur utilisation par les praticiens reste souvent complexe. En effet, plusieurs questions théoriques et pratiques restent à résoudre :

- **Problèmes d'optimisation** : Les modèles rLRA génèrent des problèmes d'optimisation multi-blocs, non convexes et souvent non lisses. Les avancées récentes en optimisation numérique, comme l'accélération, les méthodes de majoration-minimisation, les opérateurs de projection généralisés et l'optimisation alternée, peuvent être exploitées pour améliorer l'état de l'art. Il reste cependant peu clair, en pratique, quel algorithme d'optimisation choisir selon le contexte.
- **Informations complémentaires** : Des informations complémentaires sont souvent disponibles en plus des données matricielles ou tensorielles, transformant le problème d'apprentissage non supervisé rLRA en un problème semi-supervisé ou supervisé. Ces informations peuvent prendre diverses formes : une bibliothèque connue de modèles pour les sources inconnues, une tâche en aval avec des données d'entraînement disponibles, des a priori profonds pour les sources inconnues (comme des algorithmes de débruitage), la parcimonie dans des bases connues, ou encore le niveau de parcimonie des sources. Intégrer ces informations dans les modèles rLRA n'est pas toujours simple, tant sur le plan de la modélisation que de l'entraînement.
- **Multimodalité** : La multimodalité a émergé comme un sujet important dans la séparation de sources, où un même phénomène est observé via différents capteurs. Un exemple typique serait l'acquisition de l'activité cérébrale par EEG et IRMf simultanément. Le traitement conjoint de ces données peut être

réalisé par plusieurs modèles rLRA. Les questions de recherche portent généralement sur le modèle de couplage pour les différentes modalités et sur la conception d'algorithmes flexibles capables de résoudre une large gamme de problèmes de séparation de sources multimodaux, en particulier lorsque les jeux de données présentent des dynamiques, des niveaux de bruit ou des types de données différents (par exemple, flottants, entiers, catégoriques).

- **Propriétés des modèles rLRA** : Les modèles rLRA présentent encore des propriétés difficiles à cerner. Un exemple important est la décomposition de Tucker non négative, pour laquelle l'**identifiabilité** reste floue dans de nombreux cas, malgré les avancées récentes sur le sujet. En général, un utilisateur a besoin de garanties sur la nature des solutions des modèles rLRA, et ces garanties théoriques sont, en pratique, très difficiles à obtenir, d'autant plus si les hypothèses doivent être vérifiables sur le terrain.
- **Écosystème logiciel scientifique** : La dimension logicielle, souvent négligée mais cruciale en pratique, est un enjeu majeur. Les logiciels disponibles pour la décomposition de tenseurs sont principalement obsolètes : une enquête recense plus de 70 packages dispersés sur Internet. Cependant, les logiciels dédiés aux modèles rLRA ou capables de les gérer partiellement sont rares. Pour les décompositions de tenseurs, de nombreux packages se concentrent sur la décomposition elle-même sans exploiter les contractions à grande échelle sur CPU ou GPU, essentielles pour le passage à l'échelle. Un autre défi majeur réside dans l'implémentation des algorithmes rLRA, qui nécessite des optimisations non triviales (comme la mise en cache d'opérations coûteuses ou les compromis mémoire-vitesse), et qui constituent en eux-mêmes des sujets de recherche avancée.

Contributions en théorie, algorithmes et applications

Mon travail a consisté à proposer des solutions (partielles) à ces défis. J'ai structuré mes contributions en trois axes : contributions théoriques, applications et algorithmes. Ces trois aspects sont souvent interdépendants et ne forment pas une segmentation stricte de mes recherches.

Ce manuscrit ne comporte pas de chapitre dédié au développement logiciel. Les implémentations simples des méthodes étudiées sont fournies en ligne, importées depuis **Tensorly** (un package que je co-développe depuis 2019) ou depuis un ensemble local de méthodes spécialement développées pour ce manuscrit. Le développement logiciel est une composante essentielle des contributions scientifiques dans le domaine des modèles rLRA et des mathématiques appliquées. Il me semble important de montrer le code exact utilisé pour les expériences et les démonstrations, afin que le lecteur puisse mieux évaluer la fiabilité des résultats.

Les trois grandes sections de ce manuscrit sont les suivantes :

Théorie des contributions sur les modèles rLRA

J'y résume mes contributions à l'analyse de plusieurs modèles parcimonieux :

- **LRA basée sur un dictionnaire** : identifiabilité des modèles.
- **Moindres carrés non négatifs parcimonieux.**
- **Impact de l'ambiguïté d'échelle sur les solutions rLRA**, qui peut induire une parcimonie inattendue au niveau des composantes.

Contributions algorithmiques pour les modèles rLRA existants

J'y présente mes contributions à la conception d'algorithmes pour les modèles rLRA :

- **Algorithmes flexibles pour rLRA multimodale.**
- **Accélération inertielle heuristique pour les algorithmes d'optimisation alternée.**
- **Conception de surrogats serrés** pour les problèmes d'optimisation non négative, incluant la NMF dans des contextes non euclidiens.
- **Régularisation pilotée par les données pour la LRA**, notamment via le **déroulement des mises à jour multiplicatives** pour la NMF.
- **Revisite d'un travail de thèse** sur les moindres carrés projetés pour calculer des approximations de rang faible non négatives.

Applications des modèles rLRA

Mes recherches appliquées en modèles rLRA portent principalement sur :

- **L'imagerie spectrale**, où la LRA est étroitement liée aux mélanges additifs qui décrivent les techniques d'acquisition spectrale (comme la spectroscopie de fluorescence ou la télédétection hyperspectrale).
- **La transcription musicale automatique**, domaine dans lequel je m'investis personnellement en tant que pianiste et chanteur semi-professionnel.
D'autres applications, principalement liées à l'imagerie biomédicale, sont issues de collaborations ponctuelles.