

Demande de financement de thèse interdisciplinaire CNRS

Sujet de thèse :

Optimisation exacte, parcimonie et contraintes avancées pour l'analyse multi-solutions de données hyperspectrales de Mars

Contexte de la demande :

Durant les dernières décennies, la spectro-imagerie infrarouge a mené à de nombreuses découvertes sur la composition de la surface de Mars. En particulier, contraindre sa composition minéralogique par des modèles appropriés permet d'améliorer notre compréhension de l'histoire géologique et climatique de Mars. Cependant, la complexité minéralogique de certaines zones reste énigmatique : par exemple, l'origine des Interior Layered Deposits (dépôts stratifiés présents dans Valles Marineris), ainsi que d'autres régions ignées martiennes, sont toujours controversées. Pour progresser, il convient de développer une nouvelle génération d'outils multi-solutions qui donne une vision réaliste des informations contenues dans de tels jeux de données et des différentes interprétations possibles.

Les méthodes de démixage spectral utilisées classiquement en télédétection (on cherche à décomposer un spectre mesuré en une combinaison linéaire de spectres de référence pris dans un dictionnaire) s'avèrent trop limitées sur les données hyperspectrales issues d'instruments comme OMEGA ou CRISM. La surface de Mars étant partiellement recouverte par une fine couche de poussière, les signatures spectrales des éléments du sous-sol sont cachées dans un spectre au niveau de bruit élevé. D'autre part, l'exploitation d'un modèle de données trop grossier rendent les solutions souvent non interprétables.

Les travaux faisant l'objet de cette demande visent à formuler le problème d'analyse de données hyperspectrales via un modèle plus précis en incorporant des contraintes spécifiques permettant de rendre les solutions interprétables, ceci dans le formalisme de l'optimisation MIP (mixed integer programming). Cette nouvelle approche permettra la résolution exacte des problèmes d'estimation sous-jacents par des algorithmes dédiés, là où les méthodes existantes cumulent erreur de modèle et erreur d'estimation par approches sous-optimales.

Une première spécificité des travaux envisagés réside dans la formalisation du modèle de données. Les spectres mesurés subissent en effet différentes altérations, dues par exemple à la variation de la taille des grains dans les minéraux observés ou à la variation de la teneur en aérosols affectant l'atmosphère martienne. La méthodologie proposée vise à augmenter le dictionnaire de plusieurs versions ainsi déformées des spectres de référence, couplée à des contraintes de parcimonie imposant un faible nombre d'éléments actifs dans la décomposition et à des contraintes de parcimonie structurée imposant qu'au maximum un seul représentant de chaque classe soit actif.

La forte originalité de ces travaux réside en un changement de paradigme où, plutôt que de réaliser l'estimation au sens classique de l'optimisation d'un critère à solution unique (laquelle s'avère souvent ininterprétable en raison du trop fort niveau de bruit sur les données), les méthodes développées retourneront l'ensemble de solutions acceptables, par exemple l'ensemble exhaustif des solutions parcimonieuses compatibles avec le niveau de bruit donné.

Ces solutions, obtenues en chaque position spatiale de la scène observée (en chaque pixel d'un cube de données multispectrales), pourront ensuite faire l'objet d'une interprétation laissée à l'expertise du spécialiste, ou bien serviront d'entrée à une seconde étape de classification / détection, reposant sur

la prise en compte de la régularité spatiale de la présence de chaque classe de minéral (et non pas des cartes d'abondance de chaque spectre de référence, ce qui existe déjà dans la littérature).

L'ensemble de ces éléments sera formulé sous la forme de problèmes d'optimisation de type MIP, possiblement multi-objectif, pour lequel des algorithmes de résolution spécifiques seront construits reposant sur le principe des méthodes de séparation et évaluation (branch-and-bound).

Directeur de thèse : Sébastien Bourguignon, Maître de Conférences HDR (section 61), Laboratoire des Sciences du Numérique de Nantes (LS2N), Centrale Nantes

Co-directeur : Frédéric Schmidt, Professeur (section 35), Laboratoire Géosciences Paris-Saclay (GEOPS), Université Paris Saclay

Description du caractère interdisciplinaire du sujet :

Ce sujet est intrinsèquement interdisciplinaire, reposant sur un croisement de compétences en planétologie, télédétection, spectroscopie et transfert radiatif, apportées par le laboratoire GEOPS, et de savoir-faire en traitement de données et en optimisation numérique (modèles parcimonieux, méthodes branch-and-bound) apportées par le LS2N. Ainsi, la formulation des problèmes précédemment décrits, spécifique aux données de télédétection planétaire, n'a à notre connaissance jamais été envisagée pour l'analyse de telles données, vraisemblablement car leur résolution fait appel à des outils d'optimisation numérique largement méconnus dans cette communauté scientifique.

L'innovation du côté du traitement de données consiste en un développement d'un nouveau paradigme « multi-solution », en combinaison avec des outils de pointe en parcimonie exacte. Les résultats ainsi produits sont extrêmement prometteurs pour nous faire progresser sur la compréhension de la minéralogie martienne et en déchiffrer son histoire complexe. Au delà de l'application martienne, ce type d'outil sera également opportun pour analyser correctement des jeux de données hyperspectrales de télédétection terrestre, tout comme des images micro-spectroscopiques d'échantillons de laboratoire présentant le même type de limitations méthodologiques.

Unités et instituts concernés :

- **Laboratoire des Sciences du Numérique de Nantes (LS2N)**, UMR 6004, CNRS, Centrale Nantes, Nantes Université, IMT Atlantique
 - Rattachement principal : INS2I (Institut des Sciences de l'Information et de leurs Interactions)
 - Rattachements secondaires : INSIS (Institut des Sciences de l'Information et de leurs Interactions) et INSB (Institut des Sciences Biologiques)
 -
- **Laboratoire Géosciences Paris-Saclay (GEOPS)**, UMR 8048, CNRS, Université Paris Saclay
 - Rattachement principal : INSU (Institut des Sciences de l'Univers)
 - Rattachement secondaire : INEE (Institut Écologie et Environnement)