Geophysical Research Abstracts
Vol. 12, EGU2010-10027-2, 2010
EGU General Assembly 2010

Matrix Approximation Techniques for Unsupervised Hyperspectral Data Analysis

Albrecht Schmidt (1), Frédéric Schmidt (2), Erwan Treguier (1), Saïd Moussaoui (3), and Nicolas Dobigeon (4)

(1) European Space Agency, Spain (first.last@esa.int), (2) IDES, Université Paris 11, France (frederic.schmidt@u-psud.fr), (3) IRCCYN, Ecole Centrale Nantes, France (said.moussaoui@irccyn.ec-nantes.fr), (4) IRIT/INPT-ENSEEIHT, Toulouse, France (nicolas.dobigeon@enseeiht.fr)

Unsupervised analysis of hyperspectral data delivered by imaging spectrometers is interesting in many respects in the Planetary Sciences. Since there often is no ground truth to compare to, unsupervised rather than supervised methods allow to extract new information from data sets. So far, the practicability of these methods has suffered from low performance, which makes large-scale analyses almost prohibitively expensive. New research and implementation strategies for non-negative matrix factorisation make it possible to extract sources and relative abundances for typical planetary data sets with reasonable resources. In this work, we try to give an impression of some of the trade-offs and opportunities involved.

Non-negative matrix factorisation is a technique which has enjoyed considerable research and been used in many application areas, from document clustering to spectral analysis. We compare different approaches in terms of resource consumption and viability of results in terms of physical interpretation and meaningfulness in the Planetary Sciences. It turns out that there are datasets and algorithms for which consistently and efficiently meaningful results are returned. The results are meaningful in the sense that abundances as well as extracted source spectra are consistent with community opinion for well-known examples such as ices on the Martian South Poles. It is of particular importance to stress that no prerequisite assumptions are required for the applicability of the algorithms other than the approximate linearity of combinations of the source spectra. This implies that there are a great many scenarios in which the algorithms are applicable and deliver sensible results. To reduce computation times to almost real-time, we currently sample the input. We also provide estimations of how sampling is best done for real data sets and try to quantify the trade-offs.

The focus of the presentation is on implementation techniques and comparative analysis of the performance and limitations of the algorithms. Opportunities for applications and systems engineering will be discussed.