DiSCo: Dictionary-based Spectral unmixing for Coastal environmental monitoring

Supervisor: Jeremy Cohen*, CNRS, Team PANAMA, IRISA (Rennes)

Keywords: Machine Learning, Sparse Coding, Nonnegative Matrix Factorization, Applied Mathematics

Non-technical introduction

Remote sensing, and in particular hyperspectral imaging, provides a mean to monitor rapidly evolving natural environments such as forest or coasts on a large scale. In a context of ever faster degradation of such environments, it has become a crucial ecological issue to improve the performance of hyperspectral imaging processing tools.

Hyperspectral images rely on the physical properties of complex matter. A particular material, such as tree leaves, has a distinguishable response to light stimulation, a.k.a. a spectral signature. Hyperspectral sensors measure such spectral signatures on a large scale to produce an hyperspectral image, which has dimensions (pixels, pixels, wavelengths). The number of wavelength varies from 7 to several hundreds depending on the sensor, see Figure 1 for an illustration.

In the recent years, many international programs have been started to provide a wide coverage of the surface of the earth by hyperspectral imaging, such as NASA AVIRIS Next Generation[1] or Sentinel 2[2]. These airborne or satellites images however typically have a low spatial resolution, due to their high spectral resolution. Therefore, multiple materials of interest may be located on a single pixel. This is a problem for monitoring. Indeed, if multiple materials are located on a single pixel, these spectral signatures are mixed, and direct interpretation of the image is impossible.

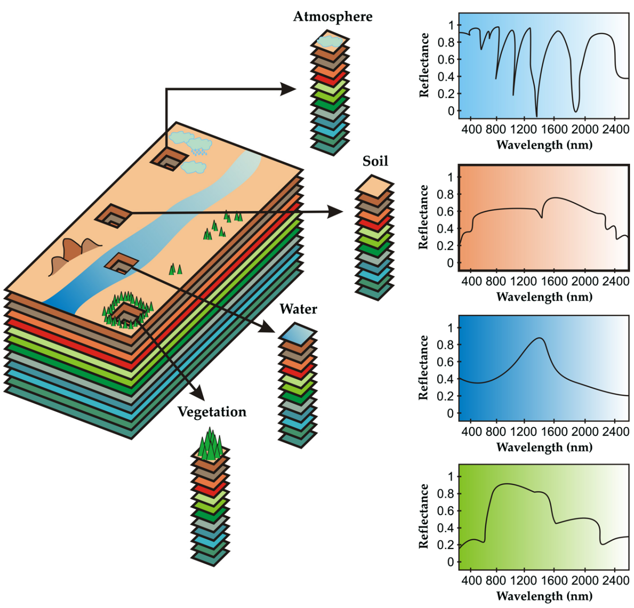

Figure 1: An hyperspectral image schematic representation

Spectral unmixing aims at leveraging machine learning and applied mathematics tools to unmix the spectral signatures, using all the information contained in the image [2]. Among various proposed methods in the literature [6], this internship aims at finding an efficient set of methods for natural images, focusing on coastal images in partnership with the startup HyTECH-imaging. Notably, in the literature, most methods suppose

*https://jeremy-e-cohen.jimdo.com
[1]https://aviris-ng.jpl.nasa.gov/newdata.html
[2]https://www.esa.int/Our_Activities/Observing_the_Earth/Copernicus/Sentinel-2

no a-priori knowledge on the materials present on the image scene, but in reality, large spectral libraries are available.

Recent works [3, 4] have focused on using such known libraries along with a workhorse framework for spectral unmixing: Nonnegative Matrix Factorization (NMF). While NMF is an algebraic model assuming the data matrix has low nonnegative rank [5], the use of libraries falls in the realm of sparse coding [7]. This joint analysis makes the problem appealing not only from a practical point of view, but also in terms of optimization strategies. Formally, the problem is the following:

$$\begin{cases} \underset{\mathcal{K},B}{argmin} \; \|X - D(:,\mathcal{K})B\|_F^2 \\ \forall i,j, \quad B_{ij} \geq 0 \end{cases} \tag{1}$$

where $X \in \mathbb{R}_+^{d\times n}$ is the hyperspectral image vectorized along the pixel mode, $D \in \mathbb{R}_+^{d\times p}$ is the known spectral library, $B \in \mathbb{R}_+^{r\times n}$ is a matrix containing the relative concentrations of materials for each pixel per column and $\mathcal{K}$ is a set of r indices selecting the right atoms in D. Typically, r, which corresponds more or less to the number of materials on the image, is much smaller than both d, the number of wavelengths, and n, the number of pixels. On the other hand, p, the number of spectra in the library, can be quite large.

Goals

Several goals can be identified for this internship:

- Evaluate the performance of dictionary-based matrix factorization methods for spectral unmixing on a wide set of real hyperspectral images. This not only calls for an efficient implementation of these methods, but also for a careful handling of the set of images which is bound to be very large.
- Benchmark existing spectral unmixing methods on coastal hyperspectal images. The wide variety of methods available in the literature are indeed very seldom compared, and there exist no fair extensive comparison in the literature. It is therefore difficult to assess whether the spectral unmixing problem is close to being solved or not, let alone which methods perform the best.
- (Optional) Propose adapted models that account for spectral variability. Indeed, the various spectral signatures may change slightly from pixel to pixel within the same class of material, and accounting for these changes is non-trivial. A solution may resort to defining a good metric to compare spectra within a same class.
- (Optional) Propose sparsity-based models that leverage the fact that only two or three materials may mix at each pixel, despite the image containing more materials. The intern may implement an efficient algorithm using this prior knowledge, and/or study the identifiability (uniqueness properties) of this model.
- (Optional) Make use of spectral unmixing for other tasks related to hyperspectral images processing, such as clustering, segmentation and detection. The comparison with neural-networks-based model will be investigated.

The goals proposed here can of course be modulated depending on the preferences of the intern. It is possible for instance to work mostly on the theoretical aspects of the machine learning models used in spectral unmixing. On the other hand, it is also possible to focus on the data processing side of the project without paying too much attention to the mathematical foundations.

Remark: The intern may pursue his work as a PhD student in the PANAMA team if he is willing to.

Academic Partnership

This project will be realized in collaboration with Nicolas Gillis[3] from the University of Mons, Belgium, who is an international expert on nonnegative matrix factorization.

The intern will also benefit from discussions with Nicolas Courty, professor at Université Bretagne Sud, who is visiting the PANAMA team for one year.

[3] https://sites.google.com/site/nicolasgillis/

Industrial Partnership

This project will benefit from a direct collaboration with the start-up HyTECH-imaging[4], which will provide a front-end expertise as well as data sets with a partially known ground truth and spectral libraries. Contact: Tristan Petit. HyTECH-imaging is based in Brest.

Moreover, the intern will make use of publicly available airborne hyperspectral images and spectral libraries from the NASA projects AVIRIS Next-Gen and ECOSTRESS [1][5].

An additional industrial connection would be to compare the proposed Dictionary-based matrix factorization models with industrial software such as ENVI[6] when it comes to spectral unmixing.

Suggested Prior Knowledge

The following list contains some data science and mathematical disciplines which knowledge, albeit not required, would make the beginning of the internship easier and faster.

- Machine Learning (principal component analysis, sparse coding, dictionary learning, covariance estimation)
- Numerical Optimization (least squares formulation, gradient descent, constrained optimization)

Acquired Skills

During the course of the internship, the following concepts will be developed:

- Matrix factorization (low rank factorization, constrained matrix factorization)
- Constrained optimization (Sparse Coding, nonnegative least squares) and Discrete Optimization (Branch and Bound)
- Hyperspectral image processing (spectral unmixing, endmember identification, spectral variability, endmember similarity measurement) and remote sensing (type of sensors, reflectance, linear mixing model)

References

[1] AM Baldridge, SJ Hook, CI Grove, and G Rivera. The ASTER spectral library version 2.0. *Remote Sensing of Environment*, 113(4):711–715, 2009.

[2] J.M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot. Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(2):354–379, 2012.

[3] J.E. Cohen and N. Gillis. A new approach to dictionary-based nonnegative matrix factorization. In *2017 Proceedings of the 25th European Signal Processing Conference (EUSIPCO)*. IEEE, 2017.

[4] J.E. Cohen and N. Gillis. Spectral unmixing with multiple dictionaries. *IEEE Geoscience and Remote Sensing Letters*, 15(2):187–191, 2018.

[5] N. Gillis. *The Why and How of Nonnegative Matrix Factorization*. PhD thesis.

[6] W. K. Ma, J. M. Bioucas-Dias, T. H. Chan, N. Gillis, P. Gader, A. J. Plaza, A. Ambikapathi, and C. Y. Chi. A signal processing perspective on hyperspectral unmixing: Insights from remote sensing. *IEEE Signal Processing Magazine*, 31(1):67–81, 2014.

[7] J.A. Tropp. Greed is good: Algorithmic results for sparse approximation. *IEEE Transactions on Information theory*, 50(10):2231–2242, 2004.

[4] https://hytech-imaging.fr/
[5] https://ecostress.jpl.nasa.gov/
[6] https://www.harrisgeospatial.com/SoftwareTechnology/ENVI.aspx