

Hyperspectral image unmixing for remote sensing

Lucas Drumetz
lucas.drumetz@imt-atlantique.fr

Copernicus Master in Digital Earth
October 11, 2021



IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom



Outline

- 1 Motivation
 - Hyperspectral images
 - Supervised classification
 - Classification vs unmixing
- 2 Mixing Model, geometric interpretation
 - Mixing Model
 - Geometric interpretation
- 3 Linear Unmixing chain
 - How many endmembers?
 - Geometrical endmember extraction
 - Abundance estimation
- 4 Limitations of the LMM
 - Non Linear mixtures
 - Endmember variability
- 5 Recent Trends
 - Multitemporal aspects

Outline

1 Motivation

- Hyperspectral images
- Supervised classification
- Classification vs unmixing

2 Mixing Model, geometric interpretation

- Mixing Model
- Geometric interpretation

3 Linear Unmixing chain

- How many endmembers?
- Geometrical endmember extraction
- Abundance estimation

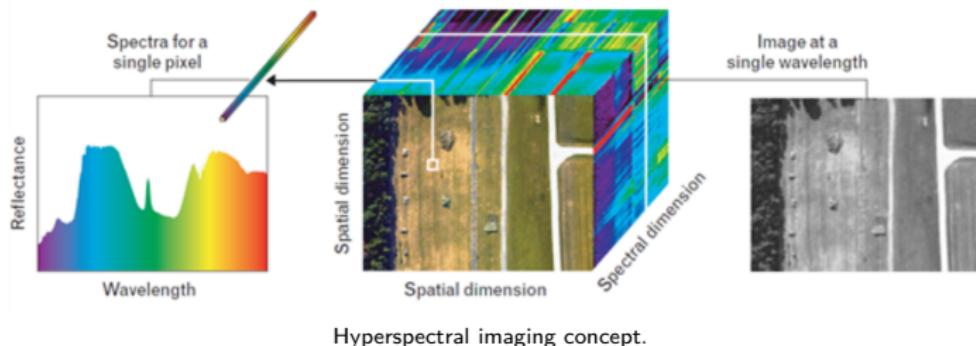
4 Limitations of the LMM

- Non Linear mixtures
- Endmember variability

5 Recent Trends

- Multitemporal aspects

Hyperspectral imaging



- Each pixel: full reflectance spectrum for many contiguous and narrow wavelengths (visible and near-IR).
- HSI: also a collection of gray-level reflectance images for each wavelength.
- The spectrum characterizes a material.

Applications in: **Remote sensing** – environment monitoring, geology, precision agriculture, planetary science, defense ...
but also food processing, chemometrics, document analysis...

We will denote as L the number of spectral bands in the image, and N the number of pixels.

Supervised classification

Supervised Classification

Assign each pixel of an image to **one** out of P predefined semantic classes, using **ground truth pixels** to **train** the algorithm

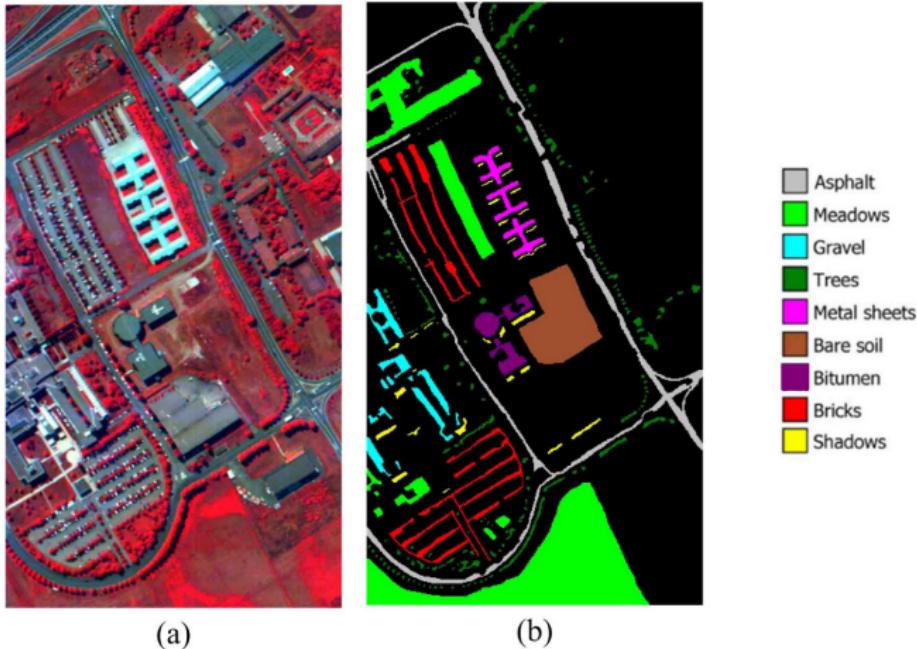


Fig. 3: (a) Three-band color composite image of AVIRIS data. (b) Reference map.

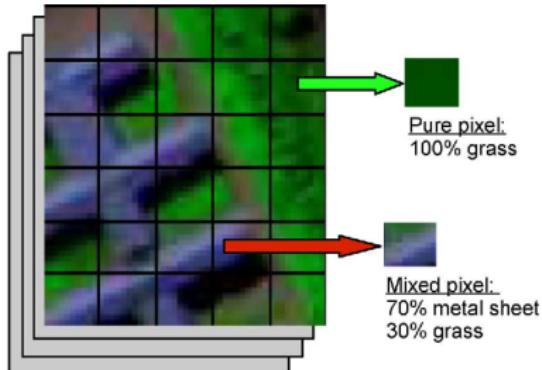
Classification pros and cons

Pros:

- Classical Machine Learning problem
- Results are easy to interpret
- Algorithms are relatively scalable

Cons:

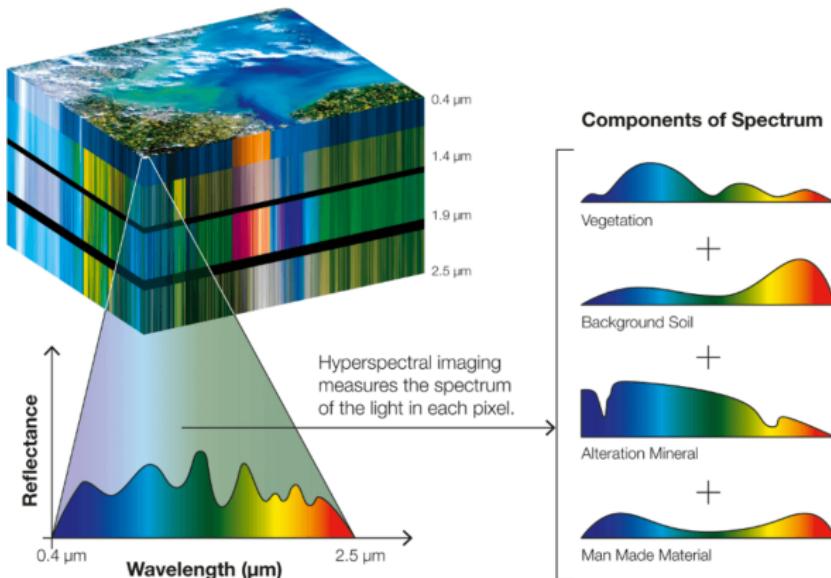
- Ground truth is needed (costly, unreliable, and rarely available in practice)
- Limited by the spatial resolution of the image
- Only outputs the majoritary material in the pixel



What happens in this case?

Notion of pure and mixed pixel

Because of the limited spatial resolution of the sensor, there can be several materials of interest in the FOV of one pixel (macroscopic mixing)



Spectral Unmixing

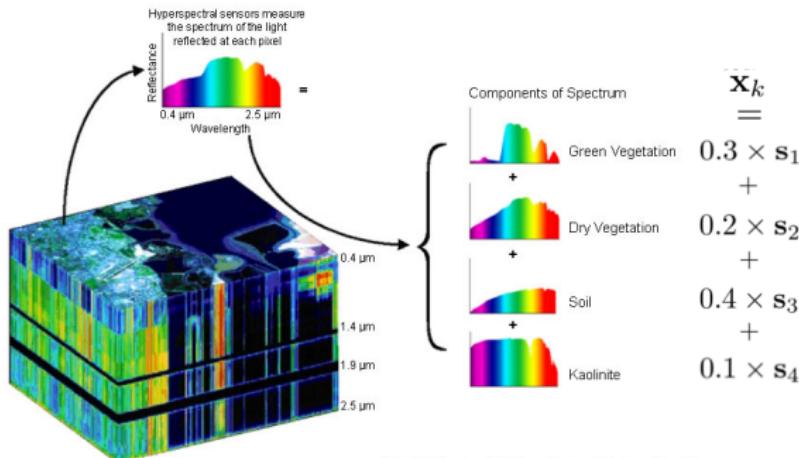
We want to address two limitations of classification: mixed pixels and absence of ground truth.

The unmixing problem

We want to automatically identify

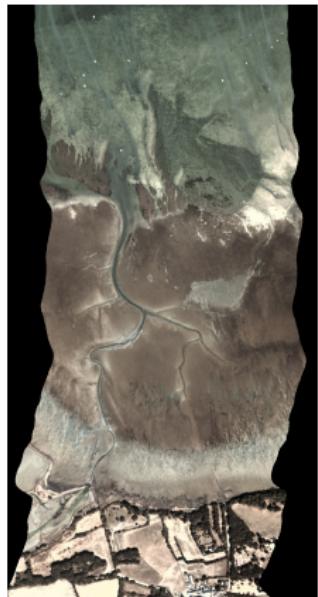
- the **spectral signatures** of the pure materials present in the image
- and their **proportions in each pixel**

In an **unsupervised** way!



A few applications

Spatial characterization of marine vegetation [[bajjouk2019](#)]. Spatial resolution: 80 cm and a spectral resolution of 5 nm with 126 bands.



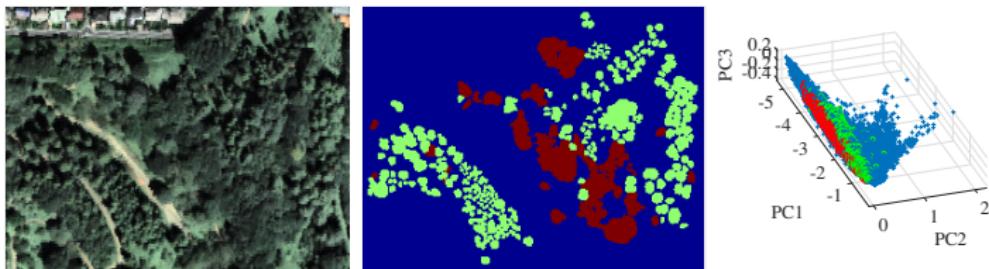
- Data acquired in the gulf of Morbihan (IFREMER)
- Marine vegetation (*zostera*) is a bio-tracer of the quality of the ecosystem
- Dictionary of spectra acquired in situ

Challenges: intra-class variability, inter-class similarity, nonlinear effects due to the water column.

A few applications (cont'd) [Drumetz et al. 2020]

Tama Forest Garden, near Tokyo, Japan. 72 bands in the visible and near IR. 1m spatial resolution.

Challenges: Shadowed areas, Intra class variability, Inter class similarity



RGB image, ground truth with conifer trees in green, and broadleaf trees in red (middle), data and ground truth scatterplot (using a PCA).

How can we differentiate between conifer and broadleaf trees with no ground truth? And obtain abundances?

Outline

1 Motivation

- Hyperspectral images
- Supervised classification
- Classification vs unmixing

2 Mixing Model, geometric interpretation

- Mixing Model
- Geometric interpretation

3 Linear Unmixing chain

- How many endmembers?
- Geometrical endmember extraction
- Abundance estimation

4 Limitations of the LMM

- Non Linear mixtures
- Endmember variability

5 Recent Trends

- Multitemporal aspects

Linear Mixing Model

If we assume that we observe a flat surface, with each of the P materials of the image occupying a fraction of the pixel's surface (no intimate mixing), we can write:

Linear Mixing Model (LMM) [Bioucas 2012]

$$\mathbf{x}_n = \sum_{p=1}^P a_{pn} \mathbf{s}_p + \mathbf{e}_n = \mathbf{S}\mathbf{a}_n + \mathbf{e}_n$$

where

$\mathbf{x}_n \in \mathbb{R}^L$ is the observed reflectance

$\mathbf{s}_p \in \mathbb{R}^L$ is the spectral signature of one of the P pure materials (and $\mathbf{S} \in \mathbb{R}^{L \times P}$ stores all of them in the columns of a matrix)

$\mathbf{a}_n \in \mathbb{R}^P$ stores the coefficients a_{pn} in a vector

$\mathbf{e}_n \in \mathbb{R}^L$ is an additive noise

Terminology

The spectra of the materials are usually referred to as *endmembers*, and the coefficients as *abundances*

Constraints on the abundances

The abundances are proportions! There cannot be negative contributions of a material and the contributions should add up to 1.

- Abundance non-negativity constraint (ANC) $a_{pn} \geq 0, \forall \{p, k\}$
- Abundance sum-to-one constraint (ASC) $\sum_{p=1}^P a_{pn} = 1, \forall n$

In a matrix form, we can rewrite:

$$\mathbf{X} = \mathbf{SA} + \mathbf{E}$$

with $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] \in \mathbb{R}^{L \times N}$, $\mathbf{E} \in \mathbb{R}^{L \times N}$, $\mathbf{S} \in \mathbb{R}^{L \times P}$ and $\mathbf{A} \in \mathbb{R}^{P \times N}$.

This is a matrix factorization problem, or a blind source separation problem.
Most approaches for those are purely statistical. Some examples:

- Principal Component Analysis
- Independent Component Analysis

But their assumptions are not compatible with hyperspectral data (constraints on the abundances, hypotheses on the decorrelation of the sources...)

An ill-posed inverse problem

An inverse problem is composed of:

- A so called "forward" model describing the relationship between the observed data and parameters of interest
- A dataset
- The goal is to retrieve the "optimal" values of the parameters by inverting the model.

In many cases, the model is not injective.

Well-posed problem (Hadamard)

- A solution exists
- It is unique
- It depends continuously on the data

Here, even if in the noiseless case we find \mathbf{S} and \mathbf{A} such that $\mathbf{X} = \mathbf{SA}$, for any invertible matrix $\mathbf{M} \in \mathbf{R}^{P \times P}$ we have $\mathbf{X} = \mathbf{SM}\mathbf{M}^{-1}\mathbf{A}$.

In particular, permuting the columns of the endmember matrix \mathbf{S} does not change the solution.

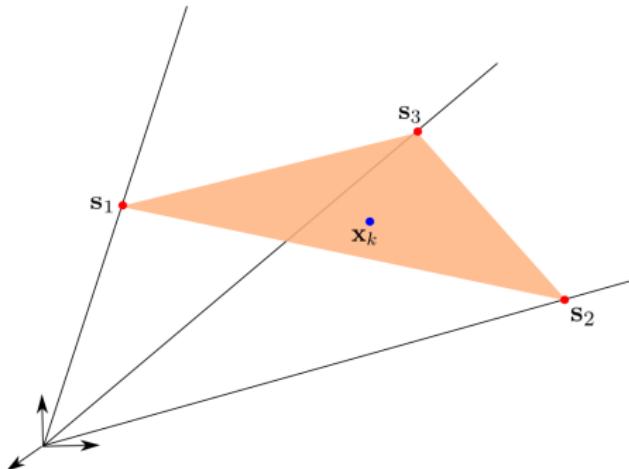
Geometric interpretation of the problem

Let us assume that we only have three spectral bands ($L = 3$, e.g. RGB images), and that there are only three materials in the image ($P = 3$). We also assume that there is no noise.

Geometric consequences of the constraints

- The ANC forces the data to live in the *convex cone* spanned by the endmembers
- The ASC forces the data to lie in the plane containing the endmembers

The intersection is the *convex hull* of the endmembers: here it is a triangle.



Notion of simplex

A P -simplex is the most simple object in P (or more) dimensions made from $P + 1$ points.

Definition

A subset $\mathcal{S} \subseteq \mathbb{R}^L$ ($L \geq P$) is a P -simplex if there exist P affinely independent points $\{\mathbf{s}_1, \dots, \mathbf{s}_P\} \in \mathbb{R}^L$ such that $\mathcal{S} = \text{conv}\{\mathbf{s}_1, \dots, \mathbf{s}_P\}$, where this denotes the convex hull of $\{\mathbf{s}_1, \dots, \mathbf{s}_P\}$, i.e. the set

$$\text{conv}\{\mathbf{s}_1, \dots, \mathbf{s}_P\} = \left\{ \mathbf{x} = \sum_{p=1}^P a_p \mathbf{s}_p, \mathbf{a} \geq \mathbf{0}, \mathbf{1}_P^\top \mathbf{a} = 1 \right\}$$

where $\mathbf{1}_L$ denotes a vector of ones whose size is given in index.

Examples:

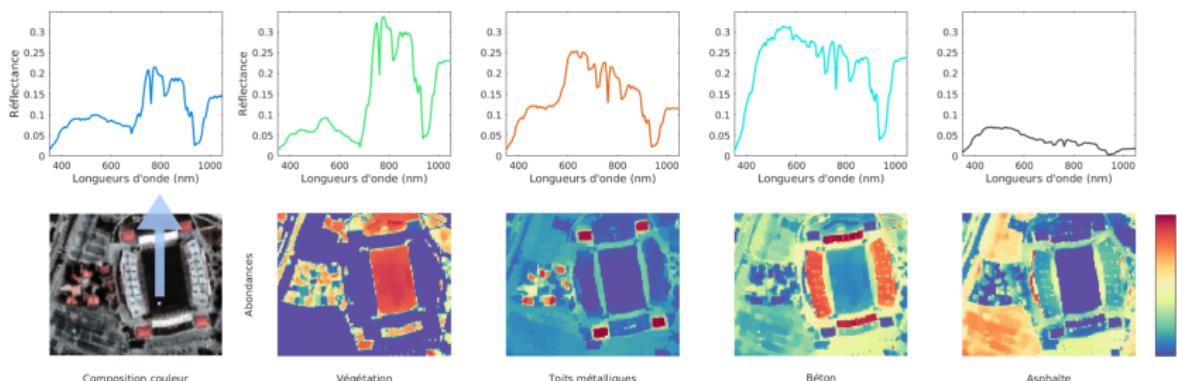
- On a line, we can form a segment with two distinct points (1-simplex)
- In the plane, we can form a triangle with three non-aligned points.
- In a 3D space, we can form a tetrahedron with 4 non-coplanar points.
- and so on...

The abundances live in the "unit" simplex, and hence the data live in a simplex too (through a linear transform)

Typical Blind Unmixing Chain

- ① Find out how many endmembers to use from the data
- ② Extract the endmember signatures from the data
- ③ Estimate the abundances

If reference spectral signatures are available for all materials (from expert knowledge, on site acquisition...) then steps 1 and 2 can be skipped ("non blind" unmixing)



An example of result. First column: one observed pixel and an RGB composition of the image. Other columns: Endmembers are on the top row, and corresponding abundance maps are on the bottom row.

Outline

1 Motivation

- Hyperspectral images
- Supervised classification
- Classification vs unmixing

2 Mixing Model, geometric interpretation

- Mixing Model
- Geometric interpretation

3 Linear Unmixing chain

- How many endmembers?
- Geometrical endmember extraction
- Abundance estimation

4 Limitations of the LMM

- Non Linear mixtures
- Endmember variability

5 Recent Trends

- Multitemporal aspects

Intrinsic Dimensionality and number of endmembers

How can we infer how many endmembers to use from the data?

Definition

We denote by \mathbf{y}_n a noiseless pixel, so that $\mathbf{x}_n = \mathbf{y}_n + \mathbf{e}_n$.

The Intrinsic Dimension (ID) of a dataset, $\mathbf{x}_1, \dots, \mathbf{x}_n$, is the dimension, d , of the vector subspace spanned by the signals, $\mathbf{y}_1, \dots, \mathbf{y}_n$.

$$d = \dim(\text{span}(\mathbf{y}_1, \dots, \mathbf{y}_n)) = \text{rank}(\mathbf{Y})$$

Example: if there are 3 endmembers (that are linearly independent in \mathbb{R}^L) and the data is generated by the LMM, then $d = 3$.

For the LMM we have $d = P$. This is not true in general for more complex models (and not always verified on real data...)

Many algorithms exist to try to estimate this number from data.

Extracting pure pixels

Pure pixels

Pixels for which the abundance is 1 for one material and 0 for all the others.

Why do we call pure spectra "endmembers"?

→ because they lie on the extreme parts of the dataset ("end-member")

Endmember extraction algorithms find the extreme points in the dataset. Most popular ones: Vertex Component Analysis (VCA) [[Nascimento 2005](#)] and Successive Projection Algorithm (SPA) [[Araujo 2001](#)].

Successive Projection Algorithm

for $i = 1, \dots, d$

- ① find the pixel index with the maximum norm and keep the corresponding pixel as an endmember
- ② project the dataset onto the subspace orthogonal to the subspace spanned by the previously found directions

Those algorithms are not naturally robust to outliers.

What is the definition of an endmember? Geometric definition vs semantic (subjective and application dependent definition)

SPA algorithm

Algorithm 1 SPA

Input: Nearly separable matrix X (Assumption 1), the number r of columns to be extracted, and a strongly convex function f (Assumption 2).

Output: Index set \mathcal{K} such that $X \approx X(:, \mathcal{K})H$ with $H \geq 0$.

- 1: Let $R = X$, $\mathcal{K} = \{\}$, $k = 1$.
- 2: **while** $R \neq 0$ and $k \leq r$ **do**
- 3: $k^* = \operatorname{argmax}_k f(R(:, k))$.
- 4: $u_j = R(:, k^*) / \|R(:, k^*)\|_2$.
- 5: $R \leftarrow (I - u_j u_j^T)R = R - u_j(u_j^T R)$.
- 6: $\mathcal{K} = \mathcal{K} \cup \{k^*\}$, $k = k + 1$.
- 7: **end while**

from [Gillis 2019]

"Nearly separable matrix": $\exists \mathbf{S} \geq \mathbf{0}, \mathbf{X} = \mathbf{S}\mathbf{A}$ with the columns of \mathbf{A} non-negative and summing to one.

"strongly convex function f ": the \mathcal{L}_2 norm $\|\cdot\|_2$ works.

We can check that the projected data at each step is orthogonal to \mathbf{u}_j (and all the previous ones).

Abundance estimation

Once the endmembers have been identified, there is only one last thing to do: estimate the abundances.

Under the LMM, this can be done by solving the following optimization problem [Heinz 2001]:

$$\arg \min_{\mathbf{A} \geq 0, \mathbf{1}_P^T \mathbf{A} = \mathbf{1}_N} \|\mathbf{X} - \mathbf{SA}\|_F^2$$

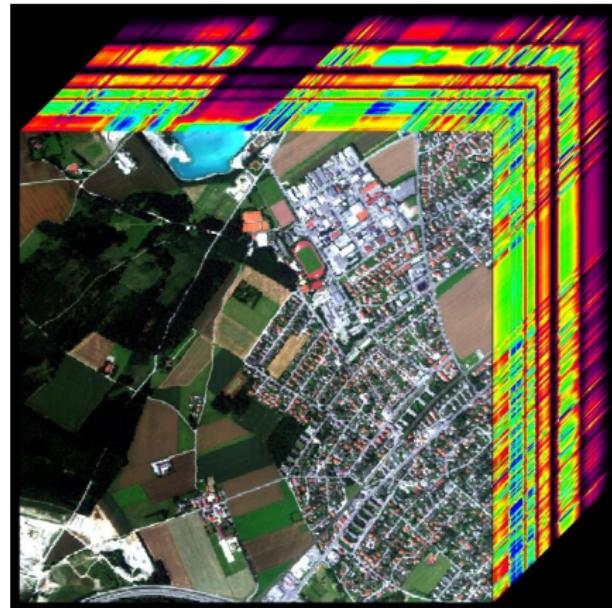
where $\|\cdot\|_F$ is the Frobenius norm, i.e. $\|\mathbf{X}\|_F = \sqrt{\sum_i \sum_j x_{ij}^2}$.

This is an optimization problem with two constraints, dedicated algorithms can be used to look for a local minimum.

We can prove there is every local minimum to the cost function is also global (convex optimization problem).

Regularizations

What information are we not taking into account so far?



- There are typically only a few materials in each pixel
- One pixel is very correlated to its neighbors

Regularizations (cont'd)

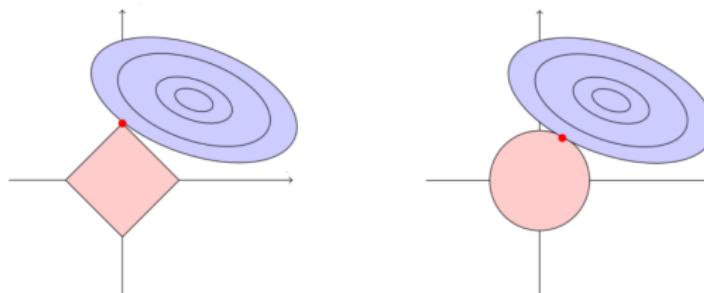
In many cases, we are going to add additional terms to the cost function to enforce desirable properties to the abundances:

$$\underset{\mathbf{A} \geq 0, \mathbf{1}_P^T \mathbf{A} = \mathbf{1}_N}{\arg \min} \|\mathbf{X} - \mathbf{S}\mathbf{A}\|_F^2 + \lambda \mathcal{R}(\mathbf{A})$$

Sparsity

A sparse solution has a certain number of zero entries for each abundance vector.

We use the \mathcal{L}_1 norm $\mathcal{R}(\mathbf{A}) = \sum_n \|\mathbf{a}_n\|_1 = \sum_n \sum_i |a_{ni}|$ [Iordache2011]
Geometrically, pixels tend to lie on facets of the simplex.



A geometric explanation of why the \mathcal{L}_1 norm (left) enforces sparsity compared to the \mathcal{L}_2 norm.

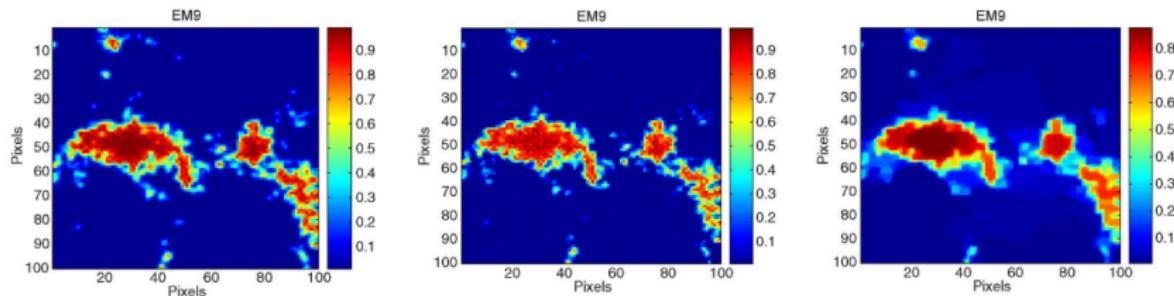
Regularizations (cont'd)

Spatial regularizations

The abundance in one pixel should look like the abundance of its neighbors.

We use the norm of the spatial gradient of the abundances $\sum_n \|\nabla \mathbf{a}_n\|$ (total variation) [Iordache2012]

This will be especially useful when the data is very noisy (e.g. SNR < 25 dB)



Ground truth abundance (left), usual linear unmixing solution (middle) and solution with TV regularization (right)

Those properties are useful, but they make the optimization problems to solve much harder and computationally expensive.

Outline

1 Motivation

- Hyperspectral images
- Supervised classification
- Classification vs unmixing

2 Mixing Model, geometric interpretation

- Mixing Model
- Geometric interpretation

3 Linear Unmixing chain

- How many endmembers?
- Geometrical endmember extraction
- Abundance estimation

4 Limitations of the LMM

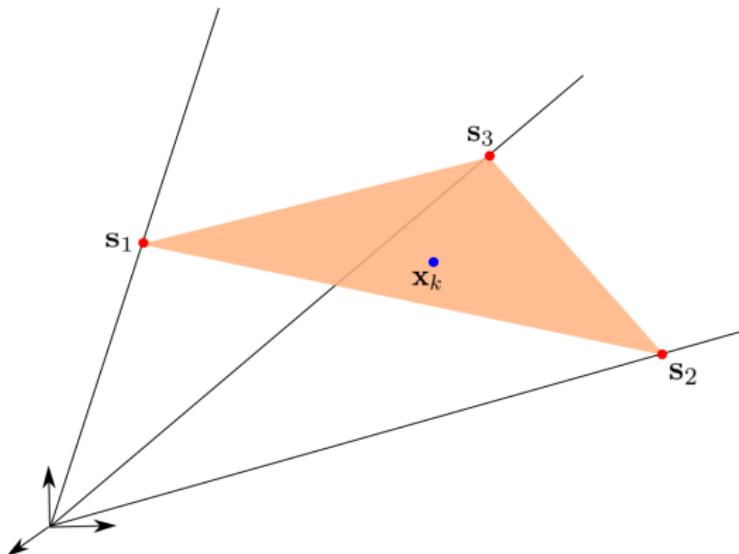
- Non Linear mixtures
- Endmember variability

5 Recent Trends

- Multitemporal aspects

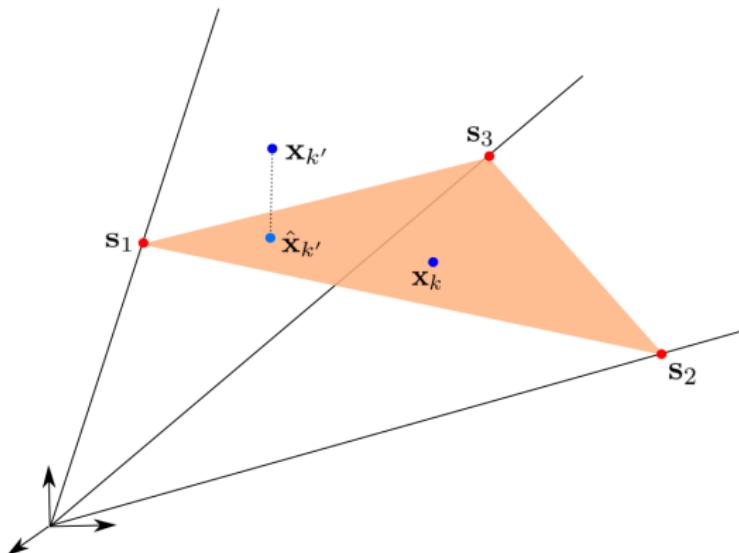
Main limitations

- Nonlinear mixtures
- Intra-class variability of the materials



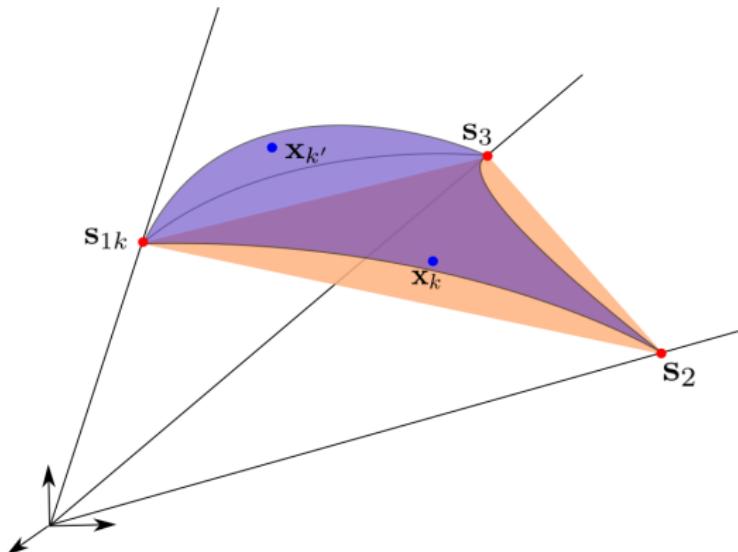
Main limitations

- Nonlinear mixtures
- Intra-class variability of the materials



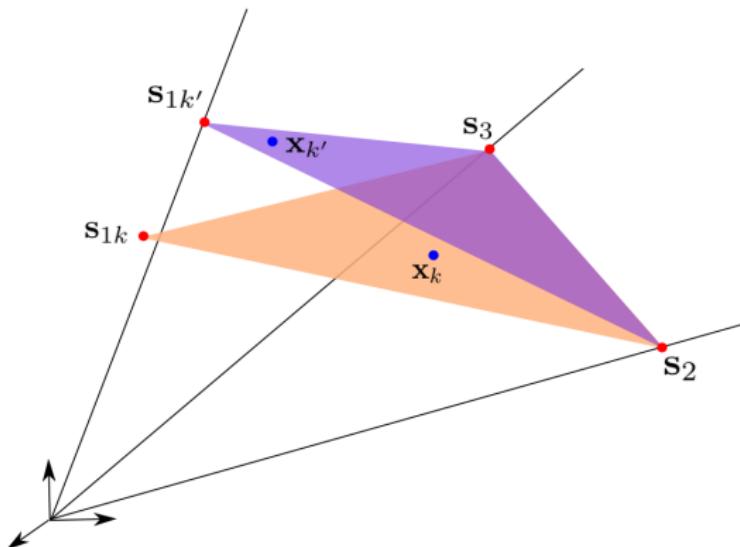
Main limitations

- Nonlinear mixtures
- Intra-class variability of the materials



Main limitations

- Nonlinear mixtures
- Intra-class variability of the materials

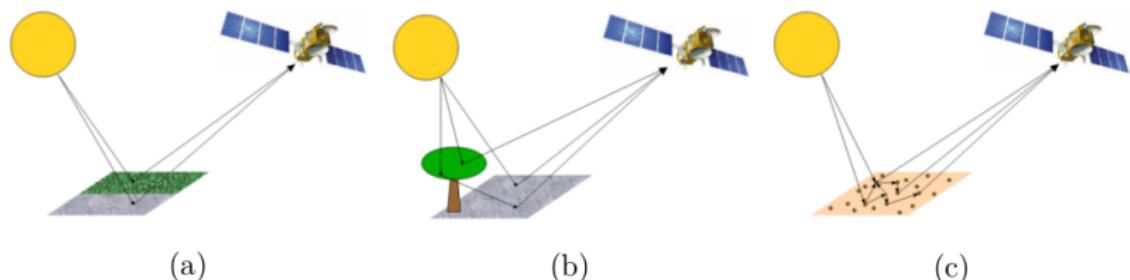


Non Linear Mixtures

In some cases, the linear mixing model does not hold, especially when the observed scene has a 3D structure, e.g. in tree canopies, or when there are shadows, or intimate mixing effects.

$$\mathbf{x} = f(\mathbf{S}, \mathbf{a}) + \mathbf{e}$$

where f is a non linear function [Heylen 2014].



Several mixing configurations. In a "checkerboard" scenario, the LMM is valid (a). Each material occupies a fraction of the (flat) surface in the FOV of the sensor. In (b) is displayed a case of multiple reflections of light before reaching the sensor. (c) shows a case of intimate mixing

For multiple reflections, bilinear or multilinear models are often used:

$$\mathbf{x} = \mathbf{S}\mathbf{a} + \sum_i \sum_j b_{ij} (\mathbf{s}_i \odot \mathbf{s}_j) + \mathbf{e}$$

with different constraints and flavors for the coefficients \mathbf{a} and \mathbf{b} . Then we must estimate both \mathbf{a} and \mathbf{b} during the optimization.

Endmember variability

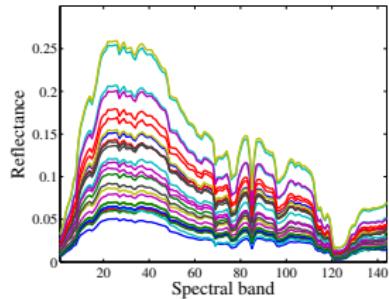
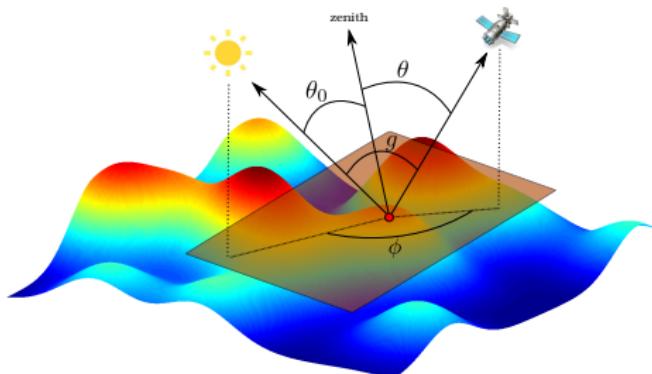
In the LMM, one assumes that a single spectral signature is completely characteristic of a material. A more accurate model would be, even if the LMM is valid:

$$\mathbf{x}_n = \mathbf{S}_n \mathbf{a}_n + \mathbf{e}_n$$

i.e. the endmember matrix is pixel dependent! [Borsig 2020]

The main causes of variability are:

- Changing illumination conditions and topography of the scene
- Intrinsic variability of the material: e.g. amount of chlorophyll in grass



Various spectra of the same material

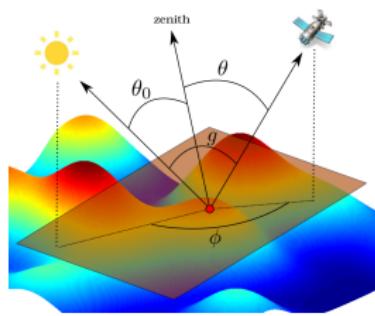
Acquisition angles for a given pixel influence the observed spectral signature.

Tackling spectral variability

Hence, we need a model on the spatial variability of the endmembers. Three main directions:

- using a physically motivated parametric variability model
- using a statistical model to allow fluctuations of the endmembers around references
- using several variants of the spectra of each material as a dictionary of candidate endmembers

Designing an Extended Linear Mixing Model



Acquisition angles for a pixel influence the observed reflectance.

The Hapke model [Hapke 2012]

$$\rho = f(\omega, \Theta, \Pi)$$

- $\Theta = [\theta, \theta_0, \phi, g]$: geometric parameters in a pixel
- Π : photometric parameters of the material.
- ρ : reflectance of the material in a given wavelength
- ω : albedo of the material in a given wavelength.



Approximation of the Hapke model [Drumetz 2019a]

The ratio between the reflectance endmember s_{kp} in a pixel, to that of a reference s_{0p} (two different geometries), for all wavelengths:

$$s_{pn} \approx \psi_n(\Theta_n) s_{0p}$$

- Each pixel is simply scaled by an unknown scaling factor:

$$\mathbf{x}_n = \sum_{p=1}^P a_{pn} \psi_n \mathbf{s}_{0p} + \mathbf{e}_n = \psi_n \sum_{p=1}^P a_{pn} \mathbf{s}_{0p} + \mathbf{e}_n = \mathbf{S}_0 \psi_n \mathbf{a}_n + \mathbf{e}_n.$$

- More flexible model with one scaling factor for each material:

$$\mathbf{x}_n = \sum_{p=1}^P a_{pn} \psi_{pn} \mathbf{s}_{0p} + \mathbf{e}_n = \mathbf{S}_0 \psi_n \mathbf{a}_n + \mathbf{e}_n,$$

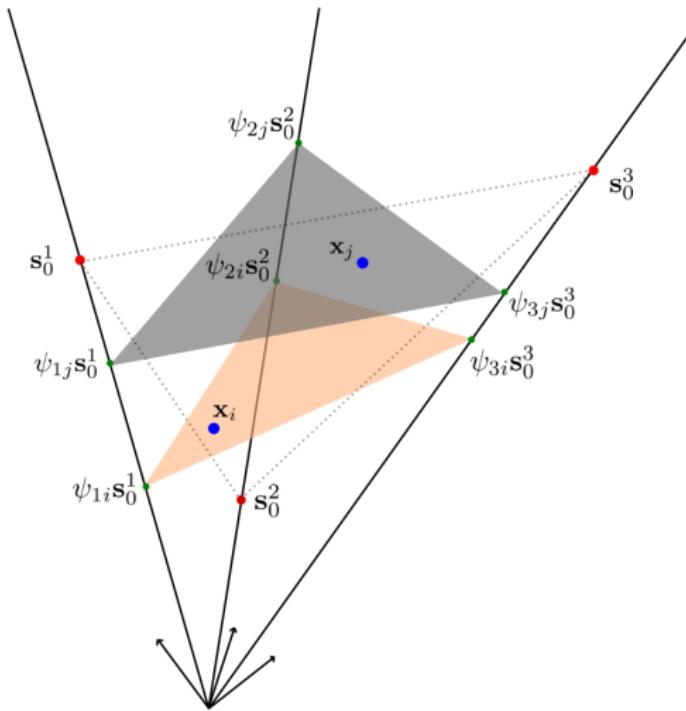
$\psi_n \in \mathbb{R}^{P \times P}$ is a diagonal matrix of scaling factors.

Matricial version of the model

$$\mathbf{X} = \mathbf{S}_0 (\mathbf{A} \odot \Psi) + \mathbf{E}$$

with $\Psi \in \mathbb{R}^{P \times N}$ gathering all the scaling factors, and \odot the Hadamard product.

This model is not identifiable! We'll need to regularize.



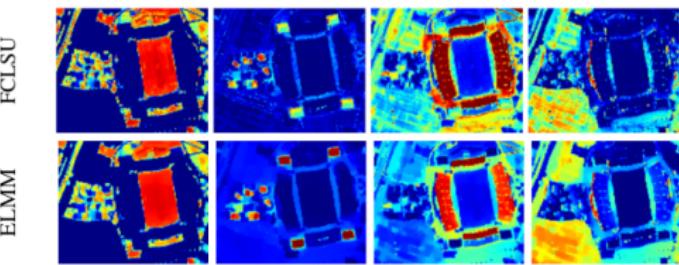
Geometric interpretation of the ELMM in the case of three endmembers, in a 3D ambient space

Real data in a simple case (cont'd) [\[drumetztip\]](#)

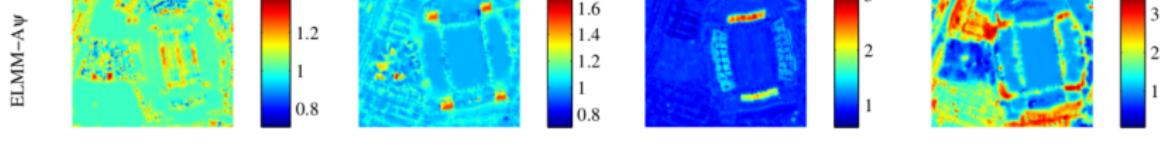
Campus of the University of Houston, Texas, USA. 144 bands in the visible and near IR, and comes with a LiDAR dataset with the same spatial resolution (2.5 m).



(a) RGB representation



(b) Abundances for LMM (top row) and the ELMM (bottom row)



(c) Scaling factors for the ELMM

Basic idea

Assign a prior distribution $p(\mathbf{S})$ to the endmembers and abundances, Make assumptions on the noise statistics and then use Bayes' theorem to compute, and maximize the posterior distribution of the endmembers and abundances.

A simple example:

- We assume a Gaussian observation noise $p(\mathbf{x}|\mathbf{S}_n, \mathbf{a}_n) = \mathcal{N}(\mathbf{S}_n \mathbf{a}_n, \sigma_e^2 \mathbf{I}_L)$
- We assign a uniform prior distribution on the simplex for the abundances $p(\mathbf{a}_n) = \mathcal{U}_{\Delta_P}(\mathbf{a}_n)$
- We assign a Gaussian distribution to each endmember, centered around a reference \mathbf{S}_0 : $p(\mathbf{s}_p) = \mathcal{N}(\mathbf{s}_{0p}, \sigma_s^2 \mathbf{I}_L)$

Then Bayes' theorem gives us:

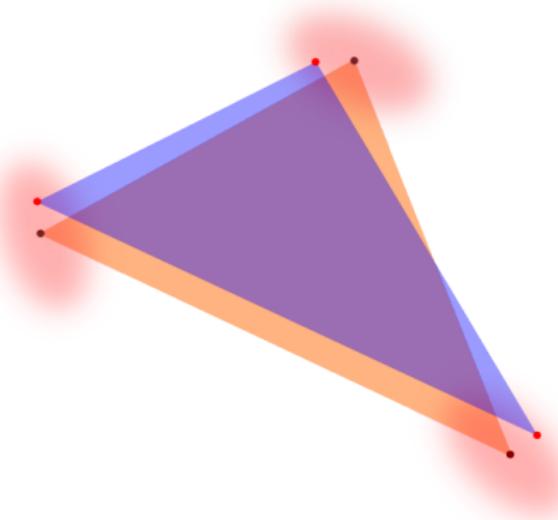
$$p(\mathbf{S}_n, \mathbf{a}_n | \mathbf{X}) \propto p(\mathbf{x} | \mathbf{S}_n, \mathbf{a}_n) p(\mathbf{a}_n) p(\mathbf{S}_n)$$

Statistical Models (cont'd)

Assuming the pixels are i.i.d., we can show that the Maximum A Posteriori estimate of (\mathbf{S}, \mathbf{A}) is given by solving:

$$\arg \min_{\mathbf{A} \geq 0, \mathbf{1}_P^T \mathbf{A} = \mathbf{1}_N, \mathcal{S}} \frac{1}{2} \sum_{n=1}^N \left(\|\mathbf{x}_n - \mathbf{S}_n \mathbf{a}_n\|^2 + \frac{\sigma_e^2}{\sigma_s^2} \|\mathbf{S}_n - \mathbf{S}_0\|_F^2 \right)$$

Now we have to optimize w.r.t. the endmembers in each pixel as well, and we may obtain local minima.



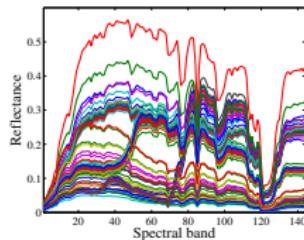
Endmembers in each pixel are Gaussian distributed around references (in brown).

Spectral bundles

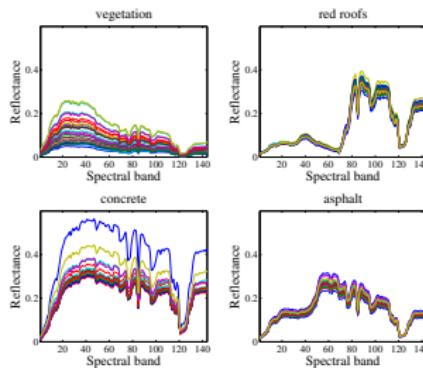
→ Recover multiple instances of the endmembers [Somers 2012].

Automated Endmember Bundles

- Endmember Extraction on random subsets of the data
- Extracted dictionary $\mathbf{B} \in \mathbb{R}^{L \times Q}$ structured in P groups



Set of extracted endmembers.



Spectral bundles after clustering into four classes.

Geometric Interpretation (LMM)

Estimate the abundances with:

$$\arg \min_{\mathbf{a}_k \in \Delta_P, \forall k=1, \dots, N} \|\mathbf{X} - \mathbf{BA}\|_F^2$$

and sum all the contributions in each class.

→ Allows to define local endmembers: convex combinations of the extracted spectra



Geometric interpretation of using FCLSU with spectral bundles.

Different flavors of sparsity can be added for a more accurate abundance estimation [Drumetz 2019b].

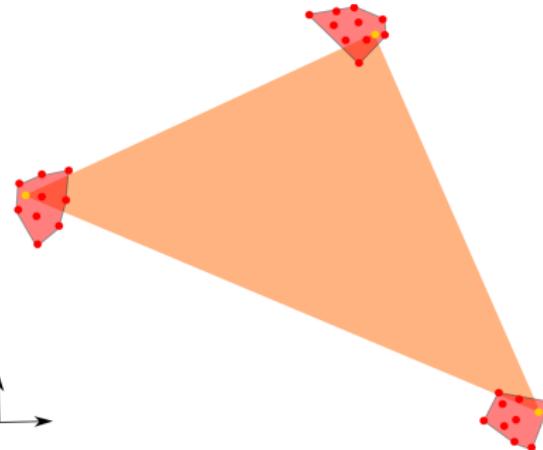
Geometric Interpretation (LMM)

Estimate the abundances with:

$$\arg \min_{\mathbf{a}_k \in \Delta_P, \forall k=1,\dots,N} \|\mathbf{X} - \mathbf{BA}\|_F^2$$

and sum all the contributions in each class.

→ Allows to define local endmembers: convex combinations of the extracted spectra



Geometric interpretation of using FCLSU with spectral bundles.

Different flavors of sparsity can be added for a more accurate abundance estimation [Drumetz 2019b].

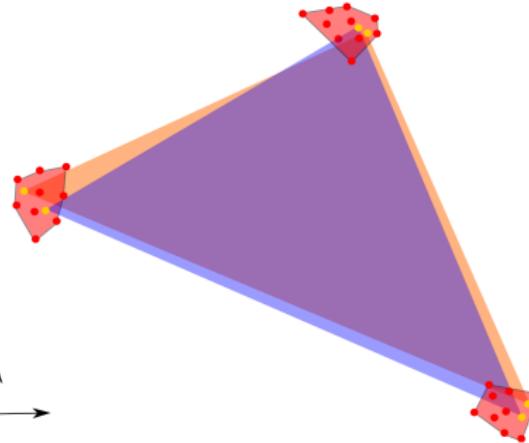
Geometric Interpretation (LMM)

Estimate the abundances with:

$$\arg \min_{\mathbf{a}_k \in \Delta_P, \forall k=1, \dots, N} \|\mathbf{X} - \mathbf{BA}\|_F^2$$

and sum all the contributions in each class.

→ Allows to define local endmembers: convex combinations of the extracted spectra



Geometric interpretation of using FCLSU with spectral bundles.

Different flavors of sparsity can be added for a more accurate abundance estimation [Drumetz 2019b].

Outline

1 Motivation

- Hyperspectral images
- Supervised classification
- Classification vs unmixing

2 Mixing Model, geometric interpretation

- Mixing Model
- Geometric interpretation

3 Linear Unmixing chain

- How many endmembers?
- Geometrical endmember extraction
- Abundance estimation

4 Limitations of the LMM

- Non Linear mixtures
- Endmember variability

5 Recent Trends

- Multitemporal aspects

Hyperspectral times series

- Highly irregular and sparse sampling in time, and data are noisy
- The more regular the sampling, the lower the spectral resolution for airborne or spaceborne data
- Potential registration issues between frames (if acquired with different satellites or with airborne sensors)

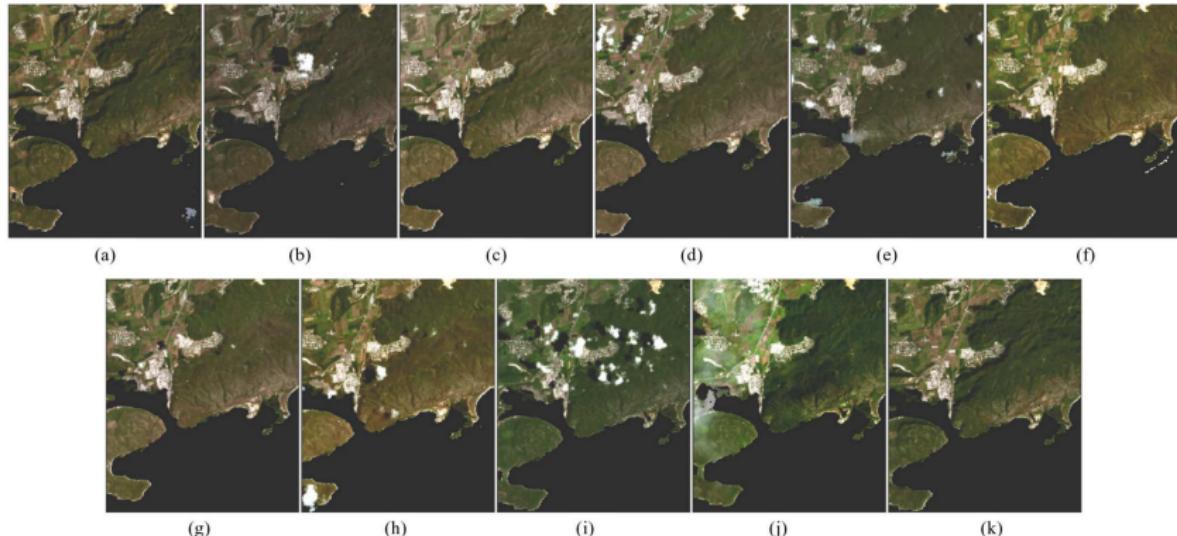


Fig. 4. True color composites of Hyperion images for the Guánica Dry Forest collected in 2008: (a) Feb. 4 (b) Mar. 21 (c) Mar. 26 (d) Apr. 18 (e) May 6 (f) June 21 (g) July 14 (h) Aug. 6 (i) Sep. 16 (j) Oct. 9 (k) Dec. 12. These RGB color composites were produced by using the bands 29 (640.5 nm), 20 (548.9 nm), and 14 (487.87 nm).

Specificity of multi/hyperspectral data to recover dynamics

- Absence of benchmarking data for the community
 - In most cases, studies are carried out frame by frame
 - Materials can appear and disappear over time
 - Endmembers but also the abundances' supports and values change dynamically
 - Endmembers are 4th order tensors: $\mathcal{S} \in \mathbb{R}^{L \times P \times N \times T}$, abundances third order tensors, and data is matricial
- Learn the dynamics of the endmembers from large volumes of data to be able to filter, interpolate, and forecast in the future... using Deep Learning techniques or complex statistical models.

A few references

- Bioucas-Dias, J. M., Plaza, A., Dobigeon, N., Parente, M., Du, Q., Gader, P., & Chanussot, J. (2012). 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.
- Robin, A., Cawse-Nicholson, K., Mahmood, A., & Sears, M. (2015). Estimation of the intrinsic dimension of hyperspectral images: Comparison of current methods. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(6), 2854-2861.
- Nascimento, J. M., & Bioucas-Dias, J. M. (2005). Vertex component analysis: A fast algorithm to unmix hyperspectral data. *IEEE transactions on Geoscience and Remote Sensing*, 43(4), 898-910.
- Araújo, M. C. U., Saldanha, T. C. B., Galvao, R. K. H., Yoneyama, T., Chame, H. C., & Visani, V. (2001). The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. *Chemometrics and Intelligent Laboratory Systems*, 57(2), 65-73.
- Heinz, D., Chang, C. I., & Althouse, M. L. (1999, June). Fully constrained least-squares based linear unmixing. In *IEEE 1999 International Geoscience and Remote Sensing Symposium. IGARSS'99* (Cat. No. 99CH36293) (Vol. 2, pp. 1401-1403). IEEE.
- Iordache, M. D., Bioucas-Dias, J. M., & Plaza, A. (2011). Sparse unmixing of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 49(6), 2014-2039.
- Iordache, M. D., Bioucas-Dias, J. M., & Plaza, A. (2012). Total variation spatial regularization for sparse hyperspectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 50(11), 4484-4502.
- Heylen, R., Parente, M., & Gader, P. (2014). A review of nonlinear hyperspectral unmixing methods. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(6), 1844-1868.

A few references (cont'd)

- Borsoi, R. A., Imbiriba, T., Bermudez, J. C. M., Richard, C., Chanussot, J., Drumetz, L., ... & Jutten, C. (2020). Spectral Variability in Hyperspectral Data Unmixing: A Comprehensive Review. arXiv preprint arXiv:2001.07307.
- Veganzones, M. A., Drumetz, L., Tochon, G., Dalla Mura, M., Plaza, A., Bioucas-Dias, J., & Chanussot, J. (2014, June). A new extended linear mixing model to address spectral variability. In 2014 6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS) (pp. 1-4). IEEE.
- Drumetz, L., Chanussot, J., & Jutten, C. (2019). Spectral unmixing: A derivation of the extended linear mixing model from the hapke model. IEEE Geoscience and Remote Sensing Letters.
- Drumetz, L., Veganzones, M. A., Henrot, S., Phlypo, R., Chanussot, J., & Jutten, C. (2016). Blind hyperspectral unmixing using an extended linear mixing model to address spectral variability. IEEE Transactions on Image Processing, 25(8), 3890-3905.
- Somers, B., Zortea, M., Plaza, A., & Asner, G. P. (2012). Automated extraction of image-based endmember bundles for improved spectral unmixing. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 5(2), 396-408.
- Drumetz, L., Meyer, T. R., Chanussot, J., Bertozzi, A. L., & Jutten, C. (2019). Hyperspectral image unmixing with endmember bundles and group sparsity inducing mixed norms. IEEE Transactions on Image Processing, 28(7), 3435-3450.
- Gillis, N. (2019, December). Successive Projection Algorithm Robust to Outliers. In 2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP) (pp. 331-335). IEEE.

Projections and subspaces

If the LMM is valid and there is no noise: then all the data lie exactly in a d -dimensional linear subspace (or equivalently, in a $d - 1$ -dimensional affine subspace.). With noise, this is only approximately true.

→ We need to remove the noise in the data. Several algorithms try to do this more or less effectively.

If we project the noisy data \mathbf{X} on a P -dimensional subspace such that the reconstruction error with the denoised data \mathbf{Y} is minimal, then we can decompose the reconstruction error into two terms:

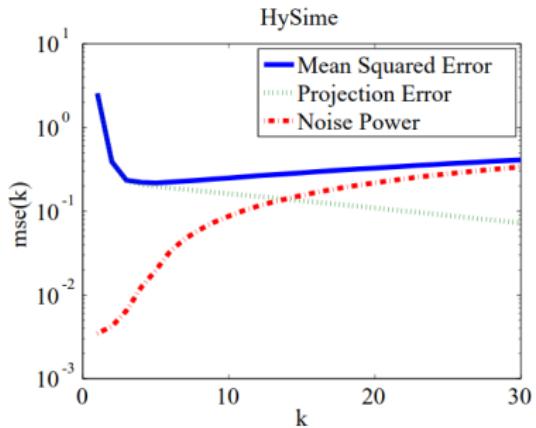
$$\|\mathbf{Y} - \hat{\mathbf{Y}}_P\|^2 = \text{projection error}(P) + \text{noise power}(P)$$

The projection error will decrease when P increases, while the noise power will increase when P increases

An example of ID estimation algorithm: HySIME

Hyperspectral Subspace Estimation by Minimum Error [Bioucas 2008]

If we project the noisy data onto a well chosen P -dimensional subspace, i.e. such that the reconstruction error $\|\mathbf{Y} - \hat{\mathbf{Y}}_P\|^2$ between \mathbf{Y}_P (a projected version of the noisy data \mathbf{X}) and the denoised data \mathbf{Y} is the smallest possible for that value of P , then this error should be minimal for $P = d$.

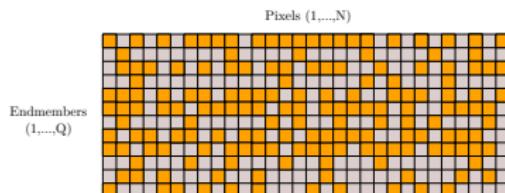


The reconstruction error can be decomposed into two terms: one term related to the projection error (residual noise included) and one to the noise power

For smaller values than P , the error will increase greatly because some points will not be reconstructed well, and for greater values, it will increase because of the noise power (projecting it to lower dimensions removes some of its energy).

Sparsity in spectral unmixing

A vector has a sparse decomposition in a dictionary when many atoms are inactive → a few endmembers active in each pixel [Iordache2011].



Sparse Unmixing with bundles

We have a large matrix of spectral bundles. To estimate the abundances, we can solve:

$$\arg \min_{\mathbf{A} \geq 0} \|\mathbf{X} - \mathbf{BA}\|_F^2 + \lambda \|\mathbf{A}\|_1,$$

with $\|\mathbf{A}\|_1 = \sum_{p=1}^Q \sum_{k=1}^N |a_{pk}|$, the (matrix) \mathcal{L}_1 norm of \mathbf{A} .

Two issues with this approach (called LASSO):

- \mathcal{L}_1 norm penalty → not compatible with the ASC: $\|\mathbf{A}\|_1 = N$ is constant.
- The structure of the dictionary, separated into P groups, is ignored.

Social norms

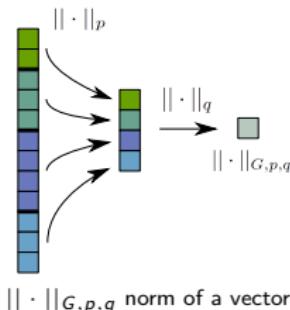
The bundle dictionary has a strong group structure, ignored by FCLSU
→ sparsity inducing *mixed norms* [kowalski2009sparse]:

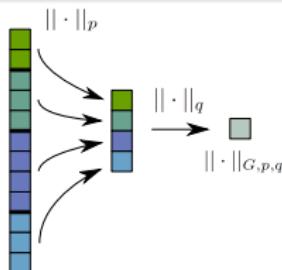
Social sparsity in abundance estimation

$$\arg \min_{\mathbf{A} \in \Delta_P} \frac{1}{2} \|\mathbf{X} - \mathbf{BA}\|_F^2 + \lambda \|\mathbf{A}\|_{G,p,q}$$

with

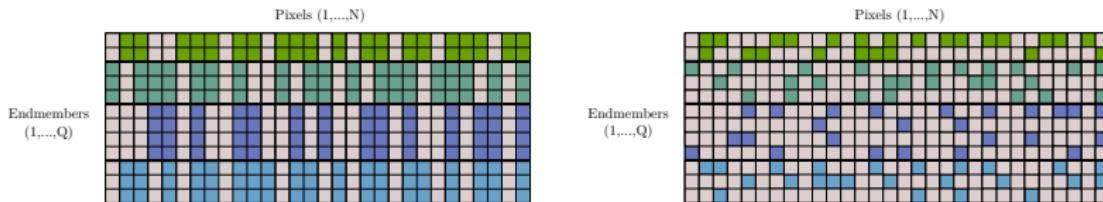
$$\|\mathbf{a}\|_{G,p,q} = \left(\sum_{i=1}^P \left(\sum_{j=1}^{m_i} |a_{G_i,j}|^p \right)^{\frac{q}{p}} \right)^{\frac{1}{q}} = \left(\sum_{i=1}^P \|\mathbf{a}_{G_i}\|_p^q \right)^{\frac{1}{q}}$$
$$\|\mathbf{A}\|_{G,p,q} = \sum_{k=1}^N \|\mathbf{a}_k\|_{G,p,q}$$





- Group sparsity $\rightarrow \mathcal{L}_{G,2,1}$ norm: only a few *materials* (not atoms) are active in each pixel
- Elitist sparsity $\rightarrow \mathcal{L}_{G,1,2}$ norm: favors within-group sparsity
- Fractional group sparsity $\rightarrow \mathcal{L}_{G,1,q}$ quasinorm, with $q < 1$: favors group and within-group sparsity in a single compact penalty, to the cost of nonconvexity

All these norms are compatible with the ASC.



Effect of the group (left) and fractional (right) penalties on the abundance matrix.

Results on real data

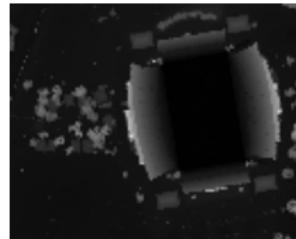
Campus of the University of Houston, Texas, USA. 144 bands in the visible and near IR, and comes with a LiDAR dataset with the same spatial resolution (2.5 m).



(a)



(b)



(c)

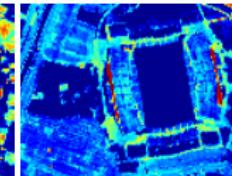
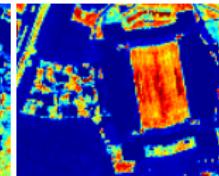
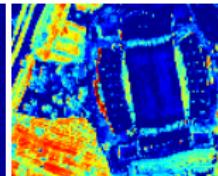
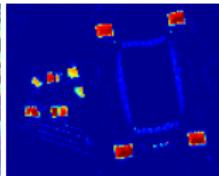
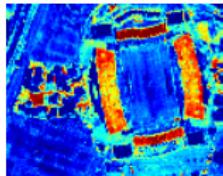
A RGB representation of the Houston hyperspectral dataset (a). High spatial resolution color image acquired over the same area at a different time (b). Associated LiDAR data (c), where black corresponds to 9.6m and white corresponds to 46.2m.

We extract 50 signatures by running VCA 10 times with $P = 5$. Then we cluster them into 5 bundles.

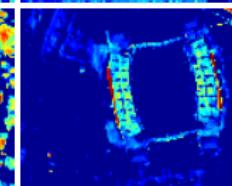
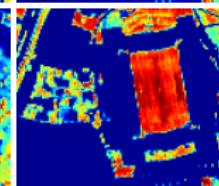
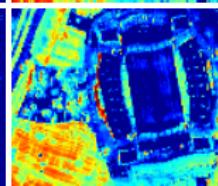
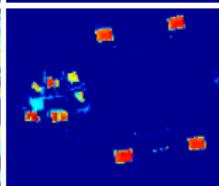
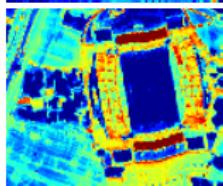
Results on real data

Abundances for the 5 identified endmembers.

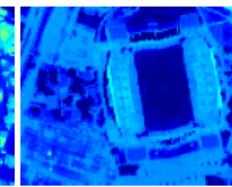
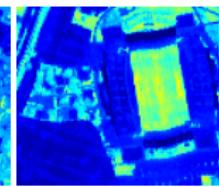
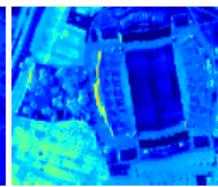
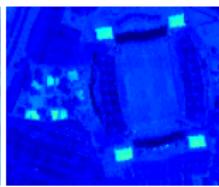
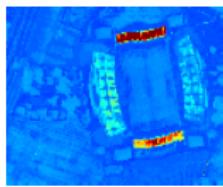
FCLSU



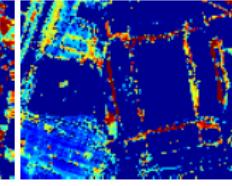
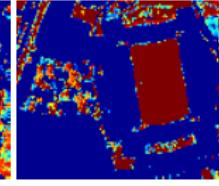
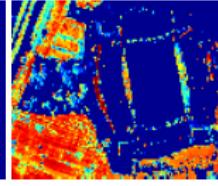
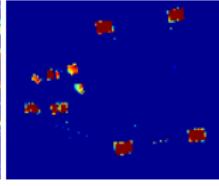
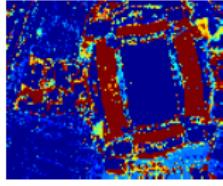
Group



Elitist



Fractional



Concrete

Red Roofs

Vegetation

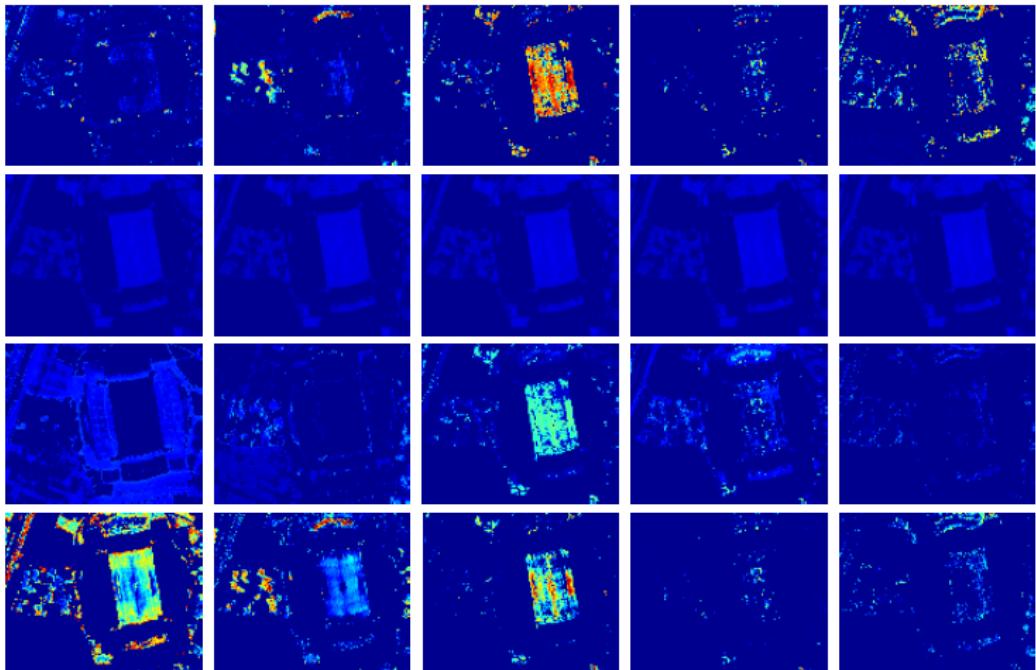
Asphalt

Colored Structures

Results on real data

Abundances for a few atoms corresponding to vegetation

FCLSU Group Elitist Fractional



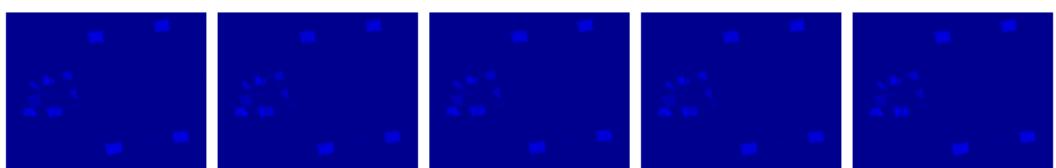
Results on real data

Abundances for a few atoms corresponding to red roofs

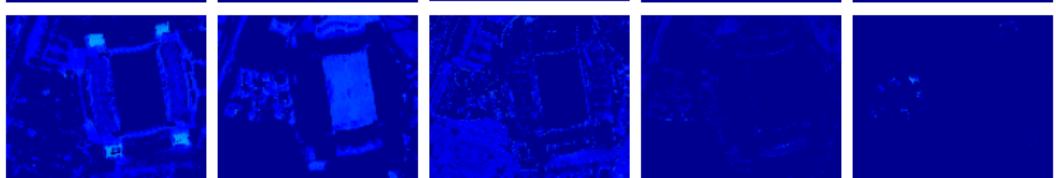
FCLSU



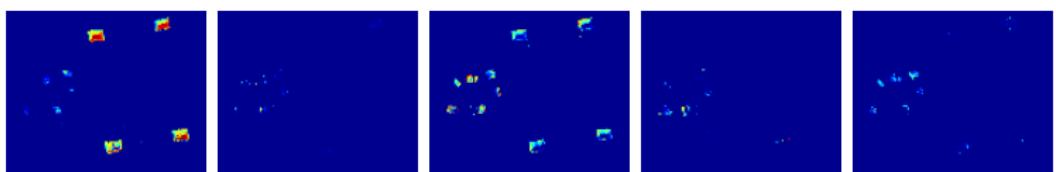
Group



Elitist



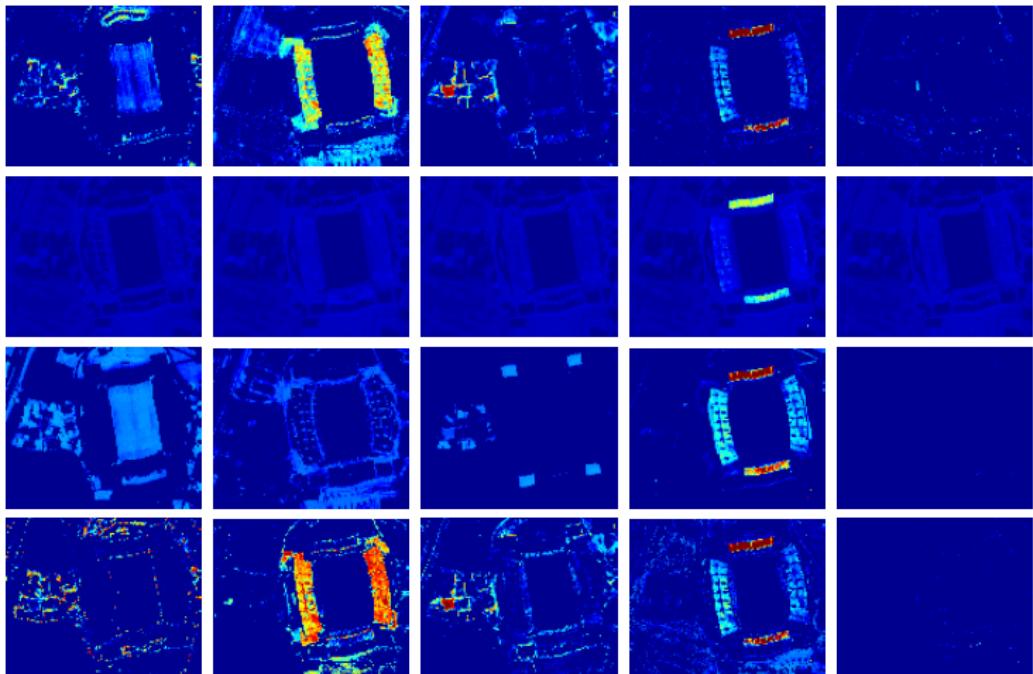
Fractional



Results on real data

Abundances for a few atoms corresponding to concrete

FCLSU Group Elitist Fractional



Modeling intrinsic variability and illumination

- Scaling factors model illumination induced variability
- Modeling of intrinsic variability is done in a statistical way. We put a Gaussian prior on \mathbf{S}_n :

$$\mathbf{S}_n \sim \mathcal{N}(\mathbf{S}_0\boldsymbol{\psi}_n, \sigma_S^2 \mathbf{I}_P)$$

Together with non informative priors on the other variables, and Gaussian noise, the MAP estimator for our Bayesian model is

$$\arg \min_{\mathbf{A} \in \Delta_P, \mathbf{S} \geq 0, \Psi \geq 0} \frac{1}{2} \sum_{n=1}^N \left(\|\mathbf{x}_n - \mathbf{S}_n \mathbf{a}_n\|_2^2 + \lambda_S \|\mathbf{S}_n - \mathbf{S}_0 \boldsymbol{\psi}_n\|_F^2 \right)$$

Globally nonconvex, but easy and relatively fast to optimize alternatively.
Other regularizers can be added, such as 2D Total variation on the abundances and scaling factors