Sparse, Smooth and Differentiable Hyperspectral Unmixing using Dispersion Model

Aymeric CÔME　　　　　　　　　　　　　　　　　　　　aymeric.come.etu@univ-lille.fr

Master Data Science
University of Lille, Centrale Lille, IMT Lille-Douai
Villeneuve d'Ascq, France

Pierre-Antoine THOUVENIN　　　　　　pierre-antoine.thouvenin@centralelille.fr

Assistant professor and member of SigMA team
Université de Lille, CNRS, Centrale Lille, CRIStAL
Villeneuve d'Ascq, France

Abstract

Hyperspectral unmixing consists in infering the proportion of presence (*abundance*) and *signatures* of pure materials composing an hyperspectral image, observed over hundreds of wavelengths. The spectra observed in each pixel indeed result from a mixture of several signatures, referrend to as *endmembers*, which characterize the materials present in the observed scene. The problem is challenging in that it is not only non-linear, but can be further affected by multiple factors which can locally affect the shape of the endmembers, the so-called spectral variability. To tackle this problem, Janiczek and al. (2020) introduced a physics-based, differentiable model to realistically capture endmembers. However, the approach investigated in Janiczek and al. (2020) exploits none of the usual prior information commonly used in the literature (e.g. spatial smoothness, sparse abundances). In this project, we propose to extend this approach by considering additional regularization terms on both the abundances and the variability, and investigate their impact on unmixing quality.

Keywords: Hyperspectral Unmixing, Linear Unmixing, Differentiable Programming, Endmember Spatial and Spectral Variability

1. Introduction

In many real-world applications such as astronomy (Kouyama et al., 2016), remote sensing (Lu and Fei, 2014) or mineral detection (Bandfield, 2002), hyperspectral images (which consist of samples of electromagnetic spectra at various wavelengths) can put into evidence some key characteristics of the location monitored. In particular, it is commonly assumed that each physical material present in the scene, called endmember (EM), have its own signature spectrum, which contributes to the HI, depending on its proportion of presence in each pixel. Hence, recovering these characteristic signatures from HI leads to a precise description and detection of materials in a scene.

However, despite the high spectral resolution of hyperspectral cameras, technical limits must be taken into account; for instance a limited spatial resolution or a varying luminosity can be expected. This implies that the pixels of a HI then contain mixtures of multiple EMs in varying abundances, and noise is injected into the measurement. That means that the signature spectra of EMs can vary spectrally and spatially. While many unmixing algorithms have already been proposed (Borsoi et al., 2020), the physics behind is often too complex to be fully modelled or the physical quantities at hand are unknown.

A first difficulty to handle is the spectral variability (SV): depending on the scene observed, an EM might have a signature different from a laboratory measurement. This kind of variability can even occurs locally, between pixels of a same HI, which actually question the hypothesis of one reference signature characteristic of one EM, rather than a class of signatures. Multiple factors can

contribute to this variability, and we distinguish two kinds: intrisic and external. Intrisic variability is due to the natural differences of the EM samples across the HI (e.g. grass is greener on the other side because the neighbor water it) while external variability refers to the environment influence, like light and shade or the quality of the hyperspectral camera used.

In the end, hyperspectral unmixing is very challenging task in that there are many unknowns and sources of variability, hence it requires prior information to regularize the problem and improve the inference process.

Linear Mixing Models (Keshava and Mustard, 2002) are a classic approach to model mixture spectra as a linear combination of the reference signatures and the abundances, plus some noise. To handle spectral variability, different signatures across the pixels of the HI can be considered. Moreover, multiplicative (Veganzones et al., 2014) or additive (Johnson and Jones, 2013) variability of the spectra have been studied in the literature. An other approach is to perform a physical analysis of the context, to try to provide an explicit model of spectral variation (Janiczek and al., 2020; Griffin and Burke, 2003). And lately, Deep approaches have been tested (Borsoi et al., 2019).

Contributions: In this report, we propose to take advantage of the physics-based dispersion model introduced in Janiczek and al. (2020) in order to perform spectral unmixing. More precisely, we incorporate important prior information about the elements of the problem (such as sparsity of the abundances or spatial smoothness for the EMs) that should improve the regularisation of the problem, and then we present a new formulation of the unmixing problem. Additionally and in the same fashion as in Janiczek and al. (2020), we keep the problem differentiable to exploit on-the-shelf algorithms, with automatic differentiation because of the complexity of the dispersion model. Finally, we run some experimentations on real-world and synthetic data, to validate the approach.

The report is organised as follows. A brief summary of different approaches to endmember variability modelling is given in Section 2. Section 3 is devoted to the formulation of the problem, introducing the a priori information considered to solve the problem. The algorithm used to address the resulting optimisation problem is briefly described in Section 4. Experiments reported in Section 5 propose a first validation of the approach on synthetic data. Conclusion and perspectives are finally reported in Section 6.

2. Variability models

In this section is presented a quick review of the literature on variability models for HI. In these applications, it is commonly assumed that the target mixture spectrum is a function of the signature spectra of the EMs present in the scene, and their abundances. Moreover, and we make this hypothesis in our contribution as well, the number of EMs is supposed to be known a priori - there exist many state-of-the-art algorithms to try to infer this information (Bioucas-Dias and Nascimento, 2008). Now the following models are designed specifically to consider SV.

2.1 Generic explicit models

The Linear Mixture Model (LMM) is defined as follows:

$$y_n = M_0 a_n + z_n$$

for a pixel n of an HI. M_0 is a $S \times P$ matrix whose columns correspond to the signatures of the P EMs, over the L spectral channels, and a_n is a P vector, giving for each EM its abundance in the pixel (so that the elements of a_n are non-negative and sum to 1). z_n is additive noise. The main culprit of this model that has been demonstrated is the proportion indeterminancy (Borsoi et al., 2020).

To consider spatial variability of the EMs, often a varying matrix of signatures is studied:

$$y_n = M_n a_n + z_n \tag{1}$$

Now the signatures of the EMs depend on the pixel, hence on the spatial location. However, the difficulty is to express M_n. Multiple approaches have been proposed to do so. Some are based on spectral libraries, consisting on trying to match reference signatures to the observations. Others create such libraries directly from the image, using pure pixels (only one EM present) for instance.

2.2 Deep models

While previous methods have shown good performances, they often depend on expensive and not always available previous informations about the EMs. Since the problem seems too complex to be described by simple models, the scientific community has designed Deep architextures, that could learn in a black-box manner the rules underlying spectral variability.

In Borsoi et al. (2019), the authors train a Variational Auto-Encoder that learns a latent representation of EM signatures, which is then used in a classic LMM optimisation framework. In particular, this works is based on the smoothness of the latent space; we will use a similar hypothesis later on as well. Yet, the method once again requires a pure pixels or libraries for training.

Another Deep approach is deployed in Hong et al. (2021): one neural network is used to learn a relevant representation of EMs signatures, and at the same time guide an unmixing network through weight sharing. Once again, this approach makes use of pure pixels in the HI. More interestingly, the networks use the convolution operation: the spatial correlation between pixels is used to perform unmixing.

Finally, the work presented in Feng et al. (2018) leverages matrix factorization with Deep networks. The observed HI is supposed to be decomposable as a successive multiplication of matrices, the last one being the abundance map, each previous matrix represented by a layer in the neural network. A major point of this approach is the addition of a sparsity penalty in the loss function: the factorization is supported by a sparsity regularization on the abundances.

Thus, Deep approaches to generate variable signatures encounters some succes; however the training can be difficult, and often require prior information to converge to a good solution.

2.3 Physics-based models

So far have been presented generic or Deep models for the variable signatures M_n (1). An interesting observation is the use of prior information about the structure and behavior of the elements of the problem, to regularise and help the algorithm find a good solution.

A last strategy to model the SV are physics-based models: an explicit description of the signatures derived from physical laws. This can be compared to the VAE, in that it is a mapping of signatures to a latent space of parameters controlling their generation, based on the current state.

Janiczek and al. (2020) presents such a model, the dispersion model, derived from the laws behind molecular vibration. Then the optimisation problem consist in infering the parameter values for each EM, and then generate the corresponding signature and unmix the target HI in accordance with the LMM model (1). One important property of the dispersion model is that it is differentiable, which allows for an easier optimisation algorithm.

In the dispersion model, every EM is characterized by a set of parameters $\rho, \omega_0, \gamma, \epsilon_r \in \mathbb{R}^K$ (band strength, resonant frequency, frictional force, relative dielectric permeability), where K is an hyper-parameter corresponding to the number of equations controlling the signature, and for every such equation (k) are involved $\rho_k, \omega_{0,k}, \gamma_k, \epsilon_{r,k}$ in a very complex - but differentiable - way. Then, the spectral emission of the Em can be computed for every wavelength ω: $\widehat{\epsilon}(\theta; \omega)$, where θ is the matrix of parameters. We denote by $E(\theta_1, \ldots, \theta_P) \in \mathbb{R}^{S \times P}$ the matrix whose columns contain the spectral signature over the S wavelengths considered, for all P EMs.

Following the LMM (1), the authors formulates the unmixing problem as:

$$\underset{\theta_{min}<\theta<\theta_{max},\ a}{\arg\min} \|b - E(\theta_1,\ldots,\theta_P)a\|_2^2 + \lambda\|a\|_p \tag{2}$$

where b is the spectra at one pixel of the HI, a is the vector of abundances (non-negative, sum to one), and a regularisation term to promote abundance sparsity with a pseudo-norm $p < 1$.

The algorithm proposed, called *Analysis-by-Synthesis*, solves (2) for every pixel. To do so, an alternating-like minimisation is performed: at each iteration the objective function is minimized on a only, then an gradient descent update of the parameters θ is done, using the chain rule to backpropagate the gradient up to the parameters (thanks to the differentiability of the dispersion model).

To avoid falling into a bad local minimum, the authors support that a good initialisation is provided by fitting each EM separately to b. This abides by the sparsity assumption made on the abundances.

In the end, the dispersion model seems to provide a good model for the EMs, and has the upside of being differentiable. However, no spatial correlation is used, and the alternating minimisation isn't justified theoretically. Hence, we come up with an other optimisation framework for performing hyperspectral unmixing with the dispersion model: the next section present it.

3. Problem statement

Based on the elements and observations we have seen so far, we come up with a formulation for the problems we are willing to tackle.

3.1 Reconstruction model

The first part of a spectral unmixing strategy is to try to reconstruct the target HI from a specified observation model, which describes the relationship between the observations and the parameters to be inferred.

In this work, we propose to make use of the dispersion model presented in Section 2.3. So, the reconstructed spectra are generated by the dispersion model from the physics parameters. More precisely, we suppose that each EM is present in a state characterised by those parameter values in every pixel; and then each EM is associated to a signature spectrum derived from the dispersion model. The SV is modeled here by the variation of the parameter values through the HI, which in turn gives different signature spectra over the pixels. Regarding the unmixing, we decided to use Linear Unmixing once again, so that our model would still compare to the one in the original paper. We then introduce abundances for each EM, accross the HI.

3.1.1 Problem statement

We consider a HI $B \in \mathbb{R}^{N \times M \times S}$ consisting of $N \times M = L$ pixels, each pixel giving a spectrum evalutated at wavelenghts $w_1,\ldots,w_S$. We assume that there are P EMs in the HI, characterised by sets of vectors of parameters $\theta_1,\ldots,\theta_P$.

Since the dispersion model from Janiczek and al. (2020) is adopted here, the parameters $\theta_p \in \mathbb{R}^{N \times M \times K \times 4}$ contain the physical parameters introduced in the previous section REF(SECTION), with different values for different pixels (for sake of lisibility, we define $T = K \times 4$ the number of parameters per EM).

Then, the corresponding abundance map $A \in [0,1]^{N \times M \times P}$ is such that $a_{i,j,p}$ is the abundance of EM p in pixel (i,j), and $\|A_{i,j,1:P}\|_1 = 1$.

Finally, $E(\theta_1,\ldots,\theta_P) = E(\Theta) \in \mathbb{R}^{N \times M \times P \times S}$ holds the signature spectrum of each EM at each pixel, derived from the dispersion model: $E(\Theta)_{i,j,p,s} = \epsilon(\theta_{i,j,p}; w_s)$.

3.1.2 MODEL OF THE HYPERSPECTRAL IMAGE

In accordance to a commonly used approach in the literature (cf (1), we model the HI as follows:

$$B = E(\Theta)A + Z$$

where Z is a white noise tensor and $E(\Theta)A \in \mathbb{R}^{N \times M \times S}$ is the sum of the signature spectra weighted by abundances, for every pixel and spectral channel.

So far it is just a reformulation of the problem from Janiczek and al. (2020), however we wish to add prior information on the elements of the problem, as discussed in the following sections. This will be expressed by additional terms in the objective function. In the end, the optimisation problem states as:

$$\underset{\Theta \in \Delta, \ A \in \Upsilon}{\arg\min} \ H(\Theta, A) + f(\Theta) + g(A) \tag{3}$$

with $H(\Theta, A) = \frac{1}{2L}\|B - E(\Theta)A\|^2$, and Υ and Δ are constraint spaces presented afterwards. The norm used on tensors, defined by $\|Q\| = \sqrt{\sum_{1 \leq i \leq N, 1 \leq j \leq M} \|Q_{i,j,1:S}\|_2^2}$, is the Froebenius norm applied on the square norm over spectrum for every pixel. $f(\Theta)$ and $g(A) = g_c(A) + g_s(A)$ are regularisation terms, enforcing priors on the parameters.

This formulation corresponds to finding the maximum likelihood in presence of white noise, with respect to priors.

3.2 Priors on the endmembers

In this section are explained the priors introduced regarding the EMs present in the HI.

3.2.1 PHYSICS-BASED CONSTRAINTS

The main prior regarding EMs is that they follow the physics-based dispersion model, controlled by the parameters Θ. These parameters actually represent physical quantities (frictional force, band strength...): in that regard, and alike what was done in Janiczek and al. (2020), their values should stay in a plausible range, determined by a physical analysis. From now on, we will design by Δ the space in which the parameters Θ live:

$$\Delta = [\rho_{min}, \rho_{max}] \times [\omega_{0,min}, \omega_{0,max}] \times [\gamma_{min}, \gamma_{max}] \times [\epsilon_{r,min}, \epsilon_{r,max}]$$

3.2.2 SMOOTHNESS

As advertised earlier, our approach to model the SV is to let the physical parameters Θ of the EMs vary across the pixels of B. Yet, to keep a spatial consistency we want to ensure a smooth variation, to represent the idea that, as the environmental conditions are roughly continuous with respect to space dimensions, so are material states and signature spectra. For instance, the quantity of water received by a lawn is close to the one received by the adjacent grass. Obviously, this hypothesis is not always appropriate, as your neighbor might water his lawn way more than you; yet it should still be a relevant hypothesis in most real-world applications.

To do so, we consider the *discrete gradient* operator, as in Condat (2014). It consists in the horizontal and vertical differences between neighbour pixels, so that for an $N \times M$ image Y we have:

$$D(Y) = (u_h(Y), \ u_v(Y)),$$
$$u_h(Y)[i,j] = Y_{i+1,j} - Y_{i,j} \text{ if } i+1 \leq N, \ 0 \text{ otherwise}$$
$$u_v(Y)[i,j] = Y_{i,j+1} - Y_{i,j} \text{ if } j+1 \leq M, \ 0 \text{ otherwise}$$

This operator is a measure of the spatial variation of the pixel values, hence in order to promote smoothness these differences should remain small. We quantify it with:

$$\|D(Y)\|_{1,2} = \sum_{i=1}^{N} \sum_{j=1}^{M} \sqrt{u_h(Y)[i,j]^2 + u_v(Y)[i,j]^2}$$

Finally, we enforce smoothness by minimizing the sum of the squared variations, so that the corresponding term is (for some penalty factor $\gamma > 0$):

$$f(\Theta) = \gamma \sum_{t=1}^{T} \sum_{p=1}^{P} \|D(\Theta_{:,:,p,t})\|_{1,2}^2 \tag{4}$$

Note that one might prefer to ensure smoothness of the spectra rather than the parameters; we would then have:

$$f(\Theta) = \gamma \sum_{t=1}^{T} \sum_{p=1}^{P} \|D(E(\Theta_{:,:,p,t}))\|_{1,2}^2$$

3.3 Priors on the abundances

We introduced the material abundances of each EM in every pixel, and, in accordance with the Linear Unmixing Model, we assume that an EM contributes to the mixture signal proportionally to its abundance. To stay in a relevant case, we need to ensure that some properties are respected for the abundances.

3.3.1 Constraints

Based on their physical meaning, abundances of EMs must obviously be positive and sum to 1 in each pixel. These constraints are usually formulated using an indicator operator:

$$\iota_\Upsilon(A) = \begin{cases} 0 \text{ if } A \in \Upsilon \\ +\infty \text{ otherwise} \end{cases}$$

where

$$\Upsilon = \{Q \in \mathbb{R}_+^{L \times P} \mid \|Q_{i,j,1:P}\|_1 = 1 \ \forall 1 \leq i \leq N, 1 \leq j \leq M\}$$

is the simplex where abundances live.

However, constraints induce non-differentiability. To circumvent this issue, we propose to relax the constraint term into a smooth penalty term, so that we strongly encourage abundances to stay close the Υ while being differentiable.

The idea is to strongly promote values close to Υ for A, thanks to an approximation to the indicator function. Such approach has already been presented and used in the literature, in the form of *log-barrier method*. In particular, we consider the *log-barrier extension* introduced in Kervadec et al. (2020). This operator is convex, continuous and twice-differentiable, and is defined as follows, for some hyper-parameter $t > 0$:

$$\forall z \in \mathbb{R}, \ \tilde{\Psi}_t(z) = \begin{cases} -\frac{1}{t} \log(-z) & \text{if } z \leq -\frac{1}{t^2} \\ tz - \frac{1}{t} \log(\frac{1}{t^2}) + \frac{1}{t} & \text{otherwise} \end{cases}$$

The higher t, the better the quality of the approximation; however one should be cautious not to overwhelm the other terms in the problems - this is discussed the next section.

To translate the sum-to-one constraint on the abundances, we rewrite it as a sum-greater-than-one and sum-lesser-than-one constraint, in addition to the positivity constraint. Finally, we come up with the following penalty term, $\zeta > 0$ and $\nu > 0$ being hyper-parameters:

$$g_c(A) = \zeta \cdot \left[\sum_{i=1}^{N} \sum_{j=1}^{M} \sum_{p=1}^{P} \tilde{\Psi}_\nu(-a_{i,j,p}) + \sum_{i=1}^{N} \sum_{j=1}^{M} \left[\tilde{\Psi}_\nu(\sum_{p=1}^{P} a_{i,j,p} - 1) + \tilde{\Psi}_\nu(1 - \sum_{p=1}^{P} a_{i,j,p}) \right] \right] \quad (5)$$

3.3.2 SPARSITY

Regarding the abundances, we whish to enforce sparsity of abundances in every pixel; we would rather have a few EMs in significant abundances than a lot of EMs in small quantities. This prior is often relevant in real-wolrd applications.

Common choices to do so are the ℓ_0 count measure, the ℓ_1 norm, or the ℓ_p ($p < 1$) quasi-norm (Natarajan, 1995). However, these terms are hard to handle during optimisation, as they introduce non-convexity and discontinuity. Hence, to keep once again our problem differentiable, we ressort to the SPOQ regularisation, introduced in Cherni et al. (2020). It gives a smooth approximation of ℓ_0.

For some hyper-parameters $p \in\ ]0, 2[$, $q \in\ [2, +\infty[$, and $(\alpha, \eta) \in\ ]0, +\infty[^2$, the SPOQ regularisation term is defined as follow, for any vector $x \in \mathbb{R}^N$:

$$SPOQ(x) = \log \left(\frac{(\ell_{p,\alpha}^p(x) + \beta^p)^{1/p}}{\ell_{q,\eta}(x)} \right)$$

where

$$\ell_{p,\alpha}(x) = \left(\sum_{n=1}^{N} (x_n^2 + \alpha^2)^{p/2} - \alpha^p \right)^p$$

$$\ell_{q,\eta}(x) = \left(\eta^q + \sum_{n=1}^{N} |x_n|^q \right)^{1/q}$$

The choice of the hyper-parameters is discussed in the next section.
We then apply the SPOQ pseudo-norm on the vector of abundances in every pixel of the HI:

$$SPOQ(A) = (SPOQ(a_{i,j,1:P}))_{i,j} \in \mathbb{R}_+^{N \times M}$$

In the end, we define the sparsity term on abundances as (with $\delta > 0$ a penalty term):

$$g_s(A) = \frac{\delta}{N \times M} \|SPOQ(A)\|_1 \quad (6)$$

It simply is a sum of the SPOQ norms over the pixels.

3.3.3 SMOOTHNESS

Finally, one could also require the abundances to vary smoothly across space. Hence, in a similar way to (2), we could add a $\beta \mathcal{L}(A)$ term to the regularisation on abundances. We will not consider it here though, for simplicity and because we value the sparsity prior more.

4. Inference algorithm

The optimisation problem presented in (3) is composed of one differentiable, non-convex block $H(\Theta, A) + f(\Theta)$ and one non-convex, non-continuous block $\delta\|A\|_{1,0} + \iota_\Upsilon(A)$, and deal with two vectors of parameters Θ and A.

The first alternative to non-differentiability is to use Moreau's proximal operator: a first-order approximation of a gradient descent step. We would then operate an alternating minimization on Θ (differentiable) and A (non-differentiable). However, algorithms involving this operator typically rely on the knowledge of a (tight) Lipschitz constant of the differentiable part to derive a fitting learning rate, and due to the complex definition of the dispersion model we do not have access to such a constant. Moreover, one can't easily derive the close form of the proximal operator of an addition of two non-differentiable, non-convex terms.

Related families of optimisation algorithms would be subgradient or bundle algorithms, unfortunately they do not seem to apply to our problem, mainly because of the non-convexity and the fact that in the constraint ι_Υ on A, Υ is a simplex rather than a convex space.

Hence, instead of testing such approaches without guarantees, we chose a Lagrangian-based method from Kervadec et al. (2020), using a *log-barrier extension* to replace the constraint term for A. Then, thanks to Cherni et al. (2020) we approximate the sparsity term on A with a smooth approximation, so that as written in (3) we put ourselves in a smooth context.

Finally, thanks to these rewritings we are able to do a Stochastic Gradient Descent-like optimisation with Adam-which has already shown to be an efficient approach for such problems. A discussion can be made on other Adam-like algorithms in a non-convex problem (cf Chen et al. (2019)), yet we will stick to the classical Adam for simplicity reasons. In particular, we shall use the Autodiff feature from PyTorch rather than deriving by hand the awfully complex gradient of the dispersion model.

To wrap-up, the inclusion of our priors raised some major optimisation issues, due to which we have been forced to resort to a sub-optimal approach. Hopefully, the behavior will still be good, thanks to the senseful priors and a good initialization provided by the dispersion model. On the other hand, Janiczek *analysis-by-synthesis* consists in an alternating minimization of the reconstruction error, on the abundances and the physics parameters. The main difference with our approach is that the parameters (and then the signature spectrum) of EMs are assumed to be the same all accross the HI.

5. Experiments

In order to validate our approach, we implemented it using PyTorch [1] with Python 3.7; the details can be found on GitHub [2]. In this section we describe the data used, the evaluation process, and make some discussion and comparison with Janiczek and al. (2020), before presenting and analysing the results.

5.1 Experimental settings

5.1.1 DATASETS

Ideally we would experiment on multiple real-world datasets for which we have access to the ground-truth abundance map and EM parameters and signature spectra in addition to a 2D HI, however such data is very hard to obtain. Additionally, we should compare our performances with the ones of Janiczek and al. (2020). Hence we resort to one dataset used in Janiczek and al. (2020) and one synthetically generated.

Feely et al. dataset (Feely and Christensen, 1999): 90 samples of thermal emission spectra in the infrared for various minerals measured in the lab, over 830 spectral channels. Ground truth was determined via optical petrography, and a labeled endmember library is provided, so that we have the true abundances for the 38 EMs. Also, as the EMs are known minerals, we can already initialize the parameters Θ with physics-based values. Janiczek and al. already tested the dispersion model

---

1. https://pytorch.org/
2. https://github.com/Javisty/spectral_ unmixing

on this data, so we want to compare how our additional priors will affect the results. Unfortunately, as the samples are actually independent pixels, we couldn't use spatial regularisation here.

Synthetic data (Thouvenin et al., 2016): synthetically generated data, with white Gaussian noise (SNR=15dB) and some random SV. The HIs have size 128×64, and are generated from 3 or 6 EMs for 413 spectral channels. We are able to test all our priors on this dataset, yet it is important to note that the spectra have been generated independently between the pixels, so that our spatial smoothness priors doesn't hold here. The code is available on GitHub [3].

5.1.2 EVALUATION CRITERIA:

To assess the performance of the approach, we use some common metrics in the literature: aSAM (average Spectral Angle Mapper) and RMSE (Root Mean Square Error) for the EMs signature spectra, MSE for abundances (per EM and global) and the reconstruction error $H(\Theta, A)$ (RE).

$$aSAM(\widehat{E}) = \frac{1}{NMP} \sum_{p=1}^{P} arccos \left(\frac{\widehat{\epsilon}_p^T \ \bar{\epsilon}_p}{\|\widehat{\epsilon}_p\|_2 \|\bar{\epsilon}_p\|_2} \right)$$

$$MSE(\widehat{A}) = \frac{1}{NMP} \|\widehat{A} - \bar{A}\|_{\mathcal{F}}^2$$

We shall as well compare those results with the ones of Janiczek and al. (2020).

5.1.3 COMPARED METHODS:

This work applys the dispersion model in an other optimisation framework, in which we promote sparsity of abundances and spatial smoothness in a differentiable way. Then, we solve the problem with a classic gradient descent algorithm. Of course, the objective function is strongly not convex, yet as in Janiczek and al. (2020) we rely on a good initialisation of the EM parameters to find a relevant minimum.

On the other hand, in the reference paper they set their dispersion model in a mixed optimisation problem, between Alternating Minimization (Bolte et al., 2014) and Block Coordinate Decent (Marecek et al., 2014). Indeed, in the analysis-by-synthesis at each iteration a full minimization is performed on the abundances by performing Sequential Least Squares Programming (Kraft, 1988), and then one step of gradient descent (Adam (Kingma and Ba, 2017)) is done on the EM parameters.

5.1.4 TRAINING AND IMPLEMENTATION DETAILS:

The final model has a lot of hyper-parameters to fix: number of EMs P, penalty factors $\delta, \zeta...$, $SPOQ$ and log-barrier $(p, \alpha, \nu...)$, Adam parameters.

The values for the SPOQ parameters were taken from Cherni et al. (2020), Section II.B: $p = 1/4$, $q = 2$, $\alpha = 7 \times 10^{-7}$, $\beta = 3 \times 10^{-3}$ and $\eta = 1 \times 10^{-1}$. Similarly, we imitated the approach described in Kervadec et al. (2020) (Appendix, Section E) regarding the log-barrier extension ν hyper-parameter: it is initialised at 5, and is multiplicatively increased by a factor 1.1 between each Adam step. The authors justify this choice by saying that in the early stages of optimisation, the model needs more relaxation with respect to the constraints, and the latters are gradually enforced as the algorithm converges. In practice, during the experiments we capped its value at $1/\nu^2 = 1 \times 10^{-3}$. Indeed, the loss rapidly explodes for bigger values of ν, as the algorithm isn't able anymore to minimize both $H(\Theta, A)$ while respecting the constraints on A with an ever greater precision.

The Adam parameters used were a learning rate of 0.01, and the standard values suggested by PyTorch[4]. The learning rate actually required some fine-tuning, as a too high value would cause the

---

3. https://github.com/pthouvenin/synthetic-hs-variability
4. https://pytorch.org/docs/stable/generated/torch.optim.Adam.html

	aSAM(E)	RMSE(E)	MSE(A)	RE
EM 1	0.337	0.266	0.120	
EM 2	0.634	0.430	0.163	
EM 3	0.547	0.415	0.0122	
EM 4	0.342	0.260	0.0112	
EM 5	0.341	0.411	0.0112	
EM 6	0.213	0.142	0.139	
Average	0.403	0.138	0.0766	0.907

Table 1: Results after 1000 epochs on the synthetic dataset with 6 endmembers.
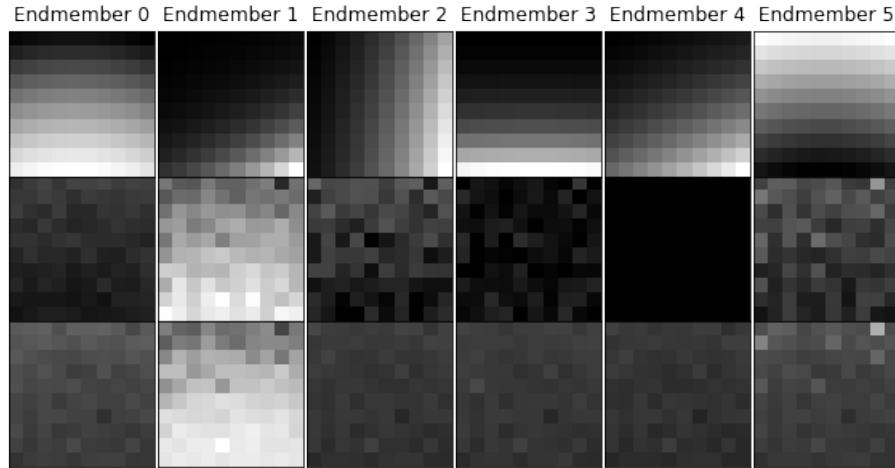
Figure 1: Truth, initial and predicted abundance maps for top, middle and bottom rows respectively. Black means 0, white means 1.

model to oscillate too much between a good reconstruction and an abundance map respecting the priors, while a low value gets the model stuck in a bad local minimum, preventing exploration.

Regarding the penalty factors, the values were chosen between 0.1 and 1. In particular, we made sure that the sparsity term $g_s(A)$ didn't take precedence over the constraining one $g_c(A)$, as the latter is more important.

The number of epochs before convergence were around 70 for the Feely dataset; on the synthetic one it changed a lot for different sets of hyper-parameters.

Finally, as mentioned earlier the initialisation is an important aspect in this optimisation problem. For the Feely dataset, like in Janiczek and al. (2020) we initialised the physical parameters with the associated EM library, which gives reference values for the materials studied. The abundance map was initialised with a homogeneous repartition. Regarding the synthetic dataset, initialisation signatures were given, so that the parameters were set by fitting these signature spectra for each EM. An initial abundance map was provided as well.
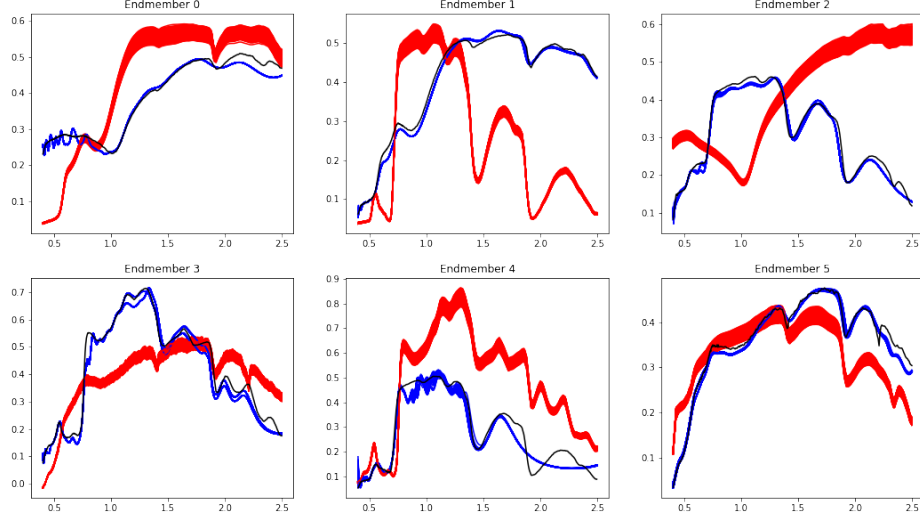
Figure 2: Spatial spectral variability. Signatures for every pixel are represented on the same figure, to visualise the variability. In blue are the predicted signatures, in red the true ones and the black line is the initial signature (same for all the pixels).

	aSAM(E)	RMSE(E)	MSE(A)	RE
EM 1	0.00624	0.0773	0.009357	
EM 2	0.00720	0.0945	0.0184	
EM 3	0.00289	0.0394	0.00187	
EM 4	0.00277	0.0361	5.29	
EM 5	0.00673	0.0835	6.80	
EM 6	0.00456	0.0569	4.31	
Average	0.00506	0.0451	0.00581	0.949

Table 2: Results after 70 epochs on 5 pixels of Feely's dataset, for 6 random endmembers among the 38. Since we don't have access to the ground truth, we compare with the predictions from Janiczek and al. (2020). The average results are computed for all the endmembers.
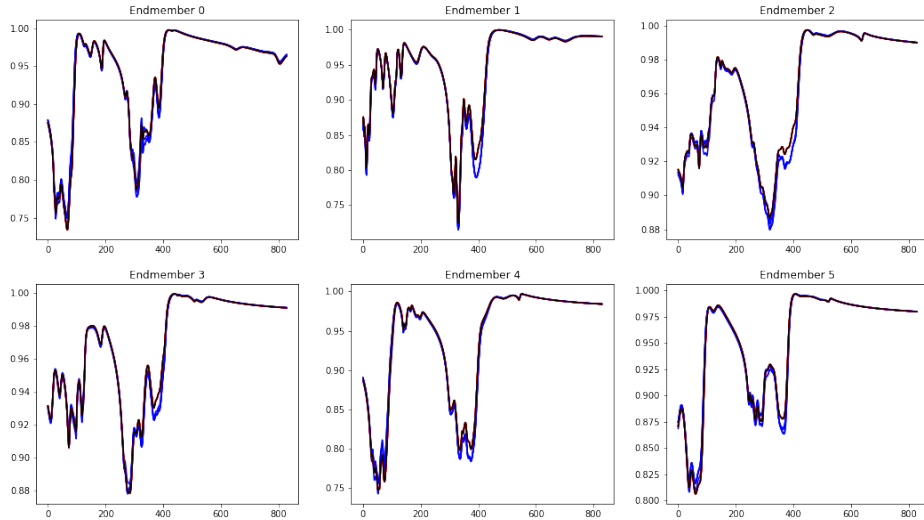
Figure 3: Spatial spectral variability. Signatures for every pixel are represented on the same figure, to visualise the variability. In blue are the predicted signatures, in red the ones for Janiczek and al. (2020) and the black line is the initial signature.

5.2 Results and Discussion

In Table 1, Figures 1 and 2 are presented the results against the ground truth for the synthetic dataset with 6 EMs, cropped to a 10×10 image. In Table 2 and Figure 3 are the results against Janiczek and al. (2020) run on 5 pixels from Feely dataset.

The results on the first experiment aren't good: in Table 1 a quick comparison of the evolution of the abundance map shows that sparsity isn't enforced, as a lot of small abundances were allowed to appear. Regarding the signatures displayed in Figure 2, it is clear that signatures haven't been able to evolve enough from the initialisation towards the target signatures, but at least we obtained a small spatial variability.

On the other hand, the second experiment show similar results to Janiczek and al. (2020) regarding the signatures, however the reconstruction is significantly worse.

For the record, it is interesting to note that the experiments on the synthetic dataset with 3 EMs only yielded worse results than for 6 EMs. And Feely dataset is actually closely related to the physical derivation of the dispersion model, so the performance of the model here should not be considered as a general case. The initialisation in particular is too good.

Overall, the approach hasn't prooved itself here. Yet, as disclaimed earlier, the lack of time to properly fine-tune the parameters might be a reason. For instance, the parameters for the sparsity were simply taken from the original paper without further consideration. Moreover, the final objective function (3) feel very difficult to minimize, as the numerous terms conflict between each other. During the training process often there was one term dominating the optimisation; and as we saw in the figures the gradient descent has difficulties to move from the initialisation.

6. Conclusion

This project presented a new approach to perform hyperspectral unmixing accounting for endmember variability. An explicit, physics-based model was combined with specific prior information on the elements of the problem, motivated by observations from the real world. This yielded a new formulation of the optimisation problem, that unfortunately wasn't differentiable. In order to take advantage of the great out-of-the-box and computationally efficient existing tools (automatic differentiation, GPU computing), and to bypass the difficulty of dealing directly with a complex model, we proposed a differentiable relaxation of the problem, along with a simple, classic algorithm to solve it. Furthermore, an implementation of the approach has been provided and is available online. Some experimentations were run to assess the relevance of the approach.

While some work has already been done, there are still a lot of perspectives to explore. First of all, to provide a good evaluation of the approach, more datasets should be considered. Additionally, there wasn't enough time to properly tune the algorithm, which is crucial considering the complexity of the problem. For instance, other minimization approach like Alternating Minimization could be considered. We could also use a way to automatically set the regularization parameters, as the experiments have highlighted a significant variability of the results in that respect.

Acknowledgments

References

Joshua L. Bandfield. Global mineral distributions on mars. *Journal of Geophysical Research*, 107: 9–1–9–20, 2002.

José M. Bioucas-Dias and José M. P. Nascimento. Hyperspectral subspace identification. *IEEE Transactions on Geoscience and Remote Sensing*, 46:2435–2445, 2008.

Jérôme Bolte, Shoham Sabach, and Marc Teboulle. Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Mathematical Programming*, 146(1-2):459–494, August 2014. ISSN 0025-5610, 1436-4646. doi: 10.1007/s10107-013-0701-9. URL http://link.springer.com/10.1007/s10107-013-0701-9. tex.ids= Bolte2014b.

Ricardo Augusto Borsoi, Tales Imbiriba, and José Carlos Moreira Bermudez. Deep Generative Endmember Modeling: An Application to Unsupervised Spectral Unmixing. *arXiv:1902.05528 [cs]*, August 2019. URL http://arxiv.org/abs/1902.05528. arXiv: 1902.05528.

Ricardo Augusto Borsoi, Tales Imbiriba, José Carlos Moreira Bermudez, Cédric Richard, Jocelyn Chanussot, Lucas Drumetz, Jean-Yves Tourneret, Alina Zare, and Christian Jutten. Spectral Variability in Hyperspectral Data Unmixing: A Comprehensive Review. *arXiv:2001.07307 [eess]*, January 2020. URL http://arxiv.org/abs/2001.07307. tex.ids: Borsoi2020 arXiv: 2001.07307.

Xiangyi Chen, Sijia Liu, Ruoyu Sun, and Mingyi Hong. On the Convergence of A Class of Adam-Type Algorithms for Non-Convex Optimization. *arXiv:1808.02941 [cs, math, stat]*, March 2019. URL http://arxiv.org/abs/1808.02941. tex.ids= Chen2019b arXiv: 1808.02941.

Afef Cherni, Emilie Chouzenoux, Laurent Duval, and Jean-Christophe Pesquet. SPOQ ℓ_p-Over-ℓ_q Regularization for Sparse Signal Recovery Applied to Mass Spectrometry. *IEEE*

Transactions on Signal Processing, 68:6070–6084, 2020. ISSN 1941-0476. doi: 10.1109/TSP.2020. 3025731. tex.ids= Cherni2020 conferenceName: IEEE Transactions on Signal Processing.

Laurent Condat. A Generic Proximal Algorithm for Convex Optimization—Application to Total Variation Minimization. *IEEE Signal Processing Letters*, 21(8):985–989, August 2014. ISSN 1558-2361. doi: 10.1109/LSP.2014.2322123. tex.ids= Condat2014 conferenceName: IEEE Signal Processing Letters.

Kimberly C. Feely and Philip R. Christensen. Quantitative compositional analysis using thermal emission spectroscopy: Application to igneous and metamorphic rocks. *Journal of Geophysical Research: Planets*, 104(E10):24195–24210, 1999. doi: https://doi.org/10.1029/1999JE001034. URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/1999JE001034`.

Xin-Ru Feng, Heng-Chao Li, Jun Li, Qian Du, Antonio Plaza, and William J. Emery. Hyperspectral Unmixing Using Sparsity-Constrained Deep Nonnegative Matrix Factorization With Total Variation. *IEEE Transactions on Geoscience and Remote Sensing*, 56(10):6245–6257, October 2018. ISSN 0196-2892, 1558-0644. doi: 10.1109/TGRS.2018.2834567.

Michael Griffin and Hsiao-hua Burke. Compensation of hyperspectral data for atmospheric effects. *Lincoln Laboratory Journal*, 14, 01 2003.

Danfeng Hong, Lianru Gao, Jing Yao, Naoto Yokoya, Jocelyn Chanussot, Uta Heiden, and Bing Zhang. Endmember-Guided Unmixing Network (EGU-Net): A General Deep Learning Framework for Self-Supervised Hyperspectral Unmixing. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–14, 2021. ISSN 2162-2388. doi: 10.1109/TNNLS.2021.3082289. Conference Name: IEEE Transactions on Neural Networks and Learning Systems.

Janiczek and al. *Differentiable Programming for Hyperspectral Unmixing Using a Physics-Based Dispersion Model*. Lecture Notes in Computer Science. Springer International Publishing, Cham, 2020. ISBN 978-3-030-58583-9. doi: 10.1007/978-3-030-58583-9_39.

Erik C. Johnson and Douglas L. Jones. Joint recovery of sparse signals and parameter perturbations with parameterized measurement models. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 5900–5904, 2013. doi: 10.1109/ICASSP.2013.6638796.

Hoel Kervadec, Jose Dolz, Jing Yuan, Christian Desrosiers, Eric Granger, and Ismail Ben Ayed. Constrained Deep Networks: Lagrangian Optimization via Log-Barrier Extensions. *arXiv:1904.04205 [cs]*, April 2020. URL `http://arxiv.org/abs/1904.04205`. arXiv: 1904.04205.

Nirmal Keshava and John Mustard. Spectral unmixing. *Signal Processing Magazine, IEEE*, 19:44 – 57, 02 2002. doi: 10.1109/79.974727.

Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. *arXiv:1412.6980 [cs]*, January 2017. URL `http://arxiv.org/abs/1412.6980`. arXiv: 1412.6980 version: 5.

Toru Kouyama, Yasuhiro Yokota, Yoshiaki Ishihara, Ryosuke Nakamura, Satoru Yamamoto, and Tsuneo Matsunaga. Development of an application scheme for the selene/sp lunar reflectance model for radiometric calibration of hyperspectral and multispectral sensors. *Planetary and Space Science*, 124:76–83, 2016. ISSN 0032-0633. doi: https://doi.org/10.1016/j.pss.2016.02.003. URL `https://www.sciencedirect.com/science/article/pii/S0032063315301008`.

D. Kraft. *A software package for sequential quadratic programming*. Deutsche Forschungs- und Versuchsanstalt für Luft- und Raumfahrt Köln: Forschungsbericht. Wiss. Berichtswesen d. DFVLR, 1988. URL `https://books.google.fr/books?id=4rKaGwAACAAJ`.

Guolan Lu and Baowei Fei. Medical hyperspectral imaging: A review. *Journal of biomedical optics*, 19:10901, 01 2014. doi: 10.1117/1.JBO.19.1.010901.

Jakub Marecek, Peter Richtárik, and Martin Takáč. Distributed block coordinate descent for minimizing partially separable functions. 06 2014. doi: 10.1007/978-3-319-17689-5_11.

Balas K. Natarajan. Sparse approximate solutions to linear systems. *SIAM J. Comput.*, 24:227–234, 1995.

Pierre-Antoine Thouvenin, Nicolas Dobigeon, and Jean-Yves Tourneret. Hyperspectral unmixing with spectral variability using a perturbed linear mixing model. *IEEE Transactions on Signal Processing*, 64(2):525–538, 2016. doi: 10.1109/TSP.2015.2486746.

M.A. Veganzones, L. Drumetz, G. Tochon, M. Dalla Mura, A. Plaza, J. Bioucas-Dias, and J. Chanussot. 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)*, pages 1–4, 2014. doi: 10.1109/WHISPERS.2014.8077595.