

Rapide introduction au démélange spectral

Guillaume TOCHON & Bashar DUDIN

1 Contexte général

La télédétection est un domaine qui consiste en l’observation d’un phénomène physique d’intérêt à distance, par opposition à une observation *in situ*. En particulier, l’imagerie satellitaire permet de cartographier l’occupation des sols pour des applications variées, allant de l’agriculture de précision au suivi environnemental, en passant par la planification urbaine. L’identification automatique des matériaux présents dans la scène observée ainsi que l’estimation de leurs proportions dans toute l’image est un problème central pour ces applications. Cependant, des images en niveau de gris ou couleur classiques sont souvent insuffisantes pour discriminer les matériaux d’intérêt (par exemple différentes espèces de végétation). On utilise donc des modalités d’imagerie non conventionnelles telles que l’imagerie hyperspectrale [1]. Ces images consistent en l’acquisition d’information dans de nombreuses longueurs d’ondes du domaine visible ou infrarouge proche (et pas seulement dans le rouge, vert et bleu comme en imagerie couleur classique) par décomposition de la lumière incidente au capteur via un prisme optique, comme illustré par la figure 1a. Chaque “tranche” du cube hyperspectral est associée à une longueur d’onde précise du spectre électromagnétique, et reflète donc la manière dont la lumière qui a frappé la scène a interagit avec les matériaux constituant cette scène. Ainsi, par opposition à l’imagerie couleur classique où chaque pixel est composé d’un triplet de valeurs, un pixel hyperspectral, appelé spectre de réflectance, est constitué de dizaines (jusqu’à quelques centaines) de valeurs, comme illustré par la figure 1b. La connaissance des spectres de réflectance permet une caractérisation bien plus précise des matériaux constituant la scène imagée, d’où le fort intérêt de ces images pour les problématiques d’imagerie satellitaire (et de plus en plus en imagerie médicale également).

2 Le démélange spectral comme problème inverse

2.1 Les modèles de mélange

Lorsqu’un capteur hyperspectral, qu’il soit aérien ou spatial, recueille une image de la Terre, il enregistre en fait la quantité de lumière qui a rebondi sur la surface et a atteint le capteur. Chaque

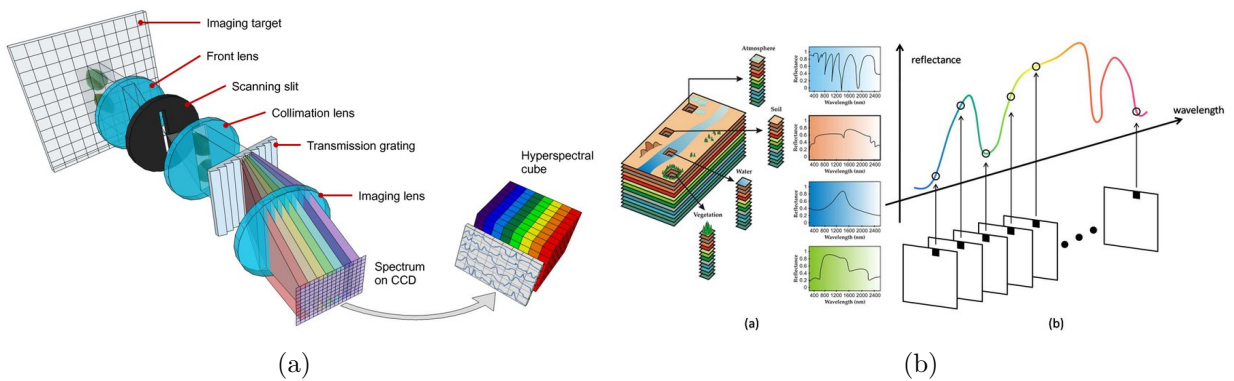


FIGURE 1 – (a) décomposition de la lumière par prisme dans un capteur hyperspectral, et (b) cube obtenu, où chaque pixel s’associe à un spectre de réflectance.

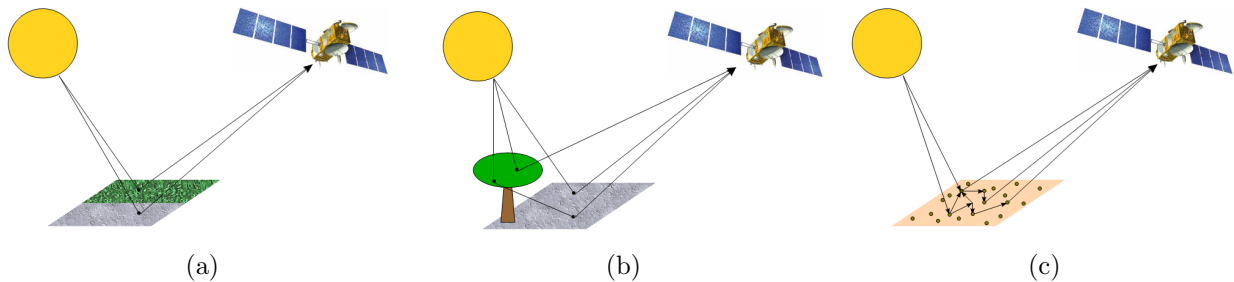


FIGURE 2 – Différents modèles d’interaction entre la lumière incidente et la surface. Les interactions peuvent être modélisées comme linéaires (a) lorsque chaque rayon de lumière ne rebondit qu’une seule fois, ou non linéaires (b) - (c) en raison de rebonds multiples ou de mélanges intimes au niveau du matériau.

rayon de lumière interagit avec les éléments présents sur le sol lors de son rebond. Le démixage spectral a pour objectif, compte tenu de ce qui a atteint le capteur, d’identifier quels étaient les éléments à la surface et comment la lumière interagissait avec eux.

Ces interactions peuvent être physiquement modélisées comme non linéaires pour plusieurs raisons, telles que la topographie du sol qui peut conduire à plusieurs rebonds (illustrés par la figure 2b) ou la granularité du matériau qui peut générer des phénomènes de mélange intimes (illustrés par la figure 2c). Cependant, en raison de la complexité de ces modèles, les effets non linéaires sont souvent négligés et approximés par un modèle de mélange linéaire (LMM, pour *linear mixing model*). Ce dernier suppose que chaque rayon de lumière ne rebondit que sur un élément au sol, de sorte que le signal optique reçu par le capteur pour chaque pixels s’écrive comme une combinaison linéaire de toutes les interactions qui se sont produites au sein de ce site (voir figure 2a). Malgré sa relative simplicité, la plupart des méthodes de démixage spectral dans la littérature de l’imagerie hyperspectrale sont basées sur le LMM [2] car il permet des interprétations géométriques simples.

2.2 Un focus sur le modèle de mélange linéaire

Étant donnée une image hyperspectral, un algorithme de démixage spectral classique génère d’abord l’ensemble des signatures spectrales des constituants purs de la scène. Ces constituants purs (on les appelle normalement des “pôles de mélanges” en bon français, mais comme c’est moche, on les appellera par leur appellation anglaise par la suite, à savoir *endmembers*) correspondent aux matériaux macroscopique composant la scène (par exemple, l’eau, la terre, la végétation, le béton, etc). Une fois les signatures spectrales des endmembers estimées, il ne reste qu’à calculer leurs *abondances fractionnelles* pour chaque pixel. Autrement dit, étant donné le spectre de réflectance d’un pixel, on “inverse le modèle” pour déterminer les proportions des constituants purs au sein de ce pixel.

Plus particulièrement, si N est le nombre de bandes/canaux de l’image hyperspectrale manipulée, et $\mathbf{e}_1, \dots, \mathbf{e}_m$ les signatures spectrales des m endmembers constituant l’image (avec $\mathbf{e}_i \in \mathbb{R}^N$), le LMM (illustré par la figure 3) indique que le spectre \mathbf{x} d’un pixel s’exprime comme une combinaison linéaire des signatures spectrales des matériaux purs présents dans l’image, pondérées par leurs abondances fractionnelles :

$$\mathbf{x} = \sum_{i=1}^m \phi_i \mathbf{e}_i + \boldsymbol{\eta}, \quad (1)$$

où ϕ_i est l’abondance fractionnelles de \mathbf{e}_i dans \mathbf{x} , et $\boldsymbol{\eta}$ est un bruit additif additionnel, permettant de modéliser les erreurs d’approximation du LMM.

Ainsi, à chaque pixel \mathbf{x} est associé un vecteur d’abondances fractionnelles $\boldsymbol{\phi} = [\phi_1, \dots, \phi_m]^T \in \mathbb{R}^m$, qui modélise la contribution (en pourcentage) de chaque endmember \mathbf{e}_i dans la signature \mathbf{x} . Sans contraintes additionnelles, le vecteur d’abondances $\boldsymbol{\phi} = [\phi_1, \dots, \phi_m]^T$ s’obtient aisément par moindres

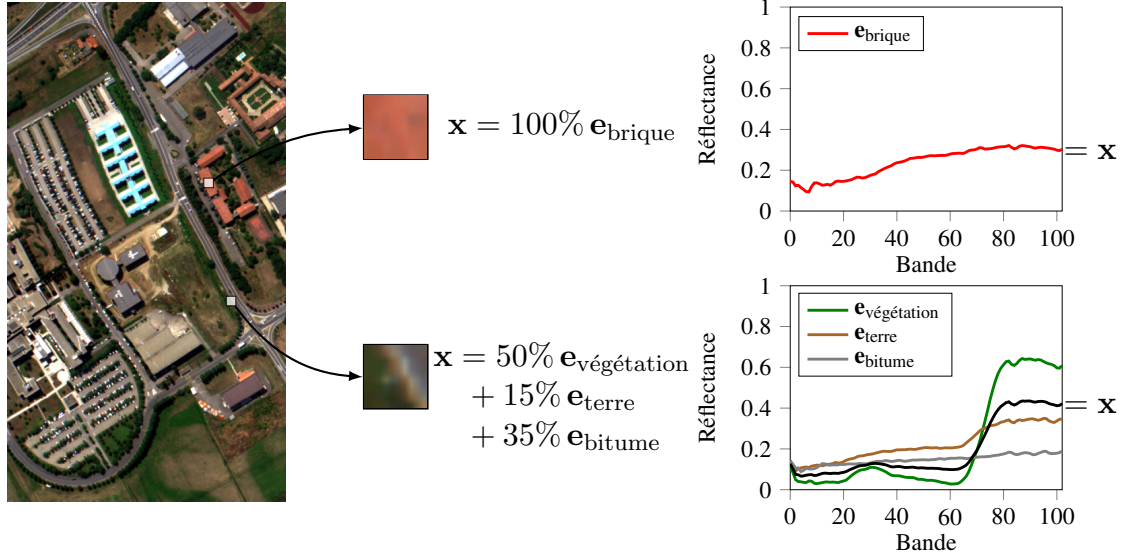


FIGURE 3 – Illustration du LMM : le spectre de chaque pixel \mathbf{x} s'exprime comme combinaison linéaire des signatures des endmembers, pondérés par leurs abondances fractionnelles au sein du pixel.

carrés ordinaires, comme solution du problème d'optimisation

$$\arg \min_{\phi \in \mathbb{R}^m} \|\mathbf{x} - \sum_{i=1}^m \phi_i \mathbf{e}_i\|_2^2, \quad (2)$$

que l'on peut réécrire

$$\arg \min_{\phi \in \mathbb{R}^m} \|\mathbf{x} - \mathbf{E}\phi\|_2^2 \quad (3)$$

avec $\mathbf{E} = [\mathbf{e}_1, \dots, \mathbf{e}_m] \in \mathbb{R}^{N \times m}$ la matrice regroupant les m endmembers stackés en colonnes.

Mais puisque les abondances fractionnelles sont des proportions, elles doivent respecter les deux contraintes de positivité et de somme à 1 pour chaque pixel :

$$\phi_i \geq 0 \quad \forall i = 1, \dots, m \quad (\text{positivité}), \quad (4)$$

$$\sum_{i=1}^m \phi_i = \mathbf{1}_m^T \phi = 1 \quad (\text{somme à 1}). \quad (5)$$

Armés de ces deux contraintes, le problème d'estimation des abondances fractionnelles d'un pixel \mathbf{x} se réécrit

$$\arg \min_{\substack{\phi \in \mathbb{R}^m \\ \phi_i \geq 0 \quad \forall i=1, \dots, m \\ \mathbf{1}_m^T \phi = 1}} \|\mathbf{x} - \mathbf{E}\phi\|_2^2 \quad (6)$$

Le LMM combiné avec les contraintes de positivité et de somme à 1 peut être interprété d'un point de vue géométrique : les m endmembers de l'image forment un $(m - 1)$ -simplexe dont les sommets sont les endmembers. Tous les pixels de l'image, s'écrivant comme combinaison linéaire des endmembers pondérés par les abondances fractionnelles, se trouvent alors à l'intérieur du simplexe, comme illustré par la figure 4. Cette interprétation ouvre notamment la voie à plusieurs méthodes géométriques pour l'identification des endmembers. Le lecteur intéressé pourra se référer à [2].

Une image hyperspectrale peut se représenter de deux manières :

- Comme un cube de données (tenseur) en 3 dimensions : $\mathcal{X} \in \mathbb{R}^{N_{row} \times N_{col} \times N}$, avec N_{row} , N_{col} et N étant le nombre de lignes, de colonnes et de bandes dans le cube 3D. Avec cette représentation, le spectre \mathbf{x} du pixel de la $i^{\text{ème}}$ ligne et $j^{\text{ème}}$ colonne s'obtient comme $\mathbf{x} = \mathcal{X}[i, j, :]$

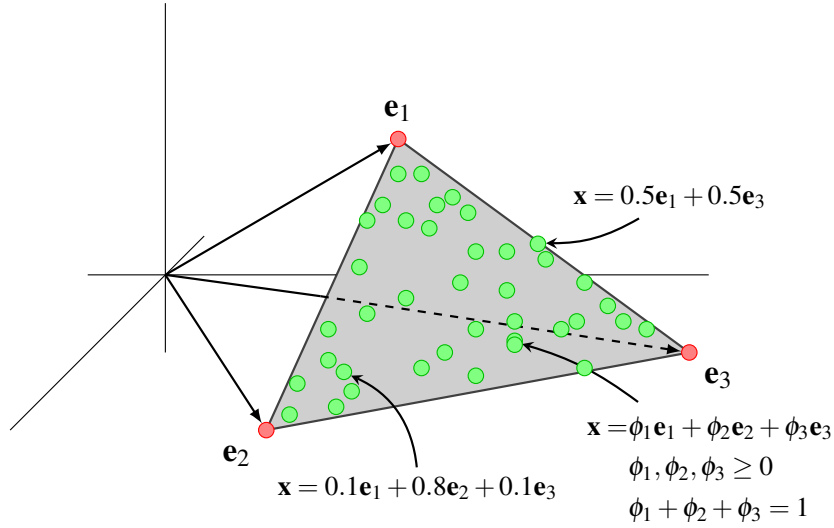


FIGURE 4 – Interprétation géométrique du LMM avec les contraintes de positivité et de somme à 1 pour un 2-simplexe constitué de $m = 3$ endmembers. Tous les pixels de l'image sont à l'intérieur du simplexe (un triangle dans ce cas, puisque constitué de 3 sommets).

- Comme une matrice $\mathbf{X} \in \mathbb{R}^{N_{pix} \times N}$ avec $N_{pix} = N_{row} \times N_{col}$ le nombre de pixels dans l'image. Pour passer de \mathcal{X} à \mathbf{X} , il suffit de vectoriser les deux premières dimensions. Avec cette représentation, le spectre \mathbf{x} du pixel de la $i^{\text{ème}}$ ligne et $j^{\text{ème}}$ colonne s'obtient comme $\mathbf{x} = \mathbf{X}[\text{idx}, :]$, où idx correspond à l'indice (i, j) vectorisé. En fait, on préfère utiliser plutôt $\mathbf{X}^T \in \mathbb{R}^{N \times N_{pix}}$ en pratique pour avoir N lignes (N étant toujours le nombre de bandes) et N_{pix} colonnes.

En fonction du traitement à appliquer à l'image, on pourra passer de l'une à l'autre de ces représentations pour manipuler la plus avantageuse des deux.

En particulier, lorsque la représentation matricielle est utilisée, donc $\mathbf{X} \in \mathbb{R}^{N \times N_{pix}}$, il est possible d'étendre l'équation (1) à l'image entière :

$$\mathbf{X} = \mathbf{E}\Phi + \boldsymbol{\eta}, \quad (7)$$

où $\mathbf{E} \in \mathbb{R}^{N \times m}$ est toujours la matrice dont les colonnes correspondent aux m endmembers, $\Phi \in \mathbb{R}^{m \times N_{pix}}$ est la matrice des abondances fractionnelles et $\boldsymbol{\eta}$ est un bruit additif indépendant (avec les dimensions qui vont bien). Cette formulation permet l'utilisation de méthodes de factorisation matricielle pour déterminer les matrices \mathbf{E} et Φ en résolvant le problème suivant :

$$\arg \min_{\substack{\mathbf{E}, \Phi \\ \Phi \succeq 0 \\ \mathbf{1}_m^T \Phi = \mathbf{1}_{N_{pix}}}} \|\mathbf{X} - \mathbf{E}\Phi\|_F^2, \quad (8)$$

où $\|\cdot\|_F$ est souvent formulé comme les normes euclidiennes ou de Frobenius. À noter que dans l'équation (8), $\Phi \succeq 0$ désigne que toutes les entrées de la matrice Φ sont positives ou nulles.

La plupart du temps, les signatures spectrales des matériaux sont inconnues et l'ensemble des endmembers doit être construit soit en sélectionnant des signatures spectrales dans une bibliothèque, soit en les estimant automatiquement à partir de l'image elle-même. La littérature de l'imagerie hyperspectral fourmille d'algorithmes d'estimation automatique des endmembers, et le lecteur intéressé pourra aller chercher dans [1] pour les approches conventionnelles pré-deep learning, ou dans [4] pour les derniers développements sur le sujet.

1. Notez ici qu'on a laissé tomber le $(\cdot)^T$, mais dans ce cas, pour obtenir \mathbf{X} à partir de \mathcal{X} , on vectorise d'abord les deux dimensions spatiales en une, puis on transpose la matrice obtenue : $\mathbf{X} = (\text{vectorize}(\mathcal{X}).axis(0, 1))^T$

Une fois en possession de la matrice des endmembers \mathbf{E} , les abondances fractionnelles Φ peuvent être estimées comme une solution à un système linéaire surdéterminé, en résolvant l'équation (8) par la méthode des moindres carrés. Lorsque les contraintes de positivité et de somme à 1 ne sont pas prises en compte dans la résolution, la solution de ce problème peut se déterminer analytiquement. En revanche, lorsque les contraintes de positivité et de somme à 1 sont actives, il n'existe pas de solution analytique au problème (8), qui se résoud donc par méthodes itératives [3].

La qualité du démixage pour un pixel donné \mathbf{x} peut être mesurée par l'erreur quadratique moyenne (RMSE, pour *root mean square error*) de la signature hyperspectrale d'origine \mathbf{x} par rapport à la signature reconstruite $\hat{\mathbf{x}} = \sum_{i=1}^m \phi_i \mathbf{e}_i$:

$$\epsilon(\mathbf{x}, \hat{\mathbf{x}}) = \sqrt{\frac{1}{N} \sum_{k=1}^N (x(k) - \hat{x}(k))^2}. \quad (9)$$

Le calcul du RMSE de chaque pixel \mathbf{x} donne une carte de reconstruction pour laquelle des valeurs faibles signifient que les pixels correspondants ont été bien reconstruits par l'ensemble des endmembers. Au contraire, un RMSE élevé pour un pixel signifie que sa signature spectrale n'est pas bien expliquée par une combinaison linéaire des endmembers. Cela peut aussi bien signifier que l'ensemble des endmembers utilisés pour le démixage est incorrect/incomplet, ou alors que le LMM n'est pas valide pour ce pixel en question.

Références

- [1] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot. Hyperspectral remote sensing data analysis and future challenges. *IEEE Geoscience and remote sensing magazine*, 1(2) :6–36, 2013.
- [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] D. C. Heinz et al. Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery. *IEEE transactions on geoscience and remote sensing*, 39(3) :529–545, 2001.
- [4] S. Zhang, Y. Su, X. Xu, J. Li, C. Deng, and A. Plaza. Recent advances in hyperspectral unmixing using sparse techniques and deep learning. In *Hyperspectral Image Analysis*, pages 377–405. Springer, 2020.