# Group Lasso avec contraintes de positivité et de somme unité pour le démélange non-supervisé des images hyperspectrales

André Ferrari, Cédric Richard, David Mary

Lab. J.-L. Lagrange, Université de Nice Sophia-Antipolis, CNRS, Observatoire de la Côte d'Azur
andre.ferrari@unice.fr, cedric.richard@unice.fr, david.mary@unice.fr

**Résumé** − Cet article présente une méthode de démélange spectral linéaire non-supervisée où ni les spectres des composants purs ni leur nombre ne sont connus. Une régularisation Group Lasso permet d'identifier l'espace engendré par les spectres des composants purs présents dans l'image, et d'estimer leurs abondances en chaque autre pixel. Un algorithme de type ADMM est mis en œuvre pour cela. Il repose notamment sur un opérateur proximal vectoriel de type $\ell_{1,2}$ sous contrainte de positivité. Une version distribuée dans laquelle chaque nœud traite une bande de fréquence est proposée. L'algorithme est appliqué à des données hyperspectrales, et comparé à une méthode de l'état de l'art.

**Abstract** − This communication proposes an algorithm for unsupervised linear unmixing where neither the endmembers nor their number are known. Joint identification of the endmembers available in the data and estimation of the abundances for each pixel is achieved solving a Group Lasso regularized convex problem. Optimization is performed with an ADMM algorithm which relies on a $\ell_{1,2}$ type proximal operator under positivity constraint. A distributed version where each node processes a frequency band is proposed. The algorithm is applied to hyperspectral data and compared to a state of the art solution.

## 1. Contexte

L'imagerie hyperspectrale consiste à acquérir des images dans des centaines de bandes spectrales contigües et identiquement géo-référencées, avec une résolution suffisante pour caractériser la variabilité naturelle de la scène observée. Le démélange spectral constitue l'un des problèmes fondamentaux du domaine. Il s'agit de décomposer un ensemble de vecteurs spectraux sur une collection de signatures spectrales de composants purs, et d'estimer la fraction d'abondance de ces derniers dans le mélange [7]. Selon que ces composés élémentaires sont connus ou non, on qualifie le problème de démélange supervisé ou non-supervisé. Notre étude traite du démélange linéaire non-supervisé.

Soit $\boldsymbol{s} = (s_1, s_2, \ldots, s_P)^\top$ un vecteur hyperspectral, avec $P$ le nombre de longueurs d'onde. On suppose que $\boldsymbol{s}$ résulte du mélange de $L$ composés purs, corrompu par un bruit blanc gaussien $\boldsymbol{n}$. Soit $\boldsymbol{M}$ la matrice de taille $(P \times L)$ regroupant les signatures spectrales des composants purs, supposés inconnus par la suite. Enfin, on note $\boldsymbol{x} = (x_1, x_2, \ldots, x_L)^\top$ le vecteur des abondances associé au pixel $\boldsymbol{s}$. En considérant le modèle de mélange linéaire, on a donc $\boldsymbol{s} = \boldsymbol{M}\boldsymbol{x} + \boldsymbol{n}$ sous hypothèses de non-négativité des abondances et de leur somme unité, ce que l'on note $\boldsymbol{x} \succeq \boldsymbol{0}$ et $\boldsymbol{1}^\top \boldsymbol{x} = 1$. La littérature rapporte de nombreuses méthodes pour l'extraction des composants purs. Elles exploitent essentiellement le fait que, étant donné les propriétés de $\boldsymbol{x}$ évoquées ci-dessus, tous les vecteurs hyperspectraux d'une scène sont contenus dans un simplexe dont les sommets sont les composants purs. Par exemple, l'algorithme N-Findr recherche le simplexe de volume maximal dont les sommets sont des pixels-vecteurs de la scène [10]. De même, l'algorithme VCA vise à maximiser la distance entre un sommet issu de la scène et le sous-espace engendré par les sommets déjà retenus [8]. Il convient de noter que ces approches de l'état de l'art supposent que des pixels-vecteurs de composés purs sont présents dans l'image analysée, les travaux relaxant cette hypothèse étant par ailleurs peu nombreux. Une estimation préalable du nombre de composés purs est enfin nécessaire pour les mettre en œuvre, en suivant [3] par exemple.

L'objet principal de cet article est de présenter une alternative aux méthodes ci-dessus sous les mêmes hypothèses où, conjointement, l'on détermine le nombre de composés purs et leur spectre, et l'on estime leurs fractions d'abondance en chaque pixel-vecteur de l'image. L'objet secondaire est de décrire une méthode d'optimisation d'un coût quadratique, régularisé par Group Lasso, sous contraintes de positivité et de somme unité. Ces problématiques ont été également étudiées dans [5, 6]. Les différences avec la solution proposée seront soulignées dans la suite.

## 2. Group Lasso avec contraintes de positivité et de somme unité pour le démélange

Soit $\boldsymbol{S} = (\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N)$ la matrice des pixels-vecteurs de l'image. On cherche conjointement à identifier les composés purs au sein de la matrice $\boldsymbol{S}$, et estimer l'ensemble des vecteurs d'abondance regroupés dans $\boldsymbol{X} = (x_1, \ldots, x_N)$. Chaque spectre $\boldsymbol{s}_k$ devant s'exprimer à partir d'un faible nombre de composés spectraux $\boldsymbol{s}_\ell$, le modèle peut s'exprimer sous la forme $\boldsymbol{S} = \boldsymbol{S}\boldsymbol{X} + \boldsymbol{E}$ où $\boldsymbol{E}$ est un bruit, et $\boldsymbol{X}$ la matrice $(N \times N)$ des abondances dont la caractéristique est d'être constituée de nombreuses lignes nulles. On note $\boldsymbol{x}_{k,:}$ la $k^e$ ligne de la matrice $\boldsymbol{X}$. Avec ce formalisme, la

matrice $\boldsymbol{M}$ des composants purs est simplement égale à $\boldsymbol{S}$ privée de toutes les colonnes $k$ associées à $\boldsymbol{x}_{k,:} = \boldsymbol{0}$.

On s'intéresse donc au problème convexe suivant

$$\begin{array}{ll} \min_{\boldsymbol{X}} & \|\boldsymbol{S}(\boldsymbol{I}-\boldsymbol{X})\|_{\mathrm{F}}^2 + \mu \sum_{k=1}^N \|\boldsymbol{x}_{k,:}\|_2 \\ \text{sujet à} & \forall k,q \quad x_{k,q} \geq 0 \\ & \forall q \quad \sum_{k=1}^N x_{k,q} = 1 \end{array} \tag{1}$$

Notons que sans régularisation Group Lasso, la solution du problème est triviale : $\boldsymbol{X} = \boldsymbol{I}$. Cette approche est similaire à celle proposée dans [5]. Les principales différences sont que d'une part les auteurs y encouragent la parcimonie des $\boldsymbol{x}_{k,:}$ par une norme infinie et que la normalisation des colonnes de $\boldsymbol{S}$ leur permet de s'affranchir de la contrainte de somme à 1.

Le problème (1) est un cas particulier du problème :

$$\begin{array}{ll} \min_{\boldsymbol{X}} & \|\boldsymbol{H}\boldsymbol{X}-\boldsymbol{Y}\|_{\mathrm{F}}^2 + \mu \sum_{k=1}^N \|\boldsymbol{x}_{k,:}\|_2 \\ \text{sujet à} & \forall k,q \quad x_{k,q} \geq 0 \\ & \forall q \quad \sum_{k=1}^N x_{k,q} = 1 \end{array} \tag{2}$$

avec $\boldsymbol{H} = \boldsymbol{S}$ et $\boldsymbol{Y} = \boldsymbol{S}$. La suite de cette section sera donc consacrée au développement d'un algorithme d'optimisation pour (2). Soulignons que [6], qui propose de réaliser le démélange à partir d'une librairie connue de spectres purs, repose sur la résolution du problème (2), sans la contrainte de somme à 1.

L'algorithme *Alternating Direction Method of Multipliers* (ADMM) [2, 4] connait actuellement une très grande popularité en raison de sa facilité de mise en œuvre et de ses performances éprouvées pour une grande variété de problèmes. Le problème (2) peut être réécrit sous la forme canonique

$$\begin{array}{ll} \min_{\boldsymbol{X},\boldsymbol{Z}} & \frac{1}{2}\|\boldsymbol{H}\boldsymbol{X}-\boldsymbol{Y}\|_{\mathrm{F}}^2 + \mu \sum_{k=1}^N \|\boldsymbol{z}_{k,:}\|_2 + \mathcal{I}(\boldsymbol{Z}) \\ \text{sujet à} & \boldsymbol{A}\boldsymbol{X} + \boldsymbol{B}\boldsymbol{Z} = \boldsymbol{C} \end{array} \tag{3}$$

avec

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{I} \\ \boldsymbol{1}^\top \end{pmatrix},\ \boldsymbol{B} = \begin{pmatrix} -\boldsymbol{I} \\ \boldsymbol{0} \end{pmatrix},\ \boldsymbol{C} = \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{1}^\top \end{pmatrix}. \tag{4}$$

où $\mathcal{I}$ est la fonction indicatrice sur l'orthant positif : $\mathcal{I}(\boldsymbol{Z}) = 0$ pour $\boldsymbol{Z} \succeq \boldsymbol{0}$ et $+\infty$ sinon. Le choix de porter la contrainte de positivité sur $\boldsymbol{Z}$ plutôt que $\boldsymbol{X}$ joue un rôle essentiel dans la formulation de l'algorithme comme on verra dans la suite. Le Lagrangien augmenté s'écrit [4]

$$L_\rho(\boldsymbol{X},\boldsymbol{Z},\boldsymbol{\Lambda}) = \frac{1}{2}\|\boldsymbol{H}\boldsymbol{X}-\boldsymbol{Y}\|_F^2 + \mu\sum_{k=1}^N \|\boldsymbol{z}_{k,:}\|_2 + \mathcal{I}(\boldsymbol{Z}) +$$
$$+ \operatorname{trace}(\boldsymbol{\Lambda}^\top(\boldsymbol{A}\boldsymbol{X}+\boldsymbol{B}\boldsymbol{Z}-\boldsymbol{C})) + \frac{\rho}{2}\|\boldsymbol{A}\boldsymbol{X}+\boldsymbol{B}\boldsymbol{Z}-\boldsymbol{C}\|_{\mathrm{F}}^2$$

où $\boldsymbol{\Lambda}$ est la matrice des facteurs de Lagrange et $\rho > 0$ un paramètre de pénalisation. Cette fonction coût est minimisée itérativement par rapport à $\boldsymbol{X}$ et $\boldsymbol{Z}$.

### a) Minimisation de $L_\rho(\boldsymbol{X},\boldsymbol{Z},\boldsymbol{\Lambda})$ par rapport à $\boldsymbol{X}$

La fonction coût est quadratique en $\boldsymbol{X}$ et admet une solution analytique, obtenue en résolvant le système linéaire

$$(\boldsymbol{H}^\top\boldsymbol{H} + \rho\boldsymbol{A}^\top\boldsymbol{A})\boldsymbol{X} = \boldsymbol{H}^\top\boldsymbol{Y} - \boldsymbol{A}^\top(\boldsymbol{\Lambda} + \rho\,(\boldsymbol{B}\boldsymbol{Z}-\boldsymbol{C})) \tag{5}$$

### b) Minimisation de $L_\rho(\boldsymbol{X},\boldsymbol{Z},\boldsymbol{\Lambda})$ par rapport à $\boldsymbol{Z}$

Il s'agit donc de résoudre à présent

$$\begin{array}{ll} \min_{\boldsymbol{Z}} & \mu\sum_{k=1}^N \|\boldsymbol{z}_{k,:}\|_2 + \operatorname{trace}(\boldsymbol{\Lambda}^\top\boldsymbol{B}\boldsymbol{Z}) + \\ & \quad \frac{\rho}{2}\|\boldsymbol{A}\boldsymbol{X}+\boldsymbol{B}\boldsymbol{Z}-\boldsymbol{C}\|_{\mathrm{F}}^2 \\ \text{sujet à} & \boldsymbol{Z} \succeq \boldsymbol{0} \end{array} \tag{6}$$

En considérant la forme particulière des matrices $\boldsymbol{A}$, $\boldsymbol{B}$ et $\boldsymbol{C}$ et en notant $\boldsymbol{\lambda}_{k,:}$ les lignes de la matrice $\boldsymbol{\Lambda}$ on obtient :

$$\operatorname{trace}(\boldsymbol{\Lambda}^\top\boldsymbol{B}\boldsymbol{Z}) = -\sum_{k=1}^N \boldsymbol{z}_{k,:}\boldsymbol{\lambda}_{k,:}^\top$$

$$\|\boldsymbol{A}\boldsymbol{X}+\boldsymbol{B}\boldsymbol{Z}-\boldsymbol{C}\|_{\mathrm{F}}^2 = \sum_{k=1}^N \|\boldsymbol{x}_{k,:}-\boldsymbol{z}_{k,:}\|_2^2 + \|\boldsymbol{1}^\top\boldsymbol{X}-\boldsymbol{1}^\top\|_2^2$$

Le problème (6) est donc séparable selon les lignes de $\boldsymbol{Z}$ et peut se réécrire pour chacune de ces lignes ainsi

$$\begin{array}{ll} \min_{\boldsymbol{z}} & \frac{1}{2}\|\boldsymbol{z}-\boldsymbol{v}\|_2^2 + \alpha\|\boldsymbol{z}\|_2 \\ \text{sujet à} & \boldsymbol{z} \succeq \boldsymbol{0} \end{array} \tag{7}$$

avec $\boldsymbol{v} = \boldsymbol{x} + \rho^{-1}\boldsymbol{\lambda}$ et $\alpha = \rho^{-1}\mu$. $\boldsymbol{\lambda}$ désigne ici la ligne $k$ de $\boldsymbol{\Lambda}$ concernée, comme $\boldsymbol{x}$ et $\boldsymbol{z}$, par le problème d'optimisation ci-dessus. Soit $(\cdot)_+ = \max(0,\cdot)$.

**Proposition 1.** *La solution du problème* (7) *est donnée par l'opérateur de seuillage multidimensionnel suivant, appelé* Positively-constrained Multidimensionnel Shrinkage-thresholding Operator *(P-MiSTO) :*

$$\begin{cases} \textit{si } \|(\boldsymbol{v})_+\|_2 < \alpha : & \boldsymbol{z}^* = \boldsymbol{0} \\ \textit{sinon :} & \boldsymbol{z}^* = \left(1 - \frac{\alpha}{\|(\boldsymbol{v})_+\|_2}\right)(\boldsymbol{v})_+ \end{cases} \tag{8}$$

*Démonstration.* Le problème étant fortement convexe, nous allons simplement vérifier le résultat obtenu dans les 2 cas $\|(\boldsymbol{v})_+\|_2 > \alpha$ et $\|(\boldsymbol{v})_+\|_2 < \alpha$. Notons pour cela $f_0(\boldsymbol{z}) = \frac{1}{2}\|\boldsymbol{z}-\boldsymbol{v}\|_2^2 + \alpha\|\boldsymbol{z}\|_2$.

Dans le premier cas le gradient de $f_0$ en $\boldsymbol{z}^* \succ \boldsymbol{0}$ est :

$$\nabla f_0(\boldsymbol{z}^*) = \left(1 + \frac{\alpha}{\|\boldsymbol{z}^*\|_2}\right)\boldsymbol{z}^* - \boldsymbol{v}$$

en remplaçant (8) on obtient :

$$\nabla f_0(\boldsymbol{z}^*) = (\boldsymbol{v})_+ - \boldsymbol{v} \geq 0 \tag{9}$$
$$\boldsymbol{z}_i^* \cdot \nabla f_0(\boldsymbol{z}^*)_i \propto ((\boldsymbol{v})_+)_i \cdot ((\boldsymbol{v})_+ - \boldsymbol{v})_i = 0 \tag{10}$$

Ces deux conditions sont les conditions d'optimalité pour que $\boldsymbol{z} \succeq \boldsymbol{0}$ soit solution du problème contraint, voir section 4.2.3 de [1].

Dans le second cas, remarquons que pour tout $\boldsymbol{z} \succeq \boldsymbol{0}$ :

$$\sum_i \boldsymbol{z}_i\boldsymbol{v}_i \leq \sum_i \boldsymbol{z}_i(\boldsymbol{v}_i)_+ \leq \|\boldsymbol{z}\|_2 \cdot \|(\boldsymbol{v})_+\|_2$$

On a donc :

$$\begin{aligned} f_0(\boldsymbol{z}) - f_0(\boldsymbol{0}) &= \frac{1}{2}\sum_i \boldsymbol{z}_i^2 - \sum_i \boldsymbol{z}_i\boldsymbol{v}_i + \alpha\|\boldsymbol{z}\|_2 \\ &\geq \frac{1}{2}\|\boldsymbol{z}\|_2^2 - \|\boldsymbol{z}\|_2 \cdot \|(\boldsymbol{v})_+\|_2 + \alpha\|\boldsymbol{z}\|_2 \\ &\geq \frac{1}{2}\|\boldsymbol{z}\|_2^2 + \|\boldsymbol{z}\|_2(\alpha - \|(\boldsymbol{v})_+\|_2) \end{aligned} \tag{11}$$

ce qui prouve que si $\alpha \geq \|(\boldsymbol{v})_+\|_2$ le minimum est bien atteint en $\boldsymbol{z}^* = \boldsymbol{0}$. □

**Algorithme 1:** GLPC

$\boldsymbol{Q} \leftarrow (\boldsymbol{H}^\top \boldsymbol{H} + \rho \boldsymbol{A}^\top \boldsymbol{A})^{-1}$;
**while** $\|\boldsymbol{R}^{[t]}\|_2 > \epsilon^{prim}$ *and* $\|\boldsymbol{P}^{[t]}\|_2 > \epsilon^{dual}$ **do**
  $\boldsymbol{X}^{[t+1]} \leftarrow \boldsymbol{Q}(\boldsymbol{H}^\top \boldsymbol{Y} - \boldsymbol{A}^\top(\rho(\boldsymbol{B}\boldsymbol{Z}^{[t]} - \boldsymbol{C}) + \boldsymbol{\Lambda}^{[t]}))$;
  **for** $k = 1 \dots N$ **do**
    $\boldsymbol{v}_{k,:}^{[t+1]} \leftarrow \left(\boldsymbol{x}_{k,:}^{[t+1]} + \frac{1}{\rho}\boldsymbol{\lambda}_{k,:}^{[t]}\right)_+$;
    **if** $\boldsymbol{v}_{k,:}^{[t+1]} = 0$ **then**
      $\boldsymbol{z}_{k,:}^{[t+1]} \leftarrow 0$
    **else**
      $\boldsymbol{z}_{k,:}^{[t+1]} \leftarrow \left(1 - \frac{\mu/\rho}{\|\boldsymbol{v}_{k,:}^{[t+1]}\|_2}\right)_+ \boldsymbol{v}_{k,:}^{[t+1]}$
    **end**
  **end**
  $\boldsymbol{R}^{[t+1]} \leftarrow \boldsymbol{A}\boldsymbol{X}^{[t+1]} + \boldsymbol{B}\boldsymbol{Z}^{[t+1]} - \boldsymbol{C}$;
  $\boldsymbol{P}^{[t+1]} \leftarrow \rho \boldsymbol{A}^t \boldsymbol{B}(\boldsymbol{Z}^{[t+1]} - \boldsymbol{Z}^{[t]})$;
  $\boldsymbol{\Lambda}^{[t+1]} \leftarrow \boldsymbol{\Lambda}^{[t]} + \rho \boldsymbol{R}^{[t+1]}$
**end**

ll convient de noter la similitude apparente entre l'opérateur (8), et l'opérateur MSTO [9] qui concerne le problème (7) privé de la contrainte de positivité.

### c) Mise-à-jour des multiplicateurs de Lagrange $\boldsymbol{\Lambda}$

La matrice $\boldsymbol{\Lambda}$ est mise à jour afin que, au terme de l'itération $[t+1]$ de l'algorithme $\boldsymbol{X}^{[t+1]} \rightarrow \boldsymbol{Z}^{[t+1]} \rightarrow \boldsymbol{\Lambda}^{[t+1]}$, le triplet résultant $(\boldsymbol{X}^{[t+1]}, \boldsymbol{Z}^{[t+1]}, \boldsymbol{\Lambda}^{[t+1]})$ vérifie la condition de faisabilité duale par rapport à $\boldsymbol{Z}$ :

$$\boldsymbol{\Lambda}^{[t+1]} = \boldsymbol{\Lambda}^{[t]} + \rho(\boldsymbol{A}\boldsymbol{X}^{[t+1]} + \boldsymbol{B}\boldsymbol{Z}^{[t+1]} - \boldsymbol{C}). \qquad (12)$$

### d) Algorithme GLPC : Group Lasso with Positive Constraint

L'algorithme GLPC consiste ainsi à itérer successivement les équations (5), (8) et (12). Les différences entre GLPC et les algorithmes proposés dans [5, 6] proviennent essentiellement de l'utilisation de l'opérateur P-MiSTO. Ainsi, [6] contourne son utilisation en utilisant plusieurs variables auxiliaires dans l'algorithme ADMM.

Les conditions de convergence sont que les fonctions coût relatives à $\boldsymbol{X}$ et $\boldsymbol{Z}$ soient fermées, propres et convexes ; l'algorithme proposé converge alors pour toute valeur de $\rho > 0$ [2]. La vitesse de convergence est évidemment fortement dépendante du choix de $\rho$. En notant $R^{[t]}$ et $P^{[t]}$ le résidu primal et dual à l'itération $t$, [2], on obtient l'Algorithme 1. Les paramètres $\epsilon^{\text{prim}}$ et $\epsilon^{\text{dual}}$ sont calculés à partir du résidu primal et dual à l'itération $t$ suivant les relations proposées dans [2] avec par exemple $\epsilon^{\text{abs}} = \epsilon^{\text{rel}} = 10^{-6}$.

Une limitation à l'utilisation de GLPC pour le démélange non-supervisé des images hyperspectrales vient de la nécessité de traiter conjointement l'ensemble du cube de données qui peut être de taille extrêmement importante. L'étape la plus critique en terme de charge de calculs est la résolution à chaque itération des $N$ équations linéaires (5). Notons que si $\boldsymbol{A} = \boldsymbol{0}$, la solution de (5) est la solution par moindres carrés de $\boldsymbol{H}\boldsymbol{X} = \boldsymbol{Y}$. Cette étape peut ainsi s'interpréter comme un démélange non contraint de $N$ spectre sur un ensemble de $N$ spectres.

De façon plus précise, considérons le modèle Bayésien hiérarchique suivant

$$\boldsymbol{Y} \sim \mathcal{MN}(\boldsymbol{H}\boldsymbol{X}, \rho\boldsymbol{I}, \boldsymbol{I}), \quad \boldsymbol{X} \sim \mathcal{MN}(\boldsymbol{M}, \boldsymbol{\Sigma}, \boldsymbol{I}) \qquad (13)$$

où $\boldsymbol{X} \sim \mathcal{MN}(\boldsymbol{M}, \boldsymbol{\Sigma}, \boldsymbol{I})$ signifie que $\boldsymbol{X}$ est de moyenne $\boldsymbol{M}$, que ses colonnes sont indépendantes entre elles et qu'elles sont individuellement Gaussiennes de covariance $\boldsymbol{\Sigma}$. La densité a posteriori de $\boldsymbol{X}|\boldsymbol{Y}$ est alors une Gaussienne dont la moyenne vérifie :

$$(\boldsymbol{H}^\top \boldsymbol{H} + \rho\boldsymbol{\Sigma}^{-1})\mathsf{E}[\boldsymbol{X}|\boldsymbol{Y}] = \boldsymbol{H}^\top \boldsymbol{Y} + \rho\boldsymbol{\Sigma}^{-1}\boldsymbol{M}$$

Comme noté dans [2] pour le Group Lasso "standard", l'étape (5) de GLPC correspond aussi à l'estimation par maximum a posteriori de $\boldsymbol{X}$ suivant un modèle Bayésien hiérarchique Gaussien (13), et pour un a priori qui dans le cas présent a pour covariance et moyenne :

$$\boldsymbol{\Sigma} = (\boldsymbol{A}^\top \boldsymbol{A})^{-1} = (\boldsymbol{I} + \boldsymbol{1}\boldsymbol{1}^\top)^{-1} = \boldsymbol{I} - (N+1)^{-1}\boldsymbol{1}\boldsymbol{1}^\top$$
$$\boldsymbol{M} = -\rho^{-1}\boldsymbol{\Sigma}\boldsymbol{A}^\top(\boldsymbol{\Lambda} + \rho\,(\boldsymbol{B}\boldsymbol{Z} - \boldsymbol{C}))$$

### e) Algorithme GLPC distribué en fréquence

Une solution naturelle pour réduire la charge de calcul consiste à distribuer l'algorithme d'optimisation. La solution classique consiste à partitionner $\boldsymbol{H}$ et $\boldsymbol{Y}$ en $d$ blocs verticaux :

$$\boldsymbol{H}^t = (\boldsymbol{H}_1^t, \dots, \boldsymbol{H}_d^t),\ \boldsymbol{Y}^t = (\boldsymbol{Y}_1^t, \dots, \boldsymbol{Y}_d^t)$$

conduisant au problème d'optimisation sous forme ADMM :

$$\begin{array}{ll} \min_{\boldsymbol{X}_i, \boldsymbol{Z}} & \frac{1}{2}\sum_{i=1}^d \|\boldsymbol{H}_i\boldsymbol{X}_i - \boldsymbol{Y}_i\|_{\text{F}}^2 + \\ & \qquad \mu \sum_{k=1}^N \|\boldsymbol{z}_{k,:}\|_2 + \mathcal{I}(\boldsymbol{Z}) \\ \text{sujet à} & \forall i = 1 \dots d,\ \boldsymbol{A}\boldsymbol{X}_i + \boldsymbol{B}\boldsymbol{Z} = \boldsymbol{C} \end{array} \qquad (14)$$

où les contraintes linéaires assurent le consensus entre les variables locales $X_i$. Dans le contexte de l'imagerie hyperspectrale, ce choix conduit à "découper" les mesures en bandes de fréquences, *chaque processeur prenant en charge les informations relatives à une bande spectrale.*

Les étapes de l'algorithme ADMM associé s'obtiennent de façon analogue au cas précédent.

- La maximisation du Lagrangien augmenté par rapport aux $X_i$, $i = 1 \dots d$, se fait en résolvant une équation analogue à (5) où $\boldsymbol{H}_i$ et $\boldsymbol{Y}_i$ se substituent à $\boldsymbol{H}$ et $\boldsymbol{Y}$.
- La maximisation par rapport à $\boldsymbol{Z}$ est encore séparable selon les lignes de $\boldsymbol{Z}$ et conduit chaque fois à résoudre un problème du type (7). La ligne $\boldsymbol{z}_k$ s'obtient en appliquant l'opérateur P-MiSTO pour $\alpha = (\rho d)^{-1}\mu$ au vecteur :
$$\boldsymbol{v} = \overline{\boldsymbol{x}}_{k,:} + \rho^{-1}\overline{\boldsymbol{\lambda}}_{k,:} \qquad (15)$$
Dans (15) $\overline{\boldsymbol{x}}_{k,:}$ et $\overline{\boldsymbol{\lambda}}_{k,:}$ représentent la moyenne de la $k^e$ ligne des $\boldsymbol{X}_i$ et des $\boldsymbol{\Lambda}_i$.
- Enfin, les $d$ multiplicateurs de Lagrange $\boldsymbol{\Lambda}_i$ sont mis à jour de façon analogue à (12) où $\boldsymbol{X}_i$ remplace $\boldsymbol{X}$.

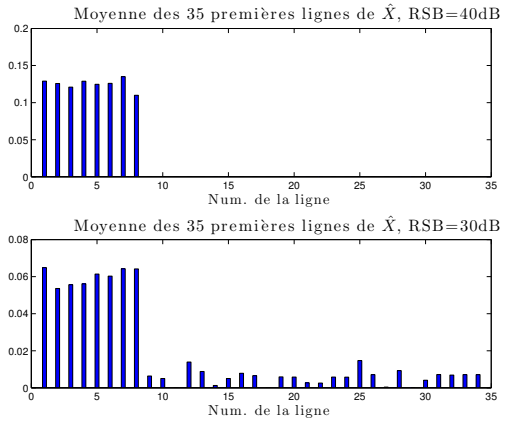


FIGURE 1 – Moyenne des 35 premières lignes de $\hat{\boldsymbol{X}}$.

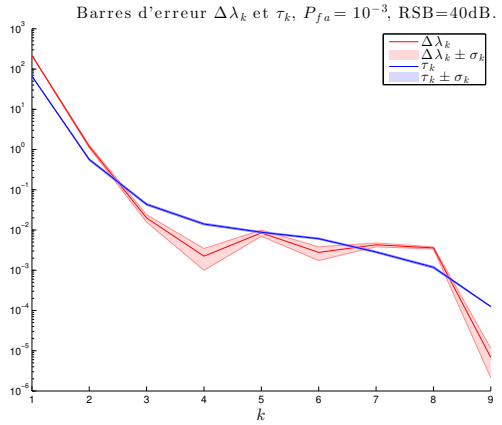


FIGURE 2 – Estimation du nombre $k$ de composantes [3].

## 4. Application au démélange des données hyperspectrales

Les données hyperspectrales considérées ont été générées à partir d'un mélange linéaire de 8 composants purs, extraits de la bibliothèque USGC. Les spectres sont constitués de 480 longueurs d'ondes. Cent spectres ont été générés, à partir de vecteurs d'abondance tirés selon une loi uniforme dans le simplexe $\boldsymbol{x} \succeq 0$, $\mathbf{1}^\top \boldsymbol{x} = 1$. Les 8 spectres des composés purs ont été introduits parmi ces réalisations, aux 8 premières places pour faciliter l'interprétation des figures. Chaque pixel a été corrompu par un bruit blanc additif gaussien. L'algorithme GLPC a été mis en œuvre sur ces données, avec $\rho = 1$ et $\mu = 0.3$. La Figure 1 présente la moyenne des 35 premières lignes de $\hat{\boldsymbol{X}}$ pour un RSB de 40 et 30 dB. Elle montre clairement que l'algorithme a été en mesure de retrouver que l'ensemble des spectres se décompose sur les 8 premières colonnes de $\boldsymbol{S}$. L'algorithme fournit par ailleurs les coefficients d'abondance, qui pourraient si besoin être aisément débiaisés dans un second temps, étant donnée la connaissance ainsi acquise de leur support.

A titre de comparaison, la méthode de l'état de l'art [3] trouve sur les mêmes réalisations respectivement 5 et 3 spectres de composés purs. Elle repose sur le seuillage adaptatif de $\Delta\lambda_k = \hat{\lambda}_k - \lambda_k$ par $\tau_k$, où $\hat{\lambda}_k$ et $\lambda_k$ désignent la $k^e$ valeur propre de la matrice de corrélation et de covariance des spectres mesurés. La Figure 2 présente les barres d'erreur sur ces paramètres obtenues par Monte Carlo. La mise en œuvre de cet algorithme questionne l'efficacité de la chaîne de traitement à suivre, notamment celle de N-Findr supposé identifier les composés purs présents dans l'image.

## 5. Conclusion et perspectives

Cette communication propose un algorithme de démélange non-supervisé d'images hyperspectrales, sous contrainte de positivité et de somme unité. Son efficacité repose sur l'utilisation d'un opérateur proximal original : P-MiSTO. Une amélioration naturelle consisterait à régulariser spatialement les abondances estimées, en incorporant par exemple un terme de variation totale.

Une seconde possibilité permettant de distribuer l'optimisation consiste à partitionner les données spatialement et non par bande de fréquence. Dans ce cas $\boldsymbol{X}$ et $\boldsymbol{Y}$ sont partitionnés en $d$ blocs horizontaux conduisant à un problème d'optimisation pour lequel le terme de démélange est $\sum_{i=1}^{d} \|\boldsymbol{H}\boldsymbol{X}_i - \boldsymbol{Y}_i\|_{\mathrm{F}}^2$. Dans le contexte du démélange non-supervisé d'images hyperspectrales cela revient donc à réaliser un démélange local mais sur la base de l'ensemble des données. Une perspective intéressante serait de développer une approche itérative réalisant ce démélange uniquement sur la base des données disponibles au noeud $i$, suivi d'une étape de "commérage" avec d'autres noeuds visant à échanger les spectres des composés purs.

## Références


[1] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, Mar. 2004.

[2] S. Boyd, N. Parikh, and E. Chu. *Distributed optimization and statistical learning via the alternating direction method of multipliers*. Now Publishers Inc, Jan. 2011.

[3] C. Chang and Q. Du. Estimation of number of spectrally distinct signal sources in hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 42(3) :608 – 619, Mar. 2004.

[4] J. Eckstein and D. P. Bertsekas. On the Douglas—Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55(1) :293–318, 1992.

[5] E. Esser, M. Möller, S. Isher, and G. Sapiro. A Convex Model for Nonnegative Matrix Factorization and Dimensionality Reduction on Physical Space. *IEEE Transactions on Image Processing*, 21(7) :3239–3252, July 2012.

[6] M.-D. Iordache, J. M. Bioucas-Dias, and A. Plaza. Collaborative Sparse Regression for Hyperspectral Unmixing. *Geoscience and Remote Sensing, IEEE Transactions on*, (99) :1–14, Feb. 2013.

[7] N. Keshava and J. F. Mustard. Spectral unmixing. *IEEE Signal Processing Magazine*, 19(1) :44–57, 2002. ISSN 1053-5888.

[8] J. M. P. Nascimento and J. M. B. Dias. Vertex component analysis : A fast algorithm to unmix hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 43(4) :898–910, Apr. 2004.

[9] A. T. Puig, A. Wiesel, G. Fleury, and A. O. Hero. Multidimensional shrinkage-thresholding operator and Group LASSO penalties. *Signal Processing Letters, IEEE*, 18(6) :363–366, Apr. 2011.

[10] M. E. Winter. N-FINDR : an algorithm for fast autonomous spectral end-member determination in hyperspectral data : an algorithm for fast autonomous spectral end-member determination in hyperspectral data. *Proc. of SPIE : Imaging Spectrometry V*, 3753(10), 1999.