

Factorisation d’images spectrales de matériaux anciens

Alexandre AIELLO¹, Serge X. COHEN², Axel DAVY¹, Agnès DESOLNEUX¹, Guillaume LE MAILLOUX¹

¹Centre de Mathématiques et Leurs Applications, CNRS, ENS Paris-Saclay, France

²IPANEMA, CNRS, ministère de la Culture, UVSQ, USR3461, Université Paris-Saclay, Gif-sur-Yvette, France

alexandre.aiello@ens-paris-saclay.fr, serge.cohen@synchrotron-soleil.fr

axel.davy@cmla.ens-cachan.fr, agnes.desolneux@cmla.ens-cachan.fr, glemail@ens-paris-saclay.fr

Résumé – Dans cet article, nous analysons la composition de matériaux anciens via des images spectrales de fluorescence X collectées au synchrotron SOLEIL. Il s’agit de proposer des méthodes permettant de reconstituer à la fois les spectres des composés atomiques présents dans le matériau étudié et leur répartition spatiale. Nous étudions d’abord un problème d’optimisation avec plusieurs *a priori* avant de montrer que l’*a priori* naturel de positivité des spectres et des cartes d’activités apporte suffisamment de bonnes propriétés pour les identifier.

Abstract – In this paper, we analyze the composition of ancient materials through spectral images gathered at synchrotron SOLEIL. The proposed approach is based on positive factorization, in order to identify both the spectrum and the location of chemical elements. First, we study an optimisation problem with some sparsity prior and then we show that the natural positivity prior attached to both spectra and activity maps is enough to perform the factorization.

1 Introduction

La caractérisation des matériaux anciens en appelle souvent à l’exploitation de données d’imagerie spectrale, c’est-à-dire d’images U où à chaque pixel n est associé un spectre $\lambda \mapsto U(n; \lambda)$. Un exemple de telle image est donné sur la figure 1. Cette image spectrale a été obtenue par fluorescence des rayons X appliquée à un fossile de poisson téléost (spécimen MHNM-KK-OT 03a). Une description plus complète de ce fossile ainsi que de l’expérience ayant menée à l’obtention de ces données est publiée dans [1].

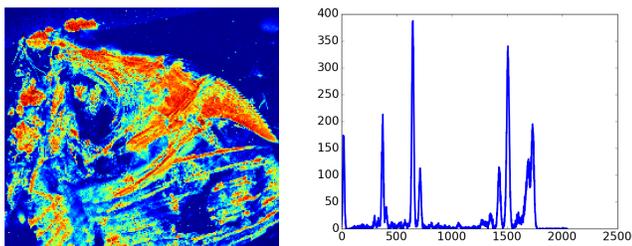


FIGURE 1 – Gauche : image correspondant à une longueur d’onde fixée $U(\cdot; 1500)$, soit 14.75 keV . Droite : spectre du pixel $(100, 100)$.

Le plus souvent il est physiquement pertinent de décomposer chaque spectre en une somme pondérée de *composantes pures* :

$$U(n; \lambda) = \sum_{k=1}^K C_k(n) S_k(\lambda) + \varepsilon(n; \lambda),$$

où $C_k(n)$ correspond à la carte d’activité du composant k au pixel n , $S_k(\lambda)$ est le *spectre pur* du composant k , et $\varepsilon(n; \lambda)$ est le terme d’erreur d’approximation due au bruit ou à l’inexactitude de la modélisation. Le but de cette étude est, par des techniques d’optimisation, de retrouver les cartes d’activité C_k et les spectres S_k . On introduira des contraintes correspondant aux connaissances *a priori* sur ces objets : les C_k sont des images positives et “régulières” par morceaux, et les S_k sont des spectres également positifs. Il est à noter que la notion de *spectre pur* ici n’est pas à prendre comme le spectre d’un composé chimiquement pur, mais plutôt par sa capacité à permettre une modélisation simple et utile des données mesurées. Et contrairement à [2] par exemple, il n’y a pas de contrainte de somme à 1 des activités C_k en un pixel.

Ce problème est relié à celui de la factorisation positive de matrices [3], ainsi qu’à des modèles de segmentation d’images spectrales. On le retrouve aussi dans des applications en imagerie médicale [4].

2 Un problème d’optimisation

Soit U une image spectrale donnée. On note Ω l’ensemble des pixels de l’image, et N leur nombre. On cherche K (nombre fixé) spectres « purs » représentés par la matrice $S = (S_k)_{1 \leq k \leq K}$ avec $\lambda \mapsto S_k(\lambda)$ spectre positif, et leurs cartes d’activités $C = (C_k)_{1 \leq k \leq K}$ avec $n \mapsto C_k(n)$ image positive, tels que la fonctionnelle

$$F(C, S) = \frac{1}{2} \|U - SC\|_2^2 + \beta \|\nabla C\|_1 + \alpha \|C\|_1 + \delta_{C \geq 0} + \delta_{S \geq 0}$$

soit minimale. Ici $\|\nabla C\|_1 = \sum_{k=1}^K \sum_{n=1}^N |\nabla C_k(n)|$, où l'on prend une norme isotrope sur le gradient en chaque pixel : $|\nabla C_k(n)| = \sqrt{(\partial_x C_k(n))^2 + (\partial_y C_k(n))^2}$, et δ vaut 0 si la condition est vérifiée et $+\infty$ sinon.

Le terme $\|U - SC\|_2^2$ est le terme d'attache aux données, alors que tous les autres sont des termes d'*a priori* sur la solution. Le coefficient β permet de rendre compte d'un *a priori* de régularité par morceaux des cartes d'activités. Le coefficient α pondère la parcimonie des données d'activité, là encore une norme L^1 semble adaptée en tant que « convexifiée » de la norme L^0 . Enfin, on force la positivité des cartes C_k et des spectres S_k .

Le choix d'une telle modélisation nous amène à utiliser dans la suite des outils d'optimisation convexe. On rappelle que la fonction duale convexe d'une fonction f est $f^*(y) = \sup_x \langle y, x \rangle - f(x)$. Si f est semi-continue inférieurement et convexe sur un espace de Hilbert, alors on a la formule de dualité : $f(x) = \sup_y \langle y, x \rangle - f^*(y)$. Dans notre cas, l'espace de Hilbert est l'ensemble des K images de taille N soit $H = (\mathbb{R}^N)^K$. Cette dualité permet de transformer un problème convexe en un problème dual qui s'avère en général plus adapté à des algorithmes d'optimisation basés sur des descentes de gradients (*cf* par exemple [5]).

3 Problème intermédiaire

Afin de mieux appréhender le rôle des coefficients α et β , une première étude intermédiaire a été menée. On suppose ici que les spectres S_k sont connus et nous cherchons à résoudre le problème d'optimisation convexe en la variable C . Le problème primal à résoudre est le suivant :

$$\text{Argmin}_C \frac{1}{2} \|U - SC\|_2^2 + \beta f(\nabla C) + \alpha g(C),$$

où $f(\nabla C) := \|\nabla C\|_1$, et $g(C) := \|C\|_1 + \delta_{C \geq 0}$. Les duales convexes de f et g sont données par : $f^*(p) = \delta_P(p)$ où $P = \{p \in H^2 \mid \|p_{k,n}\|_2 \leq 1\}$, et $g^*(v) = \delta_V(v)$ où $V = \{v \in H \mid v_{k,n} \leq 1\}$. Ainsi le problème primal peut se reformuler comme suit :

$$\begin{aligned} & \max_{p,v} \min_C \frac{1}{2} \|U - SC\|_2^2 + \langle \nabla^* p + v, C \rangle - \delta_{P \times V}(p/\beta, v/\alpha) \\ & = \frac{\|U\|_2^2}{2} - \min_{(p,v) \in \beta P \times \alpha V} \frac{1}{2} \|S^T U - (\nabla^* p + v)\|_{(S^T S)^{-1}}^2. \end{aligned}$$

Remarquons que l'adjoint de l'opérateur gradient vérifie $\nabla^* = -div$, et est donc facilement calculable. Puis l'inversion du min max en max min est justifiée d'après le théorème de dualité de Fenchel. On utilise également que $C \mapsto \frac{1}{2} \|U - SC\|_2^2 + \langle \nabla^* p + v, C \rangle$ atteint son minimum en

$$C = (S^T S)^{-1} (S^T U - (\nabla^* p + v)). \quad (1)$$

Cette relation permet de retrouver le minimiseur \hat{C} du problème primal à partir des minimiseurs duaux \hat{p} et \hat{v} . Or, le problème dual est un problème de minimisation quadratique contraint à un espace convexe, qui peut être résolu numériquement par une descente de gradient projeté. Une version accélérée est donnée par l'algorithme FISTA de Beck et Teboulle (2008), et que nous présentons ici. Notons $R = \beta P \times \alpha V$ l'espace convexe de contraintes sur lequel on désire minimiser $G(p, v) = \frac{1}{2} \|S^T U - (\nabla^* p + v)\|_{(S^T S)^{-1}}^2$.

Algorithm 1 (FISTA)

Initialisation : choisir $x_{-1} = x_0$, $t_0 \in \mathbb{R}$, $n_{it} \in \mathbb{N}$

```

for  $0 \leq k \leq n_{it}$  do
     $t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}$ 
     $y_k = x_k + \frac{t_k - 1}{t_{k+1}} (x_k - x_{k-1})$ 
     $x_{k+1} = \text{proj}_R(y_k - \tau \nabla G(y_k))$ 

```

end

return $x_{n_{it}+1}$

Des résultats de convergence sont établis dès que $\frac{2}{\tau}$ est supérieur à la constante L de Lipschitz de ∇G .

4 Tests sur données synthétiques

Pour tester l'efficacité de la méthode employée pour retrouver les cartes d'activités à partir des images spectrales, nous avons synthétisé 5 spectres S_k (avec 2048 canaux comme pour les données mesurées) comme combinaison linéaire de 2 Gaussiennes et Lorentziennes, voir Fig. 2. Le choix des Gaussiennes et Lorentziennes permet d'obtenir des spectres de même allure que ceux présents dans les données expérimentales. Puis on choisit une composition aléatoire et parcimonieuse pour chaque partie du poisson simulé (Tête, Queue, Arêtes, Oeil). Enfin on synthétise les données par échantillonnage de lois de Poisson.

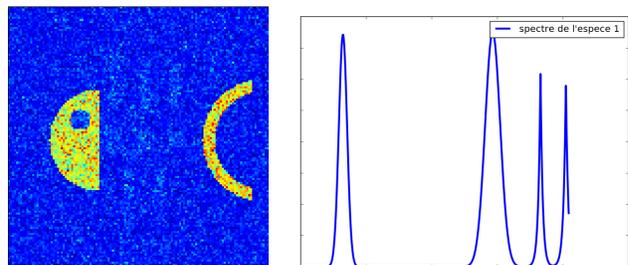


FIGURE 2 – Gauche : exemple d'image synthétique à une longueur d'onde fixée. Droite : exemple de spectre synthétique

À partir de ces données synthétiques et sur la base des spectres purs « vrais », nous pouvons observer les performances de la méthode et l'influence des paramètres de régularité β et de parcimonie α pour reconstituer les cartes

d'activités. On remarque que l'image d'activité estimée (Fig.3 centre) est d'une part très régulière comme attendu, et d'autre part les zones en bleu foncé sont des zones où l'espèce associée est absente. Ainsi on explique les données sans faire appel à cette espèce pour certaines zones : c'est la parcimonie. En revanche, de multiples expériences ont montré que le choix des paramètres α et β a relativement peu d'incidence sur le résultat. Ce qui amène naturellement à se demander : d'où vient le gain de performance ?

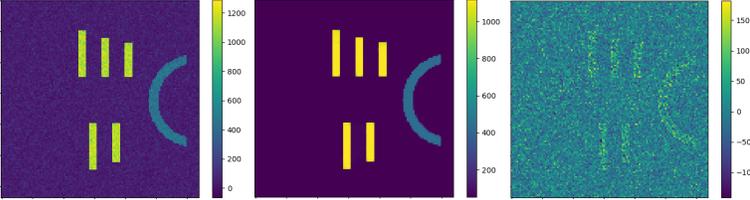


FIGURE 3 – Gauche : une carte d'activité synthétique bruitée. Centre : estimation de la même carte avec *a priori* de positivité, régularité et parcimonie. Droite : différence des deux cartes contenant essentiellement du bruit.

5 De l'importance de la positivité

Ces résultats sur les données synthétiques nous ont amenés à retravailler la manière dont on force la parcimonie et la régularité puisque les paramètres n'ont pas fait apparaître d'effet notable. En fait, on peut montrer via le principe de dualité qu'imposer la positivité va *naturellement* induire de la parcimonie. En effet, considérons le problème d'optimisation suivant avec S injective et U quelconque, pour lequel on calcule le problème dual :

$$\begin{aligned} \min_C \frac{1}{2} \|U - SC\|^2 + \delta_{C \geq 0} \quad (2) \\ = \frac{1}{2} \|U\|^2 - \min_V \frac{1}{2} \|S^T U - V\|_{(S^T S)^{-1}}^2 + \delta_{V \leq 0}. \end{aligned}$$

L'expression qui lie les variables duales et primales est semblable à (1), on obtient $\hat{C} = (S^T S)^{-1}(S^T U - \hat{V})$. Or en écrivant l'égalité des minima de (2) pour \hat{C} et \hat{V} , il vient, après calcul, la simple égalité : $\langle \hat{C}, \hat{V} \rangle = 0$. Dès lors, en utilisant le fait que \hat{C} (resp. \hat{V}) est à composantes positives (resp. négatives), on déduit que la somme $\sum_i \hat{C}_i \hat{V}_i = 0$ a tous ses termes nuls. De plus, \hat{V} est non nul sinon on aurait $\hat{C} = (S^T S)^{-1}(S^T U)$ qui est l'estimateur des moindres carrés et qui en pratique contient des coefficients négatifs, ce qui serait contradictoire. Ainsi pour toutes les composantes de \hat{V} strictement négative, la composante correspondante de \hat{C} est nulle. On comprend donc pourquoi le paramètre α des expériences n'avait que peu d'influence sur les résultats, la parcimonie étant déjà traitée naturellement par l'hypothèse de positivité.

6 Factorisation matricielle positive

L'étude du problème intermédiaire a permis de mettre en évidence l'importance de la contrainte de positivité. Ainsi, il semble justifié de simplifier le problème d'optimisation initialement proposé, en un problème biconvexe (convexe par rapport à chacun des arguments) dit de factorisation matricielle positive :

$$\text{Argmin}_{C, S} \frac{1}{2} \|U - SC\|_2^2 + \delta_{C \geq 0} + \delta_{S \geq 0}.$$

Dans le cas général, la factorisation matricielle positive est un problème NP-Hard. Néanmoins, une multitude de méthodes de résolution approchée ont été étudiées, parmi celles-ci la méthode la plus naturelle consiste à fixer une variable et à minimiser en l'autre variable, puis alternativement changer les rôles : il s'agit d'algorithmes de descentes alternées. Nous avons choisi cette méthode étant donné que chaque problème à une variable est, dans notre cas, un problème strictement convexe pour lequel il existe des descentes de gradient rapide. En effet, pour minimiser une fonction strictement convexe f dans un convexe K de \mathbb{R}^d , l'algorithme Heavy Ball permet de converger géométriquement vers un minimiseur (voir l'article [5] pour les preuves).

Algorithm 2 (Heavy Ball)

Calculer L (resp. l) la plus grande (resp. petite) valeur propre de $\nabla^2 f$ la hessienne de f .

Poser $\tau = \frac{4}{(\sqrt{L} + \sqrt{l})^2}$ et $\gamma = \left(\frac{1 - \sqrt{l}}{1 + \sqrt{l}} \right)^2$

Choisir une initialisation x_0

while Critère d'arrêt **do**

 | $x_{k+1} = \text{proj}_K(x_k - \tau \nabla f(x_k) + \gamma(x_k - x_{k-1}))$,

end

Bien que cet algorithme n'utilise pas la formulation duale du problème que l'on veut résoudre, il est utile d'utiliser la dualité de l'équation (2) afin d'établir un critère d'arrêt pertinent. En utilisant le primal dual gap (voir [5] pour une vision générale du primal dual gap), on montre le contrôle de l'erreur suivant, si $V = S^T(U - SC)$: $\|C - \hat{C}\|^2 \leq -\langle C, V \rangle + \delta_{C \geq 0} + \delta_{V \leq 0}$.

Le problème de factorisation matricielle positive étant non convexe, l'initialisation est importante. Plusieurs choix abondent dans la littérature, et pour le cas d'images hyperspectrales une initialisation utilisant l'algorithme Successive Projection Algorithm (SPA) est adaptée. En effet, en visualisant l'image hyperspectrale comme un nuage de points dans un espace de même dimension que les spectres, on désire expliquer ce nuage comme des combinaisons positives de certains points (associés à des spectres purs). Ainsi, il est naturel de chercher à estimer des points qui sont des sommets de l'enveloppe convexe du nuage. Pour plus de détails sur SPA on se référera à [4].

Algorithm 3 (Successive Projection Algorithm)

- E un nuage de N points dans \mathbb{R}^d
- p : nombre de sommets de l'enveloppe de E à estimer
- Sommets : liste des sommets, initialement vide

```
for  $1 \leq k \leq p$  do
   $x = \arg \max_{y \in E} \|y\|$ 
  Sommets = Sommets  $\cup \{x\}$ 
   $x = \frac{x}{\|x\|}$  et  $\forall y \in E, y = y - xx^T y$ 
end
```

L'algorithme SPA prend le point de plus grande norme puis projette le nuage sur l'espace orthogonal et répète cette procédure. Cet algorithme permet d'initialiser S_0 . Puis, étant donné les spectres initiaux on peut simplement initialiser les cartes d'activités comme l'approximation des moindres carrés seuillée : $C_0 = \max(0, (S_0^T S_0)^{-1} S_0^T U)$, où le maximum est pris sur les composantes.

7 Résultats et conclusions

Une implémentation de l'initialisation via SPA puis des descentes alternées via Heavy Ball ont permis d'obtenir les résultats ci-dessous. On remarque que les spectres finaux sont proches de ceux issus de l'initialisation en termes de support, mais que les intensités relatives sont fortement remodelées par la descente alternée. Par exemple pour le spectre 2, entre l'initialisation et le résultat final, il y a inversion des importances relatives des pics entre les canaux autour de 700 et ceux autour de 1600.

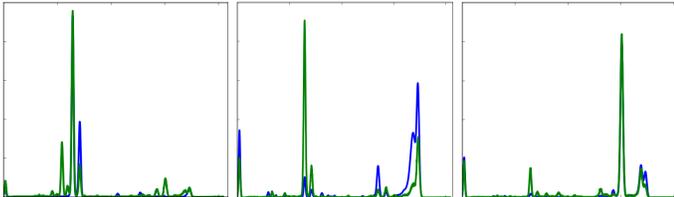


FIGURE 4 – Spectres initiaux SPA (bleu) et finaux (vert), on remarque que l'initialisation par SPA est proche en termes de support du résultat final.

En ce qui concerne les cartes d'activités associées, elles présentent de l'information additionnelle après traitement : ainsi, sur la composante 1, la descente alternée permet de montrer l'existence de structures très localisées et indépendantes en termes de localisation des composantes 2 et 3. On observe aussi que les images d'activités obtenues sont plus régulières et contrastées qu'initialement, la contrainte de positivité induisant en elle-même la parcimonie des activités et produisant des cartes bien plus lisibles.

Les perspectives et applications de cette étude sont nombreuses. Par exemple, des expériences intensives sur données synthétiques seront menées pour comprendre jusqu'à

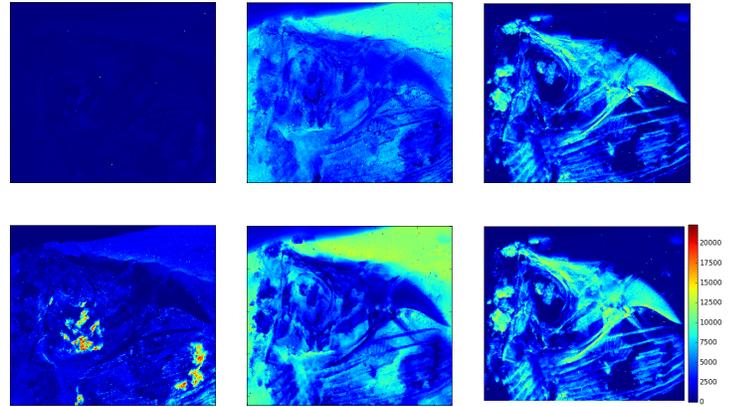


FIGURE 5 – Sur la première ligne : images initiales d'activités. Sur la deuxième : images finales d'activités.

quel niveau de bruit on peut aller tout en gardant de bons résultats. Ce point est particulièrement intéressant pour les matériaux anciens, car il permettrait de réduire le temps d'exposition de ces matériaux fragiles. Une autre piste, plus théorique, serait l'étude plus poussée du choix des paramètres α et β de parcimonie et régularité, en regardant par exemple des questions d'optimisation bi-niveau [6].

Références

- [1] P. Gueriau, C. Mocuta, D.B. Dutheil, S.X. Cohen, D. Thiaudiere, OT1 Consortium, S. Charbonnier, G. Clément and L. Bertrand, Trace Elemental Imaging of Rare Earth Elements Discriminates Tissues at Microscale in Flat Fossils, *PLoS One* 9(1), 2014.
- [2] R. Mhenni, S. Bourguignon, J. Ninin et F. Schmidt, Spectral Unmixing with sparsity and structuring constraints, HAL-02051443, 2019.
- [3] A. Cichoki, R. Zdunek, A. Huy Phan et S. Amari, *Nonnegative Matrix and Tensor Factorizations*, Wiley, 2009.
- [4] N. Sauwen, M. Acou, HN. Bharath, DM. Sima, J. Veraart, et al. , The successive projection algorithm as an initialization method for brain tumor segmentation using non-negative matrix factorization, *PLoS One* 12(8), 2017.
- [5] A. Chambolle and T. Pock, An introduction to continuous optimization for imaging, *Acta Numerica*, Vol. 25, p.161-319. 2016.
- [6] J. C. De Los Reyes, C.-B. Schönlieb, and T. Valkonen, Bilevel parameter learning for higher-order total variation regularisation models, *Journal of Mathematical Imaging and Vision*, 57(1), p.1-25, 2017.