

Restauration de données hyperspectrales astrophysiques par approximation parcimonieuse

Julien PICAUD¹, Sébastien BOURGUIGNON¹

¹LUNAM Université, École Centrale de Nantes / IRCCyN UMR CNRS 6597, 1 rue de la Noë, 44321 Nantes
Julien.Picaud@ircryn.ec-nantes.fr, Sebastien.Bourguignon@ircryn.ec-nantes.fr

Résumé – Nous envisageons la restauration de données hyperspectrales astrophysiques de grande dimension, dégradées par une réponse instrumentale à la fois spatiale et spectrale, dans un environnement très bruité. Le problème est contraint *via* une information de parcimonie au niveau spectral, exprimée dans un dictionnaire redondant de formes élémentaires. Le problème d’approximation parcimonieuse ainsi formulé est abordé par un algorithme *Orthogonal Matching Pursuit* (OMP). Afin d’optimiser le coût de calcul, nous exploitons les propriétés du modèle d’acquisition, en particulier la séparabilité de la réponse instrumentale en le produit d’une réponse spatiale par une réponse spectrale. Nous implantons également un codage spécifique des opérations impliquant le dictionnaire, permettant des calculs rapides pour les différents produits matriciels mis en jeu. L’efficacité de la mise en œuvre permet d’aborder la restauration conjointe spatiale et spectrale pour des problèmes de taille réaliste. Nous présentons un exemple de résultat sur une scène astrophysique simulée fortement bruitée, où la prise en compte du modèle d’observation spatial-spectral et d’un dictionnaire composé de formes appropriées permet de détecter des sources faibles.

Abstract – We consider the restoration of high-dimensional astrophysical hyperspectral data degraded by a spatial and spectral instrument response and a very strong noise level. The problem is constrained with sparsity prior information in the spectral domain, expressed in a redundant dictionary composed of elementary patterns. The sparse approximation problem is then tackled by the *Orthogonal Matching Pursuit* (OMP) algorithm. To optimize the computation cost, we exploit the properties of the data acquisition model, especially the separability of the instrument response into the product of a spatial response by a spectral response. We implement a specific encoding of the computations involving the dictionary, allowing fast computations for all corresponding matrix products. Implementation efficiency then enables joint spatial-spectral restoration for data with realistic size. Results are presented for a very noisy simulated astrophysical scene, where weak emitting sources are detected with such strategy.

1 Introduction

Nous considérons des observations de l’univers extragalactique en imagerie hyperspectrale, telles qu’elles seront produites par le spectro-imageur MUSE (*Multi-Unit Spectroscopic Explorer*), en construction pour le Very Large Telescope de l’Observatoire Européen Austral (ESO) au Chili [1]. Celles-ci se présentent sous la forme de cubes de 300×300 pixels (pour un champ de 1 arcmin²) et de 3 600 longueurs d’ondes dans le visible et proche infra-rouge (de 450 à 930 nm). Lors de l’acquisition de telles données, le passage de la lumière à travers l’atmosphère, le télescope et l’instrument produit des dégradations de type *étalement*, représentées par la PSF (*Point Spread Function*) de la chaîne d’acquisition. Les mesures s’effectuent par ailleurs dans un milieu très bruité, en partie dû à un signal parasite lié à l’atmosphère et au fond de ciel [2]. Afin de contraindre le problème de restauration, une approche introduisant une information *a priori* de parcimonie a été proposée [3], prenant en compte une connaissance physique des objets dans leur dimension spectrale au moyen d’un dictionnaire de spectres élémentaires. Si ce dictionnaire s’est révélé efficace en termes de modélisation pour la restauration de spectres considérés individuellement, les algorithmes d’optimisation proposés perme-

ttent difficilement le passage à l’échelle sur des problèmes incluant une dimension spatiale. Cette communication propose la mise en œuvre d’une approche de type *Orthogonal Matching Pursuit* (OMP) pour aborder ce problème [4], où les propriétés du modèle d’acquisition – en particulier, la séparabilité de la PSF en le produit d’une réponse spatiale par une réponse spectrale – et une optimisation spécifique des calculs impliquant le dictionnaire permettent d’envisager la restauration de cubes de taille réaliste. La Section 2 établit un modèle d’observation et discute en particulier la séparabilité de la PSF, puis construit le modèle d’approximation parcimonieuse, basé sur le dictionnaire introduit dans [3]. La mise en œuvre des différents types de calculs intervenant dans l’algorithme OMP est ensuite détaillée en Section 3. La dernière section présente des résultats obtenus sur une scène astrophysique simulée et discute l’évolution du coût de calcul en fonction de la taille des données.

2 Restauration jointe spatiale-spectrale et parcimonie spectrale

Un modèle direct pour la reconstruction peut être écrit sous la forme linéaire classique $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{b}$ où \mathbf{y} est le cube ob-

servé, \mathbf{H} est un opérateur linéaire représentant la PSF de l'instrument, \mathbf{x} est un cube que l'on cherche à reconstruire et \mathbf{b} représente le bruit et les erreurs de modélisation. La PSF est la réponse de l'instrument à une impulsion spatiale et spectrale. C'est une fonction tri-dimensionnelle définie en tout point du cube. La thèse d'Emma Villeneuve [2] étudie les propriétés de la PSF de l'instrument MUSE, dont nous reprenons ici certaines hypothèses. En particulier, nous la supposons *séparable* en un produit d'une réponse spatiale (FSF, *Field Spread Function*) par une réponse spectrale (LSF, *Line Spread Function*). La séparabilité est une hypothèse fondamentale dans notre cas, puisqu'elle réduit le calcul du modèle $\mathbf{H}\mathbf{x}$ et les calculs de type $\mathbf{H}^T\mathbf{y}$ (où la notation T désigne la transposition matricielle) à l'application d'un ensemble de convolutions spatiales et de convolutions spectrales.

Les dégradations instrumentales augmentent avec la longueur d'onde. Dans ce travail, suivant [2], la FSF est modélisée par une fonction de Moffat s'étalant sur 13×13 pixels et la LSF par une gaussienne sur 11 éléments spectraux où, dans chaque cas, les paramètres varient en fonction de la longueur d'onde [2]. Notons que cette variabilité spectrale reste compatible avec l'hypothèse de séparabilité : si $\text{PSF}_{\mathbf{r}_0, \lambda_0}(\mathbf{r}, \lambda)$ est la fonction d'étalement de l'instrument au pixel de position spatiale \mathbf{r}_0 et à la longueur d'onde λ_0 et de variables spatiales \mathbf{r} et spectrale λ , on a : $\text{PSF}_{\mathbf{r}_0, \lambda_0}(\mathbf{r}, \lambda) = \text{FSF}_{\lambda_0}(\mathbf{r}) \text{LSF}_{\lambda_0}(\lambda)$.

Le mauvais conditionnement de \mathbf{H} et la présence d'un bruit de niveau élevé requièrent la prise en compte de contraintes sur la solution recherchée [5]. Il a été proposé dans [3] une *approximation parcimonieuse* des spectres \mathbf{x}_k composant le cube, synthétisés à partir d'un dictionnaire \mathbf{W} de formes élémentaires : $\mathbf{x}_k = \mathbf{W}\mathbf{u}_k$, avec \mathbf{u}_k parcimonieux. Ce dictionnaire a été construit à partir de connaissances physiques sur les spectres recherchés. Les spectres sont ainsi synthétisés à partir de raies de douze différentes largeurs possibles, centrées le long de l'axe spectral (correspondant à des raies en émission ou en absorption, de largeur et de longueur d'onde centrale inconnues), d'échelons centrés le long de l'axe spectral (représentant des décrochages dans le spectre) et d'oscillations sinusoïdales paramétrant une composante basse fréquence. Plus de détails sur la composition du dictionnaire peuvent être trouvés dans [3]. La Figure 1 montre une représentation de ce dictionnaire en dimension réduite. Pour 3 600 longueurs d'onde, le dictionnaire, très redondant, comporte environ 26 000 atomes.

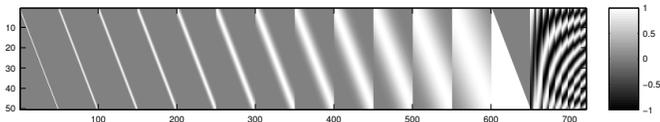


FIGURE 1 – Dictionnaire de spectres élémentaires (en dimension réduite à 50 longueurs d'onde). Celui-ci est composé de blocs correspondant à des raies de différentes largeurs, d'échelons et d'un ensemble d'oscillations basse fréquence.

Notons qu'un tel dictionnaire est extrêmement *cohérent*, *i.e.*, ses colonnes sont très corrélées. C'est notamment le cas de

raies de largeur différente et centrées au même point, ou de deux échelons voisins. De manière similaire, les raies de taille unitaire, formant le premier bloc du dictionnaire de la Figure 1, s'obtiennent par la différence de deux échelons voisins (la contrainte de parcimonie permet cependant de discriminer les deux modèles). Il a été constaté dans [3] que cette grande cohérence est responsable de la lenteur pathologique des algorithmes classiques d'optimisation en norme ℓ_1 de type gradient projeté [4]. Il nous semble en revanche plus judicieux de choisir ce modèle construit à partir de connaissances précises sur les spectres recherchés, plutôt que d'envisager une représentation parcimonieuse basée sur des transformées génériques de type ondelettes, certes moins cohérentes. Les atomes de notre dictionnaire étant par construction mieux corrélés aux formes attendues, nous pouvons envisager de meilleures performances en termes de détection, point crucial dans un milieu aussi bruité que celui des données de MUSE. Ce choix s'effectue cependant au détriment du coût de calcul. Nous verrons néanmoins en Section 3 comment d'exploiter la structure de ce dictionnaire particulier pour réaliser efficacement les calculs nécessaires.

Pour un cube hyperspectral, vu comme un ensemble de spectres, nous pouvons finalement former le modèle :

$$\mathbf{y} = \mathbf{H}\overline{\mathbf{W}}\mathbf{u} + \mathbf{b}, \text{ où :} \quad (1)$$

- \mathbf{y} est un vecteur colonne représentant le cube, ordonné comme la concaténation des différents spectres \mathbf{y}_k , $k = 1, \dots, K$ où K est le nombre de pixels dans les images ;
- \mathbf{u} est un vecteur colonne concaténant les coefficients inconnus \mathbf{u}_k associés à la décomposition du spectre \mathbf{x}_k ;
- $\overline{\mathbf{W}}$ est une matrice bloc-Toeplitz composée d'autant de blocs \mathbf{W} que de pixels dans le cube d'inconnues.

Nous cherchons donc à estimer un vecteur parcimonieux $\hat{\mathbf{u}}$ à partir des données \mathbf{y} . Notons que cette approche vise à obtenir un cube $\overline{\mathbf{W}}\hat{\mathbf{u}}$ restauré à la fois des dégradations spatiales et spectrales de l'instrument, au moyen d'une forte information *a priori* spectrale, mais sans introduire d'hypothèse sur la nature spatiale des objets – mise à part la présence de nombreux zéros dans les images, en raison de la nature creuse du dictionnaire.

3 Calculs rapides exploitant la structure du modèle

La séparabilité de la PSF permet encore d'écrire $\mathbf{H} = \mathbf{F}\mathbf{L}$, où \mathbf{F} représente la FSF et \mathbf{L} la LSF. Par la suite, nous notons $\mathbf{A} = \mathbf{F}\mathbf{L}\overline{\mathbf{W}}$. Le modèle (1) s'écrit alors $\mathbf{y} = \mathbf{A}\mathbf{u} + \mathbf{b}$. De nombreuses approches algorithmiques pour l'approximation parcimonieuse (voir par exemple [4]) requièrent des calculs de type $\mathbf{A}\mathbf{u}^{(t)}$ – où $\mathbf{u}^{(t)}$ représente le jeu de coefficients à une itération t de l'algorithme, généralement parcimonieux – et des produits de type $\mathbf{A}^T\mathbf{z}^{(t)}$, où $\mathbf{z}^{(t)}$ est de la taille des données. Nous proposons ici une mise en œuvre efficace de ces produits, permettant l'utilisation de l'algorithme OMP [4]. Chaque itération de cet algorithme peut en effet être décomposée en trois parties :

- calcul du résidu $\mathbf{e}^{(t)} = \mathbf{y} - \mathbf{A}\mathbf{u}^{(t)}$;

- sélection de l'atome qui réduit le plus le résidu énergétiquement, qui correspond à l'indice maximal dans le produit $|\mathbf{A}^T \mathbf{e}^{(t)}|^1$;
- estimation des amplitudes sur le support sélectionné par projection orthogonale.

Dans notre mise en œuvre, l'algorithme est arrêté lorsque le résidu est statistiquement au niveau du bruit, supposé connu. Notons que la plupart des étapes ci-dessous correspondent à un ensemble de calculs indépendants portant sur les différents spectres ou sur les différentes images d'un cube et sont donc fortement parallélisables.

3.1 Calculs de produits de type $\mathbf{A}\mathbf{u}$

Un produit $\mathbf{A}\mathbf{u}$ s'effectue d'abord par la synthèse de chaque spectre $\mathbf{W}\mathbf{u}_k$, auquel est appliquée la LSF. Celle-ci étant identique pour tout pixel k , nous pouvons considérer le dictionnaire équivalent \mathbf{W}_{LSF} , dont les atomes sont obtenus en appliquant la LSF aux atomes de \mathbf{W} . Ce dictionnaire est donc calculé préalablement et stocké. Les produits $\mathbf{W}_{\text{LSF}}\mathbf{u}_k$ sont alors calculés en exploitant la parcimonie de \mathbf{W}_{LSF} et celle de \mathbf{u}_k . L'application de \mathbf{F} s'effectue ensuite par une série de convolutions spatiales indépendantes – la FSF étant différente à chaque longueur d'onde – appliquées sur chaque image du cube $\mathbf{L}\mathbf{W}\mathbf{u}$, formé par l'ensemble des spectres $\mathbf{W}_{\text{LSF}}\mathbf{u}_k$.

3.2 Calculs de produits de type $\mathbf{A}^T \mathbf{z}$

Le calcul d'un produit $\mathbf{A}^T \mathbf{z} = \overline{\mathbf{W}}^T \mathbf{L}^T \mathbf{F}^T \mathbf{z}$ commence par une série de convolutions spatiales sur chaque image contenue dans \mathbf{z} . Les calculs suivants s'effectuent indépendamment sur chaque spectre du cube spatialement convolué $\mathbf{F}^T \mathbf{z}$. Il est à noter que ce dernier ne présente pas de structure parcimonieuse. Dans notre mise en œuvre, nous appliquons donc d'abord la LSF sur chaque spectre. Restent alors à calculer des produits de type $\mathbf{W}^T \mathbf{x}$, qui exploitent la forme des différents blocs du dictionnaire (cf. Figure 1) :

- chaque bloc $\mathbf{W}_i, i = 1, \dots, 12$ correspondant à une raie de largeur donnée est de Toeplitz ; le produit $\mathbf{W}_i^T \mathbf{x}$ peut donc être calculé comme la convolution discrète de \mathbf{x} par le motif de la raie correspondante.
- Pour le bloc d'échelons \mathbf{W}_{ech} , on montre que la $n^{\text{ème}}$ composante du produit $\mathbf{W}_{\text{ech}}^T \mathbf{x}$ est la somme des n derniers termes de \mathbf{x} . Le calcul est donc celui d'une somme cumulée.
- Le bloc de sinusöides \mathbf{W}_{sin} étant de petite taille, le produit $\mathbf{W}_{\text{sin}}^T \mathbf{x}$ correspondant est calculé matriciellement.

Le produit $\mathbf{W}^T \mathbf{x}$ s'obtient en concaténant l'ensemble des produits ci-dessus. Cette mise en œuvre s'est révélée plus efficace qu'un produit matriciel direct exploitant un codage creux du dictionnaire \mathbf{W}_{LSF} . En effet, à la différence des produits directs $\mathbf{W}_{\text{LSF}}\mathbf{u}$ décrits au § 3.1, ici les spectres synthétiques \mathbf{x} ne sont en général pas aussi parcimonieux que les coefficients \mathbf{u} .

3.3 Projection orthogonale

L'algorithme OMP estime par moindres carrés les amplitudes des éléments sélectionnés à chaque itération, ce qui s'écrit :

$$\hat{\mathbf{u}}_* = \arg \min_{\mathbf{u}_*} \|\mathbf{y} - \mathbf{A}_* \mathbf{u}_*\|^2 = (\mathbf{A}_*^T \mathbf{A}_*)^{-1} \mathbf{A}_*^T \mathbf{y},$$

où la notation $*$ indexe les éléments non nuls de \mathbf{u} et les colonnes correspondantes de \mathbf{A} . Le produit $\mathbf{A}^T \mathbf{y}$, duquel est extrait $\mathbf{A}_*^T \mathbf{y}$, est calculé préalablement comme expliqué au § 3.2. Dans notre mise en œuvre, nous mettons à jour récursivement la matrice $(\mathbf{A}_*^T \mathbf{A}_*)^{-1}$, dont la taille augmente d'une ligne et d'une colonne par itération : à un réordonnement près des colonnes, on a $\mathbf{A}_*^{(t)} = [\mathbf{A}_*^{(t-1)} \mathbf{a}_t]$, en notant $\mathbf{A}_*^{(t-1)}$ l'ensemble des colonnes sélectionnées jusqu'à l'itération $t-1$ et \mathbf{a}_t le nouvel atome sélectionné à l'itération t . Alors (avec $\mathbf{a}_t^T \mathbf{a}_t = 1$) :

$$\mathbf{A}_*^{(t)T} \mathbf{A}_*^{(t)} = \begin{pmatrix} \mathbf{A}_*^{(t-1)T} \mathbf{A}_*^{(t-1)} & \mathbf{d} \\ \mathbf{d}^T & 1 \end{pmatrix}$$

où $\mathbf{d} = \mathbf{A}_*^{(t-1)T} \mathbf{a}_t$. Par la formule d'inversion des matrices partitionnées, on a :

$$(\mathbf{A}_*^{(t)T} \mathbf{A}_*^{(t)})^{-1} = \frac{1}{\Delta} \begin{pmatrix} \Delta (\mathbf{A}_*^{(t-1)T} \mathbf{A}_*^{(t-1)})^{-1} + \mathbf{c}^T \mathbf{c} & -\mathbf{c}^T \\ -\mathbf{c} & 1 \end{pmatrix}$$

où $\Delta = 1 - \mathbf{d}^T (\mathbf{A}_*^{(t-1)T} \mathbf{A}_*^{(t-1)})^{-1} \mathbf{d}$ et $\mathbf{c} = \mathbf{d}^T (\mathbf{A}_*^{(t-1)T} \mathbf{A}_*^{(t-1)})^{-1}$.

La mise à jour de $(\mathbf{A}_*^T \mathbf{A}_*)^{-1}$ requiert donc essentiellement :

- le calcul explicite de la nouvelle colonne \mathbf{a}_t . Celui-ci est effectué comme au § 3.1, avec pour \mathbf{u} le vecteur de la base canonique correspondant à la coordonnée sélectionnée ;
- le calcul de \mathbf{d} . Dans la version développée, nous stockons (en format `sparse`) l'ensemble des colonnes sélectionnées \mathbf{A}_* . Cette étape correspond donc à un produit matrice-vecteur opérant en grande dimension sur des entités parcimonieuses. Il serait également possible de calculer \mathbf{d} similairement aux calculs du § 3.2 sans stocker \mathbf{A}_* , mais cette version s'est révélée plus coûteuse ici.

4 Résultats obtenus par un algorithme Orthogonal Matching Pursuit

Nous évaluons notre approche sur une scène composée de 41×41 pixels et 3 600 longueurs d'onde, simulée par les astronomes du projet MUSE [1]. Cette scène comprend notamment, au centre de l'image, deux sources noyées dans du bruit, non résolues avant convolution (*i.e.*, de taille inférieure à celle du pixel) et quasi-monochromatiques (*i.e.*, dont le spectre est composé presque exclusivement d'une raie). La Figure 2 présente une partie de ces données, déjà convoluées, sans bruit et avec bruit (respectivement, à gauche et au centre pour chaque objet). Il est clair que l'exploitation indépendante des images ou des spectres ne permet pas de détecter les deux objets faibles. En revanche, la prise en compte d'un modèle instrumental spatial-spectral (notamment de la FSF, qui provoque un étalement spa-

1. Nous supposons que les colonnes du dictionnaire sont normalisées.

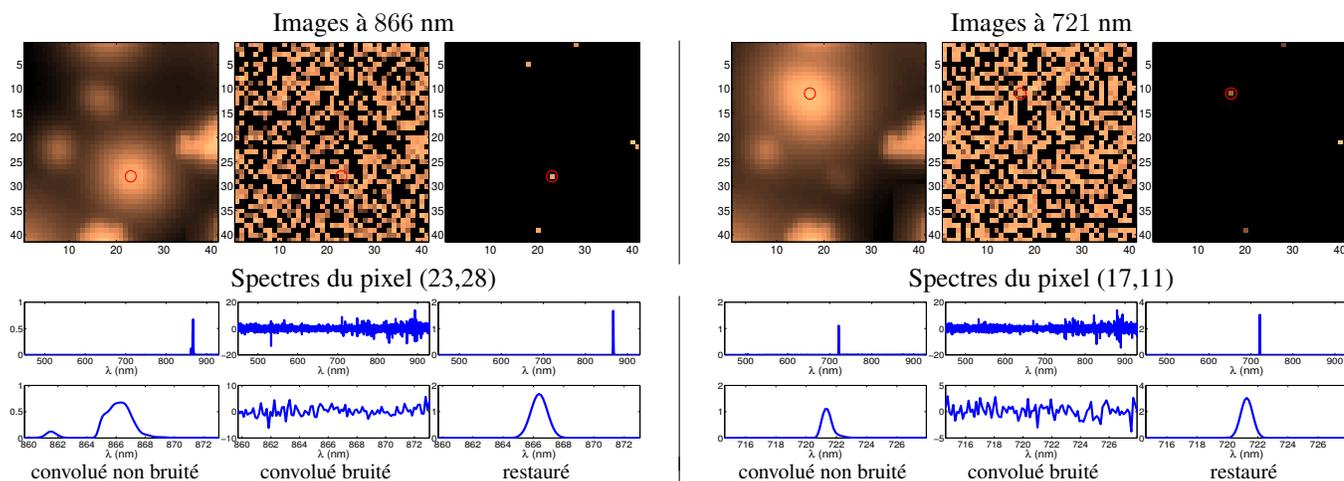


FIGURE 2 – Extraits de cubes hyperspectraux, focalisés sur les deux objets d’intérêt (respectivement, gauche et droite). Sont représentés l’image à la longueur d’onde centrale de l’objet et le spectre à la position spatiale de l’objet (repérée par un cercle sur les images). Dans chaque cas, l’intégralité du spectre et un zoom autour de la longueur d’onde d’intérêt sont représentés (respectivement en haut et en bas). Les amplitudes dans les images sont en échelle logarithmique, les pixels négatifs étant ramenés à 0. Pour chaque objet, données convoluées non bruitées (à gauche), convoluées bruitées (au centre) et reconstruites (à droite).

tial considérable de l’information), permet de reconcentrer l’information utile et de détecter les deux sources faibles : la Figure 2 (à droite pour chaque objet) montre les images et les spectres ainsi obtenus, où pour chaque objet, seul un pixel et un atome du dictionnaire spectral ont été détectés. Notons que ces détectations s’obtiennent au prix de fausses détectations, comme on peut le voir en haut de l’image à 866 nm. Celles-ci ne sont pas surprenantes puisque le flux des objets considérés est très faible, en limite de détection par rapport au bruit. La restauration de cette scène représente un problème à $41 \times 41 \times 3\,600$ données et environ $53 \times 53 \times 26\,000$ inconnues avec les hypothèses de bord considérées. Au final, seulement 309 atomes sont sélectionnés, requérant autant d’itérations de l’algorithme OMP. La mise en œuvre sous Matlab, sur un ordinateur portable possédant quatre processeurs à 2.7 GHz et 8 Go de mémoire vive, représente environ trois heures de calcul.

Nous étudions finalement la ventilation du coût de calcul sur chaque étape de l’algorithme OMP. Pour cela, nous avons considéré plusieurs scènes, de taille variant de 11×11 à 51×51 pixels. La Figure 3 représente l’évolution du coût calculatoire. Sans surprise, la partie la plus coûteuse est le calcul des produits de type $\mathbf{A}^T \mathbf{z}$ (les moins parcimonieux), les produits de forme $\mathbf{A} \mathbf{u}$ représentant seulement 10 % du coût total. Notons que le coût de l’inversion, dominé par les calculs de type $\mathbf{A}_*^T \mathbf{a}_i$ devient aussi important lorsque le nombre d’atomes augmente, laissant penser que pour des gros problèmes une version calculant ces produits sans stockage de la matrice \mathbf{A}_* (voir le § 3.3) sera plus efficace. L’application à des scènes spatiales plus grandes, et à des données réelles probablement plus riches, augmentera certes le coût de calcul. Ce dernier doit cependant être évalué en regard de celui de l’acquisition, l’obtention d’un seul cube de données par l’instrument MUSE résultant d’observations collectées sur plusieurs mois.

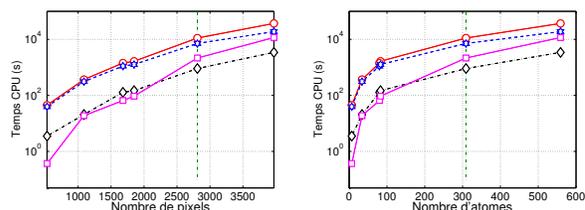


FIGURE 3 – Coût calculatoire (échelle logarithmique) des différentes étapes de l’algorithme OMP en fonction du nombre de pixels dans le cube restauré (gauche) et du nombre d’atomes dans l’approximation (droite). Coût total (\circ), coût des produits $\mathbf{A} \mathbf{u}$ (\diamond) et $\mathbf{A}^T \mathbf{x}$ (\star) et des projections orthogonales (\square).

Ce travail a été en partie financé par le projet ANR 08-BLAN-0253-01 DAHLIA, Données Astronomiques Hyperspectrales : algorithmes Avancés.

Références

- [1] R. Bacon *et al.*, “Probing unexplored territories with MUSE : a second generation instrument for the VLT,” in *Proc. SPIE 6269, Ground-based and Airborne Instrumentation for Astronomy*, pp. 62690J–62690J–9, juin 2006.
- [2] E. Villeneuve, *Déconvolution de données hyperspectrales pour l’instrument MUSE du VLT*. PhD thesis, Université de Toulouse, 2012.
- [3] S. Bourguignon, D. Mary, and É. Slezak, “Restoration of astrophysical spectra with sparsity constraints : Models and algorithms,” *IEEE J. Sel. Topics Signal Processing*, vol. 5, pp. 1002–1013, sept. 2011.
- [4] J. A. Tropp and S. Wright, “Computational methods for sparse solution of linear inverse problems,” *Proceedings of the IEEE*, vol. 98, pp. 948–958, juin 2010.
- [5] J. Idier, ed., *Approche bayésienne pour les problèmes inverses*. Paris : Traité IC2, Série traitement du signal et de l’image, Hermès, nov. 2001.