

Quelques Applications du Compressed Sensing en Astronomie

David MARY¹, Olivier J. J. MICHEL²,

¹Astronomisches Rechen Institut, Moenchofstrasse 12-14, 69120 Heidelberg, Allemagne

²Laboratoire Universitaire d'Astrophysique de Nice, UMR 6525-CNRS,
Parc Valrose, Université de Nice Sophia Antipolis, 06000 NICE

dmary@ari.uni-heidelberg.de
olivier.michel@unice.fr

Résumé – Nous proposons dans cette communication d'évaluer le potentiel des méthodes de « Compressed Sensing » récemment introduites, à travers leur application à quelques problèmes importants en astrophysique observationnelle. Après avoir rapidement décrit les bases mathématiques de ces approches, des exemples sont développés dans la cadre de la variabilité stellaire, de la reconstruction d'images satellitaire et enfin dans le cadre plus prospectif des futures possibilités offertes par les grands projets d'interféromètres à multiples bases pour l'imagerie directe dans le plan de Fourier.

Abstract – We investigate in this communication how recently introduced “Compressed Sensing” methods can be applied to some important problems in observational Astronomy. The mathematical background is first outlined. Some examples are then described in stellar variability and in image reconstruction for space-based observations. We finally illustrate the interest of such techniques for direct imaging through stellar interferometry.

1 Introduction

Enregistrer et compresser un signal requiert classiquement de l'échantillonner à la fréquence de Nyquist, d'en calculer la représentation dans une certaine base (transformation), et de retenir les coefficients les plus significatifs après transformation. Ce processus conduit à un double gaspillage : en terme d'échantillonnage (on fait l'acquisition de tout le signal) et en terme de complexité (on traite tout le signal pour n'en retenir que quelques coefficients). Les travaux récents [3, 4, 8] autour du Compressed Sensing (CS, qu'on pourrait traduire par « Acquisition compressée de données ») permettent de lever ces désavantages en modifiant le cadre de l'acquisition des données¹. L'approche exploite le fait que les signaux ayant une représentation éparsée dans une certaine base Ψ (le cas de la plupart des signaux physiques) peuvent être reconstruits à partir d'un nombre limité de projections sur une deuxième base Φ , incohérente avec la première. La base Φ est dite incohérente avec Ψ si aucune de ses lignes n'admet de représentation éparsée sur la base Ψ ². De façon très intéressante, les projections aléatoires constituent des bases incohérentes avec toutes les bases « fixes » (cas par exemple où Ψ est la matrice Identité, la matrice de la transformée de Fourier,...).

Considérons un signal x représenté par ses échantillons \underline{x} (\underline{x} est un vecteur $N \times 1$) et une base Ψ . Ψ est une matrice $N \times N$ avec $\Psi = [\Psi_1 \dots \Psi_N]$. Le signal x a une représentation éparsée dans une base Ψ si chaque vecteur \underline{x} peut être approximé par $\underline{x} \approx \sum_{i=1}^K \theta_i \Psi_i$, où les θ_i sont les coefficients de \underline{x} dans Ψ , et où $K \ll N$.

Considérons aussi une matrice $M \times N$ dite « de mesure », Φ , avec $M < N$ et dont les lignes sont incohérentes avec

les colonnes de Ψ . Comme mentionné ci-dessus, les coefficients de Φ peuvent être par exemple des variables aléatoires indépendantes et identiquement distribuées, des 0 et des 1 aléatoirement répartis, etc. Les mesures \underline{y} sont obtenues par $\underline{y} = \Phi \underline{x}$, où \underline{y} est $M \times 1$ et $M < N$ (donc de dimension réduite par rapport à \underline{x}). Les travaux ci-dessus [8] montrent que pour un vecteur \underline{x} donné, il existe un facteur $c > 1$ tel que seulement cK mesures de \underline{x} sont nécessaires pour reconstruire \underline{x} presque sûrement. En d'autres termes, cK mesures sont suffisantes pour capter toute l'information contenue dans le signal \underline{x} . Les applications sont vastes puisque de nombreux signaux « naturels » ont une représentation éparsée. Pour imposer que les signaux reconstruits soient éparsés, une approche naturelle est d'imposer que le support (ou le nombre de composantes non nulles, ou encore la norme L_0) des signaux reconstruits soit minimal. Malheureusement, cette contrainte mène à des algorithmes NP complets, dont la complexité exponentielle rend l'usage impraticable pour des valeurs réalistes de N [9]. Cependant, la contrainte sur L_0 peut être relaxée en imposant que la norme L_1 des signaux reconstruits soit minimale³. Pour des signaux K -éparsés et sous certaines conditions sur Φ et Ψ , on montre que les solutions obtenues en minimisant les normes L_1 et L_0 sont les mêmes (« relaxation convexe » [7]). La reconstruction de \underline{x} est obtenue à travers des méthodes d'optimisation sous contraintes, qui admettent des solutions en programmation linéaire. Nous imposons donc que la solution recherchée soit à la fois éparsée et en accord avec les données, et nous recherchons un ensemble de coefficients $\underline{\theta}^* = [\theta_1^* \theta_2^* \dots \theta_M^*]^t$ tels que

$$\underline{\theta}^* = \operatorname{argmin} \|\underline{\theta}\|_1, \text{ sous la contrainte } \Phi \Psi \underline{\theta}^* = \underline{y}, \quad (1)$$

et le signal reconstruit s'exprime alors $\hat{\underline{x}} = \Psi \underline{\theta}^*$.

³La norme L_p de $\underline{x} = [x_1 \dots x_N]^t$ est $\|\underline{x}\|_p = [\sum_{i=1}^N |x_i|^p]^{\frac{1}{p}}$.

¹cf <http://www.dsp.ece.rice.edu/cs/> pour références et algorithmes permettant une recherche reproductible.

²Le degré de cohérence entre deux bases Ψ_1 et Ψ_2 peut être défini comme $\max_{i \neq j} |\Psi_{1i} \cdot \Psi_{2j}|$ [7].

La Sec. 2 formalise sans entrer dans les détails le problème tel qu'il peut s'appliquer à l'étude de variabilité stellaire. Les avantages que peut apporter le CS pour la reconstruction de certaines images satellitaires (à travers l'exemple de Gaia) sont illustrés en Sec. 3. La dernière section se penche sur les aspects liés à la reconstruction d'images interférométriques, avec des mesures dans le plan de Fourier.

Il faut souligner que des méthodes prenant en compte de façon réaliste la complexité des données astrophysiques nécessitent clairement un traitement plus approfondi que les résultats présentés ici. Ce papier constitue essentiellement une brève synthèse de résultats existants dans le domaine, et un aperçu de leur potentiel pour des applications en astronomie. Pour les applications présentées il s'agit donc surtout d'une démonstration de principe. Les algorithmes de reconstruction utilisés utilisent des méthodes de programmation linéaire implémentées dans l'outil `ll-magic` disponible à <http://www.dsp.ece.rice.edu/cs/>.

2 CS et variabilité stellaire

Le problème est dans ce cas de détecter si un signal échantillonné irrégulièrement dans le temps contient ou non un signal périodique. Pour des observations au sol, l'échantillonnage irrégulier est dû par exemple à des intempéries, ou à des contraintes liées aux coordonnées géographiques des sites d'observations ne permettant pas une couverture temporelle continue. Pour des observations spatiales, l'étoile à observer peut passer dans le champ du télescope à des dates aléatoires (ce sera par exemple le cas d'un satellite comme Gaia [10]). Ce problème peut-être reformulé dans le cadre du CS : les inconnues \underline{x} sont dans ce cas le signal périodique $\underline{x} = [x_1 \dots x_N]^t$ obtenu par un échantillonnage régulier à des dates $t, t + \Delta t, \dots, t + N\Delta t$, où la résolution temporelle Δt est arbitrairement élevée. Ce signal étant périodique, il admet une représentation éparse, par exemple dans la base de Fourier (Ψ est dans ce cas la matrice de transformation de Fourier inverse). Les données \underline{y} (les « courbes de lumière ») peuvent être vues comme un échantillonnage de \underline{x} (plus du bruit) correspondant aux dates auxquelles les observations ont été effectuées, i.e. à un sous-ensemble de M dates $t + n_1\Delta t, \dots, t + n_M\Delta t$. Dans ce cas $\underline{y} = \Phi\underline{x}$, où Φ est une matrice de 0 et de 1 avec $\Phi_{i,j} = \delta_{j,n_i}$.

Pour la reconstruction, on recherche alors un ensemble de coefficients $\underline{\theta}^*$ dans l'espace de Fourier ayant une norme L_1 minimale, sous la contrainte $\Phi\Psi\underline{\theta}^* = \underline{y}$ (ou $|\Phi\Psi\underline{\theta}^* - \underline{y}|_2 < b$). Cette application ne sera pas plus détaillée ici ; mentionnons seulement qu'il est intéressant de resituer des travaux développés en analyse de variabilité stellaire (par ex. [14]) par rapport au cadre du CS : en effet, ces travaux s'appuient sur un algorithme de Matching Pursuit, qui est une méthode exploitant le caractère épars des signaux.

3 Reconstruction d'images pour le satellite Gaia

Gaia est un satellite qui fera une cartographie complète des objets célestes jusqu'à des objets très peu brillants (\approx magnitude 20 dans le visible⁴), en 3D et avec une précision astrométrique jamais atteinte (lancement fin 2011). Le

⁴Sirius, l'étoile la plus brillante du ciel a une magnitude apparente de 1.46. La différence de magnitude Δm entre deux étoiles de flux F_1 et F_2 étant $\Delta m = -2.5 \log_{10} \frac{F_2}{F_1}$, Gaia observera des étoiles $\approx 10^{-8}$ fois moins brillantes que Sirius.

grand nombre de données enregistrées (1 milliard d'étoiles en 5 ans, chacune observée en moyenne 80 fois) ne permet pas de transmettre au sol des images à haute résolution.

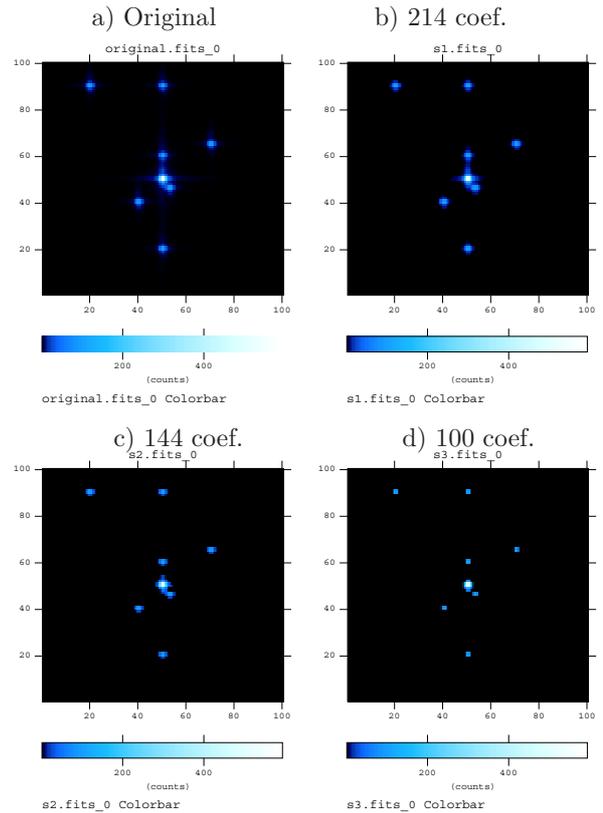


FIG. 1 – a) Image originale, et images seillées en ne retenant que les 214 (b), 144 (c) et 100 (d) pixels de plus forte intensité. L'étoile centrale est de magnitude 18, les autres de magnitude 20.

Chacune de ces 80 images représentera par conséquent une version basse résolution d'un même champ centré autour de la même étoile, convolué par la fonction d'étalement de Gaia. De plus, à cause du mouvement propre du satellite, chaque image subira d'une prise à l'autre une rotation et un décentrage aléatoires. On peut montrer [5] que le mécanisme d'acquisition de ces images peut être modélisé comme $\underline{y} = \Phi\underline{x} + \underline{b}$, où \underline{x} est l'image (vectorisée) à haute résolution recherchée, et où Φ est une matrice prenant en compte les effets de la fonction d'étalement, de rotation et de translation aléatoires, et finalement de la coaddition de pixels par paquets avant transmission au sol.

Nous illustrons ici le potentiel de méthodes basées sur une minimisation L_1 pour un système d'acquisition de données proches de celui de Gaia. Dans le cas considéré ici, la matrice Φ est une matrice aléatoire à coefficients gaussiens identiquement distribués. L'image recherchée \underline{x} est l'image à haute résolution présentée dans la Fig. 1 en haut à gauche. L'image originale (100 x 100 pixels, $N = 10000$) correspond à ce que le satellite Gaia enregistrerait avec des pixels carrés de 59 mas de côté (sans bruit). Elle montre un champ d'étoiles (pics de Dirac convolués par la PSF) d'environ 6 arcsec de côté. Ces images sont clairement éparses dans l'espace direct. Pour quantifier cette propriété, on peut considérer les images obtenues en ne retenant qu'une fraction des pixels de plus forte intensité. D'après la Fig. 1, l'information originale est contenue sur $K \approx 100$ coefficients (l'information est concentrée sur $\approx 1\%$ des pixels). Comme mentionné en Sec. 1, il faut s'attendre à ce que

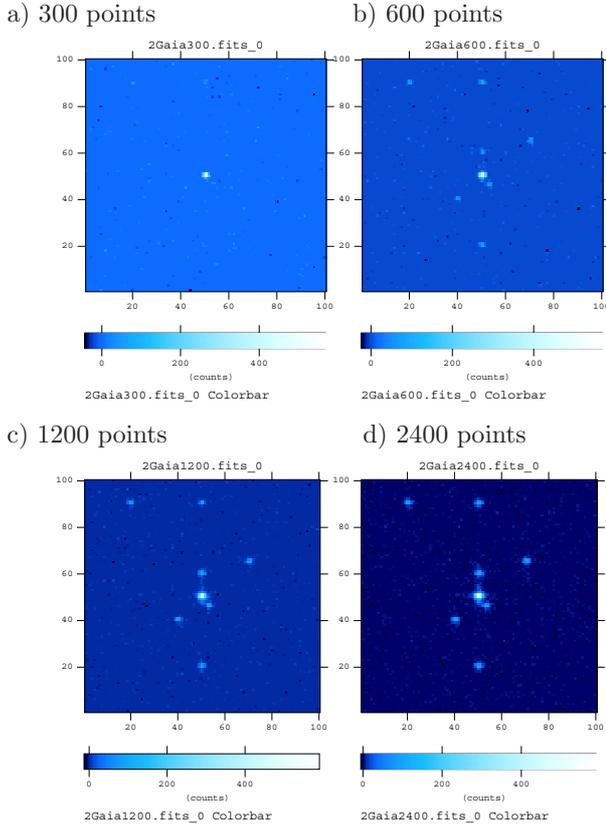


FIG. 2 – Images reconstruites à partir d’images vectorisées $\underline{y} = \Phi \tilde{\underline{x}}$, de dimensions 300, 600, 1200 et 2400 points.

des techniques basées sur le Compressed Sensing nécessitent des données de dimension cK pour reconstruire les images, où c est un facteur dépendant du signal qui reste à déterminer.

Les scènes célestes observées par Gaia étant pour la plupart éparées par nature (la grande majorité d’entre elles comportant une étoile près du centre de l’image et une ou deux autres à quelques pixels), il est pertinent d’exploiter une minimisation en norme L_1 . La Fig. 2 montre les images reconstruites en utilisant des projections aléatoires de dimensions 300 à 2400 à partir de versions bruitées $\tilde{\underline{x}}$ de l’image originale. Le bruit simulé ici est le bruit (Poissonien) de détection lié à la nature quantique de la lumière (bruit dominant pour les faibles flux considérés). La méthode se résume donc à rechercher un ensemble de coefficients $\underline{\theta}^* = \underline{x}^*$ de l’espace direct possédant une norme L_1 minimale, sous la contrainte $\Phi \tilde{\underline{x}}^* = \underline{y}$.

Les résultats de la Fig. 2 montrent que les objets seraient détectables dès lors que l’on dispose d’environ 600 points de mesure. Par ailleurs des tests ont montré que la reconstruction est peu sensible au bruit dans ce cas. Ces résultats situent empiriquement le facteur c mentionné en Sec. 1 autour de 6 (pour des méthodes basées sur la programmation linéaire, une valeur indicative pour la reconstruction exacte est $c \approx \log_2(1 + \frac{N}{K})$ soit $c \approx 7$ dans notre cas [12]). La reconstruction est de bonne qualité au-delà de 1200 points.

À titre de comparaison, les images Gaia comprendront typiquement 60 points de mesures chacune. La Fig. 2 est par conséquent comparable à une reconstruction du champ \underline{x} réalisée en utilisant les informations de 5, 10, 20 et 40 images. Néanmoins ces résultats ne sont qu’indicatifs puisque dans le cas de Gaia Φ ne sera pas exactement une matrice aléatoire (mais elle en sera proche en raison

du mécanisme d’acquisition des données). Une approche basée sur la minimisation L_1 est donc très prometteuse, et son adaptation précise au cas de Gaia est à l’étude. Les résultats seront comparés à une méthode régularisée de type Tykhonov déjà développée pour Gaia (moindres carrés régularisés avec contrainte de douceur, voir [5, 6]). L’effet du bruit (principalement le fond de ciel, le bruit Poissonien de détection et le bruit Gaussien de lecture lié à la caméra CCD) sur la reconstruction est évidemment un aspect important que l’étude devra couvrir.

Nous avons considéré ci-dessus uniquement des objets ponctuels (champs d’étoiles) parce que ce sera le cas le plus fréquents, mais un grand nombre d’objets étendus (galaxies, nébuleuses planétaires, etc...) sera également observé. La méthode de reconstruction visant une solution compacte est aussi exploitable dans ce cas : on cherche alors à exprimer l’objet étendu \underline{x} comme $\underline{x} = \Psi \underline{\theta}$, où Ψ correspond par exemple à une base d’ondelettes, ou dans le cas général à une concaténation de dictionnaires adaptés à la représentation compacte des objets astronomiques considérés. La reconstruction est obtenue par l’eq. (1).

4 CS et interférométrie

En interférométrie stellaire, les données sont composées de points (u, v) de l’espace de Fourier répartis en fonction de la configuration géométrique de la base interférométrique. L’échantillonnage ainsi réalisé est proche d’être aléatoire, menant a priori à une base de mesure universelle. Les images étant là encore dans de nombreux cas éparées (étoiles dans un champ), on peut chercher à exploiter une contrainte sur la norme L_1 pour la reconstruction. Dans ce cas les données obtenues peuvent s’écrire comme $\underline{y} = \Phi \underline{\Psi} \underline{x}$, où Φ exprime les conditions expérimentales (géométriques) de l’acquisition, et où $\underline{\Psi} \underline{x}$ est dans ce cas la transformée de Fourier de l’image à reconstruire. Le problème étant de reconstruire les images dans l’espace direct, on cherche $\underline{x}^* = \operatorname{argmin} |\underline{x}|_1$ sous la contrainte $\Phi \underline{\Psi} \underline{x}^* = \underline{y}$.

La Fig. 3 illustre ces considérations. L’image \underline{x} à reconstruire est la même qu’en Fig. 1 a). Dans le cas des deux vignettes du haut, l’« image » mesurée est obtenue à partir de l’image d’origine projetée dans l’espace de Fourier, puis sous-échantillonnée quasi-aléatoirement (ou de façon irrégulière) : on impose que les points de mesure se situent sur des lignes radiales dans le plan de Fourier. Le nombre total de points de mesure correspondant à ces lignes est indiqué. Les deux vignettes du haut sont les images reconstruites en minimisant la variation totale D (i.e. la somme du gradient, évaluée à une dimension par $D(\underline{x}) = \sum_i \sqrt{(x_{i+1} - x_i)^2}$). Cette contrainte, qui revient à favoriser les solutions constantes par morceaux, est parfois utilisée dans la littérature comme une solution alternative pour trouver des solutions compactes [11]. Il apparaît sur les vignettes de droite de la Fig. 3 que la reconstruction est acceptable avec un sous-échantillonnage d’un facteur ≈ 25 . Au-delà, des artéfacts apparaissent. Ceux-ci sont liés au type d’échantillonnage considéré : dans le cas de lignes, la Fig. 3 a) comprend deux lignes orientées à 90 degrés l’une de l’autre, structure qui se retrouve dans l’image reconstruite. De telles structures n’apparaissent pas dans le cas de points aléatoirement distribués (Fig. 3 c)). Ce dernier cas montre aussi que les étoiles environnantes sont détectables alors qu’on a peu de données.

Il est important de souligner que ces mesures multiples dans le plan (u, v) ne seraient « facilement » réalisables que dans l’hypothèse d’interféromètres multiples spatiaux.

Dans le cas d'observations interférométriques à bases multiples au sol, les perturbations de front d'onde induites par la présence de turbulence atmosphérique ne permettent pas d'accéder facilement à la phase de la visibilité complexe dans le plan (u, v) . Les relations de clôtures de phases déterminées pour chaque triplet de télescopes (un triplet permet de former trois bases) permettent d'identifier $\frac{(N_{tel}-1)(N_{tel}-2)}{2}$ équations indépendantes sur les phases des $C_{N_{tel}}^2$ visibilités mesurables dans le plan (u, v) , avec un ensemble de N_{tel} télescopes. Les mesures de phases différentielles récemment proposées par ex. dans [13] fournissent un second ensemble d'équations de contraintes sur ces phases, permettant d'envisager l'identification de l'ensemble des phases. Les variances d'estimations de ces dernières (dont l'étude est entamée dans [13]) s'expriment alors comme autant de perturbations sur les mesures éparses disponibles, et devront être soigneusement analysées dans les travaux à venir.

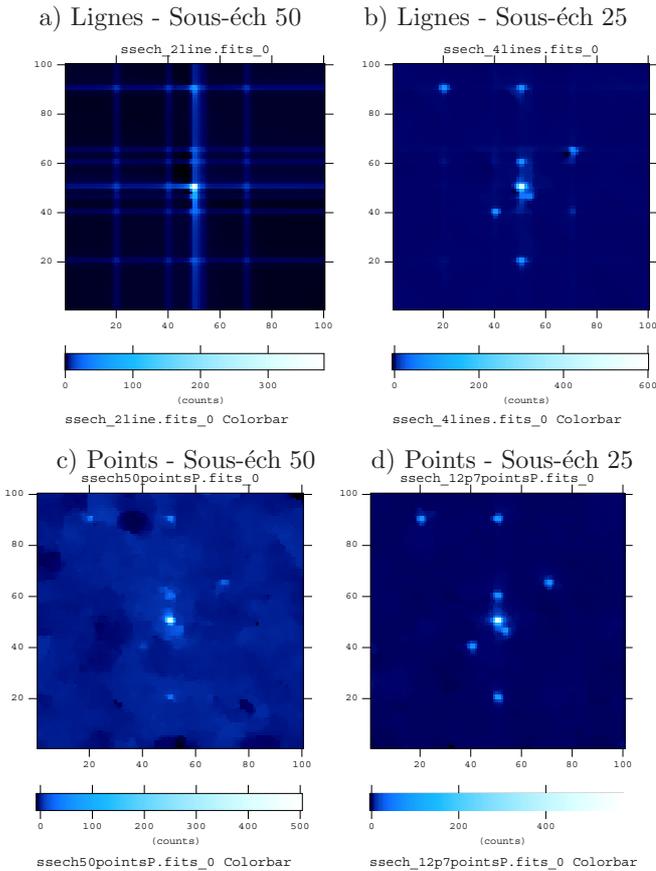


FIG. 3 – Images reconstruites pour un échantillonnage sur des lignes radiales (ligne du haut) et des points (lignes du bas), pour un sous-échantillonnage de 50 (colonne de gauche) et de 25 (colonne de droite) par rapport à la Fig. 1 a) où $N = 10000$.

5 Conclusions et perspectives

Les études envisagées ici ont en commun de ne pas offrir de degrés de liberté en ce qui concerne la conception du processus d'acquisition. Nous nous sommes limités au cas où le processus d'acquisition réalise un échantillonnage aléatoire (ou irrégulier) d'un signal qui possède une représentation éparse dans un espace transformé. Ces idées rejoignent le cadre des problèmes inverses où le caractère éparé des signaux est exploité. Les résultats présentés montrent le fort potentiel du CS pour des applications va-

riées en astronomie, même si ces études demandent à être adaptées plus finement aux contraintes expérimentales. De plus, la complexité des méthodes de minimisation L_1 reste en général élevée (cubique en N dans le cas de la programmation linéaire). Des méthodes approchées « gloutonnes » (greedy) de type Matching Pursuit sont moins complexes mais nécessitent un nombre d'observations M plus élevé. La nature du compromis complexité/qualité de la reconstruction, ainsi que celle des artefacts éventuels dans la reconstruction devront être évalués suivant le type d'application.

D'un point de vue de l'instrumentation, les études sur la théorie du CS ouvrent aussi des perspectives intéressantes pour la conception d'instruments astronomiques (par exemple un satellite qui n'enregistrerait qu'une projection aléatoire et sous-échantillonnée des signaux, voir l'exploitation d'une caméra « monopixel » [2] ou d'une « lentille aléatoire » [1]). De plus, les missions d'interférométrie spatiale à venir offrent la possibilité de définir des configurations de bases arbitraires réalisant un échantillonnage irrégulier et peu dense du plan de Fourier de l'image recherchée.

Références

- [1] R. Fergus, A. Torralba and W. T. Freeman, *Random Lens Imaging*, MIT CSAIL Technical Report, September 2006
- [2] M. Wakin, J. Laska, M. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. Kelly, and R. Baraniuk, *An Architecture for Compressive Imaging*, Proc. International Conference on Image Processing – ICIIP 2006, Atlanta, GA, Oct. 2006
- [3] E. Candès, J. Romberg and T. Tao, *Robust Uncertainty Principles : Exact Signal Reconstruction from Highly Incomplete Frequency Information*, IEEE Trans. on Information Theory, 52(2) pp. 489 - 509, February 2006
- [4] D. Donoho, *Compressed Sensing*, IEEE Trans. on Information Theory, 52(4), pp. 1289 - 1306, April 2006
- [5] C. Dollet, *Compression et restauration pour l'imagerie du ciel à haute résolution angulaire*, PhD thesis, 25 Oct. 2004, Observatoire de la côte d'Azur
- [6] D. Mary, E. Hoeg, H. Lindstroem, U. Bastian, *Updated Simulation of SM Image Reconstruction*, GAIA-C5-TN-ARI-DM-002-1, 2006-11-03, available at <http://www.rssd.esa.int/index.php>
- [7] J. A. Tropp, *Just relax : convex programming methods for identifying sparse signals in noise*, IEEE Trans. on Information Theory, vol. 52, 3, March 2006
- [8] E.J. Candès et T. Tao, *Near optimal signal recovery from random projections : universal encoding strategies ?*, IEEE Trans. on Information Theory, vol. 52, 12, December 2006
- [9] G. Davis, S. Mallat et M. Avellaneda, *Greedy adaptive approximation*, J. Construct. Approx., vol. 13, 1, 1997
- [10] M.A.C. Perryman, K. de Boer, G. Gilmore et al., *Gaia : composition, formation and evolution of the galaxy*, Astronomy and Astrophysics, 369, 339-363, 2001
- [11] L. I. Rudin, S. Osher et E. Fatemi, *Nonlinear total variation noise removal algorithm*, Physica D, 60, 1992
- [12] D. Baron, et al., *Distributed compressed sensing*, 2005, Preprint, available at <http://www.dsp.ece.rice.edu/cs/>
- [13] F. Millour : *Interférométrie différentielle avec AMBER*, Thèse de doctorat de l'UJF, Grenoble 2006.
- [14] S. Roques, B. Serre, N. Dolez, *Band-limited interpolation applied to the time series of rapidly oscillating stars*, MNRAS, 308, 876-886, 1999