

Approche bayésienne variationnelle : application à la déconvolution conjointe d'une source ponctuelle dans une source étendue.

Thomas RODET, Yuling ZHENG

Laboratoire des Signaux et Systèmes
(CNRS-Supélec-UPS)

Supélec, Plateau de Moulon, 91192 Gif-sur-Yvette Cedex, France

Thomas.Rodet@lss.supelec.fr, Yuling.Zheng@lss.supelec.fr, yu-ling.zheng@u-psud.fr

Résumé – Nous présentons dans cette communication une nouvelle méthode de déconvolution-séparation de deux sources : l'une est impulsionnelle (composée de pics), l'autre est étendue (relativement basses fréquences). Afin de séparer les sources on introduit des *a priori* différents sur les deux sources pour les discriminer : un *a priori* gaussien corrélé sur la source étendue et un *a priori* Student sur la source impulsionnelle. En remarquant que la loi de Student peut être obtenue par marginalisation du produit d'une gaussienne par une Gamma. On fait apparaître de cette manière des *a priori* conjugués, qui nous permettent d'utiliser une approche bayésienne variationnelle. Les résultats préliminaires sont présentés dans ce résumé, la qualité de la séparation des composantes astrophysique est très encourageante pour la suite de ces travaux.

Abstract –

1 Introduction

Dans cet article nous nous intéressons à un problème de déconvolution associé à un problème de séparation de composantes astrophysiques. L'objectif est de reconstruire deux cartes : l'une composée uniquement d'étoiles (pics) et l'autre représentant le «fond» du ciel à partir de données issues d'un télescope (une image floue et bruitée). Ce problème de séparation de composantes astrophysiques est très répandu dans la communauté des astrophysiciens. Par exemple, les chercheurs qui travaillent sur le fond diffus cosmologique et ceux qui s'intéressent à la structure des nuages de poussières interstellaires, essaient de retirer les étoiles présentes dans les observations.

La méthode la plus utilisée par les astrophysiciens pour enlever les étoiles de leurs cartes s'appelle «CLEAN». C'est une méthode itérative qui enlève les étoiles les unes après les autres jusqu'à ce qu'il n'y en ait plus. Cette technique possède deux limitations :

- elle ne permet pas de déconvoluer les étoiles qui ne sont pas résolues dans les données (si deux étoiles sont trop proches, elles génèrent une seule tâche dans les données) ;
- elle ne permet pas de déconvoluer le fond de la carte.

Giovannelli *et al.*, dans leur article [1], résolvent ce problème en adoptant un formalisme bayésien et en reconstruisant de manière explicite deux cartes, l'une contenant les sources ponctuelles et l'autre contenant les sources

étendues (plutôt «basses fréquences»). En introduisant les *a priori* différents sur les deux cartes ils arrivent à séparer les deux composantes astronomiques. Un *a priori* gaussien corrélé pour la carte contenant les sources étendues et un *a priori* de Laplace séparable pour la carte contenant les sources ponctuelles. Enfin, ils tiennent compte d'une information de support et une contrainte de positivité sur les différentes sources afin de contraindre le problème.

Dans ce résumé, nous reprenons la formalisation du problème de Giovannelli *et al.*, mais nous introduisons un *a priori* de Student sur l'image des sources ponctuelles. Cette loi possède un paramètre de forme qui permet d'obtenir une loi à plus lourde queue que la loi de Laplace. En utilisant cette loi on favorise les cartes les plus claires-mées. La loi *a posteriori* obtenue en utilisant un *a priori* de Student est très difficile à maximiser ou à échantillonner. En s'inspirant des travaux de G. Chantas *et al.* [2] on choisit de résoudre un problème étendu où l'on estime conjointement la carte des pics et une image d'inverse variance *a priori* de cette dernière. Nous résolvons ce problème inverse étendu à l'aide d'une approche bayésienne variationnelle [3].

2 Méthode

2.1 Modèle direct :

On considère que notre image d'intérêt \mathbf{x} contient deux composantes, une image «douce» \mathbf{s} et une image de pics \mathbf{p} . On considère que la réponse du télescope correspond à la convolution du ciel par une réponse impulsionnelle supposée connue. De plus les données sont entachées d'un bruit \mathbf{b} . La relation qui lie les données notées \mathbf{y} au ciel $\mathbf{x} = \mathbf{s} + \mathbf{p}$ est la suivante :

$$\mathbf{y} = \mathbf{H}(\mathbf{s} + \mathbf{p}) + \mathbf{b} \quad (1)$$

où \mathbf{H} est la matrice Toeplitz de convolution.

On considère ici que le bruit \mathbf{b} est gaussien iid d'inverse variance β_b . On en déduit la vraisemblance suivante :

$$\Pr(\mathbf{y}|\mathbf{p}, \mathbf{s}, \beta_b) \propto \exp\left[-\frac{1}{2}\beta_b\|\mathbf{y} - \mathbf{H}\mathbf{s} - \mathbf{H}\mathbf{p}\|^2\right]$$

2.2 Lois *a priori* :

La modélisation *a priori* de \mathbf{s} est un gaussien corrélé d'inverse variance β_s :

$$\Pr(\mathbf{s}|\beta_s) \propto \exp\left[-\frac{1}{2}\beta_s\|\mathbf{D}\mathbf{s}\|^2\right]$$

où \mathbf{D} est la matrice des différences finies.

La modélisation *a priori* de \mathbf{p} est une loi de Student iid.

$$\Pr(\mathbf{p}|\nu, \lambda) \propto \prod_i \left(1 + \frac{\lambda}{\nu} p_i^2\right)^{-\frac{\nu+1}{2}}$$

Cette loi de Student peut être obtenue par intégration du produit d'une loi normale par une loi Gamma.

$$\Pr(p_i|\nu, \lambda) = \int_{\mathbb{R}} \Pr(p_i|a_i, \lambda) \Pr(a_i|\nu) da_i$$

avec

$$\begin{aligned} \Pr(p_i|a_i, \lambda) &= \mathcal{N}(0, (\lambda a_i)^{-1}) \\ \Pr(a_i|\nu) &= \text{Gamma}\left(\frac{\nu}{2}, \frac{\nu}{2}\right) \end{aligned}$$

L'idée est d'utiliser des loi normale et Gamma à la place de la loi de Student. Nous allons donc résoudre un problème étendu où on estime conjointement $(\mathbf{s}, \mathbf{p}, \mathbf{a})$ avec \mathbf{a} qui est proportionnelle à l'inverse variance *a priori* de \mathbf{p} . Comme souvent, nous prenons le parti de résoudre un problème étendu, possédant plus d'inconnu, où nous devons estimer des variables cachées \mathbf{a} supplémentaires. Nous estimons maintenant trois fois plus d'inconnues que de données. La contre partie s'est que l'on a un *a posteriori* plus simple gaussien pour \mathbf{s}, \mathbf{p} et Gamma pour \mathbf{a} . En effet les lois ci dessus pour le problème étendu sont conjuguées, c'est à dire que l'*a posteriori* est de la même famille que la vraisemblance. En utilisant ces lois *a posteriori* simples, on peut utiliser l'approche bayésienne variationnelle présentée dans la partie 2.4

Remarque : En utilisant ce problème étendu on définit la loi de Student comme un mélange d'une infinité de loi gaussienne centrée. Il est fort probable qu'on puisse par des approches comparables définir une grande variété de lois en jouant sur la paramétrisation de la loi Gamma.

2.3 Loi *a posteriori* :

On définit une loi *a posteriori* étendue pour $(\mathbf{s}, \mathbf{p}, \mathbf{a})$

$$\Pr(\mathbf{s}, \mathbf{p}, \mathbf{a}|\mathbf{y}, \beta_b, \beta_s, \nu, \lambda) \propto \exp\left[-\frac{1}{2}\beta_b\|\mathbf{y} - \mathbf{H}\mathbf{s} - \mathbf{H}\mathbf{p}\|^2 - \frac{1}{2}\beta_s\|\mathbf{D}\mathbf{s}\|^2\right] \prod_i \sqrt{a_i} \exp\left[-\frac{1}{2}p_i^2 a_i \lambda\right] a_i^{\frac{\nu}{2}-1} \exp\left[-\frac{1}{2}\nu a_i\right] \quad (2)$$

2.4 Approche bayésienne variationnelle :

Nous évoquerons ici uniquement le principe de l'approche bayésienne variationnelle. Les personnes intéressées pourront voir par exemple la thèse de Choudrey [4]. L'idée est d'approcher la loi *a posteriori* par une loi plus simple q (avec laquelle on peut faire des calculs explicites) en minimisant la distance de Kullback-Leibler entre ces lois. Nous avons vu précédemment que nous avons réécrit notre problème inverse et que nous avons choisi des lois *a priori* pour avoir une loi *a posteriori* plus simple possible, mais ça ne suffit pas pour mener des calculs explicitement, il faut que les lois soient séparables par rapport à $(\mathbf{s}, \mathbf{p}, \mathbf{a})$. On prend une loi approchante $q(\mathbf{s}, \mathbf{p}, \mathbf{a}) = q_{\mathbf{s}}(\mathbf{s})q_{\mathbf{p}}(\mathbf{p})q_{\mathbf{a}}(\mathbf{a})$. Dans ce cas on sait que la loi approchante qui minimise la distance de Kullback-Leibler entre $\Pr(\mathbf{s}, \mathbf{p}, \mathbf{a}|\mathbf{y}, \Theta)$ et $q(\mathbf{s}, \mathbf{p}, \mathbf{a})$ est de la même famille que l'*a posteriori*. Cette minimisation fonctionnelle se fait uniquement au travers de la modification des paramètres des lois approchantes, ce qui simplifie énormément la mise en œuvre. Cette approche mène à un algorithme itératif : on choisi une première loi approchante puis on modifie les paramètres de lois à l'aide d'équations de mise à jours afin qu'à chaque itération la distance de Kullback-Leibler diminue. Vis à vis de la fonction q la distance de Kullback-Leibler est convexe, donc il n'y a qu'un minimum global. Par contre, comme on modifie la fonction q uniquement dans certaine famille de fonction au travers de paramètres le problème n'est plus convexe. Il faut donc choisir correctement l'initialisation pour s'affranchir de la présence éventuelle de minima locaux.

Remarque : Les avantages de cette approche bayésienne variationnelle sont multiples :

- La solution calculée est une loi de probabilité, ce qui permet d'en déduire immédiatement tous les estimateurs habituels (maximum *a posteriori*, moyenne *a posteriori*, médiane *a posteriori*). On peut aussi obtenir des variances *a posteriori* très utile, pour l'exploitation des résultats.

- On peut aussi obtenir numériquement, la vraisemblance du modèle. C'est à dire la probabilité des données sachant toutes les hypothèses ayant conduit à la résolution du problème inverse. À l'aide de cette vraisemblance on peut choisir entre différents modèles *a priori* ou différents modèles physiques, par exemple on peut déterminer le nombre de sources dans le cas d'un problème de séparation de sources.
- Cette approche permet aussi de faire très facilement un bon compromis entre approximation et coût de calcul. En effet, on peut choisir la granularité de la séparabilité des lois approchantes. Plus la loi est séparable plus le nombre de paramètres à estimer est faible, donc plus l'approche itérative est efficace.

2.5 Lois approchantes :

Comme nous avons choisi des *a priori* conjugués, les lois approchantes q sont de la même famille que la loi *a posteriori*. Pour que les calculs avec les lois approchantes soient plus simples qu'avec la loi *a posteriori* on prend des lois indépendantes $q(\mathbf{s}, \mathbf{p}, \mathbf{a}) = \prod_i q_{s_i}(s_i) \prod_j q_{p_j}(p_j) \prod_k q_{a_k}(a_k)$.

$$\begin{aligned} q_{s_i}(s_i) &= \mathcal{N}(ms_i, \sigma_{s_i}^2), \\ q_{p_j}(p_j) &= \mathcal{N}(mp_j, \sigma_{p_j}^2), \\ q_{a_k}(a_k) &= \text{Gamma}(\alpha_k, \gamma_k). \end{aligned}$$

2.6 Mise à jour des paramètres des lois approchantes :

Après une série de calculs que l'on ne détaillera pas dans ce résumé on obtient les équations de mise à jour suivantes :

$$\begin{aligned} (\sigma_{s_i}^2)^{k+1} &= (\beta_b \text{diag}(\mathbf{H}^t \mathbf{H})(i) + \beta_s \text{diag}(\mathbf{D}^t \mathbf{D})(i))^{-1} \\ ms_i^{k+1} &= (\sigma_{s_i}^2)^{k+1} * \beta_b [\mathbf{H}^t(\mathbf{y}(i) - \mathbf{H}ms^k - \mathbf{H}mp^k)(i) \\ &+ \text{diag}(\mathbf{H}^t \mathbf{H})(i)ms_i^k - \frac{\beta_s}{\beta_b} (\mathbf{D}^t \mathbf{D}ms^k(i) - \text{diag}(\mathbf{D}^t \mathbf{D})(i)ms_i^k)] \\ (\sigma_{p_i}^2)^{k+1} &= \left(\beta_b \text{diag}(\mathbf{H}^t \mathbf{H})(i) + \lambda \frac{\alpha_i^k}{\gamma_i^k} \right)^{-1} \\ mp_i^{k+1} &= \beta_b (\mathbf{H}^t(\mathbf{y}(i) - \mathbf{H}ms^k - \mathbf{H}mp^k)(i) \\ &+ \text{diag}(\mathbf{H}^t \mathbf{H})(i)mp_i^k) * (\sigma_{p_i}^2)^{k+1} \\ \alpha_i^{k+1} &= \frac{\nu + 1}{2} \\ \gamma_i^{k+1} &= \frac{\nu}{2} + \frac{\lambda}{2} ((mp_i^k)^2 + (\sigma_{p_i}^2)^k) \end{aligned} \quad (3)$$

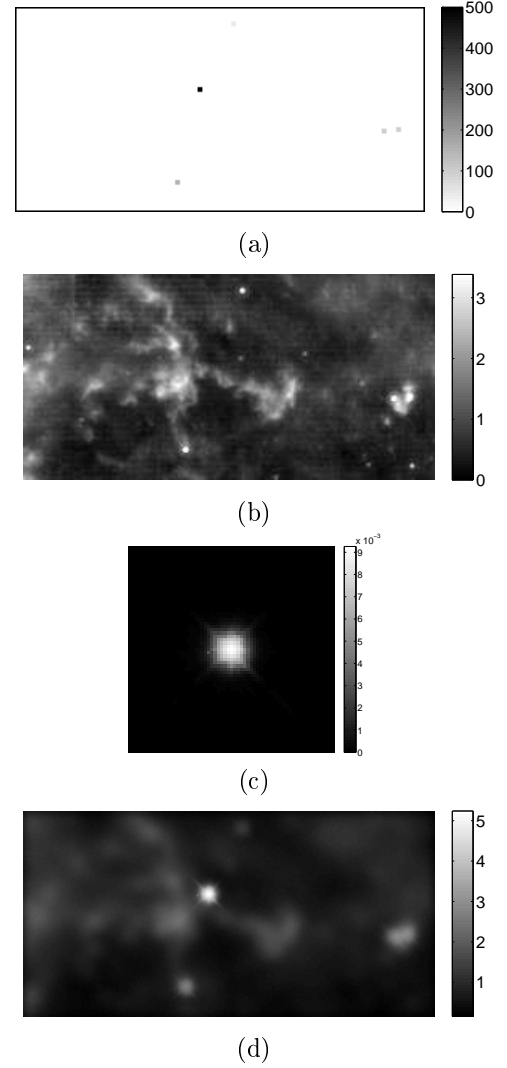


FIG. 1 – Données simulées : (a) image vraie composée de pics (les images de pics ont été affichées en couleur inversée et les pics ont été artificiellement grossis d'un facteur neuf d'être visibles); (b) image vraie de sources étendues (carte obtenue à l'aide du télescope Spitzer) (c) la réponse impulsionnelle du télescope Spitzer (d) données obtenues en ajoutant (a) et (b) en convoluant par la réponse (c) puis en ajoutant un bruit gaussien d'écart type $\sigma_b = 0.017$.

3 Résultats

L'objectif pratique de cette méthode de traitement est de séparer deux composantes astrophysiques : Les étoiles et l'émission du milieu interstellaire. En effet, lorsqu'on veut mesurer l'émission des poussières dans le milieu interstellaire, les étoiles qui sont des sources 500 à 1000 fois plus intenses réduisent la qualité de l'analyse du signal.

Pour valider notre approche nous avons essayé de nous rapprocher de ce problème de traitement des données astrophysiques. Nous avons donc simulé des données comme si elle avait été acquises par la satellite SPITZER (satellite

d’observation de l’émission infrarouge lointaine).

3.1 Simulation réaliste

Pour ce faire, nous avons construit une image du ciel qui est l’addition d’une carte d’un nuage de poussière (voir figure 1 (a)) plus une carte composée de pics de forte amplitude (voir figure 1(b)). Nous avons ensuite convolué l’image du ciel par la réponse impulsionnelle mesurée de SPITZER (voir figure 1 (c)). Notons que cette réponse impulsionnelle est de grande dimension 64×64 , ce qui rend la résolution dans le domaine de Fourier délicate. Puis on ajoute un bruit gaussien iid d’écart type $\sigma_b = 0.017$, correspondant à celui observé sur l’instrument.

3.2 Comparaison avec l’existant

Afin de démontrer l’apport de *a priori* de Student par rapport à un *a priori* de Laplace, nous avons implémenté une méthode comparable à celle de GIOVANNELLI *et al.*. On minimise le même critère, mais sans contrainte de support et de positivité et sans faire d’approximation circulante. Tous les résultats sont résumés sur la figure 2¹. Nous avons initialisé les deux approches par la solution du filtre inverse. La valeur des hyperparamètres a été fixée à : $\lambda = 0.02$, $\nu = 0.1$, $\beta_b = 1$ et $\beta_s = 10^{-5}$. Enfin nous avons obtenu une loi approchant de la loi *a posteriori* (2) en 3000 itérations et nous visualisons sur la figure 2 uniquement la moyenne de \mathbf{s} et la moyenne de \mathbf{p} .

Si on analyse les résultats des figures 2(e) et 2(f) on remarque que l’*a priori* de Student à garder un pic et à forcer les autres valeur à être très proche de 0, ce n’est pas le cas avec *a priori* de Laplace le pic n’est pas résumé à une valeur et il reste un résidu dans la carte douce figure 2(b), car les forte valeur sont quand même pénalisées.

4 Conclusions

Les premiers résultats présentés ici sont très encourageants, on a une excellente séparation des deux composantes astrophysiques. Mais il reste un travail de validation sur données réelles.

Références

- [1] J.-F. Giovannelli et A. Coulais, « Positive deconvolution for superimposed extended source and point sources », *Astron. Astrophys.*, vol. 439, pp. 401–412, 2005.
- [2] G. Chantas, N. Galatsanos, A. Likas et M. Saunders, « Variational bayesian image restoration based on a product of *t*-distributions image prior », *IEEE Trans. Image Processing*, vol. 17, n°10, pp. 1795–1805, octobre 2008.
- [3] R. Choudrey et S. Roberts, « Variational Mixture of Bayesian Independent Component Analysers », *Neural Computation*, vol. 15, n°1, 2003.

¹ Les pics des cartes Fig 2(a) et 2(c) ont été artificiellement grossis d’un facteur ζ pour qu’ils soient visibles

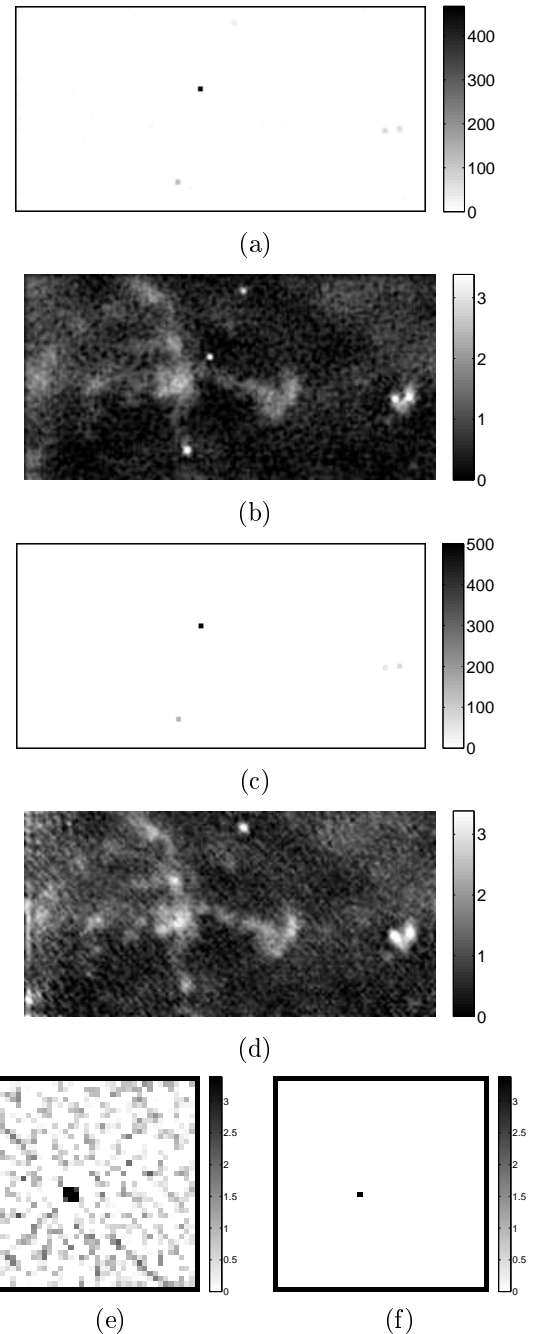


FIG. 2 – Résultats : (a) Estimation des pics avec *a priori* Laplace (paramètre de régularisation 10^{-5}); (b) Estimation du fond avec *a priori* gaussien corrélé (paramètre de régularisation 6×10^{-5}); (c) carte des pics obtenue avec notre approche¹; (d) carte de la composante étendue obtenue avec notre approche; Zoom sur la carte des pics avec la dynamique du fond (e) *a priori* Laplace, (f) *a priori* Student

- [4] R. A. Choudrey, *Variational Methods for Bayesian Independent Component Analysis*, Thèse de Doctorat, University of Oxford, 2002.