Rééchantillonnage de l'échelle dans les algorithmes MCMC pour les problèmes inverses bilinéaires

Thomas VEIT^{*}, Jérôme IDIER

Equipe ADTSI - IRCCyN / CNRS BP 92101 - 1 rue de la Noë - 44321 Nantes Cedex 3, France jerome.idier@irccyn.ec-nantes.fr, thomas.veit@inrets.fr

Résumé – Cet article présente une méthode pour améliorer le comportement des algorithmes MCMC impliqués dans la résolution de problèmes inverses bilinéaires. Les applications concernées sont en particulier la déconvolution aveugle et la séparation de sources. La méthode proposée s'intéresse à l'ambiguïté d'échelle intrinsèque aux problèmes inverses bilinéaires. La résolution de ce type de problèmes dans un cadre bayésien lève l'ambiguïté d'échelle mais les échantillons issus de l'algorithme MCMC sont fortement corrélés. L'introduction d'une étape d'échantillonnage d'un paramètre d'échelle améliore radicalement le comportement de l'algorithme MCMC. Par rapport à d'autres solutions existantes, cette approche est mathématiquement rigoureuse sans augmentation significative du coût de calcul.

Abstract – This article presents an efficient method for improving the behavior of the MCMC sampling algorithm involved in the resolution of bilinear inverse problems. Blind deconvolution and source separation are among the applications that benefit from this improvement. The proposed method addresses the scale ambiguity inherent to bilinear inverse problems. Solving this type of problem within a Bayesian framework lifts the ambiguity but the MCMC involved is badly mixing. Introducing an additional sampling step on the scale parameter drastically improves the behavior of the MCMC. The main advantage of this method, compared to other attempts to correct the pathological behavior of the MCMC, is that it is theoretically sound. The additional computational cost is negligible.

1 Introduction

Ce travail s'intéresse aux algorithmes MCMC (*Markov* Chain Monte Carlo) impliqués dans la résolution de problèmes inverses bilinéaires dans un cadre bayésien. La séparation de sources [1] et la déconvolution aveugle [2] sont des cas particuliers importants de problèmes inverses bilinéaires. Les quantités inconnues d'un problème inverse bilinéaire sont reliées aux observations par une fonction bilinéaire. Autrement dit, la fonction reliant les observations aux deux inconnues est linéaire par rapport à chacune des deux variables lorsque l'autre est constante. Nous noterons ce type de problème de manière générale $\mathbf{z} = \mathbf{x} \star \mathbf{h} + bruit$, où \mathbf{z} représente les observations et \mathbf{x} et \mathbf{h} les matrices ou vecteurs inconnus que l'ont souhaite estimer. Le symbole \star désigne un opérateur bilinéaire quelconque, par exemple un produit de convolution ou un produit matriciel.

Les problèmes inverses bilinéaires présentent une ambiguïté d'échelle intrinsèque : si le couple $(\boldsymbol{x}, \boldsymbol{h})$ est solution, tout couple $(s \times \boldsymbol{x}, \boldsymbol{h}/s)$, où $s \neq 0$, est aussi solution.

2 Contexte bayésien

Le cadre bayésien a pour but d'introduire de l'information statistique dans un problème sous-contraint. Des lois *a priori* sont affectées aux quantités inconnues du problème. On notera $Z = X \star H + B$ le problème ainsi modélisé. Les majuscules symbolisent ici des quantités aléatoires.

Dans ce cadre, les quantités inconnues \boldsymbol{x} et \boldsymbol{h} peuvent être estimées par approximation empirique de leur espérance *a posteriori*, pour \boldsymbol{x} :

$$\hat{\boldsymbol{x}} = E(\boldsymbol{X} \mid \boldsymbol{Z} = \boldsymbol{z}) = \lim_{K \to \infty} \sum_{k=1}^{K} \boldsymbol{x}_{k}.$$
 (1)

Les \boldsymbol{x}_k correspondent aux réalisations de tirages aléatoires selon la loi *a posteriori*. Il est important de noter que la résolution dans un cadre bayésien lève l'ambiguïté d'échelle à travers les lois *a priori* choisies pour \boldsymbol{X} et \boldsymbol{H} .

2.1 Echantillonneur de Gibbs

Les méthodes MCMC permettent de générer les réalisations de X et H selon leurs lois *a posteriori*. Ces réalisations sont alors utilisées pour le calcul de \hat{x} sous la forme (1). Dans le cas des problèmes inverses bilinéaires, un échantillonneur de Gibbs peut être utilisé pour simuler une chaîne de Markov dont les principales étapes sont le rééchantillonnage suivant les lois conditionnelles $f_{X|H,Z,\Theta}(x \mid h, z, \theta)$ et $f_{H|X,Z,\Theta}(h \mid x, z, \theta)$. Le vecteur θ correspond ici à des hyperparamètres ou à des variables cachées, eux-même susceptibles d'être échantillonnés.

De manière synthétique, chaque itération t de l'algorithme de Gibbs s'écrit de la manière suivante :

 $^{^{*}\}mathrm{adresse}$ actuelle: Equipe Perception, LIVIC, bâtiment 824 - 14, route de la Minière, 78000 Versailles.

Algorithme 1 Etant donné
$$\boldsymbol{x}^{(t)}, \boldsymbol{h}^{(t)}, \boldsymbol{\theta}^{(t)},$$
 générer :
1. $\boldsymbol{x}^{(t+1)} \leftarrow \boldsymbol{X} \sim f_{\boldsymbol{X}|\boldsymbol{H},\boldsymbol{Z},\boldsymbol{\Theta}}(\boldsymbol{x} \mid \boldsymbol{h}^{(t)}, \boldsymbol{z}, \boldsymbol{\theta}^{(t)})$
2. $\boldsymbol{h}^{(t+1)} \leftarrow \boldsymbol{H} \sim f_{\boldsymbol{H}|\boldsymbol{X},\boldsymbol{Z},\boldsymbol{\Theta}}(\boldsymbol{h} \mid \boldsymbol{x}^{(t+1)}, \boldsymbol{z}, \boldsymbol{\theta}^{(t)})$
3. $\boldsymbol{\theta}^{(t+1)} \leftarrow \boldsymbol{\Theta} \sim f_{\boldsymbol{\Theta}|\boldsymbol{X},\boldsymbol{H},\boldsymbol{Z}}(\boldsymbol{\theta} \mid \boldsymbol{x}^{(t+1)}, \boldsymbol{h}^{(t+1)}, \boldsymbol{z})$

2.2 Mauvaises propriétés de mélange de la chaîne; solutions existantes

Expérimentalement, on constate que l'échelle des échantillons de X et de H varie très lentement : les réalisations successives sont fortement corrélées et la convergence vers la distribution stationnaire cible est très lente. Même si on a pu constater empiriquement, sur les exemples traités, que la convergence de la forme des signaux se produit avant celle de leur échelle, nous n'avons trouvé aucun résultat mathématique qui permettrait d'assurer qu'un tel découplage est systématique, ou à défaut de savoir le prédire ou le quantifier.

Pour pallier cette lenteur de convergence, on trouve communément deux approches empiriques dans la littérature. La première consiste à calculer les estimateurs à partir de la chaîne de Markov qui n'a pas convergé. Implicitement, cette démarche repose sur le découplage hypothétique mentionné ci-dessus.

L'autre solution consiste à renormaliser à une échelle fixée le couple (x, h) à chaque itération de l'échantillonneur. Cette solution n'est pas non plus satisfaisante théoriquement, car la convergence vers la loi cible n'est plus garantie. En effet, pour conserver les propriétés de convergence, il faudrait tirer les échantillons selon des lois *conditionnées par la normalisation imposée*. Malheureusement, ces lois sont difficiles à manipuler. De plus, fixer l'échelle à l'intérieur de l'échantillonneur revient à lever doublement l'ambiguïté d'échelle.

3 Étape supplémentaire d'échantillonnage de l'échelle

La solution que nous proposons s'insère naturellement dans l'échantillonneur de Gibbs, en accélérant la convergence vers la loi cible, *sans la modifier*. L'idée est d'introduire une étape supplémentaire d'échantillonnage au sein de l'échantillonneur de Gibbs pour améliorer l'exploration de l'espace d'état. Cette nouvelle étape s'écrit

$$\boldsymbol{X} = \boldsymbol{X}^{\mathrm{old}} \times S, \quad \boldsymbol{H} = \boldsymbol{H}^{\mathrm{old}}/S,$$
 (2)

où S est un paramètre d'échelle qui est lui-même échantillonné à chaque itération.

Dans cette section, nous montrons que la distribution selon laquelle S doit être rééchantillonné s'écrit

$$f_{S|reste}(s) \propto |s|^{M-P-1} f_{\boldsymbol{X},\boldsymbol{H}|\Theta}(s\boldsymbol{x}^{\text{old}},\boldsymbol{h}^{\text{old}}/s \mid \boldsymbol{\theta}), \quad (3)$$

où M et P sont les tailles respectives de X et H. Notons que l'expression (3) se factorise sous la forme

$$f_{S \mid reste}(s) \propto |s|^{M-P-1} f_{\boldsymbol{X}\mid\Theta}(s\boldsymbol{x}^{\text{old}}\mid\boldsymbol{\theta}) f_{\boldsymbol{H}\mid\Theta}(\boldsymbol{h}^{\text{old}}/s\mid\boldsymbol{\theta}),$$

dans le cas fréquent où \boldsymbol{X} et \boldsymbol{H} sont supposés a priori
indépendants.

Dans la suite, la dépendance vis-à-vis du vecteur Θ sera omise pour alléger la démonstration.

3.1 Introduction d'un paramètre d'échelle par changement de variable

Définissons un changement de variable ϕ sur le couple $(\boldsymbol{X}, \boldsymbol{H})$ qui permet de faire apparaître explicitement le paramètre d'échelle S. Dans ce changement de variable, la quantité x_1^{old} , première composante du vecteur \boldsymbol{x} avant le changement d'échelle, joue le rôle d'une constante, supposée non nulle pour l'instant.

$$(s, \boldsymbol{v}) = \phi(\boldsymbol{x}, \boldsymbol{h})$$

= $(\underbrace{x_1/x_1^{\text{old}}}_{s}, \underbrace{x_2/x_1, \dots, x_M/x_1}_{v_1, \dots, v_{M-1}}, \underbrace{x_1h_1, \dots, x_1h_P}_{v_M, \dots, v_{M+P-1}}).$

Le vecteur \boldsymbol{v} , de taille M + P - 1, est construit de façon à ne pas être sensible à l'échelle de $(\boldsymbol{x}, \boldsymbol{h})$, au sens où seule la première des M + P composantes de $\phi(\alpha \boldsymbol{x}, \boldsymbol{h}/\alpha)$ dépend de α .

Il est facile de vérifier que ϕ est inversible, et que ϕ^{-1} s'exprime de la façon suivante :

$$\phi^{-1}(s, \boldsymbol{v}) = \left(sx_1^{\text{old}}, v_1 sx_1^{\text{old}}, \dots, v_{M-1} sx_1^{\text{old}}, \frac{v_M}{sx_1^{\text{old}}}, \dots, \frac{v_{M+P-1}}{sx_1^{\text{old}}}\right).$$

Le jacobien de ϕ^{-1} est proportionnel à s^{M-P-1} . On a donc

$$f_{S,\boldsymbol{V}}(s,\boldsymbol{v}) \propto |s|^{M-P-1} f_{\boldsymbol{X},\boldsymbol{H}}(\phi^{-1}(s,\boldsymbol{v})), \qquad (4)$$

et la loi de S conditionellement à V (c'est-à-dire à (X, H), sans tenir compte de l'échelle courante) se déduit de façon immédiate :

$$f_{S|\boldsymbol{V}}(s|\boldsymbol{v}) \propto f_{S,\boldsymbol{V}}(s,\boldsymbol{v}).$$
 (5)

3.2 $f_{S|reste}$ ne dépend pas des données

Montrons maintenant l'indépendance de S par rapport aux observations Z conditionnellement à V. Ceci résulte de la bilinéarité de l'équation d'observation, à travers l'identité $(\boldsymbol{x} \times s) \star (\boldsymbol{h}/s) = \boldsymbol{x} \star \boldsymbol{h}$. En effet on a :

$$f_{S|\boldsymbol{V},\boldsymbol{Z}}(s \,|\, \boldsymbol{v}, \boldsymbol{z}) \propto f_{\boldsymbol{Z}|S,\boldsymbol{V}}(\boldsymbol{z} \,|\, s, \boldsymbol{v}) f_{S,\boldsymbol{V}}(s, \boldsymbol{v}).$$

Mais la loi $f_{\mathbf{Z}|S,\mathbf{V}}$ ne fait pas intervenir l'échelle S puisque

$$f_{\boldsymbol{Z}|S,\boldsymbol{V}}(\boldsymbol{z}\,|\,s,\boldsymbol{v}) = f_{\boldsymbol{Z}|\boldsymbol{X},\boldsymbol{H}}(\boldsymbol{z}\,|\,\boldsymbol{x},\boldsymbol{h}) = f_{\boldsymbol{Z}|\boldsymbol{X}\star\boldsymbol{H}}(\boldsymbol{z}\,|\,\boldsymbol{x}\star\boldsymbol{h}).$$

On en déduit :

$$f_{S|\boldsymbol{V},\boldsymbol{Z}}(s \,|\, \boldsymbol{v}, \boldsymbol{z}) = f_{S|\boldsymbol{V}}(s \,|\, \boldsymbol{v}) \tag{6}$$

ce qui permet d'obtenir (3), en utilisant (6), (5) et (4).

3.3 Echantillonneur de Gibbs modifié

L'Algorithme 2 comporte à l'identique les trois étapes d'échantillonnage de Gibbs de l'Algorithme 1, auxquelles une étape supplémentaire a été adjointe, que nous interprétons également comme une étape d'échantillonnage de Gibbs, à un changement de variable près. Cette étape intercalée s'appuie sur l'expression de la loi conditionnelle de l'échelle (3).

Dans l'établissement de la loi (3), nous avons supposé que la première composante courante de x était non nulle. Justifions que cette hypothèse n'est pas limitative. Manifestement, nous aurions obtenu la même expression à l'aide d'un changement de variable équivalent, supposant que tout autre composante courante de x, ou même de h, était non nulle. Le seul cas impossible à gérer est celui où toutes les composantes de x et de h sont nulles. Dans ce cas très particulier, (3) n'a pas de sens, mais il est alors inutile de rééchantillonner l'échelle au sens de (2).

L'expression générale (3) se décline selon les lois a priori pour l'application considérée. Pour illustrer l'intérêt de cette étape supplémentaire dans l'échantillonneur de Gibbs, examinons quelques cas particuliers.

pplications $\mathbf{4}$

Application à la déconvolution aveug-4.1le de train d'impulsions

La déconvolution aveugle consiste à estimer simultanément le signal d'entrée x et la réponse impulsionnelle hdu filtre à partir d'une version filtrée z. L'opérateur bilinéaire reliant les inconnues aux observations est donc le produit de convolution. Les inconnues x et h sont ici des vecteurs, mais on pourrait s'intéresser plus généralement au cas multidimensionnel.

Le bruit B est supposé blanc et gaussien et l'a priori sur \boldsymbol{H} est choisi gaussien centré $\mathcal{N}(0, R_H)$. D'autre part, le modèle Bernoulli-Gaussien (BG) est approprié pour le signal d'entrée x dans le cas d'un train d'impulsions. Chaque composante non-nulle est modélisée par une loi gaussienne centrée $\mathcal{N}(0, \sigma_x^2)$. Le vecteur de classes binaires Q est un ensemble de variables de Bernoulli indépendantes, avec $P(Q_i = 1) = \lambda$ et $P(Q_i = 0) = 1 - \lambda$. La variable Θ contient les paramètres statistiques du bruit, du modèle de la réponse impulsionnelle et du modèle BG, mais aussi le vecteur Q de classes binaires. Pour $i \in \{1, \ldots, M\},\$

$$-\operatorname{si} q_i = 1, \, X_i \sim \mathcal{N}(0, \sigma_x^2)$$

– si $q_i = 0, X_i = 0$ avec probabilité 1.

On déduit de (3) que S^2 suit une loi conditionnelle de type Inverse Gaussienne Généralisée (GIG) de paramètres

$$\lambda = \frac{1}{2} \left(\sum_{i=1}^{M} q_i - P \right), \quad \alpha = \sum_{i=1}^{M} q_i \frac{x_i^2}{\sigma_x^2}, \quad \beta = \mathbf{h}^{\mathrm{t}} R_H \mathbf{h}.$$

La densité de la loi GIG s'écrit

$$f_{\text{GIG}}(z) = \left(\frac{\alpha}{\beta}\right)^{\frac{\Delta}{2}} \frac{z^{\lambda-1}}{2K_{\lambda}\left(\sqrt{\alpha\beta}\right)} \exp\left(-\frac{1}{2}(\beta z^{-1} + \alpha z)\right),$$
$$z > 0, \quad \lambda \in \mathbf{R}, \quad \begin{cases} \alpha > 0, \beta \ge 0 \text{ pour } \lambda > 0\\ \alpha > 0, \beta > 0 \text{ pour } \lambda = 0\\ \alpha \ge 0, \beta > 0 \text{ pour } \lambda < 0 \end{cases}$$

où $K_{\lambda}(.)$ est la fonction de Bessel de troisième espèce.

Des méthodes de simulation efficaces basées sur la méthode d'acceptation-rejet existent pour les lois GIG.

Les résultats des Figures 1 et 2 montrent l'évolution de la norme des réalisations successives de la réponse impulsionelle h au cours des itérations. Ceux de la Figure 1 ont été obtenus avec l'Algorithme 1 sans rééchantillonnage de l'échelle. On note que la norme de h évolue très lentement. La forte corrélation entre réalisations entraîne une forte dépendance par rapport à l'initialisation. La Figure 2 montre des résultats obtenus avec l'Algorithme 2. Les normes des réalisations successives sont nettement moins corrélées. L'exploration du support de la densité a *posteriori* de H est plus efficace.

4.2Séparation de sources

La séparation de sources est un autre exemple caractéristique de problème bilinéaire. Les variables du modèle s'interprètent alors de la manière suivante :

- $-\mathbf{Y} \in \mathbb{R}^{m \times N}$: observations,
- $\boldsymbol{H} \in \mathbb{R}^{m \times n}_+$: matrice de mélange, $\boldsymbol{X} \in \mathbb{R}^{n \times N}_+$: signaux source.

C

Dans le cas de signaux source et de coefficients de mélange positifs et décrits par des lois a priori gamma [3], on peut montrer que la loi conditionnelle de S est ellemême une GIG. Plus précisément, supposons que les nsignaux source sont indépendants, et que le jième signal est i.i.d. et distribué selon une loi gamma de paramètres (a_i, b_i) . De même, les colonnes de la matrice de mélange sont supposées indépendantes, et la jième colonne suit une loi gamma de paramètres (c_i, d_i) . La loi conditionnelle (3) est alors une loi GIG de paramètres

$$\lambda = N \sum_{j=1}^{n} a_j - m \sum_{j=1}^{n} b_j + 1,$$

$$\alpha = 2 \sum_{k=1}^{N} \sum_{j=1}^{n} b_j x_{j,k}, \qquad \beta = 2 \sum_{i=1}^{m} \sum_{j=1}^{n} d_j h_{i,j}$$

Dans le cas de lois *a priori* plus complexes, par exemple la loi hyperbolique généralisée (HG) utilisée pour modéliser la distribution des signaux source dans [1], le rééchantillonnage de l'échelle peut sembler ardu. En fait, la décomposition préalable de telles lois a priori sous la forme d'un mélange gaussien continu permet de se ramener au cas gaussien, conditionnellement à un jeu de variables cachées. Dans le cas des variables HG, cette décomposition est déjà utilisée, dans [1], pour rééchantillonner les signaux sources eux-même. La loi a priori des coefficients du mélange étant supposée gaussienne, la loi conditionnelle de S^2 est une GIG, comme dans le cas traité en Section 4.1.



FIG. 1 – Algorithme 1 dans le contexte de la Section 4.1 : $\|\boldsymbol{h}\|_2$ en fonction des itérations pour six initialisations différentes. Les normes des réalisations successives de \boldsymbol{H} sont fortement corrélées. De plus, cette corrélation entraîne une forte dépendance vis-à-vis de l'initialisation.

4.3 Détection-estimation de l'activité cérébrale en IRMf

Dans un contexte applicatif, la soumission conjointe [4] montre l'utilisation du rééchantillonnage de l'échelle pour accélérer une méthode MCMC en imagerie cérébrale dynamique. Plus précisément, il s'agit de la détectionestimation des amplitudes \boldsymbol{a} des réponses neuronales à des stimuli, simultanément avec l'estimation de la réponse hémodynamique \boldsymbol{h} , à partir d'observations suivant un modèle bilinéaire en $(\boldsymbol{a}, \boldsymbol{h})$.

Dans ce cas, les moyennes des lois *a priori* sont des quantités non nulles, inconnues. Pour rééchantillonner l'échelle, il est plus efficace de faire porter le changement d'échelle non seulement sur a et h, mais aussi sur les moyennes, de façon simultanée. La généralisation requise pour l'expression (3) est immédiate (voir [4]).

5 Conclusion

Le travail présenté propose d'insérer une étape de rééchantillonnage supplémentaire au sein de l'algorithme de Gibbs utilisé pour la résolution de problèmes inverses bilinéaires. Cette étape supplémentaire a pour but de corriger les mauvaises propriétés de mélange habituellement obser-



FIG. 2 – Algorithme 2 dans le contexte de la Section 4.1 : $\|\boldsymbol{h}\|_2$ en fonction des itérations pour six initialisations différentes. Les réalisations successives sont peu corrélées, et l'influence de l'initialisation devient assez rapidement négligeable. Ceci permet une exploration correcte du support de la densité *a posteriori* de \boldsymbol{H} .

vées. Par rapport aux remèdes heuristiques existants, la solution proposée a l'avantage d'être générale, rigoureuse et d'un coût de calcul négligeable.

Références

- H. Snoussi et J. Idier, « Bayesian blind separation of generalized hyperbolic processes in noisy and underdeterminate mixtures », *IEEE Trans. Signal Processing*, vol. 54, No 9, pp. 3257–3269, sept. 2006.
- [2] Q. Cheng, R. Chen et T.-H. Li, « Simultaneous wavelet estimation and deconvolution of reflection seismic signals », *IEEE Trans. Geosci. Remote Sensing*, vol. 34, pp. 377–384, mars 1996.
- [3] S. Moussaoui, D. Brie, A. Mohammad-Djafari et C. Carteret, « Separation of non-negative mixture of non-negative sources using a Bayesian approach and MCMC sampling », *IEEE Trans. Signal Processing*, vol. 54, No 11, pp. 4133–4145, nov. 2006.
- [4] P. Ciuciu, J. Idier, T. Veit et T. Vincent, « Application du rééchantillonnage stochastique de l'échelle en détection-estimation de l'activité cérébrale par IRMf », XXIe colloque GRETSI, Id. 578, sept. 2007.