

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

## Scale Resampling within MCMC Algorithms for Bilinear Inverse Problems

Thomas Veit<sup>1</sup>, Jérôme Idier<sup>2</sup> et Saïd Moussaoui<sup>2</sup>

<sup>1</sup> Équipe Perception, LIVIC, bâtiment 824 – 14, route de la Minière, 78000 Versailles  
courriel: thomas.veit@inrets.fr

<sup>2</sup> IRCCyN (UMR CNRS 6597), BP 92101 – Équipe ADTSI  
1 rue de la Noë – 44321 Nantes Cedex 3, France  
{jerome.idier, said.moussaoui}@irccyn.ec-nantes.fr

Manuscrit reçu le 22 avril 2008

Résumé et mots clés

Cet article présente une méthode pour améliorer le comportement des algorithmes MCMC impliqués dans la résolution de problèmes inverses bilinéaires tels que la déconvolution aveugle et la séparation de sources. La question plus spécifiquement abordée est celle de la gestion de l'ambiguïté d'échelle intrinsèque aux problèmes inverses bilinéaires. Le cadre bayésien lève l'ambiguïté d'échelle (pourvu que les lois manipulées soient propres) mais les échantillons obtenus par échantillonnage de Gibbs sont fortement corrélés. L'introduction d'une étape d'échantillonnage d'un paramètre d'échelle améliore radicalement le comportement de l'échantillonneur. En autorisant un déplacement supplémentaire, l'exploration de l'espace d'état est rendue plus efficace. Ceci a pour conséquences une diminution du temps de chauffe, une corrélation moins forte des échantillons et une convergence plus rapide. Par rapport à d'autres solutions envisageables, cette approche est mathématiquement rigoureuse sans augmentation significative du coût de calcul. Son apport est illustré sur un premier exemple simple, puis dans le cadre d'une application à la séparation de sources.

Problèmes inverses, Cadre bayésien, Algorithmes MCMC, Échantillonneur de Gibbs, Ambiguïté d'échelle, Déconvolution aveugle, Séparation de sources.

Abstract and key words

### Introduction

This work focuses on the Markov chain Monte Carlo (MCMC) algorithms involved in the resolution of bilinear inverse problems. This kind of problem is written  $\mathbf{z} = \mathbf{x} \star \mathbf{h} + \text{noise}$ , where  $\mathbf{z}$  stands for the observations and  $\mathbf{x}$  and  $\mathbf{h}$  stand for the unknown quantities (matrices or vectors) to be estimated. Symbol  $\star$  refers to any bilinear operator, for example a convolution or a matrix product. Source separation [11, 2] and blind deconvolution [3] are among the most important instances of bilinear inverse problems in the signal processing field.

Bilinear inverse problems are characterized by an intrinsic scale ambiguity: namely, if  $(\mathbf{x}, \mathbf{h})$  is a solution then for all scalar  $s \neq 0$ ,  $(s \times \mathbf{x}, \mathbf{h}/s)$  is also a solution.

\* Le travail présenté a débuté à l'occasion d'un post-doctorat de Thomas Veit à l'IRCCyN financé par une allocation de recherche CNRS-Région Pays de la Loire.

### Bayesian framework and MCMC

The Bayesian framework introduces statistical information for solving under-constrained problems. Prior distributions are assigned to the unknowns. The problem modeled in this way is written  $\mathbf{Z} = \mathbf{X} \star \mathbf{H} + \mathbf{B}$  where the capital letters stand for random variables. A usual Bayesian estimator of the unknown quantities  $\mathbf{x}$  and  $\mathbf{h}$  is the posterior expectation Eq. (3). Let us emphasize that the choice of the prior distribution in the Bayesian framework raises the scale ambiguity on the unknowns, at least if only *proper* priors are considered, which will be the case here.

In the usual case where the posterior distribution are not explicitly available, MCMC algorithms are well adapted to yield random realizations  $\mathbf{x}^{(t)}$  and to compute the estimate  $\hat{\mathbf{x}}$  as empirical averages.

In the bilinear case, an appealing choice to perform the sampling step is to rely on a Gibbs sampler. The two main steps are then the alternate sampling of  $\mathbf{x}$  and  $\mathbf{h}$  according to their conditional posterior distributions  $f_{\mathbf{X}|\mathbf{H},\mathbf{Z},\Theta}(\mathbf{x}|\mathbf{h},\mathbf{z},\Theta)$  and  $f_{\mathbf{H}|\mathbf{X},\mathbf{Z},\Theta}(\mathbf{h}|\mathbf{x},\mathbf{z},\Theta)$ , respectively.

### Bad mixing properties and existing solutions

In practice, a slow convergence of the Markov chain produced by the Gibbs sampler is observed. Successive sample are highly correlated. More precisely, the scale evolves slowly. To correct this behavior, two types of solutions have been proposed up to now, but none of them are satisfying, at least from a theoretical point of view. The first approach consists in estimating the unknown quantities from Markov chains (produced by Algorithm 1) that have not converged yet toward their stationary distribution. This implies that only quantities normalised with respect to scale are considered. One can reasonably expect that the convergence remains fast enough but the decoupling of scale and shape remains hypothetical. The second strategy (Algorithm 2) consists of introducing an arbitrary scaling which is in general incompatible with the prior distributions on the unknown variables. Moreover, this scaling strategy modifies the target distribution. The proposed solution rather introduces a rescaling step within the Gibbs sampler in agreement with the prior distributions. One attractive feature of this approach compared to the others is that it is theoretically sound.

### Additional sampling step and resulting Gibbs sampler

The proposed approach is inserted seamlessly into the Gibbs sampler. It speeds up the convergence toward the target distribution *without modifying it*. The main idea is to insert an additional resampling step in order to improve the state space exploration (Equation (4)). In section 3.1, it is shown that the scale parameter  $S$  has to be resampled according to Equation (5). The resulting modified Gibbs sampler is then Algorithm 3.

From Equation (5), the posterior distribution of the scale parameter can be analytically computed for various prior distributions, as for example Gaussian priors or gamma priors. In these cases, it turns out that the posterior distribution is a Generalized Inverse Gaussian (GIG) for  $S^2$  and  $S$ , respectively. Let us also remark that the Gaussian case nicely extends to continuous or discrete Gaussian mixture models. In the general case, let us emphasize that this step boils down to sampling a scalar random variable, which is assumed a fairly tractable issue.

### Experimental results

The improvements achieved by Algorithm 3 are illustrated on a low-dimensional toy example  $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$ ,  $\mathbf{h} = (h_1, h_2) \in \mathbb{R}^2$  and  $\mathbf{z} = \mathbf{x} \cdot \mathbf{h}^T + noise \in \mathbb{R}^{2 \times 2}$ . Choosing Gaussian priors for the unknowns and the noise, it is possible to derive the theoretical posterior distributions. The level lines of these distributions are plotted on Figure 1. Figures 2, 3 and 5 show the first iterations of Algorithms 1, 2, and 3, respectively. The slow convergence of Algorithm 1 can be observed on Figure 2. Figure 3 shows that Algorithm 2 does not converge toward the theoretical posterior distribution, while Figure 4 illustrates the satisfying convergence properties of the proposed scheme (Algorithm 3). The improvements of the proposed method are also illustrated in the context on non-negative source separation [2] on Figures 6 to 9. A better exploration of the state space (Figures 7 and 8) and a faster convergence is observed (Figure 9). When focusing on estimators normalized with respect to scale, Figure 10 shows that the variance of the estimators is underestimated when using Algorithm 1.

## Conclusion

This article proposes to insert an additional step in the Gibbs sampling algorithm for solving bilinear inverse problems using the MCMC approach. This additional step corrects the bad mixing properties that is generally observed. The additional sampling step applies to a scale parameter, for which the posterior distribution is theoretically established. In practice, faster convergence is observed, with almost no increase of computational cost. The performance of the proposed solution has been compared with two common alternatives. Normalizing the scale during the sampling process (Algorithm 2) is definitely not recommendable since it modifies the target distribution and introduces bias in the variance of the estimates. The basic Gibbs sampling (Algorithm 1) requires at least to use normalized estimators. However, only the proposed method (Algorithm 3) ensures fast convergence without restrictions on the estimators. Many applications could benefit from this scale resampling, the most important being source separation and blind deconvolution.

Inverse problems, Bayesian framework, MCMC algorithms, Gibbs sampler, Scale ambiguity, Blind deconvolution, Source separation.

# 1. Introduction

Cet article s'intéresse aux algorithmes MCMC (*Markov Chain Monte Carlo*) impliqués dans la résolution de problèmes inverses bilinéaires dans un cadre bayésien. 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

$$\mathbf{z} = \mathbf{x} \star \mathbf{h} + \mathbf{b}, \quad (1)$$

où  $\mathbf{z}$  représente les observations,  $\mathbf{x}$  et  $\mathbf{h}$  les vecteurs inconnus que l'on souhaite estimer, et  $\mathbf{b}$  un vecteur de bruit. Le symbole  $\star$  désigne un opérateur bilinéaire quelconque, par exemple un produit de convolution ou un produit matriciel. La séparation de sources [1,2] et la déconvolution aveugle [3] sont des cas particuliers importants de problèmes inverses bilinéaires.

Les problèmes inverses bilinéaires présentent une ambiguïté d'échelle intrinsèque: pour tout couple  $(\mathbf{x}, \mathbf{h})$  et tout  $s \neq 0$ , on a

$$(s\mathbf{x}) \star (\mathbf{h}/s) = \mathbf{x} \star \mathbf{h}. \quad (2)$$

Par défaut, nous supposons dans la suite que  $s$  et les composantes de  $\mathbf{x}$  et de  $\mathbf{h}$  sont des quantités réelles signées, mais dans certains contextes, il s'agit de quantités positives (c'est le cas dans la Section 4), voire de nombres complexes. Dans tous les cas, la démarche proposée reste valable.

Le contexte bayésien et les algorithmes MCMC sont des outils particulièrement adaptés pour la résolution de ces problèmes

inverses bilinéaires. Toutefois, on constate en pratique un comportement pathologique des chaînes de Markov issues de l'échantillonneur de Gibbs: la convergence est lente. En particulier, l'échelle des réalisations évolue de façon très progressive. Pour accélérer la convergence, les solutions proposées jusqu'à présent ne sont pas satisfaisantes d'un point de vue théorique. Elles sont de deux types. La première approche consiste à construire des estimateurs à partir de chaînes qui n'ont pas encore atteint leur état stationnaire. La seconde approche prévoit une mise à l'échelle arbitraire qui en général n'est pas en accord avec les lois *a posteriori* des quantités inconnues. Cet article propose d'insérer dans l'échantillonneur de Gibbs une étape de rééchantillonnage d'un paramètre d'échelle conforme à la loi-cible. Cette solution présente l'avantage d'être théoriquement satisfaisante et tout à fait fonctionnelle.

La résolution de problèmes inverses bilinéaires dans un cadre bayésien est présentée dans la Section 2. La Section 3 détaille le rééchantillonnage d'un paramètre d'échelle et son insertion naturelle dans un échantillonneur de Gibbs. Pour certains exemples typiques de lois *a priori*, la Sous-section 3.4 donne les expressions de la loi *a posteriori* du paramètre d'échelle. La Section 4 présente des résultats dans le cadre d'une application: la séparation de sources positives (voir également [4] pour un autre exemple, dans le cadre de la déconvolution aveugle en IRM fonctionnelle cérébrale). Enfin, la Section 5 conclut et présente certaines perspectives.

## 2. Contexte bayésien

Le cadre bayésien permet d'introduire de l'information statistique dans un problème mal posé. Des lois *a priori* (qu'on sup-

posera propres) sont affectées aux quantités inconnues du problème. Plusieurs types d'estimateurs bayésiens sont envisageables pour retrouver  $\mathbf{x}$  et  $\mathbf{h}$  à partir des données  $\mathbf{z}$ .

D'une part, le maximum *a posteriori* (MAP) pose un problème d'optimisation non trivial, car la structure bilinéaire du modèle d'observation (1) rend la vraisemblance non convexe donc éventuellement multimodale (voir [5] pour un exemple d'une telle approche en déconvolution aveugle d'image). De plus, ce cadre n'offre pas de solution systématique pour estimer les hyperparamètres de façon non supervisée.

D'autre part, les méthodes MCMC permettent d'approcher des estimateurs à coût bayésien séparable, tel que l'espérance *a posteriori* (EAP), et offrent un cadre confortable pour estimer les hyperparamètres ou les intégrer hors du problème. Pour ces raisons, c'est la démarche que nous adopterons dans la suite. Dans ce cadre, certaines quantités inconnues  $\phi(\mathbf{x}, \mathbf{h})$  liées à  $\mathbf{x}$  et  $\mathbf{h}$  peuvent être estimées par approximation empirique de leur EAP :

$$\hat{\phi} = E[\phi(\mathbf{X}, \mathbf{H}) | \mathbf{Z} = \mathbf{z}] = \frac{1}{T} \lim_{T \rightarrow \infty} \sum_{t=1}^T \phi(\mathbf{x}^{(t)}, \mathbf{h}^{(t)}), \quad (3)$$

où les majuscules symbolisent des quantités aléatoires et  $(\mathbf{x}^{(t)}, \mathbf{h}^{(t)})$  correspondent aux réalisations de tirages aléatoires dont la loi converge vers la loi *a posteriori*. Pour estimer les vecteurs  $\mathbf{x}$  et  $\mathbf{h}$  eux-mêmes, ce choix n'est pas forcément judicieux puisqu'on peut déduire de l'ambiguïté d'échelle (2) que  $E(\mathbf{H} | \mathbf{Z} = \mathbf{z})$  et  $E(\mathbf{X} | \mathbf{Z} = \mathbf{z})$  sont des vecteurs nuls si la loi *a priori* de  $(\mathbf{X}, \mathbf{H})$  est centrosymétrique. On peut comparer ce

problème à celui du *label-switching* traité dans [6], et le résoudre par la méthode proposée dans cet article, c'est pourquoi nous éluderons cette difficulté dans la suite.

Remarquons finalement que la résolution dans un cadre bayésien lève l'ambiguïté d'échelle à travers les lois *a priori* propres choisies pour  $\mathbf{X}$  et  $\mathbf{H}$ . Néanmoins, en l'absence d'information *a priori* pertinente en ce qui concerne la véritable échelle de  $\mathbf{x}$  et  $\mathbf{h}$ , il faut garder en mémoire que l'échelle ainsi obtenue reste arbitraire.

### 2.1 Exemple illustratif

Dans un objectif didactique, nous introduisons ici un exemple simple de problème inverse à structure bilinéaire. Bien que ne correspondant pas directement à une application, il permettra ensuite une représentation plus visuelle des mécanismes d'échantillonnage comparés dans le reste de l'article.

Soit  $\mathbf{z}$  un vecteur d'observation à 4 composantes réelles défini par  $\mathbf{z} = \text{vect}(\mathbf{x}\mathbf{h}') + \sigma\mathbf{b}$ , où  $\mathbf{x}$  et  $\mathbf{h}$  sont des vecteurs colonnes à deux composantes,  $\mathbf{b}$  est un bruit et la notation  $\text{vect}(\mathbf{M})$  désigne le vecteur des colonnes concaténées de la matrice  $\mathbf{M}$ . On supposera que les 8 composantes scalaires formant  $(\mathbf{x}, \mathbf{h}, \mathbf{b})$  sont des variables gaussiennes normalisées et mutuellement indépendantes, et que l'écart-type  $\sigma$  prend une valeur positive connue. La structure et la dimension réduite de ce problème permettent en particulier le calcul de densités de probabilité telles que  $f_{H_1, X_1 | Z}$  et  $f_{H | Z}$ . Celles-ci sont représentées sur la Figure 1 (seul

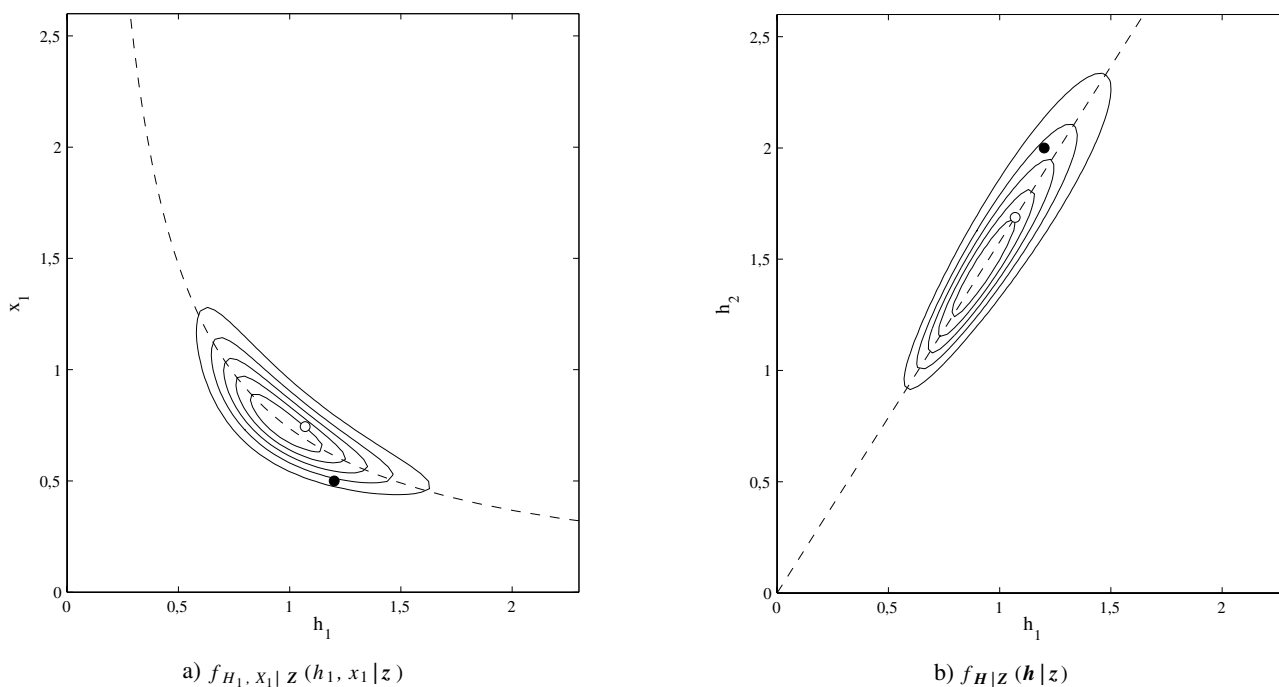


Figure 1. Densités de probabilité représentées par des lignes de niveau dans le cas de l'exemple illustratif présenté dans la Sous-section 2.1. Les 5 lignes de niveau sont équiréparties entre 0 et le maximum de la densité. Les valeurs utilisées pour générer les données sont  $\mathbf{h} = (1, 2; 2)$ ,  $\mathbf{x} = (0, 5; 1, 5)$  et  $\sigma = 0, 16$ . Les points noirs correspondent aux vraies valeurs et les points blancs à  $(E[H_1 | Z], E[X_1 | Z])$  et à  $(E[H_1 | Z], E[H_2 | Z])$ . Les lignes tiretées sont obtenues en décrivant un ensemble de couples  $(\mathbf{x}, \mathbf{h})$  qui diffèrent seulement par leur échelle.

le premier quadrant est représenté; le troisième est symétrique du premier, et la densité est très faible dans les deux autres). On notera en particulier le caractère étiré des distributions, le long d'hyperboles  $h_1 x_1 = \text{cst}$  pour  $f_{H_1, X_1 | Z}$  et le long de droites  $h_1 / h_2 = \text{cst}$  pour  $f_{H | Z}$ . Ceci reflète bien sûr le fait que l'échelle est seulement déterminée par les lois *a priori*, et ce phénomène sera d'autant plus marqué que les *a priori* sont moins informatifs que les données. La Figure 1 fait aussi apparaître les vraies valeurs de  $h_1, h_2$  et  $x_1$  et leur EAP. C'est principalement l'échelle qui distingue la vraie solution d'une estimée de type MAP ou EAP. Ceci n'est pas surprenant compte tenu du faible niveau de bruit et du fait que l'échelle de la vraie solution n'est pas contenue dans les données  $z$ .

### 2.2 Échantillonneur de Gibbs

Compte tenu de la bilinéarité du problème, il est tentant d'adopter un schéma d'échantillonnage alternant le tirage de  $h$  sachant  $x$  et le tirage de  $x$  sachant  $h$ , c'est-à-dire un échantillonneur de Gibbs par blocs, éventuellement fragmentés plus finement. L'Algorithme 1 décrit un tel échantillonneur de Gibbs par blocs incorporant un vecteur inconnu supplémentaire  $\theta$  d'hyperparamètres ou de variables cachées, lui-même échantillonné.

Dans le cas de l'exemple illustratif en dimension réduite, la Figure 2 illustre les déplacements ainsi produits dans les plans  $(h_1, x_1)$  et  $(h_1, h_2)$ . En l'absence de paramètre inconnu, l'étape 3 n'existe pas.

*Algorithme 1. Structure d'un balayage complet d'un échantillonneur de Gibbs par blocs dans le cas du modèle d'observation bilinéaire (1), en présence d'hyperparamètres ou de variables cachées  $\theta$ .*

Étant donné  $\mathbf{x}^{(t)}, \mathbf{h}^{(t)}, \boldsymbol{\theta}^{(t)}$ , générer :

1.  $\mathbf{x}^{(t+1)} \leftarrow X \sim f_{X|H,Z,\Theta}(\mathbf{x}|\mathbf{h}^{(t)}, \mathbf{z}, \boldsymbol{\theta}^{(t)})$
2.  $\mathbf{h}^{(t+1)} \leftarrow H \sim f_{H|X,Z,\Theta}(\mathbf{h}|\mathbf{x}^{(t+1)}, \mathbf{z}, \boldsymbol{\theta}^{(t)})$
3.  $\boldsymbol{\theta}^{(t+1)} \leftarrow \Theta \sim f_{\Theta|X,H,Z}(\boldsymbol{\theta}|\mathbf{x}^{(t+1)}, \mathbf{h}^{(t+1)}, \mathbf{z})$

### 2.3 Mauvaises propriétés de mélange de la chaîne; solutions existantes

La Figure 2 montre que la convergence de l'Algorithme 1 est plutôt lente: les échantillons successifs sont corrélés, en particulier en ce qui concerne leur échelle, si bien que l'ensemble des 50 premiers échantillons n'est pas encore représentatif de la loi-cible. Compte tenu de la faible dimension du problème, ce résultat est un symptôme de l'inefficacité de cet algorithme. Celle-ci devient patente pour des problèmes de plus grande taille, tel que celui présenté en Section 4. 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, mais en prenant soin de

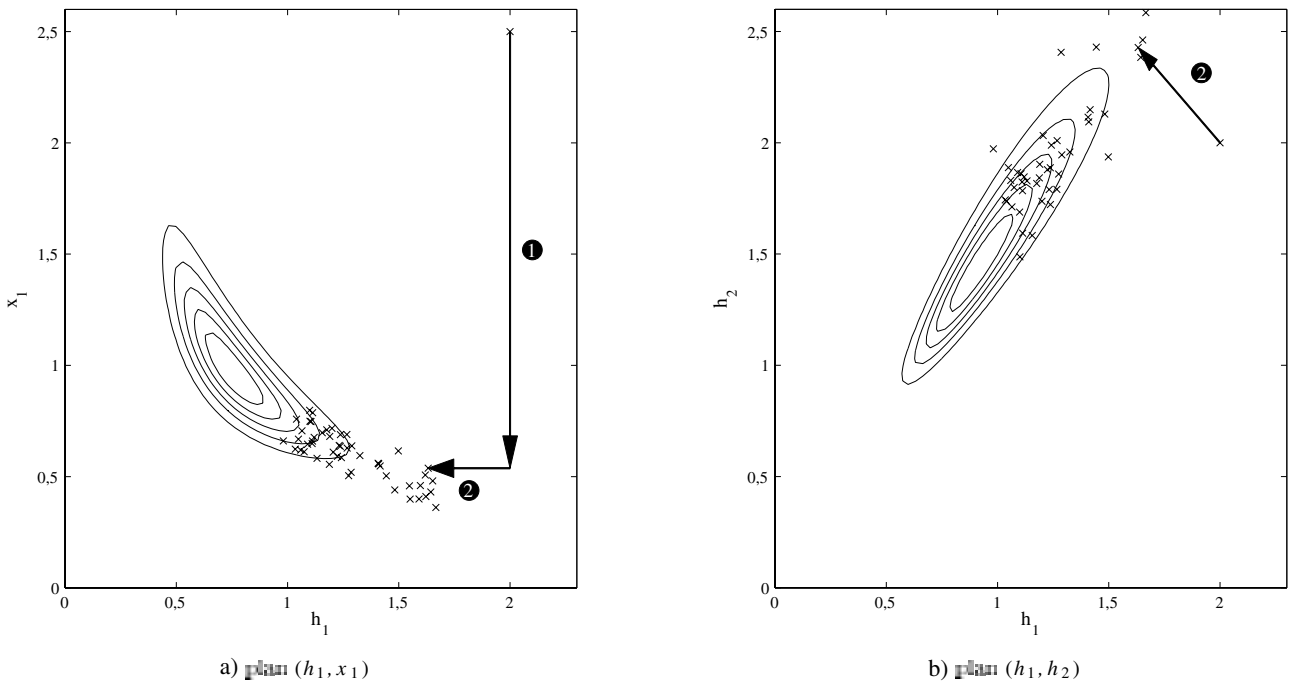


Figure 2. Résultat de l'Algorithme 1 pour l'exemple illustratif de la Figure 1. L'initialisation a été fixée arbitrairement à  $\mathbf{x}^{(1)} = (2,5; 2,5)$  et  $\mathbf{h}^{(1)} = (2; 2)$ . Les flèches 1 et 2 correspondent au premier déplacement produit par les étapes 1 et 2 de l'échantillonneur de Gibbs. Les croix représentent le résultat des 50 premières itérations. Cinq lignes de niveau des distributions a posteriori ont été superposées.

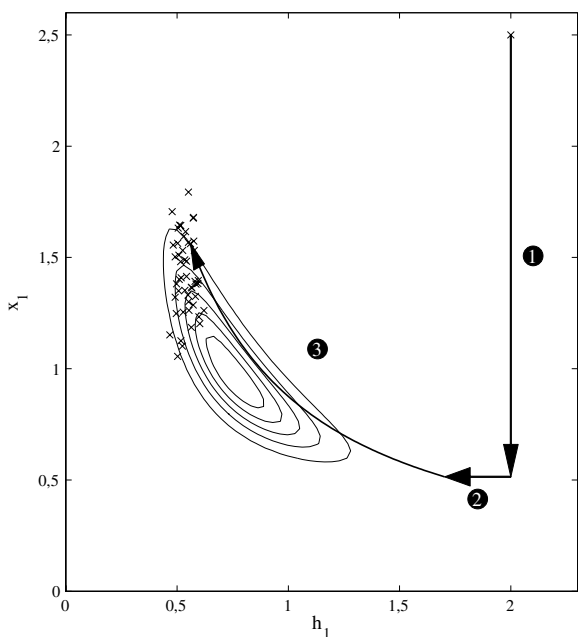
se limiter à des quantités normalisées indépendantes de l'échelle (par exemple,  $\mathbf{h}/\|\mathbf{h}\|$  ou  $\mathbf{h} \times \|\mathbf{x}\|$ ). En effet, on peut raisonnablement espérer que la convergence reste suffisamment rapide quand on fait abstraction de l'échelle (dont la valeur estimée n'est pas réellement pertinente de toute façon).

L'autre solution, plus radicale, consiste à renormaliser à une échelle fixée le couple  $(\mathbf{x}, \mathbf{h})$  à chaque itération de l'échantillonneur, conformément à l'Algorithme 2 dans le cas d'une normalisation impliquant  $\|\mathbf{h}\|$  (imposant ainsi une norme constante aux réalisations successives de  $\mathbf{h}$ ; d'autres choix de normalisation sont évidemment possibles). Pour l'exemple illustratif, la Figure 3 montre les déplacements correspondants aux étapes successives de l'Algorithme 2 pour  $\rho = 1$ . Mise à part la valeur initiale, l'ensemble des réalisations de  $\mathbf{h}^{(t)}$  (pour les valeurs entières de  $t$ ) se trouvent donc sur le cercle unité.

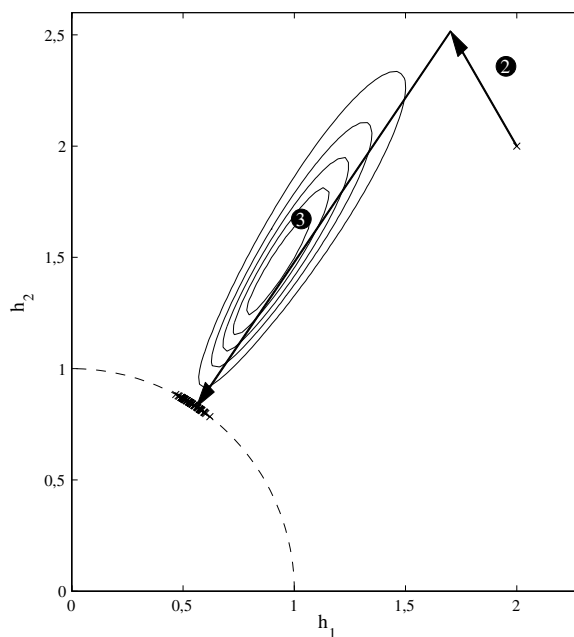
Algorithme 2. Variante de l'Algorithme 1 dans laquelle une normalisation déterministe (ici,  $\|\mathbf{h}^{(t)}\| = \rho$ ) est effectuée à chaque itération pour s'affranchir des problèmes d'échelle.

Étant donné  $\mathbf{x}^{(t)}, \mathbf{h}^{(t)}, \boldsymbol{\theta}^{(t)}$ , générer :

1.  $\mathbf{x}^{(t+1/2)} \leftarrow \mathbf{X} \sim f_{\mathbf{X}|\mathbf{H},\mathbf{Z},\boldsymbol{\theta}}(\mathbf{x}|\mathbf{h}^{(t)}, \mathbf{z}, \boldsymbol{\theta}^{(t)})$
2.  $\mathbf{h}^{(t+1/2)} \leftarrow \mathbf{H} \sim f_{\mathbf{H}|\mathbf{X},\mathbf{Z},\boldsymbol{\theta}}(\mathbf{h}|\mathbf{x}^{(t+1/2)}, \mathbf{z}, \boldsymbol{\theta}^{(t)})$
3.  $\mathbf{x}^{(t+1)} \leftarrow \mathbf{x}^{(t+1/2)} \times \|\mathbf{h}^{(t+1/2)}\|/\rho$   
 $\mathbf{h}^{(t+1)} \leftarrow \mathbf{h}^{(t+1/2)} \times \rho/\|\mathbf{h}^{(t+1/2)}\|$
4.  $\boldsymbol{\theta}^{(t+1)} \leftarrow \boldsymbol{\Theta} \sim f_{\boldsymbol{\Theta}|\mathbf{X},\mathbf{H},\mathbf{Z}}(\boldsymbol{\theta}|\mathbf{x}^{(t+1)}, \mathbf{h}^{(t+1)}, \mathbf{z})$



a) plan  $(h_1, x_1)$



b) plan  $(h_1, h_2)$

Figure 3. Algorithme 2: exemple illustratif de la Figure 1. Les flèches 1, 2 et 3 correspondent au premier déplacement produit par les étapes 1, 2 et 3 de l'Algorithme 2. Les croix représentent les résultats obtenus pour  $t = 1, \dots, 50$ .

Pour plusieurs raisons, cette solution radicale est peu recommandable.

- Elle résoud de façon approchée un problème d'échantillonnage *contraint*, le tirage aléatoire de  $(\mathbf{X}, \mathbf{H})$  sachant  $\mathbf{z}$  et  $\|\mathbf{H}\| = \rho$ . Il en résulte que la loi des échantillons ne converge plus vers une loi-cible clairement spécifiée (cf. Figure 3).

- La contrainte introduite est un nouvel arbitraire, redondant avec le choix des *a priori* du point de vue de la levée de l'indétermination de l'échelle.

- Elle biaise les calculs d'espérance.

Pour illustrer ces inconvénients, la Figure 4 analyse des résultats d'estimation de la quantité normalisée  $|h_1|/\|\mathbf{h}\|$ .

La Figure 4(a) indique clairement que l'EAP contraint par  $\|\mathbf{H}\| = \rho$  varie sensiblement avec  $\rho$ . La variabilité est encore plus marquée sur l'écart-type de la loi *a posteriori* contrainte (reporté sous forme de barres d'erreur sur la Figure 4(a)). Les espérances mathématiques ont été calculées par intégration directe, de façon partiellement analytique et partiellement numérique.

Sur la Figure 4(b), l'EAP et l'écart-type de la loi *a posteriori* contrainte ont été remplacés par des moyennes empiriques d'échantillons issus de l'Algorithme 2. Les résultats sont intermédiaires entre les valeurs obtenues sans contrainte et avec contrainte. En particulier, l'estimateur est proche de la version non contrainte pour toutes les valeurs de  $\rho$ , en revanche sa variance reste nettement surestimée pour les petites valeurs de  $\rho$ . Ceci confirme que la loi échantillonnée ne s'identifie ni à la loi-cible, ni à la loi contrainte.

On peut aussi remarquer que les résultats obtenus dans les deux cas pour  $\rho$  « suffisamment grand » s'améliorent et tendent vers



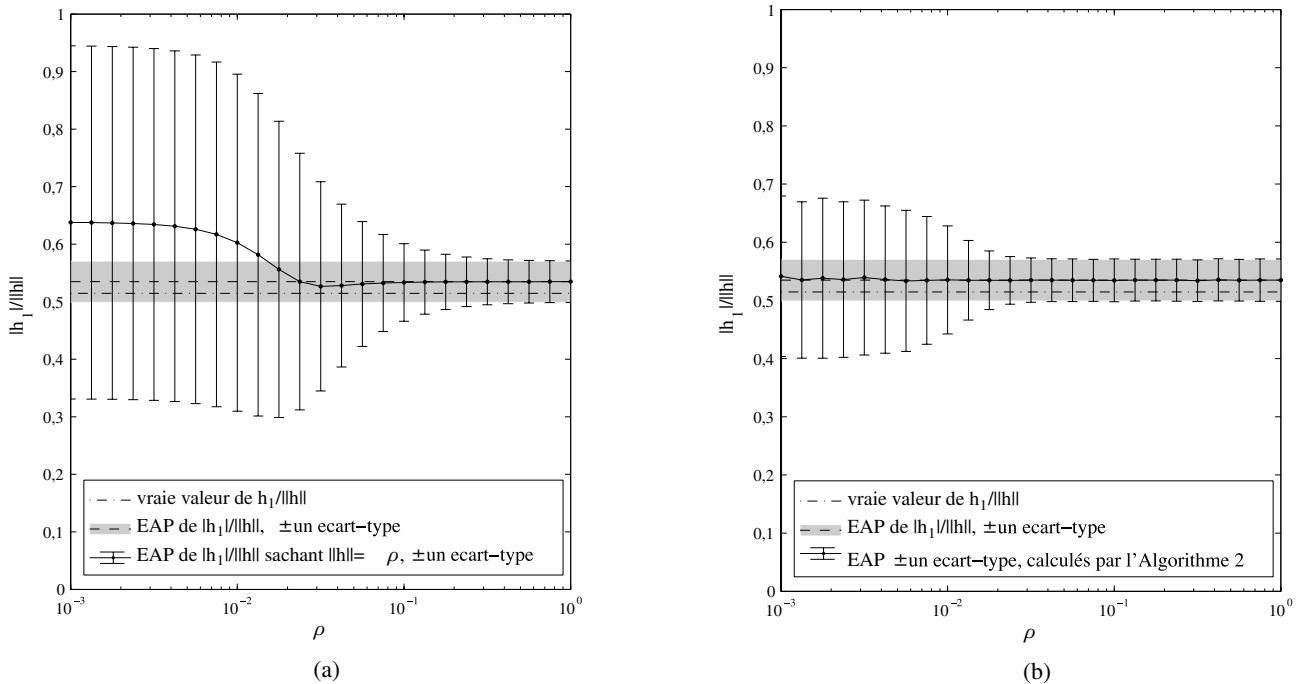


Figure 4. Comparaison d'estimateurs de la quantité normalisée  $|h_1|/||\mathbf{h}||$  : (a) EAP  $E[|h_1|/||\mathbf{h}|| | \mathbf{z}]$ , EAP contraint  $E[|h_1|/||\mathbf{h}|| | \mathbf{z}, ||\mathbf{h}|| = \rho]$ ; (b) EAP  $E[|h_1|/||\mathbf{h}|| | \mathbf{z}]$ , moyenne empirique renormalisée  $(\sum_t h^{(t)})/(\rho T)$  des échantillons issus de l'Algorithme 2 pour la contrainte  $||\mathbf{h}|| = \rho$ . Les écarts-type des lois ont également été calculés et représentés sous la forme de barres d'erreur.

ceux obtenus sans contrainte. Ce résultat peu intuitif pourrait être approfondi, mais il est probable que l'estimation de  $\mathbf{x}$  se dégrade tandis que celle de  $\mathbf{h}$  s'améliore.

### 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

$$\mathbf{X} = \mathbf{X}^{\text{old}} \times S, \quad \mathbf{H} = \mathbf{H}^{\text{old}} / S, \quad (4)$$

où  $S$  est un paramètre d'échelle qui est lui-même échantillonné à chaque itération. Sur la Figure 3 dans le cas de l'exemple simple, ce déplacement correspond à un déplacement le long d'une hyperbole  $x_1 h_1 = \text{constante}$ . Comme dans l'Algorithme 2, nous proposons d'insérer une étape de changement d'échelle. Mais plutôt que de l'imposer de manière arbitraire, le

paramètre d'échelle est rééchantillonné en accord avec les lois *a priori* choisies. Il reste à établir selon quelle loi rééchantillonner le paramètre d'échelle  $S$ .

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

$$f_{S|reste}(s) \propto |s|^{M-P-1} f_{\mathbf{X}, \mathbf{H} | \Theta}(s \mathbf{x}^{\text{old}}, \mathbf{h}^{\text{old}} / s | \Theta), \quad (5)$$

où  $M$  et  $P$  sont les tailles respectives de  $\mathbf{X}$  et  $\mathbf{H}$ . Notons que l'expression (5) se factorise sous la forme

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

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

Dans la suite, la dépendance vis-à-vis du vecteur  $\Theta$  sera omise pour alléger la démonstration dans la mesure où  $\Theta$  reste constant lors du changement d'échelle.

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

Définissons un changement de variable  $\mathcal{T}$  sur le couple  $(\mathbf{X}, \mathbf{H})$  qui permet de faire apparaître explicitement le paramètre d'échelle  $S$ . Dans ce changement de variable, la quantité  $c = x_1^{\text{old}}$ , première composante du vecteur  $\mathbf{x}$  avant le change-

ment d'échelle, joue le rôle d'une constante, supposée non nulle pour l'instant.

$$\mathcal{T} : (\mathbf{x}, \mathbf{h}) \mapsto (s, \mathbf{v}) \quad \text{avec } s = \frac{x_1}{c} \text{ et } \mathbf{v} = \begin{bmatrix} (\mathbf{x}_{2:M})/x_1 \\ \mathbf{h} \times x_1 \end{bmatrix} \quad (6)$$

en notant  $\mathbf{x}_{m:n} = [x_m, \dots, x_n]^t$ . Le vecteur  $\mathbf{v}$ , de taille  $M + P - 1$ , est construit de façon à ne pas être sensible à l'échelle de  $(\mathbf{x}, \mathbf{h})$ , au sens où seule la première des  $M + P$  composantes de  $\mathcal{T}(\alpha\mathbf{x}, \mathbf{h}/\alpha)$  dépend de  $\alpha$ .

Il est facile de vérifier que  $\mathcal{T}$  est inversible, et que  $\mathcal{T}^{-1}$  s'exprime de la façon suivante :

$$\mathcal{T}^{-1} : (s, \mathbf{v}) \mapsto (\mathbf{x}, \mathbf{h}) \quad \text{avec } \mathbf{x} = (s \cdot c) \times \begin{bmatrix} 1 \\ \mathbf{v}_{1:M-1} \end{bmatrix} \quad (7)$$

et  $\mathbf{h} = \frac{\mathbf{v}_{M:M+P-1}}{(s \cdot c)}$

Dans la suite, la matrice jacobienne et le jacobien d'un difféomorphisme  $f$  sont notés respectivement  $\mathbf{J}_f$  et  $J_f$ . En particulier,  $\mathbf{J}_{\mathcal{T}^{-1}}$  est triangulaire inférieure :

$$\mathbf{J}_{\mathcal{T}^{-1}}(s, \mathbf{v}) = \begin{bmatrix} c & \mathbf{O}_{1, M-1} & \mathbf{O}_{1, P} \\ \times & (s \cdot c) \mathbf{I}_{M-1} & \mathbf{O}_{M-1, P} \\ \times & \times & \mathbf{I}_P / (s \cdot c) \end{bmatrix}$$

en notant  $\mathbf{I}_N$  et  $\mathbf{O}_{MN}$  la matrice identité de taille  $N \times N$  et la matrice nulle de taille  $M \times N$ , respectivement. On a donc  $J_{\mathcal{T}^{-1}}(s, \mathbf{v}) = c^{M-P} s^{M-P-1}$ , d'où

$$f_{S, \mathbf{V}}(s, \mathbf{v}) = |c|^{M-P} |s|^{M-P-1} f_{\mathbf{X}, \mathbf{H}}(\mathcal{T}^{-1}(s, \mathbf{v})). \quad (8)$$

La loi de  $S$  conditionnellement à  $\mathbf{V}$  (c'est-à-dire à  $(\mathbf{X}, \mathbf{H})$ , sans tenir compte de l'échelle courante) se déduit de façon immédiate, à une constante près :

$$f_{S|\mathbf{V}}(s|\mathbf{v}) \propto f_{S, \mathbf{V}}(s, \mathbf{v}). \quad (9)$$

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

Montrons maintenant l'indépendance de  $S$  par rapport aux observations  $\mathbf{Z}$  conditionnellement à  $\mathbf{V}$ . Ceci résulte de la bilinéarité de l'équation d'observation (1), à travers l'identité  $(\mathbf{x} \times s) \star (\mathbf{h}/s) = \mathbf{x} \star \mathbf{h}$ . En effet on a :

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

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

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

On en déduit :

$$f_{S|\mathbf{V}, \mathbf{Z}}(s|\mathbf{v}, \mathbf{z}) = f_{S|\mathbf{V}}(s|\mathbf{v}) \quad (10)$$

ce qui permet d'obtenir (5), en utilisant (10), (9) et (8).

### 3.3 Échantillonneur de Gibbs modifié

L'Algorithme 3 comporte à l'identique les trois étapes d'échantillonnage de Gibbs de l'Algorithme 1, auxquelles une étape supplémentaire a été adjointe. Cette étape intercalée s'appuie sur l'expression de la loi conditionnelle de l'échelle (5). La Figure 5 montre que le mouvement supplémentaire provoque une évolution rapide de l'échelle. Il en résulte une exploration beaucoup plus efficace de la loi-cible.

**Remarque 1.** Dans ce nouvel algorithme, l'étape 3 peut s'interpréter comme la remise à jour de la variable  $S$  par un échantillonneur de Gibbs, après le changement de variable  $\mathcal{T} : (\mathbf{X}, \mathbf{H}) \mapsto (S, \mathbf{V})$ . Cette propriété permet de déduire que l'Algorithme 3 converge vers la même loi-cible  $f_{\mathbf{X}, \mathbf{H}, \Theta|\mathbf{Z}}$  que l'Algorithme 1. Une démonstration plus prosaïque consiste à interpréter l'étape 3 comme une proposition dans un algorithme de Metropolis-Hastings, puis à vérifier que la probabilité d'acceptation vaut 1. Ce calcul n'est pas aussi simple qu'il n'y paraît car la proposition a une densité de probabilité scalaire, alors que la loi-cible a une densité dans  $\mathbb{R}^{M+P}$ . Comme le démontre [7], l'expression de la probabilité d'acceptation fait alors intervenir le jacobien d'une transformation, comme dans le cas plus général des sauts réversibles. Dans un objectif didactique, le calcul est proposé en Annexe A.

**Remarque 2.** Dans l'établissement de la loi (5), nous avons supposé que la première composante courante de  $\mathbf{x}$  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  $\mathbf{x}$ , ou même de  $\mathbf{h}$ , était non nulle. Le seul cas impossible à gérer est celui où toutes les composantes de  $\mathbf{x}$  et de  $\mathbf{h}$  sont nulles. Dans ce cas très particulier, (5)

*Algorithme 3. Variante de l'Algorithme 1 proposée : la normalisation effectuée à chaque itération fait intervenir une variable d'échelle qui elle-même est échantillonnée en conformité avec la loi-cible.*

Étant donné  $\mathbf{x}^{(t)}, \mathbf{h}^{(t)}, \boldsymbol{\theta}^{(t)}$ , générer :

1.  $\mathbf{x}^{(t+1/2)} \leftarrow \mathbf{X} \sim f_{\mathbf{X}|\mathbf{H}, \mathbf{Z}, \Theta}(\mathbf{x}|\mathbf{h}^{(t)}, \mathbf{z}, \boldsymbol{\theta}^{(t)})$
2.  $\mathbf{h}^{(t+1/2)} \leftarrow \mathbf{H} \sim f_{\mathbf{H}|\mathbf{X}, \mathbf{Z}, \Theta}(\mathbf{h}|\mathbf{x}^{(t+1/2)}, \mathbf{z}, \boldsymbol{\theta}^{(t)})$
3. Si  $\mathbf{x}^{(t+1/2)} \neq \mathbf{0}$  ou  $\mathbf{h}^{(t+1/2)} \neq \mathbf{0}$   
 $s \leftarrow S \times f_{S|reste}(s)$  d'après (5), puis  
 $\mathbf{x}^{(t+1)} \leftarrow s\mathbf{x}^{(t+1/2)}$  et  $\mathbf{h}^{(t+1)} \leftarrow \mathbf{h}^{(t+1/2)}/s$
4.  $\boldsymbol{\theta}^{(t+1)} \leftarrow \Theta \sim f_{\Theta|\mathbf{X}, \mathbf{H}, \mathbf{Z}}(\boldsymbol{\theta}|\mathbf{x}^{(t+1)}, \mathbf{h}^{(t+1)}, \mathbf{z})$



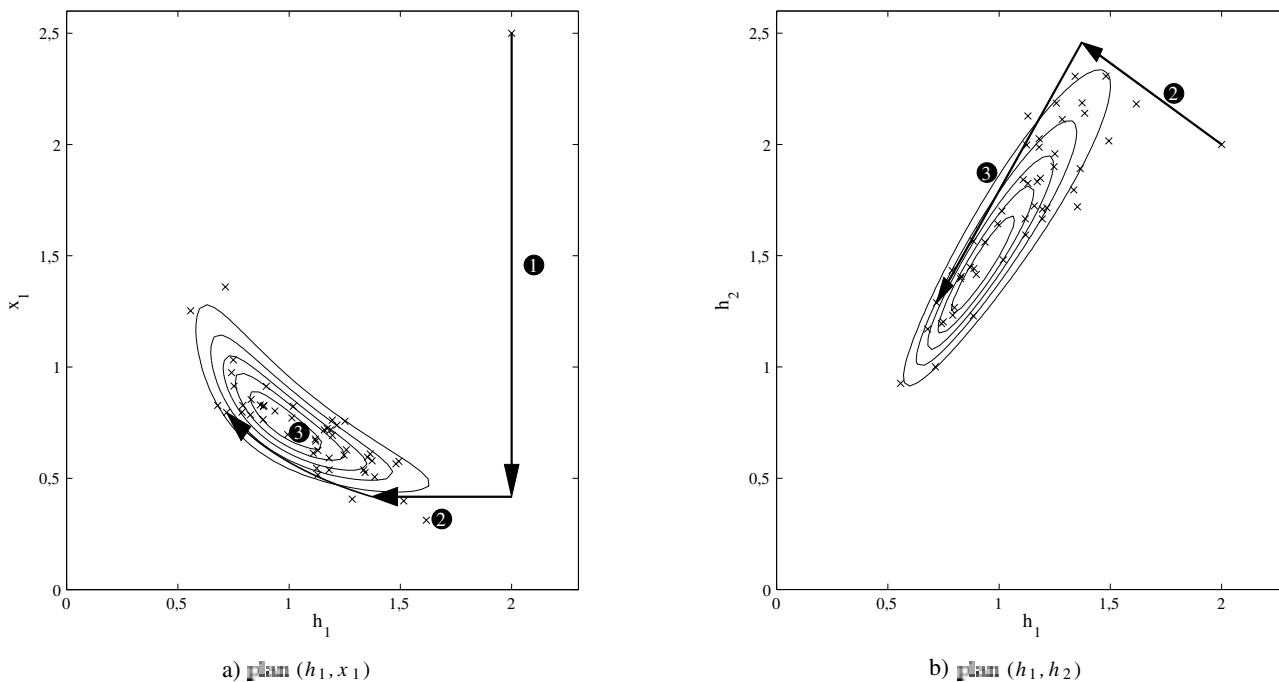


Figure 5. Algorithme 3: exemple illustratif,  $\mathbf{x} \in \mathbb{R}^2$ ,  $\mathbf{h} \in \mathbb{R}^2$ , lois a priori gaussiennes indépendantes. Les trajectoires 1, 2 et 3 correspondent aux déplacements des étapes 1 (rééchantillonnage de  $\mathbf{x}$ ), 2 (rééchantillonnage de  $\mathbf{h}$ ) et 3 (rééchantillonnage d'un paramètre d'échelle  $s$ ) de l'échantillonneur de Gibbs. Les croix représentent les 50 premières itérations de l'échantillonneur de Gibbs.

n'a pas de sens, mais il est alors inutile de rééchantillonner l'échelle au sens de (4).

La suite de l'article s'intéresse à la mise en œuvre pratique de l'étape de rééchantillonnage de l'échelle.

### 3.4 Exemple de loi a posteriori pour l'échelle

L'expression générale (5) se décline selon les lois a priori considérées. À titre d'exemple, nous examinons ici quelques cas particuliers courants, aboutissant à une loi conditionnelle de l'échelle dont l'expression et l'échantillonnage sont documentés. Dans d'autres situations, cette loi n'est pas nécessairement aussi bien maîtrisée. Il faut alors résoudre le problème de l'échantillonnage d'une loi scalaire, de forme connue à une constante près. Des techniques de type acceptation-rejet ou Metropolis-Hastings sont alors à envisager.

Cas d'un a priori gaussien

Considérons le cas d'a priori gaussiens centrés  $\mathcal{N}(0, \Sigma_X)$  sur  $\mathbf{X}$  et  $\mathcal{N}(0, \Sigma_H)$  sur  $\mathbf{H}$ , avec  $\mathbf{X}$  et  $\mathbf{H}$  indépendants a priori. À partir de l'Équation (5) et en utilisant le changement de variable  $R = S^2$ , on obtient :

$$f_{R|reste}(r) \propto r^{\frac{1}{2}(M-P)-1} \exp\left(-\frac{1}{2}(\mathbf{h}^t \Sigma_H^{-1} \mathbf{h} r^{-1} + \mathbf{x}^t \Sigma_X^{-1} \mathbf{x} r)\right)$$

On en déduit que  $S^2$  suit une loi conditionnelle de type Gaussienne Inverse Généralisée (Generalized Inverse Gaussian, GIG) de paramètres

$$\lambda = \frac{M-P}{2}, \quad \alpha = \mathbf{x}^t \Sigma_X^{-1} \mathbf{x}, \quad \beta = \mathbf{h}^t \Sigma_H^{-1} \mathbf{h}.$$

La densité de la loi GIG s'écrit :

$$f_{GIG}(z) = \left(\frac{\alpha}{\beta}\right)^{\lambda/2} \frac{z^{\lambda-1}}{2K_\lambda(\sqrt{\alpha\beta})} \exp\left(-\frac{1}{2}(\beta z^{-1} + \alpha z)\right),$$

$$z > 0, \quad \lambda \in \mathbb{R}, \quad \begin{cases} \alpha > 0, \beta \geq 0 & \text{pour } \lambda > 0 \\ \alpha > 0, \beta > 0 & \text{pour } \lambda = 0 \\ \alpha \geq 0, \beta > 0 & \text{pour } \lambda < 0 \end{cases}$$

où  $K_\lambda(\cdot)$  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 [8]. On obtient ensuite  $S$  sous la forme  $S = U \times \sqrt{R}$ , avec  $\Pr(U = -1) = \Pr(U = 1) = 1/2$  puisque la loi de  $S$  est symétrique. C'est ainsi que l'étape 3 de l'Algorithme 3 a été implantée pour générer la Figure 5, à une nuance près : de façon à maintenir les couples  $(h_1, x_1)$  et  $(h_1, h_2)$  dans le premier quadrant, le rééchantillonnage du signe de l'échelle n'a pas été effectué.

Notons que l'extension à des modèles de mélanges discret ou continu est relativement aisée. En particulier, les paramètres de la loi de  $S$  pour le modèle Bernoulli-Gaussien utilisé dans [3] sont précisés dans [9].

Cas d'*a priori* gamma

Considérons le cas de variables positives décrites par des lois *a priori* indépendantes de type gamma. Plus précisément, supposons que  $\mathbf{X}$  et  $\mathbf{H}$  sont indépendants, et que chaque composante de  $\mathbf{X}$  et de  $\mathbf{H}$  suit une loi gamma de paramètres  $(a,b)$  et  $(c,d)$ , respectivement. Rappelons que la densité d'une loi gamma de paramètre  $(a,b)$  s'écrit :

$$f(x) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx} \mathbf{1}_{\mathbb{R}_+}(x), \quad \alpha, \beta > 0.$$

Il est facile d'en déduire que  $S$  suit une loi GIG de paramètres

$$\lambda = Ma - Pc, \quad \alpha = 2b \sum_{m=1}^M x_m, \quad \beta = 2d \sum_{p=1}^P h_p.$$

La Section 4 présente une application de ce résultat dans le cas de la séparation de sources.

## 4. Application à la séparation de sources par approche bayésienne

La séparation de sources est un problème inverse bilinéaire soumis à l'indétermination d'échelle. Le modèle direct dans le cas d'un mélange linéaire instantané s'écrit sous la forme :

$$\mathbf{z}_i = \sum_{j=1}^p h_{ij} \mathbf{x}_j + \mathbf{b}_i \quad \forall i = 1, \dots, m, \quad (11)$$

où  $m$  représente le nombre d'observations  $\mathbf{z}_i \in \mathbb{R}^n$ ,  $n$  étant le nombre d'échantillons par observation et  $p$  le nombre de sources  $\mathbf{x}_j$ . On note également  $\mathbf{h}_j \in \mathbb{R}^m$  le vecteur contenant les coefficients de mélange de la  $j$ ème source dans toutes les observations. Ce modèle direct peut être réécrit sous la forme matricielle suivante :

$$\mathbf{Z} = \mathbf{H}\mathbf{X} + \mathbf{B} \quad (12)$$

où  $\mathbf{Z} \in \mathbb{R}^{m \times n}$  est la matrice des observations,  $\mathbf{H} \in \mathbb{R}^{m \times p}$  est la matrice de mélange et  $\mathbf{X} \in \mathbb{R}^{p \times n}$  est la matrice des sources.

### 4.1 Échantillonnage de l'échelle séparé pour chaque source

Nous pouvons constater dans l'équation (11) que la multiplication de chaque source  $\mathbf{x}_j$  par une constante  $s_j$  et la division des coefficients de mélange  $\mathbf{h}_j$  par les mêmes constantes permet de reproduire exactement le même résultat. Cette indétermination

d'échelle peut s'exprimer également en rajoutant une matrice diagonale  $\mathbf{S} = \text{diag}(s_1, \dots, s_p)$  et son inverse  $\mathbf{S}^{-1} = \text{diag}(1/s_1, \dots, 1/s_p)$  dans l'équation (12) de la manière suivante :

$$\mathbf{Z} = (\mathbf{H}\mathbf{S}^{-1})(\mathbf{S}\mathbf{X}) + \mathbf{B} = \tilde{\mathbf{H}}\tilde{\mathbf{X}} + \mathbf{B}. \quad (13)$$

Il existe donc une variable d'échelle par source et dont l'échantillonnage va dépendre du choix des lois *a priori* sur la distribution des signaux sources et des coefficients de mélange.

Plusieurs auteurs [10,11,12,2,13,1] ont proposé des méthodes de séparation de sources par approche bayésienne, en employant des méthodes MCMC et différents modèles probabilistes sur les sources et les coefficients de mélange. Dans le cas de lois *a priori* très complexes, par exemple la loi hyperbolique généralisée [1] ou de Student [13], 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. L'*a priori* sur les coefficients du mélange étant supposé gaussien, la loi conditionnelle de  $S^2$  est une GIG. Nous considérons, à titre d'illustration, un exemple plus simple qui est celui d'*a priori* gamma sur les sources et sur les coefficients de mélange.

### 4.2 Séparation de sources positives

Intéressons nous au cas où les signaux sources et les coefficients de mélange doivent satisfaire la contrainte de positivité (ou non-négativité). Une méthode présentée dans [2] suppose des sources i.i.d, mutuellement indépendantes et distribuées selon des lois gamma. Le même modèle est également supposé pour les coefficients de mélange :

$$\mathbf{x}_j \sim \prod_{k=1}^n \mathcal{G}(x_{jk}; a_j, b_j) \quad \text{et} \quad \mathbf{h}_j \sim \prod_{i=1}^m \mathcal{G}(h_{ij}; c_j, d_j) \quad (14)$$

À partir de ce modèle et conformément à ce qui a été présenté dans la Section 3.4, nous pouvons déduire que la loi conditionnelle de chaque variable  $S_j$  est une loi GIG dont les paramètres dépendent du quadruplet  $(a_j, b_j, c_j, d_j)$  :

$$S_j | \mathbf{x}_j, \mathbf{h}_j \sim \mathcal{GIG}(s_j; \lambda_j, \alpha_j, \beta_j)$$

avec

$$\lambda_j = na_j - mc_j, \quad \alpha_j = 2b_j \sum_{k=1}^n x_{jk}, \quad \beta_j = 2d_j \sum_{i=1}^m h_{ij}.$$

### 4.3 Résultats expérimentaux

Afin d'étudier le comportement de l'échantillonneur de Gibbs en fonction de la stratégie d'échelle dans le cas de la séparation

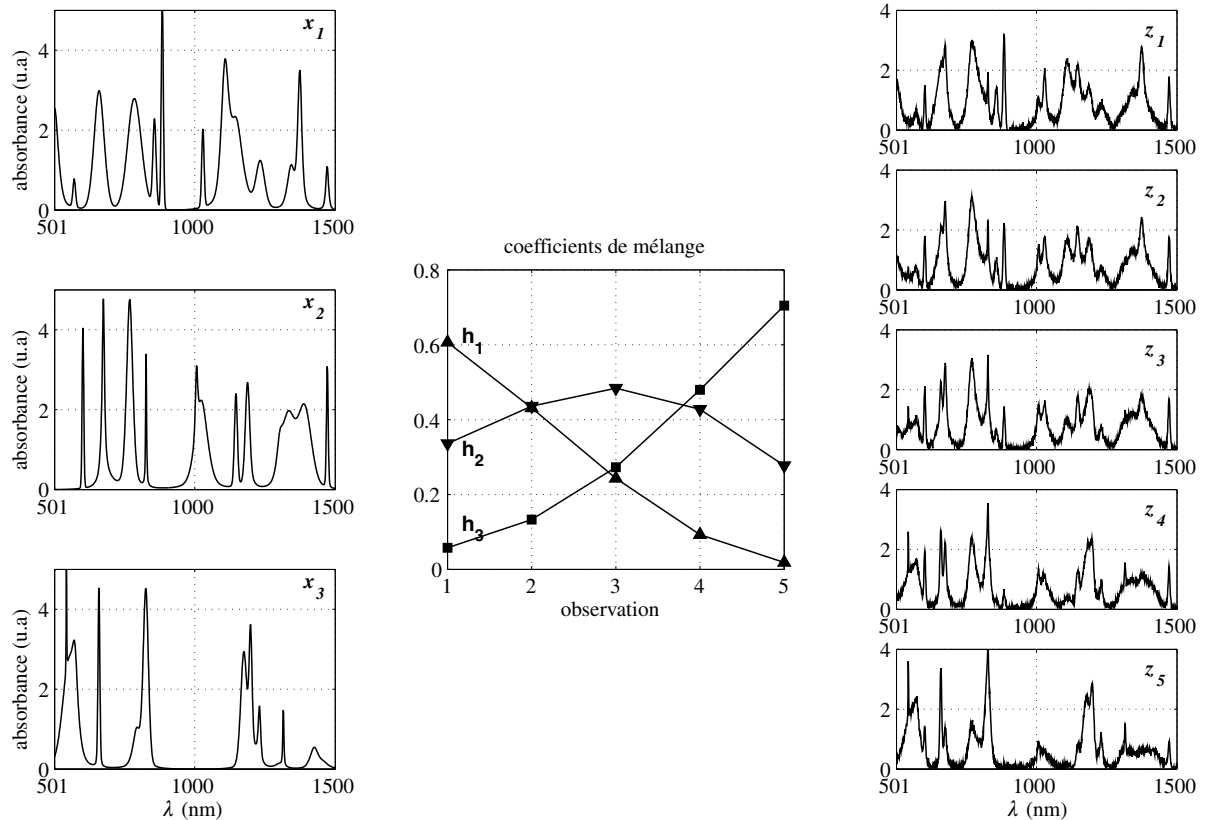


Figure 6. Exemple de mélange synthétique : (gauche) sources spectrales, (centre) coefficients de mélange et (droite) observations avec un RSB de 15 dB.

de sources positives, considérons un exemple simplifié issu de la spectrométrie. Les sources utilisées sont de formes similaires aux spectres d'absorption obtenus lors de l'analyse de substances chimiques multicomposantes [14]. La figure 6 illustre les sources et les coefficients de mélange pour  $p = 3$  sources,  $n = 1000$  échantillons et  $m = 5$  observations et un rapport signal à bruit de 15 dB sur chaque observation.

La simulation de la loi *a posteriori* est effectuée en utilisant un algorithme de Gibbs avec différentes stratégies d'échelle (Algorithmes 1, 2 et 3). L'algorithme 2 est mis en œuvre en normalisant les sources à une variance unité. Les trois algorithmes sont initialisés au même point et l'échantillonneur de Gibbs est

lancé sur  $10^5$  itérations. De plus, les lois *a priori* des hyperparamètres des sources sont ajustées de façon à obtenir une variance unité pour chaque source.

#### Analyse de chaînes de Markov

Nous limitons ici notre analyse au comportement de l'échantillonneur pour les paramètres de la première source, mais le constat reste valable pour les autres sources. La figure 7 montre l'évolution de la norme  $\ell_2$  des coefficients de mélange de la première source. Conformément aux analyses de la Sous-section 2.3, nous pouvons constater que le rééchantillonnage de l'échel-

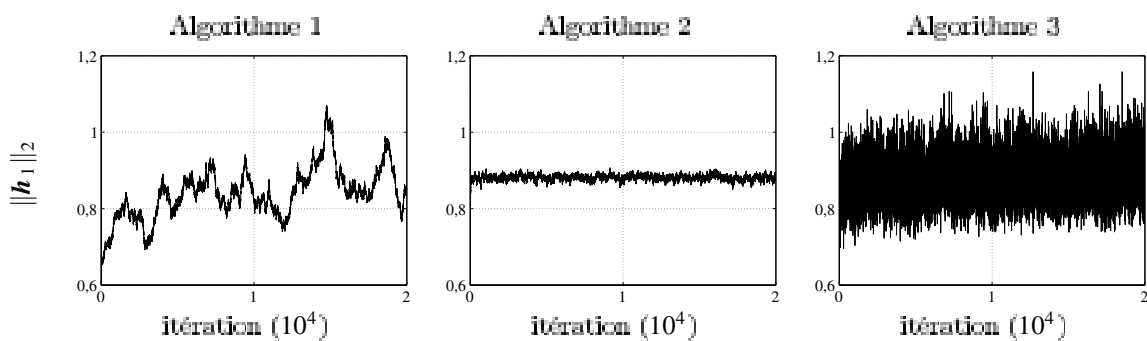


Figure 7. Évolution de la norme  $\ell_2$  des coefficients de mélange de la première source pour différentes stratégies d'échelle.

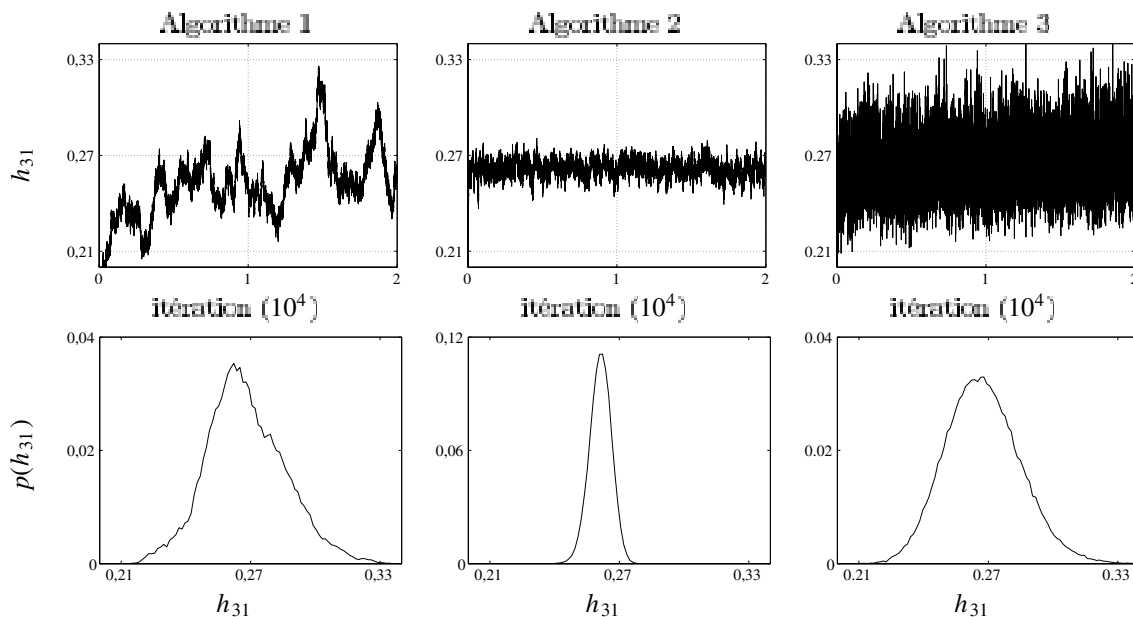


Figure 8. (haut) Chaînes de Markov du coefficient  $h_{3,1}$  simulées en utilisant les Algorithmes 1, 2 et 3. (bas) distributions empiriques estimées après  $10^5$  itérations de Gibbs.

le entraîne une dynamique plus rapide de ses réalisations et une exploration du même domaine que celui sans rééchantillonnage de l'échelle. Par ailleurs, la normalisation fixe pour avoir des sources à variance unité entraîne une réduction du domaine de variation de celle-ci, ce qui est révélateur d'une modification de la loi cible échantillonnée. La Figure 8 illustre l'évolution de la chaîne de Markov et les distributions empiriques d'un coefficient de mélange, choisi arbitrairement. Sur cette figure, nous pouvons constater que la normalisation arbitraire des sources à chaque itération (Algorithme 2) a comme conséquence une diminution de la variance de celle-ci. Par ailleurs, le rééchantillonnage de l'échelle (Algorithme 3) conduit à une chaîne de Markov moins corrélée et qui explore le même domaine que lorsque l'échelle est libre (Algorithme 1).

Convergence de l'échantillonneur et résultats de l'estimation

Un point important concerne la vitesse de convergence des algorithmes MCMC. Afin d'évaluer la vitesse de convergence nous employons le test de convergence proposé par Gelman et Rubin [15]. Il s'agit de lancer en parallèle plusieurs chaînes de Markov, de choisir un paramètre  $\phi$  à contrôler, puis de calculer à chaque itération  $t$  de l'échantillonneur la variance inter-chaines  $B_\phi(t)$  et la moyenne des variances intra-chaines  $W_\phi(t)$  pour ce paramètre afin d'en déduire le facteur :

$$\widehat{G}_\phi(t) = \sqrt{\frac{t-1}{t} + \frac{1}{t} \frac{B_\phi(t)}{W_\phi(t)}}$$

qui indique une convergence quand sa valeur est proche de 1. La Figure 9 résume le résultat du test de convergence réalisé sur le coefficient de mélange  $h_{3,1}$  dans le cas de la simulation en utilisant les Algorithmes 1 et 3. Pour chaque algorithme, 10 chaînes

sont lancées en parallèle à partir de 10 points d'initialisation différents. Le facteur  $\widehat{G}_\phi(t)$  est calculé sur les réalisations du coefficient de mélange ainsi que sur les réalisations normalisées  $h_{3,1}^{(t)} / \|h_1^{(t)}\|$  pour s'affranchir de la dépendance vis-à-vis de l'échelle. Nous pouvons constater une convergence très lente de l'Algorithme 1 dans le premier cas, tandis qu'elle est seulement un peu plus lente que celle de l'Algorithme 3 dans le second cas. L'Algorithme 1, en prenant la précaution d'estimer des quantités indépendantes de l'échelle, peut paraître une alternative viable. Toutefois, la normalisation des réalisations pour le calcul d'estimateurs indépendants de l'échelle modifie les caractéristiques des estimateurs, comme l'illustre la Figure 10.

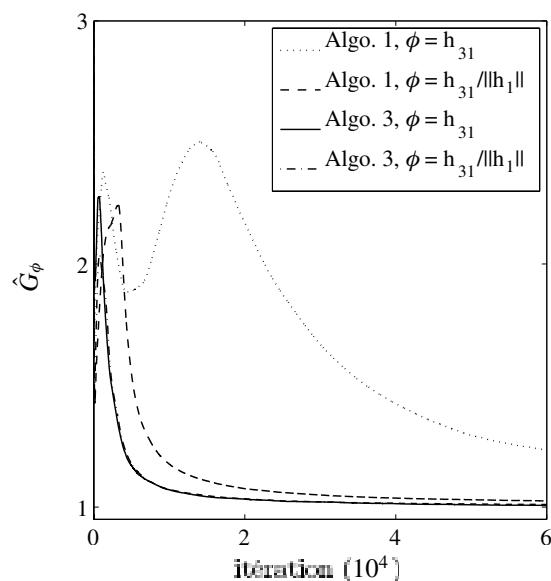


Figure 9. Courbes de diagnostic de convergence d'un coefficient de mélange dans les Algorithmes 1 et 3.

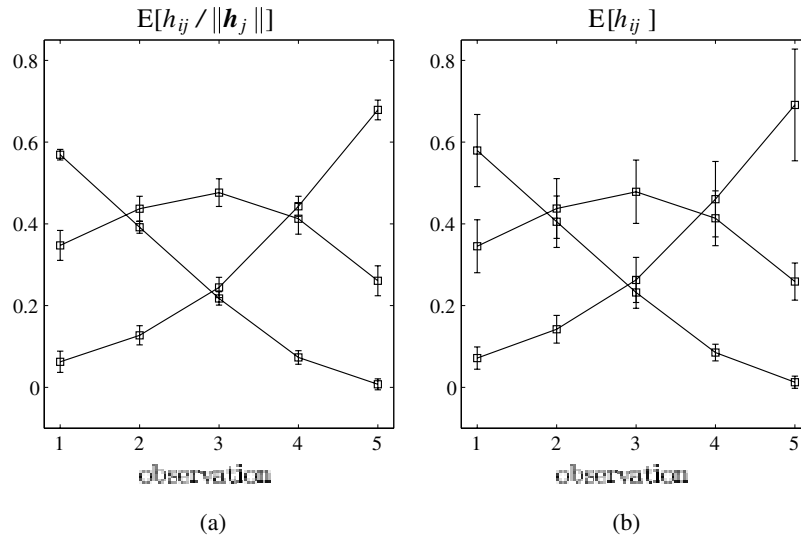


Figure 10. Coefficients de mélange estimés avec (a) et sans (b) normalisation des réalisations. Toutes les sources sont mises à la même échelle et les intervalles de confiance sont à 95 %.

Sur cet exemple, les valeurs estimées sont relativement proches (ce qui n'est pas forcément vrai en général ; sauf cas particulier,  $E[X/\|X\|]$  et  $E[X]/E[\|X\|]$  sont des quantités distinctes), tandis que la variance de l'estimateur construit à partir de réalisations normalisées est plus faible que celle d'un estimateur calculé directement sur les réalisations.

Au final, l'Algorithme 3 est le seul à permettre d'approcher l'espérance statistique de tout type de quantité, sans restriction ni précaution particulière.

## 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 observées en pratique. Elle concerne le rééchantillonnage de l'échelle d'après sa loi conditionnelle *a posteriori*, dans le but d'améliorer l'exploration de l'espace d'état de la distribution *a posteriori*. Cette étape accélère la convergence vers la loi-cible et diminue la corrélation entre échantillons successifs. Pour des lois *a priori* classiques (lois gaussiennes, lois gamma), la loi *a posteriori* de l'échelle est connue ; son échantillonnage est maîtrisé et le coût de calcul additionnel est négligeable. Les résultats sur un exemple simple ainsi que dans le cadre d'une application en séparation de sources illustrent l'intérêt de cette étape supplémentaire.

En comparaison, la solution qui consiste à renormaliser l'une des composantes inconnues au sein de l'échantillonneur de Gibbs a été étudiée (Algorithme 2). D'après nos conclusions, cette variante est à proscrire, car elle déforme la loi-cible et biaise les estimateurs et leur variance. À défaut, il est préférable de

conserver l'échantillonneur de Gibbs basique (Algorithme 1), à condition de se limiter à l'estimation de quantités normalisées. En comparaison, la solution que nous proposons (Algorithme 3) est évidemment préférable à l'Algorithme 2, car notre solution ne modifie pas la loi-cible, mais aussi à l'Algorithme 1, car la convergence de la chaîne est assurée dans son intégralité, forme et échelle ; le choix d'estimateurs peut donc être effectué sans restriction.

Les applications pouvant bénéficier du rééchantillonnage de l'échelle sont nombreuses, les plus importantes étant la séparation de sources et la déconvolution aveugle.

## A. Probabilité d'acceptation du changement d'échelle

Dans le cas général d'un saut réversible, la probabilité d'acceptation du changement d'échelle est donnée par [7, Eq. (19)]. Dans notre cas, comme dans celui de [7, Exemple 4.1], les états successifs de la chaîne sont de dimension constante (*i.e.*  $m = m' = 1$  dans les notations de Waagepetersen et Sorensen), et cette expression se simplifie sous la forme  $\min(1, \rho(\mathbf{y}, \mathbf{y}'))$ ,  $\mathbf{y}$  et  $\mathbf{y}'$  désignant respectivement les états courants et proposés et

$$\rho(\mathbf{y}, \mathbf{y}') = \frac{f(\mathbf{y}') q(\mathbf{y}', \mathbf{u}')}{f(\mathbf{y}) q(\mathbf{y}, \mathbf{u})} |J_g(\mathbf{y}, \mathbf{u})|$$

où  $f$  est la loi-cible ; les variables  $\mathbf{u}$ ,  $\mathbf{u}'$  et la fonction  $q$  sont liées au mécanisme de proposition ; enfin,  $g(\mathbf{y}, \mathbf{u}) = (\mathbf{y}', \mathbf{u}')$  [7]. Le dernier terme est une valeur absolue de déterminant jacobien dont la présence traduit le fait que le noyau de proposition est à

densité dans un espace de dimension réduite par rapport à la loicible.

Dans le cas présent, on a  $\mathbf{y} = (\mathbf{x}, \mathbf{h})$ ,  $\mathbf{y}' = (\mathbf{x} \times s, \mathbf{h}/s)$ ,  $\mathbf{u} = s$  et  $g(\mathbf{y}, s) = (\mathbf{y}', 1/s)$ . Enfin,  $q(\mathbf{y}, \cdot)$  désigne la loi  $f_{S|reste}$  suivant laquelle  $S$  est échantillonnée. Montrons alors que  $\rho(\mathbf{y}, \mathbf{y}')$  vaut 1.

D'une part, on a

$$J_g(\mathbf{y}, s) = \begin{vmatrix} s\mathbf{I}_M & \mathbf{O}_{MP} & \mathbf{O}_{M1} \\ \mathbf{O}_{PM} & \mathbf{I}_P/s & \mathbf{O}_{P1} \\ \mathbf{x}^t & -\mathbf{h}^t/s^2 & -1/s^2 \end{vmatrix} = s^{M-P-2}.$$

D'autre part, d'après (5),

$$q(\mathbf{y}, s) = \frac{1}{Z(\mathbf{y})} |s|^{M-P-1} f(\mathbf{y}')$$

avec

$$Z(\mathbf{y}) = \int |t|^{M-P-1} f(\mathbf{x} \times t, \mathbf{h}/t) dt,$$

et

$$q(\mathbf{y}', s) = \frac{1}{Z(\mathbf{y}')} |s|^{-M+P+1} f(\mathbf{y})$$

avec

$$Z(\mathbf{y}') = s \int |t|^{M-P-1} f(\mathbf{x} \times ts, \mathbf{h}/t) dt$$

$$= \frac{1}{s} \int \left| \frac{t'}{s} \right|^{M-P-1} f(\mathbf{x} \times t', \mathbf{h}/t') dt' = Z(\mathbf{y}) \times |s|^{-M+P}.$$

Finalement,

$$\rho(\mathbf{y}, \mathbf{y}') = \frac{|s|^{-M+P+1}}{|s|^{M-P-1}} \times |s|^{-M+P} \times |s|^{M-P-2} = 1.$$

## Références

- [1] H. SNOUSSI et J. IDIER, «Bayesian blind separation of generalized hyperbolic processes in noisy and underdeterminate mixtures», *IEEE Transactions on Signal Processing*, vol. 54, n° 9, pp. 3257-3269, September 2006.
- [2] 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 Transactions on Signal Processing*, vol. 54, n° 11, pp. 4133-4145, November 2006.
- [3] Q. CHENG, R. CHEN et T.-H. LI, «Simultaneous wavelet estimation and deconvolution of reflection-seismic signals», *IEEE Transactions on Geoscience and Remote Sensing*, vol. 34, n° 2, pp. 377-384, Mars 1996.
- [4] P. CIUCIU, J. IDIER, T. VEIT et T. VINCENT, «Application du rééchantillonnage stochastique de l'échelle en détection-estimation de l'activité cérébrales par IRMF», in *Actes du 21ème colloque GRETSI*, Troyes, France, pp. 373-376, 2007.
- [5] Y.-L. YOU et M. KAVEH, «A regularization approach to joint blur identification and image restoration», *IEEE Transactions on Image Processing*, vol. 5, n° 3, pp. 416-428, mars 1996.
- [6] M. STEPHENS, «Dealing with label-switching in mixture models», *Journal of the Royal Statistical Society B*, vol. 62, pp. 795-809, 2000.
- [7] R. WAAGEPETERSEN et D. SORENSEN, «A tutorial on reversible Jump MCMC with a view toward QTL-mapping », *International Statistical Review*, vol. 69, pp. 49-61, 2001.
- [8] J. DAGPUNAR, «An easily implemented generalized inverse Gaussian generator», *Communications in Statistics, Simulation and Computing*, vol. 18, n° 2, pp. 703-710, 1989.
- [9] T. VEIT et J. IDIER, «Échantillonnage de l'échelle dans les algorithmes MCMC pour les problèmes inverses bilinéaires », in *Actes du 21ème colloque GRETSI*, Troyes, France, pp. 1233-1236, septembre 2007.
- [10] S. SÉNÉCAL et P.-O. AMBLARD, «Bayesian separation of discrete sources via Gibbs sampling», in *Proceedings of International Conference on Independent Component Analysis and Blind Signal Separation (ICA'2000)*, pp. 566-572, 2000.
- [11] H. SNOUSSI et A. MOHAMMAD-DJAFARI, «Fast joint separation and segmentation of mixed images», *Journal of Electronic Imaging*, vol. 13, n° 2, pp. 349-361, 2004.
- [12] M. ICHIR et A. MOHAMMAD-DJAFARI, «Hidden Markov models for wavelet-based blind source separation», *IEEE Transactions on Image Processing*, vol. 17, n° 7, pp. 1887-1899, Juillet 2006.
- [13] C. FÉVOTTE et S. J. GODSILL, «A Bayesian approach for blind separation of sparse sources», *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, n° 6, pp. 2174-2188, novembre 2006.
- [14] S. MOUSSAOUI, C. CARTERET, D. BRIE et A. MOHAMMAD-DJAFARI, «Bayesian analysis of spectral mixture data using Markov Chain Monte Carlo methods», *Chemometrics and Intelligent Laboratory Systems*, vol. 81, n° 2, pp. 137-148, April 2006.
- [15] A. GELMAN et D. RUBIN, «Inference from iterative simulation using multiple sequences», *Statistical Science*, vol. 7, pp. 457-472, 1992.





Thomas **Veit**

Thomas Veit est chargé de recherche à l'INRETS. Ingénieur en Génie Mathématique diplômé de l'INSA de Rouen en 2001, il a obtenu un DEA de Statistique à l'Université Pierre et Marie Curie, Paris VI, en 2002. En 2005, il a soutenu une thèse de doctorat préparée à l'Irisa, Rennes puis effectué un post-doctorat à l'IRCCyN, Nantes, en 2006. Depuis novembre 2006, il effectue sa recherche au sein du Laboratoire sur les Interactions Véhicules-Infrastructure-Conducteurs, une unité de recherche mixte INRETS/LCPC. Ses thèmes de recherche sont la vision par ordinateur pour les aides à la conduite et les méthodes probabilistes et statistiques pour le traitement d'images.



Saïd **Moussaoui**

Saïd Moussaoui est ingénieur de l'école polytechnique à Alger (2001) et docteur de l'université Henri Poincaré, Nancy 1 (2005). Depuis septembre 2006 il est enseignant-chercheur à l'école centrale de Nantes. Il effectue ses activités de recherche à l'Institut de Recherche en Communications et Cybernétique de Nantes (IRCCyN UMR CNRS 6597), au sein de l'équipe ADTSI (Analyse et décision en traitement du signal et de l'image).



Jérôme **Idier**

Jérôme Idier est Directeur de recherche au CNRS. Diplômé de Supélec, Gif-sur-Yvette (1988), il a obtenu un doctorat de physique de l'Université Paris-sud, Orsay en 1991, puis une habilitation à diriger des recherches de la même université en 2001. Depuis 1991, il est chercheur au CNRS, au Laboratoire des signaux et systèmes de 1991 à 2002 et à l'Institut de Recherche en Communications et Cybernétique de Nantes depuis septembre 2002. Ses principaux sujets d'intérêt concernent les approches statistiques et l'optimisation pour la restauration de signaux et d'images et plus généralement les problèmes inverses. Les applications visées sont principalement le contrôle non destructif, l'imagerie astronomique et le traitement de signaux biomédicaux.



