

Recuit simulé par augmentation de données appliqué à la déconvolution bayésienne

Arnaud Doucet

LETI (CEA-Technologies Avancées)
DEIN - SPE CEA Saclay
91191 Gif-sur-Yvette cedex

ETIS - UCP ENSEA URA CNRS 2235
95014 Cergy-Pontoise cedex
Email : douceta@ensea.fr

RÉSUMÉ

Dans ce papier, nous proposons des algorithmes de recuit simulé, essentiellement basés sur l'augmentation de données, afin de résoudre des problèmes d'estimation bayésienne pour les modèles markoviens linéaires gaussiens à sauts. Ces algorithmes sont appliqués avec succès à un problème de déconvolution impulsionnelle.

ABSTRACT

In this paper, we propose simulated annealing algorithms mainly based on data augmentation to solve Bayesian estimation problems for jump Markov linear Gaussian systems. These algorithms are successfully applied to deconvolution of sparse spike trains.

1 Introduction

Les modèles markoviens linéaires gaussiens à sauts ont de multiples applications en radar et télécommunications. Ils sont aussi implicitement très utilisés en géophysique et en nucléaire car ils incluent les processus BG (Bernoulli-Gauss) filtrés linéairement. La structure mathématique des problèmes d'estimation optimale pour ces modèles est bien connue. Malheureusement, elle requiert des coûts de calcul et de mémoire croissant exponentiellement avec le temps d'observation. De nombreuses méthodes sous-optimales ont été proposées dans la littérature.

Dans ce papier, nous abordons des problèmes d'estimation Bayésienne (optimale) hors-ligne pour cette classe de modèles. Ces problèmes peuvent se ramener à des problèmes d'optimisation combinatoire. Nous présentons tout d'abord un algorithme MCMC (Monte Carlo par chaîne de Markov) basé sur l'augmentation de données permettant de simuler des échantillons selon la loi jointe *a posteriori*. Cet algorithme construit une chaîne de Markov convergeant uniformément géométriquement vers cette loi. Cette propriété de convergence est utilisée pour développer des algorithmes d'optimisation de type recuit simulé, i.e. des chaînes de Markov non homogènes convergeant vers les maxima globaux requis.

2 Modèle markovien linéaire gaussien à sauts

Un modèle markovien linéaire gaussien vérifie pour $k \in \mathbb{N}^*$:

$$\mathbf{x}_k = \mathbf{A}(r_k) \mathbf{x}_{k-1} + \mathbf{B}(r_k) \mathbf{v}_k + \mathbf{F}(r_k) \mathbf{u}_k$$

$$\mathbf{y}_k = \mathbf{C}(r_k) \mathbf{x}_k + \mathbf{D}(r_k) \mathbf{w}_k + \mathbf{G}(r_k) \mathbf{u}_k$$

où $\mathbf{x}_k \in \mathbb{R}^{n_x}$, $\mathbf{y}_k \in \mathbb{R}^{n_y}$, $\mathbf{u}_k \in \mathbb{R}^{n_u}$, $\mathbf{v}_k \in \mathbb{R}^{n_v}$ et $\mathbf{w}_k \in \mathbb{R}^{n_w}$, les matrices étant de dimensions appropriées. \mathbf{u}_k est un terme de contrôle connu. $\mathbf{v}_k \sim \mathcal{N}(0, \mathbf{I}_{n_v})$ et $\mathbf{w}_k \sim \mathcal{N}(0, \mathbf{I}_{n_w})$ sont deux suites i.i.d. gaussiennes et \mathbf{v}_k , \mathbf{w}_k et $\mathbf{x}_0 \sim \mathcal{N}(\hat{\mathbf{x}}_0, \mathbf{P}_0)$ sont mutuellement indépendants pour tout $k \in \mathbb{N}^*$. r_k est une chaîne de Markov homogène du premier ordre à états finis de probabilités de transition

$$p_{ij} \triangleq \Pr \{ r_{k+1} = j \mid r_k = i \}$$

où $(i, j) \in \mathcal{S} \times \mathcal{S}$, $\mathcal{S} = \{1, \dots, s\}$ et de distribution initiale $p_i \triangleq \Pr \{ r_1 = i \}$. Dans la suite du papier, on supposera que les paramètres suivants du modèle

$$\left\{ \hat{\mathbf{x}}_0, \mathbf{P}_0, p_{ij}, p_i, \mathbf{A}(i), \mathbf{B}(i), \mathbf{C}(i), \mathbf{D}(i), \mathbf{F}(i), \mathbf{G}(i) \right\}$$

sont connus. On désigne par la suite $\mathbf{Y}_{1 \rightarrow T} \triangleq (\mathbf{y}_1, \dots, \mathbf{y}_T)$, $\mathbf{R}_{1 \rightarrow T} \triangleq (r_1, \dots, r_T)$ et $\mathbf{X}_{0 \rightarrow T} \triangleq (\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_T)$ (les redondances intrinsèques au passage en représentation d'état étant éliminées).

Les hypothèses suivantes sont adoptées :

1. $p_i > 0$ et $p_{i,j} > 0$ pour tout $i, j \in \mathcal{S} \times \mathcal{S}$.
2. La distribution jointe *a priori* $p(\mathbf{X}_{0 \rightarrow T} \mid \mathbf{R}_{1 \rightarrow T})$ est une distribution gaussienne non dégénérée.
3. $\mathbf{D}^t(r_k) \mathbf{D}(r_k) > 0$ pour $k = 1, \dots, T$.

L'hypothèse 1 peut être facilement affaiblie. L'hypothèse 2 sera levée dans (5.2). Nous nous intéressons aux trois problèmes suivants :

- Estimation de la distribution jointe *a posteriori* $p(\mathbf{X}_{0 \rightarrow T}, \mathbf{R}_{1 \rightarrow T} \mid \mathbf{Y}_{1 \rightarrow T})$.

- Maximisation de la distribution jointe *a posteriori*
 $p(\mathbf{X}_{0 \rightarrow T}, \mathbf{R}_{1 \rightarrow T} | \mathbf{Y}_{1 \rightarrow T})$.
- Maximisation de la distribution marginale *a posteriori*
 $p(\mathbf{R}_{1 \rightarrow T} | \mathbf{Y}_{1 \rightarrow T})$.

Par la suite, afin d'alléger les notations, nous indexons arbitrairement les s^T réalisations possibles de la chaîne de Markov à états finis $\{\mathbf{R}_{1 \rightarrow T}^l; l = 1, \dots, s^T\}$ et notons abusivement $\int \varphi(\mathbf{R}_{1 \rightarrow T}) d\mathbf{R}_{1 \rightarrow T}$ où $\varphi : \mathcal{G}^T \rightarrow \mathbb{R}$ pour $\sum_{l=1}^{s^T} \varphi(\mathbf{R}_{1 \rightarrow T}^l)$. Nous gardons toutefois la notation $\mathbf{R}_{1 \rightarrow T}^l$ au lieu de l pour rester "homogène" avec $\mathbf{X}_{0 \rightarrow T}$.

3 Estimation bayésienne par MCMC

On peut facilement exprimer analytiquement la distribution *a posteriori* $p(\mathbf{X}_{0 \rightarrow T}, \mathbf{R}_{1 \rightarrow T} | \mathbf{Y}_{1 \rightarrow T})$. Malheureusement, cette distribution nécessite le calcul de s^T termes. En pratique, cette expression n'est donc pas utilisable. De nombreux travaux ont proposé des approximations analytiques sous-optimales pour estimer les lois marginales $p(\mathbf{x}_k | \mathbf{Y}_{1 \rightarrow T})$ et cette loi jointe [6].

Nous proposons ici une approximation basée sur une méthode MCMC dans le cadre de l'augmentation de données [11]. Les méthodes MCMC utilisées en physique depuis les années 50 et plus récemment en imagerie par Geman et Geman [7] sont des méthodes stochastiques pour simuler des échantillons selon des distributions complexes. Elles ont récemment révolutionné le domaine des statistiques appliquées. L'idée de base des méthodes MCMC est de construire une chaîne de Markov homogène $\{\mathbf{R}_{1 \rightarrow T}^{(k)}, \mathbf{X}_{0 \rightarrow T}^{(k)}; k \in \mathbb{N}\}$ qui admet comme distribution d'équilibre la distribution selon laquelle on désire simuler des échantillons. Nous utilisons ici l'algorithme dit d'augmentation de données.

Algorithme : Augmentation de Données

1. Sélectionner aléatoirement $(\mathbf{R}_{1 \rightarrow T}^{(0)}, \mathbf{X}_{0 \rightarrow T}^{(0)})$, $i = 1$.

2. Répéter :

- Simuler $\mathbf{X}_{0 \rightarrow T}^{(i)}$ selon $p(\mathbf{X}_{0 \rightarrow T} | \mathbf{Y}_{1 \rightarrow T}, \mathbf{R}_{1 \rightarrow T}^{(i-1)})$.
- Simuler $\mathbf{R}_{1 \rightarrow T}^{(i)}$ selon $p(\mathbf{R}_{1 \rightarrow T} | \mathbf{Y}_{1 \rightarrow T}, \mathbf{X}_{0 \rightarrow T}^{(i)})$.
- $i \leftarrow i + 1$

$p(\mathbf{X}_{0 \rightarrow T} | \mathbf{Y}_{1 \rightarrow T}, \mathbf{R}_{1 \rightarrow T}^{(i)})$ est une distribution gaussienne multivariée et $p(\mathbf{R}_{1 \rightarrow T} | \mathbf{Y}_{1 \rightarrow T}, \mathbf{X}_{0 \rightarrow T}^{(i)})$ est une distribution discrète multivariée. Il n'est heureusement pas nécessaire de calculer ces lois pour simuler des échantillons distribués selon elles. Pour cela, on utilise la décomposition filtrage avant/simulation arrière [2] dont la complexité est en $O(T)$. L'application de l'augmentation de données pour de tels modèles est présentée dans [2]. Ici, nous montrons que cet algorithme a de "bonnes" propriétés de convergence que nous utiliserons par la suite pour obtenir des versions recuit simulé.

Proposition 3.1 — [6] $\exists 0 \leq C < +\infty$ et $0 < \rho < 1$ tels que pour toute distribution initiale $p_0(\mathbf{X}_{0 \rightarrow T}, \mathbf{R}_{1 \rightarrow T})$, la distribution après i itérations de $(\mathbf{X}_{0 \rightarrow T}^{(i)}, \mathbf{R}_{1 \rightarrow T}^{(i)})$ notée $p_i(\mathbf{X}_{0 \rightarrow T}, \mathbf{R}_{1 \rightarrow T})$ vérifie

$$\|p_i(\mathbf{X}_{0 \rightarrow T}, \mathbf{R}_{1 \rightarrow T}) - p(\mathbf{X}_{0 \rightarrow T}, \mathbf{R}_{1 \rightarrow T} | \mathbf{Y}_{1 \rightarrow T})\|_{TV} \leq C\rho^i$$

où $\|\cdot\|_{TV}$ désigne la norme de la variation totale [11].

La démonstration s'appuie essentiellement sur le *principe de dualité* introduit par Diebolt et Robert [5]. La chaîne de Markov simulée converge uniformément géométriquement. Ce résultat implique un théorème central limite pour les moyennes ergodiques. Dans de nombreuses applications, on ne cherche pas à estimer la distribution *a posteriori* mais seulement à déterminer ses maxima joints ou marginaux. Ce résultat est aussi important pour obtenir des versions non homogènes de l'algorithme convergeant vers ces maxima.

4 Maximisation jointe

Nous désirons maximiser la distribution jointe *a posteriori*, i.e. nous cherchons à estimer

$$\arg \max_{(\mathbf{X}_{0 \rightarrow T}, \mathbf{R}_{1 \rightarrow T}^l)} p(\mathbf{X}_{0 \rightarrow T} | \mathbf{Y}_{1 \rightarrow T}, \mathbf{R}_{1 \rightarrow T}^l) p(\mathbf{R}_{1 \rightarrow T}^l | \mathbf{Y}_{1 \rightarrow T})$$

La distribution $p(\mathbf{X}_{0 \rightarrow T} | \mathbf{Y}_{1 \rightarrow T}, \mathbf{R}_{1 \rightarrow T}^l)$ est Gaussienne et non dégénérée d'après les hypothèses, i.e. $\mathbf{X}_{0 \rightarrow T} | \mathbf{Y}_{1 \rightarrow T}, \mathbf{R}_{1 \rightarrow T}^l \sim \mathcal{N}(\mathbf{m}_l, \Sigma_l)$. Il est équivalent de résoudre le problème d'optimisation combinatoire :

$$\arg \max_{\mathbf{R}_{1 \rightarrow T}^l} |\Sigma_l|^{-1/2} p(\mathbf{R}_{1 \rightarrow T}^l | \mathbf{Y}_{1 \rightarrow T})$$

où $|\Sigma_l|$ désigne le déterminant de la matrice Σ_l . On note \mathcal{M} l'ensemble de ces maxima globaux. Une fois cet estimateur obtenu, on obtient alors directement les états associés $\mathbf{X}_{0 \rightarrow T}$ par un lisseur de Kalman. Soit $\{T(i); i \in \mathbb{N}\}$ une suite décroissante de "températures". Par la suite, on désigne par $\bar{p}^{1/T(i)}(\mathbf{X}_{0 \rightarrow T} | \mathbf{Y}_{1 \rightarrow T}, \mathbf{R}_{1 \rightarrow T}^l)$ la distribution proportionnelle à $p^{1/T(i)}(\mathbf{X}_{0 \rightarrow T} | \mathbf{Y}_{1 \rightarrow T}, \mathbf{R}_{1 \rightarrow T}^l)$ et $\bar{p}^{1/T(i)}(\mathbf{R}_{1 \rightarrow T} | \mathbf{Y}_{1 \rightarrow T}, \mathbf{X}_{0 \rightarrow T})$ la distribution proportionnelle à $p^{1/T(i)}(\mathbf{R}_{1 \rightarrow T} | \mathbf{Y}_{1 \rightarrow T}, \mathbf{X}_{0 \rightarrow T})$. Nous proposons tout d'abord une version recuit simulé de l'algorithme d'augmentation de données.

Algorithme : version recuit simulé de l'augmentation de données

1. Sélectionner aléatoirement $(\mathbf{R}_{1 \rightarrow T}^{(0)}, \mathbf{X}_{0 \rightarrow T}^{(0)})$, $i = 1$.

2. Répéter :

- Simuler $\mathbf{X}_{0 \rightarrow T}^{(i)}$ selon $\bar{p}^{1/T(i)}(\mathbf{X}_{0 \rightarrow T} | \mathbf{Y}_{1 \rightarrow T}, \mathbf{R}_{1 \rightarrow T}^{(i-1)})$.
- Simuler $\mathbf{R}_{1 \rightarrow T}^{(i)}$ selon $\bar{p}^{1/T(i)}(\mathbf{R}_{1 \rightarrow T} | \mathbf{Y}_{1 \rightarrow T}, \mathbf{X}_{0 \rightarrow T}^{(i)})$.
- $i \leftarrow i + 1$

On simule des échantillons selon ces lois en utilisant la procédure filtre avant/simulation arrière. On peut obtenir le résultat suivant.

Proposition 4.1 — [6] $\exists L > 0$ tel que pour $\gamma \geq L$ et $u > 0$ si $\{T(i); i \in \mathbb{N}\}$ satisfait $T(i) = \gamma / \ln(i + u)$ alors la distribution à l'itération i de $\{\mathbf{R}_{1 \rightarrow T}^{(i)}; i \in \mathbb{N}\}$ vérifie pour toute distribution initiale $\lim_{i \rightarrow +\infty} \int |p_i(\mathbf{R}_{1 \rightarrow T}) - p_\infty(\mathbf{R}_{1 \rightarrow T})| d\mathbf{R}_{1 \rightarrow T} = 0$ où

$$p_\infty(\mathbf{R}_{1 \rightarrow T}^l) \triangleq \frac{|\Sigma_l|^{1/2}}{\sum_{\mathbf{R}_{1 \rightarrow T}^j \in \mathcal{M}} |\Sigma_j|^{1/2}} \delta_{\mathcal{M}}(\mathbf{R}_{1 \rightarrow T}^l)$$

La distribution de $\{\mathbf{R}_{1 \rightarrow T}^{(i)}; i \in \mathbb{N}\}$ converge donc vers une distribution concentrée sur les maxima globaux. Toutefois en général, cette distribution n'est pas la distribution uniforme sur

cet ensemble. L'avantage principal de cet algorithme est qu'il ne nécessite pas l'évaluation généralement numériquement coûteuse de la fonction à maximiser.

5 Maximisation marginale

On désire maintenant obtenir l'estimateur suivant :

$$\arg \max_{\mathbf{R}_{1 \rightarrow T}^l} p(\mathbf{R}_{1 \rightarrow T}^l | \mathbf{Y}_{1 \rightarrow T})$$

C'est un problème d'optimisation combinatoire avec s^T possibilités. On note \mathcal{M} l'ensemble de ces maxima globaux.

5.1 Recuit simulé utilisant l'augmentation de données

Le premier algorithme que nous proposons est basé sur l'augmentation de données. L'idée consiste à utiliser l'algorithme d'augmentation de données comme distribution de proposition pour un algorithme M-H (Metropolis-Hastings). La réversibilité de la distribution de proposition par rapport à $p(\mathbf{R}_{1 \rightarrow T} | \mathbf{Y}_{1 \rightarrow T})$ donne une expression simple pour le rapport d'acceptation des candidats.

Algorithme : recuit simulé avec augmentation de données.

1. Sélectionner aléatoirement $(\mathbf{R}_{1 \rightarrow T}^{(0)}, \mathbf{X}_{0 \rightarrow T}^{(0)})$, $i = 1$.
2. Répéter :
 - Simuler $\mathbf{X}_{0 \rightarrow T}^{(i)}$ selon $p(\mathbf{X}_{0 \rightarrow T} | \mathbf{Y}_{1 \rightarrow T}, \mathbf{R}_{1 \rightarrow T}^{(i-1)})$.
 - Simuler un candidat $\mathbf{R}_{1 \rightarrow T}^c$ selon $p(\mathbf{R}_{1 \rightarrow T} | \mathbf{Y}_{1 \rightarrow T}, \mathbf{X}_{0 \rightarrow T}^{(i)})$ et calculer

$$\alpha(\mathbf{R}_{1 \rightarrow T}^{(i-1)}, \mathbf{R}_{1 \rightarrow T}^c) = \left[\frac{p(\mathbf{R}_{1 \rightarrow T}^c | \mathbf{Y}_{1 \rightarrow T})}{p(\mathbf{R}_{1 \rightarrow T}^{(i-1)} | \mathbf{Y}_{1 \rightarrow T})} \right]^{1/T(i-1)}$$

- Simuler $u_i \sim \mathcal{U}(0, 1)$. Si $u_i < \alpha(\mathbf{R}_{1 \rightarrow T}^{(i-1)}, \mathbf{R}_{1 \rightarrow T}^c)$ alors $\mathbf{R}_{1 \rightarrow T}^{(i)} = \mathbf{R}_{1 \rightarrow T}^c$ sinon $\mathbf{R}_{1 \rightarrow T}^{(i)} = \mathbf{R}_{1 \rightarrow T}^{(i-1)}$.
- $i \leftarrow i + 1$

Cet algorithme nécessite le calcul de $p(\mathbf{R}_{1 \rightarrow T} | \mathbf{Y}_{1 \rightarrow T}) \propto p(\mathbf{Y}_{1 \rightarrow T} | \mathbf{R}_{1 \rightarrow T}) p(\mathbf{R}_{1 \rightarrow T})$ à une constante de proportionnalité près. Le premier terme est le terme de vraisemblance obtenu directement par le filtre de Kalman et $p(\mathbf{R}_{1 \rightarrow T})$ se calcule directement. Le filtre de Kalman étant calculé lors de la procédure filtrage avant/simulation arrière, la charge de calcul supplémentaire par rapport à l'augmentation de données est faible. En suivant des arguments classiques, on obtient la convergence de $p_i(\mathbf{R}_{1 \rightarrow T})$ vers la distribution uniforme sur \mathcal{M} pour une décroissance logarithmique de la température.

5.2 Recuit simulé sur l'espace discret

L'algorithme précédent maximisant la distribution marginale est très simple à implémenter à partir de de l'algorithme d'augmentation de données. Toutefois, il ne permet pas d'aborder les modèles pour lesquels l'hypothèse 2 n'est pas vérifiée, la chaîne devenant réductible [11]. C'est malheureusement le cas pour les processus BG (bien que ces processus proviennent souvent de l'approximation par

un Dirac d'une composante d'un mélange de deux Gaussiennes). On peut lever l'hypothèse 2 en construisant une chaîne de Markov uniquement sur \mathcal{S}^T . On simule alors une chaîne de Markov non homogène de distribution invariante $\bar{p}^{1/T(i)}(\mathbf{R}_{1 \rightarrow T} | \mathbf{Y}_{1 \rightarrow T})$ à l'itération i . Ceci peut être réalisé en simulant directement selon $\bar{p}^{1/T(i)}(r_k | \mathbf{R}_{1 \rightarrow T \setminus \{k\}}, \mathbf{Y}_{1 \rightarrow T}) \propto p^{1/T(i)}(r_k | \mathbf{R}_{1 \rightarrow T \setminus \{k\}}, \mathbf{Y}_{1 \rightarrow T})$ pour $k = 1, \dots, T$ auquel cas on a un échantillonneur de Gibbs [7] ou par des pas de M-H admettant $\bar{p}^{1/T(i)}(r_k | \mathbf{R}_{1 \rightarrow T \setminus \{k\}}, \mathbf{Y}_{1 \rightarrow T})$ comme loi invariante. Cette idée est séduisante mais un calcul direct de ces distributions requiert $O(T^2)$ opérations et est inutilisable pratiquement. Dans le cas d'un MA à excitation BG, Champagnat *et al.* [4] ont proposé un algorithme réduisant cette complexité. Pour une classe de modèles plus large, Carter *et al.* [3] ont proposé très récemment un algorithme permettant de simuler selon ces distributions en $O(T)$ opérations. En effet, on a

$$p(r_k | \mathbf{R}_{1 \rightarrow T \setminus \{k\}}, \mathbf{Y}_{1 \rightarrow T}) \propto p(\mathbf{Y}_{1 \rightarrow T} | \mathbf{R}_{1 \rightarrow T}) p(r_k | \mathbf{R}_{1 \rightarrow T \setminus \{k\}})$$

or la vraisemblance vérifie

$$\begin{aligned} p(\mathbf{Y}_{1 \rightarrow T} | \mathbf{R}_{1 \rightarrow T}) &= \\ p(\mathbf{Y}_{1 \rightarrow k-1} | \mathbf{R}_{1 \rightarrow k-1}) p(\mathbf{y}(k) | \mathbf{Y}_{1 \rightarrow k-1}, \mathbf{R}_{1 \rightarrow k-1}) \\ &\times \int p(\mathbf{x}(k) | \mathbf{R}_{1 \rightarrow k}, \mathbf{Y}_{1 \rightarrow k}) \\ &\times p(\mathbf{Y}_{k+1 \rightarrow T} | \mathbf{R}_{k+1 \rightarrow T}, \mathbf{x}(k)) d\mathbf{x}(k) \end{aligned}$$

En utilisant cette décomposition, Carter *et al.* propose un échantillonneur de Gibbs à balayage déterministe dans le cas où $\mathbf{A}(r_k)$ est inversible, voir [3] pour les détails. (Une légère modification est à apporter à leurs équations pour initialiser le filtre arrière). Notons qu'on peut aussi utiliser directement ce résultat pour obtenir un algorithme de maximisation déterministe du type SMLR [10] de complexité réduite. Dans le cas où la séquence $\mathbf{R}_{1 \rightarrow T}$ est indépendante, i.e. lorsque $p(\mathbf{R}_{1 \rightarrow T}) = \prod_{k=1}^T p(r_k)$, on a de plus le résultat suivant qui est un corollaire de ([8], Th. 5.1) :

Proposition 5.1 — *La corrélation maximale (voir déf. [8]) de la chaîne de Markov simulée par l'échantillonneur de Gibbs homogène est inférieure à celle obtenue par l'augmentation de données décrite dans la section (3).*

L'échantillonneur de Gibbs à balayage déterministe n'étant pas réversible, on ne peut toutefois conclure à partir de ces résultats à une plus grande vitesse de convergence de cet algorithme. Pour obtenir une telle conclusion, on peut par exemple symétriser les noyaux de transition afin d'assurer la réversibilité mais on ne garde alors qu'une valeur simulée sur deux !

On peut définir directement une version recuit simulé de l'algorithme de Carter et Kohn. Les résultats classiques de convergence s'appliquent directement dans ce cas [7].

6 Application à la déconvolution

Nous présentons ici une application de ces algorithmes à la déconvolution impulsionnelle Bayésienne. D'autres applications sont présentées dans [6].

On s'intéresse au signal issu d'un capteur neutronique. On peut modéliser ce signal par un modèle AR(2) excité par un bruit de dynamique i.i.d. modélisé par un mélange de deux Gaussiennes v'_k :

$$v'_k \sim \lambda \mathcal{N}(0, \sigma_v^2) + (1 - \lambda) \mathcal{N}(0, \bar{\sigma}_v^2) \quad (1)$$

i.e. avec une probabilité $0 < \lambda < 1$ une particule arrive sur le capteur, avec une probabilité $(1 - \lambda)$ on modélise les perturbations externes en entrée du capteur. Typiquement, on a $\bar{\sigma}_v \ll \sigma_v$. Ce signal est observé dans du bruit blanc Gaussien. Ce signal peut être modélisé ainsi :

$$\mathbf{x}_k = \begin{pmatrix} a_1 & a_2 \\ 1 & 0 \end{pmatrix} \mathbf{x}_{k-1} + \mathbf{B}(r_k) v_k \quad (2)$$

$$y_k = \begin{pmatrix} 1 & 0 \end{pmatrix} \mathbf{x}_k + \sigma_w w_k \quad (3)$$

où r_k est donc ici i.i.d. à valeurs dans $\mathcal{S} = \{1, 2\}$, $p_{11} = 1 - \lambda$, $\mathbf{B}(1) = \begin{pmatrix} \bar{\sigma}_v & 0 \end{pmatrix}^t$ et $\mathbf{B}(2) = \begin{pmatrix} \sigma_v & 0 \end{pmatrix}^t$. $\lambda = 0.1$, $\sigma_v = 0.3$, $\bar{\sigma}_v = 0.1\sigma_v$, $a_1 = 1.511$, $a_2 = -0.549$ et $\sigma_w = 0.3$ sont supposés ici connus et $T = 250$. Les hypothèses 1,2 et 3 sont vérifiées. Si on négligeait les perturbations en entrée du capteur alors $\bar{\sigma}_v = 0$ et l'hypothèse 2 n'est plus vérifiée.

La convergence des algorithmes de recuit simulé requiert en théorie des températures en $T(i) = \gamma / \ln(i + u)$, ce qui est inapplicable en pratique. Dans les simulations présentées, nous utilisons une température en $T(i) = \gamma / \alpha^i$. La figure 1 représente une réalisation de $\mathbf{X}_{1 \rightarrow T}$ et des observations $\mathbf{Y}_{1 \rightarrow T}$.

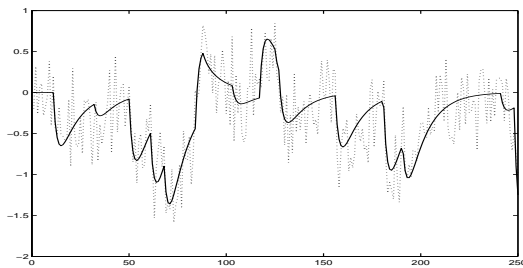


FIG. 1 — signal simulé $\mathbf{X}_{1 \rightarrow T}$ (trait plein) et observations $\mathbf{Y}_{1 \rightarrow T}$ (pointillés)

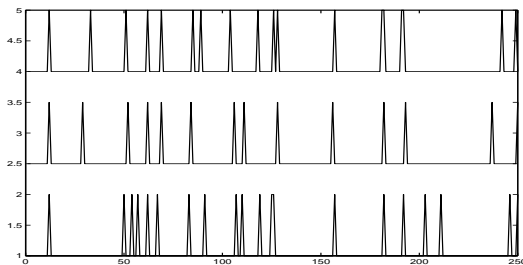


FIG. 2 — $\mathbf{R}_{1 \rightarrow T}$ simulée (haut) et estimées (milieu, bas)

La figure 2 représente la chaîne de Markov simulée $\mathbf{R}_{1 \rightarrow T}$ (haut de la figure) et l'estimation de son MAP marginal par la version recuit simulé de l'échantillonnage de Gibbs (milieu) et par le recuit simulé utilisant l'augmentation de données (bas) obtenus après 50 itérations. Pour ce modèle, les performances de l'échantillonneur de Gibbs sont expérimentalement

meilleures que celles du recuit utilisant l'augmentation de données.

7 Conclusion

Dans cet article, nous avons proposé des versions recuit simulé d'algorithmes MCMC pour résoudre des problèmes d'estimation bayésienne pour les modèles markoviens linéaires gaussiens à sauts. La convergence asymptotique de ces algorithmes vers l'ensemble des maxima globaux est assurée. En pratique, on obtient avec un faible nombre d'itérations des résultats très satisfaisants. Ces algorithmes ont aussi été utilisés dans des problèmes plus complexes de déconvolution aveugle [1].

Remerciements

L'auteur remercie Christophe Andrieu (ETIS-ENSEA) et Andrew Logothetis (Université de Melbourne).

Références

- [1] C. Andrieu, A. Doucet and P. Duvaut, "Bayesian Estimation of Filtered Point Processes using MCMC", in Proc. IEEE Conf. Asilomar, California, Nov. 1997.
- [2] C.K. Carter and R. Kohn, "On Gibbs Sampling for State Space Models", *Biometrika*, Vol. 81, No. 3, 1994, pp. 541-553.
- [3] C.K. Carter and R. Kohn, "Markov Chain Monte Carlo in Conditionally Linear Gaussian State Space Models", *Biometrika*, Vol. 83, 1996.
- [4] F. Champagnat, Y. Goussard and J. Idier, "Unsupervised Deconvolution of Sparse Spike Trains Using Stochastic Approximation", *IEEE Trans. on Sig. Proc.*, Vol. 44, No. 12, 1996, pp. 2988-2998.
- [5] J. Diebolt and C.P. Robert, "Estimation of Finite Mixture Distributions through Bayesian Sampling", *J. Roy. Stat. Soc. B*, Vol. 56, No. 2, 1994, pp. 363-375.
- [6] A. Doucet and A. Logothetis, "Stochastic Annealing Algorithms for Bayesian Estimation of Jump Markov Linear Systems", submitted, 1997.
- [7] S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distribution and the Bayesian Restoration of Images", *IEEE Trans. Patt. Ana. Mac. Int.*, Vol. 6, No. 6, 1984, pp. 721-741.
- [8] J.S. Liu, W.H. Wong and A. Kong, "Covariance Structure of the Gibbs Sampler with Applications to the Comparisons of Estimators and Augmentation Schemes", *Biometrika*, Vol. 81, No. 1, 1994, pp. 27-40.
- [9] A. Logothetis and V. Krishnamurthy, "A Bayesian Expectation-Maximization Algorithm for Estimating Jump Markov Linear Systems", submitted, 1997.
- [10] J.M. Mendel, *Maximum Likelihood Deconvolution*, Springer, New York, 1990.
- [11] A.F.M. Smith and G.O. Roberts, "Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo methods", *J. Roy. Stat. Soc. B*, Vol. 55, No. 1, 1993, pp. 3-23.