Un algorithme de lissage efficace et en ligne pour des equations différenteilles stochastiques observées avec bruit.

Pierre Gloaguen[1] & Marie-Pierre Etienne[1] & Sylvain Le Corff [2]

[1] *UMR MIA-Paris, AgroParisTech, INRA, Université Paris-Saclay, 75005, Paris, France. pierre.gloaguen@agroparistech.fr et marie.etienne@agroparistech.fr*
[2] *Laboratoire de Mathématiques d'Orsay, Université Paris-Sud, CNRS (UMR 8628), Université Paris-Saclay. sylvain.lecorff@math.u-psud.fr*

Résumé. Nous nous intéressons à un modèle de Markov caché (HMM) dont la partie cachée est une diffusion solution de l'équation différentielle stochastique (EDS) suivante :

$$\mathrm{d}X_t = \alpha^\theta(X_t)\mathrm{d}t + \mathrm{d}W_t, X_0 \sim \chi_0^\theta \,, \tag{1}$$

où α^θ est une fonction dépendant du paramètre θ, W_t est un mouvement Brownien standard issu de 0 et χ_0^θ est la distribution initiale pour la variable X_0.

La diffusion est seulement partiellement observée aux instants $t_0 = 0, \dots, t_n$ au travers d'un processus d'observations $(Y_k)_{k=0,\dots,n}$. Conditionnellement à la trajectoire $(X_t)_{0\le t\le t_n}$, les $(Y_k)_{k=0,\dots,n}$ sont indépendants et la loi de Y_k sachant $(X_t)_{0\le t\le t_n}$ a pour densité $g^\theta(X_k, \cdot)$ par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$.

Nous proposons un algorithme, linéaire en le nombre de particules et asymptotiquement sans biais, reposant sur des méthodes particulaires pour les EDS combinées avec un algorithme de lissage récent) pour approcher l'espérance de fonctionnelles additives de la forme $\mathbb{E}\left[H(X_{0:n})|Y_{0:n}\right]$, où $H(X_{0:n}) = \sum_{k=0}^{n-1} h_k\left(X_k, X_{k+1}\right)$.

Nous montrons un résultat de convergence et une illustration de l'efficacité de la méthode sur un modèle donné pour approcher la quantité pivot de la phase E d'un algorithme EM.

Mots-clés. Equation différentielle stochastique, lissage en ligne, filtre non linéaire, méthodes de Monte Carlo séquantielles.

Abstract. We consider a Hidden Markov Model (HMM) where the latent data is a continuous time process $(X_t)_{t\ge 0}$, solution to the stochastic differential equation (SDE):

$$\mathrm{d}X_t = \alpha^\theta(X_t)\mathrm{d}t + \mathrm{d}W_t, X_0 \sim \chi_0^\theta \,, \tag{2}$$

where α^θ is a function depending on the parameter θ, W_t is a standard Wiener process, and χ_0^θ is the initial distribution of the state X_0.

This continuous time process is only partially observed at times $t_0 = 0, \ldots, t_n$ throughout an observation process $(Y_k)_{k=0,\ldots,n}$, such that conditional on $(X_t)_{0\leq t\leq t_n}$, the observations $(Y_k)_{0\leq k\leq n}$ are independent and the conditional distribution of Y_k has p.d.f. $g^\theta(X_k, \cdot)$ with respect to the Lebesgue measure.

We propose a new algorithm, with linear complexity, based on on particle filtering for partially observed SDEs combined with a recent online smoother algorithm for HMM to get unbiased approximation of additive functionals given by $\mathbb{E}\left[H(X_{0:n})|Y_{0:n}\right]$, where $H(X_{0:n}) = \sum_{k=0}^{n-1} h_k\left(X_k, X_{k+1}\right)$.

In this work, we describe the key steps of this algorithm and provide convergence results along with an application to approximate the E step of the EM algorithm.

Keywords. Stochastic differential equation, online smoother algorithm, particle filter, sequential Monte Carlo methods.

1 Introduction

L'inférence des chaînes de Markov cachées (HMM) (formées d'un processus Markovien $(X) = (X_k, 0 \leq k \leq n)$ observé indirectement au travers des variables $(Y_k, 0 \leq k \leq n)$, supposées indépendantes conditionnellement à (X)) ou la reconstruction du processus caché requierent la connaissance des lois de filtrage (i.e. les lois de X_k sachant les observations $(Y_0, \ldots, Y_k)$ pour $0 \leq k \leq n$) et des lois de lissage (i.e. les lois jointes des $(X_k, \ldots, X_p)$ sachant l'ensemble des observations $(Y_0, \ldots, Y_n)$).

Ces distributions de lissage peuvent être approchées par des méthodes de Monte Carlo séquentielles (SMC) tels que l'algorithme Forward Filtering Backward Smoothing (FFBS) ou l'algorithme Forward Filtering Backward Simulation (FFBSi). Ces algorithmes nécessitent une première phase forward pour générer un ensemble de particules et de poids qui approchent les distributions de filtrage et dans un second temps une passe backward, que ce soit par simulation (algorithme FFBSi), ou pour le calcul de nouveaux poids de lissage (algorithme FFBS). Lorsque l'on cherche à approcher une espérance sous la loi de lissage d'une fonctionelle additive, Olsson et Westerborn (2016) ont proposé un algorithme en ligne de complexité linéaire, permettant l'approximation de cette espérance sans effectuer de passe backward.

Lorsque le processus $(X_k)_{0\leq k\leq n}$ est la solution d'une équation différentielle stochastique (EDS) observée aux instants $(t_k)_{0\leq k\leq n}$, la densité de transition du processus X n'est en général pas connue et la seule solution présentée dans la littérature pour résoudre ce problème du lissage dans le cadre d'EDS partiellement observée en utilisant des méthodes SMC a été proposée par Olsson et Strojby (2011) et étend l'algorithme de lissage fixed-lag proposé initialement par Olsson et all (2008). Cette solution est biaisée et ce biais ne disparait pas lorsque le nombre de particules utilisées tend vers $+\infty$.

Nous proposons un algorithme efficace et non biaisé afin d'approcher l'espérance de fonctionnelles additives (par exemple l'étape E d'un algorithme EM) en étendant l'algorithme PaRIS de Olsson et Westerborn (2016) au cas où la densité de transition du processus est inconnue. Le résultat crucial et pourtant simple de l'application de PaRIS aux EDS repose sur le fait que le mécanisme d'acceptation rejet qui assure la complexité linéaire de l'algorithme PaRIS est toujours valable lorsque les noyaux de transition sont remplacés par un estimateur non biaisé. Notre algorithme, nommé GRand PaRIS, peut être appliqué à des HMM généraux dont la dynamique est régie par une EDS mais aussi à un HMM pour lequel la densité de transition peut être estimée sans biais.

2 Algorithme GRand PaRIS

On définit $(X) = (X_t, t \geq 0)$, un processus solution faible de l'EDS dans $\mathbb{R}^d$ suivante :

$$X_0 = x_0 \quad \text{and} \quad \mathrm{d}X_t = \alpha(X_t)\mathrm{d}t + \mathrm{d}W_t \;, \tag{3}$$

où $(W_t)_{t\geq 0}$ est un mouvement Brownien standard. On suppose que α est de la forme $\alpha(x) = \nabla_x A(x)$ où $A : \mathbb{R}^d \to \mathbb{R}$ est une fonction deux fois différentiable.

Pour tout $0 \leq k \leq n$, la loi de Y_k sachant $X_k := X_{t_k}$ a une densité par rapport à une mesure de référence λ sur $\mathbb{R}^m$ donnée par $g(X_k, Y_k) = g_k(X_k)$. La distribution de X_0 a une densité par rapport à une mesure de référence μ sur $\mathbb{R}^d$ donnée par χ et pour tout $1 \leq k \leq n$, la loi de X_{k+1} sachant X_k a une densité par rapport à la mesure μ notée $q_k(X_k, \cdot)$.

Un prérequis, l'algorithme PaRIS

Soit $0 \leq k \leq k' \leq n$, la distribution de lissage jointe des états cachés $\phi_{k:k'|n}$ est définie, pour tout fonction mesurable h par $\phi_{k:k'|n}[h] = \mathbb{E}\left[h(X_k, \ldots, X_{k'})|Y_{0:n}\right]$, et $\phi_k = \phi_{k:k|k}$ est la loi de filtrage. L'algorithme PaRIS (Olsson et Westerborn, 2016) permet dans le cadre HMM classique, d'approcher sans biais des quantités de la forme

$$\phi_{0:n|n}[H_n] = \mathbb{E}\left[H_n(X_{0:n})|Y_{0:n}\right] \text{ si } H_n = \sum_{k=0}^{n-1} h_k(X_k, X_{k+1}) \;. \tag{4}$$

La distribution de filtrage et la distribution de lissage sont liées par la relation suivante :

$$\phi_{0:n|n}[h] = \phi_n[T_n[h]] \;, \text{ où } T_n[h](X_n) = \mathbb{E}\left[h(X_{0:n})|X_n,\; Y_{0:n}\right] \;. \tag{5}$$

L'algorithme PaRIS peut donc être utilisé dans le cas où, quand θ est connu, la densité de transition q^θ est connue.Cet algorithme, reposant la relation (5) (cf Olsson et Westerborn (2016) s'écrit :

- **Initialisation pour $k = 0$.**
  N particules sont tirées indépendamment $\xi_0^\ell \sim \eta_0$, avec η_0 une densité de probabilité par rapport à μ. Chaque ξ_0^ℓ est associée au poids d'importance $\omega_0^\ell =$

$\chi(\xi_0^\ell)g_0(\xi_0^\ell)/\eta_0(\xi_0^\ell)$. Pour toute fonction bornée et mesurable h définie sur $\mathbb{R}^d$, $\phi_0[h]$ peut être approché par $\phi_0^N[h] = (\Omega_0^N)^{-1} \sum_{\ell=1}^N \omega_0^\ell h\left(\xi_0^\ell\right)$ avec $\Omega_0^N := \sum_{\ell=1}^N \omega_0^\ell$.

- **Récursion de $k-1$ à $k \leq n$.** Filtre particualire
  Répéter N fois la procédure suivante :

  - Choisir un indice de particule I_k^ℓ au temps $k-1$ dans $\{1,\ldots,N\}$ avec une probabilité proportionnelle aux ω_{k-1}^j, $1 \leq j \leq N$.
  - Tirer ξ_k^ℓ en utilisant l'indice précédent grâce à un noyau de propagation p_k: $\xi_k^\ell \sim p_k(\xi_{k-1}^{I_k^\ell}, \cdot)$.
  - Associer à cette particule ξ_k^ℓ, le poids :

$$\omega_k^\ell := \frac{q_k(\xi_{k-1}^{I_k^\ell}, \xi_k^\ell) g_k(\xi_k^\ell)}{p_k(\xi_{k-1}^{I_k^\ell}, \xi_k^\ell)} \ . \tag{6}$$

  L'espérance $\phi_k[h]$ est approchée par $\phi_k^N[h] := (\Omega_t^N)^{-1} \sum_{\ell=1}^N \omega_k^\ell h\left(\xi_k^\ell\right)$ avec $\Omega_k^N := \sum_{\ell=1}^N \omega_k^\ell$.

L'algorithme PaRIS repose sur la décomposition suivante :
$T_k[H_k](X_k) = \mathbb{E}\left[T_{k-1}[H_{k-1}](X_{k-1}) + h_{k-1}(X_{k-1}, X_k) | X_k, Y_{0:k-1}\right]$.

En introduisant les statistiques suffisantes τ_k^i (partant de $\tau_0^i = 0$, $1 \leq i \leq N$) pour approcher $T_k[H_k](\xi_k^i)$, pour $1 \leq i \leq N$ et $0 \leq k \leq n$ Olsson et Westerborn (2016) remplacent ϕ_{k-1} par ϕ_{k-1}^N

$$\tau_k^i = \sum_{j=1}^N \Lambda_{k-1}^N(i,j) \left\{\tau_k^j + h_{k-1}(\xi_{k-1}^j, \xi_k^i)\right\} \ , \tag{7}$$

avec

$$\Lambda_k^N(i,\ell) = \frac{\omega_k^\ell q_k(\xi_k^\ell, \xi_{k+1}^i)}{\sum_{\ell=1}^N \omega_k^\ell q_k(\xi_k^\ell, \xi_{k+1}^i)} \ , \quad 1 \leq \ell \leq N \ . \tag{8}$$

Pour réduire le coût algorithmique de la normalisation des poids, (qui augmente quadratiquement avec N), l'algorithme PaRIS propose d'échantillonner dans l'ensemble $\{\xi_{k-1}^j\}_{j=1}^N$, par une méthode d'acceptation rejet, selon les poids $\Lambda_k^N(i,\cdot)$ pour approcher l'espérance (7). Pour ce faire, on fait l'hypothèse qu'il existe σ_+ tel que:

$$\forall x, y \in \mathbb{R}^d, q(x,y) \leq \sigma_+$$

Alors, l'algorithme de rejet suivant permet de tirer selon les poids définis en (8):

1. Tirer J avec des poids proportionels à $(\omega_k^j)_{j=1,\ldots,N}$;

2. Tirer $U \sim \mathcal{U}[0,1]$;

3. Si $U < q_k(\xi_k^J, \xi_{k+1}^i)/\sigma_+$, accepter J, sinon, retour en 1.

Ainsi, en choisissant $\tilde{N} \geq 1$ candidats selon cet algorithme, on calcule à chaque instant k une approximation de (7) de la façon suivante.

- **Récursion de $k-1$ à $k \leq n$.** Filtre et lisseur

  (i) Effectuer une étape d'un filtre particulaire pour produire $\{(\xi_k^\ell, \omega_k^\ell)\}$ pour $1 \leq \ell \leq N$.

  (ii) Pour tout $1 \leq i \leq N$, échantillonner indépendamment $J_k^{i,\ell}$ dans $\{1,\ldots,N\}$ pour $1 \leq \ell \leq \widetilde{N}$ avec pronabilités $\Lambda_k^N(i,\cdot)$, définies par (8), grâce à la procédure d'acceptation rejet.

  (iii) Poser $\tau_{k+1}^i := \widetilde{N}^{-1} \sum_{\ell=1}^{\widetilde{N}} \left\{ \tau_k^{J_k^{i,\ell}} + h_k \left(\xi_k^{J_k^{i,\ell}}, \xi_{k+1}^i \right) \right\}$.

Ainsi, à l'étape n la quantité d'intérêt (4) est approchée par $\phi_n^N[\tau_n] = \frac{1}{\Omega_n^N} \sum_{i=1}^N \omega_n^i \tau_n^i$.

Extension au cas q_k inconnu, le GRand PaRIS

L'étape de filtre particulaire est menée grâce aux on s'appuie sur les travaux de Fernhead et *al.*, où les poids de filtrage de l'équation (6) sont approchés à chaque étape par des poids aléatoires:

$$\widehat{\omega}_k^\ell := \frac{\widehat{q}_k(\xi_{k-1}^{I_k^\ell}, \xi_k^\ell; \zeta_k) g_k(\xi_k^\ell)}{p_k(\xi_{k-1}^{I_k^\ell}, \xi_k^\ell)} , \tag{9}$$

où $\widehat{q}_k(\xi_{k-1}^{I_k^\ell}, \xi_k^\ell; \zeta_k)$ est un **estimateur strictement positif et non biaisé** de $q_k(\xi_{k-1}^{I_k^\ell}, \xi_k^\ell)$ se basant sur une variable aléatoire ζ_k nécessitant la simulation de ponts Browniens. Cet estimateur est appelé GPE (Generalized Poisson Estimator).

Dans le cas où l'EDS (3) satisfait les conditions des domaines $\mathcal{D}_1$ ou $\mathcal{D}_2$ définis dans Beskos et *al.* (2006), on peut montrer qu'il existe $\hat{\sigma}_k^+$ tel que

$$\widehat{q}_k(\xi_{k-1}^{I_k^\ell}, \xi_k^\ell; \zeta_k) < \hat{\sigma}_k^+ . \tag{10}$$

Cette propriété permet naturellement d'obtenir une alternative à l'algorithme d'acceptation rejet défini plus haut, pierre angulaire de l'algorithme PaRIS. En effet, nous montrons la proposition suivante:

Proposition 1 *Soit J la variable aléatoire définie par l'algorithme d'acceptation rejet suivant:*

1. Tirer J avec des poids proportionels à $(\widehat{\omega}_k^j)_{j=1,\ldots,N}$*;*

2. Tirer ζ_k *selon l'algorithme de GPE proposé par Fernhead et al (2008);*

3. Tirer $U \sim \mathcal{U}[0,1]$*. Si* $U < \frac{\hat{q}_k(\xi_k^J,\xi_{k+1}^i,\zeta_k)}{\widehat{\sigma}_k^+}$*, accepter J, sinon, retour en 1.*

Alors, J est distribuée selon la loi discrète $\Lambda_k^N(i,\cdot)$ *sur* $\{1,\ldots,N\}$*.*

Ainsi, le mécanisme de l'algorithme PaRIS reste valable quand les poids de filtrage sont approchés sans biais grâce au GPE, et lorsque l'acceptation rejet est effectuée à l'aide de ces poids et d'un nouveau GPE. Il en découle aussitôt un nouvel estimateur de la quantité (4), noté $\hat{\phi}_n^N[\tau_n]$. On montre le résultat de convergence suivant:

Proposition 2 *Sous certaines conditions de régularité, pour tout* $0 \leq k \leq n$ *et* $\widetilde{N} \geq 1$*, il existe* $b_k, c_k > 0$ *tels que pour tout* $N \geq 1$ *et tout* $\varepsilon \in \mathbb{R}_+^\star$*,*

$$\mathbb{P}\left(\left|\hat{\phi}_k^N[\tau_k] - \phi_k\left[T_k h_k\right]\right| \geq \varepsilon\right) \leq b_k \exp\left(-c_k N \varepsilon^2\right) \ .$$

Cet algorithme est comparé à l'algorithme du fixed-lag (pour différentes valeurs de lag) pour calculer la quantité pivot de l'étape E de l'algorithme EM et, à coùt de calcul fixé, l'algorithme GRand PaRIS montre un meilleur comportement en termes de biais et une variance similaire.

Bibliographie

[1] Fearnhead, P. et Papaspiliopoulos, O. et Roberts, G. (2008), *Particle filters for partially observed diffusions*, J. Roy. Statist. Soc. Ser. B, 70(4):755–777, 2006.
[2] Olsson, J. et Cappé, O. et and Douc, R. et Moulines, E. (2008), *Sequential Monte Carlo smoothing with application to parameter estimation in nonlinear state space models*, Bernoulli.
[3] Olsson, J et Strojby, J. (2011), *Particle-based likelihood inference in partially observed diffusion processes using generalised Poisson estimators*, Electron. J. Statist., 5:1090–1122.
[4] Olsson, J et Westerborn, J. (2016) *Efficient particle-based online smoothing in general hidden Markov models: the PaRIS algorithm*, ArXiv:1412.7550.