

Etude de l'influence de la modélisation sur les performances du filtre particulaire Rao-Blackwellisé

Manuel DAVY, Thomas BRÉHARD
 CNRS/LAGIS et INRIA Futur équipe SequeL,
 40 Avenue Halley, 59650 Villeneuve d'Ascq, France
 Manuel.Davy@inria.fr, Thomas.Brehard@inria.fr

Résumé – Dans de nombreux problèmes de filtrage, le modèle dynamique pour une partie de l'état est conditionnellement linéaire et gaussien conditionnellement au reste de l'état. Il est alors possible de mettre en oeuvre un filtre rao-blackwellisé. Ici, nous étudions un cas particulier important de ce cadre général, où l'état comprend les paramètres de fonctions de bases, et les amplitudes de ces fonctions de bases, qui se combinent dans l'équation d'observation. Nous montrons que l'utilisation d'un modèle d'état simplifié et approximatif peut, sous certaines conditions, être une meilleure solution que l'utilisation d'un modèle plus précis mais plus complexes.

Abstract – In a number of filtering problems, the dynamic state-space model for a part of the state is linear and Gaussian conditionnal on the rest of the state. It is then possible to implement a Rao-Blackwellized filter. In this paper, we study an important special case of this framework, where the state is composed of some basis function parameters as well as their amplitudes. The observation is a linear combination of the amplitude weighted basis functions. We show that using a simplified model may, under some conditions, be a better solution than using a precise but more complex model.

1 Introduction

Supposons que l'on souhaite estimer séquentiellement un paramètre θ_t à l'aide d'observations via l'équation suivante :

$$\mathbf{y}_t = C(\theta_t)\beta_t + \epsilon_t \quad (1)$$

où $C(\theta_t)$ est une matrice et ϵ_t un bruit blanc gaussien de matrice de covariance Σ_{y_t} . $\{\theta_t\}_{t \in \mathbb{N}}$ est un processus markovien et son équation de transition est définie de la façon suivante :

$$\theta_t \sim k_t(\theta_t | \theta_{t-1}) . \quad (2)$$

Classiquement, le vecteur des amplitudes β_t est modélisé comme un processus markovien.

Modèle «markovien» pour β_t :

$$\beta_t \sim \mathcal{N}(\beta_t; \beta_{t-1}, \Sigma_t(\beta_{t-1})) . \quad (3)$$

Les équations (1), (2) et (3) forment un modèle conditionnellement linéaire gaussien. Ce type de modèle est utilisé dans des applications aussi diverses que la séparation de sources [6], la transcription automatique de musique [5] ou encore la déconvolution aveugle [7]. La loi *a posteriori* $p(\theta_{0:t} | \mathbf{y}_{0:t})$ peut-être approchée à l'aide d'un filtre particulaire Rao-Blackwellisé [4]. Comme nous le précisons dans la section 2, cette technique impose d'utiliser un filtre de Kalman pour calculer la vraisemblance lors de la mise à jour des poids des particules.

Récemment, Dubois a montré dans [5] qu'une modélisation plus grossière de β_t permettait d'obtenir des estimations plus robustes avec un coût algorithmique relativement faible. L'idée consiste à supprimer la dépendance temporelle de β_t en utilisant le modèle approché suivant :

Modèle «indépendant» pour β_t :

$$\beta_t \sim \mathcal{N}(\beta_t; \mu_t(\theta_t), \Sigma_t(\theta_t)) . \quad (4)$$

Cette approximation permet de s'affranchir de l'utilisation du filtre de Kalman pour calculer la vraisemblance lors de la phase de mise à jour des particules. De plus, comme l'on montré Davy *et al* [2], cette dernière peut-être calculée très rapidement.

Nous comparons dans cet article les deux algorithmes sur la base de simulations. Cette comparaison est complétée par une étude de l'influence de la modélisation sur la variance asymptotique. Cette approche s'appuie sur les résultats de convergence du filtre particulaire [1] et notamment du Théorème Central Limit (TCL).

2 Filtre particulaire Rao Blackwellisé

Nous décrivons dans cette section les filtres particuliers obtenus suivant les modélisations de β_t .

2.1 Filtre particulaire avec un modèle «markovien» de β_t

Considérons le modèle formé par les équations (1), (2) et (3). Dans ce contexte, l'estimation de la densité *a posteriori* $p(\theta_{0:t} | \mathbf{y}_{0:t})$ est basée sur la décomposition suivante :

$$p(\theta_{0:t} | \mathbf{y}_{0:t}) = \frac{p(\mathbf{y}_t | \theta_{0:t}, \mathbf{y}_{0:t-1}) k_t(\theta_t | \theta_{t-1})}{p(\mathbf{y}_t | \mathbf{y}_{0:t-1})} p(\theta_{0:t-1} | \mathbf{y}_{0:t-1})$$

Malheureusement, la vraisemblance $p(\mathbf{y}_t | \theta_{0:t}, \mathbf{y}_{0:t-1})$ ne peut pas se simplifier en $p(\mathbf{y}_t | \theta_t)$ du fait de la dépendance aux

valeurs passées de $\beta_{0:t}$. Cette remarque est une conséquence directe de l'utilisation du modèle markovien pour β_t . Néanmoins, cette vraisemblance peut-être obtenue analytiquement. En effet, $p(\mathbf{y}_t|\theta_{0:t}, \mathbf{y}_{0:t-1})$ peut être réécrit sous la forme d'une intégrale :

$$p(\mathbf{y}_t|\theta_{0:t}, \mathbf{y}_{0:t-1}) = \int p(\mathbf{y}_t|\theta_t, \beta_t)p(\beta_t|\theta_{0:t}, \mathbf{y}_{0:t-1})d\beta_t. \quad (5)$$

Le terme $p(\beta_t|\theta_{0:t}, \mathbf{y}_{0:t-1})$ peut être calculé à l'aide d'un filtre de Kalman et $p(\mathbf{y}_t|\theta_t, \beta_t)$ est directement donné par (1) de sorte qu'une forme analytique peut être obtenue pour $p(\mathbf{y}_t|\theta_{0:t}, \mathbf{y}_{0:t-1})$. L'algorithme final est décrit ci-dessous.

2.2 Filtre particulière avec un modèle «in-dépendant» de β_t

L'estimation de la densité *a posteriori* $p(\theta_{0:t}|\mathbf{y}_{0:t})$ basée sur la modélisation décrite par les équations (1), (2) et (4) est obtenue via la décomposition suivante :

$$p(\theta_{0:t}, \beta_{0:t}|\mathbf{y}_{0:t}) = \frac{p(\mathbf{y}_t|\theta_t, \beta_t)p(\beta_t|\theta_t)k_t(\theta_t|\theta_{t-1})}{p(\mathbf{y}_t|\mathbf{y}_{0:t-1})} \times p(\theta_{0:t-1}, \beta_{0:t-1}|\mathbf{y}_{0:t-1}). \quad (6)$$

En intégrant (5) par rapport à $\beta_{0:t}$, on obtient

$$p(\theta_{0:t}|\mathbf{y}_{0:t}) = \frac{p(\mathbf{y}_t|\theta_t)k_t(\theta_t|\theta_{t-1})}{p(\mathbf{y}_t|\mathbf{y}_{0:t-1})}p(\theta_{0:t-1}|\mathbf{y}_{0:t-1}). \quad (7)$$

où

$$p(\mathbf{y}_t|\theta_t) = \int p(\mathbf{y}_t|\theta_t, \beta_t)p(\beta_t|\theta_t)d\beta_t. \quad (8)$$

Du fait de la modélisation utilisée, l'intégrale (8) peut être calculée analytiquement (cf. [2]). En particulier, Davy *et al* ont montré que $p(\mathbf{y}_t|\theta_t)$ pouvait être calculé rapidement.

Algorithmes de filtrage particulière

Initialisation.

- Pour $i = 1, \dots, N$, générer $\theta_0^{(i)} \sim q_0(\theta_0)$ et calculer $\tilde{w}_0^{(i)} = p_0(\tilde{\theta}_0^{(i)})/q_0(\tilde{\theta}_0^{(i)})$

Iterations.

- Pour $t = 1, 2, \dots$, faire
 - Pour $i = 1, \dots, N$, faire
 - Tirer $\tilde{\theta}_t^{(i)} \sim q_t(\theta_t|\mathbf{y}_t, \theta_{t-1}^{(i)})$
 - Si le modèle markovien de l'équation (3) est utilisé : mettre à jour les poids

$$\tilde{w}_t^{(i)} = \tilde{w}_{t-1}^{(i)} \frac{p(\mathbf{y}_t|\tilde{\theta}_{0:t}^{(i)}, \mathbf{y}_{0:t-1})k_t(\tilde{\theta}_t^{(i)}|\theta_{t-1}^{(i)})}{q_t(\tilde{\theta}_t^{(i)}|\mathbf{y}_t, \theta_{t-1}^{(i)})} \quad (9)$$

où $p(\mathbf{y}_t|\tilde{\theta}_{0:t}^{(i)}, \mathbf{y}_{0:t-1})$ est calculé à l'aide d'un filtre de Kalman.

- Si le modèle indépendant de l'équation (4) est utilisé : mettre à jour les poids

$$\tilde{w}_t^{(i)} = \tilde{w}_{t-1}^{(i)} \frac{p(\mathbf{y}_t|\tilde{\theta}_t^{(i)})k_t(\tilde{\theta}_t^{(i)}|\theta_{t-1}^{(i)})}{q_t(\tilde{\theta}_t^{(i)}|\mathbf{y}_t, \theta_{t-1}^{(i)})} \quad (10)$$

où $p(\mathbf{y}_t|\tilde{\theta}_t^{(i)})$ est calculé via (8).

- Normaliser les poids $\tilde{w}_t^{(i)}$ de sorte que $\sum_{i=1}^N \tilde{w}_t^{(i)} = 1$
 - Si $[\sum_{i=1}^N (\tilde{w}_t^{(i)})^2]^{-1} \leq \eta N$, alors rééchantillonner, i.e., dupliquer les particules ayant un poids élevé et supprimer les autres. Les nouvelles particules sont notées $(\theta_t^{(i)})$ avec $w^{(i)} = 1/N$
 - Sinon, renommer les particules $\theta_t^{(i)} \leftarrow \tilde{\theta}_t^{(i)}$.
-

3 Comparaison des modélisations à l'aide de la variance asymptotique

Il existe plusieurs résultats sur la convergence du filtre particulière [1, 3]. Nous présentons ici un théorème Central Limit [3] et une formule pour calculer récursivement la variance asymptotique [1].

En premier lieu, introduisons les notations suivantes :

$$\begin{cases} \pi_t & \triangleq p(\theta_t|\mathbf{y}_{0:t}), \\ \tilde{\pi}_t & \triangleq p(\theta_t|\mathbf{y}_{0:t-1}), \\ \nu_t & \triangleq \frac{\pi_t}{\tilde{\pi}_t}. \end{cases} \quad (11)$$

Selon [1], pour un filtre particulière avec rééchantillonnage multinomial la variance asymptotique pour une fonction mesurable $\phi : \Theta_t \rightarrow \mathbb{R}^d$ notée $V_t(\phi)$ est obtenue récursivement via les formules suivantes :

$$\begin{cases} \tilde{V}_t(\phi) = \tilde{V}_{t-1}\{\mathbb{E}_{k_t}(\phi)\} + \mathbb{E}_{\pi_{t-1}}\{\text{Var}_{k_t}(\phi)\} & t > 0, \\ V_t(\phi) = \tilde{V}_t\{\nu_t(\phi - \mathbb{E}_{\pi_t}\phi)\} & t \geq 0, \\ \hat{V}_t(\phi) = V_t(\phi) + \text{Var}_{\pi_t}(\phi) & t \geq 0 \end{cases} \quad (12)$$

avec $\tilde{V}_0(\phi) = \text{Var}_{\tilde{\pi}_0}(\phi)$. Cette variance asymptotique apparaît directement dans le théorème de convergence du filtre particulière [3] présenté ci-dessous.

Théorème 1 *Sous des conditions techniques (cf. [3]), pour un filtre particulière avec rééchantillonnage multinomial :*

$$N^{1/2} \left\{ \frac{\sum_{n=1}^N w_t^{(n)} \phi(\theta_t^{(n)})}{\sum_{n=1}^N w_t^{(n)}} - \mathbb{E}_{\pi_t}(\phi) \right\} \rightarrow \mathcal{N}(0, V_t(\phi)) \quad (13)$$

Les variances asymptotiques associées aux algorithmes présentés dans la section 2 peuvent être obtenues récursivement via (12). Bien évidemment, elles ne peuvent pas être calculées analytiquement. Toutefois, il est possible de les comparer empiriquement, sur la base de simulations. En effet, à partir d'un nombre suffisamment grande de particules, la variance empirique, estimée à partir des particules, suit une loi en A/\sqrt{N} , où la constante A indique le taux de convergence asymptotique. Dans la section suivante, nous présentons des résultats obtenu avec cette approche.

4 Simulations

Le modèle utilisé pour les simulations est de type sinusoides bruitées. Plus précisément, la matrice $C(\theta)$ est composée de sinus et de cosinus dont les fréquences sont les éléments du vecteur θ_t , modulées par une fenêtre de Hamming (voir [2] pour une description complète de ce modèle), et \mathbf{y}_t est un signal scalaire de 400 points, représenté à la figure 1.

Les deux modèles et les deux algorithmes ont été mis en œuvre, avec 100 simulations pour divers nombres de particules. Dans les deux cas, la loi d'importance pour θ_t est choisie égale à la loi d'évolution, sous la forme d'une marche aléatoire gaussienne. Pour chaque simulation, on estime

$$\hat{\theta} = \frac{1}{T} \sum_{t=1}^T \int_{[0;0.5]} \theta_t p(\theta_t|\mathbf{y}_{1:t}) d\theta_t \quad (14)$$

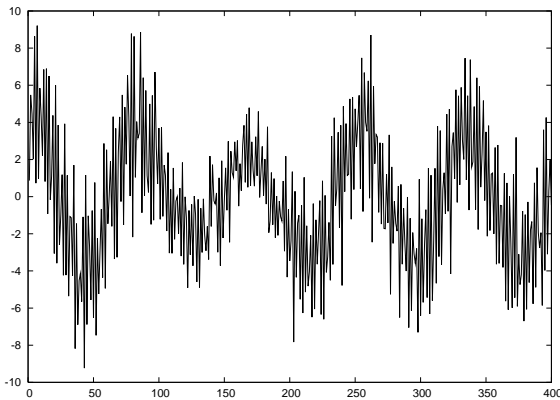


FIG. 1 – Signal d’observation traité par le filtre. Il est composé de deux sinusoides de fréquences constantes (égales à 0.01232 et 0.432) et d’amplitudes stationnaires par morceaux (pour la première composante, l’amplitude vaut 1 sur [1200] et 3 sur [201400]. pour la seconde composante, elle vaut 4 sur [1100] et 2 sur [101400].) Un bruit blanc gaussien de variance unité est ajouté.

Le tableau 1 présente le biais calculé pour les deux algorithmes et différents nombres de particules. Comme on le voit, le modèle indépendant est moins biaisé que le modèle Markovien, et cela pour tous nombres de particules testés. Le tableau 2 présente les variances estimées sur 100 simulations de l’évaluation de $\hat{\theta}$. Le modèle indépendant présente une variance plus faible pour $N = 50$ et $N = 100$, où le biais est en outre un ordre de grandeur plus faible que pour le modèle markovien. Lorsque le nombre de particules augmente, l’écart entre les deux modèles est moins sensible. Néanmoins, il faut souligner que le modèle indépendant permet une implémentation plus rapide que le modèle dépendant, et son avantage est clair à temps de calcul constant.

	$N = 50$	$N = 100$	$N = 500$	$N = 1000$
M ($\times 10^{-5}$)	202	27.1	4.8	4.19
I ($\times 10^{-5}$)	21.9	9.8	4.25	2.95

TAB. 1 – Biais calculé sous la forme de l’écart quadratique moyen entre $\hat{\theta}$ et sa valeur utilisée en simulations, moyenné sur 100 simulations pour les deux modèles (M= «markovien» I= «indépendant») avec divers nombres de particules (notés N).

5 Conclusion et perspectives

Les résultats de comparaison des variances asymptotiques obtenus par simulation montre l’intérêt de ce type

	$N = 50$	$N = 100$	$N = 500$	$N = 1000$
M ($\times 10^{-3}$)	30.3	10.8	0.85	0.52
I ($\times 10^{-3}$)	12.8	8.3	3.4	1.8

TAB. 2 – Variances moyennes sur 100 simulations calculée pour tous les pas de temps pour les deux modèles (M= «markovien» I= «indépendant») avec divers nombres de particules (notés N).

FIG. 2 – Evolution de la variance asymptotique estimée empiriquement pour le modèle indépendant

d’approche. En particulier, Ils permettent de préciser dans quelle mesure des modèles plus simples permettent d’obtenir des estimations robustes. Ces résultats sont en cours d’exploitation pour mettre au point un algorithme de filtrage particulière pour la transcription automatique de la musique à faible complexité algorithmique, dans la lignée des travaux de [5].

Références

- [1] N. Chopin. Central Limit Theorem for Sequential Monte Carlo Methods and Its Application to Bayesian Inference. *The Annals of Statistics*, 32(6), 2004.
- [2] Manuel Davy and Corentin Dubois. A Fast Filtering Approach to Bayesian Tonal Music Transcription. *NIPS 2006 Workshop “Advances in Models for Acoustic Processing”*, 2006.
- [3] R. Douc and E. Moulines. Limit theorems for weighted samples with applications to sequential Monte Carlo methods. *Annals of Statistics*, 2007. à paraître.
- [4] A. Doucet, S. Godsill, and C. Andrieu. On Sequential Monte Carlo Sampling Methods for Bayesian Filtering. Technical report, Cambridge University Engineering Department, 2000.
- [5] C. Dubois. *Méthodes de Monte Carlo séquentielles pour l’estimation de fréquences fondamentales Application à la caractérisation et à la reconnaissance de l’effet de métabole, en environnement urbain*. PhD thesis, Ecole Centrale de Nantes, Université de Nantes, 2006.
- [6] J.-R. Larocque, J.P. Reilly, and W. Ng. Particle Filters for Tracking an Unknown Number of Sources. *IEEE Trans. on Signal Processing*, 50(12), February 2002.
- [7] J.S. Liu, , and R. Chen. Blind Deconvolution via Sequential Imputation. *Journal of the American Statistical Association*, 90 :567–576, 1995.

