

Modèle d'état à coefficient aléatoire. Filtrage par densités approchées

Random Coefficient State Model. Approximated Densities Filtering

par Denis de BRUCQ, Virginie RUIZ

La31 / LACIS-ITEPEA - UFR des Sciences et Techniques
BP 118, 76821 Mont-Saint-Aignan Cedex - France.

Résumé

Le filtrage de Bucy-Kalman s'applique au modèle d'état comprenant des équations linéaires bruitées, décrivant l'évolution de l'état et des équations linéaires bruitées d'observation. Ce filtrage consiste dans le cas gaussien, à calculer de façon récursive, la loi de probabilité, a posteriori, de l'état, au vu de l'observation actuelle et des observations passées. Le filtrage par densités approchées permet de traiter des équations d'état, non linéaires ou à bruits non Gaussiens.

Pour un coefficient de rappel aléatoire, cas typique d'une situation de changements de modèles, l'article introduit une famille de lois de probabilité, paramétrées, bimodales servant, par ajustement des paramètres, à approcher les lois a posteriori de l'état aux divers instants. Les paramètres sont recalculés récursivement, lors des mises à jour et des prédictions.

Mots clés : Filtrage Non Linéaire, Détection de ruptures, Lois Exponentielles.

Abstract

The Kalman Filtering applies to state models with noisy linear equations which describe the state evolution and with noisy linear equations of observations. This filtering recursively computes the a posteriori state law given the present and past observations. The approximated densities filtering allows to process either nonlinear state equations or equations with non Gaussian noises.

For a random dynamical coefficient, typical situation of models abrupt changes, this paper introduces bi-modal parametered laws of probability which are used to approach the a posteriori state laws at any time by adapting parameters. These are recursively computed at each updating and prediction.

Key words : Nonlinear Filtering, Detection of abrupt changes, Exponential Laws.

1. Introduction

La méthode qui va être proposée étend celle classique, du filtrage de Bucy-Kalman. Nous la présenterons sur un modèle d'état explicite, simple, non linéaire et scalaire pour des raisons didactiques.

Une modélisation de l'évolution de l'état X et de la prise d'information Y , est nécessaire. Le temps t est échantillonné puisque le traitement est effectué par ordinateur. Souvent l'évolution de l'état X , au cours du temps, est approchée par une équation linéarisée à laquelle on rajoute un bruit blanc W gaussien. L'observation $Y(t)$ provient linéairement de l'état $X(t)$ auquel s'ajoute un bruit blanc, gaussien, noté $V(t)$; $t \in \mathbf{N}$

Ainsi, pour pouvoir appliquer le filtrage de Bucy-Kalman, nous observons qu'entre la réalité physique et le modèle d'état des approximations sont nécessaires.

Introduisons pour fixer les idées, le modèle d'état (échantillonné d'une équation différentielle) :

$$X(t+1) = X(t) - aX(t) + b + W(t)$$

$$Y(t) = X(t) + V(t)$$

Le coefficient a de rappel est un scalaire fixe et sera par la suite pris fonction aléatoire du temps. Afin d'avoir un système stable, $|1 - a| < 1$. Un biais constant b , a été ajouté par ailleurs.

Dans ces conditions, le filtrage [cf. Kalman et Bucy 1961] fournit la moyenne $\hat{X}(t/t)$ de l'état $X(t)$ conditionnellement à l'observation vectorielle $y(1 \rightarrow t) = (y(1), \dots, y(t))$, moyenne estimée au sens de l'erreur moyenne quadratique minimale.

Les hypothèses précédentes, pourtant bien nombreuses, sont encore insuffisantes pour assurer le caractère gaussien de la loi de la variable aléatoire $X(t)$. En effet, il faut encore supposer que la loi de l'état $X(0)$ à la date initiale, $t = 0$, est gaussienne. En ce cas, les lois de l'état $X(t)$ a priori et a posteriori sont gaussiennes.

De nombreux auteurs [cf. Fujisaki & al 1972, Zakai 1969, ...] ont cherché à étendre la validité du filtrage de Bucy-Kalman, en considérant les estimations linéaires, optimales au sens de l'erreur moyenne quadratique minimale; on parle de technique du second ordre et on s'affranchit du caractère gaussien des diverses lois de probabilité.

Les procédures approchées, les plus naturelles pour des équations non linéaires, consistent à linéariser les équations autour d'une trajectoire nominale [cf. Jazwinski 1970]. On parle de filtres de Kalman étendus. Ensuite, des approximations quadratiques ont été effectuées [cf. de Larminat 1971].

Il est possible d'introduire des équations d'état comprenant l'observation ce qui conduit à des lois d'état, gaussiennes conditionnellement à l'observation [cf. de Brucq et Rakotopara 1986]. Ensuite le filtrage de Kalman s'applique récursivement, sur les équations linéarisées autour d'un point $\hat{X}(t/t-1)$ de fonctionnement, prédit à partir des observations passées $y(1 \rightarrow t-1)$ (cf. de Brucq).

Les situations multi-modèles [cf. Lainiotis 1971] consistent à prendre comme état Z , une paire $Z^t \triangleq (M, X)$, où les modèles possibles M appartiennent à un ensemble fini $\{m_1, \dots, m_n\}$. En augmentant l'état à identifier du système, on espère rendre indépendantes les lois probabilistes du passé et du futur lorsque le présent est connu; on reste dans la théorie des chaînes de Markov. L'état X vérifie une équation d'évolution markovienne connue :

$$X(t+1) = a(M) \cdot X(t) + W(t)$$

La loi de probabilité de l'état introduit une probabilité sur chacun des M modèles puis conditionnellement à chaque modèle, une loi de probabilité sur l'état $X(t)$.

Finalement, quelques très rares systèmes relèvent de la théorie des filtres finis : les lois de probabilité de $X(t)$, a priori et a posteriori, dépendent d'un nombre fini de paramètres calculables récursivement au fur et à mesure de l'arrivée des observations. [cf. Makowski 1982, Haussmann et Pardoux 1988].

La recherche de paramètres pertinents en nombre fini, pour définir des lois de probabilité, approchées, est naturelle; Fisher et Stear 1970 approchent les densités de probabilités à l'aide de combinaison linéaire de polynômes d'Hermite.

Comme les modèles d'état conduisant à des lois de probabilité après observation exprimable de façon analytique, sont rares, les densités doivent être le plus souvent approchées.

Une méthode simple et originale d'approximation est proposée dans cet article. A la connaissance des auteurs, aucune publication antérieure n'a utilisé le maximum d'entropie pour approcher les équations du filtrage relatives à un modèle d'état à équations non-linéaires. Les lois probabilistes qui en résultent, sont de type exponentiel. Elles comprennent un nombre fini de paramètres. Le filtrage non-linéaire conduit à des équations récursives sur ces paramètres. La méthode ferme les équations du filtrage non-linéaire par approximation par maximum d'entropie.

Le coefficient a n'est dorénavant plus constant, mais constitue une chaîne de Markov ($A(t); t \in \mathbf{N}$) à deux états. En raison du produit $A(t) X(t)$, quelles que soient les autres hypothèses supposées par ailleurs, le caractère gaussien de la loi de l'état, est perdu! Pour le modèle :

$$\begin{aligned} & A(t) \text{ chaîne de Markov} \\ X(t+1) &= X(t) - A(t) \cdot X(t) + b + W(t) \\ Y(t) &= X(t) + V(t) \end{aligned}$$

les techniques d'identification de paramètres ne peuvent pas s'appliquer puisque d'un instant à l'autre, le paramètre $A(t)$ peut changer de façon aléatoire.

Pour $X(0)$ gaussien, chacune des séquences possibles $A(0) = a(0), A(1) = a(1), \dots, A(t-1) = a(t-1)$ conduit à une moyenne et à une variance de $X(t)$, gaussien conditionnellement. Ainsi, en considérant les 2^t séquences possibles de probabilité calculable, $X(t)$ est un mélange de 2^t gaussiennes dont le nombre croît avec le temps t . Les équations du filtrage non-linéaire ne sont pas en nombre fini.

L'article présente, la méthodologie du filtrage par densités approchées. Elle consiste à développer le logarithme des densités de probabilité, a priori, et, a posteriori de l'état X comme des combinaisons linéaires de fonctions $\varphi_1, \varphi_2, \dots, \varphi_k$ à choisir au mieux. Ensuite, les probabilités π de transition de la chaîne ($A(t); t \in \mathbf{N}$) de Markov, sont introduites dans le paragraphe II. Puis, les densités de probabilité de type exponentiel décrivant les lois de probabilité de l'état X , sont mises à jour et prédites, compte tenu du modèle d'état. Finalement, la pertinence de la méthode est montrée par simulation.

2. Méthodologie

Le filtrage par densités approchées consiste à utiliser l'équation d'état :

$$X(t+1) = f(X(t), W(t)) \quad (1)$$

avec f non linéaire, quelconque et W bruit blanc d'état, de loi quelconque, pour calculer à l'aide du théorème de transfert des probabilités (théorème probabiliste de changement de variables), des contraintes linéaires

$$l \triangleq E [\varphi(X(t+1))]$$

sur la variable $X(t+1)$ lorsque $Y(1 \rightarrow t) \triangleq (Y(1), \dots, Y(t))$ est connue : l'équation d'état (1) fournit précisément le changement de variable et la contrainte l s'exprime comme intégrale double sur $X(t), W(t)$

$$l \triangleq E \left[\varphi \left(f(X(t), W(t)) \right) \right]$$

Les fonction φ peuvent être des indicatrices d'ensembles, des fonctions puissances, ... La loi de probabilité prédite pour $X(t+1)$ est alors la loi rendant maximum l'entropie sous les contraintes notées l .

Observons que l'évolution d'état (1) n'est pas conservée dans son intégralité. La méthode permet ainsi, d'approcher toutes les densités de probabilité, à l'aide de densités de probabilité de type exponentiel. L'étude directe de l'équation d'état en l'absence d'observation, donne des propriétés de la loi asymptotique de probabilité : elle est concentrée autour des zéros stables de l'équation $x = f(x, 0)$; les fluctuations aléatoires dues aux bruits d'état, sont compensées par la stabilité autour de chacun des zéros. La densité de probabilité présente plusieurs maximums locaux, elle est multi-modale.

Les logarithmes des densités sont développés linéairement sur des fonctions φ , à choisir en conséquence. Il s'agit avec un nombre petit et fixé de coefficients, de décrire les phénomènes aléatoires essentiels; une attention particulière porte sur le nombre de maximums locaux des densités de probabilité.

Les coefficients des développements linéaires varient dans la famille de densités et sont recalculés pour chaque mise à jour et pour chaque prédiction. A chaque date t , l'équation d'observation est utilisée

$$Y(t) = X(t) + V(t) \quad (2)$$

avec V bruit blanc d'observation de loi de probabilité de type exponentiel.

La formule de Bayes appliquée à l'équation (2), lorsque l'état $X(t)$ et le bruit $V(t)$ sont de lois de type exponentiel, permet (cf. 11 et 12) le calcul de la loi de $X(t)$ conditionnellement à l'observation $Y(1 \rightarrow t)$.

La mise en œuvre sur l'exemple de l'article, de cette méthodologie, permet de rendre concrètes les étapes qui viennent d'être décrites.

Un autre exemple avait été présenté antérieurement en Congrès (de Brucq et al, GRETSI 1991). Il avait introduit les équations :

$$X(t+1) = \sin\left(\frac{\pi}{2} \cdot X(t)\right) + W(t)$$

$$Y(t) = X(t) + V(t)$$

3. Modèle non linéaire

Le modèle d'état (classique pour $A(t)$ constant) :

$$X(t+1) = X(t) - A(t) \cdot X(t) + b + W(t) \quad (3)$$

$$Y(t) = X(t) + V(t) \quad (4)$$

comprend en plus, la description de l'évolution aléatoire du rappel A . Nous supposons ($A(t); t \in \mathbf{N}$), processus markovien, homogène. Pour fixer les idées, supposons que $1 - A(t)$ prenne l'une des valeurs $1 - a_1 = 0.25$ et $1 - a_2 = 0.75$, de façon aléatoire. Introduisons les probabilités de transition :

$$\pi_{i,j} \hat{=} P(A(t+1) = a_j / A(t) = a_i) \quad i, j = 1, 2 \quad (5)$$

Numériquement prenons deux cas distincts :

$$\pi \hat{=} \begin{pmatrix} 0.95 & 0.05 \\ 0.05 & 0.95 \end{pmatrix} \quad \pi \hat{=} \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix} \quad (6)$$

Les deux valeurs possibles de A supposé constant (indépendant de t) conduisent à des lois de probabilités pour $X(t)$, limites différentes et par conséquent, la donnée de la loi de $X(t)$, implique des propriétés sur celle de $A(t)$; une corrélation est à attendre, à chaque instant, entre $X(t)$ et $A(t)$. La loi de $(A(t), X(t))$ conditionnellement à l'observation $y(1 \rightarrow t)$ dépend de la valeur numérique actuelle $Y(t) = y(t)$, en raison de l'équation (4) même si $A(t)$ n'y est pas explicite.

La méthodologie consiste à faire le choix d'une famille de lois de type exponentiel, pour approcher les lois, a priori, et, a posteriori, de l'état ($A(t), X(t)$) aux diverses dates. L'étape de mise à jour consiste à utiliser la formule de Bayes appliquée à l'équation d'observation (4). L'étape de prédiction consiste à partir des lois à la date t et des formules (3) et (5) d'évolution, à prédire les lois à la date $t + 1$.

4. Lois exponentielles à cinq paramètres

Les lois statistiques de type exponentiel, ont pour densité :

$$f(x) \hat{=} \exp\left(\sum_i \lambda_i \cdot \varphi_i(x)\right) \quad \lambda_i \in \mathbf{R} \quad (7)$$

avec des fonctions φ_i quelconques. Elles sont à choisir en fonction du problème traité. Les lois gaussiennes sont de type exponentiel, en effet : $\forall x \in \mathbf{R}$,

$$f(x) \hat{=} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2} \left(\frac{x-m}{\sigma}\right)^2\right) \quad (8)$$

ou

$$f(x) = \exp(\lambda_0 + \lambda_1 x + \lambda_2 x^2) \quad \text{avec} \quad \int f(x) dx = 1$$

Les paramètres λ_1 et λ_2 s'obtiennent bijectivement à partir de la moyenne m et de la variance σ^2 en écrivant l'égalité des coefficients du trinôme en x [cf. Annexe II]; le paramètre λ_0 permet la normalisation à 1; éventuellement, la normalisation sera assurée par un coefficient α multiplicatif extérieur.

La loi de probabilité de la variable aléatoire $A(t)$ est un mélange convexe de deux Dirac, qui s'écrit : $\forall a \in \mathbf{R}$,

$$P[A(t) = a_1] \cdot \delta_{a_1}(a) + P[A(t) = a_2] \cdot \delta_{a_2}(a) \\ \hat{=} p^1 \cdot \delta_{a_1}(a) + p^2 \cdot \delta_{a_2}(a) \quad (9)$$

$$p^1 + p^2 \hat{=} 1, \quad p^1 \text{ et } p^2 \in [0, 1]$$

Introduisons maintenant la loi approchée de probabilité de l'état $X(t)$. Pour $A(t) = a_j$ ($j = 1, 2$), prenons pour densité de probabilité de la variable aléatoire $X(t) : \forall x \in \mathbf{R}$

$$f^j(x) \hat{=} \exp(\lambda_0^j + \lambda_1^j x + \lambda_2^j x^2) \text{ avec } \int f^j(x) dx = 1$$

Finalement, la loi approchée retenue pour le couple $(A(t), X(t))$ est un mélange de produits de Dirac et de gaussiennes, elle est caractérisée par l'expression : $\forall a, x \in \mathbf{R}$,

$$f(a, x) = p^1 \cdot \delta_{a_1}(a) \cdot f^1(x) + p^2 \cdot \delta_{a_2}(a) \cdot f^2(x) \quad (10)$$

Pour des commodités d'écriture, nous utilisons la notation $f(a, x)$ alors que la loi de probabilité n'a pas de densité par rapport à la mesure de Lebesgue sur \mathbf{R}^2 , en raison des deux Dirac. Nous sommes en présence d'une loi de type exponentiel; cependant elle est dégénérée : il suffit de remplacer chacun des Dirac par des indicatrices d'intervalles petits autour de a_1 et a_2 pour obtenir l'écriture canonique (7) à l'aide d'une seule exponentielle.

Rappelons que la loi marginale en x , est un mélange de deux gaussiennes; elle présente deux maximums locaux; elle est bimodale.

5. Filtrage approché

La méthodologie décrite dans le paragraphe I, commence en approchant à chaque instant t , la loi du couple $(A(t), X(t))$ conditionnelle à $Y(1 \rightarrow t-1)$ ou à $Y(1 \rightarrow t)$ par une telle expression (10).

Mise à jour :

Nous calculons la probabilité a posteriori, par la formule de Bayes.

En raison de l'équation d'observation (4), la loi du couple $(A(t), X(t))$ s'affine au vu de la nouvelle valeur numérique $Y(t) = \eta$. Ainsi, avec des notations évidentes, nous obtenons, la loi de probabilité du couple $(A(t), X(t))$ à la date t , conditionnellement à l'observation $Y(1 \rightarrow t) = (y(1 \rightarrow t-1), \eta)$ par l'expression : (cf. Annexe I).

$$f_{t/t}(a, x/y) = C \cdot f_{t/t-1}(a, x/y) \cdot f_v(\eta - x) \quad (11)$$

avec

$$C^{-1} \hat{=} \int \int f_{t/t-1}(\alpha, \xi/y) \cdot f_v(\eta - \xi) \cdot d\alpha d\xi \quad (12)$$

Comme l'équation d'observation est linéaire, le conditionnement maintient le caractère de type exponentiel, de la loi a posteriori.

Pour $A(t) = a_1$ comme pour $A(t) = a_2$, les lois du bruit d'observation $V(t)$ et de l'état $X(t)$, lorsque l'observation $y(1 \rightarrow t-1)$ est donnée, sont par hypothèse, des lois gaussiennes donc de type

exponentiel, aussi les lois de densité f^1 et f^2 de $X(t)$ conditionnellement à l'observation $y(1 \rightarrow t)$ sont encore de type exponentiel : la formule (11) indique qu'une simple addition des exposants, fournit la loi conditionnelle.

$$f_{t/t}(a, x/y) = C \cdot \left[p^1 \cdot \delta_{a_1}(a) \cdot e^{\lambda_0^1 + \lambda_1^1 x + \lambda_2^1 x^2} + p^2 \cdot \delta_{a_2}(a) \cdot e^{\lambda_0^2 + \lambda_1^2 x + \lambda_2^2 x^2} \right] \cdot e^{\mu_0 + \mu_1(\eta - x) + \mu_2(\eta - x)^2}$$

où les paramètres μ_0, μ_1, μ_2 de la densité du bruit d'observation V , s'obtiennent à partir de sa moyenne $m = 0$ et de sa variance R (cf. Annexe II).

Ainsi, en additionnant les coefficients des fonctions 1, x et x^2 , nous obtenons, d'une part, pour $A(t) = a_1$, les coefficients λ_i^1 , ($i = 0, 1, 2$) de la densité $f^1(x/y)$ de $X(t)$ et d'autre part, pour $A(t) = a_2$, les coefficients λ_i^2 , ($i = 0, 1, 2$) de la densité $f^2(x/y)$ de $X(t)$, les deux conditionnellement à $y(1 \rightarrow t)$. D'où l'expression de la loi de probabilité du couple $(A(t), X(t))$, conditionnellement à l'observation $y(1 \rightarrow t)$:

$$f_{t/t}(a, x/y) = C \cdot \left[p^1 \cdot \alpha^1 \cdot \delta_{a_1}(a) \cdot \frac{1}{\alpha^1} e^{\lambda_0^1 + \lambda_1^1 x + \lambda_2^1 x^2} + p^2 \cdot \alpha^2 \cdot \delta_{a_2}(a) \cdot \frac{1}{\alpha^2} e^{\lambda_0^2 + \lambda_1^2 x + \lambda_2^2 x^2} \right] \quad (13)$$

Comme il a été dit précédemment, la normalisation à 1, peut être obtenue par un coefficient multiplicatif et pour $j = 1, 2$.

$$\frac{1}{\alpha^j} e^{\lambda_0^j + \lambda_1^j x + \lambda_2^j x^2}$$

est, par hypothèse, d'intégrale en x , égale à 1 d'où la valeur numérique de α^j . Nous obtenons de (13), par intégration en x , la loi actualisée de $A(t)$. La formule fait intervenir les probabilités a priori $p_{t/t-1}^j$ et a posteriori $p_{t/t}^j$:

$$p_{t/t}^1 \hat{=} P[A(t) = a_1/y(1 \rightarrow t)] = \frac{p_{t/t-1}^1 \cdot \alpha^1}{p_{t/t-1}^1 \cdot \alpha^1 + p_{t/t-1}^2 \cdot \alpha^2} \quad (14)$$

$$p_{t/t}^2 \hat{=} P[A(t) = a_2/y(1 \rightarrow t)] = \frac{p_{t/t-1}^2 \cdot \alpha^2}{p_{t/t-1}^1 \cdot \alpha^1 + p_{t/t-1}^2 \cdot \alpha^2}$$

La loi du couple $(A(t), X(t))$ à la date t , conditionnellement à l'observation $y(1 \rightarrow t)$ connue, est alors caractérisée par l'expression (10) avec les probabilités $p^j = p_{t/t}^j$ et les paramètres λ_i^1, λ_i^2 , ($i = 0, 1, 2$) actualisés.

Prédiction :

Introduisons des contraintes linéaires sur la variable $X(t+1)$, calculables par le théorème de transfert des probabilités, à partir des lois de probabilités de l'état et des bruits, à la date t . Ensuite, la loi approchée de $X(t+1)$ est l'unique loi qui respecte les contraintes linéaires et qui rend maximum l'entropie; la loi déterminée ainsi, est une loi de type exponentiel.

Nous avons à expliciter la densité $f_{t+1/t}(x/y)$ de la loi de $X(t+1)$ conditionnellement à l'observation $y(1 \rightarrow t)$ connue.

Commençons par la partie singulière, provenant de A . Une évolution de A a lieu entre les dates t et $t+1$ suivant l'équation (5) et par suite, les probabilités $p_{t+1/t}^1$ et $p_{t+1/t}^2$ de la loi de $A(t+1)$, valent :

$$p_{t+1/t}^1 = p_{t/t}^1 \cdot \pi_{1,1} + p_{t/t}^2 \cdot \pi_{2,1} \quad (15)$$

$$p_{t+1/t}^2 = p_{t/t}^1 \cdot \pi_{1,2} + p_{t/t}^2 \cdot \pi_{2,2}$$

Ces expressions s'entendent conditionnellement à $y(1 \rightarrow t)$ ce qui est exprimé par le symbole $/t$.

Calculons les contraintes linéaires l_0, l_1, l_2 associées aux fonctions $1, x, x^2$. Nous utilisons l'équation (3) pour effectuer le transfert des probabilités de la date $t+1$ à la date t :

$$\begin{aligned} l_0 &\hat{=} \iint f_{t+1/t}(a, x/y) dadx & (16) \\ &= \iiint f_{t/t}(a, x) f_w(w) dadxdw \end{aligned}$$

$$\begin{aligned} l_1 &\hat{=} \iint x f_{t+1/t}(a, x/y) dadx & (17) \\ &= \iiint (x - ax + b + w) f_{t/t}(a, x/y) f_w(w) dadxdw \end{aligned}$$

$$\begin{aligned} l_2 &\hat{=} \iint x^2 f_{t+1/t}(a, x/y) dadx & (18) \\ &= \iiint (x - ax + b + w)^2 f_{t/t}(a, x/y) f_w(w) dadxdw \end{aligned}$$

Les intégrations en a , se réduisent, dans cet exemple, à la simple considération des deux cas $A(t) = a_1$ et $A(t) = a_2$.

Pour $A(t) = a_1$, nous obtenons, après calcul des trois intégrales les contraintes $l_i^1, (i = 0, 1, 2)$ de la loi a posteriori de $X(t+1)$ sous l'hypothèse supplémentaire $A(t) = a_1$. Ces contraintes déterminent, au moyen des formules d'inversion (cf. Annexe II), les paramètres de Lagrange $\lambda_i^1, (i = 0, 1, 2)$ de la densité approchée de probabilité $f_{t+1/t}^1$ qui réalise le maximum d'entropie.

Nous considérons ensuite le cas $A(t) = a_2$.

A la date $t+1$, l'approximation de la loi du couple (A, X) conditionnellement à l'observation $y(1 \rightarrow t)$, s'écrit comme prévu :

$$\begin{aligned} f_{t+1/t}(a, x/y) &= p_{t+1/t}^1 \cdot \delta_{a_1}(a) \cdot e^{\lambda_0^1 + \lambda_1^1 x + \lambda_2^1 x^2} \\ &\quad + p_{t+1/t}^2 \cdot \delta_{a_2}(a) \cdot e^{\lambda_0^2 + \lambda_1^2 x + \lambda_2^2 x^2} \end{aligned}$$

où les probabilités $p_{t+1/t}^1$ et $p_{t+1/t}^2$ et les paramètres λ_i^1, λ_i^2 ($i = 0, 1, 2$) viennent d'être prédits.

6. Simulation

Des résultats ont été et peuvent être obtenus en simulation. Une étude phénoménologique s'impose pour éviter de se perdre dans de trop nombreux résultats numériques.

Sauf raison explicite, il est naturel de prendre pour loi initiale de A , la loi stationnaire provenant de l'équation d'état (6). De façon plus précise, la transition π définit une probabilité (p^1, p^2) invariante :

$$p^1 = \frac{\pi_{2,1}}{\pi_{2,1} + \pi_{1,2}} \quad p^2 = \frac{\pi_{1,2}}{\pi_{2,1} + \pi_{1,2}}$$

Pour chacune des deux valeurs, a_1, a_2 possibles de A , l'équation (3) conduit pour l'état $X(t)$, à une loi gaussienne asymptotiquement de moyenne :

$$m_1 = \frac{b}{a_1} = 1,33 \quad m_2 = \frac{b}{a_2} = 4$$

ainsi nous partons de $x(0) \hat{=} p^1 m_1 + p^2 m_2$. Donnons également les variances asymptotiques :

$$\sigma_j^2 = Q \frac{1}{1 - (1 - a_j)^2} \text{ soit } \sigma_1^2 = 0.0106 \quad \sigma_2^2 = 0.0228$$

L'observation est bruitée et fait intervenir à chaque instant t , le bruit $V(t)$. La variance R (du bruit $V(t)$ d'observation), en croissant, diminue l'impact de la connaissance de l'observation $Y(t)$ sur la connaissance de l'état $X(t)$, aléatoire. Cependant, seule une variance R nulle, conduit à une modification mathématique de la description des phénomènes. Les simulations confirment cette analyse.

Nous prenons $R = 1$.

Le système dynamique caractérisé par le coefficient a_j doit être stable, $|1 - a_j| < 1$. La loi asymptotique pour a_j constant, s'obtient d'autant plus vite que $|1 - a_j|$ est plus petit. En prenant, $1 - a_1 = 1/4$ et $1 - a_2 = 3/4$, l'effet de la condition initiale $x(0)$ s'atténue rapidement tout en restant observable durant quelques itérations.

En plus, un bruit $W(t)$ d'état s'ajoute à chaque instant t à l'évolution de $X(t)$ qui devient aléatoire. Le passage de t à $t+1$ augmente l'entropie associée à la loi de l'état $X(t)$. La forme de la loi de l'état $X(t+1)$ est également modifiée. Plus la variance Q augmente, plus la forme de la loi se rapproche de la loi du bruit $W(t)$. A la limite, l'état $X(t)$ est constamment égal au bruit $W(t)$ d'état. Les phénomènes deviennent stationnaires dans le temps; il n'y a pas transfert d'information par le modèle d'un instant t au suivant $t+1$; à chaque date t , l'observation $Y(t) = \eta$ conduit à la même réduction d'entropie. La simulation confirme cette analyse phénoménologique.

Nous prenons $Q = 0.01$

Nous présentons maintenant des résultats numériques qui illustrent le filtrage par densités approchées.

Changements rares de dynamiques :

Prenons, tout d'abord, pour matrice de transition :

$$\pi \hat{=} \begin{pmatrix} 0.95 & 0.05 \\ 0.05 & 0.95 \end{pmatrix}$$

Dans ce cas, les changements aléatoires de dynamique se produisent de façon exceptionnelle. Nous observons de longues périodes de stationnarité. La loi de X se stabilise sur l'une des lois asymptotiques en dehors des durées assez courtes suivant immédiatement les changements de dynamique.

Nous pouvons parler d'évolution linéaire bruitée avec rupture de dynamique. La détection des ruptures utilise la prédiction de l'observation. Les valeurs passées de A , constantes sur des durées longues, peuvent être identifiées. Si la prédiction est très différente de la valeur observée, un changement de modèle est décidé.

Dans les tableaux, nous fournissons :

$$\hat{X}(t+1/t) \hat{=} p_{t+1/t}^1 m_1 + p_{t+1/t}^2 m_2$$

LOIS APPROCHÉES PRÉDITES

t	$A(t)$	$X(t)$	p^1	m_1	σ_1^2	p^2	m_2	σ_2^2	$\hat{X}(t+1/t)$
0	3/4	2.66	0.71	1.33	0.0106	0.29	3.97	0.0225	2.09
1	3/4	1.67	0.95	1.33	0.0106	0.05	3.93	0.0224	1.46
2	3/4	1.28	0.95	1.33	0.0106	0.05	3.89	0.0223	1.458
3	3/4	1.37	0.95	1.33	0.0106	0.05	3.85	0.0223	1.456
4	1/4	1.39	0.95	1.33	0.0106	0.05	3.85	0.0222	1.456
5	1/4	2.20	0.95	1.33	0.0106	0.05	3.85	0.0222	3.73
6	1/4	2.69	0.07	1.34	0.0106	0.93	3.91	0.0222	3.713
7	1/4	3.04	0.08	1.34	0.0106	0.92	3.92	0.0222	3.774
8	1/4	3.31	0.06	1.34	0.0106	0.94	3.93	0.0222	3.838
9	1/4	3.62	0.05	1.34	0.0106	0.95	3.97	0.0222	3.876
10	1/4	3.76	0.05	1.34	0.0106	0.95	4.00	0.0222	3.867
11	1/4	3.95	0.05	1.34	0.0106	0.95	3.99	0.0222	3.857
12	1/4	3.85	0.05	1.34	0.0106	0.95	4.00	0.0222	3.867
55	1/4	3.76	0.09	1.34	0.0106	0.91	3.97	0.0222	3.733
56	3/4	3.83	0.08	1.34	0.0106	0.92	3.97	0.0222	3.759
57	3/4	1.94	0.31	1.33	0.0106	0.69	3.94	0.0222	3.13
58	3/4	1.57	0.78	1.33	0.0106	0.22	3.92	0.0222	1.999
59	3/4	1.44	0.95	1.33	0.0106	0.05	3.90	0.0222	1.458

Sur ce tableau, observons que les valeurs asymptotiques des moyennes et des variances se conservent pour chacune des deux composantes gaussiennes et cependant, le filtrage par densités approchées en trois itérations, explicite le changement de dynamique. Les valeurs $p_{t+1/t}^1$ et $p_{t+1/t}^2$ qui fournissent le mélange, sont très sensibles aux observations effectuées.

Voyons si pour d'autres matrices de transition, ces constatations se maintiennent.

Changements fréquents de dynamique :

Prenons maintenant une transition π qui traite à égalité les deux dynamiques :

$$\pi \hat{=} \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}$$

La connaissance sur l'évolution devient très partielle. Le filtrage par densités (approchées) permet la prédiction des deux lois de probabilité correspondant aux deux cas $a_1 = 3/4$ et $a_2 = 1/4$. En raison des symétries, les prédictions doivent maintenir ces deux cas comme équiprobables.

De plus, les deux lois approchées sont mises à jour régulièrement chaque fois que $y(t)$ est donnée ce qui permet l'estimation de l'état $X(t)$ par maximum de vraisemblance. De même, les changements de modèles proviennent lorsque $y(t)$ est donnée, du suivi des deux probabilités $p_{t/t}^1$ et $p_{t/t}^2$.

LOIS APPROCHÉES PRÉDITES

t	$A(t)$	$X(t)$	p^1	m_1	σ_1^2	p^2	m_2	σ_2^2	$\hat{X}(t+1/t)$
0	3/4	2.66	0.50	1.33	0.0106	0.50	3.97	0.0222	2.65
1	1/4	1.67	0.50	1.33	0.0106	0.50	3.93	0.0222	2.63
2	3/4	2.12	0.50	1.33	0.0106	0.50	3.90	0.0222	2.615
3	1/4	1.58	0.50	1.33	0.0106	0.50	3.87	0.0222	2.6
4	3/4	2.23	0.50	1.33	0.0106	0.50	3.87	0.0222	2.6
5	1/4	1.72	0.50	1.33	0.0106	0.50	3.86	0.0222	2.59
6	1/4	2.33	0.50	1.34	0.0106	0.50	3.91	0.0222	2.625
7	1/4	2.76	0.50	1.34	0.0106	0.50	3.91	0.0222	2.625
8	3/4	3.10	0.50	1.34	0.0106	0.50	3.92	0.0222	2.63
9	3/4	1.91	0.50	1.34	0.0106	0.50	3.93	0.0222	2.635
10	1/4	1.52	0.50	1.34	0.0106	0.50	3.94	0.0222	2.64
15	1/4	1.91	0.50	1.34	0.0106	0.50	3.94	0.0222	2.64
25	3/4	2.44	0.50	1.33	0.0106	0.50	3.87	0.0222	2.6
35	1/4	1.43	0.50	1.33	0.0106	0.50	3.87	0.0222	2.6
45	3/4	1.64	0.50	1.34	0.0106	0.50	3.91	0.0222	2.625
55	1/4	1.73	0.50	1.33	0.0106	0.50	3.90	0.0222	2.615
56	3/4	2.31	0.50	1.33	0.0106	0.50	3.89	0.0222	2.61
57	1/4	1.56	0.50	1.33	0.0106	0.50	3.88	0.0222	2.605
58	1/4	2.26	0.50	1.33	0.0106	0.50	3.89	0.0222	2.61
59	1/4	2.74	0.50	1.33	0.0106	0.50	3.89	0.0222	2.61

Observons la constance des probabilités prédites

7. Conclusion

L'article porte essentiellement sur le suivi probabiliste de phénomènes aléatoires pouvant comporter des ruptures. A une date $t+1$ non-prévisible, l'état $X(t+1)$ du système suit une loi statistique nettement différente de la loi de l'état $X(t)$ à la date t . De façon plus précise, l'évolution aléatoire des phénomènes est suivie à l'aide de l'évolution déterministe de la loi de probabilité de l'état du système; l'article porte essentiellement sur des discontinuités de l'évolution de la loi de probabilité. Les phénomènes restent markoviens mais comportent une évolution aléatoire de changement de modèles.

Les traitements antérieurs à cet article, permettent de détecter des changements par rapport au modèle prévisible provenant du passé. En introduisant plusieurs modèles concurrents, ils permettent également de tester, en parallèle, à la date $t+1$ lequel des modèles, est le plus probable.

Depuis plus de vingt ans, on sait l'impossibilité d'un calcul exact de la densité de probabilité, a posteriori, pour des équations d'état non-linéaires.

L'un des auteurs introduit dans le filtrage, l'approximation des densités de probabilité. Ainsi la méthodologie des filtres finis s'applique. Les lois a priori et a posteriori, sont paramétrées. La prédiction est effectuée par calcul de contraintes linéaires en utilisant les équations non-linéaires d'état. Ensuite, une densité de type exponentiel provient par maximisation de l'entropie sous contraintes. L'équation d'observation permet la mise à jour de la loi de probabilité.

Le suivi récursif de plusieurs (cinq dans l'article) caractéristiques de distributions statistiques conduit à une meilleure approximation que le suivi récursif des deux caractéristiques usuelles : moyenne et variance. En particulier, il est possible d'effectuer le suivi récursif du mélange aléatoire de deux lois gaussiennes. Les résultats en simulation sont concluants. La programmation en langage évolué, est aisée.

La méthodologie s'étend sans difficulté au cas de mélange d'un nombre quelconque de lois gaussiennes. De même, la dynamique aléatoire $A(t)$ du système linéaire de l'exemple peut suivre d'autres lois probabilistes; nous avons programmé le cas où la variable A appartient, de façon aléatoire, à plusieurs intervalles petits autour de a_1 et a_2 .

Que faut-il pour pouvoir appliquer la méthode? La modélisation doit permettre le transfert de la loi inconnue de probabilité de l'état $X(t+1)$, à la date $t+1$, vers celle, connue au mieux, de l'état $x(t)$ et du bruit $W(t)$, à la date t . Ainsi, la modélisation d'état peut comprendre à la fois des équations non-linéaires bruitées ainsi que des évolutions probabilistes sur des modèles.

Le choix des contraintes linéaires est important; la précision des approximations dépend de ce choix. La connaissance d'un moment de l'état $X(t+1)$ est une contrainte linéaire. De même, la connaissance de la probabilité pour $x(t+1)$ d'être dans un ensemble donné, est une contrainte linéaire.

Dans d'éventuelles généralisations ne faisant pas intervenir les lois gaussiennes, la méthodologie nécessite le calcul des contraintes linéaires l à l'aide d'intégrales multiples; de plus, le passage des contraintes l aux coefficients λ (caractérisant la loi exponentielle obtenue par maximum d'entropie à partir des contraintes l), nécessite la résolution d'un système d'équations non-linéaires.

Ainsi, au prix d'éventuels calculs menés à bien par ordinateur, le filtrage par densités approchées, s'applique à des situations plus nombreuses que le filtrage de Kalman étendu. Rappelons en particulier que la méthodologie est intéressante pour des modèles linéaires dès que les bruits ne sont pas gaussiens. Les bruits d'état W peuvent être quelconques et les bruits d'observation V peuvent être décrits par des lois de type exponentiel, constituant une famille bien plus vaste que celle des lois gaussiennes.

Annexe I

En raison de l'équation (4), la loi du triplet $(A(t), X(t), Y(t))$ conditionnellement à l'observation $Y(1 \rightarrow t-1)$ provient du

changement de variable :

$$\begin{cases} a = a \\ x = x \\ y = x + v \end{cases}$$

sur la loi du triplet $(A(t), X(t), V(t))$; rappelons que $(A(t), X(t))$ est indépendant de $V(t)$ sous $Y(1 \rightarrow t-1)$:

$$f_{A,X,V}(a, x, v) = f_{A,X}(a, x) \cdot f_V(v)$$

d'où

$$f_{A,X,Y}(a, x, y) = f_{A,X}(a, x) \cdot f_V(y - x)$$

et la loi du couple $(A(t), X(t))$ conditionnelle à $Y(t) = \eta$ et à $Y(1 \rightarrow t-1)$ est définie par :

$$f_{A,X/Y}(a, x/\eta) = C \cdot f_{A,X}(a, x) \cdot f_V(\eta - x)$$

nous introduisons C , le coefficient de normalisation vérifiant :

$$C^{-1} \hat{=} \iint f_{A,X}(\alpha, \xi) \cdot f_V(\eta - \xi) d\alpha d\xi$$

Annexe II

Rappelons qu'une loi gaussienne a pour densité : $\forall x \in \mathbb{R}$,

$$\begin{aligned} f(x) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right) \\ &= \exp\left(-\frac{1}{2}\ln(2\pi\sigma^2) - \frac{m^2}{2\sigma^2} + \frac{m}{\sigma^2} \cdot x - \frac{1}{2\sigma^2} \cdot x^2\right) \end{aligned}$$

ou

$$f(x) = \exp(\lambda_0 + \lambda_1 x + \lambda_2 x^2)$$

Ainsi, une loi gaussienne est de type exponentiel. Le logarithme de la densité se développe linéairement sur une base de trois fonctions $\varphi_0(x) \hat{=} 1$, $\varphi_1(x) \hat{=} x$, $\varphi_2(x) \hat{=} x^2$.

Dans le cas particulier d'approximation gaussienne, les contraintes l , moyennes des fonctions φ_i , ($i = 0, 1, 2$) sont connues. Ce sont précisément les moments d'ordre 0, 1 et 2 de la loi gaussienne et le calcul numérique d'intégrales du type $E[\varphi_i(X)]$ où X est une variable aléatoire, est évité.

De plus, le calcul des coefficients λ de la densité, à partir des moments l de la loi gaussienne, est simple. Il suffit de normaliser les contraintes l pour atteindre les caractéristiques statistiques de la loi et d'en déduire les coefficients λ . En effet, nous avons, dans le cas gaussien :

$$\begin{aligned} l_0 &\hat{=} \int_{\mathbb{R}} e^{\lambda_0 + \lambda_1 x + \lambda_2 x^2} dx = \sqrt{\frac{2\pi}{-2\lambda_2}} e^{\lambda_0 - \frac{\lambda_1^2}{4\lambda_2}}, \quad \lambda_2 < 0 \\ l_1 &\hat{=} \int_{\mathbb{R}} x \cdot e^{\lambda_0 + \lambda_1 x + \lambda_2 x^2} dx = -\frac{\lambda_1}{2\lambda_2} \cdot l_0 \\ l_2 &\hat{=} \int_{\mathbb{R}} x^2 \cdot e^{\lambda_0 + \lambda_1 x + \lambda_2 x^2} dx = \left(\frac{\lambda_1^2}{4\lambda_2^2} - \frac{1}{2\lambda_2}\right) \cdot l_0 \end{aligned}$$

La normalisation conduit aux caractéristiques de la loi gaussienne :

$$\begin{aligned} \text{moyenne : } m &= \frac{l_1}{l_0} \\ \text{variance : } \sigma^2 &= \frac{l_2}{l_0} - m^2 \end{aligned}$$

Et les coefficients lambdas de la densité gaussienne sont définis à partir des moments par :

$$\begin{aligned}\lambda_2 &= -\frac{1}{2\sigma^2} \\ \lambda_1 &= \frac{m}{\sigma^2} \\ \lambda_0 &= -\frac{1}{2} \ln(2\pi\sigma^2) + \ln(\alpha_0) - \frac{m^2}{2\sigma^2}\end{aligned}$$

Annexe III

Proposition :

Soient k fonctions réelles, $\varphi_0 \hat{=} 1, \varphi_1, \dots, \varphi_k$ alors le maximum d'entropie :

$$H \hat{=} \int -\ln(f(x)) f(x) dx,$$

soumis aux contraintes : $\forall i = 1, \dots, k,$

$$\int \varphi_i(x) f(x) dx = l_i$$

conduit (sous des hypothèses très générales d'intégration) à la loi exponentielle :

$$f(x) = \exp\left(\lambda_0 + \sum_{i=1}^k \lambda_i \varphi_i(x)\right)$$

où les réels $\lambda_0, \lambda_1, \dots, \lambda_k$ sont les paramètres de Lagrange, calculés à partir des contraintes.

BIBLIOGRAPHIE

- [1] I. BARRET & P. VACHER, Poursuite de cibles manœuvrantes par algorithmes markoviens hybrides, *XII Colloque GRETSI*, pp. 189-192, juin 1989.
- [2] M. BASSEVILLE, Detecting changes in signals and system-a survey, *Automatica*, Vol. 24, pp. 309-326, 1988.
- [3] V. BENES, Exact finite dimensional filters for certain diffusions with non-linear drift, *Stochastics* 5, pp. 65-92, 1981.
- [4] P. BRUNET, On the Convergence in Parameter Estimation via Suboptimal Filtering, *Proc. of the IFAC'78 World Congress Helsinki*, 1978.
- [5] A.E. BRYSON & M. FRAZIER, Smoothing for linear and non linear systems, *Proc. Patterson AFB*, Ohio TRD-ASD-TRD-63-119, pp. 354-364, Feb. 1963.
- [6] E. BUSVELLE, Immersion des systèmes stochastiques et filtres finis; filtrage particulière. Thèse Université de Rouen, 1991.
- [7] I. CSISZAR, Sanov Property, Generalized L-Projection and a conditional limit theorem, *The annals of Probability*, Vol. 12, n° 3, pp. 768-796, 1984.
- [8] D. de BRUCQ, E. BUSVELLE, P. COURTELLEMONT, V. RUIZ, Évolution d'état non-linéaire et filtrage approché par maximum d'entropie, *Treizième Colloque GRETSI*, pp. 37-40, 1991.
- [9] D. de BRUCQ & G. FOLLIOU, Théorie du signal, Masson, 1988.
- [10] D. de BRUCQ, Extension du Filtre de Kalman, Approximation quadratique, Séminaire de Mathématique, Rouen, Université de Rouen, Compte Rendu de séances 1990-1991.

- [11] D. de BRUCQ & D. RAKOTOPARA, Équations du filtrage non-linéaire pour une loi d'état conditionnellement gaussienne, *Colloque AFCET*, Toulouse, pp. 225-231, 1986.
- [12] Ph. de LARMINAT, Sur l'Identification par Filtrage non Linéaire, Thèse Nantes, 1971.
- [13] J.R. FISHER & E.B. STEAR, Near optimal non linear filtering using quasi-moment function, *Int. Jnal. of Control*, Vol. 12, n° 4, pp. 685-697, 1970.
- [14] M. FUJISAKI, G. KALLIANPUR, H. KUNITA, Stochastic differential equations for the non-linear problem, *Osaka Jnal Math.* 9, pp. 19-40, 1972.
- [15] F. GAMBOA, Méthode du maximum d'entropie sur la moyenne et applications, Doctorat en Sciences, Mathématiques, Paris Sud Centre d'Orsay, 1990.
- [16] U.G. HAUSSMANN & E. PARDOUX, A conditionally almost linear filtering problem with non-Gaussian initial condition, *Stochastics* 23, pp. 241-275, 1988.
- [17] A.H. JAZWINSKI, Stochastic processes & filtering theory, *Academic Press*, 1970.
- [18] G. JUMARIE, Non linear filtering, A weighted mean squares approach and a bayesian one via the maximum entropy principle, *Signal Processing* 21, pp. 323-328, 1990.
- [19] R.E. KALMAN & R.S. BUCY, New Results in Linear Filtering and Prediction Theory, Trans. of the ASME, *Jnal of Basic Eng.*, serie D, Vol. 83, pp. 95-108, March 1961.
- [20] M. LABARRERE, J.P. KRIEFF & B. GIMONET, Le filtrage et ses applications, CEPADUES-ÉDITIONS, 1982.
- [21] D.G. LAINIOTIS, Optimal adaptive estimation, *IEEE trans. AC* 16, n° 2, 1971.
- [22] A.M. MAKOWSKI, Results on filtering problem for linear systems with non-Gaussian initial conditions, *Proc. of the XXI IEEE conf. on decision and control*, Orlando, FL, pp. 101-104, 1982.
- [23] M. NAJIM, Modélisation et identification en traitement du signal, Masson, 1988.
- [24] M. ZAKAI, On the optimal filtering of diffusion processes, *Zeit Wahr. Verw. Geb. 11*, pp. 230-243, 1969.

LES AUTEURS

Denis de BRUCQ

Denis de Brucq est professeur à l'Université de Rouen. Ses travaux sur le filtrage non-linéaire permettent la détection de changements de modèles. Les techniques inventées s'appliquent en reconnaissance de l'écrit. Les diverses composantes des chiffres manuscrits sont identifiées. Par ailleurs, en traitement numérique de l'image pour la dermatologie, des méthodes permettant la détection des mélanomes sont mises en œuvre.

Virginie RUIZ

Virginie Ruiz Docteur en Traitement du Signal, est boursière du Ministère des Affaires Étrangères, programme Lavoisier. Durant son séjour de un an, à Londres, elle applique le filtrage non-linéaire au traitement numérique d'images médicales. Il s'agit de séparer les histogrammes provenant de divers tissus. Certains sont sains d'autres non.

Manuscrit reçu le 22 mai 1992