



## EVOLUTION D'ETAT NON-LINEAIRE

## ET FILTRAGE APPROCHE PAR MAXIMUM D'ENTROPIE

de BRUCQ D., BUSVELLE E., COURTELLEMONT P., RUIZ V.

Université de Rouen LACIS-ITEPEA BP 118  
76134 Mont-Saint-Aignan Cedex Tel 35 14 60 00

## RÉSUMÉ

La détermination de l'état d'un système après observation s'effectue usuellement par le filtre de Kalman. Pour une évolution d'état non linéaire, cette méthode ne peut pas être appliquée. L'observation linéaire est la somme d'un bruit de loi exponentielle, et d'un état de loi conditionnelle, indépendante, approchée exponentielle. L'équation d'état non-linéaire permet l'approximation numérique par maximum d'entropie sous contraintes linéaires de la loi de l'état prédit sous la forme exponentielle. Ainsi des distributions probabilistes multi-modales peuvent être introduites.

INTRODUCTION

Dans les systèmes réels, on retrouve souvent des signaux provenant de modèles internes dont l'évolution est aléatoire et non-linéaire.

Les méthodes de modélisation par modèles Auto-Régressifs sont très utilisées en traitement du signal et en automatique. Mais il s'avère, que ces méthodes sont insuffisantes pour le suivi de la dynamique des systèmes non-linéaires.

Les modélisations Auto-Régressives servent souvent pour la détection de ruptures anormales, c'est-à-dire de changements brusques dans l'évolution d'un signal [BASSEVILLE, 1988].

Les changements de modèles peuvent provenir de sauts, entre points stables de systèmes non-linéaires. Les chaos déterministes considèrent tout spécialement ces situations. Donnons un exemple pour illustrer cette idée (GRECO TDSI 1990). Soit le modèle d'état non-linéaire bruité:

$$(0.1) X(t+1) = \sin\left(\frac{\pi}{2} X(t)\right) + W(t)$$

$$(0.2) Y(t) = X(t) + V(t)$$

avec les notations usuelles. Soient P, la covariance du bruit d'état W et Q la covariance du bruit, d'observation. Le modèle est pris scalaire pour des raisons didactiques. La présence du sinus est la seule modification, par rapport au modèle d'état linéaire classique.

Rappelons que le modèle d'état linéaire avec des bruits gaussiens, permet le calcul des probabilités a posteriori, pilotées par la moyenne et la variance conditionnelles à l'observation. Le filtrage de Kalman est l'exemple type des filtres finis, où la loi de probabilité a posteriori dépend d'un nombre fini de paramètres, calculés de façon récursive au fur et à mesure de l'arrivée des observations.

L'étude et la mise en oeuvre des filtres de

## ABSTRACT

Kalman filtering is the usual technique for determination of a system state, after observation. For a non linear state evolution, this method can't be applied. The linear observation is the sum of an exponential law noise and a state with conditional, independent, exponential approximated law. The non linear state equation allows a numerical approximation by the maximum entropy, under linear constraints on the law of the predicted state with exponential form. So probabilistic, multimodal distributions can be introduced.

Kalman étendus linéaires, Kalman étendus quadratiques et les prédictions à partir du calcul effectif de l'espérance conditionnelle, visant à détecter les ruptures anormales provenant du système précédent, à évolution non linéaire, conduisent à des résultats honorables mais non convainquants (cf de BRUCQ et al GR TDSI).

Ce modèle d'état à évolution sinusoïdale, non bruité, comporte deux points fixes stables +1 et -1 et un point fixe instable 0. Les bruits font éventuellement basculer le système d'un point fixe à l'autre point fixe. Une méthode particulière (cf Busvelle et al) montre que les diverses lois a posteriori de l'état X, sont bimodales en raison du caractère stable des points fixes -1 et +1 et il est donc aberrant de se limiter, comme le fait le filtrage de Kalman, à des lois gaussiennes unimodales. La moyenne de cette loi tombe précisément dans une partie où la vraie loi a posteriori est de densité quasiment nulle.

La méthodologie introduit des densités a priori et a posteriori bimodales paramétrées, dont les coefficients sont calculables par ordinateur. La technique de calcul des intégrales du type  $E[f(X)]$ , où X est une variable aléatoire réelle, est fondée sur l'interpolation polynômiale. Nous ne reprenons pas ces calculs classiques, dans ce texte.

L'article est consacré essentiellement au traitement approché de l'exemple.

I MODELE A EVOLUTION D'ETAT NON LINEAIRE

Soit un modèle non linéaire, décrit par les équations bruitées:

$$\xi(t+1) = g(\xi(t)) + W(t)$$

$$Y(t) = h(\xi(t)) + V(t)$$

où g et h sont des fonctions non linéaires de l'état  $\xi$ , aléatoire. Soit le changement de variable:



$$X(t) \cong h(\xi(t))$$

avec la fonction  $h$  localement injective. Nous pouvons écrire à la date  $t+1$ :

$$(1.1) \quad X(t+1) = f(X(t), W(t)) \quad \text{avec}$$

$$f(X(t), W(t)) \cong h\left(g\left(h^{-1}(X(t))\right) + W(t)\right) \text{ et}$$

$$(1.2) \quad Y(t) = X(t) + V(t)$$

Rappelons que les lois exponentielles ont pour densité:

$$p(x) \cong \exp\left(\sum_i \lambda_i \varphi_i(x)\right) \quad \lambda_i \in \mathbb{R}$$

avec des fonctions  $\varphi_i$  quelconques. En raison de l'équation d'observation, la loi de  $X(t)$  s'affine au vu de la nouvelle valeur numérique  $Y(t)=y$ . Ainsi avec des notations évidentes, nous obtenons la densité de  $X(t)$  conditionnellement à l'observation,  $Y(t)=y$  et  $Y(0 \rightarrow t-1) \cong Y(0), \dots, Y(t-1)$  par l'expression:

$$(1.3) \quad f_{x/y}(x/y) = C f_x(x) f_v(y-x) \quad \text{avec}$$

$$(1.4) \quad C^{-1} \cong \int f_x(\xi) \cdot f_v(y-\xi) d\xi$$

Supposons par hypothèse, que les lois du bruit  $V(t)$  et de l'état  $X(t)$ , lorsque l'observation  $Y(0 \rightarrow t-1)$  est donnée, soient des lois exponentielles, alors la loi de  $X(t)$  conditionnellement à l'observation  $Y(0 \rightarrow t)$  est encore exponentielle.

Pour une variable aléatoire, initialement gaussienne, sa densité de probabilité  $p(x)$  est définie par:  $\forall x \in \mathbb{R}$

$$(1.5) \quad p(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right) \quad \text{ou}$$

$$p(x) = \exp\left(\lambda_0 + \lambda_1 x + \lambda_2 x^2\right)$$

Lemme:

$$\Psi(\lambda) \cong \int_{\mathbb{R}} e^{\lambda_0 + \lambda_1 x + \lambda_2 x^2} dx = \sqrt{\frac{2\pi}{-2\lambda_2}} \cdot e^{\lambda_0 - \frac{\lambda_1^2}{4\lambda_2}}$$

Le lien entre les moments (alphas) et les lambdas s'obtient aisément par dérivation de  $\Psi(\lambda)$ . Le passage des moments alphas, aux moments centrés est classique.

Les valeurs des paramètres lambdas, notés  $\lambda_0, \lambda_1, \lambda_2$  de la densité  $p_v(v)$  du bruit d'observation, s'obtiennent, à partir de la moyenne  $m=0$  et de la variance  $\sigma^2=Q$ , par les expressions précédentes. Dans (1.3), l'expression de l'exposant de l'exponentiel s'écrit:

$$(1.6) \quad \lambda_0 + \lambda_1 x + \lambda_2 x^2 + \lambda_0 + \lambda_1(y-x) + \lambda_2(y-x)^2 =$$

$$\lambda_0 + \lambda_0 + \lambda_1 y + \lambda_2 y^2 + (\lambda_1 - \lambda_1 - 2\lambda_2 y^2)x + (\lambda_2 + \lambda_2)x^2$$

Nous savons que notre modèle à points fixes  $-1$  et  $+1$ , fait apparaître des densités de probabilité bimodales. Ainsi, nous séparons l'espace  $\mathbb{R}$  d'étude en deux intervalles  $]-\infty, 0]$  et  $]0, +\infty[$ :  $\forall x \in \mathbb{R}$

$$(1.7) \quad \text{Log} p(x) = \left(\lambda_0^- + \lambda_1^- x + \lambda_2^- x^2\right) 1_{]-\infty, 0]} + \left(\lambda_0^+ + \lambda_1^+ x + \lambda_2^+ x^2\right) 1_{]0, +\infty[}$$

De plus, l'évolution (0.1) non linéaire de l'état, entraîne la perte du caractère gaussien approché, surtout au voisinage de  $+1$  et de  $-1$ . Nous introduisons dans l'expression de la densité de probabilité des paramètres  $\lambda_3^\pm$  et des fonctions  $\varphi_3^\pm(x)$  supplémentaires:

$$\varphi_3^+(x) \cong x(1-x) 1_{]0, 1]} \quad \varphi_3^-(x) \cong -x(1+x) 1_{] -1, 0]}$$

Ainsi toutes les densités de probabilité sont approchées par des lois exponentielles:

$$\exp\left[\left(\sum_i \lambda_i^- \cdot \varphi_i^-(x)\right) 1_{]-\infty, 0]} + \left(\sum_i \lambda_i^+ \cdot \varphi_i^+(x)\right) 1_{]0, +\infty[}\right]$$

Il reste évidemment à obtenir les coefficients lambdas pour chaque étape de prédiction.

Proposition:

Soient  $k$  fonctions réelles,  $\varphi_0 \cong 1, \varphi_1, \dots, \varphi_k$  alors le maximum d'entropie:

$$H \cong \int_{\mathbb{R}} -\ln f(x) f(x) dx$$

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

$$(1.8) \quad \int \varphi_i(x) f(x) d\mu(x) = \alpha_i$$

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

$$(1.9) \quad p(x) = \exp\left[\lambda_0 + \sum_{i=1}^k \lambda_i \cdot \varphi_i(x)\right]$$

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

Utilisons l'équation d'état (1.1), nous pouvons calculer numériquement les expressions:  $\forall i=1, \dots, k$

$$(1.10) \quad \alpha_j \cong E[\varphi_j(X(t+1))/Y(0 \rightarrow t)]$$

$$= \int \varphi_j(f(x, w)) P_{x(t)}(dx/Y(0 \rightarrow t)) P_{w(t)}(dw)$$

ou du moins, par la méthode d'approximation des intégrales, nous pouvons atteindre une valeur approchée, disons  $\alpha_j$ , de cette intégrale.

Ainsi au rang  $t+1$ , la connaissance a priori se limite à  $k$  contraintes linéaires sur la loi de  $X(t+1)$  conditionnellement à l'observation  $Y(0 \rightarrow t)$ . L'état prédit suit une loi exponentielle approchée, celle rendant maximum l'entropie sous contraintes

linéaires.

Nous résolvons ensuite, le système en k+1 variables  $\lambda$  à k+1 équations non-linéaires:

$$(I.11) \quad \alpha_j = \int \varphi_j \exp \left[ \lambda_0 + \sum_{i=1}^k \lambda_i \cdot \varphi_i(x) \right] dx$$

**II DETECTION DE CHANGEMENT DE MODELE ENTRE LOIS GAUSSIENNES**

Nous introduisons à chaque date  $t \in \mathbb{N}$ , deux lois de probabilité gaussienne:  $\forall x \in \mathbb{R}$

$$(II.1) \quad \text{Log } p^-(x) = \left[ \lambda_0^- + \lambda_1^- x + \lambda_2^- x^2 \right]$$

$$(II.2) \quad \text{Log } p^+(x) = \left[ \lambda_0^+ + \lambda_1^+ x + \lambda_2^+ x^2 \right]$$

Par hypothèse de récurrence, la loi de  $X(t)$  conditionnellement à  $Y(0 \rightarrow t-1)$  est approchée par l'une des deux lois

$$p^-(x/t-1) \text{ ou } p^+(x/t-1).$$

La détection consiste à calculer, les densités de  $Y(t)$  gaussiennes sous chacune des deux hypothèses  $\epsilon=+1, \epsilon=-1$ . D'où la procédure de détection, lorsque  $Y(t)=y$  est observée:

Etat  $X(t)$  sur  $]-\infty, 0]$  si  $f_y^-(y) > f_y^+(y)$ .

Etat  $X(t)$  sur  $]0, +\infty[$  si  $f_y^+(y) > f_y^-(y)$ .

Ensuite, il s'agit pour  $i=0,1,2$ , de prédire les moments  $\alpha_i^\epsilon$  de la loi de  $X(t+1)$  conditionnellement à  $Y(0 \rightarrow t)$ , sous l'hypothèse retenue:

$$(II.3) \quad \alpha_i^\epsilon \hat{=} \int \int \varphi_i^\epsilon \left[ \sin \left( \frac{\pi}{2} \cdot x \right) + w \right] \cdot p^\epsilon(x/t) dx \cdot p_w(w) dw$$

Une expression analytique existe dans le cas gaussien:

$$(II.4) \quad \begin{aligned} \alpha_0^\epsilon &= 1 \\ \alpha_1^\epsilon &= \alpha_0^\epsilon \cdot \sin \left[ \frac{\pi}{2} \cdot m^\epsilon \right] \cdot \exp \left[ - \frac{1}{2} \left[ \frac{\pi}{2} \cdot \sigma^\epsilon \right]^2 \right] \\ \alpha_2^\epsilon &= \alpha_0^\epsilon \left[ p + \frac{1}{2} \left[ 1 - \cos[\pi \cdot m^\epsilon] \exp \left[ - \frac{1}{2} (\pi \cdot \sigma^\epsilon)^2 \right] \right] \right] \end{aligned}$$

à partir des  $\lambda_i^\epsilon$ , ( $i=0,1,2$ ) actualisés qui fournissent ensuite la moyenne  $m^\epsilon$  et la variance  $\sigma^{\epsilon 2}$ .

**III FILTRE FINI A SIX PARAMETRES**

Ce cas n'est pas seulement intéressant par lui-même. Il sert également, comme premier pas, de la procédure itérative de calcul, par point fixe des coefficients pour le cas à huit paramètres.

Par hypothèse de récurrence, la loi de  $X(t)$  conditionnellement à  $Y(0 \rightarrow t-1)$  est approchée par:

$$(III.1) \quad \text{Log } p(x/t-1) = \left[ \lambda_0^- + \lambda_1^- x + \lambda_2^- x^2 \right] 1_{]-\infty, 0]} + \left[ \lambda_0^+ + \lambda_1^+ x + \lambda_2^+ x^2 \right] 1_{]0, +\infty[}$$

Mise à jour

Les coefficients  $\lambda_i^\epsilon$  ( $i=0,1,2$  et  $\epsilon=+,-$ ), sont modifiés afin de tenir compte de l'information  $Y(t)=y$ . En raison de (I.3), les coefficients  $\lambda$  actualisés vérifient:

$$(III.2) \quad \begin{cases} \lambda_0^- \leftarrow \lambda_0^- + l_0 + l_1 y + l_2 y^2 \\ \lambda_1^- \leftarrow \lambda_1^- - (l_1 + 2l_2 y) \\ \lambda_2^- \leftarrow \lambda_2^- + l_2 \\ \lambda_0^+ \leftarrow \lambda_0^+ + l_0 + l_1 y + l_2 y^2 \\ \lambda_1^+ \leftarrow \lambda_1^+ - (l_1 + 2l_2 y) \\ \lambda_2^+ \leftarrow \lambda_2^+ + l_2 \end{cases}$$

Nous calculons  $q^-$  et  $q^+$  intégrales de  $C p(x/t)$  (I.3) respectivement sur  $]-\infty, 0]$  et sur  $]0, +\infty[$

Ensuite apparaissent les deux probabilités:

$$(III.3) \quad p^- \hat{=} P\{X(t) \in ]-\infty, 0[ / Y(0 \rightarrow t)\} = \frac{q^-}{q^- + q^+}$$

$$(III.4) \quad p^+ \hat{=} P\{X(t) \in [0, +\infty[ / Y(0 \rightarrow t)\} = \frac{q^+}{q^- + q^+}$$

et le coefficient  $C$  de normalisation:

$$(III.5) \quad C = q^- + q^+$$

**Prédiction**

Il s'agit d'obtenir numériquement, pour  $i=0,1,2$  et  $\epsilon=+,-$ :

$$\int \int \varphi_i^\epsilon \left[ \sin \left( \frac{\pi}{2} \cdot x \right) + w \right] \cdot p(x/t) dx \cdot p_w(w) dw \hat{=} \alpha_i^\epsilon$$

Ensuite, nous calculons les six coefficients  $\lambda_1^-, \lambda_1^+$  pour la variable  $X(t+1)$  conditionnellement à  $Y(0 \rightarrow t)$ .

Si nous supposons les probabilités très concentrées autour de  $-1$  et  $+1$ , nous séparons les calculs d'intégrales sur  $]-\infty, 0]$  et sur  $]0, +\infty[$  et nous pouvons utiliser les relations inversibles, moments  $\alpha$ , paramètres  $\lambda$  relatives au cas gaussien. Dans le cas contraire, les méthodes numériques du cas huit paramètres sont utilisées.

**IV FILTRE FINI A HUIT PARAMETRES**

Nous approchons à chaque date  $t \in \mathbb{N}$ , la densité de probabilité  $p$  par la loi exponentielle.

Mise à jour

En raison de (I.3) et comme en (I.6), nous additionnons les coefficients des fonctions  $\varphi_i^\epsilon$ :

$$(IV.1) \quad \begin{cases} \lambda_0^- \leftarrow \lambda_0^- + l_0 + l_1 y + l_2 y^2 \\ \lambda_1^- \leftarrow \lambda_1^- - l_1 - 2l_2 y \\ \lambda_2^- \leftarrow \lambda_2^- + l_2 \\ \lambda_3^- \leftarrow \lambda_3^- \\ \lambda_0^+ \leftarrow \lambda_0^+ + l_0 + l_1 y + l_2 y^2 \\ \lambda_1^+ \leftarrow \lambda_1^+ - l_1 - 2l_2 y \\ \lambda_2^+ \leftarrow \lambda_2^+ + l_2 \\ \lambda_3^+ \leftarrow \lambda_3^+ \end{cases}$$



Les deux probabilités,  $p^-$  et  $p^+$  sont données par les définitions (III.3), (III.4) et sont calculées numériquement.

#### Prédiction

La loi prédite approchée est évidemment de même type. De plus:

$$\alpha_i^\varepsilon \cong E \left[ \varphi_1^\varepsilon [X(t+1)/Y(0 \rightarrow t)] \right] = \int \int \varphi_1^\varepsilon \left[ \sin\left(\frac{\pi}{2} \cdot x\right) + w \right] \cdot p_{X(t)}(x/t) dx \cdot p_w dw$$

Les moments ainsi définis, sont approchés par interpolation numérique. La méthode choisie,

pour approcher ensuite les paramètres  $\lambda_i^\varepsilon$ , ( $i=0,1,2,3$  et  $\varepsilon=+,-$ ) est la méthode de Newton.

#### V SIMULATION

Les résultats de simulation sont présentés pour:

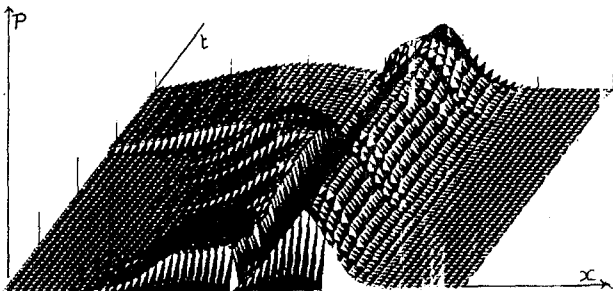
$$X(0) = \frac{\pi}{6} \quad P=0.1 \quad Q=0.25$$

#### Détection de changement de modèle

Dès la première itération, nous observons une concentration de la loi de probabilité autour de +1, avec une évolution visible de ses paramètres.

Des dates 4 à 40, nous observons des changements de modèles successifs. Lors d'une détection sur  $]-\infty, 0]$ , la loi autour de -1 évolue: concentration de la gaussienne due à une diminution de la variance.

A la date 40, nous avons un changement de modèle très net, avec une concentration autour de -1 et arrêt de l'évolution de la loi autour de +1. Lors du retour au modèle positif, nous observons une convergence progressive de la moyenne  $m^+$  vers +1. La barrière en zéro est due à la séparation en deux intervalles d'étude  $]-\infty, 0]$  et  $]0, +\infty[$ .



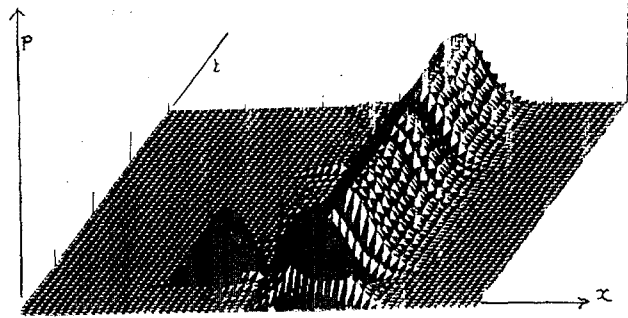
#### Filtre fini 6 paramètres

Dans le cas du filtrage fini, l'amplitude est représentative de la probabilité. La courbe obtenue est très différente de la courbe de détection.

Dès la première itération, nous avons concentration de la loi de probabilité autour de +1, avec une probabilité autour de -1 quasiment nulle.

L'évolution de la gaussienne autour de +1 se caractérise par de petites variations de la moyenne et de la variance. Tandis que la concentration autour de -1 reste nulle.

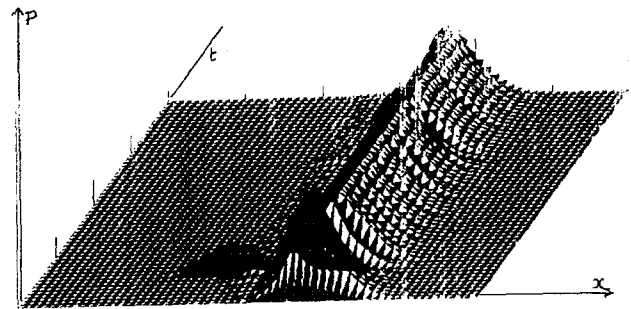
Le seul changement de modèle observé, se situe aux alentours de la date 40, avec une concentration maximale de la gaussienne autour de -1 et minimale autour de +1.



#### Filtre fini 8 paramètres

Nous avons encore concentration de la loi de probabilité autour de +1, avec un seul changement de modèle aux alentours de la date 40.

Notons que la probabilité de la loi autour de -1 est très nettement inférieure à celle obtenue par filtre fini 6 paramètres. Nous affinons ainsi le résultat par une concentration globale autour de +1 plus importante.



#### CONCLUSION

La simulation montre que le filtrage 8 paramètres est meilleur. La concentration de la loi de probabilité autour de +1 est plus importante. L'hypothèse de changement de modèles est très simplificatrice.

L'évolution non-linéaire entraînant la perte du caractère gaussien, amène des termes supplémentaires sur la densité.

Le choix des fonctions  $\varphi$  relève de développements théoriques à venir. La seule considération que nous avons prise en compte, est la création d'une dissymétrie en -1 et +1.

#### BIBLIOGRAPHIE

- BASSEVILLE M. (1988) Detecting changes in signals and systems—a survey. *Automatica*, vol.24, pp.309-326.
- BUSVELLE E., RAKOTOPARA D., de BRUCQ D. (1990) Particle Filtering Driven by Observation. Rapport SHELL-Recherche 76530 Grand Couronne.
- CSISZAR I. (1984) Sanov Property, Generalized L-Projection and a conditional limit theorem. *The annals of Probability*. Vol 12 n°3 pp.768-793.
- de BRUCQ D., BEN ABDALLAH K., BUSVELLE E., COURTELLEMONT P. (1990) Extension du filtre de Kalman, approximation quadratique. GRECO TDSI.
- GAMBOA F. (1990) Méthode du maximum d'entropie sur la moyenne et applications. Doctorat en Sciences, Mathématiques. Paris Sud, Centre d'Orsay.
- JUMARIE G. (1990) Non linear filtering. A weighted mean squares approach and a bayesian one via the maximum entropy principle. *Signal Processing* 21 pp.323-328.