

Filtrage de Kalman non linéaire à l'aide de noyaux

Liva RALAIVOLA, Florence D'ALCHÉ-BUC

Laboratoire d'Informatique de Paris 6
Université Pierre et Marie Curie
8, rue du Capitaine Scott
F-75015 Paris, France
liva.ralaivola@lip6.fr

Résumé – Dans ce travail, nous proposons une méthode originale permettant d'appréhender des séries temporelles aux caractéristiques non linéaires. Cette méthode repose sur une extension du filtrage de Kalman à l'aide de l'astuce du noyau (*kernel trick*). L'usage d'une fonction noyau permet d'employer la formulation classique du filtre Kalman dans l'espace des caractéristiques défini par la fonction noyau.

Abstract – In this work, we propose an original method which allows to address the modeling of highly nonlinear time series. This method is based on an extension of the Kalman Filter model with the help of the *kernel trick*. This algorithm has the advantage to keep the usual matrix equations valid.

1 Introduction

Nous proposons d'étendre le filtrage de Kalman [1, 10, 19] au traitement de séries dont les dépendances markoviennes ne sont pas linéaires. Notre approche se fonde sur l'introduction de fonctions noyaux qui permettent de prendre naturellement en compte les non linéarités. Le passage aux fonctions noyaux, appelé « astuce du noyau » (*kernel trick*) [6], nous permet de considérer que les données des séries étudiées sont liées linéairement mais ce, dans l'espace des caractéristiques défini par le noyau.

À l'inverse d'autres techniques non linéaires dérivées du filtre de Kalman telles que EKF (*Extended Kalman Filter* [19]) et UKF (*Unscented Kalman Filter* [9]), notre méthode permet d'englober dans le même formalisme matriciel aussi bien les phases classiques de *filtrage* et de *lissage* que la phase d'identification des paramètres du filtre.

Nous rappelons d'abord le principe du filtre de Kalman et les équations d'estimation associées. Nous décrivons ensuite notre approche et l'instancions dans le cas de l'observation d'une version simplement bruitée du processus caché. Nous présentons alors une approche *maximum a posteriori* et fournissons les paramètres optimaux de la vraisemblance pénalisée considérée. Enfin, nous procédons à des tests numériques sur la série chaotique Mackey-Glass.

2 Filtre de Kalman

2.1 Filtrage linéaire

L'algorithme proposé par Kalman [10] permet l'estimation d'un processus $\mathbf{s}_t \in \mathcal{R}^p$, auquel on n'a accès que par le biais de mesures $\mathbf{x}_t \in \mathcal{R}^d$: il se fonde sur le modèle probabiliste

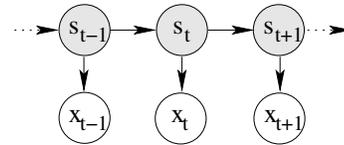


FIG. 1 – Représentation en tant que modèle à états du filtre de Kalman.

suivant (cf. figure 1) :

$$\mathbf{s}_{t+1} = A\mathbf{s}_t + \boldsymbol{\mu}_s + \mathbf{v} \quad (1)$$

$$\mathbf{x}_t = B\mathbf{s}_t + \boldsymbol{\mu}_x + \mathbf{w} \quad (2)$$

où \mathbf{v} et \mathbf{w} sont des vecteurs de bruit blanc gaussien de covariances respectives $\sigma_s^2 I$ et $\sigma_x^2 I$, et où le premier état du processus est distribué selon une loi gaussienne de moyenne $\boldsymbol{\mu}_1$ et de covariance $\sigma_s^2 I$, I étant la matrice identité.

Étant donné une série de mesures $\mathbf{x}_{1:T} = \{\mathbf{x}_1, \dots, \mathbf{x}_T\}$, l'estimation des états \mathbf{s}_t peut se faire à l'aide des procédures de filtrage et de lissage rappelées dans le tableau 1. Celles-ci supposent que les paramètres $\boldsymbol{\theta} = \{A, B, \boldsymbol{\mu}_s, \boldsymbol{\mu}_x, \boldsymbol{\mu}_1, \sigma_s^2, \sigma_x^2, \sigma_1^2\}$ sont connus et interviennent dans l'algorithme *Expectation-Maximization* (EM) [4, 5] dans le cas où l'on s'intéresse à la détermination des paramètres $\boldsymbol{\theta}$ du modèle [15].

2.2 Filtrage non linéaire

Les méthodes de filtrage non linéaires permettent d'appréhender un processus du type :

$$\mathbf{s}_{t+1} = f(\mathbf{s}_t) + \mathbf{v} \quad (3)$$

$$\mathbf{x}_t = g(\mathbf{s}_t) + \mathbf{w} \quad (4)$$

où f et g peuvent être non linéaires et où les vecteurs \mathbf{v} et \mathbf{w} sont des vecteurs de bruit dont les caractéristiques sont décrites comme précédemment.

Ce système généralise le système (1)(2) et à ce titre, deux extensions de l'algorithme d'estimation de Kalman à des dé-

TAB. 1 – Équations de filtrage et lissage. $s^0(1) = \mu_1$, $\Sigma_1^0 = \sigma_1^2 I$, $\hat{s}(T) = \mathbf{s}^T(T)$ et $\hat{\Sigma}(T) = \Sigma^T(T)$. $\hat{R}(t)$ et $\hat{R}^{t-1}(t)$ servent à la phase de maximisation de EM lorsqu'est posé le problème de l'identification des paramètres du système.

Filtrage	Lissage
$\mathbf{s}^{t-1}(t) = A\mathbf{s}^{t-1}(t-1) + \boldsymbol{\mu}_s$ $\Sigma^{t-1}(t) = A\Sigma^{t-1}(t-1)A' + \sigma_s^2 I$ $\Sigma_e(t) = B\Sigma^{t-1}(t)B' + \sigma_x^2 I$ $K_t = \Sigma^{t-1}(t)B'\Sigma_e^{-1}(t)$ $\mathbf{e}_t = \mathbf{x}_t - B\mathbf{s}^{t-1}(t) - \boldsymbol{\mu}_x$ $\mathbf{s}^t(t) = \mathbf{s}^{t-1}(t) + K_t\mathbf{e}_t$ $\Sigma^t(t) = \Sigma^{t-1}(t) - K_tB\Sigma^{t-1}(t)$	$J_{t-1} = \Sigma^{t-1}(t-1)A'(\Sigma^{t-1}(t-1))^{-1}$ $\hat{\mathbf{s}}(t-1) = \mathbf{s}^{t-1}(t-1) + J_{t-1}(\hat{\mathbf{s}}(t) - \mathbf{s}^{t-1}(t))$ $\hat{\Sigma}(t-1) = \Sigma^{t-1}(t-1) + J_{t-1}(\hat{\Sigma}(t) - \Sigma^{t-1}(t))J_{t-1}'$ $\hat{\Sigma}^{t-1}(t) = \hat{\Sigma}(t)J_{t-1}'$
	$\hat{R}(t) = \hat{\Sigma}(t) + \hat{\mathbf{s}}(t)\hat{\mathbf{s}}'(t)$ $\hat{R}^{t-1}(t) = \Sigma^{t-1}(t) + \hat{\mathbf{s}}(t)\hat{\mathbf{s}}'(t-1)$

pendances non linéaires sont connues : *Extended Kalman Filter* (EKF) [19] et *Unscented Kalman Filter* (UKF) [9]. EKF repose sur des approximations linéaires de f et g obtenues par leur développement de Taylor au premier ordre ; UKF fait une estimation des états par un processus efficace d'« échantillonnage ». Tout comme les procédures d'estimation linéaires, EKF et UKF font l'hypothèse de la connaissance de f et g .

La démarche usuelle pour l'appréhension du système non linéaire (3)(4) à l'aide des procédures dérivées de l'estimation de Kalman se décompose donc en deux étapes. D'abord, les fonctions f et g sont modélisées, par exemple à l'aide de réseaux de neurones, à partir d'une série d'états non bruités et de leurs mesures, non bruitées également. Les algorithmes EKF ou UKF sont ensuite utilisés.

3 Filtre de Kalman à noyaux

3.1 Fonctions noyaux

Depuis l'introduction des noyaux dans l'algorithme de l'hyperplan de marge optimale [6, 7, 17], de nombreux algorithmes linéaires ont été étendus à des traitements non linéaires grâce à l'astuce des noyaux (cf. [2, 11, 18] par exemple). Une fonction noyau k définie sur le produit cartésien $\mathcal{R}^p \times \mathcal{R}^p$ et satisfaisant les conditions de Mercer [16] peut s'écrire $k(\mathbf{u}, \mathbf{v}) = \langle \phi(\mathbf{u}), \phi(\mathbf{v}) \rangle$ pour tout \mathbf{u} et \mathbf{v} avec $\phi : \mathcal{R}^p \rightarrow \mathcal{F}$ opérant la transformation des données dans l'espace des caractéristiques \mathcal{F} . L'utilisation de k permet de travailler implicitement dans l'espace \mathcal{F} , et donc d'induire des non linéarités. Les noyaux classiquement utilisés sont les noyaux polynomial $k(\mathbf{u}, \mathbf{v}) = \left(\frac{\mathbf{u} \cdot \mathbf{v}}{p} + 1\right)^d$, $d > 0$ et gaussien $k(\mathbf{u}, \mathbf{v}) = \exp\left(-\frac{\|\mathbf{u} - \mathbf{v}\|^2}{2p\sigma_k^2}\right)$.

3.2 Modèle à noyau

L'approche proposée permet d'aborder le problème de l'estimation des états du processus, celui de la prévision et celui de l'identification des dynamiques f et g de (3)(4) en utilisant les équations du tableau 1. Elle repose sur l'usage de fonctions noyaux. Étant donné une série d'observations $\mathbf{x}_{1:T}$, un noyau k et la fonction ϕ correspondante, nous proposons le modèle suivant :

$$\mathbf{s}_{t+1}^\phi = A^\phi \mathbf{s}_t^\phi + \boldsymbol{\mu}_s^\phi + \mathbf{v}^\phi \quad (5)$$

$$\mathbf{x}_t^\phi = B^\phi \mathbf{s}_t^\phi + \boldsymbol{\mu}_x^\phi + \mathbf{w}^\phi \quad (6)$$

où les vecteurs ayant ϕ en exposant appartiennent à \mathcal{F} , et les lignes et les colonnes de A^ϕ et B^ϕ sont des vecteurs de \mathcal{F} ; \mathbf{v}^ϕ et \mathbf{w}^ϕ sont des vecteurs de bruit gaussiens de variances respectives σ_s^2 et σ_x^2 . Nous désignerons les paramètres de (5)(6) par $\boldsymbol{\theta}^\phi = \{A^\phi, B^\phi, \boldsymbol{\mu}_s^\phi, \boldsymbol{\mu}_x^\phi, \boldsymbol{\mu}_1^\phi, \sigma_s^2, \sigma_x^2, \sigma_1^2\}$.

Au lieu de considérer la série $\mathbf{x}_{1:T}$ nous choisissons ainsi d'étudier la série $\mathbf{x}_{1:T}^\phi = \{\mathbf{x}_1^\phi, \dots, \mathbf{x}_T^\phi\}$ des images par ϕ des \mathbf{x}_t , en notant $\mathbf{x}_t^\phi := \phi(\mathbf{x}_t)$. L'utilisation de l'application non linéaire ϕ , conduit un modèle non linéaire alors que, comme nous le montrons dans ce qui suit, le formalisme usuel du filtrage de Kalman reste applicable.

3.3 Filtrage et lissage à noyaux

Supposons que $\boldsymbol{\theta}^\phi$ est connu et que, par ailleurs, il existe un ensemble fini $\mathbf{f}_1, \dots, \mathbf{f}_m$ de vecteurs de \mathcal{R}^p tel que

$$\boldsymbol{\mu}_1^\phi = F\tilde{\boldsymbol{\mu}}_1, \quad \boldsymbol{\mu}_s^\phi = F\tilde{\boldsymbol{\mu}}_s, \quad \boldsymbol{\mu}_x^\phi = F\tilde{\boldsymbol{\mu}}_x \quad (7)$$

$$A^\phi = F\tilde{A}F', \quad B^\phi = F\tilde{B}F' \quad (8)$$

avec $\tilde{\boldsymbol{\mu}}_1, \tilde{\boldsymbol{\mu}}_s, \tilde{\boldsymbol{\mu}}_x \in \mathcal{R}^m$, $\tilde{A}, \tilde{B} \in \mathcal{R}^{m \times m}$ et F la matrice $F = [\phi(\mathbf{f}_1) \cdots \phi(\mathbf{f}_m)]$. Sous ces hypothèses, pour tout vecteur $\mathbf{y} \in \mathcal{R}^p$, on a :

$$A^\phi \mathbf{y} = F\tilde{A}F' \mathbf{y} = F\tilde{A} \mathbf{f}_y = F\tilde{\mathbf{a}}_y$$

où $\mathbf{f}_y = [\langle \phi(\mathbf{f}_1), \phi(\mathbf{y}) \rangle \cdots \langle \phi(\mathbf{f}_m), \phi(\mathbf{y}) \rangle]'$ ou encore $\mathbf{f}_y = [k(\mathbf{f}_1, \mathbf{y}) \cdots k(\mathbf{f}_m, \mathbf{y})]'$ et $\tilde{\mathbf{a}}_y$ est le vecteur résultant de la multiplication entre \tilde{A} et \mathbf{f}_y .

On peut procéder aux opérations de filtrage et lissage de la série $\mathbf{x}_{1:T}$ à partir de la série $\mathbf{x}_{1:T}^\phi$ en appliquant les équations du tableau 1 en remplaçant les paramètres de $\boldsymbol{\theta}$ par ceux de $\boldsymbol{\theta}^\phi$ leur correspondant. La seule difficulté apparaissant dans l'exécution de ces procédures se situe au niveau de l'inversion de matrices de dimension éventuellement infinie (e.g., si le noyau gaussien k est utilisé). Or, il est facile de montrer par récurrence¹ que, sous les hypothèses faites, les matrices $\Sigma^{t-1}(t)$ estimées s'écrivent sous la forme $\Sigma^{t-1}(t) = F\tilde{\Sigma}^{t-1}(t)F'$ et que, dès lors, les inversions de $\Sigma_e(t)$ et $\Sigma^{t-1}(t)$ peuvent se faire grâce à la formule de Woodbury. Pour le calcul de $\Sigma_e^{-1}(t)$ (permettant la détermination de K_t) on a par exemple :

$$\begin{aligned} \Sigma_e^{-1}(t) &= (B\Sigma^{t-1}(t)B' + \sigma_x^2 I)^{-1} \\ &= (F\tilde{B}K_F\tilde{\Sigma}^{t-1}(t)K_F\tilde{B}F' + \sigma_x^2 I)^{-1} \\ &= \frac{1}{\sigma_x^2} \left(I - \frac{1}{\sigma_x^2} F \left(I + \frac{1}{\sigma_x^2} G K_F \right)^{-1} G F' \right) \end{aligned}$$

¹Par souci de concision, nous omettons ici la démonstration qui est par ailleurs triviale.

avec $K_F = F'F = (k(\mathbf{f}_i, \mathbf{f}_j))_{i,j}$ matrice d'ordre m et $G = \tilde{B}K_F\tilde{\Sigma}^{t-1}(t)K_F\tilde{B}$. La seule inversion de matrice concerne alors la matrice $I + \frac{1}{\sigma_x^2}GK_F$. Le même type de calcul peut se faire pour l'inversion de $\Sigma^{t-1}(t-1)$ (en tenant compte du fait que K_t s'écrit aussi sous la forme $K_t = F\tilde{K}_tF'$).

Une fois définis les paramètres θ^ϕ du système (5)(6), il est donc possible de mettre en œuvre le processus d'estimation décrit dans le tableau 1.

3.4 Identification des paramètres θ^ϕ

Nous nous intéressons ici à la détermination de ces paramètres à partir de l'observation d'une séquence $\mathbf{x}_{1:T}$ dans le cas particulier où $B^\phi = I$, $\boldsymbol{\mu}^\phi = 0$ et σ_x^2 est connu. La situation correspond au cas usuel où la dynamique liant les états entre 2 instants successifs doit être apprise.

Les paramètres A^ϕ , $\boldsymbol{\mu}_s^\phi$, $\boldsymbol{\mu}_1^\phi$, σ_s^2 , σ_1^2 s'obtiennent par maximisation de la *log-vraisemblance pénalisée* \mathcal{L}^ϕ [13] définie par $2\mathcal{L}^\phi(\mathbf{x}_{1:T}; \theta^\phi) = 2\log(p(\mathbf{x}_{1:T}|\theta^\phi))p_{\gamma_a}(A^\phi)$ qui s'écrit

$$-n(T-1)\log\sigma_s^2 - h(\boldsymbol{\mu}_1^\phi, \sigma_1^2) - \frac{\gamma_a}{\sigma_s^2}\text{tr}(A^\phi A^{\phi'}) \quad (9)$$

$$- \frac{1}{\sigma_s^2} \sum_{t=2}^T \|\mathbf{s}_t^\phi - A^\phi \mathbf{s}_{t-1}^\phi - \boldsymbol{\mu}_s^\phi\|^2 \quad (10)$$

où $\text{tr}(M)$ désigne la trace de M et où la distribution p_{γ_a} (cf. équation (9)) impose de choisir la matrice A^ϕ la plus simple possible ; le même type de distributions *a priori* a été utilisé dans [3, 8] pour l'inférence bayésienne des paramètres et des états d'un filtre de Kalman linéaire. n est la dimension de l'espace \mathcal{F} et h une fonction de $\boldsymbol{\mu}_1^\phi$ et σ_1^2 qui se calcule facilement.

L'annulation de la dérivée de \mathcal{L}^ϕ conduit à :

$$A^\phi = \left(\sum_{t=2}^T \mathbf{x}_t^\phi \mathbf{x}_{t-1}^{\phi'} - \frac{1}{T-1} \sum_{t=2}^T \mathbf{x}_t^\phi \sum_{t=2}^T \mathbf{x}_{t-1}^{\phi'} \right) \quad (11)$$

$$\left(\gamma_a I + \sum_{t=2}^T \mathbf{x}_{t-1}^\phi \mathbf{x}_{t-1}^{\phi'} - \frac{1}{T-1} \sum_{t=2}^T \mathbf{x}_{t-1}^\phi \sum_{t=2}^T \mathbf{x}_{t-1}^{\phi'} \right)^{-1} \boldsymbol{\mu}_s^\phi = \frac{1}{T-1} \sum_{t=2}^T (\mathbf{x}_t^\phi - A^\phi \mathbf{x}_{t-1}^\phi) \quad (12)$$

$$\sigma_s^2 = \frac{1}{n(T-1)} \sum_{t=2}^T \|\mathbf{x}_t^\phi - A^\phi \mathbf{x}_{t-1}^\phi - \boldsymbol{\mu}_s^\phi\|^2 \quad (13)$$

en tenant compte de $B^\phi = I$ et $\boldsymbol{\mu}^\phi = 0$. L'équation (11) pose le problème de l'inversion d'une matrice d'ordre n , qui peut être inconnue (ou infinie). Tout comme pour le processus d'estimation, cet inconvénient peut être évité en introduisant la matrice $X^\phi = [\mathbf{x}_1^\phi \cdots \mathbf{x}_T^\phi]$, les vecteurs $\mathbf{f} := [0 \ 1 \cdots 1]'$, $\mathbf{g} := [1 \cdots 1 \ 0]$ de dimension T , la matrice $P = (P_{ij}) = (\delta_{i,j+1})$ d'ordre T et en définissant $J = P - \mathbf{f}\mathbf{g}'/(T-1)'$ et $M = \text{diag}(\mathbf{g}) - \mathbf{g}\mathbf{g}'/(T-1)$. Grâce à la formule de Woodbury, A^ϕ se réécrit en effet

$$\begin{aligned} A^\phi &= \left(X^\phi J X^{\phi'} \right) \left(\gamma I + X^\phi M X^{\phi'} \right)^{-1} \\ &= \frac{1}{\gamma} X^\phi J \left[I - \frac{1}{\gamma} K \left(I + \frac{1}{\gamma} M K \right)^{-1} M \right] X^{\phi'} \end{aligned}$$

avec $K = (k(\mathbf{x}_i, \mathbf{x}_j))_{i,j}$. L'équation (13) pose le problème d'une division par n . Afin de déterminer σ_s^2 , on peut utiliser une base orthonormée U de l'espace \mathcal{F}_x engendré par les $\mathbf{x}_1, \dots, \mathbf{x}_T$, obtenu en calculant les m vecteurs propres $\mathbf{u}_1, \dots, \mathbf{u}_m$ de la

TAB. 2 – Erreur quadratique moyenne en prévision MG_{17} . Deux noyaux sont considérés et le coefficient de régularisation figure en indice de FKN. En italique, les meilleurs résultats.

Méthode	gaussien		polynomial	
	1S	100S	1S	100S
SVM	<i>0.0812</i>	0.2361	0.1156	-
FKN ₀	0.0864	0.2906	0.1112	0.2975
FKN _{0.1}	0.0863	0.2893	0.1112	0.2775
FKN ₁	0.0859	0.2871	0.1117	0.2956
FKN ₁₀	0.0844	0.2140	0.1203	0.1964
FKN ₁₀₀	0.0899	<i>0.1733</i>	<i>0.0970</i>	<i>0.1744</i>

matrice $X^\phi X^{\phi'}$ de valeurs propres strictement positives (cf. [11]). Le paramètre σ_s^2 est alors [14] :

$$\sigma_s^2 = \frac{1}{m(T-1)} \sum_{t=2}^T (\mathbf{x}_t^\phi - A^\phi \mathbf{x}_{t-1}^\phi - \boldsymbol{\mu}_s^\phi)' U U' (\mathbf{x}_t^\phi - A^\phi \mathbf{x}_{t-1}^\phi - \boldsymbol{\mu}_s^\phi).$$

Ces paramètres sont bien de la forme (7)(8), avec $\mathbf{f}_t = \mathbf{x}_t$, les équations d'estimation peuvent donc être utilisées.

4 Expériences

Nous testons l'efficacité de notre approche sur les deux séries unidimensionnelles Mackey-Glass MG_{17} et MG_{30} en abordant deux tâches. La première vise à construire un modèle selon la méthode préconisée de la section 3.4 à partir de points de ces séries et à évaluer les capacités de prévision du modèle ainsi obtenu. La seconde consiste, à partir du modèle construit, à ôter le bruit d'une série.

4.1 Prévision

Nous comparons les performances de notre algorithme à celles d'une machine à vecteurs de support [12] pour laquelle les vecteurs $\mathbf{x}_t = [x_{t-6(d-1)} \cdots x_t]'$, avec $d = 6$ sont associés à la valeur cible x_{t+6} . Les capacités de prévision à 1 (1S) et 100 (100S) pas de temps sont évaluées pour ces deux modèles et indiquées dans le tableau 2. La base de validation est constituée de 100 points n'ayant pas été utilisés lors de l'apprentissage – effectué sur les 100 premiers \mathbf{x}_t . Les paramètres de la SVM sont réglés par cross-validation sur un échantillon indépendant. Les paramètres du noyau SVM sont utilisés pour le filtre de Kalman à noyaux (FKN). Les résultats, qui concernent MG_{17} , sont résumés dans le tableau 2². On constate d'une part que les performances de prévisions atteintes par notre modèle sont de très bonne qualité. D'autre part, l'utilisation d'un coefficient de régularisation permet d'éviter le phénomène de divergence qui se produit avec la SVM et le noyau polynomial pour la prévision 100S.

4.2 Soustraction de bruit

Nous nous intéressons ici à la tâche de soustraction d'un bruit blanc de variance 0.5. Le modèle FKN est appris sur les

²Le modèle (5)(6) permet seulement d'avoir des prévisions dans \mathcal{F} ; nous avons donc mis en œuvre un algorithme de détermination de préimages pour évaluer l'erreur quadratique moyenne [11, 14].

TAB. 3 – Performances pour la soustraction de bruit. Plus la valeur du noyau est proche de 1 plus la soustraction de bruit est efficace.

Méthode	Dénomination	Performance
FKN	Filtrage à noyaux	0.985
EKF	Extended Kalman Filter	0.993
UKF	Unscented Kalman Filter	0.994
CDKF	Central difference KF	0.993
SRCDKF	Square Root CDKF	0.990
SRUKF	Square Root UKF	0.991

300 premiers points de la série non bruitée en considérant les vecteurs $\mathbf{x}_t = [x_{t-5}, \dots, x_t]'$ pour extraire le signal originel de 300 points bruités grâce aux équations d'estimation. Les performances sont comparées à celle d'un réseau de neurones d'architecture 6-4-1 utilisé avec des variantes d'EKF et UKF et appris sur des vecteurs $\mathbf{x}_t = [x_{t-30}, x_{t-24}, \dots, x_t]'$ ³. On utilise un noyau gaussien avec $\sigma_k = 1$ et un coefficient de régularisation de 10. L'erreur est mesurée en calculant les noyaux entre les vecteurs \mathbf{x}_t non bruités et ceux estimés car FKN ne fournit que des estimations de vecteurs de \mathcal{F} .

On constate d'une part que la qualité d'estimation de FKN est très proche de celle des autres méthodes. Il faut cependant remarquer que le codage en vecteur utilisé par FKN n'est pas aussi optimal [12] que celui utilisé par le réseau de neurones. D'autre part, le coefficient de régularisation utilisé a été choisi intuitivement et une méthode plus robuste de détermination (e.g. par cross-validation) de ce paramètre devrait permettre à FKN d'atteindre de meilleures performances.

5 Conclusion et perspectives

Nous proposons une extension du filtre de Kalman à l'aide de noyaux. Cette approche originale permet d'appréhender des séries temporelles chaotiques et présente l'avantage de conserver le même type de calculs matriciels que ceux impliqués dans les équations d'estimation de Kalman. De plus, s'agissant de l'apprentissage des paramètres du modèle, les équations classiques utilisées par EM continuent de s'appliquer. Nous montrons qu'il est aussi possible de contrôler la complexité du modèle en maximisant la vraisemblance pénalisée. Les premières simulations ont porté sur une simplification du filtre de Kalman à noyau dans laquelle le processus observé correspond au processus caché bruité. Les premiers résultats obtenus en prévision et en identification de bruit sur deux séries non linéaires classiques sont très encourageants. Ce travail peut être prolongé dans différentes directions : tout d'abord, nous devons valider le modèle général en prévision, en identification de bruit et en lissage. Ensuite, la question du choix du noyau et surtout de la sélection de ses paramètres doit être étudiée. Enfin, cet algorithme ouvre la voie à l'analyse de séquences de données structurées (non vectorielles) pour lesquelles un noyau peut être défini.

Références

- [1] B. D. Anderson and J. B. Moore. *Optimal Filtering*. Prentice Hall, Englewood Cliffs, NJ, 1979.
- [2] F. R. Bach and M. I. Jordan. Kernel Independent Component Analysis. *Journal of Machine Learning Research*, 3 :1–48, 2002.
- [3] M. J. Beal and Z. Ghahramani. The Variational Kalman Smoother. Technical report, Gatsby Computational Neuroscience Unit, 2001.
- [4] C.M. Bishop. *Neural Networks for Pattern Recognition*, chapitre 1. Oxford University Press, Inc, New York, 1995.
- [5] J. Blimes. A gentle tutorial of the EM algorithm and its application to parameter estimation for gaussian mixture and hidden markov models. Technical report, U.C. Berkeley, 1998.
- [6] B. Boser, I. Guyon, and V. Vapnik. A Training Algorithm for Optimal Margin Classifiers. In *Proc. of the 5th Annual Workshop on Comp. Learning Theory*, volume 5, 1992.
- [7] C. Burges. A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery*, 2(2) :955–974, 1998.
- [8] Z. Ghahramani and M.J. Beal. Propagation algorithms for variational bayesian learning. In *NIPS*, pages 507–513, 2000.
- [9] S. Julier and J. Uhlmann. A New Extension of the Kalman Filter to Nonlinear Systems. In *Int. Symp. Aerospace/Defense Sensing, Simul. and Controls*, 1997.
- [10] R. E. Kalman. A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME—Journal of Basic Engineering*, 82(Series D) :35–45, 1960.
- [11] S. Mika, B. Schölkopf, A. J. Smola, K.-R. Müller, M. Scholz, and G. Rätsch. Kernel PCA and De-Noising in Feature Spaces. In *NIPS*. MIT Press, 1999.
- [12] S. Mukherjee, E. Osuna, and F. Girosi. Nonlinear prediction of chaotic time series using support vector machines. In *Proc. of IEEE NNSP'97*, 1997.
- [13] D. Ormoneit and V. Tresp. Averaging, maximum penalized likelihood and bayesian estimation for improving gaussian mixture probability density estimates. *IEEE Transactions on Neural Networks*, 9, 1998.
- [14] L. Ralaivola and F. d'Alché-Buc. Modèles dynamiques à noyaux pour la prévision de séries temporelles non linéaires. To appear in Actes de la Conférence Apprentissage CAp'03, Laval, 2003.
- [15] A-V.I. Rosti and M.J.F. Gales. Generalised linear Gaussian models. Technical Report CUED/F-INFENG/TR.420, Cambridge University Engineering Department, 2001.
- [16] B. Schölkopf, R. Herbrich, and A. J. Smola. A generalized representer theorem. Technical Report NC-TR-00-081, NeuroCOLT, 2000.
- [17] V. Vapnik. *Statistical Learning Theory*. John Wiley and Sons, inc., 1998.
- [18] A. Vinokurov, J. Shawe-Taylor, and N. Cristianini. Inferring a Semantic Representation of Text via Cross-Language Correlation Analysis. In *Advances in Neural Information Processing Systems*, volume 14, 2003.
- [19] G. Welch and G. Bishop. An introduction to the Kalman filter. Technical Report TR 95-041, University of North Carolina, 1995.

³Nous avons utilisé la librairie *ReBEL*, disponible à l'url <http://choosh.ece.ogi.edu/rebel/index.html>.