

Adaptation en ligne d'un dictionnaire pour les méthodes à noyau

Chafic SAIDE¹, Paul HONEINE¹, Régis LENGELLE¹, Cédric RICHARD², Roger ACHKAR³

¹Institut Charles Delaunay (UMR CNRS 6279), Université de technologie de Troyes, France

²Laboratoire Lagrange (UMR CNRS 7293), Observatoire de la Côte d'Azur, Université de Nice Sophia-Antipolis, France

³Université Américaine de Science et Technologie (AUST), Beirut, Liban

chafic.saide@utt.fr, paul.honeine@utt.fr, regis.lengelle@utt.fr,
cedric.richard@unice.fr, rachkar@aust.edu.lb

Résumé – Cet article traite du problème de l'identification en-ligne des systèmes non linéaires et non stationnaires par les méthodes à noyau. L'ordre des modèles est contrôlé par le critère de cohérence utilisé comme critère de parcimonie, qui mène à sélectionner les fonctions noyau les plus pertinentes au sens de ce critère, formant ainsi un dictionnaire. On exploite l'adaptation du dictionnaire en proposant une méthode de descente de gradient stochastique qui s'applique conjointement à l'estimation en ligne des coefficients du modèle à noyau. un algorithme d'identification en ligne à noyau. Pour ce dernier, sans limitation, il peut s'agir de l'algorithme de moindres carrés récursif à noyau ou de projection affine à noyau. La méthode proposée permet une diminution de l'erreur quadratique instantanée et une réduction de la complexité du modèle.

Abstract – This article tackles the online identification problem for nonlinear and nonstationary systems using kernel methods. The order of the model is controlled by the coherence criterion considered as a sparsification technique which leads to select the most relevant kernel functions to form a dictionary. We explore the dictionary adaptation using a stochastic gradient descent method along with an online kernel identification algorithm. For the latter, without limitation, it may be the kernel recursive least squares algorithm or the kernel affine projection algorithm. The proposed method leads to a reduction of the instantaneous quadratic error and to a decrease in the model's order.

1 Motivation

L'identification en-ligne des systèmes non linéaires demeure un sujet de recherche très actif. Parmi les méthodes les plus connues, les filtres de Volterra [1, 2], ou encore les réseaux de neurones [3], ont fait l'objet d'un nombre important de travaux. Chacune de ces méthodes possède ses atouts et ses limitations. Par exemple, dans le cas des filtres de Volterra, le nombre de paramètres à estimer dépend de l'ordre du filtre et de la complexité de celui-ci. Ceci implique potentiellement une grande complexité en termes de nombre de paramètres [4]. Le point faible des réseaux de neurones multicouches réside dans le choix de la structure du réseau. Par ailleurs, le critère de performance n'est pas convexe en les paramètres du réseau.

L'estimation fonctionnelle basée sur les espaces de Hilbert à noyau reproduisant (RKHS pour Reproducing Kernel Hilbert Spaces en anglais) a récemment conduit à des avancées majeures dans le domaine de l'identification des systèmes non linéaires [5]. Toutefois, l'ordre des modèles identifiés est égal au nombre de couples entrée-sortie, ce qui rend cette méthode inadéquate pour une identification en ligne. Pour surmonter cet inconvénient, plusieurs méthodes ont été proposées afin de contrôler l'ordre du modèle [6, 7, 8, 9], le modèle étant alors défini à partir d'un dictionnaire de faible taille. Dans cet article, on propose d'adapter les éléments du dictionnaire, afin de réduire l'erreur quadratique instantanée résultante et/ou

mieux contrôler l'ordre du modèle. Cet article étend en particulier l'approche que nous avons développée précédemment dans [16, 17] au cas d'un noyau polynomial.

Soient \mathcal{H} un RKHS de fonctions à valeurs réelles sur un compact $\mathcal{U} \subset \mathbb{R}^r$, $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ son produit scalaire, et $\kappa : \mathcal{U} \times \mathcal{U} \rightarrow \mathbb{R}$ la fonction noyau associée à cet espace, vérifiant ainsi la propriété reproduisante $\psi(\mathbf{u}_i) = \langle \psi(\cdot), \kappa(\cdot, \mathbf{u}_i) \rangle_{\mathcal{H}}$ pour toute fonction $\psi(\cdot)$ de \mathcal{H} et pour tout $\mathbf{u}_i \in \mathcal{U}$. Soit (\mathbf{u}_i, d_i) le couple entrée-sortie disponible à l'instant i . On considère le problème de minimisation de l'erreur quadratique entre les n sorties du modèle en $\mathbf{u}_1, \dots, \mathbf{u}_n$ et les réponses désirées d_1, \dots, d_n :

$$\min_{\psi \in \mathcal{H}} \sum_{i=1}^n (d_i - \psi(\mathbf{u}_i))^2 + \zeta \|\psi\|_{\mathcal{H}}^2 \quad (1)$$

où $\zeta \|\psi\|_{\mathcal{H}}^2$ est un terme de régularisation avec $\zeta > 0$ [18]. Par le théorème de représentation [19], la solution de ce problème est de la forme :

$$\psi_n(\cdot) = \sum_{j=1}^n \alpha_{n,j} \kappa(\cdot, \mathbf{u}_j)$$

avec $\alpha_{n,j} \in \mathbb{R}$ sont les coefficients du modèle. Ceci illustre le fait que l'ordre du modèle est égal au nombre n d'entrées et, par suite, il s'avère inadéquat pour l'identification en ligne. Pour remédier à cet inconvénient, on considère un critère de parcimonie afin de sélectionner les fonctions noyau les plus

pertinentes. On appelle *dictionnaire* à l'instant n l'ensemble $\mathcal{D}_n = \{\kappa(\cdot, \mathbf{u}_{w_j})\}_{j=1}^m$ formé par ces fonctions sélectionnées, $w_j \in \{1, 2, \dots, n\}$ et m désigne l'ordre. Le modèle à l'instant n devient :

$$\psi_n(\cdot) = \sum_{j=1}^m \alpha_{n,j} \kappa(\cdot, \mathbf{u}_{w_j}). \quad (2)$$

Depuis la communication [20] présentée au Grets 2007, beaucoup de travaux se sont focalisés sur le critère de cohérence pour contrôler la parcimonie du modèle [10, 11, 12, 13, 14, 15]. L'importance du critère de cohérence réside dans son faible coût calculatoire, tout en conservant de très bonnes performances. De plus, on démontre dans [8] que la taille m du dictionnaire reste finie lorsque n tend vers l'infini. Le critère de cohérence est défini comme suit, pour un seuil de cohérence $\mu_0 \in [0, 1]$ fixé : à l'instant n , la fonction $\kappa(\cdot, \mathbf{u}_n)$ est introduite dans le dictionnaire si

$$\max_{i=1, \dots, m} \frac{|\kappa(\mathbf{u}_n, \mathbf{u}_{w_i})|}{\sqrt{\kappa(\mathbf{u}_n, \mathbf{u}_n)} \sqrt{\kappa(\mathbf{u}_{w_i}, \mathbf{u}_{w_i})}} \leq \mu_0; \quad (3)$$

Dans le cas contraire, le dictionnaire reste inchangé. Cette expression devient $\max_i |\kappa(\mathbf{u}_n, \mathbf{u}_{w_i})| \leq \mu_0$ pour les noyaux normalisés.

2 Adaptation du dictionnaire

Une fonction noyau introduite dans le dictionnaire y demeure même si la non-stationarité du système étudié rend sa contribution faible dans l'estimation de la sortie courante. Il apparaît alors opportun d'adapter les éléments du dictionnaire \mathcal{D}_n pour obtenir un dictionnaire amélioré \mathcal{D}_n^A . A chaque instant n , le problème d'estimation fonctionnelle se fait en 2 phases. La première consiste à trouver le vecteur des coefficients optimaux $\alpha_n = (\alpha_{n,1}, \dots, \alpha_{n,m})^t$ du modèle (2) en appliquant un algorithme d'identification en-ligne. La deuxième phase a pour but d'adapter les éléments du dictionnaire sans que ce dernier ne perde en cohérence au sens de la règle (3).

L'objectif de l'adaptation des éléments du dictionnaire est la minimisation de l'erreur quadratique instantanée e_n^2 avec $e_n = d_n - \sum_{i=1}^m \alpha_{n,i} \kappa(\mathbf{u}_n, \mathbf{u}_{w_i})$, et ceci sans enfreindre la règle de cohérence fixée par (3). A l'instant n , le $i^{\text{ème}}$ élément du dictionnaire est adapté selon le schéma : $\mathbf{u}_{w_i}^A = \mathbf{u}_{w_i} - \nu_n \mathbf{g}_{w_i}$ où $\mathbf{u}_{w_i}^A$ est l'élément du dictionnaire après adaptation, \mathbf{g}_{w_i} est le gradient de l'erreur quadratique instantanée par rapport à \mathbf{u}_{w_i} et ν_n indique le pas de gradient. Evidemment, le choix de ν_n n'est pas arbitraire mais il est soumis à des contraintes strictes dépendent explicitement du noyau utilisé.

2.1 Cas de noyaux radiaux

Un noyau radial est de la forme

$$\kappa(\mathbf{u}_{w_i}, \mathbf{u}_{w_j}) = f(\|\mathbf{u}_{w_i} - \mathbf{u}_{w_j}\|^2)$$

où $f \in \mathcal{C}^\infty$, comme par exemple le noyau Gaussien avec $f(\xi) = \exp(\frac{-1}{2\sigma^2}\xi)$. Une condition suffisante pour que ce

noyau soit défini positif est sa monotonie, c'est à dire : $(-1)^k f^{(k)}(\xi) \geq 0, \forall \xi \geq 0$, $f^{(k)}(\cdot)$ indique la $k^{\text{ème}}$ dérivée de la fonction $f(\cdot)$ [21]. Après adaptation, le dictionnaire doit rester cohérent ce qui se traduit par :

$$\begin{aligned} f(\|\mathbf{u}_{w_i}^A - \mathbf{u}_{w_j}^A\|^2) &\leq \mu_0 \Rightarrow \\ f(\|\delta\mathbf{u} - \nu_n \delta\mathbf{g}\|^2) &\leq \mu_0 \quad \forall i \neq j = 1 \dots m \end{aligned}$$

où $\delta\mathbf{u} = \mathbf{u}_{w_i} - \mathbf{u}_{w_j}$ et $\delta\mathbf{g} = \mathbf{g}_{w_i} - \mathbf{g}_{w_j}$. En développant en série de Taylor le terme de gauche de cette inégalité autour de $\nu_n \sim 0$:

$$\begin{aligned} -2\|\delta\mathbf{g}\|^2 f^{(1)}(\|\delta\mathbf{u}\|^2) \nu_n^2 + \\ 2\delta\mathbf{u}^t \delta\mathbf{g} f^{(1)}(\|\delta\mathbf{u}\|^2) \nu_n + \mu_0 - f(\|\delta\mathbf{u}\|^2) &\geq 0 \end{aligned} \quad (4)$$

Afin d'adapter tous les éléments du dictionnaire, le pas du gradient ν_n est déterminé selon la valeur du discriminant Δ de (4), tout en considérant la monotonie du noyau. Si $\Delta < 0$, il n'y a pas de racines réelles et, en conséquence, il n'y a pas de contrainte sur le choix du pas. En revanche, si $\Delta \geq 0$ on obtient deux racines $\nu_{i,j-}$ et $\nu_{i,j+}$ qui sont de même signe car le terme constant du polynôme (4) est positif vu que le dictionnaire est toujours cohérent. Ces deux racines sont données par :

$$\nu_{i,j\pm} = \frac{-\delta\mathbf{u}^t \delta\mathbf{g} f^{(1)}(\|\delta\mathbf{u}\|^2) \pm \sqrt{\Delta}}{-\|\delta\mathbf{g}\|^2 f^{(1)}(\|\delta\mathbf{u}\|^2)}$$

Le domaine des valeurs admissibles de ν_n est $] -\infty, \nu_{i,j-}] \cup [\nu_{i,j+}, +\infty[$ et la valeur $\nu_n = 0$ appartient toujours à cet intervalle puisque, dans ce cas, il n'y a pas adaptation du dictionnaire et qu'il est cohérent par construction. Le choix d'un pas convenable commence par la sélection d'un pas de référence ν_0 . Après le calcul de $m(m+1)/2$ intervalles $[\nu_{i,j-}, \nu_{i,j+}]$ entre les m éléments du dictionnaire, ν_n est choisi selon la procédure décrite dans la Table 1.

2.2 Cas de noyaux projectifs : le noyau polynomial

Contrairement au cas précédent, le cas des noyaux projectifs ne peut être traité globalement. On s'intéresse donc dans la suite au noyau polynomial, défini par

$$\kappa(\mathbf{u}_{w_i}, \mathbf{u}_{w_j}) = f(\mathbf{u}_{w_i}^t \mathbf{u}_{w_j}) = (1 + \mathbf{u}_{w_i}^t \mathbf{u}_{w_j})^\beta$$

où $\beta > 0$ désigne son degré. On observe à présent qu'il est difficile d'adapter tous les éléments du dictionnaire à chaque itération, étant donné la présence d'un dénominateur dans la condition (3), le noyau n'étant plus normalisé. Par conséquent on propose d'adapter un seul élément du dictionnaire à chaque itération. L'élément choisi pour l'adaptation est celui qui correspond à la plus grande norme du gradient de l'erreur quadratique instantanée. D'autres heuristiques sont bien sûr possibles. En supposant que \mathbf{u}_{w_i} est l'élément à actualiser, la condition (3) devient :

$$\frac{(1 + \langle \mathbf{u}_{w_i}^A, \mathbf{u}_{w_j} \rangle)^2}{(1 + \|\mathbf{u}_{w_i}^A\|^2)(1 + \|\mathbf{u}_{w_j}\|^2)} \leq (\mu_0)^{2/\beta} \quad \forall j \neq i, j = 1 \dots m$$

<p>* Si $\max_{i,j} \nu_{i,j+} \leq 0 \Rightarrow \nu_n = \nu_0$</p> <p>* Si $0 \leq \min_{i,j} \nu_{i,j-} \leq \nu_0 \Rightarrow \nu_n = \min_{i,j} \nu_{i,j-}$</p> <p>* Si $0 \leq \nu_0 \leq \min_{i,j} \nu_{i,j-} \Rightarrow \nu_n = \nu_0$</p> <p>* Si $0 \leq \min_{i,j} (\nu_{i,j-})^+ \leq \nu_0 \Rightarrow \nu_n = \min_{i,j} (\nu_{i,j-})^+$</p> <p>* Si $0 \leq \nu_0 \leq \min_{i,j} (\nu_{i,j-})^+ \Rightarrow \nu_n = \nu_0$</p> <p>où $(\nu_{i,j-})^+$ indique les valeurs positives des $\nu_{i,j-}$</p>	<p>cas : $A > 0$ et $\Delta > 0$</p> <p>* Si $\nu_{i,j-} \leq \nu_{i,j+} \leq 0 \Rightarrow \nu_{ijn} = \nu_0$</p> <p>* Si $0 \leq \nu_{i,j-} \leq \nu_0 \leq \nu_{i,j+} \Rightarrow \nu_{ijn} = \nu_{i,j-}$</p> <p>* Si $0 \leq \nu_{i,j-} \leq \nu_{i,j+} \leq \nu_0 \Rightarrow \nu_{ijn} = \nu_0$ ou $\nu_{ijn} = \nu_{i,j-}$</p> <p>* Si $0 \leq \nu_0 \leq \nu_{i,j-} \leq \nu_{i,j+} \Rightarrow \nu_{ijn} = \nu_0$ ou $\nu_{ijn} = \nu_{i,j-}$</p> <p>cas : $A > 0$ et $\Delta \leq 0 \Rightarrow$ on prend $\nu_{ijn} = \nu_0$</p> <p>cas : $A < 0$ et $\Delta = 0 \Rightarrow$ on prend $\nu_{ijn} = 0$</p>	<p>cas : $A < 0$ et $\Delta > 0$</p> <p>* Si $\nu_{i,j+} \leq \nu_0 \Rightarrow \nu_{ijn} = \nu_{i,j+}$</p> <p>* Si $\nu_0 \leq \nu_{i,j+} \Rightarrow \nu_{ijn} = \nu_0$ ou $\nu_{ijn} = \nu_{i,j+}$</p> <p>cas : $A = 0$</p> <p>* Si $\frac{-C}{2B} \leq 0 \Rightarrow \nu_{ijn} = \nu_0$</p> <p>* Si $0 \leq \frac{C}{2B} \leq \nu_0 \Rightarrow \nu_{ijn} = \nu_0$</p> <p>* Si $0 \leq \frac{-C}{2B} \leq \nu_0 \Rightarrow \nu_{ijn} = \frac{-C}{2B}$</p> <p>* si $\nu_0 \leq \frac{-C}{2B} \Rightarrow \nu_{ijn} = \frac{-C}{2B}$</p>
---	---	---

TABLE 1 – Cas du noyau Gaussien

TABLE 2 – Règle d'adaptation : cas du noyau polynomial

autrement

$$\begin{aligned} & ((\mathbf{u}_{w_j}^t \mathbf{g}_{w_i})^2 - D \|\mathbf{g}_{w_i}\|^2) \nu_n^2 - \\ & 2(\kappa_{ij}^* \mathbf{u}_{w_j}^t \mathbf{g}_{w_i} - D \mathbf{u}_{w_i}^t \mathbf{g}_{w_i}) \nu_n + (\kappa_{ij}^{*2} - D \kappa_{ii}^*) \leq 0 \end{aligned} \quad (5)$$

où $\kappa_{ij}^* = \kappa^*(\mathbf{u}_{w_i}, \mathbf{u}_{w_j}) = (1 + \mathbf{u}_{w_i}^t \mathbf{u}_{w_j})$ et $D = \kappa_{jj}^* (\mu_0)^{2/\beta}$. Cette expression est de la forme : $A\nu_n^2 + 2B\nu_n + C \geq 0$ avec

$$\begin{aligned} A &= D \|\mathbf{g}_{w_i}\|^2 - (\mathbf{u}_{w_j}^t \mathbf{g}_{w_i})^2 \\ B &= \kappa_{ij}^* \mathbf{u}_{w_j}^t \mathbf{g}_{w_i} - D \mathbf{u}_{w_i}^t \mathbf{g}_{w_i} \\ C &= D \kappa_{ii}^* - \kappa_{ij}^{*2} \end{aligned}$$

avec $C \geq 0$ pour un dictionnaire cohérent. Si les racines du polynôme de l'inégalité (5) existent dans \mathbb{R} , l'intervalle des valeurs acceptables de ν_n dépend du signe de A . Si $A < 0$ on a $\nu_n \in [\nu_{i,j-}, \nu_{i,j+}]$ tel que $\nu_{i,j-} \leq 0 \leq \nu_{i,j+}$ car 0 appartient toujours à l'intervalle des valeurs admissibles de ν_n . Si $A > 0$ on a $\nu_n \in]-\infty, \nu_{i,j-}] \cup [\nu_{i,j+}, +\infty[$ et pour $A = 0$ on a une seule racine. Le choix du pas de gradient, afin d'adapter un seul élément du dictionnaire, est détaillé dans la Table 2.

3 Expérimentations

Dans cet article, et par faute d'espace, on ne considère que deux exemples pour illustrer la pertinence de l'adaptation du dictionnaire. La taille finale m ou bien la taille moyenne \bar{m} du dictionnaire, tout comme l'erreur quadratique moyenne normalisée (Normalized Mean Squared Error - NMSE), sont systématiquement utilisées comme critères de performance pour réaliser les comparaisons. Rappelons que l'erreur quadratique moyenne normalisée est calculée en utilisant les derniers 500 échantillons du signal utilisé selon la relation :

$$\text{NMSE} = E \left\{ \frac{\sum_{n=N-500}^N (d_n - \psi_n(\mathbf{u}_n))^2}{\sum_{n=N-500}^N d_n^2} \right\} \quad (6)$$

avec N est la taille de l'échantillon.

Dans le premier exemple, on prédit la série temporelle de Hénon $d_n = 1 - \gamma_1 d_{n-1}^2 + \gamma_2 d_{n-2}$ avec les conditions initiales $d_0 = -0.3$, $d_1 = 0$ et les paramètres $\gamma_1 = 1.4$ et $\gamma_2 = 0.3$. Le modèle est de la forme $\psi_n(d_{n-1}, d_{n-2})$ et utilise une suite de 2000 échantillons. Comme préconisé dans [20] avec l'algorithme KRLS (moindres carrés récursif à noyau), le noyau Gaussien de bande passante $\sigma = 0.35$ est utilisé ainsi qu'un seuil de cohérence $\mu_0 = 0.6$. Les résultats de la simulation sont

montrés dans la Figure 1, où le pas de gradient de référence est fixé à $\nu_0 = 0.05$. La taille du dictionnaire a diminué de 17 éléments sans adaptation à 16 éléments avec adaptation. Malgré une faible réduction de cette taille, on observe un gain important sur l'erreur quadratique moyenne normalisée. Celle-ci chute de 0.01036 sans adaptation à 0.001628 avec adaptation ce qui correspond à une diminution de 84.29%.

Le deuxième exemple, emprunté de [8], est défini par la série temporelle : $d_n = s_n^2$, avec $s_n = 1.1 \exp(-|s_{n-1}|) + u_n$, où u_n et d_n sont respectivement l'entrée et la sortie désirée du modèle à l'instant n . La condition initiale est $s_0 = 0.5$ et les entrées u_n sont générées selon une loi normale de moyenne nulle et d'écart type 0.25. La sortie du modèle est noyée dans un bruit blanc Gaussien de variance unité. Comme préconisé dans [8], l'algorithme KAPA (algorithme de projection affine à noyau) [22] est utilisé avec les paramètres suivants : taille de la fenêtre court-terme $p = 3$, paramètre du contrôle du pas $\eta = 0.01$ et facteur de régularisation $\zeta = 0.07$. Un ensemble de 200 séries chronologiques de 3000 échantillons chacune est utilisé pour comparer les résultats de simulation sans et avec adaptation. Le modèle est de la forme $\psi_n(u_n, u_{n-1})$, avec un noyau quadratique (polynomial de degré $\beta = 2$). Les résultats de simulation sont montrés dans la Figure 2. L'analyse des résultats de simulation nous permet de faire les conclusions suivantes :

- Pour le même seuil de cohérence, et après adaptation, une augmentation de \bar{m} de 76.23% mène à une diminution de 4.246% du NMSE.
- Pour la même taille moyenne du dictionnaire (\bar{m}), et en changeant le seuil de cohérence, après adaptation on obtient une réduction du NMSE de 67.17%.
- Pour un NMSE comparable, en faisant varier le seuil de cohérence et le pas du gradient, la taille moyenne du dictionnaire \bar{m} a diminué de 13.38%.

4 Conclusion et perspectives

Les algorithmes adaptatifs à noyau ne gèrent généralement pas la modification des éléments du dictionnaire, même si ceux-ci deviennent obsolètes du fait d'une non-stationnarité. Dans cet article, nous avons présenté une méthode pour l'adaptation des éléments du dictionnaire. Notre méthode de gradient stochastique permet itérativement à la fois l'introduction et l'adaptation des éléments de dictionnaire tout en respectant le critère de cohérence.

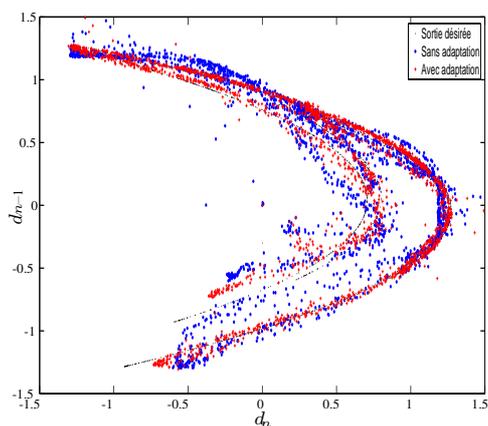


FIGURE 1 – Série de Hénon (noyau Gaussien).

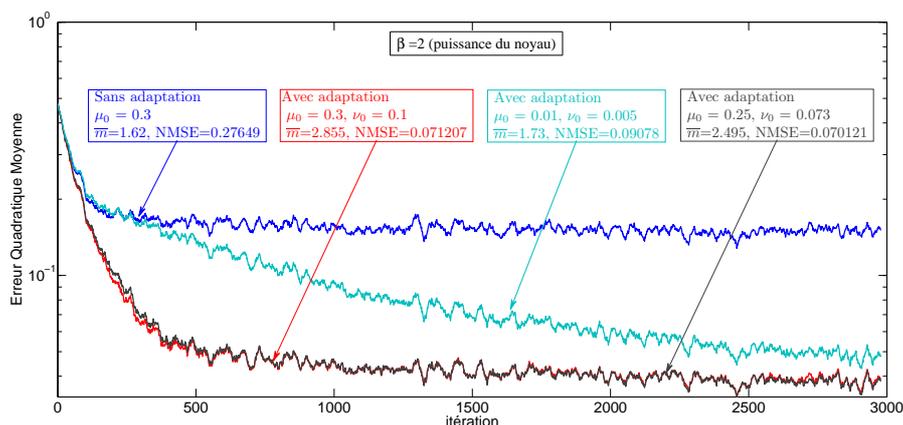


FIGURE 2 – Comparaison avec et sans adaptation (noyau polynomial de degré $\beta = 2$).

Nous avons présenté le cas d'un noyau radial et le cas d'un noyau projectif (polynomial). Nous avons également montré qu'il n'y a pas de méthode universelle pour l'adaptation du dictionnaire, ce qui signifie que chaque type de noyau nécessite une étude spécifique. Nous avons montré que, grâce à l'adaptation des éléments du dictionnaire, l'erreur d'estimation peut être considérablement réduite (pour une taille finale similaire de dictionnaire), ou le taille finale du dictionnaire peut être réduite (pour une erreur d'estimation résiduelle similaire). Évidemment, un compromis peut être adopté entre ces deux possibilités extrêmes. Parmi les extensions possibles de ce travail, nous tâcherons de réduire la complexité calculatoire qui est fortement liée au nombre d'éléments du dictionnaire.

Références

- [1] M. Schetzen. "The Volterra and Wiener Theories of Nonlinear Systems". Krieger Publishing Co., Inc., Melbourne, FL, USA, 2006.
- [2] T. Ogunfunmi. "Adaptive Nonlinear System Identification : The Volterra and Wiener Model Approaches". Springer, 2007.
- [3] S. Haykin. "Neural Networks and Learning Machines". Prentice Hall - 3rd edition, November 2008.
- [4] I. Constantin, C. Richard, R. Lengellé et L. Soufflet. "Nonlinear regularized Wiener filtering with kernels : Application in denoising MEG data corrupted by ECG". *Signal Processing, IEEE Transactions on*, vol 54, pp. 4796-4806, 2006.
- [5] P. Honeine, C. Richard, et J. C. M. Bermudez. "On-line nonlinear sparse approximation of functions". *Information Theory, 2007. ISIT 2007. IEEE International Symposium on*, pp. 956-960, 2007.
- [6] T.J. Dodd, V. Kadiramanathan et R. F. Harrison. "Function estimation in Hilbert space using sequential projections". *Conference on Intelligent Control Systems and Signal Processing*, pp.113-118, 2003.
- [7] Y. Engel, S. Mannor et R. Meir. "The kernel recursive least-squares algorithm". *IEEE Transactions on Signal Processing*, vol. 52, pp. 2275-2285, Aug. 2004.
- [8] C. Richard, J. C. M. Bermudez et P. Honeine. "Online Prediction of Time Series Data With Kernels". *IEEE Transactions on Signal Processing*, vol. 57, pp. 1058-1067, Mar. 2009.
- [9] P. Honeine. "Online kernel principal component analysis : a reduced-order model". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, issue 9, pp. 1814-1826, September 2012.
- [10] Z. Noumir, P. Honeine, and C. Richard. "One-class machines based on the coherence criterion". In *Proc. IEEE workshop on Statistical Signal Processing*, Ann Arbor, Michigan, USA, 5-8 August 2012.
- [11] Z. Noumir, P. Honeine, and C. Richard. "Online one-class machines based on the coherence criterion". In *Proc. 20th European Conference on Signal Processing*, Bucharest, Romania, 27-31 August 2012.
- [12] P. Honeine, C. Richard, J. C. M. Bermudez, H. Snoussi, M. Essoloh, and F. Vincent. "Functional estimation in hilbert space for distributed learning in wireless sensor networks." In *Proc. 34th IEEE International Conference on Acoustics, Speech and Signal Processing*, Taipei, Taiwan, April 2009.
- [13] P. Honeine, C. Richard, J. C. M. Bermudez, and H. Snoussi. "Distributed prediction of time series data with kernels and adaptive filtering techniques in sensor networks." In *Proc. 42nd Annual ASILOMAR Conference on Signals, Systems and Computers*, Pacific Grove, CA, USA, October 2008.
- [14] P. Honeine, M. Essoloh, C. Richard, and H. Snoussi. "Distributed regression in sensor networks with a reduced-order kernel model." In *Proc. 51st IEEE GLOBECOM Global Communications Conference*, New Orleans, LA, USA, 2008.
- [15] P. Honeine, C. Richard, and J. C. M. Bermudez. "On-line nonlinear sparse approximation of functions". In *Proc. IEEE International Symposium on Information Theory*, pp. 956-960, Nice, France, June 2007.
- [16] C. Saïdé, R. Lengellé, P. Honeine, C. Richard, et R. Achkar. "Dictionary adaptation for online prediction of time series data with kernels". *Proc. IEEE workshop on Statistical Signal Processing*, Ann Arbor, Michigan, USA, 5-8 August 2012.
- [17] C. Saïdé, R. Lengellé, P. Honeine, and R. Achkar. "Online kernel adaptive algorithms with dictionary adaptation for mimo models." *IEEE Signal Processing Letters*, vol. 20, issue 5, pp. 535-538, May 2013.
- [18] P. Honeine. "Méthodes à noyau pour l'analyse et la décision en environnement non-stationnaire". *Mémoire de thèse de doctorat en Optimisation et Sécurité des Systèmes - Ecole doctorale SSTO - UTT*. Troyes, France, 2007.
- [19] G. Kimeldorf, G. Wahba. "Some results on Tchebycheffian spline functions". *Journal of Mathematical Analysis and Applications*, vol. 33, pp. 82-95, Jan. 1971.
- [20] P. Honeine, C. Richard, and J. C. M. Bermudez. "Modélisation parcimonieuse non linéaire en ligne par une méthode à noyau reproduisant et un critère de cohérence". *Actes du XXI-ème Colloque GRETSI sur le Traitement du Signal et des Images*, Troyes, France, September 2007.
- [21] F. Cucker, S. Smale. "On the mathematical foundations of learning". *Bulletin of the American Mathematical Society*, vol. 39, pp. 1-49, 2002.
- [22] W. Liu, J. C. Principe et S. Haykin. "Kernel Adaptive Filtering : A Comprehensive Introduction". Wiley Publishing - 1st edition, 2010.