

# Modélisation parcimonieuse non linéaire en ligne par une méthode à noyau reproduisant et un critère de cohérence

Paul HONEINE<sup>1</sup>, Cédric RICHARD<sup>1</sup>, José Carlos M. BERMUDEZ<sup>2</sup>

<sup>1</sup>Institut Charles Delaunay (FRE CNRS 2848) - LM2S - Université de Technologie de Troyes  
12 rue Marie Curie, BP 2060, 10010 Troyes cedex - fax. 03.25.71.56.99

<sup>2</sup>Department of Electrical Engineering, Federal University of Santa Catarina  
88040-900, Florianópolis, SC - Brazil - fax. +55.48.3331.9280  
paul.honeine@utt.fr, cedric.richard@utt.fr, j.bermudez@ieee.org

**Résumé** – Cet article traite du problème d'identification de systèmes non linéaires et non stationnaires par des méthodes à noyau reproduisant. Les approches de ce type nécessitent un contrôle en ligne de l'ordre du modèle considéré. Notre approche exploite le critère de cohérence, issu des techniques de décomposition parcimonieuse. Elle permet le contrôle de l'ordre du modèle à noyau reproduisant avec un coût calculatoire linéaire par rapport à celui-ci, contrairement aux techniques existantes qui sont à complexité quadratique. On illustre cette approche par un algorithme de moindres carrés récursif, avec des simulations sur des modèles synthétiques et réels.

**Abstract** – In this article, we consider non-linear and non-stationary system identification with kernel-based methods. Such techniques require model order control, both for sparse result and online applications. For this purpose, we consider the coherence criterion, initially introduced by the sparse decomposition community. This parameter gives new insights into the model properties, with a linear computational complexity with respect to its order, whereas that of classical techniques is quadratic. We provide a recursive-least-squares algorithm, and present some results on synthetic and real time series.

## 1 Introduction

Fondés sur les travaux précurseurs d'Aronszajn [1], les espaces de Hilbert à noyau reproduisant (RKHS, reproducing kernel Hilbert spaces) et les récentes avancées dans le domaine de l'estimation fonctionnelle ont apporté des réponses convaincantes aux problèmes d'identification de systèmes non linéaires. Ces méthodes ne sont toutefois pas parcimonieuses par essence, l'ordre des modèles étant égal au nombre de couples (entrée, sortie) utilisés. Afin de surmonter ce handicap annihilant toute possibilité de traitement en ligne, différentes techniques ont été proposées afin de contrôler cet ordre. Voir [2] et références. Toutefois, souvent assujetties à une inversion matricielle, ces approches peuvent s'avérer coûteuses en temps de calcul.

Dans cet article, on s'intéresse au contrôle de l'ordre de modèles à noyau dans le cadre d'une procédure d'identification de systèmes en ligne. On propose d'appliquer un critère directement inspiré de la littérature relative à l'approximation parcimonieuse de fonctions : la cohérence du dictionnaire utilisé. Au-delà des nombreuses propriétés de cette grandeur, que l'on revisite dans le contexte des méthodes à noyau, on montre que notre approche permet un contrôle en ligne de l'ordre du modèle avec un coût calculatoire linéaire par rapport à celui-ci, contrairement aux techniques existantes qui sont à complexité quadratique.

Cet article est organisé comme suit. Dans la prochaine section, on présente brièvement les fondements de la modélisation parcimonieuse dans un RKHS. On décrit

dans la section 3 un critère de cohérence pour le contrôle de l'ordre de modèles. Après avoir présenté un algorithme de moindres carrés récursifs à noyau pour l'estimation des paramètres des modèles, on illustre cette approche au travers de simulations sur des séries temporelles synthétiques et réelles.

## 2 Éléments sur la modélisation parcimonieuse dans un RKHS

Soient  $\mathcal{H}$  un RKHS de fonctions à valeurs réelles sur un compact  $\mathcal{U} \subset \mathbb{R}^p$ , et  $\langle \cdot, \cdot \rangle_{\mathcal{H}}$  son produit scalaire. On définit par  $\kappa: \mathcal{U} \times \mathcal{U} \rightarrow \mathbb{R}$  le noyau de 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  $\mathbf{u}_i$  de  $\mathcal{U}$ . Étant donné cet espace, on considère le problème de minimisation de l'erreur quadratique entre les  $n$  sorties du modèle  $\psi(\mathbf{u}_i)$  et les réponses désirées  $d_i$ . Ce problème d'optimisation s'écrit

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

où on a introduit un terme de régularisation tel que  $\eta > 0$ . Le Théorème de Représentation [3] établit que la solution de ce problème peut être recherchée dans l'espace engendré par les  $n$  fonctions noyau  $\kappa(\cdot, \mathbf{u}_j)$ , ce qui permet d'écrire  $\psi(\cdot) = \sum_{j=1}^n \alpha_j \kappa(\cdot, \mathbf{u}_j)$ . Ce résultat montre que l'ordre du modèle recherché correspond au nombre de données  $n$  disponibles, réduisant à néant toute perspective

d'application de ce résultat dans le cadre d'un traitement en ligne.

Dans un contexte d'identification en ligne de systèmes, on cherche à déterminer à l'itération  $n$  un modèle réduit de la forme

$$\psi_n(\cdot) = \sum_{\omega_j \in \mathcal{J}_n} \alpha_{n,j} \kappa(\cdot, \mathbf{u}_{\omega_j}), \quad (2)$$

où  $\mathcal{J}_n$  est un sous-ensemble de  $m$  indices de  $\{1, \dots, m\}$ . L'ordre du modèle est contrôlé explicitement par un choix approprié des fonctions noyau dans (2), parmi les  $n$  fonctions noyau des observations précédentes. En désignant par dictionnaire l'ensemble  $\mathcal{D}_m$  des  $m$  fonctions  $\kappa(\cdot, \mathbf{u}_{\omega_j})$  retenues, le problème du choix des éléments constituant ce dictionnaire se pose alors. Une technique courante consiste à retenir, à l'itération  $n$ , la fonction noyau  $\kappa(\cdot, \mathbf{u}_n)$  si elle ne peut être représentée par une combinaison linéaire des éléments du dictionnaire à une erreur  $\epsilon_0$  près, c'est-à-dire

$$\min_{\gamma} \|\kappa(\cdot, \mathbf{u}_n) - \sum_{\omega_j \in \mathcal{J}_{n-1}} \gamma_j \kappa(\cdot, \mathbf{u}_{\omega_j})\|_{\mathcal{H}}^2 > \epsilon_0^2. \quad (3)$$

Dans cette expression,  $\kappa$  est un noyau de norme unité tel que  $\kappa(\mathbf{u}_k, \mathbf{u}_k) = 1$  pour tout  $\mathbf{u}_k \in \mathcal{U}$ , sinon  $\kappa(\cdot, \mathbf{u}_k)$  est remplacé par  $\kappa(\cdot, \mathbf{u}_k) / \sqrt{\kappa(\mathbf{u}_k, \mathbf{u}_k)}$  ci-dessus. Dans le cas où la règle (3) n'est pas vérifiée,  $\kappa(\cdot, \mathbf{u}_n)$  n'est pas ajoutée au dictionnaire. Le seuil  $\epsilon_0$  conditionne le niveau de parcimonie du modèle. Cette approche s'avère très coûteuse en temps de calcul car elle nécessite l'inversion d'une matrice de taille  $m \times m$ . En effet, la complexité de cette opération est en  $\mathcal{O}(m^2)$  si elle est effectuée récursivement. L'objet de cet article est de proposer une stratégie de complexité  $\mathcal{O}(m)$  et d'établir un lien formel avec la règle (3).

### 3 Critère de cohérence

La cohérence est un paramètre principalement utilisé dans le cadre des techniques d'approximation parcimonieuse de fonctions, et permet de caractériser le dictionnaire adopté [4]. Adaptée à un dictionnaire de  $m$  fonctions noyau  $\kappa(\cdot, \mathbf{u}_{\omega_1}), \dots, \kappa(\cdot, \mathbf{u}_{\omega_m})$  de norme unité, elle est définie par

$$\mu = \max_{i=j} |\langle \kappa(\cdot, \mathbf{u}_{\omega_i}), \kappa(\cdot, \mathbf{u}_{\omega_j}) \rangle_{\mathcal{H}}| = \max_{i=j} |\kappa(\mathbf{u}_{\omega_i}, \mathbf{u}_{\omega_j})|,$$

pour tout  $i, j = 1, \dots, m$ . La cohérence correspond donc à la plus grande corrélation entre deux éléments du dictionnaire et, en conséquence, est nulle pour toute base orthonormale. Plutôt que de résoudre le problème quadratique (3), nous proposons ici d'introduire  $\kappa(\cdot, \mathbf{u}_n)$  dans le dictionnaire si la cohérence de celui-ci, après ajout, ne dépasse pas un seuil  $\mu_0 \in [0, 1]$  donné, à savoir

$$\max_{\omega_j \in \mathcal{J}_{n-1}} |\kappa(\mathbf{u}_n, \mathbf{u}_{\omega_j})| \leq \mu_0. \quad (4)$$

Le niveau de parcimonie est alors déterminé par  $\mu_0$ . Bien qu'il s'agisse d'un critère simple à évaluer puisque de complexité calculatoire linéaire en  $m$ , il vérifie un certain nombre de propriétés fondamentales.

On montre par exemple que les  $m$  fonctions noyau d'un dictionnaire  $\mu_0$ -cohérent sont linéairement indépendantes

si on a  $(m-1)\mu_0 < 1$ . Pour démontrer cette condition suffisante, on a recouru à un résultat très connu en algèbre linéaire, qui est qu'une matrice à diagonale dominante est non singulière. Une matrice est dite à diagonale dominante si chaque élément de la diagonale est supérieur, en valeur absolue, à la somme des valeurs absolues des autres éléments de la même ligne. Dans le cas de la matrice de Gram associée au dictionnaire, ceci se traduit par  $|\kappa(\mathbf{u}_j, \mathbf{u}_j)| > \sum_{i=j}^m |\kappa(\mathbf{u}_i, \mathbf{u}_j)|$ , pour tout  $j = 1, \dots, m$ . Pour un noyau de norme unité, on obtient la condition suivante :  $(m-1)\mu_0 < 1$ .

Dans un article précédent sur le filtrage adaptatif non linéaire [2], on a démontré d'autres propriétés de dictionnaires issus du critère de cohérence, notamment :

- Parce que  $\mathcal{U}$  est compact par hypothèse, le dictionnaire ainsi obtenu est de taille finie. Ceci implique que l'ordre du modèle  $\psi_n(\cdot)$  est fini et dépend de  $\mu_0$ .
- Tout dictionnaire déterminé à l'aide de la règle de cohérence vérifie le critère d'approximation linéaire (3) avec  $\epsilon_0^2 > 1 - (m-1)\mu_0^2 / (1 - (m-1)\mu_0)$ . Ainsi garantit-on, avec notre approche, un résultat équivalent avec un coût calculatoire moindre.
- L'entropie quadratique de Rényi du dictionnaire est bornée inférieurement par une fonction croissante de  $\mu_0$  et décroissante de  $m$ . Pour un noyau de norme unité, cette borne vaut  $C_0 - \log((1 + (m-1)\mu_0)/m)$  avec  $C_0$  une constante dépendant du noyau.

### 4 Méthode de moindres carrés récurrente a noyau

Dans ce qui suit, on considère la mise en œuvre du critère de cohérence dans le cadre d'une technique de modélisation récursive non linéaire par moindres carrés, à partir du modèle (2). On montre que le problème d'optimisation (1) s'écrit sous la forme  $\min_{\alpha} \|\mathbf{d}_n - \mathbf{H}_n \alpha\|^2 + \eta \alpha^t \mathbf{K}_n \alpha$ , où  $\mathbf{K}_n$  est la matrice de Gram du dictionnaire, c'est-à-dire  $\mathbf{K}_n(i, j) = \kappa(\mathbf{u}_{\omega_i}, \mathbf{u}_{\omega_j})$ , et  $\mathbf{H}_n$  la matrice de taille  $(n \times m)$  telle que  $\mathbf{H}_n(i, j) = \kappa(\mathbf{u}_i, \mathbf{u}_{\omega_j})$ . A l'itération  $n$ , en supposant que  $\mathbf{P}_n = (\mathbf{H}_n^t \mathbf{H}_n + \eta \mathbf{K}_n)^{-1}$  existe, la solution de ce problème s'écrit  $\alpha_n = \mathbf{P}_n \mathbf{H}_n^t \mathbf{d}_n$ . A l'itération  $n+1$ , étant donné une nouvelle observation  $\kappa(\cdot, \mathbf{u}_{n+1})$ , deux cas peuvent se présenter selon que la règle de cohérence (4) est satisfaite ou pas.

**Premier cas :**  $\max_{j=1, \dots, m} |\kappa(\mathbf{u}_{n+1}, \mathbf{u}_{\omega_j})| > \mu_0$

Ce test indique que la fonction noyau candidate  $\kappa(\cdot, \mathbf{u}_{n+1})$  peut être représentée convenablement par les éléments du dictionnaire, et pour cette raison ne nécessite pas d'être ajoutée à celui-ci. On complète la matrice des observations  $\mathbf{H}_n$  avec la ligne supplémentaire  $\mathbf{h}_{n+1} = [\kappa(\mathbf{u}_{n+1}, \mathbf{u}_{\omega_1}) \cdots \kappa(\mathbf{u}_{n+1}, \mathbf{u}_{\omega_m})]^t$ , et le vecteur des sorties  $\mathbf{d}_n$  avec la dernière sortie désirée  $d_{n+1}$ . La mise à jour des coefficients du modèle s'effectue alors selon

$$\alpha_{n+1} = \alpha_n + \frac{\mathbf{P}_n \mathbf{h}_{n+1}}{1 + \mathbf{h}_{n+1}^t \mathbf{P}_n \mathbf{h}_{n+1}} (d_{n+1} - \mathbf{h}_{n+1}^t \alpha_n)$$

avec

$$\mathbf{P}_{n+1} = \mathbf{P}_n - \frac{\mathbf{P}_n \mathbf{h}_{n+1} \mathbf{h}_{n+1}^t \mathbf{P}_n}{1 + \mathbf{h}_{n+1}^t \mathbf{P}_n \mathbf{h}_{n+1}}$$

**Second cas :**  $\max_{j=1, \dots, m} |\kappa(\mathbf{u}_{n+1}, \mathbf{u}_{\omega_j})| \leq \mu_0$

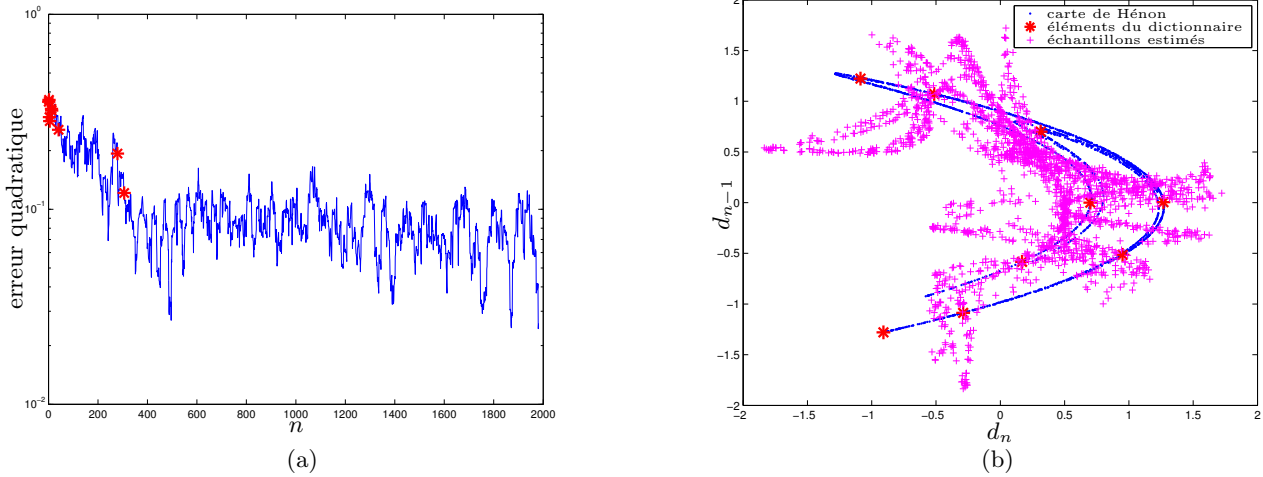


FIG. 1 – Evolution de l’erreur quadratique (a) entre la sortie désirée  $d_n$  et celle estimée  $\psi(d_{n-1}, d_{n-2})$ , pour  $\mu_0 = 0.1$ . Les deux séries sont comparées en (b) dans le plan de phase. Les éléments du dictionnaire sont indiqués par des étoiles.

Dans ce cas,  $\kappa(\cdot, \mathbf{u}_{n+1})$  ne peut être représentée convenablement par les éléments du dictionnaire et est donc insérée dans celui-ci. Il s’en suit une augmentation de l’ordre du modèle, qui devient égal à  $m + 1$ . On montre que la mise à jour des paramètres s’effectue en deux temps. On applique tout d’abord les expressions présentées dans le premier cas ci-dessus afin d’obtenir ce que l’on note  $\tilde{\alpha}_{n+1}$  et  $\tilde{\mathbf{P}}_{n+1}$ , puis on utilise les expressions de mise à jour suivantes

$$\alpha_{n+1} = \begin{bmatrix} \tilde{\alpha}_{n+1} \\ 0 \end{bmatrix} + \frac{d_{n+1} - \mathbf{h}_{n+1}^t \tilde{\alpha}_{n+1}}{1 - \mathbf{h}_{n+1}^t \tilde{\mathbf{P}}_{n+1} \mathbf{h}_{n+1}} \begin{bmatrix} -\tilde{\mathbf{P}}_{n+1} \mathbf{h}_{n+1} \\ 1/h_0 \end{bmatrix}$$

$$\mathbf{P}_{n+1} = \begin{bmatrix} \tilde{\mathbf{P}}_{n+1} & \mathbf{0}_n \\ \mathbf{0}_n^t & 0 \end{bmatrix} + \frac{1}{1 - \mathbf{h}_{n+1}^t \tilde{\mathbf{P}}_{n+1} \mathbf{h}_{n+1}} \times \begin{bmatrix} -\tilde{\mathbf{P}}_{n+1} \mathbf{h}_{n+1} \\ 1/h_0 \end{bmatrix} \begin{bmatrix} -(\tilde{\mathbf{P}}_{n+1} \mathbf{h}_{n+1})^t & 1/h_0 \end{bmatrix},$$

où  $h_0 = \kappa(\mathbf{u}_{n+1}, \mathbf{u}_{n+1})$ .

## 5 Expérimentations

Dans une première expérimentation, nous nous intéressons au système non linéaire de Hénon décrit par l’équation

$$d_n = 1 - \gamma_1 d_{n-1}^2 + \gamma_2 d_{n-2} \quad d_0 = -0.3 \quad d_1 = 0.$$

Pour  $\gamma_1 = 1.4$  et  $\gamma_2 = 0.3$ , la série temporelle présente un comportement chaotique. On souhaite la modéliser à partir de 2000 échantillons, par une expression non-linéaire de la forme  $d_n = \psi(d_{n-1}, d_{n-2})$ . Après avoir testé plusieurs noyaux, dont on a ajusté les paramètres par validation croisée, nous avons observé que le noyau Gaussien  $\kappa(\mathbf{u}_i, \mathbf{u}_j) = \exp(-\|\mathbf{u}_i - \mathbf{u}_j\|^2 / 2\beta_0^2)$  avec  $\beta_0 = 0.35$  menait à des performances satisfaisantes en terme d’erreur de prédiction. Dans un premier temps, nous avons fixé le seuil de cohérence à  $\mu_0 = 0.1$ , ce qui a conduit à un dictionnaire de  $m = 9$  éléments vérifiant (4). On note que le dictionnaire obtenu satisfait à la condition suffisante d’indépendance linéaire, soit  $\mu_0 < 1/(m - 1)$ . L’évolution de l’erreur quadratique de prédiction est représentée à la

Fig. 1(a), où l’on indique par des étoiles les éléments retenus pour le dictionnaire. On présente en Fig. 1(b) la carte de Hénon, c’est-à-dire les données dans le plan  $(d_n, d_{n-1})$ , les prédictions ainsi que les éléments du dictionnaire. Si ces derniers semblent choisis de manière pertinente sur les paraboles marquant les lieux de Hénon, on remarque que les prédictions s’écartent sensiblement de ce qui est attendu. Afin d’améliorer les résultats, nous avons choisi d’augmenter la taille du dictionnaire en rehaussant le seuil de cohérence  $\mu_0$  à 0.6. Bien que la condition suffisante d’indépendance linéaire ne soit plus vérifiée, on note une amélioration des performances grâce aux 52 éléments retenus pour le dictionnaire. En Fig. 2, on représente (a) l’évolution de l’erreur quadratique résultante, et (b) les cartes de Hénon des observations et de leur prédiction. On aboutit à une erreur de prédiction plus faible, au prix d’un ordre de modèle plus élevé, donc d’une complexité plus importante.

Dans une seconde application, on considère un signal électrocardiographique (ECG) de 3000 échantillons correspondant à environ 10 cycles cardiaques. Dans cette expérimentation, on propose d’utiliser le noyau de Laplace  $\kappa(\mathbf{u}_i, \mathbf{u}_j) = \exp(-\|\mathbf{u}_i - \mathbf{u}_j\|/\beta_0)$  avec  $\beta_0 = 0.35$ , en fixant le seuil de cohérence à  $\mu_0 = 0.9$ . Dans un premier temps, notre approche consiste à identifier un modèle autorégressif de ce signal de la forme  $d_n = \psi(d_{n-1}, d_{n-2})$ . Ceci produit un dictionnaire de  $m = 9$  éléments. On représente à la Fig. 3(a) en haut, le signal ECG et le signal obtenu selon le modèle. Au-delà d’une erreur de convergence élevée pour les premiers 500 échantillons, un décrochage pour les ondes négatives de l’ECG est visible. Dans un second temps, on s’intéresse à un modèle de la forme  $d_n = \psi(d_{n-1}, d_{n-2}, d_{n-3})$ . Ceci conduit à une augmentation de la taille du dictionnaire, qui est portée à  $m = 12$ . En étudiant les résultats obtenus dans la Fig. 3(a) en bas, on remarque que l’on modélise mieux le signal ECG que dans le premier cas. On illustre à la Fig 3(b) le signal ECG, les observations obtenues à partir du modèle  $\psi(d_{n-1}, d_{n-2}, d_{n-3})$  et les éléments du dictionnaire dans l’espace de phase, à savoir  $(d_n, d_{n-1}, d_{n-2})$ . Dans les deux cas, il faut noter que la condition suffisante d’indépendance linéaire n’est pas satisfaite.

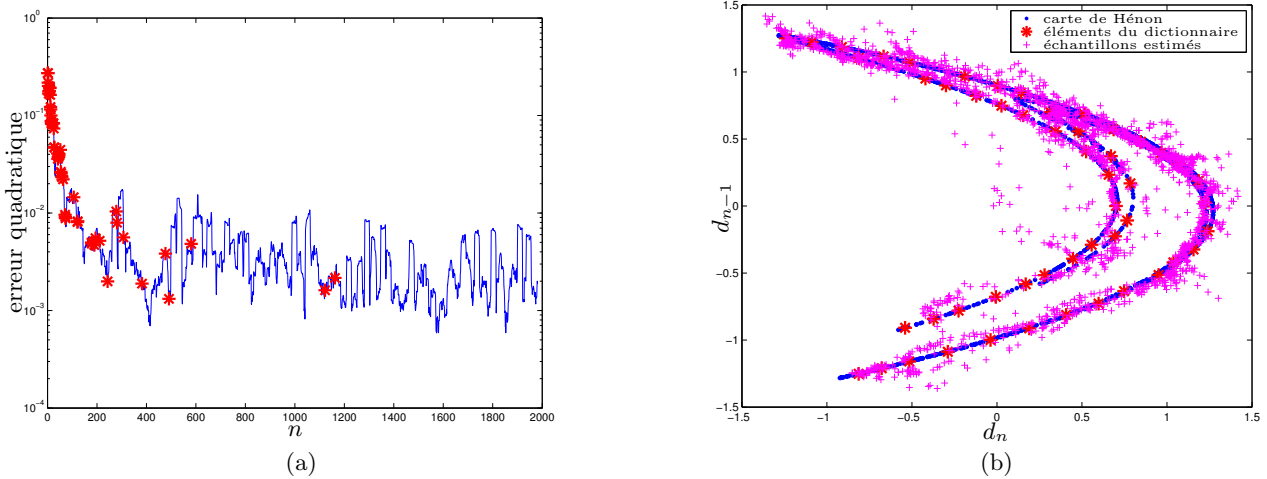


FIG. 2 – Résultats obtenus comme en figure 1 pour un seuil de cohérence  $\mu_0 = 0.6$

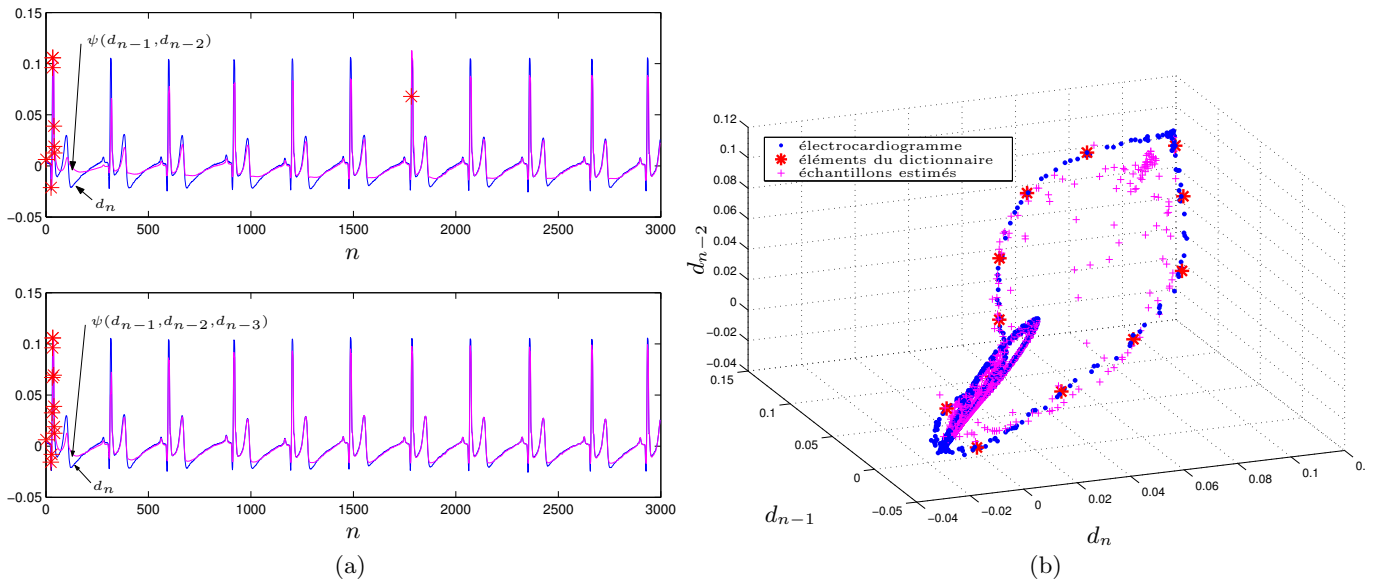


FIG. 3 – On représente en (a) le signal ECG à identifier, et le signal résultant du modèle retenu, de la forme  $\psi(d_{n-1}, d_{n-2})$  en haut et  $\psi(d_{n-1}, d_{n-2}, d_{n-3})$  en bas. L'espace de phase est illustré en (b) pour ce dernier cas. Les éléments du dictionnaire sont représentés par les étoiles.

## 6 Conclusion

Dans ce document, nous avons proposé une approche pour l'identification en ligne de systèmes non linéaires par des méthodes à noyau reproduisant. Afin de contrôler l'ordre du modèle résultant, nous avons recouru au critère de cohérence, initialement proposé dans le cadre des techniques de décomposition parcimonieuse. Cette approche a été couplée à un algorithme de type moindres carrés récursif, et illustrée au travers d'expérimentations.

Dans de futurs travaux, on envisage d'adapter des critères plus avancés issus des méthodes de décompositions parcimonieuses, dont la fonction de cohérence cumulée. D'autre part, on souhaite développer une stratégie permettant aussi d'exclure des éléments du dictionnaire.

## Références

- [1] N. Aronszajn, "Theory of reproducing kernels," *Trans. Amer. Math. Soc.*, vol. 68, pp. 337–404, 1950.
- [2] C. Richard, J. C. M. Bermudez, and P. Honeine, "Nonlinear kernel-based adaptive filtering with order controlled by a coherence criterion," *submitted to IEEE Transactions on Signal Processing*, 2007.
- [3] G. Kimeldorf and G. Wahba, "Some results on tchebycheffian spline functions," *Journal of Mathematical Analysis and Applications*, vol. 33, pp. 82–95, 1971.
- [4] R. Gribonval and P. Vandergheynst, "On the exponential convergence of Matching Pursuits in quasi-incoherent dictionaries," *IEEE Transactions on Information Theory*, vol. 52, no. 1, pp. 255–261, 2006.