

Structures à noyau reproduisant pour le filtrage adaptatif

Cédric RICHARD¹, Régis LENGELLÉ¹, Ibtissam CONSTANTINE², Laurent SOUFFLET²

¹LM2S - Laboratoire de Modélisation et Sûreté des Systèmes
Université de Technologie de Troyes, 12 rue Marie Curie, BP 2060, 10010 Troyes cedex, France

²FORENAP - Institut de recherche en neurosciences, neuropharmacologie et psychiatrie
Centre hospitalier de Rouffach, 27 rue du 4ème RSM, 68250 Rouffach, France

cedric.richard@utt.fr
http://www-lm2s.utt.fr/~richard

Résumé –

Dans cet article, on montre que les espaces de Hilbert à noyau reproduisant offrent un cadre intéressant pour l'élaboration de filtres non-linéaires à temps discret et réponse impulsionnelle finie, optimaux ou adaptatifs. L'intérêt majeur de ces espaces, dans lesquels s'effectue la recherche d'une solution, est que cette dernière s'y exprime sous forme d'une décomposition en série de noyaux reproduisants. Le nombre de paramètres à estimer est indépendant de l'ordre du filtre et de la *complexité* de celui-ci, que caractérise par exemple le degré du polynôme dans le cas particulier d'un filtre de Volterra. Les avantages liés au caractère linéaire de la structure par rapport aux paramètres à estimer perdurent.

Abstract –

In this paper, it is shown how the ideas of reproducing kernel Hilbert spaces can be applied to provide an alternative solution to the design of nonlinear optimum or adaptive filters. The main ideas behind this approach is to determine a Hilbert space corresponding to the solution space. The particular advantage of such a space is that the nonlinear filter can be expanded in terms of the reproducing kernel. The number of parameters to be estimated is independent of the memory length and of the *complexity* of the filter, e.g., the polynomial degree in the case of a Volterra filter. The advantage of linear-in-the parameters estimation remains.

1 Présentation du problème

On qualifie de *filtre* toute structure matérielle ou logicielle appliquée à des données bruitées en vue d'en extraire une information pertinente au sens d'un critère donné. Ces données peuvent par exemple provenir de capteurs bruyants, ou encore être issues de canaux de communication soumis à des perturbations. La forme canonique du problème de filtrage considéré ici est proposée en figure 1. Elle comporte une entrée $x[n]$ et une sortie désirée $d[n]$, que l'on suppose centrées et réelles sans perte de généralité. On note $e[n] = d[n] - y[n]$ l'erreur commise. La question posée est celle du choix d'un modèle w et de la mise en œuvre d'une technique opérationnelle permettant d'en déterminer les paramètres caractéristiques par optimisation d'un critère de performance. La théorie de Wiener s'applique aux processus stationnaires du second ordre, et consiste en la recherche de la structure linéaire [Hay91]

$$y[n] = \sum_{k=0}^{p-1} w_k x[n-k] \quad (1)$$

minimisant la variance de l'erreur $\mathcal{E}\{e[n]^2\}$. En introduisant les notations vectorielles suivantes

$$\mathbf{x}_n = (x[n] \dots x[n-p+1])^t$$

$$\mathbf{w} = (w_0 \dots w_{p-1})^t,$$

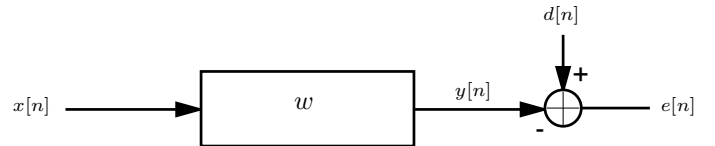


Figure 1: Système considéré et notations.

la solution du problème est obtenue en résolvant les équations $\mathcal{E}\{\mathbf{x}_n(d[n] - \mathbf{w}^t \mathbf{x}_n)\} = 0$, c'est-à-dire le système

$$\mathbf{R}_x \mathbf{w} = \mathbf{R}_{xd} \quad (2)$$

avec \mathbf{R}_x et \mathbf{R}_{xd} désignant $\mathcal{E}\{\mathbf{x}_n \mathbf{x}_n^t\}$ et $\mathcal{E}\{d[n] \mathbf{x}_n\}$.

Force est de constater que le caractère linéaire de la structure recherchée rend son élaboration aisée, au détriment de sa capacité à fournir des solutions satisfaisantes à tout problème. Afin de pallier cette limite tout en conservant une structure comparable à (1), on peut envisager de plonger préalablement l'observation \mathbf{x}_n dans un espace de représentation plus adéquat grâce au concours d'une application non-linéaire $\phi(\cdot)$, puis à considérer le filtre $y[n] = \mathbf{w}^t \phi(\mathbf{x}_n)$. Comme précédemment, \mathbf{w} peut alors être déterminé en résolvant l'équation de Wiener-Hopf définie à partir de

$$\mathcal{E}\{\phi(\mathbf{x}_n)(d[n] - \mathbf{w}^t \phi(\mathbf{x}_n))\} = 0. \quad (3)$$

Pour l'élaboration d'un filtre polynômial de degré 2 par exemple, ce concept en apparence basique préconise ainsi la mise

en œuvre d'un filtre linéaire opérant sur l'observation $\phi(\mathbf{x}_n)$ constituée des composantes de \mathbf{x}_n et de leurs produits croisés. Le nombre de paramètres à estimer, égal à $p(p+3)/2$ et déjà prohibitif au regard de la relative simplicité de la structure obtenue, laisse présager des difficultés pratiques rencontrées lorsqu'une telle stratégie est adoptée sans précautions. Dans le domaine de la reconnaissance des formes, des considérations sur les espaces de Hilbert à noyau reproduisant ont pourtant rendu des pratiques analogues possibles. En autorisant la synthèse de statistiques linéaires généralisées sans jamais évaluer explicitement $\phi(\mathbf{x}_n)$, ces structures algébriques ont en effet permis de dégager de nouvelles perspectives dans le cadre des méthodes à noyau, en particulier avec les Support Vector Machines [Vap95], pour la résolution de problèmes de décision [Ric02, Sho99] et de régression [Smo98, Dod02].

Dans cet article, on se propose d'appliquer les mêmes principes afin d'étendre au cas non-linéaire des techniques classiques de filtrage optimal et adaptatif à temps discret et réponse impulsionnelle finie. Pour ce faire, des considérations sur les espaces de Hilbert à noyau reproduisant sont proposées au préalable.

2 Espaces à noyau reproduisant et condition de Mercer

2.1 Éléments de théorie

Soit \mathcal{H} un espace fonctionnel hilbertien réel de produit scalaire $\langle \cdot ; \cdot \rangle_{\mathcal{H}}$, composé de fonctions ψ continues sur un ensemble \mathcal{X} . D'après le théorème de représentation de Riesz, il existe une fonction unique $\kappa(\mathbf{x}_i, \mathbf{x}_j)$ de la variable \mathbf{x}_i , avec \mathbf{x}_j fixé, telle que

$$\psi(\mathbf{x}_j) = \langle \psi ; \kappa(\cdot, \mathbf{x}_j) \rangle_{\mathcal{H}}, \quad \forall \psi \in \mathcal{H}. \quad (4)$$

Dans cette expression, $\kappa(\cdot, \mathbf{x}_j)$ désigne une fonction définie sur \mathcal{X} , obtenue en fixant le second argument de κ à \mathbf{x}_j . Il en résulte que l'ensemble $\{\kappa(\cdot, \mathbf{x}) : \mathbf{x} \in \mathcal{X}\}$ engendre \mathcal{H} , et que le produit scalaire $\langle \cdot ; \cdot \rangle_{\mathcal{H}}$ ne nécessite d'être défini que sur cet ensemble de générateurs. Au vu de cette propriété, κ est appelé *noyau reproduisant* de \mathcal{H} . L'équation (4) implique de plus

$$\langle \kappa(\cdot, \mathbf{x}_i) ; \kappa(\cdot, \mathbf{x}_j) \rangle_{\mathcal{H}} = \kappa(\mathbf{x}_j, \mathbf{x}_i), \quad (5)$$

pour tout $\mathbf{x}_i, \mathbf{x}_j \in \mathcal{X}$. Ce résultat signifie que $\kappa(\mathbf{x}_i, \mathbf{x}_j)$ fournit le produit scalaire des images dans \mathcal{H} de toute paire d'éléments de l'ensemble \mathcal{X} . En autorisant la mise en œuvre de ce concept sans nécessairement connaître explicitement \mathcal{H} , la condition de Mercer a contribué aux plus récents développements des structures à noyau. Celle-ci stipule qu'à tout noyau κ vérifiant

$$\iint \kappa(\mathbf{x}_1, \mathbf{x}_2) \psi(\mathbf{x}_1) \psi(\mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 \geq 0, \quad \forall \psi \in \mathcal{H}, \quad (6)$$

on peut associer un espace fonctionnel hilbertien réel \mathcal{H} à noyau reproduisant. On a alors en particulier

$$\kappa(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i) ; \phi(\mathbf{x}_j) \rangle_{\mathcal{H}} \quad (7)$$

où $\phi(\mathbf{x})$ désigne la fonction $\kappa(\cdot, \mathbf{x})$ de \mathcal{H} étant donné \mathbf{x} un élément de \mathcal{X} .

2.2 Éventail de noyaux reproduisants

L'expression (7) montre qu'il n'est pas nécessaire de connaître $\phi(\cdot)$ pour l'évaluation de produits scalaires dans l'espace \mathcal{H} associé. Cette application est implicitement définie par le choix d'un noyau reproduisant κ , à partir duquel il est possible de couvrir une large classe de non-linéarités. En voici quelques exemples classiques, une liste plus complète pouvant être consultée dans [Can02, Vap95].

2.2.1 Noyaux polynômiaux

Le noyau polynômial défini ainsi

$$\kappa(\mathbf{x}_1, \mathbf{x}_2) = (1 + \mathbf{x}_1 \cdot \mathbf{x}_2)^q \quad (8)$$

vérifie la condition de Mercer. Ceci induit qu'il fournit à moindre coup de calcul le produit scalaire des images de \mathbf{x}_i et \mathbf{x}_j par une application $\phi(\cdot)$. Dans ce cas, on montre que les composantes de $\phi(\mathbf{x})$ sont les monômes de degré inférieur ou égal à q constitués des composantes de \mathbf{x} , ce qui ne va pas sans rappeler les filtres de Volterra [Duv87]. Parce qu'ils sont fonction du produit scalaire des observations, de tels noyaux sont dits *projectifs*.

2.2.2 Noyaux exponentiels radiaux

Les noyaux de type *radial* dépendent de $\|\mathbf{x}_1 - \mathbf{x}_2\|$. Ils ont fait l'objet d'une attention particulière dans la littérature en raison du rôle central qu'ils jouent dans les méthodes d'estimation et de classification à base de noyaux ou de potentiels [Dev96]. On compte parmi eux le noyau gaussien, défini par

$$\kappa(\mathbf{x}_1, \mathbf{x}_2) = \exp(-\|\mathbf{x}_1 - \mathbf{x}_2\|^2 / \beta_0), \quad (9)$$

où β_0 est appelé *largeur de bande*. Ce noyau est caractérisé par un continuum de valeurs propres, ce qui signifie que les composantes de ϕ ne sont pas en nombre fini comme dans l'exemple (8). Enfin, dans la famille des noyaux radiaux, on peut encore citer le noyau exponentiel

$$\kappa(\mathbf{x}_1, \mathbf{x}_2) = \exp(-\|\mathbf{x}_1 - \mathbf{x}_2\| / \beta_0). \quad (10)$$

2.2.3 Noyaux sigmoïdaux

On peut élaborer un réseau de neurones à une couche cachée en choisissant le noyau sigmoïdal

$$\kappa(\mathbf{x}_1, \mathbf{x}_2) = \tanh(\alpha_0 (\mathbf{x}_1 \cdot \mathbf{x}_2) + \beta_0). \quad (11)$$

La qualité reproduisante de κ dépend des paramètres α_0 et β_0 sélectionnés, contrairement aux noyaux polynômiaux et radiaux présentés ci-dessus. S'ils ne sont pas convenablement choisis, il en résulte la perte du cadre rigoureux offert par les espaces de Hilbert à noyau reproduisant. Toutefois, les errements de la pratique font que cette contrainte se trouve parfois relaxée [Can02].

3 Techniques de filtrage à noyau

3.1 Filtrage optimal

Soit \mathcal{H} un espace de Hilbert à noyau reproduisant κ . On note $\phi(\mathbf{x})$ la fonction de \mathcal{X} dans \mathcal{H} associée. Avant d'aborder

le thème du filtrage adaptatif, on s'intéresse dans un premier temps au problème de filtrage optimal. Celui-ci consiste en la recherche d'une fonction ψ de \mathcal{H} minimisant la variance de l'erreur entre la sortie désirée $d[n]$ et la sortie obtenue, définie à présent par

$$y[n] = \langle \psi ; \phi(\mathbf{x}_n) \rangle_{\mathcal{H}}. \quad (12)$$

Compte tenu de son appartenance à \mathcal{H} , la fonction ψ recherchée s'exprime sous la forme

$$\psi = \sum_k a_k \phi(\mathbf{x}_k), \quad a_k \in \mathbb{R} \quad (13)$$

où $\mathcal{X} \triangleq \{\mathbf{x}_k\}_k$ désigne l'ensemble des entrées du filtre. En combinant les deux expressions précédentes, on aboutit directement à la forme duale

$$y[n] = \sum_k a_k \kappa(\mathbf{x}_n, \mathbf{x}_k), \quad (14)$$

dont il s'agit d'optimiser les paramètres a_k . Le caractère linéaire de cette structure permet d'appliquer directement la théorie de Wiener. Il convient de remarquer que (14) ne repose pas explicitement sur $\phi(\cdot)$, dont il n'est d'ailleurs pas nécessaire de connaître l'expression analytique. Celle-ci est en effet implicitement définie par le choix d'un noyau reproduisant κ . A titre d'exemple, l'usage du noyau (8) mène à un filtre de Volterra de degré q , dont la formulation duale est fournie par

$$y[n] = \sum_k a_k (1 + \mathbf{x}_n \cdot \mathbf{x}_k)^q. \quad (15)$$

Celle-ci ne nécessite pas le calcul de $\phi(\mathbf{x}_n)$, dont on rappelle que les composantes sont les monômes de degré inférieur ou égal à q constitués des composantes de \mathbf{x}_n .

La difficulté majeure rencontrée avec la formulation (14) est qu'elle repose sur une somme infinie de termes ainsi qu'une connaissance explicite de l'espace des observations \mathcal{X} ! En pratique, cet obstacle peut être contourné en limitant la recherche d'une solution ψ à un sous-espace $\mathcal{H}_p \subset \mathcal{H}$ de dimension p , engendré donc par des fonctions $\{\phi(\mathbf{x}_k)\}_{0 \leq k \leq p-1}$. Les observations \mathbf{x}_k désignent p réalisations de l'entrée \mathbf{x} , jouant ainsi le rôle de base d'apprentissage. Elles peuvent être arbitrairement choisies à défaut de solution plus adéquate. Dans ces conditions, on aboutit à la formulation duale $y[n] = \sum_{k=0}^{p-1} a_k \kappa(\mathbf{x}_n, \mathbf{x}_k)$, dont il s'agit de déterminer les paramètres a_k afin de minimiser la variance de l'erreur $e[n]$. La solution du problème est obtenue en résolvant un système à p inconnues analogue à (2) dans lequel $\tilde{\mathbf{\kappa}}(\mathbf{x}_n) = (\kappa(\mathbf{x}_n, \mathbf{x}_0) \dots \kappa(\mathbf{x}_n, \mathbf{x}_{p-1}))^t$ est substitué à \mathbf{x}_n . On le note

$$\mathbf{R}_x^\phi \mathbf{a} = \mathbf{R}_{xd}^\phi. \quad (16)$$

On remarque que le nombre de paramètres à estimer est indépendant de l'ordre du filtre et de la *complexité* de celui-ci, que caractérise par exemple le degré du polynôme dans le cas particulier d'un filtre de Volterra. Les avantages liés au caractère linéaire de la structure par rapport aux paramètres à estimer perdurent.

3.2 Filtres adaptatifs

La résolution du système (16) par une méthode d'inversion directe peut s'avérer très coûteuse en temps de calculs, puisque

en $o(p^3)$. La méthode du gradient constitue une approche itérative à même de pallier cet inconvénient. Celle-ci consiste à incrémenter une solution candidate $\mathbf{a}^{(n)}$ à l'étape n dans la direction opposée au gradient de la fonction coût, ici la variance de l'erreur, afin d'obtenir $\mathbf{a}^{(n+1)}$. Plus précisément, la forme explicite de cet algorithme classique [Hay91] est la suivante dans le cas de l'approche à noyau présentée :

$$\mathbf{a}^{(n+1)} = \mathbf{a}^{(n)} + \eta^{(n)} (\mathbf{R}_{xd}^\phi - \mathbf{R}_x^\phi \mathbf{a}^{(n)}), \quad \eta^{(n)} > 0, \quad (17)$$

où les $\eta^{(n)}$ définissent une série de coefficients décroissants appelés pas d'adaptation. Sans surprise, les conditions de convergence de cet algorithme vers la solution de (16) demeurent inchangées par rapport à la méthode du gradient classique s'adressant aux filtres linéaires. Il en est de même pour la caractérisation de la vitesse de convergence et le calcul du pas optimal.

Les méthodes de gradient stochastique consistent à remplacer le gradient de la fonction coût par une approximation de celui-ci calculée à partir des données disponibles. En particulier, la méthode LMS repose sur les estimées instantanées suivantes des corrélations \mathbf{R}_{xd}^ϕ et \mathbf{R}_x^ϕ

$$\hat{\mathbf{R}}_x^\phi = \tilde{\mathbf{\kappa}}(\mathbf{x}_n) \tilde{\mathbf{\kappa}}^t(\mathbf{x}_n), \quad \hat{\mathbf{R}}_{xd}^\phi = d[n] \tilde{\mathbf{\kappa}}(\mathbf{x}_n), \quad (18)$$

ce qui revient à minimiser le critère instantané $e[n]^2$ à l'aide de l'algorithme de gradient. La relation permettant la mise à jour des coefficients du filtre est à présent

$$\mathbf{a}^{(n+1)} = \mathbf{a}^{(n)} + \eta^{(n)} \tilde{\mathbf{\kappa}}(\mathbf{x}_n) (d[n] - y[n]), \quad \eta^{(n)} > 0, \quad (19)$$

où $y[n] = \tilde{\mathbf{\kappa}}^t(\mathbf{x}_n) \mathbf{a}^{(n)}$ désigne la sortie du filtre à l'instant n . Comme précédemment, les discussions relatives à la convergence et à la stabilité de cet algorithme demeurent inchangées par rapport à l'approche classique [Mac95]. Il est à noter que cette formulation peut être déclinée en algorithme du signe, qui consiste à adopter la variante simplifiée suivante de l'algorithme LMS :

$$\mathbf{a}^{(n+1)} = \mathbf{a}^{(n)} + \eta^{(n)} \tilde{\mathbf{\kappa}}(\mathbf{x}_n) \text{sgn}(e[n]), \quad \eta^{(n)} > 0, \quad (20)$$

Il en est de même pour l'algorithme LMS normalisé, qui consiste à prendre un pas d'adaptation inversement proportionnel à la puissance instantanée de l'entrée, soit

$$\eta^{(n)} = \frac{1}{\|\tilde{\mathbf{\kappa}}(\mathbf{x}_n)\|^2}. \quad (21)$$

4 Mise en œuvre

Afin d'illustrer l'approche proposée, on considère le système non-linéaire défini par [Bil86]

$$d[n] = 0.5 d[n-1] + 0.3 d[n-1] x[n-1] + 0.2 x[n-1] + 0.05 d^2[n-1] + 0.6 x^2[n-1],$$

où $x[n]$ et $d[n]$ font respectivement fonction d'entrée du système et de sortie désirée. On suppose que cette dernière est par ailleurs corrompue par un bruit blanc qui suit une loi gaussienne $\mathcal{N}(0, 0.05)$. L'équation récurrente précédente est initialisée avec $y[0] = 1$, et $x[n]$ suit une loi $\mathcal{N}(0.2, 0.1)$. Les figures 2.(a) et 2.(b) représentent les sorties désirées $d[n]$ et obtenues $y[n]$ avec un noyau polynômial de degré 3 et un noyau

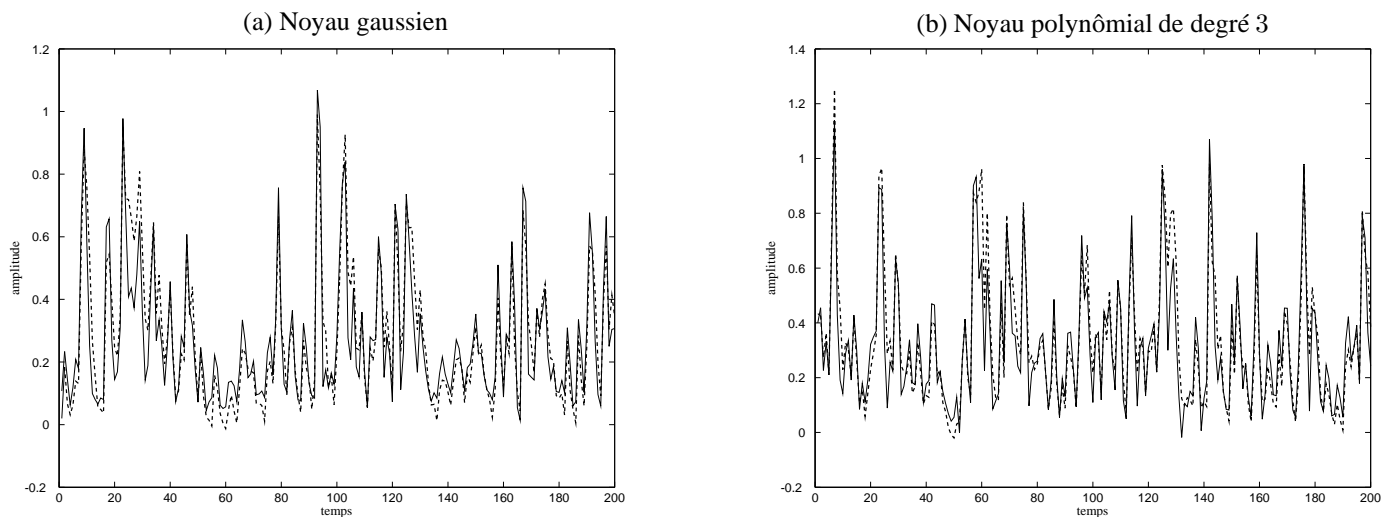


Figure 2: Illustration de la technique de filtrage à noyau sur des données test. La sortie désirée $d[n]$ figure en pointillés tandis que la sortie obtenue est en trait plein.

gaussien de largeur de bande β_0 unitaire, grâce à la méthode du gradient. Dans chacune de ces expérimentations, la structure a été recherchée dans le sous-espace \mathcal{H}_{50} engendré par les fonctions $\{\phi(\mathbf{x}_k)\}_{0 \leq k \leq 49}$ associées aux noyaux sélectionnés, avec $\mathbf{x}_k = (x[k] \dots x[k+2])$. Aussi les résultats obtenus s'avèrent-ils très satisfaisants, ce qui illustre la capacité de l'approche proposée à élaborer des filtres non-linéaires. Il convient également de noter que le coût de calcul demeure modeste puisqu'aucun calcul n'est explicitement fait dans l'espace transformé \mathcal{H} .

5 Conclusion et perspectives

Des considérations sur les espaces de Hilbert ont d'abord été présentées dans ce chapitre. Celles-ci constituent les bases du *kernel trick*, sur lequel reposent les Support Vector Machines et autres méthodes à noyaux pour la classification. Ce principe, aujourd'hui communément utilisé en Reconnaissance des Formes, est ici mis en œuvre pour l'élaboration de filtres non-linéaires selon un formalisme équivalent à celui de Wiener et de ses extensions adaptatives. Il en résulte une méthode particulièrement souple puisqu'elle permet de synthétiser de multiples formes de filtres par simple changement de noyau. Malheureusement, la complexité des modèles considérés invitent rapidement l'expérimentateur à se positionner au mieux par rapport au compromis biais-variance. Afin de compléter ce travail, il convient donc de s'intéresser à présent à cette question. Le caractère embryonnaire de la discussion sur le choix du sous-espace \mathcal{H}_p dans lequel une solution est recherchée n'aura pas échappé au lecteur. Il s'avère également nécessaire de creuser ce sujet, et d'envisager éventuellement une technique adaptative pour le suivi du sous-espace de solutions le plus favorable possible.

References

[Bil86] S. A. Billings, W. S. F. Voon. Correlation-based model validity tests for non-linear models. *International Journal of Control*, vol. 44, p. 235-244, 1986.

- [Can02] S. Canu. Modèles connexionistes et SVM pour la décision. In R. Lengellé, (éd.). *Décision et Reconnaissance des Formes en Signal*. Paris : Hermès Sciences, Traité IC2, 2002.
- [Dev96] L. Devroye, L. Györfi, G. Lugosi. *A Probabilistic Theory of Pattern Recognition*. New York : Springer-Verlag, 1996.
- [Dod02] T. Dodd, R. Harrison. A new solution to volterra series estimation. *Proc. IFAC World Congress*, 2002.
- [Duv87] P. Duvaut, B. Picinbono. Détection, estimation et filtres de Volterra. *Proc. Onzième colloque GRETSI*, 1987.
- [Hay91] S. Haykin. *Adaptive filter theory*. Englewood Cliffs : Prentice Hall, 1991.
- [Mac95] O. Macchi. *Adaptive processing: the LMS approach with applications in transmission*. New York : Wiley, 1995.
- [Ric02] C. Richard. *Méthodes à noyau et critères de contraste pour la détection à structure imposée*. Habilitation à diriger des recherches, Université de Technologie de Compiègne, 2002.
- [Sho99] B. Schölkopf, C. Burges, A. Smola. *Advances in Kernel Methods-Support Vector Learning*. Cambridge : MIT Press, 1999.
- [Smo98] A. Smola, B. Schölkopf. *A tutorial on support vector regression*. NeuroCOLT2 Technical Report NC2-TR-1998-030, 1998.
- [Vap95] V. Vapnik. *The Nature of Statistical Learning Theory*. New York : Springer-Verlag, 1995.