

Filtrage adaptatif non-linéaire par méthodes de gradient stochastique court-terme à noyau

Cédric RICHARD

ISTIT (FRE CNRS 2732) - Équipe Modélisation et Sûreté des Systèmes
Université de Technologie de Troyes, 12 rue Marie Curie, BP 2060, 10010 Troyes cedex, France
cedric.richard@utt.fr

Résumé –

Les algorithmes à noyau tels que les Support Vector Machines suscitent un intérêt sans cesse grandissant depuis près de dix ans, essentiellement dans le cadre de tâches de classification et de régression. Ils reposent pour la plupart sur le coup du noyau, selon l'appellation consacrée, et le théorème de représentation. Ces deux éléments permettent de créer des versions non-linéaires de méthodes de traitement de données originellement linéaires, avec un surcoût calculatoire modique et sans qu'il soit nécessaire de modifier l'architecture générale des algorithmes. Dans cet article, on montre que les noyaux reproduisants offrent des perspectives intéressantes pour l'élaboration de filtres adaptatifs non-linéaires de type gradient stochastique. Les avantages liés au caractère linéaire de la structure par rapport aux paramètres à estimer perdurent.

Abstract –

Kernel-based algorithms such as Support Vector Machines have been a topic of considerable interest over the last ten years, most notably for classification and regression. The use of kernels, through the kernel trick and the representer theorem, is an attractive computational shortcut to create nonlinear versions of conventional linear algorithms. Here we demonstrate the versatility and utility of this family of methods to develop nonlinear adaptive filtering algorithms, especially of the least-mean square type. The advantage of linear-in-the parameters estimation remains.

1 Introduction

Le thème du filtrage adaptatif a suscité un intérêt considérable au cours des trente dernières années car il vise à apporter des éléments de réponse à deux problèmes récurrents et souvent associés en traitement du signal, la modélisation de systèmes non-stationnaires et la carence en information statistique *a priori* [1, 2]. Le caractère linéaire des modèles généralement retenus est inhérent à leur interprétabilité et leur identification aisées. Nul n'ignore toutefois que les modèles non-linéaires sont la promesse de performances accrues, hélas au prix de difficultés conceptuelles et de mise en œuvre [3]. Les filtres polynômiaux et réseaux de neurones, souvent retenus en filtrage adaptatif non-linéaire, n'échappent pas à la règle. Les premiers sont généralement dotés de paramètres à déterminer en nombre rédhibitoire [4], tandis que les seconds représentent le stade ultime du modèle de type boîte noire et pâtissent d'une optimisation assez périlleuse [5]. Confronté depuis toujours à des difficultés similaires, le domaine de la reconnaissance des formes a récemment apporté des éléments de solution convaincants au travers des méthodes à noyau [6]. Elles sont pour la plupart issues d'algorithmes originellement linéaires auxquels on a pu appliquer deux principes clés : le coup du noyau [7] et le théorème de représentation [8].

L'objectif de cet article est d'adapter la stratégie mise en œuvre par les méthodes à noyau à la problématique particulière du filtrage adaptatif. Si la forme retenue est celle d'une méthode de gradient stochastique à noyau, l'approche proposée se veut générique afin de ne pas tomber dans les travers d'un algorithme ad'hoc. Elle offre en revanche des perspectives vers d'autres stratégies telles que RLS, ou encore QR-LS et inverse QR-LS à noyau. L'article est organisé ainsi. Dans un premier

temps, nous présentons les principes de base communs à la majorité des méthodes à noyau. Nous illustrons ceux-ci sur le problème du filtrage optimal, point de départ de la solution adaptative présentée dans un second temps. Des simulations et quelques perspectives viennent conclure cette étude.

2 Méthodes à noyau reproduisant : les figures imposées

Le coup du noyau et le théorème de représentation sont des principes récurrents dans le domaine des méthodes à noyau. Leur présentation fait ici suite à celle de la notion de noyau.

2.1 Notions fondamentales de noyaux

Si les notions de noyau défini positif et de noyau reproduisant peuvent être envisagées dans une perspective plus large [9, 10], le sujet traité dans cet article conduit à restreindre l'espace \mathcal{U} des observations à \mathbb{R}^p ou tout sous-espace non-vide de celui-ci. Un noyau est une fonction symétrique $\kappa(\mathbf{u}_i, \mathbf{u}_j)$ de $\mathcal{U} \times \mathcal{U}$ dans \mathbb{R} . On rappelle la définition d'un noyau défini positif.

Définition 1. On dit d'un noyau κ qu'il est défini positif sur \mathcal{U} s'il vérifie

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j \kappa(\mathbf{u}_i, \mathbf{u}_j) \geq 0 \quad (1)$$

pour tout $n \in \mathbb{N}$, $\mathbf{u}_1, \dots, \mathbf{u}_n \in \mathcal{U}$ et $a_1, \dots, a_n \in \mathbb{R}$.

On note $(\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{H}})$ un espace de Hilbert de fonctions de \mathcal{U} dans \mathbb{R} . Le caractère reproduisant d'un noyau au sein de \mathcal{H} , lorsqu'il a lieu, couvre les notions suivantes.

Noyaux projectifs		Noyaux radiaux	
sigmoïdal	$\tanh(\alpha_0 + \beta_0 \langle \mathbf{u}_i, \mathbf{u}_j \rangle)$	multi-quadratique	$\sqrt{\alpha_0^2 + \ \mathbf{u}_i - \mathbf{u}_j\ ^2}$
polynomial homogène	$(\langle \mathbf{u}_i, \mathbf{u}_j \rangle)^q$	multi-quadratique inverse	$1/\sqrt{\alpha_0^2 + \ \mathbf{u}_i - \mathbf{u}_j\ ^2}$
polynomial non-homogène	$(\alpha_0 + \langle \mathbf{u}_i, \mathbf{u}_j \rangle)^q$	Thin plate spline	$\ \mathbf{u}_i - \mathbf{u}_j\ ^{2\gamma_0} \ln \ \mathbf{u}_i - \mathbf{u}_j\ $

Table 1: Exemples de noyaux reproduisants

Définition 2. Une fonction $\kappa(\mathbf{u}_i, \mathbf{u}_j)$ de $\mathcal{U} \times \mathcal{U}$ dans \mathbb{R} est le noyau reproduisant de \mathcal{H} , s'il en admet un, si et seulement si

- la fonction $\kappa_{u_i} : \mathbf{u}_j \mapsto \kappa_{u_i}(\mathbf{u}_j) = \kappa(\mathbf{u}_i, \mathbf{u}_j)$ appartient à \mathcal{H} , quel que soit $\mathbf{u}_i \in \mathcal{U}$ fixé ;
- on a $\psi(\mathbf{u}_i) = \langle \psi, \kappa_{u_i} \rangle_{\mathcal{H}}$ pour tout $\mathbf{u}_i \in \mathcal{U}$ et $\psi \in \mathcal{H}$.

On peut montrer que tout noyau défini positif κ est le noyau reproduisant d'un espace de Hilbert de fonctions de \mathcal{U} dans \mathbb{R} en exhibant directement celui-ci. On considère pour cela l'espace vectoriel \mathcal{H}_0 engendré par les fonctions $\{\kappa_u\}_{u \in \mathcal{U}}$, auquel on associe le produit scalaire

$$\langle \psi, \phi \rangle_{\mathcal{H}_0} = \sum_{i=1}^n \sum_{j=1}^m a_i b_j \kappa(\mathbf{u}_i, \mathbf{u}_j), \quad (2)$$

avec $\psi = \sum_{i=1}^n a_i \kappa_{u_i}$ et $\phi = \sum_{j=1}^m b_j \kappa_{u_j}$ appartenant à l'espace \mathcal{H}_0 . Il s'agit donc là d'un espace pré-hilbertien, que l'on complète conformément à [9] de sorte que toute suite de Cauchy y converge. On aboutit ainsi à l'espace de Hilbert \mathcal{H} à noyau reproduisant κ recherché. Réciproquement, on démontre que tout noyau reproduisant est défini positif en considérant l'élément $\psi = \sum_{i=1}^n a_i \kappa_{u_i}$ de \mathcal{H} , et en notant que le membre de gauche de (1) correspond alors à $\|\psi\|_{\mathcal{H}}^2$. Le tableau 1 présente quelques noyaux reproduisants, d'autres exemples ainsi que des règles permettant de les combiner étant consultables dans [6, 11].

2.2 Coup du noyau, théorème de représentation

De la propriété $\psi(\mathbf{u}_i) = \langle \psi, \kappa_{u_i} \rangle_{\mathcal{H}}$ issue de la définition 2 résulte une propriété fondamentale des espaces de Hilbert à noyau reproduisant. En remplaçant ψ par κ_{u_j} , on aboutit en effet à la relation

$$\kappa(\mathbf{u}_i, \mathbf{u}_j) = \langle \kappa_{u_j}, \kappa_{u_i} \rangle_{\mathcal{H}} \quad (3)$$

quels que soient \mathbf{u}_i et $\mathbf{u}_j \in \mathcal{U}$. Le noyau κ fournit donc le produit scalaire des images dans \mathcal{H} de toute paire d'éléments de \mathcal{U} , sans qu'il soit nécessaire de les expliciter. Ce *coup du noyau*, selon l'appellation consacrée, permet d'élaborer des méthodes non-linéaires de traitement de données à partir d'approches linéaires sous réserve que celles-ci puissent s'exprimer uniquement en fonction de produits scalaires des observations. Il suffit en effet de remplacer chacun d'eux par un noyau non-linéaire. La structure des algorithmes demeure alors inchangée et le surcoût calculatoire dû à l'évaluation des noyaux faible.

Pour qu'il soit opérationnel, le coup du noyau nécessite souvent d'être associé au théorème de représentation [8, 12]. Ce théorème établit que toute fonction ψ^* d'un espace de Hilbert à noyau reproduisant \mathcal{H} qui minimise un coût

$$J((\mathbf{u}_1, y_1, \psi(\mathbf{u}_1)), \dots, (\mathbf{u}_n, y_n, \psi(\mathbf{u}_n))) + g(\|\psi\|_{\mathcal{H}}^2), \quad (4)$$

impliquant n sorties $\psi(\mathbf{u}_i)$ obtenues pour des entrées \mathbf{u}_i et éventuellement n sorties désirées y_i , avec g une fonction

monotone croissante sur \mathbb{R}_+ , peut s'écrire sous la forme

$$\psi^* = \sum_{i=1}^n a_i^* \kappa_{u_i}. \quad (5)$$

On démontre ceci en notant que toute fonction ψ de \mathcal{H} se décompose selon $\psi = \sum_{i=1}^n a_i \kappa_{u_i} + \psi^\perp$, avec $\langle \psi^\perp, \kappa_{u_i} \rangle_{\mathcal{H}} = 0$ pour tout $i = 1, \dots, n$. Puisque $\psi(\mathbf{u}_j) = \langle \psi, \kappa_{u_j} \rangle_{\mathcal{H}}$, la valeur de $\psi(\mathbf{u}_j)$ n'est donc pas affectée par ψ^\perp , pour $j = 1, \dots, n$.

3 Applications au filtrage

Le coup du noyau et le théorème de représentation trouvent une application dans le problème du filtrage optimum. Elle est ici décrite avant de s'intéresser à une technique de filtrage adaptatif à noyau basée sur un algorithme de gradient stochastique.

3.1 Filtrage optimum à noyau

Soit κ un noyau reproduisant et \mathcal{H} l'espace de Hilbert qui lui est associé. On s'intéresse préalablement à la recherche d'une fonction $\hat{\psi}$ de \mathcal{H} minimisant l'erreur quadratique

$$\hat{\psi} = \arg \min_{\psi \in \mathcal{H}} \sum_{i=1}^n |d(i) - \psi(\mathbf{u}_i)|^2 \quad (6)$$

entre n sorties désirées notées $d(i)$ et celles fournies par un modèle non-linéaire $y(i) = \psi(\mathbf{u}_i)$ excité par des entrées \mathbf{u}_i , pour $i = 1, 2, \dots, n$. En vertu du théorème de représentation, on sait que la fonction recherchée peut s'exprimer sous la forme d'une série de noyaux dépendant des données disponibles, soit $\psi(\cdot) = \sum_{k=1}^n \alpha(k) \kappa(\cdot, \mathbf{u}_k)$. On aboutit donc à la formulation duale

$$y(i) = \sum_{k=1}^n \alpha(k) \kappa(\mathbf{u}_i, \mathbf{u}_k) = \boldsymbol{\alpha}^t \boldsymbol{\kappa}_i \quad (7)$$

où $\boldsymbol{\alpha}$ et $\boldsymbol{\kappa}_i = (\kappa(\mathbf{u}_i, \mathbf{u}_1) \dots \kappa(\mathbf{u}_i, \mathbf{u}_n))^t$ appartiennent à \mathbb{R}^n . La résolution du problème (6) se réduit donc à la minimisation de la fonction coût $J(\boldsymbol{\alpha}) = \|\mathbf{d} - \mathbf{K} \boldsymbol{\alpha}\|^2$, où \mathbf{K} est la matrice de Gram telle que $\mathbf{K}(i, j) = \kappa(\mathbf{u}_i, \mathbf{u}_j)$. La solution $\hat{\boldsymbol{\alpha}}$ est obtenue en résolvant le système linéaire de taille $(n \times n)$

$$(\mathbf{K}^t \mathbf{K}) \boldsymbol{\alpha} = \mathbf{K}^t \mathbf{d}. \quad (8)$$

Il est à noter qu'à un facteur $\frac{1}{n}$ près, les deux membres de cette expression intègrent une estimation de la matrice de corrélation du vecteur $\boldsymbol{\kappa}$, et de l'intercorrélation de $\boldsymbol{\kappa}$ et \mathbf{d} .

3.2 Filtrage adaptatif à noyau

Le filtrage adaptatif pose la question du traitement d'un flux de données et de la mise à jour itérative de la solution $\hat{\psi}(\cdot)$ à l'arrivée de chaque nouvelle observation. La méthode optimale esquissée ci-dessus ne peut clairement être mise en œuvre

noyau	polynôme $q = 2$	polynôme $q = 3$	polynôme $q = 4$	gaussien $\sigma_0^2 = 0.1$	gaussien $\sigma_0^2 = 0.55$	gaussien $\sigma_0^2 = 1.0$	algorithme NLMS
$\hat{\sigma}_e^2$	0.0436	0.0565	0.0940	0.0394	0.0385	0.0392	0.1637

Table 2: Récapitulatif des résultats obtenus en simulation

dans ce contexte. Elle repose en effet sur l'inversion d'une matrice et une formulation duale de la solution dont les dimensions sont précisément le nombre de données parvenues à l'instant considéré. Afin de pallier cette difficulté, de précédents travaux ont suggéré de contrôler l'extension du modèle (7) avant de procéder à la détermination de α [13, 14, 15]. Supposons qu'à l'instant n , le modèle s'écrive sous la forme $\psi_n(\cdot) = \sum_{k \in \mathcal{K}_n} \alpha(k) \kappa(\cdot, \mathbf{u}_k)$ où \mathcal{K}_n désigne un ensemble d'indices. A l'instant suivant, le noyau $\kappa(\cdot, \mathbf{u}_{n+1})$ n'est introduit dans l'expression de $\psi_{n+1}(\cdot)$ que s'il apporte une richesse jugée suffisante. Ceci se traduit dans [13] par la vérification de la condition $\min_{\gamma} \|\kappa(\cdot, \mathbf{u}_{n+1}) - \sum_{k \in \mathcal{K}_n} \gamma_k \kappa(\cdot, \mathbf{u}_k)\|_{\mathcal{H}}^2 > \nu$, où ν est un seuil fixé au préalable. Un critère équivalent, au facteur $\hat{\alpha}(n+1)$ près, est considéré dans [14, 15] sous une forme différente. Les mises en œuvre diffèrent également puisque dans [14, 15], par exemple, chaque itération s'achève par la recherche puis la suppression éventuelle du noyau contribuant le moins à la décroissance de $J(\alpha)$. Dans [13] en revanche, les auteurs laissent croître librement l'ordre du modèle pour avoir démontré que l'ensemble \mathcal{K}_{∞} est de cardinal fini. En environnement non-stationnaire, ceci pose toutefois la question de la pertinence et de la parcimonie d'un modèle dont le dictionnaire $\{\kappa(\cdot, \mathbf{u}_k)\}_{k \in \mathcal{K}_{\infty}}$ est élaboré au cours des premières itérations, solution que nous avons par ailleurs adoptée dans de précédents travaux [16, 17]. Outre l'importante surcharge calculatoire occasionnée par la régulation de la taille du modèle, un reproche s'adressant à [13] et [14, 15] réside dans le fait que les algorithmes de mise à jour des coefficients du filtre sont uniquement de type "itératif sur l'ordre", privant par là-même l'utilisateur d'accéder à une plus grande variété de méthodes de filtrage adaptatif.

Afin de s'affranchir des trois inconvénients mentionnés ci-dessus, nous proposons d'adopter une approche court-terme en reformulant le critère (6) ainsi :

$$\hat{\psi}_n = \arg \min_{\psi \in \mathcal{H}_n} \sum_{i=1}^n |d(i) - \psi(\mathbf{u}_i)|^2 \quad (9)$$

où \mathcal{H}_n est le sous-espace de \mathcal{H} engendré par les m fonctions $\kappa(\cdot, \mathbf{u}_k)$ les plus récentes, c'est-à-dire telles que $k = n - m + 1, \dots, n$. Parce que \mathcal{H}_n est lui-même un espace de Hilbert à noyau reproduisant κ , nous pouvons rechercher une solution au problème (9) sous la forme

$$\psi_n(\mathbf{u}_i) = \sum_{k=n-m+1}^n \alpha_n(k) \kappa(\mathbf{u}_i, \mathbf{u}_k) = \alpha_n^t \boldsymbol{\kappa}_{i,n} \quad (10)$$

La fonction $\hat{\psi}_n(\cdot)$ est finalement obtenue en considérant le système linéaire de taille $(m \times m)$ suivant

$$(\mathbf{K}_n^t \mathbf{K}_n) \alpha_n = \mathbf{K}_n^t \mathbf{d}_n, \quad (11)$$

où \mathbf{K}_n est la matrice de taille $(n \times m)$ dont la ligne i correspond à $\boldsymbol{\kappa}_{i,n}^t$, et \mathbf{d}_n est le vecteur des sorties désirées disponibles de l'origine à l'instant n . Par rapport à l'approche (6)-(8), la

stratégie court-terme (9)-(11) ne nécessite pas l'inversion d'une matrice dont la taille croît avec le nombre de données reçues. S'il s'agit déjà là d'un avantage décisif dans un contexte de filtrage adaptatif, on évitera toutefois de répéter à chaque pas de temps la coûteuse résolution du système (11). Dans ce document, nous portons notre attention sur des méthodes de gradient stochastique même si la formulation proposée se prête également à d'autres stratégies classiques telles que RLS, ou encore QR-LS et inverse QR-LS. 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 l'approximation de $\mathbf{K}_n^t \mathbf{K}_n$ et $\mathbf{K}_n^t \mathbf{d}_n$ par les grandeurs instantanées respectives $\boldsymbol{\kappa}_{n,n} \boldsymbol{\kappa}_{n,n}^t$ et $d(n) \boldsymbol{\kappa}_{n,n}$. La relation de mise à jour des paramètres de la solution est alors donnée par

$$\hat{\alpha}_n = \hat{\alpha}_{n-1} + \eta_n \boldsymbol{\kappa}_{n,n} (d(n) - \boldsymbol{\kappa}_{n,n}^t \hat{\alpha}_{n-1}), \quad (12)$$

avec $\eta^{(n)} > 0$. Remarquons qu'en multipliant à gauche chacun des membres de cette expression par $\boldsymbol{\kappa}_{n,n}^t$, on trouve

$$\hat{\psi}_n(\mathbf{u}_n) = \hat{\psi}_{n-1}(\mathbf{u}_n) + \eta_n \|\boldsymbol{\kappa}_{n,n}\|^2 (d(n) - \hat{\psi}_{n-1}(\mathbf{u}_n)). \quad (13)$$

Les discussions sur la convergence et à la stabilité de cet algorithme demeurent inchangées par rapport à l'approche linéaire. Enfin, cette formulation peut être déclinée sans difficulté selon les classiques variantes ϵ -NLMS, sign-error LMS et autre leaky-LMS.

3.3 Simulations

Nous nous intéressons au système non-linéaire suivant

$$z(n) = 0.5 z(n-1) + 0.3 z(n-1) u(n-1) + 0.2 u(n-1) + 0.05 z^2(n-1) + 0.6 u^2(n-1) \quad (14)$$

initialisée par $z(1) = 1$. On considère que l'entrée $u(n)$ suit la loi normale $\mathcal{N}(0, 0.1)$, et que la sortie observée est donnée par $d(n) = z(n) + b(n)$ où $b(n)$ est un bruit blanc de loi $\mathcal{N}(0, 0.1)$. La méthode NLMS court-terme à noyau, obtenue en posant $\eta_n = 1/\|\boldsymbol{\kappa}_{n,n}\|^2$ dans la relation (12), a été mise en œuvre afin d'estimer un modèle de la forme $d(n) = \psi(u(n-1), d(n-1))$. La largeur m de l'horizon d'observation a été arbitrairement fixée à 2, signifiant que $\alpha_n \in \mathbb{R}^2$, et les noyaux polynômiaux et gaussiens rappelés ci-dessous ont été successivement testés :

$$\kappa(\mathbf{u}_i, \mathbf{u}_j) = (1 + \mathbf{u}_i^t \mathbf{u}_j)^q \quad (15)$$

$$\kappa(\mathbf{u}_i, \mathbf{u}_j) = \exp(-\|\mathbf{u}_i - \mathbf{u}_j\|^2 / 2\sigma_0^2), \quad (16)$$

où $q \in \mathbb{N}^*$ et $\sigma_0^2 \in \mathbb{R}^*$. A titre de remarques, rappelons que le noyau polynômial confère au modèle estimé la structure d'un filtre de Volterra de degré q , et que l'espace \mathcal{H} associé au noyau gaussien est de dimension infinie. Ainsi ces noyaux permettent-ils d'accéder à des modèles d'une complexité certaine, au prix d'une charge calculatoire modeste. Le tableau 2 présente la variance σ_e^2 de l'erreur de prédiction pour différents paramètres de noyaux. Celle-ci a été estimée à partir de 100

réalisations de signaux de 1000 échantillons. On remarque que le noyau gaussien de variance 0.55 a conduit aux meilleures performances, parmi tous les choix de σ_0^2 testés entre 0.1 et 1.4, avec un pas de 0.05. Celles-ci sont supérieures à celles obtenues avec un noyau polynomial, plus sensible au bruit, ainsi qu'à celles de l'algorithme NLMS classique, linéaire donc inapproprié pour modéliser le système (14).

4 Conclusion

Dans cet article, nous avons montré que les espaces de Hilbert à noyau reproduisant offrent un cadre intéressant pour l'élaboration de filtres non-linéaires, 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 alors indépendant de la *complexité* du filtre, que caractérise par exemple le degré du polynôme dans le cas particulier d'une série de Volterra. Les avantages liés au caractère linéaire de la structure par rapport aux paramètres à estimer perdurent. Dans ce contexte, nous nous sommes plus particulièrement intéressés à une approche adaptative court-terme de gradient stochastique à noyau. Nous avons montré sur un exemple qu'elle offre des performances supérieures à la méthode LMS dont elle s'inspire. La stratégie présentée offre de nombreuses perspectives d'extension non-linéaires d'autres méthodes de filtrage adaptatif conventionnelles.

Références

- [1] A. H. Sayed, *Fundamentals of adaptive filtering*. Hoboken, NJ: John Wiley & Sons, 2003.
- [2] S. Haykin, *Adaptive filter theory*. Englewood Cliffs, NJ: Prentice Hall, 1991.
- [3] G. B. Giannakis et E. Serpedin, "A bibliography on nonlinear system identification," *Signal Processing*, vol. 81, pp. 553–580, 2001.
- [4] V. J. Mathews et G. L. Sicuranze, *Polynomial signal processing*. New York, NY: John Wiley & Sons, 2000.
- [5] J. Sjöberg, Q. Zhang, L. Ljung, A. Benveniste, B. Deylon, P.-Y. Glorennec, H. Hjalmarsson, et A. Juditsky, "Nonlinear black-box modeling in system identification: a unified overview," *Automatica*, vol. 31, no. 12, pp. 1691–1724, 1995.
- [6] V. N. Vapnik, *The nature of statistical learning theory*. New York, NY: Springer, 1995.
- [7] B. Boser, I. Guyon, et V. Vapnik, "A training algorithm for optimal margin classifiers," in *Proc. Fifth Annual Workshop on Computational Learning Theory*, pp. 144–152.
- [8] B. Schölkopf, R. Herbrich, et R. Williamson, "A generalized representer theorem," NeuroCOLT, Royal Holloway College, University of London, UK, Tech. Rep. NC2-TR-2000-81, 2000.
- [9] N. Aronszajn, "Theory of reproducing kernels," *Transactions of the American Mathematical Society*, vol. 68, pp. 337–404, 1950.
- [10] F. Cucker et S. Smale, "On the mathematical foundations of learning," *Bulletin of the American Mathematical Society*, vol. 39, no. 1, pp. 1–49, 2002.
- [11] R. Herbrich, *Learning kernel classifiers. Theory and algorithms*. Cambridge, MA: The MIT Press, 2002.
- [12] G. Kimeldorf et G. Wahba, "Some results on Tchebycheffian spline functions," *Journal of Mathematical Analysis and Applications*, vol. 33, pp. 82–95, 1971.
- [13] Y. Engel, S. Mannor, et R. Meir, "Kernel recursive least squares," *IEEE Transactions on Signal Processing*, vol. 52, no. 8, pp. 2275–2285, 2004.
- [14] T. J. Dodd, V. Kadiramanathan, et R. F. Harrison, "Function estimation in Hilbert space using sequential projections," in *Proc. IFAC Conference on Intelligent Control Systems and Signal Processing*, 2003, pp. 113–118.
- [15] T. J. Dodd, B. Mitchinson, et R. F. Harrison, "Sparse stochastic gradient descent learning in kernel models," in *Proc. Second International Conference on Computational Intelligence, Robotics and Autonomous Systems*, 2003.
- [16] C. Richard, R. Lengellé, I. Constantin, et L. Soufflet, "Structures à noyau reproduisant pour le filtrage adaptatif," in *Proc. Colloque Grets*, 2003.
- [17] I. Constantin, C. Richard, R. Lengellé, et L. Soufflet, "Regularized kernel-based Wiener filtering. application to magnetoencephalographic signals denoising," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2005.