

Reconnaissance des formes par méthodes à noyau dans le domaine temps-fréquence

Paul HONEINE¹, Cédric RICHARD², Patrick FLANDRIN³

¹Sonalyse, Pist Oasis, 131 impasse des palmiers, 30319 Alès, France

²ISTIT (FRE CNRS 2732) - Équipe Modélisation et Sûreté des Systèmes
Université de Technologie de Troyes, 12 rue Marie Curie, BP 2060, 10010 Troyes cedex, France

³Laboratoire de Physique (UMR CNRS 5672), École Normale Supérieure de Lyon
46 allée d'Italie, 69364 Lyon, France

paul.honeine@mail.sonalyse.fr, cedric.richard@utt.fr, patrick.flandrin@ens-lyon.fr

Résumé –

Au cours de la dernière décennie, de nombreuses méthodes pour l'analyse et la classification de données fondées sur la théorie des noyaux reproduisants ont été proposées. Elles constituent une source de progrès, tant au niveau de la complexité algorithmique que des performances atteintes. On montre dans cet article qu'un choix approprié de noyau reproduisant permet à ces méthodes d'opérer sur des représentations temps-fréquence. Ce lien ouvre de multiples perspectives au domaine de l'analyse des signaux non-stationnaires puisqu'il lui permet d'accéder aux plus récentes avancées méthodologiques et théoriques en matière de reconnaissance des formes et de théorie de l'apprentissage.

Abstract –

During the last decade, many important advances have been made in the field of pattern recognition with the theory of reproducing kernels. This unified view has led to new algorithms with improved performance and lower computational complexity. In this paper, we show that some specific reproducing kernels allow these algorithms to operate on time-frequency representations. This link offers new perspectives in the field of non-stationary signal analysis since it provides an access to the most recent methodological and theoretical developments of pattern recognition and statistical learning theory.

1 Introduction

L'analyse des signaux non-stationnaires nécessite la mise en œuvre d'outils spécifiques parmi lesquels émergent les distributions temps-fréquence et temps-échelle. Celles-ci constituent une alternative intéressante aux techniques d'analyse spectrale classiques en incluant explicitement une dimension supplémentaire d'évolution temporelle. Elles offrent de plus une grande variété d'espaces de représentation, à la mesure de la diversité des objectifs visés par les utilisateurs. Par exemple, il est possible de privilégier l'intelligibilité de l'information délivrée par une représentation en y limitant les manifestations de bruits et autres termes interférentiels nuisant à sa lisibilité [1, 2, 3]. On peut encore être à la recherche d'un espace de représentation favorisant la résolution d'un problème de décision [4], point de rencontre entre l'analyse des signaux non-stationnaires et les théories de la décision et de l'apprentissage.

Au cours de la dernière décennie, de multiples algorithmes pour l'analyse et la classification de données fondés sur la théorie des espaces de Hilbert à noyau reproduisant ont été développés [5]. Ils reposent sur un principe central appelé *coup du noyau*, exploité pour la première fois avec les Support Vector Machines ou SVM [6], qui confère un caractère non-linéaire à nombre de méthodes originellement linéaires [7]. Parmi elles, on compte par exemple des généralisations non-linéaires de l'analyse en composantes principales [8] et des analyses dis-

criminantes de Fisher [9] et généralisée [10], respectivement désignées dans la littérature par kernel-PCA, kernel-FDA et kernel-GDA. Les méthodes à noyau constituent une source de progrès, tant au niveau du coût calculatoire que des performances atteintes, ces dernières étant par ailleurs encadrées par des bornes théoriques sans cesse affinées dans le cadre de la théorie statistique de l'apprentissage [11, 12]. Clairement, le domaine de l'analyse temps-fréquence ne s'est pas encore approprié ces techniques et la théorie qui les accompagne bien que, paradoxalement, la recherche d'espaces de représentation mettant en évidence une information pertinente y fasse l'objet de toutes les attentions. Aujourd'hui encore, peu de travaux combinent en effet méthodes à noyau et analyse temps-fréquence. Citons toutefois [13, 14] où, pour résoudre un problème de classification de signaux non-stationnaires, les auteurs associent les SVM à un noyau défini à partir de représentations temps-fréquence paramétrées selon [2].

A caractère plus générique, le présent article a pour objectif de montrer que les méthodes à noyau les plus performantes et les plus diverses peuvent être mises en œuvre dans le plan temps-fréquence grâce à un choix approprié de noyau reproduisant. L'analyse temps-fréquence des informations extraites par ces algorithmes est également abordée, contribuant ainsi à justifier le choix de ces espaces de représentation. Mais au préalable, il convient de rappeler les principaux éléments de théorie soutenant les méthodes à noyau.

2 Rappels du cadre théorique

Les méthodes à noyau sont pour la plupart issues d'algorithmes linéaires auxquels on a pu appliquer les deux résultats clés que sont le coup du noyau et le théorème de représentation [15]. Leur présentation fait ici suite à celle de la notion de noyau.

2.1 Noyau défini positif, noyau reproduisant

On considère un sous-espace \mathcal{X} de $\mathcal{L}_2(\mathbb{C})$, l'espace des signaux d'énergie finie à valeurs dans \mathbb{C} , que l'on munit du produit scalaire canonique $\langle x_i, x_j \rangle = \int_t x_i(t) \bar{x}_j(t) dt$ et de la norme associée. Un noyau désigne une fonction $\kappa(x_i, x_j)$ de $\mathcal{X} \times \mathcal{X}$ dans \mathbb{C} , à symétrie hermitienne. On rappelle les deux définitions fondamentales suivantes [5, 11].

Définition 1. *Un noyau κ est dit défini positif sur \mathcal{X} s'il vérifie*

$$\sum_{i=1}^n \sum_{j=1}^n a_i \bar{a}_j \kappa(x_i, x_j) \geq 0 \quad (1)$$

pour tout $n \in \mathbb{N}$, $x_1, \dots, x_n \in \mathcal{X}$ et $a_1, \dots, a_n \in \mathbb{C}$.

Définition 2. *Soit $(\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{H}})$ un espace de Hilbert constitué de fonctions de \mathcal{X} dans \mathbb{C} . La fonction $\kappa(x_i, x_j)$ de $\mathcal{X} \times \mathcal{X}$ dans \mathbb{C} est le noyau reproduisant de \mathcal{H} , sous réserve qu'il en admette un, si et seulement si*

- la fonction $\kappa_{x_i} : x_j \mapsto \kappa_{x_i}(x_j) = \kappa(x_i, x_j)$ appartient à \mathcal{H} , quel que soit $x_i \in \mathcal{X}$ fixé ;
- on a $\psi(x_i) = \langle \psi, \kappa_{x_i} \rangle_{\mathcal{H}}$ pour tout $x_i \in \mathcal{X}$ et $\psi \in \mathcal{H}$.

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

$$\langle \psi, \phi \rangle_{\mathcal{H}_0} = \sum_{i=1}^n \sum_{j=1}^m a_i \bar{b}_j \kappa(x_i, x_j), \quad (2)$$

avec $\psi = \sum_{i=1}^n a_i \kappa_{x_i}$ et $\phi = \sum_{j=1}^m b_j \kappa_{x_j}$ appartenant à l'espace \mathcal{H}_0 . Il s'agit donc là d'un espace pré-hilbertien, que l'on complète conformément à [5] 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 peut démontrer que tout noyau reproduisant est défini positif [5]. A titre d'exemple, parmi les noyaux reproduisants classiques, citons le noyau gaussien $\kappa(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / 2\sigma_0^2)$, où σ_0^2 est un paramètre à fixer. D'autres noyaux, ainsi que des règles permettant de les combiner, sont présentés dans [12, 16].

2.2 Coup du noyau, théorème de représentation

Du second point de la définition 2 résulte une propriété fondamentale des espaces de Hilbert à noyau reproduisant. En remplaçant ψ par κ_{x_j} , on aboutit en effet à

$$\kappa(x_i, x_j) = \langle \kappa_{x_j}, \kappa_{x_i} \rangle_{\mathcal{H}} \quad (3)$$

quels que soient x_i et $x_j \in \mathcal{X}$. Ainsi le noyau κ fournit-il le produit scalaire des images dans \mathcal{H} de toute paire d'éléments de \mathcal{X} , sans qu'il soit nécessaire d'explicitier ces images. Ce principe est appelé *coup du noyau*. Il 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 alors chacun de ces produits scalaires par un noyau non-linéaire. Ainsi la structure des algorithmes demeure-t-elle inchangée et le surcoût calculatoire dû à l'évaluation des noyaux négligeable.

Pour qu'il soit opérationnel, le coup du noyau nécessite souvent d'être associé au théorème de représentation [15]. Il en est par exemple ainsi pour les méthodes kernel-PCA, kernel-FDA et kernel-GDA. Ce théorème établit que toute fonction ψ^* d'un espace de Hilbert à noyau reproduisant \mathcal{H} qui minimise un coût

$$J((x_1, y_1, \psi(x_1)), \dots, (x_n, y_n, \psi(x_n))) + g(\|\psi\|_{\mathcal{H}}^2), \quad (4)$$

impliquant n sorties $\psi(x_i)$ obtenues pour des entrées x_i et éventuellement n sorties désirées y_i , avec g une fonction monotone croissante sur \mathbb{R}^+ , peut s'écrire sous la forme

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

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

3 Mises en œuvre temps-fréquence

L'objectif de cette section est de montrer que les méthodes à noyau peuvent être mise en œuvre dans le domaine temps-fréquence grâce à un choix approprié de noyau reproduisant. On s'intéresse d'abord à la distribution de Wigner, par soucis de clarté, avant d'étendre l'étude à d'autres distributions. On suppose disposer dans la suite d'un ensemble \mathcal{A}_n comprenant n signaux $x_i \in \mathcal{X}$, éventuellement accompagnés d'une étiquette ou autre sortie désirée $y_i \in \mathcal{Y}$.

3.1 Distribution de Wigner

On désigne par W_x la distribution de Wigner de x , c'est-à-dire

$$W_x(t, f) = \int x(t + \tau/2) \bar{x}(t - \tau/2) e^{-2j\pi f\tau} d\tau. \quad (6)$$

Les classiques méthodes paramétriques de reconnaissance des formes, appliquées ici dans le plan temps-fréquence, reposent sur l'estimation de Ψ^* de sorte que la statistique

$$\psi^*(x) = \langle \Psi^*, W_x \rangle = \iint \Psi^*(t, f) W_x(t, f) dt df \quad (7)$$

optimise un critère J donné, par exemple celui de Fisher pour une analyse factorielle discriminante. En pratique,¹ la résolution d'un tel problème est rendue difficile par la taille des représentations W_{x_i} manipulées. Le coup du noyau et le théorème de représentation peuvent ici jouer un rôle déterminant. On considère pour cela le noyau suivant

$$\kappa_W(x_i, x_j) = \langle W_{x_i}, W_{x_j} \rangle. \quad (8)$$

¹Les méthodes de discrétisation de la distribution de Wigner étant multiples, on a pris le parti de raisonner sur la définition (6) dans cet article. En pratique, on rappelle toutefois que les représentations temps-fréquence discrètes de signaux composés de l échantillons sont des matrices de taille $(l \times l)$.

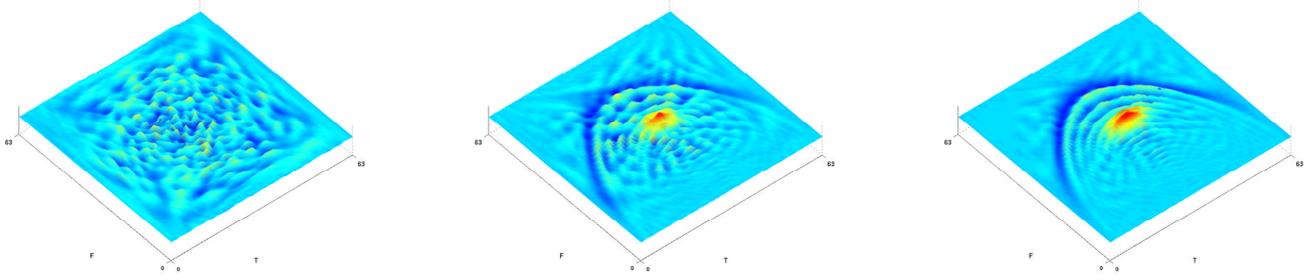


Figure 1: Problème de classification concernant deux familles de 200 signaux de longueur 64, gaussiens (\mathcal{C}_1) et à modulation de fréquence parabolique (\mathcal{C}_2), noyés dans un bruit blanc. Résultats issus d'une kernel-FDA, sans régularisation à gauche et avec régularisation au milieu, et de l'application d'un algorithme SVM à droite. Le noyau quadratique (9) et la relation de reconstruction (10) ont été utilisés.

Notons immédiatement qu'il ne nécessite aucun calcul de distributions de Wigner puisque, par la relation de Moyal, on a

$$\kappa_W(x_i, x_j) = |\langle x_i, x_j \rangle|^2. \quad (9)$$

On vérifie qu'il s'agit d'un noyau défini positif. La condition (1), qui peut s'exprimer sous la forme $\|\sum_i a_i W_{x_i}\|^2 \geq 0$, est en effet satisfaite. Il lui correspond donc un espace de Hilbert à noyau reproduisant \mathcal{H}_W unique. On l'obtient en complétant l'espace fonctionnel \mathcal{H}_0 défini ci-dessous de sorte que toute suite de Cauchy y converge, ce que l'on note $\mathcal{H}_W = \overline{\mathcal{H}_0}$.

$$\mathcal{H}_0 = \{\psi : \mathcal{X} \rightarrow \mathbb{R} \mid \psi = \sum_i a_i |\langle \cdot, x_i \rangle|^2, a_i \in \mathbb{R}, x_i \in \mathcal{X}\}$$

Rien ne s'oppose désormais à la mise en œuvre du coup du noyau et à l'application du théorème de représentation, ou plus généralement à l'association du noyau (8)-(9) à tout algorithme à noyau décrit dans la littérature. Le résultat (5) admet alors une interprétation temps-fréquence implicite, que l'on exhibe en identifiant Ψ^* figurant dans l'expression (7). Ainsi, on a

$$\Psi^* = \sum_{i=1}^n a_i^* W_{x_i} \quad (10)$$

où les échantillons x_i désignent les éléments de l'ensemble d'apprentissage \mathcal{A}_n . On insiste sur le fait que l'estimation des paramètres a_i ne nécessite aucun calcul de distributions temps-fréquence puisqu'elle repose sur le noyau (9). En revanche, si on le souhaite, un calcul itératif permet de déterminer Ψ^* dans un second temps sans qu'il ait jamais été nécessaire de conserver en mémoire l'ensemble des distributions de Wigner des éléments de la base d'apprentissage.

3.2 Extension à d'autres distributions

3.2.1 Transformée de Fourier à court-terme

On s'intéresse tout d'abord à une distribution temps-fréquence linéaire : la transformée de Fourier à court-terme, notée F_x . Elle est unitaire à condition que la fenêtre d'analyse soit d'énergie unité, ce qui signifie

$$\langle F_{x_i}, F_{x_j} \rangle = \langle x_i, x_j \rangle. \quad (11)$$

Le noyau associé, $\kappa_F(x_i, x_j) = \langle x_i, x_j \rangle$, est évidemment défini positif puisque la condition $\|\sum_i a_i x_i\|^2 \geq 0$ est satisfaite. Toute méthode à noyau peut donc être mise en œuvre sur la transformée de Fourier à court-terme à partir de celui-ci. Dans un second temps, on peut accéder à l'information temps-fréquence extraite grâce à l'expression (5) dans laquelle W_{x_i} aura fait place à F_{x_i} .

3.2.2 Classe de Cohen

On se concentre à présent sur les distributions C_x de la classe de Cohen, caractérisées par une fonction de paramétrisation Φ que l'on a choisie de considérer dans le plan Doppler-retard

$$C_x(t, f) = \iint \Phi(\nu, \tau) A_x(\nu, \tau) e^{-2j\pi(f\tau + \nu t)} d\nu d\tau, \quad (12)$$

où A_x désigne la fonction d'ambiguïté à bande étroite du signal x . Notons en passant que la distribution A_x est unitaire. Tout comme la distribution de Wigner, lui est associé le noyau quadratique (9), et par conséquent le même espace de Hilbert à noyau reproduisant. En outre, leurs signatures Ψ^* respectives sont liées par une transformée de Fourier bidimensionnelle, les paramètres a_k^* obtenus au terme de la minimisation du critère (4) étant identiques dans les deux cas en vertu de l'égalité de Parseval-Plancherel appliquée à (7). Concernant les distributions de la classe de Cohen en général, le noyau reproduisant associé s'exprime ainsi

$$\kappa(x_i, x_j) = \iint |\Phi(\nu, \tau)|^2 A_{x_i}(\nu, \tau) \overline{A_{x_j}(\nu, \tau)} d\nu d\tau. \quad (13)$$

On retrouve à nouveau le noyau quadratique (9) si la condition d'unitarité $|\Phi(\nu, \tau)| = 1$ est satisfaite. Parmi les distributions concernées, on peut citer par exemple celles de Page-Levin et de Rihaczek. Dans les autres cas, les calculs ne se présentent pas aussi favorablement puisqu'il est nécessaire d'explicitier les distributions C_{x_i} pour évaluer leurs produits scalaires deux à deux préalablement à l'application d'une méthode à noyau.

4 Applications

La première application concerne un problème de classification de deux familles de 200 signaux de longueur 64, gaussiens (\mathcal{C}_1) et à modulation de fréquence parabolique (\mathcal{C}_2), noyés dans un bruit blanc gaussien. La figure 1 présente les signatures Ψ^* obtenues au terme de la mise en œuvre de deux algorithmes de classification non-linéaire, kernel-FDA et SVM, sur la base du noyau quadratique (9). Comme le montre la figure de gauche, l'application directe de la méthode kernel-FDA ne met pas en évidence la configuration temps-fréquence des signaux étudiés. Des problèmes d'instabilité numérique identifiés dans [9] en sont la cause, que l'on peut contenir à l'aide d'une technique de régularisation de type Tikhonov comme l'illustre la figure du centre. La figure de droite met en évidence la propriété régularisante des SVM, propriété intégrante du critère de marge

Méthode	Taux d'erreur (%)
kernel-FDA non-régularisée	50
kernel-FDA régularisée	7
SVM	1.5

Figure 2: Comparaison des taux d'erreur obtenus avec les méthodes SVM et kernel-FDA, non-régularisée et régularisée, pour le problème décrit en figure 1.

sur lequel elles reposent et qui a fait leur succès. Le tableau en figure 2 fournit les taux d'erreur associés.

La seconde application est relative à la mise en œuvre de la méthode kernel-PCA, qui effectue une analyse en composantes principales non-linéaires d'un ensemble d'individus. Celle-ci a été appliquée à une famille de 300 signaux de 256 composantes, à modulation de fréquence sinusoïdale et noyés dans un bruit blanc. La figure 3 représente la première composante principale obtenue à partir du noyau (11) et de la transformée de Fourier à court-terme. Les caractéristiques temps-fréquence des signaux étudiés sont clairement identifiables. Notons qu'une analyse en composantes principales appliquée directement aux transformées de Fourier à court-terme des signaux aurait fourni le même résultat. Elle aurait cependant été très difficile à mettre en œuvre puisqu'elle aurait nécessité la manipulation de 300 individus constitués de 256^2 composantes, et notamment la diagonalisation de leur matrice de covariance de taille $256^2 \times 256^2$. L'approche proposée a permis de surmonter cet inconvénient puisqu'elle n'a nécessité que la diagonalisation d'une matrice de Gram, ici de taille 300×300 .

5 Conclusion

La théorie des noyaux reproduisants a permis le développement fulgurant d'une classe d'algorithmes de reconnaissance des formes dont la formulation ne dépend pas de la nature des données traitées, ni de l'espace de représentation adopté pour résoudre les problèmes. Au delà de ce caractère universel, celles que l'on range désormais sous le qualificatif de méthodes à noyau doivent également leur succès à l'essor de la théorie statistique de l'apprentissage, au sein de laquelle la prédiction de leurs performances en généralisation fait aujourd'hui encore l'objet d'études approfondies. L'analyse temps-fréquence traitant d'espaces de représentation particuliers, nous avons montré dans cet article qu'elle peut bénéficier des plus récents développements des méthodes à noyau. Il s'agit à présent de poursuivre cet effort et de profiter de l'éclairage nouveau apporté sur elles par la théorie de l'apprentissage et celle des noyaux reproduisants.

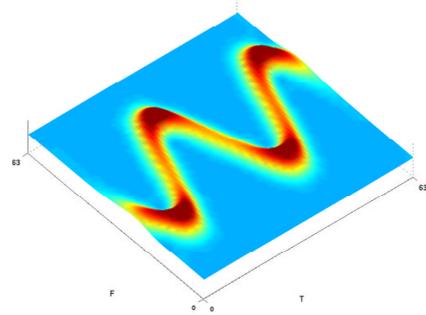


Figure 3: Module-carré de la 1^{ère} composante principale issue d'un algorithme kernel-PCA appliqué à 300 signaux à modulation de fréquence sinusoïdale, noyés dans un bruit blanc.

Références

- [1] F. Auger et P. Flandrin, "Improving the readability of time-frequency and time-scale representations by reassignment methods," *IEEE Transactions on Signal Processing*, vol. 43, no. 5, pp. 1068–1089, 1995.
- [2] D. Jones et R. Baraniuk, "An adaptive optimal-kernel time-frequency representation," *IEEE Transactions on Signal Processing*, vol. 43, no. 10, pp. 2361–2371, 1995.
- [3] J. Gosme, C. Richard, et P. Gonçalves, "Adaptive diffusion of time-frequency and time-scale representations: a review," *IEEE Transactions on Signal Processing (à paraître)*, 2005.
- [4] C. Doncarli et N. Martin, *Décision dans le plan temps-fréquence*. Paris: Hermès Sciences, Traité IC2, 2004.
- [5] N. Aronszajn, "Theory of reproducing kernels," *Transactions of the American Mathematical Society*, vol. 68, pp. 337–404, 1950.
- [6] B. Boser, I. Guyon, et V. Vapnik, "An training algorithm for optimal margin classifiers," in *Proc. Fifth Annual Workshop on Computational Learning Theory*, 1992, pp. 144–152.
- [7] J. Vert, K. Tsuda, et B. Schölkopf, "A primer on kernel methods," in *Kernel Methods in Computational Biology*, B. Schölkopf, K. Tsuda, and J. Vert, Eds. Cambridge, MA: MIT Press, 2004, pp. 35–70.
- [8] B. Schölkopf, A. J. Smola, et K. R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation*, vol. 10, no. 5, pp. 1299–1319, 1998.
- [9] S. Mika, G. Rätsch, J. Weston, B. Schölkopf, et K. R. Müller, "Fisher discriminant analysis with kernels," in *Advances in neural networks for signal processing*, Y. H. Hu, J. Larsen, E. Wilson, and S. Douglas, Eds. San Mateo, CA: Morgan Kaufmann, 1999, pp. 41–48.
- [10] G. Baudat et F. Anouar, "Generalized discriminant analysis using a kernel approach," *Neural Computation*, vol. 12, no. 10, pp. 2385–2404, 2000.
- [11] 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.
- [12] V. N. Vapnik, *The nature of statistical learning theory*. New York, NY: Springer, 1995.
- [13] M. Davy, A. Gretton, A. Doucet, et P. Rayner, "Optimised support vector machines for nonstationary signal classification," *IEEE Signal Processing Letters*, vol. 9, no. 12, pp. 442–445, 2002.
- [14] A. Gretton, M. Davy, A. Doucet, et P. Rayner, "Nonstationary signal classification using support vector machine," in *Proc. IEEE Workshop on Statistical Signal Processing*, 2001.
- [15] 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.
- [16] R. Herbrich, *Learning kernel classifiers. Theory and algorithms*. Cambridge, MA: The MIT Press, 2002.