Contribution à la théorie des noyaux conditionnellement définis positifs et applications

Yves Auffray, Pierre Barbillon, Jean-Michel Marin

Contribution à la théorie des noyaux conditionnellement définis positifs et applications

Yves Auffray*† & Pierre Barbillon*‡ & Jean-Michel Marin§

*Université Paris-Sud, Laboratoire de Mathématiques d'Orsay, Orsay Cedex, F-91405
§Université Montpellier II, I3M, Case Courrier 51, place Eugène Bataillon, 34095 Montpellier cedex 5
†Dassault Aviation, 78 quai Marcel Dassault, Cedex 300, 92552 Saint-Cloud Cedex
‡ INRIA Saclay, équipe Select

Résumé long

Les noyaux conditionnellement définis positifs sont une généralisation des noyaux définis positifs. Le cadre théorique qui est généralement mis en avant pour les justifier est celui développé par Schaback (Schaback, 1999) sous le nom d'espaces natifs (native spaces).
Cependant, cet outil ne supporte pas une généralisation complète du cas des noyaux définis positifs : la définition de conditionnellement défini positif usuellement utilisée exclut de son champ une partie des noyaux définis positifs.
Nous proposons un cadre rénové qui généralise naturellement et totalement le cas défini positif.
Prenons comme point de départ le théorème d'Aronszajn (Aronszajn, 1950) qui est la propriété centrale dans le cas défini positif.
Soit E un ensemble quelconque et $K : E \times E \to \mathbb{R}$ un noyau symétrique, défini positif :

$$\forall (\lambda_1, \mathbf{x}_1) \dots (\lambda_N, \mathbf{x}_N) \in \mathbb{R} \times E, \sum_{1 \leq l,m \leq N} \lambda_l \lambda_m K(\mathbf{x}_l, \mathbf{x}_m) \geq 0$$

Notons

- pour tout $\mathbf{x} \in E$, $K_{\mathbf{x}}$ la fonction partielle $\mathbf{x}' \in E \mapsto K(\mathbf{x}, \mathbf{x}') \in \mathbb{R}$,
- $\mathcal{F}_K$ l'espace vectoriel des combinaisons linéaires (finies) des fonctions partielles $\{K_{\mathbf{x}}, \mathbf{x} \in E\}$.

La forme bilinéaire $<,>_{\mathcal{F}_K}$ induite par K sur $\mathcal{F}_K$ par la formule :

$$< \sum_{k=1}^{L} \lambda_l K_{\mathbf{x}_l}, \sum_{m=1}^{M} \mu_m K_{\mathbf{x}'_m} >_{\mathcal{F}_K} = \sum \sum \lambda_l \mu_m K(\mathbf{x}_l, \mathbf{x}'_m)$$

est symétrique positive.
Le théorème d'Aronszajn s'énonce alors

Théorème (Aronszajn) 1. *$<,>_{\mathcal{F}_K}$, en tant que forme bilinéaire, est* **definie** *positive*

2. *Il existe un unique espace de Hilbert $\mathcal{H}_K$ de fonctions réelles définies sur E tel que*

   - *$(\mathcal{F}_K, <,>_{\mathcal{F}_K})$ est un sous-espace préhilbertien de $(\mathcal{H}_K, <,>_{\mathcal{H}_K})$*
   - *la propriété de reproduction suivante est satisfaite*

$$\forall f \in \mathcal{H}_K,\ \mathbf{x} \in E,\ f(\mathbf{x}) = < f, K_{\mathbf{x}} >_{\mathcal{H}_K}\ . \tag{1}$$

Ce théorème a plusieurs conséquences algorithmiques importantes.
La propriété de reproduction (1) fournit une caractérisation simple et utile de la projection orthogonale $S_{K,\mathbf{X}}(f)$ de $f \in \mathcal{H}_K$ sur $\mathcal{F}_K(\mathbf{X}) = Vect(\{\sum_{l=1}^{N} \lambda_l K_{\mathbf{x}_l}, \lambda_1, \ldots, \lambda_N \in \mathbb{R}\})$: c'est l'interpolateur de f aux points de $\mathbf{X}$ de $\mathcal{H}_K$-norme minimale.
Découle de cela la propriété dite *théorème du représentant* (Kimeldorf and Wahba, 1971) qui, dans le contexte de régression régularisée, s'énonce : $(\mathbf{x}_1, \mathbf{y}_1), \ldots, (\mathbf{x}_N, \mathbf{y}_N) \in E \times \mathbb{R}$ et $\lambda > 0$ étant donnés, toute solution du problème

$$\min_{f \in \mathcal{H}_K} \sum_{k=1}^{N} (\mathbf{y}_k - f(\mathbf{x}_k))^2 + \lambda ||f||_{\mathcal{H}_K} \tag{2}$$

appartient à $\mathcal{F}_K(\mathbf{X})$.
Enfin, l'observation suivante sera utile dans la suite : le qualificatif *défini* dans *noyau défini positif* renvoie au fait que la forme bilinéaire $<,>_{\mathcal{F}_K}$ estdéfinie positive, comme il est spécifié au point 1 du théorème, et non pas au fait que les matrices $(K(\mathbf{x}_l, \mathbf{x}_m))_{1 \leq l,m \leq N}, N \in \mathbb{N}, (\mathbf{x}_1, \ldots, \mathbf{x}_N) \in E^N$ le sont (ce qui est faux en général).

Notre but est maintenant de produire une notion de noyau *conditionnellement* positif qui généralise pleinement le cas défini positif et pour laquelle on puisse établir un analogue du théorème d'Aronszajn.
On suppose que, $\mathbb{P} \subset \mathbb{R}^E$ étant un espace vectoriel de fonctions réelles définies sur E de dimension finie n, on dispose d'un noyau $K : E \times E \to \mathbb{R}$ symétrique et $\mathbb{P}$-conditionnellement positif : $\sum_{1 \leq k,l \leq L} \lambda_l \lambda_k K(\mathbf{x}_l, \mathbf{x}_k) \geq 0$ pour tout $L \in \mathbb{N}$, $\mathbf{x}_1, \ldots, \mathbf{x}_L \in E$, $\lambda_1, \ldots, \lambda_L \in \mathbb{R}$ tels que $\forall p \in \mathbb{P}$, $\sum_{l=1}^{L} \lambda_l p(\mathbf{x}_l) = 0$.

Une partie finie $\boldsymbol{\Upsilon}$ de E, telle que la seule fonction de $\mathbb{P}$ qui s'annule sur $\boldsymbol{\Upsilon}$ est la fonction nulle, est dite $\mathbb{P}$-unisolvante. Les ensembles $\mathbb{P}$-unisolvants minimaux sont ceux qui contiennent exactement n éléments. $\boldsymbol{\Xi} = \{\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n\}$ étant un tel ensemble $\mathbb{P}$-unisolvant minimal, l'application linéaire $L^{\boldsymbol{\Xi}} : p \in \mathbb{P} \mapsto (p(\boldsymbol{\xi}_1), \ldots, p(\boldsymbol{\xi}_n)) \in \mathbb{R}^n$ est bijective. Il en résulte qu'il existe une base de $\mathbb{P}$, $\{h_1^{\boldsymbol{\Xi}}, \ldots, h_n^{\boldsymbol{\Xi}}\}$ satisfaisant $h_i^{\boldsymbol{\Xi}}(\boldsymbol{\xi}_k) = \begin{cases} 1 & \text{si} \quad i = k \\ 0 & \text{sinon} \end{cases}$.

À $\boldsymbol{\Xi}$, on associe le projecteur $\boldsymbol{\pi}^{\boldsymbol{\Xi}} : \mathbb{R}^E \to \mathbb{P}$, $\boldsymbol{\pi}^{\boldsymbol{\Xi}}(f) = \displaystyle\sum_{i=1}^{n} f(\boldsymbol{\xi}_i) h_i^{\boldsymbol{\Xi}}$.

Soit $\mathcal{F}_{K,\mathbb{P}}$ le sous-espace de $(\mathcal{F}_K, <,>_{\mathcal{F}_K})$ constitué des $f = \sum_{l=1}^{L} \lambda_l K_{\mathbf{x}_l} \in \mathcal{F}_K$ tels que $\sum_{l=1}^{L} \lambda_l p(\mathbf{x}_l) = 0$ pour tout $p \in \mathbb{P}$.
L'hypothèse que K est $\mathbb{P}$-conditionnellement positif se traduit par le fait que la restriction de $<,>_{\mathcal{F}_K}$ à $\mathcal{F}_{K,\mathbb{P}}$ est bilinéaire symétrique et positive.

Nous montrons que $\forall f \in \mathcal{F}_{K,\mathbb{P}}, < f, f >_{\mathcal{F}_K} = 0 \iff f \in \mathbb{P}$.
Cela nous conduit naturellement à la définition suivante :

Définition *Un noyau K $\mathbb{P}$-conditionnellement positif est dit $\mathbb{P}$-conditionnellement **défini** positif si $\mathbb{P} \cap \mathcal{F}_{K,\mathbb{P}} = \{0\}$.*

Par exemple,

1. Si $\mathbb{P} = \{0\}$, alors, évidemment, $\mathbb{P} \cap \mathcal{F}_{K,\mathbb{P}} = \{0\}$. On retrouve ainsi la définition usuelle d'un noyau défini positif.

2. Si pour tout $L \in \mathbb{N}$, et pour tout $\mathbf{x}_1, \ldots, \mathbf{x}_L$ deux à deux distincts

$$\left\{ \begin{array}{c} \sum_{1 \le k,l=L} \lambda_l \lambda_k K(\mathbf{x}_l, \mathbf{x}_k) = 0 \\ \text{et} \\ \sum_{l=1}^{L} \lambda_l p(\mathbf{x}_l) = 0, \forall p \in \mathbb{P} \end{array} \right. \Rightarrow \lambda_1 = \ldots = \lambda_L = 0 \,,$$

   on montre facilement que $\mathbb{P} \cap \mathcal{F}_{K,\mathbb{P}} = \{0\}$.
   C'est la définition de noyau $\mathbb{P}$-conditionnellement défini positif généralement adoptée (voir par exemple Schaback (1999)). On voit qu'elle est construite sur l'idée, restrictive comme la remarque qui suit l'énoncé du théorème d'Aronszajn le signale, que les matrices $(K(\mathbf{x}_l, \mathbf{x}_m))_{1 \le l,m \le N}, N \in \mathbb{N}, (\mathbf{x}_1, \ldots, \mathbf{x}_N) \in E^N$ devraient être nécessairement ($\mathbb{P}$-)définies positives pour que le formalisme fonctionne. Ce qui est erroné.

Dans le cadre de cette définition, on peut énoncer un théorème qui généralise le théorème d'Aronszajn.
On définit d'abord, étant donnés $\mathbb{P} \subset \mathbb{R}^E$, $\mathbf{\Xi}$ ensemble $\mathbb{P}$-unisolvant minimal et K noyau $\mathbb{P}$-conditionnellement défini positif, le noyau $K^{\mathbf{\Xi}}$ par :

$$K^{\mathbf{\Xi}}(\mathbf{x}, \mathbf{x}') = K(\mathbf{x}, \mathbf{x}') - \boldsymbol{\pi}^{\mathbf{\Xi}}(K_{\mathbf{x}})(\mathbf{x}') - \boldsymbol{\pi}^{\mathbf{\Xi}}(K_{\mathbf{x}'})(\mathbf{x}) + \sum_{i,j=1}^{n} h_i^{\mathbf{\Xi}}(\mathbf{x}) h_j^{\mathbf{\Xi}}(\mathbf{x}') K(\boldsymbol{\xi}_i, \boldsymbol{\xi}_j) \,.$$

Il est immédiat de constater que c'est un noyau (inconditionnellement) défini positif : on note $\mathcal{H}_{K^{\mathbf{\Xi}}}$ son RKHS associé.

L'analogue du théorème d'Aronszajn s'énonce alors :

Théorème *Il existe un unique espace semi-hilbertien $\left(\mathcal{H}_{K,\mathbb{P}}, <,>_{\mathcal{H}_{K,\mathbb{P}}}\right)$ de fonctions de $\mathbb{R}^E$, satisfaisant les propriétés suivantes :*

- $\left(\mathcal{F}_{K,\mathbb{P}}, <,>_{\mathcal{F}_{K,\mathbb{P}}}\right)$ *est un sous-espace pré-hilbertien de* $\left(\mathcal{H}_{K,\mathbb{P}}, <,>_{\mathcal{H}_{K,\mathbb{P}}}\right)$ *;*

- $\mathbb{P} \subset \mathcal{H}_{K,\mathbb{P}}$ *est l'espace nul de* $<,>_{\mathcal{H}_{K,\mathbb{P}}}$ *;*

- *quel que soit* $\boldsymbol{\Xi}$*, ensemble* $\mathbb{P}$*-unisolvant minimal, la propriété de reproduction suivante est vérifiée :*

$$\forall f \in \mathcal{H}_{K,\mathbb{P}},\ \mathbf{x} \in E,\ f(\mathbf{x}) = \boldsymbol{\pi}^{\boldsymbol{\Xi}}(f)(\mathbf{x}) + < f, K_{\mathbf{x}}^{\boldsymbol{\Xi}} >_{\mathcal{H}_{K,\mathbb{P}}} .$$

Nous appelons l'espace $\mathcal{H}_{K,\mathbb{P}}$ $\mathbb{P}$-espace semi-hilbertien à noyau reproduisant ou $\mathbb{P}$-RKSHS associé à K.
De plus, quelque soit le choix de $\boldsymbol{\Xi}$, ensemble $\mathbb{P}$-unisolvant minimal, on obtient l'identification suivante :

$$\mathcal{H}_{K,\mathbb{P}} = \mathbb{P} \bigoplus \mathcal{H}_{K^{\boldsymbol{\Xi}}}$$

$\boldsymbol{\pi}^{\boldsymbol{\Xi}}$ et $\left(\mathrm{Id}_{\mathcal{H}_{K,\mathbb{P}}} - \boldsymbol{\pi}^{\boldsymbol{\Xi}}\right)$ étant les projecteurs associés à cette décomposition en somme directe et

$$< f, g >_{\mathcal{H}_{K,\mathbb{P}}} = < f - \boldsymbol{\pi}^{\boldsymbol{\Xi}}(f), g - \boldsymbol{\pi}^{\boldsymbol{\Xi}}(g) >_{\mathcal{H}_{K^{\boldsymbol{\Xi}}}} .$$

Fixons un couple $(K, \mathbb{P})$ tel que K soit $\mathbb{P}$-conditionnellement défini positif. Le résultat suivant, qui caractérise géométriquement le meilleur interpolateur, au sens de la semi-norme $||\ ||_{\mathcal{H}_{K,\mathbb{P}}}$, aux points d'un ensemble $\mathbb{P}$-unisolvant $\mathbf{X}$ d'une fonction $f \in \mathcal{H}_{K,\mathbb{P}}$, est établi dans le cadre général que nous proposons.

Proposition *Soit* f *dans* $\mathcal{H}_{K,\mathbb{P}}$.
Pour $\mathbf{X}$*,* $\mathbb{P}$*-unisolvant, on note*

$$\mathcal{F}_{K,\mathbb{P}}(\mathbf{X}) = \{\sum_{l=1}^{N} \lambda_l K_{\mathbf{x}_l}, \lambda_1, \ldots, \lambda_N \in \mathbb{R}\ : \forall p \in \mathbb{P},\ \sum_{i=1}^{N} \lambda_i p(\mathbf{x}_i) = 0\} .$$

1. *Le problème suivant*
$$\min_{g \in \mathbb{P} + \mathcal{F}_{K,\mathbb{P}}(\mathbf{X})} ||f - g||_{\mathcal{H}_{K,\mathbb{P}}} \tag{3}$$
a parmi ses solutions un unique interpolateur de f *sur* $\mathbf{X}$*. Notons* $S_{K,\mathbb{P},\mathbf{X}}(f)$ *cet interpolateur.*

2. *Etant donné un ensemble* $\mathbb{P}$*-unisolvant minimal* $\boldsymbol{\Xi} \subset \mathbf{X}$*, on a*
$$S_{K,\mathbb{P},\mathbf{X}}(f) = \boldsymbol{\pi}^{\boldsymbol{\Xi}}(f) + S_{K^{\boldsymbol{\Xi}},\mathbf{X}}(f - \boldsymbol{\pi}^{\boldsymbol{\Xi}}(f)) \tag{4}$$
où $S_{K^{\boldsymbol{\Xi}},\mathbf{X}} : \mathcal{H}_{K^{\boldsymbol{\Xi}}} \mapsto \mathcal{F}_{K^{\boldsymbol{\Xi}}}(\mathbf{X})$ *est le projecteur orthogonal sur* $\mathcal{F}_{K^{\boldsymbol{\Xi}}}(\mathbf{X}) = \{\sum_{l=1}^{N} \lambda_l K_{\mathbf{x}_l}^{\boldsymbol{\Xi}}, \lambda_1, \ldots, \lambda_N \in \mathbb{R}\}$.

3. $S_{K,\mathbb{P},\mathbf{X}}(f)$ *est l'interpolateur de* f *sur* $\mathbf{X}$ *de semi-norme minimale.*

Les preuves complètes de ce résultat, ainsi que la formulation explicite de l'interpolateur sont présentées dans Auffray and Barbillon (2009).
Le théorème du représentant est une conséquence immédiate de ce qui précède : le problème de régression

$$\min_{f \in \mathcal{H}_{K,\mathbb{P}}} \sum_{k=1}^{N} (\mathbf{y}_k - f(\mathbf{x}_k))^2 + \lambda ||f||^2_{\mathcal{H}_{K,\mathbb{P}}}$$

régularisée au moyen de la semi-norme $|| \ ||_{\mathcal{H}_{K,\mathbb{P}}}$, qui ne pénalise pas les fonctions de $\mathbb{P}$, a ses solutions dans $\mathbb{P} + \mathcal{F}_{K,\mathbb{P}}(\mathbf{X})$, si $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$.
Enfin rappelons le lien qui existe entre l'interpolation dans les RKSHS et la technique du krigeage, très en vogue en géostatistiques et en "computer experiment" (Koehler and Owen, 1996). Le but du krigage est d'approcher une fonction $f \in \mathbb{R}^E$ connue en les point d'un *design* $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\} \subset E$. Sous sa forme élémentaire, elle postule que f est la réalisation d'un processus gaussien de la forme

$$F(\mathbf{x}) = \sum_{i=1}^{n} \beta_i p_i(\mathbf{x}) + Z(\mathbf{x}) \tag{5}$$

où $\mathbb{P} \subset \mathbb{R}^E$ est un espace vectoriel de dimension finie dont une base est $(p_1, \ldots, p_n)$ et Z un processus gaussien centré. Elle propose d'approcher $f(\mathbf{x})$ par le meilleur estimateur linéaire non biaisé (BLUP en anglais).
Mais il est remarquable que le BLUP ne dépende de la fonction aléatoire F que par un processus centré associé $F_{\mathbb{P}}$ défini sur l'espace $\mathcal{M}_{\mathbb{P}}$ des mesures à support fini qui s'annulent sur $\mathbb{P}$, par : $F_{\mathbb{P}}(\sum_{m=1}^{M} \mu_m \delta_{\mathbf{x}_m}) = \sum_{m=1}^{M} \mu_m F(\mathbf{x}_m)$.
D'où l'idée du krigeage dit *intrinsèque* (voir Matheron (1973), Vasquez (2005)) d'oublier la modélisation (5) et de partir d'une fonction aléatoire G à paramètre dans $\mathcal{M}_{\mathbb{P}}$, spécifiée par un noyau $\mathbb{P}$-conditionnellement défini positif, puis de résoudre les équations du BLUP où G joue le rôle précédemment dévolu à $F_{\mathbb{P}}$.
Nous montrons que cette démarche aboutit à la résolution du même système linéaire que celui obtenu pour le calcul de l'interpolateur de f sur $\mathbf{X}$ et ainsi le BLUP se réinterprète comme l'interpolateur décrit dans la proposition précédente.
Mots clés : Noyau (Conditionnellement) Défini positif, RKHS, Espace Natif, Interpolation, Krigeage, Régression Régularisée.

Abstract

Since Aronszajn, it is well known that a functional Hilbert space, called Reproducing Kernel Hilbert Space (RKHS) asociated to K, can be associated to any positive definite kernel K. This correspondance is the basis of many useful algorithms. In the more general

context of conditionally positive definite kernels the *native spaces* are the usual theoretical frame. However, due to a restrictive definition of *conditionally positive definite*, that framework does not provide a full generalisation of the positive definite case. We propose a more natural and general definition from which we state a full generalisation of Aronszajn's theorem. It states that for every couple $(K, \mathbb{P})$ such that K is a $\mathbb{P}$-conditionally definite positive kernel, there is a unique functional semi-Hilbert space of functions $\mathcal{H}_{K,\mathbb{P}}$ satisfying a generalized reproducing property.
Eventually, we verify that this tool, as native spaces, leads to the same interpolation operator than the one provided by the Kriging method and that, using *representer theorem*, we can identify the solution of a regularized regression problem in $\mathcal{H}_{K,\mathbb{P}}$.
Key words : (Conditionally) Positive Definite Kernel, R.K.H.S, Native Space, Interpolation, Kriging, Regularized Regression.

References

Aronszajn, N. (1950). Theory of reproducing kernel. *Transactions of American Mathematical Society*, 68(3):337–404.

Auffray, Y. and Barbillon, P. (2009). Conditionally positive definite kernels : theoritical contribution, application to interpolation and approximation. Technical report, Rapport de recherche INRIA.

Kimeldorf, G. and Wahba, G. (1971). Some results on tchebycheffian spline functions. *Journal of Mathematical Analysis and Applications*, 33(1):82–95.

Koehler, J. R. and Owen, A. B. (1996). Computer experiments. In *Design and analysis of experiments*, volume 13 of *Handbook of Statist.*, pages 261–308. North-Holland, Amsterdam.

Matheron, G. (1973). The intrinsic random function. *Adv. Appi Prob*, 5:439–468.

Schaback, R. (1999). Native Hilbert spaces for radial basis functions. I. In *New developments in approximation theory (Dortmund, 1998)*, volume 132 of *Internat. Ser. Numer. Math.*, pages 255–282. Birkhäuser, Basel.

Vasquez, E. (2005). *Modélisation comportementale de systèmes non-linéaires multivariables par méthodes à noyaux et applications.* PhD thesis, Université Paris-sud.