Chapter 1

# THEORY

## 1. INTRODUCTION

A Reproducing Kernel Hilbert Space (RKHS) is first of all a Hilbert space, that is, the most natural extension of the mathematical model for the actual space where everyday life takes place (the Euclidean space $\mathbb{R}^3$). When studying elements of some abstract set $\mathcal{S}$ it is convenient to consider them as elements of some other set $\mathcal{S}'$ on which is already defined a structure relevant to the problem to be treated. It can be for instance an order structure, a vector structure, a metric structure or a mixing of algebraic and topological structures. For this we need an "imbedding theorem" or a "representation theorem". Through this kind of theorem the study of elements of $\mathcal{S}$ is transferred to their "representers" in $\mathcal{S}'$ and can be carried out using the structure on $\mathcal{S}'$. For their richness and simplicity Hilbert spaces are introduced as often as possible when a vector structure and an inner product can be exploited. They provide powerful mathematical tools and geometric concepts on which our intuition can rest. The phrase "RKHS method" is generic to name a method based on the embedding of the abstract set $\mathcal{S}$ into some RKHS $\mathcal{S}'$.

We will see that RKHS are spaces of functions with the nice property that if a function $f$ is close to a function $g$ in the sense of the distance derived from the inner product, then the values $f(x)$ are close to the values $g(x)$. This property has consequences which are desirable in a wide variety of applications as we shall see throughout this book. When a space of functions is endowed with a "sup" distance on some set $E$

$$d(f, g) = \sup_{x \in E} |f(x) - g(x)| ,$$

the closeness of two functions implies closeness of their values but it is not the rule in general. Consider, for instance, the space of polynomials over $[0, 1]$, endowed with the $L^p$ distance

$$d(P_1, P_2) = \left( \int_0^1 |P_1(x) - P_2(x)|^p \, d\lambda(x) \right)^{1/p},$$

where $p > 0$ and $\lambda$ denotes the Lebesgue measure on the set $\mathbb{R}$ of real numbers. In this space, the sequence of polynomials $(Q_n)_{n \geq 0}$, with $Q_n(x) = x^n$, tends to the null function because

$$d(Q_n, 0) = \left( \int_0^1 x^{np} \, d\lambda(x) \right)^{1/p} = (np+1)^{(-1/p)}$$

while the sequence $(Q_n(1))_{n \geq 0}$ is constant and equal to 1 (Figure 1).

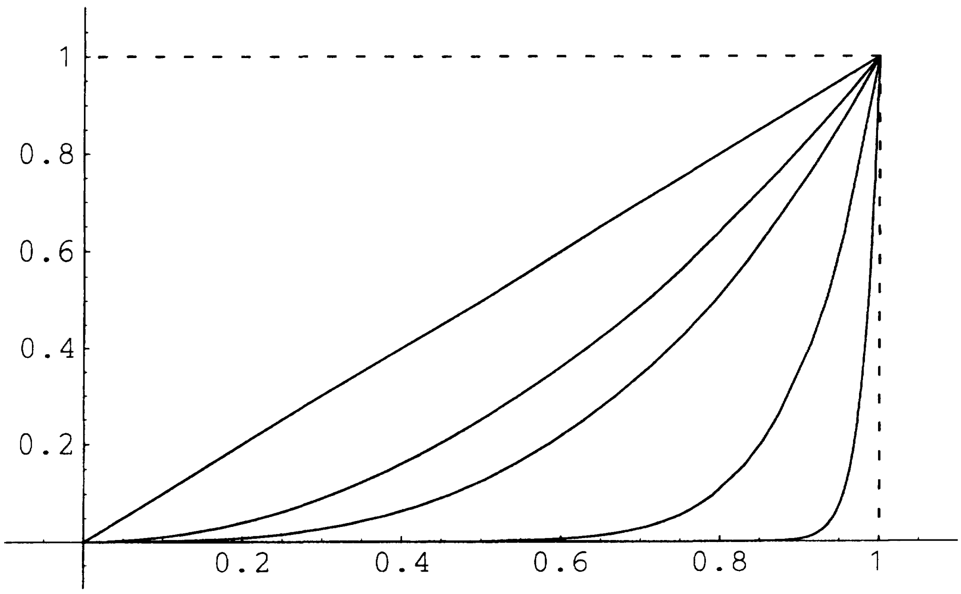


Figure 1.1: The functions $x, x^2, x^3, x^{10}, x^{50}$ over $[0, 1]$

On this space of polynomials the evaluation of functions at the point 1, that is the application

$$P \longmapsto P(1)$$

is not continuous.

Among Hilbert spaces of functions, RKHS are characterized by the property that the evaluation of functions at a fixed point $x$, $f \longmapsto f(x)$ is a continuous mapping. The reproducing kernel of such a space $\mathcal{H}$ is a function of two variables $K(y, x)$ with the property that for fixed $x$, the function of $y$, $K(y,x)$, denoted by $K(.,x)$ belongs to $\mathcal{H}$ and represents the evaluation function at the point $x$ (this will be made precise in Definition 1). Lemma 2 in Subsection 3 claims that reproducing kernels are positive type functions. One of the most remarkable theorem of this theory is that the converse is true. To prove this we first characterize subspaces of functions endowed with an inner product, that are embedded in RKHS. Then, using this characterization and considering the space spanned by the functions $(K(.,x))_{x\in E}$, where $K$ is a positive type function, we exhibit a Hilbert space with reproducing kernel $K$. In Subsection 4 we deal with operations on reproducing kernels, sum of reproducing kernels, restriction of a reproducing kernel, reproducing kernel of a closed subspace. Then we look at some particular cases, separable RKHS and spaces of continuous functions. Section 6 and 7 are devoted to extensions of Aronszajn's theory which will be used later on. They can be skipped at first reading.

## 2. NOTATION AND BASIC DEFINITIONS

Let $E$ be a non empty abstract set. Let $\mathcal{H}$ be a vector space of functions defined on $E$ and taking their values in the set $\mathbb{C}$ of complex numbers. $\mathcal{H}$ is endowed with the structure of Hilbert space defined by an inner product $< .,. >_{\mathcal{H}}$

$$\begin{array}{rcl} \mathcal{H}\times\mathcal{H} & \longrightarrow & \mathbb{C} \\ (\varphi,\psi) & \longmapsto & <\varphi,\psi>_{\mathcal{H}}\,. \end{array}$$

Let $||.||_{\mathcal{H}}$ denote the associated norm:

$$\forall\varphi\in\mathcal{H},\quad ||\varphi||_{\mathcal{H}} = <\varphi,\varphi>_{\mathcal{H}}^{1/2}\,.$$

For any $t\in E$, we will denote by $e_t$ the evaluation functional at the point $t$, *i.e.* the mapping

$$\begin{array}{rcl} \mathcal{H} & \longrightarrow & \mathbb{C} \\ g & \longmapsto & e_t(g) = g(t). \end{array}$$

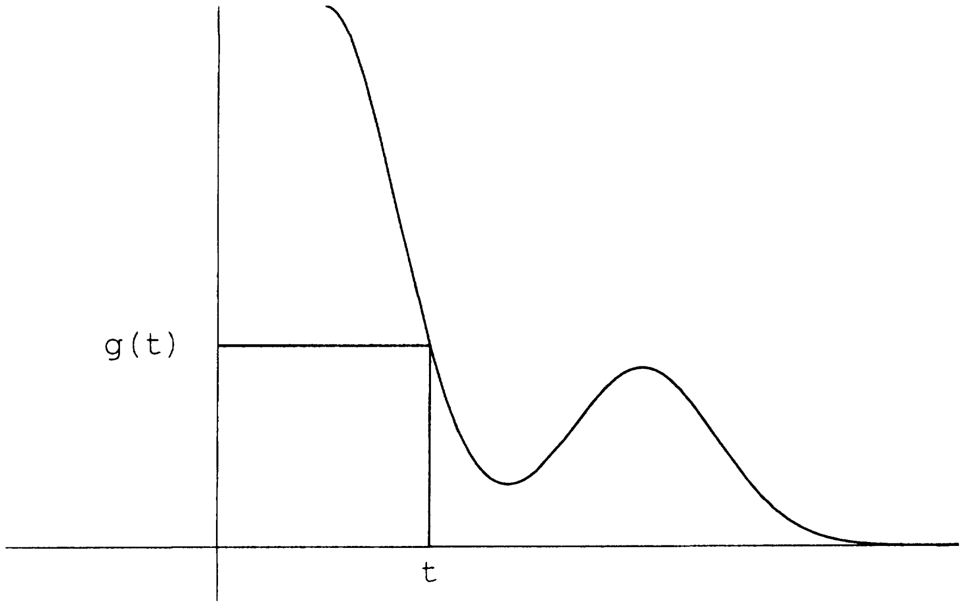


Figure 1.2: The evaluation functional $e_t$ associates with any function $g$ its value $g(t)$ at the point $t$.

A complex number is written $z = x + \mathrm{i}y$ where

$$\begin{aligned} x &= \Re(z) \quad \text{is the real part of } z \\ \text{and } y &= \Im(z) \quad \text{is the imaginary part.} \end{aligned}$$

The conjugate of $z$ is $\overline{z} = x - \mathrm{i}y$. A complex-valued function defined on $E$ will be denoted either by a single letter e.g. $f$ or $\varphi$, or by $f(.)$ or $\varphi(.)$. The conjugate of a complex number or matrix or function is written with a bar $(\overline{x}, \overline{M}, \overline{f})$, and the adjoint or transconjugate of a matrix $M$ is written $M^*$. For any function $K$ on $E \times E$ the mappings $s \longmapsto K(s,t)$ with fixed $t$ (resp. $t \longmapsto K(s,t)$ with fixed $s$) will be denoted by $K(.,t)$ (resp. $K(s,.)$).
The notation $\mathbb{C}^E$ will be used for the set of complex functions defined on $E$.
For basic definitions about Hilbert or pre-Hilbert spaces and applications to Functional Analysis see for instance the books by Bourbaki, Dieudonné, Dudley (1989) or Rudin. Let us now introduce some examples of such spaces which will be used in the sequel.
**Example 1** Let $\mathcal{H}$ be a finite dimensional complex vector space of functions with basis $(f_1, f_2, \ldots, f_n)$. Any vector of $\mathcal{H}$ can be written in a unique way as a linear combination of $f_1, f_2, \ldots, f_n$. Therefore an inner product $< .,. >_{\mathcal{H}}$ on $\mathcal{H}$ is entirely defined by the numbers

$$g_{ij} = < f_i, f_j > , 1 \leq i, j \leq n.$$

If

$$v = \sum_{i=1}^{n} v_i f_i \quad \text{and} \quad w = \sum_{j=1}^{n} w_j f_j,$$

then

$$< v, w >_{\mathcal{H}} = < \sum_{i=1}^{n} v_i f_i, \sum_{i=1}^{n} w_i f_i >_{\mathcal{H}} = \sum_{i=1}^{n} \sum_{j=1}^{n} v_i \overline{w}_j g_{ij}.$$

The matrix $G = (g_{ij})$ is called the Gram matrix of the basis. $G$ is hermitian ($G = G^*$) and positive definite ($v^* G v > 0$ whenever $v \neq 0$). A finite dimensional space endowed with any inner product is always complete (any Cauchy sequence is convergent) and therefore it is a Hilbert space.

A particular finite dimensional space will play a key role in the theory of higher order kernels (Chapter 3), the space $\mathbb{P}_r$ of polynomials of degree at most $r$. Let $K_0$ be a probability density on $\mathbb{R}$, that is a nonnegative integrable function satisfying

$$\int_{\mathbb{R}} K_0(x)\, d\lambda(x) = 1.$$

Suppose that $K_0$ has finite moments up to order $2r$ ($r$ nonnegative integer). Then the space $\mathbb{P}_r$ is a Hilbert space with the inner product

$$< P, Q >_{K_0} = \int_{\mathbb{R}} P(x) Q(x) K_0(x)\, d\lambda(x). \tag{1.1}$$

(See Exercise 1). The Gram matrix of the canonical basis $1, x, x^2, \ldots, x^r$ is the $(r+1) \times (r+1)$ matrix $(g_{ij})$ defined by

$$g_{ij} = \int_{\mathbb{R}} x^{i+j-2} K_0(x)\, d\lambda(x), \quad 1 \leq i \leq r+1, 1 \leq j \leq r+1,$$

that is the Hankel matrix of moments of $K_0$.

**Example 2** Let $E = \mathbb{N}^*$ be the set of positive integers and let $\mathcal{H} = l^2(\mathbb{C})$ be the set of complex sequences $(x_i)_{i \in \mathbb{N}^*}$ such that $\sum_{i \in \mathbb{N}^*} |x_i|^2 < \infty$. $\mathcal{H}$ is a Hilbert space with the inner product

$$< x, y >_{l^2(\mathbb{C})} = \sum_{i \in \mathbb{N}} x_i \overline{y}_i \qquad \text{if} \quad x = (x_i) \quad \text{and} \quad y = (y_i).$$

**Example 3** Let $E = (a, b), -\infty \leq a < b \leq \infty$ and $\mathcal{L}^2(a, b)$ be the set of complex measurable functions over $(a, b)$ such that

$$\int_a^b |f(x)|^2\, d\lambda(x) < \infty.$$

Identifying two functions $f$ and $g$ of $\mathcal{L}^2(a,b)$ which are equal except on a set of Lebesgue measure equal to zero, we get a vector space $L^2(a,b)$ which is a Hilbert space with the inner product

$$< f, g >_{L^2(a,b)} = \int_a^b f(x)\overline{g(x)}\, d\lambda(x).$$

**Example 4** Let $E = (0,1)$ and

$$\mathcal{H} = \left\{\varphi \mid \varphi(0) = 0,\ \varphi \text{ is absolutely continuous and } \varphi' \in L^2(0,1)\right\},$$

where $\varphi'$ is defined almost everywhere as the derivative of $\varphi$. $\mathcal{H}$ is a Hilbert space with the inner product

$$< \varphi, \psi > = \int_0^1 \varphi' \overline{\psi}' d\lambda.$$

$\mathcal{H}$ belongs to the class of Sobolev spaces (Adams, 1975). Basic definitions and properties of these spaces are given in the appendix.
**Example 5** Let $E = \mathbb{R}$ and

$$\mathcal{H} = H^1(\mathbb{R}) = \left\{\varphi | \varphi \text{ is absolutely continuous, } \varphi \text{ and } \varphi' \text{ are in } L^2(\mathbb{R})\right\},$$

where $\varphi'$ is (almost everywhere) the derivative of $\varphi$.
$\mathcal{H}$ is a Hilbert space with the inner product

$$< \varphi, \psi >_{\mathcal{H}} = \int_{\mathbb{R}} (\varphi \overline{\psi} + \varphi' \overline{\psi}')d\lambda.$$

As in Example 4, $\mathcal{H}$ belongs to the class of Sobolev spaces.
**Example 6** Let $(\Omega, \mathcal{A}, P)$ be a probability space, let $\mathcal{F}$ be a Hilbert space with inner product $< .,. >_{\mathcal{F}}$ and associated norm $||.||_{\mathcal{F}}$ and let $\mathcal{L}^2(\Omega, \mathcal{A}, P)$ be the set of random variables $X$ with values in $\mathcal{F}$ such that

$$E_P\left(||X||^2_{\mathcal{F}}\right) = \int ||X||^2_{\mathcal{F}}\ dP < \infty.$$

Identifying two random variables $X$ and $Y$ such that $P(X \neq Y) = 0$ we get the space $L^2_{\mathcal{F}}(\Omega, \mathcal{A}, P)$ which is a Hilbert space when endowed with the inner product

$$< X, Y > = E_P\left(< X, Y >_{\mathcal{F}}\right).$$

Let us now introduce the definition of a reproducing kernel.

DEFINITION 1 (REPRODUCING KERNEL) *A function*

$$\begin{array}{rcl} K : E \times E & \longrightarrow & \mathbb{C} \\ (s,t) & \longmapsto & K(s,t) \end{array}$$

*is a reproducing kernel of the Hilbert space $\mathcal{H}$ if and only if*
*a)* $\forall t \in E, \quad K(.,t) \in \mathcal{H}$
*b)* $\forall t \in E, \quad \forall \varphi \in \mathcal{H} \quad < \varphi, K(.,t) >= \varphi(t).$

This last condition is called "the reproducing property": the value of the function $\varphi$ at the point $t$ is reproduced by the inner product of $\varphi$ with $K(.,t)$. From a) and b) it is clear that

$$\forall (s,t) \in E \times E \qquad K(s,t) =< K(.,t), K(.,s) > .$$

A Hilbert space of complex-valued functions which possesses a reproducing kernel is called

*a reproducing kernel Hilbert space (RKHS)*

or

*a proper Hilbert space.*

The first terminology will be adopted in the sequel. Let us examine the Hilbert spaces listed in Examples 1 to 6 above.
**Example 1** Let $(e_1, e_2, \ldots, e_n)$ be an orthonormal basis in $\mathcal{H}$ and define

$$K(x,y) = \sum_{i=1}^{n} e_i(x)\overline{e}_i(y).$$

Then for any $y$ in $E$ ,

$$K(.,y) = \sum_{i=1}^{n} \overline{e}_i(y)\, e_i(.)$$

belongs to $\mathcal{H}$ and for any function

$$\varphi(.) = \sum_{i=1}^{n} \lambda_i\, e_i(.)$$

in $\mathcal{H}$, we have

$$\forall y \in E \qquad < \varphi, K(.,y) >_{\mathcal{H}} \;=\; < \sum_{i=1}^{n} \lambda_i\, e_i(.), \sum_{j=1}^{n} \overline{e}_j(y)\, e_j(.) >_{\mathcal{H}}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \bar{\bar{e}}_j(y) < e_i, e_j >_{\mathcal{H}}$$
$$= \sum_{i=1}^{n} \lambda_i e_i(y) = \varphi(y).$$

Any finite dimensional Hilbert space of functions has a reproducing kernel.

**Example 2** Let $K(i,j) = \delta_{ij}$ (*delta* function or Kronecker symbol, equal to 1 if $i = j$, to 0 otherwise). Then

$$\forall j \in \mathbb{N} \qquad K(.,j) = (0,0,\ldots,0,1,0,\ldots) \in \mathcal{H} \text{ (1 at the } j\text{-th place)}$$
$$\forall j \in \mathbb{N} \qquad \forall x = (x_i)_{i\in\mathbb{N}} \in \mathcal{H} \qquad < x, K(.,j) >_{\mathcal{H}} = \sum_{i\in\mathbb{N}} x_i \bar{\delta}_{ij} = x_j.$$

$K$ is the reproducing kernel of $\mathcal{H}$.

**Example 3** $\mathcal{H} = L^2(a,b)$ is rather a space of classes of functions than a space of functions, thus Definition 1 does not strictly apply in that case. However we could wonder whether there exists, for $t \in (a,b)$, a class of functions $K(.,t)$ such that

$$\forall \varphi \in L^2(a,b) \qquad \int_{[a,b]} \varphi\, \overline{K(.,t)}\, d\lambda = \varphi(t) \qquad \text{a.s.} \tag{1.2}$$

The answer is negative. A theorem of Yosida states that the identity is not an integral operator in $L^2(a,b)$ *i.e.* (1.2) cannot be satisfied. Applied to nonnegative functions $\varphi$ integrating to one over the interval $(a,b)$ and belonging to $L^2(a,b)$, the above formula means that we would be able to estimate unbiasedly a probability density function from one observation of a random variable with this density. In Subsection 7 it will be shown that this is impossible.

**Example 4** $\mathcal{H}$ has reproducing kernel $K(x,y) = \min(x,y)$. The weak derivative (see Appendix) of $\min(.,y)$ is the function $\mathbf{1}_{(0,y)}$ and

$$< \varphi, K(.,y) >_{\mathcal{H}} = \int_0^y \varphi'(x)\, d\lambda(x) = \varphi(y).$$

**Example 5** A simple integration by parts shows that $\mathcal{H}$ has the reproducing kernel

$$K(x,y) = \frac{1}{2}\exp(-|x-y|).$$

We have

$$\frac{\partial}{\partial x} K(x,y) = \begin{cases} -K(x,y) & \text{if} \quad x > y \\ K(x,y) & \text{if} \quad x < y \end{cases}$$

and

$$\frac{\partial^2}{\partial x^2} K(x,y) = K(x,y) \quad \text{if} \quad x \neq y.$$

For $\varphi$ and $\psi$ in $\mathcal{H}$, with $\psi$ twice differentiable except, possibly, at the point $y$, we have

$$\begin{aligned}
\int_{\mathbb{R}} \varphi' \overline{\psi}' \, d\lambda &= \int_{-\infty}^{y} \varphi' \overline{\psi}' \, d\lambda + \int_{y}^{\infty} \varphi' \psi' \, d\lambda \\
&= [\varphi \overline{\psi}']_{-\infty}^{y} - \int_{-\infty}^{y} \varphi' \overline{\psi}'' \, d\lambda + [\varphi \overline{\psi}']_{y}^{\infty} - \int_{y}^{\infty} \varphi \overline{\psi}'' \, d\lambda.
\end{aligned}$$

As $K(y,y) = 1/2$, taking $\psi(.) = K(.,y)$ and using the above formulas for the derivatives of $K(.,y)$, one gets

$$\int_{\mathbb{R}} \varphi' \overline{\psi}' \, d\lambda = \varphi(y) - \int_{\mathbb{R}} \varphi \overline{\psi}'' \, d\lambda = \varphi(y) - \int_{\mathbb{R}} \varphi \overline{\psi}' \, d\lambda$$

Hence,

$$< \varphi, \psi >_{\mathcal{H}} = \int_{\mathbb{R}} \varphi \overline{\psi} \, d\lambda + \int_{\mathbb{R}} \varphi' \overline{\psi}' \, d\lambda = \varphi(y),$$

and $K$ is the reproducing kernel of $\mathcal{H}$.
**Example 6** See the particular case $\mathcal{F} = \mathbb{C}$ in Example 3.
Theorem 15 below, its Corollary and Exercise 11 give ways of constructing Hilbert spaces of functions with no reproducing kernel. Despite the fact that all infinite dimensional separable Hilbert spaces are isomorphic to $l^2(\mathbb{C})$ which has a reproducing kernel (Example 2), this property gives no guarantee that a given separable Hilbert space of functions has a reproducing kernel.
Riesz's representation theorem will provide the first characterization of RKHS.

THEOREM 1 *A Hilbert space of complex valued functions on $E$ has a reproducing kernel if and only if all the evaluation functionals $e_t, t \in E$, are continuous on $\mathcal{H}$.*

**Proof.** If $\mathcal{H}$ has a reproducing kernel $K$ then for any $t \in E$, we have

$$\forall \varphi \in \mathcal{H} \qquad e_t(\varphi) = < \varphi, K(.,t) >_{\mathcal{H}} .$$

Thus the evaluation functional $e_t$ is linear and, by the Cauchy-Schwarz inequality, continuous:

$$|e_t(\varphi)| = | < \varphi, K(.,t) >_{\mathcal{H}} | \quad \leq \quad \|\varphi\| \, \|K(.,t)\| = \|\varphi\| \, [K(t,t)]^{1/2}.$$

Moreover, for $\varphi = K(.,t)$, the upper bound is obtained so that the norm of the continuous linear functional $e_t$ is given by

$$\|e_t\| = \sup_{\|\varphi\| \neq 0} \frac{|e_t(\varphi)|}{\|\varphi\|} = [K(t,t)]^{1/2}.$$

Conversely, from Riesz's representation theorem, if the linear mapping

$$\begin{array}{ccc} \mathcal{H} & \longrightarrow & \mathbb{C} \\ \varphi & \longmapsto & e_t(\varphi) = \varphi(t) \end{array}$$

is continuous, there exists a function $N_t(.)$ in $\mathcal{H}$ such that

$$\forall \varphi \in \mathcal{H} < \varphi, N_t >= \varphi(t).$$

If this property holds for any $t \in E$, then it is clear that $K(s,t) = N_t(s)$ is the reproducing kernel of $\mathcal{H}$. ■

COROLLARY 1 *In a RKHS a sequence converging in the norm sense converges pointwise to the same limit.*

**Proof.** If $(\varphi_n)$ converges to $\varphi$ in the norm sense we have, for any $t \in E$,

$$|\varphi_n(t) - \varphi(t)| = e_t(\varphi_n) - e_t(\varphi)$$

and $(e_t(\varphi_n))$ converges to $e_t(\varphi)$ by continuity of $e_t$. ■

Now the question arises of characterizing reproducing kernels.

| When is a complex-valued function $K$ defined on $E \times E$ a reproducing kernel? |
|---|

The aim of the next subsection is to prove that the set of positive type functions and the set of reproducing kernels on $E \times E$ are identical. For this purpose a definition is needed.

DEFINITION 2 (POSITIVE TYPE FUNCTION) *A function $K : E \times E \longrightarrow \mathbb{C}$ is called a positive type function (or a positive definite function) if*

$$\forall n \geq 1,\ \forall (a_1, \ldots, a_n) \in \mathbb{C}^n,\ \forall (x_1, \ldots, x_n) \in E^n, \tag{1.3}$$

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i \overline{a}_j K(x_i, x_j) \in \mathbb{R}^+,$$

*where* $\mathbb{R}^+$ *denotes the set of nonnegative real numbers.*

It is worth noting that Condition (1.3) given in Definition 2 is equivalent to the positive definiteness of the matrix

$$(K(x_i, x_j))_{1 \leq i,j \leq n}$$

for any choice of $n \in \mathbb{N}^*$ and $(x_1, \ldots, x_n) \in E^n$.

**Examples of positive type functions**

- Any constant non negative function on $E \times E$ is of positive type since

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i \overline{a}_j = \left| \sum_{i=1}^{n} a_i \right|^2 \in \mathbb{R}^+.$$

- The delta function

$$(x, y) \longmapsto \delta_{xy} = \begin{cases} 1 & if \quad x = y \\ 0 & if \quad x \neq y \end{cases}$$

is of positive type.

**Proof.** Let $n \geq 1$, $(a_1, \ldots, a_n) \in \mathbb{C}^n$, $(x_1, \ldots, x_n) \in E^n$ and $\{\alpha_1, \ldots, \alpha_p\}$ the set of different values among $x_1, \ldots, x_n$. We can write

$$\begin{aligned} \sum_{i=1}^{n} \sum_{j=1}^{n} a_i \overline{a}_j \delta_{x_i x_j} &= \sum_{i=1}^{n} \sum_{x_j = x_i} a_i \overline{a}_j \\ &= \sum_{k=1}^{p} \sum_{x_i = x_j = \alpha_k} a_i \overline{a}_j \\ &= \sum_{k=1}^{p} \left| \sum_{x_i = \alpha_k} a_i \right|^2 \in \mathbb{R}^+. \end{aligned}$$

∎

- The product $\alpha K$ of a positive type function $K$ with a non negative constant $\alpha$ is a positive type function.

| How to prove that a given function is of positive type? |
| --- |

Direct verification of (1.3) is often untractable. Another possibility is to use the following lemma.

LEMMA 1 *Let $\mathcal{H}$ be some Hilbert space with inner product $< ., . >_{\mathcal{H}}$ and let $\varphi : E \to \mathcal{H}$. Then, the function $K$*

$$\begin{array}{rcl} E \times E & \longrightarrow & \mathbb{C} \\ (x, y) & \longmapsto & K(x, y) =< \varphi(x), \varphi(y) >_{\mathcal{H}} \end{array}$$

*is of positive type.*

**Proof.** The conclusion easily follows from the following equalities

$$\begin{aligned} \sum_{i=1}^{n}\sum_{j=1}^{n} a_i \bar{a}_j K(x_i, x_j) &= \sum_{i=1}^{n}\sum_{j=1}^{n} < a_i\varphi(x_i), a_j\varphi(x_j) >_{\mathcal{H}} \\ &= \left\| \sum_{i=1}^{n} a_i\varphi(x_i) \right\|_{\mathcal{H}}^2 . \end{aligned}$$

■

Lemma 1 tells us that writing

$$K(x, y) =< \varphi(x), \varphi(y) >_{\mathcal{H}}$$

in some space $\mathcal{H}$ is sufficient to prove positive definiteness of $K$.
A particular space $\mathcal{H}$ used in Probability Theory is the space $L^2(\Omega, \mathcal{A}, P)$ of square integrable random variables on some probability space $(\Omega, \mathcal{A}, P)$ (see Example 6). In this context, to prove that a function $K$ under consideration is of positive type one proves that it is the covariance function of some complex valued zero mean stochastic process $(X_t)_{t\in E}$.
As we have

$$0 \leq \mathrm{Var}\left(\sum_{i=1}^{n} a_i X_{t_i}\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i \bar{a}_j E(X_{t_i}\overline{X}_{t_j}),$$

the covariance function

$$\begin{array}{rcl} E \times E & \longrightarrow & \mathbb{C} \\ E(X_t\overline{X}_s) =< X_t, X_s >_{L^2(\Omega,\mathcal{A},P)} & & \end{array}$$

is of positive type. It is enough to prove that for any choice of $n \in \mathbb{N}^*$ and $(t_1, \ldots, t_n) \in E^n$, the matrix

$$(K(t_i, t_j))_{1 \leq i,j \leq n}$$

is the covariance matrix of some zero mean random vector. To appreciate the usefulness of the above lemma, try to prove directly (using Definition 2) that

$$(t, s) \longmapsto \min(t, s)$$

is of positive type and then see Exercise 7.
An important question in practice is to determine whether a given function belongs to a given RKHS. Roughly speaking, a function $f$ belongs to a RKHS only if it is at least as smooth as the kernel, since any function of the RKHS is either a linear combination of kernels or a limit of such combinations. More precise arguments will be given in Theorem 43 of Chapter 2.

## 3. REPRODUCING KERNELS AND POSITIVE TYPE FUNCTIONS

In the equivalence between reproducing kernels and positive type functions the following implication is clear.

LEMMA 2 *Any reproducing kernel is a positive type function.*

**Proof.** If $K$ is the reproducing kernel of $\mathcal{H}$ we have

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i\overline{a}_j K(x_i, x_j) = \|\sum_{i=1}^{n} \overline{a}_i K(.,x_i)\|^2 \in \mathbb{R}^+.$$

∎

Before proving the converse let us begin with some properties of positive type functions.

LEMMA 3 *Let $L$ be any positive type function on $E \times E$. Then*
*a) $\forall x \in E \qquad L(x,x) \in \mathbb{R}^+$*
*b) $\forall (x,y) \in E \times E \qquad L(x,y) = \overline{L(y,x)}$*
*c) $\overline{L}$ is a positive type function*
*d) $|L(x,y)|^2 \le L(x,x)\ L(y,y)$.*

**Proof.** a) is clear. Take $n = 1$ and $a_1 = 1$ in Definition 2.
b) Let $(x,y) \in E \times E$. From (1.3) the number

$$C(\alpha,\beta) = |\alpha|^2 L(x,x) + \alpha\overline{\beta}L(x,y) + \beta\overline{\alpha}L(y,x) + |\beta|^2 L(y,y)$$

is, for any $(\alpha,\beta) \in \mathbb{C}^2$, a nonnegative real number. Thus, putting first $\alpha = \beta = 1$ and then $\alpha = \mathbf{i}$ and $\beta = 1$, one gets

$$L(x,y) + L(y,x) = C(1,1) - L(x,x) - L(y,y) = A$$

and

$$\mathbf{i}L(x,y) - \mathbf{i}L(y,x) = C(\mathbf{i},1) - L(x,x) - L(y,y) = B.$$

Hence

$$\begin{array}{rrcll} & L(x,y) & + & L(y,x) & = A \in \mathbb{R}. \\ \text{and} & \mathbf{i}L(x,y) & - & \mathbf{i}L(y,x) & = B \in \mathbb{R}. \end{array}$$

It follows that

$$\begin{array}{rrcl} & A+\mathbf{i}B & = & 2\,L(y,x) \\ \text{and} & A-\mathbf{i}B & = & 2\,L(x,y) \end{array}$$

hence $L(y,x)$ is the conjugate of $L(x,y)$.
c) taking the conjugate in (1.3) one gets

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\overline{a}_i\,\overline{\overline{a}}_j\;\overline{L}(x_i,x_j)\in\mathbb{R}^+.$$

d) From b) we have, for any real number $\alpha$,

$$0 \le C\,(\alpha, L(x,y)) = \alpha^2 L(x,x) + 2\alpha\; |L(x,y)|^2 + |L(x,y)|^2\,L(y,y).$$

So, $C\,(\alpha, L(x,y))$ is a nonnegative polynomial of degree at most 2 in $\alpha$. Hence we have

$$|L(x,y)|^4 \le |L(x,y)|^2\,L(x,x)L(y,y).$$

If $L(x,y)\neq 0$, the conclusion follows. Otherwise it is clear from a). ■

LEMMA 4 *A* <u>*real*</u> *function $L$ defined on $E\times E$ is a positive type function if and only if*
*a) $L$ is symmetric*
*b) (1.3) is satisfied with $\mathbb{C}^n$ replaced with $\mathbb{R}^n$.*

**Proof.** A real-valued positive type function clearly satisfies a) and b). Conversely, let $n\ge 1$, let $(a_1,\ldots,a_n)\in\mathbb{C}^n$ and $(x_1,\ldots,x_n)\in E^n$. Writing $a_j=\alpha_j+\mathbf{i}\beta_j, 1\le j\le n$, and using the equality

$$\sum_{j=1}^{n}\sum_{k=1}^{n}\alpha_j\beta_k L(x_j,x_k)=\sum_{j=1}^{n}\sum_{k=1}^{n}\beta_j\alpha_k L(x_k,x_j)$$

we have

$$\begin{aligned}\sum_{j=1}^{n}\sum_{k=1}^{n}a_j\overline{a}_k L(x_j,x_k) &= \sum_{j=1}^{n}\sum_{k=1}^{n}\alpha_j\alpha_k L(x_j,x_k)+\sum_{j=1}^{n}\sum_{k=1}^{n}\beta_j\beta_k L(x_j,x_k)\\ &+ \sum_{j=1}^{n}\sum_{k=1}^{n}\mathbf{i}\beta_j\alpha_k(L(x_j,x_k)-L(x_k,x_j)).\end{aligned}$$

This last sum is a nonnegative real number whenever a) and b) are satisfied. ■

Other properties of positive type functions will be easily derived from the fact that the converse of Lemma 2 is true.
**Remark** As we have, for $n \geq 1$, $(a_1, \ldots, a_n) \in \mathbb{R}^n$ and $(x_1, \ldots, x_n) \in E^n$,

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i\, a_j\, L(x_i, x_j) = \sum_{i=1}^{n} a_i^2\, L(x_i, x_i) + \sum_{i<j} a_i\, a_j\, \left(L(x_i, x_j) + L(x_j, x_i)\right),$$

any real function $L$ satisfying

$$\forall (x, y) \in E^2 \; L(x, y) = -L(y, x)$$

also satisfies (1.3) with $\mathbb{C}^n$ replaced with $\mathbb{R}^n$. Such a function is identically 0 on the diagonal of $E \times E$ but is not symmetric unless it is 0 everywhere. See also Exercise 4.
We are now in a position to state the main result of this paragraph, from which the converse of Lemma 2 will appear as a consequence (see Theorem 3 below).

THEOREM 2 *Let $\mathcal{H}_0$ be any subspace of $\mathbb{C}^E$, the space of complex functions on $E$, on which an inner product $< .,. >_{\mathcal{H}_0}$ is defined, with associated norm $\|.\|_{\mathcal{H}_0}$.*
*In order that there exists a Hilbert space $\mathcal{H}$ such that*

*a) $\mathcal{H}_0 \subset \mathcal{H} \subset \mathbb{C}^E$ and the topology defined on $\mathcal{H}_0$ by the inner product $< .,. >_{\mathcal{H}_0}$ coincides with the topology induced on $\mathcal{H}_0$ by $\mathcal{H}$*
*b) $\mathcal{H}$ has a reproducing kernel $K$*
*it is necessary and sufficient that*

*c) the evaluation functionals $(e_t)_{t\in E}$ are continuous on $\mathcal{H}_0$*
*d) any Cauchy sequence $(f_n)$ in $\mathcal{H}_0$ converging pointwise to 0 converges also to 0 in the norm sense.*

**Remark** Assumption d) is equivalent to
*d') for any function $f$ in $\mathcal{H}_0$ and any Cauchy sequence $(f_n)$ in $\mathcal{H}_0$ converging pointwise to $f$, $(f_n)$ converges also to $f$ in the norm sense.*
**Proof. Direct part** If $\mathcal{H}$ does exist with conditions a) and b) satisfied the evaluation functionals are continuous on $\mathcal{H}$ (by Theorem 1) and therefore on $\mathcal{H}_0$. Now, let $(f_n)$ be a Cauchy sequence in $\mathcal{H}_0$ converging pointwise to 0. As $\mathcal{H}$ is complete, $(f_n)$ converges in the norm sense to some $f \in \mathcal{H}$. Thus we have

$$\forall x \in E \qquad f(x) = e_x(f) = \lim_{n\to\infty} e_x(f_n)$$

$$= \lim_{n\to\infty} f_n(x) = 0$$

and $f \equiv 0$.

**Converse.** Suppose c) and d) hold.
Define $\mathcal{H}$ as being the set of functions $f$ in $\mathbb{C}^E$ for which there exists a Cauchy sequence $(f_n)$ in $\mathcal{H}_0$ converging pointwise to $f$. Obviously

$$\mathcal{H}_0 \subset \mathcal{H} \subset \mathbb{C}^E.$$

The proof of Theorem 2 will be completed by Lemmas 5 to 9 below. Lemmas 5 and 6 enable us to extend to $\mathcal{H}$ the inner product on $\mathcal{H}_0$. In Lemmas 5 to 9 we use the hypotheses and notation of Theorem 2.

LEMMA 5 *Let $f$ and $g$ belong to $\mathcal{H}$. Let $(f_n)$ and $(g_n)$ be two Cauchy sequences in $\mathcal{H}_0$ converging pointwise to $f$ and $g$.*
*Then the sequence $< f_n, g_n >_{\mathcal{H}_0}$ is convergent and its limit only depends on $f$ and $g$.*

**Proof.** First recall that any Cauchy sequence is bounded.
$\forall (n,m) \in \mathbb{N}^2$,

$$\begin{aligned} | < f_n, g_n >_{\mathcal{H}_0} - < f_m, g_m >_{\mathcal{H}_0} | &= | < f_n - f_m, g_n >_{\mathcal{H}_0} \\ &\quad + < f_m, g_n - g_m >_{\mathcal{H}_0} | \end{aligned}$$

$$\leq \|f_n - f_m\|_{\mathcal{H}_0} \|g_n\|_{\mathcal{H}_0} + \|f_m\|_{\mathcal{H}_0} \|g_n - g_m\|_{\mathcal{H}_0},$$

by Cauchy-Schwarz inequality. This shows that $(< f_n, g_n >_{\mathcal{H}_0})$ is a Cauchy sequence in $\mathbb{C}$ and therefore convergent. In the same way, if $(f'_n)$ and $(g'_n)$ are two other Cauchy sequences in $\mathcal{H}_0$ converging pointwise respectively to $f$ and $g$, we have

$$\forall n \in \mathbb{N}, \quad \begin{aligned} | < f_n, g_n >_{\mathcal{H}_0} - < f'_n, g'_n > |_{\mathcal{H}_0} &\leq \|f_n - f'_n\|_{\mathcal{H}_0} \|g_n\|_{\mathcal{H}_0} \\ &\quad + \|f'_n\|_{\mathcal{H}_0} \|g_n - g'_n\|_{\mathcal{H}_0}. \end{aligned}$$

$(f_n - f'_n)$ and $(g_n - g'_n)$ are Cauchy sequences in $\mathcal{H}_0$ converging pointwise to 0. From assumption d) they also converge to 0 in the norm sense. It follows that $(< f_n, g_n >_{\mathcal{H}_0})$ and $(< f'_n, g'_n >_{\mathcal{H}_0})$ have the same limit. ■

LEMMA 6 *Suppose that $(f_n)$ is a Cauchy sequence in $\mathcal{H}_0$ converging pointwise to $f$ and that $\lim_{n\to\infty} < f_n, f_n >_{\mathcal{H}_0} = 0$ ($(f_n)$ tends to 0 in the norm sense). Then $f \equiv 0$.*

**Proof.**

$$\forall x \in E, \qquad f(x) = \lim_{n\to\infty} f_n(x) = \lim_{n\to\infty} e_x(f_n) = 0 \text{ by assumption c).}$$

■

Thus we can define an inner product on $\mathcal{H}$ by setting

$$< f, g >_{\mathcal{H}} = \lim_{n\longrightarrow\infty} < f_n, g_n >_{\mathcal{H}_0}$$

where $(f_n)$(resp. $(g_n)$) is a Cauchy sequence in $\mathcal{H}_0$ converging pointwise to $f$ (resp. $g$). The positivity, the hermitian symmetry and the linearity of $< .,. >_{\mathcal{H}}$ in the first variable are clear because $< .,. >_{\mathcal{H}_0}$ has those properties. Lemma 6 entails that $f \equiv 0$ whenever $< f, f >= 0$. Condition a) in Theorem 2 is satisfied for the topology on $\mathcal{H}$ associated with this inner product.

LEMMA 7 *Let $f \in \mathcal{H}$ and $(f_n)$ be a Cauchy sequence in $\mathcal{H}_0$ converging pointwise to $f$. Then $(f_n)$ converges to $f$ in the norm sense.*

**Proof.** Let $\epsilon > 0$ and let $N(\epsilon)$ be such that

$$(m > N(\epsilon) \text{ and } n > N(\epsilon)) \Rightarrow \|f_n - f_m\|_{\mathcal{H}_0} < \epsilon.$$

Fix $n > N(\epsilon)$. The sequence $(f_m - f_n)_{m\in\mathbb{N}}$ is a Cauchy sequence in $\mathcal{H}_0$ converging pointwise to $(f - f_n)$. Therefore

$$\|f - f_n\|_{\mathcal{H}} = \lim_{m\longrightarrow\infty} \|f_m - f_n\|_{\mathcal{H}_0} \leq \epsilon$$

Thus $(f_n)$ converges to $f$ in the norm sense. ■

COROLLARY 2 *$\mathcal{H}_0$ is dense in $\mathcal{H}$.*

**Proof.** By definition, for any $f \in \mathcal{H}$ there exists a Cauchy sequence $(f_n)$ in $\mathcal{H}_0$ converging pointwise to $f$. By Lemma 7 $(f_n)$ converges to $f$ in the norm sense. The corollary follows. ■

LEMMA 8 *The evaluation functionals are continuous on $\mathcal{H}$.*

**Proof.** As the evaluation functionals are linear it suffices to show that they are continuous at 0.
Let $x \in E$. The evaluation functional $e_x$ is continuous on $\mathcal{H}_0$ (assumption c) in Theorem 2). Fix $\epsilon > 0$ and let $\eta$ such that

$$(f \in \mathcal{H}_0 \text{ and } \|f\|_{\mathcal{H}_0} < \eta) \Rightarrow |f(x)| < \frac{\epsilon}{2}.$$

For any function $\varphi$ in $\mathcal{H}$ with $\|\varphi\|_{\mathcal{H}} < \frac{\eta}{2}$ there exists by Lemma 7 a function $g$ in $\mathcal{H}_0$ such that

$$|g(x) - \varphi(x)| < \frac{\epsilon}{2} \quad \text{and} \quad \|g - \varphi\|_{\mathcal{H}} < \frac{\eta}{2}.$$

This entails

$$\|g\|_{\mathcal{H}_0} = \|g\|_{\mathcal{H}} \leq \|g - \varphi\|_{\mathcal{H}} + \|\varphi\|_{\mathcal{H}} < \eta$$

hence $|g(x)| < \frac{\epsilon}{2}$ and $|\varphi(x)| < \epsilon$. Thus $e_x$ is continuous on $\mathcal{H}$. ∎

LEMMA 9 *$\mathcal{H}$ is a reproducing kernel Hilbert space.*

**Proof.** In view of Lemma 8 it remains to prove that $\mathcal{H}$ is complete. Let $(f_n)$ be a Cauchy sequence in $\mathcal{H}$ and let $x \in E$. As the evaluation functional $e_x$ is linear and continuous (Lemma 8), $(f_n(x))$ is a Cauchy sequence in $\mathbb{C}$ and thus converges to some $f(x)$. One has to prove that such defined $f$ belongs to $\mathcal{H}$. Let $(\epsilon_n)$ be any sequence of positive numbers tending to zero as n tends to $\infty$. As $\mathcal{H}_0$ is dense in $\mathcal{H}$

$$\forall i \in \mathbb{N}^* \qquad \exists g_i \in \mathcal{H}_0 \qquad \text{such that} \qquad \|f_i - g_i\|_{\mathcal{H}} < \epsilon_i.$$

From the inequalities

$$\begin{aligned} | g_i(x) - f(x) | &\leq | g_i(x) - f_i(x) | + | f_i(x) - f(x) | \\ &\leq |e_x(g_i - f_i)| + |f_i(x) - f(x)| \end{aligned}$$

and from the properties of $e_x$ (Lemma 8) it follows that $(g_n(x))$ tends to $f(x)$ as $n$ tends to $\infty$. We have

$$\begin{aligned} \|g_i - g_j\|_{\mathcal{H}_0} = \|g_i - g_j\|_{\mathcal{H}} &\leq \|g_i - f_i\|_{\mathcal{H}} + \|f_i - f_j\|_{\mathcal{H}} + \|f_j - g_j\|_{\mathcal{H}} \\ &\leq \epsilon_i + \epsilon_j + \|f_i - f_j\|_{\mathcal{H}} \end{aligned}$$

Thus $(g_n)$ is a Cauchy sequence in $\mathcal{H}_0$ tending pointwise to $f$, and so $f \in \mathcal{H}$. By Lemma 7 $(g_n)$ tends to $f$ in the norm sense. Now,

$$\|f_i - f\|_{\mathcal{H}} \leq \|f_i - g_i\|_{\mathcal{H}} + \|g_i - f\|_{\mathcal{H}}.$$

Therefore $(f_n)$ converges to $f$ in the norm sense and $\mathcal{H}$ is complete. ∎

**Remark** As $\mathcal{H}_0$ is dense in $\mathcal{H}$, $\mathcal{H}$ is isomorphic to the completion of $\mathcal{H}_0$. It is the smallest Hilbert space of functions on $E$ satisfying a) in Theorem 2. $\mathcal{H}$ is called the **functional** completion of $\mathcal{H}_0$.
Now we can prove the converse of Lemma 2 and give some properties of the Hilbert space having a given positive type function as reproducing kernel (see Moore (1935) and Aronszajn (1943)).

THEOREM 3 (MOORE-ARONSZAJN THEOREM) *Let $K$ be a positive type function on $E \times E$. There exists only one Hilbert space $\mathcal{H}$ of functions on $E$ with $K$ as reproducing kernel. The subspace $\mathcal{H}_0$ of $\mathcal{H}$ spanned by the functions $(K(.,x)_{x\in E})$ is dense in $\mathcal{H}$ and $\mathcal{H}$ is the set of functions on $E$ which are pointwise limits of Cauchy sequences in $\mathcal{H}_0$ with the inner product*

$$< f,g >_{\mathcal{H}_0} = \sum_{i=1}^{n}\sum_{j=1}^{m} \alpha_i \overline{\beta}_j K(y_j, x_i) \qquad (1.4)$$

*where $f = \sum_{i=1}^{n} \alpha_i K(.,x_i)$ and $g = \sum_{j=1}^{m} \beta_j K(.,y_j)$.*

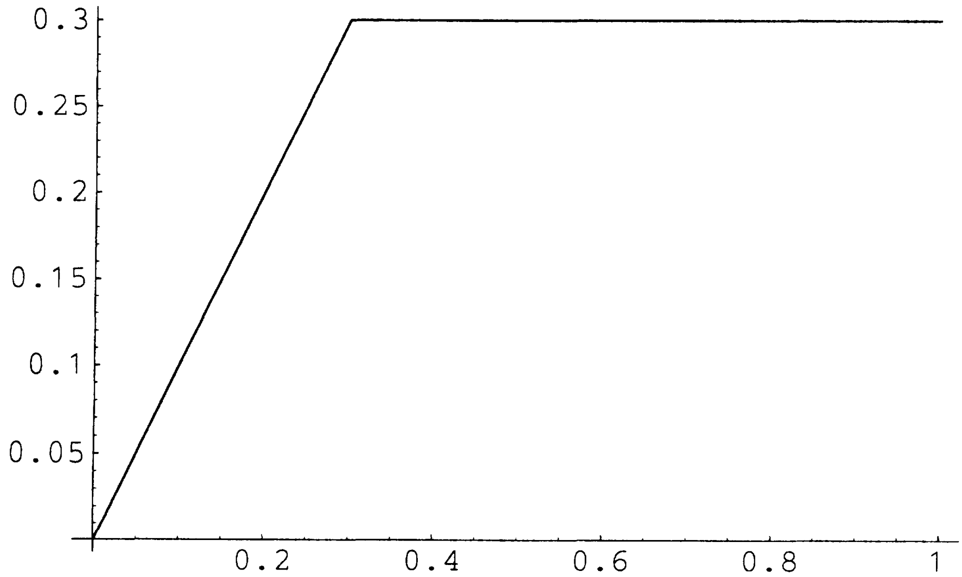


Figure 1.3: The function $\min(.,0.3)$ (See Example 4 page 12 and 15). The functions $\min(.,x)$ span a dense subspace of $\mathcal{H}$.

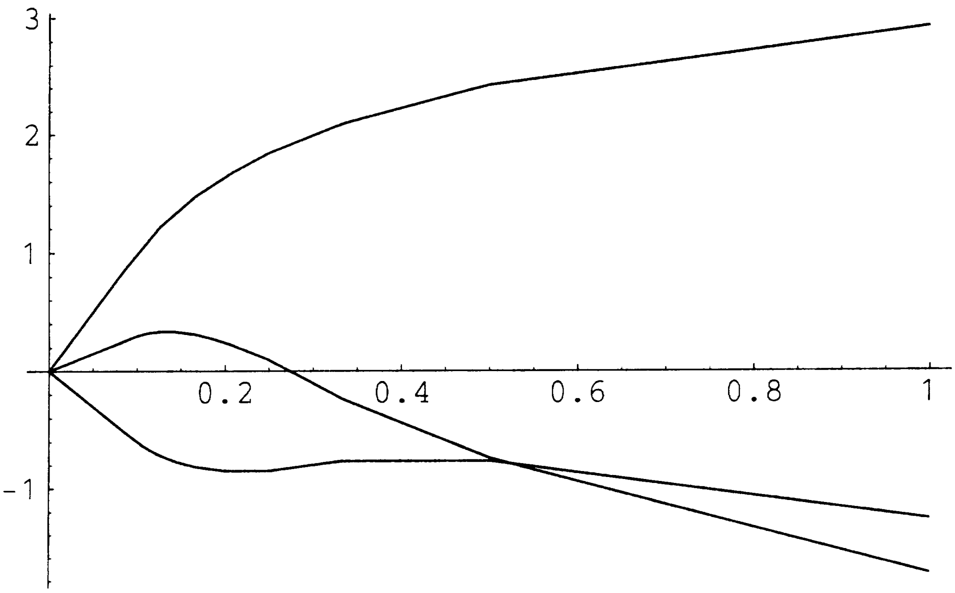


Figure 1.4: Three linear combinations of functions $\min(.,x)$ (See Example 4 page 12 and 15).

**Proof.** First remark that the complex number $< f, g >$ defined by (1.4) does not depend on the representations not necessarily unique of $f$ and $g$:

$$< f, g >_{\mathcal{H}_0} = \sum_{i=1}^{n} \alpha_i \overline{g(x_i)} = \sum_{j=1}^{m} \overline{\beta}_j f(y_j),$$

this shows that $< f, g >_{\mathcal{H}_0}$ depends on $f$ and $g$ only through their values. Then, taking

$$f = \sum_{i=1}^{n} \alpha_i \; K(., x_i) \quad \text{and} \quad g = K(., x)$$

we get

$$< f, K(., x) > = \sum_{i=1}^{n} \alpha_i \; \overline{g(x_i)} = \sum_{i=1}^{n} \alpha_i \; K(x, x_i) = f(x).$$

Thus the inner product with $K(., x)$ "reproduces" the values of functions in $\mathcal{H}_0$. In particular

$$\|K(., x)\|^2_{\mathcal{H}_0} = < K(., x), K(., x) > = K(x, x).$$

As $K$ is a positive type function, $< ., . >_{\mathcal{H}_0}$ is a semi-positive hermitian form on $\mathcal{H}_0 \times \mathcal{H}_0$. Now, suppose that $< f, f >_{\mathcal{H}_0} = 0$. From the Cauchy-Schwarz inequality we have

$$\forall x \in E \qquad |f(x)| = | < f, K(., x) >_{\mathcal{H}_0} | \leq < f, f >^{1/2}_{\mathcal{H}_0} [K(x, x)]^{1/2} = 0$$

and $f \equiv 0$.
Let us consider $\mathcal{H}_0$ endowed with the topology associated with the inner product $< .,. >_{\mathcal{H}_0}$ and check conditions c) and d) in Theorem 2. Let $f$ and $g$ in $\mathcal{H}_0$

$$\begin{aligned} \forall x \in E \qquad |e_x(f) - e_x(g)| &= |< f - g, K(.,x) >_{\mathcal{H}_0}| \\ &\leq \|f - g\|_{\mathcal{H}_0}[K(x,x)]^{1/2}. \end{aligned}$$

Therefore the evaluation functionals are continuous on $\mathcal{H}_0$ and c) is satisfied.
Let us now check Condition d). Let $(f_n)$ be a Cauchy sequence (hence bounded) in $\mathcal{H}_0$ converging pointwise to 0 and let $A > 0$ be an upper bound for $(\|f_n\|_{\mathcal{H}_0})$. Let $\epsilon > 0$ and $N(\epsilon)$ such that

$$n > N(\epsilon) \Rightarrow \|f_{N(\epsilon)} - f_n\|_{\mathcal{H}_0} < \frac{\epsilon}{A}.$$

Fix $k, \alpha_1, \ldots, \alpha_k$ and $x_1, \ldots, x_k$ such that

$$f_{N(\epsilon)} = \sum_{i=1}^{k} \alpha_i K(.,x_i).$$

As

$$\|f_n\|^2_{\mathcal{H}_0} =< f_n - f_{N(\epsilon)}, f_n >_{\mathcal{H}_0} + < f_{N(\epsilon)}, f_n >_{\mathcal{H}_0},$$

we have, for $n > N(\epsilon)$,

$$\|f_n\|^2_{\mathcal{H}_0} < \epsilon + \sum_{i=1}^{k} \alpha_i f_n(x_i),$$

hence $\limsup_{n \longrightarrow \infty} \|f_n\|^2 \leq \epsilon$. As $\epsilon$ is arbitrary this entails that $(f_n)$ converges to 0 in the norm sense. We are now in a position to apply Theorem 2 to $\mathcal{H}_0$: there exists a Hilbert space $\mathcal{H}$ of functions on $E$ satisfying a) and b) in Theorem 2. $\mathcal{H}$ is the set of functions $f$ for which there exists a Cauchy sequence $(f_n)$ in $\mathcal{H}_0$ converging pointwise to $f$. From Lemma 7 such a sequence $(f_n)$ is also converging to $f$ in the norm sense: $\mathcal{H}_0$ is dense in $\mathcal{H}$. Therefore $\mathcal{H}$ is unique and

$$\begin{aligned} \forall x \in E \qquad f(x) = \lim_{n\to\infty} f_n(x) &= \lim_{n\to\infty} < f_n, K(.,x) >_{\mathcal{H}_0} \\ &= < f, K(.,x) >_{\mathcal{H}} \end{aligned}$$

thus $K$ is the reproducing kernel of $\mathcal{H}$. ∎

Theorem 3 claims that a RKHS of functions on a set $E$ is characterized by its kernel $K$ on $E \times E$ and that the property for $K$ of being a reproducing kernel is equivalent to the property of being a positive type function. Actually, as stated in the following theorem, the notions of positive type function, reproducing kernel or RKHS can reduce to a very simple one, the notion of sequence in a space $\ell^2(X)$, with suitable index set. Recall that the space $\ell^2(X)$ is the set of complex sequences $\{x_\alpha, \alpha \in X\}$ satisfying

$$\sum_{\alpha \in X} |x_\alpha|^2 < \infty$$

endowed with the inner product

$$< (x_\alpha), (y_\alpha) >= \sum_{\alpha \in X} x_\alpha \overline{y}_\alpha .$$

Theorem 4 provides a characterization of ALL reproducing kernels on an abstract set $E$. It turns out that the definition of a positive type function or of a reproducing kernel on $E \times E$ or of a RKHS of functions on $E$ is equivalent to the definition of a mapping on $E$ with values in some space $\ell^2(X)$ (see Fortet, 1995).
At first sight this characterization can appear mainly theoretical. It is not the case. Indeed this theorem provides an effective way of constructing reproducing kernels or of proving that a given function is a reproducing kernel.

THEOREM 4 *A complex function $K$ defined on $E \times E$ is a reproducing kernel or a positive type function if and only if there exists a mapping $T$ from $E$ to some space $\ell^2(X)$ such that*

$$\begin{aligned} \forall (x,y) \in E \times E \qquad K(x,y) &= < T(x), T(y) >_{\ell^2(X)} \\ &= \sum_{\alpha \in X} (T(x))_\alpha \, (T(y))_\alpha . \end{aligned}$$

**Proof.** Let $\mathcal{H}$ be a RKHS of functions on a set $E$ with kernel $K$. Consider the mapping

$$\begin{array}{rccl} \Psi_K : & E & \longrightarrow & \mathcal{H} \\ & x & \longmapsto & K(.,x) \end{array}$$

Like any Hilbert space, $\mathcal{H}$ is isometric to some space $\ell^2(X)$. If $\varphi$ denotes any isometry from $\mathcal{H}$ to $\ell^2(X)$, the mapping $T = \varphi \circ \Psi_K$ meets the requirements. Conversely, the mapping

$$T : \quad E \longrightarrow \ell^2(X)$$

being given from a set $E$ to some space $\ell^2(X)$ the mapping

$$\begin{array}{rccl} K: & E \times E & \longrightarrow & \mathbb{C} \\ & (x,y) & \longmapsto & < T(x), T(y) >_{\ell^2(X)} \end{array}$$

is by Lemma 1 a positive type function. ∎

Particularizing to a set $E$, a pre-Hilbert space $\mathcal{H}$ (which can be considered through a suitable isomorphism as a part of a space $\ell^2(X)$ and a mapping $T$ from $E$ to $\mathcal{H}$, one can construct as many reproducing kernels as desired.

**Example 1** Let $E = [0,1]$, $\mathcal{H} = L^2(-1,1)$ and $T(x) = \cos(x.)$. By Theorem 4 we get that $K$ defined on $E \times E$ by

$$\begin{aligned} K(x,y) &= < T(y), T(x) >_{\mathcal{H}} = \int_{-1}^{1} \cos(xt)\cos(yt)d\lambda(t) \\ &= \frac{\sin(x-y)}{x-y} + \frac{\sin(x+y)}{x+y} \quad \text{if} \quad x \neq y, \end{aligned}$$

$$K(x,x) = 1 + \frac{\sin(2x)}{2x} \quad \text{if} \quad x \neq 0$$

and

$$K(0,0) = 2$$

is a reproducing kernel.

In the same way one easily proves that a given function is a reproducing kernel. This is illustrated in the following example.

**Example 2.** To prove that the function $K$ defined on $\mathbb{R}^+ \times \mathbb{R}^+$ by

$$K(x,y) = \min(x,y)$$

note that (Neveu (1968), Chapter 4)

$$\begin{aligned} \forall (x,y) \in \mathbb{R}^+ \times \mathbb{R}^+, \qquad K(x,y) &= \int_{\mathbb{R}^+} \mathbf{1}_{[0,y]}(t)\mathbf{1}_{[0,x]}(t)d\lambda(t) \\ &= < T(y), T(x) >_{\mathcal{H}} \end{aligned}$$

where $\mathcal{H} = L^2(\mathbb{R}^+, \lambda)$ and

$$\begin{array}{rccl} T: & E & \longrightarrow & \mathcal{H} \\ & x & \longmapsto & \mathbf{1}_{[0,x]}(.). \end{array}$$

# 4. BASIC PROPERTIES OF REPRODUCING KERNELS

## 4.1. SUM OF REPRODUCING KERNELS

THEOREM 5 *Let $K_1$ and $K_2$ be reproducing kernels of spaces $\mathcal{H}_1$ and $\mathcal{H}_2$ of functions on $E$ with respective norms $\|.\|_{\mathcal{H}_1}$ and $\|.\|_{\mathcal{H}_2}$. Then $K = K_1 + K_2$ is the reproducing kernel of the space $\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2 = \{f | f = f_1 + f_2,\ f_1 \in \mathcal{H}_1,\ f_2 \in \mathcal{H}_2\}$ with the norm $\|.\|_{\mathcal{H}}$ defined by*

$$\forall f \in \mathcal{H} \qquad \|f\|^2_{\mathcal{H}} \quad = \min_{\substack{f = f_1 + f_2,\\ f_1 \in \mathcal{H}_1,\ f_2 \in \mathcal{H}_2}} \left(\|f_1\|^2_{\mathcal{H}_1} + \|f_2\|^2_{\mathcal{H}_2}\right).$$

**Proof.** Let $\mathcal{F} = \mathcal{H}_1 \oplus \mathcal{H}_2$ be the Hilbert sum of the spaces $\mathcal{H}_1$ and $\mathcal{H}_2$. $\mathcal{F}$ is the set $\mathcal{H}_1 \times \mathcal{H}_2$ endowed with the norm defined by

$$\|(f_1, f_2)\|^2_{\mathcal{F}} = \|f_1\|^2_{\mathcal{H}_1} + \|f_2\|^2_{\mathcal{H}_2}.$$

Let

$$\begin{array}{rcl} u : \mathcal{F} & \longrightarrow & \mathcal{H} \\ (f_1, f_2) & \longmapsto & f_1 + f_2 \end{array}$$

and let $N = u^{-1}(\{0\})$ be the kernel of $u$. The map $u$ is linear and onto. Its kernel $N$ is a subspace of $\mathcal{F}$. Let $((f_n, -f_n))$ be a sequence of elements of $N$ converging to $(f_1, f_2)$. Then $(f_n)$ converges to $f_1$ in $\mathcal{H}_1$ and pointwise and $(-f_n)$ converges to $f_2$ in $\mathcal{H}_2$ and pointwise. Therefore $f_1 = -f_2$ and $N$ is a closed subspace of $\mathcal{F}$.

Let $N^{\perp}$ be the orthogonal complement of $N$ in $\mathcal{F}$ and let $v$ be the restriction of $u$ to $N^{\perp}$. The map $v$ is one-to-one hence one can define a inner product on $\mathcal{H}$ by setting

$$< f, g >_{\mathcal{H}} = < v^{-1}(f), v^{-1}(g) >_{\mathcal{F}}$$

Endowed with this inner product $\mathcal{H}$ is a Hilbert space of functions. It is clear that for any $y$ in $E$, $K(., y)$ is a element of $\mathcal{H}$. Its remains to check the reproducing property of $K$ and to express the norm of $\mathcal{H}$ in terms of $\|.\|_{\mathcal{H}_1}$ and $\|.\|_{\mathcal{H}_2}$.

Let $f \in \mathcal{H}$, $(f', f'') = v^{-1}(f)$ and $(K'(., y), K''(., y)) = v^{-1}(K(., y))$, $y \in E$.

As

$$K'(., y) - K_1(., y) + K''(., y) - K_2(., y) = K(., y) - K(., y) = 0,$$

$(K'(., y) - K_1(., y), K''(., y) - K_2(., y)) \in N$, its inner product in $\mathcal{F}$ with $(f', f'')$ is 0. Thus

$$< f', K'(., y) >_{\mathcal{H}_1} + < f'', K''(., y) >_{\mathcal{H}_2}$$

is equal to

$$< f', K_1(.,y) >_{\mathcal{H}_1} + < f'', K_2(.,y) >_{\mathcal{H}_2}$$

$$\begin{aligned} \text{and } < f, K(.,y) >_{\mathcal{H}} &= < v^{-1}(f), v^{-1}(K(.,y)) >_{\mathcal{F}} \\ &= < (f', f''), (K'(.,y), K''(.,y)) >_{\mathcal{F}} \\ &= f'(y) + f''(y) = f(y) \end{aligned}$$

and the reproducing property is proved. Now let $(f_1, f_2) \in \mathcal{F}^2$, $f = f_1 + f_2$ and $(g_1, g_2) = (f_1, f_2) - v^{-1}(f)$. On the one hand, by definition of the norm in $\mathcal{F}$,

$$\|(f_1, f_2)\|^2_{\mathcal{F}} = \|f_1\|^2_{\mathcal{H}_1} + \|f_2\|^2_{\mathcal{H}_2}.$$

On the other hand, as $(g_1, g_2)$ belongs to $N$ and $v^{-1}(f)$ belongs to $N^{\perp}$, we have

$$\begin{aligned} \|(f_1, f_2)\|^2_{\mathcal{F}} &= \left\|v^{-1}(f)\right\|^2_{\mathcal{F}} + \|(g_1, g_2)\|^2_{\mathcal{F}} \\ &= \left\|v^{-1}(f)\right\|^2_{\mathcal{F}} + \|g_1\|^2_{\mathcal{H}_1} + \|g_2\|^2_{\mathcal{H}_2}. \end{aligned}$$

Therefore, for the decomposition $f = f_1 + f_2$, we always have

$$\|f\|^2_{\mathcal{H}} = \left\|v^{-1}(f)\right\|^2_{\mathcal{F}} \leq \|f_1\|^2_{\mathcal{H}_1} + \|f_2\|^2_{\mathcal{H}_2} \tag{1.5}$$

and the equality holds in (1.5) if and only if $(f_1, f_2) = v^{-1}(f)$. ■

## 4.2. RESTRICTION OF THE INDEX SET

THEOREM 6 *Let $\mathcal{H}$ be a Hilbert space of functions defined on $E$ with reproducing kernel $K$ and norm $\|.\|_{\mathcal{H}}$ and let $E_1$ be a non empty subset of $E$. The restriction $K_1$ of $K$ to $E_1 \times E_1$ is the reproducing kernel of the space $\mathcal{H}_1$ of restrictions of elements of $\mathcal{H}$ to $E_1$ endowed with the norm $\|.\|_{\mathcal{H}_1}$ defined by*

$$\forall f_1 \in \mathcal{H}_1 \qquad \|f_1\|_{\mathcal{H}_1} = \min_{\substack{f \in \mathcal{H} \\ f|_{E_1} = f_1}} \|f\|_{\mathcal{H}}$$

*where $f|_{E_1}$ stands for the restriction of $f$ to the subset $E_1$.*

**Proof.** Let $u : \mathcal{H} \longrightarrow \mathcal{H}_1 \quad f \longmapsto f|_{E_1}$ and let $N = u^{-1}(\{0\})$. The map $u$ is linear and onto and $N$ is a subspace of $\mathcal{H}$. If a sequence $(f_n)$ in $N$ tends to $f \in \mathcal{H}$, we have

$$\forall x \in E_1, \qquad \forall n \geq 1, \qquad f_n(x) = 0,$$

and, as the convergence in norm implies the pointwise convergence, the limit $f$ satisfies $f|_{E_1} \equiv 0$. Therefore $N$ is closed. Let $N^\perp$ be its orthogonal complement in $\mathcal{H}$ and let $v$ be the restriction of $u$ to $N^\perp$. The map $v$ is one-to-one hence one can define an inner product on $\mathcal{H}_1$ by setting

$$< f, g >_{\mathcal{H}_1} \quad = \quad < v^{-1}(f), v^{-1}(g) >_{\mathcal{H}} .$$

Endowed with this inner product $\mathcal{H}_1$ is a Hilbert space of functions. It is clear that for any $y$ in $E_1$, $K_1(.,y)$ is a element of $\mathcal{H}_1$ and that $[K(.,y) - v^{-1}(K_1(.,y))] \in N$. Thus

$$\begin{aligned}\forall y \in E_1,\ \forall f \in \mathcal{H}_1,\ < f, K_1(.,y) >_{\mathcal{H}_1} &= < v^{-1}(f), v^{-1}(K_1(.,y)) >_{\mathcal{H}} \\ &= < v^{-1}(f), K(.,y) >_{\mathcal{H}} \\ &= v^{-1}(f)(y) = f(y).\end{aligned}$$

This shows that $K_1$ is the reproducing kernel of $\mathcal{H}_1$.
Let $g \in \mathcal{H}$. If $g|_{E_1} = f_1$, then $(g - v^{-1}(f_1))$ belongs to $N$ and $v^{-1}(f_1)$ belongs to $N^\perp$. By the Pythagore identity,

$$||g||^2_{\mathcal{H}} = ||g - v^{-1}(f_1)||^2_{\mathcal{H}} + ||v^{-1}(f_1)||^2_{\mathcal{H}}$$

therefore

$$||f_1||_{\mathcal{H}_1} = ||v^{-1}(f_1)||_{\mathcal{H}} \leq ||g||_{\mathcal{H}}$$

and the equality occurs if and only if $g = v^{-1}(f_1)$. The conclusion follows.

■

## 4.3. SUPPORT OF A REPRODUCING KERNEL

In this subsection we introduce a notion which is of great importance in the search for bases in RKHS. It is the notion of support of a function of two or several variables, first introduced by Duc-Jacquet (1973).

DEFINITION 3 *Let $K$ be a non null complex function defined on $E \times E$. A subset $A$ of $E$ is said to be binding for $K$ if and only if there exist elements $x_1, \ldots, x_n$ in $A$ such that the functions $K(.,x_1), \ldots, K(.,x_n)$ are linearly dependent in the vector space $\mathbb{C}^E$.*

Then we have the following theorem allowing to define the notion of support of $K$.

THEOREM 7 *The set $\mathcal{F}_K$ of non-binding sets for $K$ partially ordered by inclusion is inductive and therefore admits at least a maximal element.*

**Proof.** Let $\{A_i : i \in I\}$ be a set of elements of $\mathcal{F}_K$ linearly ordered by inclusion. It is clear that the set $\bigcup_{i\in I} A_i$ belongs to $\mathcal{F}_K$ and is the upper bound of the chain $(A_i)_{i\in I}$. Thus $\mathcal{F}_K$ partially ordered by inclusion is inductive. By Zorn's Lemma (Dudley, 1989) it has at least a maximal element. ■

DEFINITION 4 *Let $K$ be a non null complex function defined on $E \times E$. A subset $S$ of $E$ is called a support of $K$ if and only if $S$ is a maximal element of the set $\mathcal{F}_K$ of non-binding sets for $K$.*

The link between support of reproducing kernel and basis of $\mathcal{H}_0$ is expressed in the following theorem.

THEOREM 8 *Let $\mathcal{H}$ be a RKHS with kernel $K$ on $E \times E$. Let $\mathcal{H}_0$ be the subspace of $\mathcal{H}$ spanned by $\{K(.,x) : x \in E\}$. If a subset $S$ of $E$ is a support of $K$ then $\{K(.,x) : x \in S\}$ is a basis of $\mathcal{H}_0$. Conversely if $K(.,x_1)$, ..., $K(.,x_n)$ are linearly independent, there exists a support $S$ of $K$ containing $\{x_1, \ldots, x_n\}$.*

**Proof.** The set $\mathcal{S} = \{K(.,x) : x \in S\}$ is a set of linearly independent elements of $\mathcal{H}_0$. If $x_0$ belongs to $X \setminus S$, consider the set $\mathcal{S}\bigcup\{K(.,x_0)\}$ and use the maximality of $S$ to get that $K(.,x_0)$ can be written as a linear conbination of elements of $\mathcal{S}$. Therefore the set $\mathcal{S}$ spans $\mathcal{H}_0$.
Now, if $K(.,x_1), \ldots, K(.,x_n)$ are linearly independent, the set $\{x_1, \ldots, x_n\}$ is a non-binding set for $K$, hence it is included in a support of $K$. ■

## 4.4. KERNEL OF AN OPERATOR

DEFINITION 5 *Let $E$ be a pre-Hilbert space of functions defined on $E$ and let $u$ be an operator in $E$. A function $U : E \times E \longrightarrow \mathbb{C} \quad (x,y) \longmapsto U(x,y)$ is said to be a kernel of $u$ if and only if*

$$\forall y \in E, \qquad U(.,y) \in \mathcal{E}$$
$$\forall y \in E, \quad \forall f \in \mathcal{E}, \quad u(f)(y) =< f, U(.,y) >_{\mathcal{E}} .$$

If $u$ has two kernels $U_1$ and $U_2$ one has

$$\forall y \in E, \quad \forall f \in \mathcal{E}, \quad < f, U_1(.,y) - U_2(y) >_{\mathcal{E}}= u(f)(y) - u(f)(y) = 0.$$

Thus $\forall y \in E, \quad U_1(.,y) = U_2(y)$ and $U_1 = U_2$.
So, for any operator there is at most one kernel. It is also clear from Definition 5 that a Hilbert space of functions $\mathcal{H}$ has a reproducing kernel $K$ if and only if $K$ is the kernel of the identity operator in $\mathcal{H}$.

THEOREM 9 *In a Hilbert space $\mathcal{H}$ of functions with reproducing kernel $K$ any continuous operator $u$ has a kernel $U$ given by*

$$U(x,y) = [u^*(K(.,y))](x) \tag{1.6}$$

*where $u^*$ denotes the adjoint operator of $u$.*

**Proof.** By Riesz's theorem, in the Hilbert space $\mathcal{H}$ any continuous operator $u$ has an adjoint defined by

$$\forall (f,g) \in \mathcal{H} \times \mathcal{H} \quad < u(f), g >_{\mathcal{H}} = < f, u^*(g) >_{\mathcal{H}} .$$

Thus we have

$$\begin{aligned} \forall y \in E, \quad \forall f \in \mathcal{H}, \quad < f, u^*(K(.,y)) >_{\mathcal{H}} &= < u(f), K(.,y) >_{\mathcal{H}} \\ &= u(f)(y). \end{aligned}$$

■

**Example: covariance operator**

As will be seen in Chapter 2 in a much more general setting, a fundamental tool in the study of stochastic processes is the covariance operator. Let $X$ be a random variable defined on some probability space $(\Omega, \mathcal{A}, P)$ with values in $(\mathcal{H}, \mathcal{B})$ where $\mathcal{H}$ is a RKHS of functions on a set $E$ and $\mathcal{B}$ is its Borel $\sigma$-algebra. For any $\omega \in \Omega$, $X(\omega) = X_.(\omega)$ is the function defined on $E$ by

$$\begin{array}{rcl} E & \longrightarrow & \mathbb{C} \\ t & \longmapsto & X_t(\omega) \end{array}$$

called trajectory associated with $\omega$. In other words $(X_t)_{t \in E}$ is a stochastic process on $(\Omega, \mathcal{A}, P)$ with trajectories in $\mathcal{H}$.
Suppose that $X$ is a second order random variable, *i.e.* that

$$E_P\left(\|X\|^2_{\mathcal{H}}\right) < \infty.$$

Then the covariance operator of $X$ is defined by

$$C_X(f) = E_P\left(< X, f >_{\mathcal{H}} X\right)$$

where the expectation is taken in the sense of Bochner integral of $\mathcal{H}-$valued random variables (see Chapter 4). It can be defined equivalently as the unique operator $C_X$ satisfying

$$< C_X f, g >_{\mathcal{H}} = E\left(< X, f >_{\mathcal{H}} < X, g >_{\mathcal{H}}\right).$$

The operator $C_X$ is self-adjoint, positive, continuous and compact. From Theorem 9 its kernel is given by

$$\begin{aligned} U(t,s) &= [C_X(K(.,s))](t) \\ &= < C_X(K(.,s)), K(.,t) >_{\mathcal{H}} \\ &= E(< X, K(.,s) >_{\mathcal{H}} < X, K(.,t) >_{\mathcal{H}}) \\ &= E(X_t X_s). \end{aligned}$$

$U$ is the second moment function of $X$.
We have proved the following result.

THEOREM 10 *The covariance operator $C_X$ of a second order random variable $X$ with values in a RKHS $\mathcal{H}$ of functions on a set $E$ has a kernel which is the second moment function of $X$. This means that for any $t$ in $E$ the function*

$$\begin{aligned} E(X_\bullet X_s) = C_X(K(.,s)) : E &\longrightarrow C \\ t &\longmapsto E(X_t X_s) \end{aligned}$$

*belongs to $\mathcal{H}$ and we have*

$$\forall f \in \mathcal{H} \qquad \forall s \in E \qquad C_X(f)(s) =< f, E(X_\bullet X_s) >$$

Kernels in the sense of Definition 5 provide nice representations of operators in RKHS, useful in many fields of application.
A very useful property to find the reproducing kernel of a subspace is that it is the kernel of the projection operator onto this subspace.

THEOREM 11 (**Kernel of a closed subspace**) *Let $V$ be a closed subspace of a Hilbert space $\mathcal{H}$ with reproducing kernel $K$. Then $V$ is a reproducing kernel Hilbert space and its kernel $K_V$ is given by*

$$K_V(x,y) = [\Pi_V(K(.,y))](x)$$

*where $\Pi_V$ denotes the orthogonal projection onto the space $V$.*

**Proof.** As $\Pi_V$ is a self-adjoint operator ($\Pi_V = \Pi_V^*$), by Theorem 9 $K_V$ is the kernel of $\Pi_V$. Now, the restriction of $\Pi_V$ to the subspace $V$ is the identity of $V$. Therefore $K_V$ is the reproducing kernel of $V$. ∎

The Riesz's representation theorem guarantees that any continuous linear functional $u$

$$\begin{aligned} \mathcal{H} &\longrightarrow \mathbb{C} \\ f &\longmapsto u(f) \end{aligned}$$

has a representer $\tilde{u}$ in $\mathcal{H}$ in the sense that

$$\forall f \in \mathcal{H}, \ < f, \tilde{u} >= u(f). \tag{1.7}$$

In the case of a RKHS, $\tilde{u}$ can be expressed easily through the kernel as stated in the following lemma.

LEMMA 10 *In a Hilbert space $\mathcal{H}$ of functions with reproducing kernel $K$ any continuous linear form $u : \mathcal{H} \longrightarrow \mathbb{C}$ has a* **Riesz representer** $\tilde{u}$ *given by*

$$\tilde{u}(x) = u(K(.,x))$$

**Proof.** Put $f(.) = K(.,x)$ in Equation (1.7). ■

## 4.5. CONDITION FOR $\mathcal{H}_K \subset \mathcal{H}_R$.

Denote by $\mathcal{H}_K$ the RKHS corresponding to a given reproducing kernel $K$, as given by the Moore-Aronszajn theorem. When the index set is a separable metric space, Aronszajn (1950) proves the following.

THEOREM 12 *Let $K$ be a continuous nonnegative kernel on $T \times T$ and $R$ be a continuous positive kernel on $T \times T$. The following statements are equivalent:*
*(i) $\mathcal{H}_K \subset \mathcal{H}_R$*
*(ii) There exists a constant $B$ such that $B^2 R - K$ is a nonnegative kernel*

Ylvisaker (1962) gives an alternative condition which is that if $\sum_{j=1}^{N(n)} c_{j_n} R(., t_{j_n})$ is a Cauchy sequence in $\mathcal{H}_R$, then $\sum_{j=1}^{N(n)} c_{j_n} K(., t_{j_n})$ must be a Cauchy sequence in $\mathcal{H}_K$.
Driscoll (1973) proves that any of these conditions is equivalent to
*(iii) There exists an operator $L : \mathcal{H}_R \longrightarrow \mathcal{H}_K$ such that $\| L \| \leq B$ and $LR(t,.) = K(t,.), \forall t \in T_0$ where $T_0$ is a countable dense subset of $T$.*
Moreover $(i)$ implies that there exists a constant $B$ such that

$$\forall g \in \mathcal{H}_K, \qquad \| g \|_R \leq B \| g \|_K,$$

and either of these conditions implies that there exists a self adjoint operator $L : \mathcal{H}_R \longrightarrow \mathcal{H}_K$ such that $\| L \| \leq B$ and

$$\forall t \in T, \qquad LR(t,.) = K(t,.).$$

## 4.6. TENSOR PRODUCTS OF RKHS

Products of functions and kernels play an important role in multidimensional settings. This is why we recall in this subsection some basic facts about tensor products of vector spaces of functions. Direct product RKHS are considered in Parzen (1963). We refer to Neveu (1968) for a nice introduction to tensor products of Hilbert spaces and more particularly of RKHS and for the proofs of the results given hereafter. Let $H_1$ and $H_2$ be two vector spaces of complex functions respectively defined on $E_1$ and $E_2$. The tensor product $H_1\tilde{\otimes}H_2$ is defined as the vector space generated by the functions

$$\begin{array}{rccl} f_1 \otimes f_2 : & E_1 \times E_2 & \longrightarrow & \mathbb{C} \\ & (x_1, x_2) & \longmapsto & f_1(x_1)f_2(x_2) \end{array}$$

where $f_1$ varies in $H_1$ and $f_2$ varies in $H_2$. If $< .,. >_1$ and $< .,. >_2$ are inner products respectively on $H_1$ and $H_2$, it can be shown that the mapping

$$\begin{array}{ccc} H_1\tilde{\otimes}H_2 & \longrightarrow & \mathbb{C} \\ (f_1 \times f_2, f_1' \times f_2') & \longmapsto & < f_1, f_2 >_1 < f_1', f_2' >_2 \end{array}$$

is an inner product on $H_1\tilde{\otimes}H_2$ which is therefore a pre-Hilbert space. Its completion is called the tensor product of the Hilbert spaces $H_1$ and $H_2$ and is denoted by $H_1 \otimes H_2$.

If $H_1$ is a RKHS with kernel $K_1$ and $H_2$ is a RKHS with kernel $K_2$, it is then clear that the mapping

$$\begin{array}{rccl} K_1 \otimes K_2 : & (E_1 \times E_2)^2 & \longrightarrow & \mathbb{C} \\ & ((x_1, x_2), (y_1, y_2)) & \longmapsto & K_1(x_1, y_1)\; K_2(x_2, y_2) \end{array}$$

is a positive type function on $(E_1 \times E_2)^2$. More precisely we have the following theorem (Neveu, 1968).

THEOREM 13 *Let $H_1$ and $H_2$ be two RKHS with respective reproducing kernels $K_1$ and $K_2$. Then the tensor product $H_1\tilde{\otimes}H_2$ of the vector spaces $H_1$ and $H_2$ admits a functional completion $H_1 \otimes H_2$ which is a RKHS with reproducing kernel $K = K_1 \otimes K_2$.*

It follows from this theorem that the product of a finite family of reproducing kernels on the same set $E^2$ is a reproducing kernel on $E^2$.

## 5. SEPARABILITY. CONTINUITY

Let us first prove a lemma which is useful in the study of functionals on RKHS.

LEMMA 11 *In a separable RKHS $\mathcal{H}$ there is a countable set $\mathcal{D}_0$ of finite linear combinations of functions $K(., x), x \in E$, which is dense in $\mathcal{H}$.*

**Proof.** By Theorem 3 the subspace $\mathcal{H}_0$ of $\mathcal{H}$ spanned by the functions $K(., x), x \in E$, is dense in $\mathcal{H}$. Let $\{x_p : p \in \mathbb{N}\}$ be a countable subset dense in $\mathcal{H}$ and let $n$ be a positive integer. As $\overline{\mathcal{H}_0} = \mathcal{H}$, for any $p$ in $\mathbb{N}$ there exists $y_p^n$ in $\mathcal{H}_0$ such that

$$\|y_p^n - x_p\| < \frac{1}{n}.$$

Then the countable set

$$\mathcal{D}_0 = \bigcup_{n>0} \{y_p^n : p \in \mathbb{N}\} \subset \mathcal{H}_0$$

satisfies the requirement. To see this, consider $y$ in $\mathcal{H}$, $\varepsilon > 0$ and $n > (2/\varepsilon)$. There exists $p$ in $\mathbb{N}$ such that

$$\|y - x_p\| < \frac{\varepsilon}{2}.$$

Therefore

$$\|x_p - y_p^n\| < \frac{1}{n} < \frac{\varepsilon}{2} \text{ and } \|y - y_p^n\| < \varepsilon.$$

We can conclude that $\mathcal{D}_0$ is dense in $\mathcal{H}$. ■

Indeed the above Lemma particularizes the property that in a separable metric space $\mathcal{E}$ any dense subset contains a countable subset which is dense in $\mathcal{E}$.

In separable Hilbert spaces, countable orthonormal systems are used to expand any element as an infinite sum. In separable RKHS the reproducing kernel can be expressed through orthonormal systems as stated in the following theorem.

THEOREM 14 *Let $\mathcal{H} \subset \mathbb{C}^E$ be a separable (i.e. with a countable dense subset) Hilbert space with reproducing kernel $K$. For any complete orthonormal system $(e_i)_{i\in\mathbb{N}}$ in $\mathcal{H}$ we have*

$$\forall t \in E, \qquad K(.,t) = \sum_{i=0}^{\infty} \overline{e}_i(t)\, e_i(.) \qquad \textit{(convergence in } \mathcal{H}). \qquad (1.8)$$

*Conversely if (1.8) holds for an orthonormal system $(e_i)_{i\in\mathbb{N}}$ then this system is complete and $\mathcal{H}$ is separable.*
*Moreover, (1.8) implies that*

$$\forall s \in E, \qquad \forall t \in E, \qquad K(s,t) = \sum_{i=0}^{\infty} \overline{e}_i(t)\, e_i(s) \qquad \textit{(convergence in } \mathbb{C}).$$

**Proof.** Fix $t$ in $E$. The function $K(.,t)$, as an element of $\mathcal{H}$, has a Fourier series expansion

$$\sum_{i=0}^{\infty} c_i\, e_i(.)$$

where

$$c_i = < K(.,t), e_i > = \overline{< e_i, K(.,t) >} = \overline{e}_i(t).$$

Conversely if (1.8) holds for an orthonormal system $(e_i)_{i \in \mathbb{N}}$ then

$$\forall \varphi \in \mathcal{H}, \qquad \forall t \in E, \qquad \varphi(t) = < \varphi, K(.,t) > = \sum_{i=0}^{\infty} e_i(t) < \varphi, e_i >$$

thus

$$\forall \varphi \in \mathcal{H}, \qquad \varphi = \sum_{i=0}^{\infty} < \varphi, e_i > e_i \qquad \text{(convergence in } \mathcal{H}).$$

Therefore the system $(e_i)_{i \in \mathbb{N}}$ is complete and $\mathcal{H}$ is separable. The last property follows from (1.8) by computing $K(s,t)$ as $< K(.,s), K(.,t) >$ .

■

The next theorem provides a criterion for the separability of a Hilbert space of functions.

THEOREM 15 *Let $\mathcal{H}$ be a Hilbert space of functions on $E$ with reproducing kernel $K$. Suppose that $E$ contains a countable subset $E_0$ such that*

$$\forall g \in \mathcal{H}, \qquad (g|_{E_0} = 0 \Leftrightarrow g = 0).$$

*Then $\mathcal{H}$ is separable.*

**Proof.** Consider an element $g$ in $\mathcal{H}$ orthogonal to the family $(K(.,y))_{y \in E_0}$. For any $y \in E_0$, one has

$$g(y) = < g, K(.,y) >_{\mathcal{H}} = 0,$$

thus $g|_{E_0} = 0$ and $g = 0$ by the hypothesis. It follows that the subspace $V^{\perp}$ orthogonal to the closed subspace $V$ generated by the family $(K(.,y))_{y \in E_0}$ is equal to $\{0\}$. Hence $V$ is equal to $\mathcal{H}$. The countable family $(K(.,y))_{y \in E_0}$ is total in $\mathcal{H}$ (it generates a dense subspace of $\mathcal{H}$) and therefore $\mathcal{H}$ is separable. ■

An easy corollary exhibits a class of Hilbert spaces of continuous functions for which there is no hope to find a reproducing kernel.

COROLLARY 3 *A non separable Hilbert space $\mathcal{H}$ of continuous functions on a separable topological space $E$ has no reproducing kernel.*

**Proof.** Let $E_0$ be a countable dense subset of $E$. As any element of $\mathcal{H}$ is continuous the condition of Theorem 15 is satisfied. Therefore if $\mathcal{H}$ had a reproducing kernel it would be separable. This would contradict the hypothesis. ■

Fortet(1973) proves the following criterion.

THEOREM 16 *A RKHS $\mathcal{H}$ with kernel $K$ is separable if and only if for any $\varepsilon > 0$, there exists a countable partition $B_j, j \in \mathbb{N}$ of $E$ such that:*

$$\forall j, \forall t_1, t_2 \in B_j, K(t_1, t_1) + K(t_2, t_2) - K(t_1, t_2) - K(t_2, t_1) < \varepsilon \quad (1.9)$$

Let us now turn to a characterization of RKHS of continuous functions.

THEOREM 17 **(Reproducing kernel Hilbert space of continuous functions)** *Let $\mathcal{H}$ be a Hilbert space of functions defined on a metric space $(E, d)$ with reproducing kernel $K$. Then any element of $\mathcal{H}$ is continuous if and only if $K$ satisfies the following conditions*
*a)$\forall y \in E$, $K(., y)$ is continuous*
*b)$\forall x \in E$, $\exists r > 0$, such that the function*

$$\begin{array}{ccc} E & \longrightarrow & \mathbb{R}^+ \\ y & \longmapsto & K(y, y) \end{array}$$

*is bounded on the open ball $B(x, r)$.*

**Proof.** If any element of $\mathcal{H}$ is continuous, a) is clearly satisfied. Suppose that b) does not hold true. Then there exists $x \in E$ such that

$$\forall n \in \mathbb{N}^*, \quad \exists x_n \in B(x, 1/n), \quad \text{such that } K(x_n, x_n) \geq n.$$

As the sequence $(x_n)$ converges to $x$ we have for any (continuous) function $\varphi$ in $\mathcal{H}$

$$< \varphi, K(., x) > = \varphi(x) = \lim_{n \longrightarrow \infty} \varphi(x_n) = \lim_{n \longrightarrow \infty} < \varphi, K(., x_n) > .$$

Therefore the sequence $(K(., x_n))$ converges weakly to $K(., x)$ whereas

$$||K(., x_n)||^2 = K(x_n, x_n)$$

tends to infinity. This is a contradiction since any weakly convergent sequence in a Hilbert space is bounded. Hence b) is satisfied.
Conversely suppose that a) and b) hold true. Let $(x_n)$ be a convergent sequence in $E$ with limit $x$, let $\varphi$ be a element of $\mathcal{H}$ and let $(r, M)$ such that

$$\sup_{y \in B(x,r)} K(y, y) \leq M.$$

For $n$ large enough $x_n$ belongs to $B(x, r)$ hence we have

$$||K(., x_n)||^2 = K(x_n, x_n) \leq M.$$

Let $\mathcal{H}_0$ be the dense subspace of $\mathcal{H}$ spanned by the functions $(K(.,y))_{y\in E}$. Any element of $\mathcal{H}_0$ can be written as a finite linear combination

$$\sum_{i=1}^{k} a_i \, K(.,y_i),$$

so it is, by a), a continuous function. Let $(\varphi_m)_m$ be a sequence in $\mathcal{H}_0$ converging to $\varphi$ in the norm sense. By Corollary 1 $(\varphi_m)_m$ also converges pointwise to $\varphi$. Let $\epsilon > 0$. Fix $m$ large enough to have

$$|\varphi_m(x) - \varphi(x)| < \epsilon$$

and

$$||\varphi_m - \varphi||_{\mathcal{H}} < \epsilon.$$

As $\varphi_m$ is continuous, for $n$ large enough we have

$$|\varphi_m(x_n) - \varphi_m(x)| < \epsilon.$$

Therefore, for $n$ large enough,

$$\begin{aligned} |\varphi(x_n) - \varphi(x)| &\leq |\varphi(x_n) - \varphi_m(x_n)| + |\varphi_m(x_n) - \varphi_m(x)| \\ &\quad + |\varphi_m(x) - \varphi(x)| \\ &\leq |< K(.,x_n), \varphi - \varphi_m >_{\mathcal{H}}| + 2\,\epsilon \\ &\leq (K(x_n,x_n))^{1/2}\, ||\varphi - \varphi_m||_{\mathcal{H}} + 2\,\epsilon \\ &\leq (M+2)\,\epsilon. \end{aligned}$$

This shows that $\varphi$ is continuous at $x$. ∎

In particular, if $E$ is a bounded interval of $\mathbb{R}$ (or if $E$ is unbounded but $\int\int K^2(s,t)d\lambda(s)d\lambda(t) < \infty$) and if the kernel is continuous on $E \times E$, then $\mathcal{H}$ is a space of continuous functions.The following corollaries apply respectively to bounded kernels continuous in each variable and to continuous kernels on compact sets.

COROLLARY 4 *Let $\mathcal{H}$ be a Hilbert space of functions defined on a metric space $(E,d)$ with reproducing kernel $K$. If $K$ is bounded and if, for any $y \in E$, $K(.,y)$ is continuous (this implies, by symmetry, that for any $x \in E$, $K(x,.)$ is continuous) then $\mathcal{H}$ is a space of continuous functions. If, moreover, $E$ is separable, then $\mathcal{H}$ is separable and*

$$\forall s \in E, \qquad \forall t \in E, \qquad K(s,t) = \sum_{i=0}^{\infty} \overline{e}_i(t)\, e_i(s),$$

*where $(e_i)$ is any orthonormal system in $\mathcal{H}$.*

COROLLARY 5 *Let $\mathcal{H}$ be a Hilbert space of functions defined on a compact metric space $(E, d)$ with reproducing kernel $K$. If $K$ is continuous then $\mathcal{H}$ is a separable space of continuous functions and*

$$\forall s \in E, \qquad \forall t \in E, \qquad K(s,t) = \sum_{i=0}^{\infty} \overline{e}_i(t)\, e_i(s), \tag{1.10}$$

*where the convergence is uniform on $E \times E$ and $(e_i)$ is any orthonormal system in $\mathcal{H}$. The functions $e_i$ are uniformly continuous and bounded by $(\sup_t K(t,t))^{1/2}$.*

**Proof.** $E$ is compact, hence separable and $K$ is continuous on $E \times E$ and therefore bounded. Thus, Corollary 4 applies. $\mathcal{H}$ is a separable space of continuous functions and

$$\forall s \in E, \qquad \forall t \in E, \qquad K(s,t) = \sum_{i=0}^{\infty} \overline{e}_i(t)\, e_i(s),$$

where the functions $(e_i)$ are continuous (therefore uniformly continuous) and orthonormal in $\mathcal{H}$.
For any $t \in E$,

$$|e_i(t)| = | < e_i, K(.,t) >_{\mathcal{H}} | \leq \|e_i\| \|K(.,t)\| = [K(t,t)]^{1/2}.$$

It remains to prove that the convergence in (1.10) holds not only in the pointwise sense but also uniformly on $E \times E$. The sequence

$$\left\{ \sum_{i=0}^{n} |e_i|^2(t) : n \in \mathbb{N} \right\}$$

is an increasing sequence of continuous functions of the variable $t$ converging pointwise to the continuous function

$$K(t,t) = \sum_{i=0}^{\infty} |e_i|^2(t)$$

on the compact set $E$. By Dini's theorem the convergence is uniform. As we have

$$\left| \sum_{i=n}^{\infty} \overline{e}_i(t)\, e_i(s) \right|^2 \leq \left| \sum_{i=n}^{\infty} |e_i| \right|^2 (t) \left| \sum_{i=n}^{\infty} |e_i| \right|^2 (s),$$

the convergence of

$$\sum_{i=0}^{n} \overline{e}_i(t)\, e_i(s) \text{ to } K(s,t)$$

is uniform on $E \times E$. ∎

Under the hypotheses of Corollary 5 we will prove in Chapter 4 Section 7 another property of orthonormal systems when they characterize signed measures.

## 6. EXTENSIONS

Section 6 and Section 7 are devoted to extensions of Aronszajn's theory which will be used later on. They can be skipped at first reading.

### 6.1. SCHWARTZ KERNELS

The last extension of the formalism dates from 1962 when Schwartz introduced the notion of hilbertian subspace of a topological vector space and pointed out the correspondence between hilbertian subspaces and kernels generalizing Aronszajn's ones. Schwartz considered vector spaces over the complex field $\mathbb{C}$. For the sake of simplicity we will restrict ourselves to real vector spaces. Let $\mathcal{E}$ be a topological vector space over the real field $\mathbb{R}$ locally convex, separated and quasi-complete (any bounded closed subset of $\mathcal{E}$ is complete; when $\mathcal{E}$ is a metric space quasi-complete is equivalent to complete).

DEFINITION 6 *A subspace $\mathcal{H}$ of $\mathcal{E}$ is called a hilbertian subspace of $\mathcal{E}$ if and only if $\mathcal{H}$ is a Hilbert space and the natural embedding*

$$\begin{aligned} I : \mathcal{H} &\longrightarrow \mathcal{E} \\ h &\longmapsto h \end{aligned}$$

*is continuous.*

This means that any sequence $(f_n)$ in $\mathcal{H}$ converging to some $f$ in the sense of the norm of $\mathcal{H}$ also converges to $f$ in the sense of the topology defined on $\mathcal{E}$. In other words, the Hilbert space topology on $\mathcal{H}$ is finer than the topology induced on $\mathcal{H}$ by the one on $\mathcal{E}$. Let $Hilb(\mathcal{E})$ be the set of hilbertian subspaces of $\mathcal{E}$. $Hilb(\mathcal{E})$ has a remarkable structure.

**1) External product by a positive real number.**

For any $\lambda \geq 0$ and any $\mathcal{H}$ in $Hilb(\mathcal{E})$, $\lambda\,\mathcal{H}$ is $\{0\}$ if $\lambda = 0$, otherwise $\lambda\,\mathcal{H}$ is the space $\mathcal{H}$ endowed with the inner product

$$< h, k >_{\lambda\mathcal{H}} = \frac{1}{\lambda} < h, k >_{\mathcal{H}} .$$

**2) Internal addition.**

$\mathcal{H}_1 + \mathcal{H}_2$ is the sum of the vector spaces $\mathcal{H}_1$ and $\mathcal{H}_2$ endowed with the norm defined by

$$||h||^2_{\mathcal{H}_1+\mathcal{H}_2} = \inf_{\substack{h_1 + h_2 = h \\ h_1 \in \mathcal{H}_1,\, h_2 \in \mathcal{H}_2}} \left(||h_1||^2_{\mathcal{H}_1} + ||h_2||^2_{\mathcal{H}_2}\right) .$$

$\mathcal{H}_1 + \mathcal{H}_2 \in Hilb(\mathcal{E})$. The internal addition in $Hilb(\mathcal{E})$ is associative and commutative, $\{0\}$ is a neutral element and we have

$$\begin{aligned}(\lambda + \mu)\,\mathcal{H} &= \lambda\,\mathcal{H} + \mu\,\mathcal{H} \\ \lambda\,(\mathcal{H}_1 + \mathcal{H}_2) &= \lambda\mathcal{H}_1 + \lambda\mathcal{H}_2.\end{aligned}$$

**3) Order structure.**

$$\mathcal{H}_1 \le \mathcal{H}_2 \Longleftrightarrow \begin{cases} \mathcal{H}_1 \subset \mathcal{H}_2 \\ \text{and } \forall h \in \mathcal{H}_1 \quad \|h\|_{\mathcal{H}_1} \ge \|h\|_{\mathcal{H}_2}. \end{cases}$$

In other words, $\mathcal{H}_1$ is a subset of $\mathcal{H}_2$ and the norm on $\mathcal{H}_1$ is finer than the norm on $\mathcal{H}_2$. We have

$$\mathcal{H}_1 \le \mathcal{H}_1 + \mathcal{H}_2,\, \mathcal{H}_2 \le \mathcal{H}_1 + \mathcal{H}_2,$$

$$\text{if } \mathcal{H} \ne \{0\} \qquad (\mathcal{H} \le \lambda\mathcal{H} \Longleftrightarrow \lambda \ge 1)\,.$$

Let $\mathcal{E}'$ be the topological dual of $\mathcal{E}$, that is the set of continuous linear mappings from $\mathcal{E}$ to $\mathbb{R}$, and let $L^+(\mathcal{E})$ be the set of linear mappings from $\mathcal{E}'$ to $\mathcal{E}$ that are

$$\begin{array}{lll} \text{symmetric} & \forall e' \in \mathcal{E}' & \forall f' \in \mathcal{E}' \langle\langle L(e'), f'\rangle\rangle = \langle\langle e', L(f')\rangle\rangle \\ \text{and positive} & \forall e' \in \mathcal{E}' & \langle\langle L(e'), e'\rangle\rangle \in \mathbb{R}^+ \end{array}$$

where $\langle\langle .,.\rangle\rangle$ is the duality bracket between $\mathcal{E}$ and $\mathcal{E}'$.
The elements of the set $L^+(\mathcal{E})$ are called Schwartz kernels relative to $\mathcal{E}$. $L^+(\mathcal{E})$ is also equipped with three structures induced by the usual operations (external product by a positive real number and internal addition) and the order relation between positive operators:

$$L_1 \le L_2 \Longleftrightarrow (L_2 - L_1) \text{ is a positive operator.}$$

The main result of the theory is stated in the following theorem.

THEOREM 18 *$L^+(\mathcal{E})$ and $Hilb(\mathcal{E})$ each equipped with the structures defined above are isomorphic.*
*The Schwartz kernel $\Xi$ of a hilbertian subspace $\mathcal{H}$ is equal to $I\,I^*$ where $I$ is the natural embedding: $\mathcal{H} \longrightarrow \mathcal{E}$ and $I^*$ is its adjoint: $\mathcal{E}' \longrightarrow \mathcal{H}$.*
*$\Xi$ is the unique mapping: $\mathcal{E}' \longrightarrow \mathcal{H}$ such that*

$$\forall e' \in \mathcal{E}', \quad \forall h \in \mathcal{H} \quad < h, \Xi(e') >_{\mathcal{H}} = \langle\langle h, e'\rangle\rangle.$$

*In other words the restriction of the mapping $\langle\langle ., e'\rangle\rangle$ to the space $\mathcal{H}$ is represented in $\mathcal{H}$ by $\Xi(e')$.*

**Example 1** Let E be the set $\mathbb{R}^E$ of $\mathbb{R}$-valued mappings defined on $E$ endowed with the pointwise convergence topology. In that case Schwartz 's theory coincides with Aronszajn's one. $\mathcal{E}'$ is equal to the space of measures with finite support in $E$. The reproducing kernel $K$ of a hilbertian subspace $\mathcal{H}$ is given by

$$K(.,y) = \Xi(\delta_y), \qquad y \in E$$

where $\Xi$ is the Schwartz kernel of $\mathcal{H}$. This is equivalent to

$$K(x,y) = < \Xi(\delta_y), \delta_x >_{\mathcal{H}} .$$

The expression of $\Xi$ in terms of $K$ is then given by, for all $L \in \mathcal{E}'$:

$$\Xi(L)(t) = LK(.,t).$$

**Example 2** $\mathcal{E} = L^2(\mu)$ (some additional conditions are needed to deal with classes of functions). The kernel $\Xi$ of a hilbertian subspace $\mathcal{H}$ is the unique mapping: $L^2(\mu) \longrightarrow \mathcal{H}$ such that

$$\forall e' \in L^2(\mu), \qquad \forall h \in \mathcal{H}, \qquad < h, \Xi(e') >_{\mathcal{H}} = \int_E h(x)e'(x)\, d\mu(x).$$

**Example 3 RKHS of Schwartz distributions** Meidan (1979) concentrates on the case of subspaces of the set of Schwartz distributions $\mathcal{D}'(\mathbb{R}^d)$ (see the appendix) and gives several stochastic applications. The originality of this approach is that, in contrast with the case of subspaces of $\mathbb{C}^E$, every Hilbert space of distributions is actually a hilbertian subspace of $\mathcal{D}'(\mathbb{R}^d)$. Let $T$ be an open subset of $\mathbb{R}^d$. $\langle\langle .,. \rangle\rangle_{\mathcal{D},\mathcal{D}'}$ will denote the duality bracket between $\mathcal{D}(T)$ and $\mathcal{D}'(T)$.
The "evaluation functionals" here are the maps $e_\phi$ indexed by the elements $\phi$ of $\mathcal{D}(T)$ and defined by:

$$\mathcal{H} \longrightarrow \mathbb{C} \tag{1.11}$$
$$f \longmapsto e_\phi(f) = \langle\langle f, \phi \rangle\rangle_{\mathcal{D},\mathcal{D}'} = f(\phi) \tag{1.12}$$

Meidan (1979) proves an equivalent of Moore's theorem with a correspondence between Schwartz kernels of $\mathcal{D}'(\mathbb{R}^d)$ and hilbertian subspaces of $\mathcal{D}'(\mathbb{R}^d)$.
Ylvisaker (1964) considers a family of kernels $\{K_\omega, \omega \in \Omega\}$ on $E$ and a probability space $(\Omega, \mathcal{B}, P)$ such that the kernels are measurable and integrable with respect to $\mu$ for each $(s,t) \in E \times E$. Then

$$K(s,t) = \int_\Omega K_\omega(t,s) d\mu(\omega)$$

is a reproducing kernel on $E$. Let $f = (f_1, \ldots, f_n)$ be a family of functions on $E$ linearly independent and $G_f^{K_\omega}$ be the Gram matrix $< f_i, f_j >_{K_\omega}$ . Let $G_f^{K_\omega} = 0$ if there exists an $i$ such that $f_i \notin \mathcal{H}_{K_\omega}$.

THEOREM 19 *Assume moreover that* $(G_f^K)^{-1}$ *is measurable and integrable with respect to* $\mu$, *then*

$$(G_f^K)^{-1} >> \int_\Omega (G_f^{K_\omega})^{-1} d\mu(\omega) \tag{1.13}$$

This inequality reduces to a simpler statement on matrices in the case $T$ is finite.

## 6.2. SEMI-KERNELS

The concept of semi-kernel occurs naturally in some RKHS problems. To quote two examples, they are useful in the characterization of spline functions (see Chapter 3) and they are related to variograms in the same fashion as kernels are related to covariance functions (see Chapter 2). They first appeared in Duchon (1975). Different frameworks are used though for their definition as in Laurent (1991) or in Bezhaev and Vasilenko (1993). We will follow Laurent's presentation here even though it is more difficult because we will need this degree of generality later.
Let $\mathcal{E}$ be a locally convex topological vector space and $\mathcal{E}'$ its dual space. In our applications, we will find $\mathcal{E} = \mathbb{R}^E$ for $E \subset \mathbb{R}^d$ in which case $\mathcal{E}'$ will be the space of finite combinations of Dirac functionals, or $\mathcal{E} = \mathcal{D}'(\mathbb{R}^d)$ with the obvious dual. Let us consider a linear subspace $\mathcal{H}$ of $\mathcal{E}$ endowed with a semi-inner product and induced semi-norm respectively denoted by $[.,.]_N$ and $|\ .\ |_N$, where $N$ is the kernel subspace of the semi-norm: $|\ x\ |_N = 0 \Leftrightarrow x \in N$. The semi-norm induces a natural inner product and norm on the factor space $\mathcal{H}/N$. If an element of $\mathcal{H}/N$ is denoted by $u + N$, then we have, for $u \in \mathcal{H} \setminus N$

$$< u + N, v + N >_{\mathcal{H}/N} = [u, v]_N \text{ and } \|\ u + N\ \|_{\mathcal{H}/N} = |\ u\ |_N\ .$$

$N^0$ denotes the set of linear functionals vanishing on $N$.

DEFINITION 7 *The space* $\mathcal{H}$ *endowed with the semi-norm* $|\ .\ |_N$ *is called a semi-Hilbert space if*
*1) the null space is finite dimensional*
*2) the factor space* $\mathcal{H}/N$ *endowed with the induced norm is complete*
*3) the factor space* $\mathcal{H}/N$ *is topologically included in* $\mathcal{E}/N$ *with the weak topology.*

Condition 3) is equivalent to
$| x_n |_N \longrightarrow 0$ *implies that* $\langle\langle x_n, u\rangle\rangle \longrightarrow 0$ *for all* $u \in N^0$.
Let $\Lambda$ be a linear subspace of $\mathcal{E}'$.

DEFINITION 8 *A linear mapping* $\mathbf{K}$ *from* $\Lambda$ *to* $\mathcal{E}$ *is called a semi-kernel operator for* $\mathcal{H}$ *and* $\Lambda$ *relatively to the semi-norm* $| \ . \ |_N$ *if for all* $u \in N^0 \cap \Lambda$ *and* $v \in \mathcal{H}$ *, we have* $\mathbf{K}u \in \mathcal{H}$ *and* $\langle\langle u, v\rangle\rangle = [\mathbf{K}u, v]_N$.

The last property expresses a "restricted" (or semi) reproducing property of the semi-kernel operator via the semi-inner product, restricted in the sense that it only reproduces functionals that vanish on the subspace $N$.
We observe that $\mathbf{K}$ is not uniquely defined since if $L$ is any linear map from $\Lambda$ into $N$, then $\mathbf{K} + L$ also satisfies the semi-reproducing property. However the map induced by $\mathbf{K}$ from $N^0$ into $\mathcal{H}/N$ is unique. It corresponds to the Schwartz 's kernel of the hilbertian subspace $\mathcal{H}/N$. In this sense, $\mathbf{K}$ is the semi-counterpart of the Schwartz kernel operator.
The map $\mathbf{K}$ has a "semi-symmetry" property in the sense that, for all $m$ and $l$ in $N^0 \cap \Lambda$:

$$\langle\langle l, \mathbf{K}m\rangle\rangle = [\mathbf{K}l, \mathbf{K}m]_N = [\mathbf{K}l, \mathbf{K}m]_N = \langle\langle \mathbf{K}l, m\rangle\rangle \tag{1.14}$$

Similarly, it has a "semi-positivity" property in the sense that, for all $m \in N^0 \cap \Lambda$, we have

$$\langle\langle \mathbf{K}m, m\rangle\rangle = [\mathbf{K}m, \mathbf{K}m]_N \geq 0 \tag{1.15}$$

It is important to note however that the last two properties are not necessarily satisfied when $m$ or $l$ vary in the whole space $\mathcal{H}$.
In the case $\mathcal{H} \subset \mathbb{R}^E$, this leads to the definition of semi-reproducing kernel, by the correspondence:

$$K^*(t, s) =< \mathbf{K}(K(t, .)), K(s, .) >_{\mathcal{H}} \qquad \text{for all } t, s \in E, \tag{1.16}$$

or equivalently,

$$K^*(t, .) = \mathbf{K}K(t, .) \tag{1.17}$$

Let us write the semi-reproducing property in this case. We have, as in formula (1.6)

$$\mathbf{K}u(.) =< u(..), K^*(., ..) >_{\mathcal{H}} \tag{1.18}$$

Combined with definition (8), this yields, for all $u \in N^0$ and $v \in \mathcal{H}$:

$$< u, v >_{\mathcal{H}} = [< u(..), K^*(., ..) >_{\mathcal{H}}, v(.)]_N \tag{1.19}$$

It is equivalent to the following: for any finite linear combination of Dirac functionals vanishing on $N$ (*i.e.* $\sum \lambda_i v(t_i) = 0 \forall v \in N$), we have

$$\sum \lambda_i x(t_i) = [x, \sum \lambda_i K^*(., t_i)]_N \tag{1.20}$$

**Example** $\mathcal{H} = H^m(0,1)$ is endowed with the norm:

$$\| u \|^2 = \sum_{j=0}^{m-1} u^{(j)}(0)^2 + | u |_N \tag{1.21}$$

and the semi-norm: $| u |_N = \int_0^1 u^{(m)}(t)^2 d\lambda(t)$. Its null space is clearly the set of functions on $(0,1)$ which coincide with polynomials of degree less than or equal to $m-1$. The factor space $\mathcal{H}/N$ is clearly isometrically isomorphic to the subset $\{u \in H^m(0,1) : u^{(j)}(0) = 0\}$. To check 3) we need to prove that, if $| u_n | \longrightarrow 0$ , then for all $\lambda_i$ such that $\sum \lambda_i t_i^j = 0, j = 1, \ldots, m-1$, we have $\sum \lambda_i u_n(t_i) \longrightarrow 0$. It suffices to use the reproducing property of $H^m(0,1)$ with the norm, denoting by $K$ its reproducing kernel. Then

$$\begin{aligned}\sum_{i=1}^{p} \lambda_i u_n(t_i) &= \sum_{i=1}^{p} \lambda_i \sum_{j=1}^{m-1} u_n^{(j)}(0) \frac{\partial^j K}{s^j}(t_i, 0) \\ &+ \sum_{i=1}^{p} \lambda_i \int_0^1 u_n^{(m)}(s) \frac{\partial^m K}{s^m}(t_i, s) d\lambda(s)\end{aligned}$$

The first sum vanishes because of the condition on the $\lambda_i$ and the fact that $\frac{\partial^j K}{s^j}(t_i, 0)$ is a polynomial in $t_i$. To see that the second sum tends to 0 use the Cauchy Schwarz inequality in each integral and the assumption $| u_n | \longrightarrow 0$.

In the case $\mathcal{H} \subset \mathbb{R}^E$ and finite dimensional $N$, Bezhaev and Vasilenko prove an important relationship between kernels and semi-kernels. This formula can be found in particular cases in Meinguet(1979) for kernels related to thin plate splines and in Thomas-Agnan (1991) for kernels related to alpha-splines.

# 7. POSITIVE TYPE OPERATORS

## 7.1. CONTINUOUS FUNCTIONS OF POSITIVE TYPE

A function of one variable $f : E \longrightarrow \mathbb{C}$ (E open set of $\mathbb{R}^d$) is said to be of positive type if the function of two variables defined on $E \times E$ by

$$(x, y) \longmapsto f(x - y)$$

is a function of positive type according to Definition 1.3. The following Bochner theorem (Bochner, 1932) characterizes the class of continuous functions of positive type by the behavior of their Fourier transform.

THEOREM 20 *The Fourier transform of a bounded positive measure on $\mathbb{R}^d$ is a continuous function of positive type. Conversely, any function of positive type is the Fourier transform of a bounded positive measure.*

Let us give the proof of the direct statement, which is the easiest and refer the reader to Gelfand and Vilenkin (1967) for details of the converse.

**Proof.** Let $\mu$ be a bounded positive measure. One can easily see that the Fourier transform of the corresponding distribution is the continuous function $f$:

$$f(x) = \int_{\mathbb{R}^d} \exp(2\pi \mathbf{i} < \omega, x >) d\mu(\omega).$$

Then, for all $(a_1, \ldots, a_n) \in \mathbb{C}^n$ and all $(x_1, \ldots, x_n) \in (\mathbb{R}^d)^n$, we have

$$\begin{aligned}\sum_{i,j=1}^{n} a_i \bar{a}_j f(x_i - x_j) &= \textstyle\sum_{i,j=1}^{n} a_i \bar{a}_j \int_{\mathbb{R}^d} \exp(2\pi \mathbf{i} < \omega, x_i - x_j >) d\mu(\omega) \\ &= \textstyle\int_{\mathbb{R}^d} \mid \sum_{j=1}^{n} a_j \exp\left(2\pi \mathbf{i} < x_j, \omega >\right) \mid^2 d\mu(\omega).\end{aligned}$$

■

Micchelli(1986) gives sufficient conditions on a function $F : \mathbb{R}^{*+} \longrightarrow \mathbb{R}$ so that the function $f(x) = F(\| x \|^2)$ be conditionally positive definite of order $k$ for all dimensions $s$.

**Examples** $f(x) = \exp(-\lambda \| x \|^\tau)$ is positive definite for all $\lambda > 0$ and $0 < \tau \leq 2$ (it is the characteristic function of a stable law, see Luckas (1970)).

The following families of functions are known to be of positive type (Micchelli, 1986):

$$f(x) = \frac{1}{(r^2 + \| x \|^2)^\alpha}$$

for all $r > 0$ and all $\alpha > 0$.

$$f(x) = \frac{1}{(1 + \| x \|^\tau)^\alpha}$$

for all $\alpha > 0$ and $0 < \tau \leq 2$.

Another characterization of positive definiteness in the continuous case is the following theorem (see Stewart, 1976).

THEOREM 21 *A continuous function $f$ is positive definite if and only if for all continuous function with compact support $\phi$, we have*

$$\int\!\!\int f(s-t)\phi(s)\phi(t)\, d\lambda(s)\, dlambda(t) \geq 0. \qquad (1.22)$$

Note that if $f$ is not restricted to be continuous, there are unbounded functions that satisfy (1.22). A characterization of functions that satisfy (1.22) for various conditions on $\phi$ is found in Stewart(1976).
Note that if $f$ is positive definite on $\mathbb{R}^d$, then it is also positive definite on $\mathbb{R}^p$ for all $p \leq d$, but the converse does not hold.

## 7.2. SCHWARTZ DISTRIBUTIONS OF POSITIVE TYPE OR CONDITIONALLY OF POSITIVE TYPE

**SCHWARTZ DISTRIBUTIONS OF POSITIVE TYPE.**
Schwartz (1964) proved an extension of this result to the case of Schwartz distributions of positive type. Let us deduce the definition of Schwartz distributions of positive type from the definition of positive operator (see paragraph 6.1). Let $T$ be an open subset of $\mathbb{R}^d$. $D$ will denote the derivation operator.
A Schwartz 's distribution $f$ induces an operator $L_f$ from $\mathcal{D}(\mathcal{T})$ to $\mathcal{D}'(\mathcal{T})$ as follows: for $\phi \in \mathcal{D}(T)$, $L_f(\phi)$ is the element of $\mathcal{D}'(T)$ defined by

$$\psi \in \mathcal{D}(T) \longmapsto < f, \phi * \overline{\psi} >$$

where $\phi * \overline{\psi}$ is the convolution product:

$$\phi * \overline{\psi}(t) = \int_T \phi(s)\overline{\psi}(t-s)d\lambda(s)$$

A Schwartz 's distribution $f$ is then of positive type if the operator $L_f$ is a positive operator. This is equivalent to state that

$$\forall \phi \in \mathcal{D}(T), \qquad \langle\langle f, \phi * \overline{\phi}\rangle\rangle \geq 0 \tag{1.23}$$

Note that for a continuous function, we have

$$< f, \phi * \overline{\phi} >= \int\int \overline{f(x-y)}\phi(y)\overline{\phi(x)}d\lambda(x)d\lambda(y)$$

and one can show that (1.23) is then equivalent to the ordinary definition. However there exist discontinuous unbounded functions that satisfy (1.23). The Bochner-Schwartz theorem characterizes the class of Schwartz distributions of positive type, and uses the following notion of slowly increasing measure.

DEFINITION 9 *A positive measure $\mu$ is said to be slowly increasing if there exists an integer p such that the function* $(1+ \mid x \mid^2)^{-p}$ *is integrable with respect to $\mu$.*

THEOREM 22 (BOCHNER-SCHWARTZ THEOREM) *The Fourier transform of a slowly increasing positive measure is a Schwartz distribution of positive type.*
*Conversely, any Schwartz distribution of positive type is the Fourier transform of a slowly increasing positive measure.*

**Remark** The class of tempered distributions (elements of $\mathcal{S}'(\mathcal{T})$) of positive type also coincides with the class of Fourier transforms of slowly increasing positive measures. This theorem allows us to find easy examples of distributions of positive type. For a positive integer $m$, the distribution $(-1)^m\delta^{(2m)}$ and the distribution $(1 - D^2)^m\delta$ are of positive type, their Fourier transform being respectively $(x/(2\pi))^{2m}$ and $(x^2+1)^m$. Similarly, if $f$ is a distribution of positive type, then $D\overline{D}f$ is also of positive type (the converse is false). An easy consequence of this theorem yields another characterization.

THEOREM 23 *A distribution $f \in \mathcal{D}'(T)$ is of positive type if and only if there exists a continuous function of positive type $u$ and an integer $p$ such that $f = (1 - \Delta)^p u$, where $\Delta$ is the Laplace operator*

$$\Delta = \sum_{k=1}^{d} \frac{\partial^2}{\partial x_k^2}.$$

**SCHWARTZ DISTRIBUTIONS CONDITIONALLY OF POSITIVE TYPE.**

We adopt Gelfand and Vilenkin's definition of conditionally positive definiteness.

DEFINITION 10 *A Schwartz distribution $f$ is conditionally of positive type of order $m$ if for all $\phi \in \mathcal{D}(T)$, and all linear differential operator $L$ homogeneous of order $m$ with constant coefficients, $< L\overline{L}f, \phi * \overline{\phi} > \geq 0$.*

Note that if $f$ is conditionally of positive type of order $m$, then $f$ is also conditionally of positive type of order $p$ for all $p \geq m$.
Let $P_L$ be the polynomial associated with $L$

$$\text{if} \quad L = \sum_{|k|=m} a_k D^k \quad \text{then} \quad P_L(s) = \sum_{|k|=m} a_k(-2\pi \mathbf{i} s)^k$$

By the Bochner-Schwartz theorem it is easy to see that, if $f$ is an element of $\mathcal{S}'(\mathcal{T})$, then $f$ is conditionally of positive type of order $m$ if and only if for all polynomial $P$ homogeneous of degree $m$, $L\overline{L}f$ is of positive type, *i.e.* $| P |^2 \mathcal{F}f$ is a positive slowly increasing measure.
Gelfand and Vilenkin (1967) give an equivalent to the Bochner-Schwartz theorem in the conditionally positive type case.

THEOREM 24 *Let $f$ be an element of $\mathcal{D}'(T)$ conditionally of positive type of order $m$. There exists a slowly increasing positive measure $\mu$ such that $\int_{0<|\lambda|<1} |\lambda|^{2s}\, d\mu(\lambda) < \infty$ and that for all $\phi \in \mathcal{D}(T)$*

$$< f, \phi > = \int_{T\setminus\{0\}} \left( \mathcal{F}\phi(s) - \beta(s) \sum_{|k|=0}^{2m-1} \frac{\mathcal{F}\phi^{(k)}(0)}{k!} s^k \right) d\mu(s) + \sum_{|k|=0}^{2m} a_k \frac{\mathcal{F}\phi^{(k)}(0)}{k!},$$

*where $\beta(.)$ is a function such that $\beta(.) - 1$ has a zero of order $2m+1$ at the origin and the complex numbers $a_k$ for $|k| = 2s$ are such that $\sum_{|i|=|j|=m} a_{i+j} z_i \overline{z}_j$ defines a positive definite hermitian form.*

There exists a simpler equivalent condition for a function to be conditionally of positive type of order $m$ in the case of distributions which are represented by continuous functions. The equivalence is demonstrated in Gelfand and Vilenkin (p 274) for the case $m = 1$ and one variable and in Madych and Nelson (1990) for the general case.

THEOREM 25 *A (continuous) function $f$ is conditionally of positive type of order $m$ if and only if*

$$\sum_{i,j=1}^{n} f(x_i - x_j) z_i \overline{z}_j \geq 0 \tag{1.24}$$

*for all $(z_1, \ldots, z_n)$ such that $\sum_{i=1}^{n} z_i P(x_i) = 0$, for all $P$ polynomial of degree less than or equal to $m$.*

A Bochner-Schwartz equivalent in that case is found in Cressie (1993).

THEOREM 26 *If $f$ is a continuous function on $\mathbb{R}^d$ satisfying $f(0) = 0$, then $-f$ is conditionally of positive type if and only if*

$$f(h) = Q(h) + \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{1 - \cos(\omega' h)}{\| \omega \|^2} G(d\omega) \tag{1.25}$$

*where $Q$ is a quadratic form and $G$ is a positive symmetric measure without atom at the origin that satisfies $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (1 + \| \omega \|^2)^{-1} G(d\omega) < \infty$.*

The case $m = 0$ corresponds to the condition $\sum_{i=1}^{n} z_i = 0$ and is sometimes referred to without mention of the order. For $m = -1$ corresponding to no condition, the above definition reduces to the usual definition

of function of positive type. Micchelli (1986) gives sufficient conditions for a function $f$ on $\mathbb{R}^s$ of the form: $f(x) = F(\| x \|^2)$, where $F$ is a function from $(0, \infty)$ into $\mathbb{R}$, to be conditionally positive definite of order $k$ for any dimension $s$. There are extensions of this definition for example in Myers (1992) and Atteia(1992). In Atteia(1992), for a closed linear subspace of $\mathcal{D}'(\mathbb{R}^d)$, an element $f$ of $\mathcal{D}'(\mathbb{R}^d)$ is $N$-conditionally of positive type if for all $\phi$ in $N^0$, we have $\langle\langle f, \phi * \phi' \rangle\rangle \geq 0$.
To our knowledge, the relationship between semi-kernels and functions conditionally of positive type has not been explored completely. It is clear that when the kernel subspace of the semi-norm is a space of polynomials, semi-kernels are conditionally of positive type in the classical sense. Atteia(1992) explores a Moore's equivalent for semi-hilbertian kernels and $N$-conditionally distributions of positive type.
Conditionally positive definite functions are related to the exchangeability of random variables (Johansen, 1960).

## 8. EXERCISES

1 Let $K_0$ be a probability density on $\mathbb{R}$. Prove that the mapping

$$(P,Q) \longmapsto < P,Q >_{K_0} = \int_{\mathbb{R}} P(x)Q(x)K_0(x)\,d\lambda(x)$$

defines an inner product on the space $\mathbb{P}_r$ of polynomials of degree at most $r$.

2 Prove that the spaces defined in Examples 1 to 5 are Hilbert spaces.

3 A function $f:\ \mathbb{R} \longrightarrow \mathbb{C}$ of one single variable is said to be of positive type if the function $K$ :

$$\begin{array}{ccc} \mathbb{R}\times\mathbb{R} & \longrightarrow & \mathbb{C} \\ (x,y) & \longmapsto & K(x,y) = f(x-y) \end{array}$$

is a positive type function (see subsection 7.1). Show that
a) If $f$ is of positive type, then $\overline{f}$ is of positive type and

$$\max_{x\in\mathbb{R}} |f(x)| = f(0).$$

b) Any finite linear combination of functions of positive type with non negative coefficients is also of positive type.
c) The pointwise limit of a sequence of functions of positive type is of positive type.
d) The product of a finite number of functions of positive type is also of positive type.

4 Let $L$ be nonnegative on the diagonal $D = \{(x,x) : x \in E\}$ of $E \times E$ and such that

$$L(x,y) = -L(y,x) \text{ if } x \neq y.$$

Prove that $L$ satisfies (1.3) with $\mathbb{C}^n$ replaced with $\mathbb{R}^n$ but that $L$ is not a reproducing kernel if it is not identically 0 outside $D$.

5 Let $V_k$ $(k \geq 0)$ be the vector space of functions $f : \mathbb{R} \longrightarrow \mathbb{R}$ spanned by $e_0, \ldots, e_k$, with

$$\begin{aligned} \forall x \in \mathbb{R},\ e_0(x) &= \frac{1}{\sqrt{2\pi}} \\ \forall k \in \mathbb{N}^*,\ \forall x \in \mathbb{R},\ e_{2k}(x) &= \frac{1}{\sqrt{\pi}}\cos(kx) \\ e_{2k-1}(x) &= \frac{1}{\sqrt{\pi}}\sin(kx) \end{aligned}$$

a) Show that

$$(f,g) \longrightarrow < f,g >= \int_0^{2\pi} f(x)g(x)\, d\lambda(x)$$

defines a Hilbert space structure on $V_k$ for which $(e_0,\ldots,e_k)$ is an orthonormal basis.
b) For $q \in \mathbb{N}$, let $D_q$ be the Dirichlet kernel of order $q$ :

$$\begin{aligned} D_q(x-y) &= \frac{1}{2\pi}\frac{\sin[(q+1/2)(x-y)]}{\sin[(x-y)/2]} \text{ if } (x-y) \notin 2\pi\mathbb{Z}, \\ D_q(0) &= \frac{2q+1}{2\pi}. \end{aligned}$$

Let $H_k$ be the reproducing kernel of $V_k$. Show that

$$\begin{aligned} H_{2q}(x,y) &= D_q(x-y) \\ H_{2q+1}(x,y) &= D_q(x-y) + \frac{1}{\pi}\sin[(q+1)x]\sin[(q+1)y]. \end{aligned}$$

6 For any sequence $u = (u_i)_{i\in\mathbb{N}^*}$ of real numbers denote by $\Delta u$ the sequence $(u_{i+1}-u_i)_{i\in\mathbb{N}^*}$. Let $0 < \theta < 1$ and $H_\theta$ be the set of sequences $(u_i)_{i\in\mathbb{N}^*}$ such that

$$\begin{aligned} \lim_{i\to\infty} u_i &= 0 \\ \text{and } \sum_{i=1}^{\infty} \frac{(\Delta u)_i^2}{\theta^i} &< \infty. \end{aligned}$$

Show that the mapping $< .,. >_\theta$ defined by

$$< u,v >_\theta = \sum_{i=1}^{\infty} \frac{(\Delta u)_i(\Delta v)_i}{\theta^i},$$

where $u = (u_i)_{i\in\mathbb{N}^*}$ and $v = (v_i)_{i\in\mathbb{N}^*}$, is an inner product on $H_\theta$. Show that, endowed with this inner product, $H_\theta$ is a Hilbert space with reproducing kernel $G$ such that

$$\forall (i,j) \in \mathbb{N}^* \times \mathbb{N}^*, \qquad G(i,j) = \frac{\theta^{\max(i,j)}}{1-\theta}.$$

7 Let $E$ be a subset of $\mathbb{R}^{+*}$, $n \geq 1$ and $t_1 < t_2 < \ldots < t_n$ be $n$ elements of $E$. Let $Y_1, Y_2, \ldots, Y_n$ be $n$ independent zero mean gaussian

variables with respective variances $t_1, t_2 - t_1, \ldots, t_n - t_{n-1}$. Consider the random vector $(X_1, X_2, \ldots, X_n)$ where

$$\forall i \in \{1, \ldots, n\} \qquad X_i = \sum_{k=1}^{i} Y_k.$$

1) Show that any variable $X_i$ is a zero mean gaussian random variable with variance $t_i$ and that, for $i < j$,

$$E(X_i X_j) = t_i.$$

Deduce from 1) that the function

$$\begin{array}{rcl} E \times E & \longrightarrow & \mathbb{R} \\ (t, s) & \longmapsto & \min(t, s) \end{array}$$

is of positive type.

8 Prove that $\exp(sf(.))$ is of positive type for all values of $s$ if and only if $f$ is conditionally of positive type of order 1. Prove that this condition is also equivalent to the fact that for all $x_0$,

$$f(x-y) - f(x - x_0) - f(x_0 - x) + f(0)$$

is of positive type.

9 Prove that for any kernel $K$ on $E \times E$, a function $f$ on $E$ belongs to the RKHS $\mathcal{H}_K$ with kernel $K$ if and only if there exists $\lambda > 0$ such that $K(s,t) - \lambda f(s) f(t)$ is positive definite.

10 Considering the spaces $l^2\mathrm{N}$ and $L^2(0,1)$, prove that the RKHS property is not invariant under Hilbert space isomorphism.

11 [Any Hilbert space is isomorphic to a RKHS] Let $\mathcal{E}$ be a Hilbert space with the inner product

$$\begin{array}{rcl} \mathcal{E} \times \mathcal{E} & \longrightarrow & \mathbb{C} \\ (\alpha, \beta) & \longrightarrow & <\alpha, \beta>_{\mathcal{E}} = K(\alpha, \beta). \end{array}$$

Show that $K$ is a positive type function on $\mathcal{E} \times \mathcal{E}$. Let $\mathcal{H}$ be the Hilbert space of functions on $\mathcal{E} \times \mathcal{E}$ with reproducing kernel $K$. Show that $\mathcal{H}$ is equal to the set of functions

$$\{K(., \beta); \beta \in \mathcal{E}\}$$

and that $\mathcal{E}$ and $\mathcal{H}$ are isomorphic.

12 Prove that if $K$ is bounded on the diagonal of $E \times E$, then any function of the RKHS $\mathcal{H}_K$ with kernel $K$ is also bounded. In that case, prove that if a sequence converges in the $\mathcal{H}_K$-norm sense, then it converges also in the uniform norm sense.

13 Let $\mathcal{H}$ be a Hilbert space, $D$ a set and $h$ a map from $D$ to $\mathcal{H}$. Let $\mathcal{H}_1$ be the closed linear span of $\{h(t); t \in D\}$. Prove that the set $\mathcal{F}$ of functions $s \longrightarrow < x, h(s) >$ when $x$ ranges in $\mathcal{H}$ is a RKHS with kernel $R$ given by

$$R(s,t) = < h(t), h(s) > .$$

14 [A Hilbert space of functions with no reproducing kernel] Consider the RKHS of Example 4: $E = (0,1)$ and

$$\mathcal{H} = \left\{\varphi \mid \varphi(0) = 0,\ \varphi \text{is absolutely continuous and } \varphi' \in L^2(0,1)\right\}$$

The reproducing kernel of $\mathcal{H}$ is $\min(x,y)$ and the inner product is given by

$$< \varphi, \psi > = \int_0^1 \varphi' \overline{\psi}' d\lambda.$$

Show that the functions $(\min(.,x))_{x \in E}$ are linearly independent on $(0,1)$ and that the same property holds for the functions $(\cos(.\ x))_{x \in E}$.
Let $\mathcal{H}_0$ be the subset of $\mathcal{H}$ spanned by the functions $(\min(.,x))_{x \in E}$ and consider an algebraic complement $\mathcal{S}_0$ of $\mathcal{H}_0$ in $\mathcal{H}$. Any element $f$ of $\mathcal{H}$ can be uniquely written as

$$f = f_0 + s_0$$

where $f_0 \in \mathcal{H}_0$ and $s_0 \in \mathcal{S}_0$.
Let $\mathcal{I}$ be the mapping:

$$\begin{array}{ccc} \mathcal{H} & \longrightarrow & \mathbb{C}^E \\ f & \longmapsto & \mathcal{I}(f) \end{array}$$

be defined in the following way: if $f = f_0 + s_0$ with $f_0 \in \mathcal{H}_0$ and $s_0 \in \mathcal{S}_0$ then
either $f_0 = 0$ and $\mathcal{I}(f) = s_0$
or there exist unique vectors $(a_1, \ldots, a_n)$ and $(x_1, \ldots, x_n)$ such that

$$f_0 = \sum_{i=1}^{n} a_i \min(., x_i).$$

In the latter case let

$$\mathcal{I}(f) = \sum_{i=1}^{n} a_i \cos(.\ x_i) + s_0.$$

Show that $\mathcal{I}$ is an algebraic isomorphism from $\mathcal{H}$ onto $\mathcal{M} = \mathcal{I}(\mathcal{H})$. Define the mapping $B$:

$$\begin{array}{ccc} \mathcal{M} \times \mathcal{M} & \longrightarrow & \mathbb{C} \\ (f,g) & \longmapsto & B(f,g) = < \mathcal{I}^{-1}(f), \mathcal{I}^{-1}(g) >_{\mathcal{H}} . \end{array}$$

Show that $\mathcal{M}$ is a vector space in which $\mathcal{I}(\mathcal{H}_0)$ and $\mathcal{I}(\mathcal{S}_0)$ are complementary and that $B$ is an inner product on $\mathcal{M}$.
Compute the norm of $\cos(.\ /n), n \in \mathbb{N}^*$, and show that the evaluation functional at the point 0 is not continuous on $\mathcal{M}$.
Show that $\mathcal{M}$ is a Hilbert space of functions with no reproducing kernel.

15 Let $(G(i,j))_{i,j \geq 1}$ be an infinite real matrix. Let

$$G_n = (G(i,j))_{1 \leq i,j \leq n}$$

be the $n \times n$ matrix obtained from $G$ by truncation. Show that $G$ is the reproducing kernel of a Hilbert space of real sequences if and only if for any $n \in \mathbb{N}^*$, $G_n$ is a symmetric non negative definite matrix.

16 Let $V$ be a closed subspace of a Hilbert space $\mathcal{H}$ with kernel $K$. Let $P_V$ be the projection operator of $\mathcal{H}$ onto $V$.
a) Prove that the kernel of $P_V$ (see Definition 3) coincides with $K_V$, the reproducing kernel of $V$.
b) Prove that for all $t \in E$, we have $K_V(t,t) \leq K(t,t)$.
c) Define explicitly the projection operator of $L^2(-\pi, \pi)$ onto the subspace of trigonometric polynomials of order less than or equal to $m$.
d) Define explicitly the projection operator of $L^2(\mathbb{R})$ onto the Paley-Wiener space, subset of $L^2(\mathbb{R})$ of elements whose Fourier transform has support included in $(-\pi, \pi)$.

17 Let $\mu$ be a probability measure on the interval $(a,b)$, absolutely continuous with respect to Lebesgue measure on this interval. Let $N(s,t) = 1_{(a,b)}(s)$.
a) prove that:

$$\mu((a, \min(x,y)) = \int_a^b N(x,s)N(y,s)d\mu(s)$$

b) use a) to prove that $K(x,y) = \mu((a,\min(x,y))$ is a function of positive type.
c) Prove that $\mathcal{H}_K$ is the set of continuous functions $F$ on the interval $(a,b)$ such that there exists a measure $\nu_F$, absolutely continuous with respect to Lebesgue measure, satisfying

$$F(x) = \nu_F((a,x)) \text{ and } \frac{d\nu_F}{d\mu} \in L^2(\mu),$$

endowed with the inner product:

$$< F, G >= \int_a^b \frac{d\nu_F}{d\mu}\frac{d\nu_G}{d\mu} d\mu.$$

18 Try to generalize the construction of Exercise 14 to any RKHS, using a suitable family of functions in place of $(\cos(.\ x))_{x\in E}$.

19 Let $K$ be a reproducing kernel on the set $T$ and let $t_1, \ldots, t_n$ be $n$ fixed distinct points of $T$. Let $E$ be a positive constant and let $H_E$ be the set of functions $f$ in $\mathcal{H}_K$ such that $\| f \|_K^2 \leq E$. Let $\lambda$ and $\theta$ be respectively the largest eingenvalue of the $n \times n$ matrix $(K(t_i,t_j))$.
1) using the Schwarz inequality, prove that for any $f \in \mathcal{H}_K$,

$$(\sum_{i=1}^{n} f^2(t_i))^2 \leq \| f \|_K^2 \sum_{i=1}^{n}\sum_{j=1}^{n} f(t_i)f(t_j)K(t_i,t_j) \tag{1.26}$$

2) using 1), prove that

$$\frac{(\sum_{i=1}^{n} f^2(t_i))^2}{\| f \|^2} \leq \lambda \tag{1.27}$$

3) let $\mu$ be the eigenvector corresponding to $\lambda$ such that $\sum_{i=1}^{n} \mu_i^2 = E$ and let $h(s) = \sum_{i=1}^{n} \mu_i K(s,t_i)$. Prove that $h$ maximizes $(\sum_{i=1}^{n} f^2(t_i))^2$ among elements of $H_E$.

20 Let $L_1$ and $L_2$ be bounded linear operators defined on a RKHS $\mathcal{H}_K$, and let $\Lambda_1(x,y)$ and $\Lambda_2(x,y)$ be their associated kernels in the sense that $Lf(y) =< f(x), \Lambda(x,y) >_K, \forall f \in \mathcal{H}_K$.
1) prove that if $\alpha_1$ and $\alpha_2$ are two reals, the kernel associated to $\alpha_1 L_1 + \alpha_2 L_2$ is $\alpha_1\Lambda_1 + \alpha_2\Lambda_2$.
2) prove that the operator associated with $L_1^*$ is $\Lambda(y,x)$
3) prove that the operator associated with $L_1 L_2$ is

$$\Lambda(x,y) =< \Lambda_1(.,y), \Lambda_2(x,.) > .$$

4) $L_1$ is a positive operator if and only if $\Lambda_1$ is positive definite.
5) If $\Lambda$ is a symmetric kernel, then $\Lambda$ is the kernel of a bounded self adjoint operator with lower bound $m$ and upper bound $M$ if and only if $MK - \Lambda$ is positive definite and $\Lambda - mK$ is positive definite.

21 Let $R$ and $K$ be two reproducing kernels on $E$. Assume that $K$ belongs to the tensor product $\mathcal{H}_R \bigotimes \mathcal{H}_R$. Prove that $\mathcal{H}_{R+K}$ and $\mathcal{H}_R$ consist of the same functions and that we have for all $f$ and $g$ in $\mathcal{H}_R$:

$$< f, g >_{R+K} = < (I + \mathbb{K}^{-1}) f, g >_R \tag{1.28}$$

where $\mathbb{K}$ is the operator $\mathcal{H}_R \longrightarrow \mathcal{H}_R$ defined by

$$\mathbb{K} f(t) = < K(., t), f >_R .$$

22 Prove that a function $f$ belongs to $\mathcal{H}_K$ if and only if there exists a constant $C$ such that for all $s$ and $t$ in $E$, $f(s)f(t) \leq C^2 K(s,t)$ and that the minimum of such $C$ coincides with $\| f \|_K$.