Boletín de la Sociedad Matemática Mexicana Vol. 28, No. 2, 1983

POSITIVE DEFINITE KERNELS*

BY S. D. CHATTERJI

§0. Introduction

We indicate a very general factorization theorem valid for positive definite operator valued kernels in §2. This is preceded in §1 by an introductory survey of the uses of scalar valued positive definite kernels. In the last section §3, we make a few historical remarks and refer to articles where further details, mathematical as well as bibliographical, can be obtained. An attempt is made, however, to make this paper reasonably self-contained.

§1. Scalar valued kernels: an introductory survey

Let S be an arbitrary set and let $K: S \times S \to \mathbb{C}$ *be such that*

(1) $$\sum_{j,k} K(t_j, t_k)\alpha_j\bar{\alpha}_k \geqslant 0$$

for every choice of points $t_1, \cdots, t_n$ in S and $\alpha_1, \cdots, \alpha_n$ in $\mathbb{C}$ and $n = 1, 2, \cdots$. Such a K is called a ($\mathbb{C}$-valued) positive definite (p.d.) kernel on S. If in the above $\mathbb{C}$ is replaced by $\mathbb{R}$ then an additional requirement of symmetry $(K(s, t) = K(t, s))$ is imposed on K; for $\mathbb{C}$-valued kernels K, (1) implies automatically that $K(s, t) = \overline{K(t, s)}$. In the sequel, we shall discuss only the $\mathbb{C}$-valued case; we assume henceforth also that all the vector spaces which intervene in the ensuing discussion are over the field $\mathbb{C}$. The $\mathbb{R}$-valued case is perfectly analogous and is not spelled out in detail for the sake of brevity; the only change needed is that a certain symmetry hypothesis has to be imposed to start with.

The main structure theorem for a p.d. kernel K is that there is a vector subspace $\mathscr{H}$ of the vector space $\Phi(S, \mathbb{C})$ of all $\mathbb{C}$-valued functions on S, which is such that (i) $\mathscr{H}$ is a Hilbert space with respect to a certain inner product $\langle \cdot \mid \cdot \rangle$ (We assume linearity in the second argument as in the physical literature.); (ii) for all $s \in S$, $K_s \in \mathscr{H}$ where $K_s(t) = K(s, t)$; (iii) for all $f \in \mathscr{H}$, $\langle K_s \mid f \rangle = f(s)$. It follows that the set $\{K_s : s \in S\}$ is total in $\mathscr{H}$ and the map $s \mapsto K_s$ from S to $\mathscr{H}$ gives the factorization $K(s, t) = \langle K_t \mid K_s \rangle$. The space $\mathscr{H}$ is called the reproducing kernel Hilbert space associated with K and the theorem itself is often called the Moore-Aronszajn (reproducing kernel) theorem.

The most straight-forward proof of the above theorem is to introduce an inner product in the linear manifold M generated by the elements K_s, $s \in S$, via the formula $\langle K_t \mid K_s \rangle = K(s, t)$ and the sesquilinearity of $\langle \cdot \mid \cdot \rangle$. The relation (1) then gives the positivity of the inner product; it turns out further that $\langle f \mid f \rangle = \| f \|^2 = 0$ for $f \in M$ if and only if $f = 0$. It remains then only to

* Presented at the "Workshop on the prediction theory on non-stationary processes and related topics", held at Centro de Investigación en Matemáticas (CIMAT), Guanajuato, México, June 20–26, 1982.

complete M under $\langle \cdot \mid \cdot \rangle$ to obtain $\mathscr{A}$. Of course, it has to be verified that $\mathscr{A}$ consists of elements of $\Phi(S, \mathbb{C})$; this is actually not too hard to do.

Although the theorem is easy to prove, its consequences can be surprisingly far-reaching. Indeed, it may be argued that wherever positive definiteness plays any role, the above theorem can be used profitably. For instance, the existence of a Gaussian process with K as covariance is a corollary. We need only take an orthonormal basis $\{e_\alpha, \alpha \in I\}$ in the associated reproducing kernel space $\mathscr{A}$ and construct a probability space (Ω, Σ, P) which can support a family $\{\eta_\alpha, \alpha \in I\}$ of independent, real, Gaussian random variables with zero mean and unit variance. This is easy via the construction of infinite products of probability spaces. If $K_s = \sum_\alpha a_\alpha(s)e_\alpha$, we may define $\xi_s = \sum_\alpha a_\alpha(s)\eta_\alpha$ (since $\sum_\alpha |a_\alpha(s)|^2 < \infty$). It is easily verified that $\{\xi_s, s \in S\}$ is a Gaussian process with covariance K.

It is true that a very general theorem of Kolmogorov gives the existence of the Gaussian process also, from whence the reproducing kernel space $\mathscr{A}$ can be easily constructed. Indeed, if $\{\xi_s, s \in S\}$ is a process (defined on the probability space (Ω, Σ, P)) whose covariance is K, we may define $\mathscr{A}$ to be the space of all functions f_ξ in $\Phi(S, \mathbb{C})$ of the form $f_\xi(s) = E\{\xi \cdot \overline{\xi_s}\}$ for ξ in the Hilbert space $\mathscr{W}$ spanned by the ξ_s, $s \in S$ in $L^2(\Omega, \Sigma, P)$. Since $\xi \mapsto f_\xi$ is an injection, one obtains the desired Hilbertian structure in $\mathscr{A}$ from that in $\mathscr{W}$.

However, the more interesting applications of the Moore-Aronszajn theorem seem to arise from p.d. kernels $K: S \times S \to \mathbb{C}$ where a semigroup G acts on S in a way convenient for the analysis of K. For instance, let us suppose that G is a $*$-semigroup (i.e., there is a map in G, $g \mapsto g^*$ (called involution) such that $(g_1g_2)^* = g_2^*g_1^*$ and $(g^*)^* = g$) and let us denote the result of $g \in G$ acting on $s \in S$ by $g.s$. It is tempting to define a linear operator $T_g: M \to M$ (M = linear manifold spanned by K_s, $s \in S$, $K_s(t) = K(s, t)$) via the formula $T_gK_s = K_{g.s}$. For this to be successful, it is clear that we need to have that

$$\sum_{j=1}^n \alpha_j K_{s_j} = 0 \Rightarrow \sum_{j=1}^n \alpha_j K_{g\cdot s_j} = 0 \tag{2}$$

for all acceptable choices of α_j, s_j, g and n. A little experimentation shows that the simple condition

$$K(gs, t) = K(s, g^*t) \tag{3}$$

for all choices of g, s, t will suffice. Indeed, from (3), we have that

$$\begin{aligned} \langle K_s \mid \textstyle\sum_{j=1}^n \alpha_j K_{g\cdot s_j} \rangle &= \textstyle\sum_{j=1}^n \alpha_j \langle K_s \mid K_{g\cdot s_j} \rangle \\ &= \textstyle\sum_{j=1}^n \alpha_j K(g\cdot s_j, s) \\ &= \textstyle\sum_{j=1}^n \alpha_j K(s_j, g^*s) \\ &= \textstyle\sum_{j=1}^n \alpha_j \langle K_{g^*s} \mid K_{s_j} \rangle \\ &= \langle K_{g^*s} \mid \textstyle\sum_{j=1}^n \alpha_j K_{s_j} \rangle; \end{aligned}$$

this means that if $\sum_j \alpha_j K_{s_j} = 0$ then $\langle K_s \mid \sum_j \alpha_j K_{g\cdot s_j} \rangle = 0$ for all $s \in S$. Since the set $\{K_s : s \in S\}$ is total in $\mathscr{A}$ we conclude that $\sum_j \alpha_j K_{g \cdot s_j} = 0$ i.e. (2) holds. Thus, if (3) holds then the linear operators $T_g : M \to M$ are well-defined.

It is much more subtle to guess a simple condition on K which will guarantee the continuity of the operators T_g. It turns out that the following is enough:

$$K(gs, gs) \leqslant \gamma(g) \cdot K(s, s) \tag{4}$$

for all $g \in G$ and $s \in S$ where $\gamma(g)$ is a positive, finite constant depending on g. Condition (4) as well as (3) appear in print in Masani's 1978 paper (cf. §3) in a much more general setting which we shall consider later. But, even in the case of scalar-valued kernels K which we have considered up to now, it is not a trivial matter to show that (4) (along with (3)) implies the continuity of T_g. However, once this is established, it is quite easy to show that T_g can be extended to be a bounded linear operator from $\mathscr{A}$ to $\mathscr{A}$ (i.e., $T_g \in \mathscr{L}(\mathscr{A})$) and that $g \mapsto T_g$ is a *-homomorphism (representation) of G in $\mathscr{L}(\mathscr{A})$.

Suppose now that the representations of G in a Hilbert space are known or understood; then the structure of K satisfying (3) and (4) can be revealed with greater transparency. Let us consider a few important special cases. In each of them, the set S coincides with the *-semigroup G and the kernel K is of the form $K(s, t) = p(t^* s)$ where $p : S \to \mathbb{C}$ is a p.d. function on the *-semigroup S i.e.

$$\textstyle\sum_{j,k=1}^{n} p(s_k{}^* s_j) \alpha_j \bar{\alpha}_k \geqslant 0$$

for all choices of $s_j \in S$, $\alpha_j \in \mathbb{C}$, $n = 1, 2, \cdots$. It is clear in this situation that (3) is always satisfied.

Suppose first that $G = S$ is a group and $g^* = g^{-1}$. Then

$$K(gs, gs) = p(s^{-1} g^{-1} gs) = p(1) = K(s, s)$$

and (4) is trivially satisfied. Actually, in this case, the boundedness, indeed the isometric character of T_g is immediate and does not require any subtle reasoning. Further $T_{g^*} = T_{g^{-1}} = T_g{}^*$ implies that T_g is unitary and we get

$$\begin{aligned} p(s) = K(s, 1) &= \langle K_1 \mid K_s \rangle \\ &= \langle K_1 \mid T_s K_1 \rangle. \end{aligned}$$

This last is simply the Gelfand-Raikov theorem (for a discrete group G) stating that any p.d. function on G comes from a unitary representation (with a cyclic vector). Further, if G is abelian, we can deduce the Bochner theorem from the spectral representation of T_g, $g \in G$. We recall that the latter can be deduced from the Gelfand-Naimark theorem on commutative C^*-algebras. If we assume that $G = S$ is a topological group and p is continuous then we obtain easily the corresponding topological versions.

Suppose next that $G = S = \mathscr{A}$ is a C^*-algebra with a unit; actually, this discussion can be extended without any important changes to the case of an involutive Banach algebra with unit (or even only a bounded approximate unit). Let φ now be a linear functional on $\mathscr{A}$ to $\mathbb{C}$ which is positive (i.e.,

$\varphi(x^*x) \geq 0$ for $\forall\ x \in \mathscr{A}$). It is easily seen then that φ is p.d. (since $\sum_{j,k} \varphi(x^*_k x_j)\alpha_j \bar{\alpha}_k = \varphi(y^*y) \geq 0$ with $y = \sum_j \alpha_j x_j$) so that we can again introduce a p.d. kernel $K:\mathscr{A} \times \mathscr{A} \to \mathbb{C}$ with $K(x, y) = \varphi(y^*x)$. We have already remarked that K then satisfies (3); the fact that K fulfills (4) as well follows from a known (easy) inequality for positive linear functionals φ viz.

$$|\varphi(a^*xa)| \leq \varphi(a^*a) \cdot \|x\|$$

so that

$$\begin{aligned} K(xa, xa) &= \varphi(a^*x^*xa) \\ &\leq \|x^*x\| \cdot \varphi(a^*a) \\ &= \|x\|^2 \cdot K(a, a). \end{aligned}$$

Hence we deduce the existence of the operators $T_x \in \mathscr{L}(\mathscr{A})$ with $x \mapsto T_x$ a *-homomorphism between $\mathscr{A}$ and $\mathscr{L}(\mathscr{A})$. This gives a version of the so-called GNS-construction.

Finally, take G to be an *abelian* *-semigroup; note, that any abelian semigroup has at least one obvious involution viz. $g = g^*$. If again $S = G$ and $K(s, t) = p(t^*s)$ with $p:S \to \mathbb{C}$ a bounded function as defined above, then again the foregoing theory applies and gives rise to a commuting family of normal operators (indeed contractions); this, via spectral theory, leads to a solution of a general moment problem (cf. the paper of Lindahl and Maserick as well as that of Berg, Christiansen and Ressel referred to in section §3; also, cf. Masani (1981) cited there).

§2. Operator valued kernels

Let S be an abstract set as before and let, for each $s \in S$, E_s be a vector space and F_s a linear subspace of E_s^*, the space of all semi-linear functionals on E_s i.e. $\varphi \in E_s^*$ if $\varphi:E_s \to \mathbb{C}$ with $\varphi(x + y) = \varphi(x) + \varphi(y)$, $\varphi(\lambda x) = \bar{\lambda}\varphi(x)$ for $\lambda \in \mathbb{C}$. We note $\varphi(x)$ by $\langle x, \varphi \rangle$. For each $(s, t) \in S \times S$, let us be given an operator $K(s, t) \in \mathrm{Hom}(E_s, F_t)$ with the property that

$$\sum_{j,k} \langle x_k, K(s_j, s_k)x_j \rangle \geq 0 \tag{1}$$

for all choices of $s_1, \cdots, s_n$ in S and of x_j in E_{s_j}, $1 \leq j \leq n$, $n = 1, 2, \cdots$. Let us call such an object an operator valued p.d. kernel. We describe in this section the structure of such kernels in a way which generalizes exactly the main structure theorem for the scalar valued kernels. Roughly speaking, we show that there is a Hilbert space $(\mathscr{H}, \langle \cdot \mid \cdot \rangle)$ whose elements are sections of the "vector bundle" $\Pi_{s \in S} E_s^*$ and there are linear operators $T_s \in \mathrm{Hom}(E_s, \mathscr{H})$, $s \in S$, such that $K(s, t)$ factorizes as $K(s, t) = T_t' T_s$ where T_t', the adjoint of T_t, is considered as an element of $\mathrm{Hom}(\mathscr{H}, E_t^*)$. A little thought will show that this factorization reduces to the main structure theorem for scalar valued p.d. kernels in case $E_s = F_t = \mathbb{C}$ for all $s, t \in S$.

Further, if the E_s, $s \in S$, are locally convex vector spaces then the $T_s \in \mathscr{L}(E_s, \mathscr{H})$, the set of all continuous linear operators from E_s to $\mathscr{H}$, provided that the E_s satisfy some natural conditions (e.g. it suffices that the E_s be barrelled). We proceed to state these facts more formally.

Let E_s, F_t, K be as described above. We define first $\bar{F}_t$, a sequential enlargement of F_t, as follows:

$$\bar{F}_t = \{\varphi \in E_t^*: \exists \varphi_n \in F_t,\ n = 1, 2, \cdots, \text{ with } \lim_n \langle x, \varphi_n \rangle = \langle x, \varphi \rangle \ \forall\ x \in E\}.$$

Now let $K_{s,x}(t) = K(s, t)x$ with $x \in E_s$. For fixed s and fixed $x \in E_s$, $K_{s,x} \in \Pi_{\alpha \in S} F_\alpha$. Let $\mathscr{H}_0$ be the liner space spanned by the $K_{s,x}$, $s \in S$, $x \in E_s$. If $f = \sum_j a_j K_{s_j,x_j}$, $g = \sum_j b_j K_{s_j,x_j}$ are two elements of $\mathscr{H}_0$, we define

$$\langle f \mid g \rangle = \textstyle\sum_{j,k} \bar{a}_j b_k \langle x_k, K(s_j, s_k) x_j \rangle.$$

It turns out that this is a well-defined Hermitian, strictly positive definite, sequilinear from on $\mathscr{H}_0$. Let us now define $T_s \in \mathrm{Hom}(E_s, \mathscr{H}_0)$ by $T_s x = K_{s,x}$.

With the above notation, we can now state the following theorem:

THEOREM (1).
(*A*) *The pre-Hilbert space* $(\mathscr{H}_0, \langle \cdot \mid \cdot \rangle)$ *can be completed to a Hilbert space* $(\mathscr{H}, \langle \cdot \mid \cdot \rangle)$ *where* $\mathscr{H}$ *is a linear subspace of* $\Pi_{s \in S} \bar{F}_s$. *Further*

$$\langle K_{s,x} \mid f \rangle = \langle x, f(s) \rangle$$

for all $f \in \mathscr{H}$, $s \in S$, $x \in E_s$.
(*B*) *The operators* T_s *defined above have the following property*: T_s' *maps* $\mathscr{H}_0$ *into* F_s *and* T_s' *can also be considered as elements of* $\mathrm{Hom}(\mathscr{H}, \bar{F}_s)$. *Further,* $T_s' f = f(s)$ *and* $K(s, t) = T_t' T_s$.

Remark. The factorization above is unique upto unitary isomorphism in the sense that if $\tilde{\mathscr{H}}$ is any other Hilbert space and $\tilde{T}_s \in \mathrm{Hom}(E_s, \tilde{\mathscr{H}})$ with $K(s, t) = \tilde{T}_t' \tilde{T}_s$ and if $\tilde{\mathscr{H}}$ is minimal in the sense that $\tilde{\mathscr{H}}$ is generated by the elements $\tilde{T}_s x$, $s \in S$, $x \in E_s$ then there is a unitary isomorphism V: $\tilde{\mathscr{H}} \rightarrow \mathscr{H}$ such that $V\tilde{T}_s = T_s$. Note also that conversely if $\tilde{T}_s \in \mathrm{Hom}(E_s, \tilde{\mathscr{H}})$, $s \in S$, are given (where $\tilde{\mathscr{H}}$ is a Hilbert or a pre-Hilbert space) then $K(s, t) = \tilde{T}_t' \tilde{T}_s$ defines a p.d. kernel with values in $\mathrm{Hom}(E_s, F_t)$ where F_t is a suitable linear subspace of E_t^*.

Suppose now that the E_s, $s \in S$, are separated locally convex spaces and F_s are linear subspaces of E_s', the space of continuous semi-linear forms on E_s. For each pair (s, t) we suppose as before that $K(s, t) \in \mathrm{Hom}(E_s, F_t)$. Although we are making no continuity assumptions here, it turns out that if we assume that K is a p.d. kernel then each $K(s, t)$ is automatically continuous if E_s is given its $\sigma(E_s, E_s')$-topology and F_t the $\sigma(F_t, E_t)$-topology. This is an immediate consequence of the fact that the p.d. property implies the equality

$$\langle y, K(s, t)x \rangle = \langle x, K(t, s)y \rangle$$

for all s, t in S and $x \in E_s$, $y \in E$. Standard theorems (e.g. in Boubaki's "Espaces vectoriels topologiques") on transposed maps then give us that $K(s,$

t) is even continuous in the original topology of E_s and the strong topology on F_t that the original topology of E_s coincides with that of its Mackey topology; this happens if, for example, E_s is barrelled or bornological. We can now state the following in case the E_s, $s \in S$ are locally convex and F_t, $t \in S$ are linear subspaces of E_t'.

THEOREM 2.
(A) *There exists a Hilbert space* $\mathscr{H}$ *and operators* $T_s \in \mathscr{L}(E_s, \mathscr{H})$ *so that* $K(s, t) = T_t' T_s$ *for all s, t in S if and only if the maps* $m_s: E_s \to \mathbb{C}$ *defined by* $m_s(x) = \langle x, K(s, s)x \rangle$ *are continuous at* $x = 0$ *(and hence for all x) for all* $x \in S$.
(B) *If* E_x, $s \in S$ *are barrelled then there exists a Hilbert space* $\mathscr{H}$ *and* $T_s \in \mathscr{L}(E_s, \mathscr{H})$, $s \in S$, *such that* $K(s, t) = T_t' T_s$ *for all s, t in S. The same holds if* E_s, $s \in S$, *are bornological and further* $\bar{F}_s \subset E_s'$, $s \in S$. *In each case,* $K(s, t)$ *is continuous from* E_s *to* F_t *where* E_s *has the original topology and* F_t *the strong topology.*

Remark. Conversely, if $K(s, t) = T_t' T_s$ with $T_s \in \mathscr{L}(E_s, \mathscr{H})$ (for some Hilbert space $\mathscr{H}$) then $\{K(s, t), (s, t) \in S \times S\}$, gives a p.d. kernel. Here and in Theorem 2, we can choose $\mathscr{H}$ to be of the special form described in Theorem 1.

§3. Remarks and references

P. Masani's paper ("Dilations as propagators of Hilbertian varieties" SIAM, (1978), p. 414–456) contains exact references of papers we shall refer to briefly in the sequel. My paper ("Factorization of positive definite operator-valued kernels" forthcoming in a special volume entitled "Prediction theory and harmonic analysis" ed. V. Mandrekar and H. Salehi, North-Holland, Amsterdam (1982–83)) contains detailed proofs of the factorization theorems in §2 in case $E_s = E$ and $F_s \subset E^*$ for all $s \in S$. The fact that the work can be generalized as here, was pointed out to me by Dr. G. Vincent-Smith; indeed the proofs of my cited paper need no changes whatsoever. Further historical remarks can be found in my paper and in that of Masani. In particular, in my paper, the related studies of Górniak (Lecture Notes in Mathematics, Springer Verlag, Vol. **656**, 1978), Górniak and Weron (Bull. Acad. Pol. Sci., 1976; Studia Mathematica, Warsaw, 1980), Weron (Lecture Notes in Mathematics, Springer Verlag, Vol. **472**, 1975) are referred to. Also, an unpublished report of Pedrick (exact reference in Masani's SIAM paper) treats factorization theorems in the case of $E_s = E$, a locally convex space. Masani refers also to the classical papers of Aronszajn and Kolmogorov and also to works of Allen, Narcowich and Williams. More importantly, Masani sets up a programme of obtaining various dilation theorems via factorization theorems of the type in §2. In this connexion, the pioneering work of Nagy (cf. Masani (1978) for reference) must be mentioned. Nagy's remark that the kernel $K(m, n) = T^{(m-n)}$ if $(m - n) \geqslant 0$ and $K(m, n) = (T^*)^{(m-n)}$ if $(m - n) < 0$ with$T \in \mathscr{L}(\mathscr{H}, \mathscr{H})$, $\mathscr{H}$ a Hilbert space, is a p.d. kernel on the abelian group $\mathbb{Z}$ if and only if $\| T \| \leqslant 1$ is the starting point of one of his theories of dilation. To him is also due the fruitful idea of working with *-semigroups. Masani reworks all of this in a wider context of $K(s, t) \in \mathscr{L}(\mathscr{W}, \mathscr{W}')$, $\mathscr{W}$ a Banach space and derives a large variety of theorems as consequences of his general work. In this connexion, we mention his recent survey paper (Masani: "An outline of the spectral theory of propagators" in Functional analysis and approximation, Birkhäuser, 1981) where the work of Lindahl and Maserick (Duke Math. J. **38** (1971), p. 771–782) and that of Berg, Christiansen and Ressel (Math. Ann. **223** (1976), p. 253–272) on general moment problems are fitted into the general framework of what Masani calls the theory of propagators. We do not enter into the latter here; perhaps, an application of the very general factorization theorem of the present paper in the propagator framework of Masani would lead to interesting new results.

ECOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
CH-1015 LAUSANNE, SWITZERLAND