Positive Definite Functions on Hilbert Space

B. J. C. Baxter
Department of Mathematics and Statistics,
Birkbeck College, University of London,
Malet Street, London WC1E 7HX,
England
b.baxter@bbk.ac.uk

Let H be any real, separable, infinite dimensional Hilbert space; thus H is isomorphic to $\ell^2(\mathbb{Z})$. A function $f : [0,\infty) \to \mathbb{R}$ for which the quadratic form

$$\sum_{j=1}^{n} \sum_{k=1}^{n} a_j a_k f(\|x_j - x_k\|^2)$$

is always non-negative, for *any* positive integer n and all points $x_1, \ldots, x_n$ in H is said to be *positive definite on Hilbert space*. In Schoenberg (1938), it was shown that a function is positive definite on Hilbert space if and only if it is completely monotonic, and this characterization is of central importance in the theory of radial basis functions and learning theory. In this paper, we present a short geometric proof of this beautiful fact.

1. Introduction

Let H be any real, separable, infinite dimensional Hilbert space and let $f : [0,\infty) \to \mathbb{R}$ be a function for which the quadratic form

$$\sum_{j=1}^{n} \sum_{k=1}^{n} a_j a_k f(\|x_j - x_k\|^2)$$

is non-negative, for *any* positive integer n and any points $x_1, \ldots, x_n \in H$, where $\|\cdot\|$ denotes the Hilbert space norm. We shall say that the function f is *positive definite on Hilbert space*. Such functions were characterized by I. J. Schoenberg in a remarkable series of seminal papers collected in Schoenberg (1988), wherein he was able to show that the class of functions positive definite on Hilbert space is precisely the set of *completely monotonic* functions; we recall that $f : (0,\infty) \to \mathbb{R}$ is completely monotonic provided that

$$(-1)^k f^{(k)}(x) \geq 0, \qquad \text{for every } k = 0, 1, \ldots, \text{ and for } 0 < x < \infty.$$

The theory of positive definite functions on Hilbert space has had great impact in many fields; see Micchelli (1986) and the survey of metric geometry in Schoenberg (1988). In particular, if $f : [0, \infty) \to \mathbb{R}$ is a non-constant completely monotonic function, then the $n \times n$ symmetric matrix

$$\Big(f(\|x_j - x_k\|^2)\Big)_{j,k=1}^{n}$$

is positive definite for any set of distinct points $x_1, \ldots, x_n \in \mathbb{R}^d$, where the dimension d can be any positive integer. Thus non-constant completely monotonic functions can be used for radial basis function interpolation in any dimension d; see, for instance Buhmann (2003), Powell (1992) or Light (2000). The theorem is also of crucial important to the emerging discipline of learning theory Poggio (2003).

Unfortunately, Schoenberg's proof was rather complicated. Specifically, he first derived an integral representation for positive definite functions on *finite dimensional* Hilbert space (that is, copies of $\mathbb{R}^d$). A careful limiting argument, similar in flavour to the Central Limit Theorem, was then used to prove that positive definite functions on Hilbert space are Laplace transforms of finite positive Borel measures defined on the half-line $[0, \infty)$ (a lucid exposition of this may be found in Chapter 41 of Donoghue (1969)). Such Laplace transforms are known to be the completely monotonic functions, by a celebrated theorem of Bernstein (for which a classic reference is Widder (1946), but a beautiful proof recently appeared in Mattner (1993)).

In this paper, we present a short direct proof that postive definite functions on Hilbert space are completely monotonic. Our argument is based on a geometric construction first used in Baxter (1991).

2. Positive Definite Functions on Hilbert Space

In this note, H can be any Hilbert space isomorphic to $\ell^2(\mathbb{Z})$.

Definition 2.1. A function $f\colon [0, \infty) \to \mathbb{R}$ will be called *positive definite on Hilbert space* (HPD) if the matrix

$$\Big(f(\|x_j - x_k\|^2)\Big)_{j,k=1}^{n} \tag{2.1}$$

is non-negative definite for every positive integer n and any points $x_1, \ldots, x_n \in H$. We shall call any matrix of the form (2.1) a *distance matrix*.

The classical theory of positive definite functions provides the well-known inequality

$$|f(t)| \le f(0),$$

but a stronger bound holds for HPD functions.

Proposition 2.1. Every HPD function is non-negative.

Proof. Let $u_1, u_2, \ldots$ be any orthonormal sequence in H and choose any nonzero real number λ. Using the relation $f(\|\lambda u_j - \lambda u_k\|^2) = f(2\lambda^2)$, for $j \neq k$, and defining

$$A_{jk} = f(\|\lambda u_j - \lambda u_k\|^2), \qquad 1 \le j, k \le n,$$

we find

$$0 \le \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}^T A \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = nf(0) + \frac{1}{2}n(n-1)f(2\lambda^2). \tag{2.2}$$

Hence, for $n \ge 2$, we have

$$f(2\lambda^2) + \frac{2f(0)}{n-1} \ge 0. \tag{2.3}$$

Letting $n \to \infty$, we conclude $f(2\lambda^2) \ge 0$ for every nonzero real number λ. $\square$

In fact, HPD functions satisfy a *much* stronger property. We recall that, for any $h > 0$, the *forward difference operator* Δ_h is defined by the equation

$$\Delta_h f(t) := f(t+h) - f(t), \qquad t \ge 0. \tag{2.4}$$

Of course, we define higher order differences in the usual way: $\Delta_h^m f := \Delta_h(\Delta_h^{m-1} f)$.

Theorem 2.2. If $f\colon [0,\infty) \to \mathbb{R}$ is HPD, then $(-1)^m \Delta_h^m f$ is also HPD, for every $h > 0$ and positive integer m.

Proof. It is sufficient to prove that $-\Delta_h f$ is HPD. To this end, let $x_1, \ldots, x_n$ be any vectors in H, and choose any unit vector $y \in H$ that is orthogonal to $x_1, \ldots, x_n$. We now let A denote the $2n \times 2n$ distance matrix generated by the points

$$x_1, \ldots, x_n, x_1 + h^{1/2}y, \ldots, x_n + h^{1/2}y.$$

It is easily checked that

$$A = \begin{pmatrix} B & C \\ C & B \end{pmatrix},$$

where the $n \times n$ matrices B and C are given by the equations

$$B_{jk} = f(\|x_j - x_k\|^2), \qquad 1 \le j, k \le n,$$

and

$$C_{jk} = f(\|x_j - (x_k + h^{1/2}y)\|^2) = f(\|x_j - x_k\|^2 + h), \qquad 1 \le j, k \le n.$$

Now, given any vector $a \in \mathbb{R}^n$, and since f is HPD, we have

$$0 \le \begin{pmatrix} a \\ -a \end{pmatrix}^T A \begin{pmatrix} a \\ -a \end{pmatrix} = 2a^T(B - C)a \equiv 2a^T Da,$$

where

$$D_{jk} = -\Delta_h f(\|x_j - x_k\|^2), \qquad 1 \le j, k \le n.$$

Hence $-\Delta_h f$ is HPD. []

A standard limiting argument then allows us to deduce that f is completely monotonic. The slick construction of Theorem 2.2 was, in fact, distilled from a more elaborate geometric artifice to be found in Section 3 of Baxter (1991).

We first choose any points $x_1, \ldots, x_n \in H$ and let $a_1, \ldots, a_n$ be real numbers. We now pick a positive integer m and any set of orthonormal vectors $u_1, \ldots, u_m \in H$ that are orthogonal to $x_1, \ldots, x_n$. The vertices of the cube generated by the closed convex hull of $u_1, \ldots, u_m$ will be denoted $y_1, y_2, \ldots, y_{2^m}$, their order being irrelevant. We shall also write

$$y_k = \sum_{\ell=1}^{m} y_k(\ell) u_\ell, \qquad 1 \le k \le 2^m.$$

Let us now introduce the new configuration of $2^m n$ points

$$\left\{ x_j + h^{1/2} y_k : 1 \le j \le n, \quad 1 \le k \le 2^m \right\}. \tag{2.5}$$

Thus each point x_j now has an associated *cubic cloud* of points, The coefficient associated with the point $x_j + h^{1/2} y_j$ will be

$$\alpha(j, k) = a_j (-1)^{y_k(1) + \cdots + y_k(m)}, \qquad 1 \le j \le n, \quad 1 \le k \le 2^m. \tag{2.6}$$

Thus, since f is HPD, we have the inequality

$$0 \le Q := \sum_{j_1=1}^{n} \sum_{k_1=1}^{2^m} \sum_{j_2=1}^{n} \sum_{k_2=1}^{2^m} \alpha(j_1, k_1) \alpha(j_2, k_2) A(j_1, k_1; j_2, k_2), \tag{2.7}$$

where the $2^m n \times 2^m n$ distance matrix A is given by

$$A(j_1, k_1; j_2, k_2) = f(\|(x_{j_1} + h^{1/2} y_{k_1}) - (x_{j_2} + h^{1/2} y_{k_2})\|^2), \tag{2.8}$$

for $1 \le j_1, j_2 \le n$ and $1 \le k_1, k_2 \le 2^m$. However, by construction,

$$\|(x_{j_1} + h^{1/2} y_{k_1}) - (x_{j_2} + h^{1/2} y_{k_2})\|^2 = \|x_{j_1} - x_{j_2}\|^2 + h\|y_{k_1} - y_{k_2}\|^2, \tag{2.9}$$

and the squared distances $\{\|y_{k_1} - y_{k_2}\|^2 : 1 \le k_1, k_2 \le n\}$ take the values $0, 1, 2, \ldots, m$, the distance ℓ occuring with frequency $\binom{m}{\ell}$. Further, if $\|y_{k_1} - y_{k_2}\|^2 = \ell$, then

$$(-1)^{y_{k_1}(1) + \cdots + y_{k_1}(m)} (-1)^{y_{k_2}(1) + \cdots + y_{k_2}(m)} = (-1)^\ell,$$

so that

$$\begin{aligned} 0 \le Q &= \sum_{j_1=1}^{n}\sum_{j_2=1}^{n} a_{j_1}a_{j_2}\sum_{\ell=0}^{m}(-1)^{\ell}\binom{m}{\ell} f(\|x_{j_1}-x_{j_2}\|^2+h\ell) \\ &= \sum_{j_1=1}^{n}\sum_{j_2=1}^{n} a_{j_1}a_{j_2}(-1)^m\Delta_h^m f(\|x_{j-1}-x_{j_2}\|^2). \qquad (2.10) \end{aligned}$$

Finally, we recall the elementary formula (see, for example, Davis (1975))

$$(-1)^m\Delta_h^m f(t) = \sum_{\ell=0}^{m}(-1)^{\ell}\binom{m}{\ell} f(t+h\ell).$$

Thus we have shown that the function

$$(0,\infty) \ni t \mapsto (-1)^m\Delta_h^m f(t)$$

is HPD, for every $h > 0$. In fact, this implies that f is, in fact, infinitely differentiable:

Theorem 2.3. Let $f : [0,\infty) \to \mathbb{R}$ be HPD. Then f is completely monotonic.

Proof. This is an immediate consequence of the single theorem proved in Boas and Widder (1940). □

REFERENCES

B. J. C. Baxter (1991), "Conditionally positive functions and p-norm distance matrices", *Constr. Approx.* **7**: 427–440.

R. P. Boas and D. V. Widder (1940), "Functions with positive differences", Duke Mathematical Journal **7**:496–503.

M. D. Buhmann (2003), *Radial Basis Functions: Theory and Implementations*, Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press.

P. J. Davis (1975), *Interpolation and Approximation*, Dover.

W. F. Donoghue, Jr., *Distributions and Fourier Transforms*, Academic Press, New York.

W. A. Light and E. W. Cheney (2000), *A Course in Approximation Theory*, Brooks/Cole.

L. Mattner (1993), "Bernstein's theorem, inversion formula of Post and Widder, and the uniqueness theorem for Laplace transforms", *Expositiones Mathematicae* **11**: 137–140.

C. A. Micchelli (1986), "Interpolation of scattered data: distance matrices and conditionally positive functions", *Constr. Approx.* **2**: 11-22.

T. Poggio and S. Smale (2003), "The Mathematics of Learning: Dealing with Data", Notices of the AMS, May 2003: 537–544.

M. J. D. Powell (1992), "The Theory of Radial Basis Functions in 1990", in *Advances in Numerical Analysis*, vol. II, ed. W. A. Light, Oxford University Press, pp. 105–210.

I. J. Schoenberg (1938), "Metric spaces and completely monotone functions", *Ann. of Math.* **39**: 811-841.

I. J. Schoenberg (1988), *Selected Papers*, Volume 1, ed. C. deBoor, Birkhaüser.

D. V. Widder (1946), *The Laplace Transform*, Princeton University Press.