Positive definite functions, completely monotone functions, the Bernstein-Widder theorem, and Schoenberg's theorem

Jordan Bell

June 26, 2015

1 Linear operators

For a complex Hilbert space H let $\mathscr{L}(H)$ be the bounded linear operators $H \to H$. It is a fact that $A \in \mathscr{L}(H)$ is self-adjoint if and only if $\langle Ah, h \rangle \in \mathbb{R}$ for all $h \in H$.[1] For a bounded self-adjoint operator A it is a fact that[2]

$$\|A\| = \sup_{\|h\|=1} |\langle Ah, h \rangle|.$$

$A \in \mathscr{L}(H)$ is called **positive** if it is self-adjoint and

$$\langle Ah, h \rangle \geq 0, \qquad h \in H;$$

because we have taken H to be a complex Hilbert space, for A to be positive it suffices that the inequality is satisfied.

For $A, B \in \mathscr{L}(\mathbb{C}^n)$, we define their **Hadamard product** $A * B \in \mathscr{L}(H)$ by

$$(A * B)e_i = \sum_{j=1}^{n} \langle Ae_i, e_j \rangle \langle Be_i, e_j \rangle e_j.$$

So,

$$\langle (A * B)e_i, e_j \rangle = \langle Ae_i, e_j \rangle \langle Be_i, e_j \rangle .$$

The **Schur product theorem** states that if $A, B \in \mathscr{L}(\mathbb{C}^n)$ are positive then their Hadamard product $A * B$ is positive.[3]

[1] John B. Conway, *A Course in Functional Analysis*, second ed., p. 33, Proposition 2.12.
[2] John B. Conway, *A Course in Functional Analysis*, second ed., p. 34, Proposition 2.13.
[3] Ward Cheney and Will Light, *A Course in Approximation Theory*, p. 81, chapter 12.

2 Positive definite functions

Let X be a real or complex linear space, let $f : X \to \mathbb{C}$ be a function, and for $x_1, \ldots, x_n \in X$, define $F_{f;x_1,\ldots,x_n} \in \mathscr{L}(\mathbb{C}^n)$ by

$$F_{f;x_1,\ldots,x_n} e_i = \sum_{j=1}^{n} f(x_i - x_j) e_j,$$

where $\{e_1, \ldots, e_n\}$ is the standard basis for $\mathbb{C}^n$. Thus for $u = \sum_{i=1}^{n} u_i e_i \in \mathbb{C}^n$,

$$\begin{aligned}\langle F_{f;x_1,\ldots,x_n} u, u\rangle &= \left\langle \sum_{i=1}^{n} u_i \sum_{j=1}^{n} f(x_i - x_j) e_j, \sum_{k=1}^{n} u_k e_k \right\rangle \\ &= \sum_{i=1}^{n} u_i \sum_{j=1}^{n} f(x_i - x_j) \left\langle e_j, \sum_{k=1}^{n} u_k e_k \right\rangle \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} u_i \overline{u_j} f(x_i - x_j).\end{aligned}$$

We call f **positive definite** if for all $x_1, \ldots, x_n \in X$, $F_{f;x_1,\ldots,x_n}$ is a positive operator, i.e. for $u \in \mathbb{C}^n$,

$$\langle F_{f;x_1,\ldots,x_n} u, u\rangle \geq 0.$$

We call f **strictly positive definite** for all distinct $x_1, \ldots, x_n \in X$ and nonzero $u \in \mathbb{C}^n$,

$$\langle F_{f;x_1,\ldots,x_n} u, u\rangle > 0.$$

3 Completely monotone functions

A function $f : [0, \infty) \to \mathbb{R}$ is called **completely monotone** if

1. $f \in C[0, \infty)$
2. $f \in C^{\infty}(0, \infty)$
3. $(-1)^k f^{(k)}(x) \geq 0$ for $k \geq 0$ and $x \in (0, \infty)$

Because a completely monotone function is continuous, $f(x)$ tends to $f(0)$ as $x \downarrow 0$. Because a completely monotone function is nonincreasing and convex, $f(x)$ has a limit, which we call $f(\infty)$, as $x \uparrow \infty$.

The **Bernstein-Widder theorem** states that a function f satisfying $f(0) = 1$ is completely monotone if and only if it is the Laplace transform of a Borel probability measure on $[0, \infty)$.[4]

[4] Peter D. Lax, *Functional Analysis*, p. 138, chapter 14, Theorem 3; `http://djalil.chafai.net/blog/2013/03/23/the-bernstein-theorem-on-completely-monotone-functions/`

Theorem 1 (Bernstein-Widder theorem)**.** A function $f : [0, \infty) \to \mathbb{R}$ satisfies $f(0) = 1$ and is completely monotone if and only if there is a Borel probability measure μ on $[0, \infty)$ such that

$$f(x) = \int_0^\infty e^{-xt} d\mu(t), \qquad x \in [0, \infty).$$

Proof. If f is the Laplace transform of some probability measure μ on $\mathscr{B}_{[0,\infty)}$, then using the dominated convergence theorem yields that f is continuous and by induction that $f \in C^\infty(0, \infty)$. For $k \geq 0$ and for $x \in (0, \infty)$,

$$f^{(k)}(x) = \int_0^\infty (-t)^k e^{-xt} d\mu(t),$$

as $\int_0^\infty t^k e^{-xt} d\mu(t) \geq 0$ so $(-1)^k f^{(k)}(x) \geq 0$. Hence f is completely monotone, and $f(0) = \int_0^\infty d\mu(t) = 1$.

If f satisfies $f(0) = 1$ and is completely monotone, then for each $k \geq 0$, the function $(-1)^k f^{(k)} : (0, \infty) \to \mathbb{R}$ is nonnegative and is nonincreasing, so for $k \geq 1$ and $t \in (0, \infty)$, using that $(-1)^k f^{(k)}$ is nondecreasing and that $(-1)^{k-1} f^{(k-1)}$ is nonnegative,

$$\begin{aligned}
(-1)^k f^{(k)}(t) &\leq \frac{2}{t} \int_{t/2}^t (-1)^k f^{(k)}(u) du \\
&= \frac{2}{t} (-1)^k \left(f^{(k-1)}(t) - f^{(k-1)}(t/2) \right) \\
&\leq \frac{2}{t} (-1)^{k-1} f^{(k-1)}(t/2).
\end{aligned}$$

Doing induction, for any $k \geq 1$,

$$\begin{aligned}
(-1)^k f^{(k)}(t) &\leq \prod_{j=1}^{k-1} \left(\frac{2^j}{t} \right) \cdot f' \left(\frac{t}{2^{k-1}} \right) \\
&\leq \prod_{j=1}^{k-1} \left(\frac{2^j}{t} \right) \cdot \frac{2^k}{t} \left(f \left(\frac{t}{2^{k-1}} \right) - f \left(\frac{t}{2^k} \right) \right) \\
&= t^{-k} 2^{k(k-1)/2} \left(f \left(\frac{t}{2^{k-1}} \right) - f \left(\frac{t}{2^k} \right) \right).
\end{aligned}$$

Because $f(x) \to f(0)$ as $x \downarrow 0$,

$$f \left(\frac{t}{2^{k-1}} \right) - f \left(\frac{t}{2^k} \right) \to 0, \qquad t \downarrow 0,$$

and because $f(x) \to f(\infty)$ as $x \uparrow \infty$,

$$f \left(\frac{t}{2^{k-1}} \right) - f \left(\frac{t}{2^k} \right) \to 0, \qquad t \uparrow \infty.$$

Hence for each $k \geq 1$,

$$|f(t)| = o_k(t^{-k}), \qquad t \downarrow 0, \tag{1}$$

and

$$|f(t)| = o_k(t^{-k}), \qquad t \uparrow \infty. \tag{2}$$

Furthermore, for any $x \in (0, \infty)$, $f^{(k)}(t) \to f^{(k)}(x)$ as $t \to x$, so it is immediate that

$$(t-x)^k f^{(k)}(t) \to 0, \qquad t \to x. \tag{3}$$

For $x \geq 0$ and $k \geq 1$, integrating by parts, using (2) and (1) or (3) respectively as $x = 0$ or $x > 0$,

$$\begin{aligned} f(x) - f(\infty) &= -\int_x^\infty f'(t)dt \\ &= -(t-x)f'(t)\Big|_x^\infty + \int_x^\infty f''(t)(t-x)dt \\ &= \int_x^\infty f''(t)(t-x)dt \\ &= \frac{(t-x)^2}{2} f''(t)\Big|_x^\infty - \int_x^\infty f'''(t)\frac{(t-x)^2}{2}dt \\ &= -\int_x^\infty f'''(t)\frac{(t-x)^2}{2}dt \\ &= (-1)^k \int_x^\infty f^{(k)}(t)\frac{(t-x)^{k-1}}{(k-1)!}dt. \end{aligned}$$

Hence for $x \geq 0$ and $n \geq 0$,

$$f(x) - f(\infty) = \frac{(-1)^{n+1}}{n!}\int_x^\infty f^{(n+1)}(t)(t-x)^n dt.$$

Define

$$\phi_n(y) = (1 - y/n)^n 1_{[0,n]}(y).$$

For $n \geq 1$, by change of variables,

$$\begin{aligned} f(x) - f(\infty) &= \frac{(-1)^{n+1}}{n!}\int_{x/n}^\infty f^{(n+1)}(nu)(nu-x)^n n du \\ &= \frac{(-1)^{n+1}}{(n-1)!}\int_{x/n}^\infty f^{(n+1)}(nu)(nu)^n \left(1 - \frac{x}{nu}\right)^n du \\ &= \frac{(-1)^{n+1}}{(n-1)!}\int_0^\infty (nu)^n \phi_n(x/u) f^{(n+1)}(nu)du \\ &= \frac{(-1)^{n+1}}{(n-1)!}\int_0^\infty (n/t)^n \phi_n(xt) f^{(n+1)}(n/t)t^{-2}dt. \end{aligned}$$

For $t \geq 0$, define

$$s_n(t) = \frac{(-1)^{n+1}}{(n-1)!}\int_{1/t}^\infty (nu)^n f^{(n+1)}(nu)du,$$

where $s_n(0) = 0$, and for $t < 0$ let $s_n(t) = 0$. s_n is continuous and because f is completely monotone, s_n is nondecreasing, so there is a unique positive measure σ_n on $\mathscr{B}_{\mathbb{R}}$ such that[5]

$$\sigma_n((a, b]) = s_n(b) - s_n(a), \qquad a \leq b.$$

On the other hand, s_n is absolutely continuous, so σ_n is absolutely continuous with respect to Lebesgue measure λ_1, and for λ_1-almost all $t \in \mathbb{R}$,[6]

$$\frac{d\sigma_n}{d\lambda_1}(t) = s_n'(t).$$

Now for $t > 0$, by the fundamental theorem of calculus and the chain rule,

$$s_n'(t) = \frac{(-1)^{n+1}}{(n-1)!}(n/t)^n f^{(n+1)}(n/t) \cdot t^{-2},$$

and therefore

$$\begin{aligned} f(x) - f(\infty) &= \int_0^\infty \phi_n(xt) s_n'(t) d\lambda_1(t) \\ &= \int_0^\infty \phi_n(xt) \frac{d\sigma_n}{d\lambda_1}(t) d\lambda_1(t) \\ &= \int_0^\infty \phi_n(xt) d\sigma_n(t). \end{aligned}$$

The **total variation** of σ_n is equal to the total variation of s_n, and because s_n is nondecreasing,

$$\|\sigma_n\| = \int_0^\infty |s_n'(t)| dt = \int_0^\infty s_n'(t) dt = s_n(\infty) - s_n(0) = s_n(\infty),$$

which is

$$\|\sigma_n\| = \frac{(-1)^{n+1}}{(n-1)!} \int_0^\infty (nu)^n f^{(n+1)}(nu) du = f(0) - f(\infty),$$

showing that $\{\sigma_n : n \geq 1\}$ is bounded for the total variation norm. We claim that $\{\sigma_n : n \geq 1\}$ is **tight**: for each $\epsilon > 0$ there is a compact subset K_ϵ of $\mathbb{R}$ such that $\sigma_n(K_\epsilon^c) < \epsilon$ for all n. Taking this for granted, **Prokhorov's theorem**[7] states that there is a subsequence σ_{k_n} of σ_n that converges **narrowly** to some positive measure σ on $\mathscr{B}_{\mathbb{R}}$. Finally, the sequence $t \mapsto \phi_n(xt)$ tends in $C_b([0, \infty))$ to $t \mapsto e^{-xt}$, and it thus follows that[8]

$$\int_0^\infty \phi_n(xt) d\sigma_n(t) \to \int_0^\infty e^{-xt} d\sigma(t),$$

[5] Charalambos D. Aliprantis and Kim C. Border, *Infinite Dimensional Analysis: A Hitchhiker's Guide*, third ed., p. 393, Theorem 10.48.

[6] H. L. Royden, *Real Analysis*, third ed., p. 303, Exercise 16.

[7] V. I. Bogachev, *Measure Theory*, volume II, p. 202, Theorem 8.6.2.

[8] cf. Charalambos D. Aliprantis and Kim C. Border, *Infinite Dimensional Analysis: A Hitchhiker's Guide*, third ed., p. 511, Corollary 15.7.

so

$$f(x) - f(\infty) = \int_0^\infty e^{-xt} d\sigma(t).$$

Let

$$\mu = \sigma + f(\infty)\delta_0,$$

with which

$$\int_0^\infty e^{-xt} d\mu(t) = \int_0^\infty e^{-xt} d\sigma(t) + f(\infty),$$

hence

$$f(x) = \int_0^\infty e^{-xt} d\mu(t).$$

Because $f(0) = 1$, $\int_0^\infty d\mu(t) = 1$, showing that μ is a probability measure. $\square$

4 Fourier transforms

For a topological space X and a positive Borel measure μ on X, $F \subset X$ is called a support of μ if (i) F is closed, (ii) $\mu(F^c) = 0$, and (iii) if G is open and $G \cap F \neq \emptyset$ then $\mu(G \cap F) > 0$. If F_1 and F_2 are supports of μ, it is straightforward that $F_1 = F_2$. It is a fact that if X is second-countable then μ has a support, which we denote by $\operatorname{supp} \mu$.[9]

Lemma 2. If μ is a Borel measure on a topological space X and μ has a support $\operatorname{supp} \mu$, if $f : X \to [0, \infty)$ is continuous and $\int_X f d\mu = 0$ then $f(x) = 0$ for all $x \in \operatorname{supp} \mu$.

Proof. Let $F = \operatorname{supp} \mu$ and let $E = \{x \in X : f(x) \neq 0\}$. E is an open subset of X. Suppose by contradiction that there is some $x \in E \cap F$, i.e. that $E \cap F \neq \emptyset$. Because f is continuous and $f(x) > 0$, there is some open neighborhood G of x for which $f(y) > f(x)/2$ for $y \in U$. Then $x \in G \cap F$, so $G \cap F \neq \emptyset$ and because F is the support of μ, $\mu(G \cap F) > 0$ and a fortiori $\mu(G) > 0$. Then

$$0 = \int_X f d\mu \geq \int_G f(y) d\mu(y) \geq \int_G \frac{f(x)}{2} d\mu(y) = \frac{f(x)}{2}\mu(G) > 0,$$

a contradiction. Therefore $E \cap F = \emptyset$, i.e. for all $x \in F$, $f(x) = 0$. $\square$

The following lemma asserts that a certain function is nonzero λ_d-almost everywhere, where λ_d is Lebesgue measure on $\mathbb{R}^d$.[10]

[9] Charalambos D. Aliprantis and Kim C. Border, *Infinite Dimensional Analysis: A Hitchhiker's Guide*, third ed., p. 442, Theorem 12.14.

[10] Ward Cheney and Will Light, *A Course in Approximation Theory*, p. 91, chapter 13, Lemma 6.

Lemma 3. Let $x_1, \ldots, x_n$ be distinct points in $\mathbb{R}^d$, let $u \in \mathbb{C}^n$ not be the zero vector, and define

$$g(y) = \sum_{j=1}^{n} u_j e^{-2\pi i x_j \cdot y}, \qquad y \in \mathbb{R}^d.$$

For λ_d-almost all $y \in \mathbb{R}^d$, $g(y) \neq 0$.

The following theorem gives conditions under which the Fourier transform of a Borel measure on $\mathbb{R}^d$ is strictly positive definite.[11]

Theorem 4. If μ is a finite Borel measure on $\mathbb{R}^d$ and $\lambda_d(\operatorname{supp}\mu) > 0$, then $\hat{\mu} : \mathbb{R}^d \to \mathbb{C}$ is strictly positive definite.

Proof. For distinct $x_1, \ldots, x_n \in \mathbb{R}^d$ and for nonzero $u \in \mathbb{C}^n$,

$$\begin{aligned}
\sum_{j=1}^{n}\sum_{k=1}^{n} u_j \overline{u_k} \hat{\mu}(x_j - x_k) &= \sum_{j=1}^{n}\sum_{k=1}^{n} u_j \overline{u_k} \int_{\mathbb{R}^d} e^{-2\pi i (x_j - x_k)\cdot y} d\mu(y) \\
&= \int_{\mathbb{R}^d} \left(\sum_{j=1}^{n} u_j e^{-2\pi i x_j \cdot y}\right) \overline{\left(\sum_{k=1}^{n} u_k e^{-2\pi i x_k \cdot y}\right)} d\mu(y) \\
&= \int_{\mathbb{R}^d} \left|\sum_{j=1}^{n} u_j e^{-2\pi i x_j \cdot y}\right|^2 d\mu(y) \\
&= \int_{\mathbb{R}^d} |g(y)|^2 d\mu(y).
\end{aligned}$$

It is apparent that this is nonnegative. If it is equal to 0 then because g is continuous we obtain from Lemma 2 that $|g(y)|^2 = 0$ for all $y \in \operatorname{supp}\mu$, i.e. $g(y) = 0$ for all $y \in \operatorname{supp}\mu$. In other words,

$$\operatorname{supp}\mu \subset \{y \in \mathbb{R}^d : g(y) = 0\}.$$

But by Lemma 3, $\lambda_d(\{y \in \mathbb{R}^d : g(y) = 0\}) = 0$, so $\lambda_d(\operatorname{supp}\mu) = 0$, contradicting the hypothesis $\lambda_d(\operatorname{supp}\mu) > 0$. Therefore

$$\int_{\mathbb{R}^d} |g(y)|^2 d\mu(y) > 0,$$

which shows that $\hat{\mu}$ is strictly positive definite. $\square$

5 Schoenberg's theorem

Let $(X, \langle \cdot, \cdot \rangle)$ be a real inner product space. We call a function $F : X \to \mathbb{R}$ **radial** when $\|x\| = \|y\|$ implies that $F(x) = F(y)$.

[11] Ward Cheney and Will Light, *A Course in Approximation Theory*, p. 92, chapter 13, Theorem 3.

An identity that is worth memorizing is that for $y \in \mathbb{R}$,

$$\int_{\mathbb{R}} e^{-\pi x^2} e^{-2\pi i x y} dx = e^{-\pi y^2}.$$

Using this and Fubini's theorem yields, $y \in \mathbb{R}^d$,

$$\int_{\mathbb{R}^d} e^{-\pi |x|^2} e^{-2\pi \langle x, y \rangle} = e^{-\pi |y|^2}.$$

Lemma 5. For $\alpha > 0$ and $y \in \mathbb{R}^d$,

$$\int_{\mathbb{R}^d} \left(\frac{\pi}{\alpha}\right)^{d/2} \exp\left(-\frac{\pi^2}{\alpha}|x|^2\right) e^{-2\pi i \langle x, y \rangle} dx = e^{-\alpha |y|^2}.$$

Proof. Define $T : \mathbb{R}^d \to \mathbb{R}^d$ by

$$T(x) = \sqrt{\frac{\pi}{\alpha}} x, \qquad x \in \mathbb{R}^d.$$

$T'(x) = \sqrt{\frac{\pi}{\alpha}} I \in \mathscr{L}(\mathbb{R}^d)$ and $J_T(x) = \det T'(x) = \left(\frac{\pi}{\alpha}\right)^{d/2}$. Let $u \in \mathbb{R}^d$ and define $f(x) = e^{-\pi |x|^2} e^{-2\pi i \langle x, u \rangle}$. By the change of variables formula,[12]

$$\int_{\mathbb{R}^d} (f \circ T) \cdot |J_T| d\lambda_d = \int_{T(\mathbb{R}^d)} f d\lambda_d,$$

and because T is self-adjoint this is

$$\int_{\mathbb{R}^d} e^{-\pi |T(x)|^2} e^{-2\pi i \langle x, Tu \rangle} \left(\frac{\pi}{\alpha}\right)^{d/2} dx = \int_{\mathbb{R}^d} e^{-\pi |x|^2} e^{-2\pi i \langle x, u \rangle} dx,$$

and therefore

$$\int_{\mathbb{R}^d} \left(\frac{\pi}{\alpha}\right)^{d/2} \exp\left(-\frac{\pi^2}{\alpha}|x|^2\right) e^{-2\pi i \langle x, Tu \rangle} dx = e^{-\pi |u|^2}.$$

For $u = T^{-1}(y) = \sqrt{\frac{\alpha}{\pi}} y$ this is

$$\int_{\mathbb{R}^d} \left(\frac{\pi}{\alpha}\right)^{d/2} \exp\left(-\frac{\pi^2}{\alpha}|x|^2\right) e^{-2\pi i \langle x, y \rangle} dx = e^{-\alpha |y|^2},$$

proving the claim. □

We now prove that on a real inner product space, $x \mapsto e^{-\alpha \|x\|^2}$ is strictly positive definite whenever $\alpha > 0$.[13]

---

[12] Charalambos D. Aliprantis and Owen Burkinshaw, *Principles of Real Analysis*, third ed., p. 393, Theorem 40.7.

[13] Ward Cheney and Will Light, *A Course in Approximation Theory*, p. 104, chapter 15, Theorem 2.

Theorem 6. Let $(X, \langle\cdot,\cdot\rangle)$ be a real inner product space. If $\alpha > 0$, then

$$x \mapsto e^{-\alpha\|x\|^2}, \qquad x \in X,$$

is radial and strictly positive definite.

Proof. Let $x_1, \ldots, x_n$ be distinct points in X. There is an n-dimensional linear subspace V of X that contains $x_1, \ldots, x_n$. By the Gram-Schmidt process, V has an orthonormal basis $\{v_1, \ldots, v_n\}$. Define $T : V \to \mathbb{R}^n$ by $Tv_j = e_j$, where $\{e_1, \ldots, e_n\}$ is the standard basis for $\mathbb{R}^n$, which is an orthogonal transformation, and define

$$f(u) = e^{-\alpha|u|^2}, \qquad u \in \mathbb{R}^d.$$

For $u \in \mathbb{C}^n$, $u \neq 0$,

$$\begin{aligned}\sum_{j=1}^n \sum_{k=1}^n u_j \overline{u_k} e^{-\alpha\|x_j - x_k\|^2} &= \sum_{j=1}^n \sum_{k=1}^n u_j \overline{u_k} \exp\left(-\alpha|T(x_j - x_k)|^2\right) \\ &= \sum_{j=1}^n \sum_{k=1}^n u_j \overline{u_k} f(Tx_j - Tx_k).\end{aligned}$$

Now, let μ be the Borel measure on $\mathbb{R}^d$ whose density with respect to λ_d is

$$y \mapsto \left(\frac{\pi}{\alpha}\right)^{d/2} \exp\left(-\frac{\pi^2}{\alpha}|y|^2\right).$$

Because μ is absolutely continuous with respect to λ_d, $\lambda_d(\operatorname{supp}\mu) > 0$, so Theorem 4 states that the Fourier transform $\hat{\mu} : \mathbb{R}^d \to \mathbb{C}$ is strictly positive definite. Applying Lemma 5, the Fourier transform of μ is

$$\hat{\mu}(u) = \int_{\mathbb{R}^d} \left(\frac{\pi}{\alpha}\right)^{d/2} \exp\left(-\frac{\pi^2}{\alpha}|y|^2\right) e^{-2\pi i\langle y,u\rangle} dy = e^{-\alpha|u|^2} = f(u),$$

so f is strictly positive definite. Because T is an orthogonal transformation it is in particular one-to-one, so $Tx_1, \ldots, Tx_n$ are distinct points in $\mathbb{R}^d$. Thus the fact that f is strictly positive definite means that

$$\sum_{j=1}^n \sum_{k=1}^n u_j \overline{u_k} e^{-\alpha\|x_j - x_k\|^2} = \sum_{j=1}^n \sum_{k=1}^n u_j \overline{u_k} f(Tx_j - Tx_k) > 0,$$

which establishes that $x \mapsto e^{-\alpha\|x\|^2}$ is strictly positive definite. □

The following is **Schoenberg's theorem**.[14]

[14] Ward Cheney and Will Light, *A Course in Approximation Theory*, p. 101, chapter 15, Theorem 1; René L. Schilling, Renming Song, and Zoran Vondraček, *Bernstein Functions: Theory and Applications*, p. 142, Theorem 12.14; William F. Donoghue Jr., *Distributions and Fourier Transforms*, p. 205, §41.

Theorem 7 (Schoenberg's theorem)**.** Let $(X, \langle \cdot, \cdot \rangle)$ be a real inner product space. If $f : [0, \infty) \to \mathbb{R}$ is completely monotone, $f(0) = 1$, and f is not constant, then

$$x \mapsto f(\|x\|^2), \qquad X \to [0, \infty),$$

is radial and strictly positive definite.

Proof. Because f is completely monotone, the Bernstein-Widder theorem (Theorem 1) tells us that there is a Borel probability measure μ on $[0, \infty)$ such that

$$f(t) = \int_0^\infty e^{-st} d\mu(s), \qquad t \in [0, \infty),$$

that is, f is the Laplace transform of μ. Now, the Laplace transform of δ_0 is $t \mapsto 1$, and because f is not constant, the Laplace transform of μ is not equal to the Laplace transform of δ_0, which implies that $\mu \neq \delta_0$.[15] Therefore $\mu((0, \infty)) > 0$.

Let $x_1, \ldots, x_n$ be distinct points in X and let $u \in \mathbb{C}^n$, $u \neq 0$. Then, because $\sum_{j=1}^n \sum_{k=1}^n u_j \overline{u_k} \geq 0$,

$$\begin{aligned}
\sum_{j=1}^n \sum_{k=1}^n u_j \overline{u_k} f(\|x_j - x_k\|^2) &= \sum_{j=1}^n \sum_{k=1}^n u_j \overline{u_k} \int_0^\infty \exp\left(-s \|x_j - x_k\|^2\right) d\mu(s) \\
&= \int_0^\infty \sum_{j=1}^n \sum_{k=1}^n u_j \overline{u_k} \exp\left(-s \|x_j - x_k\|^2\right) d\mu(s) \\
&= \sum_{j=1}^n \sum_{k=1}^n u_j \overline{u_k} \mu(\{0\}) \\
&\quad + \int_0^\infty 1_{(0,\infty)}(s) \sum_{j=1}^n \sum_{k=1}^n u_j \overline{u_k} \exp\left(-s \|x_j - x_k\|^2\right) d\mu(s) \\
&\geq \int_0^\infty 1_{(0,\infty)}(s) \sum_{j=1}^n \sum_{k=1}^n u_j \overline{u_k} \exp\left(-s \|x_j - x_k\|^2\right) d\mu(s) \\
&= \int_0^\infty g(s) d\mu(s).
\end{aligned}$$

Assume by contradiction that $\int_0^\infty g(s) d\mu(s) = 0$. Because $g \geq 0$, this implies that $\mu(\{s \in [0, \infty) : g(s) > 0\}) = 0$.[16] By Theorem 6, for each $s > 0$,

$$\sum_{j=1}^n \sum_{k=1}^n u_j \overline{u_k} \exp\left(-s \|x_j - x_k\|^2\right) > 0,$$

[15] Bert Fristedt and Lawrence Gray, *A Modern Approach to Probability Theory*, p. 218, §13.5, Theorem 6.

[16] Charalambos D. Aliprantis and Kim C. Border, *Infinite Dimensional Analysis: A Hitchhiker's Guide*, third ed., p. 411, Theorem 11.16.

so $g(s) > 0$ when $s > 0$. Thus $\mu((0, \infty)) = 0$, a contradiction. Therefore,

$$\sum_{j=1}^{n} \sum_{k=1}^{n} u_j \overline{u_k} f(\|x_j - x_k\|^2) = \int_0^\infty g(s) d\mu(s) > 0,$$

which shows that $x \mapsto f(\|x\|^2)$ is strictly positive definite. $\square$