MERCER'S THEOREM

JOHN THICKSTUN

INNER PRODUCTS

Let V be a finite dimensional vector space equipped with an inner product $\langle\cdot,\cdot\rangle_V$. The vector space of linear maps $V \to V$ is isomorphic to the bilinear functionals $V \times V \to \mathbb{R}$. The subspace of self-adjoint maps is isomorphic to the symmetric bilinear functionals. Therefore an inner product on V must have the form $(x, y) \mapsto \langle x, My\rangle_V$ for some self-adjoint linear map $M : V \to V$. Denote this map by $\langle\cdot,\cdot\rangle_M$. We show in the following proposition that the inner products on a finite dimensional vector space are characterized by the positive definite linear operators.

Proposition 1. *Let $M : V \to V$ be a self-adjoint linear map. Then M is positive definite iff $\langle\cdot,\cdot\rangle_M$ is an inner product.*

Proof. ($\Longrightarrow$) By the spectral theorem, $M = UDU^*$ for unitary U, diagonal D. If M is positive definite then D admits a unique square root $D^{\frac{1}{2}}$ and

$$\langle x, x\rangle_M = \langle x, UDU^*x\rangle_V = \langle D^{\frac{1}{2}}U^*x, D^{\frac{1}{2}}U^*x\rangle_V \geq 0.$$

Because $D^{\frac{1}{2}}U^*$ is full-rank, $\langle x, x\rangle_M = 0$ iff $x = 0$.

($\Longleftarrow$) Let y be an eigenvector of M with eigenvalue λ; i.e. $My = \lambda y$. If $\langle\cdot,\cdot\rangle_M$ is an inner product then

$$\lambda\langle y, y\rangle_V = \langle y, \lambda y\rangle_V = \langle y, My\rangle_V = \langle y, y\rangle_M > 0.$$

Therefore $\lambda > 0$. $\square$

KERNELS

Let $\mathcal{X}$ be a measure space with $k : L^2(\mathcal{X} \times \mathcal{X}) \to \mathbb{R}$. We call k a kernel iff there is some feature map $\varphi : \mathcal{X} \to \mathcal{H}$ into a separable Hilbert space $\mathcal{H}$ such that

$$k(x, x') \equiv \langle\varphi(x), \varphi(x')\rangle_{\mathcal{H}}.$$

I.e. k is a kernel iff for some space $\mathcal{H}$ and map φ, the following diagram commutes:

$$\begin{tikzcd}
\mathcal{X} \times \mathcal{X} \arrow[d, "\varphi"'] \arrow[dr, "k"] & \\
\mathcal{H} \times \mathcal{H} \arrow[r, "\langle\cdot,\cdot\rangle_{\mathcal{H}}"'] & \mathbb{R}
\end{tikzcd}$$

Note that φ and $\mathcal{H}$ need not be unique. If $\mathcal{X} = \mathbb{R}$ and $k(x, x') = xx'$ then k factors through both $\varphi_1(x) = x$ and $\varphi_2(x) = \frac{x}{\sqrt{2}}(e_1 + e_2)$ where $e_1, e_2 \in \mathbb{R}^2$ are the standard basis.

FINITE SETS

Suppose $\mathcal{X}$ is a finite set of size n equipped with the counting measure. Consider the set of functions $\mathcal{X} \to \mathbb{R}$. This is an n-dimensional Hilbert space V. We can associate a symmetric function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ with a self-adjoint linear map $K : V \to V$,

$$(Ku)_i = \sum_{j=1}^{n} k(x_i, x_j)u_j.$$

We say k is positive definite iff the corresponding transformation K is positive definite.

Proposition 2. *A symmetric function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is positive definite iff k is a kernel.*

Proof. ($\Longrightarrow$) Suppose k is positive definite (i.e. K is positive definite) and let $K = UDU^*$ be the spectral decomposition of K. Because K is positive definite, D is positive definite and we may define $\varphi(x) = D^{\frac{1}{2}}U^*\mathbb{1}_x$. Then

$$\langle \varphi(x), \varphi(x') \rangle_V = \langle \mathbb{1}_x, \mathbb{1}_{x'} \rangle_K = \sum_{i,j} \mathbb{1}_x(x_i)k(x_i, x_j)\mathbb{1}_{x'}(x_j) = k(x, x').$$

It follows that k is a kernel.

($\Longleftarrow$) Conversely, suppose k is a kernel. Then

$$\langle u, Ku \rangle_V = \sum_{i=1}^{n} u_i(Ku)_i = \sum_{i=1}^{n} u_i \sum_{j=1}^{n} \langle \varphi(x_i), \varphi(x_j) \rangle_{\mathcal{H}} u_j$$

$$= \left\langle \sum_{i=1}^{n} u_i\varphi(x_i), \sum_{j=1}^{n} u_j\varphi(x_j) \right\rangle_{\mathcal{H}} = \left\| \sum_{i=1}^{n} u_i\varphi(x_i) \right\|_{\mathcal{H}}^2 \geq 0.$$

Therefore K is positive definite. $\square$

HILBERT SPACES

Proposition 3. *The vector space of bounded linear maps on a real Hilbert space $\mathcal{H}$ are isomorphic to its bounded bilinear forms:*

$$\mathcal{H} \to \mathcal{H} \cong \mathcal{H} \times \mathcal{H} \to \mathbb{R}.$$

Proof. Intuitively, because Hilbert spaces are self-dual, we can capture this identity with some algebraic manipulations:

$$\mathcal{H} \times \mathcal{H} \to \mathbb{R} \cong \mathcal{H} \to (\mathcal{H} \to \mathbb{R}) \cong \mathcal{H} \to \mathcal{H}^* \cong \mathcal{H} \to \mathcal{H}.$$

We now make a more careful argument. Suppose B is a bounded bilinear form on $\mathcal{H}$. If $u \in \mathcal{H}$ then $B(u, \cdot)$ is a bounded linear functional. By Riesz representation, $B(u, \cdot) = \langle Ku, \cdot \rangle_{\mathcal{H}}$ for some unique $Ku \in \mathcal{H}$. The map $K : \mathcal{H} \to \mathcal{H}$ is a bounded linear operator. Conversely, any bounded linear $K : \mathcal{H} \to \mathcal{H}$ defines a bounded bilinear form $\langle f, Kg \rangle_{\mathcal{H}}$. $\square$

In light of this result, we will again introduce the notation $\langle \cdot, \cdot \rangle_K$ to indicate the bilinear form associated with an operator K. The subspace of bounded self-adjoint operators is isomorphic to the symmetric bilinear forms. Therefore, like in the finite setting, we see that any inner product must have this form. The following statement is probably true:

Theorem. *Let $K : \mathcal{H} \to \mathcal{H}$ be a bounded, self-adjoint linear map. Then K is positive definite iff $\langle \cdot, \cdot \rangle_K$ is an inner product.*

However, the proof of this would require some heavy machinery from spectral operator theory, and we don't need such a general statement. We would need a general definition of a positive definite linear operator, and we would also need to work with a spectral theorem for non-compact operators. Instead we will turn our attention to a particular class of Hilbert spaces.

Mercer's theorem

Let $\mathcal{X}$ be a measure space and let $L^2(\mathcal{X})$ indicate the Hilbert space of square-integrable functions $\mathcal{X} \to \mathbb{R}$. Let $k \in L^2(\mathcal{X} \times \mathcal{X})$ and consider the corresponding Hilbert-Schmidt operator $K : L^2(\mathcal{X}) \to L^2(\mathcal{X})$,

$$Kf(x) = \int_{\mathcal{X}} k(x, x') f(x') \, dx'.$$

If k is symmetric, i.e. $k(x, x') = k(x', x)$ then by Fubini's theorem,

$$\langle f, Kg \rangle = \int_{\mathcal{X}} f(x) \int_{\mathcal{X}} k(x, x') g(x') \, dx' \, dx = \int_{\mathcal{X}} \int_{\mathcal{X}} k(x', x) f(x) \, dx \, g(x') \, dx' = \langle Kf, g \rangle.$$

So symmetric k correspond to self-adjoint Hilbert-Schmidt operators. We say that k satisfies *Mercer's condition* iff

$$\int_{\mathcal{X} \times \mathcal{X}} k(x, x') f(x) f(x') \, dx \, dx' \geq 0 \text{ for all } f \in L^2(\mathcal{X}).$$

This is the natural generalization of positive definiteness to Hilbert-Schmidt operators.

Theorem. *(Mercer) A symmetric $k \in L^2(\mathcal{X} \times \mathcal{X})$ is Mercer iff k is a kernel.*

Proof. ($\Longrightarrow$) Let K be the self-adjoint Hilbert-Schmidt operator corresponding to k. The theory of Hilbert-Schmidt operators ensures that K is compact. By the spectral theorem for compact self-adjoint operators, K has a countable collection of orthonormal eigenvectors $U_i \in L^2(\mathcal{X})$ with eigenvalues λ_i such that for all $f \in L^2$,

$$Kf = \sum_{i=1}^{\infty} \lambda_i \langle f, U_i \rangle_{L^2} U_i.$$

By Mercer's condition and Fubini's theorem,

$$\lambda_i \langle U_i, U_i \rangle_{L^2} = \langle U_i, KU_i \rangle_{L^2} = \int_{\mathcal{X}} U_i(x) \int_{\mathcal{X}} k(x, x') U_i(x') \, dx' \, dx \geq 0.$$

Because $\langle \cdot, \cdot \rangle_{L^2}$ is positive definite, $\lambda_i \geq 0$. Define $\varphi_i \in L^2(\mathcal{X})$ by $x \mapsto \sqrt{\lambda_i} U_i(x)$. Then

$$\int_{\mathcal{X}} \langle \varphi(x), \varphi(x') \rangle_{\ell^2} f(x')\, dx' = \int_{\mathcal{X}} \sum_{i=1}^{\infty} \lambda_i U_i(x) U_i(x') f(x')\, dx'$$

$$= \sum_{i=1}^{\infty} \lambda_i U_i(x) \int_{\mathcal{X}} U_i(x') f(x')\, dx = \sum_{i=1}^{\infty} \lambda_i \langle f, U_i \rangle_{L^2} U_i(x) = Kf(x).$$

This holds for all f and therefore

$$k(x, x') = \langle \varphi(x), \varphi(x') \rangle_{\ell^2}.$$

($\Longleftarrow$) Conversely, if k is a kernel then

$$\int_{\mathcal{X} \times \mathcal{X}} k(x, x') f(x) f(x')\, dx\, dx' = \int_{\mathcal{X} \times \mathcal{X}} \langle f(x)\varphi(x), f(x')\varphi(x') \rangle_{\mathcal{H}}\, dx\, dx'$$

$$= \left\langle \int_{\mathcal{X}} f(x)\varphi(x)\, dx, \int_{\mathcal{X}} f(x')\varphi(x')\, dx' \right\rangle_{\mathcal{H}} \geq 0.$$

Therefore k is Mercer. □

Commuting the integral and inner product commute in the final step above can be justified by writing the vectors $\varphi(x)$ and $\varphi(x')$ in a Schauder basis (guaranteed to exist by the separability of $\mathcal{H}$), writing the inner product as a countable sum of element-wise products in this basis, and applying Fubini's theorem. I'm not sure what happens here if $\mathcal{H}$ isn't required to be separable (we don't require separability at any other point in the argument). Maybe there's better argument that avoids this complication.