Mercer's theorem

From Wikipedia, the free encyclopedia

In mathematics, specifically functional analysis, Mercer's theorem is a representation of a symmetric positive-definite function on a square as a sum of a convergent sequence of product functions. This theorem, presented in (Mercer 1909), is one of the most notable results of the work of James Mercer. It is an important theoretical tool in the theory of integral equations; it is used in the Hilbert space theory of stochastic processes, for example the Karhunen–Loève theorem; and it is also used to characterize a symmetric positive semi-definite kernel.[1]

Contents

Introduction

To explain Mercer's theorem, we first consider an important special case; see below for a more general formulation. A *kernel*, in this context, is a symmetric continuous function

$$K : [a, b] \times [a, b] \to \mathbb{R}$$

where symmetric means that $K(x, s) = K(s, x)$.

K is said to be *non-negative definite* (or positive semidefinite) if and only if

$$\sum_{i=1}^{n} \sum_{j=1}^{n} K(x_i, x_j) c_i c_j \geq 0$$

for all finite sequences of points $x_1, \ldots, x_n$ of $[a, b]$ and all choices of real numbers $c_1, \ldots, c_n$ (cf. positive definite kernel).

Associated to K is a linear operator (more specifically a Hilbert–Schmidt integral operator) on functions defined by the integral

$$[T_K \varphi](x) = \int_a^b K(x,s)\varphi(s)\,ds.$$

For technical considerations we assume φ can range through the space $L^2[a, b]$ (see Lp space) of square-integrable real-valued functions. Since T is a linear operator, we can talk about eigenvalues and eigenfunctions of T.

Theorem. Suppose K is a continuous symmetric non-negative definite kernel. Then there is an orthonormal basis $\{e_i\}_i$ of $L^2[a, b]$ consisting of eigenfunctions of T_K such that the corresponding sequence of eigenvalues $\{\lambda_i\}_i$ is nonnegative. The eigenfunctions corresponding to non-zero eigenvalues are continuous on $[a, b]$ and K has the representation

$$K(s,t) = \sum_{j=1}^\infty \lambda_j \, e_j(s) \, e_j(t)$$

where the convergence is absolute and uniform.

Details

We now explain in greater detail the structure of the proof of Mercer's theorem, particularly how it relates to spectral theory of compact operators.

- The map $K \to T_K$ is injective.
- T_K is a non-negative symmetric compact operator on $L^2[a,b]$; moreover $K(x, x) \geq 0$.

To show compactness, show that the image of the unit ball of $L^2[a,b]$ under T_K equicontinuous and apply Ascoli's theorem, to show that the image of the unit ball is relatively compact in C([a,b]) with the uniform norm and *a fortiori* in $L^2[a,b]$.

Now apply the spectral theorem for compact operators on Hilbert spaces to T_K to show the existence of the orthonormal basis $\{e_i\}_i$ of $L^2[a,b]$

$$\lambda_i e_i(t) = [T_K e_i](t) = \int_a^b K(t,s) e_i(s)\,ds.$$

If $\lambda_i \neq 0$, the eigenvector e_i is seen to be continuous on $[a,b]$. Now

$$\sum_{i=1}^\infty \lambda_i |e_i(t) e_i(s)| \leq \sup_{x \in [a,b]} |K(x,x)|^2,$$

which shows that the sequence

$$\sum_{i=1}^{\infty} \lambda_i e_i(t) e_i(s)$$

converges absolutely and uniformly to a kernel K_0 which is easily seen to define the same operator as the kernel K. Hence $K=K_0$ from which Mercer's theorem follows.

Finally, to show non-negativity of the eigenvalues one can write $\lambda\langle f, f\rangle = \langle f, T_K f\rangle$ and expressing the right hand side as an integral well approximated by its Riemann sums, which are non-negative by positive definiteness of K, implying $\lambda\langle f, f\rangle \geq 0$, implying $\lambda \geq 0$.

Trace

The following is immediate:

Theorem. Suppose K is a continuous symmetric non-negative definite kernel; T_K has a sequence of nonnegative eigenvalues $\{\lambda_i\}_i$. Then

$$\int_a^b K(t,t)\, dt = \sum_i \lambda_i.$$

This shows that the operator T_K is a trace class operator and

$$\operatorname{trace}(T_K) = \int_a^b K(t,t)\, dt.$$

Generalizations

Mercer's theorem itself is a generalization of the result that any positive semidefinite matrix is the Gramian matrix of a set of vectors.

The first generalization replaces the interval [*a*, *b*] with any compact Hausdorff space and Lebesgue measure on [*a*, *b*] is replaced by a finite countably additive measure μ on the Borel algebra of *X* whose support is *X*. This means that μ(*U*) > 0 for any nonempty open subset *U* of *X*.

A recent generalization replaces this conditions by that follows: the set *X* is a first-countable topological space endowed with a Borel (complete) measure μ. *X* is the support of μ and, for all *x* in *X*, there is an open set *U* containing *x* and having finite measure. Then essentially the same result holds:

Theorem. Suppose *K* is a continuous symmetric positive definite kernel on *X*. If the function κ is $L^1_\mu(X)$, where κ(x)=K(x,x), for all x in X, then there is an orthonormal set $\{e_i\}_i$ of $L^2_\mu(X)$ consisting of eigenfunctions of T_K such that corresponding sequence of eigenvalues $\{\lambda_i\}_i$ is nonnegative. The eigenfunctions corresponding to non-zero eigenvalues are continuous on X and K has the representation

$$K(s,t) = \sum_{j=1}^\infty \lambda_j \, e_j(s) \, e_j(t)$$

where the convergence is absolute and uniform on compact subsets of X.

The next generalization deals with representations of *measurable* kernels.

Let (X, M, μ) be a σ-finite measure space. An L^2 (or square integrable) kernel on X is a function

$$K \in L^2_{\mu \otimes \mu}(X \times X).$$

L^2 kernels define a bounded operator T_K by the formula

$$\langle T_K \varphi, \psi \rangle = \int_{X \times X} K(y,x) \varphi(y) \psi(x) \, d[\mu \otimes \mu](y,x).$$

T_K is a compact operator (actually it is even a Hilbert–Schmidt operator). If the kernel K is symmetric, by the spectral theorem, T_K has an orthonormal basis of eigenvectors. Those eigenvectors that correspond to non-zero eigenvalues can be arranged in a sequence $\{e_i\}_i$ (regardless of separability).

Theorem. If K is a symmetric positive definite kernel on(X, M, μ), then

$$K(y,x) = \sum_{i \in \mathbb{N}} \lambda_i e_i(y) e_i(x)$$

where the convergence in the L^2 norm. Note that when continuity of the kernel is not assumed, the expansion no longer converges uniformly.

Mercer's condition

In mathematics, a real-valued function *K(x,y)* is said to fulfill Mercer's condition if for all square integrable functions $g(x)$ one has

$$\iint g(x) K(x,y) g(y) \, dx \, dy \geq 0.$$

Discrete analog

This is analogous to the definition of a positive-semidefinite matrix. This is a matrix K of dimension N, which satisfies, for all vectors g, the property

$$(g, Kg) = g^T \cdot Kg = \sum_{i=1}^{N} \sum_{j=1}^{N} g_i \, K_{ij} \, g_j \geq 0.$$

Examples

A positive constant function

$$K(x, y) = c$$

satisfies Mercer's condition, as then the integral becomes by Fubini's theorem

$$\iint g(x) \, c \, g(y) \, dxdy = c \int g(x) \, dx \int g(y) \, dy = c \left(\int g(x) \, dx \right)^2$$

which is indeed non-negative.

See also

- Kernel trick
- Representer theorem
- Spectral theory
- Mercer's condition

References

- Adriaan Zaanen, *Linear Analysis*, North Holland Publishing Co., 1960,
- Ferreira, J. C., Menegatto, V. A., *Eigenvalues of integral operators defined by smooth positive definite kernels*, Integral equation and Operator Theory, 64 (2009), no. 1, 61–81. (Gives the generalization of Mercer's theorem for metric spaces. The result is easily adapted to first countable topological spaces)
- Konrad Jörgens, *Linear integral operators*, Pitman, Boston, 1982,
- Richard Courant and David Hilbert, *Methods of Mathematical Physics*, vol 1, Interscience 1953,
- Robert Ash, *Information Theory*, Dover Publications, 1990,
- Mercer, J. (1909), "Functions of positive and negative type and their connection with the theory of integral equations", *Philosophical Transactions of the Royal Society A*, 209 (441–458): 415–446, doi:10.1098/rsta.1909.0016,
- Hazewinkel, Michiel, ed. (2001), "Mercer theorem", *Encyclopedia of Mathematics*, Springer, ISBN 978-1-55608-010-4

- H. König, *Eigenvalue distribution of compact operators*, Birkhäuser Verlag, 1986. (Gives the generalization of Mercer's theorem for finite measures μ.)

Notes

1. http://www.cs.berkeley.edu/~bartlett/courses/281b-sp08/7.pdf

Categories: Theorems in functional analysis
