1 Mercer's Theorem and Feature Map

1.1 Mercer's Theorem

Definition 1.1 (Definite Kernel)**.** *The function $k : [a,b] \times [a,b] \to \mathbb{R}$ is a definite kernel where the following double integral:*

$$J(f) = \int_a^b \int_a^b k(x,y) f(x) f(y) dx dy, \tag{1}$$

satisfies $J(f) > 0$ for all $f(x) \neq 0$.

Mercer improved over Hilbert's work to propose his theorem, the Mercer's theorem, introduced in the following.

Theorem 1.1 (Mercer's Theorem)**.** *Suppose $k : [a,b] \times [a,b] \to \mathbb{R}$ is a continuous symmetric positive semi-definite kernel which is bounded:*

$$\sup_{x,y} k(x,y) < \infty. \tag{2}$$

Assume the operator T_k

$$T_k f(x) := \int_a^b k(x,y) f(y) dy, \tag{3}$$

is positive semi-definite:

$$\iint k(x,y) f(y) dx dy \geq 0. \tag{4}$$

1) Then, there is a set of orthonormal bases $\{\psi_i(.)\}_{i=1}^{\infty}$ of $L^2(a,b)$ consisting of eigenfunctions of T_K such that the corresponding sequence of eigenvalues $\{\lambda_i\}_{i=1}^{\infty}$ are non-negative:

$$\int k(x,y) \psi_i(y) dy = \lambda_i \psi_i(x). \tag{5}$$

2) The eigenfunctions corresponding to the non-zero eigenvalues are continuous on $[a,b]$ and k can be represented as:

$$k(x,y) = \sum_{i=1}^{\infty} \lambda_i \psi_i(x) \psi_i(y), \tag{6}$$

where the convergence is absolute and uniform.

Proof. The proof relies on establishing that the operator T_k is symmetric and compact, which allows the application of the Spectral Theorem for compact self-adjoint operators.

Step 1: The Operator T_k is Symmetric and Compact. According to the theorem's assumptions, the Hilbert-Schmidt integral operator T_k is a symmetric operator on the Hilbert space $L^2(a,b)$ and is also a compact operator.

- **Symmetry:** The symmetry of T_k follows from the symmetry of the kernel $k(x,y)$, i.e., $k(x,y) = k(y,x)$. For any $f, g \in L^2(a,b)$, we must show $\langle T_k f, g \rangle_{L^2} = \langle f, T_k g \rangle_{L^2}$. Expanding the left side:

$$\begin{aligned} \langle T_k f, g \rangle_{L^2} &= \int_a^b (T_k f(x)) g(x) dx \\ &= \int_a^b \left(\int_a^b k(x,y) f(y) dy \right) g(x) dx \end{aligned}$$

Since $k(x, y)$ is continuous and bounded, the integrand is absolutely integrable, allowing us to interchange the order of integration using **Fubini's Theorem**:

$$\begin{aligned}
\langle T_k f, g \rangle_{L^2} &= \int_a^b \int_a^b k(x, y) f(y) g(x) dx dy \\
&= \int_a^b f(y) \left(\int_a^b k(y, x) g(x) dx \right) dy \quad (\text{using } k(x, y) = k(y, x)) \\
&= \int_a^b f(y)(T_k g(y)) dy = \langle f, T_k g \rangle_{L^2}.
\end{aligned}$$

Thus, T_k is symmetric (self-adjoint).

- **Compactness:** The compactness of T_k is guaranteed because $k(x, y)$ is continuous and bounded on the compact domain $[a, b] \times [a, b]$. The operator T_k maps bounded sets in $L^2(a, b)$ to relatively compact sets in $C([a, b])$. This relies on the **Arzelà-Ascoli Theorem**, which requires that the image set $\mathcal{F} = \{T_k f \mid \|f\|_{L^2} \leq 1\}$ is uniformly bounded and equicontinuous. Both conditions are satisfied due to the uniform continuity and boundedness of $k(x, y)$.

Step 2: Application of the Spectral Theorem and Eigenvalue Decomposition. Since T_k is a symmetric (self-adjoint) and compact operator on the Hilbert space $L^2(a, b)$, the **Spectral Theorem for compact self-adjoint operators** applies. This theorem guarantees:

1. The existence of an orthonormal basis $\{\psi_i(.)\}_{i=1}^{\infty}$ for $L^2(a, b)$ consisting of the eigenfunctions of T_k.
2. That the operator satisfies the eigenvalue decomposition:

$$T_k \psi_i(x) = \lambda_i \psi_i(x), \tag{7}$$

where $\{\lambda_i\}_{i=1}^{\infty}$ are the corresponding real eigenvalues.

Substituting the definition of T_k from Eq. (3) into Eq. (7) yields the eigenfunction decomposition:

$$\int k(x, y) \psi_i(y) dy \overset{(3)}{=} T_k \psi_i(x) \overset{(7)}{=} \lambda_i \psi_i(x). \tag{8}$$

Furthermore, since $k(x, y)$ is positive semi-definite (Eq. (4)), it guarantees that all eigenvalues λ_i must be non-negative ($\lambda_i \geq 0$).

Step 3: Derivation of the Kernel Expansion (Mercer's Series). The functions $\{\psi_i(.)\}_{i=1}^{\infty}$ form a **complete orthonormal basis** for $L^2(a, b)$ due to the Spectral Theorem. This completeness means that any function in $L^2(a, b)$, including $g_x(y) = k(x, y)$ (for a fixed x), can be perfectly represented by the basis. This is formalized by **Parseval's Identity** (the equality case of Bessel's Inequality).

1. We expand $k(x, y)$ (as a function of y) in the orthonormal basis $\{\psi_i(y)\}$:

$$k(x, y) = \sum_{i=1}^{\infty} \langle k(x, \cdot), \psi_i \rangle_{L^2} \psi_i(y)$$

2. The coefficients are the inner products, which by the eigenvalue equation (Eq. (8)) are:

$$\langle k(x, \cdot), \psi_i \rangle_{L^2} = \int_a^b k(x, y) \psi_i(y) dy = \lambda_i \psi_i(x). \tag{9}$$

3. Substituting the coefficients gives Mercer's Series:

$$k(x,y) = \sum_{i=1}^{\infty}(\lambda_i\psi_i(x))\psi_i(y) = \sum_{i=1}^{\infty}\lambda_i\psi_i(x)\psi_i(y), \tag{10}$$

which is Eq. (6).

Step 4: Proving Absolute and Uniform Convergence. The final step is crucial as it elevates the convergence from L^2 (guaranteed by the Spectral Theorem) to the stronger **absolute and uniform convergence** required by the theorem.

- **Meaning of Convergence:**
  1. **Absolute Convergence:** The series $\sum_{i=1}^{\infty}|\lambda_i\psi_i(x)\psi_i(y)|$ must converge.
  2. **Uniform Convergence:** The rate of convergence of the partial sums $S_n(x,y) = \sum_{i=1}^{n}\lambda_i\psi_i(x)\psi_i(y)$ to $k(x,y)$ must be independent of the location (x,y) in the domain. Uniform convergence is required to ensure that the sum $k(x,y)$ inherits the **continuity** property from its continuous terms (eigenfunctions $\psi_i(x)$).

- **The Truncated Kernel $r_n(x,y)$:** To prove this strong form of convergence, we must analyze the remainder of the series. We define the truncated kernel r_n (with parameter n) as the remainder:

$$r_n(x,y) := k(x,y) - \sum_{i=1}^{n}\lambda_i\psi_i(x)\psi_i(y) = \sum_{i=n+1}^{\infty}\lambda_i\psi_i(x)\psi_i(y). \tag{11}$$

  The definition of the truncated kernel is essential because it allows us to utilize the positive semi-definite property of $k(x,y)$. Since $k(x,y)$ is positive semi-definite and each term $\lambda_i\psi_i(x)\psi_i(y)$ (with $\lambda_i \geq 0$) is also a rank-1 positive semi-definite kernel, their difference $r_n(x,y)$ is **also a positive semi-definite kernel**.

- **Establishing the Bound:** The positive semi-definite property implies that $r_n(x,x) \geq 0$ for every $x \in [a,b]$, which leads to:

$$r_n(x,x) = k(x,x) - \sum_{i=1}^{n}\lambda_i\psi_i(x)^2 \geq 0$$

  which establishes the critical upper bound for the partial sums:

$$\sum_{i=1}^{n}\lambda_i\psi_i(x)^2 \leq k(x,x) \leq \sup_{x\in[a,b]} k(x,x). \tag{12}$$

  Since $k(x,x)$ is bounded (Eq. (2)), the partial sums for the series $\sum_{i=1}^{\infty}\lambda_i\psi_i(x)^2$ are uniformly bounded, ensuring its convergence.

- **Conclusion of Convergence and Uniform Bound:** The uniform boundedness of the partial sums for the diagonal terms, $\sum_{i=1}^{n}\lambda_i\psi_i(x)^2$ (Eq. (12)), is the key. Since the eigenvalues λ_i are non-negative, the series $\sum_{i=1}^{\infty}\lambda_i\psi_i(x)^2$ converges pointwise and is bounded by $\sup_x k(x,x)$.

  We now apply the **Cauchy-Schwarz inequality for series** to bound the magnitude of the remainder term $r_n(x,y)$ (the sum from $i = n+1$ to ∞):

$$\begin{aligned}
|r_n(x,y)| &= \left|\sum_{i=n+1}^{\infty}\lambda_i\psi_i(x)\psi_i(y)\right| \\
&\leq \sum_{i=n+1}^{\infty}\lambda_i|\psi_i(x)\psi_i(y)| \quad \text{(Since } \lambda_i \geq 0, \text{ the absolute value moves inside)} \\
&\leq \sqrt{\left(\sum_{i=n+1}^{\infty}\lambda_i\psi_i(x)^2\right)\left(\sum_{i=n+1}^{\infty}\lambda_i\psi_i(y)^2\right)}
\end{aligned}$$

Let $R_n(x) = \sum_{i=n+1}^{\infty} \lambda_i \psi_i(x)^2$ be the remainder of the convergent series for $k(x, x)$. Since $k(x, y)$ is continuous on the compact domain, the convergence of the diagonal series $k(x, x) = \sum_{i=1}^{\infty} \lambda_i \psi_i(x)^2$ is itself uniform. This implies that $R_n(x) \to 0$ uniformly for all $x \in [a, b]$ as $n \to \infty$.

Since both $R_n(x) \to 0$ and $R_n(y) \to 0$ uniformly, their product also goes to zero uniformly:

$$|r_n(x, y)| \leq \sqrt{R_n(x)R_n(y)} \to 0 \quad \text{uniformly as } n \to \infty.$$

Because the remainder term $r_n(x, y)$ converges uniformly to zero, the original series $\sum_{i=1}^{\infty} \lambda_i \psi_i(x)\psi_i(y)$ converges **uniformly**. Furthermore, the boundedness of the series of absolute values (established via the Cauchy-Schwarz bound) proves **absolute convergence**.

This strong form of convergence (absolute and uniform) is necessary to ensure the resulting function $k(x, y)$ is continuous, validating the series representation of the kernel, and concluding the proof.

$\square$