# Kernel Method: Data Analysis with Positive Definite Kernels

## 5. Theory on Positive Definite Kernel and Reproducing Kernel Hilbert Space

Kenji Fukumizu

The Institute of Statistical Mathematics.
Graduate University of Advanced Studies /
Tokyo Institute of Technology

Nov. 17-26, 2010
Intensive Course at Tokyo Institute of Technology

# Outline

# Review on Positive Definite Kernels I

**Proposition 1**

*If $k_i : \mathcal{X} \times \mathcal{X} \to \mathbb{C}$ $(i = 1, 2, \ldots)$ are positive definite kernels, then so are the following:*

1. *(positive combination)* $\quad ak_1 + bk_2 \qquad (a, b \geq 0)$.
2. *(product)* $\quad k_1 k_2 \qquad (k_1(x, y)k_2(x, y))$ .
3. *(limit)* $\quad lim_{i\to\infty} k_i(x, y)$, *assuming the limit exists.*

Remark. Proposition 1 says that the set of all positive definite kernels is closed (w.r.t. pointwise convergence) convex cone stable under multiplication.

Example: If $k(x, y)$ is positive definite,

$$e^{k(x,y)} = 1 + k + \frac{1}{2}k^2 + \frac{1}{3!}k^3 + \cdots$$

is also positive definite.

# Review on Positive Definite Kernels II

## Proposition 2

*Let* $k : \mathcal{X} \times \mathcal{X} \to \mathbb{C}$ *be a positive definite kernel and* $f : \mathcal{X} \to \mathbb{C}$ *be an arbitrary function. Then,*

$$\tilde{k}(x, y) = f(x)k(x, y)\overline{f(y)}$$

*is positive definite. In particular,*

$$f(x)\overline{f(y)}$$

*is a positive definite kernel.*

Example. Normalization:

$$\tilde{k}(x, y) = \frac{k(x, y)}{\sqrt{k(x, x)k(y, y)}}$$

# Negative Definite Kernel

Definition. A function $\psi : \mathcal{X} \times \mathcal{X} \to \mathbb{C}$ is called a negative definite kernel if it is Hermitian i.e. $\psi(y, x) = \overline{\psi(x, y)}$, and

$$\sum_{i,j=1}^{n} c_i \overline{c_j} \psi(x_i, x_j) \leq 0$$

for any $x_1, \ldots, x_n$ ($n \geq 2$) in $\mathcal{X}$ and $c_1, \ldots, c_n \in \mathbb{C}$ with $\sum_{i=1}^{n} c_i = 0$.

Note: a negative definite kernel is not necessarily minus positive definite kernel, because we need the condition $\sum_{i=1}^{n} c_i = 0$.

# Properties of negative definite kernels

## Proposition 3

1. *If $k$ is positive definite, $\psi = -k$ is negative definite.*
2. *Constant functions are negative definite.*

Proof. (2) $\sum_{i,j=1}^{n} c_i c_j = \sum_{i=1}^{n} c_i \sum_{j=1}^{n} c_j = 0$.

## Proposition 4

*If $\psi_i : \mathcal{X} \times \mathcal{X} \to \mathbb{C}$ $(i = 1, 2, \ldots)$ are negative definite kernels, then so are the following:*

1. *(positive combination)* $a\psi_1 + b\psi_2 \quad (a, b \geq 0)$.
2. *(limit)* $lim_{i\to\infty} \psi_i(x, y)$, *assuming the limit exists.*

- The set of all negative definite kernels is a closed convex cone.
- Multiplication does not preserve negative definiteness.

# Example of Negative Definite Kernel

## Proposition 5

*Let $V$ be an inner product space, and $\phi : \mathcal{X} \to V$. Then,*

$$\psi(x, y) = \|\phi(x) - \phi(y)\|^2$$

*is a negative definite kernel on $\mathcal{X}$.*

Proof. [Exercise]

# Relation Between Positive and Negative Definite Kernels

## Lemma 6

*Let $\psi(x, y)$ be a hermitian kernel on $\mathcal{X}$. Fix $x_0 \in \mathcal{X}$ and define*

$$\varphi(x, y) = -\psi(x, y) + \psi(x, x_0) + \psi(x_0, y) - \psi(x_0, x_0).$$

*Then, $\psi$ is negative definite if and only if $\varphi$ is positive definite.*

Proof. "If" part is easy (exercise). Suppose $\psi$ is neg. def. Take any $x_i \in \mathcal{X}$ and $c_i \in \mathbb{C}$ $(1 = 1, \ldots, n)$. Define $c_0 = -\sum_{i=1}^n c_i$. Then,

$$\begin{aligned} 0 &\geq \textstyle\sum_{i,j=0}^n c_i \overline{c_j} \psi(x_i, x_j) \qquad\qquad [\text{for } x_0, x_1, \ldots, x_n] \\ &= \textstyle\sum_{i,j=1}^n c_i \overline{c_j} \psi(x_i, x_j) + \overline{c_0} \sum_{i=1}^n c_i \psi(x_i, x_0) + c_0 \sum_{j=1}^n c_i \psi(x_0, x_j) \\ &\qquad + |c_0|^2 \psi(x_0, x_0) \\ &= \textstyle\sum_{i,j=1}^n c_i \overline{c_j} \{\psi(x_i, x_j) - \psi(x_i, x_0) - \psi(x_0, x_j) + \psi(x_0, y_0)\} \\ &= -\textstyle\sum_{i,j=1}^n c_i \overline{c_j} \varphi(x_i, x_j). \end{aligned}$$

# Schoenberg's Theorem

**Theorem 7 (Schoenberg's theorem)**

*Let $\mathcal{X}$ be a nonempty set, and $\psi : \mathcal{X} \times \mathcal{X} \to \mathbb{C}$ be a kernel. $\psi$ is negative definite if and only if $\exp(-t\psi)$ is positive definite for all $t > 0$.*

Proof.

If part:

$$\psi(x, y) = \lim_{t \downarrow 0} \frac{1 - \exp(-t\psi(x, y))}{t}.$$

Only if part: We can prove only for $t = 1$. Take $x_0 \in \mathcal{X}$ and define

$$\varphi(x, y) = -\psi(x, y) + \psi(x, x_0) + \psi(x_0, y) - \psi(x_0, x_0).$$

$\varphi$ is positive definite (Lemma 6).

$$e^{-\psi(x,y)} = e^{\varphi(x,y)} e^{-\psi(x,x_0)} \overline{e^{-\psi(y,x_0)}} e^{\psi(x_0,x_0)}.$$

This is also positive definite. □

# Generating New Kernels I

## Proposition 8

*If* $\psi : \mathcal{X} \times \mathcal{X} \to \mathbb{C}$ *is negative definite and* $\psi(x, x) \geq 0$*. Then, for any* $0 < p \leq 1$*,*

$$\psi(x, y)^p$$

*is negative definite.*

Proof. Use the formula

$$z^p = \frac{p}{\Gamma(1-p)} \int_0^\infty t^{-p-1}(1 - e^{-tz})dt \quad (z \in \mathbb{C}).$$

Then,

$$\psi(x, y)^p = \frac{p}{\Gamma(1-p)} \int_0^\infty t^{-p-1}(1 - e^{-t\psi(x,y)})dt$$

The integrand is negative definite for all $t > 0$. $\square$.

# Generating New Kernels II

- For any $0 \leq p \leq 2$,

$$\|x - y\|^p$$

  is negative definite on $\mathbb{R}^m$.

- For any $0 \leq p \leq 2$ and $\alpha > 0$,

$$\exp(-\alpha\|x - y\|^p)$$

  is positive definite on $\mathbb{R}^n$.

  - $\alpha = 2 \Rightarrow$ Gaussian kernel.
  - $\alpha = 1 \Rightarrow$ Laplacian kernel.

# Generating New Kernels III

## Proposition 9

*If $\psi : \mathcal{X} \times \mathcal{X} \to \mathbb{C}$ is negative definite and $\mathrm{Re}\psi(x,y) \geq 0$. Then, for any $a > 0$,*

$$\frac{1}{\psi(x,y)+a}$$

*is positive definite.*

Proof.

$$\frac{1}{\psi(x,y)+a} = \int_0^\infty e^{-t(\psi(x,y)+a)} dt.$$

The integrand is positive definite for all $t > 0$. □.

For any $0 < p \leq 2$,

$$\frac{1}{1+\|x-y\|^p}$$

is positive definite on $\mathbb{R}^m$. ($p = 2$: Cauchy kernel.)

# Positive definite functions

Definition. Let $\phi : \mathbb{R}^n \to \mathbb{C}$ be a function. $\phi$ is called a positive definite function (or function of positive type) if

$$k(x, y) = \phi(x - y)$$

is a positive definite kernel on $\mathbb{R}^n$, *i.e.*,

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

for any $x_1, \ldots, x_n \in \mathcal{X}$ and $c_1, \ldots, c_n \in \mathbb{C}$.

- A positive definite kernel of the form $\phi(x - y)$ is called shift invariant (or translation invariant).
- Examples: Gaussian and Laplacian kernels.

# Bochner's theorem I

The Bochner's theorem characterizes *all* the continuous shift-invariant kernels on $\mathbb{R}^n$.

**Theorem 10 (Bochner)**

*Let $\phi$ be a continuous function on $\mathbb{R}^n$. Then, $\phi$ is positive definite if and only if there is a finite non-negative Borel measure $\Lambda$ on $R^n$ such that*

$$\phi(x) = \int e^{\sqrt{-1}\omega^T x} d\Lambda(\omega).$$

- $\phi$ is the inverse Fourier (or Fourier-Stieltjes) transform of $\Lambda$.
- Roughly speaking, the shift invariant functions are the class that have non-negative Fourier transform.

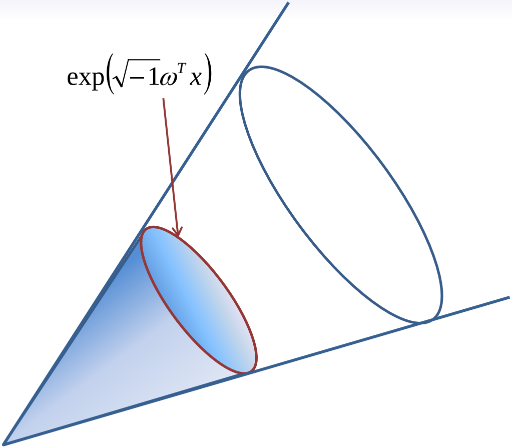
# Bochner's Theorem II

- The Fourier kernel $e^{\sqrt{-1}x^T\omega}$ is a positive definite function for every $\omega \in \mathbb{R}^n$.

$$\exp(\sqrt{-1}(x-y)^T\omega) = \exp(\sqrt{-1}x^T\omega)\overline{\exp(\sqrt{-1}y^T\omega)}.$$

- The set of all positive definite functions is a convex cone, which is closed under the pointwise-convergence topology.

- The generator of the convex cone is the Fourier kernels $\{e^{\sqrt{-1}x^T\omega} \mid \omega \in \mathbb{R}^n\}$.




The closed cone of positive definite functions.

# Bochner's theorem III

- Example on $\mathbb{R}$: (positive constants are neglected)

| p.d. function | Fourier transform |
|---|---|
| $\exp(-\frac{1}{2\sigma^2}x^2)$ | $\exp(-\frac{\sigma^2}{2}\lvert\omega\rvert^2)$ |
| $\exp(-\alpha\lvert x\rvert)$ | $\frac{1}{\omega^2+\alpha^2}$ |
| $\frac{1}{x^2+\alpha^2}$ | $\exp(-\alpha\lvert\omega\rvert)$ |

- Bochner's theorem can be extended to toplogical groups and semigroups [BCR84].

# Explicit Realization of RKHS by Bochner's Theorem I

Assume in the Bochner's theorem $d\Lambda = \rho(\omega)d\omega$, *i.e.*,

$$k(x,y) = \int e^{\sqrt{-1}\omega^T(x-y)}\rho(\omega)d\omega,$$

$\rho(\omega)$ is continuous for every $\omega$, and $\int \rho(\omega)d\omega < \infty$.
(*e.g.* Gaussian, Laplacian, Cauchy.)

Then, the RKHS $\mathcal{H}_k$ is given by[1]

$$\mathcal{H}_k = \left\{ f \in L^2(\mathbb{R}^m, dx) \;\middle|\; \int \frac{|\hat{f}(t)|^2}{\rho(t)} dt < \infty \right\},$$

$$\langle f, g\rangle_{\mathcal{H}_k} = \int \frac{\hat{f}(t)\overline{\hat{g}(t)}}{\rho(t)} dt.$$

[1] $\hat{f}$ denotes the Fourier transform defined by $\hat{f} = \frac{1}{(2\pi)^m}\int e^{-\sqrt{1}\omega^T x} f(x)dx$.

# Explicit Realization of RKHS by Bochner's Theorem II

- $\mathcal{H}_k$ is a Hilbert space consisting of functions on $\mathbb{R}^m$.
- $\langle \cdot, \cdot \rangle_{\mathcal{H}_k}$ defines an inner product on $\mathcal{H}_k$.
- $k(\cdot, x)$ is the reproducing kernel of $\mathcal{H}_k$:

  Proof. From

$$k(x, y) = \int e^{\sqrt{-1}\omega^T (x-y)} \rho(\omega) d\omega = \int e^{\sqrt{-1}\omega^T x} e^{-\sqrt{-1}\omega^T y} \rho(\omega) d\omega,$$

  the Fourier transform of $k(\cdot, y)$ ($y$ fixed) is given by

$$\widehat{k(\cdot, y)}(\omega) = e^{-\sqrt{-1}\omega^T y} \rho(\omega).$$

  Thus,

$$\langle f, k(\cdot, y) \rangle_{\mathcal{H}_k} = \int \frac{\hat{f}(\omega) e^{\sqrt{-1}\omega^T y} \rho(\omega)}{\rho(\omega)} d\omega = \int \hat{f}(\omega) e^{\sqrt{-1}\omega^T y} d\omega = f(y).$$

# Examples

- Gaussian RBF kernel: $\rho(t) = \frac{1}{2\pi} \exp\{-\frac{\sigma^2}{2}\omega^2\}$,

$$\mathcal{H}_k = \left\{ f \in L^2(\mathbb{R}, dx) \;\middle|\; \int |\hat{f}(\omega)|^2 \exp\bigl(\frac{\sigma^2}{2}\omega^2\bigr) d\omega < \infty \right\},$$

$$\langle f, g \rangle_{\mathcal{H}_k} = \int \hat{f}(\omega)\overline{\hat{g}(\omega)} \exp\bigl(\frac{\sigma^2}{2}\omega^2\bigr) d\omega$$

- Laplacian kernel: $\rho(\omega) = \frac{1}{2\pi}\frac{1}{\omega^2+\beta^2}$,

$$\mathcal{H}_k = \left\{ f \in L^2(\mathbb{R}, dx) \;\middle|\; \int |\hat{f}(\omega)|^2 (\omega^2 + \beta^2) dt < \infty \right\},$$

$$\langle f, g \rangle = \int \hat{f}(\omega)\overline{\hat{g}(\omega)} (\omega^2 + \beta^2) d\omega.$$

- Cauchy kernel: $\rho(\omega) = \frac{1}{2\pi} e^{-\alpha|\omega|}$,

$$\mathcal{H}_k = \Big\{ f \in L^2(\mathbb{R}, dx) \,\Big|\, \int |\hat{f}(\omega)|^2 e^{\alpha|\omega|} d\omega < \infty \Big\},$$

$$\langle f, g \rangle_{\mathcal{H}_k} = \int \hat{f}(\omega) \overline{\hat{g}(\omega)} e^{\alpha|\omega|} d\omega.$$

- Note in the above three examples the RKHS's admits different decay rates of frequency.

# Complete Orthonormal System

- A subset $\{u_i\}_{i \in I}$ of $\mathcal{H}$ is called an orthonormal system (ONS) if $(u_i, u_j) = \delta_{ij}$ ($\delta_{ij}$ is Kronecker's delta).
- A subset $\{u_i\}_{i \in I}$ of $\mathcal{H}$ is called a complete orthonormal system (CONS) (orthonormal basis) if it is ONS and if $(x, u_i) = 0$ ($\forall i \in I$) implies $x = 0$.
- A Hilbert space is called separable if it has a countable CONS.

# Fourier Expansion

**Theorem 11 (Fourier series expansion)**

*Let $\{u_i\}_{i=1}^{\infty}$ be a CONS of a separable Hilbert space. For each $x \in \mathcal{H}$,*

$$x = \sum_{i=1}^{\infty}(x, u_i)u_i, \qquad \textit{(Fourier expansion)}$$

$$\|x\|^2 = \sum_{i=1}^{\infty}|(x, u_i)|^2. \qquad \textit{(Parseval's equality)}$$

Proof omitted.

Example: CONS of $L^2([0\ 2\pi], dx)$

$$u_n(t) = \frac{1}{\sqrt{2\pi}}e^{\sqrt{-1}nt} \qquad (n = 0, 1, 2, \dots)$$

Then,

$$f(t) = \frac{1}{\sqrt{2\pi}}\sum_{n=-\infty}^{\infty}a_n e^{\sqrt{-1}nt}$$

is the (ordinary) Fourier expansion of a periodic function.

# Bounded Operator I

Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be Hilbert spaces. A linear transform $T : \mathcal{H}_1 \to \mathcal{H}_2$ is often called operator.

Definition. A linear operator $\mathcal{H}_1$ and $\mathcal{H}_2$ is called bounded if

$$\sup_{\|x\|_{\mathcal{H}_1}=1} \|Tx\|_{\mathcal{H}_2} < \infty.$$

The operator norm of a bounded operator $T$ is defined by

$$\|T\| = \sup_{\|x\|_{\mathcal{H}_1}=1} \|Tx\|_{\mathcal{H}_2} = \sup_{x \neq 0} \frac{\|Tx\|_{\mathcal{H}_2}}{\|x\|_{\mathcal{H}_1}}.$$

(Corresponds to the largest singular value of a matrix.)

Fact. If $T : \mathcal{H}_1 \to \mathcal{H}_2$ is bounded,

$$\|Tx\|_{\mathcal{H}_2} \leq \|T\| \|x\|_{\mathcal{H}_1}.$$

# Bounded Operator II

## Proposition 12

*A linear operator is bounded if and only if it is continuous.*

Proof. Assume $T : \mathcal{H}_1 \to \mathcal{H}_2$ is bounded. Then,

$$\|Tx - Tx_0\| = \|T(x - x_0)\| \leq \|T\|\|x - x_0\|$$

means continuity of $T$.

Assume $T$ is continuous. For any $\varepsilon > 0$, there is $\delta > 0$ such that $\|Tx\| < \varepsilon$ for all $x \in \mathcal{H}_1$ with $\|x\| < 2\delta$.
Then,

$$\sup_{\|x\|=1} \|Tx\| = \sup_{\|x\|=\delta} \frac{1}{\delta}\|Tx\| \leq \frac{\varepsilon}{\delta}.$$

□

# Hilbert-Schmidt Operator I

$\mathcal{H}$: separable Hilbert space.

Definition. An operator $T$ on $\mathcal{H}$ is called Hilbert-Schmidt if for a CONS $\{\varphi_i\}_{i=1}^{\infty}$

$$\sum_{i=1}^{\infty} \|T\varphi_i\|^2 < \infty,$$

and its Hilbert-Schmidt norm $\|T\|_{HS}$ is defined by

$$\|T\|_{HS} = \left(\sum_{i=1}^{\infty} \|T\varphi_i\|^2\right)^{1/2}.$$

- $\|T\|_{HS}$ does not depend on the choice of a CONS.

  $\because$) From Parseval's equality, for a CONS $\{\psi_j\}_{j=1}^{\infty}$,

$$\|T\|_{HS}^2 = \sum_{i=1}^{\infty} \|T\varphi_i\|^2 = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} |(\psi_j, T\varphi_i)|^2$$
$$= \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} |(T^*\psi_j, \varphi_i)|^2 = \sum_{j=1}^{\infty} \|T^*\psi_j\|^2.$$

# Hilbert-Schmidt Operator II

- Fact: $\|T\| \leq \|T\|_{HS}$.

  (∵) Let $u_1$ be the unit vector such that $\|Tu_1\| \geq \|T\| - \varepsilon$. Make CONS including $u_1$ and compute $\|T\|_{HS}^2$. )

- Hilbert-Schmidt norm is an extension of Frobenius norm of a matrix:

$$\|T\|_{HS}^2 = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} |(\psi_j, T\varphi_i)|^2.$$

  $(\psi_j, T\varphi_i)$ is the component of the matrix expression of $T$ with the CONS's $\{\varphi_i\}$ and $\{\psi_j\}$.

# Integral Kernel

$(\Omega, \mathcal{B}, \mu)$: measure space.

$K(x, y)$: measurable function on $\Omega \times \Omega$ such that

$$\int_\Omega \int_\Omega |K(x,y)|^2 d\mu(x) d\mu(y) < \infty. \qquad \text{(square integrability)}$$

Def. Operator $T_K : L^2(\Omega, \mu) \to L^2(\Omega, \mu)$ by

$$(T_K f)(x) = \int_\Omega K(x,y) f(y) d\mu(y) \qquad (f \in L^2(\Omega, \mu)).$$

$T_K$: integral operator with integral kernel $K$.

Fact: $T_K f \in L^2(\Omega, \mu)$.

$$\begin{aligned} \because) \quad \int |T_K f(x)|^2 dx &= \int \left| \int K(x,y) f(y) d\mu(y) \right|^2 d\mu(x) \\ &\leq \int \int |K(x,y)|^2 d\mu(y) \int |f(y)|^2 d\mu(y) d\mu(x) \\ &= \int \int |K(x,y)|^2 d\mu(x) d\mu(y) \|f\|_{L^2}^2. \end{aligned}$$

# Hilbert-Schmidt Operator and Integral Kernel I

## Theorem 13

*Let $(\Omega, \mathcal{B}, \mu)$ be a measure space, and assume $L^2(\Omega, \mu)$ is separable. Then, $T_K$ is a Hilbert-Schmidt operator, and*

$$\|T_K\|_{HS}^2 = \int\int_{\Omega\times\Omega} |K(x,y)|^2 d\mu(x)d\mu(y).$$

Proof. Let $\{\varphi_i\}$ be a CONS. From Parseval's equality,

$$\int |K(x,y)|^2 dy = \sum_i \left|(K(x,\cdot), \varphi_i)_{L^2}\right|^2 = \sum_i \left|\int K(x,y)\overline{\varphi_i(y)}dy\right|^2 = \sum_i |T_K\overline{\varphi_i}(x)|^2.$$

Integrate w.r.t. $x$, ($\{\overline{\varphi_i}\}$ is also a CONS)

$$\int\int |K(x,y)|^2 dxdy = \sum_i \|T_K\overline{\varphi_i}\|^2 = \|T_K\|_{HS}^2.$$

□

# Hilbert-Schmidt Operator and Integral Kernel II

Converse is also true!

**Theorem 14**

*Let $(\Omega, \mathcal{B}, \mu)$ be a measure space, and assume $L^2(\Omega, \mu)$ is separable. For any Hilbert-Schmidt operator $T$ on $L^2(\Omega, \mu)$, there is a square integrable kernel $K(x, y)$ such that*

$$T\varphi = \int K(x,y)\varphi(y)d\mu(y) \;\; \big(= T_K\varphi\big).$$

Proof omitted.

# Hilbert-Schmidt Expansion I

- $(\Omega, \mathcal{B}, \mu)$: measure space.
- $K(x, y)$: Hermitian ($K(y, x) = \overline{K(x, y)}$) square integrable kernel.

- Fact: $T_K$ is self-adjoint, *i.e.*,

$$(T_K f, g) = (f, T_K g) \qquad (\forall f, g \in L^2(\Omega, \mu)).$$

Proof.

$$(T_K f, g) = \int\int K(x, y) f(y) \overline{g(x)} d\mu(x) d\mu(Y)$$
$$= \int f(y) \overline{\int K(y, x) g(x) d\mu(x)} d\mu(y) = (f, T_K g).$$

- For Hermite kernels, $T_K$ admits eigendecompsition in an analogous way to Hermitian (or symmetric) matrices.

# Hilbert-Schmidt Expansion II

A self-adjoint Hilbert-Schmidt operator admits Hilbert-Schmidt expansion:

- Every eigenvalue of $T_K$ is a real value.
- The eigenspace of each eigenvalue is finite dimensional.
- Let

$$|\lambda_1| \geq |\lambda_2| \geq \cdots > 0$$

be the non-zero eigenvalues (counted as multiplicity).

- Let $\phi_i$ be the unit eigenvector w.r.t. $\lambda_i$.
- Hilbert-Schmidt expansion

$$T_K f = \sum_{i=1}^{\infty} \lambda_i (f, \phi_i) \phi_i.$$

# Hilbert-Schmidt Expansion III

## Theorem 15

*Let $K$ be a Hermitian square integrable kernel, and $\lambda_i, \phi_i$ as above.*

$$K(x,y) = \sum_{i=1}^{\infty} \lambda_i \phi_i(x) \overline{\phi_i(y)}$$

*in $L^2(\Omega \times \Omega, \mu \times \mu)$.*

(Proof omitted.) This is a generalization of eigendecomposition.

*c.f.* $A$: $m \times m$ Hermitian (or symmetric) matrix.
$\{\lambda_i\}_{i=1}^m$: the eigenvalues of $A$. $u_i$: unit eigenvector w.r.t. $\lambda_i$.
Then,

$$A = \sum_{i=1}^{m} \lambda_i u_i u_i^*,$$

$$Av = \sum_{i=1}^{m} \lambda_i (v, u_i) u_i.$$

# Integral Kernel and Positive Definiteness I

Consider positive definite $K(x, y)$.

## Proposition 16 (Positive definiteness)

*Let $D$ be a compact subset of $\mathbb{R}^m$, and $K(x, y)$ be a continuous symmetric kernel on $\Omega \times \Omega$.*
*$K(x, y)$ is positive definite on $\Omega$ if and only if*

$$\int \int_{D \times D} K(x, y) f(x) \overline{f(y)} dx dy \geq 0$$

*for any $f \in L^2(D)$.*

*c.f.* Definition of positive definiteness:

$$\sum_{i,j} K(x_i, x_j) c_i \overline{c_j} \geq 0.$$

# Integral Kernel and Positive Definiteness II

Proof.

($\Rightarrow$). For a continuous function $f$, a Riemann sum satisfies

$$\sum_{i,j} K(x_i, x_j) f(x_i) \overline{f(x_j)} |E_i||E_j| \geq 0.$$

The integral is the limit of such sums, thus non-negative. For $f \in L^2(\Omega, \mu)$, approximate it by a continuous function.

($\Leftarrow$). Omitted. See [Fuk10, Sec. 6.3]

# Integral Operator by Positive Definite Kernel

$D$: compact subset of $\mathbb{R}^m$.

$K(x,y)$: continuous positive definite kernel on $D$.

$$(T_K f)(x) = \int_D K(x,y) f(y) dy \qquad (f \in L^2(D))$$

Fact: Recall from Proposition 16

$$(T_K f, f)_{L^2(D)} \geq 0 \qquad (\forall f \in L^2(D)).$$

In particular, every eigenvalue of $T_K$ is non-negative.

# Mercer's Theorem

$\{\lambda_i\}_{i=1}^{\infty}$ $(\lambda_1 \geq \lambda_2 \geq \cdots > 0)$, $\{\phi_i\}_{i=1}^{\infty}$: the positive eigenvalues and unit eigenfunctions of $T_K$.

From Hilbert-Schmidt expansion,

$$K(x,y) = \textstyle\sum_{i=1}^{\infty} \lambda_i \phi_i(x) \overline{\phi_i(y)},$$

in $L^2(D \times D)$.

**Theorem 17 (Mercer)**

*Let $K$ be a continuous positive definite kernel on a compact subset $D$ in $\mathbb{R}^m$. Then,*

$$K(x,y) = \sum_{i=1}^{\infty} \lambda_i \phi_i(x) \overline{\phi_i(y)},$$

*where the convergence is absolute and uniform over $D \times D$.*

Proof is omitted. See [RSN65], Sec. 98, or [Ito78], Chap. 13.

# Explicit Expression of RKHS

Let $K(x, y)$ be a continuous positive definite kernel on a compact subset $D$ in $\mathbb{R}^m$.

$\{\lambda_i\}_{i=1}^{\infty}$ ($\lambda_1 \geq \lambda_2 \geq \cdots > 0$), $\{\phi_i\}_{i=1}^{\infty}$: the positive eigenvalues and unit eigenfunctions of $T_K$.

By adding the northonormal basis of $\mathcal{N}(T_K)$, we have a CONS $\{\phi_i\}$ of $\mathcal{H}_K$ consisting of eigenvectors of $T_K$.

## Theorem 18

$$\mathcal{H}_k = \left\{ f \in L^2(D) \;\middle|\; f = \sum_{i=1}^{\infty} a_i \phi_i, \sum_{i=1}^{\infty} \frac{|a_i|^2}{\lambda_i} < \infty \right\},$$

*and for* $f = \sum_{i=1}^{\infty} a_i \phi_i$ *and* $g = \sum_{i=1}^{\infty} b_i \phi_i$,

$$\langle f, g \rangle_{\mathcal{H}_k} = \sum_{i=1}^{\infty} \frac{a_i \overline{b_i}}{\lambda_i},$$

*where* $a_i$ *and* $b_i$ *are set 0 if* $\lambda_i = 0$.

- It is not difficult to show $\mathcal{H}_k$ is a Hilbert space.
- Reproducing property:
  First note that by Mercer's theorem,

  $$\sum_{i=1}^{\infty} \lambda_i |\phi_i(x)|^2 < \infty,$$

  which means $K(\cdot, x) = \sum_{i=1}^{\infty} \lambda_i \phi_i(\cdot) \overline{\phi_i(x)} \in \mathcal{H}_k$.
  For arbitrary $f = \sum_{i=1} a_i \phi_i \in \mathcal{H}_k$

  $$\langle f, K(\cdot, x) \rangle = \sum_{i=1}^{\infty} \frac{a_i \lambda_i \phi_i(x)}{\lambda_i} = \sum_{i=1}^{\infty} a_i \phi_i(x) = f(x).$$

- *C.f.*, RKHS on a finite set.

# Restriction of RKHS

$k$: positive definite kernel on a set $\mathcal{X}$. $\mathcal{H}_k$: corresponding RKHS.
$\mathcal{Y}$: subset of $\mathcal{X}$.

$\tilde{k}$: restriction of $k$ to $\mathcal{Y} \times \mathcal{Y}$ $\Rightarrow$ positive definite kernel on $\mathcal{Y}$.

**Theorem 19**

*The RKHS corresponding to $\tilde{k}$ is $\{f|_{\mathcal{Y}} \mid f \in \mathcal{H}_k\}$.*

# Sum of RKHS

$k_1, k_2$: positive definite kernels on a set $\mathcal{X}$.
$\mathcal{H}_1, \mathcal{H}_2$: corresponding RKHS's.

$k_1 + k_2$: positive definite kernel on $\mathcal{X}$.

**Theorem 20**

*The RKHS corresponding to $k_1 + k_2$ is given by*

$$\mathcal{H} = \{f_1 + f_2 : \mathcal{X} \to \mathbb{R} \mid f_1 \in \mathcal{H}_1, f_2 \in \mathcal{H}_2\},$$

*and its norm is given by*

$$\|f\|_{\mathcal{H}}^2 = \min\{\|f_1\|_{\mathcal{H}_1}^2 + \|f_2\|_{\mathcal{H}_2}^2 \mid f = f_1 + f_2, f_1 \in \mathcal{H}_1, f_2 \in \mathcal{H}_2\}.$$

# Product of RKSH

$k_1, k_2$: positive definite kernels on set $\mathcal{X}_1, \mathcal{X}_2$, resp.
$\mathcal{H}_1, \mathcal{H}_2$: corresponding RKHS's.

$k((x_1, x_2), (y_1, y_2)) := k_1(x_1, y_1)k_2(x_2, y_2)$: positive definite kernel on $\mathcal{X}_1 \times \mathcal{X}_2$.

## Theorem 21

*The RKHS corresponding to $k$ is the tensor product $\mathcal{H}_1 \otimes \mathcal{H}_2$.*

## Corollary 22

*If $k_1$ and $k_2$ are positive definite kernels on $\mathcal{X}$, the RKHS corresponding to $k(x, y) = k_1(x, y)k_2(x, y)$ is the restriction of $\mathcal{H}_1 \otimes \mathcal{H}_2$ to the diagonal set $\{(x, x) \in \mathcal{X} \times \mathcal{X} \mid x \in \mathcal{X}\}$.*

# Tensor Product

Define inner product space $\mathcal{H}_1 \tilde{\otimes} \mathcal{H}_2$ by

$$\mathcal{H}_1 \tilde{\otimes} \mathcal{H}_2 := \left\{ \sum_{i=1}^{n} f_i^{(1)} \otimes f_i^{(2)} \;\middle|\; f_i^{(1)} \in \mathcal{H}_1, f_i^{(2)} \in \mathcal{H}_2, i = 1, \ldots, n \right\}.$$

$$\left\langle \sum_{i=1}^{n} f_i^{(1)} \otimes f_i^{(2)}, \sum_{j=1}^{m} g_j^{(1)} \otimes g_j^{(2)} \right\rangle := \sum_{i=1}^{n} \sum_{j=1}^{m} \langle f_i^{(1)}, g_j^{(1)} \rangle_{\mathcal{H}_1} \langle f_i^{(2)}, g_j^{(2)} \rangle_{\mathcal{H}_2}.$$

The tensor product $\mathcal{H}_1 \otimes \mathcal{H}_2$ is the completion of $\mathcal{H}_1 \tilde{\otimes} \mathcal{H}_2$.

# Summary of Section 5

- Negative definite kernels and positive definite kernels are related by Schoenberg's theorem.
- Various examples of positive definite kernels can be derived by functional operations.
- Bochner's theorem: characterization of continuous shift-invariant kernels on $\mathbb{R}^m$ by Fourier transform.
- Based on Bochner's theorem, RKHS for shift-invariant kernels can be written explicitly by Fourier transform.
- Mercer's theorem: eigendecomposition of positive definite kernel.

# References

[BCR84] Christian Berg, Jens Peter Reus Christensen, and Paul Ressel.
*Harmonic Analysis on Semigroups*.
Springer-Verlag, 1984.

[Fuk10] Kenji Fukumizu.
*Introduction to Kernel Method (in Japanese)*.
Asakura Shoten, 2010.

[Ito78] Seizo Ito.
*Kansu-Kaiseki III (Iwanami kouza Kiso-suugaku)*.
Iwanami Shoten, 1978.

[RSN65] Frigyes Riesz and Béla Sz.-Nagy.
*Functional Analysis (2nd ed.)*.
Frederick Ungar Publishing Co, 1965.