· We can build up new kernels from old kernels for example through some transformations.

Theorem. (Sums of kernels are kernels)
Given $\alpha > 0$, and k, k_1, k_2 are kernels on Ω, then αk, $k_1 + k_2$ are kernels on Ω.

Proof. Note from the Remark after Theorem 2.2.3, kernels are positive definite. Consider two kernels $k_1(x, x')$ and $k_2(x, x')$. Then

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \left[k_1(x_i, x_j) + k_2(x_i, x_j) \right]$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j k_1(x_i, x_j) + \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j k_2(x_i, x_j)$$

$$\geq 0.$$

In a similar way we can show that αk is a kernel on Ω. ※

Theorem (Products of kernels are kernels)
Given k_1 on Ω_1 and k_2 on Ω_2, then $k_1 \times k_2$ is a kernel on $\Omega_1 \times \Omega_2$. If $\Omega_1 = \Omega_2 = \Omega$, then $k = k_1 \times k_2$ is a kernel on Ω.

Proof, for a simpler case where $\mathcal{H}_1$ corresponding to k_1 is $\mathbb{R}^m$, and $\mathcal{H}_2$ corresponding to k_2 is $\mathbb{R}^n$.
For a general proof which uses tensor product

in [Ingo Steinwart and Andreas Christmann, Support Vector Machines, Information Science and Statistics, Springer 2008). We will use linear algebra to prove the theorem.

Define $k_1 := u^T v$ for $u, v \in \mathbb{R}^m$ and $k_2 := p^T q$ for $p, q \in \mathbb{R}^n$. We know that the inner product between matrices $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{m \times n}$ is

$$\langle A, B \rangle = \mathrm{trace}(A^T B) \qquad (*)$$

Then
$$\begin{aligned} k_1 k_2 &= k_1 (p^T q) \\ &= k_1\, \mathrm{trace}(q p^T) \\ &= \mathrm{trace}(q \underbrace{u^T v}_{k_1} p^T) \\ &= \langle A, B \rangle \end{aligned}$$

when $A := u q^T$, $B = v p^T$

In other words the product $k_1 k_2$ defines a valid inner product as stated in $(*)$. ✗.

polynomial kernel: Let $x, y \in \mathbb{R}^m$ for $m \geq 1$, and let $d \geq 1$ be an integer and $c \geq 0$ be a positive real. Then

$$k(x, y) := (\langle x, y \rangle + c)^d$$

is a kernel.

Proof: If we expand the sum, we obtain the ~~non~~negative combination of kernels $\langle x, y \rangle$ raised to integer powers. which are kernels by product rule. Thus it is a kernel using the sum rule. ✗.

In particular for $d=2$ we have

$$k(x,y) = \langle x,y\rangle^2 + 2c\langle x,y\rangle + c^2$$
$$= \left(\sum_{i=1}^{m} x_i y_i\right)^2 + 2c\sum_{i=1}^{m} x_i y_i + c^2$$
$$= \cdots$$

(Your text)

From this ...

$\phi(x) = \cdots$:

Note that ~~in~~ the case of $d=1$ is a linear kernel.

(Harmonic analysis on semigroups. Springer, 1984)

[Christian Berg, Jens Peter Reus Christensen, Paul Ressel].

(give a reference e.g. wikipedia ...)

Corallary (to the theorem of product kernels:)

Let φ be a positive definite on $\Omega\times\Omega$ $(\to \mathbb{C})$ such that $|\varphi(x,y)|<\rho$ for all $(x,y)\in\Omega\times\Omega$. Then if $f(z)=\sum_{n=0}^{\infty} a_n z^n$ is holomorphic in $\{z\in\mathbb{C} : |z|<\rho\}$ and $a_n\geq 0$ for all $n\geq 0$ then composed kernel $f\circ\varphi$ is positive definite.

In particular, if φ is positive definite, then so is $\exp(\varphi)$.

<u>Proof</u>. By the above theorem, for each $n\geq$, the kernel φ^n is positive definite; thus $\sum_{n=0}^{N} a_n\varphi^n$ is positive definite for all $N\geq 0$ and so is its pointwise limit $f\circ\varphi$. ✗.

The Gaussian kernel: $K_\sigma(x,y) = e^{-\frac{\|x-y\|^2}{2\sigma^2}}$

$$= \left[e^{-\frac{\|x\|^2}{2\sigma^2}} \cdot e^{-\frac{\|y\|^2}{2\sigma^2}} \right] \cdot \left[\sum_{i=0}^{\infty} \frac{(x\cdot y)^i}{i!} \right]$$

kernel by the corollary

By the product rule $K_\sigma(x,y)$ is a kernel.

— × —

Exponential kernel: $K(x,y) = \exp\left(-\frac{\|x-y\|}{2\sigma^2}\right)$

~~Here~~ We postpone the proof.

Let us turn to a discussion on explicit realization of RKHS. First reconsider the linear kernel, we would like to know ~~how~~ what RKHS of linear kernel is.

The RKHS of the linear kernel:

$$K(x,y) = \langle x, y \rangle_{\mathbb{R}^d}$$

<u>Fact.</u> ~~Theorem~~: The RKHS of the linear kernel is the set of linear functions of the form

$$f_w(x) = \langle w, x \rangle_{\mathbb{R}^d} \quad \text{for } w \in \mathbb{R}^d$$

endowed with the norm

$$\|f\|_{\mathcal{H}} = \|w\|_2$$

<u>Proof</u>. The RKHS of the linear kernel consists of functions

$$x \in \mathbb{R}^d \longmapsto f(x) = \sum_i a_i \langle x_i, x \rangle_{\mathbb{R}^d} = \langle w, x \rangle_{\mathbb{R}^d}$$

with $w = \sum_i a_i x_i$

Hence the RKHS is the set of linear forms endowed with the following inner product:

$$\langle f, g \rangle_{\mathcal{H}_K} = \langle w, v \rangle_{\mathbb{R}^d}$$

when $f(x) = w \cdot x$ and $g(x) = v \cdot x$.

So the linear K, $K(x,y) = x^T y$

$$f(x) = w^T x$$

$$\|f\|_{\mathcal{H}} = \|w\|_2 . \#$$

· Next we discuss the RKHS of polynomial kernel.

For simplicity we study polynomial kernel on $\mathbb{R}$:

$$k(x,y) = (xy+c)^d \quad (c>0,\ d\in\mathbb{N})$$

Fact: Let H_k be the RKHS corresponding the polynomial kernel k. Then H_k is $d+1$ dimensional vector space with a basis $\{1, x, x^2, \cdots, x^d\}$.

Proof. Let $G = \text{span}\{1, x, \cdots, x^d\}$.

Since $k(x,z) = z^d x^d + \binom{d}{1} c\, z^{d-1} x^{d-1} + \binom{d}{2} c^2 z^{d-2} x^{d-2} + \cdots$

$$+ \binom{d}{d-1} c^{d-1} z x + c^d$$

we can conclude that

$$\text{span}\{k(\cdot, z) : z \in \mathbb{R}^m\} \subset G.$$

On the other hand, any polynomial of degree d lies in H_k, which follows from the following fact: for any $(a_0, \cdots, a_d)$ the linear equation

$$\begin{pmatrix} z_0^d & \cdots & z_0 & 1 \\ z_1^d & \cdots & z_1 & 1 \\ \vdots & & & \\ z_d^d & \cdots & z_d & 1 \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_d \end{pmatrix} = \begin{pmatrix} \frac{a_0}{c^d} \\ \frac{a_1}{c^{d-1}\binom{d}{d-1}} \\ \vdots \\ a_d \end{pmatrix}$$

is solvable (where $z_0, \cdots, z_d$ are zeroes of $k(x,z)$)

Then $\sum_{i=0}^{d} b_i k(x, z_i) = \sum_{i=0}^{d} a_i x^i$ #

~~To prove the~~

As shown the RKHS is a Hilbert subspace.

Recall that Ω is a set, $\mathbb{C}^{\Omega}$ is a set of all functions on Ω with pointwise topology. Let $\mathcal{G} = L^2(\cancel{X}T, \mu)$ where $(\cancel{X}T, \mathcal{B}, \mu)$ is a measurable space. Assume

$$H(\cdot\,; x) \in L^2(\cancel{X}T, \mu) \quad \text{for all } x \in \Omega.$$

~~Next~~ Construct a continuous embedding

$$j : L^2(T, \mu) \to \mathbb{C}^{\Omega}$$

by $\quad F \longmapsto f(x) = \int F(t)\,\overline{H(t;x)}\,d\mu(t) = (F, H(\cdot,x))_{\mathcal{G}}.$

Clearly j is an injection if $\operatorname{span}\{H(t;x) : x \in \Omega\}$ is dense in $L^2(T,\mu)$. Let the image of j be $\mathcal{H}$ with an inner product

$$\langle f, g\rangle_{\mathcal{H}} := (F, G)_{\mathcal{G}} \quad \text{where } f = j(F),\ g = j(G)$$

Then we have the isomorphism (given by j) between $L^2(T,\mu)$ and $\mathcal{H}$. so $\mathcal{H}$ is a Hilbert space and

$$\mathcal{H} = \{ f \in \mathbb{C}^{\Omega} : \exists\, F \in L^2(T,\mu),\ f(x) = \int F(t)\,\overline{H(t;x)}\,d\mu(t) \}.$$

Since $\quad f(x) = (F, H(\cdot,x))_{\mathcal{G}} = \langle f, j(H(\cdot,x))\rangle_{\mathcal{H}}$

we have the following proposition:

<u>Proposition</u>: $\mathcal{H}$ is a RKHS and its reproducing kernel is

$$k(x,y) = \langle j(H(\cdot\,;x)), j(H(\cdot\,;y))\rangle_{\mathcal{H}} = \int H(t;x)\,\overline{H(t;y)}\,d\mu(t).$$

· Consider now $\Omega = J = \mathbb{R}$, $\mathcal{G} = L^2(\mathbb{R}, \rho(t)dt)$ with the properties that $\rho(t)$ is continuous, $\rho(t) > 0$ and $\int \rho(t)\, dt < \infty$

Take $H(t;x) = e^{-\sqrt{-1}\,xt}$. In this case, $\mathrm{Span}\{H(t;x) : x \in \Omega\}$ is dense in $L^2(\mathbb{R}, \rho(t)dt)$.

<u>Fact</u>: $\mathcal{H} = \{ f \in L^2(\mathbb{R}, dx) : \int \frac{\hat{f}(t)}{\rho(t)} dt < \infty \}$

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

$$k(x,y) = \int e^{-\sqrt{-1}(x-y)t} \rho(t)\, dt \quad \text{(a positive defiute kernel)}$$

<u>Proof</u>: Let $f = j(F)$. By definition, the Fourier transform

$$f(x) = \int F(t) \cdot e^{\sqrt{-1}\, tx} \rho(t)\, dt$$

Because $\int \rho(t)\, dt < \infty$, $F \in L^2(\mathbb{R}, \rho(t)dt)$ means $|F(t)|^2 \rho(t)^2 \in L^1(\mathbb{R}, dt)$

Then $F(t)\rho(t) \in L^1(\mathbb{R}, dt) \cap L^2(\mathbb{R}, dt)$. So the Fourier isometry of $L^2(\mathbb{R}, dt)$ implies

$$f(x) \in L^2(\mathbb{R}, dx) \text{ and } \hat{f}(t) = \frac{1}{2\pi} \int f(x) e^{-\sqrt{-1}\, tx} dx = F(t)\rho(t)$$

which gives $F(t) = \dfrac{\hat{f}(t)}{\rho(t)}$.

Then

$$\langle f, g \rangle_{\mathcal{H}} = \langle j(F), j(G) \rangle_{\mathcal{H}} = (F, G)_{\mathcal{G}}$$

$$= \int \frac{\hat{f}(t)}{\rho(t)} \overline{\frac{\hat{g}(t)}{\rho(t)}} \rho(t)\, dt = \int \frac{\hat{f}(t)\,\overline{\hat{g}(t)}}{\rho(t)} dt$$

Moreover $F \in L^2(\mathbb{R}, \rho(t)dt) \iff \dfrac{\hat{f}(t)}{\rho(t)} \in L^2(\mathbb{R}, \rho(t)dt)$

$$\iff \int \frac{|\hat{f}(t)|^2}{\rho(t)} dt < \infty. \;\#$$

Example (Gaussian RBF kernel, revisited):

$$k(x,y) = \exp\left(-\frac{1}{2\sigma^2}|x-y|^2\right)$$

Let $\rho(t) = \frac{1}{2\pi}\exp\left(-\frac{\sigma^2}{2}t^2\right)$. i.e. $\mathcal{G} = L^2\left(\mathbb{R}; \frac{1}{2\pi}e^{-\frac{\sigma^2}{2}t^2}dt\right)$

So the reproducing kernel (= Gaussian RBF kernel) is

$$k(x,y) = \frac{1}{2\pi}\int e^{\sqrt{-1}(x-y)t}\, e^{-\frac{\sigma^2}{2}t^2}dt = \frac{1}{\sigma}\exp\left(-\frac{1}{2\sigma^2}(x-y)^2\right)$$

and $\underbrace{\mathcal{H}}_{\text{the RKHS}} = \{ f \in L^2(\mathbb{R}, dx) : \int |\hat{f}(t)|^2 \exp\left(\frac{\sigma^2}{2}t^2\right)dt < \infty \}$

with inner product

$$\langle f, g\rangle = \int \hat{f}(t)\overline{\hat{g}(t)}\exp\left(\frac{\sigma^2}{2}t^2\right)dt.$$

Example (Laplacian kernel / exponential kernel):

$$k(x,y) = \exp(-\beta|x-y|)$$

In this case $\rho(t) = \frac{1}{2\pi}\frac{1}{t^2+\beta^2}$ i.e. $\mathcal{G} = L^2\left(\mathbb{R}, \frac{dt}{2\pi(t^2+\beta^2)}\right)$.

The reproducing kernel is the Laplacian kernel:

$$k(x,y) = \frac{1}{2\pi}\int e^{\sqrt{-1}(x-y)t}\frac{1}{t^2+\beta^2}dt$$

$$= \frac{1}{2\beta}\exp(-\beta|x-y|)$$

The RKHS is

$$\mathcal{H} = \{ f \in L^2(\mathbb{R}, dx) : \int |\hat{f}(t)|^2 (t^2+\beta^2)dt < \infty \}$$

equipped with the inner product

$$\langle f, g\rangle = \int \hat{f}(t)\overline{\hat{g}(t)}(t^2+\beta^2)dt$$

Example Cauchy kernel: $\rho(t) = \frac{1}{2\pi}e^{-\alpha|t|}$, $k(x,y) = \frac{1}{2\pi}\int e^{\sqrt{-1}(x-y)t}e^{-\alpha|t|}dt = \frac{1}{1+|x-y|^2}$

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

$$\langle f, g\rangle = \int \hat{f}(t)\hat{g}(t)\, e^{\alpha|t|}dt.$$

· Sometimes it is useful to have the concept of <u>negative definite kernel</u>.

[Berg et al]

<u>Definition</u>: A function $\psi: \Omega \times \Omega \to \mathbb{C}$ is called a negative definite kernel if it is Hermitian, i.e.

$$\psi(y,x) = \overline{\psi(x,y)} \quad \text{and}$$

$$\sum_{i,j=1}^{n} c_i \bar{c}_j \psi(x_i, x_j) \leq 0 \quad \forall x_1, \dots, x_n \ (n \geq 2) \text{ in } \Omega$$

and $c_1 + \dots + c_n = 0$. $c_i \in \mathbb{C}$.

<u>Remark</u>: We have to be careful. Unlike negative definite ~~matrix~~ quadratic form ~~is the~~ which is minus positive definite quadratic form, a negative definite kernel is not necessarily minus positive definite kernel because there is a constraint on $c_i \in \mathbb{C}$: $\sum_{i=1}^{n} c_i = 0$.

<u>Proposition 1</u>: If k is positive definite, $\psi = -k$ is negative definite. Furthermore, constant functions are negative definite.

As for ~~pd~~ positive definite kernels, positive combination of negative definite kernels is negative definite, limit of $\psi_i(x,y)$ is negative definite, but ~~multip~~
negative definite kernels
multiplication does not preserve negative definiteness.

~~Proposition~~ Def. Let φ on $\Omega\times\Omega$ is ~~negative definite~~ a real kernel. Then it is negative definite if and only if φ is symmetric and $\sum_{i,j=1}^{n} c_i c_j \varphi(x_i, x_j) \le 0$ $\forall n \in \mathbb{N}$ $\{x_1, \dots, x_n\} \subseteq \Omega$ and $\{c_1, \dots, c_n\} \subseteq \mathbb{R}$ with $c_1 + \dots + c_n = 0$.

Proposition 2. Let V be ~~an~~ a real inner product space, $\Phi : \Omega \to V$. Then

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

is a negative definite kernel on Ω.

Proof. Let $c = (c_1, \dots, c_n)^T$

Then

$$c^T \psi(x_i, x_j) c = \sum_{i,j=1}^{n} c_i c_j (\Phi(x_i) - \Phi(x_j))^T (\Phi(x_i) - \Phi(x_j))$$

$$= 2\sum_{i,j} c_i c_j \Phi(x_i)^T \Phi(x_i) - 2\sum_{i,j} c_i c_j \Phi(x_i)^T \Phi(x_j)$$

$$= 2\left(\sum_{j=1}^{n} c_j\right)\left(\sum_{i=1}^{n} c_i \Phi(x_i)^T \Phi(x_i)\right) - 2\sum c_i c_j \Phi(x_i)^T \Phi(x_j)$$

$$= -2\left\|\sum_{i=1}^{n} c_i \Phi(x_i)\right\|^2 \le 0. \quad \#$$

[Berg et al]

Theorem (Schoenberg). Let Ω be a nonempty set and $\psi : \Omega\times\Omega \to \mathbb{C}$ be a kernel. Then ψ is negative definite if and only if $\exp(-t\psi)$ is positive definite for all $t > 0$.

[Berg et al].

Proposition 3 If $\psi : \Omega\times\Omega \to \mathbb{C}$ is negative definite and $\psi(x,x) \ge 0$, Then for any $0 < p < 1$, $\psi(x,y)^p$ is negative definite.

· Application to the listed kernel:

— For any $0 \le p \le 2$, $\|x-y\|^p$ is negative definite on $\mathbb{R}^m$.

Proof. By prop. 2 where we choose $\phi(x) = x$ we have $\|x-y\|^2$ is negative definite, hence $(\|x-y\|^2)^\alpha$ is negative definite by prop. 3 for $0 < \alpha \le 1$ so $\|x-y\|^p$ is negative for all $0 \le p \le 2$. ~~By Schoenberg's theorem~~.

— For any $0 \le p \le 2$ and $\alpha > 0$

$$\exp(-\alpha\|x-y\|^p) \text{ is positive definite on } \mathbb{R}^n$$

The statement follows by the Schonberg theorem and $\|x-y\|^p$ is negative definite for $0 \le p \le 2$.

In particular, $p = 1$ we get the Laplacian kernel is positive definite and $p = 2$ the Gaussian kernel is positive definite.

Proposition 4. If $\psi : \Omega \times \Omega \to \mathbb{C}$ is negative definite and $\operatorname{Re} \psi(x,y) \ge 0$. Then for any $\alpha > 0$

$$\frac{1}{\psi(x,y)+\alpha} \text{ is positive definite.}$$

Proof: $\displaystyle \frac{1}{\psi(x,y)+\alpha} = \int_0^\infty e^{-t(\psi(x,y)+\alpha)}\, dt > 0$ $\quad \forall t > 0$ #.

~~Ide~~ Fact: For any $0 < p \le 2$, $\frac{1}{1+\|x-y\|^p}$ is positive definite on $\mathbb{R}^m$; when $p = 2$ it is the Cauchy kernel.