Final Homework

Exercise 1. Positive definiteness ($\mathcal{X} = \mathbb{R}^+$)

- $K_1(x,x') = 2^{x-x'}$ is not positive definite
  indeed, take $x=0$, $x'=1 \in \mathbb{R}_+$. $K_1(x,x') = 1/2 \neq 2 = K_1(x',x)$
  so K_1 is not symmetric.

- $K_2(x,x') = 2^{x+x'}$ is positive definite.
  indeed: - $\forall x,x' \in \mathcal{X}$, $K_2(x,x') = 2^{x+x'} = 2^{x'+x} = K_2(x',x)$
  - Let $N \in \mathbb{N}$, $(x_1,\dots,x_N) \in \mathcal{X}^N$, $(a_1,\dots,a_N) \in \mathbb{R}^N$, one has:
  $$\sum_{i=1}^{N}\sum_{j=1}^{N} a_i a_j K_2(x_i,x_j) = \sum_{i=1}^{N}\sum_{j=1}^{N} a_i a_j 2^{x_i} 2^{x_j} = \left(\sum_{i=1}^{N} a_i 2^{x_i}\right)^2 \geq 0$$

- $K_3(x,x') = 2^{xx'}$ is positive definite
  indeed, $K(x,x') = xx'$ is positive definite (linear kernel in $\mathbb{R}^+$)
  so $\underbrace{\ln(2)}_{>0} xx'$ is positive definite and $e^{\ln(2)xx'} = 2^{xx'}$
  is positive definite as an exponential of a positive definite kernel.

- $K_4(x,x') = \log(1+xx')$ is not positive definite
  indeed, take $x_1 = 0{,}5$ and $x_2 = 2 \in \mathbb{R}^+$, and $a_1 = -2$, $a_2 = 1$, one has $\sum_{i=1}^{2}\sum_{j=1}^{2} a_i a_j K_4(x_i,x_j) = 4\log\left(\frac{5}{4}\right) + 2(-2)\log(2) + \log(5)$
  $$= 5\log(5) - 6\log(4) = \log\left(\frac{3125}{4096}\right) < 0$$

- $K_5(x,x') = \max(x,x')$ is not positive definite

indeed, take $x_1 = 0$, $x_2 = 1 \in \mathbb{R}^+$ and $a_1 = -a_2 = 1$, one has:

$$\sum_{i=1}^{2}\sum_{j=1}^{2} a_i a_j K_5(x_i, x_j) = 2 a_1 a_2 + a_2^2 = -2 + 1 = -1 < 0$$

- $K_6(x,x') = \min(f(x)g(x'), f(x')g(x))$ where f and g are non-negative functions is positive definite.

indeed, denote by

$$\varphi : \mathbb{R}_+ \to \mathbb{R}_+$$
$$x \mapsto \begin{cases} f(x)/g(x) & \text{if } g(x) \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

We will prove that $K_6(x,x') = g(x)g(x') \times \min(\varphi(x), \varphi(x'))$ and prove that the 2 parts of the product are positive definite kernels.

- Let $x, x' \in \mathbb{R}_+$. If $g(x) = 0$, then $K_6(x,x') = 0$ since $f(x')g(x) = 0$ and $f(x)g(x') \geq 0$. Similarly if $g(x') = 0$, $K_6(x,x') = 0$. In both cases $K_6(x,x') = g(x)g(x') \times \min(\varphi(x), \varphi(x'))$

If $g(x)$ and $g(x') \neq 0$; since $f(x), g(x), g(x'), f(x') \geq 0$,

$$\min(f(x)g(x'), f(x')g(x)) = \min\Big(\underbrace{\frac{f(x)}{g(x)}}_{\geq 0} \underbrace{g(x')g(x)}_{\geq 0}, \underbrace{\frac{f(x')}{g(x')}}_{\geq 0} \underbrace{g(x)g(x')}_{\geq 0}\Big)$$
$$= g(x)g(x') \min(\varphi(x), \varphi(x'))$$

So $\forall x, x' \in \mathcal{X}$, $K_6(x,x') = g(x)g(x') \min(\varphi(x), \varphi(x'))$

- Let $x, x' \in \mathbb{R}_+$, $g(x)g(x') = \langle g(x), g(x') \rangle_{\mathbb{R}_+}$ so $g(x)g(x')$ is positive definite (according to Aronszajn theorem for $\mathbb{R}_+$)

- We introduce

$$\Phi : \mathbb{R}_+ \to L^2(\mathbb{R}_+)$$
$$x \mapsto \mathbb{1}_{[0,\varphi(x)]}$$

where $\mathbb{1}_{[0,\varphi(x)]} =: t \mapsto \begin{cases} 1 & \text{if } t \leq \varphi(x) \\ 0 & \text{otherwise} \end{cases}$

Let $x, x' \in \mathbb{R}_+$. Since $\varphi(x), \varphi(x') \geq 0$, one has

$$\int_0^{+\infty} \Phi(x) = \int_0^{+\infty} 1\!\!1_{[0,\varphi(x)]} = \varphi(x)$$

and

$$\min(\varphi(x), \varphi(x')) = \int_0^{+\infty} \Phi(\min(\varphi(x), \varphi(x'))) = \int_0^{+\infty} 1\!\!1_{[0, \min(\varphi(x), \varphi(x'))]}$$

$$= \int_0^{+\infty} 1\!\!1_{[0,\varphi(x)]} 1\!\!1_{[0,\varphi(x')]}$$

$$= \langle \Phi(x), \Phi(x') \rangle_{L^2(\mathbb{R}_+)}$$

So $\min(\varphi(x), \varphi(x'))$ is positive definite according to Aronszajn's theorem for the Hilbert space $L^2(\mathbb{R}_+)$.

- Finally, as a product of 2 positive definite kernels, K_G is positive definite.

Exercise 2. Kernels encoding equivalence classes

($\Rightarrow$) Assume K is p.d.

Let $x, x', x'' \in \mathcal{X}$

- $K(x, x') = K(x', x)$ so $K(x, x') = 1 \iff K(x', x) = 1$
- if $K(x, x') = K(x', x'') = 1$ and $K(x, x'') = 0$, one has $\sum_{i=1}^{3} \sum_{j=1}^{3} a_i a_j K(x_i, x_j)$ (where $x_1 := x$, $x_2 = x'$, $x_3 = x''$)

using that K is symmetric

$$= \sum_{i=1}^{3} a_i^2 + 2 a_1 a_2 K(x_1, x_2) + 2 a_2 a_3 K(x_2, x_3) + 2 a_1 a_3 K(x_1, x_3)$$

and $K(x,x) = 1 \; \forall x \in \mathcal{X}$

$$= \sum_{i=1}^{3} a_i^2 + 2 a_1 a_2 + 2 a_2 a_3 + 0$$

now, if we choose $a_1 = a_3 = 1$ and $a_2 = -1$, we get that $\sum_{i=1}^{3} \sum_{j=1}^{3} a_i a_j K(x_i, x_j) = -1 < 0$, which is absurd since K is p.d.

Therefore $\underline{K(x,x')=1=K(x',x'') \Rightarrow K(x,x'')=1}$

($\Leftarrow$) Assume that $\forall x, x', x'' \in \mathcal{X}$: $\overset{(1)}{\cdot}\ K(x,x')=1 \Leftrightarrow K(x',x)=1$

$\overset{(2)}{\cdot}\ K(x,x')=K(x',x'')=1 \Rightarrow K(x,x'')=1$

- Let $x, x' \in \mathcal{X}$

  if $K(x,x')=1$ then $K(x',x)=1$ according to (1), so

  $K(x,x')=K(x',x)$

  if $K(x,x')=0$, then we rewrite (1) as

  $(1) \Leftrightarrow (K(x,x') \neq 1 \Leftrightarrow K(x',x) \neq 1)$

  $\Leftrightarrow (K(x,x')=0 \Leftrightarrow K(x',x)=0)$

  so $K(x',x)=0 \Rightarrow K(x',x)=K(x,x')$

  therefore in all cases, $\underline{K(x,x')=K(x',x)}$

- Let $N \in \mathbb{N}$, $(x_1,\dots,x_N) \in \mathcal{X}^N$, $(a_1,\dots,a_N) \in \mathbb{R}^N$

  For $x_i, x_j \in \mathcal{X}^N$ we note $x_i \sim x_j$ if $K(x_i,x_j)=1$

  and $x_i \not\sim x_j$ otherwise

  Let $x, x', x'' \in \{x_i\}_{i \in [N]}$; we have that:

  - $K(x,x)=1$ so $x \sim x$
  - $K(x,x')=1 \Leftrightarrow K(x',x)=1$ so $x \sim x' \Leftrightarrow x' \sim x$
  - $K(x,x')=K(x',x'')=1 \Rightarrow K(x,x'')=1$ so $x \sim x', x' \sim x'' \Rightarrow x \sim x''$

  So $\underline{\sim \text{ is an equivalence relation}}$ (reflexive, symmetric and transitive). Therefore, we know that $\sim$ $\underline{\text{partitions the set } \{x_i\}_{i \in [N]}}$ (with the set of all equivalent classes of $\{x_i\}_{i \in [N]}$). We note k the cardinal of this partition and $[1], \dots, [k]$ the k equivalence classes.

  One has:

$$\underline{\sum_{i=1}^{N}\sum_{j=1}^{N} a_i a_j K(x_i, x_j)} = \sum_{i=1}^{N} \sum_{j\sim i} a_i a_j = \sum_{l=1}^{k} \sum_{i\in[l]} \sum_{j\sim i} a_i a_j$$

$$= \sum_{l=1}^{k} \sum_{i\in[l]} \sum_{j\in[l]} a_i a_j = \sum_{l=1}^{k} \Big(\sum_{i\in[l]} a_i\Big)^2 \underline{\geq 0}$$

So $\underline{K \text{ is p.d.}}$

We have proven: K is p.d $\iff \begin{cases} \forall x, x', x'' \in \mathcal{X} \quad K(x,x') = 1 \iff K(x',x) = 1 \\ \text{and} \quad K(x,x') = K(x',x'') = 1 \Rightarrow K(x,x'') = 1 \end{cases}$

Exercise 3. Kernel mean embedding

1. Let T the linear operator $Tf = \mathbb{E}_{X\sim P}[f(X)] \in \mathbb{R}$
   We have $|\mathbb{E}_{X\sim P}[f(X)]| \leq \mathbb{E}_{X\sim P}[|f(X)|]$ using Jensen's inequality.
   Therefore, $|\mathbb{E}_{X\sim P}[f(X)]| \leq \mathbb{E}_{X\sim P}[|\langle f, K_X\rangle_{\mathcal{H}}|]$ (reproducing prop)

$$\leq \|f\|_{\mathcal{H}} \, \mathbb{E}_{X\sim P}[\|K_X\|_{\mathcal{H}}] \quad \text{(Cauchy Schwarz)}$$

$$\leq \|f\|_{\mathcal{H}} \, \underbrace{\mathbb{E}_{X\sim P}\left[\sqrt{K(X,X)}\right]}_{<\infty \text{ because } K \text{ is bounded}}$$

The Riesz representation theorem can be applied since T is bounded and we have that there exists a unique $h \in \mathcal{H}$ such that $Tf = \langle f, h\rangle_{\mathcal{H}} \iff \mathbb{E}_{X\sim P}[f(X)] = \langle f, h\rangle_{\mathcal{H}}, \forall f \in \mathcal{H}$.

Let $y \in \mathcal{X}$, and choose $f = K_y$. $h(y) = \langle h, K_y\rangle_{\mathcal{H}}$

$$= \mathbb{E}_{X\sim P}[K_y(X)]$$

$$= \mu(P)(y)$$

Therefore $\underline{\mu(P) = h} \in \mathcal{H}$.

We have proven that $\mu(P)$ is in $\mathcal{H}$ and that

$\mathbb{E}_{X\sim P}[f(X)] = \langle f, \mu(P)\rangle_{\mathcal{H}}$ for any $f \in \mathcal{H}$.

2. We note $\mathcal{F} = \{ f \in \mathcal{H} : \|f\|_{\mathcal{H}} \leq 1 \}$. We will show that

$\mathbb{E}_S[\|\mu(P) - \mu(P_S)\|_{\mathcal{H}}] \leq 2 \, \mathrm{Rad}_m(\mathcal{F})$

where $\mathrm{Rad}_m(\mathcal{F})$ is the Rademacher complexity of $\mathcal{F}$.

- First, for $h \in \mathcal{H}$ we can write $\|h\|_{\mathcal{H}} = \sup\limits_{f \in \mathcal{F}} \langle f, h \rangle$

(using Cauchy-Schwarz inequality and the fact that for $h \neq 0$ $\langle h/\|h\|, h\rangle_{\mathcal{H}} = \|h\|_{\mathcal{H}}$)

- So $\underbrace{\mathbb{E}_S\left[\|\mu(P) - \mu(P_S)\|_{\mathcal{H}}\right]}_{\text{We note } L \text{ this quantity}} = \mathbb{E}_S\left[\sup\limits_{f\in\mathcal{F}} \langle f, \mu(P) - \mu(P_S)\rangle\right]$

- We have that $\mu(P) = \mathbb{E}_{X\sim P}[K_X(.)]$

and $\mu(P_S) = \mathbb{E}_{X\sim P_S}[K_X(.)] = \sum\limits_{i=1}^{m} \mathbb{P}(X = x_i) \, K_{x_i}(.) = \frac{1}{m}\sum\limits_{i=1}^{m} K_{x_i}(.)$

Let S' another iid sample $(x_1', \dots, x_m')$

$\mathbb{E}_{S'}[\mu(P_{S'})] \underset{(*)}{=} \frac{1}{m}\sum\limits_{i=1}^{m} \mu(P) = \mu(P)$ (*) using the linearity of $\mathbb{E}_{S'}$ and $\mathbb{E}_{S'} K_{x_i'} = \mu(P)$.

Therefore we can write $\mathbb{E}_S[\sup \langle f, \mu(P) - \mu(P_S)\rangle]$

$= \mathbb{E}_S[\sup \langle f, \mathbb{E}_{S'}(\mu(P_{S'})) - \mu(P_S)\rangle]$

$\leq \mathbb{E}_{S,S'}[\sup \langle f, \mu(P_{S'}) - \mu(P_S)\rangle]$

(using the fact that the expectation of the supremum is bigger than the supremum of the expectation)

- the question 3 (that will be proven without using question 2) gives us that:

$$L \leq \mathbb{E}_{S,S'}\left[\sup_{f\in\mathcal{F}}\left\{\frac{1}{m}\sum_{i=1}^{m} f(x_i') - f(x_i)\right\}\right]$$

- Let $i \in [m]$ and σ_i a Rademacher variable ($\sigma_i = 1$ with probability $1/2$ and $\sigma_i = -1$ with same probability) one has:

$$\mathbb{E}_{S,S',\sigma_i}\left[\sup_{f\in\mathcal{F}} \sigma_i\left(f(x_i') - f(x_i)\right) + \sum_{j\neq i}\left(f(x_j') - f(x_j)\right)\right]$$

$$= \frac{1}{2}\mathbb{E}_{S,S'}\left[\sup_{f}\left\{f(x_i') - f(x_i) + \sum_{j\neq i}\left(f(x_j') - f(x_j)\right)\right\}\right]$$

$$+ \frac{1}{2}\mathbb{E}_{S,S'}\left[\sup_{f}\left\{\underbrace{f(x_i) - f(x_i')} + \sum_{j\neq i}\left(f(x_j') - f(x_j)\right)\right\}\right]$$

x_i and x_i' are iid therefore, when averaging, we retrieve the same result as with $f(x_i') - f(x_i)$

$$= \mathbb{E}_{S,S'}\left[\sup\left\{\sum_{j=1}^{m} f(x_j') - f(x_j)\right\}\right]$$

We can repeat this operation for $i = 1, \dots, n$ and we get: $\mathbb{E}_{S,S'}\left[\sup_{f}\left\{\frac{1}{m}\sum_{i=1}^{m} f(x_i') - f(x_i)\right\}\right] = \mathbb{E}_{S,S',\sigma}\left[\sup_{f}\left\{\frac{1}{m}\sum_{i=1}^{m}\sigma_i\left(f(x_i') - f(x_i)\right)\right\}\right]$:

- Then by triangular inequality

$$\underline{L} \leq \mathbb{E}_{S,S',\sigma}\left[\sup_{f\in\mathcal{F}}\left\{\frac{1}{m}\sum_{i=1}^{m}\sigma_i f(x_i')\right\} + \sup_{f\in\mathcal{F}}\left\{\frac{1}{m}\sum_{i=1}^{m}-\sigma_i f(x_i)\right\}\right]$$

$$\leq \mathbb{E}_{S,S',\sigma}\left[\sup_{f\in\mathcal{F}}\left\{\frac{1}{m}\sum_{i=1}^{m}\sigma_i f(x_i')\right\} + \sup_{f\in\mathcal{F}}\left\{\frac{1}{m}\sum_{i=1}^{m}\sigma_i f(x_i)\right\}\right]$$

$$\leq \mathbb{E}_{S,\sigma}\left[\sup_{f\in\mathcal{F}}\left\{\frac{1}{m}\sum_{i=1}^{m}\sigma_i f(x_i)\right\}\right] + \mathbb{E}_{S',\sigma}\left[\sup_{f\in\mathcal{F}}\left\{\frac{1}{m}\sum_{i=1}^{m}\sigma_i f(x_i')\right\}\right]$$

$$= \underline{2\ \mathrm{Rad}_m(\mathcal{F})}$$

Then, using the capacity control of RKHS balls theorem, we know that $\underline{\mathrm{Rad}_m(\mathcal{F}) \leq 2\frac{\sqrt{\mathbb{E}\,k(X,X)}}{\sqrt{m}}}$

Finally, $\underline{\mathbb{E}_S[\|\mu(P)-\mu(P_S)\|_{\mathcal{H}}] \leq 4\frac{\sqrt{\mathbb{E}\,k(X,X)}}{\sqrt{m}}}$

3. One has $\|\mu(P_{S_1})-\mu(P_{S_2})\|_{\mathcal{H}} = \sup_{\substack{f\in\mathcal{H}\\ \|f\|_{\mathcal{H}}\leq 1}} \langle f, \mu(P_{S_1})-\mu(P_{S_2})\rangle$

(same reas. in question 2)

Therefore, $MMD(S_1,S_2) = \left(\sup_{\|f\|_{\mathcal{H}}\leq 1}\left\{\langle f,\mu(P_{S_1})\rangle_{\mathcal{H}} - \langle f,\mu(P_{S_2})\rangle_{\mathcal{H}}\right\}\right)^2$

$$= \left(\sup_{\|f\|_{\mathcal{H}}\leq 1}\left\{\mathbb{E}_{X\sim P_{S_1}}(f(X)) - \mathbb{E}_{X\sim P_{S_2}}(f(X))\right\}\right)^2$$

$$= \left(\sup_{\|f\|_{\mathcal{H}}\leq 1}\left\{\sum_{i=1}^{m}\mathbb{P}_{S_1}(X=x_i)\,f(x_i) - \sum_{j=1}^{m}\mathbb{P}_{S_2}(X=y_j)\,f(y_j)\right\}\right)^2$$

$$\Rightarrow \underline{MMD(S_1,S_2) = \left(\sup_{\|f\|_{\mathcal{H}}\leq 1}\left\{\frac{1}{m}\sum_{i=1}^{m}f(x_i) - \frac{1}{m}\sum_{j=1}^{m}f(y_j)\right\}\right)^2}$$

and since $f(x) = \langle f, K_x(.)\rangle_{\mathcal{H}}$

$$MMD(S_1,S_2) = \left(\sup_{\|f\|_{\mathcal{H}}\leq 1}\langle f, \frac{1}{m}\sum_{i=1}^{m}K_{x_i}(.) - \frac{1}{m}\sum_{j=1}^{m}K_{y_j}(.)\rangle_{\mathcal{H}}\right)^2$$

So $MMD(S_1, S_2) = \left\| \frac{1}{m} \sum_{i=1}^{m} K_{x_i} - \frac{1}{m} \sum_{j=1}^{m} K_{y_j} \right\|_{\mathcal{H}}^2$

$$= \left\langle \frac{1}{m} \sum_{i=1}^{m} K_{x_i} - \frac{1}{m} \sum_{j=1}^{m} K_{y_j}, \frac{1}{m} \sum_{i=1}^{m} K_{x_i} - \frac{1}{m} \sum_{j=1}^{m} K_{y_j} \right\rangle_{\mathcal{H}}$$

$$= \frac{1}{m^2} \sum_{i,j=1}^{m} K(x_i, x_j) + \frac{1}{m^2} \sum_{i,j=1}^{m} K(y_i, y_j) - \frac{2}{mm} \sum_{i=1}^{m} \sum_{j=1}^{m} K(x_i, y_j)$$

4. $\mathcal{X} = \mathbb{R}^d$, $K(x,y) = e^{-\frac{\|x-y\|^2}{2\sigma^2}}$

Let $S_1 = (x_1, \dots, x_n)$ and $S_2 = (y_1, \dots, y_m)$

- Consider $\sigma_1 \leqslant \sigma_2$ and their associated Gaussian Kernels K_1, K_2. K_1's RKHS is $\mathcal{H}_1 = \{ f : \int |\hat{f}(\omega)| e^{\sigma_1^2 \omega^2/2} d\omega < +\infty \}$ according to the results on shift invariant kernels (slide 314), and the same holds for $\mathcal{H}_2$, but for σ_2.
Moreover, $\|f\|_{\mathcal{H}_1}^2 = \int 1/(2\pi)^d \, |\hat{f}(\omega)|^2 e^{\sigma_1^2 \omega^2/2} d\omega$ (same for $\mathcal{H}_2$)

- Now let $f \in \mathcal{H}_2$ such that $\|f\|_{\mathcal{H}_2} \leqslant 1$.
One has $\sigma_1 < \sigma_2$ so $\forall \omega \in \mathbb{R}^d$:
$0 \leqslant e^{\sigma_1 \omega^2/2} \leqslant e^{\sigma_2 \omega^2/2}$
$\Rightarrow 0 \leqslant |\hat{f}(\omega)|^2 e^{\sigma_1^2 \omega^2/2} \leqslant |\hat{f}(\omega)|^2 e^{\sigma_2 \omega^2/2}$
$\Rightarrow \int 1/(2\pi)^d \, |\hat{f}(\omega)|^2 e^{\sigma_1^2 \omega^2/2} d\omega \leqslant \int 1/(2\pi)^d \, |\hat{f}(\omega)| e^{\sigma_2^2 \omega^2/2} d\omega$

Therefore $\|f\|_{\mathcal{H}_1} \leqslant \|f\|_{\mathcal{H}_2} \leqslant 1$, so $\|f\|_{\mathcal{H}_2} \leqslant 1 \Rightarrow \|f\|_{\mathcal{H}_1} \leqslant 1$

- Combining this to question 3, we get that:
$$MMD_{\sigma_1}(S_1, S_2) = \sup_{\|f\|_{\mathcal{H}_1} \leqslant 1} \left\{ \frac{1}{m} \sum_{i=1}^{m} f(x_i) - \frac{1}{m} \sum_{j=1}^{m} f(y_j) \right\}$$
$$\geqslant \sup_{\|f\|_{\mathcal{H}_2} \leqslant 1} \left\{ \frac{1}{m} \sum_{i=1}^{m} f(x_i) - \frac{1}{m} \sum_{j=1}^{m} f(y_j) \right\} = MMD_{\sigma_2}(S_1, S_2)$$

- So MMD is a decreasing function of σ

Exercise 4. Properties of the dot-product kernel

1. - Let $x, x' \in \mathcal{X} = \mathbb{S}^{p-1}$. $K_1(x,x') = K(\langle x,x'\rangle) = K(\langle x',x\rangle) = K_1(x',x)$

- Let $N \in \mathbb{N}$, $(x_1, \dots, x_N) \in \mathcal{X}^N$, $(\alpha_1, \dots, \alpha_N) \in \mathbb{R}^N$, one has:

$$\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j K_1(x_i, x_j) = \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j \sum_{k=0}^{+\infty} a_k \langle x_i, x_j\rangle^k$$

$$= \sum_{k=0}^{+\infty} a_k \underbrace{\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j \langle x_i, x_j\rangle^k}_{\geq 0 \text{ since } K(x,x') = \langle x,x'\rangle^k \text{ is p.d. (polynomial kernel)}}$$

≥ 0 if $a_k \geq 0$ for any $k \in \mathbb{N}$

Therefore if all coefficients a_i are non negative, then K_1 is p.d.

2. - Let $x, x' \in \mathcal{X}$. $K_2(x,x') = \begin{cases} \|x\| \|x'\| K(\langle x,x'\rangle / \|x\| \|x'\|) & \text{if } \|x\|\|x'\| \neq 0 \\ 0 & \text{otherwise} \end{cases}$

$$= \begin{cases} \|x'\| \|x\| K(\langle x',x\rangle / \|x'\| \|x\|) & \text{if } \|x\|\|x'\| \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$= K_2(x', x)$$

- Let $N \in \mathbb{N}$, $(x_1, \dots, x_N) \in \mathcal{X}^N$, $(\alpha_1, \dots, \alpha_N) \in \mathbb{R}^N$, one has:

$$\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j K_2(x_i, x_j) = \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j \|x_i\| \|x_j\| \sum_{k=0}^{+\infty} a_k \frac{\langle x_i, x_j\rangle^k}{\|x_i\|^k \|x_j\|^k}$$

$$= \sum_{i=1}^{N}\sum_{j=1}^{N} \underbrace{\alpha_i \|x_i\|}_{\in \mathbb{R}} \underbrace{\alpha_j \|x_j\|}_{\in \mathbb{R}} K_1\Big(\underbrace{\frac{x_i}{\|x_i\|}}_{\in \mathcal{X}}, \underbrace{\frac{x_j}{\|x_j\|}}_{\in \mathcal{X}}\Big) \geq 0$$

≥ 0 since K_1 is p.d.

Therefore if K_1 is p.d. so is K_2

3. Let $x, x' \in \mathbb{S}^{p-1}$ one has:

$$\|\varphi(x) - \varphi(x')\|_{\mathcal{H}}^2 = \langle \varphi(x) - \varphi(x'), \varphi(x) - \varphi(x') \rangle$$

$$= \langle \varphi(x), \varphi(x) \rangle + \langle \varphi(x'), \varphi(x') \rangle - 2\langle \varphi(x), \varphi(x') \rangle$$

$$= K_1(x,x) + K_1(x',x') - 2K_1(x,x')$$

$$= \sum_{k=0}^{+\infty} a_k \left[\langle x,x\rangle^k + \langle x',x'\rangle^k - 2\langle x,x'\rangle^k \right]$$

$$= \sum_{k=0}^{+\infty} a_k \left[2\left(1 - \langle x,x'\rangle^k\right) \right]$$

$$= \sum_{k=0}^{+\infty} a_k \left[2\left(1 - \langle x,x'\rangle\right) \overbrace{\sum_{i=0}^{k-1} \langle x',x\rangle^i}^{=0 \text{ if } k=0} \right]$$

using $2 - 2\langle x,x'\rangle = \|x\|^2 + \|x'\|^2 - 2\langle x,x'\rangle = \|x - x'\|^2$ $(\|x\| = \|x'\| = 1)$

$$= \sum_{k=0}^{+\infty} a_k \|x - x'\|^2 \sum_{i=0}^{k-1} \langle x',x\rangle^i$$

because Cauchy-Schwarz inequality gives $\langle x',x\rangle \leq 1$

$$\leq \|x - x'\|^2 \underbrace{\sum_{k=0}^{+\infty} a_k k}_{= K'(1) = 1}$$

$$\leq \|x - x'\|^2$$

Therefore, $\underline{\|\varphi(x) - \varphi(x')\|_{\mathcal{H}} \leq \|x - x'\|}$

4. Let $x, y \in \mathbb{S}^{p-1}$,

$$K_1(x,y) = \sum_{k=0}^{+\infty} a_k \langle x,y\rangle^k = \sum_{k=0}^{+\infty} a_k \operatorname{tr}\left\{(x^T y)^k\right\}$$

$$= \sum_{k=0}^{+\infty} a_k \operatorname{tr}\left\{x^{\otimes k}, y^{\otimes k}\right\} = \underline{\sum_{k=0}^{+\infty} a_k \langle x^{\otimes k}, y^{\otimes k}\rangle_F}$$

$$= \sum_{k=0}^{+\infty} \sum_{i_1,\dots,i_k=1}^{p} x_{i_1} \cdots x_{i_k} y_{i_1} \cdots y_{i_k} a_k$$

$$= \underline{\sum_{k=0}^{+\infty} \sum_{i_1,\dots,i_k=1}^{p} \sqrt{a_k}\, x_{i_1} \cdots x_{i_k} \times \sqrt{a_k}\, y_{i_1} \cdots y_{i_k}}$$

Then we can reorder the elements of the sum to obtain

$$K_1(x,y) = \sum_{k'=0}^{+\infty} \phi_{k'}(x)\, \phi_{k'}(y)$$

Where $\phi_{k'}(x) = \sqrt{a_k}\, x_{i_1} \cdots x_{i_k}$ for $k, i_1, \dots, i_k$ defined as:

- $k' \in \left[\sum_{i=0}^{k-1} p^i , \sum_{i=0}^{k} p^i \right[$ i.e., for $p > 1$,

$$\frac{1-p^k}{1-p} \leq k' \leq \frac{1-p^{k+1}}{1-p} \iff p^k \leq 1 - k'(1-p) < p^{k+1}$$

$$\iff k \leq \frac{\ln(1+k'(p-1))}{\ln(p)} < k+1 \iff k = \left\lfloor \frac{\ln(1+k'(p-1))}{\ln(p)} \right\rfloor$$

- $i_1 = 1 + \left(k' - \sum_{i=0}^{k-1} p^i\right) // p^{k-1}$

quotient

$$i_2 = 1 + \left[\left(k' - \sum_{i=0}^{k-1} p^i\right) \% \, p^{k-1}\right] // p^{k-2}$$

remainder

$\vdots$

$$i_k = 1 + \left[\left(\left(k' - \sum_{i=0}^{k-1} p^i\right) \% \, p^{k-1}\right) \dots\right) \% \, p$$

Therefore, define $\psi : \mathbb{S}^{p-1} \to \ell^2$

$$x \mapsto (\phi_k(x))_{k \in \mathbb{N}}$$

we have $K_1(x,y) = \langle \psi(x), \psi(y) \rangle_{\ell_2}$ for any x, y in $\mathbb{S}^{p-1}$

(well defined on $\mathbb{S}^{p-1}$ because $\sum_k (\phi_k(x))^2 = K_1(x,x) < +\infty$

so $\phi_k \in \ell^2$ for any k)

5. Let $(w_z)_{k \in \mathbb{N}}$ be defined as $(w_z)_{k'} = \begin{cases} 0 & \text{if } a_k = 0 \\ \frac{b_k}{\sqrt{a_k}} z_{i_1} \dots z_{i_k} & \text{where } k \end{cases}$

and $i_1, \dots, i_k$ are defined as in 4.

Assume that $b_k = 0$ for any k such that $a_k = 0$
and that $\sum_{k=0}^{+\infty} \frac{b_k^2}{a_k} < +\infty$ (with $b_k^2 / a_k = 0$ if $a_k = 0$)

One has:

$$
\begin{aligned}
g_z(x) &= \sum_{k=0}^{+\infty} b_k \langle z, x \rangle^k \\
&= \sum_{k=0}^{+\infty} b_k \langle z^{\otimes k}, x^{\otimes k} \rangle \quad \text{(see 4.)} \\
&= \sum_{k=0}^{+\infty} b_k \sum_{i_1, \dots, i_k = 1}^{p} z_{i_1} \dots z_{i_k} x_{i_1} \dots x_{i_k} \quad \text{since } \underbrace{a_k = 0}_{} \Rightarrow b_k = 0 \\
&= \sum_{\substack{k=0 \\ a_k \neq 0}}^{+\infty} \sum_{i_1, \dots, i_k = 1}^{p} \frac{b_k}{\sqrt{a_k}} z_{i_1} \dots z_{i_k} \sqrt{a_k} \, x_{i_1} \dots x_{i_k} + \sum_{\substack{k=0 \\ a_k = 0}}^{+\infty} 0 \\
&= \langle \psi(x), w_z \rangle_{\ell_2}
\end{aligned}
$$

and $w_z \in \ell_2$ because $\langle w_z, w_z \rangle_{\ell_2} = \sum_{k=0}^{+\infty} \frac{b_k^2}{a_k} \sum_{i_1, \dots, i_k = 1}^{p} (z_{i_1} \dots z_{i_k})^2$

$$
= \sum_{k=0}^{+\infty} \frac{b_k^2}{a_k} \underbrace{\langle z, z \rangle^k}_{=1} < +\infty
$$

Therefore with the previous assumptions, $\exists w \in \ell_2$, $g_z(x) = \langle w, \psi(x) \rangle_{\ell_2}$
so $g_z \in \mathcal{H}$

Finally, $\begin{cases} a_k = 0 \Rightarrow b_k = 0 & \text{for any } k \in \mathbb{N} \\ \sum_{k=0}^{+\infty} \frac{b_k^2}{a_k} < +\infty \end{cases}$ are sufficient conditions for g_z to be in $\mathcal{H}$