Contents

CHAPTER I

[The lone chapter]

Conventions.

- $K \in \{\mathbb{R}, \mathbb{C}\}$.
- Vector spaces will be over K.
- Abusing the notation slightly, the same notation will be used to denote the restriction to $\mathbb{R} \to \mathbb{R}$ of Re, Im and complex conjugation.
- $\mathscr{H}$ will denote a Hilbert space over K.
- X will denote a generic set.
- Evaluation functionals on K^X will be denoted by δ_x's.
- We'll use $\|x\| := \sqrt{\langle x, x \rangle}$ for $x \in \mathscr{H}$.
- Whenever $\langle \cdot, \cdot \rangle$ is a p.s.d. conjugate symmetric bilinear form on a vector space,[1] then we'll use the usual $\|\cdot\|$ to denote the induced seminorm.

1. Basics

Definition 1.1 (Positive semi-definite kernel)**.** A p.s.d. kernel on a set X is a function $k\colon X \times X \to K$ that is

(i) *conjugate symmetric*, *i.e.*, $k(y, x) = \overline{k(x, y)}$; and,

(ii) *positive semi-definite*, *i.e.*, for any $x_1, \ldots, x_n \in X$ and any $\alpha_1, \ldots, \alpha_n \in K$, we

[1] That is, it's almost a norm except not possibly satisfying the positive definiteness.

have that

$$\sum_{i,j=1}^{n} \overline{\alpha_i}\, k(x_i, x_j)\, \alpha_j \geq 0.$$

Definition 1.2 (Reproducing kernel)**.** Let $\mathscr{H} \subseteq K^X$. Then an r.k. for $\mathscr{H}$ is a function $k\colon X \times X \to K$ such that the following hold:

(i) $k(\cdot, x) \in \mathscr{H}$ for each $x \in X$.

(ii) *(Reproducing property)***.** $f(x) = \langle f, k(\cdot, x)\rangle$ for each $f \in \mathscr{H}$ and each $x \in X$.

Corollary 1.3. *There exists at most one r.k. for a Hilbert space.*

Proof. Suppose k_1 and k_2 are r.k.'s for $\mathscr{H} \subseteq K^X$. Then for any $f \in \mathscr{H}$ and $x \in X$, we have $\langle f, k_1(\cdot, x) - k_2(\cdot, x)\rangle = f(x) - f(x) = 0$ so that $k_1(\cdot, x) = k_2(\cdot, x)$. Since x was arbitrary, $k_1 = k_2$. □

Corollary 1.4. *An r.k. is a p.s.d. kernel.*

Proof. Let k be an r.k. of $\mathscr{H} \subseteq K^X$.

- Conjugate symmetry: $k(y, x) = \langle k(\cdot, x), k(\cdot, y)\rangle = \overline{\langle k(\cdot, y), k(\cdot, x)\rangle} = \overline{k(x, y)}$.
- Positive semi-definite: Let $x_1, \ldots, x_n \in X$ and $\alpha_1, \cdots, \alpha_n \in K$. Then setting $k_x := k(\cdot, x)$, we have

$$\begin{aligned} \sum\nolimits_{i,j} \overline{\alpha_i}\, k(x_i, x_j)\, \alpha_j &= \sum\nolimits_{i,j} \overline{\alpha_i}\, \langle k_{x_j}, k_{x_i}\rangle \alpha_j \\ &= \sum\nolimits_{i,j} \langle \alpha_j k_{x_j}, \alpha_i k_{x_i}\rangle \\ &= \left\| \sum\nolimits_i \alpha_i k_{x_i} \right\|^2 \\ &\geq 0. \end{aligned}$$ □

Remark. The converse is the content of 2.3.

Definition 1.5 (Reproducing kernel Hilbert space)**.** If $\mathscr{H} \subseteq K^X$ and admits a reproducing kernel, then it's called an RKHS.

Corollary 1.6 (Characterizing RKHS's)**.** *$\mathscr{H} \subseteq K^X$ is an RKHS $\iff$ evaluation functionals on it are continuous.*

Proof. "$\Rightarrow$": Let k be the r.k. of $\mathscr{H}$. Then $|\delta_x(f)| = |f(x)| = |\langle f, k(\cdot, x)\rangle| \leq \|f\|\|k(\cdot, x)\| = \sqrt{k(x, x)}\,\|f\|$. (Note that $k(x, x) \geq 0$ since k is positive semi-definite.) Thus $\|\delta_x\| \leq \sqrt{k(x, x)} < +\infty$.

"$\Leftarrow$": By Riesz, we can define $\ell\colon X \to \mathscr{H}$ such that $\delta_x = \langle \cdot, \ell_x \rangle$. Define $k\colon X \times X \to K$ by $k(\cdot, x) := \ell_x$. Then $f(x) = \delta_x(f) = \langle f, \ell_x \rangle = \langle f, k(\cdot, x) \rangle$. □

Corollary 1.7. *Convergence in an RKHS $\implies$ pointwise convergence.*

Proof. Let $\mathscr{H} \subseteq K^X$ be an RKHS with k being the r.k. Let $f_n \to f$ in $\mathscr{H}$. Let $x \in X$. Since δ_x is continuous, $\delta_x(f_n) \to \delta_x(f)$ in K, *i.e.*, $f_n(x) \to f(x)$ in K. □

Remark. Note that completeness of $\mathscr{H}$ was of no consequence in the results so far. Thus it might seem that we must generalize the definition of r.k. to include inner product spaces. However, as 2.3 will show, if k is an r.k. for an inner product space, then k will also be an r.k. for an **?????????**

2. Moore-Aronszajn

Lemma 2.1. *Let k be a p.s.d. kernel. Define $\mathscr{H}_0$ to be the span in K^X of all $k_x := k(\cdot, x)$'s for $x \in X$. Then the following hold:*

(i) $\mathscr{H}_0$ admits an inner product such that the following hold:
  - *(a) $\langle k_x, k_y \rangle = k(y, x)$.*
  - *(b) $f(x) = \langle f, k_x \rangle$ for any $f \in \mathscr{H}_0$.*

(ii) Cauchy sequences in $\mathscr{H}_0$ have pointwise limits.

Proof. (i) We show that

$$\langle f, g \rangle = \sum\nolimits_{i.j} \overline{\beta_j}\, k(y_j, x_i)\, \alpha_i \tag{2.1}$$

defines an inner product on $\mathscr{H}_0$ for $f = \sum_{i=1}^m \alpha_i k_{x_i}$ and $g = \sum_{j=1}^n \beta_j k_{y_j}$. That it's well-defined follows because

$$\sum\nolimits_j \overline{\beta_j} f(y_j) = \sum\nolimits_{i.j} \overline{\beta_j}\, k(y_j, x_i)\, \alpha_i = \sum\nolimits_i \overline{g(x_i)} \alpha_i \tag{2.2}$$

where the second equality follows since <u>k is conjugate symmetric</u>. It's immediate from 2.1 that $\langle \cdot, \cdot \rangle$ is p.s.d. and conjugate symmetric (since <u>k is a p.s.d. kernel</u>), and from 2.2 that it's bilinear with $\langle f, k_x \rangle = f(x)$ for any $x \in X$ (take $g = k_x$). Only positive definiteness remains to be shown:

Firstly, note that $\langle \cdot, \cdot \rangle$ is a p.s.d. conjugate symmetric bilinear form so that it obeys Cauchy-Schwarz and induces a seminorm. Now, let $\|f\| = 0$. Then for any $x \in X$, we have $|f(x)| = |\langle f, k_x \rangle| \leq \|f\| \|k_x\| = 0$.

(ii) Let (f_n) be Cauchy in $\mathscr{H}_0$. Then for any $x \in X$, we $|f_m(x) - f_n(x)| = |\delta_x(f_m - f_n)| \leq \|\delta_x\| \|f_m - f_n\| \xrightarrow{w} 0$ as $m, n \to 0$ so that $(f_n(x))$ is Cauchy in K and hence convergent. $\square$

Lemma 2.2. *Continuing 2.1, define $\mathscr{H}$ to be the set of pointwise limits of Cauchy sequences in $\mathscr{H}_0$. Clearly, this is a vector subspace of K^X. There exists an inner product on $\mathscr{H}$ given by*

$$\langle f, g \rangle =$$

Theorem 2.3 (Moore-Aronszajn)**.** *Let k be a p.s.d. kernel on X. Then there exists a **unique??** RKHS whose kernel is k.*