Contents

CHAPTER I

[the lone chapter]

Conventions. Unless stated otherwise, assume the following: *Filler text...*

- $K \in \{\mathbb{R}, \mathbb{C}\}$.
- X will denote a generic set.
- Vector spaces will be over K.
- Evaluation functions on any subset of K^X will be denoted by δ_x's. These are clearly linear maps in case the domain is a subspace of K^X.
- $\mathscr{H}$ will denote a Hilbert space 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.
- Whenever $\langle \cdot, \cdot \rangle$ is semi-inner-product on a vector space,[1] we'll use the usual $\|\cdot\|$ to denote the induced seminorm.[2]
- For any function $k \colon X \times X \to K$, we'll use k_x to stand for $k(\cdot, x) \colon X \to K$

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,

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

[2]Note that a semi-inner-product obeys the Cauchy-Schwarz inequality (just follow Schwarz's proof using the quadratic polynomial).

(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 have that

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

Definition 1.2 (Reproducing kernel and RKHS's)**.** A *reproducing kernel* for $\mathscr{H} \subseteq K^X$ is a function $k\colon X \times X \to K$ such that for any $x \in X$, we have that $k_x \in \mathscr{H}$ and that it obeys the *reproducing property*, *i.e.*,

$$f(x) = \langle f, k_x \rangle$$

for any $f \in \mathscr{H}$.

Further, if $\mathscr{H}$ admits a reproducing kernel, it's called a *reproducing kernel Hilbert space.*

Lemma 1.3. *A reproducing kernel is a p.s.d. kernel.*

Proof. Let k be a reproducing kernel of $\mathscr{H} \subseteq K^X$.

- Conjugate symmetry: $k(y, x) = k_x(y) = \langle k_x, k_y \rangle = \overline{\langle k_y, k_x \rangle} = \overline{k_y(x)} = \overline{k(x, y)}$.
- 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 2.4.

Lemma 1.4 (Characterizing RKHS's)**.** *$\mathscr{H} \subseteq K^X$ is an RKHS $\iff$ evaluation functionals on it are continuous.*

Proof. "⇒": Let k be a reproducing kernel of $\mathscr{H}$. Then $|\delta_x(f)| = |f(x)| = |\langle f, k_x \rangle| \leq \|k_x\| \|f\|$ (note that $k_x \in \mathscr{H}$). Thus $\|\delta_x\| \leq \|k_x\| < +\infty$.

"⇐": By Riesz, we can define $k\colon X \times X \to K$ such that $\delta_x = \langle \cdot, k_x \rangle$ (since δ_x's are continuous). Then $f(x) = \delta_x(f) = \langle f, k_x \rangle$. □

Corollary 1.5. *Convergence in an RKHS $\implies$ pointwise convergence.* What about the converse?

Theorem 1.6 (RKHS $\mapsto$ r.k. is injective)**.**

(i) An RKHS's reproducing kernel is unique.

(ii) Conversely, a reproducing kernel makes at most one Hilbert space into an RKHS.

Proof. (i) Suppose k and k' are reproducing kernels for $\mathscr{H} \subseteq K^X$. Then for any $f \in \mathscr{H}$ and $x \in X$, we have $\langle f, k_x - k'_x \rangle = f(x) - f(x) = 0$ so that $k_x = k'_x$. Since x was arbitrary, $k = k'$.

(ii) Let k be a reproducing kernel for $\mathscr{H} \subseteq K^X$. It suffices to show that $\mathscr{H}$ is uniquely determined (along with its inner product[3]) by k. Let $\mathscr{H}_0$ be the subspace of $\mathscr{H}$ spanned by k_x's for $x \in X$. Note that $\mathscr{H}_0^\perp = \{0\}$ (because of k's reproducing property) so that $\overline{\mathscr{H}_0} = (\mathscr{H}_0^\perp)^\perp = \mathscr{H}$.[4] Since $\mathscr{H}_0$ is determined by k, it suffices to have that the topology on $\mathscr{H}$ is also determined by k $\overset{\text{w}}{\Longleftarrow}$ the norm on $\mathscr{H}$ is determined by k. Note that this will also imply that the inner product is uniquely determined (due to polarization).
It suffices to have that the norm is determined on $\mathscr{H}_0$ for it is dense in $\mathscr{H}$. Indeed, for $f = \sum_{i=\alpha_i}^n \alpha_i k_{x_i}$, we have

$$\begin{aligned}\|f\|^2 &= \sum_{i,j=1}^n \alpha_i \overline{\alpha_j} \langle k_{x_i}, k_{x_j} \rangle \\ &= \sum_{i,j=1}^n \overline{\alpha_j}\, k(x_j, x_i)\, \alpha_i.\end{aligned}$$ □

Remark. (i) Note that completeness of $\mathscr{H}$ was used in writing $\overline{\mathscr{H}_0} = (\mathscr{H}_0^\perp)^\perp$.

An example to show the necessity of this?

(ii) The converse is the content of 2.4, which together with the above gives the following correspondence for any set X:

$$\{\} \longleftrightarrow \{\} \longleftrightarrow \{\}$$

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 reproducing kernel to include inner product spaces. However, as 2.4 will show, if k is an reproducing kernel for an inner product space, then k will also be an reproducing kernel for an **?????????**

[3]Note that addition and scalar multiplication are already determined (namely, pointwise) by definition.

[4]A naïve glance suggests that this fixes $\mathscr{H}$, but it need *not*! We haven't yet reduced the description of $\mathscr{H}$ to only that of k. The closure of $\mathscr{H}_0$ (which *does* only depend on k) is dependent on the norm topology induced from $\mathscr{H}$, whose description hasn't been reduced yet.

2. Moore-Aronszajn

We now ask the converse of 1.3. Given a p.s.d. kernel k on X, we construct an RKHS $\mathscr{H}$ whose reproducing kernel is precisely k. We do so in the following steps:

(i) Each k_x must lie in $\mathscr{H}$. Thus, we are motivated to first define a vector space $\mathscr{H}_0$ spanned by k_x's.

(ii) We show that there's a unique inner product on $\mathscr{H}_0$ with respect to which k has the reproducing property.

(iii) Finally, we complete $\mathscr{H}_0$, and verify that it's the required RKHS.

Lemma 2.1. *Let k be a p.s.d. kernel. Define $\mathscr{H}_0$ to be the span in K^X of k_x's for $x \in X$. Then $\mathscr{H}_0$ admits a unique inner product such that $f(x) = \langle f, k_x \rangle$ for any $f \in \mathscr{H}_0$.*

Proof. 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:

Note that $\langle \cdot, \cdot \rangle$ is a semi-inner-product so that it obeys Cauchy-Schwarz and induces a seminorm. Now, let $\|f\| = 0$. Then for any $x \in X$, we have $|f(x)| = |\langle f, k_x \rangle| \leq \|f\| \|k_x\| = 0$.

□

Lemma 2.2. *Continuing 2.1, let $S := \{$Cauchy sequences in $\mathscr{H}_0\}$. Then there exists a linear map $\phi \colon S \to K^X$ that maps Cauchy sequences to their pointwise limits, the kernel of which consists precisely of sequences that converge to 0 in $\mathscr{H}_0$.*

Proof. First we show that ϕ is indeed well-defined:

Let (f_n) be Cauchy in $\mathscr{H}_0$. Then for any $x \in X$, we have $|f_m(x) - f_n(x)| = |\langle f_m - f_n, k_x \rangle| \leq \|f_n - f_m\| \|k_x\| \overset{\text{w}}{\to} 0$ as $m, n \to \infty$ so that $(f_n(x))$ is Cauchy in K and hence convergent. Thus, pointwise limits of Cauchy sequences in $\mathscr{H}_0$ do exist, and since K is Hausdorff, these are unique.

Linearity of ϕ is easy. We now compute $\ker\phi$. If $f_n \to 0$ in $\mathscr{H}_0$, then for any $x \in X$, we have $f_n(x) = \langle f_n, k_x\rangle \xrightarrow{w} 0$. Conversely, let (f_n) be a Cauchy sequence in $\mathscr{H}_0$ that converges to 0 pointwise. We show that it converges to 0 in $\mathscr{H}_0$ as well:

Let $\varepsilon > 0$. Fix an N and write $f_N = \sum_{i=1}^n \alpha_i k_{x_i}$. Now,

$$\begin{aligned}\|f_n\|^2 &= \left|\langle f_n - f_N, f_n\rangle + \langle f_N, f_n\rangle\right| \\ &\le \left|\langle f_n - f_N, f_n\rangle\right| + \left|\sum_{i=1}^n \overline{\alpha_i} f_n(x_i)\right| \\ &\le \|f_n - f_N\|\|f_n\| + \sum_{i=1}^n |\alpha_i||f_n(x_i)|\end{aligned}$$

so that taking N large enough ensures that the above is eventually less than any arbitrary $\varepsilon > 0$. □

Note that $S/\ker\phi$ is precisely the[5] metric completion of $\mathscr{H}_0$.

Lemma 2.3. *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 2.4 (Moore-Aronszajn)**.** *Let k be a p.s.d. kernel on X. Then there exists a* ***unique??*** *RKHS whose kernel is k.*

[5] "The" is up to isometries.