The Moore-Aronszajn Theorem

Conventions. Unless stated otherwise, assume the following:

- $K \in \{\mathbb{R}, \mathbb{C}\}$.
- 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. RKHS's and Kernels

Definition 1.1 (p.s.d. kernels)**.** A positive semi-definite kernel on 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 $x_1, \dots, x_n \in X$ and any $\alpha_1, \dots, \alpha_n \in K$, we have

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

Definition 1.2 (r.k.'s and RKHS's)**.** Let $\mathscr{H}$ be a vector subspace of K^X and $k\colon X \times X \to K$ be such that for any $x \in X$, we have that $k_x \in \mathscr{H}$ and that it obeys the *reproducing property*, *i.e.*,[3]

$$\delta_x = \langle \cdot, k_x \rangle.$$

Then we say that "$\mathscr{H}$ is a *reproducing kernel Hilbert space* with a *reproducing kernel* k".

Remark. As is usual, we'll use the following terminology:

(i) "$\mathscr{H}$ is an RKHS of functions on X" iff $\mathscr{H}$ is a vector subspace of K^X and there exists a $k\colon X \times X \to K$ such that $\mathscr{H}$ becomes an RKHS with a reproducing kernel k.

(ii) "k is a reproducing kernel on X" iff there exists a Hilbert space $\mathscr{H}$ which is a vector subspace of K^X such that $\mathscr{H}$ becomes an RKHS with a reproducing kernel k.

Corollary 1.3 (Immediate consequences)**.**

(i) A reproducing kernel is a p.s.d. kernel.

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

(iii) Convergence in an RKHS $\implies$ *pointwise convergence.*

What about the converse?

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

Proof. (i) 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 $x_1, \dots, x_n \in X$ and $\alpha_1, \cdots, \alpha_n \in K$. Then

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

[3]Equivalently, $f(x) = \langle f, k_x \rangle$ for any $f \in \mathscr{H}$.

(ii) If k is a reproducing kernel for $\mathscr{H}$, then $\delta_x = \langle \cdot, k_x \rangle$, which is continuous. Conversely, if the evaluations are continuous, then by Riesz, we can define $k \colon X \times X \to K$ such that $k_x \in \mathscr{H}$ and $\delta_x = \langle \cdot, k_x \rangle$.

(iii) Follows from (ii).

(iv) 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, $k = k'$. □

Remark. The converse of (i) is the content of Theorem 2.3. Also note how completeness of $\mathscr{H}$ is used in "$\Leftarrow$" of **??** (ii).

Theorem 1.4 (RKHS $\mapsto$ r.k. is injective)**.** *Distinct RKHS's have distinct reproducing kernels.*

Proof. Let k be the reproducing kernel for $\mathscr{H} \subseteq K^X$. It suffices to show that $\mathscr{H}$ is uniquely determined (along with its inner product[4]) 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 $\overline{\mathscr{H}_0} = (\mathscr{H}_0^\perp)^\perp = \mathscr{H}$.[5] This determines $\mathscr{H}$ as a set, for a general element of $\mathscr{H}$ is a limit of the Cauchy sequence in $\mathscr{H}_0$, which is just its pointwise limit (which is existent due to **??** (ii) of Corollary 1.3). All that now remains is to determine is the norm on $\mathscr{H}$, which is determined once it's determined on its dense subset $\mathscr{H}_0$. Note that this will also determine the inner product on $\mathscr{H}_0$ (due to polarization).

Indeed, for $f = \sum_{i=1}^n \alpha_i k_{x_i}$, 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. Note how the completeness of $\mathscr{H}$ was used in writing $\overline{\mathscr{H}_0} = (\mathscr{H}_0^\perp)^\perp$.

Any example to show the necessity of this?

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

[5] A naïve glance suggests that this fixes $\mathscr{H}$ as a set, but it *doesn't*!—at least yet. 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 $\mathscr{H}$, which might still depend on the choice of $\mathscr{H}$, not just k.

We summarize our results so far:

$$\left\{\begin{matrix}\text{RKHS's of}\\ \text{functions on } X\end{matrix}\right\} \underset{\text{Corollary 1.3 (iv)}}{\overset{\text{Theorem 1.4}}{\longleftrightarrow}} \{\text{r.k.'s on } X\} \underset{\text{(inclusion)}}{\overset{\text{Corollary 1.3 (i)}}{\hookrightarrow}} \left\{\begin{matrix}\text{p.s.d. kernels}\\ \text{on } X\end{matrix}\right\}$$

To complete the circle of ideas, we show in the next section that any p.s.d. kernel is a reproducing kernel for some (and hence unique) RKHS which, among other things, will lead the inclusion above to become equality.

2. Moore-Aronszajn

The upcoming lemmas are geared towards the following goal: Given a p.s.d. kernel k on X, we find 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 on X. Define $\mathscr{H}_0$ to be the subspace of K^X generated by k_x's for $x \in X$. Then $\mathscr{H}_0$ admits a unique inner product with respect to which k has the reproducing property.*

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 Eq. (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 Eq. (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 $|f(x)| = |\langle f, k_x\rangle| \le \|f\|\|k_x\| = 0$. □

Lemma 2.2. *Continuing Lemma 2.1, let*[6] $S := \{\text{Cauchy sequences in } \mathscr{H}_0\}$. *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 $|f_m(x) - f_n(x)| = |\langle f_m - f_n, k_x\rangle| \le \|f_n - f_m\|\|k_x\| \overset{\text{w}}{\to} 0$ 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.

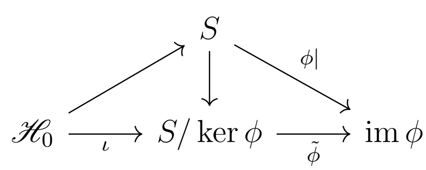
Linearity of ϕ is easy. We now compute $\ker\phi$. If $f_n \to 0$ in $\mathscr{H}_0$, then for any $x \in X$, we have $f_n(x) = \langle f_n, k_x\rangle \overset{\text{w}}{\to} 0$. Conversely, let (f_n) be a Cauchy sequence in $\mathscr{H}_0$ that converges to 0 pointwise. We show that it converges to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

Fix an N and write $f_N = \sum_i \alpha_i k_{x_i}$ for finitely many i's. Now,

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

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

Thus, we have the following commutative diagram:



The map $\mathscr{H}_0 \to S$ represents the function $f \mapsto (f, f, \ldots)$.

Now we make our final arguments:

(i) $\iota\colon f \mapsto \overline{(f, f, \ldots)}$ is a metric completion with the usual metric on $S/\ker\phi$, namely $d(\overline{(f_i)}, \overline{(g_i)}) = \lim_i d(f_i, g_i) \overset{\text{w}}{=} \lim_i \|f_i - g_i\|$.

(ii) Thus, the metric space $S/\ker\phi$ admits a (unique) Hilbert space structure such that ι becomes a norm completion with the norm recovering the metric.

(iii) The vector space structure thus endowed on $S/\ker\phi$ is precisely the one due to the algebraic quotient:

[6] “S” for “sequences”.

Note that any two generic elements of $S/\ker\phi$ are given by $\overline{(f_i)}$ and $\overline{(g_i)}$ where $(f_i), (g_i)$ are Cauchy sequences in $\mathscr{H}_0$. Note that $\iota(f_n) \overset{n}{\to} \overline{(f_i)}$ and $\iota(g_n) \overset{n}{\to} \overline{(g_i)}$ (easy). Continuity of addition and linearity of ι ensure that $\iota(f_n + g_n) \overset{n}{\to} \overline{(f_i)} + \overline{(g_i)}$. Finally, note that $\iota(f_n + g_n) \overset{n}{\to} \overline{(f_i + g_i)}$ as well so that we indeed have $\overline{(f_i)} + \overline{(g_i)} = \overline{(f_i + g_i)}$, which is precisely the definition of vector addition in the algebraic quotient. Similarly, one can verify for scalar multiplication.

(iv) $\tilde{\phi}$ is a vector space isomorphism.[7] Thus, the inner product on $S/\ker\phi$ can be transported to $\operatorname{im}\phi$ without altering the latter's vector space structure, making $\tilde{\phi}$ an isometric isomorphism.

(v) Note that $\tilde{\phi} \circ \iota$ is precisely the inclusion $\mathscr{H}_0 \hookrightarrow \operatorname{im}\phi$ (just traverse along the top arrows in the commutative diagram above) which is thus an isometric linear map.

(vi) $\operatorname{im}\phi$ is complete since $S/\ker\phi$ is, and thus is a Hilbert space.

(vii) Finally, we show that k still has the reproducing property on $\operatorname{im}\phi$:

Let $f \in \operatorname{im}\phi$ be the pointwise limit of the Cauchy sequence (f_i) in $\mathscr{H}_0$. Then $f_i \to f$ in $\operatorname{im}\phi$ as well:

Note that $f = \tilde{\phi}(\,\overline{(f_i)}\,)$ and $f_j = \tilde{\phi} \circ \iota(f_j)$. Thus it suffices to have $\iota(f_j) \overset{j}{\to} \overline{(f_i)}$ in $S/\ker\phi$ which is indeed true.

Thus, for any $x \in X$, one has $\langle f, k_x \rangle = \lim_i \langle f_i, k_x \rangle = \lim_i f_i(x) = f(x)$ as claimed.

We have thus constructed a Hilbert space, namely $\operatorname{im}\phi$, whose reproducing kernel is precisely k, proving the following:

Theorem 2.3 (Moore-Aronszajn). *Any p.s.d. kernel is a reproducing kernel.*

3. A Toy Example

3.1 For finite X

Here, we ask the question: *If k is a p.s.d. kernel on a finite X, what is the associated RKHS?*

[7] With respect to the algebraic vector space structure on $S/\ker\phi$, not neceassrily the vector space structure coming from completion. Thus, it was crucial to show that these two structures are exactly the same.

Without loss of generality, let $X = \{1, \ldots, n\}$ for $n \in \mathbb{N}$. Now, any p.s.d. kernel k on X is simply a p.s.d. and conjugate symmetric $n \times n$ matrix and our inner product space $\mathscr{H}_0$ (in the language of Lemma 2.1) is the column space $\operatorname{col}(k)$ of k. It's complete being finite dimensional and thus itself is the required RKHS. In this case, Eq. (2.1) becomes

$$\begin{aligned}
\langle k\alpha, k\beta \rangle &= \sum_{i,j=1}^{n} \alpha_i k(j,i) \overline{\beta_j} \\
&= \sum_{i,j=1}^{n} \alpha_i \overline{k(i,j)\beta_j} \\
&= \langle \alpha, k\beta \rangle_e
\end{aligned}$$

where $\langle \cdot, \cdot \rangle_e$ denotes the usual Euclidean inner product on K^n.

Noting that $\operatorname{col} k = K^n$ for positive definite k's we recover the familiar correspondence between inner products on K^n and positive definite $n \times n$ matrices.

3.2 Finite-dimensional spaces

Let $\mathscr{H}$ be finite-dimensional subspace of K^X. We ask: *Is V an RKHS? If so, what is the associated reproducing kernel?*

The answer is an easy yes. Suppose it indeed is with the reproducing kernel being k. Let $f_1, \ldots, f_n$ form an orthonormal basis for $\mathscr{H}$. Then we must have

$$\begin{aligned}
k_x &= \sum_{i=1}^{n} \langle k_x, f_i \rangle f_i \\
&= \sum_{i=1}^{n} \overline{f_i(x)} f_i
\end{aligned}$$

yielding

$$k(x,y) = \sum_{i=1}^{n} f_i(x) \overline{f_i(y)}. \tag{3.1}$$

Now, it's straightforward to check that k *defined* by Eq. (3.1) indeed is a reproducing kernel for $\mathscr{H}$:

- Firstly, note that each $k_x \in \mathscr{H}$ being a linear combination of f_i's.

- Secondly, for $g \in \mathscr{H}$, we have $\langle g, k_x \rangle = \sum_i f_i(x)\langle g, f_i \rangle = \left(\sum_i \langle g, f_i \rangle f_i\right)(x) = g(x)$.

Taking $\mathscr{H}$ to be the set of all polynomials of degree less than or equal to an $n \in \mathbb{N}$ with the inner product being such that the monomials form an orthonormal basis, Eq. (3.1) becomes

$$k(x, y) = \begin{cases} \frac{1-(x\overline{y})^{n+1}}{1-x\overline{y}}, & x\overline{y} \neq 1 \\ n+1, & \text{otherwise} \end{cases}.$$

4. References

1. *Aronszajn's theorem* by Jean-Philippe Vert. Link.
2. *Uniqueness of the RKHS* by Jean-Philippe Vert. Link.