C-MInDS Winter Workshop, IIT Bombay **5–7 February 2026**

A Function-Space Tour of Data Science: Hilbert Spaces and Kernel Methods

`rahul@ucsd.edu`

Disclaimer: *These notes carry no warranty of correctness, completeness, or proper citation of sources.*

1 Riesz Representation Theorem

Theorem 1.1 (Riesz Representation)**.** *Let $(\mathcal{F}, \langle\cdot,\cdot\rangle)$ be a (real) Hilbert space and let $\Phi : \mathcal{F} \to \mathbb{R}$ be linear. If Φ is bounded (or, equivalently, continuous), i.e., there exists $B > 0$ such that*

$$|\Phi(f)| \leq B\,\|f\|_{\mathcal{F}} \text{ for all } f \in \mathcal{F}, \tag{1}$$

then there exists a unique *$\varphi \in \mathcal{F}$ such that*

$$\Phi(f) = \langle \varphi, f\rangle \text{ for all } f \in \mathcal{F}. \tag{2}$$

Remark 1.2. We refer to $\varphi \in \mathcal{F}$ as the *Riesz representer* of the functional $\Phi : \mathcal{F} \to \mathbb{R}$.

2 Moore–Aronszajn Theorem

Let $\Omega \subseteq \mathbb{R}^d$ be a domain and let $k : \Omega \times \Omega \to \mathbb{R}$ be a kernel.

Definition 2.1 (Valid kernel)**.** A kernel k is called *valid* if it is symmetric and positive semidefinite, i.e.,

$$k(x, x') = k(x', x), \quad \text{and} \quad \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j\, k(x_i, x_j) \geq 0, \tag{3}$$

for all $n \in \mathbb{N}$, all $x_1, \ldots, x_n \in \Omega$, and all $a_1, \ldots, a_n \in \mathbb{R}$.

Theorem 2.2 (Moore–Aronszajn)**.** *For every valid kernel $k : \Omega \times \Omega \to \mathbb{R}$, there exists a unique reproducing kernel Hilbert space (RKHS) $(\mathcal{F}, \langle\cdot,\cdot\rangle)$ of functions $\Omega \to \mathbb{R}$ such that for every $x \in \Omega$,*

$$k(\cdot, x) \in \mathcal{F}, \qquad \text{and} \qquad \langle k(\cdot, x), f\rangle = f(x) \text{ for all } f \in \mathcal{F}. \tag{4}$$

Proof. Define

$$\mathcal{F}_0 \coloneqq \operatorname{span}\{k(\cdot, x) \,:\, x \in \Omega\} = \left\{ \sum_{i=1}^{n} a_i\, k(\cdot, x_i) \,:\, n \in \mathbb{N},\ a_i \in \mathbb{R},\ x_i \in \Omega \right\} = \big\{\text{kernel machines}\big\}. \tag{5}$$

Define the inner product between two kernels as

$$\langle k(\cdot, x), k(\cdot, x')\rangle \coloneqq k(x, x'). \tag{6}$$

Next, we extend this inner product bilinearly to all of $\mathcal{F}_0$. In particular, for $f = \sum_{i=1}^{n} a_i k(\cdot, x_i)$ and $g = \sum_{j=1}^{m} b_j k(\cdot, x'_j)$, we have that[1]

$$\langle f, g\rangle \coloneqq \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j\, k(x_i, x'_j). \tag{7}$$

[1] It can be shown that this bilinear extension is independent of the representation used to implement f and g (if multiple representations exist).

With $f \in \mathcal{F}_0$ as above, observe that the reproducing property holds on $\mathcal{F}_0$ since

$$\langle k(\cdot,x), f\rangle = \sum_{i=1}^{n} a_i \langle k(\cdot,x), k(\cdot,x_i)\rangle = \sum_{i=1}^{n} a_i\, k(x,x_i) = f(x). \tag{8}$$

The validity of k (symmetry and positive semidefiniteness) ensures that this satisfies the properties of an inner product. In particular, if we define $\|f\|_{\mathcal{F}} \coloneqq \sqrt{\langle f,f\rangle}$ for $f \in \mathcal{F}_0$, the Cauchy–Schwarz inequality holds. One nuance that often goes overlooked is how the kernel being merely positive semidefinite is enough to ensure that that inner product is positive definite. In other words, we need to ensure that for $f \in \mathcal{F}_0$, if $\langle f,f\rangle = 0$, then $f = 0$. Suppose that $\langle f,f\rangle = 0$. Then,

$$|f(x)|^2 = |\langle f, k(\cdot,x)\rangle|^2 \le \|f\|_{\mathcal{F}}^2 \|k(\cdot,x)\|_{\mathcal{F}}^2 = \langle f,f\rangle\, k(x,x) = 0, \tag{9}$$

which implies that $f = 0$.

In general $\mathcal{F}_0$ may not be complete. Let $\mathcal{F}$ denote the completion[2] of $\mathcal{F}_0$ under $\|\cdot\|_{\mathcal{F}}$, i.e.,

$$\mathcal{F} \coloneqq \overline{\mathcal{F}_0}^{\|\cdot\|_{\mathcal{F}}} \coloneqq \left\{ f : \Omega \to \mathbb{R} : \begin{array}{l} \{f_n\}_{n\in\mathbb{N}} \text{ is Cauchy with respect to } \|\cdot\|_{\mathcal{F}} \\ f_n \to f \text{ as } n \to \infty \text{ in } \|\cdot\|_{\mathcal{F}} \end{array} \right\}. \tag{10}$$

The reproducing property then extends to $\mathcal{F}$ by continuity.

To complete the proof, suppose that there exists a larger RKHS $\mathcal{G} \supset \mathcal{F}$ whose reproducing kernel is $k(\cdot,\cdot)$. Since $\mathcal{F}$ is closed (by nature of being a completion), there exists a complementary subspace $\mathcal{F}^{\perp}$ in $\mathcal{G}$ such that

$$\mathcal{G} = \mathcal{F} \oplus \mathcal{F}^{\perp}. \tag{11}$$

Thus, for every $f \in \mathcal{G}$, we can write $f = f_1 + f_2$, where $f_1 \in \mathcal{F}$ and $f_2 \in \mathcal{F}^{\perp}$. For any $f \in \mathcal{G}$, we have that

$$\langle k(\cdot,x), f\rangle = \langle k(\cdot,x), f_1 + f_2\rangle = \langle k(\cdot,x), f_1\rangle + \langle k(\cdot,x), f_2\rangle = f_1(x) + 0. \tag{12}$$

In other words $\mathcal{F}^{\perp} = \{0\}$ and so $\mathcal{G} = \mathcal{F}$. □

Remark 2.3. We have a bijection between (valid) kernels and RKHSs. Every RKHS has a unique kernel (Riesz Representation Theorem) and every kernel induces a unique RKHS (Moore-Aronszajn Theorem).

Exercise 2.4. *This exercise justifies the completion step in* (10). *Suppose* $\{f_n\}_{n\in\mathbb{N}} \subset \mathcal{F}_0$ *is Cauchy in* $\|\cdot\|_{\mathcal{F}}$*:*

$$\|f_m - f_n\|_{\mathcal{F}} \to 0 \text{ as } \min\{m,n\} \to \infty. \tag{13}$$

Prove that, for each $x \in \Omega$*, the scalar sequence* $\{f_n(x)\}$ *is Cauchy in* $\mathbb{R}$ *(with the usual metric), and hence defines a pointwise limit*

$$f(x) \coloneqq \lim_{n\to\infty} f_n(x). \tag{14}$$

Proof. For each $x \in \Omega$,

$$\begin{aligned} |f_m(x) - f_n(x)| &= |\langle k(\cdot,x), f_m - f_n\rangle| \\ &\le \|k(\cdot,x)\|_{\mathcal{F}} \|f_m - f_n\|_{\mathcal{F}} \\ &= \sqrt{k(x,x)}\, \|f_m - f_n\|_{\mathcal{F}}. \end{aligned} \tag{15}$$

Thus, the scalar sequence $\{f_n(x)\}$ is Cauchy in $\mathbb{R}$ (since $\sqrt{k(x,x)} = \|k(\cdot,x)\|_{\mathcal{F}} < +\infty$ for each $x \in \Omega$). □

[2]Technically, the completion is made up from equivalence classes of Cauchy sequences and arguing that pointwise limits exist as in Exercise 2.4.

3 The Representer Theorem

Let $(\mathcal{F}, \langle\cdot,\cdot\rangle)$ be an RKHS on Ω with kernel $k(\cdot,\cdot)$. Given data $\{(x_i, y_i)\}_{i=1}^n \subset \Omega \times \mathbb{R}$ and a loss function $\mathcal{L} : \mathbb{R} \times \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$, consider the objective

$$J(f) \coloneqq \sum_{i=1}^{n} \mathcal{L}(y_i, f(x_i)) + \lambda \|f\|_{\mathcal{F}}^2, \quad \lambda > 0. \tag{16}$$

We are interested in studying the optimization problem

$$\min_{f \in \mathcal{F}} J(f). \tag{17}$$

Before focusing on the case when J is given by (16), we first investigate general ideas to ensure existence of minimizers to problems of the form (17).

3.1 Existence of Minimizers in Infinite Dimensions

In infinite-dimensional normed spaces, closed and bounded sets are generally not compact in the norm topology (corollary of Riesz's lemma), so "direct method" existence proofs often switch to weaker topologies where bounded sets can be compact.

Definition 3.1 (Strong and weak convergence)**.** Let $\mathcal{F}$ be a Hilbert space. A sequence $\{f_n\}_{n\in\mathbb{N}} \subset \mathcal{F}$ converges *strongly* (in norm) to $f \in \mathcal{F}$ if

$$\|f_n - f\| \to 0 \text{ as } n \to \infty.$$

A sequence $\{f_n\}_{n\in\mathbb{N}} \subset \mathcal{F}$ converges *weakly* to $f \in \mathcal{F}$ if

$$\langle f_n, g\rangle \to \langle f, g\rangle \text{ as } n \to \infty \text{ for all } g \in \mathcal{F}.$$

Remark 3.2. Strong convergence implies weak convergence, but the reverse is not true. A standard counterexample is the canonical basis $\{e_n\}_{n\in\mathbb{N}}$ of $\ell^2(\mathbb{N})$. We have that $\langle e_n, g\rangle \to 0$ as $n \to \infty$ for all $g \in \ell^2(\mathbb{N})$, but $\|e_n - 0\|_{\ell^2} = 1$ for all $n \in \mathbb{N}$.

Fact 3.3. *All (strongly) continuous* linear *functionals on a Hilbert space are weakly continuous.*

Fact 3.4 (special case of the Banach–Alaoglu Theorem)**.** *The closed unit ball*

$$B_{\mathcal{F}} \coloneqq \{f \in \mathcal{F} : \|f\| \le 1\} \tag{18}$$

of a Hilbert space $\mathcal{F}$ is weakly compact.

Remark 3.5 (Direct method in the calculus of variations). To establish existence of minimizers to (17), it suffices[3] to show that J is proper and there exists some topology τ for which

1. J is τ-lower-semicontinuous (τ-l.s.c.) and
2. J has τ-compact sublevel sets.

Indeed, properness ensures that the effective domain of J, denoted by

$$\operatorname{dom} J \coloneqq \{f \in \mathcal{F} : J(f) < +\infty\} \tag{19}$$

is nonempty. Next, let α denote the infimal value of (17). Pick any minimizing sequence $\{f_n\}_{n\in\mathbb{N}}$ such that $J(f_n) \downarrow \alpha$. For $\beta > \alpha$, there exists $N \in \mathbb{N}$ such that for any $n \ge N$, $J(f_n) \le \beta$, i.e., eventually, f_n lies in the *sublevel set*

$$S_\beta \coloneqq \{f \in \mathcal{F} : J(f) \le \beta\}. \tag{20}$$

Since S_β is τ-compact, we can extract a convergent subsequence from $\{f_n\}_{n\ge N}$ denoted by $\{f_{n_j}\}_{j\in\mathbb{N}}$ such that $f_{n_j} \to f^\star$, for some $f^\star \in S_\beta$. Since J is τ-l.s.c., we have that

$$J(f^\star) \le \liminf_{j\to\infty} J(f_{n_j}) = \alpha, \tag{21}$$

so $f^\star$ is a minimizer.

[3] Once you already have τ-compact sublevel sets (by any method), you do not need to assume that the objective is bounded from below as an additional assumption.

3.2 The RKHS Representer Theorem

We now return to J of the form (16) and prove the famous RKHS representer theorem.

Theorem 3.6 (Representer Theorem)**.** *If, for every $y \in \mathbb{R}$, the loss function $\mathcal{L}(y, \cdot) : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ is lower semicontinuous (l.s.c.), proper, and bounded from below, the solution set*

$$S \coloneqq \operatorname*{arg\,min}_{f \in \mathcal{F}} \left\{ J(f) = \sum_{i=1}^{n} \mathcal{L}(y_i, f(x_i)) + \lambda \|f\|_{\mathcal{F}}^2 \right\} \tag{22}$$

is nonempty. Furthermore, every $f^\star \in S$ is of the form

$$f^\star = \sum_{i=1}^{n} a_i k(\cdot, x_i). \tag{23}$$

In particular, if $\mathcal{L}(y, \cdot) : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ is also convex, then S is singleton.

Proof. Since $\mathcal{L}(y, \cdot)$ is l.s.c., J is weakly l.s.c. Indeed,

- the point evaluation $f \mapsto f(x)$ is weakly continuous for all $x \in \Omega$ and
- the norm $\|\cdot\|_{\mathcal{F}}$ is weakly l.s.c. on $\mathcal{F}$ since

$$\|f\|_{\mathcal{F}} = \sup_{\substack{g \in \mathcal{F} \\ \|g\|_{\mathcal{F}} = 1}} |\langle f, g \rangle|, \tag{24}$$

  where we recall that the sup of weakly continuous functions is weakly l.s.c.[4]

Since $\mathcal{L}(y, \cdot)$ is bounded from below, J is coercive.[5] Indeed, J is the sum of something bounded from below and something coercive (all norms are obviously coercive). Since J is weakly l.s.c., it has weakly closed sublevel sets. Furthermore, since J is coercive, it has bounded sublevel sets. Thus, it has weakly compact sublevel sets and hence minimizers exist.

To establish the form of any solution, we first note that $\operatorname{span}\{k(\cdot, x_i)\}_{i=1}^n$ is a closed subspace of $\mathcal{F}$. Hence, for any $f \in \mathcal{F}$, we can write

$$f = f_1 + f_2 = \begin{cases} f_1 \in \operatorname{span}\{k(\cdot, x_i)\}_{i=1}^n \\ f_2 \perp \operatorname{span}\{k(\cdot, x_i)\}_{i=1}^n \end{cases} \tag{25}$$

Since $f_2(x_i) = \langle k(\cdot, x_i), f_2 \rangle = 0$ for $i = 1, \dots, n$, we have that

$$\sum_{i=1}^{n} \mathcal{L}(y_i, f(x_i)) = \sum_{i=1}^{n} \mathcal{L}(y_i, f_1(x_i)), \tag{26}$$

i.e., the data-fitting term is "blind" to the f_2-component. Thus, only the regularization term penalizes the f_2-component in the objective. Suppose that $f^\star \in \mathcal{F}$ is a solution with (25) decomposition $f^\star = f_1^\star + f_2^\star$ such that $f_2^\star \neq 0$ in which case $\|f_2\|_{\mathcal{F}} > 0$. Then, $\|f^\star\|_{\mathcal{F}}^2 = \|f_1^\star + f_2^\star\|_{\mathcal{F}}^2 = \|f_1^\star\|_{\mathcal{F}}^2 + \|f_2^\star\|_{\mathcal{F}}^2$ which implies $\|f_1^\star\|_{\mathcal{F}} < \|f^\star\|_{\mathcal{F}}$, a contradiction. Therefore, every solution $f^\star \in \operatorname{span}\{k(\cdot, x_i)\}_{i=1}^n$, i.e.,

$$f^\star = \sum_{i=1}^{n} a_i k(\cdot, x_i). \tag{27}$$

When $\mathcal{L}(y, \cdot)$ is convex, J is strictly convex as the sum of something convex and something strictly convex (since norms on Hilbert spaces are always strictly convex). Thus, the solution is unique. $\square$

---

[4] In any topological space, the sup of continuous functions is l.s.c.

[5] Recall that coercivity means $J(f) \to \infty$ as $\|f\|_{\mathcal{F}} \to \infty$.