Foundations of Reproducing Kernel Hilbert Spaces – Advanced Topics in Machine Learning

Zoltán Szabó

slides: http://www.gatsby.ucl.ac.uk/~szabo/teaching.html

Gatsby Unit

March 22, 2016

Overview

1. Concepts from functional analysis:
   - normed-, inner product space,
   - convergent-, Cauchy sequence,
   - complete spaces: Banach-, Hilbert space,
   - continuous/bounded linear operators.

Overview

2. RKHS:
   - different views:
     1. continuous evaluation functional,
     2. reproducing kernel,
     3. positive definite function,
     4. feature view (kernel).
   - equivalence, explicit construction.

Normed space

We define the 'length' of a vector.

$\mathcal{F}$: vector space over $\mathbb{R}$. $\|\cdot\| : \mathcal{F} \to [0, \infty)$ is norm on $\mathcal{F}$, if

1. $\|f\| = 0$ iff. $f = 0$ (norm separates points),

Normed space

We define the 'length' of a vector.

$\mathcal{F}$: vector space over $\mathbb{R}$. $\|\cdot\| : \mathcal{F} \to [0, \infty)$ is norm on $\mathcal{F}$, if

1. $\|f\| = 0$ iff. $f = 0$ (norm separates points),
2. $\|\lambda f\| = |\lambda| \, \|f\| \; \forall \lambda \in \mathbb{R}, \forall f \in \mathcal{F}$ (positive homogenity),

Normed space

We define the 'length' of a vector.

$\mathcal{F}$: vector space over $\mathbb{R}$. $\|\cdot\| : \mathcal{F} \to [0, \infty)$ is norm on $\mathcal{F}$, if

1. $\|f\| = 0$ iff. $f = 0$ (norm separates points),
2. $\|\lambda f\| = |\lambda| \, \|f\| \; \forall \lambda \in \mathbb{R}, \forall f \in \mathcal{F}$ (positive homogenity),
3. $\|f + g\| \leq \|f\| + \|g\| \; \forall f, g \in \mathcal{F}$ (triangle inequality).

Normed space

We define the 'length' of a vector.

$\mathcal{F}$: vector space over $\mathbb{R}$. $\|\cdot\| : \mathcal{F} \to [0, \infty)$ is norm on $\mathcal{F}$, if

1. $\|f\| = 0$ iff. $f = 0$ (norm separates points),
2. $\|\lambda f\| = |\lambda| \, \|f\| \; \forall \lambda \in \mathbb{R}, \forall f \in \mathcal{F}$ (positive homogenity),
3. $\|f + g\| \leq \|f\| + \|g\| \; \forall f, g \in \mathcal{F}$ (triangle inequality).

Note:

- norm $\Rightarrow$ metric: $d(f, g) = \|f - g\| \Rightarrow$
- study continuity, convergence.

Normed space: examples

- $(\mathbb{R}, |\cdot|)$,
- $\left(\mathbb{R}^d, \|\mathbf{x}\|_p = [\sum_i |x_i|^p]^{\frac{1}{p}}\right)$, $1 \le p$.
  - $p = 1$: $\|\mathbf{x}\|_1 = \sum_i |x_i|$ (Manhattan),
  - $p = 2$: $\|\mathbf{x}\|_2 = \sqrt{\sum_i x_i^2}$ (Euclidean),
  - $p = \infty$: $\|\mathbf{x}\|_\infty = \max_i |x_i|$ (maximum norm).
- $\left(C[a, b], \|f\|_p = \left[\int_a^b |f(x)|^p \mathrm{d}x\right]^{\frac{1}{p}}\right)$, $1 \le p$.

Inner product space

$\mathcal{F}$: vector space over $\mathbb{R}$. $\langle \cdot, \cdot \rangle : \mathcal{F} \times \mathcal{F} \to \mathbb{R}$ is an inner product on $\mathcal{F}$ if for $\forall \alpha_i \in \mathbb{R}, f_i, f, g \in \mathcal{F}$

1. $\langle \alpha_1 f_1 + \alpha_2 f_2, g \rangle = \alpha_1 \langle f_1, g \rangle + \alpha_2 \langle f_2, g \rangle$ (linearity),

Inner product space

$\mathcal{F}$: vector space over $\mathbb{R}$. $\langle \cdot, \cdot \rangle : \mathcal{F} \times \mathcal{F} \to \mathbb{R}$ is an inner product on $\mathcal{F}$ if for $\forall \alpha_i \in \mathbb{R}, f_i, f, g \in \mathcal{F}$

1. $\langle \alpha_1 f_1 + \alpha_2 f_2, g \rangle = \alpha_1 \langle f_1, g \rangle + \alpha_2 \langle f_2, g \rangle$ (linearity),
2. $\langle f, g \rangle = \langle g, f \rangle$ (symmetry),

Inner product space

$\mathcal{F}$: vector space over $\mathbb{R}$. $\langle \cdot, \cdot \rangle : \mathcal{F} \times \mathcal{F} \to \mathbb{R}$ is an inner product on $\mathcal{F}$ if for $\forall \alpha_i \in \mathbb{R}, f_i, f, g \in \mathcal{F}$

1. $\langle \alpha_1 f_1 + \alpha_2 f_2, g \rangle = \alpha_1 \langle f_1, g \rangle + \alpha_2 \langle f_2, g \rangle$ (linearity),
2. $\langle f, g \rangle = \langle g, f \rangle$ (symmetry),
3. $\langle f, f \rangle \geq 0$; $\langle f, f \rangle = 0 \Leftrightarrow f = 0$.

Inner product space

$\mathcal{F}$: vector space over $\mathbb{R}$. $\langle\cdot,\cdot\rangle : \mathcal{F} \times \mathcal{F} \to \mathbb{R}$ is an inner product on $\mathcal{F}$ if for $\forall \alpha_i \in \mathbb{R}, f_i, f, g \in \mathcal{F}$

1. $\langle \alpha_1 f_1 + \alpha_2 f_2, g\rangle = \alpha_1 \langle f_1, g\rangle + \alpha_2 \langle f_2, g\rangle$ (linearity),
2. $\langle f, g\rangle = \langle g, f\rangle$ (symmetry),
3. $\langle f, f\rangle \geq 0$; $\langle f, f\rangle = 0 \Leftrightarrow f = 0$.

Notes:

- 1, 2 $\Rightarrow$ bilinearity.
- inner product $\Rightarrow$ norm: $\|f\| = \sqrt{\langle f, f\rangle}$.
- 1,2,3' ($\langle f, f\rangle \geq 0$) is called semi-inner product.

Inner product space: examples

- $\left(\mathbb{R}^d, \langle \mathbf{x}, \mathbf{y} \rangle = \sum_i x_i y_i\right)$.
- $\left(\mathbb{R}^{d_1 \times d_2}, \langle \mathbf{A}, \mathbf{B} \rangle_F = tr(\mathbf{A}^T \mathbf{B}) = \sum_{ij} A_{ij} B_{ij}\right)$.
- $\left(C[a,b], \langle f, g \rangle = \int_a^b f(x) g(x) \mathrm{d}x\right)$.

Norm vs inner product

Relations:

- $|\langle f, g\rangle| \leq \|f\| \cdot \|g\|$ (CBS),
- $4\langle f, g\rangle = \|f+g\|^2 - \|f-g\|^2$ (polarization identity),
- $\|f+g\|^2 + \|f-g\|^2 = 2\|f\|^2 + 2\|g\|^2$ (parallelogram law).

Norm vs inner product

Relations:

- $|\langle f, g \rangle| \leq \|f\| \cdot \|g\|$ (CBS),
- $4\langle f, g \rangle = \|f+g\|^2 - \|f-g\|^2$ (polarization identity),
- $\|f+g\|^2 + \|f-g\|^2 = 2\|f\|^2 + 2\|g\|^2$ (parallelogram law).

Notes:

- CBS holds for semi-inner products.
- parallelogram law = characterization of '$\|\cdot\| \leftarrow \langle \cdot, \cdot \rangle$'.

Convergent-, Cauchy sequence

$\mathcal{F}$: normed space, $\{f_n\}_{n=1}^{\infty} \subset \mathcal{F}, f \in \mathcal{F}$,

- Convergent sequence: $f_n \xrightarrow{\mathcal{F}} f$ if $\forall \epsilon > 0$ $\exists N = N(\varepsilon) \in \mathbb{N}$, s.t. $\forall n \geq N$, $\|f_n - f\|_{\mathcal{F}} < \epsilon$.

Convergent-, Cauchy sequence

$\mathcal{F}$: normed space, $\{f_n\}_{n=1}^{\infty} \subset \mathcal{F}, f \in \mathcal{F}$,

- Convergent sequence: $f_n \xrightarrow{\mathcal{F}} f$ if $\forall \epsilon > 0$ $\exists N = N(\varepsilon) \in \mathbb{N}$, s.t. $\forall n \geq N$, $\|f_n - f\|_{\mathcal{F}} < \epsilon$.
- Cauchy sequence: $\{f_n\}_{n=1}^{\infty}$ is a Cauchy sequence if $\forall \epsilon > 0$ $\exists N = N(\varepsilon) \in \mathbb{N}$, s.t. $\forall n, m \geq N$, $\|f_n - f_m\|_{\mathcal{F}} < \epsilon$.

Convergent-, Cauchy sequence

$\mathcal{F}$: normed space, $\{f_n\}_{n=1}^{\infty} \subset \mathcal{F}, f \in \mathcal{F}$,

- Convergent sequence: $f_n \xrightarrow{\mathcal{F}} f$ if $\forall \epsilon > 0$ $\exists N = N(\varepsilon) \in \mathbb{N}$, s.t. $\forall n \geq N$, $\|f_n - f\|_{\mathcal{F}} < \epsilon$.
- Cauchy sequence: $\{f_n\}_{n=1}^{\infty}$ is a Cauchy sequence if $\forall \epsilon > 0$ $\exists N = N(\varepsilon) \in \mathbb{N}$, s.t. $\forall n, m \geq N$, $\|f_n - f_m\|_{\mathcal{F}} < \epsilon$.

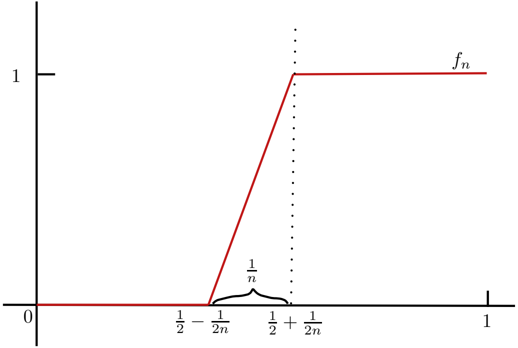
Note:

- convergent $\Rightarrow$ Cauchy: $\|f_n - f_m\|_{\mathcal{F}} \leq \|f_n - f\|_{\mathcal{F}} + \|f - f_m\|_{\mathcal{F}}$.

Not every Cauchy sequence converges

Examples:

- $1, 1.4, 1.41, 1.414, 1.4142, \ldots$: Cauchy in $\mathbb{Q}$, but $\sqrt{2} \notin \mathbb{Q}$.
- $\left(C[0,1], \|\cdot\|_{L^2[0,1]}\right)$:



But a Cauchy sequence is bounded.

Banach space, Hilbert space

- Complete space: $\forall$ Cauchy sequence converges.

Banach space, Hilbert space

- Complete space: $\forall$ Cauchy sequence converges.
- Banach space = complete normed space, e.g.
  1. Let $p \in [1, \infty)$, $L^p(\mathcal{X}, \mathcal{A}, \mu) :=$

$$\left\{ f : (\mathcal{X}, \mathcal{A}) \to \mathbb{R} \text{ measurable} : \|f\|_p = \left[\int_{\mathcal{X}} |f(x)|^p \mathrm{d}\mu(x)\right]^{1/p} < \infty \right\}.$$

  2. $\left(C[a,b], \|f\|_\infty = \max_{x \in [a,b]} |f(x)|\right)$.

Banach space, Hilbert space

- Complete space: $\forall$ Cauchy sequence converges.
- Banach space = complete normed space, e.g.
  1. Let $p \in [1, \infty)$, $L^p(\mathcal{X}, \mathcal{A}, \mu) :=$
  $$\left\{ f : (\mathcal{X}, \mathcal{A}) \to \mathbb{R} \text{ measurable} : \|f\|_p = \left[\int_{\mathcal{X}} |f(x)|^p \mathrm{d}\mu(x) \right]^{1/p} < \infty \right\}.$$
  2. $\left(C[a, b], \|f\|_\infty = \max_{x \in [a,b]} |f(x)|\right)$.
- Hilbert space = complete inner product space; $L^2(\mathcal{X}, \mathcal{A}, \mu)$.

Linear-, bounded operator

$\mathcal{F}$, $\mathcal{G}$: normed spaces. $A : \mathcal{F} \to \mathcal{G}$ is called

- linear operator:
  1. $A(\alpha f) = \alpha (Af) \quad \forall \alpha \in \mathbb{R}, f \in \mathcal{F}$, (homogeneity),
  2. $A(f+g) = Af + Ag \quad \forall f, g \in \mathcal{F}$ (additivity).

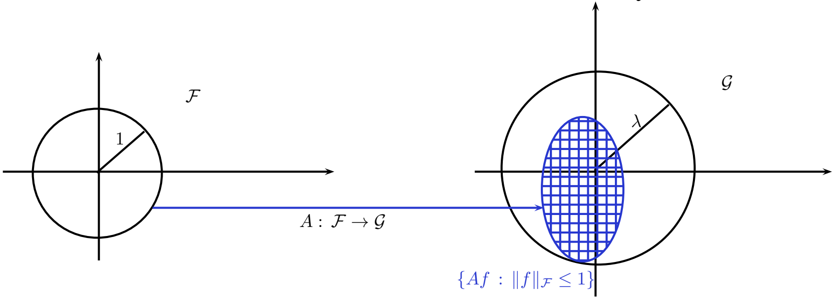
  $\mathcal{G} = \mathbb{R}$: linear functional.

Linear-, bounded operator

$\mathcal{F}$, $\mathcal{G}$: normed spaces. $A : \mathcal{F} \to \mathcal{G}$ is called

- linear operator:
  1. $A(\alpha f) = \alpha (Af) \quad \forall \alpha \in \mathbb{R}, f \in \mathcal{F}$, (homogeneity),
  2. $A(f+g) = Af + Ag \quad \forall f, g \in \mathcal{F}$ (additivity).

  $\mathcal{G} = \mathbb{R}$: linear functional.
- bounded operator: A is linear & $\|A\| = \sup_{f \in \mathcal{F}} \frac{\|Af\|_{\mathcal{G}}}{\|f\|_{\mathcal{F}}} < \infty$.



Unbounded linear functional: example

$\left(C^1[0,1], \|f\|_\infty := \max_{x\in[0,1]} |f(x)|\right)$, $A(f) = f'(0) \in \mathbb{R}$:

1. A: linear $\Leftarrow$ differentiation & evaluation are linear,
2. $f_n(x) = e^{-nx}$ $(n \in \mathbb{Z}^+)$:
   - $\|f_n\|_\infty \le 1$, but
   - $|A(f_n)| = |f_n'(0)| = \left| -ne^{-nx}\big|_{x=0}\right| = |-n| = n \to \infty$.

Continuous operator

- Def.: A is
  - continuous at $f_0 \in \mathcal{F}$: $\forall \epsilon > 0$ $\exists \delta = \delta(\epsilon, f_0) > 0$, s.t.

$$\|f - f_0\|_{\mathcal{F}} < \delta \qquad \text{implies} \qquad \|Af - Af_0\|_{\mathcal{G}} < \epsilon.$$

  - continuous: if it is continuous at $\forall f_0 \in \mathcal{F}$.

Continuous operator

- Def.: A is
  - continuous at $f_0 \in \mathcal{F}$: $\forall \epsilon > 0\ \exists \delta = \delta(\epsilon, f_0) > 0$, s.t.

$$\|f - f_0\|_{\mathcal{F}} < \delta \qquad \text{implies} \qquad \|Af - Af_0\|_{\mathcal{G}} < \epsilon.$$

  - continuous: if it is continuous at $\forall f_0 \in \mathcal{F}$.

- Example:
  - Let $A_g(f) := \langle f, g\rangle_{\mathcal{F}} \in \mathbb{R}$, where $f, g \in \mathcal{F}$.
  - A_g is Lipschitz continuous:

$$\left|A_g(f_1) - A_g(f_2)\right| \overset{\langle\cdot,\cdot\rangle_{\mathcal{F}}\text{: lin.}}{=} |\langle f_1 - f_2, g\rangle_{\mathcal{F}}| \overset{\text{CBS}}{\leq} \|g\|_{\mathcal{F}} \|f_1 - f_2\|_{\mathcal{F}}.$$

Continuous-bounded relations

Theorem:

- A: linear operator. Equivalent: A is
  1. continuous,
  2. continuous at one point,
  3. bounded.

Continuous-bounded relations

Theorems:

- A: linear operator. Equivalent: A is
  1. continuous,
  2. continuous at one point,
  3. bounded.
- Riesz representation ($\mathcal{F}$: Hilbert, $\mathcal{G} = \mathbb{R}$):

$$\boxed{\text{continuous linear functionals} = \{\langle \cdot, g \rangle_{\mathcal{F}} : g \in \mathcal{F}\}}.$$

Let us switch to RKHS-s!

View-1: continuous evaluation.

- Let $\mathcal{H} \subset \mathbb{R}^{\mathcal{X}}$ be a Hilbert space.
- Consider for fixed $x \in \mathcal{X}$ the $\delta_x : f \in \mathcal{H} \mapsto f(x) \in \mathbb{R}$ map.

View-1: continuous evaluation.

- Let $\mathcal{H} \subset \mathbb{R}^{\mathcal{X}}$ be a Hilbert space.
- Consider for fixed $x \in \mathcal{X}$ the $\delta_x : f \in \mathcal{H} \mapsto f(x) \in \mathbb{R}$ map.
- The (Dirac) evaluation functional is linear:

$$\begin{aligned}\delta_x(\alpha f + \beta g) &= (\alpha f + \beta g)(x) = \alpha f(x) + \beta g(x)\\ &= \alpha \delta_x(f) + \beta \delta_x(g) \quad (\forall \alpha, \beta \in \mathbb{R}, f, g \in \mathcal{H}).\end{aligned}$$

View-1: continuous evaluation.

- Let $\mathcal{H} \subset \mathbb{R}^{\mathcal{X}}$ be a Hilbert space.
- Consider for fixed $x \in \mathcal{X}$ the $\delta_x : f \in \mathcal{H} \mapsto f(x) \in \mathbb{R}$ map.
- The (Dirac) evaluation functional is linear:

$$\begin{aligned}\delta_x(\alpha f + \beta g) &= (\alpha f + \beta g)(x) = \alpha f(x) + \beta g(x) \\ &= \alpha \delta_x(f) + \beta \delta_x(g) \quad (\forall \alpha, \beta \in \mathbb{R}, f, g \in \mathcal{H}).\end{aligned}$$

- Def.: $\mathcal{H}$ is called RKHS if δ_x is continuous $\forall x \in \mathcal{X}$.

Example for non-continuous δ_x

$\mathcal{H} = L^2[0,1] \ni f_n(x) = x^n$:

1. $f_n \to 0 \in \mathcal{H}$ since

$$\lim_{n\to\infty} \|f_n - 0\|_2 = \lim_{n\to\infty} \left(\int_0^1 x^{2n} \mathrm{d}x\right)^{1/2} = \lim_{n\to\infty} \frac{1}{\sqrt{2n+1}} = 0,$$

Example for non-continuous δ_x

$\mathcal{H} = L^2[0,1] \ni f_n(x) = x^n$:

1. $f_n \to 0 \in \mathcal{H}$ since

$$\lim_{n\to\infty} \|f_n - 0\|_2 = \lim_{n\to\infty} \left(\int_0^1 x^{2n} \mathrm{d}x \right)^{1/2} = \lim_{n\to\infty} \frac{1}{\sqrt{2n+1}} = 0,$$

2. but $\delta_1(f_n) = 1 \nrightarrow \delta_1(0) = 0$.

In L^2: norm convergence $\not\Rightarrow$ pointwise convergence.

View-1: convergence

In RKHS: convergence in norm $\Rightarrow$ pointwise convergence!

- Result: $f_n \xrightarrow{\mathcal{H}} f \Rightarrow f_n \xrightarrow{\forall x} f$.

View-1: convergence

In RKHS: convergence in norm $\Rightarrow$ pointwise convergence!

- Result: $f_n \xrightarrow{\mathcal{H}} f \Rightarrow f_n \xrightarrow{\forall x} f$.
- Proof: For any $x \in \mathcal{X}$,

$$|f_n(x) - f(x)| \overset{\delta_x \text{ def}}{=} |\delta_x(f_n) - \delta_x(f)| \overset{\delta_x \text{ lin}}{=} |\delta_x(f_n - f)|$$

$$\overset{\delta_x\text{: bounded}}{\leq} \underbrace{\|\delta_x\|}_{<\infty} \underbrace{\|f_n - f\|_{\mathcal{H}}}_{\to 0}.$$

View-2: reproducing $\Rightarrow$ elements, kernel trick.

- Let $\mathcal{H}$ be a Hilbert space of $\mathcal{X} \to \mathbb{R}$ functions.
- $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is called a reproducing kernel of $\mathcal{H}$ if for $\forall x \in \mathcal{X}$
  1. $k(\cdot, x) \in \mathcal{H}$ ('generators'),

View-2: reproducing $\Rightarrow$ elements, kernel trick.

- Let $\mathcal{H}$ be a Hilbert space of $\mathcal{X} \to \mathbb{R}$ functions.
- $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is called a reproducing kernel of $\mathcal{H}$ if for $\forall x \in \mathcal{X}, f \in \mathcal{H}$
  1. $k(\cdot, x) \in \mathcal{H}$ ('generators'),
  2. $\langle f, k(\cdot, x) \rangle_{\mathcal{H}} = f(x)$ (reproducing property).

View-2: reproducing ⇒ elements, kernel trick.

- Let $\mathcal{H}$ be a Hilbert space of $\mathcal{X} \to \mathbb{R}$ functions.
- $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is called a reproducing kernel of $\mathcal{H}$ if for $\forall x \in \mathcal{X}, f \in \mathcal{H}$
  1. $k(\cdot, x) \in \mathcal{H}$ ('generators'),
  2. $\langle f, k(\cdot, x)\rangle_{\mathcal{H}} = f(x)$ (reproducing property).

  Specifically: $\forall x, y \in \mathcal{X}$,

$$k(x, y) = \langle k(\cdot, x), k(\cdot, y)\rangle_{\mathcal{H}}.$$

View-2: reproducing ⇒ elements, kernel trick.

- Let $\mathcal{H}$ be a Hilbert space of $\mathcal{X} \to \mathbb{R}$ functions.
- $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is called a reproducing kernel of $\mathcal{H}$ if for $\forall x \in \mathcal{X}, f \in \mathcal{H}$
  1. $k(\cdot, x) \in \mathcal{H}$ ('generators'),
  2. $\langle f, k(\cdot, x)\rangle_{\mathcal{H}} = f(x)$ (reproducing property).

  Specifically: $\forall x, y \in \mathcal{X}$,

$$k(x, y) = \langle k(\cdot, x), k(\cdot, y)\rangle_{\mathcal{H}}.$$

Questions

Uniqueness, existence?

Reproducing kernel: uniqueness

Reproducibility & norm definition ⇒ uniqueness.

- Let k_1, k_2 be r.k.-s of $\mathcal{H}$. Then for $\forall f \in \mathcal{H}, \forall x \in \mathcal{X}$

$$\langle f, k_1(\cdot,x) - k_2(\cdot,x)\rangle_{\mathcal{H}} \overset{\langle\cdot,\cdot\rangle_{\mathcal{H}} \text{ lin, } k_i \text{ r.k.}}{=} f(x) - f(x) = 0.$$

Reproducing kernel: uniqueness

Reproducibility & norm definition $\Rightarrow$ uniqueness.

- Let k_1, k_2 be r.k.-s of $\mathcal{H}$. Then for $\forall f \in \mathcal{H}, \forall x \in \mathcal{X}$

$$\langle f, k_1(\cdot,x) - k_2(\cdot,x)\rangle_{\mathcal{H}} \overset{\langle\cdot,\cdot\rangle_{\mathcal{H}} \text{ lin, } k_i \text{ r.k.}}{=} f(x) - f(x) = 0.$$

- Choosing $f = k_1(\cdot,x) - k_2(\cdot,x)$, we get

$$\|k_1(\cdot,x) - k_2(\cdot,x)\|_{\mathcal{H}}^2 = 0, \quad (\forall x \in \mathcal{X})$$

i.e., $k_1 = k_2$.

View-2 (r.k.) $\Leftrightarrow$ view-1 (RKHS)

- Result: $\mathcal{H}$ has a r.k. (k) $\Leftrightarrow$ $\mathcal{H}$ is a RKHS.

View-2 (r.k.) $\Leftrightarrow$ view-1 (RKHS)

- Result: $\mathcal{H}$ has a r.k. (k) $\Leftrightarrow$ $\mathcal{H}$ is a RKHS.
- Proof ($\Rightarrow$):

$$|\delta_x(f)| \leq \sqrt{k(x,x)}\,\|f\|_{\mathcal{H}},$$

i.e. $\delta_x : \mathcal{H} \to \mathbb{R}$ is bounded (hence continuous).

- Result: $\mathcal{H}$ has a r.k. (k) $\Leftrightarrow$ $\mathcal{H}$ is a RKHS.
- Proof ($\Rightarrow$):

$$|\delta_x(f)| \overset{\delta_x \text{ def}}{=} |f(x)| \overset{k\text{: r.k.}}{=} |\langle f, k(\cdot, x)\rangle_{\mathcal{H}}| \leq \sqrt{k(x,x)}\, \|f\|_{\mathcal{H}},$$

i.e. $\delta_x : \mathcal{H} \to \mathbb{R}$ is bounded (hence continuous).

View-2 (r.k.) $\Leftrightarrow$ view-1 (RKHS)

- Result: $\mathcal{H}$ has a r.k. (k) $\Leftrightarrow$ $\mathcal{H}$ is a RKHS.
- Proof ($\Rightarrow$):

$$|\delta_x(f)| \overset{\delta_x \text{ def}}{=} |f(x)| \overset{k\text{: r.k.}}{=} |\langle f, k(\cdot,x)\rangle_{\mathcal{H}}| \overset{CBS}{\leq} \|k(\cdot,x)\|_{\mathcal{H}} \|f\|_{\mathcal{H}} \overset{k\text{: r.k.}}{=} \sqrt{k(x,x)}\, \|f\|_{\mathcal{H}},$$

i.e. $\delta_x : \mathcal{H} \to \mathbb{R}$ is bounded (hence continuous).

View-2 (r.k.) $\Leftrightarrow$ view-1 (RKHS)

- Result: $\mathcal{H}$ has a r.k. (k) $\Leftrightarrow$ $\mathcal{H}$ is a RKHS.
- Proof ($\Rightarrow$):

$$|\delta_x(f)| \overset{\delta_x \text{ def}}{=} |f(x)| \overset{k\text{: r.k.}}{=} |\langle f, k(\cdot, x)\rangle_{\mathcal{H}}| \overset{CBS}{\leq} \|k(\cdot, x)\|_{\mathcal{H}} \|f\|_{\mathcal{H}} \overset{k\text{: r.k.}}{=} \sqrt{k(x, x)} \|f\|_{\mathcal{H}},$$

i.e. $\delta_x : \mathcal{H} \to \mathbb{R}$ is bounded (hence continuous).

Convergence in RKHS $\Rightarrow$ uniform convergence! (k: bounded).

View-2 (r.k.) $\Leftrightarrow$ view-1 (RKHS): $\Leftarrow$, existence of r.k.

Proof ($\Leftarrow$): Let δ_x be continuous for all $x \in \mathcal{X}$.

1. By the Riesz repr. theorem $\exists f_{\delta_x} \in \mathcal{H}$

$$\delta_x(f) = \langle f, \underbrace{f_{\delta_x}}_{=k(\cdot,x)?} \rangle_{\mathcal{H}}, \ \forall f \in \mathcal{H}.$$

Proof ($\Leftarrow$): Let δ_x be continuous for all $x \in \mathcal{X}$.

1. By the Riesz repr. theorem $\exists f_{\delta_x} \in \mathcal{H}$

$$\delta_x(f) = \langle f, \underbrace{f_{\delta_x}}_{=k(\cdot,x)?} \rangle_{\mathcal{H}}, \ \forall f \in \mathcal{H}.$$

2. Let $k(x', x) = f_{\delta_x}(x')$, $\forall x, x' \in \mathcal{X}$, then

$$k(\cdot, x) = f_{\delta_x} \in \mathcal{H},$$
$$\langle f, k(\cdot, x)\rangle_{\mathcal{H}} = \delta_x(f) = f(x).$$

Thus, k is the reproducing kernel.

View-3: positive definiteness.

- Let $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a symmetric function.

View-3: positive definiteness.

- Let $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a symmetric function.
- $\mathbf{G} := [k(x_i, x_j)]_{i,j=1}^n$: Gram matrix.

View-3: positive definiteness.

- Let $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a symmetric function.
- $\mathbf{G} := [k(x_i, x_j)]_{i,j=1}^n$: Gram matrix.
- k is called positive definite, if

$$\mathbf{a}^T \mathbf{G} \mathbf{a} \geq 0$$

for $\forall n \geq 1$, $\forall \mathbf{a} \in \mathbb{R}^n$, $\forall (x_1, \ldots, x_n) \in \mathcal{X}^n$.

View-4: 'kernel as inner product' view.

- Def.: A $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ function is called kernel, if
  1. $\exists \phi : \mathcal{X} \to \mathcal{F}$, where $\mathcal{F}$ is a Hilbert space s.t.
  2. $k(x, y) = \langle \phi(x), \phi(y) \rangle_{\mathcal{F}}$.
- Intuition: k is inner product in $\mathcal{F}$.

Reproducing kernel ⇒ kernel ⇒ positive definiteness

- Every r.k. is a kernel: $\phi(x) := k(\cdot, x)$, $k(x, y) = \langle k(\cdot, x), k(\cdot, y)\rangle_{\mathcal{H}}$.
- Every kernel is positive definite:

$$\mathbf{a}^T\mathbf{G}\mathbf{a} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j k(x_i, x_j)$$

$$\overset{k \text{ def }, \langle\cdot,\cdot\rangle_{\mathcal{F}} \text{ lin}}{=} \left\langle \sum_{i=1}^{n} a_i \phi(x_i), \sum_{j=1}^{n} a_j \phi(x_j) \right\rangle_{\mathcal{F}}$$

$$\overset{\|\cdot\|_{\mathcal{F}} = \sqrt{\langle\cdot,\cdot\rangle_{\mathcal{F}}}}{=} \left\| \sum_{i=1}^{n} a_i \phi(x_i) \right\|_{\mathcal{F}}^2 \geq 0.$$

Until now

- Result-1 (proved):

  RKHS (δ_x continuous) $\Leftrightarrow$ reproducing kernel.
- Result-2 (proved):

  reproducing kernel $\Rightarrow$ kernel $\Rightarrow$ positive definite.

Until now

- Result-1 (proved):

  RKHS (δ_x continuous) $\Leftrightarrow$ reproducing kernel.

- Result-2 (proved):

  reproducing kernel $\Rightarrow$ kernel $\Rightarrow$ positive definite.

Moore-Aronszajn theorem (follows)

positive definite $\Rightarrow$ reproducing kernel.

Until now

- Result-1 (proved):

  RKHS (δ_x continuous) $\Leftrightarrow$ reproducing kernel.

- Result-2 (proved):

  reproducing kernel $\Rightarrow$ kernel $\Rightarrow$ positive definite.

Moore-Aronszajn theorem (follows)

positive definite $\Rightarrow$ reproducing kernel.

$\Rightarrow$ the 4 notions are *exactly* the same!

Moore-Aronszajn construction: high-level view

- Given: a $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ positive definite function.
- We construct a pre-RKHS $\mathcal{H}_0$:

$$\mathcal{H}_0 = \left\{ f = \sum_{i=1}^{n} \alpha_i k(\cdot, x_i) : \alpha_i \in \mathbb{R}, x_i \in \mathcal{X} \right\} \supseteq \{k(\cdot, x) : x \in \mathcal{X}\},$$

$$\langle f, g \rangle_{\mathcal{H}_0} = k(x, y),$$

where $f = k(\cdot, x)$, $g = k(\cdot, y)$.

Moore-Aronszajn construction: high-level view

- Given: a $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ positive definite function.
- We construct a pre-RKHS $\mathcal{H}_0$:

$$\mathcal{H}_0 = \left\{ f = \sum_{i=1}^{n} \alpha_i k(\cdot, x_i) : \alpha_i \in \mathbb{R}, x_i \in \mathcal{X} \right\} \supseteq \{k(\cdot, x) : x \in \mathcal{X}\},$$

$$\langle f, g \rangle_{\mathcal{H}_0} = \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i \beta_j k(x_i, y_j),$$

where $f = \sum_{i=1}^{n} \alpha_i k(\cdot, x_i)$, $g = \sum_{j=1}^{m} \beta_j k(\cdot, y_j)$.

Moore-Aronszajn construction: high-level view

- $\mathcal{H}_0$ will satisfy:
  0. linear space (✓); $\langle f, g \rangle_{\mathcal{H}_0}$: well-defined & inner product.
  1. δ_x-s are continuous on $\mathcal{H}_0$ ($\forall x$).
  2. For any $\{f_n\} \subset \mathcal{H}_0$ Cauchy seq.:

$$f_n \xrightarrow{\forall x} 0 \quad \Rightarrow \quad f_n \xrightarrow{\mathcal{H}_0} 0.$$

- From $\mathcal{H}_0$ we construct $\mathcal{H}$ as:
  1. $\mathcal{H} \subset \mathbb{R}^{\mathcal{X}}$, for which
  2. $\exists \{f_n\}$ $\mathcal{H}_0$-Cauchy seq. such that $f_n \xrightarrow{\forall x} f$.

Moore-Aronszajn construction: high-level view

- Let

$$\langle f, g \rangle_{\mathcal{H}} := \lim_{n \to \infty} \langle f_n, g_n \rangle_{\mathcal{H}_0}, \tag{1}$$

  where $f_n \xrightarrow{\forall x} f$, $g_n \xrightarrow{\forall x} g$ $\mathcal{H}_0$-Cauchy sequences.
- $\mathcal{H}$ will satisfy:
  - $\mathcal{H}_0 \subset \mathcal{H}$: ✓ $[f_n \equiv f \in \mathcal{H}_0]$.
  - $\mathcal{H}$ is a RKHS with r.k. k:
    - -1 $\mathcal{H}$: linear space (✓),
    - 0 $\langle f, g \rangle_{\mathcal{H}}$: well-defined & inner product.
    - 1 $\mathcal{H}$ is complete.
    - 2 δ_x-s are continuous on $\mathcal{H}$ ($\forall x$).
    - 3 $\mathcal{H}$ has r.k. k (used to define $\mathcal{H}_0$).

$\langle \cdot, \cdot \rangle_{\mathcal{H}_0}$: well-defined, k reproducing on $\mathcal{H}_0$

- Recall: if $f = \sum_{i=1}^{n} \alpha_i k(\cdot, x_i)$, $g = \sum_{j=1}^{m} \beta_j k(\cdot, y_j)$, then

$$\langle f, g \rangle_{\mathcal{H}_0} = \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i \beta_j k(x_i, y_j).$$

- $\langle \cdot, \cdot \rangle_{\mathcal{H}_0}$ is independent of the particular $\{\alpha_i\}$ and $\{\beta_j\}$:

$$\langle f, g \rangle_{\mathcal{H}_0} = \sum_{i=1}^{n} \alpha_i \sum_{j=1}^{m} \beta_j k(x_i, y_j) = \sum_{i=1}^{n} \alpha_i g(x_i) \left[= \sum_{j=1}^{m} \beta_j f(y_j) \right].$$

$\langle \cdot, \cdot \rangle_{\mathcal{H}_0}$: well-defined, k reproducing on $\mathcal{H}_0$

- Recall: if $f = \sum_{i=1}^{n} \alpha_i k(\cdot, x_i)$, $g = \sum_{j=1}^{m} \beta_j k(\cdot, y_j)$, then

$$\langle f, g \rangle_{\mathcal{H}_0} = \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i \beta_j k(x_i, y_j).$$

- $\langle \cdot, \cdot \rangle_{\mathcal{H}_0}$ is independent of the particular $\{\alpha_i\}$ and $\{\beta_j\}$:

$$\langle f, g \rangle_{\mathcal{H}_0} = \sum_{i=1}^{n} \alpha_i \sum_{j=1}^{m} \beta_j k(x_i, y_j) = \sum_{i=1}^{n} \alpha_i g(x_i) \left[= \sum_{j=1}^{m} \beta_j f(y_j) \right].$$

- $\Rightarrow$ reproducing property on $\mathcal{H}_0$:

$$\langle f, k(\cdot, x) \rangle_{\mathcal{H}_0} = \sum_{i=1}^{n} \alpha_i k(x_i, x) = f(x).$$

$\langle \cdot, \cdot \rangle_{\mathcal{H}_0}$: inner product

- The 'tricky' property to check:

$$\|f\|_{\mathcal{H}_0} := \langle f, f \rangle_{\mathcal{H}_0} = 0 \implies f = 0.$$

- This holds by CBS (for the semi-inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}_0}$): $\forall x$

$$|f(x)| \overset{k \text{ r.k. on } \mathcal{H}_0}{=} \left| \langle f, k(\cdot, x) \rangle_{\mathcal{H}_0} \right| \overset{\text{CBS}}{\leq} \underbrace{\|f\|_{\mathcal{H}_0}}_{=0} \sqrt{k(x, x)} = 0.$$

Pre-RKHS: main property-1

δ_x is continuous on $\mathcal{H}_0$ $(\forall x)$: Let $f, g \in \mathcal{H}_0$, then

$$|\delta_x(f) - \delta_x(g)| \overset{\delta_x \text{ def}, k \text{ r.k.}, \langle \cdot,\cdot \rangle_{\mathcal{H}_0} \text{lin}}{=} \left| \langle f - g, k(\cdot, x) \rangle_{\mathcal{H}_0} \right| \overset{\text{CBS}, k \text{ r.k.}}{\leq} \sqrt{k(x,x)} \, \|f - g\|_{\mathcal{H}_0}.$$

$f_n : \mathcal{H}_0$-Cauchy $\xrightarrow{(\forall x)} 0 \Rightarrow f_n \xrightarrow{\mathcal{H}_0} 0$:

- f_n: Cauchy $\Rightarrow$ bounded, i.e. $\|f_n\|_{\mathcal{H}_0} < A$.
- f_n: Cauchy $\Rightarrow n, m \geq \exists N_1$: $\|f_n - f_m\|_{\mathcal{H}_0} < \epsilon/(2A)$.
- Let $f_{N_1} = \sum_{i=1}^{r} \alpha_i k(\cdot, x_i)$. $n \geq \exists N_2$: $|f_n(x_i)| < \frac{\epsilon}{2r|\alpha_i|}$ $(i = 1, \ldots, r)$.

For $n \geq \max(N_1, N_2)$:

$$\|f_n\|_{\mathcal{H}_0}^2 < \epsilon.$$

Pre-RKHS: main property-2

$f_n : \mathcal{H}_0$-Cauchy $\xrightarrow{(\forall x)} 0 \Rightarrow f_n \xrightarrow{\mathcal{H}_0} 0$:

- f_n: Cauchy $\Rightarrow$ bounded, i.e. $\|f_n\|_{\mathcal{H}_0} < A$.
- f_n: Cauchy $\Rightarrow n, m \geq \exists N_1$: $\|f_n - f_m\|_{\mathcal{H}_0} < \epsilon/(2A)$.
- Let $f_{N_1} = \sum_{i=1}^r \alpha_i k(\cdot, x_i)$. $n \geq \exists N_2$: $|f_n(x_i)| < \frac{\epsilon}{2r|\alpha_i|}$ $(i = 1, \ldots, r)$.

For $n \geq \max(N_1, N_2)$:

$$\|f_n\|_{\mathcal{H}_0}^2 = \langle f_n, f_n\rangle_{\mathcal{H}_0} \leq \left|\left\langle f_n - f_{N_1}, f_n\right\rangle_{\mathcal{H}_0}\right| + \left|\left\langle f_{N_1}, f_n\right\rangle_{\mathcal{H}_0}\right|$$

$$\leq \underbrace{\left\|f_n - f_{N_1}\right\|_{\mathcal{H}_0} \|f_n\|_{\mathcal{H}_0}}_{<[\epsilon/(2A)]A = \frac{\epsilon}{2}} + \sum_{i=1}^r \underbrace{|\alpha_i f_n(x_i)|}_{<|\alpha_i| \frac{\epsilon}{2r|\alpha_i|}} < \epsilon.$$

$\langle \cdot, \cdot \rangle_{\mathcal{H}}$: well-defined

$\alpha_n = \langle f_n, g_n \rangle_{\mathcal{H}_0}$ is convergent by Cauchyness in $\mathbb{R}$:

$$|\alpha_n - \alpha_m| < \epsilon$$

$\langle \cdot, \cdot \rangle_{\mathcal{H}}$: well-defined

$\alpha_n = \langle f_n, g_n \rangle_{\mathcal{H}_0}$ is convergent by Cauchyness in $\mathbb{R}$:

$$\begin{aligned}
|\alpha_n - \alpha_m| &= \left| \langle f_n, g_n \rangle_{\mathcal{H}_0} - \langle f_m, g_m \rangle_{\mathcal{H}_0} \right| \\
&= \left| \langle f_n, g_n \rangle_{\mathcal{H}_0} - \langle f_m, g_n \rangle_{\mathcal{H}_0} + \langle f_m, g_n \rangle_{\mathcal{H}_0} - \langle f_m, g_m \rangle_{\mathcal{H}_0} \right| \\
&= \left| \langle f_n - f_m, g_n \rangle_{\mathcal{H}_0} + \langle f_m, g_n - g_m \rangle_{\mathcal{H}_0} \right| \\
&\le \left| \langle f_n - f_m, g_n \rangle_{\mathcal{H}_0} \right| + \left| \langle f_m, g_n - g_m \rangle_{\mathcal{H}_0} \right| \\
&\le \underbrace{\|g_n\|_{\mathcal{H}_0}}_{<A} \underbrace{\|f_n - f_m\|_{\mathcal{H}_0}}_{<\frac{\epsilon}{2A}} + \underbrace{\|f_m\|_{H_0}}_{<B} \underbrace{\|g_n - g_m\|_{\mathcal{H}_0}}_{<\frac{\epsilon}{2B}} < \epsilon.
\end{aligned}$$

f_n, g_n: Cauchy $\Rightarrow$ bounded, i.e. $\|f_n\|_{\mathcal{H}_0} < A$, $\|g_n\|_{\mathcal{H}_0} < B$.

$\langle \cdot, \cdot \rangle_{\mathcal{H}}$: well-defined

The limit is independent of the Cauchy seq. chosen: let

- $f_n, f'_n \xrightarrow{\forall x} f$; $g_n, g'_n \xrightarrow{\forall x} g$: $\mathcal{H}_0$-Cauchy seq.-s,
- $\alpha_n = \langle f_n, g_n \rangle_{\mathcal{H}_0}$, $\alpha'_n = \langle f'_n, g'_n \rangle_{\mathcal{H}_0}$.

$\langle \cdot, \cdot \rangle_{\mathcal{H}}$: well-defined

The limit is independent of the Cauchy seq. chosen: let

- $f_n, f'_n \xrightarrow{\forall x} f$; $g_n, g'_n \xrightarrow{\forall x} g$: $\mathcal{H}_0$-Cauchy seq.-s,
- $\alpha_n = \langle f_n, g_n \rangle_{\mathcal{H}_0}$, $\alpha'_n = \langle f'_n, g'_n \rangle_{\mathcal{H}_0}$.
- 'Repeating' the previous argument:

$$|\alpha_n - \alpha'_n| \leq \underbrace{\|g_n\|_{\mathcal{H}_0}}_{\text{bounded}} \underbrace{\|f_n - f'_n\|_{\mathcal{H}_0}}_{\to 0} + \underbrace{\|f'_n\|_{\mathcal{H}_0}}_{\text{bounded}} \underbrace{\|g_n - g'_n\|_{\mathcal{H}_0}}_{\to 0}.$$

$\langle\cdot,\cdot\rangle_{\mathcal{H}}$: well-defined

The limit is independent of the Cauchy seq. chosen: let

- $f_n, f_n' \xrightarrow{\forall x} f$; $g_n, g_n' \xrightarrow{\forall x} g$: $\mathcal{H}_0$-Cauchy seq.-s,
- $\alpha_n = \langle f_n, g_n\rangle_{\mathcal{H}_0}$, $\alpha_n' = \langle f_n', g_n'\rangle_{\mathcal{H}_0}$.
- 'Repeating' the previous argument:

$$|\alpha_n - \alpha_n'| \leq \underbrace{\|g_n\|_{\mathcal{H}_0}}_{\text{bounded}} \underbrace{\|f_n - f_n'\|_{\mathcal{H}_0}}_{\to 0} + \underbrace{\|f_n'\|_{\mathcal{H}_0}}_{\text{bounded}} \underbrace{\|g_n - g_n'\|_{\mathcal{H}_0}}_{\to 0}.$$

- '$\to 0$': $f_n, f_n' \xrightarrow{\forall x} f \Rightarrow f_n - f_n' \xrightarrow{\forall x} 0 \Rightarrow f_n - f_n' \xrightarrow{\mathcal{H}_0} 0$ ($g_n - g_n'$ similarly).

The 'tricky' bit:

$$\langle f, f \rangle_{\mathcal{H}} = 0 \Rightarrow f = 0.$$

- Let $f_n \xrightarrow{\forall x} f$ $\mathcal{H}_0$-Cauchy, and $\langle f, f \rangle_{\mathcal{H}} = \lim_n \|f_n\|^2_{\mathcal{H}_0} = 0$. Then

$$|f(x)| = \lim_{n\to\infty} |f_n(x)| = \lim_{n\to\infty} |\delta_x(f_n)| \overset{(*)}{\leq} \lim_{n\to\infty} \underbrace{\|\delta_x\|}_{<\infty} \underbrace{\|f_n\|_{\mathcal{H}_0}}_{\to 0} = 0,$$

$(*)$: δ_x is continuous on $\mathcal{H}_0$.

Until now: $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ is well-defined & inner product.

Remains:

1. δ_x-s are continuous on $\mathcal{H}$ ($\forall x$).
2. $\mathcal{H}$ is complete.
3. The reproducing kernel on $\mathcal{H}$ is k.

δ_x-s are continuous on $\mathcal{H}$: lemma

$\mathcal{H}_0$ is dense in $\mathcal{H}$.

- Sufficient to show: $f_n \xrightarrow{\forall x} f$ $\mathcal{H}_0$-Cauchy $\Rightarrow f_n \xrightarrow{\mathcal{H}} f$.

δ_x-s are continuous on $\mathcal{H}$: lemma

$\mathcal{H}_0$ is dense in $\mathcal{H}$.

- Sufficient to show: $f_n \xrightarrow{\forall x} f$ $\mathcal{H}_0$-Cauchy $\Rightarrow f_n \xrightarrow{\mathcal{H}} f$.
- Proof: Fix $\epsilon > 0$,
  - f_n: $\mathcal{H}_0$-Cauchy $\Rightarrow \exists N \leq \forall m, n$: $\|f_m - f_n\|_{\mathcal{H}_0} < \epsilon$.
  - Fix $n^* \geq N$, then $f_m - f_{n*} \xrightarrow{\forall x} f - f_{n^*}$.
  - By the definition of $\|\cdot\|_{\mathcal{H}}$:

$$\|f - f_{n^*}\|_{\mathcal{H}}^2 = \lim_{m\to\infty} \|f_m - f_{n^*}\|_{\mathcal{H}_0}^2 \leq \epsilon^2,$$

  i.e., $f_n \xrightarrow{\mathcal{H}} f$.

δ_x-s are continuous on $\mathcal{H}$

Sufficient to show: δ_x linear is continuous at $f \equiv 0$. Fix $x \in \mathcal{X}$.

- We have seen: δ_x is continuous on $\mathcal{H}_0$, i.e. $\exists \eta$

$$\|g - 0\|_{\mathcal{H}_0} = \|g\|_{\mathcal{H}_0} < \eta \Rightarrow |\delta_x(g) - \delta_x(0)| = |\delta_x(g) - 0| = |g(x)| < \epsilon/2.$$

δ_x-s are continuous on $\mathcal{H}$

Sufficient to show: δ_x linear is continuous at $f \equiv 0$. Fix $x \in \mathcal{X}$.

- We have seen: δ_x is continuous on $\mathcal{H}_0$, i.e. $\exists \eta$

$$\|g - 0\|_{\mathcal{H}_0} = \|g\|_{\mathcal{H}_0} < \eta \Rightarrow |\delta_x(g) - \delta_x(0)| = |\delta_x(g) - 0| = |g(x)| < \epsilon/2.$$

- Take $f \in \mathcal{H}$: $\|f\|_{\mathcal{H}} < \eta/2$. Since $\mathcal{H}_0 \subset \mathcal{H}$ dense, $\exists f_n$ $\mathcal{H}_0$-Cauchy, $\exists N$

$$|f(x) - f_N(x)| < \epsilon/2 \quad [\Leftarrow f_n \xrightarrow{\forall x} f],$$

$$\|f - f_N\|_{\mathcal{H}} < \eta/2 \quad [\Leftarrow f_n \xrightarrow{\mathcal{H}} f] \Rightarrow$$

$$\|f_N\|_{\mathcal{H}_0} = \|f_N\|_{\mathcal{H}} \leq \underbrace{\|f\|_{\mathcal{H}}}_{<\frac{\eta}{2}} + \underbrace{\|f - f_N\|_{\mathcal{H}}}_{<\frac{\eta}{2}} < \eta.$$

δ_x-s are continuous on $\mathcal{H}$

Sufficient to show: δ_x linear is continuous at $f \equiv 0$. Fix $x \in \mathcal{X}$.

- We have seen: δ_x is continuous on $\mathcal{H}_0$, i.e. $\exists \eta$

$$\|g - 0\|_{\mathcal{H}_0} = \|g\|_{\mathcal{H}_0} < \eta \Rightarrow |\delta_x(g) - \delta_x(0)| = |\delta_x(g) - 0| = |g(x)| < \epsilon/2.$$

- Take $f \in \mathcal{H}$: $\|f\|_{\mathcal{H}} < \eta/2$. Since $\mathcal{H}_0 \subset \mathcal{H}$ dense, $\exists f_n$ $\mathcal{H}_0$-Cauchy, $\exists N$

$$|f(x) - f_N(x)| < \epsilon/2 \quad [\Leftarrow f_n \xrightarrow{\forall x} f],$$

$$\|f - f_N\|_{\mathcal{H}} < \eta/2 \quad [\Leftarrow f_n \xrightarrow{\mathcal{H}} f] \Rightarrow$$

$$\|f_N\|_{\mathcal{H}_0} = \|f_N\|_{\mathcal{H}} \leq \underbrace{\|f\|_{\mathcal{H}}}_{<\frac{\eta}{2}} + \underbrace{\|f - f_N\|_{\mathcal{H}}}_{<\frac{\eta}{2}} < \eta.$$

- With $g = f_N$ we get $|f_N(x)| < \frac{\epsilon}{2} \Rightarrow |f(x)| \leq \underbrace{|f(x) - f_N(x)|}_{<\frac{\epsilon}{2}} + \underbrace{|f_N(x)|}_{<\frac{\epsilon}{2}} < \epsilon.$

$\mathcal{H}$ is complete

High-level idea: let $\{f_n\} \subset \mathcal{H}$ be any Cauchy seq.,

- $\exists f(x) := \lim_n f_n(x)$ since
  - δ_x cont. on $\mathcal{H} \Rightarrow \{f_n(x)\} \subset \mathbb{R}$ Cauchy seq. $\Rightarrow$ convergent.

$\mathcal{H}$ is complete

High-level idea: let $\{f_n\} \subset \mathcal{H}$ be any Cauchy seq.,

- $\exists f(x) := \lim_n f_n(x)$ since
  - δ_x cont. on $\mathcal{H} \Rightarrow \{f_n(x)\} \subset \mathbb{R}$ Cauchy seq. $\Rightarrow$ convergent.
- Question: is the point-wise limit $f \in \mathcal{H}$?

$\mathcal{H}$ is complete

High-level idea: let $\{f_n\} \subset \mathcal{H}$ be any Cauchy seq.,

- $\exists f(x) := \lim_n f_n(x)$ since
  - δ_x cont. on $\mathcal{H} \Rightarrow \{f_n(x)\} \subset \mathbb{R}$ Cauchy seq. $\Rightarrow$ convergent.
- Question: is the point-wise limit $f \in \mathcal{H}$?
- Idea:
  1. $\mathcal{H}_0$ dense in $\mathcal{H} \Rightarrow \exists g_n \in \mathcal{H}_0$ s.t. $\|g_n - f_n\|_{\mathcal{H}} < \frac{1}{n}$.
  2. We show
     - $g_n \xrightarrow{\forall x} f$; $\{g_n\} \subset \mathcal{H}_0$: Cauchy seq.$\} \Rightarrow f \in \mathcal{H}$.
     - $f_n \xrightarrow{\mathcal{H}} f$.

- $g_n \xrightarrow{\forall x} f$:

$$|g_n(x) - f(x)| \le |g_n(x) - f_n(x)| + |f_n(x) - f(x)| = \underbrace{|\delta_x(g_n - f_n)|}_{\to 0;\ \delta_x \text{ cont. on } \mathcal{H}} + \underbrace{|f_n(x) - f(x)|}_{\to 0;\ f \text{ def.}}.$$

- $\{g_n\} \subset \mathcal{H}_0$ is Cauchy sequence:

$$\begin{aligned}\|g_m - g_n\|_{\mathcal{H}_0} &= \|g_m - g_n\|_{\mathcal{H}} \\ &\le \|g_m - f_m\|_{\mathcal{H}} + \|f_m - f_n\|_{\mathcal{H}} + \|f_n - g_n\|_{\mathcal{H}} \\ &\le \underbrace{\frac{1}{m} + \frac{1}{n}}_{g_m, g_n \text{ def.}} + \underbrace{\|f_m - f_n\|_{\mathcal{H}}}_{\to 0; f_n: \mathcal{H}\text{-Cauchy}}.\end{aligned}$$

- Finally, $f_n \xrightarrow{\mathcal{H}} f$:

$$\|f - f_n\|_{\mathcal{H}} \le \|f - g_n\|_{\mathcal{H}} + \|g_n - f_n\|_{\mathcal{H}} \le \overbrace{\|f - g_n\|_{\mathcal{H}}}^{\to 0 \text{: shown at '}\mathcal{H}_0\text{ dense in }\mathcal{H}\text{'}} + \overbrace{\frac{1}{n}}^{g_n \text{ def.}}.$$

Final property: the reproducing kernel on $\mathcal{H}$ is k

- Let $f \in \mathcal{H}$, and $f_n \xrightarrow{\forall x} f$ $\mathcal{H}_0$-Cauchy sequence.
- Then,

$$\langle f, k(\cdot, x)\rangle_{\mathcal{H}} \overset{(a)}{=} \lim_{n\to\infty} \langle f_n, k(\cdot, x)\rangle_{\mathcal{H}_0} \overset{(b)}{=} \lim_{n\to\infty} f_n(x) \overset{(c)}{=} f(x),$$

where

- (a): definition of $\langle \cdot, \cdot \rangle_{\mathcal{H}}$,
- (b): k reproducing kernel on $\mathcal{H}_0$,
- (c): $f_n \xrightarrow{\forall x} f$.

Summary

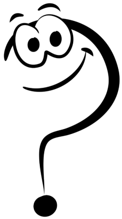
We have shown that

- RKHS (δ_x continuous) $\Leftrightarrow$ reproducing kernel $\Leftrightarrow$ kernel (feature view) $\Leftrightarrow$ positive definite.

$\backsim$

- *Moore-Aronszajn theorem*:
  - RKHS construction for a k pos. def. function.
  - Idea:
    1. pre-RKHS: $\mathcal{H}_0 = span\left[\{k(\cdot, x)\}_{x \in \mathcal{X}}\right]$,
    2. $\mathcal{H} :=$ pointwise limit of $\mathcal{H}_0$-Cauchy sequences.

Thank you for the attention!



Vector space axioms

$(V, +, \lambda\cdot)$ is vector space if $[\forall \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v} \in V$, $a, b \in \mathbb{R}]$:

$$\begin{aligned}
(\mathbf{v}_1 + \mathbf{v}_2) + \mathbf{v}_3 &= \mathbf{v}_1 + (\mathbf{v}_2 + \mathbf{v}_3), \text{ (associativity)}\\
\mathbf{v}_1 + \mathbf{v}_2 &= \mathbf{v}_2 + \mathbf{v}_1, \text{ (commutativity)}\\
\exists \mathbf{0} : \mathbf{v} + \mathbf{0} &= \mathbf{v},\\
\exists -\mathbf{v} : \mathbf{v} + (-\mathbf{v}) &= \mathbf{0},\\
a(b\mathbf{v}) &= (ab)\mathbf{v},\\
1\mathbf{v} &= \mathbf{v},\\
a(\mathbf{v}_1 + \mathbf{v}_2) &= a\mathbf{v}_1 + a\mathbf{v}_2,\\
(a+b)\mathbf{v} &= a\mathbf{v} + b\mathbf{v}.
\end{aligned}$$

$\mathcal{H}$ is a vector space

$\mathcal{H} \subset \mathbb{R}^{\mathcal{X}} \Rightarrow$ Needed:

1. $f \in \mathcal{H} \Rightarrow \lambda f \in \mathcal{H}$: $\exists \{f_n\} \subset \mathcal{H}_0$-Cauchy, $f_n \xrightarrow{\forall x} f$.

$$\{\lambda f_n\} \subset \mathcal{H}_0 \ (\Leftarrow \mathcal{H}_0\text{: vector space}), \ \text{Cauchy},$$

$$(\lambda f_n)(x) \xrightarrow{\forall x} (\lambda f)(x).$$

$\mathcal{H}$ is a vector space

$\mathcal{H} \subset \mathbb{R}^{\mathcal{X}} \Rightarrow$ Needed:

1. $f \in \mathcal{H} \Rightarrow \lambda f \in \mathcal{H}$: $\exists \{f_n\} \subset \mathcal{H}_0$-Cauchy, $f_n \xrightarrow{\forall x} f$.

$$\{\lambda f_n\} \subset \mathcal{H}_0 \ (\Leftarrow \mathcal{H}_0\text{: vector space}), \ \text{Cauchy},$$
$$(\lambda f_n)(x) \xrightarrow{\forall x} (\lambda f)(x).$$

2. $f, g \in \mathcal{H} \Rightarrow f + g \in \mathcal{H}$: $\exists \{f_n\}, \{g_n\} \subset \mathcal{H}_0$-Cauchy, $f_n \xrightarrow{\forall x} f$, $g_n \xrightarrow{\forall x} g$

$$\{f_n + g_n\} \subset \mathcal{H}_0 \ (\Leftarrow \mathcal{H}_0\text{: vector space}), \ \text{Cauchy},$$
$$(f_n + g_n)(x) \xrightarrow{\forall x} (f + g)(x).$$

$\langle\cdot,\cdot\rangle_{\mathcal{H}_0}$: inner product

Needed: for $\forall f = \sum_i \alpha_i k(\cdot, x_i)$, $g = \sum_j \beta_j k(\cdot, y_j) \in \mathcal{H}_0$

1. $\langle f, g\rangle_{\mathcal{H}_0} = \langle g, f\rangle_{\mathcal{H}_0}$:

$$\langle f, g\rangle_{\mathcal{H}_0} = \sum_i \sum_j \alpha_i \beta_j k(x_i, y_j) = \sum_j \sum_i \beta_j \alpha_i k(y_j, x_i) = \langle g, f\rangle_{\mathcal{H}_0}.$$

$\langle \cdot, \cdot \rangle_{\mathcal{H}_0}$: inner product

Needed: for $\forall \lambda \in \mathbb{R}$, $f = \sum_i \alpha_i k(\cdot, x_i)$, $g = \sum_j \beta_j k(\cdot, y_j) \in \mathcal{H}_0$

1. $\langle f, g \rangle_{\mathcal{H}_0} = \langle g, f \rangle_{\mathcal{H}_0}$:

$$\langle f, g \rangle_{\mathcal{H}_0} = \sum_i \sum_j \alpha_i \beta_j k(x_i, y_j) = \sum_j \sum_i \beta_j \alpha_i k(y_j, x_i) = \langle g, f \rangle_{\mathcal{H}_0}.$$

2. $\langle \lambda f, g \rangle_{\mathcal{H}_0} = \lambda \langle f, g \rangle_{\mathcal{H}_0}$:

$$\langle \lambda f, g \rangle_{\mathcal{H}_0} = \sum_{i,j} (\lambda \alpha_i) \beta_j k(x_i, y_j) = \lambda \sum_{i,j} \alpha_i \beta_j k(x_i, y_j) = \lambda \langle f, g \rangle_{\mathcal{H}_0}.$$

$\langle \cdot, \cdot \rangle_{\mathcal{H}_0}$: inner product

Needed: for $\forall \lambda \in \mathbb{R}$, $f = \sum_i \alpha_i k(\cdot, x_i)$, $g = \sum_j \beta_j k(\cdot, y_j) \in \mathcal{H}_0$

1. $\langle f, g \rangle_{\mathcal{H}_0} = \langle g, f \rangle_{\mathcal{H}_0}$:

$$\langle f, g \rangle_{\mathcal{H}_0} = \sum_i \sum_j \alpha_i \beta_j k(x_i, y_j) = \sum_j \sum_i \beta_j \alpha_i k(y_j, x_i) = \langle g, f \rangle_{\mathcal{H}_0}.$$

2. $\langle \lambda f, g \rangle_{\mathcal{H}_0} = \lambda \langle f, g \rangle_{\mathcal{H}_0}$:

$$\langle \lambda f, g \rangle_{\mathcal{H}_0} = \sum_{i,j} (\lambda \alpha_i) \beta_j k(x_i, y_j) = \lambda \sum_{i,j} \alpha_i \beta_j k(x_i, y_j) = \lambda \langle f, g \rangle_{\mathcal{H}_0}.$$

3. $\langle f_1 + f_2, g \rangle_{\mathcal{H}_0} = \langle f_1, g \rangle_{\mathcal{H}_0} + \langle f_2, g \rangle_{\mathcal{H}_0}$ $[f_1 \leftrightarrow \alpha'_i, x'_i,\ f_2 \leftrightarrow \alpha''_i, x''_i]$:

$$\text{l.h.s} = \sum_{i,j} \alpha_i \beta_j k(x_i, y_j) = \sum_{i,j} \alpha'_i \beta_j k(x'_i, y_j) + \sum_{i,j} \alpha''_i \beta_j k(x''_i, y_j) = \text{r.h.s.},$$

where $f_1 + f_2 \leftrightarrow \alpha'_i, \alpha''_i,\ x'_i, x''_i$.

$\langle \cdot, \cdot \rangle_{\mathcal{H}_0}$: inner product

Needed: for $\forall \lambda \in \mathbb{R}$, $f = \sum_i \alpha_i k(\cdot, x_i)$, $g = \sum_j \beta_j k(\cdot, y_j) \in \mathcal{H}_0$

1. $\langle f, g \rangle_{\mathcal{H}_0} = \langle g, f \rangle_{\mathcal{H}_0}$:

$$\langle f, g \rangle_{\mathcal{H}_0} = \sum_i \sum_j \alpha_i \beta_j k(x_i, y_j) = \sum_j \sum_i \beta_j \alpha_i k(y_j, x_i) = \langle g, f \rangle_{\mathcal{H}_0}.$$

2. $\langle \lambda f, g \rangle_{\mathcal{H}_0} = \lambda \langle f, g \rangle_{\mathcal{H}_0}$:

$$\langle \lambda f, g \rangle_{\mathcal{H}_0} = \sum_{i,j} (\lambda \alpha_i) \beta_j k(x_i, y_j) = \lambda \sum_{i,j} \alpha_i \beta_j k(x_i, y_j) = \lambda \langle f, g \rangle_{\mathcal{H}_0}.$$

3. $\langle f_1 + f_2, g \rangle_{\mathcal{H}_0} = \langle f_1, g \rangle_{\mathcal{H}_0} + \langle f_2, g \rangle_{\mathcal{H}_0}$ $[f_1 \leftrightarrow \alpha_i', x_i',\ f_2 \leftrightarrow \alpha_i'', x_i'']$:

$$\text{l.h.s} = \sum_{i,j} \alpha_i \beta_j k(x_i, y_j) = \sum_{i,j} \alpha_i' \beta_j k(x_i', y_j) + \sum_{i,j} \alpha_i'' \beta_j k(x_i'', y_j) = \text{r.h.s.},$$

where $f_1 + f_2 \leftrightarrow \alpha_i', \alpha_i'', x_i', x_i''$.

4. $f = 0 \Rightarrow \langle f, f \rangle_{\mathcal{H}_0} = 0$:

$$f = 0 \times k(\cdot, x) \Rightarrow \langle f, f \rangle_{\mathcal{H}_0} = 0 \times 0 \times k(x, x) = 0.$$

$\langle \cdot, \cdot \rangle_{\mathcal{H}}$: semi inner product

Needed: for $\forall f, f_1, f_2, g \in \mathcal{H}$

1. $\langle f, g \rangle_{\mathcal{H}} = \langle g, f \rangle_{\mathcal{H}}$:

$$\langle f, g \rangle_{\mathcal{H}} = \lim_n \langle f_n, g_n \rangle_{\mathcal{H}_0} \overset{\mathcal{H}_0:\checkmark}{=} \lim_n \langle g_n, f_n \rangle_{\mathcal{H}_0} = \langle g, f \rangle_{\mathcal{H}}.$$

$\langle \cdot, \cdot \rangle_{\mathcal{H}}$: semi inner product

Needed: for $\forall f, f_1, f_2, g \in \mathcal{H}, \lambda \in \mathbb{R}$

1. $\langle f, g \rangle_{\mathcal{H}} = \langle g, f \rangle_{\mathcal{H}}$:

$$\langle f, g \rangle_{\mathcal{H}} = \lim_n \langle f_n, g_n \rangle_{\mathcal{H}_0} \overset{\mathcal{H}_0:\checkmark}{=} \lim_n \langle g_n, f_n \rangle_{\mathcal{H}_0} = \langle g, f \rangle_{\mathcal{H}}.$$

2. $\langle \lambda f, g \rangle_{\mathcal{H}} = \lambda \langle f, g \rangle_{\mathcal{H}}$:

$$\langle \lambda f, g \rangle_{\mathcal{H}} = \lim_n \langle \lambda f_n, g_n \rangle_{\mathcal{H}_0} \overset{\mathcal{H}_0:\checkmark}{=} \lim_n \lambda \langle f_n, g_n \rangle_{\mathcal{H}_0} = \lambda \lim_n \langle f_n, g_n \rangle_{\mathcal{H}_0} = \lambda \langle f, g \rangle_{\mathcal{H}}.$$

Needed: for $\forall f, f_1, f_2, g \in \mathcal{H}, \lambda \in \mathbb{R}$

1. $\langle f, g\rangle_{\mathcal{H}} = \langle g, f\rangle_{\mathcal{H}}$:

$$\langle f, g\rangle_{\mathcal{H}} = \lim_n \langle f_n, g_n\rangle_{\mathcal{H}_0} \overset{\mathcal{H}_0:\checkmark}{=} \lim_n \langle g_n, f_n\rangle_{\mathcal{H}_0} = \langle g, f\rangle_{\mathcal{H}}.$$

2. $\langle \lambda f, g\rangle_{\mathcal{H}} = \lambda\langle f, g\rangle_{\mathcal{H}}$:

$$\langle \lambda f, g\rangle_{\mathcal{H}} = \lim_n \langle \lambda f_n, g_n\rangle_{\mathcal{H}_0} \overset{\mathcal{H}_0:\checkmark}{=} \lim_n \lambda\langle f_n, g_n\rangle_{\mathcal{H}_0} = \lambda \lim_n \langle f_n, g_n\rangle_{\mathcal{H}_0} = \lambda\langle f, g\rangle_{\mathcal{H}}.$$

3. $\langle f_1 + f_2, g\rangle_{\mathcal{H}} = \langle f_1, g\rangle_{\mathcal{H}} + \langle f_2, g\rangle_{\mathcal{H}}$:

$$\langle f_1 + f_2, g\rangle_{\mathcal{H}} = \lim_n \langle f_{1,n} + f_{2,n}, g\rangle_{\mathcal{H}_0} \overset{\mathcal{H}_0:\checkmark}{=} \lim_n [\langle f_{1,n}, g\rangle_{\mathcal{H}_0} + \langle f_{2,n}, g\rangle_{\mathcal{H}_0}]$$
$$= \lim_n \langle f_{1,n}, g\rangle_{\mathcal{H}_0} + \lim_n \langle f_{2,n}, g\rangle_{\mathcal{H}_0} = \langle f_1, g\rangle_{\mathcal{H}} + \langle f_2, g\rangle_{\mathcal{H}}.$$

$\langle \cdot, \cdot \rangle_{\mathcal{H}}$: semi inner product

Needed: for $\forall f, f_1, f_2, g \in \mathcal{H}, \lambda \in \mathbb{R}$

1. $\langle f, g\rangle_{\mathcal{H}} = \langle g, f\rangle_{\mathcal{H}}$:

$$\langle f, g\rangle_{\mathcal{H}} = \lim_n \langle f_n, g_n\rangle_{\mathcal{H}_0} \overset{\mathcal{H}_0:\checkmark}{=} \lim_n \langle g_n, f_n\rangle_{\mathcal{H}_0} = \langle g, f\rangle_{\mathcal{H}}.$$

2. $\langle \lambda f, g\rangle_{\mathcal{H}} = \lambda\langle f, g\rangle_{\mathcal{H}}$:

$$\langle \lambda f, g\rangle_{\mathcal{H}} = \lim_n \langle \lambda f_n, g_n\rangle_{\mathcal{H}_0} \overset{\mathcal{H}_0:\checkmark}{=} \lim_n \lambda \langle f_n, g_n\rangle_{\mathcal{H}_0} = \lambda \lim_n \langle f_n, g_n\rangle_{\mathcal{H}_0} = \lambda\langle f, g\rangle_{\mathcal{H}}.$$

3. $\langle f_1 + f_2, g\rangle_{\mathcal{H}} = \langle f_1, g\rangle_{\mathcal{H}} + \langle f_2, g\rangle_{\mathcal{H}}$:

$$\begin{aligned}\langle f_1 + f_2, g\rangle_{\mathcal{H}} &= \lim_n \langle f_{1,n} + f_{2,n}, g\rangle_{\mathcal{H}_0} \overset{\mathcal{H}_0:\checkmark}{=} \lim_n [\langle f_{1,n}, g\rangle_{\mathcal{H}_0} + \langle f_{2,n}, g\rangle_{\mathcal{H}_0}] \\ &= \lim_n \langle f_{1,n}, g\rangle_{\mathcal{H}_0} + \lim_n \langle f_{2,n}, g\rangle_{\mathcal{H}_0} = \langle f_1, g\rangle_{\mathcal{H}} + \langle f_2, g\rangle_{\mathcal{H}}.\end{aligned}$$

4. $f = 0 \Rightarrow \langle f, f\rangle_{\mathcal{H}} = 0$: Let $f_n \equiv 0$

$$\langle f, f\rangle_{\mathcal{H}} = \lim_n \langle 0, 0\rangle_{\mathcal{H}_0} \overset{\mathcal{H}_0:\checkmark}{=} \lim_n 0 = 0.$$