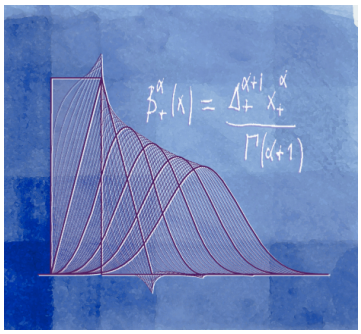
ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Sparse stochastic processes

RKHS, splines, and Gaussian processes

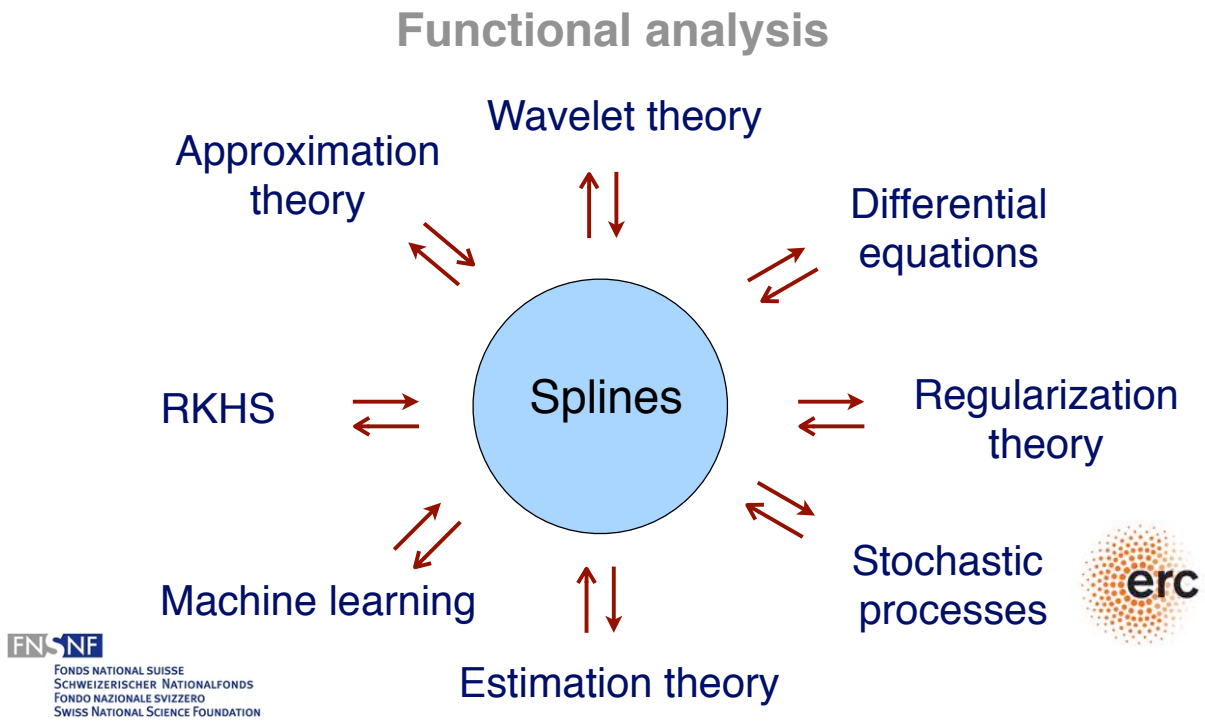
Prof. Michael Unser, LIB



March 2017

EDEE Course

Splines as a unifying mathematical concept



FNSNF
Fonds national suisse
Schweizerischer Nationalfonds
Fondo nazionale svizzero
Swiss National Science Foundation

erc

CONTENT

- **1. Preliminaries**
  - Brief overview of continuous-domain operators ✔
  - Inner products vs. duality products ✔
  - Self-adjoint operators and positive definite kernels
- **2. Reproducing kernel Hilbert spaces**
- **3. Variational splines and representer theorems**
- **4. Gaussian processes**

Linear operators, adjoints and kernels

Pair of vector spaces $\mathcal{X}, \mathcal{Y}$ with topological duals $\mathcal{X}', \mathcal{Y}'$

$$\mathcal{S}(\mathbb{R}^d) \subseteq \mathcal{X}, \mathcal{Y}, \mathcal{X}', \mathcal{Y}' \subseteq \mathcal{S}'(\mathbb{R}^d)$$

- Linear operator $\mathrm{G} : \mathcal{X} \to \mathcal{Y}$ and its adjoint $\mathrm{G}^* : \mathcal{Y}' \to \mathcal{X}'$

$$\boxed{\langle f, \mathrm{G}\varphi \rangle = \langle \mathrm{G}^* f, \varphi \rangle}$$

- "Integral" representation: $g \in \mathcal{S}'(\mathbb{R}^d \times \mathbb{R}^d)$ = Schwartz kernel of G

$$\mathrm{G} : \varphi \mapsto \int_{\mathbb{R}^d} g(\cdot, \boldsymbol{y}) \varphi(\boldsymbol{y}) \mathrm{d}\boldsymbol{y} \qquad \mathrm{G}^* : \varphi \mapsto \int_{\mathbb{R}^d} g(\boldsymbol{y}, \cdot) \varphi(\boldsymbol{y}) \mathrm{d}\boldsymbol{y}$$

- Self-adjoint operator

The operator H is said to be **self-adjoint** if $\mathrm{H}\{\varphi\} = \mathrm{H}^*\{\varphi\}$ for all $\varphi \in \mathcal{S}(\mathbb{R}^d)$

$\Leftrightarrow$ the kernel of the operator is symmetric

Positive-definite kernel

Definition

A symmetric kernel function $h : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is said to be positive semi-definite (or **positive definite**, for short) if

$$\sum_{m=1}^{N} \sum_{n=1}^{N} z_m h(\boldsymbol{x}_m, \boldsymbol{x}_n) z_n \geq 0$$

for any $N \in \mathbb{N}$, $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N \in \mathbb{R}^d$, and $z_1, \ldots, z_N \in \mathbb{R}$.

- Equivalent conditions
  - The square matrix $\mathbf{H} \in \mathbb{R}^{N \times N}$ with entries $[\mathbf{H}]_{m,n} = h(\boldsymbol{x}_m, \boldsymbol{x}_n)$ is positive definite for any choice of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N \in \mathbb{R}^d$; i.e.,

    $\mathbf{z}^T \mathbf{H} \mathbf{z} = \langle \mathbf{H}\mathbf{z}, \mathbf{z} \rangle \geq 0$ for all $\mathbf{z} \in \mathbb{R}^N$ $\quad \Leftrightarrow \quad$ The eigenvalues of $\mathbf{H}$ are non-negative

  - The operator $\mathrm{H} : \varphi \mapsto \int_{\mathbb{R}^d} h(\cdot, \boldsymbol{y}) \varphi(\boldsymbol{y}) \mathrm{d}\boldsymbol{y}$ is positive-definite; i.e.,

    $\langle \mathrm{H}\{\varphi\}, \varphi \rangle \geq 0$ for all $\varphi \in \mathcal{S}(\mathbb{R}^d)$

Aronszajn's theorem

Theorem (Moore-Aronszajn)

The kernel function $h : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is symmetric positive (semi-)definite if and only if there exists some Hilbert space $\mathcal{H}$ and a families of elements $\{f_{\boldsymbol{x}}\}_{\boldsymbol{x} \in \mathbb{R}^d}$ in $\mathcal{H}$ such that

$$h(\boldsymbol{x}, \boldsymbol{y}) = \langle f_{\boldsymbol{x}}, f_{\boldsymbol{y}} \rangle_{\mathcal{H}}. \tag{1}$$

In particular, there is a unique Hilbert space $\mathcal{H}_{\mathrm{rep}}$—the reproducing kernel Hilbert space of h—such that (1) holds with $f_{\boldsymbol{x}} = h(\cdot, \boldsymbol{x})$.

Sketch of proof (direct part):

$$\begin{aligned}
\sum_{m=1}^{N} \sum_{n=1}^{N} z_m h(\boldsymbol{x}_m, \boldsymbol{x}_n) z_n &= \sum_{m=1}^{N} \sum_{n=1}^{N} z_m \langle f_{\boldsymbol{x}_m}, f_{\boldsymbol{x}_n} \rangle_{\mathcal{H}} z_n \\
&= \langle \sum_{i=1}^{n} z_m f_{\boldsymbol{x}_m}, \sum_{j=1}^{n} z_n f_{\boldsymbol{x}_n} \rangle_{\mathcal{H}} \qquad \text{(bilinearity of } \langle \cdot, \cdot \rangle\text{)} \\
&= \| \sum_{i=1}^{n} z_m f_{\boldsymbol{x}_m} \|_{\mathcal{H}}^2 \geq 0
\end{aligned}$$

CONTENT

- **1. Preliminaries ✔**
- **2. Reproducing kernel Hilbert spaces**
  - Definition RKHS
  - Decay and continuity properties
  - RKHS associated with an invertible operator
  - Operators with non-trivial null spaces
- **3. Variational splines and representer theorems**
- **4. Gaussian processes**

2. RKHS

In essence, any Hilbert space $\mathcal{H} = \{f : \mathbb{R}^d \to \mathbb{R} : \|f\|_{\mathcal{H}} = \sqrt{\langle f, f \rangle_{\mathcal{H}}} < \infty\} \subseteq \mathcal{S}'(\mathbb{R}^d)$ whose members f are "ordinary"—but, not necessarily bounded—functions on $\mathbb{R}^d$ is a RKHS.

Definition
The Hilbert space $\mathcal{H} \subseteq \mathcal{S}'(\mathbb{R}^d)$ is said to be a ***reproducing kernel Hilbert space*** (RKHS) if the shifted Dirac impulse $\delta(\cdot - \boldsymbol{r}_0) \in \mathcal{H}'$ for any $\boldsymbol{r}_0 \in \mathbb{R}^d$.

2.1 Definition of reproducing kernel

$\delta_{\boldsymbol{x}_0} = \delta(\cdot - \boldsymbol{x}_0) \in \mathcal{H}'$ $\Rightarrow$ existence of $\delta^*_{\boldsymbol{x}_0} = \mathrm{R}\{\delta_{\boldsymbol{x}_0}\} \in \mathcal{H}$ such that

$$f(\boldsymbol{x}_0) = \delta_{\boldsymbol{x}_0}(f) = \langle \delta(\cdot - \boldsymbol{x}_0), f\rangle = \langle \delta^*_{\boldsymbol{x}_0}, f\rangle_{\mathcal{H}},$$

for all $f \in \mathcal{H}$ and any $\boldsymbol{x}_0 \in \mathbb{R}^d$

Definition

The **reproducing kernel** of a RKHS on $\mathbb{R}^d$ is the function $h : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ such that

(i) $h(\boldsymbol{x}_0, \cdot) \in \mathcal{H}$ for all $\boldsymbol{x}_0 \in \mathbb{R}^d$

(ii) $f(\boldsymbol{x}_0) = \langle h(\boldsymbol{x}_0, \cdot), f\rangle_{\mathcal{H}}$ for all $f \in \mathcal{H}$ and $\boldsymbol{x}_0 \in \mathbb{R}^d$.

Reproducing kernel: fundamental properties

Proposition: Let $\mathcal{H}$ be a RKHS on $\mathbb{R}^d$. Then, its reproducing kernel $h : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ has the following properties.

1. It is unique.
2. $h(\boldsymbol{x}, \boldsymbol{y}) = \langle h(\boldsymbol{x}, \cdot), h(\cdot, \boldsymbol{y})\rangle_{\mathcal{H}}$
3. Symmetry: $h(\boldsymbol{x}, \boldsymbol{y}) = h(\boldsymbol{y}, \boldsymbol{x})$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$
4. Positive definiteness.
5. The linear span of $\{h(\boldsymbol{x}, \cdot), \boldsymbol{x} \in \mathbb{R}^d\}$ is dense in $\mathcal{H}$.
6. Link with the Riesz map: The operator $\mathrm{R} : \varphi \mapsto \mathrm{R}\{\varphi\} = \int_{\mathbb{R}^d} h(\cdot, \boldsymbol{y})\varphi(\boldsymbol{y})\mathrm{d}\boldsymbol{y}$ is a unitary mapping $\mathcal{H}' \to \mathcal{H}$ with the property that $\langle u, \mathrm{R}u\rangle = \|u\|^2_{\mathcal{H}'}$ for all $u \in \mathcal{H}'$ (the dual space of $\mathcal{H}$).
7. Invertibility: The operator R admits a unique inverse $\mathrm{R}^{-1} : \mathcal{H} \to \mathcal{H}'$ with the property that $\langle \mathrm{R}^{-1} f, f\rangle = \|f\|^2_{\mathcal{H}}$ for all $f \in \mathcal{H}$.

Elements for the proof:

- Symmetry: $h(\boldsymbol{x},\boldsymbol{y}) = \langle h(\boldsymbol{x},\cdot), h(\boldsymbol{y},\cdot)\rangle_{\mathcal{H}} = \langle h(\boldsymbol{y},\cdot), h(\boldsymbol{x},\cdot)\rangle_{\mathcal{H}} = h(\boldsymbol{y},\boldsymbol{x})$

- Unicity: h is the kernel of the Riesz map $\mathcal{H}' \to \mathcal{H}$

$$f(\boldsymbol{x}_0) = \langle \delta(\cdot - \boldsymbol{x}_0), f\rangle = \langle \mathrm{R}\{\delta(\cdot - \boldsymbol{x}_0)\}, f\rangle_{\mathcal{H}} = \langle h(\cdot,\boldsymbol{x}_0), f\rangle_{\mathcal{H}}$$
$$\Leftrightarrow \quad h(\cdot,\boldsymbol{x}_0) = \mathrm{R}\{\delta(\cdot - \boldsymbol{x}_0)\}$$

- Positive definiteness: link with inner product

- Denseness in $\mathcal{H}$: Consider any $g \in \mathcal{H}$ that is orthogonal to the linear span of $\{h(\boldsymbol{x},\cdot)\}_{\boldsymbol{x}\in\mathbb{R}^d}$. Then, $\langle g, h(\boldsymbol{x},\cdot)\rangle = 0$ for every $\boldsymbol{x} \in \mathbb{R}^d$, which, by the reproducing kernel property, is equivalent to $g = 0$.

- Link with Riesz map and invertibility follow from Riesz' theorem

$$\langle v, u\rangle = \langle v^*, u\rangle_{\mathcal{H}} = \langle \mathrm{R}v, u\rangle_{\mathcal{H}} = \langle v, \mathrm{R}^{-1}u\rangle_{\mathcal{H}'} = \langle v, u^*\rangle_{\mathcal{H}}$$

for all $v \in \mathcal{H}'$ and $u \in \mathcal{H}$

2.2 Decay and continuity properties

- Control of the rate of algebraic decay/growth: $\alpha \in \mathbb{R}$

$$\|f\|_{\infty,\alpha} = \operatorname{ess\,sup}_{\boldsymbol{x}\in\mathbb{R}^d} \left(|f(\boldsymbol{x})|(1+\|\boldsymbol{x}\|)^{\alpha}\right)$$
$$= \sup_{\boldsymbol{x}\in\mathbb{R}^d} \left(|f(\boldsymbol{x})|(1+\|\boldsymbol{x}\|)^{\alpha}\right) \quad \text{(if } f \text{ is continuous)}$$

$$L_{\infty,\alpha}(\mathbb{R}^d) = \left\{f \in \mathcal{S}'(\mathbb{R}^d) : \|f\|_{\infty,\alpha} < +\infty\right\}$$

$$C_{\mathrm{b},\alpha}(\mathbb{R}^d) = \left\{f : \mathbb{R}^d \to \mathbb{R} \text{ continuous and s.t. } \|f\|_{\infty,\alpha} < +\infty\right\}.$$

Both are Banach spaces with

$$\mathcal{S}(\mathbb{R}^d) \subseteq \quad C_{\mathrm{b},\alpha+\beta}(\mathbb{R}^d) \subseteq C_{\mathrm{b},\alpha}(\mathbb{R}^d) \subseteq L_{\infty,\alpha}(\mathbb{R}^d) \quad \subseteq \mathcal{S}'(\mathbb{R}^d),$$

for any $\beta \geq 0$

Continuity and decay (Cont'd)

- Continuous functions of slow growth

  For any continuous function $f \in \mathcal{S}'(\mathbb{R}^d)$, there exists some (critical) exponent $\alpha_0 \in \mathbb{R}$ such that $f \in C_{\mathrm{b},\alpha}(\mathbb{R}^d)$ for all $\alpha \leq \alpha_0$

  $$\Leftrightarrow \quad |f(\boldsymbol{x})| \leq \frac{\|f\|_{\infty,\alpha_0}}{(1+\|\boldsymbol{x}\|)^{\alpha_0}}$$

- Kernel spaces

  $$C_{\mathrm{b},\alpha}(\mathbb{R}^d \times \mathbb{R}^d) \subseteq L_{\infty,\alpha}(\mathbb{R}^d \times \mathbb{R}^d) \subseteq \mathcal{S}'(\mathbb{R}^d \times \mathbb{R}^d)$$

  $$\|h(\cdot,\cdot)\|_{\infty,\alpha} = \sup_{\boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^d} |h(\boldsymbol{x},\boldsymbol{y})|(1+\|\boldsymbol{x}\|)^{\alpha}(1+\|\boldsymbol{y}\|)^{\alpha}$$

$$C_{\mathrm{b},\alpha}(\mathbb{R}^d \times \mathbb{R}^d) = \left\{ h : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R} \text{ (jointly) continuous and s.t. } \|h(\cdot,\cdot)\|_{\infty,\alpha} < +\infty \right\}$$

RKHS of continuous functions

$$\mathcal{H} \subseteq C_{\mathrm{b},\alpha}(\mathbb{R}^d) \quad \Rightarrow \quad \|f\|_{\infty,\alpha} \leq C\|f\|_{\mathcal{H}}, \text{ for all } f \in \mathcal{H} \quad \text{(continuous embedding)}$$

$$\Downarrow$$

$$|f(\boldsymbol{x}_0)| = |\langle f, \delta(\cdot - \boldsymbol{x}_0)\rangle| \leq (1+\|\boldsymbol{x}_0\|)^{-\alpha}\|f\|_{\infty,\alpha} \leq C_{\boldsymbol{x}_0}\|f\|_{\mathcal{H}}$$

$$\Downarrow$$

$$\delta(\cdot - \boldsymbol{x}_0) : f \mapsto f(\boldsymbol{x}_0) \qquad \delta(\cdot - \boldsymbol{x}_0) \in \mathcal{H}' \qquad \text{(RKHS property)}$$

Theorem 1: A bivariate function $h : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is the reproducing kernel of a RKHS $\mathcal{H} \subseteq C_{\mathrm{b},\alpha}(\mathbb{R}^d)$ with $\alpha \in \mathbb{R}$ **if and only if** it is positive-definite, separately continuous in each variable, and such that $h(\cdot,\cdot) \in L_{\infty,\alpha}(\mathbb{R}^d \times \mathbb{R}^d)$. In particular, this implies that

1. $h(\boldsymbol{x}_0,\cdot) \in C_{\mathrm{b},\alpha}(\mathbb{R}^d)$ for any $\boldsymbol{x}_0 \in \mathbb{R}^d$
2. $\|h(\cdot,\cdot)\|_{\infty,\alpha} = \sup_{\boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^d} |h(\boldsymbol{x},\boldsymbol{y})|\ (1+\|\boldsymbol{x}\|)^{\alpha}(1+\|\boldsymbol{y}\|)^{\alpha} < \infty$
3. $A_{\alpha,h} = \sup_{\boldsymbol{x} \in \mathbb{R}^d} h(\boldsymbol{x},\boldsymbol{x})\ (1+\|\boldsymbol{x}\|)^{2\alpha} < \infty$

with Conditions 2. and 3. being equivalent.

Equivalence of Conditions 2. and 3.

Key property: $h(\boldsymbol{x},\boldsymbol{y}) = \langle h(\boldsymbol{x},\cdot), h(\cdot,\boldsymbol{y})\rangle_{\mathcal{H}}$

PD-kernel equivalent of Cauchy-Schwarz Inequality: $|h(\boldsymbol{x},\boldsymbol{y})|^2 \leq h(\boldsymbol{x},\boldsymbol{x})h(\boldsymbol{y},\boldsymbol{y})$

$$\Downarrow$$

$$\begin{aligned}\|h(\cdot,\cdot)\|^2_{\infty,\alpha} &= \sup_{\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^d} |(1+\|\boldsymbol{x}\|)^{\alpha} h(\boldsymbol{x},\boldsymbol{y})(1+\|\boldsymbol{y}\|)^{\alpha}|^2 \\ &\leq \sup_{\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^d} (1+\|\boldsymbol{x}\|)^{2\alpha} h(\boldsymbol{x},\boldsymbol{x})(1+\|\boldsymbol{y}\|)^{2\alpha} h(\boldsymbol{y},\boldsymbol{y}) = A^2_{h,\alpha}\end{aligned}$$

Conversely,

$$(1+\|\boldsymbol{x}\|)^{\alpha} h(\boldsymbol{x},\boldsymbol{y})(1+\|\boldsymbol{y}\|)^{\alpha} \leq \|h(\cdot,\cdot)\|_{\infty,\alpha}$$

so that

$$A_{\alpha,h} = \sup_{\boldsymbol{x}\in\mathbb{R}^d} (1+\|\boldsymbol{x}\|)^{\alpha} h(\boldsymbol{x},\boldsymbol{x})(1+\|\boldsymbol{x}\|)^{\alpha} \leq \|h(\cdot,\cdot)\|_{\infty,\alpha},$$

from which we deduce that $A_{\alpha,h} = \|h(\cdot,\cdot)\|_{\infty,\alpha}$.

Simplified version of theorem: $h \in C_{\mathrm{b},\alpha}(\mathbb{R}^d \times \mathbb{R}^d) \quad \Rightarrow \quad \mathcal{H} \subseteq C_{\mathrm{b},\alpha}(\mathbb{R}^d)$

- Boundedness: $\mathcal{H} \subseteq L_{\infty,\alpha}(\mathbb{R}^d)$

$$\begin{aligned}|f(\boldsymbol{x})| &= |\langle h(\boldsymbol{x},\cdot), f\rangle_{\mathcal{H}}| \\ &\leq \|h(\boldsymbol{x},\cdot)\|_{\mathcal{H}} \|f\|_{\mathcal{H}} = \sqrt{h(\boldsymbol{x},\boldsymbol{x})}\, \|f\|_{\mathcal{H}} \qquad \text{(Cauchy-Schwarz)}\end{aligned}$$

$$\|f\|^2_{\infty,\alpha} = \sup_{\boldsymbol{x}\in\mathbb{R}^d} |f(\boldsymbol{x})|^2 (1+\|\boldsymbol{x}\|)^{2\alpha} \leq \sup_{\boldsymbol{x}\in\mathbb{R}^d} (1+\|\boldsymbol{x}\|)^{2\alpha} h(\boldsymbol{x},\boldsymbol{x})\, \|f\|^2_{\mathcal{H}} = A_{h,\alpha}\, \|f\|^2_{\mathcal{H}} < \infty$$

- Continuity: $\mathcal{H} \subseteq C_{\mathrm{b},\alpha}(\mathbb{R}^d) \subseteq L_{\infty,\alpha}(\mathbb{R}^d)$

$$\begin{aligned}|f(\boldsymbol{x}) - f(\boldsymbol{x}_0)|^2 &= |\langle h(\boldsymbol{x},\cdot) - h(\boldsymbol{x}_0,\cdot), f\rangle_{\mathcal{H}}|^2 \\ &\leq \|h(\boldsymbol{x},\cdot) - h(\boldsymbol{x}_0,\cdot)\|^2_{\mathcal{H}}\, \|f\|^2_{\mathcal{H}} \qquad \text{(Cauchy-Schwarz)} \\ &= \big(h(\boldsymbol{x},\boldsymbol{x}) + h(\boldsymbol{x}_0,\boldsymbol{x}_0) - 2h(\boldsymbol{x},\boldsymbol{x}_0)\big)\, \|f\|^2_{\mathcal{H}}.\end{aligned}$$

$$h \in C_{\mathrm{b},\alpha}(\mathbb{R}^d \times \mathbb{R}^d) \quad \Rightarrow \quad \lim_{\boldsymbol{x}\to\boldsymbol{x}_0} h(\boldsymbol{x},\boldsymbol{x}) + h(\boldsymbol{x}_0,\boldsymbol{x}_0) - 2h(\boldsymbol{x},\boldsymbol{x}_0) = 0$$

$$\Rightarrow \quad \lim_{\boldsymbol{x}\to\boldsymbol{x}_0} |f(\boldsymbol{x}) - f(\boldsymbol{x}_0)|^2 = 0$$

Turning $L_2(\mathbb{R}^d)$ into a RKHS

- $L_2(\mathbb{R}^d)$ is NOT a reproducing kernel Hilbert space ?

  Candidate reproducing kernel $\quad \delta(\cdot - \boldsymbol{x}_0) : \varphi \mapsto \langle \delta(\cdot - \boldsymbol{x}_0), \varphi \rangle = \varphi(\boldsymbol{x}_0)$

  - $\delta(\cdot - \boldsymbol{x}_0) \notin \big(L_2(\mathbb{R}^d)\big)' = L_2(\mathbb{R}^d)$
  - $f(\boldsymbol{x}_0)$ is not well defined pointwise for all $f \in L_2(\mathbb{R}^d)$

- Solution = Smoothing

  "Smoothing" operator $\mathrm{G} : L_2(\mathbb{R}^d) \to \mathcal{H} \subseteq C_{\mathrm{b},\alpha}(\mathbb{R}^d)$

Proposition
Let $g : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a kernel such that $g(\boldsymbol{x}, \cdot) \in L_2(\mathbb{R}^d)$ for any fixed $\boldsymbol{x} \in \mathbb{R}^d$. Then, the output of the linear operator $\mathrm{G} : w \mapsto f = \int_{\mathbb{R}^d} g(\cdot, \boldsymbol{y}) w(\boldsymbol{y}) \mathrm{d}\boldsymbol{y}$ is well defined pointwise for any $w \in L_2(\mathbb{R}^d)$. If, in addition, there is some $\alpha \in \mathbb{R}$ such that

$$\sup_{\boldsymbol{x} \in \mathbb{R}^d} (1 + \|\boldsymbol{x}\|)^{\alpha} \|g(\boldsymbol{x}, \cdot)\|_{L_2(\mathbb{R}^d)} < \infty,$$

then G is bounded from $L_2(\mathbb{R}^d) \to L_{\infty,\alpha}(\mathbb{R}^d)$.

2.4 RKHS associated with invertible operator

- Native space

$$\mathcal{H}_{\mathrm{L}} = \big\{ f : \mathbb{R}^d \to \mathbb{R} \text{ s.t. } \|f\|_{\mathcal{H}_{\mathrm{L}}} \triangleq \|\mathrm{L}f\|_{L_2(\mathbb{R}^d)} < \infty \big\} \subseteq L_2(\mathbb{R}^d)$$

Coercive regularisation operator: $\mathrm{L} : \mathcal{H}_{\mathrm{L}} \to L_2(\mathbb{R}^d)$

$$c\|f\|_{L_2(\mathbb{R}^d)} \leq \|\mathrm{L}f\|_{L_2(\mathbb{R}^d)} \quad = \|f\|_{\mathcal{H}_{\mathrm{L}}}$$

coercivity ↗ — ↖ continuity ⇔ specifies a valid norm

$\Leftrightarrow$ Existence of (stable) inverse operator: $\mathrm{G} = \mathrm{L}^{-1} : L_2(\mathbb{R}^d) \to \mathcal{H}_{\mathrm{L}}$

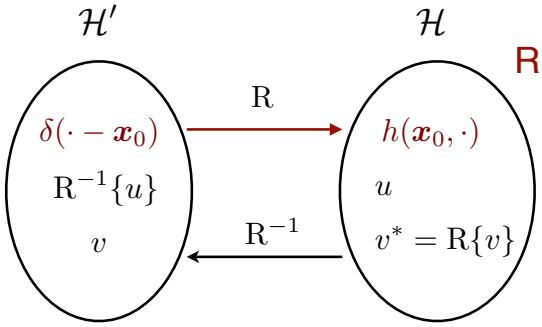
- Inner product: $\langle f, g \rangle_{\mathcal{H}_{\mathrm{L}}} = \langle \mathrm{L}f, \mathrm{L}g \rangle \quad = \langle \underbrace{(\mathrm{L}^*\mathrm{L})f}_{f^* \in \mathcal{H}'_{\mathrm{L}}}, g \rangle$

- Riesz maps

$$\mathrm{R}^{-1} = \mathrm{L}^*\mathrm{L} : \mathcal{H}_{\mathrm{L}} \to \mathcal{H}'_{\mathrm{L}}$$

$$\mathrm{R} = \mathrm{G}^*\mathrm{G} = (\mathrm{L}^*\mathrm{L})^{-1} : \mathcal{H}'_{\mathrm{L}} \to \mathcal{H}_{\mathrm{L}}$$

Reminder: Dual of a Hilbert space



Riesz isomorphism

$$\langle u, v^* \rangle_{\mathcal{H}} = \langle u, v \rangle = \langle \mathrm{R}^{-1} u, v \rangle_{\mathcal{H}'}$$

The dual of a Hilbert space $\mathcal{H}$ is a Hilbert space $\mathcal{H}'$ with $\mathcal{H}' = \mathrm{R}^{-1}(\mathcal{H})$.

Determination of reproducing kernel

$h(\boldsymbol{x}, \boldsymbol{y}) = \mathrm{R}\{\delta(\cdot - \boldsymbol{y})\}(\boldsymbol{x})$ with $\mathrm{R} = (\mathrm{L}^*\mathrm{L})^{-1} : \mathcal{H}'_{\mathrm{L}} \to \mathcal{H}_{\mathrm{L}}$

Hypothesis: $h(\boldsymbol{x}, \cdot) \in C_{\mathrm{b},0}(\mathbb{R}^d)$ (continuous and bounded) for any (fixed) $\boldsymbol{x} \in \mathbb{R}^d$

$\Rightarrow$ $\mathcal{H}_{\mathrm{L}} \subseteq C_{\mathrm{b}}(\mathbb{R}^d)$ is a RKHS (by Theorem 1)

- Explicit determination for LSI (=convolution) operator

  $\hat{L}(\boldsymbol{\omega}) = \mathcal{F}\{\mathrm{L}\delta\}(\boldsymbol{\omega})$: frequency response of operator

  $$h(\boldsymbol{x}, \boldsymbol{y}) = \rho_{\mathrm{L}^*\mathrm{L}}(\boldsymbol{x} - \boldsymbol{y}) \quad \text{where} \quad \rho_{\mathrm{L}^*\mathrm{L}}(\boldsymbol{x}) = \mathcal{F}^{-1}\left\{ \frac{1}{|\widehat{L}(\boldsymbol{\omega})|^2} \right\}(\boldsymbol{x})$$

Necessary (and sufficient) condition for coercivity: $c = \operatorname{ess\,inf}_{\boldsymbol{\omega} \in \mathbb{R}^d} |\hat{L}(\boldsymbol{\omega})| > 0$

$$\|\mathrm{L}f\|_{L_2}^2 = \int_{\mathbb{R}^d} \left|\hat{L}(\boldsymbol{\omega})\hat{f}(\boldsymbol{\omega})\right|^2 \frac{\mathrm{d}\boldsymbol{\omega}}{(2\pi)^d} \geq c^2 \int_{\mathbb{R}^d} \left|\hat{f}(\boldsymbol{\omega})\right|^2 \frac{\mathrm{d}\boldsymbol{\omega}}{(2\pi)^d} = c^2 \|f\|_{L_2}^2$$

2.5 Factorization of reproducing kernel

- Motivation: Extension for **non-coercive regularization** operator
  - Riesz map: $\mathrm{R} = \mathrm{G}\mathrm{G}^* : \mathcal{H}' \to \mathcal{H}$ $\quad\quad$ $\mathcal{H}' \to \mathcal{H}_0 \to \mathcal{H}$ with $\mathcal{H}_0 \subseteq L_2(\mathbb{R}^d)$
  - Inverse operator: $\mathrm{G} = \mathrm{L}^{-1}$ with Schwartz kernel $g(\boldsymbol{x}, \boldsymbol{y}) = \mathrm{G}\{\delta(\cdot - \boldsymbol{y})\}(\boldsymbol{x})$

Theorem 2

Let $\mathcal{H} \subseteq C_{\mathrm{b},\alpha}(\mathbb{R}^d)$ be a RKHS equipped with the **inner product** $\langle f, g \rangle_{\mathcal{H}} = \langle \mathrm{L}f, \mathrm{L}g \rangle$. Then, there exits a unique continuous operator $\mathrm{L}^{-1} : L_2(\mathbb{R}^d) \to L_{\infty,\alpha}(\mathbb{R}^d)$ and a Hilbert space $\mathcal{H}_0 \subseteq L_2(\mathbb{R}^d)$ such that the Riesz map from $\mathcal{H}' \to \mathcal{H}$ factors though $\mathcal{H}_0$ as $\mathrm{R} = \mathrm{L}^{-1}\mathrm{L}^{-1*} : \mathcal{H}' \to \mathcal{H}_0 \to \mathcal{H}$.

The Schwartz kernel of L^{-1} denoted by $g(\cdot, \cdot)$ satisfies the estimate

$$\sup_{\boldsymbol{x} \in \mathbb{R}^d} (1 + \|\boldsymbol{x}\|)^{\alpha} \|g(\boldsymbol{x}, \cdot)\|_{L_2(\mathbb{R}^d)} = \sqrt{A_{\alpha,h}} < \infty$$

and is linked to the reproducing kernel $h(\cdot, \cdot)$ of $\mathcal{H}$ by

$$h(\boldsymbol{x}, \boldsymbol{y}) = \int_{\mathbb{R}^d} g(\boldsymbol{x}, \boldsymbol{z}) g(\boldsymbol{y}, \boldsymbol{z}) \mathrm{d}\boldsymbol{z}$$

$$g(\boldsymbol{y}, \boldsymbol{x}) = \mathrm{L}\{h(\cdot, \boldsymbol{y})\}(\boldsymbol{x}).$$

Example 1: RKHS associated with an ortho-basis

- Orthonormal system: $\{\phi_n\}_{n \in \mathbb{N}}$ with $\phi_n \in C_{\mathrm{b}}(\mathbb{R}^d) \cap L_2(\mathbb{R}^d)$
- Inner product: $\langle f, g \rangle_{\mathcal{V}} = \sum_{n \in \mathbb{N}} \langle \phi_n, f \rangle \langle \phi_n, g \rangle$
- $\mathcal{V} = \overline{\mathrm{span}\{\phi_n\}}_{n \in \mathbb{Z}}$ equipped with $\langle \cdot, \cdot \rangle_{\mathcal{V}}$ is a Hilbert space

Simplified setting

$\mathcal{V} = \mathcal{V}'$ with $\mathcal{V} \subseteq L_2(\mathbb{R}^d)$

- Generic form of reproducing kernel

$$h_{\mathcal{V}}(\boldsymbol{x}, \boldsymbol{y}) = \sum_{n \in \mathbb{N}} \phi_n(\boldsymbol{x}) \phi_n(\boldsymbol{y})$$

- Orthogonal projector $\mathrm{Proj}_{\mathcal{V}} : L_2(\mathbb{R}^d) \to \mathcal{V}$

$$\mathrm{Proj}_{\mathcal{V}}\{f\}(\boldsymbol{x}) = \langle f, h_{\mathcal{V}}(\boldsymbol{x}, \cdot) \rangle = \sum_{n \in \mathbb{N}} \phi_n(\boldsymbol{x}) \langle f, \phi_n \rangle$$

- Factorization (trivial): $\mathrm{Proj}_{\mathcal{V}} \circ \mathrm{Proj}_{\mathcal{V}} = \mathrm{Proj}_{\mathcal{V}}$

$$h_{\mathcal{V}}(\boldsymbol{x}, \boldsymbol{y}) = \langle h_{\mathcal{V}}(\boldsymbol{x}, \cdot), h_{\mathcal{V}}(\cdot, \boldsymbol{y}) \rangle$$

Example 2: RKHS of bandlimitted functions

- Subspace of bandlimited functions in $L_2(\mathbb{R})$ with Nyquist frequency $\omega_{\max} = \pi$

$$\mathcal{H}_{\text{sinc}} = \{f \in \mathcal{S}'(\mathbb{R}) : (\text{sinc} * f) \in L_2(\mathbb{R})\}$$

- Regularization operator

$$\mathrm{L}_{\text{sinc}} : \varphi \mapsto \text{sinc} * \varphi$$

$$\text{sinc}(x) = \frac{\sin(\pi x)}{(\pi x)} = \mathcal{F}^{-1}\{\mathbb{1}_{[-\pi,\pi]}\}(x)$$

- Reproducing kernel

$$h_{\text{sinc}}(x, y) = \text{sinc}(x - y) = \sum_{k \in \mathbb{Z}} \text{sinc}(x - k)\text{sinc}(y - k)$$

- Reproduction formula = Shannon's Sampling theorem

$$\forall f \in \mathcal{H}_{\text{sinc}} : \quad f(x) = (\text{sinc} * f)(x) = \sum_{k \in \mathbb{Z}} \langle f, \text{sinc}(\cdot - k)\rangle \, \text{sinc}(x - k)$$
$$= \sum_{k \in \mathbb{Z}} f(k) \, \text{sinc}(x - k)$$

Example 3: Subspace of polynomials

- $\mathcal{N} = \text{span}\{p_1\} \subseteq C_{\text{b}}(\mathbb{R})$ with $p_1(x) = 1$
- Biorthogonal system: (δ, p_1) with $\langle \delta, p_1 \rangle = 1$
- Inner product: $\langle f, g \rangle_{\mathcal{N}} = f(0)g(0) = \langle \delta, f \rangle \langle \delta, g \rangle$
- $\mathcal{N} = \text{span}\{p_1\}$ equipped with $\langle f, g \rangle_{\mathcal{N}}$ is a RKHS
- Reproducing kernel

$$h_{\mathcal{N}}(x, y) = p_1(x)p_1(y) = 1$$

- Projection operator: $\text{Proj}_{\mathcal{N}}\{f\} = \langle \delta, f \rangle p_1$

Alternative solution(s)

(ϕ_1, p_1) with $\langle \phi_1, p_1 \rangle = 1$

$\langle f, g \rangle_{\mathcal{N}} = \langle \phi_1, f \rangle \langle \phi_1, g \rangle$

$\text{Proj}_{\mathcal{N}}\{f\} = \langle \phi_1, f \rangle p_1$

but, there is no factorization ... ?!?

2.6 RKHS associated with derivative operator

- Key characteristics $\mathrm{L} = \mathrm{D} = \dfrac{\mathrm{d}}{\mathrm{d}x}$
  - Green's function(s) ρ_{D} such that $\mathrm{D}\{\rho_{\mathrm{D}}\} = \delta$
    - Causal solution: $\mathbb{1}_+(x)$
    - Canonical solution: $\frac{1}{2}\mathrm{sign}(x) = \mathcal{F}\left\{\frac{1}{\mathrm{j}\omega}\right\}(x) = \mathbb{1}_+(x) - \frac{1}{2}$
  - Null space of dimension $N_0 = 1$:

    $\mathcal{N}_{\mathrm{D}} = \{q \in \mathcal{S}'(\mathbb{R}) : \mathrm{D}\{q\} = 0\} = \mathrm{span}\{p_1\}$ with $p_1(x) = 1$

- Related operators
  - Adjoint: $\mathrm{D}^* = -\mathrm{D}$

    $\langle \mathrm{D}f, \varphi\rangle = \underbrace{f(x)\varphi(x)|_{-\infty}^{+\infty}}_{=0} - \int_{\mathbb{R}} f(x)\varphi'(x)\mathrm{d}x = \langle f, -\mathrm{D}\varphi\rangle$
  - $\mathrm{D}^*\mathrm{D} = -\mathrm{D}^2$
  - Green's function of $\mathrm{D}^*\mathrm{D}$:

    $\rho_{\mathrm{D}^*\mathrm{D}}(x) = \mathcal{F}\left\{\dfrac{1}{\omega^2}\right\}(x) = -\frac{1}{2}|x| \in C_{\mathrm{b},-1}(\mathbb{R})$

Derivative operator (Cont'd)

- Native space

  $\mathcal{H}_{\mathrm{D}} = \{f : \mathbb{R} \to \mathbb{R} \text{ s.t. } \mathrm{D}f \in L_2(\mathbb{R})\} \quad \subseteq C_{\mathrm{b},-1}(\mathbb{R}^d)$

Proposition
$\mathcal{H}_{\mathrm{D}}$ equipped with the inner product $\langle f, g\rangle_{\mathcal{H}_{\mathrm{D}}} = \langle \mathrm{D}f, \mathrm{D}g\rangle + f(0)g(0)$ is a RKHS whose reproducing kernel is

$h_{\mathrm{D}}(x, y) = \frac{1}{2}\big(|x| + |y| - |x - y|\big) + 1.$

$h_{\mathrm{D}}(\cdot, y) \in C_{\mathrm{b},-1}(\mathbb{R})$ for any $y \in \mathbb{R}$ (first-order of growth)

Proof by verification

1) $\mathcal{H}_{\mathrm{D}}$ is a Hilbert space

- $\langle \mathrm{D}f, \mathrm{D}g \rangle$ is a semi-inner product
- $\langle \mathrm{D}f, \mathrm{D}f \rangle = 0 \quad \Leftrightarrow \quad f = b_1 p_1 \in \mathcal{N}_{\mathrm{D}}$ with $b_1 \in \mathbb{R}$
- $\|f\|^2_{\mathcal{H}_{\mathrm{D}}} = \langle \mathrm{D}f, \mathrm{D}f \rangle + |f(0)|^2 = 0 \quad \Leftrightarrow \quad f \in \mathcal{N}_{\mathrm{D}}$ and $f(0) = 0 \quad \Leftrightarrow \quad f = 0$ (unicity)

2) Check **reproducing kernel** property of $h_{\mathrm{D}}(x, y) = \frac{1}{2}\big(|x| + |y| - |x - y|\big) + 1$

$$\begin{aligned} \mathrm{D}\{h_{\mathrm{D}}(\cdot, y)\}(x) &= \tfrac{1}{2}\big(\mathrm{sign}(x) - \mathrm{sign}(x - y)\big) = \mathbb{1}_{[0,y)}(x) \in L_2(\mathbb{R}) \Rightarrow h_{\mathrm{D}}(\cdot, y) \in \mathcal{H}_{\mathrm{D}} \\ \mathrm{D}^*\mathrm{D}\{h_{\mathrm{D}}(\cdot, y)\}(x) &= -\mathrm{D}\{\tfrac{1}{2}\big(\mathrm{sign}(\cdot) - \mathrm{sign}(\cdot - y)\big)\}(x) = -\delta(x) + \delta(x - y) \\ h_{\mathrm{D}}(0, y) &= \tfrac{1}{2}\left(|0| + |y| - |0 - y|\right) + 1 = 1 \end{aligned}$$

Hence, for all $f \in \mathcal{H}_{\mathrm{D}}$,

$$\begin{aligned} \langle f, h_{\mathrm{D}}(\cdot, y) \rangle_{\mathcal{H}_{\mathrm{D}}} &= \langle \mathrm{D}f, \mathrm{D}\{h_{\mathrm{D}}(\cdot, y)\} \rangle + f(0) \times 1 \\ &= \langle f, \mathrm{D}^*\mathrm{D}\{h_{\mathrm{D}}(\cdot, y)\} \rangle + f(0) \\ &= \big(-f(0) + f(y)\big) + f(0) = f(y). \end{aligned}$$

Native space for derivative operator

$\mathcal{H}_{\mathrm{D}}$: RKHS equipped with the inner product $\langle f, g \rangle_{\mathcal{H}_{\mathrm{D}}} = \langle \mathrm{D}f, \mathrm{D}g \rangle + f(0)g(0)$.

- Direct sum structure
  - "Quotient" space $\mathcal{H}_{\mathrm{D}}/\mathcal{N}_{\mathrm{L}}$: $\quad \mathcal{H}_{\mathrm{D},\delta} = \{f \in \mathcal{H}_{\mathrm{D}} : \langle \delta, f \rangle = f(0) = 0\}$
    $\Rightarrow$ RKHS with inner product $\langle \mathrm{D}f, \mathrm{D}g \rangle$
  - Null space $\mathcal{N}_{\mathrm{D}} = \{f \in \mathcal{H}_{\mathrm{D}} : \mathrm{D}\{f\} = 0\} = \mathrm{span}\{p_1\}$
    $\Rightarrow$ RKHS of dim $N_0 = 1$ with inner product $\langle f, g \rangle_{\mathcal{N}_{\mathrm{L}}} = f(0)g(0)$

$\Rightarrow$ $\boxed{\mathcal{H}_{\mathrm{D}} = \mathcal{H}_{\mathrm{D},\delta} \oplus \mathcal{N}_{\mathrm{L}}}$

Every function $f \in \mathcal{H}_{\mathrm{D}}$ has a unique decomposition as $\quad f = f_1 + f_2$

- $f_1 = (\mathrm{Id} - \mathrm{Proj}_{\mathcal{N}_{\mathrm{D}}})\{f\} = f - f(0)p_1 \in \mathcal{H}_{\mathrm{D},\delta}$
- $f_2 = \mathrm{Proj}_{\mathcal{N}_{\mathrm{D}}}\{f\} = \langle \delta, f \rangle p_1 = f(0)p_1 \in \mathcal{N}_{\mathrm{D}}$

2.7 Operators with non-trivial null spaces

- Specification of native space: first attempt

$$\mathrm{L} : \mathcal{H}_{\mathrm{L}} \to L_2(\mathbb{R}^d)$$

$$\mathcal{H}_{\mathrm{L,ext}} = \{f \in \mathcal{S}'(\mathbb{R}^d) : \|\mathrm{L}f\|^2_{L_2(\mathbb{R}^d)} < \infty\}$$

"Extended" null space: $\mathcal{N}_{\mathrm{L,ext}} = \{q \in \mathcal{H}_{\mathrm{L,ext}} : \mathrm{L}\{q\} = 0\}$.

Potential problem for $d > 1$ as $\mathcal{N}_{\mathrm{L,ext}}$ may not be **finite-dimensional** !!!

- Solution: Restriction of null space to $\mathcal{N}_{\mathrm{L}} = \mathrm{span}\{p_n\}_{n=1}^{N_0} \subseteq \mathcal{N}_{\mathrm{L,ext}}$

  Fixed quotient space: $\mathcal{H}_Q = \mathcal{H}_{\mathrm{L,ext}}/\mathcal{N}_{\mathrm{L,ext}} = \mathcal{H}_{\mathrm{L}}/\mathcal{N}_{\mathrm{L}}$

Proposition
The native space of L, $\mathcal{H}_{\mathrm{L}} = \mathcal{H}_Q \oplus \mathcal{N}_{\mathrm{L}}$, where $\mathcal{N}_{\mathrm{L}}$ is a **finite-dimensional null space** of L endowed with the inner product $\langle \cdot, \cdot \rangle_{\mathcal{N}_{\mathrm{L}}}$, is a **Hilbert space** for the inner product

$$\langle f, g \rangle_{\mathcal{H}_{\mathrm{L}}} = \langle \mathrm{L}f, \mathrm{L}g \rangle_{L_2} + \langle \mathrm{Proj}_{\mathcal{N}_{\mathrm{L}}}\{f\}, \mathrm{Proj}_{\mathcal{N}_{\mathrm{L}}}\{g\} \rangle_{\mathcal{N}_{\mathrm{L}}}$$

where $\mathrm{Proj}_{\mathcal{N}_{\mathrm{L}}}$ is a projection operator from $\mathcal{H}_{\mathrm{L}} \to \mathcal{N}_{\mathrm{L}}$.

2.7.1 Hilbert-space structure of null space

- Motivation: $\mathrm{L}s = w \quad \Rightarrow \quad s = \mathrm{L}^{-1}w$?

Remove the null-space ambiguity by imposing N_0 linear boundary conditions: $\boldsymbol{\phi}(s) = \mathbf{0}$.

- Admissible operator $\mathrm{L} : \mathcal{H}_{\mathrm{L}} \to L_2(\mathbb{R}^d)$
- Null space: $\mathcal{N}_{\mathrm{L}} = \mathrm{span}\{p_n\}_{n=1}^{N_0}$ equipped with a norm $\|\cdot\|_{\mathcal{N}} : \mathcal{H}_{\mathrm{L}} \to \mathbb{R}^+$
- Linear "boundary" functionals $\boldsymbol{\phi} : \mathcal{H}_{\mathrm{L}} \to \mathbb{R}^{N_0}$

$$f \mapsto \boldsymbol{\phi}(f) = (\langle \phi_1, f \rangle, \ldots, \langle \phi_{N_0}, f \rangle)$$

Definition
Let $\mathrm{L} : \mathcal{H}_{\mathrm{L}} \to L_2(\mathbb{R}^d)$ be a spline-admissible differential operator with a null space $\mathcal{N}_{\mathrm{L}}$ of dimension N_0. The pair $(\mathrm{L}, \boldsymbol{\phi})$ is said to be **admissible** if there exist two constants $B \geq A > 0$ such that

$$A\|f\|_{\mathcal{N}} \leq \|\boldsymbol{\phi}(f)\|_2 \leq B\left(\|\mathrm{L}f\|_{L_2} + \|f\|_{\mathcal{N}}\right), \quad \forall f \in \mathcal{H}_{\mathrm{L}}.$$

Lower bound: $\boldsymbol{\phi}(q) = \mathbf{0} \quad \Leftrightarrow \quad q = 0 \quad \forall q \in \mathcal{N}_{\mathrm{L}}$ Upper bound $\Leftrightarrow \quad \phi_n \in \mathcal{H}'_{\mathrm{L}}$

Biorthogonal system

- Basis for $\mathcal{N}_{\mathrm{L}}$: $\boldsymbol{p} = (p_1, \ldots, p_{N_0})$
- Analysis functionals: $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_{N_0})$

- Biorthogonality relation

$$\langle \phi_m, p_n \rangle = \delta_{m,n} \quad \Leftrightarrow \quad \boldsymbol{\phi}(p_n) = \mathbf{e}_n \quad \Leftrightarrow \quad \langle \boldsymbol{\phi}, \boldsymbol{p}^T \rangle = \mathbf{I}_{N_0}$$

- Projector onto $\mathcal{N}_{\mathrm{L}} = \mathrm{span}\{p_n\}_{n=1}^{N_0} \quad \perp \quad \mathcal{N}_{\mathrm{L}}' = \mathrm{span}\{\phi_n\}_{n=1}^{N_0}$

$$\begin{aligned} \mathrm{Proj}_{\mathcal{N}_{\mathrm{L}}} : \mathcal{H}_{\mathrm{L}} &\to \mathcal{N}_{\mathrm{L}} \\ f &\mapsto \sum_{n=1}^{N_0} \langle \phi_n, f \rangle p_n \end{aligned}$$

RKHS associated with null space

- Biorthogonal system: $\{(\phi_n, p_n)\}_{n=1}^{N_0}$

Proposition

Let $(\boldsymbol{\phi}, \boldsymbol{p})$ be a biorthogonal system for $\mathcal{N}_{\mathrm{L}} \subseteq C_{\mathrm{b},\alpha}(\mathbb{R}^d)$. Then, $\mathcal{N}_{\mathrm{L}} = \mathrm{span}\{p_n\}_{n=1}^{N_0}$ equipped with the inner product

$$\langle f, g \rangle_{\mathcal{N}_{\mathrm{L}}} = \sum_{n=1}^{N_0} \langle \phi_n, f \rangle \langle \phi_n, g \rangle = \langle \mathrm{R}_{\boldsymbol{\phi}}\{f\}, g \rangle, \quad f, g \in \mathcal{N}_{\mathrm{L}}$$

is a RKHS with reproducing kernel

$$\mathrm{R}_{\boldsymbol{p}}\{\delta(\cdot - \boldsymbol{y})\}(\boldsymbol{x}) = \sum_{n=1}^{N_0} p_n(\boldsymbol{x}) p_n(\boldsymbol{y}).$$

$$\Leftrightarrow \quad \forall f \in \mathcal{N}_{\mathrm{L}} : \quad \|f\|_{\mathcal{N}_{\mathrm{L}}} = \|\boldsymbol{\phi}(f)\|_2$$

Dual of null space

- Biorthogonal system: $\{(\phi_n, p_n)\}_{n=1}^{N_0}$

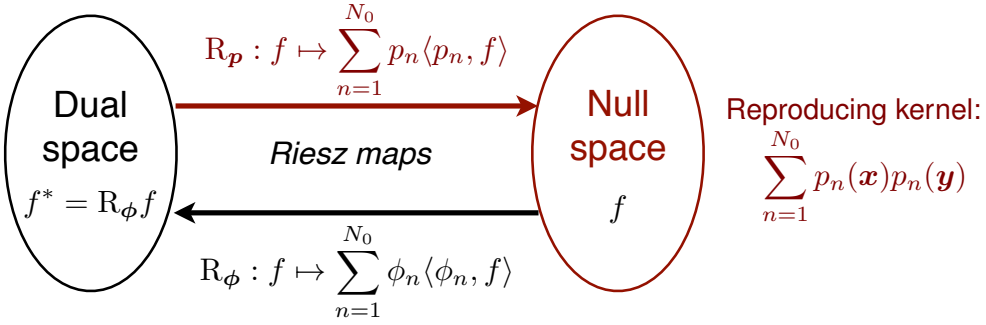
Proposition

The dual of $\mathcal{N}_{\mathrm{L}} = \mathrm{span}\{p_n\}_{n=1}^{N_0}$ is the Hilbert space $\mathcal{N}'_{\mathrm{L}} = \mathrm{span}\{\phi_n = p_n^*\}_{n=1}^{N_0}$ equipped with the inner product

$$\langle f^*, g^* \rangle_{\mathcal{N}'_{\mathrm{L}}} = \sum_{n=1}^{N_0} \langle p_n, f^* \rangle \langle p_n, g^* \rangle = \langle \underbrace{\mathrm{R}_{\boldsymbol{p}} f^*}_{=f}, g^* \rangle, \quad f^*, g^* \in \mathcal{N}'_{\mathrm{L}}.$$

$$\mathcal{N}'_{\mathrm{L}} = \mathrm{span}\{\phi_n = \mathrm{R}_{\boldsymbol{\phi}} p_n\}$$

$$\mathcal{N}_{\mathrm{L}} = \mathrm{span}\{p_n = \mathrm{R}_{\boldsymbol{p}} \phi_n\}$$



Reproducing kernel:

$$\sum_{n=1}^{N_0} p_n(\boldsymbol{x}) p_n(\boldsymbol{y})$$

Summary of operators

Description	Operator	Kernel
Riesz map $\mathcal{N}'_{\mathrm{L}} \to \mathcal{N}_{\mathrm{L}}$	$\mathrm{R}_{\boldsymbol{p}}$	$\sum_{n=1}^{N_0} p_n(\boldsymbol{x}) p_n(\boldsymbol{y})$ reproducing kernel
Riesz map $\mathcal{N}_{\mathrm{L}} \to \mathcal{N}'_{\mathrm{L}}$	$\mathrm{R}_{\boldsymbol{\phi}}$	$\sum_{n=1}^{N_0} \phi_n(\boldsymbol{x}) \phi_n(\boldsymbol{y})$
Projector $\mathcal{H}_{\mathrm{L}} \to \mathcal{N}_{\mathrm{L}}$	$\mathrm{Proj}_{\mathcal{N}_{\mathrm{L}}}$	$\sum_{n=1}^{N_0} p_n(\boldsymbol{x}) \phi_n(\boldsymbol{y})$
Projector $\mathcal{H}'_{\mathrm{L}} \to \mathcal{N}'_{\mathrm{L}}$	$\mathrm{Proj}_{\mathcal{N}'_{\mathrm{L}}}$	$\sum_{n=1}^{N_0} \phi_n(\boldsymbol{x}) p_n(\boldsymbol{y})$

- Relations
  - $\mathrm{R}_{\boldsymbol{\phi}}\{p_n\} = p_n^* = \phi_n$
  - $\mathrm{R}_{\boldsymbol{p}}\mathrm{R}_{\boldsymbol{\phi}}\{f\} = f$ for all $f \in \mathcal{N}_{\mathrm{L}}$
  - $\mathrm{Proj}_{\mathcal{N}_{\mathrm{L}}} = (\mathrm{Proj}_{\mathcal{N}'_{\mathrm{L}}})^*$
  - $\mathrm{R}_{\boldsymbol{p}}\{\phi_n\} = p_n, \quad n = 1, \ldots, N_0$
  - $\mathrm{R}_{\boldsymbol{\phi}}\mathrm{R}_{\boldsymbol{p}}\{f^*\} = f^*$ for all $f^* \in \mathcal{N}'_{\mathrm{L}}$
  - $\mathrm{Proj}_{\mathcal{N}'_{\mathrm{L}}} = (\mathrm{Proj}_{\mathcal{N}_{\mathrm{L}}})^*$

2.7.2 Conditional positive-definiteness

Classical definition

Let $\mathcal{N} = \mathrm{span}\{p_n\}_{n=1}^{N_0}$ be a finite-dimensional subspace of $C_{\mathrm{b},\alpha}(\mathbb{R}^d)$. Then, the kernel function $h : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is said to $\mathcal{N}$-**conditionally positive-definite** if

$$\sum_{m=1}^{N}\sum_{n=1}^{N} z_m h(\boldsymbol{x}_m, \boldsymbol{x}_n) z_n \geq 0$$

for any $N \in \mathbb{N}$, $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N \in \mathbb{R}^d$, and $z_1, \ldots, z_N \in \mathbb{R}$, subject to the condition

$$\sum_{m=1}^{N} z_m p_n(\boldsymbol{x}_m) = 0$$

for $n = 1, \ldots, N_0$. The conditional positive-definiteness is said to be **strict** if

$$\sum_{m=1}^{N}\sum_{n=1}^{N} z_m h(\boldsymbol{x}_m, \boldsymbol{x}_n) z_n > 0,$$

under the same conditions with $(z_1, \ldots, z_N) \in \mathbb{R}^N \backslash \{\mathbf{0}\}$ and the $\boldsymbol{x}_n$ all being distinct.

$\Leftrightarrow$ $\boldsymbol{p}$-**conditional positivity** where $\boldsymbol{p} = (p_1, \ldots, p_{N_0})$ represents a basis of $\mathcal{N}$

(Conditionally) positive-definite operator

Definition

Let A be a continuous operator $\mathcal{S}(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$ and $\mathcal{N}$ some finite-dimensional subspace of $\mathcal{S}'(\mathbb{R}^d)$ spanned by $\boldsymbol{p} = (p_1, \ldots, p_{N_0})$. The operator A is said to be:

- **Symmetric** or **self-adjoint** if, for all $\varphi_1, \varphi_2 \in \mathcal{S}(\mathbb{R}^d)$,

$$\langle \mathrm{A}\varphi_1, \varphi_2 \rangle = \langle \mathrm{A}\varphi_2, \varphi_1 \rangle.$$

- **Positive-definite** if, for any $\varphi \in \mathcal{S}(\mathbb{R}^d)$,

$$\langle \mathrm{A}\varphi, \varphi \rangle \geq 0.$$

- $\mathcal{N}$-**conditionally positive-definite** (or $\boldsymbol{p}$-conditionally positive), if

$$\langle \mathrm{A}\varphi, \varphi \rangle \geq 0$$

for any $\varphi \in \mathcal{S}_{\boldsymbol{p}}(\mathbb{R}^d) = \mathcal{S}(\mathbb{R}^d) \cap \mathcal{N}^{\perp} = \{\varphi \in \mathcal{S}(\mathbb{R}^d) : \boldsymbol{p}(\varphi) = \mathbf{0}\}$.

- **Strictly** $\mathcal{N}$-**conditionally positive-definite** if, for all $\varphi \in \mathcal{S}_{\boldsymbol{p}}(\mathbb{R}^d) \backslash \{0\}$,

$$\langle \mathrm{A}\varphi, \varphi \rangle > 0.$$

Positive-definiteness: equivalences

Theorem

$\mathrm{A} : \varphi \mapsto \int_{\mathbb{R}^d} a(\cdot, \boldsymbol{y})\varphi(\boldsymbol{y})\mathrm{d}\boldsymbol{y}$ with $a(\boldsymbol{x}_0, \cdot) = a(\cdot, \boldsymbol{x}_0) \in C_{\mathrm{b},\alpha}(\mathbb{R}^d)$ for any $\boldsymbol{x}_0 \in \mathbb{R}^d$

($\boldsymbol{p}$-conditional) positive definiteness of $\mathrm{A} : \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$

$\Updownarrow$

($\boldsymbol{p}$-conditional) positive-definiteness of **continuous** kernel $a : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$.

Sketch of proof:

Direct part

Sequence of test functions $\varphi_k = \sum_{n=1}^{N} z_n u_k(\cdot - \boldsymbol{x}_n) \in \mathcal{S}(\mathbb{R}^d)$ such that

$$\lim_{k\to\infty} \langle f, \varphi_k(\cdot - \boldsymbol{x}_n)\rangle = \sum_{n=1}^{N} z_n f(\boldsymbol{x}_m) \qquad u_k = k^d \mathrm{e}^{-\frac{1}{2}\|k\boldsymbol{x}\|^2} \xrightarrow[k\to\infty]{} \delta$$

Indirect part

Making the link with Riemann integrals; i.e.,

$$\langle f, \varphi\rangle = \lim_{i\to\infty} \frac{1}{i^d} \sum_{\boldsymbol{m}=-(i^2,\ldots,i^2)}^{(i^2,\ldots,i^2)} f(\boldsymbol{x}_{\boldsymbol{m}}) \underbrace{\varphi(\boldsymbol{x}_{\boldsymbol{m}})}_{z_{\boldsymbol{m}}}, \qquad \text{with } \boldsymbol{x}_{\boldsymbol{m}} = \tfrac{\boldsymbol{m}}{i} \in \mathbb{R}^d$$

2.7.3 Spline-admissible operator

Definition

A linear operator $\mathrm{L} : \mathcal{H}_{\mathrm{L}} \to L_2(\mathbb{R}^d)$ is called **spline-admissible** if there exists a symmetric kernel $G_{\mathrm{L}^*\mathrm{L}} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, a finite-dimensional subspace $\mathcal{N}_{\mathrm{L}} = \mathrm{span}\{p_n\}_{n=1}^{N_0}$ and an order $\alpha \in \mathbb{R}$ of algebraic growth such that :

1. $G_{\mathrm{L}^*\mathrm{L}}$ is a **Green's function** of $(\mathrm{L}^*\mathrm{L})$ with the property that

$$\mathrm{L}\{G_{\mathrm{L}^*\mathrm{L}}(\cdot, \boldsymbol{y})\}(\boldsymbol{x}) = G_{\mathrm{L}^*}(\boldsymbol{x}, \boldsymbol{y}) = G_{\mathrm{L}}(\boldsymbol{y}, \boldsymbol{x})$$
$$\mathrm{L}^*\mathrm{L}\{G_{\mathrm{L}^*\mathrm{L}}(\cdot, \boldsymbol{y})\} = \mathrm{L}^*\{G_{\mathrm{L}^*}(\cdot, \boldsymbol{y})\} = \delta(\cdot - \boldsymbol{y})$$

2. **Null-space property**: $\mathrm{L}\{q\} = 0$ for all $q \in \mathcal{N}_{\mathrm{L}} \subseteq \mathcal{H}_{\mathrm{L}}$.

3. $G_{\mathrm{L}^*\mathrm{L}}$ is **strictly $\mathcal{N}_{\mathrm{L}}$-conditionally positive-definite**.

4. **Continuity** and polynomial growth:
   $G_{\mathrm{L}^*\mathrm{L}}(\cdot, \boldsymbol{y}_0) \in C_{\mathrm{b},\alpha}(\mathbb{R}^d)$ for any $\boldsymbol{y}_0 \in \mathbb{R}^d$ and $\mathcal{N}_{\mathrm{L}} \subseteq C_{\mathrm{b},\alpha}(\mathbb{R}^d)$.

5. **Boundedness** on the diagonal: $\sup_{\boldsymbol{x}\in\mathbb{R}^d} |G_{\mathrm{L}^*\mathrm{L}}(\boldsymbol{x}, \boldsymbol{x})| \, (1 + \|\boldsymbol{x}\|)^{2\alpha} < \infty$.

Green's function

Definition

The kernel $G_{\mathrm{L}} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is a Green's function of L if $\mathrm{L}\{G_{\mathrm{L}}(\cdot, \boldsymbol{y})\} = \delta(\cdot - \boldsymbol{y})$.

$\Leftrightarrow \quad \mathrm{L}\mathrm{L}^{-1}\{\varphi\} = \varphi$ for any $\varphi \in \mathcal{S}(\mathbb{R}^d)$ where $\mathrm{L}^{-1} : \varphi \mapsto \displaystyle\int_{\mathbb{R}^d} G_{\mathrm{L}}(\cdot, \boldsymbol{y})\varphi(\boldsymbol{y})\mathrm{d}\boldsymbol{y}$

- Preferred scenario: L is LSI with requency response $\widehat{L}(\boldsymbol{\omega})$

$$G_{\mathrm{L}}(\boldsymbol{x}, \boldsymbol{y}) = \rho_{\mathrm{L}}(\boldsymbol{x} - \boldsymbol{y}) \quad \text{with} \quad \rho_{\mathrm{L}}(\boldsymbol{x}) = \mathcal{F}^{-1}\left\{\frac{1}{\widehat{L}(\boldsymbol{\omega})}\right\}(\boldsymbol{x}) \quad = \text{canonical } \textbf{Green's} \text{ function}$$

- Example of admissible operator: $\mathrm{D}^m = \frac{\mathrm{d}^m}{\mathrm{d}x^m} \qquad \overset{\mathcal{F}}{\longleftrightarrow} \qquad (\mathrm{j}\omega)^m$
  - Green's function (causal): $\mu_m(x) = \frac{x_+^{m-1}}{(m-1)!}$
  - Green's function (canonical): $\rho_{\mathrm{D}^m}(x) = \mathcal{F}^{-1}\left\{\frac{1}{(\mathrm{j}\omega)^m}\right\}(x) = \frac{1}{2}\mathrm{sign}(x)\frac{x^{m-1}}{(m-1)!}$
  - Null space = polynomials of degree $m-1$: $\mathcal{N}_{\mathrm{D}^m} = \mathrm{span}\{\frac{x^{n-1}}{(n-1)!}\}_{n=1}^{N_0}$

$\mathcal{N}_{\mathrm{D}^m}$-conditionally-positive kernel: $G(x,y) = \dfrac{(-1)^m}{2}\dfrac{|x-y|^{2m-1}}{(2m-1)!}$

2.7.4 RKHS associated with admissible operator

- "Orthogonal" complement of $\mathcal{N}_{\mathrm{L}}$ in $\mathcal{H}_{\mathrm{L}}$

$$\mathcal{H}_{\mathrm{L},\boldsymbol{\phi}} = \big\{f \in \mathcal{H}_{\mathrm{L}} : \boldsymbol{\phi}(f) = \mathbf{0}\big\},$$

Theorem 3

Let L be an admissible operator and $(\boldsymbol{\phi}, \boldsymbol{p})$ a corresponding biorthogonal system for $\mathcal{N}_{\mathrm{L}}$. Then, $\mathcal{H}_{\mathrm{L},\boldsymbol{\phi}}$ is a **Hilbert space** equipped with the **inner product** $\langle f, g\rangle_{\mathrm{L}} = \langle \mathrm{L}\{f\}, \mathrm{L}\{g\}\rangle$. Moreover, there exists an isometric map $\mathrm{L}_{\boldsymbol{\phi}}^{-1} : L_2(\mathbb{R}^d) \to \mathcal{H}_{\mathrm{L},\boldsymbol{\phi}}$ such that

$$\mathcal{H}_{\mathrm{L},\boldsymbol{\phi}} = \big\{f = \mathrm{L}_{\boldsymbol{\phi}}^{-1} w : w \in L_2(\mathbb{R}^d)\big\}.$$

The operator $\mathrm{L}_{\boldsymbol{\phi}}^{-1}$ (see next slide) is uniquely specified through the following properties

1. **Right-inverse property**: $\mathrm{L}\mathrm{L}_{\boldsymbol{\phi}}^{-1} w = w$ for all $w \in L_2(\mathbb{R}^d)$
2. **Boundary conditions**: $\boldsymbol{\phi}(\mathrm{L}_{\boldsymbol{\phi}}^{-1} w) = \mathbf{0}$ for all $w \in L_2(\mathbb{R}^d)$.

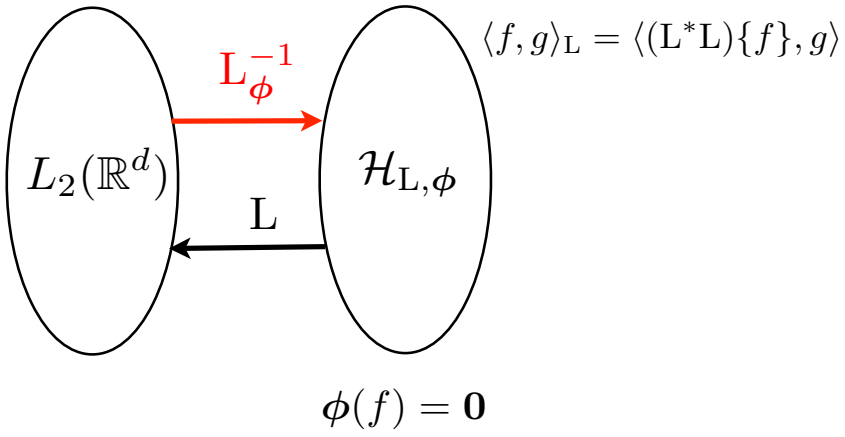
Construction of stable right-inverse

- Ingredients for construction of $\mathrm{L}_{\boldsymbol{\phi}}^{-1} : L_2(\mathbb{R}^d) \to \mathcal{H}_{\mathrm{L},\boldsymbol{\phi}} \subseteq C_{\mathrm{b},\alpha_0}(\mathbb{R}^d)$
  - Basis for $\mathcal{N}_{\mathrm{L}}$: $\boldsymbol{p} = (p_1, \ldots, p_{N_0})$
  - Biorthogonal analysis functionals: $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_{N_0})$
  - Green's function $G_{\mathrm{L}} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ of L $\quad\Leftrightarrow\quad$ $\mathrm{L}\{G_{\mathrm{L}}(\cdot, \boldsymbol{y})\} = \delta(\cdot - \boldsymbol{y})$
  - $q_n(\boldsymbol{y}) = \langle \phi_n, G_{\mathrm{L}}(\cdot, \boldsymbol{y}) \rangle = \mathrm{G}^*\{\phi_n\}(\boldsymbol{y}), \quad n = 1, \ldots, N_0$

$$\mathrm{L}_{\boldsymbol{\phi}}^{-1} : \varphi \mapsto \int_{\mathbb{R}^d} g_{\boldsymbol{\phi}}(\cdot, \boldsymbol{y}) \varphi(\boldsymbol{y})$$

$$g_{\boldsymbol{\phi}}(\boldsymbol{x}, \boldsymbol{y}) = G_{\mathrm{L}}(\boldsymbol{x}, \boldsymbol{y}) - \sum_{n=1}^{N_0} p_n(\boldsymbol{x}) q_n(\boldsymbol{y})$$

$$\mathcal{H}_{\mathrm{L},\boldsymbol{\phi}} \subseteq C_{\mathrm{b},\alpha_0}(\mathbb{R}^d) \quad \Leftrightarrow \quad \sup_{\boldsymbol{x} \in \mathbb{R}^d} (1 + \|\boldsymbol{x}\|)^{\alpha_0} \|g_{\boldsymbol{\phi}}(\boldsymbol{x}, \cdot)\|_{L_2(\mathbb{R}^d)} < \infty$$



2.7.5 Determination of reproducing kernel

Generalized impulse response of $\mathrm{A}_{\boldsymbol{\phi}} = \mathrm{L}_{\boldsymbol{\phi}}^{-1}\mathrm{L}_{\boldsymbol{\phi}}^{-1*}$

Theorem 4

The **reproducing kernel** of the Hilbert space $\mathcal{H}_{\mathrm{L},\boldsymbol{\phi}}$ specified in Theorem 3 is

$$a_{\boldsymbol{\phi}}(\boldsymbol{x},\boldsymbol{y}) = \int_{\mathbb{R}^d} g_{\boldsymbol{\phi}}(\boldsymbol{x},\boldsymbol{z}) g_{\boldsymbol{\phi}}(\boldsymbol{y},\boldsymbol{z}) \mathrm{d}\boldsymbol{z}$$

where $g_{\boldsymbol{\phi}}(\boldsymbol{x},\boldsymbol{y}) = \mathrm{L}_{\boldsymbol{\phi}}^{-1}\{\delta(\cdot - \boldsymbol{y})\}(\boldsymbol{x})$. Moreover, if $G_{\mathrm{L}^*\mathrm{L}}(\boldsymbol{x},\boldsymbol{y})$ is the symmetric Green's function of $\mathrm{L}^*\mathrm{L}$, then the reproducing kernel can be expressed as

$$a_{\boldsymbol{\phi}}(\boldsymbol{x},\boldsymbol{y}) = G_{\mathrm{L}^*\mathrm{L}}(\boldsymbol{x},\boldsymbol{y}) - \sum_{n=1}^{N_0} p_n(\boldsymbol{x}) v_n(\boldsymbol{y}) - \sum_{n=1}^{N_0} v_n(\boldsymbol{x}) p_n(\boldsymbol{y}) + \sum_{m=1}^{N_0}\sum_{n=1}^{N_0} r_{m,n} p_m(\boldsymbol{x}) p_n(\boldsymbol{y})$$

with

$$v_n(\boldsymbol{y}) = \langle \phi_n, G_{\mathrm{L}^*\mathrm{L}}(\cdot,\boldsymbol{y})\rangle = \int_{\mathbb{R}^d} \phi_n(\boldsymbol{z}) G_{\mathrm{L}^*\mathrm{L}}(\boldsymbol{z},\boldsymbol{y}) \mathrm{d}\boldsymbol{z}$$

$$r_{m,n} = \langle \phi_m \otimes \phi_n, G_{\mathrm{L}^*\mathrm{L}}\rangle = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \phi_m(\boldsymbol{x})\phi_n(\boldsymbol{y}) G_{\mathrm{L}^*\mathrm{L}}(\boldsymbol{x},\boldsymbol{y}) \mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y}.$$

Reproducing kernel and Riesz map

$$a_{\boldsymbol{\phi}}(\boldsymbol{x},\boldsymbol{y}) = G_{\mathrm{L}^*\mathrm{L}}(\boldsymbol{x},\boldsymbol{y}) - \sum_{n=1}^{N_0} p_n(\boldsymbol{x}) v_n(\boldsymbol{y}) - \sum_{n=1}^{N_0} v_n(\boldsymbol{x}) p_n(\boldsymbol{y}) + \sum_{m=1}^{N_0}\sum_{n=1}^{N_0} r_{m,n} p_m(\boldsymbol{x}) p_n(\boldsymbol{y})$$

- Riesz map $\mathcal{H}'_{\mathrm{L},\boldsymbol{\phi}} \to \mathcal{H}_{\mathrm{L},\boldsymbol{\phi}}$

$$\mathrm{A}_{\boldsymbol{\phi}} = \mathrm{L}_{\boldsymbol{\phi}}^{-1}\mathrm{L}_{\boldsymbol{\phi}}^{-1*} : \varphi \mapsto \int_{\mathbb{R}^d} a_{\boldsymbol{\phi}}(\cdot,\boldsymbol{y})\varphi(\boldsymbol{y})\mathrm{d}\boldsymbol{y}$$

- Link with unregularized inverse A : $\varphi \mapsto \int_{\mathbb{R}^d} G_{\mathrm{L}^*\mathrm{L}}(\boldsymbol{x},\boldsymbol{y})\varphi(\boldsymbol{y})\mathrm{d}\boldsymbol{y}$
  - $v_n(\boldsymbol{y}) = \langle \phi_n, G_{\mathrm{L}^*\mathrm{L}}(\cdot,\boldsymbol{y})\rangle = \mathrm{A}^*\{\phi_n\}(\boldsymbol{y}) = \mathrm{A}\{\phi_n\}(\boldsymbol{y})$
  - $r_{m,n} = \langle \phi_m \otimes \phi_n, G_{\mathrm{L}^*\mathrm{L}}\rangle = \langle \mathrm{A}\{\phi_m\}, \phi_n\rangle$

$$(\mathrm{L}^*\mathrm{L})\mathrm{A}\{\varphi\} = (\mathrm{L}^*\mathrm{L})\mathrm{A}_{\boldsymbol{\phi}}\{\varphi\} = \varphi \quad \text{for all } \varphi \in \mathcal{S}(\mathbb{R}^d)$$

but A (unlike $\mathrm{A}_{\boldsymbol{\phi}}$) **does not admit** a stable extension $\mathcal{H}'_{\mathrm{L}} \to \mathcal{H}_{\mathrm{L}}$

Characterization of dual space

Proposition

Let L be an admissible operator and $(\boldsymbol{\phi}, \boldsymbol{p})$ a corresponding biorthogonal system for $\mathcal{N}_{\mathrm{L}}$. Then, the continuous dual of the RKHS $\mathcal{H}_{\mathrm{L},\boldsymbol{\phi}}$ is the Hilbert space

$$\mathcal{H}'_{\mathrm{L},\boldsymbol{\phi}} = \{f^* = \mathrm{L}^* w : w \in L_2(\mathbb{R}^d)\}$$

equipped with the inner product

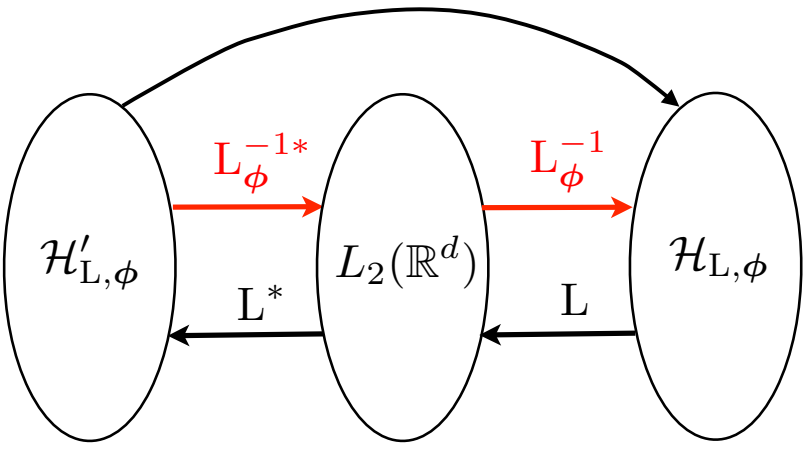
$$\langle f, g \rangle_{\mathcal{H}'} = \langle \mathrm{L}_{\boldsymbol{\phi}}^{-1*} f, \mathrm{L}_{\boldsymbol{\phi}}^{-1*} g \rangle_{L_2(\mathbb{R}^d)} = \langle \mathrm{A} f^*, g^* \rangle$$

where the operators $\mathrm{L}_{\boldsymbol{\phi}}^{-1}$ and A are as previously defined. The operator $(\mathrm{L}^*\mathrm{L})$ is the **Riesz map** $\mathcal{H}_{\mathrm{L},\boldsymbol{\phi}} \to \mathcal{H}'_{\mathrm{L},\boldsymbol{\phi}}$ so that any $f^* = \mathrm{L}^*\mathrm{L}\{f\} \in \mathcal{H}'_{\mathrm{L},\boldsymbol{\phi}}$ can be viewed as the Riesz conjugate of some corresponding $f \in \mathcal{H}_{\mathrm{L},\boldsymbol{\phi}}$. Finally, we have the "**orthogonality**" property

$$\boldsymbol{p}(f^*) = \mathbf{0} \quad \Leftrightarrow \quad \boldsymbol{p}\big(\mathrm{L}^*\mathrm{L}\{f\}\big) = \mathbf{0}$$

for any $f^* \in \mathcal{H}'_{\mathrm{L},\boldsymbol{\phi}}$ and/or $f \in \mathcal{H}_{\mathrm{L},\boldsymbol{\phi}}$.

Riesz map: $\mathrm{A}_{\boldsymbol{\phi}} = \mathrm{L}_{\boldsymbol{\phi}}^{-1}\mathrm{L}_{\boldsymbol{\phi}}^{-1*}$



$\boldsymbol{p}(f^*) = \mathbf{0}$ $\qquad$ $\boldsymbol{\phi}(f) = \mathbf{0}$

Properties of positive operator $\mathrm{A}_{\boldsymbol{\phi}} = \mathrm{L}_{\boldsymbol{\phi}}^{-1}\mathrm{L}_{\boldsymbol{\phi}}^{-1*}$

Proposition

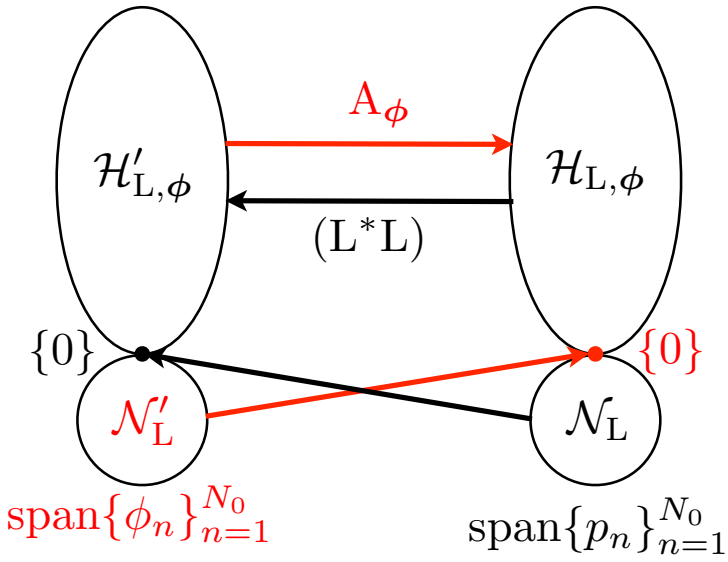
The operator $\mathrm{A}_{\boldsymbol{\phi}} : \varphi \mapsto \int_{\mathbb{R}^d} a_{\boldsymbol{\phi}}(\cdot, \boldsymbol{y})\varphi(\boldsymbol{y})\mathrm{d}\boldsymbol{y}$, where $a_{\boldsymbol{\phi}}(\cdot, \boldsymbol{y})$ is given in Theorem 3, has the following properties:

1. It is the Riesz map $\mathcal{H}'_{\mathrm{L},\boldsymbol{\phi}} \to \mathcal{H}_{\mathrm{L},\boldsymbol{\phi}} = \{f \in \mathcal{H}_{\mathrm{L}} : \boldsymbol{\phi}(f) = \mathbf{0}\}$.
2. It is bounded $\mathcal{H}'_{\mathrm{L}} \to \mathcal{H}_{\mathrm{L}} = \mathcal{H}_{\mathrm{L},\boldsymbol{\phi}} \oplus \mathcal{N}_{\mathrm{L}}$.
3. It has a finite-dimensional null space $\mathrm{span}\{\phi_n\}_{n=1}^{N_0} = \mathcal{N}'_{\mathrm{L}}$.
4. It imposes the boundary conditions: $\boldsymbol{\phi}(\mathrm{A}_{\boldsymbol{\phi}}\{f^*\}) = \mathbf{0}$ for all $f^* \in \mathcal{H}'_{\mathrm{L}}$. In other words, $\mathrm{A}_{\boldsymbol{\phi}}$ continuously maps $\mathcal{H}'_{\mathrm{L}} \to \mathcal{H}_{\mathrm{L},\boldsymbol{\phi}}$.
5. Relation with unregularized operator $\mathrm{A} : \varphi \mapsto \int_{\mathbb{R}^d} G_{\mathrm{L}^*\mathrm{L}}(\cdot, \boldsymbol{y})\varphi(\boldsymbol{y})\mathrm{d}\boldsymbol{y}$

$$\forall f^* \in \mathcal{H}'_{\mathrm{L},\boldsymbol{\phi}} : \qquad \mathrm{A}_{\boldsymbol{\phi}}\{f^*\} = \mathrm{A}\{f^*\} - \sum_{n=1}^{N_0} \langle \mathrm{A}\{\phi_n\}, f^* \rangle p_n$$

$$\forall \phi = p^* \in \mathcal{N}'_{\mathrm{L}} : \quad \mathrm{A}_{\boldsymbol{\phi}}\{\phi\} = 0 \quad \text{and} \quad \mathrm{A}\{\phi\} \in \mathcal{H}_{\mathrm{L}}$$

$$\mathcal{H}'_{\mathrm{L}} = \mathcal{H}'_{\mathrm{L},\boldsymbol{\phi}} \oplus \mathcal{N}'_{\mathrm{L}} \qquad \mathcal{H}_{\mathrm{L}} = \mathcal{H}_{\mathrm{L},\boldsymbol{\phi}} \oplus \mathcal{N}_{\mathrm{L}}$$



RHKS structure of native space $\mathcal{H}_{\mathrm{L}}$

Theorem 5

Any $f \in \mathcal{H}_{\mathrm{L}} = \mathcal{H}_{\mathrm{L},\boldsymbol{\phi}} \oplus \mathcal{N}_{\mathrm{L}}$ has a unique representation as

$$f = \mathrm{L}_{\boldsymbol{\phi}}^{-1} w + q$$

where $w = \mathrm{L}f \in L_2(\mathbb{R}^d)$ and $q = \mathrm{Proj}_{\mathcal{N}_{\mathrm{L}}}\{f\} = \sum_{n=1}^{N_0} \langle f, \phi_n \rangle p_n \in \mathcal{N}_{\mathrm{L}}$. Moreover, $\mathcal{H}_{\mathrm{L}}$ equipped with the inner product

$$\langle f, g \rangle_{\mathcal{H}_{\mathrm{L}}} = \langle \mathrm{L}f, \mathrm{L}g \rangle + \sum_{n=1}^{N_0} \langle \phi_n, f \rangle \langle \phi_n, g \rangle$$

is a **RKHS** whose reproducing kernel is

$$h_{\boldsymbol{\phi}}(\boldsymbol{x}, \boldsymbol{y}) = a_{\boldsymbol{\phi}}(\boldsymbol{x}, \boldsymbol{y}) + \sum_{n=1}^{N_0} p_n(\boldsymbol{x}) p_n(\boldsymbol{y}).$$

Finally, if the operator A meets the admissibility conditions (strict conditional positivity and α-boundedness) and the ϕ_n are such that $\mathrm{A}\{\phi_n\} \in C_{\mathrm{b},\alpha}(\mathbb{R}^d)$, then $\mathcal{S}(\mathbb{R}^d) \subseteq \mathcal{H}_{\mathrm{L}} \subseteq C_{\mathrm{b},\alpha}(\mathbb{R}^d) \subseteq \mathcal{S}'(\mathbb{R}^d)$ with the insurance that $\mathcal{H}_{\mathrm{L}}$ is **dense** in $\mathcal{S}'(\mathbb{R}^d)$.

Summary of operators

Description	Operator	Kernel	
Right-inverse of L	G	$G_{\mathrm{L}}(\boldsymbol{x}, \boldsymbol{y})$	$= \mathrm{L}\{G_{\mathrm{L}^*\mathrm{L}}(\cdot, \boldsymbol{x})\}(\boldsymbol{y})$???
Right-inverse of $(\mathrm{L}^*\mathrm{L})$	$\mathrm{A} = \mathrm{G}\mathrm{G}^*$	$G_{\mathrm{L}^*\mathrm{L}}(\boldsymbol{x}, \boldsymbol{y})$	
Stable inverse of L	$\mathrm{L}_{\boldsymbol{\phi}}^{-1}$	$G_{\mathrm{L}}(\boldsymbol{x}, \boldsymbol{y}) - \sum_{n=1}^{N_0} p_n(\boldsymbol{x}) \mathrm{G}^*\{\phi_n\}(\boldsymbol{y})$	
Riesz map $\mathcal{H}'_{\mathrm{L},\boldsymbol{\phi}} \to \mathcal{H}_{\mathrm{L},\boldsymbol{\phi}}$	$\mathrm{A}_{\boldsymbol{\phi}}$	$a_{\boldsymbol{\phi}}(\boldsymbol{x}, \boldsymbol{y})$	reproducing kernel
Riesz map $\mathcal{H}_{\mathrm{L},\boldsymbol{\phi}} \to \mathcal{H}'_{\mathrm{L},\boldsymbol{\phi}}$	$(\mathrm{L}^*\mathrm{L})$		
Riesz map $\mathcal{H}'_{\mathrm{L}} \to \mathcal{H}_{\mathrm{L}}$	$\mathrm{A}_{\boldsymbol{\phi}} + \mathrm{R}_{\boldsymbol{p}}$	$a_{\boldsymbol{\phi}}(\boldsymbol{x}, \boldsymbol{y}) + \sum_{n=1}^{N_0} p_n(\boldsymbol{x}) p_n(\boldsymbol{y})$	reproducing kernel
Riesz map $\mathcal{H}_{\mathrm{L}} \to \mathcal{H}'_{\mathrm{L}}$	$(\mathrm{L}^*\mathrm{L}) + \mathrm{R}_{\boldsymbol{\phi}}$		

- Direct sum decomposition
  - $\forall \phi = p^* \in \mathcal{N}'_{\mathrm{L}}: \quad \mathrm{A}_{\boldsymbol{\phi}}\{\phi\} = 0$
  - $\forall \tilde{f}^* \in \mathcal{H}'_{\mathrm{L},\boldsymbol{\phi}}: \quad \mathrm{R}_{\boldsymbol{p}}\{\tilde{f}^*\} = 0$
  - $\forall p \in \mathcal{N}_{\mathrm{L}}: \quad (\mathrm{L}^*\mathrm{L})\{p\} = 0$
  - $\forall \tilde{f} \in \mathcal{H}_{\mathrm{L},\boldsymbol{\phi}}: \quad \mathrm{R}_{\boldsymbol{\phi}}\{\tilde{f}\} = 0$

Summary of key relations

$\mathrm{L}_{\boldsymbol{\phi}}^{-1} : L_2(\mathbb{R}^d) \to \mathcal{H}_{\mathrm{L},\boldsymbol{\phi}} \subseteq \mathcal{H}_{\mathrm{L}}$

- $\mathrm{L}_{\boldsymbol{\phi}}^{-1} = \big(\mathrm{Id} - \mathrm{Proj}_{\mathcal{N}_{\mathrm{L}}'}\big)^* \mathrm{G} = \big(\mathrm{Id} - \mathrm{Proj}_{\mathcal{N}_{\mathrm{L}}}\big)\mathrm{G}$

  $\forall w \in L_2(\mathbb{R}^d): \quad \mathrm{L}\mathrm{L}_{\boldsymbol{\phi}}^{-1}\{w\} = w,$

  $\forall w \in L_2(\mathbb{R}^d): \quad \boldsymbol{\phi}(\mathrm{L}_{\boldsymbol{\phi}}^{-1} w) = \mathbf{0}$

$\mathrm{L}_{\boldsymbol{\phi}}^{-1*} : \mathcal{H}_{\mathrm{L},\boldsymbol{\phi}}' \to L_2(\mathbb{R}^d)$

- $\mathrm{L}_{\boldsymbol{\phi}}^{-1*} = \mathrm{G}^*(\mathrm{Id} - \mathrm{Proj}_{\mathcal{N}_{\mathrm{L}}'})$

  $\forall w \in L_2(\mathbb{R}^d): \quad \mathrm{L}_{\boldsymbol{\phi}}^{-1*}\mathrm{L}^*\{w\} = w$

  $\forall \phi \in \mathcal{N}_{\mathrm{L}}': \quad \mathrm{L}_{\boldsymbol{\phi}}^{-1*}\{\phi\} = 0$

Riesz map $\mathrm{A}_{\boldsymbol{\phi}} : \mathcal{H}_{\mathrm{L},\boldsymbol{\phi}}' \to \mathcal{H}_{\mathrm{L},\boldsymbol{\phi}} \subseteq \mathcal{H}_{\mathrm{L}}$

- $\mathrm{A}_{\boldsymbol{\phi}} = \big(\mathrm{Id} - \mathrm{Proj}_{\mathcal{N}_{\mathrm{L}}}\big)\big(\mathrm{G}\mathrm{G}^*\big)\big(\mathrm{Id} - \mathrm{Proj}_{\mathcal{N}_{\mathrm{L}}'}\big) = \mathrm{L}_{\boldsymbol{\phi}}^{-1}\mathrm{L}_{\boldsymbol{\phi}}^{-1*}$

  $\forall \phi \in \mathcal{N}_{\mathrm{L}}': \quad \mathrm{A}_{\boldsymbol{\phi}}\{\phi\} = 0$

  $\forall f^* \in \mathcal{H}_{\mathrm{L}}': \quad \boldsymbol{\phi}\big(\mathrm{A}_{\boldsymbol{\phi}}\{f^*\}\big) = \mathbf{0}$

  $\forall \tilde{f}^* \in \mathcal{H}_{\mathrm{L},\boldsymbol{\phi}}': \quad (\mathrm{L}^*\mathrm{L})\mathrm{A}_{\boldsymbol{\phi}}\{\tilde{f}^*\} = \tilde{f}^*$

  $\forall \tilde{f}^* \in \mathcal{H}_{\mathrm{L},\boldsymbol{\phi}}': \quad \mathrm{A}_{\boldsymbol{\phi}}\{\tilde{f}^*\} = \mathrm{A}\{\tilde{f}^*\} + \sum_{n=1}^{N_0} \langle \mathrm{A}\{\phi_n\}, \tilde{f}^* \rangle p_n$

Native space: the bottom line

- The native space of L is a **Hilbert space** $\mathcal{H}_{\mathrm{L}} \subseteq C_{\mathrm{b},\alpha}(\mathbb{R}^d)$ equipped with the inner product

  $$\langle f, g \rangle_{\mathcal{H}_{\mathrm{L}}} = \langle \mathrm{L}f, \mathrm{L}g \rangle + \sum_{n=1}^{N_0} \langle \phi_n, f \rangle \langle \phi_n, g \rangle$$

  with $\phi_n \in \mathcal{H}_{\mathrm{L}}'$ where $(\boldsymbol{\phi}, \boldsymbol{p}) = \big((\phi_n), (p_n)\big)$ is a biothonormal system for $\mathcal{N}_{\mathrm{L}} = \mathrm{span}\{p_n\}_{n=1}^{N_0} \subseteq \mathcal{H}_{\mathrm{L}}$.

- **Direct sum structure**: $\mathcal{H}_{\mathrm{L}} = \mathcal{H}_{\mathrm{L},\boldsymbol{\phi}} \oplus \mathcal{N}_{\mathrm{L}} = \mathcal{H}_{\mathrm{L},\tilde{\boldsymbol{\phi}}} \oplus \mathcal{N}_{\mathrm{L}}$ where $\mathcal{H}_{\mathrm{L},\boldsymbol{\phi}} = \{f \in \mathcal{H}_{\mathrm{L}} : \boldsymbol{\phi}(f) = \mathbf{0}\}$ with the norms induced by the **biorthogonal systems** $(\boldsymbol{\phi}, \boldsymbol{p})$ and $(\tilde{\boldsymbol{\phi}}, \tilde{\boldsymbol{p}})$ being equivalent.

- If there exists a symmetric $\mathcal{N}_{\mathrm{L}}$-conditionally-positive operator $\mathrm{A} : \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$ such that $(\mathrm{L}^*\mathrm{L})\mathrm{A}\{\varphi\} = \varphi$ for all $\varphi \in \mathcal{S}(\mathbb{R}^d)$, then:
  - $\mathcal{H}_{\mathrm{L}}$ can be defined as the completion of $\mathcal{S}(\mathbb{R}^d)$ is the $\|\cdot\|_{\mathcal{H}_{\mathrm{L}}}$-norm.
  - $\mathcal{H}_{\mathrm{L}}$ is **dense** in $\mathcal{S}'(\mathbb{R}^d)$.
  - Any function $f \in C_{\mathrm{b},\alpha}(\mathbb{R}^d)$ can be approximated as closely as desired by a **kernel expansion** of the form

    $$f_K = \sum_{k=1}^{K} a_k G_{\mathrm{L}^*\mathrm{L}}(\cdot, \boldsymbol{x}_k) + \sum_{n=1}^{N_0} b_n p_n$$

    with a finite number $(K + N_0)$ of terms and some (possibly adaptive) centers $\boldsymbol{x}_k \in \mathbb{R}^d$.

CONTENT

- **1. Preliminaries** ✔
- **2. Reproducing kernel Hilbert spaces** ✔
- **3. Variational splines and representer theorems**
  - Regularization induced by an inner product
  - Non-coercive regularization functionals
  - Discretization and numerical solutions
- **4. Gaussian processes**

3. VARIATIONAL SPLINES

- Generalized **interpolation** problem

  Recover $f \in \mathcal{H}_{\mathrm{L}}$ from the measurements $y_m = \langle \nu_m, f \rangle, (m = 1, \ldots, M)$

  Traditional setting: $y_m = f(x_m) = \langle \delta(\cdot - x_m), f \rangle$

- Variational (=minimum energy) solution

  $$f_0 = \arg \min_{f \in \mathcal{H}_{\mathrm{L}}} \|\mathrm{L}f\|_{L_2}^2 \quad \text{s.t.} \quad (f(x_m) = y_m)_{m=1}^M$$

- Minimization of least-squares functional

  Perturbed measurements model: $y_m = \langle \nu_m, f \rangle + \epsilon_m \quad (m = 1, \ldots, M)$

  $$J_{\mathrm{LS}}(f|\mathbf{y}, \lambda) = \sum_{m=1}^{M} \left(y_m - \langle \nu_m, f \rangle\right)^2 + \lambda \|\mathrm{L}f\|_{L_2}^2$$

3.1 Regularization induced by an inner product

Optimization is performed over a Hilbert space $\mathcal{H}$

Effect of measurement functional $\nu_m \in \mathcal{H}'$: $\quad \langle \nu_m, f \rangle = \langle \nu_m^*, f \rangle_{\mathcal{H}}$

$$J_{\text{LS}}(f|\mathbf{y}, \lambda) = \sum_{m=1}^{M} \left(y_m - \langle \nu_m^*, f \rangle_{\mathcal{H}} \right)^2 + \lambda \|f\|_{\mathcal{H}}^2 = \tfrac{1}{2} a(f, f) - v(f) + C_0$$

Linear functional: $v(f) = 2 \sum_{m=1}^{M} y_m \langle \nu_m^*, f \rangle_{\mathcal{H}}$

Bilinear form: $a(f_1, f_2) = 2\lambda \langle f_1, f_2 \rangle_{\mathcal{H}} + 2 \sum_{m=1}^{M} \langle \nu_m^*, f_1 \rangle_{\mathcal{H}} \langle \nu_m^*, f_2 \rangle_{\mathcal{H}}$

$$f_0 = \arg\min_{f \in \mathcal{H}} J(f|\mathbf{y}, \lambda) \quad \Leftrightarrow \quad v(f) = a(f_0, f) \quad \text{(Condition of optimality)}$$

$$\Rightarrow \quad \boxed{f_0 \in \text{span}\{\nu_m^*\}_{m=1}^{M}}$$

with guarantee of unicity (Hilbert's projection theorem)

Regularized least-squares fit: discretization

$\mathcal{H}$: RKHS with reproducing kernel $r_{\mathcal{H}} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$

Solution space parametrized by $\mathbf{a} = (a_1, \ldots, a_M) \in \mathbb{R}^M$

$$f(\boldsymbol{x}) = \sum_{m=1}^{M} a_m \nu_m^*(\boldsymbol{x}), \qquad \nu_m^*(\boldsymbol{x}) = \int_{\mathbb{R}^d} r_{\mathcal{H}}(\boldsymbol{x}, \boldsymbol{y}) \nu_m(\boldsymbol{y}) \mathrm{d}\boldsymbol{y}$$

System/Gram matrix : $\quad [\mathbf{G}]_{m,n} = \langle \nu_m, \nu_n^* \rangle = \langle \nu_m^*, \nu_n^* \rangle_{\mathcal{H}} = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \nu_m(\boldsymbol{x}) r_{\mathcal{H}}(\boldsymbol{x}, \boldsymbol{y}) \nu_n(\boldsymbol{y}) \mathrm{d}\boldsymbol{x} \mathrm{d}\boldsymbol{y}$

$$\Rightarrow \quad J_{\text{LS}}(f|\mathbf{y}, \lambda) = J(\mathbf{a}|\mathbf{y}, \lambda) = \|\mathbf{y} - \mathbf{G}\mathbf{a}\|_2^2 + \lambda \mathbf{a}^T \mathbf{G} \mathbf{a}$$

Condition for optimality: $\quad \dfrac{\partial J(\mathbf{a}|\mathbf{y}, \lambda)}{\partial \mathbf{a}} = 2\mathbf{G}\big(\mathbf{y} - (\mathbf{G} + \lambda \mathbf{I}_M)\mathbf{a}\big) = 0$

\+ unicity

$$\Rightarrow \quad \mathbf{y} = (\mathbf{G} + \lambda \mathbf{I}_M)\mathbf{a} \quad \Leftrightarrow \quad \mathbf{a} = (\mathbf{G} + \lambda \mathbf{I}_M)^{-1} \mathbf{y}$$

Generalization: Abstract representer theorem

- $\mathcal{H}$: Hilbert space with inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ and Riesz map $\mathrm{R} : \mathcal{H}' \to \mathcal{H}$
- $\boldsymbol{\nu} : f \mapsto \boldsymbol{\nu}(f) = (\langle \nu_1, f \rangle, \ldots, \langle \nu_M, f \rangle)$ is a continuous linear operator $\mathcal{H}_{\mathrm{L}} \to \mathbb{R}^M$ that extracts M measurements from the signal f;
- $\mathcal{C}$: closed convex subset of $\mathbb{R}^M$ such that its preimage in $\mathcal{H}$ is nonempty.

Theorem

The solution of the general convex optimization problem

$$\arg\min_{f \in \mathcal{H}} \|f\|_{\mathcal{H}}^2 \quad \text{s.t.} \quad \boldsymbol{\nu}(f) \in \mathcal{C}$$

is unique and of the form

$$f_0 = \sum_{m=1}^{M} a_m \nu_m^*$$

with $\nu_m^* = \mathrm{R}\{\nu_m\} \in \mathcal{H}$ and suitable linear weights $\mathbf{a} = (a_m) \in \mathbb{R}^M$.

Special case: $\mathcal{C}_{\mathrm{LS}} = \{\mathbf{z} \in \mathbb{R}^d : \|\mathbf{z} - \mathbf{y}\|_2^2 \leq \sigma^2\}$

Proof (step by step)

- Unicity of the solution (by Hilbert's projection theorem)
  $\mathcal{C}$ convex & closed, $\boldsymbol{\nu}$ linear & continuous $\quad \Rightarrow \quad$ $\mathcal{U} = \boldsymbol{\nu}^{-1}(\mathcal{C})$ convex & closed
- Riesz' representation theorem: $\langle \nu_m, f \rangle = \langle \nu_m^*, f \rangle_{\mathcal{H}}$ with $\nu_m^* = \mathrm{R}\{\nu_m\} \in \mathcal{H}$
- Identification of reconstruction space: $\mathcal{V} = \mathrm{span}\{\nu_m^*\}_{m=1}^M \subseteq \mathcal{H}$
- Optimization space: $\mathcal{H} = \mathcal{V} \oplus \mathcal{V}^\perp$

  $\Rightarrow \quad f = u + u^\perp$ with $u \in \mathcal{V}$ and $u^\perp \in \mathcal{V}^\perp$
- Null space of $\boldsymbol{\nu}$: $\mathcal{V}^\perp = \{f \in \mathcal{H} : \langle f, \nu_m \rangle = \langle f, \nu_m^* \rangle_{\mathcal{H}} = 0, \ \ m = 1, \ldots, M\}$

  $\Rightarrow \quad f_0 = u_0 + u_0^\perp$ with $\boldsymbol{\nu}(f_0) = \boldsymbol{\nu}(u_0)$ and $\boldsymbol{\nu}(u_0^\perp) = 0$
- f_0 is the minimal-norm solution

  $\|f_0\|_{\mathcal{H}}^2 = \|u_0 + u_0^\perp\|_{\mathcal{H}}^2 \leq \|u_0\|_{\mathcal{H}}^2 \quad \Leftrightarrow \quad u_0^\perp = 0 \quad \Leftrightarrow \quad f_0 \in \mathcal{V}$ QED

Smoothing splines and/or ridge regression

$\mathcal{H}$: RKHS with reproducing kernel $r_{\mathcal{H}} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$

"Ideal" sampling functionals: $\nu_m = \delta(\cdot - \boldsymbol{x}_m), \quad (m = 1, \ldots, M)$

$$f_{\mathrm{LS}} = \arg\min_{f \in \mathcal{H}} \left(\sum_{m=1}^{M} \big(y_m - f(\boldsymbol{x}_m)\big)^2 + \lambda \|f\|_{\mathcal{H}}^2 \right)$$

Solution space parametrized by $\mathbf{a} = (a_1, \ldots, a_M) \in \mathbb{R}^M$

$$f(\boldsymbol{x}) = \sum_{m=1}^{M} a_m r_{\mathcal{H}}(\boldsymbol{x}, \boldsymbol{x}_m)$$

⟵ **kernel estimator**

- Regularized least-squares fit

autocorrelation matrix, ridge

$$\Rightarrow \quad \mathbf{a}_{\mathrm{LS}} = (\mathbf{R} + \lambda \mathbf{I}_M)^{-1} \mathbf{y}$$

Statistical linear regression: $\lambda \propto \sigma^2$

with $\quad [\mathbf{R}]_{m,n} = \langle r_{\mathcal{H}}(\cdot, \boldsymbol{x}_m), r_{\mathcal{H}}(\cdot, \boldsymbol{x}_n) \rangle_{\mathcal{H}} = r_{\mathcal{H}}(\boldsymbol{x}_m, \boldsymbol{x}_n)$

Application to machine learning

- The "machine learning" problem
  - Find a function $f : \mathbb{R}^d \to \mathbb{R}$ such that $f(\boldsymbol{x}_m) \approx y_m$ (with a **training set** of size M)
  - Proximity between $\mathbf{z}(f) = \big(f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_M)\big)$ and $\mathbf{y} = (y_1, \ldots, y_M)$ is measured by some **loss function** $F : \mathbb{R}^M \times \mathbb{R}^M \to \mathbb{R}$

Definition

A multivariate function $g : \mathbb{R}^M \to \mathbb{R}$ is **convex** if

$$g\big(\tau \mathbf{z}_1 + (1-\tau)\mathbf{z}_2\big) \leq \tau g\big(\mathbf{z}_1\big) + (1-\tau) g\big(\mathbf{z}_2, \mathbf{y}\big)$$

for all $\mathbf{z}_1, \mathbf{z}_2 \in \mathbb{R}^M$ and all $\tau \in [0, 1]$. It is **strictly convex** if the order relation holds with a strict inequality for $\mathbf{z}_1 \neq \mathbf{z}_2$ and $\tau \in (0, 1)$.

A multivariate function $g : \mathbb{R}^M \to \mathbb{R}$ is said to be **coercive** if $\lim_{\|\boldsymbol{z}\| \to \infty} g(\mathbf{z}) = \infty$.

The loss function $F : \mathbb{R}^M \times \mathbb{R}^M \to \mathbb{R}$ is said to be convex (resp., coercive) if $F(\cdot, \mathbf{y}) : \mathbb{R}^M \to \mathbb{R}$ is convex (resp., coercive) for any fixed $\mathbf{y} \in \mathbb{R}^d$

Examples of loss functions

Representer theorem for machine learning

- $\mathcal{H} \subseteq C_{\mathrm{b},\alpha}(\mathbb{R}^d)$: RKHS with inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ and reproducing kernel $r_{\mathcal{H}} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$
- Sampling operator $f \mapsto \mathbf{z}(f) = \big(f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_M)\big)$ with $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M \in \mathbb{R}^d$
- Corresponding data values $\mathbf{y} = (y_1, \ldots, y_M) \in \mathbb{R}^M$

Theorem

The solution of the generic minimization problem

$$\arg\min_{f \in \mathcal{H}} \big(F\big(\mathbf{z}(f), \mathbf{y}\big) + \lambda \|f\|_{\mathcal{H}}^2\big)\,,$$

where the cost function F is strictly convex, is unique and of the form

$$f(\boldsymbol{x}) = \sum_{m=1}^{M} a_m r_{\mathcal{H}}(\boldsymbol{x}, \boldsymbol{x}_m)$$

with suitable linear weights $\mathbf{a} = (a_m) \in \mathbb{R}^M$.

Theoretical justification for **kernel** methods

(Schoenberg 1964, Kimeldorf-Wahba 1971) (Schölkopf-Smola 2001)

$$F(\mathbf{z}(f), \mathbf{y}) = \|\mathbf{z}(f) - \mathbf{y}\|_2^2$$

Proof

- $f \mapsto \mathbf{z}(f)$: linear and continuous on $\mathcal{H}$
- $f \mapsto F\big(\mathbf{z}(f), \mathbf{y}\big)$: strictly convex and continuous on $\mathcal{H}$
- $f \mapsto \lambda \|f\|_{\mathcal{H}}^2$: strictly convex, continuous and trivially coercive on $\mathcal{H}$

> **Theorem** (Classical existence result from convex analysis)
> Let $\mathcal{X}$ be a reflexive Banach space and $J : \mathcal{X} \to \mathbb{R}$ a **convex**, **(lower semi-) continuous**, and **coercive** functional on $\mathcal{X}$. Then, the problem
>
> $$S = \arg \inf_{x \in \mathcal{X}} J(x)$$
>
> admits at least one solution. Moreover, the solution set S is a convex subset of $\mathcal{X}$ which reduces to a single point (unique solution) when J is **strictly convex**.

$\Rightarrow$ **Unique solution** f_0 with "optimal" data term $F(\mathbf{y}_0, \mathbf{y})$ where $\mathbf{y}_0 = \mathbf{z}(f_0) \in \mathbb{R}^M$

Since f_0 achieves the minimum, it is also the solution of the interpolation problem

$$f_0 = \min_{f \in \mathcal{H}} \|f\|_{\mathcal{H}}^2 \quad \text{s.t.} \quad \mathbf{z}(f) = \mathbf{y}_0$$

$$\Rightarrow \quad f_0 \in \operatorname{span}\{r_{\mathcal{H}}(\cdot, \boldsymbol{x}_m)\}_{m=1}^{M}$$

3.2 Non-coercive regularization functionals

- Functional context
  - Non-trivial null space $\mathcal{N}_{\mathrm{L}} = \{f \in \mathcal{H}_{\mathrm{L}} : \mathrm{L}\{f\} = 0\}$
  - Biorthogonal system $(\boldsymbol{\phi}, \boldsymbol{p})$ for $\mathcal{N}_{\mathrm{L}} = \operatorname{span}\{p_n\}_{n=1}^{N_0}$ with $\phi_n \in \mathcal{H}'_{\mathrm{L}}$
  - Native space $\mathcal{H}_{\mathrm{L}} \subseteq C_{\mathrm{b},\alpha}(\mathbb{R}^d)$ is the Hilbert space equipped with the composite norm
    $\|f\|_{\mathcal{H}_{\mathrm{L}}} = \sqrt{\|\mathrm{L}f\|_{L_2}^2 + \|\boldsymbol{\phi}(f)\|_2^2}$
- Variational formulation of linear inverse problems

$$\arg \min_{f \in \mathcal{H}_{\mathrm{L}}} \Big(F\big(\boldsymbol{\nu}(f), \mathbf{y}\big) + \lambda \|\mathrm{L}f\|_{L_2(\mathbb{R}^d)}^2 \Big)$$

  - Motivation: favors ("smooth") solution with large null-space component
  - Theoretical challenge: $\|\mathrm{L}f\|_{L_2}^2$ is only a semi-norm !

Proper regularization operator

Definition

The operator $\mathrm{L} : \mathcal{H}_{\mathrm{L}} \to L_2(\mathbb{R}^d)$ with null space $\mathcal{N}_{\mathrm{L}} = \mathrm{span}\{p_n\}_{n=1}^{N_0} \subseteq \mathcal{H}_{\mathrm{L}}$ is a proper regularization operator for the measurement operator $\boldsymbol{\nu} : f \mapsto \boldsymbol{\nu}(f) = (\langle \nu_1, f \rangle, \ldots, \langle \nu_M, f \rangle)$ if the following technical conditions are met:

1. L is spline-admissible
2. $\nu_1, \ldots, \nu_M \in \mathcal{H}'_{\mathrm{L}}$
3. For all $q \in \mathcal{N}_{\mathrm{L}}$, $\|\boldsymbol{\nu}(q)\|_2 \geq 0$ with equality if and only if $q = 0$.

$$\Leftrightarrow \quad \exists c > 0 \quad \text{s.t.} \quad \|q\|^2_{\mathcal{N}_{\mathrm{L}}} = \|\boldsymbol{\phi}(q)\|_2^2 \leq \tfrac{1}{c}\|\boldsymbol{\nu}(q)\|_2^2.$$

- Criterion

  The singular values of $\mathbf{P} = [\boldsymbol{\nu}(p_1) \ \cdots \ \boldsymbol{\nu}(p_{N_0})]$ are strictly positive and bounded

Representer theorem for linear inverse problems

- $\boldsymbol{\nu} : f \mapsto \boldsymbol{\nu}(f) = (\langle \nu_1, f \rangle, \ldots, \langle \nu_M, f \rangle)$ is a continuous linear operator $\mathcal{H}_{\mathrm{L}} \to \mathbb{R}^M$ that extracts M measurements from the signal f;
- $\mathrm{L} : \mathcal{H}_{\mathrm{L}} \to L_2(\mathbb{R}^d)$ is a proper regularization operator;
- $\{p_n\}_{n=1}^{N_0}$ is a basis of the null space of the regularization operator;
- $F : \mathbb{R}^M \times \mathbb{R}^M \to \mathbb{R}$ is a strictly convex and coercive loss function;
- $\mathbf{y} \in \mathbb{R}^M$ is a given data vector and $\lambda \in \mathbb{R}^+$ an adjustable regularization parameter.

Theorem

The solution of the general minimization problem

$$\arg\min_{f \in \mathcal{H}_{\mathrm{L}}} F(\boldsymbol{\nu}(f), \mathbf{y}) + \lambda \|\mathrm{L}f\|^2_{L_2(\mathbb{R}^d)}$$

is **unique** and has the generic **linear parametric form**

$$f_{(\lambda)} = \sum_{m=1}^{M} a_m \varphi_m + \sum_{n=1}^{N_0} b_n p_n$$

with $\varphi_m = \mathrm{A}\{\nu_m\} = \displaystyle\int_{\mathbb{R}^d} G_{\mathrm{L}^*\mathrm{L}}(\cdot, \boldsymbol{y}) \nu_m(\boldsymbol{y}) \mathrm{d}\boldsymbol{y}, \quad \mathbf{a} = (a_m) \in \mathbb{R}^M, \quad \mathbf{b} = (b_n) \in \mathbb{R}^{N_0},$

subject to the "**orthogonality**" constraint: $\langle \mathbf{a}, \boldsymbol{\nu}(p_n) \rangle = 0$ for $n = 1, \ldots, N_0$.

Proof

Minimization of $J(f|\mathbf{y},\lambda) = F(\boldsymbol{\nu}(f),\mathbf{y}) + \lambda\|\mathrm{L}f\|^2_{L_2(\mathbb{R}^d)}$

$\mathcal{H}_\mathrm{L} = \mathcal{H}_{\mathrm{L},\boldsymbol{\phi}} \oplus \mathcal{H}_\mathrm{L} \quad \Leftrightarrow \quad f = \mathrm{L}^{-1}_{\boldsymbol{\phi}} w + q \quad$ with $\quad w = \mathrm{L}f,\ q = \mathrm{Proj}_{\mathcal{N}_\mathrm{L}}\{f\}$

- Step 1: Establish existence and unicity
  - **Continuity** of J: directly follows from continuity of $f \mapsto \|\mathrm{L}f\|^2_{L_2}$, $\boldsymbol{\nu}$, and $F(\cdot,\mathbf{y})$.
  - **Strict-convexity** of J: follows from strict convexity of J and $F(\cdot,\mathbf{y})$, and linearity of $\boldsymbol{\nu}$.
  - **Coercivity** of J: $\|\mathrm{L}f\|^2_{L_2} \to \infty$ as $\|w\|_{L_2} \to \infty$,
    while $F\big(\nu(\mathrm{L}^{-1}_{\boldsymbol{\phi}} w) + \boldsymbol{\nu}(q),\mathbf{y}\big) \to \infty$ as $\|\boldsymbol{\phi}(q)\|_2 \to \infty$

Key relations: $\|f\|^2_{\mathcal{H}_\mathrm{L}} = \|w\|^2_{L_2} + \|\boldsymbol{\phi}(q)\|^2_2$ and $c\|\boldsymbol{\phi}(q)\|_2 \le \|\boldsymbol{\nu}(q)\|_2$

Proof (Cont.)

- Step 2: Generic form of solution

Minimizer: f_0 such that $\boldsymbol{\nu}(f_0) = \mathbf{y}_0 \in \mathbb{R}^M$ and $\boldsymbol{\phi}(f_0) = \mathbf{c}_0 = (c_1,\dots,c_{N_0}) \in \mathbb{R}^{N_0}$

Direct sum decomposition in $\mathcal{H}_\mathrm{L} = \mathcal{H}_{\mathrm{L},\boldsymbol{\phi}} \oplus \mathcal{H}_\mathrm{L}$

$f_0 = \tilde{f}_0 + q_0 \quad$ with $\quad \tilde{f}_0 = f_0 - q_0 \in \mathcal{H}_{\mathrm{L},\boldsymbol{\phi}}, \quad q_0 = \mathrm{Proj}_{\mathcal{N}_\mathrm{L}}\{f_0\} \in \mathcal{N}_\mathrm{L}$

subject to $\boldsymbol{\phi}(\tilde{f}_0) = \mathbf{0} \ \Rightarrow \ \boldsymbol{\phi}(q_0) = \mathbf{c}_0, \quad$ and $\quad \boldsymbol{\nu}(\tilde{f}_0) = \mathbf{y}_0 - \boldsymbol{\nu}(q_0)$

$\Rightarrow \quad q_0 = \sum_{n=1}^{N_0} c_n p_n \quad$ and $\quad \tilde{f}_0 = \arg\min_{\tilde{f}\in\mathcal{H}_{\mathrm{L},\boldsymbol{\phi}}} \|\tilde{f}\|^2_{\mathcal{H}_{\mathrm{L},\boldsymbol{\phi}}}$ s.t. $\boldsymbol{\nu}(\tilde{f}) = \mathbf{y}_0 - \boldsymbol{\nu}(q_0)$

$\Rightarrow \quad f_0 = \sum_{m=1}^{M} a_m \mathrm{A}_{\boldsymbol{\phi}}\{\nu_m\} + \sum_{n=1}^{N_0} c_n p_n$ (by Abstract representer theorem)

- Orthogonality property: $\tilde{f}^*_0 = (\mathrm{L}^*\mathrm{L})\{\tilde{f}_0\} = \sum_{m=1}^{M} a_m \nu_m \in \mathcal{H}'_{\mathrm{L},\boldsymbol{\phi}} \perp \mathcal{N}_\mathrm{L}$

$$f_0 = \mathrm{A}_{\boldsymbol{\phi}}\left\{\sum_{m=1}^{M} a_m\nu_m\right\} + \sum_{n=1}^{N_0} c_n p_n = \mathrm{A}\left\{\sum_{m=1}^{M} a_m\nu_m\right\} + \sum_{n=1}^{N_0} b_n p_n$$

(using relation between $\mathrm{A}_{\boldsymbol{\phi}}$ and A for $f^* \in \mathcal{H}'_{\mathrm{L},\boldsymbol{\phi}}$)

3.3 Discretization and numerical solution

- Finite-dimensional solution space

$$\mathcal{V}_{\mathrm{L},\boldsymbol{\nu}} = \{f = \boldsymbol{\varphi}^T\mathbf{a} + \boldsymbol{p}^T\mathbf{b} : \mathbf{a} \in \mathbb{R}^M, \mathbf{b} \in \mathbb{R}^{N_0}\} \subseteq \mathcal{H}_{\mathrm{L}}$$

where $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_M)$ with $\varphi_n = \mathrm{A}\{\nu_n\}$ and $\boldsymbol{p} = (p_1, \ldots, p_{N_0})$

$$\Rightarrow \quad J(f|\mathbf{y}, \lambda) = J(\mathbf{a}, \mathbf{b}|\mathbf{y}, \lambda) = F(\mathbf{G}\mathbf{a} + \mathbf{P}\mathbf{b}, \mathbf{y}) + \lambda \mathbf{a}^T \mathbf{G}\mathbf{a}$$

- Exact discretization

System/**Gram** matrix $\mathbf{G} \in \mathbb{R}^{M \times M}$: $[\mathbf{G}]_{m,n} = \langle \nu_m, \varphi_n \rangle = \langle \nu_m, \mathrm{A}\{\nu_n\} \rangle = \langle \varphi_m, \varphi_n \rangle_{\mathcal{H}}$

$$= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \nu_m(\boldsymbol{x}) G_{\mathrm{L}^*\mathrm{L}}(\boldsymbol{x}, \boldsymbol{y}) \nu_n(\boldsymbol{y}) \mathrm{d}\boldsymbol{x} \mathrm{d}\boldsymbol{y}$$

Additional **system** matrix $\mathbf{P} \in \mathbb{R}^{M \times N_0}$: $[\mathbf{P}]_{m,n} = \langle \nu_m, p_n \rangle = \int_{\mathbb{R}^d} \nu_m(\boldsymbol{x}) p_n(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}$

Discrete optimization problem

$$\arg \min_{\mathbf{a} \in \mathbb{R}^M, \, \mathbf{b} \in \mathbb{R}^{N_0}} \left\{ F(\mathbf{G}\mathbf{a} + \mathbf{P}\mathbf{b}, \mathbf{y}) + \lambda \, \mathbf{a}^T \mathbf{G}\mathbf{a} \right\}$$

- Steepest descent algorithm

Initialization: $(\mathbf{a}_0, \mathbf{b}_0)$

$$\mathbf{a}_{k+1} = \mathbf{a}_k - \tau \mathbf{G} \left(\nabla F(\mathbf{G}\mathbf{a}_k + \mathbf{P}\mathbf{b}_k, \mathbf{y}) + 2\lambda \mathbf{a}_k \right)$$
$$\mathbf{b}_{k+1} = \mathbf{b}_k - \tau \mathbf{P}^T \nabla F(\mathbf{G}\mathbf{a}_k + \mathbf{P}\mathbf{b}_k, \mathbf{y})$$

Until stop criterion

Least-squares approximation problem

$$J(\mathbf{a},\mathbf{b}|\mathbf{y},\lambda) = \|\mathbf{G}\mathbf{a}+\mathbf{P}\mathbf{b}-\mathbf{y}\|_2^2 + \lambda\,\mathbf{a}^T\mathbf{G}\mathbf{a}$$

$$\frac{\partial J(\mathbf{a},\mathbf{b}|\mathbf{y},\lambda)}{\partial \mathbf{a}} = 2\mathbf{G}\Big((\mathbf{G}+\lambda\mathbf{I})\mathbf{a}+\mathbf{P}\mathbf{b}-\mathbf{y}\Big) = \mathbf{0}$$

$$\frac{\partial J(\mathbf{a},\mathbf{b}|\mathbf{y},\lambda)}{\partial \mathbf{b}} = 2\mathbf{P}^T(\mathbf{G}\mathbf{a}+\mathbf{P}\mathbf{b}-\mathbf{y}) = 0$$

$$\Rightarrow \quad \begin{pmatrix}\mathbf{a}\\ \mathbf{b}\end{pmatrix} = \begin{pmatrix}(\mathbf{G}+\lambda\mathbf{I}) & \mathbf{P}\\ \mathbf{P}^T & \mathbf{0}\end{pmatrix}^{-1}\begin{pmatrix}\mathbf{y}\\ \mathbf{0}\end{pmatrix}$$

- Extreme case: $\lambda \to \infty$

$$\mathbf{a}_{(\infty)} = \mathbf{0} \quad \text{and} \quad \mathbf{b}_{(\infty)} = (\mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T\mathbf{y} \qquad = \arg\min_{\mathbf{b}} \|\mathbf{P}\mathbf{b}-\mathbf{y}\|_2^2$$

CONTENT

- **1. Preliminaries** ✔
- **2. Reproducing kernel Hilbert spaces** ✔
- **3. Variational splines and representer theorems** ✔
- **4. Gaussian processes**
  - Generalized stochastic processes (GSP)
  - Mean and covariance forms
  - The characteristic functional
  - Characterization of Gaussian processes
  - MMSE solution of inverse problems