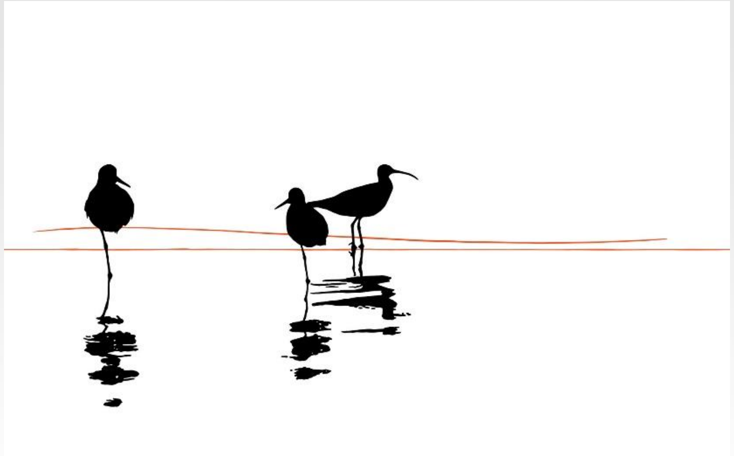


Functional Analysis through Applications: Reproducing Kernel Hilbert Spaces

Hagen Knaf, Hochschule RheinMain

What is it all about?

- For many problems / tasks in (applied) mathematics good linear methods are available.
- »Linear« essentially means »using Linear Algebra«.
- However applied mathematicians are often forced to solve nonlinear problems.
- Instead of designing new methods for the solution one can
  - try to »transform« the problem into a linear one,
  - solve the transformed problem,
  - and take the »inverse transform« as a solution of the original problem.
- In the next two lectures one way to bring this very vague approach into an applicable form is presented.

What is it all about?

$$\begin{aligned} x_1^2 x_2^3 x_3^4 &= 2 \\ x_1^4 x_2^4 x_3^{-1} &= 1 \\ x_1^3 x_2^5 x_3^2 &= 4 \end{aligned}$$

$$\begin{aligned} 2u_1 + 3u_2 + 4u_3 &= \log(2) \\ 4u_1 + 4u_2 - u_3 &= 0 \\ 3u_1 + 5u_2 + 2u_3 &= \log(4) \end{aligned}$$

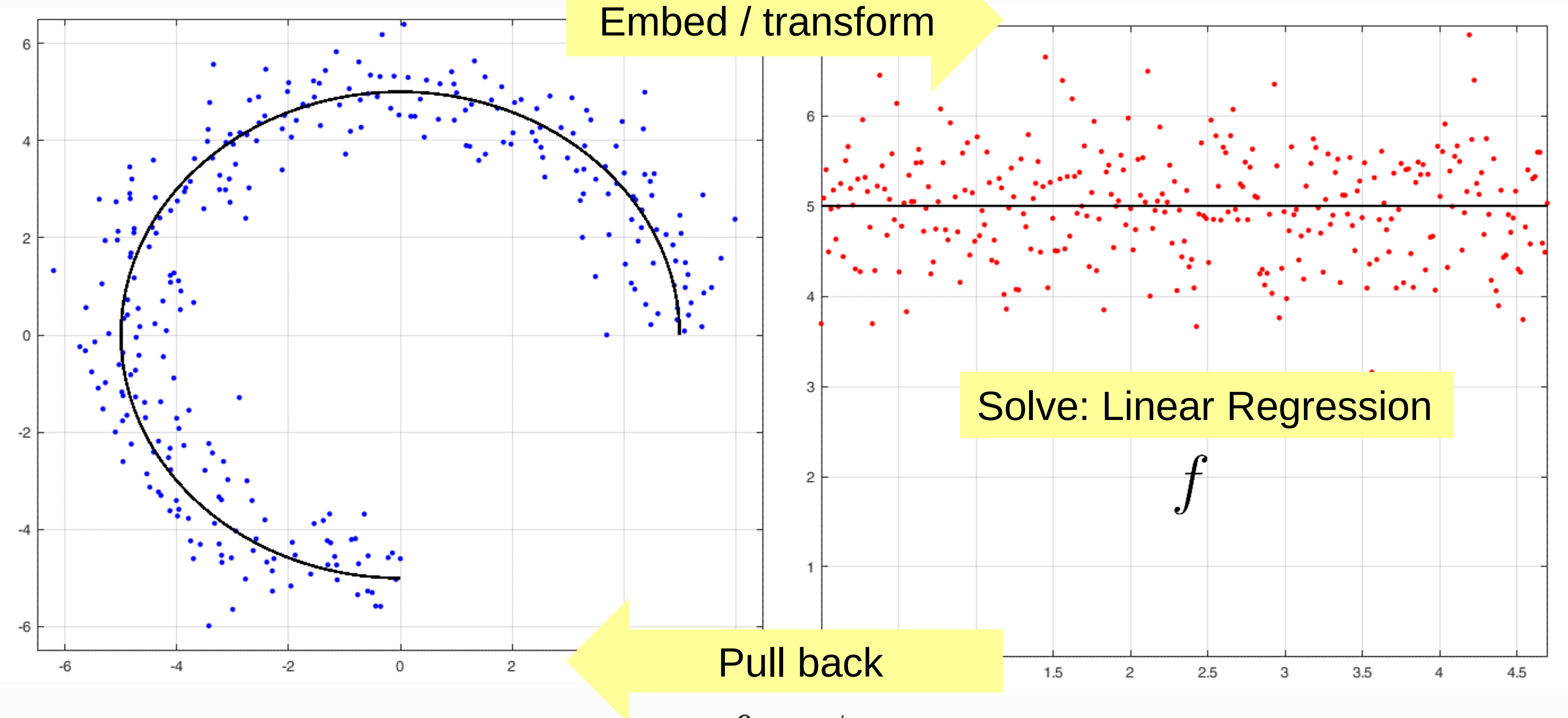
$$\begin{aligned} \phi : (\mathbb{R}^{>0})^3 &\rightarrow \mathbb{R}^3 \\ (x_1, x_2, x_3) &\mapsto (\log(x_1), \log(x_2), \log(x_3)) \end{aligned}$$

$$\begin{aligned} u_1 &= -\log(2) \\ u_2 &= \log(2) \\ u_3 &= 0 \end{aligned}$$

$$\begin{aligned} x_1 &= \tfrac{1}{2} \\ x_2 &= 2 \\ x_3 &= 1 \end{aligned}$$

What is it all about?

$$\phi : X \to H$$



$$f \circ \phi$$

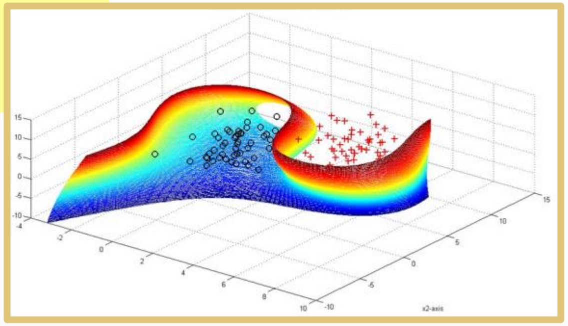
What is it all about?

Tasks appearing in the transform-solve-pull back approach:

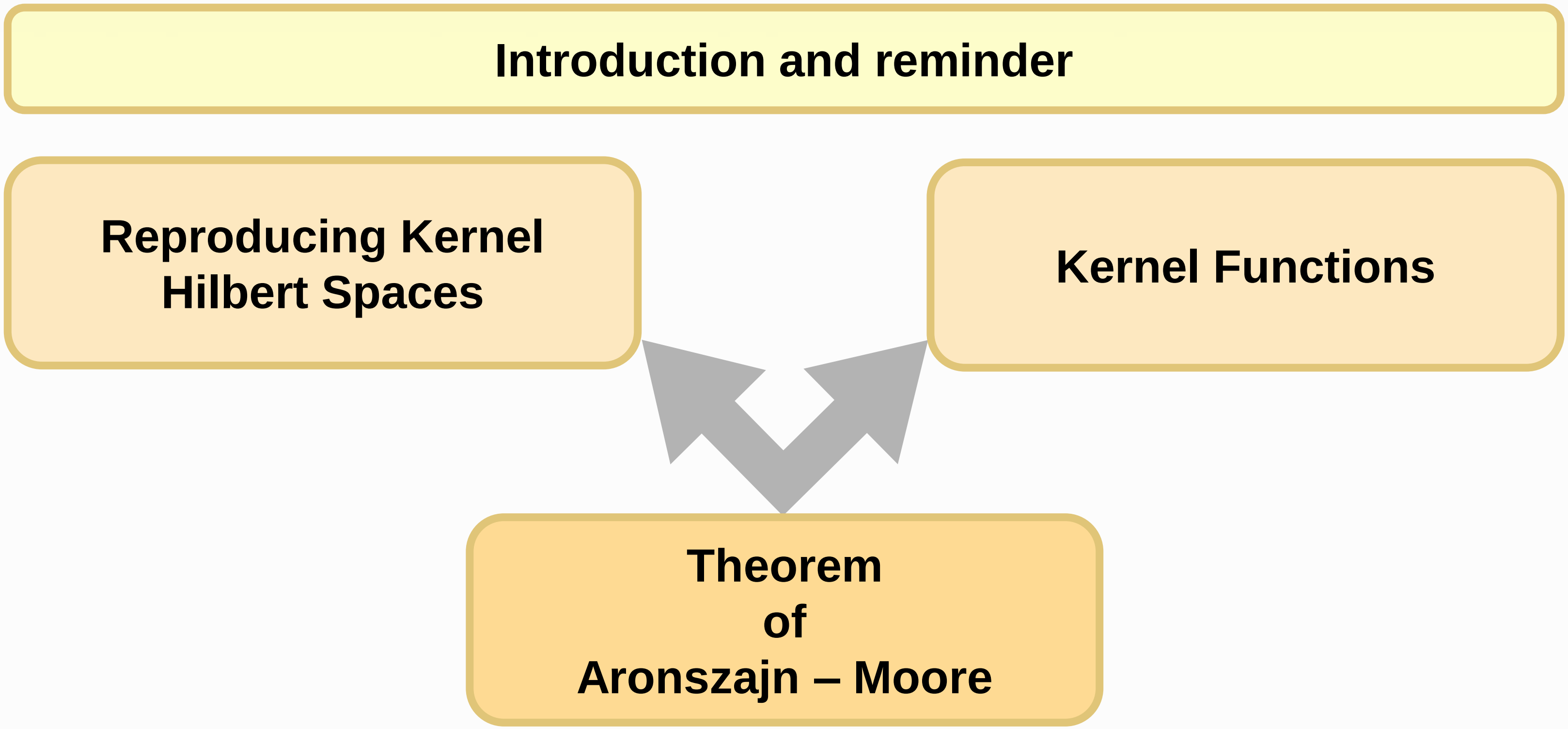
- Find a good transform mapping by checking the »quality« of many candidates.
- Effectively parameterize sets / families of candidate mappings.
- Determine sets / families of candidate mappings, such that it is not too difficult / slow to perform computations in the transformed space H.

Outline of the two lectures

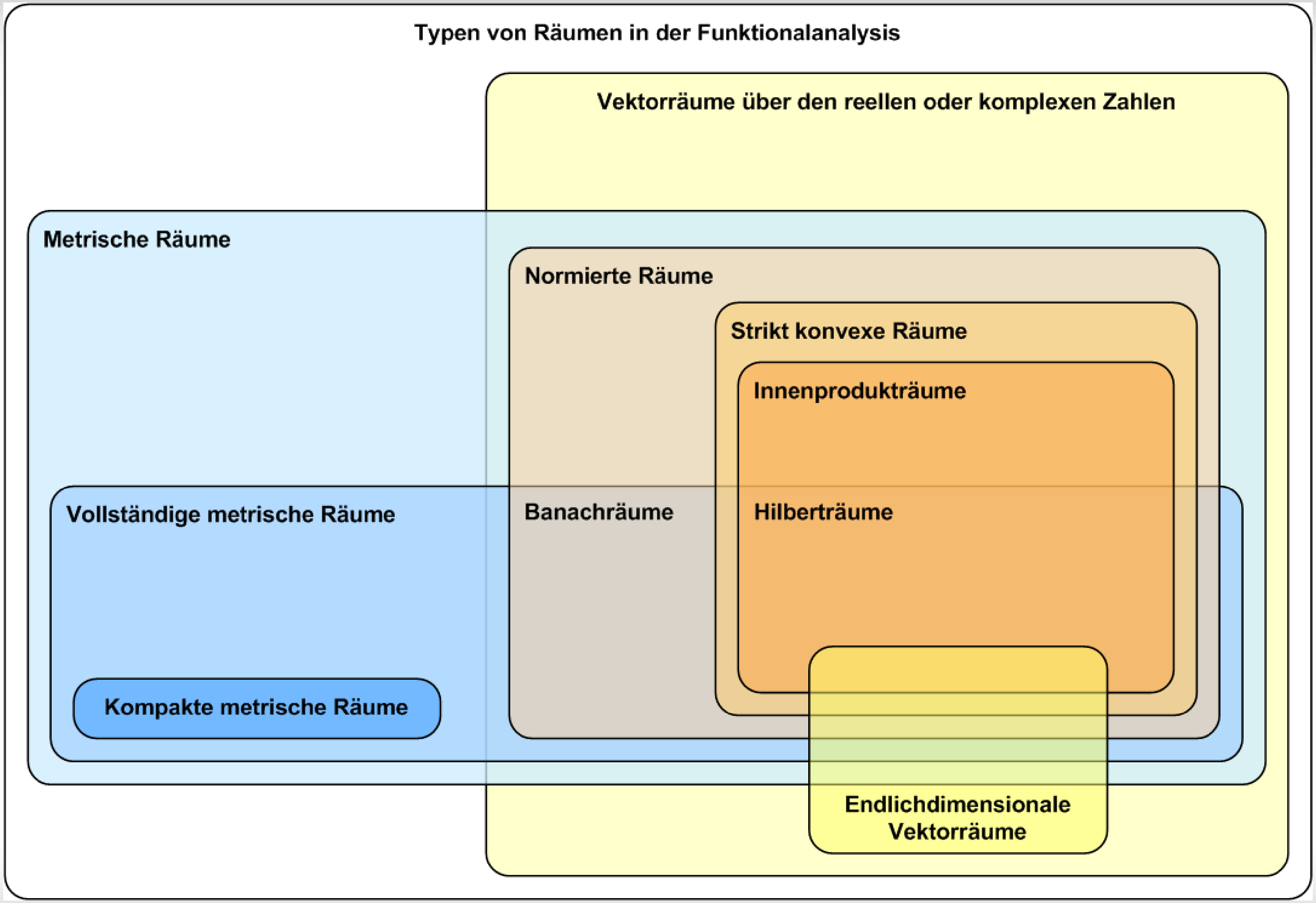
1. Introduction and reminder
2. Reproducing Kernel Hilbert Spaces
3. Kernel Functions
4. The Theorem of Aronszajn – Moore
5. What is Data Mining?
6. An Overview of Discriminant Analysis
7. Kernel – Fisher – Discriminant Analysis
8. The Kernel Method in general



Outline of the first lecture



Introduction and reminder



Vector spaces of functions

- For a set $X \neq \emptyset$ consider the *functions on* X:

$$\mathrm{Fun}(X, \mathbb{R}) := \{f : X \to \mathbb{R} : f \text{ arbitrary}\}.$$

- Pointwise addition and scalar multiplication turns $\mathrm{Fun}(X, \mathbb{R})$ into a real vector space:

$$(f + g)(x) := f(x) + g(x) \text{ for } f, g \in \mathrm{Fun}(X, \mathbb{R}),$$

$$(\lambda f)(x) := \lambda f(x)) \text{ for } f \in \mathrm{Fun}(X, \mathbb{R}),\ \lambda \in \mathbb{R}.$$

- The dimension of $\mathrm{Fun}(X, \mathbb{R})$ is finite if and only if X is a finite set.

REMARK: Virtually everything explained in this talk can be done over the complex numbers as well.

Vector spaces of functions

Frequently we don't want to consider *all* functions on a set X:

- Let P be a property of functions such that: if f, g possess property P, then $f+g$ and λf possess property P as well. Then

$$P(X,\mathbb{R}) := \{f : X \to \mathbb{R} : f \text{ posesses property } P\}$$

  is a vector subspace of $\mathrm{Fun}(X,\mathbb{R})$.

- Examples of such properties P:
  - continuity (X a metric space),
  - (partial) differentiability ($X \subseteq \mathbb{R}^n$ open),
  - integrability (X a measurable set),
  - analyticity ($X \subseteq \mathbb{R}$ open).

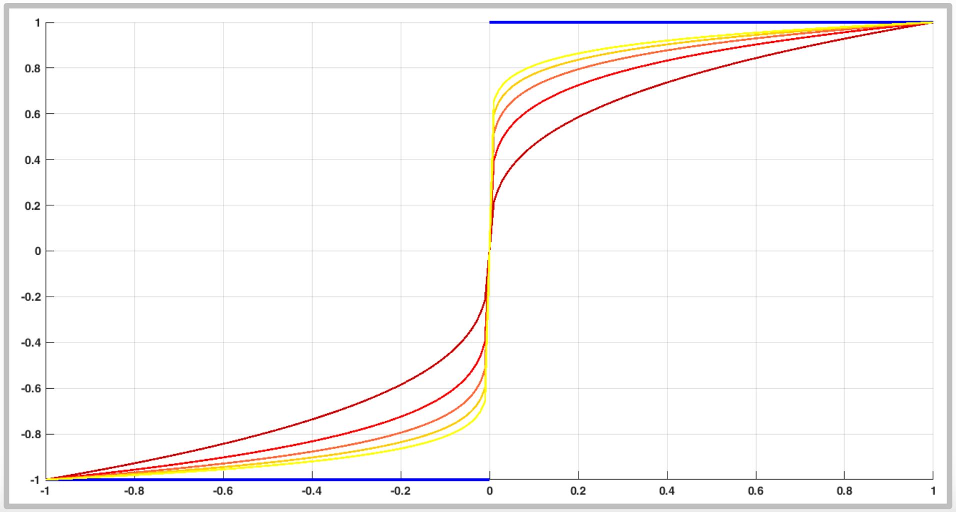
Pointwise convergence

- A sequence $(f_i)_{i\in\mathbb{N}}$ in $\mathrm{Fun}(X,\mathbb{R})$ is said to be *pointwise convergent* if the limit
$$\lim_{i\to\infty} f_i(x)$$
exists for every $x \in X$. The function
$$f(x) := \lim_{i\to\infty} f_i(x)$$
is then called the *pointwise limit* of $(f_i)_{i\in\mathbb{N}}$.
- If V is a vector subspace of $\mathrm{Fun}(X,\mathbb{R})$ and f is the pointwise limit of the sequence $(f_i)_{i\in\mathbb{N}}$, where $f_i \in V$ for all i, then f needs not be an element of V.

Pointwise convergence

EXAMPLE:

- $V = C([a,b],\mathbb{R}) := \{f : [a,b] \to \mathbb{R} : \ f \text{ is continous}\}$,
- $f_i := \sqrt[2i+1]{t}$, $i = 1, 2, 3, \ldots$,
- $f(x) = 1$ for $x > 0$, $f(0) = 0$, $f(x) = -1$ for $x < 0$.



Inner Product Spaces and Hilbert Spaces

DEFINITION: *An inner product space is a vector space H over the reals $\mathbb{R}$ equipped with a symmetric, positive definite, bilinear form*

$$H \times H \to \mathbb{R}^{\geq 0},\ (x, y) \mapsto \langle x, y \rangle$$

called scalar product.

INEQUALITY OF CAUCHY-SCHWARZ: *Every positive semidefinite, bilinear form $\langle \cdot, \cdot \rangle$ on a real vector space H has the property*

$$\forall x, y \in H \quad \langle x, y \rangle^2 \leq \langle x, x \rangle \langle y, y \rangle$$

In an inner product space $(H, \langle \cdot, \cdot \rangle)$ equality holds if and only if x, y are linearly dependent.

Inner Product Spaces and Hilbert Spaces

- The scalar product gives rise to the *norm* $\|x\| := \sqrt{\langle x, x\rangle}$ and hence to the *metric* $d(x,y) := \|x - y\|$.
- The scalar product, the norm and the metric are continuous functions.
- Addition and scalar multiplication are continuous maps

$$+ : H \times H \to H, \quad s : \mathbb{R} \times H \to H;$$

  here $H \times H$ and $\mathbb{R} \times H$ are equipped with the relevant product metrics.
- In an inner product space H the notion of *orthogonality* of elements $x, y \in H$ is defined:

$$x \perp y \;:\Leftrightarrow\; \langle x, y\rangle = 0.$$

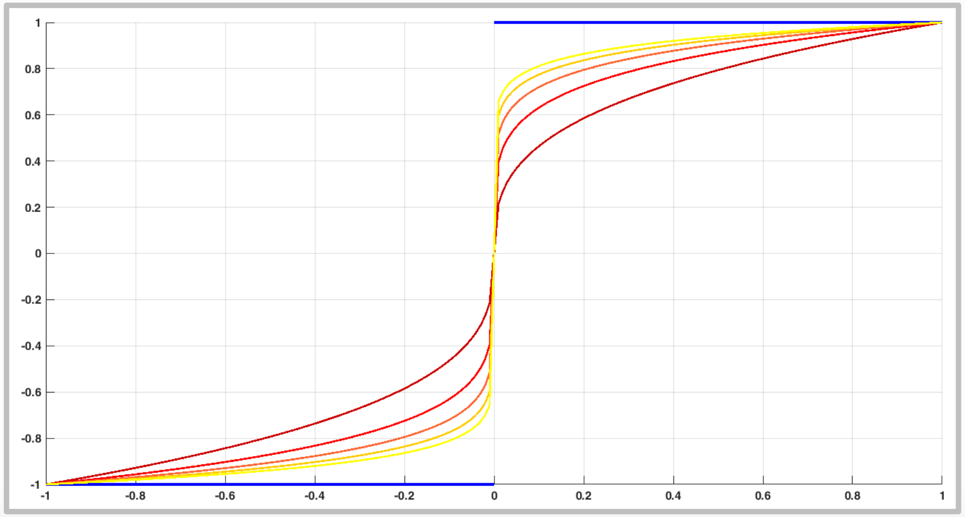
Inner Product Spaces and Hilbert Spaces

EXAMPLE:

- $C([a,b],\mathbb{R}) := \{f : [a,b] \to \mathbb{R} : \ f \text{ is continuous}\}$,
- $\langle f, g\rangle := \int\limits_a^b f g \, dt$,
- In $C([-1,1],\mathbb{R})$ the sequence

  $$f_k := \sqrt[2k+1]{t}, \; k = 1,2,3,\ldots,$$

  is Cauchy but not convergent.

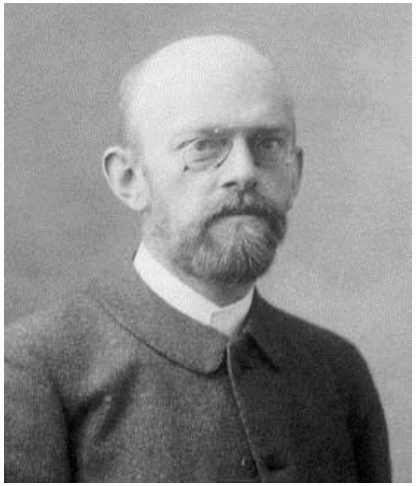


Inner Product Spaces and Hilbert Spaces

DEFINITION: *A Hilbert space is an inner product space* $(H, \langle\cdot,\cdot\rangle)$ *that is complete with respect to the norm* $\|x\| := \sqrt{\langle x, x\rangle}$*: every Cauchy sequence* $(x_k)_{k\in\mathbb{N}}$ *in* H *has a limit.*

EXAMPLE:

- $\ell^2 := \{f : \mathbb{N} \to \mathbb{R} : \sum_{k=1}^{\infty} f(k)^2 \text{ existiert}\}$,
- $\langle f, g\rangle := \sum_{k=1}^{\infty} f(k)g(k)$,
- $(\frac{1}{k})_{k\in\mathbb{N}} \in \ell^2$ because $\sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi}{6}$.



D. Hilbert
1862 – 1943

Inner Product Spaces and Hilbert Spaces

EXAMPLE:

- $L^2([a,b],\mathbb{R}) := \{\overline{f} : f^2 : [a,b] \to \mathbb{R}$ Lebesgue-integrable$\}$, with

$$\overline{f} := \{g : [a,b] \to \mathbb{R}) : f, g \text{ coincide on a set of measure } 0\}.$$

- $\langle \overline{f}, \overline{g} \rangle := \int_a^b f g \, dt.$

- Note that the map

$$C([a,b],\mathbb{R}) \to L^2([a,b],\mathbb{R}), \;\; f \mapsto \overline{f}$$

is linear, injective and continuous.

Completion

THEOREM: *For every inner product space* $(H, \langle\cdot,\cdot\rangle)$ *there exists a Hilbert space* $(\widehat{H}, [\cdot,\cdot])$ *possessing the properties*

- H *is a dense vector subspace of* $\widehat{H}$, *that is every element of* $\widehat{H}$ *is the limit of a Cauchy sequence in* H.
- *The scalar product* $[\cdot,\cdot]$ *is an extension of the scalar product* $\langle\cdot,\cdot\rangle$.

$(\widehat{H}, [\cdot,\cdot])$ *is called the completion of* $(H, \langle\cdot,\cdot\rangle)$; *it is uniquely determined by* $(H, \langle\cdot,\cdot\rangle)$.

EXAMPLE: The completion of $C([a,b],\mathbb{R})$ is $L^2([a,b],\mathbb{R})$.

Representation Theorem of Riesz: *Let* $(H, \langle \cdot, \cdot \rangle)$ *be a Hilbert space. Then every continuous linear mapping* $T : H \to \mathbb{R}$ *has the form*

$$T(x) = \langle x, v \rangle$$

for some $v \in H$ *uniquely determined by* T.

Remark: For $v \in H$ such that $\|v\| = 1$ the mapping

$$p(x) = \langle x, v \rangle v$$

is the orthogonal projection onto the line $\mathbb{R}v$.

Reproducing Kernel Hilbert Spaces

DEFINITION: *Let $X \neq \emptyset$ be a set.*
A reproducing kernel Hilbert space (RKHS) on X is a Hilbert space $(H, \langle\cdot,\cdot\rangle)$ possessing the properties

1. *H ist a vector subspace of $\mathrm{Fun}(X, \mathbb{R})$.*
2. *For every $x \in X$ the evaluation functional $e_x : H \to \mathbb{R},\ h \mapsto h(x)$ is continuous.*

- In a RKHS addition and scalar multiplication are the pointwise operations of functions on X.
- Convergence in H implies pointwise convergence:

$$\forall x \in X \quad (\lim_{k\to\infty} h_k)(x) = \lim_{k\to\infty} h_k(x).$$

Reproducing Kernel Hilbert Spaces

EXAMPLE:

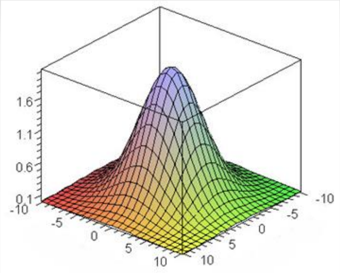
- $\ell^2 := \{f : \mathbb{N} \to \mathbb{R} : \sum_{k=1}^{\infty} f(i)^2 \text{ exists}\}$ is a vector subspace of $\mathrm{Fun}(\mathbb{N}, \mathbb{R})$.
- For every $n \in \mathbb{N}$ the evaluation functional $e_n(f) := f(n)$ is continuous:

$$\|f\| = \sqrt{\sum_{i=1}^{\infty} f(i)^2} \geq \sqrt{f(n)^2}.$$

hence

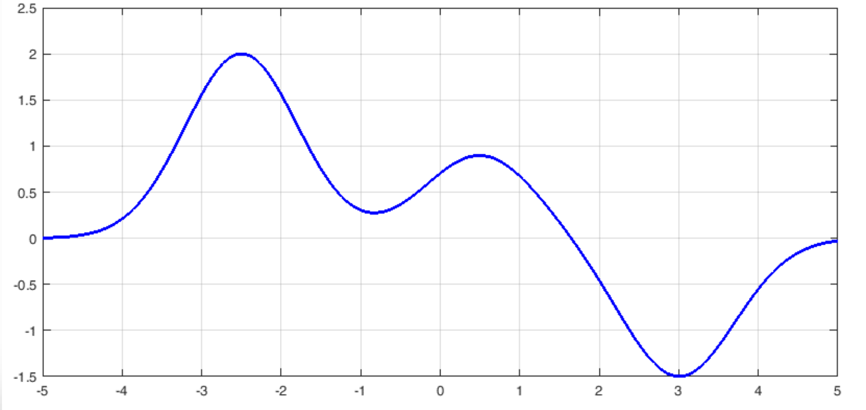
$$\|e_n\|_{\mathrm{op}} := \sup\left(\frac{|e_n(f)|}{\|f\|} : f \in \ell^2 \setminus \{0\}\right) = 1.$$

Reproducing Kernel Hilbert Spaces



EXAMPLE: Gaussian RKHS

- Let $X = \{x_1, \ldots, x_n\} \subset \mathbb{R}^m$ be a set with n elements.
- $H := \sum_{i=1}^{n} \mathbb{R} k_i$, $k_i(x) := e^{-\frac{\|x - x_i\|_2^2}{h^2}}$, $h > 0$, $\|\cdot\|_2$ the Euclidean norm on $\mathbb{R}^m$.
- $\{k_1, \ldots, k_n\}$ forms a basis of H.
- Scalar product: bilinear extension of $\langle k_i, k_j \rangle := e^{-\frac{\|x_i - x_j\|_2^2}{h^2}}$.



An element of H in the case $m = 1$:

$$f(x) = 2e^{-(x+2.5)^2} + 0.9e^{-(x-0.5)^2} - 1.5e^{-(x-3)^2}$$

Reproducing Kernel Hilbert Spaces

EXAMPLE: Gaussian RKHS

- $\langle \sum_{i=1}^{n} \lambda_i k_i, \sum_{j=1}^{n} \lambda_j k_j \rangle \leq n^2 \max(|\lambda_i| : i = 1, \ldots, n)^2$.
- The linear map $T : \mathbb{R}^n \to H$, $(\lambda_1, \ldots, \lambda_n) \mapsto \sum_{i=1}^{n} \lambda_i k_i$ is continuous, using the 1-norm on $\mathbb{R}^n$: $\|T\|_{\text{op}} \leq n$.
- Therefore the minimum $\mu := \min(\| \sum_{i=1}^{n} \lambda_i k_i \| : \sum_{i=1}^{n} |\lambda_i| = 1)$ exists.
- The inequality $\sum_{i=1}^{n} |\lambda_i| \leq \mu^{-1} \| \sum_{i=1}^{n} \lambda_i k_i \|$ yields the completeness of H.
- The evaluation functionals e_x are continuous: $\|e_x\|_{\text{op}} \leq \frac{1}{n}$.

Reproducing Kernel Hilbert Spaces

Intermezzo: The last results are not specific to the Gaussian RKHS.

THEOREM: *A normed space $(H, \|\cdot\|)$ of finite dimension is complete.*
More specific: for every basis $(b_1, \ldots, b_n)$ of H convergence of a sequence

$$(h_i)_{i\in\mathbb{N}} = (\sum_{j=1}^{n} \lambda_{ij} b_i)_{i\in\mathbb{N}}$$

is equivalent to the convergence of the sequences $(\lambda_{ij})_{i\in\mathbb{N}}$ of coefficients.

Moreover every linear map $T : H \to Y$ into an arbitrary normed space $(Y, \|\cdot\|_Y)$ is continuous.

REMARK / HINT: To prove the theorem just rewrite the arguments given for the Gaussian RKHS in general form using the continuity of the norm function, of addition and of scalar multiplication. (recommendable exercise).

Reproducing Kernel Hilbert Spaces

An almost non-example:

- $C([0,1],\mathbb{R}) := \{f : [0,1] \to \mathbb{R} : \ f \text{ is continuous}\}$, $\langle f, g\rangle := \int_0^1 f g \, dt$.
- The evaluation functional e_1 ist not continuous: for $f_i := \sqrt{2i+1}x^i$ we have $\|f_i\| = 1$ and $e_1(f_i) = \sqrt{2i+1}$, thus $\|e_1\|_{\text{op}} = \infty$.
- Of course $C([0,1],\mathbb{R})$ is not complete, hence no Hilbert space ... grumpf.
- Proper non-examples of RKHS are hard to write down explicitely.

Reproducing Kernel Hilbert Spaces

THEOREM: *Let H be a RKHS on X.*
Then there exists a unique function $K : X \times X \to \mathbb{R}$ such that

- $\forall y \in X \quad k_y := K(\cdot, y) \in H$,
- $\forall x \in X \quad e_x = \langle \cdot, k_x \rangle$.

The function K is called reproducing kernel of H and has the properties:

1. $\forall x, y \in X \quad K(x, y) = K(y, x)$,
2. *For every n-tupel $(x_1, \ldots, x_n) \subseteq X^n$ of elements of X the matrix $K[x_1, \ldots, x_n] := (K(x_i, x_j))_{i,j} \in \mathbb{R}^{n \times n}$ is positive semidefinite:*

$$\forall v \in \mathbb{R}^n \quad v^t K[x_1, \ldots, x_n] v \geq 0.$$

Reproducing Kernel Hilbert Spaces

PROOF:

- By the theorem of Riesz for every $x \in X$ there exists a function k_x such that $h(x) = e_x(h) = \langle h, k_x \rangle$ for all $h \in H$.
- Define $K(x, y) := k_y(x)$.
- For $v = (v_1, \ldots, v_n) \in \mathbb{R}^n$:

$$\begin{aligned} v^t K[x_1, \ldots, x_n] v &= \sum_{i=1}^{n} \sum_{j=1}^{n} v_i K(x_j, x_i) v_j \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} v_i \langle k_i, k_j \rangle v_j \\ &= \langle \sum_{i=1}^{n} v_i k_i, \sum_{j=1}^{n} k_j v_j \rangle \geq 0. \end{aligned}$$

Reproducing Kernel Hilbert Spaces

EXAMPLES:

- The reproducing kernel of ℓ^2 is the function

$$K : \mathbb{N} \times \mathbb{N} \to \mathbb{R}, \ (n,m) \mapsto 0 \text{ for } n \neq m, \ (n,n) \mapsto 1.$$

- The reproducing kernel of the Gaussian RKHS is the function

$$K(x,y) := e^{-\frac{\|x-y\|_2^2}{h^2}}.$$

- For pairwise distinct points $x_1, \ldots, x_n \in \mathbb{R}^m$ the matrices

$$(e^{-\frac{\|x_i-x_j\|_2^2}{h^2}})_{i,j\in\{1,\ldots,n\}}$$

are positive definite.

Reproducing Kernel Hilbert Spaces

Embedding the set X:
Let H be a RKHS on X with reproducing kernel K and consider the map

$$\phi : X \to H, \; y \mapsto k_y = K(\cdot, y).$$

- $\forall x, y \in X \quad \langle \phi(x), \phi(y) \rangle = K(x, y).$
- The map ϕ is injective if and only if for all points $x_1, x_2 \in X$, $x_1 \neq x_2$, there exists a function $h \in H$ such that $h(x_1) \neq h(x_2)$.
- If ϕ is injective the equation

  $$d(x, y) := \|k_x - k_y\| = \sqrt{K(x,x) + K(y,y) - 2K(x,y)}$$

  defines a metric on X.

Reproducing Kernel Hilbert Spaces

EXAMPLE: Gaussian RKHS (continued)

The embedding (injectivity !)

$$\phi : X = \{x_1, \dots, x_n\} \to H,\ x_i \mapsto k_i = e^{-\frac{\|x-x_i\|_2^2}{h^2}}$$

leads to the metric

$$d(x_i, x_j) = \sqrt{2 - 2e^{-\frac{\|x_i-x_j\|_2^2}{h^2}}}.$$

Since $X \subset \mathbb{R}^m$ is arbitrary, the formula actually defines a metric on $\mathbb{R}^m$.
The question arises whether one can define a Gaussian RKHS on $\mathbb{R}^m$.

- Note that the distance geometry of X with respect to the Euclidean distance is different from the one given by d.
- Even if $x_j = \lambda x_i$ the images k_i and k_j are linearly independent; the map ϕ thus is highly nonlinear.

Reproducing Kernel Hilbert Spaces

PROPOSITION: *The functions k_x, $x \in X$, in a RKHS H are linearly independent if and only if the reproducing kernel K of H is positive definite.*
In particular: $\phi : X \to H$ is injective if $K[x_1, x_2]$ is positive definite for all $x_1, x_2 \in X$, $x_1 \neq x_2$.

PROOF:

- The linear relation $\sum_{i=1}^{n} \lambda_{x_i} k_{x_i} = 0$ is equivalent to

$$
\begin{aligned}
0 &= \langle \sum_{i=1}^{n} \lambda_{x_i} k_{x_i}, \sum_{j=1}^{n} \lambda_{x_j} k_{x_j} \rangle \\
&= \sum_{i,j} \lambda_i \lambda_j K(x_j, x_i) \\
&= (\lambda_1, \ldots, \lambda_n) K[x_1, \ldots, x_n] (\lambda_1, \ldots, \lambda_n)^t.
\end{aligned}
$$

- Apply that for $n = 2$.

Reproducing Kernel Hilbert Spaces

Theorem: *For a RKHS H on X the vector space*

$$U := \sum_{x \in X} \mathbb{R} k_x$$

lies dense in H: the closure $\overline{U}$ of U equals H.

In particular: If X is finite, then $H = U$ and $\dim(H) \leq |X|$.

Proof: If $\overline{U} \neq H$ there exists $h \in H$ such that $h \perp u$ for all $u \in \overline{U}$.
In particular $h(x) = \langle h, k_x \rangle = 0$ for all $x \in X$. Hence $h = 0$ – contradiction.

Remarks:

- This result is the reason for the term *reproducing* kernel.
- In general the functions $\{k_x : x \in X\}$ are linearly dependent.
- The upper bound for the dimension is attained in the case of a Gaussian RKHS on a finite set.

Kernel Functions

DEFINITION: *Let $X \neq \emptyset$ be a set.*
A function

$$K : X \times X \to \mathbb{R}$$

is called kernel function on X if it is symmetric, and if for every $n \in \mathbb{N}$ and for every n-tupel $(x_1, \ldots, x_n) \subseteq X^n$ of elements of X the matrix

$$K[x_1, \ldots, x_n] := (K(x_i, x_j))_{i,j} \in \mathbb{R}^{n \times n}$$

is positive semidefinite:

$$\forall v \in \mathbb{R}^n \quad v^t K[x_1, \ldots, x_n] v \geq 0.$$

REMARK: It suffices to check the required positive semidefiniteness for n-tupels of pairwise distinct elements $x_i \in X$.

Kernel Functions

Kernel functions on a finite set X

Let $X = \{x_1, \ldots, x_n\}$, $n \in \mathbb{N}$.
Every kernel function on X can be obtained by chosing a positive semidefinite matrix $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ and defining

$$K(x_i, x_j) := a_{ij}.$$

It is not necessary to consider subsets of X.

Although in applications one usually considers finite sets X only, it is necessary to know kernel functions on infinite sets for the following reason:

- frequently $X \subset \mathbb{R}^m$, the values $K(x_i, x_j)$ should then be related to the components of the $x_i \in \mathbb{R}^m$,
- the position of X in $\mathbb{R}^m$ however is usually to some extend arbitrary.

Kernel Functions

THEOREM:

- *The pointwise sum $K_1 + K_2$ of two kernel functions on X is a kernel function on X.*
- *The pointwise product $K_1 \cdot K_2$ of two kernel functions on X is a kernel function on X.*
- *The pointwise product λK of a kernel function K on X with a non-negative real number λ is a kernel function on X.*

In particular: the set $\mathcal{K}(X)$ of kernel functions on $X \neq \emptyset$ together with pointwise addition and multiplication forms a commutative semiring.

REMARK: Only the proof of the second statement is not straightforward. One has to show that the Schur product $(a_{ij}) \odot (b_{ij}) = (a_{ij}b_{ij})$ of positive semidefinite matrices is positive semidefinite.

Kernel Functions

COROLLARY: *For* $K \in \mathcal{K}(X)$ *and every polynomial* $p(X) = \sum_{i=0}^{n} a_i X^i$ *with non-negative coefficients, the function*

$$p(K) = \sum_{i=0}^{n} a_i K^i$$

is a kernel function on X.

According to the corollary the functions

$$K(x, y) := (\langle x, y \rangle_2 + c)^d, \ c > 0, d \in \mathbb{N}$$

are kernel functions on $X = \mathbb{R}^m$; here $\langle x, y \rangle_2$ denotes the standard scalar product on $\mathbb{R}^m$, which is a kernel function.
They are widely used in Data Mining and called *polynomial kernels of degree* d.

Kernel Functions

PROPOSITION: *Let* $(K_i)_{i\in\mathbb{N}}$ *be a pointwise convergent sequence of kernel functions on* X, *then the (pointwise) limit*

$$K(x,y) := \lim_{i\to\infty} K_i(x,y)$$

is a kernel function on X.

COROLLARY: *Let* $K \in \mathcal{K}(X)$ *be a kernel function with values in the set* $U \subseteq \mathbb{R}$. *Let* $f : U \to \mathbb{R}$ *be a function defined by a power series with non-negative coefficients:* $f(u) = \sum_{i=0}^{\infty} a_i u^i$. *Then the pointwise limit*

$$f(K) := \sum_{i=0}^{\infty} a_i K^i.$$

is a kernel function on X.

Kernel Functions

An application of the last corollary yields another kernel function important in Data Mining: for every $h \in \mathbb{R}$ the function

$$K(x,y) := e^{-\frac{\|x-y\|^2}{h^2}}$$

is a kernel function on $X = \mathbb{R}^m$ called the *Gauß kernel of bandwidth h*.

In the proof the following auxiliary result is used:

PROPOSITION: *For every kernel function $K \in \mathcal{K}(X)$ and every function $f : X \to \mathbb{R}$ the function*

$$K_f(x,y) := f(x)K(x,y)f(y)$$

is a kernel function.

Kernel Functions

PROOF:

$$(\lambda_1, \ldots, \lambda_n) K_f[x_1, \ldots, x_n](\lambda_1, \ldots, \lambda_n)^t = \sum_{i,j} \lambda_i f(x_i) K(x_j, x_i) f(x_j) \lambda_j$$

$$= \quad (f(x_1)\lambda_1, \ldots, f(x_n)\lambda_n) K[x_1, \ldots, x_n](f(x_1)\lambda_1, \ldots, f(x_n)\lambda_n)^t.$$

PROOF:

- $e^{-\frac{\|x-y\|^2}{h^2}} = e^{-\frac{\langle x,x \rangle}{h^2}} e^{2\frac{\langle x,y \rangle}{h^2}} e^{-\frac{\langle y,y \rangle}{h^2}}.$
- e^x can be expressed as a power series with positive coefficients,
- hence $e^{2\frac{\langle x,y \rangle}{h^2}}$ is a kernel function.
- Use the proposition to get the result.

Kernel Functions

Ordering the set of kernel functions

- The pointwise difference $K_1 - K_2$ of kernel functions on X in general is not a kernel function.
- Partial ordering on $\mathcal{K}(X)$: $K_1 \leq K_2 \;:\Leftrightarrow\; K_2 - K_1 \in \mathcal{K}(X)$.
  - $\forall K \in \mathcal{K}(X) \quad K \leq K$.
  - $\forall K_1, K_2 \in \mathcal{K}(X) \quad K_1 \leq K_2 \wedge K_2 \leq K_1 \;\Rightarrow\; K_1 = K_2$.
  - $\forall K_1, K_2, K_3 \in \mathcal{K}(X) \quad K_1 \leq K_2 \wedge K_2 \leq K_3 \;\Rightarrow\; K_1 \leq K_3$.
- The ordering is compatible with the algebraic operations:
  - $\forall K, K_1, K_2 \in \mathcal{K}(X) \quad K_1 \leq K_2 \;\Rightarrow\; K_1 + K \leq K_2 + K$.
  - $\forall K, K_1, K_2 \in \mathcal{K}(X) \quad K_1 \leq K_2 \;\Rightarrow\; K_1 \cdot K \leq K_2 \cdot K$.

The Theorem of Aronszajn – Moore

THEOREM (E.H.MOORE, N.ARONSZAJN (1935/1950)): *Let* $X \neq \emptyset$ *be a set.*

For every kernel function K *on* X *there exists a unique reproducing kernel Hilbert space* $H(K)$ *on* X *having* K *as its reproducing kernel.*

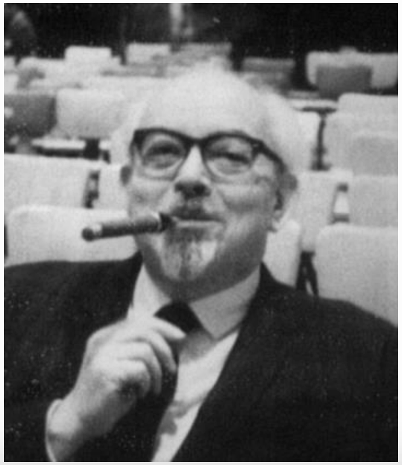
Let $\mathcal{H}(X)$ *be the set of RKHS on* X. *The map*

$$\mathcal{K}(X) \to \mathcal{H}(X), \; K \mapsto H(K)$$

is bijective.

E. H. Moore
(1862 – 1932)

N. Aronszajn
(1907 – 1980)



The Theorem of Aronszajn – Moore

The main steps of the proof of the existence of $H(K)$.

- Consider the vector space $U := \sum_{x \in X} \mathbb{R} k_x$, $k_x := K(\cdot, x)$, and define a bilinear form via

$$\langle \sum_{i=1}^{n} \lambda_i k_i, \sum_{j=1}^{n} \mu_j k_j \rangle := \sum_{i,j}^{n} \lambda_i \mu_j K(x_i, x_j).$$

- Show that this bilinear form is positive definite.
- Show that the completion H of $(U, \langle \cdot, \cdot \rangle)$ is a RKHS.
- Show that the reproducing kernel of H equals K.

The Theorem of Aronszajn – Moore

EXAMPLE: Gaussian RKHS on $\mathbb{R}^m$

According to the proof of the theorem of Aronszajn-Moore one can contruct the Gaussian RKHS on $\mathbb{R}^m$ by taking the completion H of the vector space

$$U := \sum_{y \in \mathbb{R}^m} \mathbb{R} e^{-\frac{\|x-y\|_2^2}{h^2}}$$

with respect to the inner product

$$\langle \sum_{i=1}^{n} \lambda_i e^{-\frac{\|x-y_i\|_2^2}{h^2}}, \sum_{j=1}^{n} \mu_j e^{-\frac{\|x-y_j'\|_2^2}{h^2}} \rangle := \sum_{i,j}^{n} \lambda_i \mu_j e^{-\frac{\|y_i-y_j'\|_2^2}{h^2}}.$$

The first concrete description of the functions $f \in H$ thus obtained seems to have been given as recently as in the year 2006.

The Theorem of Aronszajn – Moore

EXAMPLE: Polynomial RKHS on $X \subseteq \mathbb{R}^m$

- Consider $K(x,y) := (\langle x,y\rangle_2 + c)^d$, $c > 0$, $d \in \mathbb{N}$, $x = (x_1, \ldots, x_m), y = (y_1, \ldots, y_m) \in \mathbb{R}^m$.

- The functions $k_y = K(\cdot, y)$ are linear combinations of monomial functions $m(x) = x_1^{e_1} \cdot \ldots \cdot x_m^{e_m}$ of degree $\le d$, therefore

$$\dim(H(K)) \le \sum_{e=0}^{d} \binom{m+e-1}{e} = \sum_{e=0}^{d} \frac{(m+e)!}{(m-1)!\, e!}.$$

- In the case $c = 0$: $\dim(H(K)) \le \binom{m+e-1}{e}$ with equality for $X = \mathbb{R}^m$.

- If X contains a nonempty open subset of $\mathbb{R}^m$, then $\phi : X \to H(K)$ is injective.

The Theorem of Aronszajn – Moore

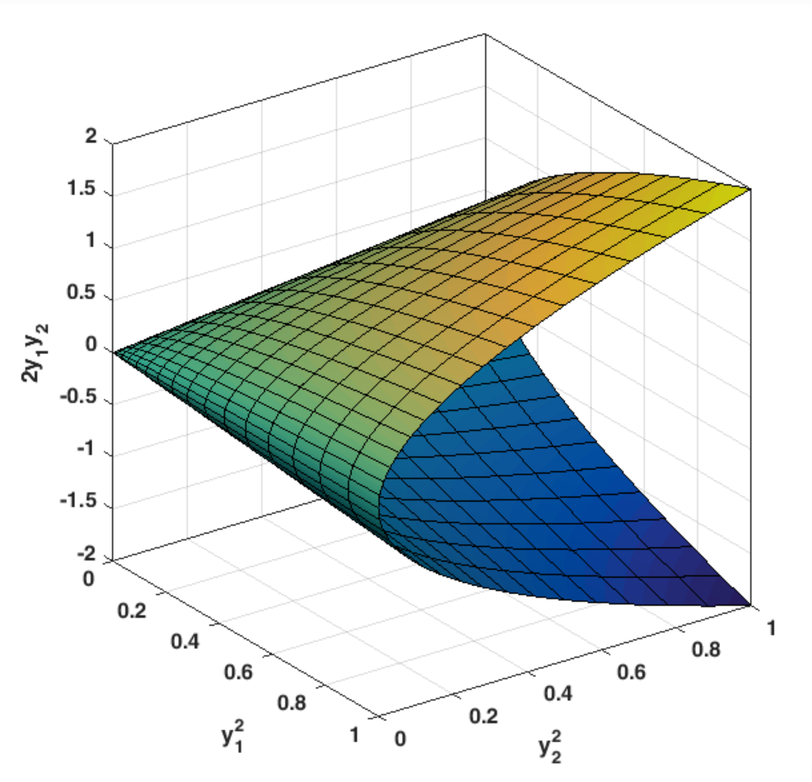
EXAMPLE: Polynomial RKHS of degree $d = 2$ on $\mathbb{R}^2$

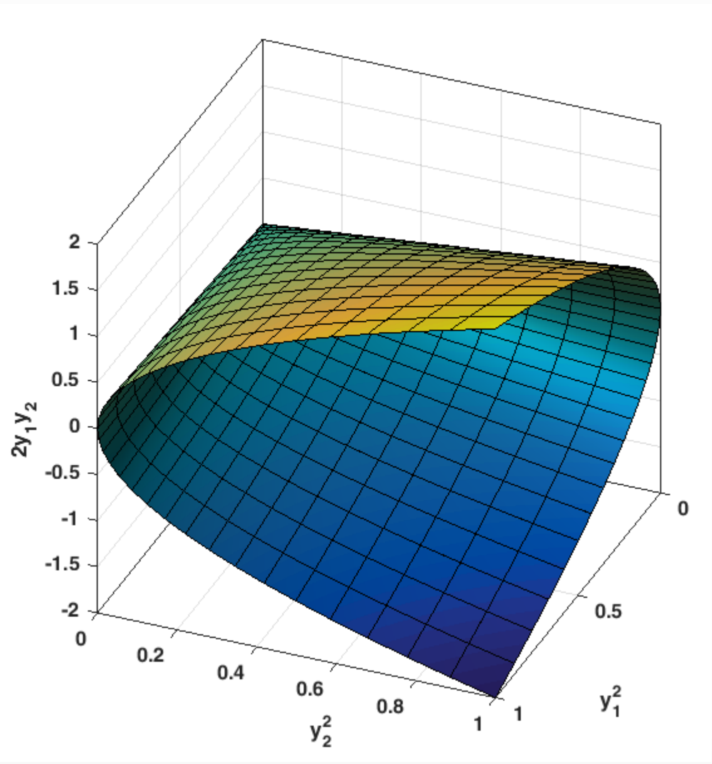
- Consider $K(x, y) := \langle x, y\rangle_2^2$, $x = (x_1, x_2), y = (y_1, y_2) \in \mathbb{R}^2$.
- The monomial functions x_1^2, x_2^2, $x_1 x_2$ form a basis of $H(K)$.
- $k_y = y_1^2 x_1^2 + y_2^2 x_2^2 + 2y_1 y_2 x_1 x_2$ is the unique linear combination of k_y with respect to the monomial basis..
- The embedding $\phi : \mathbb{R}^2 \to H(K)$ is therefore essentially equal to the map:

$$\mathbb{R}^2 \to \mathbb{R}^3, \ (y_1, y_2) \mapsto (y_1^2, y_2^2, 2y_1 y_2).$$

EXAMPLE: Polynomial RKHS of degree $d = 2$ on $\mathbb{R}^2$ The image of to the map:

$$\mathbb{R}^2 \to \mathbb{R}^3, \ (y_1, y_2) \mapsto (y_1^2, y_2^2, 2y_1y_2).$$





Further Reading – introduction to the field

- H. Knaf, Kernel Fisher discriminant functions – a concise and rigorous introduction, Berichte des ITWM 117 (2007).
  - Full proofs of all results mentioned in the present slides (except the ones in the introduction) can be found in this report.
- J.H. Manton, P.-O. Amblard: *A Primer on Reproducing Kernel Hilbert Spaces*, Foundations and Trends in Signal Processing Vol. 8 (2015). (Preprint in arXiv)
- V. I. Paulsen, M. Raghupathi: *An Introduction to the Theory of Reproducing Kernel Hilbert Spaces*, Cambridge Studies in Advanced Mathematics 152 (2016). (Preprint in arXiv)
- J.S. Taylor, N. Cristianini, *Kernel Methods in Pattern Analysis*, Cambridge University Press 2004.

Further Reading – scientific articles

- N. Aronszajn: *Theory of reproducing kernels*, Trans. Amer. Math. Soc. 68, (1950).
- I. Steinwart et al.: *An explicit description of the reproducing kernel Hilbert spaces of Gaussian RBF kernels*, Los Alamos Report LA-UR 04-8274 (2006).