RKHS, Mercer's theorem, Unbounded domains, Frames and Wavelets

9.520 Class 22, 2004

Tomaso Poggio and Sayan Mukherjee

About this class

Goal To introduce an alternate perspective of RKHS via integral operators and Mercer's theorem. Formulate the problem of RKHS on unbounded domains. Construct RKHS from frames and wavelets.

Integral operators

Consider the integral operator L_K on $L_2(X, \nu)$ defined by

$$\int_X K(\mathbf{s}, \mathbf{t}) f(\mathbf{s}) d\nu(\mathbf{s}) = g(\mathbf{t})$$

where X is a compact subset of $\mathbb{R}^n$ and ν a Borel measure.*

If K is pd then L_K is positive that is

$$\int_X K(\mathbf{t}, \mathbf{s}) f(\mathbf{t}) f(\mathbf{s}) d\nu(\mathbf{t}) d\nu(\mathbf{s}) \geq 0$$

for all $f \in L_2(X, \nu)$. The converse is also true, that is L_K positive implies that K is pd.

*We assume K to be continuous which implies the integral operator is compact.

Mercer's theorem

A symmetric, *pd* kernel $K : X \times X \to \mathbb{R}$, with X a compact subset of $\mathbb{R}^n$ has the expansion

$$K(\mathbf{s}, \mathbf{t}) = \sum_{q=1}^{\infty} \mu_q \phi_q(\mathbf{s}) \phi_q(\mathbf{t})$$

where the convergence is in $L_2(X, \nu)$.*

The ϕ_q are the orthonormal eigenfunctions of the integral equation

$$\int_X K(\mathbf{s}, \mathbf{t}) \phi(\mathbf{s}) d\nu(\mathbf{s}) = \mu\, \phi(\mathbf{t}).$$

*If the measure ν on X is non-degenerate in the sense that open sets have positive measure everywhere, then the convergence is absolute and uniform and the $\phi(\mathbf{x})$ are continuous on X (see Cucker and Smale and their correction to the Bull Am Math paper).

RKHS can be very rich

For many kernels K the number of positive eigenvalues, μ_q, is infinite. The corresponding RKHS are dense in L_2. This means is that these hypothesis spaces are "big", since they can approximate arbitrarily well a very rich class of functions can be approximated.

One reason why we work with RKHS because they can represent a very large set of hypotheses.

Another view of RKHS

Consider the space of functions associated with a pd kernel K by Mercer's theorem, that is spanned by the $\phi_p(s)$,

$$\mathcal{H}_K = \{f \mid f(s) = \sum_{p=1}^{\infty} c_p \phi_p(s)\}.$$

The space $\mathcal{H}$ is the RKHS associated with K if I assume that $\|f\|_K < \infty$, where $\|f\|_K$ is the norm induced by the dot product defined as

$$\langle f, g \rangle_K = \left\langle \sum_{p=1}^{\infty} c_p \phi_p(s), \sum_{q=1}^{\infty} d_q \phi_q(s) \right\rangle_K \equiv \sum_{p=1}^{\infty} \frac{d_p c_p}{\mu_p},$$

$$\|\mathbf{f}\|_K^2 = \left\langle \sum_{p=1}^{\infty} c_p \phi_p(s), \sum_{q=1}^{\infty} c_q \phi_q(s) \right\rangle_K \equiv \sum_{p=1}^{\infty} \frac{c_p^2}{\mu_p}.$$

Another view of RKHS (consistency checks)

One can check that the dot product defined above gives the reproducing property of K

$$\begin{aligned}
\langle f(\cdot), K(\cdot, x)\rangle_K &= \sum_p \frac{c_p \phi_p(x) \mu_p}{\mu_p} \\
&= \sum_p c_p \phi_p(x) \\
&= f(x).
\end{aligned}$$

One can also check that the dot product defined above is the same as the dot product defined earlier, that is $\langle f, g\rangle_K = \sum_{i,j} \alpha_i \beta_j K(x_i, x_j)$ by using Mercer's theorem. The RKHS defined in these two ways is the same (for rigorous proofs see Cucker and Smale).

Another proof of the representer theorem

Let us now derive the main result of an earlier class using Mercer theorem. The functional to be minimized can be written as

$$\sum_{i=1}^{\ell}(y_i - f(x_i))^2 + \lambda \sum_{p=1}^{\infty} \frac{c_p^2}{\lambda_p}.$$

Since we know that $f(x) = \sum_{q=1}^{\infty} c_q \phi_q(x)$, taking the derivative with respect to c_q gives

$$c_q = \lambda_q \sum_{i=1}^{\ell} \alpha_i \phi_q(x_i),$$

with $\alpha_i = (y_i - f(x_i))/\lambda$, from which we find

$$f(x) = \sum_{q=1}^{\infty} \sum_{i=1}^{\ell} \alpha_i \lambda_q \phi_q(x_i) \phi_q(x) = \sum_{i=1}^{\ell} \alpha_i K(x, x_i).$$

Note that unlike the case of translation invariance, the kernel function can also be a true function of two variables.

Unbounded domains: 1-D Linear Splines

If

$$\|f\|_K^2 \equiv \int |f'|^2(x)dx = \frac{1}{2\pi}\int \omega^2|\tilde{f}(\omega)|^2 d\omega$$

then

$$\tilde{K}(\omega) = \frac{1}{\omega^2} \text{ and } K(x-y) \propto |x-y|.$$

So

$$f(x) = \sum_{i=1}^{\ell} \alpha_i |x - x_i| + d_1.$$

The solution is a **piecewise linear polynomial**, i.e. a spline of order 1.

Mercer's theorem on unbounded domains

If the kernel is symmetric but defined over an unbounded domain, say $L_2([-\infty, \infty] \times [\infty, \infty])$, the eigenvalues of the equation

$$\int_{-\infty}^{\infty} K(s,t)\phi(s)ds = \lambda\phi(t)$$

are not necessarily countable and Mercer theorem does not apply.

Let us consider the special case, of considerable interest in the learning context, in which the kernel is translation invariant, or

$$K(s,t) = K(s-t).$$

As we will see, this implies that we will have to consider *Fourier hypotheses spaces*! BTW, the many different Fourier hypotheses spaces all have the same "features"!

Fourier Transform

The *Fourier Transform* of a real valued function $f \in L_1$ is the complex valued function $\tilde{f}(\omega)$ defined as

$$\tilde{f}(\omega) = \int_{-\infty}^{+\infty} f(x)\ e^{-j\omega x} dx$$

The original function f can be obtained through the *inverse Fourier Transform* as

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \tilde{f}(\omega)\ e^{j\omega x} d\omega$$

Notice that periodic functions can be expanded in a Fourier series. This can be shown from the periodicity condition $f(x+T) = f(x)$ for $T = \frac{2\pi}{\omega_0}$. Taking the Fourier transform of both side yields $\tilde{f}(\omega)\ e^{-j\omega T} = \tilde{f}(\omega)$. This is possible if $\tilde{f}(\omega) \neq 0$ only when $\omega = n\omega_0$. This implies for nontrivial f $\tilde{f}(\omega) = \sum_n \beta_n \delta(\omega - n\omega_0)$, which is a Fourier series.

Parseval's formula

If f is **also** square integrable, the *Fourier Transform* leaves the norm of f unchanged. Parseval formula gives

$$\int_{-\infty}^{+\infty} |f(x)|^2 dx = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |\tilde{f}(\omega)|^2 d\omega.$$

A Mercer-like theorem for translation invariant kernels

For shift invariant kernels

$$K(s-t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \tilde{K}(\omega)e^{j\omega(s-t)}d\omega$$
$$= \frac{1}{2\pi}\int_{-\infty}^{+\infty} \tilde{K}(\omega)e^{j\omega(s)}e^{j\omega(-t)}d\omega.$$

For kernels to which Mercer theorem applies (those with a bounded domain)

$$K(s,t) = \sum_{p=1}^{\infty} \lambda_p \phi_p(s)\phi_p^*(t).$$

This suggests the following correspondence

$$\tilde{K}(\omega) \to \lambda_p \quad \text{and} \quad e^{j\omega s} \to \phi_p(s).$$

A Mercer-like theorem for translation invariant kernels

The associated eigenvalue problem is then:

$$\int_{-\infty}^{\infty} K(s-t)\phi(s)ds = \lambda\phi(t).$$

Notice that (because of the convolution theorem) the Fourier transform of the above is $\tilde{K}(\omega)\tilde{\phi}(\omega) = \lambda\tilde{\phi}(\omega)$.

Bochner theorem

A function $K(s-t)$ is **positive definite** if and only if it is the Fourier transform of a symmetric, positive function $\tilde{K}(\omega)$ decreasing to 0 at infinity.

This sounds familiar and it is necessary to make consistent the previous correspondance.

RKHS for shift-invariant kernels

If we consider a positive definite function $K(s-t)$ and define, in the Fourier domain, the scalar product

$$\langle f(s), g(s) \rangle_K \equiv \frac{1}{2\pi} \int \frac{\tilde{f}(\omega)\tilde{g}^*(\omega)}{\tilde{K}(\omega)} d\omega,$$

the subspace $\mathcal{H}_K$ of L_2 of the functions f for which

$$\|f\|_K^2 = \frac{1}{2\pi} \int \frac{|\tilde{f}(\omega)|^2}{\tilde{K}(\omega)} d\omega < +\infty.$$

is a RKHS.

Reproducing property

The reproducing property can be easily verified. Taking the product $\langle K(s-t), f(s)\rangle_K$ gives

$$\langle K(s-t), f(s)\rangle_K \equiv \frac{1}{2\pi}\int \frac{\tilde{K}(\omega)\tilde{f}^*(\omega)e^{-j\omega t}}{\tilde{K}(\omega)}d\omega = f(t).$$

Note that the Fourier domain is natural for shift invariant kernels.

Regularizers

In the regularization formulation of radial basis functions or SVMs, the regularizers induced by the previous RKHS norm have thus the form

$$\|f\|_K^2 = \frac{1}{2\pi} \int \frac{|\tilde{f}(\omega)|^2}{\tilde{K}(\omega)} d\omega < +\infty.$$

defined via the Fourier transform of the kernel K. They are called smoothness functionals for reasons that will become even more obvious.

Examples of smoothness functionals, i.e. RKHS norms, in the Fourier domain

Consider

$$\Phi_1[f] = \int_{-\infty}^{+\infty} |f'(x)|^2 dx \; = \; \frac{1}{2\pi} \int \omega^2 |\tilde{f}(\omega)|^2 d\omega$$

$$\Phi_2[f] = \int_{-\infty}^{+\infty} |f''(x)|^2 dx \; = \; \frac{1}{2\pi} \int \omega^4 |\tilde{f}(\omega)|^2 d\omega$$

Examples of smoothness functionals, i.e. RKHS norms, in the Fourier domain (cont.)

Note again that both functionals are of the form

$$\Phi[f] = \frac{1}{2\pi} \int \frac{|\tilde{f}(\omega)|^2}{\tilde{K}(\omega)} d\omega = |f|_K^2$$

for some positive, symmetric function $\tilde{K}(\omega)$ decreasing to zero at infinity. In particular, we have

- $\tilde{K}(\omega) = 1/\omega^2$ for Φ_1,

- $\tilde{K}(\omega) = 1/\omega^4$ for Φ_2.

Examples of smoothness functionals, i.e. RKHS norms, in the Fourier domain (cont.)

Considering the FT in the distribution sense we have

$$K(x) = -|x|/2 \quad \Longleftrightarrow \quad \tilde{K}(\omega) = 1/\omega^2$$
$$K(x) = -|x|^3/12 \quad \Longleftrightarrow \quad \tilde{K}(\omega) = 1/\omega^4$$

For both kernels, the singularity of the FT for $\omega = 0$ is due to the seminorm property and, as we will see later, to the fact that the kernel is only conditionally positive definite. Notice that these kernels are symmetric and positive definite (therefore) with a real, positive Fourier transform.

Other examples of kernel functions

Other possible kernel functions are, for example,

$$K(x) = e^{-x^2/2\sigma^2} \quad \Longleftrightarrow \quad \tilde{K}(\omega) = e^{-\omega^2\sigma^2/2}$$

$$K(x) = \frac{1}{2}e^{-\gamma|x|} \quad \Longleftrightarrow \quad \tilde{K}(\omega) = \frac{1}{1+\omega^2}$$

$$K(x) = \sin(\Omega x)/(\pi x) \quad \Longleftrightarrow \quad \tilde{K}(\omega) = U(\omega + \Omega) - U(\omega - \Omega).$$

Note that all these are Fourier pairs in the ordinary sense.

Corresponding hypothesis spaces

- As it can easily be seen through power series expansion, the hypothesis space for the Gaussian kernel consists of the set of square integrable functions whose derivatives of all orders are square integrable.

- The hypothesis space of the kernel

$$K(x) = 1/2e^{-\gamma|x|}$$

consists of the square integrable functions whose first derivative is square integrable (Sobolev space).

- The hypothesis space of the kernel

$$K = \sin(\Omega x)/(\pi x)$$

is the space of *band-limited* functions whose FT vanishes for $|\omega| > \Omega$.

The representer theorem on unbounded domains for shift invariant kernels

We are now ready to state an important result (Duchon, 1977; Meinguet, 1979; Wahba, 1977; Madych and Nelson, 1990; Poggio and Girosi, 1989; Girosi, 1992)

Theorem: Let $\tilde{K}(\omega)$ be the FT (in the ordinary sense) of a kernel function $K(x)$. The function f_λ minimizing the functional

$$\frac{1}{\ell}\sum_{i=1}^{\ell}(y_i - f(x_i))^2 + \lambda\frac{1}{2\pi}\int\frac{|\tilde{f}(\omega)|^2}{\tilde{K}(\omega)}d\omega$$

has the form

$$f_\lambda(x) = \sum_{i=1}^{\ell}\alpha_i K(x - x_i)$$

for some suitable values of the coefficients α_i, $i = 1, ..., \ell$.

The proof is in Appendix 1 of this class.

RKHS and smoothness functionals

Summing up, smoothness functionals can be seen as controlling the norm of functions in some RKHS.

These functionals can be built directly in the original space and amounts to finding a positive definite function inducing an appropriate norm in the space.

With some care, the analysis can be extended to the case in which the function is only conditionally positive definite.

Conditionally positive definite functions

Let $r = \|\mathbf{x}\|$ with $\mathbf{x} \in \mathbb{R}^n$. A continuous function $K = K(r)$ is **conditionally (strictly) positive definite** of order m on $\mathbb{R}^n$, if and only if for any distinct points $\mathbf{t}_1, \mathbf{t}_2, ..., \mathbf{t}_\ell \in \mathbb{R}^n$ and scalars $c_1, c_2, ..., c_\ell$ such that $\sum_{i=1}^{\ell} c_i p(\mathbf{t}_i) = 0$ for all $p \in \pi_{m-1}(\mathbb{R}^n)$, the quadratic form

$$\sum_{i=1}^{\ell} \sum_{j=1}^{\ell} c_i c_j K(\|\mathbf{t}_i - \mathbf{t}_j\|)$$

is (positive) nonnegative.

$$K(x) = -\frac{|x|^{2m-1}}{2m(2m-1)} \rightarrow \tilde{K}(\omega) = 1/\omega^{2m}$$

is an example of **conditionally positive definite** function of order m (Madych and Nelson, 1990).

Norms and seminorms

If K is a conditionally positive definite function of order m, then

$$\Phi[f] = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{|\tilde{f}(\omega)|^2}{\tilde{K}(\omega)}$$

is a **seminorm** whose null space is the set of polynomials of degree $m - 1$.

If K is strictly positive definite, then Φ is a norm.

(Madych and Nelson, 1990)

Computation of the coefficients

For a positive definite kernel, we have shown that

$$f_\lambda(x) = \sum_{i=1}^{\ell} \alpha_i K(x, x_i).$$

The coefficients α_i can be found by solving the linear system

$$(K + \ell\lambda I)\alpha = \mathbf{y}.$$

with $K_{ij} = K(x_i, x_j), \quad i, j = 1, 2, ..., \ell,$

$\alpha = (\alpha_1, \alpha_2, ..., \alpha_\ell),$

$\mathbf{y} = (y_1, y_2, ..., y_\ell)$ and

I the $\ell \times \ell$ identity matrix.

Computation of the coefficients (cont.)

For a conditionally positive definite kernel of order m, it can be shown that

$$f_\lambda(x) = \sum_{i=1}^{\ell} \alpha_i K(x, x_i) + \sum_{k=1}^{m-1} d_k \gamma_k(x),$$

where the coefficients α and $\mathbf{d} = (d_1, d_2, ..., d_m)$ are found by solving the linear system

$$(K + \ell\lambda I)\alpha + \Gamma^\top \mathbf{d} = \mathbf{y}$$

$$\Gamma\alpha = 0.$$

with

$$\Gamma_{ik} = \gamma_k(x_i).$$

Example 1: 1-D Linear Splines (again)

$$\|f\|_K^2 = \int |f'(x)|^2 dx = \frac{1}{2\pi} \int \omega^2 |\tilde{f}(\omega)|^2 d\omega$$

$$\tilde{K}(\omega) = \frac{1}{\omega^2}$$

$$K(x-y) \propto |x-y|$$

$$f(x) = \sum_{i=1}^{\ell} \alpha_i |x - x_i| + d_1.$$

The solution is a **piecewise linear polynomial**.

Example 2: 1-D Cubic Splines

$$\|f\|_K^2 = \Phi[f] = \int |f''(x)|^2 dx = \frac{1}{2\pi} \int \omega^4 |\tilde{f}(\omega)|^2 d\omega$$

$$\tilde{K}(\omega) = \frac{1}{\omega^4}$$

$$K(x-y) \propto |x-y|^3$$

$$f(x) = \sum_{i=1}^{\ell} \alpha_i |x - x_i|^3 + d_2 x + d_1.$$

The solution is a **piecewise cubic polynomial**.

Example 3: 2-D Thin Plate Splines

$$\|f\|_K^2 = \Phi[f] = \int\int dx_1 dx_2 \left(\left(\frac{\partial^2 f}{\partial x_1^2} \right)^2 + 2 \left(\frac{\partial^2 f}{\partial x_1 \partial x_2} \right)^2 + \left(\frac{\partial^2 f}{\partial x_2^2} \right)^2 \right)$$

$$\|f\|_K^2 = \Phi[f] = \frac{1}{(2\pi)^2} \int\int d\omega \ \|\omega\|^4 |\tilde{f}(\omega)|^2$$

$$\tilde{K}(\omega) = \frac{1}{\|\omega\|^4}$$

$$K(\mathbf{x}) \propto \|\mathbf{x}\|^2 \ln \|\mathbf{x}\|$$

$$f(\mathbf{x}) = \sum_{i=1}^{\ell} \alpha_i \|\mathbf{x} - \mathbf{x}_i\|^2 \ln \|\mathbf{x} - \mathbf{x}_i\| + \mathbf{d}_2 \cdot \mathbf{x} + d_1$$

Notes on stabilizers and splines

If the stabilizer is $|Pf\|^2 = \|D^n f\|^2$ where D^n is derivative of order n then the null space of the stabilizer is the space of polynomes of order $n-1$. The kernel – and therefore the order of the splines – is determined by the degree of $|Pf\|^2$: for the n derivative the order is $2n$. Furthermore, the kernel corresponding to D^{2n} is $|x|^{2n-1}$. Thus for $n = 1$ one has $|Pf\|^2 = \|D^1 f\|^2$ with a null space of constants and a kernel $|x|$; for $n = 2$ one has $|Pf\|^2 = \|D^2 f\|^2$ with a null space of linear functions and a kernel $|x|^3$.

Example 4: Gaussian RBFs

$$\|f\|_K^2 = \Phi[f] = \frac{1}{2\pi} \int d\omega \; e^{\|\boldsymbol{\omega}\|^2 \sigma^2/2} |\tilde{f}(\omega)|^2$$

$$\tilde{K}(\omega) = e^{-\|\omega\|^2 \sigma^2/2}$$

$$K(\mathbf{x}) = e^{-\|\mathbf{x}\|^2/2\sigma^2}$$

$$f(\mathbf{x}) = \sum_{i=1}^{\ell} \alpha_i e^{-\|\mathbf{x}-\mathbf{x}_i\|^2/2\sigma^2}$$

Example 5: Radial Basis Functions

Define $r = \|\mathbf{x}\|$, $\mathbf{x} \in \mathbb{R}^n$.

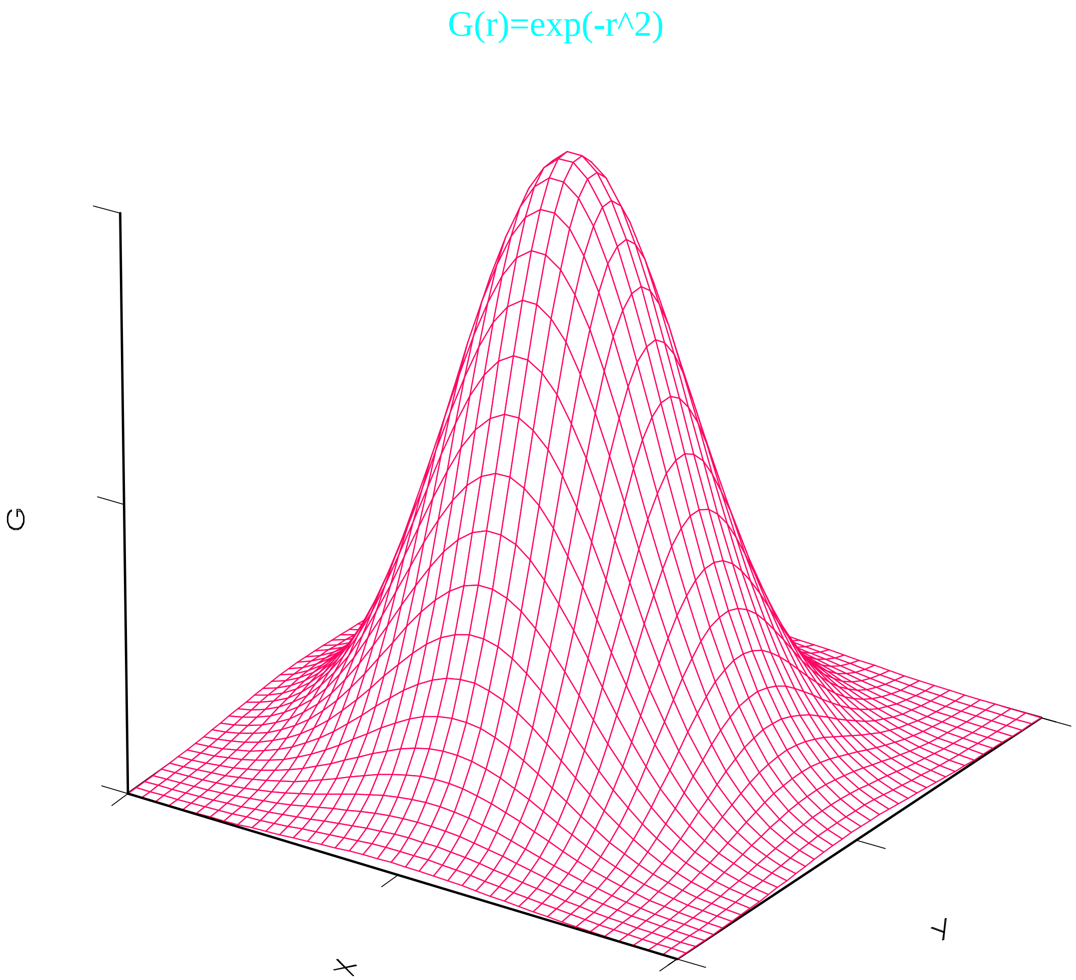
$K(r) = e^{-r^2/2\sigma^2}$ Gaussian

$K(r) = \sqrt{r^2 + c^2}$ multiquadric

$K(r) = \frac{1}{\sqrt{c^2+r^2}}$ inverse multiquadric

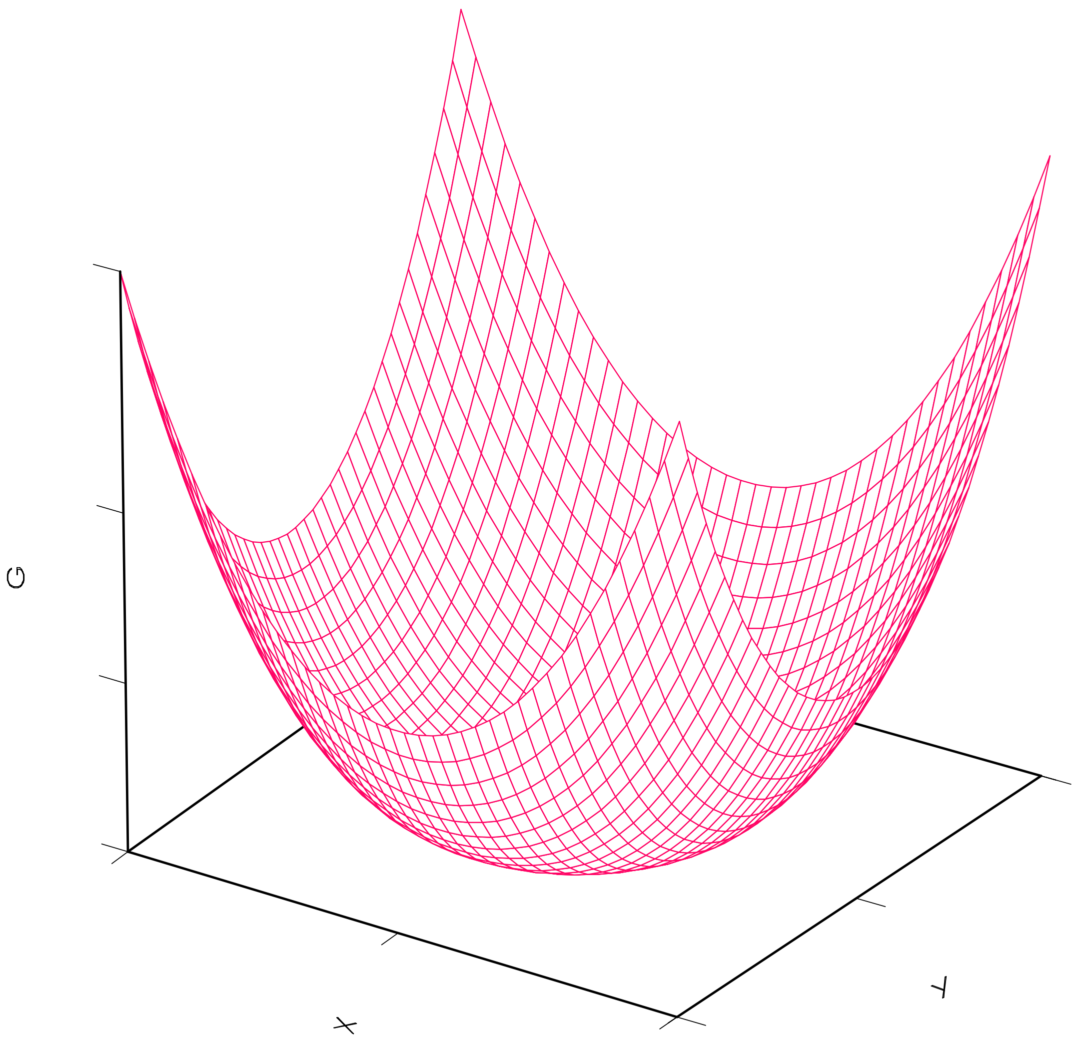
$K(r) = r^{2m-n} \ln r$ multivariate splines (n even)

$K(r) = r^{2m-n}$ multivariate splines (n odd)



G(r)=exp(-r^2)
G
X
Y

G(r)=r^2 ln(r)



Frames

A family of functions $(\phi_j)_{j\in J}$ in an Hilbert space H is called a *frame* if there exist $0 < A$ and $A \leq B < \infty$ so that, for all $f \in H$:

$$A\|f\|^2 \leq \sum_{j\in J} |\langle f, \phi_j\rangle|^2 \leq B\|f\|^2$$

We call A and B the *frame bounds*.

Tight frames

If the frame bounds A and B are such that $A = B$ then the frame is a *tight* frame.

For a tight frame we have

$$f(\mathbf{x}) = \frac{1}{A} \sum_{j \in J} \langle f, \phi_j \rangle \phi_j(\mathbf{x})$$

If $A = 1$ the frame is an orthonormal basis.

Frame Operator

Our goal is to find a "reconstruction" formula, that allows us to reconstruct f from its projection on the frame.

We need to introduce the *frame operator*:

$$F: \; H \to \ell_2$$

such that

$$(Ff)_j = \langle f, \phi_j \rangle$$

(remember that ℓ_2 is the Hilbert space of square summable sequences).

Adjoint of the Frame Operator

The adjoint operator F^* is the operator that satisfies:

$$\langle F^* c, f \rangle = \langle c, Ff \rangle = \sum_{j \in J} c_j \langle f, \phi_j \rangle$$

and therefore:

$$F^* c = \sum_{j \in J} c_j \phi_j(\mathbf{x})$$

The Dual Frame

If $(F^*F)^{-1}$ exists, then the frame is said to be a *Riesz frame* with linearly independent ϕ. For a Riesz frame, applying the operator $(F^*F)^{-1}$ to the frame functions ϕ_j gives another set of basis functions:

$$\tilde{\phi}_j = (F^*F)^{-1}\phi_j$$

The family of functions $\left(\tilde{\phi}_j\right)_{j\in J}$ is a frame, with frame constants B^{-1} and A^{-1}.

$$B^{-1}\|f\|^2 \leq \sum_{j\in J} |\langle f, \tilde{\phi}_j\rangle|^2 \leq A^{-1}\|f\|^2$$

The frame and dual frames are biorthogonal

$$\langle \tilde{\phi}_j, \phi_i\rangle = \delta_{i,j}.$$

Frame example

The set

$$\phi_1 = (0,1); \quad \phi_2 = (-1,0); \quad \phi_3 = (\sqrt{2}/2, -\sqrt{2}/2)$$

is a frame in $\mathbb{R}^2$, because for any $\mathbf{v} \in \mathbb{R}^2$ we have

$$\|\mathbf{v}\|^2 \leq \sum_{j=1}^{3} (\mathbf{v} \cdot \phi_j)^2 \leq 2\|\mathbf{v}\|^2$$

Thus, $A = 1$ and $B = 2$.

$$F = \begin{pmatrix} 0 & 1 \\ -1 & 0 \\ \sqrt{2}/2 & -\sqrt{2}/2 \end{pmatrix}.$$

$F^* = F^\top$

Frame example (dual frames)

$$F^*F = \begin{pmatrix} 3/2 & -1/2 \\ -1/2 & 3/2 \end{pmatrix}$$

and

$$(F^*F)^{-1} = \frac{1}{2} \begin{pmatrix} 3/2 & 1/2 \\ 1/2 & 3/2 \end{pmatrix}.$$

Thus

$$\tilde{\phi}_1 = (1/4, 3/4), \quad \tilde{\phi}_2 = (-3/4, -1/4),$$

$$\tilde{\phi}_3 = (\sqrt{2}/4, -\sqrt{2}/4).$$

Biorthogonality

If a set of frames $\tilde{\phi}_j$ and ϕ_j are biorthogonal then

$$\langle \phi_i, \tilde{\phi}_j \rangle = \langle \phi_i, ((F^*F)^{-1}\Phi)_j \rangle = \delta_{i,j},$$

or

$$(\Phi, (F^*F)^{-1}\Phi) = \mathrm{Id},$$

since

$$\langle \Phi, (F^*F)^{-1}\Phi \rangle = \Phi^T\Phi(F^*F)^{-1},$$

in operator notation this can be written

$$\Phi^T\Phi(F^*F)^{-1} = (F^*F)(F^*F)^{-1} = \mathrm{Id}.$$

Reconstruction Formula for Frames

The biorthogonality due to the frame operator and its adjoint results in the following reconstruction

$$\sum_{j \in J} \langle f, \phi_j \rangle \tilde{\phi}_j = f = \sum_{j \in J} \langle f, \tilde{\phi}_j \rangle \phi_j$$

Therefore f can be reconstructed from its frame coefficients *if the dual frame is used for the reconstruction* (frame and dual frame can also be exchanged).

RKHS and Frames

Previously we looked at a RKHS spanned by a space of functions $\phi_p(s)$ where

$$\mathcal{H}_K = \{f \mid f(s) = \sum_{p=1}^{\infty} c_p \phi_p(s)\}.$$

The functions $\{\phi_p(s)\}_{p=1}^{\infty}$ are now a frame in the RKHS rather than the eigenvectors of an integral equation.

One gets the following analog of "Mercer's" theorem

$$K(s,t) = \sum_{i=1}^{\infty} \tilde{\phi}_i(s) \phi_i(t),$$

where $\{\phi_i(s)\}_{i=1}^{\infty}$ and $\{\tilde{\phi}_i(s)\}_{i=1}^{\infty}$ are the frames and dual frames.

RKHS and Frames (Consistency)

We can write

$$
\begin{aligned}
f(x) &= \sum_{j=1}^{N} c_j K(x_j, x) \\
&= \sum_{j=1}^{N} c_j \sum_{p=1}^{\infty} \tilde{\phi}_p(x_j)\phi_p(x) \\
&= \sum_{p=1}^{\infty} \left(\sum_{j=1}^{N} c_j \tilde{\phi}_p(x_j) \right) \phi_p(x) \\
&= \sum_{p=1}^{\infty} d_p \phi_p(x),
\end{aligned}
$$

where $d_p = \sum_{j=1}^{N} c_j \tilde{\phi}_p(x_j)$.

Constructing a RKHS from frames

Let $\{\phi_n\}_{n=1}^{L}$ be a finite set of non-zero functions of a Hilbert Space $\mathcal{H}$ so that:

$$\forall n \ \ ||\phi_n|| < \infty$$

and

$$\forall x \ \ |\phi_n(x)| \leq M.$$

Let $\mathcal{H}$ be the set of functions:

$$\mathcal{H}_K = \{f \mid f(s) = \sum_{n=1}^{L} c_n \phi_n(s)\}.$$

$\langle \mathcal{H}_K, \langle \cdot, \cdot \rangle_K \rangle$ is a RKHS and the Reproducing Kernel is

$$K(s,t) = \sum_{n=1}^{L} \tilde{\phi}_n(s) \phi_n(t).$$

Symmetry of the Kernel

It is not immediately obvious that the Kernel

$$K(s,t) = \sum_{n=1}^{L} \tilde{\phi}_n(s)\phi_n(t),$$

is symmetric because in general $\tilde{\phi}_n(t) \neq \phi_n(t)$.

By the reproducing property

$$f(x) = \langle f(\cdot), K(\cdot,t)\rangle_K = \langle f(\cdot), K(s,\cdot)\rangle_K,$$

so

$$0 = \langle f(\cdot), K(\cdot,t) - K(s,\cdot)\rangle_K$$

and

$$K(s,t) = K(t,s).$$

The Frame operator and the evaluation functional

For any function $f \in \mathcal{H}$

$$\langle F^* F K, f \rangle = f(x),$$

or

$$\langle K, f \rangle_K = f(x),$$

so the evaluation functional can be written in terms of the frame operator.

Appendix 1: proof of the representer theorem

We first rewrite the functional in term of the FT $\tilde{f}$ and using property 1 to obtain

$$\sum_{i=1}^{\ell} \left(y_i - \frac{1}{2\pi} \int \tilde{f}(\omega) e^{j\omega x_i} d\omega \right)^2 + \lambda \frac{1}{2\pi} \int \frac{\tilde{f}(-\omega)\tilde{f}(\omega)}{\tilde{K}(\omega)} d\omega.$$

Taking the functional derivative w.r.t $\tilde{f}(\xi)$ gives

$$-\frac{1}{\pi} \sum_{i=1}^{\ell} (y_i - f(x_i)) \int \frac{\mathcal{D}\tilde{f}(\omega)}{\mathcal{D}\tilde{f}(\xi)} e^{j\omega x_i} d\omega + \frac{2}{2\pi} \lambda \int \frac{\tilde{f}(-\omega)}{\tilde{K}(\omega)} \frac{\mathcal{D}\tilde{f}(\omega)}{\mathcal{D}\tilde{f}(\xi)} d\omega =$$

$$-\frac{1}{\pi} \sum_{i=1}^{\ell} (y_i - f(x_i)) \int \delta(\omega - \xi) e^{j\omega x_i} d\omega + \frac{1}{\pi} \lambda \int \frac{\tilde{f}(-\omega)}{\tilde{K}(\omega)} \delta(\omega - \xi) d\omega.$$

From the definition of δ we have

$$-\frac{1}{\pi} \sum_{i=1}^{\ell} (y_i - f(x_i)) e^{j\xi x_i} + \frac{1}{\pi} \lambda \frac{\tilde{f}(-\xi)}{\tilde{K}(\xi)}$$

Proof (cont.)

Equating the derivative to zero and changing the sign of ξ we find

$$\tilde{f}_\lambda(\xi) = \tilde{K}(\xi) \sum_{i=1}^{\ell} \frac{y_i - f(x_i)}{\lambda} e^{-j\xi x_i}.$$

Defining the coefficients

$$\alpha_i = \frac{y_i - f(x_i)}{\lambda},$$

taking the inverse FT and using property 2, we finally obtain

$$f_\lambda(x) = \sum_{i=1}^{\ell} \alpha_i K(x - x_i).$$

Appendix 2: A second route for the extension to unbounded domain: Evaluation functionals

As we have seen already, a more general way to introduce RKHS and derive its properties (which includes what we have seen so far as special cases and does not require the analysis of the spectrum of the kernel function) depends on the fundamental assumption that in the considered space all the linear evaluation functionals for each f are continuous and bounded.

A linear evaluation functional is a functional $\mathcal{F}_t$ that *evaluates* each function in the space at the point t, or

$$\mathcal{F}_t[f] = f(t).$$

Appendix 2: Continuous and bounded functionals

In a normed space a linear functional is *continuous* if and only if it is *bounded* on the unit sphere.

The norm of a functional $\mathcal{F}$ is given by

$$\|\mathcal{F}\| = \sup_{\|f\| \leq 1} |\mathcal{F}[f]|$$

Appendix 2: Example in infinite dimensional spaces

In $C[a, b]$ (with the *sup* norm) the linear evaluation functional

$$\delta_t[f] = f(t)$$

is continuous because

$$\|\delta_t[f]\| \leq \|f\|.$$

Clearly, its norm is $\|\delta_t[f]\| = 1$.

Appendix 2: RKHS: a reminder

Given a Hilbert space with the property that all the linear evaluation functionals for each function f are continuous and bounded, through Riesz–Fisher representation theorem it can be proven that one can always find a positive definition function $K(s,t)$ with the reproducing property

$$f(t) = \langle K(s,t), f(s) \rangle_K.$$

Appendix 3: Example: bandlimited functions again

We met the smoothness functional

$$\Phi[f] = \frac{1}{2\pi} \int_{-\Omega}^{+\Omega} |\tilde{f}(\omega)|^2 d\omega$$

induced by the positive definite function

$$K(s-t) = \frac{\sin(\Omega(s-t))}{\pi(s-t)}$$

whose Fourier Transform equals 1 in the interval $[-\Omega, \Omega]$ and 0 otherwise. The space of functions identified by this smoothness functionals, the set of Ω-bandlimited functions, can be seen as RKHS.

Appendix 3: Evaluation functional for bandlimited functions

In agreement with the correspondence of above, from the convolution theorem it is easy to see that all Ω-bandlimited functions are eigenfunctions of the integral equation

$$\int \frac{\sin(\Omega(s-t))}{\pi t} f(s)ds = \lambda f(t)$$

with 1 as unique eigenvalue. From the same equation and the fact the kernel does not appear explicitly in the definition of the norm of f, we see that the evaluation functional for each fixed t can be written as

$$\mathcal{F}_t[f(s)] = \int \frac{\sin(\Omega(s-t))}{\pi t} f(s)ds = f(t).$$

Appendix 3: The Gaussian example

In general, the evaluation functionals are not easy to write. In the Gaussian case, for example, we know how to write the evaluation functional for each **fixed** t in the Fourier domain,

$$\langle f(s), e^{-(s-t)^2/2\sigma^2} \rangle_K = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\tilde{f}(\omega) e^{-\omega^2\sigma^2/2} e^{j\omega t}}{e^{-\omega^2\sigma^2/2}} d\omega = f(t),$$

but not in the original domain.

Appendix 4: Another interpretation: Feature Space

Another interpretation of the kernel is that one maps into a higher dimensional space called a *feature space* ($\Phi : \mathbf{x} \to \mathbf{\Phi}(\mathbf{x})$). The kernel is simply an inner product in the feature space.

The normalized bases in this higher dimensional space are

$$\left\{ \frac{1}{\sqrt{\mu_1}} \phi_1, .., \frac{1}{\sqrt{\mu_N}} \phi_N \right\}$$

where N is the number of eigenfunctions (possibly infinite) in the series expansion of the kernel.

So the inner product between two points mapped into feature space is

$$K(\mathbf{x}, \mathbf{y}) = \langle \Phi(\mathbf{x}), \Phi(\mathbf{y}) \rangle = \sum_{p=1}^{N} \frac{c_p d_p}{\mu_p}.$$

Representer Theorem (revisited)

Theorem. The solution to the Tikhonov regularization problem

$$\min_{f \in \mathcal{H}} \frac{1}{\ell} \sum_{i=1}^{\ell} V(y_i, f(x_i)) + \lambda \|f\|_K^2$$

can be written in the form

$$f(x) = \sum_{i=1}^{\ell} c_i K(x, x_i).$$

Representer Theorem Proof II, Preliminaries

Instead of writing

$$f(x) = \sum_{i=1}^{\ell} a_i K(x, x_i),$$

we can write

$$f = \sum_{i=1}^{\ell} a_i \Phi(x_i).$$

With this notation,

$$f(x) = \langle f, \Phi(x) \rangle = \left\langle \sum_{i=1}^{\ell} a_i \Phi(x_i), \Phi(x) \right\rangle = \sum_{i=1}^{\ell} a_i K(x, x_i).$$

Representer Theorem Proof II, Part 1

(Schölkopf et. al 2001) Suppose that we **cannot** write the solution to a Tikhonov problem in the form

$$f = \sum_{i=1}^{\ell} a_i \Phi(x_i).$$

Then, clearly, we can write it in the form

$$f = \sum_{i=1}^{\ell} a_i \Phi(x_i) + v,$$

where v satisfies

$$\langle v, \Phi(x_i) \rangle = 0$$

for all points in the training set.

Representer Theorem Proof II, Part 2

Applying f to an arbitrary training point x_j shows that

$$\begin{aligned} f(x_j) = \langle f, \Phi(x_j) \rangle &= \left\langle \sum_{i=1}^{\ell} a_i \Phi(x_i) + v, \Phi(x_j) \right\rangle \\ &= \left\langle \sum_{i=1}^{\ell} a_i \Phi(x_i), \Phi(x_j) \right\rangle . \end{aligned}$$

Therefore, the choice of v has no effect on $f(x_j)$ or on

$$\sum_{i=1}^{\ell} V(y_i, f(x_i)).$$

Representer Theorem Proof II, Part 3

Now, let's consider $\|f\|_K^2$:

$$\begin{aligned}
\|f\|_K^2 &= \|\sum_{i=1}^{\ell} a_i \Phi(x_i) + v\|_K^2 \\
&= \|\sum_{i=1}^{\ell} a_i \Phi(x_i)\|_K^2 + \|v\|_K^2 \\
&\geq \|\sum_{i=1}^{\ell} a_i \Phi(x_i)\|_K^2.
\end{aligned}$$

Representer Theorem Proof II, Part 4

Given a Tikhonov regularization problem

$$\min_{f \in H} \frac{1}{\ell} \sum_{i=1}^{\ell} V(y_i, f(x_i)) + \lambda \|f\|_K^2,$$

we write the solution in the form:

$$f(x) = \sum_{i=1}^{\ell} a_i \Phi(x_i) + v,$$

where v satisfies the orthogonality condition discussed above. Suppose v is not equal to 0. Consider the new function

$$f_2(x) = \sum_{i=1}^{\ell} a_i \Phi(x_i).$$

Then $V(y_i, f(x_i)) = V(y_i, f_2(x_i))$ for all training points, but $\|f\|_K^2 > \|f_2\|_K^2$, contradicting our assumption that f was optimal.

Representer Theorem (remarks)

This second proof allows for cross-talk between the empirical terms, and a monotonic function g on the complexity term; using the same proof, the solution to

$$\min_{f \in H} \; h\left[\sum_{i=1}^{\ell} V(y_i, f(x_i))\right] + g(\|f\|_K^2)$$

has the same form, where h is arbitrary and g is monotonically increasing.