An Introduction to Generalized Sobolev Spaces

Michael McCourt
Department of Mathematical and Statistical Sciences
University of Colorado Denver

Meshfree Seminar
Illinois Institute of Technology
July 21, 2014
Disclaimer

Today’s talk was thrown together very quickly and stolen liberally from old talks and chapter 7 of the as-yet unpublished book “Kernel-based Approximation Methods in MATLAB”.

As a result, I am inviting comments and questions to better understand how we can present this content effectively.

We are also skipping some material for the sake of time and simplicity, but if you do not understand something please stop me so that we can fix it.
Fundamental Application (Scattered Data Fitting)

Given data \((x_j, y_j), j = 1, \ldots, N,\) with \(x_j \in \mathbb{R}^d,\ y_j \in \mathbb{R},\) find a (continuous) function \(s_f\) such that \(s_f(x_j) = y_j,\ j = 1, \ldots, N.\)

Consider here multivariate kernel-based interpolation using a data-dependent linear function space (basis)

\[
s(x) = \sum_{j=1}^{N} c_j K(x, x_j) = k(x)^T c, \quad x \in \Omega \subseteq \mathbb{R}^d
\]

with \(K : \Omega \times \Omega \to \mathbb{R}\) a positive definite kernel.

To find \(c_j\) solve the interpolation equations

\[
s(x_i) = f(x_i) = y_i, \quad i = 1, \ldots, N
\]

which gives a linear system \(Kc = y;\) \(K\) is symmetric positive definite.
Kernel Flexibility: Goodness and Badness

The shape parameter $\varepsilon$ associated with some RBFs plays an important role in the quality of associated approximations.

Example Gaussian support vector machine decision contour

$\varepsilon = 10$, Too Erratic   $\varepsilon = 1$, Nice   $\varepsilon = .1$, Insensitive

This freedom to choose your kernel basis is a challenge when no guidance is available (e.g., MLE, CV).

Is it possible to “design” the right kernel?
Reproducing Kernel Hilbert Spaces

Each positive definite kernel $K$ has a unique Hilbert space associated with it called its Native Space, $\mathcal{H}_K$. For some domain $\Omega$,

$$\mathcal{H}_K = \left\{ f \left| f = \sum_{k=1}^{N_K} c_k K(\cdot, x_k), \quad x_k \in \Omega \right. \right\},$$

where, in most cases, $N_K = \infty$.

It can be very difficult to classify functions in this space, but all such functions satisfy the reproducing property:

$$f(z) = (f, K(\cdot, z))_{\mathcal{H}_K}, \quad z \in \Omega,$$

for the native space inner product $(f, g)_{\mathcal{H}_K}$.

Of course, as Dr. Erickson recently discussed, the Native space inner product is itself difficult to understand. But some of its properties are very useful.
Reproducing Kernel Optimality

One important consequence of the reproducing property is the interpolant error bound

$$|s(x) - f(x)| \leq \|f\|_{\mathcal{H}_K} \sqrt{K(x, x) - k(x)^T K^{-1} k(x)}.$$ 

Another significant result is the orthogonality property

$$\|f\|_{\mathcal{H}_K}^2 = \|s - f\|_{\mathcal{H}_K}^2 + \|s\|_{\mathcal{H}_K}^2,$$

which is proved in Fasshauer’s book. This has the important consequence of implying the optimality of the interpolant $s$:

$$\|s\|_{\mathcal{H}_K} \leq \|v\|_{\mathcal{H}_K}, \quad v \in \mathcal{H}_K, \quad v(x_i) = y_i$$
Interpolation through Optimization with Regularization

Infinitely many interpolants exist for any given scattered data input. The RKHS interpolant $s = k(\cdot)^T c$ is the interpolant that solves

$$\min_c \left\| \begin{pmatrix} k(x_1)^T c \\ \vdots \\ k(x_N)^T c \end{pmatrix} - \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} \right\| + \mu c^T K c.$$

because $\|s\|_{\mathcal{H}_K}^2 = c^T K c$.

Goal

- Design an inner product which measures a function’s “crumminess”
- Find the kernel whose RKHS involves that inner product
- Interpolate with that kernel to minimize “crumminess”
What is a “Good” Interpolant?

The norm minimization we achieve for free through RKHS will hopefully steer us towards a good interpolant.

The $100$ question is: how does one define good?

We believe that the answer to this question is application-dependent.

- Positive (Pressure, density, etc)
- Monotone (Increasing field)
- Integrates to 1 (Cumulative distribution function)

Often times, the real metric to be minimized is “Dollars/Science”, but such a metric is often ill-defined so we will suffice to work with some proxy.

We will focus on the relationship between function values and derivatives to define our inner product, and only work on $[0, 1]$. 
Generalized Sobolev Spaces

Drs. Ye and Fasshauer have developed Hilbert spaces whose reproducing kernels are of the form

\[ K(x, z) = G(x, z) + R(x, z), \quad R(x, z) = \sum_{k=1}^{n_a} a_k \psi_k(x) \psi_k(z), \quad a_k \geq 0 \]

where

\[ \mathcal{L} G(x, z) = \delta(x - z), \quad x, z \in \Omega, \]
\[ \mathcal{B} G(x, z) = 0, \quad x \in \partial \Omega, \]

and

\[ \mathcal{L} R(x, z) = 0 \quad (\text{implying } \psi \in \text{null}(\mathcal{L})), \]

for differential operator \( \mathcal{L} = \mathcal{P}^* \mathcal{P} \) and boundary condition operator \( \mathcal{B} \).
A Specific Generalized Sobolev Space

So that things fit on slides, we are going to consider only the case Dr. Erickson spoke on last week: \( L = -D^2 \) and \( P = D \) with \( B = I \).

The standard Sobolev space inner product, without boundary conditions, would involve only the \( P \) operator:

\[
(f, g)_P = \int_0^1 (Pf)(x)(Pg)(x) \, dx = \int_0^1 f'(x)g'(x) \, dx.
\]

Our GSS inner product will also include contributions from the boundary condition operator:

\[
(f, g)_B = f(0)g(0) + f(1)g(1).
\]

Functions in this generalized Sobolev space include

\[
\mathcal{H}^R_{P,B} = \{ f \mid (f, f)_P, (f, f)_B < \infty \},
\]

although other restrictions may apply depending on the design of \( R \).
The Components of a Generalized Sobolev Space Kernel

The kernel $G(x, z) = \min(x, z) - xz$ is called the “Brownian-Bridge” kernel and serves as half of our GSS kernel

$$K(x, z) = G(x, z) + R(x, z).$$

To determine $R(x, z) = \sum_{k=1}^{n} a_k \psi_k(x) \psi_k(z)$ we must find $\psi_k \in \text{null}(\mathcal{L}) = \text{span}\{x, 1-x\}$,

$$\psi_1(x) = c_1 x + c_2 (1 - x),$$

$$\psi_2(x) = c_3 x + c_4 (1 - x),$$

which are $(\cdot, \cdot)_B$ orthonormal (recall $(f, g)_B = f(0)g(0) + f(1)g(1)$),

$$c_1^2 + c_2^2 = 1,$$

$$c_3^2 + c_4^2 = 1,$$

$$c_1 c_3 + c_2 c_4 = 0.$$
The Components of a Generalized Sobolev Space Kernel

Plugging in these \( \psi \in \text{null}(\mathcal{L}) \) functions, and using the \( B \)-ON to simplify things, gives us

\[
K(x, z) = \min(x, z) + xz(-1 + a_1 + a_2 - 2c_1 c_2(a_1 - a_2)) \\
+ (x + z)(c_1 c_2(a_1 - a_2) - a_1 + c_2^2(a_1 - a_2)) \\
+ a_1 - c_1^2(a_1 - a_2),
\]

which is the reproducing kernel for \( \mathcal{H}_R^{P,B} \) with inner product

\[
(f, g)_{\mathcal{H}_R^{P,B}} = (f, g)_P \\
+ f(0)g(0) \left[ (1 - c_1^2)/a_1 + c_1^2/a_2 - 1 \right] \\
+ (f(0)g(1) + f(1)g(0)) \left[ c_1 c_2 (1/a_1 - 1/a_2) - 1 + 8c_1^2 - 8c_1^4 \right] \\
+ f(1)g(1) \left[ c_1^2/a_1 + (1 - c_1^2)/a_2 - 1 \right].
\]

We’ve assumed \( a_1, a_2 > 0 \); other formulas exist as well.
The Components of a Generalized Sobolev Space Kernel

These results suggest that, because of the choice of \( \mathcal{P}, \mathcal{B} \) and the design of the \( \partial \Omega \) inner product, the Native space inner product must be of the form

\[
(f, g)_{\mathcal{H}^R_{\mathcal{P}, \mathcal{B}}} = (f, g)_{\mathcal{P}} + p_1 f(0)g(0) + p_2 f(0)g(1) + f(1)g(0)) + p_3 f(1)g(1).
\]

At this point, we may choose \( p_1, p_2, p_3 \) and solve

\[
\begin{align*}
(1 - c_1^2)/a_1 + c_1^2/a_2 - 1 &= p_1 \\
c_1 c_2 (1/a_1 - 1/a_2) - 1 + 8c_1^2 - 8c_1^4 &= p_2 \\
c_1^2/a_1 + (1 - c_1^2)/a_2 - 1 &= p_3 \\
c_1^2 + c_2^2 &= 1
\end{align*}
\]

Note: Not all possible \( p_i \) values are acceptable.
Samples of Generalized Sobolev Space Kernel

\[ (f, g)_{H^R_{\mathcal{P}, \mathcal{B}}} = (f, g)_{\mathcal{P}} + 2.875f_0g_0 - 0.471(f_0g_1 + f_1g_0) + 1.125f_1g_1, \]

\[ K_1(x, z) = \min(x, z) - 0.129xz - 0.3261(x + z) + 0.2656. \]

\[ (f, g)_{H^R_{\mathcal{P}, \mathcal{B}}} = (f, g)_{\mathcal{P}} + 7f_0g_0 + 1(f_0g_1 + f_1g_0) + 2f_1g_1, \]

\[ K_2(x, z) = \min(x, z) - 0.3598xz - 0.2055(x + z) + 0.1376. \]
Samples of Generalized Sobolev Space Kernel

\[(f, g)_{H^{R_3}} = (f, g)_P + 100f_0g_0 + 101(f_0g_1 + f_1g_0) + 102f_1g_1,\]

\[K_3(x, z) = \min(x, z) + 3.3171xz - 2.1692(x + z) + 1.1006.\]

\[(f, g)_{H^{R_4}} = (f, g)_P + .001f_0g_0 + .002(f_0g_1 + f_1g_0) + .0005f_1g_1,\]

\[K_4(x, z) = \min(x, z) + .998xz - .9988(x + z) + .9999.\]
Numerical Example of Interpolation

For this simple example, there is not enough freedom to see any difference between interpolants when boundary values are provided:

\[ N = 10 \text{ evenly spaced points}, \quad f(x) = \sin(2\pi x) + x. \]
Reference of GSS Kernels in Use

\[(f, g)_{H^R_1} = (f, g)_P + 2.875f_0g_0 - 0.0471(f_0g_1 + f_1g_0) + 1.125f_1g_1,\]

\[K_1(x, z) = \min(x, z) - 0.129xz - 0.3261(x + z) + 0.2656.\]

\[(f, g)_{H^R_2} = (f, g)_P + 7f_0g_0 + 1(f_0g_1 + f_1g_0) + 2f_1g_1,\]

\[K_2(x, z) = \min(x, z) - 0.3598xz - 0.2055(x + z) + 0.1376.\]

\[(f, g)_{H^R_3} = (f, g)_P + 100f_0g_0 + 101(f_0g_1 + f_1g_0) + 102f_1g_1,\]

\[K_3(x, z) = \min(x, z) + 3.3171xz - 2.1692(x + z) + 1.1006.\]

\[(f, g)_{H^R_4} = (f, g)_P + 0.001f_0g_0 + 0.002(f_0g_1 + f_1g_0) + 0.0005f_1g_1,\]

\[K_4(x, z) = \min(x, z) + 0.998xz - 0.9988(x + z) + 0.9999.\]
Numerical Example of Interpolation **Without** Boundary Data

If boundary data is removed, we see some difference in the interpolants, attributable to the design of the inner product.

\[ N = 10 \text{ evenly spaced points (no boundary)}, \quad f(x) = \sin(2\pi x) + x. \]
Boundary Conditions are a Pain

Not every choice of $p_1, p_2, p_3$ in

$$(f, g)_{\mathcal{H}_P^R, B} = (f, g)_P + p_1 f(0)g(0) + p_2 (f(0)g(1) + f(1)g(0)) + p_3 f(1)g(1).$$

allows for interpolation in the full Sobolev space $\mathcal{H}_P, B = \{ f | (f, f)_P, (f, f)_B < \infty \}$.

For instance, if we work with $p_1 = p_2 = p_3 = 0$, then $K(x, z) = \min(x, z)$ and only functions with $f(0) = 0$ are in the Native space, as discussed by Dr. Erickson last week.

Similarly, if we choose $p_1 = p_3 = 1$ and $p_2 = 0$, then $K(x, z) = \min(x, z) - xz + 1/2$ and only periodic functions are in the Native space.
Expanding the Inner Product Structure

Some of you may find the structure

$$(f, g)_{\mathcal{H}_P^R} = (f, g)_P + p_1 f(0)g(0) + p_2 (f(0)g(1)+f(1)g(0)) + p_3 f(1)g(1).$$

of the inner product rather limiting.

The involvement of the $P$ operator is fixed and to change the interaction between derivatives throughout the domain we would need to change $P$.

It would be possible to change the structure of the boundary contributions, though this would require a low-level change to

$$(f, g)_{\partial\Omega} = f(0)g(0) + f(1)g(1).$$
Alternate Boundary Inner Products

The definition of the boundary inner product could be modified to, e.g.,

\[(f, g)_{\partial \Omega} = f(0)g(0) + \sigma^2 f(1)g(1)\]

\[(f, g)_{\partial \Omega} = f(0)g(1) + f(1)g(0)\]

\[(f, g)_{\partial \Omega} = f(a)g(a) + \sigma^2(f(0)g(0) + f(1)g(1)), \quad a \in (0, 1)\]

\[(f, g)_{\partial \Omega} = f(0)g(0) + f(1)g(1) + \sigma^2(f'(0)g'(0) + f'(1)g'(1))\]

\[(f, g)_{\partial \Omega} = \int_0^1 f(x) \, dx \int_0^1 g(x) \, dx + \sigma^2(f(0)g(0) + f(1)g(1))\]

Each of these choices would yield different \(\psi\) functions to form

\[R(x, z) = \sum_{k=1}^{n_a} a_k \psi_k(x) \psi_k(z),\]

and different structure in the Native space inner product

\[(f, g)_{\mathcal{H}_{\mathcal{P}, \mathcal{B}}^{\mathcal{R}}} = (f, g)_{\mathcal{P}} + \sum_{k=1}^{n_a} \frac{\hat{f}_k \hat{g}_k}{a_k} - \sum_{i=1}^{n_a} \sum_{j=1}^{n_a} \hat{f}_k \hat{g}_k(\psi_i, \psi_j)_{\mathcal{P}},\]

where \(\hat{f}_k = (f, \psi_k)_{\mathcal{B}}\) are Fourier-type coefficients.