Published in Image Processing On Line on 2026–02–13.
Submitted on 2022–12–12, accepted on 2023–01–19.
ISSN 2105–1232 © 2026 IPOL & the authors CC–BY–NC–SA
This article is available online with supplementary materials,
software, datasets and online demo at
https://doi.org/10.5201/ipol.2026.451

Thin-plate Splines on the Sphere for Interpolation, Computing Spherical Averages, and Solving Inverse Problems

Max Dunitz

Centre Borelli, Université Paris-Saclay, Université Paris Cité, ENS Paris-Saclay, CNRS, SSA, INSERM, Gif-sur-Yvette, France
Advanced Track and Trace, Rueil-Malmaison, France
max.dunitz@ens-paris-saclay.fr

Communicated by Jean-Michel Morel and Miguel Colom *Demo edited by* Max Dunitz

Abstract

In many applications, planar spline interpolations of scattered data on the sphere are unsatisfactory; spherical splines are desired. Wahba (1981) defined the thin-plate splines on the sphere by analogy with the polynomial splines on the circle and the thin-plate splines in $\mathbb{R}^d$. The thin-plate spline fit to a scattered data set on the sphere is the solution to an empirical risk minimization problem that penalizes the infidelity of the fit to the data as well as its "wiggliness". This latter term is the square of a seminorm penalty based on the Laplace-Beltrami operator. The minimization problem is posed in a reproducing kernel Hilbert space (RKHS) of functions of finite wiggliness, whose reproducing kernel is isotropic and, due to a result by Schoenberg (1942), given by a Legendre series. A closed-form expression (in terms of the polylogarithm) for the kernel was found by Wendelberger (1982) and re-discovered by Beatson and zu Castell (2018). These closed-form expressions make not just spline interpolation but also downstream signal-processing tasks, such as cubature or resolution of inverse problems, more tractable in fields where scattered data and spherical models are common, such as remote sensing, geostatistics, motion planning, graphics, and medical imaging. In this paper, we present a tutorial on spline methods in RKHSs and show how they can be used to interpolate, smooth, and numerically integrate scattered data on the sphere and solve related inverse problems. The accompanying demo compares thin-plate spline interpolation over the sphere with thin-plate splines on an equirectangular projection and natural cubic splines on a one-dimensional latitudinal projection used in greenhouse gas monitoring. Global mean values of the interpolation surfaces are presented as well, to illustrate how this isotropic spherical kernel–which penalizes interpolant wiggliness without concern for application-specific factors like atmospheric winds–affects the computation of global averages.

Source Code

A Python implementation of the algorithms described in this article is available at the associated web page[1]. Usage instructions are included in the `README.txt` file of the archive. The associated online demo is accessible through the web site.

Keywords: spherical signals; geostatistics; cubature; approximation; scattered data; inverse problems; interpolation; RKHS; reproducing kernel Hilbert space; splines; thin-plate splines

[1] https://doi.org/10.5201/ipol.2026.451

1 Introduction

Among signal-processing practitioners, the word "spline" often conjures up a limited set of techniques for interpolating data that are sampled at regular locations (such as images). In the statistics and inverse-problems communities, spline models are used to solve a richer set of variational data-fitting problems, such as interpolating data, smoothing data, and solving inverse problems based on observations of bounded linear functionals. They can be applied in diverse settings, including on compact Riemannian manifolds such as the sphere, or in finite simple graphs. Their flexibility makes them suitable for scattered data applications–that is, on data sets that are irregularly sampled–without the need for gridding the data. While certain less flexible methods for interpolation used in image processing do not generally require a matrix inversion, this relative advantage often vanishes in the context of an inverse-problem processing chain that requires an inversion anyway.

In the reproducing kernel Hilbert space (RKHS, introduced in Section 2.1) framework, splines are not just an element of a signal-processing chain but also a language in which to express the solution to the empirical risk minimization problem the processing chain seeks to resolve. The positive-definite kernel associated with each RKHS model space expresses the similarity between points in the index set (which can be arbitrary). Accordingly, the curve-fitting properties of spline models depend on how the kernel expresses similarity on the index set. Kernels can be defined using the geometric properties of the index set, statistical models expressing similarity (covariance) between elements the data set, or computed features. For problems posed on index sets with geometric structure–such as Euclidean space, compact Riemannian manifolds, and graphs–and in the absence of additional information apart from a preference for smoothness, the thin-plate splines are a natural choice of interpolant. In Euclidean space, they represent the bending energy of a thin sheet in the linear elastic regime (see Section 2.6.4). They are, moreover, based on the Laplacian, which yields desirable invariance properties. The use of the Laplacian, which maps functions to functions and possesses a spectrum that exposes geometric and topological properties of the index set, also lends the approach interpretability and generalizability.

Derived using a smoothness seminorm penalty involving the iterated Laplacian, the iterated Laplace-Beltrami operator of a compact Riemannian manifold, or the iterated Laplacian matrix of a graph, the space of thin-plate splines is equipped with a notion of "wiggliness" that is adapted to the metric of the index set and that possesses the same isometry-invariance properties of the Laplacian. The minimizer of empirical risk, therefore, is a function over the Euclidean space, manifold, or full set of vertices in the graph that minimizes disagreement with the scattered observations as well as this measure of wiggliness.

1.1 What Do We Mean by "Spline Model"?

Spline models are, in general, solutions to an empirical risk minimization problem formulated over a hypothesis space of functions over an index set. These problems penalize disagreement with a set of observations (often scattered pointwise evaluations) and prior knowledge.

In geostatistics, this prior knowledge usually takes the form of a Gaussian process. To each point on the index set (typically identified with time or Euclidean space), we associate a random variable corresponding to the real variable we wish to interpolate. We assume the joint density of any finite set of these variables is Gaussian. The mean of the observed value depends only on the observed location – and is often constant. The covariance of any two random variables assumes a parametric form that depends on the displacements between the two corresponding index locations – for isotropic models, on their Euclidean distance alone. Non-isotropy may be introduced to incorporate knowledge about prevailing winds, ocean currents, and so forth (with care to ensure valid covariance matrices). Specifying the parametric form of the model amounts to choosing a function space and a distribution

thereon. High covariance at small displacements favors smooth interpolants. An interpolating spline is found by conditioning the Gaussian process on the values of noise-free observations; a smoothing spline is found by incorporating a likelihood function modeling observation noise. In either case, the mean value of the posterior distribution at a set of evaluation points is easily computed using linear algebra (see, e.g., [164], Equation 2.23). In the kriging community, the parametric model of covariance is called a variogram or semi-variogram; in the Gaussian process community, a covariance function or kernel. With cokriging, auxiliary data, often sampled at different scattered locations, can lend further credibility to the interpolant. For instance, humidity measurements can enhance the estimates of atmospheric temperature between scattered observations. Introducing prior knowledge requires choosing a parametric family of covariance function (and, for cokriging, cross-covariance functions) and selecting the parameters.

In applications where such prior knowledge is inaccessible or difficult to model accurately, smoothing splines are a common tool. These penalize the wiggliness of the function using the geometry of the index set. Since, by Stokes's theorem, $-\mathrm{div}$ and ∇ are formally adjoint, penalizing the Dirichlet energy–the squared Euclidean norm of the gradient field of the function over the index set–is equivalent to penalizing a Laplacian-based term: $u \cdot \Delta u$

$$\int_{\mathcal{X}} ||\nabla u(x)||^2 \,\mathrm{d}x = \int_{\mathcal{X}} \langle \nabla u(x), \nabla u(x) \rangle \,\mathrm{d}x = \int_{\mathcal{X}} -\mathrm{div}(\nabla u(x)) \cdot u(x) \,\mathrm{d}x = \int_{\mathcal{X}} u(x) \cdot \Delta u(x) \,\mathrm{d}x \,.$$

More generally, the smoothing-spline wiggliness penalty of order m takes the following form:

$$J_{m,\mathcal{X}}(f) = \begin{cases} \int_{\mathcal{X}} (\Delta^{m/2} f(x))^2 \,\mathrm{d}x \,, & \text{if } m \text{ is even;} \\ \int_{\mathcal{X}} ||\nabla(\Delta^{(m-1)/2} f(x))||^2 \,\mathrm{d}x \,, & \text{if } m \text{ is odd.} \end{cases}$$

This penalty can be written, where boundary conditions allow,

$$J_{m,\mathcal{X}}(f) = (-1)^m \int_{\mathcal{X}} f(x) \cdot \Delta^m f(x) \,\mathrm{d}x \,. \tag{1}$$

This sensible notion of wiggliness yields penalties that are invariant to isometries. That polynomials of the Laplacian are translation- and rotation-invariant differential operators in $\mathbb{R}^d$ (indeed, the only ones!) follows easily from the properties of the Fourier transform (see, e.g., [49], Theorem 2.1). The wiggliness seminorm of thin-plate splines in $\mathbb{R}^d$, defined using the iterated Laplacian, is invariant to isometries: the wiggliness of spline interpolants can be defined in terms of a radial basis function that depends only on the Euclidean distances between spline knots. On compact Riemannian manifolds, the Laplace-Beltrami operator, defined using the metric, commutes with isometries (in fact, the only diffeomorphisms that leave the Laplace-Beltrami operator invariant are isometries; see [64], Proposition 2.4). On the sphere in particular, all positive-definite kernels that depend only on the geodesic distance between points have a simple characterization in terms of their expansion on the eigenfunctions of the spherical Laplacian. The wiggliness of a spline interpolant can be given in terms of such a kernel. By a theorem of Schoenberg, expansions of such kernels on these eigenfunctions, the spherical harmonics, must weight equally every harmonic of the same Dirichlet energy; see Section 2.2.3. This isotropy ensures that rotating the sphere will not affect the measured wiggliness of a function thereon. A similar situation arises on graphs: wiggliness penalties based on the iterated Laplacian are invariant to automorphisms (vertex permutations that preserve the edge structure); the Laplace matrix of the transformed graph is permutation-similar to the original Laplace matrix and thus has the same spectrum [57, 139].

The form (1) of the penalty is particularly useful for index sets $\mathcal{X}$ that are compact manifolds like the sphere (and for graphs). Since the eigenfunctions of the Laplace-Beltrami operator form a complete orthonormal system for $L^2(\mathcal{X})$, with corresponding eigenvalues giving the Dirichlet energy

of each mode, we can represent this penalty "on the Fourier side," as an infinite series of wiggliness-weighted Fourier coefficients. The space of functions for which this series converges–the functions with finite wiggliness penalty–is an RKHS. For certain manifolds like the circle and sphere, no truncation of the infinite series is required in practice as the series has a closed-form expression in terms of special functions.

1.2 Natural Cubic Splines

To make things a bit more concrete, let us take a look at one of the best-known examples of these splines, the natural cubic splines. The thin-plate splines in $\mathbb{R}^d$ or on the sphere can be seen as generalizations of these splines. The natural cubic splines live in a space of well-behaved functions that we call $\mathcal{H}$, as it is an RKHS. The space $\mathcal{H}$ can be written as the direct sum of two RKHSs $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$:

- $\mathcal{H}_1$ is an infinite-dimensional space of continuous functions[2] with continuous ordinary derivative and ordinary second derivative that exists almost everywhere and is square-integrable. The squared norm of this space is the wiggliness penalty, designed to enforce smoothness by measuring curvature on some index set $\mathcal{X} = [a, b]$, given by

$$J_{2,\mathcal{X}}(u) = \int_{\mathcal{X}} \left(u^{(2)}(x)\right)^2 \mathrm{d}x\,. \tag{2}$$

  To ensure the wiggliness penalty is a definite norm on $\mathcal{H}_1$, we require that all nonzero functions in $\mathcal{H}_1$ be wiggly: $u \in \mathcal{H}_1$ and $u \neq 0 \implies J_{2,\mathcal{X}}(u) > 0$. In Section 2.6.1, we will use boundary conditions to enforce definiteness.

- $\mathcal{H}_0$ is the finite-dimensional null space of $J_{2,\mathcal{X}}$: $u \in \mathcal{H}_0 \implies J_{2,\mathcal{X}}(u) = 0$. $\mathcal{H}_0$ contains the functions that are sufficiently well-behaved to live in $\mathcal{H}$ and sufficiently non-wiggly to be assigned 0 by $J_{2,\mathcal{X}}$, which is a seminorm on $\mathcal{H}$. In Section 2.6.1, we will see that $\mathcal{H}_0$ is the space of affine functions on $\mathcal{X}$.

We will give precise definitions of the spaces $\mathcal{H}, \mathcal{H}_0$, and $\mathcal{H}_1$ associated with the natural cubic splines in Section 2.6.1. The natural cubic splines are functions $\sigma \in \mathcal{H}$ that solve an empirical risk minimization problem. Specifically, for each data set $(\{(x_i, y_i)\}_{i=1}^n)$ with $x_i \in \mathcal{X}$ and $y_i \in \mathbb{R}$ and for each choice of regularization parameter $\lambda > 0$, the associated natural cubic smoothing spline is the function $\sigma \in \mathcal{H}$ that minimizes, over $\mathcal{H}$, the empirical risk

$$R_{2,\mathcal{X},\lambda}(u) = \underbrace{\frac{1}{n}\sum_{i=1}^{n} (u(x_i) - y_i)^2}_{\text{adherence to training data}} + \lambda \underbrace{J_{2,\mathcal{X}}(u)}_{\substack{\text{wiggliness} \\ \text{penalty}}}.$$

By the Wahba-Kimeldorf representer theorem [79, 131, 159] (see Section 2.5), the spline may be written sparsely, or at least in a finite manner, as follows:

$$\sigma = \sum_{j=1}^{\dim \mathcal{H}_0} \alpha_j \phi_j + \sum_{i=1}^{n} \beta_i k_{x_i},$$

[2]The elements of $\mathcal{H}_1$ are equivalence classes of functions that agree almost everywhere. Here, by a Sobolev embedding theorem, we can choose a unique representer for $\mathcal{H}_1$ that is absolutely continuous, has absolutely continuous (ordinary) derivative, and has an (ordinary) second derivative that is defined almost everywhere and square integrable. See Theorem 129 of [17] and Theorem 10.45 of [162].

where the ϕ_j are a basis for the finite-dimensional RKHS $\mathcal{H}_0$ and the k_{x_i} are the *representers of evaluation* at the scattered data, that is, the Riesz representations of the evaluation functionals at the x_i. (RKHSs are Hilbert spaces on which all evaluation functionals–that is, linear functionals on $\mathcal{H}_1$ associated with a point $x \in \mathcal{X}$ that map a function f to its pointwise evaluation $f(x)$–are bounded and therefore have Riesz representations.) By the *kernel trick*, we do not need direct access to the representers of evaluation; our interaction with them is mediated by the Gram matrix of their inner products $(\mathbf{K})_{i,j} = \langle k_{x_i}, k_{x_j} \rangle_{\mathcal{H}_1}$. We will recall these details in Section 2.1.

1.3 Thin-plate Splines Are Generalizations of the Natural Cubic Splines

Natural cubic splines and their planar equivalent, the thin-plate splines on the plane, were first introduced using integration by parts. The kernel can be computed using the Green's function of the Laplacian. Since the Green's function of the planar Laplacian would vary with the geometry of any choice of bounded domain over which the Laplacian-based wiggliness penalty is applied, typically $\mathcal{X}$ is set to all of $\mathbb{R}^d$, with the constraint that $m > d/2$.

We rewrite the penalty (2) to place it in a form that is consistent with the thin-plate splines in Euclidean d-space. With appropriate boundary conditions[3], two integrations of (2) by parts yield

$$J_{2,\mathcal{X}}(u) = \int_{\mathcal{X}} u(x) \cdot u^{(4)}(x) \, \mathrm{d}x = \int_{\mathcal{X}} u(x) \cdot \Delta^2 u(x) \, \mathrm{d}x \,, \tag{3}$$

where $\Delta = \frac{\mathrm{d}^2}{\mathrm{d}x^2}$ is the Laplacian operator on $\mathbb{R}$. The iterated Laplacian operator contributes similarly to the definition of thin-plate splines in Euclidean space $\mathbb{R}^d$. On the sphere, the iterated Laplace-Beltrami operator Δ_S plays the part of the Laplacian in defining thin-plate splines. This operator is the restriction of the Euclidean Laplacian to the surface of the sphere; its isometry invariance can be established without results from differential geometry by appealing to the isometry invariance of the iterated Euclidean Laplacian and the restriction operator; see [68], Chapter 3.1, or [52], page 5. By analogy, splines can be defined on graphs using the Laplace matrix; the wiggliness penalty is invariant to edge-preserving vertex relabelings, which leave the Laplace matrix's spectrum alone. Our generalizations of the natural cubic splines to $\mathbb{R}^d$ use wiggliness penalties of the form

$$J_{m,\mathcal{X}}(u) = (-1)^m \int_{\mathcal{X}} u(x) \cdot \Delta^m u(x) \, \mathrm{d}x \,,$$

which have finite-dimensional null space. On a restricted space of functions $\mathcal{H}_1$ of nonzero wiggliness, these penalties are definite, and induced by a definite inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}_1}$ given by

$$\langle f, g \rangle_{\mathcal{H}_1} = (-1)^m \int_{\mathcal{X}} f(x) \Delta^m g(x) \, \mathrm{d}x \,.$$

We can use the Green's function $E_m(x,t)$ of the m-iterated Laplacian, which satisfies

$$\Delta^m E_m(t,x) = \delta(t-x),$$

to establish the *reproducing property*

$$f(t) = (-1)^m \int_{\mathcal{X}} f(x) \cdot \underbrace{\Delta^m E_m(t,x)}_{\delta(t-x)} \mathrm{d}x = \langle f, E_m(t,\cdot) \rangle_{\mathcal{H}_1},$$

[3] In defining the natural cubic splines–and, more generally, the natural polynomial splines (see [17], Theorem 68)–our model space $\mathcal{H}_1$ will be defined using boundary conditions $u(x_1) = u(x_n) = \ldots = u^{(m-1)}(x_1) = u^{(m-1)}(x_n) = 0$, which do not necessarily permit the integration by parts to rewrite (2) as (3). However, it turns out that the functions in the model space that minimize (2) also satisfy the natural, or Neumann, boundary conditions and are linear beyond the scattered data: if $\mathcal{X} = [x_1, x_n]$, requiring that $u^{(n)}(x_1) = u^{(n)}(x_n) = 0$, for $n = m, m+1, \ldots, 2m-1$, does not change the solution [158].

via the "sifting" property of the Dirac δ. Taking the inner product of a function f with a Green's function $E_m(t, \cdot)$ of the m-iterated Laplacian, evaluated at one argument t, evaluates f at t; the Green's function acts as a Riesz representation of evaluation at t in this restricted space of functions and can be used to compute a reproducing kernel. We can then extend the solution to the direct sum of $\mathcal{H}_1$ and $\mathcal{H}_0$, the null space of $J_{m,\mathcal{X}}$. We cover this approach in detail in Section 2.6.

While the natural polynomial splines on $\mathbb{R}$ and the thin-plate splines on $\mathbb{R}^d$ are introduced using a Green's function, thin-plate splines on compact manifolds such as the sphere and circle are derived "on the Fourier side" using the eigenfunctions of the Laplace-Beltrami operator.

1.4 Certain Thin-Plate Splines Are Derived on the "Fourier Side"

It is well-known (see, e.g., [123], Theorem 1.29) that for any compact connected oriented Riemannian manifold $\mathcal{X}$, there exists a complete orthonormal basis of $L^2(\mathcal{X})$ consisting of eigenfunctions $\{\phi_n\}_{n=1}^{\infty}$ of the Laplace-Beltrami operator whose eigenvalues $\{\lambda_n\}_{n=1}^{\infty}$ are nonnegative, each with finite multiplicity, accumulating only at infinity. Zero is an eigenvalue of multiplicity one, whose associated eigenspace consists of the constant functions, i.e., span $\{1\}$. This holds on a manifold without boundary (as is the case for the sphere), or with boundary, provided we impose the Neumann or Dirichlet boundary conditions. In Section 2.6.2, we define the polynomial splines on the circle as splines on the compact interval $[0, 1]$ using Dirichlet boundary conditions.

Moreover, each eigenvalue λ_n of the Laplace-Beltrami operator gives the Dirichlet energy (a common measurement of wiggliness) of the corresponding eigenfunction ϕ_n. Expanding a function in $f \in L^2(\mathcal{X})$ as a Fourier series on this basis,

$$f \sim \sum_{n=1}^{\infty} (f)_n \phi_n \qquad \text{(where convergence is in the } L^2(\mathcal{X}) \text{ norm)},$$

we can, using the orthonormality of the basis functions, write a seminorm penalty like

$$J_{m,\mathcal{X}}(f) = (-1)^m \int_{\mathcal{X}} f(x) \cdot (\Delta^m f)(x) \, \mathrm{d}x$$

in series form

$$\begin{aligned} J_{m,\mathcal{X}}(f) &= (-1)^m \int_{\mathcal{X}} f(x) \cdot \Delta^m f(x) \, \mathrm{d}x = (-1)^m \int_{\mathcal{X}} \left(\sum_{n=1}^{\infty} (f)_n \phi_n(x) \right) \cdot \left(\sum_{n=1}^{\infty} (f)_n \lambda_n^m \phi_n(x) \right) \mathrm{d}x \\ &= \sum_{n=1}^{\infty} (f)_n^2 \lambda_n^m \underbrace{\int_{\mathcal{X}} \phi_n(x)^2 \, \mathrm{d}x}_{1} \\ &= \sum_{n=1}^{\infty} (f)_n^2 \lambda_n^m. \end{aligned}$$

The series converges whenever the Fourier components $\{(f)_n\}_{n=1}^{\infty}$ decay sufficiently quickly to overcome the weighting the iterated Laplace-Beltrami operator places on the wiggly, high-Dirichlet-energy (high-λ_n) components. In such situations, as we will see with the thin-plate splines on the circle and on the sphere, the functions of finite wiggliness penalty are so well-behaved that they can be evaluated pointwise and constitute an RKHS. The Fourier expansion $f \sim \sum_{n=1}^{\infty} (f)_n \phi_n$ in fact converges pointwise for any spline of order $m > 0$.

We develop these ideas more explicitly and rigorously in Section 2.2. The process of constructing an RKHS and corresponding reproducing kernel by penalizing the functions' components on certain the eigenfunctions is described in Proposition 2.38. To see how weighting spherical harmonics by their Dirichlet energy yields an isotropic kernel, see Section 2.2.3.

1.5 Outline

The remainder of this article is organized as follows. In the next section, we give a tutorial on RKHSs and on solving certain variational problems in these spaces–problems where the regularization penalty has a null space of finite dimension. The solutions to these problems are what we call splines. We develop the motivation given by Wahba for the thin-plate splines on the sphere [154, 158] in part by enumerating their kin, the splines that belong to the family of, in the words of Duchon, "splines minimizing rotation-invariant seminorms in Sobolev spaces"–the thin-plate splines [40].

Those interested only in implementation details of thin-plate splines on the sphere may safely skip to Section 3, where the kernel for thin-plate splines on the sphere is given, along with pseudocode for finding the thin-plate spline interpolants of scattered data on the sphere. In Section 4, we stress that the (nearly) closed-form expressions for the reproducing kernels of the thin-plate splines on the sphere allow us not just to learn interpolants of scattered data, but also to estimate the values of a linear functional applied to an unknown continuous function on the sphere from its scattered samples by applying the functional to the interpolant. More generally, we can solve inverse problems where our scattered data need not be mere (possibly noisy) *evaluations* of an unknown function on the sphere, but can be *measurements* produced by arbitrary bounded linear measurement functionals. As an illustration, we estimate the global mean of the atmospheric CO_2 concentration based on scattered measurements. This example can be run in the IPOL demo, which is described in Section 5. Finally, in Section 6, we provide some brief discussion with pointers to extensions and other implementations of the thin-plate splines on the sphere in the literature.

2 Thin-plate Splines on the Sphere: An Overview of the Mathematical Background

The solutions to norm-minimization problems that arise in approximation and inference contexts, such as

$$\arg\min_{x\in\mathbb{R}^n} ||\mathbf{A}x - b||_{\mathbb{R}^m},$$

can often be characterized with local criteria, since norm objectives are convex. When formulated in a Hilbert space $\mathcal{H}$, these local criteria take the form of orthogonality constraints. In such cases, the norm objectives $\|\cdot\|$ are induced by an inner product $\langle\cdot,\cdot\rangle = \|\cdot\|^2$. Squaring each norm objective term $||\cdot||$, we arrive at a problem that is equivalent to our original problem (when the objective consists of a single term) or that serves as a tractable proxy for it[4] (when the objective consists of multiple summed norm terms). In our example, we get the following reformulation

$$\arg\min_{x\in\mathbb{R}^n} ||\mathbf{A}x - b||^2_{\mathbb{R}^m} = \arg\min_{x\in\mathbb{R}^n} \langle \mathbf{A}x - b, \mathbf{A}x - b\rangle_{\mathbb{R}^m},$$

In this case, the objective's (Fréchet) derivative is proportional to the inner product[5]. The derivative, taken at x, is the following bounded linear functional

$$\nabla_x ||\mathbf{A}x - b||^2_{\mathbb{R}^m} = v \mapsto 2\langle \mathbf{A}x - b, \mathbf{A}v\rangle_{\mathbb{R}^m} = 2\langle \mathbf{A}^T(\mathbf{A}x - b), v\rangle_{\mathbb{R}^n}.$$

[4]For example, the problem of finding a k-dimensional subspace of a Euclidean space $\mathbb{R}^d$ that approximates a data set in $\mathbb{R}^d$ by minimizing its sum of *squared* Euclidean residual norms after orthogonal projections is solved by taking the span of the first k principal components found by Principal Components Analysis (PCA). This problem is an example of *benign non-convexity* as the problem is formulated over the non-convex manifold of subspaces, the Grassmannian. However, replacing the sum of squared residual Euclidean norms with a sum of residual Euclidean norms eliminates the benignity in the worst case (the problem becomes the **NP**-complete 2-1 norm matrix approximation problem [98]).

[5]In any *real* Hilbert space, the Fréchet derivative of the map $x \mapsto ||x||^2_{\mathcal{H}}$, evaluated at $x \in \mathcal{H}\setminus\{0\}$, is the bounded linear functional on $\mathcal{H}'$ given by $h \mapsto 2\langle h, x\rangle_{\mathcal{H}}$. Hilbert spaces over $\mathbb{C}$ require greater care.

As a result, norm-minimization problems reduce to an orthogonality constraint (called the normal equations) via the first-order condition[6]

$$\text{find } x \text{ such that } \nabla_x ||\mathbf{A}x - b||^2_{\mathbb{R}^m} \equiv 0 \iff \forall v \in \mathbb{R}^n, \langle \mathbf{A}^T(\mathbf{A}x - b), v\rangle_{\mathbb{R}^n} = 0 \iff \mathbf{A}^T(\mathbf{A}x - b) = 0.$$

In Euclidean space, these equations $\mathbf{A}^T\mathbf{A}x = \mathbf{A}^T b$ always have a solution. Indeed, a solution always exists when we replace the $m \times n$ matrix $\mathbf{A}$ with an arbitrary bounded linear map A from any Hilbert space $\mathcal{H}_1$ to any Hilbert space $\mathcal{H}_2$, provided the range of A is closed in $\mathcal{H}_2$ (i.e., if A is bounded below): we write $\mathrm{A}^T\mathrm{A}x = \mathrm{A}^T b$, where A^T is the adjoint of A (see [90], Section 6.9). In effect, this is the Hilbert space projection theorem applied to the range of A; the solution applies to b the Moore-Penrose pseudoinverse of A[7].

When optimizing over spaces of large or infinite dimensions, we can still struggle to express the solution (or a Cauchy sequence that converges to it quickly) on a computer. However, if we know the solution lies in a finite-dimensional vector or affine subspace, or in the orthogonal complement of a finite dimensional space[8], we can write the solution using finite-dimensional linear algebra.

Using the theory of reproducing kernel Hilbert spaces, we can take full advantage of these results about norm-minimization problems in Hilbert spaces to solve interpolation problems. In a reproducing kernel Hilbert space $\mathcal{H}$, a constraint set consisting of a finite set of pointwise evaluation equalities, as in the interpolation problem

$$\arg\min_{f \in \mathcal{H}} ||f||^2_{\mathcal{H}} \text{ subject to } f(x_1) = a_1, \ldots, f(x_n) = a_n,$$

is an affine subspace that can be expressed in terms of the inner product $\langle f, k(\cdot, x_i)\rangle_{\mathcal{H}} = a_i$, where k is the kernel associated with the space. The solution to the interpolation problem can be found by inverting the Gram matrix $(\mathbf{K})_{i,j} = k(x_i, x_j)$ associated with the kernel. Even if we relax the interpolation problem into an empirical risk minimization problem

$$\arg\min_{f \in \mathcal{H}} \sum_{i=1}^{n} \text{loss}(f(x_i), a_i) + \lambda ||f||^2_{\mathcal{H}},$$

with arbitrary loss (see [131], Theorem 1) or replace the function evaluations $f(x_i)$ with other bounded linear functionals, the solution remains a finite-dimensional linear algebra problem, even if $\mathcal{H}$ is infinite-dimensional. We can see that any minimizer of the above loss must lie in span $\{k(\cdot, x_i)\}_{i=1}^n$, as projecting any potential solution onto this finite-dimensional subspace cannot affect the data-adherence loss term but can reduce the wiggliness penalty $||f||^2_{\mathcal{H}}$. This striking result–that the minimizer over a large or infinite-dimensional space of a norm-minimization problem lies in a finite-dimensional space spanned by what are called the Riesz representations or "representers" of evaluation at the data locations–is known as the representer theorem (see Section 2.5). It was introduced by Grace Wahba and George Kimeldorf in the context of L-splines [79, 80], which include the splines in this article as a special case. In short, for norm-minimization problems in Hilbert spaces, completeness and convexity guarantees us the solution's existence; Hilbert space theory allows us to characterize the solution in terms of inner products; and RKHS theory gives us an expression for the solution in terms of the Gram matrix of the kernel. It is the last step that (in theory) brings tractability.

[6] Since the objective is convex and the set over which we are optimizing is convex, the Euler inequality is necessary and sufficient for a global optimum. Since we are optimizing over the space $\mathcal{H}$ itself, every point is an interior point, and the Euler inequality becomes the first-order condition.

[7] Which always exists in Hilbert spaces when the range of A is closed and, more generally, for von Neumann-regular operators in C^*-algebras [62].

[8] After all, minimizing the distance (induced by the Hilbert norm) between a function f and a Hilbert subspace M–i.e., finding its orthogonal projection f_M on M–is equivalent to maximizing the alignment between the difference vector $f - f_M$ and any vector in $M^\perp$.

This section is organized as follows. In Section 2.1, we review RKHS theory. In Section 2.2, we restrict our attention to index sets that are closed regions in Euclidean space and synthesize kernels for such index sets using an ℓ^1 sequence and a complete orthonormal system for the index set; we give particular attention to the sphere, on which we characterize all positive-definite functions that are isotropic. Then in 2.3, we return to a general context, where index sets can be arbitrary, and show how norm-minimization problems like interpolation and smoothing can be solved in an RKHS using finite-dimensional linear algebra, for which pseudocode is provided. Sections 2.4-2.5 are devoted to translating these results to the case of seminorm-minimization problems, where the seminorm has null space of finite dimension. Finally, in 2.6, we introduce the thin-plate splines over different index sets, including the sphere, as solutions to seminorm-minimization problems.

2.1 Reproducing Kernel Hilbert Spaces (RKHSs)

Much of the material in this section is standard and may be found, for instance, in [4, 17, 78, 93, 95, 132, 134, 141].

2.1.1 RKHS Basics

A Hilbert space[9] is an inner product space that is complete with respect to the norm induced by its inner product. The theory of reproducing kernel Hilbert space depends on the Riesz representation theorem, which identifies an isometric isomorphism between a Hilbert space $\mathcal{H}$ and its ("continuous" or "topological") dual space $\mathcal{H}'$, that is, the space of bounded linear functionals from $\mathcal{H}$ to a complete field (which we take to be $\mathbb{R}$, rather than $\mathbb{C}$). Through this isomorphism, any bounded linear functional on $\mathcal{H}$ can be expressed as an inner product between the input and a fixed element of $\mathcal{H}$, often called the representer of the functional.

Theorem 2.1 (Riesz-Fréchet representation theorem)**.** *Let* $\mathrm{E} : \mathcal{H} \to \mathbb{R}$ *be a linear functional on a Hilbert space* $\mathcal{H}$*. Suppose that* E *is bounded (or, equivalently, since it is linear, continuous). That is, suppose there is a number* $M > 0$ *such that, for all* $u \in \mathcal{H}$*, we have that*

$$||\mathrm{E}u||_{\mathbb{R}} \leq M||u||_{\mathcal{H}}.$$

Then there exists a unique element η_{E} *in* $\mathcal{H}$*, called a representer of the functional* E*, such that, for all* $u \in \mathcal{H}$*,*

$$\mathrm{E}u = \langle u, \eta_{\mathrm{E}} \rangle_{\mathcal{H}};$$

moreover, $||\eta_{\mathrm{E}}||_{\mathcal{H}} = ||\mathrm{E}||_{\mathcal{H}'}$.

Proof. See, for example, [36], Theorem 3.7.7. □

Remark 2.2. *This theorem tells us that every bounded linear functional has a representer. Its proof involves the construction of a linear*[10] *isometric isomorphism that maps the bounded functional* E *to its representer* η_{E}*. Conversely, every* $u \in \mathcal{H}$ *is a representer of the bounded linear functional* $\mathrm{E}_u = \cdot \mapsto \langle \cdot, u \rangle_{\mathcal{H}}$.

In a reproducing kernel Hilbert space (RKHS), evaluation functionals have representers.

[9] While we work with real Hilbert spaces, the key results presented here all generalize to complex Hilbert spaces when conjugated accordingly, except where stated otherwise. In proofs, replace words like "bilinear" with "sesquilinear", "symmetry" with "conjugate symmetry", and so forth.

[10] Antilinear (conjugate-linear) if we take the field to be $\mathbb{C}$.

Definition 2.3 (RKHS: when you're here, your evaluation functionals are bounded)**.** *A reproducing kernel Hilbert space (RKHS) is a Hilbert space $\mathcal{H} \subseteq \mathbb{R}^{\mathcal{X}}$ over an index set $\mathcal{X}$ such that, for all $x \in \mathcal{X}$ the evaluation functional at x*

$$\begin{aligned} \mathrm{E}_x : \mathcal{H} &\to \mathbb{R} \\ u &\mapsto u(x), \end{aligned}$$

is a bounded linear functional.

Remark 2.4. *This definition tells us that as functions in an RKHS approach each other in the RKHS norm, their pointwise evaluations approach each other as well. Indeed, given an RKHS $\mathcal{H}$ with inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ and induced norm $|| \cdot ||_{\mathcal{H}}$, we know that, since for all $x \in \mathcal{X}$ the evaluation functional at x, E_x, is bounded, by the Riesz representation theorem, there is a unique representer $k_x = \eta_{\mathrm{E}_x}$ of E_x such that, for all $f \in \mathcal{H}$*

$$f(x) = \langle f, k_x \rangle_{\mathcal{H}}.$$

Using the Cauchy-Schwarz inequality, we can confirm that a sequence of functions $\{f_n\}_{n=0}^{\infty}$ that converges in $\mathcal{H}$ to a function f also converges pointwise at every $x \in \mathcal{X}$

$$\text{for all } \epsilon > 0,\ ||f_n - f||_{\mathcal{H}} < \delta_x = \frac{\epsilon}{||k_x||_{\mathcal{H}}} \implies |f_n(x) - f(x)| = |\langle f_n - f, k_x \rangle_{\mathcal{H}}| \leq ||f_n - f||_{\mathcal{H}} ||k_x||_{\mathcal{H}} < \epsilon.$$

If there exists an M for which, for all $x \in \mathcal{X}$, $||k_x||_{\mathcal{H}} \leq M$, then this convergence is uniform.

Pointwise convergence is assured in an RKHS even if f_n converges to f only weakly in $\mathcal{H}$.

$$\text{for all } h \in \mathcal{H},\ \langle f_n, h \rangle_{\mathcal{H}} \longrightarrow \langle f, h \rangle_{\mathcal{H}}$$

implies that, in particular, for all $x \in \mathcal{X}$,

$$f_n(x) = \langle f_n, k_x \rangle_{\mathcal{H}} \longrightarrow \langle f, k_x \rangle_{\mathcal{H}} = f(x).$$

Before we present an example of an RKHS, let us recall the following definition of a Sobolev space of positive integral order.

Definition 2.5 (Sobolev space of positive integral order m)**.** *Let $\mathcal{X}$ be a bounded interval of the real line. The Sobolev space $W^{m,2}(\mathcal{X}) = \{u \in \mathcal{D}'(\mathcal{X}) \mid u^{(i)} \in L^2(\mathcal{X}) \text{ for } i = 0, 1, \ldots, m\}$, where $u^{(i)}$ is the ith weak (distributional) derivative. We can simplify this definition by noting (see [17], Theorem 129) that any distribution is in $W^{m,2}(\mathcal{X})$ if and only if it has a unique representer u that satisfies the following:*

1. *its ordinary derivatives $u^{(i)}$ are absolutely continuous and square integrable on $\mathcal{X}$ for $i = 0, \ldots, m-1$;*
2. *its ordinary derivative $u^{(m)}$ is defined almost everywhere and is square integrable on $\mathcal{X}$.*

Note that the spline literature uses nonstandard definitions of Sobolev norms[11]. The classical inner product given to a Sobolev space is

$$\langle f, g \rangle_{\mathcal{H}} = \sum_{i=0}^{m} \langle f^{(i)}, g^{(i)} \rangle_{L^2(\mathcal{X})}.$$

[11] For an open, bounded subset of $\mathbb{R}$ such as an interval (a, b), the space of functions assigned a finite value by this norm–the Beppo Levi norm–coincides algebraically with the Sobolev spaces because the Poincaré identity tells us these norms are equivalent on the smooth test functions compactly supported on this subset [37]. Splines derived from wiggliness penalties corresponding to standard Sobolev 2-norms on $\mathcal{X} = \mathbb{R}^d$, that is, $||u||_{H_m} = \sum_{i=1}^{m} ||u^{(i)}||^2_{L^2(\mathcal{X})}$ (with $m > d/2$) are called Matérn kernels and can be expressed in terms of the modified Bessel function of the second kind. A variety of other Sobolev-like penalties have been considered for constructing splines such as the splines *with tension* that minimize a difference between the $m+1$-iterated Laplacian and the weighted m-iterated Laplacian [21].

However, the wiggliness penalties of thin-plate splines are symmetric bilinear forms that exclude the first $m-1$ derivatives from the penalty

$$\langle f^{(m)}, g^{(m)} \rangle_{L^2(\mathcal{X})}, \tag{4}$$

which does not possess definiteness on the model spaces we care about, such as the Sobolev space of order 2 on $\mathcal{X} = [0,1]$ (the model space for the natural cubic splines). We call any such semidefinite bilinear form a semi-inner product or an indefinite inner product. In Section 2.6.1, we will use the decomposition principle (see Section (2.4)) to complement the indefinite inner product with an inner product over its null space, rendering it definite. The Sobolev spaces of integral order $m \geq 1$ are RKHSs [17, 100] with this extension to the inner product (4), as we will see in Section 2.6.1.

The first example of an RKHS we consider is the Sobolev space of order $m = 1$ on $\mathcal{X} = [0,1]$. We impose boundary conditions on the space so as to make (4) strictly definite over the space. In Section 2.6.1, we show how to remove these boundary conditions by using the decomposition principle to complement the indefinite inner product (4) with an inner product on its null space.

Example 2.6. *Let $\mathcal{H}$ be the Sobolev space of absolutely continuous functions $f : [0,1] \to \mathbb{R}$ with derivative $f' \in L^2([0,1])$ and for which $f(0) = 0$, with inner product*

$$\langle f, g \rangle_{\mathcal{H}} = \int_0^1 f'(u) g'(u) \, \mathrm{d}u = \langle f', g' \rangle_{L^2([0,1])}.$$

Thus, the induced norm

$$||f||_{\mathcal{H}}^2 = \langle f, f \rangle_{\mathcal{H}} = \int_0^1 (f'(u))^2 \, \mathrm{d}u = ||f'||_{L^2([0,1])}$$

is a sort of measure of wiggliness.

Note that the condition of absolute continuity and boundary condition $f(0) = 0$ together guarantee the positive definiteness of $\langle \cdot, \cdot \rangle_{\mathcal{H}}$: $\forall f \in \mathcal{H}$, if $||f||_{\mathcal{H}}^2 = \int_0^1 (f'(u))^2 \, \mathrm{d}u = 0$, then $f = 0$. The full proof that $\mathcal{H}$ is a Hilbert space can be found in [17].

For any $x \in [0,1]$ and $f \in \mathcal{H}$, the evaluation functional at x

$$\mathrm{E}_x f = f(x) = \int_0^x f'(u) \, \mathrm{d}u = \int_0^1 f'(u) \mathbb{1}_{u \leq x}(u) \, \mathrm{d}u = \langle f', \mathbb{1}_{\cdot \leq x} \rangle_{L^2([0,1])} = \langle f, \min(\cdot, x) \rangle_{\mathcal{H}}$$

is bounded since it is expressible as an inner product between f and a fixed element of $\mathcal{H}$ ($\forall x \in [0,1]$, the function $\min(\cdot, x) \in \mathcal{H}$ since it is absolutely continuous, has square-integrable derivative $\mathbb{1}_{\cdot \leq x}$, and satisfies the boundary condition: $\min(x, 0) = 0$). Indeed,

$$|\mathrm{E}_x f| = |f(x)| = |\langle f, \min(\cdot, x) \rangle_{\mathcal{H}}| \leq ||f||_{\mathcal{H}} \sqrt{\int_0^1 (\mathbb{1}_{\cdot \leq x}(u))^2 \, \mathrm{d}u} = \sqrt{x} \cdot ||f||_{\mathcal{H}}.$$

Since $\mathcal{H}$ is a Hilbert space whose evaluation functional is bounded (with representer of evaluation at x given by $\min(\cdot, x)$), it is an RKHS.

The Fourier coefficients $\hat{f}$ of functions f in the space $\mathcal{H}$ in Example 2.6 decay in a manner concomitant with the smoothness of its functions; in fact $\mathcal{H}$ can be defined [17, 162]

$$\mathcal{H} = \left\{ f \in L^2([0,1]) \;\middle|\; \sum_{n=-\infty}^{\infty} (1 + n^2) |\hat{f}(n)|^2 < \infty \right\}.$$

The Fourier coefficients are weighted or "low-pass filtered" by an ℓ^1 sequence given by $\lambda_n = \frac{1}{1+n^2}$ so as to excise from $\mathcal{H}$ all functions whose Fourier coefficients do not decay sufficiently quickly. Only functions for which

$$\sum_{n=0}^{\infty}(1+n^2)|\hat{f}(n)|^2 < \infty.$$

are kept. We will see in Section 2.2.1 that this Fourier characterization can be used to construct RKHSs, including that of the thin-plate splines on the sphere. As we observe in our first example of a Hilbert space that fails to be an RKHS, this filtering by the sequence λ_n is needed to enforce regularity.

Non-example 2.7. *The Hilbert space $L^2([0,1])$ is not an RKHS, since pointwise evaluation is not well-defined in $L^2([0,1])$. Moreover, while the Dirac delta's "sifting property" allows it to formally play the part of a Riesz representation of the evaluation functional*

$$\forall x \in (0,1),\ f(x) = \int_0^1 \delta(u-x)f(u)\,\mathrm{d}u = \langle f, \delta(\cdot - x)\rangle_{L^2([0,1])},$$

the tempered distribution $\delta(\cdot - x)$ is neither bounded nor in $L^2([0,1])$.

RKHSs–and the Riesz representations of bounded linear evaluation functionals that reside therein–are closely associated with functions called positive-definite kernels.

Definition 2.8 (Positive-definite kernel)**.** *A positive-definite kernel on a set $\mathcal{X}$ is a function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ that is symmetric*

$$\forall (x, x') \in \mathcal{X}^2, k(x, x') = k(x', x)$$

and definite–that is,

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j k(x_i, x_j) \geq 0 \tag{5}$$

holds for all $n \in \mathbb{N}, (x_1, x_2, \ldots, x_n) \in \mathcal{X}^n$, and $a = (a_1, a_2, \ldots a_n) \in \mathbb{R}^n$.

When the inequality (5) is strict for all n and choices of $(x_1, x_2 \ldots, x_n) \in \mathcal{X}^n$ and nonzero weight vector $a \in \mathbb{R}^n$, we call the positive-definite kernel *strictly* positive-definite. This convention is, unfortunately, not aligned with the one we use for matrices. We call a symmetric $n \times n$ matrix $\mathbf{M}$ positive-definite only if the quadratic form $x^T\mathbf{M}x > 0$ for all nonzero $x \in \mathbb{R}^n$–that is, only if the symmetric bilinear form $\langle x, y\rangle = x^T\mathbf{M}y$ is a (definite) inner product.

Remark 2.9. *Define the Gram matrix $\mathbf{K}$ of k on any collection of n points $(x_1, x_2, \ldots x_n)$ in $\mathcal{X}^n$ by*

$$(\mathbf{K})_{i,j} = k(x_i, x_j).$$

Then k is a positive-definite kernel if and only if every Gram matrix $\mathbf{K}$ based on k is a symmetric positive-semidefinite matrix; k is strictly positive-definite if and only if every associated Gram matrix $\mathbf{K}$ is symmetric positive-definite.

Remark 2.10. *While $\mathcal{X}$ is an arbitrary set, and not necessarily an inner product space, a positive-definite kernel k defined on $\mathcal{X}^2$ nevertheless behaves a bit like an inner product. In particular, it obeys the Cauchy-Schwarz inequality: for all $(x, x') \in \mathcal{X}^2$,*

$$k(x, x')^2 \leq k(x, x)k(x', x'),$$

since the Gram matrix associated with the points x and x',

$$\mathbf{K} = \begin{pmatrix} k(x,x) & k(x,x') \\ k(x',x) & k(x',x') \end{pmatrix},$$

has only nonnegative eigenvalues, since the Rayleigh quotient

$$\lambda = \frac{x^T \mathbf{K} x}{x^T x} \geq 0 \; \forall x \in \mathbb{R}^2 \text{ (and, in particular, for eigenvectors of } \mathbf{K})$$

and thus nonnegative determinant

$$k(x,x)k(x',x') - k(x,x')^2 \geq 0,$$

by the symmetry of k.

In fact, positive-definite kernels *are* inner products–just not on $\mathcal{X}$. As shown by Kolmogorov [137] (for countable index sets $\mathcal{X}$) and Mercer [102] (for compact $\mathcal{X}$) and later extended by Aronszajn to arbitrary index sets $\mathcal{X}$ [3], a positive-definite kernel k defined on an index set $\mathcal{X}$ expresses an inner product in a Hilbert space $\mathcal{H}$ associated with k and $\mathcal{X}$.

Proposition 2.11 (Aronszajn-Moore theorem)**.** *A function*

$$k : \mathcal{X}^2 \to \mathbb{R}$$

is a positive-definite kernel if and only if there is a Hilbert space $\mathcal{H}$ and a mapping

$$\phi : \mathcal{X} \to \mathcal{H}$$

such that

$$\forall\, (x,x') \in \mathcal{X}^2,\; \langle \phi(x), \phi(x') \rangle_{\mathcal{H}} = k(x,x').$$

Thus, each positive-definite kernel k takes in pairs of values from the index set $\mathcal{X}$ and outputs an inner product between pairs of functions in the Hilbert space $\mathcal{H}$ that Aronszajn associated with k.

Before proving Proposition 2.11, let us first introduce an alternate characterization of an RKHS.

Definition 2.12 (RKHS: we have the reproducing kernel)**.** *Let $\mathcal{X}$ be a set and $\mathcal{H} \subseteq \mathbb{R}^{\mathcal{X}}$ be a Hilbert space of functions on $\mathcal{X}$ with inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$. Then $\mathcal{H}$ is an RKHS if there exists a reproducing kernel, that is, a function*

$$k : \mathcal{X}^2 \to \mathbb{R}$$

for which

- *$\mathcal{H}$ contains, for all $x \in \mathcal{X}$, the function $k_x \stackrel{\text{def}}{=} k(\cdot, x)$*

$$\begin{aligned} k_x : \mathcal{X} &\to \mathbb{R} \\ y &\mapsto k(y,x). \end{aligned}$$

- *For all $x \in \mathcal{X}$ and $f \in \mathcal{H}$ the reproducing property holds*

$$f(x) = \langle f, k_x \rangle_{\mathcal{H}}.$$

The reproducing property is another way of saying that $k_x = k(\cdot, x)$ acts as the Riesz representer of the $\mathrm{E}_x : \mathcal{H} \to \mathbb{R}$. A reproducing kernel, therefore, allows us to evaluate any function $f \in \mathcal{H} \subseteq \mathbb{R}^{\mathcal{X}}$ at a point $x \in \mathcal{X}$ simply by taking an inner product with a function determined entirely by the kernel: $k_x = k(\cdot, x) \in \mathcal{H}$.

We can see immediately the equivalence of Definition 2.12 with Definition 2.3. Since for all $x \in \mathcal{X}$, we have that $k_x \in \mathcal{H}$ and $f(x) = \langle f, k_x \rangle_{\mathcal{H}}$, if $\mathcal{H}$ is associated with a reproducing kernel, then all evaluation functionals are linear and bounded

$$|\mathrm{E}_x f| = |f(x)| = |\langle f, k_x \rangle_{\mathcal{H}}| \leq ||f||_{\mathcal{H}} ||k_x||_{\mathcal{H}} = \sqrt{k(x,x)} ||f||_{\mathcal{H}},$$

by the Cauchy-Schwarz inequality. Conversely, if all evaluation functionals are bounded, their Riesz representers all exist in $\mathcal{H}$ and can be used, via their inner products, to define a reproducing kernel

$$\forall (x, x') \in \mathcal{X}^2,\ k(x, x') \stackrel{\text{def}}{=} \langle k_x, k_{x'} \rangle_{\mathcal{H}} = k_x(x') = k_{x'}(x).$$

It is straightforward to use the reproducing property to show that if a Hilbert space has a reproducing kernel, it is unique.

Lemma 2.13. *If k and k' are both reproducing kernels associated with an RKHS $\mathcal{H} \subset \mathbb{R}^{\mathcal{X}}$, then $k = k'$.*

Proof. Suppose k and k' are both reproducing kernels. Then by the bilinearity of $\langle \cdot, \cdot \rangle_{\mathcal{H}}$, we must have, for all $x \in \mathcal{X}$ that

$$||k_x - k'_x||^2_{\mathcal{H}} = \langle k_x - k'_x, k_x - k'_x \rangle_{\mathcal{H}} = \langle k_x - k'_x, k_x \rangle_{\mathcal{H}} - \langle k_x - k'_x, k'_x \rangle_{\mathcal{H}}. \tag{6}$$

But since k and k' are both reproducing kernels, $k_x = k(\cdot, x)$ and $k'_x = k'(\cdot, x)$ both reproduce evaluation at x; the difference (6) becomes

$$(k_x - k'_x)(x) - (k_x - k'_x)(x) = 0.$$

Since $||\cdot||_{\mathcal{H}}$ is definite, $k_x = k'_x$ for all $x \in \mathcal{X}$. For all $(x, y) \in \mathcal{X}^2$,

$$k(x, y) = \langle k_y, k_x \rangle_{\mathcal{H}} = \langle k'_y, k'_x \rangle_{\mathcal{H}} = k'(x, y).$$

□

Moreover, any reproducing kernel $k : \mathcal{X}^2 \to \mathbb{R}$ and index set $\mathcal{X}$ is associated with a unique RKHS. We can therefore speak of "the" reproducing kernel of an RKHS or "the" RKHS of a reproducing kernel.

By the reproducing property, we can evaluate a reproducing kernel k on a pair $(x, y) \in \mathcal{X}^2$ by taking the inner product $\langle k_x, k_y \rangle_{\mathcal{H}}$. A reproducing kernel must be symmetric, then, by the symmetry of the inner product

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

Since a reproducing kernel k is symmetric, we must have, for all $x \in \mathcal{X}$, $k_x = k(x, \cdot) = k(\cdot, x)$. A reproducing kernel is, moreover, positive-definite.

Proposition 2.14. *A function $k : \mathcal{X}^2 \to \mathbb{R}$ is positive definite if and only if it is a reproducing kernel associated with an RKHS $\overline{\mathcal{H}} \subset \mathbb{R}^{\mathcal{X}}$.*

Proof. We start by showing that a reproducing kernel is positive definite. (We have just seen that it is symmetric.) Let k be a reproducing kernel on index set $\mathcal{X}$ associated with some RKHS $\mathcal{H}$. To see that k is positive definite, let $(x_1, x_2, \dots x_n) \in \mathcal{X}^n$ for some $n \in \mathbb{N}$ and let $(a_1, a_2, \dots, a_n) \in \mathbb{R}^n$. Then by the bilinearity of the inner product

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j k(x_i, x_j) = \sum_{i=1}^{N}\sum_{j=1}^{n} a_i a_j \langle k_{x_i}, k_{x_j}\rangle_{\mathcal{H}} = \left\langle \sum_{i=1}^{n} a_i k_{x_i}, \sum_{j=1}^{n} a_j k_{x_j} \right\rangle_{\mathcal{H}} = \left\| \sum_{i=1}^{n} a_i k_{x_i} \right\|_{\mathcal{H}}^2 \geq 0.$$

To show the converse, suppose $k : \mathcal{X}^2 \to \mathbb{R}$ is a positive-definite kernel. We will construct the space $\overline{\mathcal{H}}$ whose functions k reproduces. Form the linear manifold $\mathcal{H}$ by taking all finite linear combinations of the $k_x = k(\cdot, x)$ for $x \in \mathcal{X}$

$$\mathcal{H} = \text{span } \{k_x\}_{x\in\mathcal{X}}.$$

We can therefore express any $f \in \mathcal{H}$ and $g \in \mathcal{H}$ as linear combinations of the k_x functions

$$f = \sum_{i=1}^{m} a_i k_{x_i} \text{ and } g = \sum_{j=1}^{n} b_j k_{y_j} \text{ for } (x_1, \dots, x_m) \in \mathcal{X}^m, \text{ and } (y_1, \dots, y_n) \in \mathcal{X}^n. \tag{7}$$

We endow this space with the following inner product

$$\langle f, g\rangle_{\mathcal{H}} = \left\langle \sum_{i=1}^{m} a_i k_{x_i}, \sum_{j=1}^{n} b_j k_{y_j} \right\rangle_{\mathcal{H}} = \sum_{i=1}^{m}\sum_{j=1}^{n} a_i b_j \langle k_{x_i}, k_{y_j}\rangle_{\mathcal{H}} = \sum_{i=1}^{m}\sum_{j=1}^{n} a_i b_j k(x_i, y_j).$$

We can see that $\langle f, g\rangle_{\mathcal{H}}$ does not depend on the choice of expansion in (7)

$$\langle f, g\rangle_{\mathcal{H}} = \sum_{i=1}^{m}\sum_{j=1}^{n} a_i b_j k(x_i, y_j) = \sum_{j=1}^{n} b_j \underbrace{\left(\sum_{i=1}^{m} a_j k_{x_i}(y_j) \right)}_{f(y_j)} = \sum_{j=1}^{n} b_j f(y_j), \text{ and}$$

$$\langle f, g\rangle_{\mathcal{H}} = \sum_{i=1}^{m}\sum_{j=1}^{n} a_i b_j k(y_j, x_i) = \sum_{i=1}^{m} a_i \underbrace{\left(\sum_{j=1}^{n} b_j k_{y_j}(x_i) \right)}_{g(x_i)} = \sum_{i=1}^{m} a_i g(x_i).$$

Thus, $\langle\cdot,\cdot\rangle_{\mathcal{H}}$ is a symmetric bilinear form. Moreover, the k_x reproduce evaluation at x: letting $f = \sum_{i=1}^{m} a_i k_{x_i}$ and $k_x = \sum_{i=1}^{1} 1 k_x$,

$$\langle f, k_x\rangle_{\mathcal{H}} = \sum_{j=1}^{1}\sum_{i=1}^{m} 1 \cdot a_i k(x_i, x) = \sum_{i=1}^{m} a_i k(x_i, x) = \left(\sum_{i=1}^{m} a_i k_{x_i} \right)(x) = f(x).$$

That $||f||_{\mathcal{H}} \geq 0$ follows directly from the bilinearity of $\langle\cdot,\cdot\rangle_{\mathcal{H}}$ and the positive definiteness of k

$$||f||_{\mathcal{H}}^2 = \langle f, f\rangle_{\mathcal{H}} = \left\langle \sum_{i=1}^{m} a_i k_{x_i}, \sum_{i=1}^{m} a_i k_{x_i} \right\rangle_{\mathcal{H}} = \sum_{i=1}^{m}\sum_{j=1}^{m} a_i a_j k(x_i, x_j) \geq 0.$$

Then the Cauchy-Schwarz relation holds and, in particular,

$$|f(x)| = |\langle f, k_x\rangle_{\mathcal{H}}| \leq ||f||_{\mathcal{H}} \sqrt{k(x, x)}.$$

Thus $||f||_{\mathcal{H}} = 0$ entails that $f(x) = 0$ for all $x \in \mathcal{X}$ and thus $f = 0$.

This linear manifold $\mathcal{H}$ is therefore an inner product space as $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ is a strictly positive-definite symmetric bilinear form. Moreover, in this space, norm convergence implies pointwise convergence. Let $\{f_n\}_{n\in\mathbb{N}}$ be a Cauchy sequence of functions in $\mathcal{H}$. We know that for all $m \in \mathbb{N}$, $n \in \mathbb{N}$, and $x \in \mathcal{X}$, the Cauchy-Schwarz inequality and reproducing property guarantee that

$$|f_m(x) - f_n(x)| = |(f_m - f_n)(x)| = |\langle f_m - f_n, k_x \rangle_{\mathcal{H}}| \leq ||f_m - f_n||_{\mathcal{H}} \sqrt{k(x,x)}.$$

Since $\{f_n\}_{n\in\mathbb{N}}$ is Cauchy in the norm $||\cdot||_{\mathcal{H}}$, for any $\epsilon > 0$ there is an $N_\epsilon > 0$ such that for all $m, n > N_\epsilon$, $||f_m - f_n||_{\mathcal{H}} < \epsilon$. For every $x \in \mathcal{X}$ the sequence of values $\{f_n(x)\}_{n\in\mathbb{N}}$ is Cauchy in $\mathbb{R}$: for all $\epsilon > 0$, we can choose $\epsilon_x = \frac{\epsilon}{\sqrt{k(x,x)}}$; then for $m, n > N_{\epsilon_x}$,

$$|f_m(x) - f_n(x)| \leq \epsilon_x \sqrt{k(x,x)} = \epsilon.$$

This Cauchy sequence $\{f_n(x)\}_{n\in\mathbb{N}}$ of real values therefore converges in $\mathbb{R}$ to some value $f(x)$. We define f in this manner to be the pointwise limit of the Cauchy sequence of functions $\{f_n\}_{n\in\mathbb{N}}$.

Let $\overline{\mathcal{H}}$ be the result of adding the pointwise limit functions of all Cauchy sequences in $\mathcal{H}$. It is straightforward, but tedious, to show that the inner product continues to be well-defined, that k remains a reproducing kernel, that $\mathcal{H} = \operatorname{span}\{k_x\}_{x\in\mathcal{X}}$ is dense in $\overline{\mathcal{H}}$, and that the evaluation functionals remain bounded. $\overline{\mathcal{H}}$ is the RKHS for which the positive-definite function k acts as a reproducing kernel. □

Remark 2.15. *There is another way to see that the span of the representers of evaluation at the points in the index set* $\operatorname{span}\{k_x\}_{x\in\mathcal{X}}$ *must be dense in an RKHS* $\mathcal{H}$*. The sequence* $\{k_x\}_{x\in\mathcal{X}}$ *forms a complete system for* $\mathcal{H}$*, since any* $f \in \mathcal{H}$ *orthogonal to the representers of evaluation at all* $x \in \mathcal{X}$ *must be identically 0:* $f(x) = \langle f, k_x \rangle_{\mathcal{X}} = 0$ *for all* $x \in \mathcal{X}$*.*

We can now prove Proposition 2.11.

Proof. Suppose k is a positive-definite kernel defined on the set $\mathcal{X}$. We just showed that there is an RKHS $\mathcal{H}$ corresponding[12] to k. Consider the map from the points in the index set to their Riesz representers of evaluation

$$\begin{aligned} \phi : \mathcal{X} &\to \mathcal{H} \\ x &\mapsto k_x. \end{aligned}$$

Then for all pairs $(x, x') \in \mathcal{X}^2$ we have that $k(x, x') = \langle k_x, k_{x'} \rangle_{\mathcal{H}} = \langle \phi(x), \phi(x') \rangle_{\mathcal{H}}$ is a reproducing kernel. Indeed, the reproducing property follows from the fact that k_x is a Riesz representer of evaluation at x: $\forall f \in \mathcal{H}, \forall x \in \mathcal{X}, f(x) = \langle f, k_x \rangle_{\mathcal{H}}$. Conversely, suppose there is a Hilbert space $\mathcal{H}$ and a mapping

$$\begin{aligned} \phi : \mathcal{X} &\to \mathcal{H} \\ x &\mapsto k_x, \end{aligned}$$

such that for all $(x, x') \in \mathcal{X}^2$, we can define a reproducing kernel $k(x, x') = \langle \phi(x), \phi(x') \rangle_{\mathcal{H}}$. Then for any $n \in \mathbb{N}$ and $(a_1, a_2, \ldots a_n) \in \mathcal{X}^n$, we see, by the bilinearity of $\langle \cdot, \cdot \rangle_{\mathcal{H}}$, that k is positive definite

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j k(x_i, x_j) = \left\langle \sum_{i=1}^{n} a_i \phi(x_i), \sum_{j=1}^{n} a_j \phi(x_j) \right\rangle_{\mathcal{H}} = \left\| \sum_{i=1}^{N} a_i \phi(x_i) \right\|_{\mathcal{H}}^2 \geq 0.$$

□

[12] In the proof of Proposition 2.14, we used a bar to emphasize that $\overline{\mathcal{H}}$ was the completion of the span of the representers of evaluation at each point in the index set; here we jettison the bar as $\mathcal{H}$ is understood to be complete.

Remark 2.16 (The kernel trick)**.** *We have now seen that a kernel k maps pairs of points in an index set $\mathcal{X}$ to the inner product between functions–the representers of evaluation at these points–in $\mathcal{H} \subseteq \mathbb{R}^{\mathcal{X}}$. These functions are often called "feature vectors" by machine learning practitioners, and the map ϕ the "feature map". To solve many learning and approximation problems in practice, one does not need access to the representers of evaluation k_x or the space $\mathcal{H}$ in which they live, so long as one is confident they are suited to the application and one can compute the kernel k on pairs of inputs in the space. One of the most commonly used kernels has a native space $\mathcal{H}$ that is difficult to characterize [142], and the feature map need not be uniquely specified given an RKHS and its kernel [34]. The machine learning literature calls "the kernel trick" this ability to solve problems in infinite-dimensional spaces without full access to the functions that live there, using only the inner products $k(x, x') = \langle k_x, k_{x'} \rangle_{\mathcal{H}}$ between select functions in the space. In the context of approximating functions with splines, this trick was identified by Wahba and Kimeldorf with the representer theorem (see Section 2.5).*

Positive-definite kernels also assesses the similarity between points in $\mathcal{X}$. With respect to this similarity metric, the RKHS norm $||\cdot||_{\mathcal{H}}$ indicates a function's smoothness or regularity (the smaller the norm of a function, the smoother it is). The RKHS $\mathcal{H}$ and its norm are therefore useful to consider in applications such as spline smoothing, even when one can make use of the kernel trick.

Remark 2.17. *Positive-definite kernels encode a metric or pseudometric on the index set $\mathcal{X}$ according to which any function f in the associated RKHS $\mathcal{H}$ is Lipschitz continuous with Lipschitz constant $M = ||f||_{\mathcal{H}}$*

$$d_{\mathbb{R}}(f(x), f(x')) = |f(x) - f(x')| = |\langle f, k_x - k_{x'} \rangle_{\mathcal{H}}| \leq \underbrace{||f||_{\mathcal{H}}}_{M} ||k_x - k_{x'}||_{\mathcal{H}} = M d_{\mathcal{X}}(x, x'),$$

where

$$d_{\mathcal{X}}(x, x') = d_{\mathcal{H}}(k_x, k_{x'}) = ||k_x - k_{x'}||_{\mathcal{H}} = \sqrt{\langle k_x - k_{x'}, k_x - k_{x'} \rangle_{\mathcal{H}}} = \sqrt{k(x, x) + k(x', x') - 2k(x, x')}.$$

Positivity of $d_{\mathcal{X}}$ need not hold, so $d_{\mathcal{X}}$ is in general a pseudometric. However, if k is strictly positive-definite, its Gram matrix on x and $x' \neq x$ has a positive determinant: $k(x, x)k(x', x') > k(x, x')^2$. But then, by the inequality of arithmetic and geometric means, we have that

$$\frac{k(x, x) + k(x', x')}{2} \geq \sqrt{k(x, x)k(x', x')} > k(x, x'),$$

so for any $x \neq x'$, we can establish the positivity of $d_{\mathcal{X}}$

$$\frac{1}{2} d_{\mathcal{X}}(x, x')^2 = \frac{k(x, x) + k(x', x')}{2} - k(x, x') > 0.$$

Strict positive definiteness is a sufficient but not necessary condition for $d_{\mathcal{X}}$ to be a metric. Consider the positive-definite kernel of Example 2.6, which is not strictly positive-definite (the Gram matrix on $\{0, 1\} \subseteq \mathcal{X}$ has eigenvalues $\{0, 1\}$). Nevertheless, the associated distance metric exhibits positivity

$$d_{\mathcal{X}}(x, x') = ||\min(\cdot, x) - \min(\cdot, x')||_{\mathcal{H}} = \sqrt{\min(x, x) + \min(x', x') - 2\min(x, x')} = \sqrt{|x - x'|}.$$

This metric $d_{\mathcal{X}}$ illustrates, moreover, that the distances between elements of the index set possess different properties from elements of the RKHS and its reproducing kernel. With one argument fixed, $d_{\mathcal{X}}$ is not sufficiently well-behaved to reside in the RKHS $\mathcal{H}$ of Example 2.6 as its derivative is not square integrable on $[0, 1]$. Moreover, it is not a positive-definite kernel (the distance matrix of $d_{\mathcal{X}}$ on $\{0, 1\}$ has eigenvalues $\{1, -1\}$).

Remark 2.18. *If $\mathcal{X}$ is a real inner product space, its dual space is an RKHS. For each $x \in \mathcal{X}$, the representation of evaluation at x is the functional $\phi(x) = \langle \cdot, x \rangle_{\mathcal{X}}$; the kernel is the inner product: $k(x,y) = \langle k_x, k_y \rangle_{\mathcal{H}} = \langle x, y \rangle_{\mathcal{X}}$. Solutions of spline interpolation problems in this RKHS are given by ordinary linear regression; smoothing problems, ridge regression*[13]*. The metric on $\mathcal{X}$ induced by the linear kernel is the same as the metric induced by inner product of $\mathcal{X}$*

$$d_{\mathcal{X}}(x,y) = \sqrt{||x||_{\mathcal{X}}^2 + ||y||_{\mathcal{X}}^2 - 2\langle x, y \rangle_{\mathcal{X}}} = ||x - y||_{\mathcal{X}}.$$

Linear kernels defined over a Euclidean inner product space, for instance, induce a Euclidean metric over the index set.

On the other hand, kernels can endow an inner product space with metrics that disagree sharply with the metric induced by the inner product. While the Euclidean space $\mathcal{X} = \mathbb{R}$ can be endowed with a Euclidean metric by the linear kernel, different kernels, associated with different RKHSs of functions on $\mathcal{X}$, can equip $\mathcal{X}$ with vastly different metrics. The Paley-Wiener space of finite-energy bandlimited signals

$$\mathcal{PW}_{\pi w} = \{f \in L^2(\mathbb{R}) \mid \operatorname{support}(\widehat{f}) \subseteq [-\pi w, \pi w]\}$$

is the RKHS induced by the sinc kernel: $k(x,y) = \frac{\sin(\pi w(x-y))}{\pi(x-y)}$. This strictly positive-definite kernel induces a bounded, oscillating metric on $\mathbb{R}$

$$d_{\mathbb{R}}(x,y) = ||k_x - k_y||_{\mathcal{PW}_{\pi w}} = \sqrt{2\left(w - \frac{\sin(\pi w(x-y))}{\pi(x-y)}\right)}.$$

Norm-minimizing solutions to interpolation and smoothing problems over this space exhibit characteristic wiggles that are related to the nature of this distance metric.

When there is additional structure on $\mathcal{X}$, there is more to say about the relationship between the properties of k and those of the functions in $\mathcal{H}$. The boundedness (and, if $\mathcal{X}$ is a topological space, the continuity) of the kernel k depends on the boundedness (respectively, continuity) of the feature map.

Definition 2.19 (Feature map)**.** *We shall call the map*

$$\begin{aligned} \phi : \mathcal{X} &\to \mathcal{H} \\ x &\mapsto k_x, \end{aligned}$$

introduced in the proof of Proposition 2.11 the feature map of the kernel k or RKHS $\mathcal{H}$.

Definition 2.20. *We say a kernel is bounded if*

$$\sup_{x \in \mathcal{X}} k(x,x) < \infty.$$

Remark 2.21. *Note that the maximum absolute value of k must occur on its "diagonal". Clearly, we have that imposing the diagonal constraint cannot increase the maximum kernel absolute value*

$$\sup_{(x,x') \in \mathcal{X}^2} |k(x,x')| \geq \sup_{x \in \mathcal{X}} k(x,x).$$

(We omitted the absolute value sign on the right-hand side because $k(x,x) = ||k_x||_{\mathcal{H}}^2$.)

[13] To pose the problem over the richer space of affine functions, not just the dual space, the direct-sum decomposition principle can be used (see Section 2.4).

On the other hand, by the Cauchy-Schwarz inequality and monotone continuity of the square root function, we have the opposite relation

$$\sup_{(x,x')\in\mathcal{X}^2} |k(x,x')| = \sup_{(x,x')\in\mathcal{X}^2} |\langle k_x, k_{x'}\rangle_{\mathcal{H}}| \leq \sup_{x\in\mathcal{X}} \sqrt{k(x,x)} \cdot \sup_{x'\in\mathcal{X}} \sqrt{k(x',x')} = \sup_{x\in\mathcal{X}} k(x,x). \tag{8}$$

Thus,

$$\sup_{x\in\mathcal{X}} k(x,x) = \sup_{(x,x')\in\mathcal{X}^2} |k(x,x')|.$$

Proposition 2.22. *A kernel is bounded if and only if its feature map is bounded. In this case, every function in $\mathcal{H}$ is bounded.*

Proof. The first statement holds because

$$\forall x \in \mathcal{H},\ ||\phi(x)||^2_{\mathcal{H}} = \langle k_x, k_x\rangle_{\mathcal{H}} = k(x,x).$$

The second since, for all $f \in \mathcal{H}$ and all $x \in \mathcal{X}$,

$$|f(x)| = |\langle f, k_x\rangle_{\mathcal{H}}| \leq ||f||_{\mathcal{H}}\sqrt{k(x,x)}.$$

□

We have already seen that convergence in an RKHS implies pointwise convergence. We can expand on this, with the following result from Aronszajn [3].

Proposition 2.23 (Convergence in $\mathcal{H}$ and pointwise convergence)**.** *Let $\mathcal{H} \subseteq \mathbb{R}^{\mathcal{X}}$ be an RKHS. Then*

1. *The sequence of functions $\{f_n\}_{n=0}^{\infty}$ converges weakly in $\mathcal{H}$ to f if and only if (a) for each $x \in \mathcal{X}$, $f_n(x)$ converges to $f(x)$ in $\mathbb{R}$ and (b) $\{||f_n||_{\mathcal{H}}\}_{n=0}^{\infty}$ is bounded.*
2. *The sequence $\{f_n\}_{n=0}^{\infty}$ converges strongly to f if and only if (a) holds and (b') $\lim_{n\to\infty} ||f_n||_{\mathcal{H}} = ||f||_{\mathcal{H}}$.*

Proof. 1. Suppose $\{f_n\}_{n=0}^{\infty}$ converges weakly to f in $\mathcal{H}$. Then for all $g \in \mathcal{H}$

$$\lim_{n\to\infty} \langle f_n, g\rangle_{\mathcal{H}} = \langle f, g\rangle_{\mathcal{H}},$$

and in particular, for all $x \in \mathcal{X}$, we have that

$$\lim_{n\to\infty} f_n(x) = \lim_{n\to\infty} \langle f_n, k_x\rangle_{\mathcal{H}} = \langle f, k_x\rangle_{\mathcal{H}} = f(x).$$

The uniform boundedness principle for Hilbert spaces, a consequence of the Banach-Steinhaus theorem (see, e.g., [36], Theorem 3.3.15), allows us to immediately establish (b).

Now suppose (a) and (b) are established. We want to show that for any $h \in \mathcal{H}$,

$$\lim_{n\to\infty} \langle f_n - f, h\rangle_{\mathcal{H}} = 0.$$

From (a), we know that, for all $x \in \mathcal{X}$,

$$\lim_{n\to\infty} \langle f_n - f, k_x\rangle_{\mathcal{H}} = \lim_{n\to\infty} (f_n(x) - f(x)) = 0.$$

Using (b), we can choose an $M < \infty$ such that $||f_n||_{\mathcal{H}} < M$ for all $n \in \mathbb{N}$.

Because $\{k_x\}_{x\in\mathcal{X}}$ is dense in $\mathcal{H}$, for all $\epsilon > 0$, property (a) allows us to choose a finite sequence of representers of evaluation $\{k_{x_1}, \ldots, k_{x_m}\}$ and weights $\alpha_1, \ldots, \alpha_m$ such that the function h' defined as the linear combination

$$h' = \sum_{i=1}^{m} \alpha_i k_{x_i} \tag{9}$$

is within an ϵ-neighborhood of h in the norm of $\mathcal{H}$

$$||h - h'||_{\mathcal{H}} < \epsilon.$$

In particular, with enough terms, the expansion (9) will satisfy

$$||h - h'||_{\mathcal{H}} < \frac{\epsilon}{M + ||f||_{\mathcal{H}}},$$

for any $\epsilon > 0$ and $f \in \mathcal{H}$. Then

$$\begin{aligned}
\lim_{n\to\infty} |\langle f_n - f, h\rangle_{\mathcal{H}}| &\leq \lim_{n\to\infty} |\langle f_n - f, h - h'\rangle_{\mathcal{H}}| + \lim_{n\to\infty} |\langle f_n - f, h'\rangle_{\mathcal{H}}| \\
&\leq \lim_{n\to\infty} ||f_n - f||_{\mathcal{H}} ||h - h'||_{\mathcal{H}} + \underbrace{\lim_{n\to\infty} \left| \sum_{i=1}^{m} \alpha_i \langle f_n - f, k_{x_i}\rangle_{\mathcal{H}} \right|}_{0, \text{ since } f_n \xrightarrow{\text{pointwise}} f} \\
&\leq \lim_{n\to\infty} (||f_n|| + ||f||) \cdot \frac{\epsilon}{M + ||f||_{\mathcal{H}}} < \epsilon,
\end{aligned}$$

by the triangle inequality.

2. Weak convergence plus (b') is equivalent to strong convergence in Hilbert spaces (see, e.g., [36], Theorem 3.3.13).

□

Aronszajn also considered the continuity properties of the kernel when $\mathcal{X}$ is a topological space.

Proposition 2.24 (Kernel and RKHS continuity over a topological space)**.** *Let $\mathcal{H} \subseteq \mathbb{R}^{\mathcal{X}}$ be an RKHS with reproducing kernel k, and $\mathcal{X}$ a topological space. Then the following statements related to the continuity of k hold.*

1. *The map $\phi : x \mapsto k_x$ is continuous if and only if k is continuous along the diagonal.*
2. *Every function $f : \mathcal{X} \to \mathbb{R}$ in $\mathcal{H}$ is continuous in $\mathcal{X}$ if and only if every representer of evaluation k_x is continuous and the map $x \mapsto k(x, x)$ is locally bounded.*
3. *If $\mathcal{X}$ is locally compact and the feature map ϕ is continuous, then for every sequence of functions $\{f_n\}_{n=0}^{\infty}$ that converges weakly in $\mathcal{H}$, $\{f_n\}_{n=0}^{\infty}$ also converges uniformly over any compact set in $\mathcal{X}$. Thus, by the uniform limit theorem, if the map ϕ is continuous, then any weakly convergent sequence of functions that are continuous on $\mathcal{X}$ converges to a continuous function.*
4. *Every family of bounded functions $B_M = \{f \in \mathcal{H} \,|\, ||f||_{\mathcal{H}} \leq M\}$ is equicontinuous if and only if the map $x \mapsto k(x, x)$ is continuous and each representer of evaluation is continuous. In this case, k is continuous.*
5. *If $\mathcal{X}$ is separable and every function in $\mathcal{H}$ is continuous, then $\mathcal{H}$ is separable.*

Proof. 1. Observe that $||k_x||_{\mathcal{H}}^2 = \langle k_x, k_x\rangle_{\mathcal{H}} = k(x, x)$.

2. Suppose $f \in \mathcal{H}$ is continuous and consider a sequence $\{y_n\}_{n=0}^{\infty}$ of points in $\mathcal{X}$. Since representers of evaluation are in $\mathcal{H}$, they must be continuous; we therefore need only show (b), the local boundedness of the map $x \mapsto k(x,x)$. For any continuous function $f \in \mathcal{H}$, the reproducing property implies that the representers of convergent sequences in $\mathcal{X}$ are weakly convergent in $\mathcal{H}$

$$\lim_{n\to\infty} y_n = y \implies \lim_{n\to\infty} \langle f, k_{y_n}\rangle_{\mathcal{H}} = \lim_{n\to\infty} f(y_n) = f(y) = \langle f, k_y\rangle_{\mathcal{H}}.$$

Since every function in $\mathcal{H}$ is continuous, k_{y_n} converges weakly to k_y in $\mathcal{H}$, and we can apply part 1 of Proposition 2.23 to conclude that $k(y,y) = ||k_y||^2_{\mathcal{H}}$ is locally bounded.

Conversely, suppose $x \mapsto k(x,x)$ is locally bounded and every representer of evaluation is continuous, and consider any sequence $\{z_n\}_{n=0}^{\infty}$ in $\mathcal{X}$ that converges to a limit z. Then for all $n > N$, for some N sufficiently large, the quantity $||k_z||_{\mathcal{H}} + ||k_{z_n}||_{\mathcal{H}}$ must be bounded above by twice the local bound $M < \infty$ of the map $x \mapsto k(x,x)$, evaluated at z; in other words, $||k_z||_{\mathcal{H}} + ||k_{z_n}||_{\mathcal{H}} \leq 2M$. Since the (bounded) representers of evaluation are a dense subset of $\mathcal{H}$, we have that for any $f \in \mathcal{H}$, there is a sequence of (bounded) representers of evaluation $\{f_m\}_{m=0}^{\infty}$ strongly converging to f in $\mathcal{H}$; then, for any sequence $\{z_n\}_{n=0}^{\infty}$ converging to z in $\mathcal{X}$,

$$\begin{aligned}
|f(z) - f(z_n)| &= |(f(z) - f_m(z)) + (f_m(z) - f_m(z_n)) + (f_m(z_n) - f(z_n))| \\
&\leq |f(z) - f_m(z)| + |f_m(z) - f_m(z_n)| + |f_m(z_n) - f(z_n)| \\
&= |f_m(z) - f_m(z_n)| + |\langle f - f_m, k_z\rangle_{\mathcal{H}}| + |\langle f_m - f, k_{z_n}\rangle_{\mathcal{H}}| \\
&\leq |f_m(z) - f_m(z_n)| + ||f - f_m||_{\mathcal{H}} \left(||k_z||_{\mathcal{H}} + ||k_{z_n}||_{\mathcal{H}}\right),
\end{aligned}$$

where on the third line we swapped the first and second terms in addition to applying the reproducing property. For any $\epsilon > 0$, we can choose m so that

$$||f - f_m||_{\mathcal{H}} < \frac{\epsilon}{4M}.$$

Since each f_m is a representer of evaluation and therefore continuous, we can choose N' such that for $n > N'$, $|f_m(z) - f_m(z_n)| < \frac{\epsilon}{2}$. Then for m sufficiently large and $n > \max(N, N')$, we have that

$$|f(z) - f(z_n)| \leq |f_m(z) - f_m(z_n)| + ||f - f_m||_{\mathcal{H}} \left(||k_z||_{\mathcal{H}} + ||k_{z_n}||_{\mathcal{H}}\right) < \frac{\epsilon}{2} + \frac{\epsilon}{4M} \cdot 2M = \epsilon.$$

Thus, since f was arbitrary, all functions in $\mathcal{H}$ are continuous by the local boundedness, continuity, and density in $\mathcal{H}$ of the representers of evaluation.

See also [132], Proposition 24, and [17], Theorem 17.

3. The continuous ϕ maps a compact set $\mathcal{X}' \subseteq \mathcal{X}$ to a compact subset $\mathcal{H}' \subseteq \mathcal{H}$. Suppose the sequence of functions $f_n \longrightarrow f$ weakly while remaining in $\mathcal{H}'$. By Proposition 2.23 (the uniform boundedness principle), $\{f_n\}_{n=0}^{\infty}$ is bounded; choose an integer M so that $\sup\{||f_n||_{\mathcal{H}}\}_{n=0}^{\infty} < M$. In fact, $\{f_n\}_{n=0}^{\infty}$ converges strongly[14]. By the compactness of $\mathcal{X}'$ and continuity of ϕ, for every $\epsilon > 0$, there is a finite subcover of open balls of radius $\delta_{\epsilon/4M}$ centered at $C = \{y_1, \ldots, y_m\}$ such that for every $y \in \mathcal{X}'$ there exists some center $y_l \in \mathcal{X}'$ satisfying

$$d(y, y_l) < \delta_{\epsilon/4M} \implies ||\phi(y) - \phi(y_l)||_{\mathcal{H}} = ||k_y - k_{y_l}||_{\mathcal{H}} < \frac{\epsilon}{4M}.$$

[14] Any subsequence of $\{f_n\}_{n=0}^{\infty}$ has a convergent subsequence whose limit is necessarily f by the weak convergence. But saying that every subsequence of a sequence in a metric space itself has a subsequence that converges to a fixed limit implies that f_n converges in the metric $d(f_n, f) = ||f_n - f||_{\mathcal{H}}$!

Then, choosing N such that for all $n > N$ and $y_l \in C$, $|f(y_l) - f_n(y_l)| < \frac{\epsilon}{2}$, we see that

$$\begin{aligned}|f(y) - f_n(y)| &= |(f(y) - f(y_l)) + (f(y_l) - f_n(y_l)) + (f_n(y_l) - f_n(y))| \\ &= |\langle f - f_n, k_y - k_{y_l}\rangle_{\mathcal{H}} + f(y_l) - f_n(y_l)| \leq |\langle f - f_n, k_y - k_{y_l}\rangle_{\mathcal{H}}| + |f(y_l) - f_n(y_l)| \\ &\leq \underbrace{||f - f_n||_{\mathcal{H}}}_{\substack{\leq ||f||_{\mathcal{H}} + ||f_n||_{\mathcal{H}} \\ < M+M}} \underbrace{||k_y - k_{y_l}||_{\mathcal{H}}}_{\leq \frac{\epsilon}{4M}} + \underbrace{|f(y_l) - f_n(y_l)|}_{\leq \frac{\epsilon}{2}} < \epsilon.\end{aligned}$$

Since N does not depend on y (i.e., no matter which open ball y falls in, its center point $y_l \in C$ satisfies $|f(y_l) - f_n(y_l)| < \frac{\epsilon}{2}$), f_n converges to f uniformly.

4. Suppose every set of bounded functions in $\mathcal{H}$ is equicontinuous. Then every function in $\mathcal{H}$ is continuous and by part 2, k is locally bounded on the diagonal. In particular, the representer of evaluation at x, k_x, is continuous for any $x \in \mathcal{X}$. Since the function k_x is continuous, we have that for any sequence $\{x_n\}_{n=0}^{\infty}$ converging to x within the neighborhood of x in which $x \mapsto k(x, x)$ is locally bounded,

$$\begin{aligned}|k(x_n, x_n) - k(x, x)| &= |(k(x_n, x_n) - k(x_n, x)) + (k(x_n, x) - k(x, x))| \\ &\leq |k(x_n, x_n) - k(x_n, x)| + |k(x_n, x) - k(x, x)| \\ &= |k_{x_n}(x_n) - k_{x_n}(x)| + |k_x(x_n) - k_x(x)|,\end{aligned}$$

by the symmetry of k and triangle inequality. Both terms tend to 0 as $n \longrightarrow \infty$ by the continuity of the representers of evaluation. Hence, k is continuous along the diagonal.

Now suppose all representers of evaluation k_x are continuous and that k is continuous along the diagonal. Consider any sequences $\{x_n\}_{n=0}^{\infty}$ and $\{y_n\}_{n=0}^{\infty}$ converging to x and y, respectively, for any $(x, y) \in \mathcal{X}^2$. Accordingly, by the symmetry of k, reproducing property, and triangle inequality,

$$\begin{aligned}|k(x_n, y_n) - k(x, y)| &= |k(x_n, y_n) - k(x_n, y) + k(x_n, y) - k(x, y)| \\ &\leq |k_{x_n}(y_n) - k_{x_n}(y)| + |k_y(x_n) - k_y(x)| \\ &= |\langle k_{x_n}, k_{y_n} - k_y\rangle_{\mathcal{H}}| + |\langle k_y, k_{x_n} - k_x\rangle_{\mathcal{H}}| \\ &\leq ||k_{y_n} - k_y||_{\mathcal{H}}||k_{x_n}||_{\mathcal{H}} + ||k_{x_n} - k_x||_{\mathcal{H}}||k_y||_{\mathcal{H}}.\end{aligned}$$

Since k is continuous along the diagonal, $||k_{x_n}||_{\mathcal{H}} = k(x_n, x_n) \longrightarrow k(x, x) = ||k_x||_{\mathcal{H}}$ as $n \longrightarrow \infty$. Thus, we need only show that the sequences of representers of evaluation $\{k_{x_n}\}_{n=0}^{\infty}$ and $\{k_{y_n}\}_{n=0}^{\infty}$ converge in the norm to k_x and k_y, respectively. But both terms of

$$\begin{aligned}||k_{x_n} - k_x||_{\mathcal{H}}^2 &= \langle k_{x_n} - k_x, k_{x_n} - k_x\rangle_{\mathcal{H}} \\ &= (k(x_n, x_n) - k(x_n, x)) + (k(x, x_n) - k(x, x)) \\ &= (k_{x_n}(x_n) - k_{x_n}(x)) + (k_x(x_n) - k_x(x)),\end{aligned}$$

converge to 0 as $n \longrightarrow \infty$ by the continuity of the representers of evaluation k_{x_n} and k_x. The strong convergence of $\{k_{y_n}\}_{n=0}^{\infty}$ to k_y can be shown in the same way. Thus, k is continuous in both arguments simultaneously, since for all $(x, y) \in \mathcal{X}^2$ and all sequences $\{x_n\}_{n=0}^{\infty}$ and $\{y_n\}_{n=0}^{\infty}$ converging to x and y, respectively, we have that $k(x_n, y_n)$ must converge to $k(x, y)$.

Finally, suppose k is continuous in both arguments simultaneously. Consider the family of functions $B_M = \{f \in \mathcal{H} \,|\, ||f||_{\mathcal{H}} \leq M\}$. We want to show that for all $\epsilon > 0$ and all $x \in \mathcal{X}$, there exists a $\delta_x > 0$ such that for all $f \in B_M$ and all $y \in \mathcal{X}$ satisfying $d_{\mathcal{X}}(x, y) < \delta_x$, we have $|f(x) - f(y)| < \epsilon$. But because k is continuous in both arguments, we can choose, for any $\epsilon > 0$

and $x \in \mathcal{X}$, a $\delta_x > 0$ such that for any y and z satisfying $d_{\mathcal{X}}(x,y) < \delta_x$ and $d_{\mathcal{X}}(x,z) < \delta_x$, we have that $|k(x,x) - k(y,z)| < \frac{\epsilon^2}{3M^2}$. Then for all y satisfying $d_{\mathcal{X}}(x,y) < \delta_x$, we have that

$$\begin{aligned} ||k_x - k_y||_{\mathcal{H}}^2 &= \langle k_x - k_y, k_x - k_y \rangle_{\mathcal{H}} = (k(x,x) - k(x,y)) + (k(y,y) - k(y,x)) \\ &= (k(x,x) - k(x,y)) + (k(y,y) - k(x,x)) + (k(x,x) - k(y,x)) \\ &\leq |k(x,x) - k(x,y)| + |k(x,x) - k(y,y)| + |k(x,x) - k(y,x)| < \frac{\epsilon^2}{M^2}. \end{aligned}$$

In that case, for any $f \in B_M$,

$$|f(x) - f(y)| = |\langle f, k_x - k_y \rangle_{\mathcal{H}}| \leq ||f||_{\mathcal{H}} \cdot ||k_x - k_y||_{\mathcal{H}} < M \cdot \frac{\epsilon}{M} = \epsilon.$$

5. In this case, the representers of evaluation $R = \{k_x \,|\, x \in \mathcal{X}'\}$ at a countable, dense subset $\mathcal{X}' \subseteq \mathcal{X}$ form a complete system for $\mathcal{H}$. Indeed, if $f \in \mathcal{H}$ is orthogonal to each function in R, it evaluates to 0 at each point in $\mathcal{X}'$. By the continuity of f, this means it must be identically 0. For any $x \in \mathcal{X}$, let $\{x_n\}_{n=0}^{\infty}$ be a sequence of points in $\mathcal{X}'$ converging to x. Then, since f is continuous,
$$f(x) = \lim_{n \to \infty} f(x_n) = \lim_{n \to \infty} \langle f, k_{x_n} \rangle_{\mathcal{H}} = 0.$$

□

For more examples of how constraints on $\mathcal{X}$ and k can endow the functions in the associated RKHS $\mathcal{H}$ with desirable properties (such as differentiability), see, for instance, [132], Section 9; [3], Section 5; and [4], Section 2.

We end this section by introducing two types of kernels on Euclidean spaces that arise in many applications: the shift-invariant and radial kernels. These are related to the Fourier and Laplace transforms of nonnegative Borel measures, respectively. In particular, the former can be seen to be infinite linear combinations of complex exponentials of different frequencies; the latter, infinite linear combinations of Gaussians of different scales.

Definition 2.25 (Shift-invariant and radial kernels)**.** *We say a kernel $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is* shift invariant *if it satisfies, for all $(x,y) \in \mathbb{R}^d \times \mathbb{R}^d$, $k(x,y) = k'(x-y)$ for some fixed $k' : \mathbb{R}^d \to \mathbb{R}$. A function $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ that satisfies $k(x,y) = k'(||x-y||_{\mathbb{R}^d})$ for some $k' : [0,\infty) \to \mathbb{R}$ is called a* radial function*. If, for all $d \in \mathbb{N}_{\geq 1}$, k' determines a positive-definite kernel $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, we say that k' is a* radial basis function*.*

Proposition 2.26. *1. A continuous shift-invariant function $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is positive definite on $\mathcal{X} = \mathbb{R}^d$ if and only if it can be expressed $k(x,y) = k'(x-y)$, where k' is the Fourier-Stieltjes transform of a finite nonnegative Borel measure μ*

$$k'(\tau) = \widehat{\mu}(\tau) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} e^{-i\langle \tau, x \rangle_{\mathbb{R}^d}} \, \mathrm{d}\mu(x) \, .$$

2. For a fixed $d \in \mathbb{N}_{\geq 1}$, a continuous radial function $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is positive definite if and only if it can be written $k(x,y) = k'(||x-y||_{\mathbb{R}^d})$, where k' is the Hankel transform of a finite nonnegative Borel measure on $[0,\infty)$ not concentrated at the origin, μ. A function $k' : [0,\infty) \to \mathbb{R}$ yields a positive-definite radial function $k(x,y) = k'(||x-y||_{\mathbb{R}^d})$ on $\mathbb{R}^d$ for all $d \in \mathbb{N}_{\geq 1}$ if and only if

$$k'(r) = \int_0^{\infty} e^{-r^2 t^2} \, \mathrm{d}\mu(t)$$

for some finite *nonnegative Borel measure μ. Since the* completely monotone functions *are the Laplace-Stieltjes transforms of nonnegative Borel measures, this is equivalent to saying $k'(r) = f(r^2)$ for some completely monotone function.*

Proof. 1. This was shown by Bochner in a book on Fourier integrals in 1932 [19] and an article [18] in 1933, first handling $d = 1$ and then the general case. (The theorem is sometimes also attributed to Khinchin [136] and is closely related to the Wiener-Khinchin theorem.) In the full proof, the exponential must be handled with care given the improper boundary. However, the only direction we need is easy: for any $n \in \mathbb{N}_{\geq 1}$ and $c \in \mathbb{R}^n$, we have that

$$\begin{aligned}\sum_{j=1}^{n}\sum_{k=1}^{n} c_j c_k k(x_j - x_k) &= \sum_{j=1}^{n}\sum_{k=1}^{n} c_j c_k \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} e^{-i\langle (x_j - x_k), x\rangle_{\mathbb{R}^d}} \,\mathrm{d}\mu(x) \\ &= \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \left(\sum_{j=1}^{n} c_j e^{-i\langle x_j, x\rangle_{\mathbb{R}^d}} \sum_{k=1}^{n} \overline{c_k e^{-i\langle x_k, x\rangle_{\mathbb{R}^d}}} \right) \mathrm{d}\mu(x) \\ &= \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \left| \sum_{j=1}^{n} c_j e^{-i\langle x_j, x\rangle_{\mathbb{R}^d}} \right|^2 \mathrm{d}\mu(x) \geq 0.\end{aligned}$$

(This result is usually stated for complex-valued kernels, due to the conjugate used in the proof.)

For the other direction, see [162], Theorem 6.6. It is generalized to characterize *conditionally* positive-definite functions in Theorems 8.12 and 8.14 of that work, which also provides sufficient conditions for *strict* positive definiteness (e.g., if the carrier of μ has nonzero Lebesgue measure).

2. See [128], Theorems 2-3.

□

Remark 2.27 (Certain familiar one-dimensional shift-invariant kernels are not radial kernels in higher dimensions.)**.** *Suppose $\varphi_r(\cdot)$ determines a positive-definite kernel in Euclidean space of every dimension–that is, for all $d \in \mathbb{N}_{\geq 1}$, the function $\varphi : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ defined by $\varphi(x, y) = \varphi_r(||x - y||_{\mathbb{R}^d})$ is positive definite. But then, by the second part of Proposition 2.26, there is a nonnegative Borel measure μ for which*

$$\varphi_r(||x - y||_{\mathbb{R}^d}) = \int_0^\infty e^{-||x-y||^2_{\mathbb{R}^d} t^2} \,\mathrm{d}\mu(t)\,.$$

In particular, φ_r can never be negative or equal zero by the positivity of the exponential on real arguments and nonnegativity of the measure: for any distance $d_0 \geq 0$, $\varphi_r(d_0) = \int_0^\infty e^{-d_0^2 t^2} \,\mathrm{d}\mu(t) > 0$ unless μ is the zero measure (in which case φ is the zero kernel).

This is why familiar reproducing kernels on the real line like the cardinal B-spline of even order (which is compactly supported) or Paley-Wiener sinc kernel of the Shannon-Whittaker-Kotel'nikov series (whose oscillations around zero bring its value "beyond the zero") are not radial basis functions. However, one can use Bochner's theorem to define sinc-like kernels in a given dimension by taking the Fourier transform of the indicator function over a domain such as a rectangle, a disk, or a hexagon in the plane (or extensions thereof in higher dimensions).

2.1.2 Mercer Kernels

Historically, the first proof of Proposition 2.11 for a non-finite index set $\mathcal{X}$ is due to Mercer; in this case, $\mathcal{X}$ was a real interval, though the result is readily generalized to any compact Hausdorff space endowed with a strictly positive, finite Borel measure [93]. We give a version of this result now, as it aids in the design of kernels on the sphere.

Proposition 2.28 (Mercer). *Let $\mathcal{X}$ be a closed, bounded region in $\mathbb{R}^d$. Consider a* Hilbert-Schmidt integral operator *on $L^2(\mathcal{X})$, that is, an operator L_k*

$$\mathrm{L}_k : L^2(\mathcal{X}) \to L^2(\mathcal{X})$$
$$f \mapsto \mathrm{L}_k f, \text{ with } (\mathrm{L}_k f)(x) = \int_{\mathcal{X}} k(x,y) f(y)\,\mathrm{d}y \text{ and } \iint_{\mathcal{X}\times\mathcal{X}} |k(x,y)|^2\,\mathrm{d}x\,\mathrm{d}y = B < \infty. \tag{10}$$

Suppose k is symmetric. Then

1. *The operator L_k is compact and self-adjoint.*
2. *The eigenfunctions of L_k $\{\phi_n\}_{n=0}^{\infty}$ form a complete orthonormal system for $L^2(\mathcal{X})$. Moreover, we can expand k into products of these eigenfunctions, weighted by the corresponding real eigenvalues λ_n*
$$k(x,y) = \sum_{n=0}^{\infty} \lambda_n \phi_n(x)\phi_n(y);$$
this series expansion converges to k in $L^2(\mathcal{X}\times\mathcal{X})$.
3. *If, moreover, k is continuous in both arguments, and $\lambda_n > 0$ for all $n \in \mathbb{N}$,*[15] *then the expansion of the kernel in the eigenfunctions*
$$k(x,y) = \sum_{n=0}^{\infty} \lambda_n \phi_n(x)\phi_n(y)$$
converges uniformly, not just in the mean.

Proof. We first observe that the operator L_k is bounded

$$\begin{aligned} ||\mathrm{L}_k f||^2_{L^2(\mathcal{X})} &= \int_{\mathcal{X}} |(\mathrm{L}_k f)(x)|^2\,\mathrm{d}x = \int_{\mathcal{X}} \left| \int_{\mathcal{X}} k(x,y) f(y)\,\mathrm{d}y \right|^2 \mathrm{d}x \\ &\le \iint_{\mathcal{X}\times\mathcal{X}} |k(x,y)|^2\,\mathrm{d}x\,\mathrm{d}y \int_{\mathcal{X}} |f(y)|^2\,\mathrm{d}y = B||f||_{\mathcal{H}}, \end{aligned}$$

by the Cauchy-Schwarz inequality, the integral absolute value inequality, the kernel boundedness condition (10), and Fubini's theorem, which guarantees that $\int_{\mathcal{X}} |k(x,y)|\,\mathrm{d}y$ is finite a.e. and integrable in x.

1. Then compactness follows, for instance, via [78], Theorem 5.1, and the finite norm of k in $L^2(\mathcal{X}\times\mathcal{X})$ stipulated above. If $\{\phi_n\}_{n=0}^{\infty}$ is complete orthonormal sequence in $L^2(\mathcal{X})$, then $\{\xi_l\}_{l=1}^{\infty} \overset{\text{def}}{=} \{\phi_n\phi_m\}_{n,m=1}^{\infty}$ is a complete orthonormal sequence in $L^2(\mathcal{X}\times\mathcal{X})$ and the Fourier coefficients of k on this basis are in ℓ^2. Indeed, by the Parseval relation, the Fourier coefficients
$$\lambda_l = \langle k, \xi_l \rangle_{L^2(\mathcal{X}\times\mathcal{X})} = \iint_{\mathcal{X}\times\mathcal{X}} k(x,y)\xi_l(x,y)\,\mathrm{d}x\,\mathrm{d}y > 0 = \langle \mathrm{L}_k\phi_m, \phi_n \rangle_{L^2(\mathcal{X})}$$
satisfy
$$\sum_{l=0}^{\infty} \lambda_l^2 = \iint_{\mathcal{X}\times\mathcal{X}} |k(x,y)|^2\,\mathrm{d}x\,\mathrm{d}y = B < \infty.$$

[15] Or, more generally, all but a finite number of nonzero eigenvalues have the same sign [121].

When the Mercer assumption of kernel continuity holds, compactness can also be shown more directly via Ascoli's theorem, using the uniform boundedness and equicontinuity of the image of any bounded sequence of functions under the operator L_k.

Self-adjointness follows from the kernel's symmetry

$$\langle \mathrm{L}_k f, g\rangle_{L^2(\mathcal{X})} = \int_{\mathcal{X}} g(x) \int_{\mathcal{X}} k(x,y) f(y)\,\mathrm{d}y\,\mathrm{d}x = \int_{\mathcal{X}} f(y) \int_{\mathcal{X}} k(y,x) g(x)\,\mathrm{d}x\,\mathrm{d}y = \langle f, \mathrm{L}_k g\rangle_{L^2(\mathcal{X})}$$

(see [121], Section 92), since L_k is bounded.

2. Thanks to item 1, we can invoke the spectral theory of compact self-adjoint operators ([36], Theorem 4.10.1 and Corollary 4.10.2) to show 2. The key is to use the orthonormal system $\{\phi_n\}_{n=0}^{\infty}$ of $L^2(\mathcal{X})$ to form the orthonormal system $\{\xi_l\}_{l=1}^{\infty}$ for $L^2(\mathcal{X}\times\mathcal{X})$ and expand k as a Fourier series on that basis; the Fourier coefficients of this expansion are the eigenvalues of the integral operator (10). The result was first stated in Erhard Schmidt's thesis [127], and proofs can be found in [121], Section 97, or [78], Theorem 6.2.

3. Originally shown by Mercer in 1909 [102]. A more concise proof uses Dini's theorem to show that the monotonically increasing partial sums of nonnegative continuous terms converging pointwise to the continuous $k(x,x)$ thereby converge uniformly

$$\forall x \in \mathcal{X},\ \sum_{n=0}^{N} \lambda_n \phi_n(x)^2 \overset{\text{uniformly}}{\longrightarrow} k(x,x).$$

The Cauchy-Schwarz relation

$$|k(x,y)| \le k(x,x)^{1/2} k(y,y)^{1/2}$$

allows this convergence result to be generalized to all pairs $(x,y) \in \mathcal{X}^2$. Details are given in Riesz and Sz.-Nagy [121], Section 98, and Jörgens [74], Theorem 8.11. A similar proof in a more general context is available in the appendix of [86].

□

Definition 2.29 (Mercer kernel)**.** *We will call any symmetric positive-definite kernel defined on a compact set $\mathcal{X}$ a* Mercer kernel *if it is continuous in both arguments (ensuring that the square integrability constraint (10) is satisfied) and the eigenvalues λ_n of the associated Hilbert-Schmidt integral operator L_k are all nonnegative.*

The assumption of eigenvalue nonnegativity turns out to be unnecessary; that k is a continuous positive-definite kernel on a compact set implies that $\lambda_n \ge 0$ and $\{\lambda_n\}_{n=0}^{\infty} \in \ell^1$ (see [34], Chapter III, Proposition 2 and Corollary 3; [46], Theorem 1.1; or [86], Lemma 1).

Remark 2.30. *The continuity of k in both arguments and compactness of $\mathcal{X}$ are sufficient to ensure that the image of any $L^2(\mathcal{X})$ function under L_k is continuous; in particular, the eigenfunctions $\{\phi_n\}_{n=0}^{\infty}$ associated with eigenvalues $\lambda_n > 0$ are all continuous [34]. With μ denoting the Lebesgue measure, we have, by the Cauchy-Schwarz inequality, that*

$$\begin{aligned}
|(\mathrm{L}_k f)(x) - (\mathrm{L}_k f)(x')| &= \left| \int_{\mathcal{X}} \left(k(x,y) - k(x',y)\right) f(y)\,\mathrm{d}y \right| = \left| \langle k_x - k_{x'}, f\rangle_{L^2(\mathcal{X})} \right| \\
&\le ||k_x - k_{x'}||_{L^2(\mathcal{X})} \cdot ||f||_{L^2(\mathcal{X})} \\
&\le \sqrt{\mu(\mathcal{X})} \cdot \underbrace{\max_{y\in\mathcal{X}} |k(x,y) - k(x',y)|}_{M} \cdot ||f||_{L^2(\mathcal{X})}.
\end{aligned}$$

Since k is continuous and $\mathcal{X}$ is compact, k is uniformly continuous, and the product $\sqrt{\mu(\mathcal{X})}\cdot M < \infty$. The image of any $L^2(\mathcal{X})$ function under L_k is therefore continuous[16].

Remark 2.31. *Any non-degenerate Borel measure may be used in the definition of the integral operator; this will affect the eigenfunctions and eigenvalues, as well as the representers of evaluation, but not the unique RKHS associated with the kernel on $\mathcal{X}$ [93]. Thus, although there is a unique RKHS associated with a kernel, the feature map needs not be unique.*

Remark 2.32. *Note that the condition (10) guarantees that the sequence of eigenvalues of the operator L_k resides in ℓ^2, since, by the orthonormality of the $\{\phi_n\}_{n=0}^{\infty}$ (and Tonelli's theorem)*

$$\iint\limits_{\mathcal{X}\times\mathcal{X}} |k(x,y)|^2 \,\mathrm{d}x\,\mathrm{d}y = \iint\limits_{\mathcal{X}\times\mathcal{X}} \left(\sum_{n=0}^{\infty} \lambda_n \phi_n(x)\phi_n(y)\right)\left(\sum_{n=0}^{\infty} \lambda_n \phi_n(x)\phi_n(y)\right) \mathrm{d}x\,\mathrm{d}y = \sum_{n=0}^{\infty} \lambda_n^2.$$

In fact, for Mercer kernels, the sequence $\{\lambda_n\}_{n=0}^{\infty}$ resides in ℓ^1, and L_k has finite trace, since

$$k(x,x) = \sum_{n=0}^{\infty} \lambda_n \phi_n(x)^2 \implies \sum_{n=0}^{\infty} |\lambda_n| = \sum_{n=0}^{\infty} \lambda_n = \sum_{n=0}^{\infty} \lambda_n \int_{\mathcal{X}} \phi_n(x)^2 \,\mathrm{d}x = \int_{\mathcal{X}} k(x,x)\,\mathrm{d}x < \infty.$$

The spectral theory of compact self-adjoint operators requires only that the real $\lambda_n \to 0$ as $n \to \infty$. This condition placed on the operator eigenvalues is strengthened greatly in the case of Hilbert-Schmidt integral operators.

It is this condition–that the sequence of eigenvalues of L_k decays sufficiently quickly to remain in ℓ^2–that gives us a valid inner product when all eigenvalues $\lambda_n > 0$.

Corollary 2.33. *Consider a Hilbert-Schmidt integral operator L_k of a Mercer kernel k. Let $\{\phi_n\}_{n=0}^{\infty}$ be its continuous eigenfunctions and $\{\lambda_n\}_{n=0}^{\infty}$ the corresponding eigenvalues. Then for all $(x,y) \in \mathcal{X}^2$, the Fourier expansion of $k(x,y)$ is an inner product in ℓ^2 of the images of x and y under the continuous feature map*

$$\begin{aligned} \phi : \mathcal{X} &\to \ell^2 \\ x &\mapsto \phi(x) = (\sqrt{\lambda_0}\phi_0(x), \sqrt{\lambda_1}\phi_1(x), \ldots). \end{aligned} \tag{11}$$

Proof. By our construction, $k(x,y) = \sum_{n=0}^{\infty} \lambda_n \phi_n(x)\phi_n(y) = \langle \phi(x), \phi(y)\rangle_{\ell^2}$. Every $x \in \mathcal{X}$ is mapped to an element of ℓ^2 since

$$||\phi(x)||_{\ell^2}^2 = \sum_{n=0}^{\infty} \lambda_n \phi_n(x)^2 = k(x,x) < \infty.$$

The feature map is continuous because $x_n \longrightarrow x$ implies

$$||\phi(x_n) - \phi(x)||_{\ell^2}^2 = \langle \phi(x_n) - \phi(x), \phi(x_n) - \phi(x)\rangle_{\ell^2} = k(x_n,x_n) + k(x,x) - 2k(x,x_n) \longrightarrow 0,$$

by the continuity of k. □

This feature map ϕ, depends on our choice of measure in the definition of the integral operator (10), as do the eigenvalues $\{\lambda_n\}_{n=0}^{\infty}$ and eigenvectors $\{\phi_n\}_{n=0}^{\infty}$. Nevertheless, we can use these eigenvalues and eigenvectors to identify the unique RKHS associated with the continuous positive-definite kernel k on the compact region $\mathcal{X}$.

[16] Some authors define the Mercer integral operator as the composition of L_k with the inclusion map, in which case it maps $L^2(\mathcal{X})$ equivalence classes to $L^2(\mathcal{X})$ equivalence classes of functions that coincide μ-a.e. with a continuous function.

Corollary 2.34. *Let k be Mercer kernel defined on a compact $\mathcal{X}$ and L_k the associated Hilbert-Schmidt integral operator, with eigenfunctions $\{\phi_n\}_{n=0}^{\infty}$ and corresponding eigenvalues $\{\lambda_n\}_{n=0}^{\infty}$. Then the RKHS $\mathcal{H}$ associated with k (and $\mathcal{X}$) is defined as*

$$\mathcal{H} = \left\{ f \in L^2(\mathcal{X}) \;\middle|\; f \overset{L^2(\mathcal{X})}{\sim} \sum_{n=0}^{\infty} (f)_n \phi_n \text{ such that } \sum_{n=0}^{\infty} \frac{(f)_n^2}{\lambda_n} < \infty \right\}, \tag{12}$$

and is endowed with the inner product between f with Fourier coefficients $(f)_n = \langle f, \phi_n \rangle_{L^2(\mathcal{X})}$ and g with Fourier coefficients $(g)_n = \langle g, \phi_n \rangle_{L^2(\mathcal{X})}$

$$\langle f, g \rangle_{\mathcal{H}} = \sum_{n=0}^{\infty} \frac{\langle f, \phi_n \rangle_{L^2(\mathcal{X})} \langle \phi_n, g \rangle_{L^2(\mathcal{X})}}{\lambda_n} = \sum_{n=0}^{\infty} \frac{(f)_n (g)_n}{\lambda_n}. \tag{13}$$

Proof. First note that $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ is indeed a valid inner product; in particular, it is definite, as $\lambda_n > 0$ for all n by the Mercer condition[17]. Observe that the operator L_k is positive by the Mercer condition. Thus, we can define the operator[18]

$$\mathrm{L}_k^{1/2} : L^2(\mathcal{X}) \to \mathcal{H}$$
$$f \overset{L^2(\mathcal{X})}{\sim} \sum_{n=0}^{\infty} \underbrace{\langle f, \phi_n \rangle_{L^2(\mathcal{X})}}_{(f)_n} \phi_n \mapsto \mathrm{L}_k^{1/2} f = \sum_{n=0}^{\infty} (f)_n \sqrt{\lambda_n} \phi_n, \tag{14}$$

which is an isomorphism (by construction), so $\mathcal{H}$ is a separable Hilbert space since $L^2(\mathcal{X})$ is a separable Hilbert space. The series on the right-hand side converges pointwise since it converges in the norm (by the compactness of L_k and thus $\mathrm{L}_k^{1/2}$) and thus pointwise (since $\mathcal{H}$ is an RKHS, as we will confirm).

We know, moreover, that, being the eigenfunctions of the Hilbert-Schmidt integral operator of a Mercer kernel, the ϕ_n are all continuous (see Remark 2.30). For any $x \in \mathcal{X}$, define the representer of evaluation at x as

$$k_x = \sum_{n=0}^{\infty} \underbrace{\lambda_n \phi_n(x)}_{(k_x)_n} \phi_n; \tag{15}$$

then we can recover the Hilbert-Schmidt expansion using

$$\forall (x, y) \in \mathcal{X}^2,\ k(x, y) = \langle k_x, k_y \rangle_{\mathcal{H}} = \sum_{n=0}^{\infty} \frac{(k_x)_n (k_y)_n}{\lambda_n} = \sum_{n=0}^{\infty} \lambda_n \phi_n(x) \phi_n(y), \tag{16}$$

and the representers of evaluation k_x inhabit $\mathcal{H}$

$$||k_x||_{\mathcal{H}}^2 = \sum_{n=0}^{\infty} \frac{(k_x)_n^2}{\lambda_n} = \sum_{n=0}^{\infty} \lambda_n \phi_n(x)^2 = k(x, x) < \infty.$$

The reproducing property then follows

$$\forall f \in \mathcal{H},\ \langle f, k_x \rangle_{\mathcal{H}} = \sum_{n=0}^{\infty} \frac{(f)_n (k_x)_n}{\lambda_n} = \sum_{n=0}^{\infty} \frac{\langle f, \phi_n \rangle_{L^2(\mathcal{X})} \cdot \lambda_n \phi_n(x)}{\lambda_n} = \sum_{n=0}^{\infty} (f)_n \phi_n(x) = f(x). \tag{17}$$

□

[17] This can be generalized to $\lambda \geq 0$ by excluding the eigenfunctions ϕ_n associated with $\lambda_n = 0$ from our definition of $\mathcal{H}$, to avoid a division by 0 in the inner product definition without breaking the reproducing property or altering the Hilbert-Schmidt expansion. Then $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ will still be a valid inner product over $\mathcal{H}$. In fact, we need only require that all but finitely many of the λ_n have the same sign [121]. See [34], Chapter 3, Remark 3.

[18] Here we use continuous functions as representers of equivalence classes of functions that coincide almost everywhere with them. See [34] for a presentation where $\mathrm{L}_k^{1/2}$ is composed with the inclusion map, as is suggested by (12) rather than our expansion (14).

In summary, we started with a Mercer kernel k, which gave us a complete orthonormal system $\{\phi_n\}_{n=0}^{\infty}$ for $L^2(\mathcal{X})$, where the ϕ_n are continuous eigenfunctions (with eigenvalues $\lambda_n > 0$) of the corresponding Hilbert-Schmidt integral operator L_k. Associated with this kernel is the unique RKHS $\mathcal{H}$ with representers of evaluation at x given by (15). Thanks to the nonnegativity of the eigenvalues, we can write a complete orthonormal system for $\mathcal{H}$ using the eigenvalues and eigenfunctions of L_k: $\{\sqrt{\lambda_n}\phi_n\}_{n=0}^{\infty}$ (orthonormality is immediate; using the definition of $\langle\cdot,\cdot\rangle_{\mathcal{H}}$, completeness follows since any function $f \in \mathcal{H}$ orthogonal to each of the ϕ_n has $L^2(\mathcal{X})$ Fourier coefficients $(f)_n = 0$ for all $n \in \mathbb{N}$ and thus is identically 0). We found two ways to express the kernel evaluations $k(x, y)$ as an inner product: in ℓ^2, between $\phi(x)$ and $\phi(y)$ (defined in (11)), and in $\mathcal{H}$, between the representers of evaluation, k_x and k_y (given in Equation (16)).

2.2 Synthesizing Mercer Kernels on the Sphere

In this subsection, we derive positive-definite functions on the sphere by running the Mercer theorem (Proposition 2.28) in reverse: we start with the sequences $\{\lambda_n\}_{n=0}^{\infty}$ of nonnegative eigenvalues and the continuous eigenfunctions $\{\phi_n\}_{n=0}^{\infty}$ and use them to synthesize a continuous kernel k. We begin this work while continuing to let $\mathcal{X}$ be an arbitrary closed region in Euclidean space; later, we will introduce results that are specific to $\mathcal{X} = \mathbb{S}^2$.

2.2.1 Mercer Synthesis

We saw how a continuous kernel on a compact Euclidean domain can be expressed as the *uniformly convergent* series

$$k(x, y) = \sum_{n=0}^{\infty} \lambda_n \phi_n(x)\phi_n(y) \,\forall (x, y) \in \mathcal{X} \times \mathcal{X},$$

where the weights λ_n are nonnegative eigenvalues in ℓ^1 and orthonormal basis functions ϕ_n continuous eigenfunctions of the associated Hilbert-Schmidt integral operator.

In this section, we consider the converse: can we instead start with a family $\{\phi_n\}_{n=0}^{\infty}$ of continuous functions defined on the compact Euclidean domain $\mathcal{X}$ that form a complete orthonormal system for $L^2(\mathcal{X})$, as well as an ℓ^1 sequence of weights $\{\lambda_n\}_{n=0}^{\infty}$ with $\lambda_n > 0$ for all $n \in \mathbb{N}$ and derive a continuous kernel? We require a few lemmas.

If $\mathcal{X}$ is compact, we can prove a result that follows from parts 1 and 3 of Proposition 2.24, but that is nice to revisit in light of (13).

Lemma 2.35 (Strong convergence implies uniform convergence when $\mathcal{X}$ compact and k diagonally continuous)**.** *Suppose $\mathcal{H} \subseteq \mathbb{R}^{\mathcal{X}}$ is an RKHS of the form (12) with inner product (13) and "diagonally continuous" kernel (i.e., the map $x \mapsto k(x, x)$ is continuous). Then convergence in the RKHS norm implies uniform convergence.*

Proof. Suppose $\{f_m\}_{m=0}^{\infty}$ converges to f in $\mathcal{H}$. Letting $\{(f_m)_n\}_{n=0}^{\infty}$ be the Fourier coefficients of f_m on the orthonormal system $\{\phi_n\}_{n=0}^{\infty}$ for $L^2(\mathcal{X})$, then for any $x \in \mathcal{X}$,

$$\begin{aligned} |f_m(x) - f(x)| &= \left|\sum_{n=0}^{\infty}((f_m)_n - (f)_n)\phi_n(x)\right| = \left|\sum_{n=0}^{\infty} \frac{(f_m)_n - (f)_n}{\sqrt{\lambda_n}}\sqrt{\lambda_n}\phi_n(x)\right| \\ &\leq \left(\sum_{n=0}^{\infty} \frac{((f_m)_n - (f)_n)^2}{\lambda_n}\right)^{1/2} \left(\sum_{n=0}^{\infty} \lambda_n \phi_n(x)\phi_n(x)\right)^{1/2} = ||f_m - f||_{\mathcal{H}}\sqrt{k(x, x)}; \end{aligned}$$

thus, for all $\epsilon > 0$, there exists N such that for all $m > N$ and all $x \in \mathcal{X}$,

$$m > N \implies ||f_m - f||_{\mathcal{H}} < \frac{\epsilon}{\max\limits_{x \in \mathcal{X}} \sqrt{k(x,x)}} \implies |f_m(x) - f(x)| < \epsilon.$$

The result holds because the the continuous map $x \mapsto k(x,x)$ attains a maximum value on the compact $\mathcal{X}$. □

Lemma 2.36 (Fourier expansion of a reproducing kernel)**.** *Given an RKHS $\mathcal{H} \subseteq \mathbb{R}^{\mathcal{X}}$ with associated reproducing kernel k and an orthonormal system (not necessarily complete) $\{\psi_n\}_{n=0}^{\infty}$ in $\mathcal{H}$, then*

$$\forall x \in \mathcal{X}, \sum_{n=0}^{\infty} \psi_n(x)^2 \leq k(x,x).$$

Proof. By the reproducing property, the Fourier expansion of the representer of evaluation at any $x \in \mathcal{X}$ on $\{\psi_n\}_{n=0}^{\infty}$ can be written

$$k_x = \sum_{n=0}^{\infty} \langle k_x, \psi_n \rangle_{\mathcal{H}} \psi_n = \sum_{n=0}^{\infty} \psi_n(x) \psi_n.$$

Thus, Bessel's inequality gives

$$\sum_{n=0}^{\infty} \psi_n(x)^2 = \sum_{n=0}^{\infty} |\langle k_x, \psi_n \rangle_{\mathcal{H}}|^2 \leq ||k_x||_{\mathcal{H}}^2 = k(x,x).$$

This becomes equality when the system $\{\psi_n\}_{n=0}^{\infty}$ is complete. In the Fourier-weighted Hilbert space, $\psi_n = \sqrt{\lambda_n}\phi_n$, so

$$k(x,x) = \sum_{n=0}^{\infty} \lambda_n \phi_n(x)^2,$$

where $\{\phi_n\}_{n=0}^{\infty}$ is a complete orthonormal system for $L^2(\mathcal{X})$. □

Remark 2.37 (Pointwise convergence of Riesz-Fischer limits in an RKHS)**.** *The Riesz-Fischer theorem*[19] *states that, given an orthonormal system $\{\psi_n\}_{n=0}^{\infty}$ (not necessarily complete) in a Hilbert space $\mathcal{H}$ and a sequence of weights $\{(g)_n\}_{n=0}^{\infty}$ in ℓ^2, the sum*

$$\sum_{n=0}^{\infty} (g)_n \psi_n \longrightarrow g$$

converges strongly in $\mathcal{H}$ to a limit g. If $\mathcal{H}$ is, moreover, an RKHS, over an index set $\mathcal{X}$, then by Remark 2.4, this convergence is also pointwise, and for all $x \in \mathcal{X}$,

$$\begin{aligned} |g(x)| &= \left| \left\langle \sum_{n=0}^{\infty} (g)_n \psi_n, k_x \right\rangle_{\mathcal{H}} \right| = \left| \sum_{n=0}^{\infty} (g)_n \langle \psi_n, k_x \rangle_{\mathcal{H}} \right| = \left| \langle \{(g)_n\}_{n=0}^{\infty}, \{\psi_n(x)\}_{n=0}^{\infty} \rangle_{\ell^2} \right| \\ &\leq \left(\sum_{n=0}^{\infty} (g)_n^2 \right)^{1/2} \left(\sum_{n=0}^{\infty} \psi_n(x)^2 \right)^{1/2} \leq \sqrt{k(x,x)} \left(\sum_{n=0}^{\infty} (g)_n^2 \right)^{1/2}, \end{aligned}$$

by the Cauchy-Schwarz inequality and Lemma 2.36.

[19] The theorem we refer to was given in [47, 48, 119] in the context of $L^2(\mathbb{R})$ and quickly generalized to arbitrary separable Hilbert spaces, as in [82], Section 16, Theorem 9, or [36], Theorem 3.4.10.

Proposition 2.38 (Fourier-weighted Hilbert space synthesis of a Mercer kernel)**.** *Suppose we are given a compact region in $\mathbb{R}^d$ $\mathcal{X}$, a sequence of weights $\{\lambda_n\}_{n=0}^{\infty} \in \ell^1$ with $\lambda_n > 0$ for all $n \in \mathbb{N}$, and a sequence of functions $\{\phi_n\}_{n=0}^{\infty}$ that are continuous on $\mathcal{X}$ and form a complete orthonormal system for $L^2(\mathcal{X})$. Moreover, to ensure the local boundedness of*

$$k(x,x) = \sum_{n=0}^{\infty} \lambda_n \phi_n(x)^2$$

for each $x \in \mathcal{X}$, we make the sufficient (but not necessary) assumption that the family $\{\phi_n\}_{n=0}^{\infty}$ is locally bounded, i.e., for all $x \in \mathcal{X}$ there exists a neighborhood $B(x)$ such that for all $n \in \mathbb{N}$ and all $y \in B(x)$, we have that $|\phi_n(y)| < M_x$.

Then the weights $\{\lambda_n\}_{n=0}^{\infty}$ and complete orthonormal system $\{\phi_n\}_{n=0}^{\infty}$ synthesize a Mercer kernel in the sense that the kernel

$$k(x,y) = \sum_{n=0}^{\infty} \lambda_n \phi_n(x)\phi_n(y),$$

which converges in $L^2(\mathcal{X}\times\mathcal{X})$ by the Riesz-Fischer theorem, also converges uniformly to a continuous kernel.

Proof. Define[20]

$$\langle f, g\rangle_{\mathcal{H}} = \sum_{n=0}^{\infty} \frac{(f)_n (g)_n}{\lambda_n},$$

and

$$\mathcal{H} = \left\{ f \in L^2(\mathcal{X}) \,\middle|\, f \overset{L^2(\mathcal{X})}{\sim} \sum_{n=0}^{\infty} (f)_n \phi_n \text{ and } \sum_{n=0}^{\infty} \frac{(f)_n^2}{\lambda_n} < \infty \right\}.$$

By construction, the Fourier-weighting operator

$$\begin{aligned} \mathrm{L}_k^{1/2} &: L^2(\mathcal{X}) \to \mathcal{H} \\ f \overset{L^2(\mathcal{X})}{\sim} \sum_{n=0}^{\infty} (f)_n \phi_n &\mapsto \mathrm{L}_k^{1/2} f = \sum_{n=0}^{\infty} (f)_n \sqrt{\lambda_n}\phi_n, \end{aligned}$$

is an isometry

$$||f||_{L^2(\mathcal{X})}^2 = \sum_{n=0}^{\infty} (f)_n^2 = \sum_{n=0}^{\infty} \frac{(\sqrt{\lambda_n}(f)_n)^2}{\lambda_n} = \left\| \mathrm{L}_k^{1/2} f \right\|_{\mathcal{H}}^2,$$

the Fourier weighting injection an isomorphism. Thus, $\mathcal{H}$ is a separable Hilbert space, with complete orthonormal system $\{\sqrt{\lambda_n}\phi_n\}_{n=0}^{\infty}$.

The reproducing property (17) still holds, by construction, and the representers of evaluation are given by

$$\forall x \in \mathcal{X},\ k_x = \sum_{n=0}^{\infty} \lambda_n \phi_n(x)\phi_n.$$

The ϕ_n are continuous, and therefore uniformly continuous (since $\mathcal{X}$ is compact). We now show the kernel is continuous. First, observe that for all $x \in \mathcal{X}$, the representer of evaluation at x

$$k_x = \sum_{n=0}^{\infty} \lambda_n \phi_n(x)\phi_n = \sum_{n=0}^{\infty} \left(\sqrt{\lambda_n}\phi_n(x)\right)(\sqrt{\lambda_n}\phi_n),$$

[20] This generalizes immediately to the case where we merely require that the $\lambda_n \geq 0$: we simply exclude those terms where $\lambda_n = 0$ from the following sums.

is in $\mathcal{H}$ by the Riesz-Fischer theorem, since $\{\sqrt{\lambda_n}\phi_n(x)\}_{n=0}^{\infty} \in \ell^2$, and, moreover, is continuous, since for any sequence of points $\{x_m\}_{m=0}^{\infty}$ in $\mathcal{X}$ converging to $x \in \mathcal{X}$ and any $y \in \mathcal{X}$, we have that

$$\lim_{m\to\infty} x_m = x \implies \lim_{m\to\infty} k_y(x_m) = \lim_{m\to\infty} \sum_{n=0}^{\infty} \lambda_n \phi_n(y)\phi_n(x_m) = \sum_{n=0}^{\infty} \lim_{m\to\infty} \lambda_n \phi_n(y)\phi_n(x_m)$$
$$= \sum_{n=0}^{\infty} \lambda_n \phi_n(y)\phi_n(x) = k_y(x),$$

by Tannery's theorem (since the set $\{\phi_n\}_{n=0}^{\infty}$ is locally bounded, $|\lambda_n\phi_n(y)\phi_n(x_m)| \leq \frac{M_y M_x}{n} \longrightarrow 0$ for sufficiently large m and n, where M_x is a local bound of $\{\phi_n\}_{n=0}^{\infty}$ in a neighborhood of x).

We now make the recognition that $x \mapsto k_x(x) = k(x,x)$ is locally bounded, since

$$\sum_{n=0}^{\infty} \lambda_n \phi_n(x)^2 \leq M_x^2 \sum_{n=0}^{\infty} \lambda_n,$$

and $\{\lambda_n\}_{n=0}^{\infty} \in \ell^1$.

Thus, by Proposition 2.24, part 2, the functions in $\mathcal{H}$ are all continuous on $\mathcal{X}$. Using this, we can see that k_{x_m} converges weakly in $\mathcal{H}$ to k_x,

$$\forall\, f \in \mathcal{H},\ \lim_{m\to\infty} \langle f, k_{x_m}\rangle_{\mathcal{H}} = \lim_{m\to\infty} f(x_m) = f(x) = \langle f, k_x\rangle_{\mathcal{H}}$$

and, moreover,

$$\lim_{m\to\infty} ||k_{x_m}||_{\mathcal{H}}^2 = \lim_{m\to\infty} \sum_{n=0}^{\infty} \lambda_n \phi_n(x_m)^2 = \sum_{n=0}^{\infty} \lim_{m\to\infty} \lambda_n \phi_n(x_m)^2 = ||k_x||_{\mathcal{H}}^2;$$

thus, k_{x_m} converges strongly to k_x, so that the map $\phi : x \mapsto k_x$ is continuous and k is continuous on the diagonal (by Proposition 2.24, part 1). It remains to be seen that the convergence is uniform.

We notice that the pointwise convergence of the monotonically increasing sequence of functions

$$k_M(x,x) = \sum_{n=0}^{M} \lambda_n \phi_n(x)^2$$

to the continuous function

$$k(x,x) = \sum_{n=0}^{\infty} \lambda_n \phi_n(x)^2$$

is uniform, by Dini's theorem. By the Cauchy-Schwarz inequality, we have also that

$$|k(x,y)| = \left|\sum_{n=0}^{\infty} \lambda_n \phi_n(x)\phi_n(y)\right| \leq \sum_{n=0}^{\infty} \sqrt{\lambda_n}\,|\phi_n(x)|\,\sqrt{\lambda_n}\,|\phi_n(y)|$$
$$\leq \left(\sum_{n=0}^{\infty} \lambda_n \phi_n(x)^2\right)^{1/2} \left(\sum_{n=0}^{\infty} \lambda_n \phi_n(y)^2\right)^{1/2} \leq \sqrt{k(x,x)} \cdot \sqrt{k(y,y)}.$$

Thus, the sequence of partial sums $\sum_{n=0}^{N} \lambda_n \phi_n(x)\phi_n(y)$ converges uniformly to $k(x,y)$ on $\mathcal{X} \times \mathcal{X}$. By the uniform limit theorem, k is continuous on $\mathcal{X} \times \mathcal{X}$. □

Remark 2.39. *The authors of [78] call the space $\mathcal{H}$ derived this way a "Fourier weighted Hilbert space" because it applies selection to the elements of $L^2(\mathcal{X})$ via the sequence of nonnegative weights $\{\lambda_n\}$, which must penalize large Fourier coefficients $(f)_n$ for large n, so that the sequence of coefficients $\{(f)_n/\sqrt{\lambda_n}\}_{n=0}^{\infty}$ may remain in ℓ^2 and thus give, by the Riesz-Fischer theorem, the expansion weights of an element of $\mathcal{H}$. ($L^2(\mathcal{X})$ itself, which is not an RKHS, is (re)constructed by a much less selective sequence of weights, not in ℓ^1, given by $\lambda_n = 1\,\forall n \in \mathbb{N}$; nonnegative weight sequences in ℓ^1 yield a Fourier weighted Hilbert space that is in fact an RKHS.)*

The construction of a Fourier weighted Hilbert space via Mercer synthesis is summarized in Algorithm 1.

Algorithm 1: Synthesis of a Mercer kernel and Fourier-weighted Hilbert space.

Data: An ℓ^1 sequence $\{\lambda_n\}_{n=0}^{\infty}$ of nonnegative weights $\lambda_n \geq 0$ and a locally bounded complete orthonormal system $\{\phi_n\}_{n=0}^{\infty}$ for $L^2(\mathcal{X})$. The ϕ_n are continuous on the compact set $\mathcal{X}$.

Result: A positive-definite kernel k on $\mathcal{X}$ and an associated RKHS $\mathcal{H}$.

Define the kernel

$$\forall (x,y) \in \mathcal{X}^2,\; k(x,y) = \sum_{n=0}^{\infty} \lambda_n \phi_n(x)\phi_n(y).$$

The corresponding RKHS $\mathcal{H}$

- consists of continuous functions on $\mathcal{X}$;
- has inner product $\langle f, g\rangle_{\mathcal{H}} = \sum_{\substack{n=0\\ \lambda_n>0}}^{\infty} \frac{\langle f,\phi_n\rangle_{L^2(\mathcal{X})}\langle g,\phi_n\rangle_{L^2(\mathcal{X})}}{\lambda_n} = \sum_{\substack{n=0\\ \lambda_n>0}}^{\infty} \frac{(f)_n (g)_n}{\lambda_n}$;
- satisfies $f \in \mathcal{H} \iff ||f||_{\mathcal{H}}^2 = \sum_{\substack{n=0\\ \lambda_n>0}}^{\infty} \frac{(f)_n^2}{\lambda_n} < \infty$;
- has complete orthonormal system $\{\sqrt{\lambda_n}\phi_n\}_{n=0}^{\infty}$;
- has representers of evaluation at $x \in \mathcal{X}$ given by $k_x = \sum_{\substack{n=0\\ \lambda_n>0}}^{\infty} \lambda_n \phi_n(x)\phi_n$.

Return the Mercer kernel k and RKHS $\mathcal{H}$.

2.2.2 The Real Spherical Harmonics: A Complete Orthonormal System for $L^2(\mathbb{S}^2)$

To synthesize kernels on "Fourier side" we require a complete orthonormal system for $L^2(\mathbb{S}^2)$. A convenient choice is the real spherical harmonics $\{Y_l^n \mid l \in \mathbb{N} \text{ and } n \in [\![-l, l]\!]\}$, which form a complete orthonormal system for $L^2(\mathbb{S}^2)$ with respect to the inner product

$$\langle f, g\rangle_{L^2(\mathbb{S}^2)} = \frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi} f(x)g(x)\sin(\theta)\,\mathrm{d}\theta\,\mathrm{d}\phi\,,$$

and act as eigenfunctions of the 2-sphere Laplace-Beltrami operator Δ_S

$$\Delta_S Y_l^n(\theta,\phi) = -\left[l(l+1)\right] Y_l^n(\theta,\phi). \tag{18}$$

(We have adopted the notational convention $[\![a, b]\!] = \{a, a+1, \ldots, b\}$.) In Section 2.7, it is this latter property (Equation (18)) that makes this choice of complete orthonormal system particularly suited to the thin-plate spherical splines.

To define the real spherical harmonics, we first define the complex spherical harmonics [160] as

$$Z_l^n(\theta,\phi) = \sqrt{\frac{2l+1}{4\pi}\frac{(l-n)!}{(l+n)!}} P_l^n(\cos(\theta))e^{in\phi},$$

where the P_l^n are *associated Legendre polynomials* [87, 126], which can be defined by the Rodrigues formula

$$\text{for } x\in[-1,1],\ P_l^n(x) \stackrel{\text{def}}{=} \begin{cases} \frac{(-1)^n}{2^l l!}(1-x^2)^{n/2}\frac{\mathrm{d}^{l+n}}{\mathrm{d}x^{l+n}}(x^2-1)^l, & \text{if } n\in[\![0,l]\!]; \\ \frac{(l+n)!}{(l-n)!}\frac{1}{2^l l!}(1-x^2)^{n/2}\frac{\mathrm{d}^{l+n}}{\mathrm{d}x^{l+n}}(x^2-1)^l, & \text{if } n\in[\![-l,-1]\!]. \end{cases}$$

Observe that for n even, $P_l^n(x)$ is a polynomial; that $P_l^{-n}(x)$ differs from $P_l^n(x)$ by a scale factor; and that $P_l^0(x)$ is the standard (not "associated") Legendre polynomial. The associated Legendre polynomials are the canonical solutions of the general Legendre equation of degree l and order n,

$$\frac{\mathrm{d}}{\mathrm{d}x}\left[(1-x^2)\frac{\mathrm{d}}{\mathrm{d}x}P_l^n(x)\right] + \left[l(l+1) - \frac{n^2}{1-x^2}\right]P_l^n(x) = 0,$$

which reduces to the Legendre equation if $n=0$. (The other solutions to this equation, called the associated Legendre functions of the second kind, have singularities at ± 1.)

From the complex spherical harmonics Z_l^n, we define the (real) spherical harmonics of degree $l\in\mathbb{N}$ and order $n\in[\![-l,l]\!]$ by extracting the real and imaginary parts and renormalizing [78]

$$Y_l^n(\theta,\phi) = \begin{cases} \sqrt{2}\mathrm{Re}\left[Z_l^n(\theta,\phi)\right] = \sqrt{2}\sqrt{\frac{2l+1}{4\pi}\frac{(l-n)!}{(l+n)!}}\cos(n\phi)P_l^n\left(\cos(\theta)\right), & n\in[\![-l,-1]\!]; \\ \sqrt{2}\mathrm{Im}\left[Z_l^n(\theta,\phi)\right] = \sqrt{2}\sqrt{\frac{2l+1}{4\pi}\frac{(l-n)!}{(l+n)!}}\sin(n\phi)P_l^n\left(\cos(\theta)\right), & n\in[\![1,l]\!]; \\ \sqrt{\frac{2l+1}{4\pi}}P_l^0(\cos(\theta)), & n=0. \end{cases}$$

A useful property of the spherical harmonics is the *addition theorem* for real or complex spherical harmonics [70, 163], an analogue of the addition formulas for sinusoids

$$\frac{4\pi}{2l+1}\sum_{n=-l}^{l} Y_l^n(\theta,\phi)Y_l^n(\theta',\phi') = P_l^0(\cos(\sphericalangle(p,p'))) \tag{19}$$

where P_l^0 are, as before, the (standard, un-associated) Legendre polynomials and $\sphericalangle(p,p')$ is the angle between $p=(\theta,\phi)$ and $p'=(\theta',\phi')$. Setting $p=p'$ and noting that $P_l^0(\cos(0)) = P_l^0(1) = 1$, we get a corollary, sometimes called Unsöld's theorem [163]

$$\sum_{n=-l}^{l} Y_l^n(\theta,\phi)^2 = \frac{2l+1}{4\pi}.$$

Let x be the map from spherical to Euclidean coordinates on the unit circle

$$\begin{aligned} x : [0,\pi]\times[0,2\pi) &\to \mathbb{R}^3 \\ (\theta,\phi) &\mapsto (\sin(\theta)\cos(\phi), \sin(\theta)\sin(\phi), \cos(\theta))^T, \end{aligned}$$

and $p=(\theta,\phi)^T$ and $p'=(\theta',\phi')^T$ be two points on the sphere. Then $x(\theta,\phi)$ and $x(\theta',\phi')$ are two Euclidean vectors on the unit sphere in $\mathbb{R}^3$, and the cosine of the angle $\sphericalangle(p,p')$ between them is, by the addition theorem for sinusoids, the following

$$\begin{aligned} \cos(\sphericalangle(p,p')) &= x(\theta,\phi)^T x(\theta',\phi') \\ &= \sin(\theta)\cos(\phi)\sin(\theta')\cos(\phi') + \sin(\theta)\sin(\phi)\sin(\theta')\sin(\phi') + \cos(\theta)\cos(\theta') \\ &= \cos(\theta)\cos(\theta') + \sin(\theta)\sin(\theta')\cos(\phi-\phi'). \end{aligned}$$

The spherical harmonics are not locally bounded. If a function f lies on the $(2l+1)$-dimensional space H_l of homogeneous polynomials of total degree l, i.e., span $\{Y_l^{-l}, Y_l^{-l+1}, \ldots, Y_l^l\}$, then

$$\sup_{p\in\mathbb{S}^2} |f(p)| \leq \sqrt{\frac{2l+1}{4\pi}} ||f||_{L^2(\mathbb{S}^2)};$$

moreover, every one of these spaces H_l has an element for which this upper bound is attained on some $t \in \mathbb{S}^2$ (see [52], Section 6).

However, the Legendre polynomials are bounded

$$\forall u \in [-1,1],\ |P_l^0(u)| \leq 1 \text{ ([105], Lemma 2, p. 43).}$$

The addition theorem therefore allows us to rewrite an expansion on a basis that is not locally bounded (i.e., the spherical harmonics) as an expansion on a basis (the Legendre polynomials) that is bounded everywhere in magnitude by unity.

Thus, even though the Riesz-Fischer theorem requires that, for every ℓ^2 sequence of weights $\{\lambda_{l,n}\}_{l=0,\,n=-l}^{\infty,\,l}$, the weighted sum of spherical harmonics must converge in $L^2(\mathbb{S}^2 \times \mathbb{S}^2)$

$$\sum_{l=0}^{\infty}\sum_{n=-l}^{l} \lambda_{l,n} Y_l^n(\theta,\phi) \xrightarrow{L^2(\mathbb{S}^2\times\mathbb{S}^2)} f,$$

we cannot apply Proposition 2.38 to synthesize a kernel using these weights

$$k(p,p') = \sum_{l=0}^{\infty}\sum_{n=-l}^{l} \lambda_{l,n} Y_l^n(\theta,\phi) Y_l^n(\theta',\phi'),$$

that converges for all pairs $(p,p') \in \mathbb{S}^2 \times \mathbb{S}^2$. Nevertheless, if we require that the weights $\{\lambda_{l,n}\}_{l=0,\,n=-l}^{\infty,\,l}$ be constant over each degree–i.e., for all $l \in \mathbb{N}$, $\lambda_{l,n} = \alpha_l \geq 0$–then we have that

$$\sum_{l=0}^{\infty}\sum_{n=-l}^{l} \lambda_{l,n} Y_l^n(\theta,\phi)^2 = \sum_{l=0}^{\infty} \alpha_l \sum_{n=-l}^{l} Y_l^n(\theta,\phi)^2 = \sum_{l=0}^{\infty} \alpha_l \frac{2l+1}{4\pi}$$

converges for all $p = (\theta,\phi) \in \mathbb{S}^2$ if and only if $\{\alpha_l(2l+1)\}_{l=0}^{\infty} \in \ell^1$. Then by the Cauchy-Schwarz inequality

$$k(p,p') = \sum_{l=0}^{\infty}\sum_{n=-l}^{l} \lambda_{l,n} Y_l^n(\theta,\phi) Y_l^n(\theta',\phi') \text{ converges for all } (p,p') \in \mathbb{S}^2 \iff \{\alpha_l(2l+1)\}_{l=0}^{\infty} \in \ell^1.$$

Even though the spherical harmonics are not locally bounded, we can apply Proposition 2.38 after requiring that the weights on the spherical harmonics exhibit a particular multiplicity, enabling the application of the addition theorem, which restates the kernel expansion in terms of the Legendre polynomials P_l^0.

As we will see in the next section, it turns out that this seemingly stringent restriction in the form of the kernel–our choosing, in the Hilbert-Schmidt synthesis sum, the spherical harmonics as the complete orthonormal sequence of $L^2(\mathbb{S}^2)$, as well as requiring that the ℓ^1 sequence of weights $\lambda_{l,n}$ in the sum adopt a constant value $\alpha_l \geq 0$ over each $(2l+1)$-multiplicity eigenspace H_l of the Laplace-Beltrami operator–is not limiting in practice, if we desire a property called isotropy.

2.2.3 Mercer Synthesis and Isotropic Kernels on the Sphere

On the sphere $\mathbb{S}^2$, the spherical harmonics (see Section 2.2.2) are eigenfunctions for the Hilbert-Schmidt integral operator of any continuous kernel that depends only on the cosine of the geodesic (great circle) angle between the two points being considered. To see this, let us first recall a formula that greatly simplifies the Hilbert-Schmidt expansion of any such kernel.

Proposition 2.40 (Funk-Hecke formula)**.** *In spherical coordinates, let the colatitude and longitude of any two points p and p' be given by (θ, ϕ) and (θ', ϕ'), respectively, and let x be the map from spherical coordinates to Euclidean coordinates, so that $||x(\theta,\phi)||_{\mathbb{R}^3} = ||x(\theta',\phi')||_{\mathbb{R}^3} = 1$. Consider any kernel of the form*

$$k\left((\theta,\phi),(\theta',\phi')\right) = \varphi(x(\theta,\phi)^T x(\theta',\phi')), \text{ with } \varphi \text{ continuous on } [-1,1].$$

Let Y_l^n be a spherical harmonic of degree l and order n. Then Y_l^n is an eigenfunction of the operator

$$\begin{aligned} \mathrm{L}_\varphi : L^2(\mathbb{S}^2) &\to L^2(\mathbb{S}^2) \\ f &\mapsto \mathrm{L}_\varphi f = \int_0^{2\pi}\int_0^{\pi} \varphi(x(\theta,\phi)^T x(\theta',\phi')) f(\theta',\phi') \sin(\theta')\,\mathrm{d}\theta'\,\mathrm{d}\phi', \end{aligned}$$

with an eigenvalue α_l that depends only on the degree l and the function φ

$$\begin{aligned} \alpha_l &= 2\pi \int_{-1}^{1} \varphi(x) P_l^0(x)\,\mathrm{d}x \\ &= \frac{2\pi}{2^l l!}\int_{-1}^{1} \varphi^{(k)}(x)(1-x)^k\,\mathrm{d}x \quad \textit{(if } \varphi \textit{ is } k \textit{ times continuously differentiable on [-1,1])}, \end{aligned} \tag{20}$$

where P_l^0 is the Legendre polynomial of degree l; the second equality holds by the Rodrigues formula.

Proof. Originally given in [50, 63]. Any continuous function on $[-1,1]$ has an almost everywhere pointwise convergent[21] Legendre expansion

$$\varphi \overset{\text{a.e.}}{=} \sum_{l=0}^{\infty} c_l P_l^0 = \sum_{l=0}^{\infty} \langle \varphi, P_l^0 \rangle_{L^2(-1,1)} P_l^0 = \sum_{l=0}^{\infty} \left(\frac{2l+1}{2} \int_{-1}^{1} \varphi(t) P_l^0(t)\,\mathrm{d}t \right) P_l^0. \tag{21}$$

Then for any degree l_0 and order n_0, we have that

$$\begin{aligned} \mathrm{L}_\varphi Y_{l_0}^{n_0} &= \int_0^{2\pi}\int_0^{\pi} \varphi(x(\theta,\phi)^T x(\theta',\phi')) Y_{l_0}^{n_0}(\theta',\phi') \sin(\theta')\,\mathrm{d}\theta'\,\mathrm{d}\phi' \\ &= \int_0^{2\pi}\int_0^{\pi} \sum_{l=0}^{\infty} c_l \underbrace{P_l^0(x(\theta,\phi)^T x(\theta',\phi'))}_{\sum_{n=-l}^{l} Y_l^n(\theta,\phi) Y_l^n(\theta',\phi') \frac{4\pi}{2l+1}} Y_l^n(\theta',\phi') \sin(\theta')\,\mathrm{d}\theta'\,\mathrm{d}\phi' \\ &= \sum_{l=0}^{\infty} \frac{4\pi c_l}{2l+1} \sum_{n=-l}^{l} Y_l^n(\theta,\phi) \underbrace{\int_0^{2\pi}\int_0^{\pi} Y_l^n(\theta',\phi') Y_{l_0}^{n_0}(\theta',\phi') \sin(\theta')\,\mathrm{d}\theta'\,\mathrm{d}\phi'}_{\delta_{n,n_0}\delta_{l,l_0}} = \frac{4\pi c_{l_0}}{2l_0+1} Y_{l_0}^{n_0}, \end{aligned}$$

[21] Note that a continuous function is square integrable on the compact interval $[-1,1]$. Then apply [117], which gives a Carleson-Hunt-like theorem for Legendre expansions: Legendre expansions of functions in $L^p(-1,1)$ converge pointwise almost everywhere if $p > 4/3$.

by the addition theorem (19). Expanding out the Fourier coefficient c_{l_0} of the Legendre expansion of φ (21), we see that

$$\alpha_{l_0} = \frac{4\pi c_{l_0}}{2l_0+1} = \frac{4\pi}{2l_0+1}\left(\frac{2l_0+1}{2}\int_{-1}^{1}\varphi(t)P_{l_0}^0(t)\,\mathrm{d}t\right) = 2\pi\int_{-1}^{1}\varphi(t)P_{l_0}^0(t)\,\mathrm{d}t\,,$$

and $\mathrm{L}_\varphi Y_{l_0}^{n_0} = \alpha_{l_0} Y_{l_0}^{n_0}$. Furthermore, the Hilbert-Schmidt expansion of the kernel is its Legendre expansion

$$\begin{aligned} k((\theta,\phi),(\theta',\phi')) = \varphi(x(\theta,\phi)^T x(\theta',\phi')) &= \sum_{l=0}^{\infty}\sum_{n=-l}^{l}\alpha_l Y_l^n(\theta,\phi)Y_l^n(\theta',\phi') \\ &= \sum_{l=0}^{\infty}\alpha_l\frac{2l+1}{4\pi}P_l^0(x(\theta,\phi)^T x(\theta',\phi')) \\ &= \sum_{l=0}^{\infty}\left(\frac{2l+1}{2}\int_{-1}^{1}\varphi(t)P_l^0(t)\,\mathrm{d}t\right)P_l^0(x(\theta,\phi)^T x(\theta',\phi')), \end{aligned} \tag{22}$$

which agrees with the Legendre expansion (21). See also [105], Section 4, Lemma 1. □

Remark 2.41. *Notice that $x(\theta,\phi)^T x(\theta',\phi')$ is precisely the cosine of the geodesic angle $\sphericalangle(p,p')$ between p and p'. Thus, we can write $k(p,p') = \varphi(\cos(\sphericalangle(p,p')))$.*

Definition 2.42 (Isotropic kernel)**.** *Let us call any kernel of this form*

$$k(p,p') = \varphi(\cos(\sphericalangle(p,p'))),$$

an isotropic kernel.

Example 2.43. *Let k be the isotropic kernel $k(p,p') = \cos(\sphericalangle(p,p')) = x(\theta,\phi)^T x(\theta',\phi') = x(p)^T x(p')$, so that $\varphi(x) = x$. Clearly k is continuous and positive-definite, since for any p and p' in $\mathbb{S}^2$, $k(p,p')$ gives an inner product in the Euclidean space $\mathbb{R}^3$; indeed, for any $n \in \mathbb{N}$ and any choice of $\alpha \in \mathbb{R}^n$ and of points on the sphere $\{p_1,\ldots,p_n\}$,*

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j k(p_i,p_j) = \sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j\left\langle x(p_i),x(p_j)\right\rangle_{\mathbb{R}^3} = \left\|\sum_{i=1}^{n}\alpha_i x(p_i)\right\|_{\mathbb{R}^3}^2 \geq 0.$$

We will verify that the Funk-Hecke formula (20) gives the coefficients of its Hilbert-Schmidt expansion. Observe that $\varphi(x) = P_1^0(x)$ on $[-1,1]$ is the Legendre polynomial of degree 1. By the orthogonality of the Legendre polynomials, the Funk-Hecke formula produces the eigenvalue sequence

$$\alpha_l = 2\pi\underbrace{\int_{-1}^{1}\varphi(x)P_l^0(x)\,\mathrm{d}x}_{\frac{2}{2l+1}\delta_{l,1}} = \begin{cases}\frac{4\pi}{2l+1}, & \text{if } l=1;\\ 0, & \text{otherwise.}\end{cases}$$

Indeed, we can confirm that this eigenvalue sequence synthesizes the Mercer kernel k. By the addition theorem for spherical harmonics (19), the series

$$\begin{aligned} k(p,p') = \sum_{l=0}^{\infty}\sum_{n=-l}^{l}\alpha_l Y_l^n(\theta,\phi)Y_l^n(\theta',\phi') &= \sum_{l=0}^{\infty}\alpha_l\frac{2l+1}{4\pi}P_l^0(\cos(\sphericalangle(p,p'))) \\ &= P_1^0(\cos(\sphericalangle(p,p'))) = \cos(\sphericalangle(p,p')). \end{aligned}$$

The synthesized kernel k is isotropic and continuous, and the convergence is (trivially) uniform.

We wish to find isotropic positive-definite kernels on the sphere. One approach is to do Mercer synthesis (see Section 2.2.1): start with a complete orthonormal system for $\mathbb{S}^2$ (namely, the spherical harmonics) and a sequence of eigenvalues (in ℓ^1 and nonnegative, so that they converge to a positive-definite kernel; with multiplicity equal to the degree to guarantee isotropy, by the Funk-Hecke formula (20)). Working on "the Fourier side" is a great way to guarantee the positive definiteness and continuity of the kernel, but deriving a kernel that can be expressed in closed form requires care.

We could instead choose a continuous function φ and verify that the eigenvalues $\{\lambda_{l,n}\}_{l=0,\, n=-l}^{\infty,\, l}$ of L_φ given by the Funk-Hecke formula (Proposition 2.40) $\lambda_{l,n} = \alpha_l$ are nonnegative and in ℓ^1 (i.e., $\{\alpha_l(2l+1)\}_{l=0}^\infty \in \ell^1$). Checking these criteria can be tedious, but if we start with a kernel that is easy to compute, we will not be stuck evaluating the kernel via infinite series.

Whichever approach we take, we can be confident that we can recover any isotropic positive-definite kernel on the sphere. In other words, the choices we made–using the spherical harmonics, with nonnegative eigenvalues in ℓ^1 with Funk-Hecke multiplicities (Proposition 2.40)–are not limiting. I.J. Schoenberg showed that all isotropic positive-definite functions on the sphere admit series expansions on the spherical harmonics with nonnegative, summable weights, whose multiplicities are specified by the Funk-Hecke formula (20).

Proposition 2.44 (Schoenberg, 1942 (Theorem 1))**.** *A continuous function $\varphi : [-1,1] \to \mathbb{R}$ gives rise to a positive-definite isotropic kernel $k(p,p') = \varphi(\cos(\sphericalangle(p,p')))$ on the sphere $\mathbb{S}^2$ if and only if its expansion in the Legendre polynomials P_l^0 has nonnegative weights that are in ℓ^1. In other words, in the Fourier expansion of φ on the complete orthogonal system $\{P_l^0\}_{l=0}^\infty$ for $L^2(-1,1)$*

$$\varphi(\cos(\sphericalangle(p,p'))) = \sum_{l=0}^\infty \frac{\langle \varphi, P_l^0 \rangle_{L^2(-1,1)}}{||P_l^0||^2_{L^2(-1,1)}} P_l^0(\cos(\sphericalangle(p,p'))) = \sum_{l=0}^\infty c_l P_l^0(\cos(\sphericalangle(p,p'))), \tag{23}$$

the weights all satisfy

$$c_l = \left(\frac{2l+1}{2} \int_{-1}^1 \varphi(u) P_l^0(u)\, \mathrm{d}u \right) \geq 0 \text{ and } \sum_{l=0}^\infty c_l < \infty.$$

Proof. See [129], Theorem 1. The key to the proof is the recognition that the Legendre polynomials P_l^0, interpreted as isotropic kernels, are all positive-definite functions on the sphere (the case with $l = 1$ is explored in Example 2.43); this is an easy consequence of the addition theorem, and an inductive proof is given in [129]; see also [140], Chapter 4. The remaining details follow.

First, suppose that the weights $\{c_l\}_{l=0}^\infty \in \ell^1$ and $c_l \geq 0$ for all $l \in \mathbb{N}$. Since the Legendre polynomials P_l^0 are all continuous and bounded in absolute value by 1 on $[-1,1]$, the Legendre expansion (23) converges uniformly (by the Weierstrass M-test, since the sequence $\{c_l\}_{l=0}^\infty$ is in ℓ^1) to a continuous limit (by the uniform limit theorem, since the P_l^0 are continuous). This continuous limit φ of positive-definite functions must also be positive definite. Indeed, for any $n \in \mathbb{N}$, choice of points $\{p_i\}_{i=1}^n$ on the sphere, and $\alpha \in \mathbb{R}^n$, we have

$$\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j P_l^0(\cos(\sphericalangle(p_i,p_j))) \geq 0.$$

Hence,

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j\varphi(\cos(\sphericalangle(p_i,p_j))) = \sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j\sum_{l=0}^{\infty}c_l P_l^0(\cos(\sphericalangle(p_i,p_j))) = \sum_{l=0}^{\infty}\underbrace{c_l}_{\geq 0}\underbrace{\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j P_l^0(\cos(\sphericalangle(p_i,p_j)))\right)}_{\geq 0} \geq 0,$$

and the continuous φ is positive definite.

Conversely, suppose φ is a positive-definite isotropic kernel on the sphere and is continuous on $[-1,1]$. Then for any choice of $n \in \mathbb{N}$, points $\{p_i\}_{i=1}^n$ on the sphere $\mathbb{S}^2$, and weights $\alpha \in \mathbb{R}^n$, φ satisfies

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j\varphi(\cos(\sphericalangle(p_i,p_j))) \geq 0.$$

This is equivalent to the integral inequality for any continuous function $h : \mathbb{S}^2 \mapsto \mathbb{R}$

$$\int_{\mathbb{S}^2} h(p)\int_{\mathbb{S}^2} h(p')\varphi(\cos(\sphericalangle(p,p')))\,\mathrm{d}S(p')\,\mathrm{d}S(p) \geq 0,$$

where, at $p = (\theta,\phi)$, the unit sphere's surface area differential $\mathrm{d}S(p) = \sin(\theta)\,\mathrm{d}\theta\,\mathrm{d}\phi$. Setting $h \equiv 1$, we get the requirement that

$$4\pi\int_{\mathbb{S}^2}\varphi(\cos(\sphericalangle(p,p')))\,\mathrm{d}S(p') \geq 0.$$

Then

$$\int_0^{2\pi}\int_0^{\pi}\varphi(\cos(\sphericalangle((\theta,\phi),(\theta',\phi'))))\sin(\theta')\,\mathrm{d}\theta'\,\mathrm{d}\phi' \geq 0, \tag{24}$$

and the lth coefficient of the expansion of φ in Legendre polynomials can be written, via the substitution $u = \cos(\theta')$,

$$\begin{aligned} c_l &= \frac{2l+1}{2}\int_{-1}^{1}\varphi(u)P_l^0(u)\,\mathrm{d}u = \frac{2l+1}{2}\int_{1}^{-1}\varphi(u)P_l^0(u)\cdot(-\,\mathrm{d}u) \\ &= \frac{2l+1}{2}\int_0^{\pi}\varphi(\cos(\theta'))P_l^0(\cos(\theta'))\sin(\theta')\,\mathrm{d}\theta' \\ &= \frac{2l+1}{4\pi}\int_0^{2\pi}\int_0^{\pi}P_l^0(\cos(\sphericalangle(p_1,p')))\varphi(\cos(\sphericalangle(p_1,p')))\sin(\theta')\,\mathrm{d}\theta'\,\mathrm{d}\phi', \end{aligned}$$

where $p_1 = (\theta_1,\phi_1)$ is the North Pole, so that for any $p' = (\theta',\phi')$ on $\mathbb{S}^2$,

$$\cos(\sphericalangle(p_1,p')) = \underbrace{\cos(\theta_1)}_{1}\cos(\theta') + \underbrace{\sin(\theta_1)}_{0}\sin(\theta')\cos(\phi_1-\phi') = \cos(\theta').$$

But then $c_l \geq 0$ by (24), since the product $P_l^0\varphi$ of two positive-definite functions is positive definite.

We now show the absolute summability of the $\{c_l\}_{l=0}^{\infty}$ and the uniform convergence of the weighted sum of Legendre polynomials, with $c_l \geq 0$, to the kernel φ. We recall the result (see [105]; [114]; [129]; [172], remarks to Theorem 3.1) that the decomposition

$$\varphi \sim \sum_{l=0}^{\infty}c_l P_l^0 \tag{25}$$

is everywhere Abel-summable if φ is everywhere continuous. Then, in particular, φ is Abel-summable at 1

$$\lim_{r\to 1^-}\sum_{l=0}^{\infty} c_l r^l \underbrace{P_l^0(1)}_{1} = \lim_{r\to 1^-}\sum_{l=0}^{\infty} c_l r^l = A < \infty.$$

But since, for all $(p,p') \in \mathbb{S}^2\times\mathbb{S}^2$ and $m\in\mathbb{N}$,

$$\sum_{l=0}^{m} |c_l P_l^0(\cos(\sphericalangle(p,p')))| \le \sum_{l=0}^{m} c_l \underbrace{P_l^0(1)}_{1} = \sum_{l=0}^{m} c_l,$$

and since $c_l \ge 0$, we have that

$$\sum_{l=0}^{m} c_l r^l \le \sum_{l=0}^{\infty} c_l r^l, \text{ and } \lim_{r\to 1^-}\sum_{l=0}^{m} c_l r^l = \sum_{l=0}^{m} c_l \le \lim_{r\to 1^-}\sum_{l=0}^{\infty} c_l r^l = A.$$

As m was arbitrary, we can conclude

$$\sum_{l=0}^{\infty} |c_l P_l^0(\cos(\sphericalangle(p,p')))| \le \sum_{l=0}^{\infty} c_l \le A.$$

Conversely, since for any $r\in(0,1)$

$$\sum_{l=0}^{\infty} c_l r^l \le \sum_{l=0}^{\infty} c_l, \text{ we see that } A = \lim_{r\to 1^-}\sum_{l=0}^{\infty} c_l r^l \le \sum_{l=0}^{\infty} c_l.$$

Thus,

$$\varphi(1) = \sum_{l=0}^{\infty} c_l = A < \infty,$$

and (25) converges absolutely and uniformly for all $\sphericalangle(p,p') \in [0,\pi]$. Hence, the Hilbert-Schmidt kernel sum of the isotropic kernel $k : \mathbb{S}^2\times\mathbb{S}^2 \to \mathbb{R}$,

$$k(p,p') = \varphi(\cos(\sphericalangle(p,p'))) = \sum_{l=0}^{\infty} c_l P_l^0(\cos(\sphericalangle(p,p'))) = \sum_{l=0}^{\infty} \underbrace{\frac{4\pi c_l}{2l+1}}_{\alpha_l} \sum_{n=-l}^{l} Y_l^n(\theta,\phi) Y_l^n(\theta',\phi'),$$

converges uniformly.

Note that we have adapted Schoenberg's more general Gegenbauer ("ultraspherical") polynomial expansion of an isotropic kernel in $\mathbb{S}^m$ (2.11) to the specific case of the Legendre polynomials, which is appropriate for $\mathbb{S}^2$ (set $m=2, \lambda=\frac{1}{2}$, and in the integrals $u=\cos(\theta')$). □

Remark 2.45. *Schoenberg's theorem tells us that an isotropic kernel applies the same weight to each spherical harmonic of the same wiggliness. Isotropic kernels are good candidates to serve as wiggliness penalties. Suppose an interpolant is a weighted sum of the representations of evaluation, using an isotropic kernel k whose RKHS is $\mathcal{H}$, of a finite data set: $I(\theta,\phi) = \sum_{i=1}^{n} \alpha_i k(\cdot,(\theta_i,\phi_i))$. The Hilbert norm of this interpolant is $\alpha^T\mathbf{K}\alpha$, where $\alpha = (\alpha_1,\dots,\alpha_n)^T$ and $(\mathbf{K})_{i,j} = k((\theta,\phi_i),(\theta_j,\phi_j))$. Since k is isotropic, the wiggliness of the interpolant depends only on the weights α and the pairwise geodesic distances between points in the data set.*

Remark 2.46. *For the Legendre expansion (25) to converge everywhere pointwise, the continuity of φ is ordinarily insufficient; continuous differentiability is typically the sufficient criterion that is most convenient to show [125]. Schoenberg's theorem (Proposition 2.44) shows that the Legendre series converges uniformly if φ is continuous and positive-definite (i.e., if we enforce the rather strong constraint that all Fourier coefficients c_l in its expansion on the Legendre polynomials satisfy $c_l \geq 0$).*

Example 2.47. *The positive sequence $\{\lambda_{l,n}\}_{n=0,\,l=-n}^{\infty,\,l}$ given by $\lambda_{l,n} = \alpha_l = \frac{4\pi}{2l+1}\beta^l$, with $\beta \in (0,1)$, is in ℓ^1 by the convergence of the geometric series. We take the Hilbert-Schmidt expansion of this sequence of eigenvalues on the spherical harmonics and apply the addition theorem to synthesize the corresponding kernel*

$$k(p,p') = \sum_{l=0}^{\infty}\sum_{n=-l}^{l} \alpha_l Y_l^n(\theta,\phi)Y_l^n(\theta',\phi') = \sum_{l=0}^{\infty} \alpha_l \frac{2l+1}{4\pi} P_l^0(\cos(\sphericalangle(p,p'))) = \sum_{l=0}^{\infty} P_l^0(\cos(\sphericalangle(p,p')))\beta^l.$$

The Legendre polynomials P_l^0 can be defined, inter alia, by the generating function identified by Legendre while investigating $1/r$ potentials [22]

$$\frac{1}{\sqrt{1-2\beta z+\beta^2}} = \sum_{l=0}^{\infty} P_l^0(z)\beta^l.$$

For any $z \in [-1,1]$ and $\beta \in (0,1)$, the function on the left-hand side is easily seen to be continuous in both arguments at (z,β). Since for $z \in [-1,1]$, the $|P_l^0(z)| \leq 1$, we can apply the Weierstrass M-test to confirm that the right-hand side, as a function of z, converges uniformly to the left-hand side on $[-1,1]$. Thus, the Mercer kernel associated with this positive eigenvalue sequence is

$$k(p,p') = \frac{1}{\sqrt{1-2\beta\cos(\sphericalangle(p,p'))+\beta^2}}.$$

Let $\mathcal{H}$ be the RKHS associated with k. The reproducing property can be verified using (17)

$$\begin{aligned}\langle k(\cdot,s), f\rangle_{\mathcal{H}} &= \sum_{l=0}^{\infty}\sum_{n=-l}^{l} \frac{(k(\cdot,s))_{n,l}(f)_{n,l}}{\alpha_l} \\ &= \sum_{l=0}^{\infty}\sum_{n=-l}^{l} \frac{\alpha_l Y_l^n(\theta_s,\phi_s)(f)_{n,l}}{\alpha_l} = \sum_{l=0}^{\infty}\sum_{n=-l}^{l} (f)_{n,l}Y_l^n(\theta_s,\phi_s) = f(s);\end{aligned}$$

the Fourier expansion converges uniformly and thus pointwise since k is continuous and $f \in \mathcal{H}$.

Let us summarize our findings. Using Proposition 2.38, we synthesized a kernel as a Hilbert-Schmidt weighted sum of spherical harmonics. We used the eigenvalue conditions of Mercer kernels–nonnegativity and bounded ℓ^1 norm–and added a new one by imposing equality on the eigenvalues of the $2l+1$ spherical harmonics of the same degree. We found the synthesized positive-definite kernels were isotropic. By the addition theorem, the kernel sum can be expressed in terms of the (bounded) Legendre polynomials $\{P_l^0\}_{l=0}^{\infty}$. This process can synthesize all isotropic positive-definite kernels on the sphere, by Schoenberg's theorem (Proposition 2.44).

Moreover, the Fourier expansion of any such $\varphi \in L^2(-1,1)$ on the Legendre polynomials

$$\varphi \sim \sum_{l=0}^{\infty} c_l P_l^0, \text{ where } c_l = \left\langle \varphi, \frac{2l+1}{2}P_l^0 \right\rangle_{L^2(-1,1)} = \frac{2l+1}{2}\int_{-1}^{1} \varphi(t)P_l^0(t)\,\mathrm{d}t$$

converges not only in $L^2(-1,1)$ but uniformly and absolutely.

The condition that the sequence $\{\lambda_{l,n}\}_{l=0,\,n=-l}^{\infty,\,l}$ of nonnegative weights on the spherical harmonics be absolutely summable can be rewritten using the multiplicity ($\lambda_{l,n} = \alpha_l$) guaranteed by the Funk-Hecke formula: (Proposition 2.40) $\{(2l+1)\alpha_l\}_{l=0}^{\infty} \in \ell^1$. By the addition theorem, this obliges the weights c_l on the Legendre polynomial expansion of φ to be absolutely summable. The α_l are nonnegative precisely when the c_l are

$$\alpha_l = 2\pi \underbrace{\int_{-1}^{1} \varphi(t) P_l^0(t)\,\mathrm{d}t}_{\frac{2}{2l+1}c_l} \geq 0 \iff c_l \geq 0,$$

by the Funk-Hecke formula (20). The conditions (nonnegativity and summability) Schoenberg's theorem (Proposition 2.44) places on the weights of the expansion of $\varphi : [-1,1] \to \mathbb{R}$ on the Legendre polynomials are the same as the conditions Mercer synthesis (Proposition 2.38) places on the weights of the Hilbert-Schmidt expansion of a kernel $k : \mathbb{S}^2 \times \mathbb{S}^2 \to \mathbb{R}$ on a complete orthonormal system for $L^2(\mathbb{S}^2)$; the Legendre expansion follows from the Hilbert-Schmidt expansion if we impose the constraint of isotropy (or, equivalently, Funk-Hecke eigenvalue multiplicities) by the addition theorem for spherical harmonics (19). We could not use Proposition 2.38 because the spherical harmonics are not locally bounded. However, isotropy and the addition theorem make it clear that the Hilbert-Schmidt expansion of the isotropic kernel on the spherical harmonics converges uniformly whenever the expansion of φ on the Legendre polynomials does

$$\begin{aligned} k(p,p') = \varphi(\cos(\sphericalangle(p,p'))) &= \sum_{l=0}^{\infty}\sum_{n=-l}^{l} \underbrace{\lambda_{l,n}}_{\alpha_l} Y_l^n(\theta,\phi) Y_l^n(\theta',\phi') \\ &= \sum_{l=0}^{\infty} \underbrace{\alpha_l}_{\frac{4\pi}{2l+1}c_l} \sum_{n=-l}^{l} Y_l^n(\theta,\phi) Y_l^n(\theta',\phi') = \sum_{l=0}^{\infty} \alpha_l \frac{2l+1}{4\pi} P_l^0(\cos(\sphericalangle(p,p'))) = \sum_{l=0}^{\infty} c_l P_l^0(\cos(\sphericalangle(p,p'))). \end{aligned}$$

The RKHS $\mathcal{H}$ associated with our Mercer kernel (Definition 2.29) can be defined exactly as suggested in Section 2.2.1, with inner product

$$\langle f, g\rangle_{\mathcal{H}} = \sum_{\substack{l=0 \\ \lambda_{l,n}>0}}^{\infty}\sum_{n=-l}^{l} \frac{(f)_{l,n}(g)_{l,n}}{\lambda_{l,n}} = \sum_{\substack{l=0 \\ \alpha_l>0}}^{\infty} \frac{1}{\alpha_l} \sum_{n=-l}^{l} (f)_l^n (g)_l^n,$$

where $(f)_{l,n}$ and $(g)_{l,n}$ are the Fourier coefficients of the expansion of f and g on the spherical harmonics, respectively. The inclusion criterion of our Fourier-weighted Hilbert space remains the same

$$f \in \mathcal{H} \iff ||f||_{\mathcal{H}}^2 = \sum_{\substack{l=0 \\ \alpha_l>0}}^{\infty} \frac{1}{\alpha_l} \sum_{n=-l}^{l} \left((f)_l^n\right)^2 < \infty.$$

While the weights $\{\lambda_{l,n}\}_{l=0,\,n=-l}^{\infty,\,l}$ of the Hilbert-Schmidt expansion of an isotropic positive-definite kernel on the spherical harmonics assume a constant value α_l over each space of spherical polynomials of degree l, the Fourier coefficients of functions in the RKHS $\mathcal{H}$ need not exhibit this multiplicity; the addition theorem cannot be applied, so a double sum remains in the inner product. However, like the Hilbert-Schmidt expansion of the kernel, the expansion of any $f \in \mathcal{H}$

$$f(p) = \langle f, k_p\rangle_{\mathcal{H}} = \sum_{\substack{l=0 \\ \alpha_l>0}}^{\infty}\sum_{n=-l}^{l} \frac{(f)_{n,l}\alpha_l Y_l^n(\theta,\phi)}{\alpha_l} = \sum_{\substack{l=0 \\ \alpha_l>0}}^{\infty}\sum_{n=-l}^{l} (f)_{n,l} Y_l^n(\theta,\phi)$$

converges uniformly[22] by Proposition 2.24, part 3. By the definition of $\langle \cdot, \cdot \rangle_{\mathcal{H}}$, the Fourier expansion of any $f \in \mathcal{H}$ on the spherical harmonics converges weakly to f in $\mathcal{H}$

$$\forall g \in \mathcal{H}, \left\langle \sum_{\substack{l=0 \\ \alpha_l > 0}}^{\infty} \sum_{n=-l}^{l} (f)_{n,l} Y_l^n, g \right\rangle_{\mathcal{H}} = \sum_{\substack{l=0 \\ \alpha_l > 0}}^{\infty} \sum_{n=-l}^{l} \frac{(f)_{n,l}(g)_{n,l}}{\alpha_l} = \langle f, g \rangle_{\mathcal{H}}.$$

Hence, that convergence is uniform.

2.3 Solving Norm-Minimizing Interpolation and Smoothing Problems in an RKHS

The development of RKHS theory was motivated by interpolation applications in a rather particular setting (in particular, finding holomorphic interpolants of scattered data in the unit disk [15]). Nachman Aronszajn generalized the notion of reproducing kernels to arbitrary index sets [3]. Grace Wahba revisited interpolation with this more general perspective. The algorithms Wahba derived for solving interpolating problems in an RKHS [79, 158] require neither uniform convergence of the Hilbert-Schmidt expansion nor kernel continuity; in fact, the index set $\mathcal{X}$ need not be topological.

As we turn our attention to solving interpolation and smoothing problems in an RKHS, we return to the full generality of the Aronszajn theorem (Proposition 2.11) and Wahba's work. In this section, $\mathcal{X}$ can be an arbitrary set.

Definition 2.48 (The exact interpolation problem and the smoothing problem)**.** *Suppose we have a sequence of n sample locations $\{x_i\}_{i=1}^n$ in $\mathcal{X}$ and corresponding values $\{y_i\}_{i=1}^n$ in $\mathbb{R}$. An* interpolating function *f in a reproducing kernel Hilbert space $\mathcal{H} \subseteq \mathbb{R}^{\mathcal{X}}$ (with reproducing kernel k) is any function $f \in \mathcal{H}$ for which $f(x_i) = y_i$ for each $i = 1, \ldots, n$. The* exact interpolation problem *is solved by finding the interpolating function of minimal norm*

$$\arg\min_{f \in \mathcal{H}} ||f||_{\mathcal{H}} \text{ subject to } f(x_i) = y_i \text{ for } i = 1, \ldots, n. \tag{26}$$

The closely related smoothing problem *(often called in the machine learning literature* kernel ridge regression*) seeks the function in $\mathcal{H}$ that minimizes the empirical risk*

$$\arg\min_{f \in \mathcal{H}} R_\lambda(f) = \frac{1}{n} \sum_{i=1}^{n} (f(x_i) - y_i)^2 + \lambda ||f||_{\mathcal{H}}^2. \tag{27}$$

This latter problem is readily generalized to incorporate other losses besides the square loss and other bounded linear functionals besides evaluation at the x_i. We will later consider the case where the desired penalty is not a definite norm, but a seminorm.

In both cases, the solution lies in a finite-dimensional subspace of $\mathcal{H}$, namely, in the span of the representers of evaluation $\{k_{x_1}, \ldots, k_{x_n}\}$.

2.3.1 Solving the Exact Interpolation Problem

Given a vector of sample locations $x = (x_1, \ldots, x_n)^T$, it is far from guaranteed that $\mathcal{H}$ is sufficiently rich to include a function that interpolates any possible set of corresponding sample values $y =$

[22] It is often misstated in the literature that the Fourier series expansion of continuous functions on the spherical harmonics is uniformly convergent (e.g., in [91]). In fact, this condition is not sufficient, but uniform convergence always holds for Fourier expansions of continuously differentiable functions on $\mathbb{S}^2$ (as on $\mathbb{S}^1$) [75]. In our case, we have a new criterion: $f \in \mathcal{H}$.

$(y_1, \ldots, y_n)^T$. The exact interpolation problem has no solution when there exists no $f \in \mathcal{H}$ that interpolates the data–that is, if y is not in the range of the vectorized evaluation map

$$\begin{aligned} \mathrm{E} : \mathcal{H} &\to \mathbb{R}^n \\ f &\mapsto (f(x_1), \ldots, f(x_n))^T. \end{aligned}$$

Luckily, we can easily make the diagnosis that there exist vectors $y \in \mathbb{R}^n$ for which the RKHS $\mathcal{H}$ contains no corresponding interpolator by looking at the Gram matrix. Intuitively, the linear relationship that exists between the representers of evaluation $k_{x_1}, \ldots, k_{x_n}$ is reproduced in the Gram matrix and constrains the values every function in the space attains at the evaluation points.

Proposition 2.49. *Consider a reproducing kernel Hilbert space $\mathcal{H} \subseteq \mathbb{R}^{\mathcal{X}}$ with reproducing kernel k. Let $\mathbf{K}$ be the Gram matrix associated with data points $x = (x_1, \ldots, x_n)^T$: that is, $(\mathbf{K})_{i,j} = k(x_i, x_j)$. Then $\mathbf{K}$ is singular if there exists some $y \in \mathbb{R}^n$ for which no solution to (26) exists.*

Proof. Suppose there exists some vector of sample values $y \in \mathbb{R}^n$ that cannot be interpolated at the points x by any $f \in \mathcal{H}$. Thus, the vectorized evaluation map E is not onto: $\operatorname{range} \mathrm{E} \subsetneq \mathbb{R}^n$. Choose any nonzero element α of the orthogonal complement of the range of E, that is, $\alpha \in \operatorname{null} \mathrm{E}^*$, where * indicates the adjoint. With this choice, $\mathrm{E}^*\alpha = 0$, and, for all $f \in \mathcal{H}$, we have that $\langle \mathrm{E}^*\alpha, f \rangle_{\mathcal{H}} = \langle \alpha, \mathrm{E}f \rangle_{\mathbb{R}^n} = 0$. We can then write, using the reproducing property,

$$0 = \langle \alpha, \mathrm{E}f \rangle_{\mathbb{R}^n} = \sum_{i=1}^n \alpha_i f(x_i) = \left\langle f, \sum_{i=1}^n \alpha_i k_{x_i} \right\rangle_{\mathcal{H}} \quad \forall f \in \mathcal{H}.$$

Since only the function that is identically zero[23] may satisfy this, we conclude

$$\sum_{i=1}^n \alpha_i k_{x_i} \equiv 0 \text{ and yet } \alpha \neq 0.$$

Moreover, this $\alpha \in \operatorname{null} \mathbf{K}$ by the fundamental theorem of linear algebra [144] and symmetry of $\mathbf{K}$. We see this because α is orthogonal to the ith row of the Gram matrix for $i = 1, \ldots, n$

$$0 = \langle k_{x_i}, 0 \rangle_{\mathcal{H}} = \left\langle k_{x_i}, \sum_{j=1}^n \alpha_j k_{x_j} \right\rangle_{\mathcal{H}} = \sum_{j=1}^n \alpha_j k(x_i, x_j) = \mathbf{K}[i, :]\alpha.$$

This nonzero vector in $\alpha \in \operatorname{null} \mathbf{K}$ establishes the result. □

The converse of this result follows immediately once we decompose $\mathcal{H}$ as the direct sum of $\mathcal{S} = \operatorname{span}\{k_{x_1}, \ldots, k_{x_n}\}$ and $\mathcal{S}^\perp$, its orthogonal complement[24].

Proposition 2.50. *Consider a reproducing kernel Hilbert space $\mathcal{H} \subseteq \mathbb{R}^{\mathcal{X}}$ with reproducing kernel k. Write $\mathcal{H} = \mathcal{S} \oplus \mathcal{S}^\perp$, where $\mathcal{S} = \operatorname{span}\{k_{x_1}, \ldots, k_{x_n}\}$. Then $\operatorname{null} \mathrm{E} = \mathcal{S}^\perp$, and $\operatorname{range} \mathrm{E} = \operatorname{range} \mathbf{K}$ has the same dimension as $\mathcal{S}$.*

Proof. We can see that $f \in \mathcal{S}^\perp \iff f \in \operatorname{null} \mathrm{E}$ by recognizing that f is orthogonal to each representer of evaluation k_{x_i} in $\mathcal{S}$ whenever the map E annihilates f: that is, by the reproducing property, $f(x_i) = \langle f, k_{x_i} \rangle_{\mathcal{H}} = 0$ for all $i = 1, \ldots, n$.

[23] It is orthogonal to itself, after all, and thus has norm zero!

[24] Note that $\mathcal{S}$ is closed (it is finite-dimensional). $\mathcal{S}^\perp$ is closed (since it is an orthogonal complement, by the continuity of the inner product) and $(\mathcal{S}^\perp)^\perp = \mathcal{S}$.

Now consider $f \in \mathcal{S}$. We can write $f = \sum_{i=1}^n \alpha_i k_{x_i}$[25]. By the reproducing property, the vector

$$\begin{aligned}(f(x_1), \ldots, f(x_n))^T &= \left(\left\langle k_{x_1}, \sum_{i=1}^n \alpha_i k_{x_i} \right\rangle_{\mathcal{H}}, \ldots, \left\langle k_{x_n}, \sum_{i=1}^n \alpha_i k_{x_i} \right\rangle_{\mathcal{H}} \right)^T \\ &= \left(\sum_{i=1}^n \alpha_i k(x_1, x_i), \ldots, \sum_{i=1}^n \alpha_i k(x_n, x_i) \right)^T \\ &= \mathbf{K}\alpha.\end{aligned}$$

The restriction of E to $\mathcal{S}$ is therefore the range of $\mathbf{K}$. The result follows by noting that functions in $\mathcal{S}^\perp$ evaluate to 0. □

Since $f \in \mathcal{S}^\perp$ if and only if $f(x_i) = 0$ for each $i = 1, \ldots, n$, the orthogonal projection operator $\mathrm{P}_{\mathcal{S}}$ from $\mathcal{H}$ onto $\mathcal{S}$ does not change the evaluation of a function on the data points x. Writing the unique decomposition of $f = f_{\mathcal{S}} + f_{\mathcal{S}^\perp}$ onto the orthogonal subspaces $\mathcal{S}$ and $\mathcal{S}^\perp$, respectively, we have

$$\forall f \in \mathcal{H},\ f(x_i) = \langle f, k_{x_i} \rangle_{\mathcal{H}} = \langle f_{\mathcal{S}}, k_{x_i} \rangle_{\mathcal{H}} + \underbrace{\langle f_{\mathcal{S}^\perp}, k_{x_i} \rangle_{\mathcal{H}}}_{0} = \langle \mathrm{P}_{\mathcal{S}} f, k_{x_i} \rangle_{\mathcal{H}} = (\mathrm{P}_{\mathcal{S}} f)(x_i) \text{ for } i = 1, \ldots, n.$$

Though this projection operator does not change a function's evaluations at the points $\{x_1, \ldots, x_n\}$, it can change its norm.

Proposition 2.51. *Consider a reproducing kernel Hilbert space $\mathcal{H} \subseteq \mathbb{R}^{\mathcal{X}}$ with reproducing kernel k. Write $\mathcal{H} = \mathcal{S} \oplus \mathcal{S}^\perp$, where $\mathcal{S} = \operatorname{span}\{k_{x_1}, \ldots, k_{x_n}\}$. Given a vector of sample locations x and sample values y in $\mathbb{R}^n$, if there exists a function $f \in \mathcal{H}$ such that f interpolates these data, then the solution to the minimum-norm exact interpolation problem (26) is $P_{\mathcal{S}} f$.*

Proof. Suppose the minimum-norm interpolant f, when projected on the orthogonal subspaces $\mathcal{S}$ and $\mathcal{S}^\perp$ as $f_{\mathcal{S}} + f_{\mathcal{S}^\perp}$, has nonzero component $f_{\mathcal{S}^\perp}$. Since for all ϵ, the function $f_{\mathcal{S}} + \epsilon f_{\mathcal{S}^\perp}$ also interpolates the data, we see that if $\epsilon \in [0, 1)$, $f_{\mathcal{S}} + \epsilon f_{\mathcal{S}^\perp}$ is an interpolant with smaller norm

$$||f_{\mathcal{S}} + \epsilon f_{\mathcal{S}^\perp}||_{\mathcal{H}}^2 = ||f_{\mathcal{S}}||_{\mathcal{H}}^2 + \epsilon^2 ||f_{\mathcal{S}^\perp}||_{\mathcal{H}}^2 < ||f_{\mathcal{S}}||_{\mathcal{H}}^2 + ||f_{\mathcal{S}^\perp}||_{\mathcal{H}}^2 = ||f||_{\mathcal{H}}^2.$$

We conclude that the minimum-norm interpolant f has no component in $\mathcal{S}^\perp$.

Observe that the difference between any two functions $f_1 \in \mathcal{H}$ and $f_2 \in \mathcal{H}$ that interpolate the data lies entirely in $\mathcal{S}^\perp$, since for $i = 1, \ldots, n$, we have that $f_1 - f_2 \perp k_{x_i}$

$$(f_1 - f_2)(x_i) = \langle f_1 - f_2, k_{x_i} \rangle_{\mathcal{H}} = f_1(x_i) - f_2(x_i) = 0.$$

It follows that $\mathrm{P}_{\mathcal{S}} f_1 = \mathrm{P}_{\mathcal{S}} f_2$, and that this function is the unique minimum-norm interpolant. □

Remark 2.52. *This result is a special case of Wahba's representer theorem, as stated in Section 2.5 (Proposition 2.60). The interpolant basis functions $\{k_{x_i}\}_{i=1}^n$ vary with the sample locations $\{x_i\}_{i=1}^n$. It is possible on the real line to find interpolation basis functions that depend on the number of sample points but not their locations (there is a unique interpolant of n data points in $\operatorname{span}\{1, x, \ldots, x^{n-1}\}$). Such a construction is not possible on domains in $\mathbb{R}^d$ for $d \geq 2$ due to the Haar-Mairhuber-Curtis theorem [94]. Basis functions therefore need vary with the sample locations themselves and much of the interpolation literature makes the association explicit. This orthogonality argument yields a solution with a "knot placement" pattern that resembles those of most methods in the literature.*

[25] If $\mathbf{K}$ is singular, the choice of α is not unique because the representers of evaluation are linearly dependent; nevertheless all such representations evaluate to the same function.

Corollary 2.53. *Consider a reproducing kernel Hilbert space $\mathcal{H} \subseteq \mathbb{R}^{\mathcal{X}}$ with reproducing kernel k. If the Gram matrix $\mathbf{K}$ is nonsingular given sample locations x, any map $x \to y \in \mathbb{R}^n$ can be interpolated by*

$$f = \sum_{i=1}^{n} \alpha_i k_{x_i}, \text{ where } \alpha = \mathbf{K}^{-1} y.$$

The norm of our minimum-norm solution to the exact interpolation problem (26) is therefore given by

$$||f||_{\mathcal{H}}^2 = \left\langle \sum_{i=1}^{n} \alpha_i k_{x_i}, \sum_{j=1}^{n} \alpha_j k_{x_j} \right\rangle_{\mathcal{H}} = \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j k(x_i, x_j) = \alpha^T \underbrace{\mathbf{K}\alpha}_{y} = \alpha^T y = y^T \mathbf{K}^{-1} y.$$

Corollary 2.54. *Under the assumptions of Corollary 2.53, if the Gram matrix $\mathbf{K}$ is singular given sample locations x and if $y \in \operatorname{range} \mathbf{K}$ there exist infinitely many vectors $\alpha \in \mathbb{R}^n$ for which $\mathbf{K}\alpha = y$. All of these coordinate vectors characterize the same function*

$$f = \sum_{i=1}^{n} \alpha_i k_{x_i}.$$

Proof. If $\mathbf{K}\alpha = \mathbf{K}\alpha' = y$, then $\alpha - \alpha' \in \operatorname{null} \mathbf{K}$. We know that the coordinate vector $(\alpha - \alpha')$ identifies the function that is identically zero, since the function

$$g = \sum_{i=1}^{n} (\alpha - \alpha')_i k_{x_i}$$

has norm

$$||g||_{\mathcal{H}}^2 = \langle g, g \rangle_{\mathcal{H}} = (\alpha - \alpha') \underbrace{\mathbf{K}(\alpha - \alpha')}_{0} = 0.$$

Hence,

$$f = \sum_{i=1}^{n} \alpha_i k_{x_i} = \sum_{i=1}^{n} \alpha_i' k_{x_i}$$

is the unique minimum-norm solution to (26) even if it can be written in multiple ways as a linear combination of the (linearly dependent) k_{x_i}. □

Remark 2.55. *If $\mathbf{K}$ is nonsingular, the columns of $\mathbf{K}^{-1}$ give a* partition of unity, *that is, a set of functions $\{f_1, \ldots, f_n\}$ such that $f_i(x_j) = \delta_{i,j}$ for $i, j = 1, \ldots, n$, where δ is the Kronecker delta. In constructing the functions f_i, we choose as the coordinate of f_i on the k_{x_j} the ith column of $\mathbf{K}^{-1}$*

$$f_i = \sum_{j=1}^{n} \mathbf{K}^{-1}[j, i] k_{x_j}.$$

Let e_i be the ith standard basis function of $\mathbb{R}^n$. Each f_i is then the unique function in $\mathcal{S}$ that interpolates e_i, and therefore the minimum-norm function in $\mathcal{H}$ that interpolates e_i. Let us observe how our minimum-norm interpolant of y can be expressed in terms of the functions in the partition of unity: if $\alpha = \mathbf{K}^{-1} y$, then

$$f = \sum_{j=1}^{n} (\mathbf{K}^{-1} y)_j k_{x_j} = \sum_{j=1}^{n} \left(\mathbf{K}^{-1} \left(\sum_{i=1}^{n} y_i e_i \right) \right)_j k_{x_j} = \sum_{i=1}^{n} y_i \left(\sum_{j=1}^{n} \mathbf{K}^{-1}[j, i] k_{x_j} \right) = \sum_{i=1}^{n} y_i f_i.$$

If $\mathbf{K}$ is singular, there exist vectors of values $y \in \mathbb{R}^n$ that cannot be interpolated by any $f \in \mathcal{H}$ and there can be no partition of unity.

Remark 2.56. *When* $\mathbf{K}$ *is singular, we can solve the closely related* best-approximation problem, *wherein we seek*

$$f^* = \arg\min_{f \in \mathcal{H}} \sum_{i=1}^{n} (f(x_i) - y_i)^2 . \tag{28}$$

By the reasoning we used to prove Proposition 2.51, projecting any solution of (28) *into* $\mathcal{S}$ *leaves its evaluations at* $x = (x_1, \ldots, x_n)^T$*–and hence the sum-of-squared-errors loss in* (28)*–unchanged. We can therefore take* f^* *as lying in* $\mathcal{S}$*. Thus, we can write*

$$f^* = \sum_{i=1}^{n} \alpha_i k_{x_i}.$$

Our problem (28) *reduces to a least-squares approximation problem in* $\mathbb{R}^n$

$$\arg\min_{\alpha \in \mathbb{R}^n} \frac{1}{n} \|\mathbf{K}\alpha - y\|_{\mathbb{R}^n}^2 .$$

Taking the gradient of the (convex) objective and setting it to zero, we get the normal equations

$$\mathbf{K}^T\mathbf{K}\alpha = \mathbf{K}^T y,$$

and observe the problem is solved using the Moore-Penrose pseudo-inverse to find the $\alpha \in (\operatorname{null}\mathbf{K})^\perp$ *that* $\mathbf{K}$ *maps to the projection of* y *onto* $\operatorname{range}\mathbf{K}$

$$\alpha = (\mathbf{K}^T\mathbf{K})^{-1}\mathbf{K}^T y. \tag{29}$$

2.3.2 Solving the Smoothing Problem

Consider again a reproducing kernel Hilbert space $\mathcal{H} \subseteq \mathbb{R}^{\mathcal{X}}$ with reproducing kernel k. Write $\mathcal{H} = \mathcal{S} \oplus \mathcal{S}^\perp$, where $\mathcal{S} = \operatorname{span}\{k_{x_1}, \ldots, k_{x_n}\}$. We now solve the smoothing problem (27)

$$\arg\min_{f \in \mathcal{H}} R_\lambda(f) = \frac{1}{n}\sum_{i=1}^{n}(f(x_i) - y_i)^2 + \lambda||f||_{\mathcal{H}}^2.$$

When $\lambda = 0$, this reduces to the best-approximation problem considered in the previous remark. Accordingly, we assume $\lambda > 0$.

Because projecting any $f \in \mathcal{H}$ onto $\mathcal{S}$ does not change the values at the sample points and thus the error in (27) but reduces the norm if f has a nonzero component in $\mathcal{S}^\perp$, any solution will be of the form

$$f = \sum_{i=1}^{n} \alpha_i k_{x_i},$$

for otherwise the norm penalty could be reduced without reducing the data-adherence loss. In this case, the vectorized evaluation operator E is effected by the matrix $\mathbf{K}$ if we represent a function f by its (not necessarily unique) coordinates α, and $||f||_{\mathcal{H}^2}^2 = \alpha^T\mathbf{K}\alpha$. Our problem (27) can be rewritten as the quadratic form

$$\arg\min_{\alpha \in \mathbb{R}^n} \frac{1}{n}(\mathbf{K}\alpha - y)^T(\mathbf{K}\alpha - y) + \lambda\alpha^T\mathbf{K}\alpha. \tag{30}$$

Noting that the term we are optimizing is differentiable and convex in α, we set the derivative to 0 and find

$$\begin{aligned} 0 &= \frac{2}{\lambda}\mathbf{K}^T(\mathbf{K}\alpha - y) + 2\lambda\mathbf{K}\alpha \\ &= \mathbf{K}\left((\mathbf{K} + \lambda n\mathbf{I})\alpha - y\right), \end{aligned}$$

as $\mathbf{K} = \mathbf{K}^T$. Because $\lambda > 0$, the matrix $\mathbf{K} + \lambda n \mathbf{I}$ is invertible, we can always force $(\mathbf{K} + \lambda n \mathbf{I})\alpha - y$ to reside in $\operatorname{null} \mathbf{K}$ by setting

$$\alpha = (\mathbf{K} + \lambda n \mathbf{I})^{-1} y. \tag{31}$$

If $\mathbf{K}$ is singular, this is not the only choice of α that gives a solution to (27), but it does indeed solve the problem. Algorithm 2 summarizes this solution to the spline smoothing problem.

Algorithm 2: The solution to the spline smoothing problem in an RKHS $\mathcal{H} \subset \mathbb{R}^{\mathcal{X}}$, with wiggliness penalty given by the squared norm $|| \cdot ||_{\mathcal{H}}^2$

$$u^* = \operatorname*{arg\,min}_{u \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^{n} (u(x_i) - y_i)^2 + \lambda ||u||_{\mathcal{H}}^2$$

lies in the span of the representers of evaluation $k_{x_i} = k(\cdot, x_i)$ at the data points $\{x_i\}_{i=1}^n$. Consequently, writing any such function as $u = \sum_{i=1}^n \alpha_i k_{x_i}$, we can find u^* by optimizing over the vector of weights $\alpha \in \mathbb{R}^n$

$$\alpha^* = \operatorname*{arg\,min}_{\alpha \in \mathbb{R}^n} \frac{1}{n} (y - \mathbf{K}\alpha)^T (y - \mathbf{K}\alpha) + \lambda \alpha^T \mathbf{K} \alpha,$$

where $\mathbf{K}$ is the Gram matrix of k on the data points, using (31) (or, if $\mathbf{K}$ is singular and $\lambda = 0$, (29)). Having found α^*, we may evaluate u^* at any $x \in \mathcal{X}$ with n calls to the kernel

$$u^*(x) = \sum_{i=1}^{n} \alpha_i^* k(x, x_i).$$

Data: A set of n sample locations $\{x_i\}_{i=1}^n$ in $\mathcal{X}$ and n corresponding sample values $y_i \in \mathbb{R}$.
Parameters: A regularization penalty $\lambda \geq 0$ and (implicitly) a choice of model space $\mathcal{H}$ and kernel k, plus squared data adherence loss and wiggliness penalty $|| \cdot ||_{\mathcal{H}}^2$.
Result: A set of weights $\alpha \in \mathbb{R}^n$ specifying the empirical risk minimizing function u^*.
Compute the $n \times n$ Gram matrix $\mathbf{K}$

$$(\mathbf{K})_{i,j} \leftarrow k(x_i, x_j);$$

Solve $(\mathbf{K} + \lambda n \mathbf{I})\alpha = y$

$$\alpha \leftarrow (\mathbf{K} + \lambda n \mathbf{I})^{-1} y \text{ (or, if } \mathbf{K} \text{ is singular and } \lambda = 0, \alpha \leftarrow \mathbf{K}^\dagger y);$$

Return α;

2.4 The Decomposition Principle: Seminorm-Minimizing Interpolation and Smoothing Where the Penalty Null Space Has Finite Dimension

We know how to solve interpolation and smoothing problems in an RKHS where regularity is enforced by penalizing the norm. However, many classic wiggliness penalties involve a seminorm penalty. For instance, the natural polynomial spline penalty of the form

$$J_{m,\mathcal{X}}(f) = \int_{\mathcal{X}} (f^{(m)}(x))^2 \, \mathrm{d}x \,,$$

is indefinite over the model spaces that interest us, and hence not a norm. There is a rich mathematical literature[26] describing how to solve optimization problems in such indefinite inner product spaces, which we can place beyond the scope of this article by assuming that the null space of our penalty is of finite dimension.

Assumption 2.57. *Suppose that over the model space $\mathcal{H}$, assumed to be an RKHS, the penalty $J_{m,\mathcal{X}}$ used to enforce regularity is the square of a seminorm penalty whose null space is of finite dimension $m \geq 1$.*

With this assumption, we can use what Berlinet and Thomas-Agnan call "the decomposition principle" ([17], Section 6.1.3) to write $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$, where $\mathcal{H}_0$ is the finite-dimensional null space of the penalty $J_{m,\mathcal{X}}$, and in $\mathcal{H}_1$, $J_{m,\mathcal{X}}$ is the square of a definite norm. We will modify the indefinite penalty $J_{m,\mathcal{X}}$ to turn it into the square of a definite norm $||\cdot||_{\mathcal{H}}$ over all of $\mathcal{H}$.

To this end, we begin by turning $\mathcal{H}_0$ into an RKHS by endowing it with an inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}_0}$ and induced norm. The choice of norm on $\mathcal{H}_0$ is irrelevant to the optimization problems we consider. It is absent from empirical risk minimization problems that penalize wiggliness; for constrained optimization problems, we note that, since $\mathcal{H}_0$ is a finite-dimensional, any choice of norm defines the same topology. Thus, we may arbitrarily choose a norm $||\cdot||_{\mathcal{H}_0}$ on $\mathcal{H}_0$. A common choice is the norm induced by

$$\langle f, g \rangle_{\mathcal{H}_0} = \sum_{j=1}^{m} f(x_j) g(x_j), \tag{32}$$

where $\{x_1, \ldots, x_m\}$ is a unisolvent set for $\mathcal{H}_0$–that is, a set of $m = \dim \mathcal{H}_0$ distinct points for which the only function in $\mathcal{H}_0$ that evaluates to 0 at each point in the set is the zero function. More generally, we can use any collection of m linearly independent functionals, and need not limit ourselves to pointwise evaluation.

Lemma 2.58. *Let $\mathcal{H}_0$ be any finite-dimensional vector space of functions defined on a set $\mathcal{X}$, and let $m = \dim \mathcal{H}_0$ be the dimension of $\mathcal{H}_0$. Letting $U = \{u_1, \ldots, u_m\}$ be any linearly independent set of m bounded linear functionals on (and thus a basis for) the dual space of $\mathcal{H}_0$, endow $\mathcal{H}_0$ with the inner product*

$$\langle f, g \rangle_{\mathcal{H}_0} \stackrel{\text{def}}{=} \sum_{i=1}^{m} u_i(f) u_i(g).$$

Then $\mathcal{H}_0$ is an RKHS.

Proof. Since $\mathcal{H}_0$ is a finite-dimensional inner product space, it is a Hilbert space. It remains to be seen that it has a reproducing kernel. Since the functionals in U are linearly independent, the vectorized application map of these functionals (between two spaces of dimension $m < \infty$)

$$\begin{aligned} E : \mathcal{H}_0 &\to \mathbb{R}^m \\ f &\mapsto (u_1(f), \ldots, u_m(f))^T, \end{aligned}$$

is invertible, and on $\mathcal{H}_0$, only the zero function evaluates to 0 by all of the $u_i \in U$. We construct a partition of unity with respect to these functionals. For $i = 1, \ldots, m$, let f_i be the inverse image

[26] The bijection between RKHSs and positive-definite functions can be extended [132] to a surjection between *differences* of positive-definite functions and what are now called reproducing kernel Krein spaces (RKKSs). An example of this *multiplicité*, that is, of a pair of positive kernels whose difference engenders multiple RKKSs is given in [132], p. 247. Despite the fact that we can no longer talk about *the* reproducing kernel space associated with the pair of positive-definite kernels, and other technical challenges involved, this perspective has been useful to applied mathematicians [24, 29, 55, 104, 124]; indeed, as Schwartz writes, *"Néanmoins c'est peut-être là, non pas une monstruosité, mais une nouveauté pleine d'intérêt."* Interested readers are invited to turn to [20].

under this map of the ith standard basis function $e_i \in \mathbb{R}^m$ (i.e., the functions $\{f_i\}_{i=1}^m$ are chosen so as to make the set U their standard dual set, so $u_j(f_i) = \delta_{i,j}$, where δ is the Kronecker delta). With this choice of inner product (and induced norm), the $\{f_j\}_{j=1}^m$ form an orthonormal basis for $\mathcal{H}_0$, for

$$||f_j||_{\mathcal{H}_0}^2 = \sum_{i=1}^m u_i(f_j)u_i(f_j) = \sum_{i=1}^m \delta_{i,j}^2 = 1, \text{ and } \langle f_i, f_j\rangle_{\mathcal{H}_0} = \delta_{i,j}.$$

For all $x \in \mathcal{X}$, let

$$k_x^0 \stackrel{\text{def}}{=} \sum_{i=1}^m f_i(x)f_i;$$

we claim the kernel for $\mathcal{H}_0$ is given by

$$\begin{aligned} k^0(x,y) = \langle k_x^0, k_y^0\rangle_{\mathcal{H}_0} &= \sum_{i=1}^m u_i\left(\sum_{j=1}^m f_j(x)f_j\right) u_i\left(\sum_{j=1}^m f_j(y)f_j\right) \\ &= \sum_{i=1}^m \left(\sum_{j=1}^m f_j(x)\underbrace{u_i(f_j)}_{\delta_{i,j}}\right)\left(\sum_{j=1}^m f_j(y)\underbrace{u_i(f_j)}_{\delta_{i,j}}\right) = \sum_{i=1}^m f_i(x)f_i(y). \end{aligned}$$

Clearly, for all $x \in \mathcal{X}$,

$$k_x^0 = k^0(\cdot, x) = \sum_{i=1}^m f_i(x)f_i$$

is in $\mathcal{H}_0$, for it is a linear combination of the basis functions f_i, and

$$k^0(x,x) = ||k_x^0||_{\mathcal{H}_0}^2 = \sum_{i=1}^m f_i(x)^2 < \infty.$$

We can verify that k_x^0 is indeed a representer of evaluation at x, since for any $f \in \mathcal{H}_0$, we can write $f = \sum_{j=1}^m \alpha_j f_j$; hence,

$$\begin{aligned} \langle f, k_x^0\rangle_{\mathcal{H}_0} &= \sum_{i=1}^m u_i\left(\sum_{j=1}^m \alpha_j f_j\right) u_i\left(\sum_{j=1}^m f_j(x)f_j\right) \\ &= \sum_{i=1}^m \left(\sum_{j=1}^m \alpha_j u_i(f_j)\right)\left(\sum_{j=1}^m f_j(x)u_i(f_j)\right) = \sum_{i=1}^m \alpha_i f_i(x) = f(x). \end{aligned}$$

Consequently, $\mathcal{H}_0$ is an RKHS. □

We are now ready to complement the semi-inner product $\langle\cdot,\cdot\rangle_{\mathcal{H}_1}$ that induces the penalty $J_{m,\mathcal{X}}$ with this inner product $\langle\cdot,\cdot\rangle_{\mathcal{H}_0}$ to form a definite inner product over $\mathcal{H}$ with respect to which $\mathcal{H}$ remains an RKHS. In doing so, we write $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$. The space $\mathcal{H}_1$ may contain equivalence classes of functions, but we might also choose a representative element of each equivalence class, for instance, by imposing boundary conditions as in Example 2.6 or in Section 2.6.1.

Proposition 2.59. *Let $\mathcal{H}$ be a Hilbert space on which the penalty $J_{m,\mathcal{X}} = ||\cdot||_{\mathcal{H}_1}^2$ is the square of a seminorm $||\cdot||_{\mathcal{H}_1}$ in the space $\mathcal{H}$. We maintain our assumption that the (nontrivial) null space $\mathcal{H}_0$ of the seminorm penalty has finite dimension m. Moreover, we retain the inner product $\langle\cdot,\cdot\rangle_{\mathcal{H}_0}$ defined in Lemma 2.58. Then we can "complete" the seminorm penalty by defining the definite inner product on $\mathcal{H}$*

$$\langle f, g\rangle_{\mathcal{H}} = \langle \mathrm{P}_0 f, \mathrm{P}_0 g\rangle_{\mathcal{H}_0} + \langle \mathrm{P}_1 f, \mathrm{P}_1 g\rangle_{\mathcal{H}_1},$$

where P_0 *is the orthogonal projection operator onto* $\mathcal{H}_0$*, and* P_1 *onto* $\mathcal{H}_1 = \mathcal{H}/\mathcal{H}_0$*. In* $\mathcal{H}$*,* $||\mathrm{P}_1 \cdot||^2_{\mathcal{H}_1} = J_{m,\mathcal{X}}(\cdot)$*.*

Proof. Since $\mathcal{H}$ is complete and $\mathcal{H}_0$ is closed and therefore a Hilbert space, $\mathcal{H}_1 = \mathcal{H}/\mathcal{H}_0$ is closed and complete [85]. We need only observe that the two RKHSs $\mathcal{H}_0$ and $\mathcal{H}_1$ are orthogonal with respect to the inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$. It is easy to see that the map

$$\begin{aligned} \mathrm{P}_0 : \mathcal{H} &\to \mathcal{H}_0 \\ f &\mapsto \sum_{i=1}^{m} \langle f, f_i \rangle_{\mathcal{H}_0} f_i = \sum_{i=1}^{m} \left(\sum_{j=1}^{m} u_j(f) u_j(f_i) \right) f_i = \sum_{i=1}^{m} u_i(f) f_i, \end{aligned}$$

defines the orthogonal projection operator from $\mathcal{H}$ to $\mathcal{H}_0$, and $\mathrm{P}_1 = \mathrm{I} - \mathrm{P}_0$ from $\mathcal{H}$ to $\mathcal{H}_1$. By our construction, which set $\mathcal{H}_0$ to be the null space of the bounded linear penalty functional $u \mapsto J_{m,\mathcal{X}}(u)$, we have that $u \in \mathcal{H}_0 \iff J_{m,\mathcal{X}}(u) = 0 \iff ||u||_{\mathcal{H}_1} = 0$.

We observe that $\mathcal{H}_1$ is the space of functions that every functional in U maps to zero

$$f \in \mathcal{H}_1 \implies (\mathrm{I} - \mathrm{P}_0) f = f, \text{ or } \mathrm{P}_0 f = 0, \text{ i.e., } \sum_{i=1}^{m} u_i(f) f_i = 0,$$

which implies that $u_i(f) = 0$ for each $u_i \in U$ (since the f_i are linearly independent). Consequently, any function $f \in \mathcal{H}_1$ satisfies

$$||f||^2_{\mathcal{H}_0} = \sum_{i=1}^{m} u_i(f) u_i(f) = 0,$$

and indeed $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$, with $\mathcal{H}_0 \perp \mathcal{H}_1$. Thus, for all $f \in \mathcal{H}$, the norm

$$||f||^2_{\mathcal{H}} = ||\mathrm{P}_0 f||^2_{\mathcal{H}_0} + ||\mathrm{P}_1 f||^2_{\mathcal{H}_1} = ||f||^2_{\mathcal{H}_0} + ||f||^2_{\mathcal{H}_1}$$

is definite since $||f||_{\mathcal{H}} = 0$ implies that both the null space component $\mathrm{P}_0 f$ and the wiggly component $\mathrm{P}_1 f$ have zero norm in their respective Hilbert spaces. □

In summary, working with an indefinite penalty in an RKHS $\mathcal{H}$ with finite null space involves constructing two orthogonal spaces: the space of functions that are "beyond reproach" $\mathcal{H}_0$ and the space $\mathcal{H}_1$, all of whose nonzero functions' comportment earns them a nonzero wiggliness penalty. $\mathcal{H}_0$ is an RKHS; if one of $\mathcal{H}_1$ or $\mathcal{H}_0$ is an RKHS, so is the other. In this case, the kernel k of $\mathcal{H}$ is the sum of the kernels k^0 of $\mathcal{H}_0$ and k^1 of $\mathcal{H}_1$ ([4], Section 6).

2.5 Wahba's Representer Theorem: Using Finite-Dimensional Matrix Algebra to Solve an Empirical Risk Minimization Problem with Seminorm Penalty over $\mathcal{H}$

As in the previous section, we can write our model space $\mathcal{H}$ as the direct sum of a penalty null space $\mathcal{H}_0$, with dimension $m = \dim \mathcal{H}_0$ and basis $\{\phi_1, \ldots, \phi_m\}$, and a space $\mathcal{H}_1$ of wiggly functions. We are given an empirical risk minimization problem over $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$

$$\text{find } u^* = \arg\min_{u \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^{n} (\mathrm{L}_i u - y_i)^2 + \lambda ||\mathrm{P}_1 u||^2_{\mathcal{H}_1}, \tag{33}$$

where the L_i are a set of bounded linear operators (such as pointwise evaluation in an RKHS) with Riesz representers [121] $\eta_i \in \mathcal{H}$ so that $\mathrm{L}_i f = \langle f, \eta_i \rangle_{\mathcal{H}}$ for any $f \in \mathcal{H}$ and P_1 is the orthogonal projection operator onto $\mathcal{H}_1$. For those interested in inverse problems apart from signal approximation

and reconstruction, we stress that each L_i may be any bounded linear operator applied to the signal, such as pointwise derivative evaluations, integrals over a domain, or observations through a linear instrument with known system response[27]. They need only have Riesz representers η_i.

The representers can be evaluated pointwise by applying the operators they represent to the kernel. By the reproducing property,

$$\eta_i(t) = \langle \eta_i, k_t \rangle_{\mathcal{H}} = \mathrm{L}_i k_t = \mathrm{L}_i k(\cdot, t).$$

Moreover, since the operator P_1 is self-adjoint, the projection of the representer onto the penalized subspace $\xi_i = \mathrm{P}_1 \eta_i$ can be evaluated pointwise using the kernel k^1 of $\mathcal{H}_1$

$$\xi_i(t) = \langle \xi_i, k_t \rangle_{\mathcal{H}} = \langle \mathrm{P}_1 \eta_i, k_t \rangle_{\mathcal{H}} = \langle \eta_i, \mathrm{P}_1 k_t \rangle_{\mathcal{H}} = \langle \eta_i, k_t^1 \rangle_{\mathcal{H}} = \mathrm{L}_i k^1(\cdot, t).$$

Consequently, the inner product between any two functionals $\xi_i = \mathrm{P}_1 \eta_i$ and $\xi_j = \mathrm{P}_1 \eta_j$, which are projections of the representers of evaluation η_i and η_j onto the penalized space $\mathcal{H}_1$, can be computed as follows:

$$\langle \xi_i, \xi_j \rangle_{\mathcal{H}} = \langle \mathrm{P}_1 \eta_i, \mathrm{P}_1 \eta_j \rangle_{\mathcal{H}} = \langle \eta_i, \mathrm{P}_1 \eta_j \rangle_{\mathcal{H}} = \mathrm{L}_i \xi_j = \mathrm{L}_i(\mathrm{L}_j(k_s^1)) = \mathrm{L}_i(s \mapsto (\mathrm{L}_j(t \mapsto k^1(s,t)))),$$

since P_1 is an orthogonal projection operator, meaning it is self-adjoint and $\mathrm{P}_1^2 = \mathrm{P}_1$. In this way, the inner product between ξ_i and ξ_j is the number that results from applying L_i to the function $s \mapsto \mathrm{L}_j k_s^1$. The inner product $\langle \xi_i, \xi_j \rangle_{\mathcal{H}}$ can therefore be computed without explicit knowledge of the η_i and η_j, just the functionals L_i and L_j they represent, as well as the kernel. (For this, it certainly helps to have a kernel in closed form that can be evaluated quickly!)

We can write $\mathcal{H}_1$ as the direct sum of $\mathcal{S} = \operatorname{span}\{\xi_1, \ldots, \xi_n\}$ and its orthogonal complement $\mathcal{S}^\perp$ since $\mathcal{S}$ is closed (it's a finite-dimensional space) and the orthogonal complement of $\mathcal{S}^\perp$ is $(\mathcal{S}^\perp)^\perp = \mathcal{S}$.

Any element of $\mathcal{H}$, and thus the solution to (33), can be written (uniquely) as

$$u^* = \underbrace{\sum_{j=1}^{m} d_j \phi_j}_{u_0 \in \mathcal{H}_0} + \underbrace{\sum_{i=1}^{n} c_i \xi_i}_{u_{\mathcal{S}} \in \mathcal{S} \subset \mathcal{H}_1} + \underbrace{\rho}_{u_{\mathcal{S}^\perp} \in \mathcal{S}^\perp \subset \mathcal{H}_1}.$$

Let us define $\mathbf{\Sigma}$ to be the $n \times n$ matrix whose ith row and jth column contains $(\mathbf{\Sigma})_{i,j} = \langle \xi_i, \xi_j \rangle_{\mathcal{H}}$, and let the matrix $\mathbf{T}$ be the $n \times m$ matrix defined by $(T)_{i,j} = \phi_j(x_i)$. We can, by orthogonality, write (33) (nearly!) as a finite-dimensional linear algebra problem by writing the solution as $u^* = u_0 + s + \rho$, with $u_0 \in \mathcal{H}_0$, $s \in \mathcal{S}$, and $\rho \in \mathcal{S}^\perp$

$$c^*, d^*, \rho^* = \underset{c \in \mathbb{R}^n,\, d \in \mathbb{R}^m\, \rho^* \in \mathcal{S}^\perp}{\arg\min} \frac{1}{n} ||y - (\mathbf{\Sigma} c + \mathbf{T} d)||_{\mathbb{R}^n}^2 + \lambda \left(c^T \mathbf{\Sigma} c + ||\rho||_{\mathcal{H}_1}^2 \right). \tag{34}$$

Wahba and Kimeldorf employed an elementary orthogonality argument to remove ρ from the above problem, making it a (convex) matrix algebra problem.

Proposition 2.60 (The representer theorem (Wahba and Kimeldorf [79]))**.** *In the solution u^* to the empirical risk minimization problem (33), the component $\rho \in \mathcal{S}^\perp \subset \mathcal{H}_1$ must be 0.*

[27] It is a well-known fact from linear systems theory [25, 109] that (as a consequence of Young's inequality) the convolution operator $u \mapsto g * u$, where g is the absolutely summable or integrable impulse response of a linear, time-invariant system, maps bounded signals to bounded signals and is in fact a bounded linear operator whose operator norm equals the ℓ^1 or L^1 norm of g.

Proof. We observe that $u \in \mathcal{S}^\perp \subset \mathcal{H}_1$ if and only if, for each $i = 1, \ldots, n$,

$$0 = \langle u, \xi_i \rangle_{\mathcal{H}} = \langle u, \mathrm{P}_1 \eta_i \rangle_{\mathcal{H}} = \langle \mathrm{P}_1 u, \eta_i \rangle_{\mathcal{H}} = \langle u, \eta_i \rangle_{\mathcal{H}} = \mathrm{L}_i u,$$

and thus, u is mapped to 0 by each functional in $\{L_1, \ldots, L_n\}$. Consequently, the orthogonal projection operator $\mathrm{P}_{\mathcal{S}}$ from $\mathcal{H}_1$ onto $\mathcal{S}$ does not change the measurement-fidelity penalty

$$\frac{1}{n} \sum_{i=1}^{n} (\mathrm{L}_i u - y_i)^2.$$

Now suppose the solution to the empirical risk minimization problem u^*, when projected on the orthogonal subspaces $\mathcal{S}$ and $\mathcal{S}^\perp$ of the penalized space $\mathcal{H}_1$ as $u_{\mathcal{S}} + \rho$, has nonzero component ρ, so that $||\rho||^2_{\mathcal{H}_1} > 0$. Since the functions $u_{\mathcal{S}}$ and $u_{\mathcal{S}} + \rho$ share the same measurement fidelity penalty, we see that $u_{\mathcal{S}}$ has smaller risk in (33), since, by the Pythagorean theorem,

$$||u_{\mathcal{S}}||^2_{\mathcal{H}_1} < ||u_{\mathcal{S}}||^2_{\mathcal{H}_1} + ||\rho||^2_{\mathcal{H}_1} = ||u_{\mathcal{S}} + \rho||^2_{\mathcal{H}_1}.$$

Then $u_{\mathcal{S}} + \rho$ is in fact not the minimum-risk solution to (33), and we conclude u^* has no component in $\rho \in \mathcal{S}^\perp$. □

Since ρ must be 0, the problem (34) really is a finite-dimensional linear algebra problem, and since $\mathbf{\Sigma}$ is a matrix of inner products, we can see that $\mathbf{\Sigma}$ is a symmetric, positive-definite matrix

$$z^T \mathbf{\Sigma} z = \sum_{i=1}^{n} \sum_{j=1}^{n} z_i z_j \langle \xi_i, \xi_j \rangle_{\mathcal{H}_1} = \left\langle \sum_{i=1}^{n} z_i \xi_i, \sum_{j=1}^{n} z_j \xi_j \right\rangle_{\mathcal{H}_1} = \left\| \sum_{i=1}^{n} z_i \xi_i \right\|^2_{\mathcal{H}_1} > 0 \text{ whenever } z \neq 0.$$

Hence, the problem (34) can be simplified

$$c^*, d^* = \underset{c \in \mathbb{R}^n,\, d \in \mathbb{R}^m}{\arg\min} \frac{1}{n} (y - \mathbf{\Sigma} c - \mathbf{T} d)^T (y - \mathbf{\Sigma} c - \mathbf{T} d) + \lambda c^T \mathbf{\Sigma} c. \tag{35}$$

The problem is convex in both c and d (note the positive-semidefiniteness of the matrices $\mathbf{\Sigma}$, $\mathbf{\Sigma}^2$, and $\mathbf{T}^T\mathbf{T}$). Multiplying the objective by n, we see that

$$c^*, d^* = \underset{c \in \mathbb{R}^n,\, d \in \mathbb{R}^m}{\arg\min}\ c^T \mathbf{\Sigma}^2 c + d^T \mathbf{T}^T \mathbf{T} d + 2 d^T \mathbf{T}^T \mathbf{\Sigma} c - 2(d^T \mathbf{T}^T + c^T \mathbf{\Sigma}) y + n \lambda c^T \mathbf{\Sigma} c. \tag{36}$$

First setting to 0 the objective's gradient with respect to c and letting $\mathbf{M} = \mathbf{\Sigma} + n\lambda \mathbf{I}$, we obtain

$$\mathbf{\Sigma}(\underbrace{\mathbf{\Sigma} c + n\lambda c}_{\mathbf{M} c} + \mathbf{T} d - y) = 0,$$

ensuring that at the solution to (36),

$$y = \mathbf{M} c + \mathbf{T} d + \epsilon = \mathbf{M}(c + \mathbf{M}^{-1} \epsilon) + \mathbf{T} d, \text{ with } \epsilon \in \operatorname{null} \mathbf{\Sigma}. \tag{37}$$

It is not hard to see that ϵ can be assumed to be 0[28]. Now setting to 0 the objective's gradient with respect to d, we get that

$$\mathbf{T}^T (\mathbf{T} d - y + \mathbf{\Sigma} c) = 0.$$

[28] Orthogonally diagonalize the symmetric positive semidefinite (and hence normal) matrix $\mathbf{\Sigma}$ in its eigenvectors: $\mathbf{\Sigma} = \mathbf{U}\mathbf{S}\mathbf{U}^T$. Then since $\mathbf{M}^{-1} = \mathbf{U}(\mathbf{S} + n\lambda \mathbf{I})^{-1}\mathbf{U}^T$, we see that $\operatorname{null} \mathbf{\Sigma}$ is invariant under $\mathbf{M}^{-1}$ and $\mathbf{M}^{-1}\epsilon \in \operatorname{null} \mathbf{\Sigma}$. This means that updating the weights on the $\{\xi_i\}_{i=1}^n$ from c^* to $c^* + \mathbf{M}^{-1}\epsilon$, with $\epsilon \in \operatorname{null} \mathbf{\Sigma}$, cannot affect the wiggliness penalty, since $(c^* + \mathbf{M}^{-1}\epsilon)^T \mathbf{\Sigma} (c^* + \mathbf{M}^{-1}\epsilon) = (c^*)^T \mathbf{\Sigma} c^*$. Substituting (37) into n times the data-adherence penalty of

Substituting in our expression for y (37) (and recalling that $\mathbf{M} = \mathbf{\Sigma} + n\lambda\mathbf{I}$), we arrive at

$$\mathbf{T}^T(-n\lambda\mathbf{I})\,c = 0.$$

The solution to (36) can therefore be computed by solving the following linear system [153]

$$\begin{aligned} \mathbf{M}c + \mathbf{T}d &= y \\ \mathbf{T}^T c &= 0. \end{aligned} \tag{38}$$

Note that if we penalize all functions in the space $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$ with the definite penalty $||\cdot||^2_{\mathcal{H}} = ||\cdot||^2_{\mathcal{H}_0} + ||\cdot||^2_{\mathcal{H}_1}$, every nonzero function in $\mathcal{H}$ receives nonzero regularity penalty and the two terms involving $\mathbf{T}$ disappear. Then (36) becomes the spline smoothing problem (30) in $\mathcal{H}$, and (38) may be rewritten

$$\mathbf{M}c = y,$$

with solution (31).

Remark 2.61. *The representer theorem may be profitably applied to weakly nonlinear bounded functionals by approximating them through linearization [157]. It can be applied without error in more general settings, for instance, to allow our empirical risk minimization problem (33) to incorporate arbitrary loss functions (not just mean squared error) and any norm penalty term $g(||\cdot||_{\mathcal{H}_1})$ where $g : [0,\infty) \to \mathbb{R}$ is strictly monotonically increasing (not just $g(x) = \lambda x^2$) [131]. Recent work has sought to extend the theorem beyond the setting of RKHSs associated with a regularization penalty functional to other Banach spaces of finite penalty. For instance, Unser et al. solve spline smoothing problems and inverse problems whose formulation applies the regularization penalty in the Banach space $\mathcal{M}$ of regular Borel measures with total variation norm [151]. The Riesz-Markov theorem (sometimes called, where context permits it, the Riesz representation theorem) guarantees the existence of a unique regular Borel measure ν to represent* any *bounded linear functional Φ on C_0 (the Banach space of continuous functions that vanish at infinity, with sup norm) in the sense that integrating any continuous function $f \in C_0$ with respect to ν gives Φf; the operator norm of Φ equals the total variation of ν [120]. A shift-invariant regularization operator maps a native space of slowly growing functions to measures whose total variation (called gTV) provides a seminorm on the native space. A generalized decomposition principle can make it a norm. As with Wahba's representer theorem, this framework allows certain inverse problems posed in infinite-dimensional spaces to be*

(35) and minimizing the penalty with respect to ϵ subject to the constraint $\mathbf{\Sigma}\epsilon = 0$, we obtain

$$\epsilon^* = \underset{\epsilon\in\mathbb{R}^n}{\arg\min}\,(\mathbf{\Sigma}c + n\lambda c + \mathbf{T}d + \epsilon - \mathbf{\Sigma}c - \mathbf{T}d)^T(\mathbf{\Sigma}c + n\lambda c + \mathbf{T}d + \epsilon - \mathbf{\Sigma}c - \mathbf{T}d) \text{ subject to } \mathbf{\Sigma}\epsilon = 0,$$

from which we compute the Lagrangian

$$n^2\lambda^2 c^T c + 2n\lambda c^T\epsilon + \epsilon^T\epsilon + \mu^T\mathbf{\Sigma}\epsilon,$$

where we have introduced the Lagrangian dual variable μ. The first-order conditions with respect to ϵ and μ yield the system

$$\begin{pmatrix} \mathbf{\Sigma} & \mathbf{0} \\ 2\mathbf{I} & \mathbf{\Sigma} \end{pmatrix}\begin{pmatrix} \epsilon \\ \mu \end{pmatrix} = \begin{pmatrix} 0 \\ -2n\lambda c \end{pmatrix}.$$

Since $\mathbf{\Sigma}$ is invertible, so too is the matrix on the left-hand side; we can therefore solve this system by left-multiplying both sides by this inverse:

$$\begin{pmatrix} \epsilon \\ \mu \end{pmatrix} = \begin{pmatrix} \mathbf{\Sigma}^{-1} & \mathbf{0} \\ -2\mathbf{\Sigma}^{-2} & \mathbf{\Sigma}^{-1} \end{pmatrix}\begin{pmatrix} 0 \\ -2n\lambda c \end{pmatrix} = \begin{pmatrix} 0 \\ -2n\lambda\mathbf{\Sigma}^{-1}c \end{pmatrix};\ \text{in particular, } \epsilon = 0.$$

transformed into finite-dimensional linear algebra problems; as with the thin-plate splines on an interval and in Euclidean space, the solution can be written in terms of a finite number of evaluations of a slowly growing Green's function of the regularization operator and a basis of the finite-dimensional regularization penalty null space. This work differs from the Wahba-Kimeldorf representer theorem in the knot placement and in its use of Green's function that need not be a reproducing kernel, whose native space need not be an RKHS.

2.5.1 A Corollary of the Representer Theorem: Spline Smoothing Using a Seminorm Wiggliness Penalty with a Finite-Dimensional Null Space

When the $\{\mathrm{L}_i\}_{i=1}^n$ are evaluation functionals, the matrix $\mathbf{\Sigma}$ of inner products in $\mathcal{H}_1$ of the projection of the representers of evaluation onto $\mathcal{H}_1$ becomes the Gram matrix $\mathbf{K}$ of the RKHS $\mathcal{H}_1$ of penalized functions associated with the n sample locations of the data one wishes to smooth. Algorithm 3 describes how to set up and solve the linear system (38) for the spline smoothing problem in a space $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$, where the penalty applies only to the function component in $\mathcal{H}_1$.

2.6 Examples of Laplacian-Based Wiggliness Penalties

A natural choice of wiggliness penalty involves the Laplacian, Laplace matrix, or Laplace-Beltrami operator. These penalties are popular because the Laplacian-based penalties give natural notions of wiggliness, and the Laplacian commutes with isometries: in Euclidean space, the penalties are unaffected by rotations and translations; on graphs, by vertex relabelings that preserve the edge structure. The penalty can also be motivated using optimal transport; see [27], Section 2.2.

The penalties $J_{m,\mathcal{X}}$ we consider take the form

$$\int_{\mathcal{X}} \left(\Delta^{m/2} f(x)\right)^2 \mathrm{d}x \;\text{ (if } m \text{ is even), or } \int_{\mathcal{X}} \left\|\nabla\left(\Delta^{(m-1)/2} f(x)\right)\right\|^2 \mathrm{d}x \;\text{ (if } m \text{ is odd),}$$

which can also be written, where boundary conditions permit it, as

$$(-1)^m \int_{\mathcal{X}} f(x)\Delta^m f(x)\,\mathrm{d}x\,.$$

For example, with $m = 0$, the penalty $\int_{\mathcal{X}}(f(x))^2\,\mathrm{d}x$ measures regularity (or signal energy); setting $m = 1$, the penalty $\int_{\mathcal{X}} ||\nabla f(x)||^2\,\mathrm{d}x$ measures the Dirichlet energy, closely related to total variation[29].

Such penalty functionals $J_{m,\mathcal{X}}$, on the corresponding model space $\mathcal{H}$, have finite-dimensional null spaces that induce the structure $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$ described in the previous section. In particular, on a compact manifold, the null space of the Laplace-Beltrami operator contains only the functions that assume constant values.

They are also convenient choices for constructing splines via Mercer synthesis. For any bounded domain $\mathcal{X}$ in $\mathbb{R}^d$, the eigenfunctions of the Laplacian form a complete orthonormal system for $L^2(\mathcal{X})$ and are naturally sorted in increasing Dirichlet energy. The eigenfunctions of the Laplacian can be seen as local extrema of the Dirichlet energy functional subject to a normality constraint[30], and the corresponding eigenvalue is the Dirichlet energy of the eigenfunction (see Lemma 2.2 of [107]).

We consider several examples of splines using this Laplacian-based penalty functional in this section, before seeing how they make congenial company with the thin-plate splines on the sphere.

[29] See [44], p. 42. In Sobolev spaces, this penalty is related to that of the $m = 0$ case by Poincaré's inequalities [116]. The version on the circle (i.e., on [0,1] with periodic boundary conditions) is called the Poincaré-Wirtinger inequality and can be proved using the zero-mean Fourier expansion (46); the version on the sphere, by spherical harmonic expansion [16].

[30] Posed in the Hilbert space $W^{1,2} = H^1$, i.e., the Sobolev space of order 1. See Appendices A.1 and A.2 of [107].

Algorithm 3: Given an RKHS $\mathcal{H} \subset \mathbb{R}^{\mathcal{X}}$ and a seminorm wiggliness penalty $J_{m,\mathcal{X}}(\cdot) = ||\mathrm{P}_1 \cdot ||^2_{\mathcal{H}_1}$, Wahba's representer theorem (Proposition 2.60) permits us to write the solution to the spline smoothing empirical risk minimization problem in $\mathcal{H}$

$$u^* = \operatorname*{arg\,min}_{u \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^{n} (u(x_i) - y_i)^2 + \lambda ||\mathrm{P}_1 u||^2_{\mathcal{H}_1}$$

as follows:

$$u^* = \sum_{j=1}^{m} d_j \phi_j + \sum_{i=1}^{n} c_i k^1(\cdot, x_i),$$

where k^1 is the reproducing kernel of $\mathcal{H}_1$. Consequently, u^* may be found directly from the vectors of weights $c \in \mathbb{R}^n$ and $d \in \mathbb{R}^m$

$$c^*, d^* = \operatorname*{arg\,min}_{c \in \mathbb{R}^n,\, d \in \mathbb{R}^m} \frac{1}{n} (y - \mathbf{K}_1 c - \mathbf{T} d)^T (y - \mathbf{K}_1 c - \mathbf{T} d) + \lambda c^T \mathbf{K}_1 c,$$

where $\mathbf{K}_1$ is the Gram matrix of k^1 on the data points, by solving the linear system (38). This algorithm sets up and solves that system. Given c and d, u^* may be evaluated at any $x \in \mathcal{X}$ with n evaluations of the kernel and m evaluations of ϕ_j

$$u^*(x) = \sum_{j=1}^{m} d_j \phi_j(x) + \sum_{i=1}^{n} c_i k^1(x, x_i).$$

Be warned: for notational simplicity, in this pseudocode, we use 1-indexing.

Data: A set of n sample locations $\{x_i\}_{i=1}^n$ in $\mathcal{X}$ and n corresponding sample values $y_i \in \mathbb{R}$.
Parameters: A regularization penalty $\lambda \geq 0$ and (implicitly) a choice of model space $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$ with reproducing kernel $k = k^0 + k^1$ and seminorm wiggliness penalty $||\mathrm{P}_1 \cdot ||^2_{\mathcal{H}_1}$, whose finite-dimensional null space $\mathcal{H}_0$ has basis $\{\phi_1, \ldots, \phi_m\}$.
Result: A set of basis function weights $c \in \mathbb{R}^n$ and $d \in \mathbb{R}^m$ specifying the empirical risk minimizing function u^*.

Compute the $n \times n$ Gram matrix $\mathbf{K}_1$ in whose ith row and jth column reposes the value

$$(\mathbf{K}_1)_{i,j} \leftarrow k^1(x_i, x_j);$$

Compute the $n \times m$ matrix $\mathbf{T}$, which satisfies

$$(\mathbf{T})_{i,j} \leftarrow \phi_j(x_i);$$

Augment the Gram matrix of k^1 on our data set with null-space basis function matrix $\mathbf{T}$ to form an $(n+m) \times (n+m)$ matrix $\mathbf{K}$ and set y accordingly

$$\mathbf{K} \leftarrow \begin{pmatrix} \mathbf{K}_1 + \lambda n \mathbf{I}_{n \times n} & \mathbf{T} \\ \mathbf{T}^T & \mathbf{0}_{m \times m} \end{pmatrix} \text{ and } y \leftarrow \begin{pmatrix} y \\ \mathbf{0}_m \end{pmatrix};$$

Solve $\mathbf{K}\alpha = y$,

$$\alpha \leftarrow \mathbf{K}^{-1} y \text{ (or, if } \mathbf{K}_1 \text{ is not strictly positive-definite and } \lambda = 0,\ \alpha \leftarrow \mathbf{K}^{\dagger} y);$$

Return the spline weights $c \leftarrow \alpha[1:n]$ and $d \leftarrow \alpha[n+1:n+m]$;

The rest of the section is organized as follows. We begin by considering the polynomial splines on $[0,1]$ with natural boundary conditions. We find a closed-form expression for the kernel using the Green's function of the associated differential operator (see Equation (43)). The penalty null space consists of the polynomials of degree at most $m-1$. Next, we consider the polynomial splines on $[0,1]$ with periodic boundary conditions. We use Mercer synthesis (see Section 2.2.3) to construct the kernel using the Fourier basis as our complete orthonormal system for $L^2([0,1])$. No matter the order of the spline, the null space $\mathcal{H}_0$ always consists only of constant functions. As our third example, we define the thin-plate splines on graphs and obtain the kernel by taking the Moore-Penrose pseudoinverse of the Laplacian matrix; a Mercer-like expansion is given in Equation (51). Here too the null space consists only of constant functions no matter the spline order. We follow up with the thin-plate splines in Euclidean space $\mathbb{R}^d$. Mercer's theorem (Proposition 2.28) does not apply here ($\mathbb{R}^d$ is not compact); we find the kernel using the Green's function of the differential operator. The penalty null space here consists of polynomials of degree at most $m-1$. Finally, we use Mercer synthesis (Proposition 2.38) to define the thin-plate splines on the sphere using the spherical harmonics. The penalty null space contains the constant functions[31]. We summarize this roadmap in Table 1.

spline	$\mathcal{X}$	null space $\mathcal{H}_0$	derivation of kernel
natural polynomial (natural bdy)	$[0,1]$	polynomials $\mathcal{P}_{m-1}$	Green's function
circular polynomial (periodic bdy)	$[0,1]$	span $\{1\}$	"Fourier side" synthesis
thin-plate on graph	$\{1,\dots,n\}$	span $\{1\}$	Moore-Penrose `pinv`
Euclidean thin-plate	$\mathbb{R}^d$	polynomials $\mathcal{P}_{m-1}$	Green's function
thin-plate on the sphere	$\mathbb{S}^2$	span $\{1\}$	"Fourier side" synthesis

Table 1: The splines we consider, each of order m, as well as their index set, their null space (either the constant functions or the polynomials of degree at most $m-1$), and the approach we take to derive their kernel ("Fourier side" synthesis or via use of the Green's function of the differential operator; in the case of the thin-plate splines on a graph, we use the Moore-Penrose pseudoinverse of the Laplacian matrix). The order m can be as low as 1 (or even 0; see Example 2.63) for the thin-plate splines on graphs, and 1 for the splines on [0,1] and $\mathbb{S}^2$; for thin-plate splines in Euclidean d-space, we have the technical restriction that $m > d/2$.

2.6.1 Natural Polynomial Splines

We derive[32] the natural polynomial splines [130] (also called D^m splines [17]) on $\mathcal{X} = [0,1]$ using RKHS theory [3, 6, 34, 132], reproducing the presentation in [79, 80, 158] so that we can use Wahba's representer theorem [79, 80, 131, 159] (Proposition 2.60). The same results can be found using only integration by parts and an elementary proof of optimality [1, 58].

Model space: Let $\mathcal{H} = W^{m,2}$ be the Sobolev space of functions u such that $u, u', \dots, u^{(m-1)}$ are absolutely continuous, $u^{(m)}$ is defined almost everywhere, and $u^{(m)} \in L^2(0,1)$. (Here, $u^{(m)}$ is the ordinary derivative. See Definition 2.5 for motivation for this definition.) This Sobolev space is an RKHS with respect to its canonical inner product (see Definition 2.5); however, we wish to endow it with an inner product related to the seminorm penalty

$$\int_0^1 (f^{(m)}(x))^2 \,\mathrm{d}x\,, \text{ induced by the semidefinite bilinear form } \int_0^1 f^{(m)}(x) g^{(m)}(x)\,\mathrm{d}x\,. \tag{39}$$

[31] As we will see in the IPOL demo, this null space of constant functions can be profitably employed to estimate spherical averages of scattered data on the sphere.

[32] Note that this derivation yields polynomial splines of odd degree only: natural cubic splines, natural quintic splines, and so forth.

Observe that our definition of $\mathcal{H}$ allows us to invoke Taylor's theorem with the Lagrange remainder term and write, for all $t \in [0,1]$,

$$u(t) = \underbrace{\sum_{n=0}^{m-1} \frac{t^n}{n!} u^{(n)}(0)}_{u_0(t) \in \mathcal{H}_0} + \underbrace{\int_0^1 \frac{(t-x)_+^{m-1}}{(m-1)!} u^{(m)}(x)\,\mathrm{d}x}_{u_1(t) \in \mathcal{H}_1}, \tag{40}$$

where $(\cdot)_+ = \max(\cdot, 0)$ is the rectifier. We make the recognition that the first term u_0 is in the null space of the penalty seminorm (39). We will show that these two terms are the two components of a unique decomposition of $u \in \mathcal{H}$ into $u_0 \in \mathcal{H}_0$ and $u_1 \in \mathcal{H}_1$, respectively, where $\mathcal{H}_0$ is the space of polynomials of degree $\leq m-1$ and $\mathcal{H}_1$ the space of functions u for which $u(0) = u'(0) = \ldots = u^{(m-1)}(0)$ [158].

Null space: The null space $\mathcal{H}_0$ of (39) in $\mathcal{H}$ is a finite-dimensional RKHS spanned by polynomials of degree at most $m-1$. The decomposition principle works with any choice of separating set U of m functionals in the definition of the inner product on $\mathcal{H}_0$. To better match Taylor's theorem, we define[33]

$$\langle f, g \rangle_{\mathcal{H}_0} = \sum_{n=0}^{m-1} f^{(n)}(0) g^{(n)}(0).$$

Letting $\phi_n(t) = \frac{t^n}{n!}$, we can define a kernel for $\mathcal{H}_0$ via the expansion

$$k^0(s,t) = \sum_{n=0}^{m-1} \phi_n(s)\phi_n(t).$$

Clearly, for all $t \in [0,1]$, the representer of evaluation $k_t^0 = k^0(\cdot, t) \in \mathcal{H}_0$. Since $\frac{\partial^m}{\partial x^m}\phi_n(x)\big|_{x=0} = \delta_{m,n}$, where δ is the Kronecker δ, we can easily verify that, for any $f \in \mathcal{H}_0$, the reproducing property holds. Let $f = \sum_{n'=0}^{m-1} \alpha_{n'}\phi_{n'}$. Then

$$\begin{aligned}
\langle k_t^0, f \rangle_{\mathcal{H}_0} &= \left\langle \sum_{n=0}^{m-1} \phi_n(t)\phi_n(\cdot), \sum_{n'=0}^{m-1} \alpha_{n'}\phi_{n'}(\cdot) \right\rangle_{\mathcal{H}_0} \\
&= \sum_{i=0}^{m-1} \frac{\partial^i}{\partial x^i} \left(\sum_{n=0}^{m-1} \phi_n(t)\phi_n(x) \right) \Bigg|_{x=0} \frac{\partial^i}{\partial x^i} \left(\sum_{n'=0}^{m-1} \alpha_{n'}\phi_{n'}(x) \right) \Bigg|_{x=0} \\
&= \sum_{i=0}^{m-1} \alpha_i \phi_i(t) = f(t).
\end{aligned}$$

Wiggly space: For a penalty (39) of order m, let

$$\mathcal{H}_1 = \left\{ u \in \mathcal{H} \mid u(0) = u'(0) = \ldots = u^{(m-1)}(0) = 0 \right\}. \tag{41}$$

These boundary conditions[34] enforce the requirement that $||u||_{\mathcal{H}_0} = \sum_{i=0}^m (u^{(i)}(0))^2 = 0$; hence the

[33] For other choices, see [7].

[34] Note that these boundary conditions are not the natural or Neumann boundary conditions, that is

$$u^{(m)}(0) = \ldots = u^{(2m-1)}(0) = u^{(m)}(1) = \ldots = u^{(2m-1)}(1) = 0.$$

Even though many functions in our model space $\mathcal{H}$ do not satisfy these constraints, functions that solve smoothing problems (27) over this space–that is, splines–do [80, 81, 158]. Thus, these empirical risk minimizing splines satisfy

$$\int_0^1 (f^{(m)}(x))^2\,\mathrm{d}x = (-1)^m \int_0^1 f(x)(\Delta^m f)(x)\,\mathrm{d}x\,,$$

where $\Delta = \frac{\partial^2}{\partial x^2}$ is the one-dimensional Laplacian.

Taylor expansion (40) can be rewritten, for any $u \in \mathcal{H}_1$,

$$u(t) = \int_0^1 \frac{(t-x)_+^{m-1}}{(m-1)!} u^{(m)}(x)\,\mathrm{d}x\,.$$

We can simplify this integrand by recalling that the Green's function for the problem $D^m f = g$ with boundary conditions $f(0) = f'(0) = \ldots = f^{(m-1)}(0) = 0$ is

$$G_m(t,x) = \frac{(t-x)_+^{m-1}}{(m-1)!},$$

so that

$$D^m G_m(t,x) = \delta(t-x); \tag{42}$$

thus, for $u \in \mathcal{H}_1$, we get a sort of generalization of the Dirac δ's sifting property

$$u(t) = \int_0^1 G_m(t,x) u^{(m)}(x)\,\mathrm{d}x\,.$$

We define the inner product on $\mathcal{H}_1$ so that its induced norm, squared, is the order-m spline wiggliness penalty

$$\langle f, g\rangle_{\mathcal{H}_1} \stackrel{\text{def}}{=} \int_0^1 f^{(m)}(x) g^{(m)}(x)\,\mathrm{d}x\,, \text{ and } ||f||^2_{\mathcal{H}_1} = \langle f, f\rangle_{\mathcal{H}_1} = J_{m,\mathcal{X}}(f).$$

The reproducing property for this inner product on $\mathcal{H}_1$ follows from our recognizing that, if $u \in \mathcal{H}_1$–that is, if the $\mathcal{H}_0$ component $u_0 \equiv 0$–then function evaluations of u at a point t look like an inner product of u with the Green's function with one argument fixed at t

$$u(t) = \int_0^1 \frac{(t-x)_+^{m-1}}{(m-1)!} u^{(m)}(x)\,\mathrm{d}x = \int_0^1 G_m(t,x) u^{(m)}(x)\,\mathrm{d}x = \langle G_m(t,\cdot), u\rangle_{\mathcal{H}_1}.$$

To show that the Green's functions represent evaluation at points via the inner product $\langle\cdot,\cdot\rangle_{\mathcal{H}_1}$, we need to define a kernel. The reproducing kernel k^1 associated with the RKHS $\mathcal{H}_1$ is

$$k^1(s,t) = \int_0^1 G_m(s,x) G_m(t,x)\,\mathrm{d}x\,. \tag{43}$$

The *representers of evaluation at* t, $k^1_t \stackrel{\text{def}}{=} k^1(\cdot,t)$, for $t \in [0,1]$, are all in $\mathcal{H}_1$ for $m \geq 1$, since, by (42),

$$\frac{\partial^m}{\partial t^m} k^1(s,t) = \int_0^1 G_m(s,x) \left(\frac{\partial^m}{\partial t^m} G_m(t,x)\right) \mathrm{d}x = \int_0^1 G_m(s,x)\delta(t-x)\,\mathrm{d}x = G_m(s,t); \tag{44}$$

hence $G_m(\cdot,t) \in L^2([0,1])$ and its antiderivatives are absolutely continuous[35] on $[0,1]$. Furthermore, for all $t \in [0,1]$,

$$k^1_t(s) = \int_0^1 G_m(s,x) G_m(t,x)\,\mathrm{d}x = \int_0^1 G_m(s,x) \left(\frac{\partial^m}{\partial x^m} k^1_t(x)\right) \mathrm{d}x\,; \tag{45}$$

comparison with (40) confirms that the component of k^1_t in $\mathcal{H}_0$ is 0.

[35]For $m \geq 1$. If $m = 1$, $\frac{\partial^1}{\partial s^1} k^1_t(s) = 1 - H(s-t)$, where H is the Heavyside step function, and its antiderivative, the inverted ramp $k^1_t = \min(\cdot, t)$, is absolutely continuous on $[0,1]$, since $\min(s,t) = \min(0,t) + \int_0^s (1 - H(u-t))\,\mathrm{d}u$ for all $s \in [0,1]$. For $m > 1$, $\frac{\partial^m}{\partial s^m} k^1_t(s) = G_m(s,t)$ is already absolutely continuous.

The space $\mathcal{H}_1$ is an RKHS since the evaluation functionals $u \mapsto u(t)$ are all bounded; indeed, by the Cauchy-Schwarz inequality,

$$|u(t)| = \int_0^1 G_m(t,x)u^{(m)}(x)\,\mathrm{d}x \leq \sqrt{\int_0^1 (G_m(t,x))^2\,\mathrm{d}x}\sqrt{\int_0^1 u^{(m)}(x)u^{(m)}(x)} = \sqrt{k^1(t,t)}||u||_{\mathcal{H}_1}.$$

The reproducing property can be verified for $u \in \mathcal{H}_1$. Using Equation (44), we see that

$$\langle u, k_t^1\rangle_{\mathcal{H}_1} = \int_0^1 \left(\frac{\partial^m}{\partial x^m}u(x)\right)\left(\frac{\partial^m}{\partial x^m}k^1(x,t)\right)\mathrm{d}x = \int_0^1 G_m(x,t)u^{(m)}(x)\,\mathrm{d}x = u(t).$$

Thus, evaluation of a function $u \in \mathcal{H}_1$ at $t \in [0,1]$ can be performed by inner product of u and the representer of evaluation at t, k_t^1, which is the Green's function with one argument fixed $G_m(\cdot,t)$, integrated m times. $\mathcal{H}_1$, a closed subset of $\mathcal{H}$, is therefore a reproducing kernel Hilbert space with reproducing kernel k^1. Further technical details can be found in [80, 158], along with a Bayesian interpretation of the natural splines of order m.

The reproducing kernel for $\mathcal{H}$: We can verify the orthogonality of $\mathcal{H}_0$ and $\mathcal{H}_1$. Clearly, if $u \in \mathcal{H}_0$, then $u^{(m)}(t) \equiv 0$, and

$$||u||_{\mathcal{H}_1} = \sqrt{\langle u,u\rangle_{\mathcal{H}_1}} = \int_0^1 (u^{(m)}(x))^2\,\mathrm{d}x = 0.$$

If $u \in \mathcal{H}_1$, $u(0) = \ldots = u^{(m-1)}(0)$, and

$$||u||_{\mathcal{H}_0} = \sqrt{\langle u,u\rangle_{\mathcal{H}_0}} = \sum_{n=0}^{m-1}\left(u^{(n)}(0)\right)^2 = 0.$$

By the decomposition principle, $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$, with inner product and induced norm

$$\langle f,g\rangle_{\mathcal{H}} = \langle \mathrm{P}_0 f, \mathrm{P}_0 g\rangle_{\mathcal{H}_0} + \langle \mathrm{P}_1 f, \mathrm{P}_1 g\rangle_{\mathcal{H}_1} \text{ and } ||u||_{\mathcal{H}}^2 = ||u_0||_{\mathcal{H}_0}^2 + ||u_1||_{\mathcal{H}_1}^2,$$

where the orthogonal projection operators

$$\mathrm{P}_0 : \mathcal{H} \to \mathcal{H}_0 \qquad\qquad \mathrm{P}_1 : \mathcal{H} \to \mathcal{H}_1$$
$$u \mapsto \sum_{n=0}^{m-1}\frac{u^{(n)}(0)}{n!}\phi_n\,, \qquad\qquad u \mapsto u - \sum_{n=0}^{m-1}\frac{u^{(n)}(0)}{n!}\phi_n\,,$$

are as in Section 2.4, and with reproducing kernel equal to the the sum of k^0 and k^1 (since the kernel of the direct sum space is the sum of the kernels [3])

$$k(s,t) = k^0(s,t) + k^1(s,t) = \sum_{n=0}^{m-1}\frac{s^n}{n!}\frac{t^n}{n!} + \int_0^1 G_m(s,x)G_m(t,x)\,\mathrm{d}x\,.$$

Thus, $k_t = k_t^0 + k_t^1$ and $\mathrm{P}_0 k_t = k_t^0$ and $\mathrm{P}_1 k_t = k_t^1$. That the reproducing property holds follows immediately from Taylor's theorem with the Lagrange remainder term

$$\begin{aligned}
u(t) &= \sum_{n=0}^{m-1}\frac{t^n}{n!}u^{(n)}(0) + \int_0^1 \frac{(t-x)_+^{m-1}}{(m-1)!}u^{(m)}(x)\,\mathrm{d}x \\
&= \sum_{n=0}^{m-1}\frac{\partial^n}{\partial x^n}\left(\sum_{i=0}^{m-1}\phi_i(t)\phi_i(x)\right)\Bigg|_{x=0}\frac{\partial^n}{\partial x^n}\Bigg(\underbrace{\sum_{i=0}^{m-1}u^{(i)}(0)\phi_i(x)}_{(\mathrm{P}_0 u)(x) = u_0(x)}\Bigg)\Bigg|_{x=0} + \\
&\qquad \int_0^1 \left(\frac{\partial^m}{\partial x^m}k_t^1(x)\right)u^{(m)}(x)\,\mathrm{d}x \\
&= \langle k_t^0, \mathrm{P}_0 u\rangle_{\mathcal{H}_0} + \langle \mathrm{P}_1 u, k_t^1\rangle_{\mathcal{H}_1} = \langle u, k_t\rangle_{\mathcal{H}},
\end{aligned}$$

since

$$\frac{\partial^m}{\partial x^m}u_1(x) = \frac{\partial^m}{\partial x^m}u(x) - \frac{\partial^m}{\partial x^m}\left(\sum_{n=0}^{m-1}\frac{u^{(n)}(0)}{n!}\frac{x^n}{n!}\right) = \frac{\partial^m}{\partial x^m}u(x).$$

In fact, following the same process, a generalized Taylor's expansion may be derived using other differential operators, using the corresponding Green's function to define the kernel of $\mathcal{H}_1$ and remainder term and extended Chebyshev system to define the polynomial term [79].

Solving the Spline Smoothing Problem with Polynomial Splines. We can use Algorithm 4 to solve system (38) for the natural polynomial splines. Here $\mathbf{T}$ contains samples of the polynomials ϕ_j, which span $\mathcal{H}_0$, at our scattered data, and $\mathbf{K}_1$ the kernel defined in (45).

Remark 2.62 (Adapting the algorithm to other intervals)**.** *This algorithm in fact works for any interval $[a,b]$ whose left endpoint $a = 0$. To adapt this algorithm to data in some interval $[a,b]$, with $a \neq 0$, one need only reparameterize the data with the map $t \mapsto t - a$ that sets the left endpoint to 0. However, the transformation $t \mapsto \frac{t-a}{b-a}$ of $[a,b]$ to the unit interval $[0,1]$ is more commonly used (see, e.g., [148], Proposition 2).*

Consider the natural cubic spline in this case. Let $\mathcal{X} = [a,b]$ and $\mathcal{H} \subseteq \mathbb{R}^{\mathcal{X}}$ be the set of absolutely continuous functions on $[a,b]$ whose first derivatives are absolutely continuous and second derivatives square integrable on $[a,b]$. Write $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$, with $\mathcal{H}_0 = \operatorname{span}\{1, x-a\}$ and $\langle f,g\rangle_{\mathcal{H}_0} = f(a)g(a) + f'(a)g'(a)$. The reproducing kernel k^0 for $\mathcal{H}_0$ is $k^0(x,u) = 1 + (u-a)(x-a)$: indeed, for $f \in \mathcal{H}_0$,

$$f(x) = f(a)\cdot 1 + f'(a)(x-a) = \langle f, k^0_x\rangle_{\mathcal{H}_0}.$$

Let $\mathcal{H}_1$ be the functions $f \in \mathcal{H}$ such that $f(a) = f'(a) = 0$. Define $\langle f,g\rangle_{\mathcal{H}_1} = \langle f'', g''\rangle_{L^2([a,b])}$, so that $f \in \mathcal{H}_0 \implies ||f||^2_{\mathcal{H}_1} = 0$. The reproducing kernel k^1 for $\mathcal{H}_1$ is

$$\begin{aligned} k^1(x,y) &= \int_a^b (k^1_x)''(u)(k^1_y)''(u)\,\mathrm{d}u = \int_a^b (x-u)_+(y-u)_+\,\mathrm{d}u \\ &= \frac{1}{3}(\min(x,y)^3 - a^3) - \frac{x+y}{2}(\min(x,y)^2 - a^2) + xy(\min(x,y) - a), \end{aligned}$$

as

$$k^1_x(y) = \begin{cases} -\frac{y^3}{6} + \frac{xy^2}{2} - axy + a^2\frac{x+y}{2} - \frac{a^3}{3}, & \text{if } y \le x \\ -\frac{x^3}{6} + \frac{x^2y}{2} - axy + a^2\frac{x+y}{2} - \frac{a^3}{3}, & \text{if } y > x, \end{cases}$$

with

$$(k^1_x)'(y) = \begin{cases} -\frac{y^2}{2} + xy - ax + \frac{a^2}{2}, & \text{if } y \le x \\ \frac{x^2}{2} - ax + \frac{a^2}{2}, & \text{if } y > x, \end{cases}$$

and

$$(k^1_x)''(y) = \begin{cases} x - y, & \text{if } y \le x \\ 0, & \text{if } y > x, \end{cases}$$

satisfies $k^1_x(a) = (k^1_x)'(a) = 0$; and k^1_x and $(k^1_x)'$ absolutely continuous; and $(k^1_x)'' \in L^2([a,b])$. Thus, $f \in \mathcal{H}_1 \implies ||f||^2_{\mathcal{H}_0} = 0$. The reproducing property in $\mathcal{H}_1$ for all $f \in \mathcal{H}_1$ and all $x \in [a,b]$ can be verified using integration by parts (by the absolute continuity of $f \in \mathcal{H}_1$ and f') as follows:

$$\begin{aligned} \langle f, k^1_x\rangle_{\mathcal{H}_1} &= \int_a^b f''(u)(k^1_x)''(u)\,\mathrm{d}u = \int_a^b f''(u)(x-u)_+\,\mathrm{d}u \\ &= x\int_a^x f''(u)\,\mathrm{d}u - \int_a^x f''(u)u\,\mathrm{d}u = x(f'(x) - f'(a)) - (uf'(u)|_a^x) + \int_a^x f'(u)\,\mathrm{d}u \\ &= x(f'(x) - \underbrace{f'(a)}_{0}) - (xf'(x) - a\underbrace{f'(a)}_{0}) + f(x) - \underbrace{f(a)}_{0} = f(x). \end{aligned}$$

Algorithm 4: Here $\mathcal{X} = [0,1]$ and our RKHS $\mathcal{H} = W^{m,2}(\mathcal{X})$ (see Definition 2.5). $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$, where $\mathcal{H}_0$ is the m-dimensional space of polynomials of degree $\leq m-1$ and $\mathcal{H}_1$ the space of functions $u \in \mathcal{H}$ for which $u(0) = u'(0) = \ldots = u^{(m-1)}(0) = 0$. The wiggliness penalty seminorm on $\mathcal{H}$ is $J_{m,\mathcal{X}}(u) = ||\mathrm{P}_1 u||^2_{\mathcal{H}_1} = \int_0^1 (u^{(m)}(x))^2 \,\mathrm{d}x$, since P_1 removes from u a polynomial of degree at most $m-1$, whose mth derivative is 0. Given a set of sample points $\{x_i\}_{i=1}^n$ in $[0,1]$ and corresponding values $\{y_i\}_{i=1}^n$ in $\mathbb{R}$, the representer theorem (Proposition 2.60) locates the solution to the spline smoothing empirical risk minimization problem in $\mathcal{H}$

$$u^* = \operatorname*{arg\,min}_{u \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^n (u(x_i) - y_i)^2 + \lambda \int_0^1 (u^{(m)}(x))^2 \,\mathrm{d}x\,,$$

in $\mathcal{H}_0$ and a finite-dimensional subspace of $\mathcal{H}_1$

$$u^*(x) = \sum_{j=1}^m d_j \frac{x^j}{j!} + \sum_{i=1}^n c_i \int_0^1 G_m(x,x') G_m(x_i,x') \,\mathrm{d}x'\,.$$

This algorithm recovers c and d by solving the linear system (38). Note: we use 1-indexing.

Data: A set of n sample locations $\{x_i\}_{i=1}^n$ in $\mathcal{X} = [0,1]$ and n corresponding sample values $y_i \in \mathbb{R}$.

Parameters: A regularization penalty $\lambda \geq 0$ and (implicitly) a choice of model space $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$ and seminorm wiggliness penalty $J_{m,\mathcal{X}}$.

Result: A set of basis function weights $c \in \mathbb{R}^n$ and $d \in \mathbb{R}^m$ specifying the empirical risk minimizing function u^*.

Compute the $n \times n$ Gram matrix $\mathbf{K}_1$, which satisfies

$$(\mathbf{K})_{i,j} \leftarrow k^1(x_i,x_j) = \int_0^1 G_m(x_i,u) G_m(x_j,u) \,\mathrm{d}u\,;$$

For reference, the natural linear, cubic, and quintic spline kernels on $[0,1]$ are given below:

m	$k^1(x,y)$
1	$\min(x,y)$
2	$xy\min(x,y) - \frac{x+y}{2}\min(x,y)^2 + \frac{1}{3}\min(x,y)^3$
3	$\frac{x^2y^2\min(x,y)}{4} - \frac{xy(x+y)\min(x,y)^2}{4} + \frac{(x^2+4xy+y^2)}{12}\min(x,y)^3 - \frac{x+y}{8}\min(x,y)^4 + \frac{1}{20}\min(x,y)^5$

Fill the $n \times m$ matrix $\mathbf{T}$ with the basis functions of $\mathcal{H}_0$ evaluated at the sample locations

$$(\mathbf{T})_{i,j} \leftarrow \frac{(x_i)^j}{j!};$$

Augment the Gram matrix of k^1 on our data set with null-space basis function matrix $\mathbf{T}$ to form an $(n+m) \times (n+m)$ matrix $\mathbf{K}$ and set y accordingly

$$\mathbf{K} \leftarrow \begin{pmatrix} \mathbf{K}_1 + \lambda n \mathbf{I}_{n\times n} & \mathbf{T} \\ \mathbf{T}^T & \mathbf{0}_{m\times m} \end{pmatrix} \text{ and } y \leftarrow \begin{pmatrix} y \\ \mathbf{0}_m \end{pmatrix};$$

Solve $\mathbf{K}\alpha = y$,

$$\alpha \leftarrow \mathbf{K}^{-1} y \text{ (or, if } \mathbf{K}_1 \text{ has redundant samples and } \lambda = 0,\ \alpha \leftarrow \mathbf{K}^\dagger y);$$

Return the spline weights $c \leftarrow \alpha[1:n]$ and $d \leftarrow \alpha[n+1:n+m]$;

The reproducing property on $\mathcal{H}$ follows from the fact that $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$. It can be verified using

$$\begin{aligned}\langle f, k_x\rangle_{\mathcal{H}} = \langle f, k_x^0\rangle_{\mathcal{H}_0} + \langle f, k_x^1\rangle_{\mathcal{H}_1} &= f(a)\underbrace{k_x^0(a)}_{1} + f'(a)\underbrace{(k_x^0)'(a)}_{x-a} + \int_a^b f''(u)\underbrace{(k_x^1)''(u)}_{(x-u)_+}\mathrm{d}u \\ &= f(a) + f'(a)(x-a) + xf'(u)|_a^x - \left[uf'(u)|_a^x - \int_a^x f'(u)\,\mathrm{d}u\right] = f(x).\end{aligned}$$

Similarly, the reproducing kernel k^1 for the order 1 natural polynomial spline is $k^1(x,y) = \min(x,y) - a$ (the null space remains $\mathcal{H}_0 = span\{1\}$); and for the order 3 natural polynomial spline, the kernel

$$\begin{aligned}k^1(x,y) = \frac{x^2y^2}{4}(\min(x,y) - a) - \frac{xy(x+y)}{4}(\min(x,y)^2 - a^2) + \frac{(x^2+4xy+y^2)}{12}(\min(x,y)^3 - a^3) - \\ \frac{x+y}{8}(\min(x,y)^4 - a^4) + \frac{1}{20}(\min(x,y)^5 - a^5),\end{aligned}$$

with null space $\mathcal{H}_0 = span\{1, x-a, \frac{(x-a)^2}{2}\}$, so that $k^0(x,u) = 1 + (u-a)(x-a) + \frac{(u-a)^2}{2}\frac{(x-a)^2}{2}$ is a reproducing kernel with inner product $\langle f, g\rangle_{\mathcal{H}_0} = f(a)g(a) + f'(a)g'(a) + f''(a)g''(a)$.

2.6.2 Polynomial Splines on the Circle

To form splines on the circle, we proceed in much the same way as before, setting the index set $\mathcal{X} = [0,1]$ and defining an identical wiggliness penalty seminorm [158]. However, we change the model space to enforce periodicity through boundary conditions. By changing the RKHS boundary conditions, we change the corresponding reproducing kernel. We define this kernel on the "Fourier side", using Mercer synthesis (Proposition 2.38) rather than the Green's function of the penalty differential operator (i.e., the iterated Laplacian) with periodic boundary conditions.

Model space: Let $\mathcal{H}$ be the space of functions u on $[0,1]$ for which $u, u', \ldots, u^{(m-1)}$ are absolutely continuous, $u^{(m)} \in L^2(0,1)$, and u satisfies the periodic boundary conditions: namely, u and its first $m-1$ derivatives agree at the boundary.

$$u^{(k-1)}(1) = u^{(k-1)}(0) \text{ for } k = 1, \ldots, m.$$

Since $u^{(k-1)}$ is absolutely continuous, the boundary conditions are equivalent to

$$\int_0^1 u^{(k)}(x)\,\mathrm{d}x = 0 \text{ for } k = 1, \ldots, m.$$

Null space: Let $\mathcal{H}_0$ be the space of constant functions: $\mathcal{H}_0 = \text{span}\{1\}$. Note that constant functions satisfy the boundary conditions and reside in $\mathcal{H}$. To this space, we can give an inner product $\langle f, g\rangle_{\mathcal{H}_0} = \int_0^1 f(x)\,\mathrm{d}x \int_0^1 g(x)\,\mathrm{d}x$ and turn into an RKHS with reproducing kernel $k^0(s,t) = 1$; clearly $k_t^0 = k^0(\cdot, t) = 1 \in \mathcal{H}_0$. Indeed, for any $f \in \mathcal{H}_0$, we have that

$$\forall t \in [0,1],\ f(t) = \langle f, k_t^0\rangle_{\mathcal{H}_0} = \int_0^1 f(x)\,\mathrm{d}x \cdot \int_0^1 1\,\mathrm{d}x\,.$$

Wiggly space: Let $\mathcal{H}_1 = \{u \in \mathcal{H} \mid \int_0^1 u(x)\,\mathrm{d}x = 0\}$, when endowed with the norm

$$||u||_{\mathcal{H}_1}^2 = \left\|u^{(m)}\right\|_{L^2([0,1])}^2 = \int_0^1 \left(u^{(m)}(t)\right)^2 \mathrm{d}t\,,$$

be the space of zero-mean periodic functions. An elementary fact of Fourier series is that the 1-periodic sinusoids form a complete orthonormal system for $\mathcal{H}_1$ [84, 149]; that is, any function $u \in \mathcal{H}_1$ can be represented as the zero-mean Fourier series

$$\forall t \in [0,1],\ u(t) = \sqrt{2}\sum_{\nu=1}^{\infty} \alpha_\nu \cos(2\pi\nu t) + \sqrt{2}\sum_{\nu=1}^{\infty} \beta_\nu \sin(2\pi\nu t), \tag{46}$$

which converges not just in $L^2([0,1])$ but also, as indicated by the notation, pointwise–in fact, absolutely and uniformly–for any $u \in \mathcal{H}_1$[36], and has finite $\mathcal{H}_1$ norm. Thus, by the square integrability of the mth derivative of u and orthonormality of our Fourier series basis,

$$\begin{aligned}
||u||_{\mathcal{H}_1}^2 &= \int_0^1 \left(\frac{\mathrm{d}^m}{\mathrm{d}t^m}\left(\sqrt{2}\sum_{\nu=1}^{\infty} \alpha_\nu \cos(2\pi\nu t) + \sqrt{2}\sum_{\nu=1}^{\infty} \beta_\nu \sin(2\pi\nu t)\right)\right)^2 \mathrm{d}t \\
&= \int_0^1 \left(\sqrt{2}\sum_{\nu=1}^{\infty} (2\pi\nu)^{2m} \begin{cases} (-1)^{\frac{m}{2}} \alpha_\nu \cos(2\pi\nu t) + (-1)^{\frac{m}{2}} \beta_\nu \sin(2\pi\nu t), & \text{if } m \text{ even;} \\ (-1)^{\frac{m+1}{2}} \alpha_\nu \sin(2\pi\nu t) + (-1)^{\frac{m-1}{2}} \beta_\nu \cos(2\pi\nu t), & \text{if } m \text{ odd.} \end{cases}\right)^2 \mathrm{d}t \\
&= \sum_{\nu=1}^{\infty} (2\pi\nu)^{2m} \int_0^1 2 \begin{cases} \alpha_\nu^2 \cos^2(2\pi\nu t) + \beta_\nu^2 \sin^2(2\pi\nu t), & \text{if } m \text{ even;} \\ \alpha_\nu^2 \sin^2(2\pi\nu t) + \beta_\nu^2 \cos^2(2\pi\nu t), & \text{if } m \text{ odd.} \end{cases} \mathrm{d}t + \underbrace{0+0}_{\substack{\text{orthogonal} \\ \text{cross-terms}}} \\
&= \sum_{\nu=1}^{\infty} (2\pi\nu)^{2m} (\alpha_\nu^2 + \beta_\nu^2) < \infty.
\end{aligned}$$

This wiggliness penalty is a definite norm on $\mathcal{H}_1$. Writing u as the uniformly convergent zero-mean Fourier series (46), we see that

$$||u||_{\mathcal{H}_1}^2 = \sum_{\nu=1}^{\infty} (2\pi\nu)^{2m} (\alpha_\nu^2 + \beta_\nu^2) = 0 \implies (\forall \nu,\ \alpha_\nu = \beta_\nu = 0) \implies u \equiv 0.$$

The bilinear form on $\mathcal{H}_1$ that induces the root wiggliness penalty as its norm is

$$\langle f, g \rangle_{\mathcal{H}_1} = \int_0^1 f^{(m)}(x) g^{(m)}(x)\, \mathrm{d}x\,.$$

The reproducing kernel for $\mathcal{H}_1$ can be written [158]

$$k^1(s,t) = \sum_{\nu=1}^{\infty} \frac{2}{(2\pi\nu)^{2m}} \cos(2\pi\nu(s-t)), \tag{47}$$

from which the reproducing property

$$\begin{aligned}
\langle u, k_t^1 \rangle_{\mathcal{H}_1} &= \int_0^1 \left(\sqrt{2}\sum_{\nu=1}^{\infty} (2\pi\nu)^{2m} \begin{cases} (-1)^{\frac{m}{2}} \alpha_\nu \cos(2\pi\nu t) + (-1)^{\frac{m}{2}} \beta_\nu \sin(2\pi\nu t), & \text{if } m \text{ even;} \\ (-1)^{\frac{m+1}{2}} \alpha_\nu \sin(2\pi\nu t) + (-1)^{\frac{m-1}{2}} \beta_\nu \cos(2\pi\nu t), & \text{if } m \text{ odd.} \end{cases}\right) \cdot \\
&\qquad \left(\sum_{\eta=1}^{\infty} \frac{2}{(2\pi\eta)^{2m}} \begin{cases} (-1)^{\frac{m}{2}} \cos(2\pi\eta(s-t)), & \text{if } m \text{ even;} \\ (-1)^{\frac{m+1}{2}} \sin(2\pi\eta(s-t)), & \text{if } m \text{ odd.} \end{cases}\right) \mathrm{d}s \\
&= \sum_{\nu=1}^{\infty} \sqrt{2} \begin{cases} (-1)^m \alpha_\nu \cos(2\pi\nu t) + (-1)^m \beta_\nu \sin(2\pi\nu t), & \text{if } m \text{ even;} \\ (-1)^{m+1} \alpha_\nu \cos(2\pi\nu t) - (-1)^m \beta_\nu \sin(2\pi\nu t), & \text{if } m \text{ odd.} \end{cases} \\
&= \sqrt{2}\sum_{\nu=1}^{\infty} \alpha_\nu \cos(2\pi\nu t) + \beta_\nu \sin(2\pi\nu t),
\end{aligned}$$

[36] Since u and u' are absolutely continuous on the interval; see [149], Section 11, and [84], Theorem 33.7.

can be verified by the pointwise-convergent zero-mean Fourier series (46). The simplification on the second line follows from a simple application of the product formula; indeed, for $(\eta, \nu) \in \mathbb{N}^2_{\geq 1}$,

$$\int_0^1 2\cos(2\pi\eta(s-t))\cos(2\pi\nu s)\,\mathrm{d}s = \begin{cases}\cos(2\pi\nu t), & \text{if } \eta = \nu\\ 0, & \text{otherwise.}\end{cases}$$
$$\int_0^1 2\cos(2\pi\eta(s-t))\sin(2\pi\nu s)\,\mathrm{d}s = \begin{cases}\sin(2\pi\nu t), & \text{if } \eta = \nu\\ 0, & \text{otherwise.}\end{cases}$$
$$\int_0^1 2\sin(2\pi\eta(s-t))\cos(2\pi\nu s)\,\mathrm{d}s = \begin{cases}-\sin(2\pi\nu t), & \text{if } \eta = \nu\\ 0, & \text{otherwise.}\end{cases}$$
$$\int_0^1 2\sin(2\pi\eta(s-t))\sin(2\pi\nu s)\,\mathrm{d}s = \begin{cases}\cos(2\pi\nu t), & \text{if } \eta = \nu\\ 0, & \text{otherwise.}\end{cases}$$

A closed-form expression for this kernel (47) was given in terms of the Bernoulli polynomials in [17, 32]

$$k^1(s,t) = \frac{(-1)^{m-1}}{(2m)!}B_{2m}\left(s - t - \lfloor s - t\rfloor\right),$$

where $\lfloor\cdot\rfloor$ indicates the floor function and the Bernoulli polynomials $B_r(t)$ can be defined on $[0,1]$ recursively [158]

$$B_r(t)\begin{cases}= 1, & r = 0;\\ \text{solves } \frac{1}{r}\frac{\mathrm{d}}{\mathrm{d}t}B_r(t) = B_{r-1}(t) \text{ with periodic boundary conditions}, & \text{otherwise}\end{cases}$$

and defined explicitly [89] as

$$B_r(t) = \sum_{n=0}^{r}\frac{1}{n+1}\sum_{k=0}^{n}(-1)^k\binom{n}{k}(t+k)^r.$$

The reproducing kernel for $\mathcal{H}$: Define the orthogonal projection operators

$$\mathrm{P}_0 : \mathcal{H} \to \mathcal{H}_0 \qquad \mathrm{P}_1 : \mathcal{H} \to \mathcal{H}_1$$
$$u \mapsto \int_0^1 u(t)\,\mathrm{d}t\ , \qquad u \mapsto u - \int_0^1 u(t)\,\mathrm{d}t\,.$$

With respect to the inner product

$$\langle f, g\rangle_{\mathcal{H}} = \langle \mathrm{P}_0 f, \mathrm{P}_0 g\rangle_{\mathcal{H}_0} + \langle \mathrm{P}_1 f, \mathrm{P}_1 g\rangle_{\mathcal{H}_1},$$

the subspaces $\mathcal{H}_0$ and $\mathcal{H}_1$ are evidently orthogonal for $m \geq 1$, since the mth derivative of any function in $\mathcal{H}_0$ is 0, and since the area $\int_0^1 u_1(x)\,\mathrm{d}x$ of any function $u_1 \in \mathcal{H}_1$ is 0.

Our model space $\mathcal{H}$ is the direct sum of the two perpendicular spaces $\mathcal{H}_0$ of constant (and therefore periodic) functions and $\mathcal{H}_1$ of zero-mean periodic functions. Since the reproducing kernel of the direct sum of two perpendicular subspaces $\mathcal{H}_0$ and $\mathcal{H}_1$ is the sum of the kernels [3], we have that our reproducing kernel for $\mathcal{H}$

$$\begin{aligned}k(s,t) = k^0(s,t) + k^1(s,t) &= 1 + \sum_{\nu=1}^{\infty}\frac{2}{(2\pi\nu)^{2m}}\cos(2\pi\nu(s-t))\\ &= 1 + \frac{(-1)^{m-1}}{(2m)!}B_{2m}(s - t - \lfloor s - t\rfloor).\end{aligned}$$

Algorithm 5: An algorithm that fits periodic splines on $\mathcal{X} = [0,1]$ based on the seminorm penalty $J_{m,\mathcal{X}}(u) = \int_0^1 (u^{(m)}(x))^2 \,\mathrm{d}x$. By the representer theorem (Proposition 2.60), the solution to

$$u^* = \operatorname*{arg\,min}_{u \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^{n} (u(x_i) - y_i)^2 + \lambda \int_0^1 (u^{(m)}(x))^2 \,\mathrm{d}x\,,$$

takes the form

$$u^* = d + \sum_{i=1}^{n} c_i \frac{(-1)^{m-1}}{(2m)!} B_{2m}\left(\cdot - x_j - \lfloor s - t \rfloor\right).$$

This algorithm recovers $c \in \mathbb{R}^n$ and d from samples $\{y_i\}_{i=1}^n$, $y_i \in \mathbb{R}$, taken at scattered values $\{x_i\}_{i=1}^n$, $x_i \in [0,1]$.

Data: A set of n sample locations $\{x_i\}_{i=1}^n$ in $[0,1]$ and n corresponding sample values $y_i \in \mathbb{R}$.
Parameters: A regularization penalty parameter $\lambda \geq 0$ and (implicitly) a choice of model space $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$ and seminorm wiggliness penalty $J_{m,\mathcal{X}}$, whose one-dimensional null space $\mathcal{H}_0 = \operatorname{span}\{1\}$.
Result: A set of basis function weights $c \in \mathbb{R}^n$ and $d \in \mathbb{R}$ specifying the empirical risk minimizing function u^*.

Compute the $n \times n$ Gram matrix $\mathbf{K}_1$ in whose ith row and jth column reposes the value

$$(\mathbf{K})_{i,j} \leftarrow k^1(x_i, x_j) = \frac{(-1)^{m-1}}{(2m)!} B_{2m}\left(x_i - x_j - \lfloor s - t \rfloor\right),$$

where

$$B_r(t) = \sum_{n=0}^{r} \frac{1}{n+1} \sum_{k=0}^{n} (-1)^k \binom{n}{k} (t+k)^r.$$

Augment the Gram matrix of k^1 on our data set with null-space basis function matrix $\mathbf{T} = \mathbf{1}_n$ to form an $(n+m) \times (n+m)$ matrix $\mathbf{K}$ and set y accordingly

$$\mathbf{K} \leftarrow \begin{pmatrix} \mathbf{K}_1 + \lambda n \mathbf{I}_{n \times n} & \mathbf{1}_n \\ \mathbf{1}_n^T & 0 \end{pmatrix} \text{ and } y \leftarrow \begin{pmatrix} y \\ 0 \end{pmatrix};$$

Solve $\mathbf{K}\alpha = y$,

$$\alpha \leftarrow \mathbf{K}^{-1} y \text{ (or, if } \mathbf{K}_1 \text{ is not strictly positive-definite and } \lambda = 0,\ \alpha \leftarrow \mathbf{K}^{\dagger} y);$$

Return the spline weights $c \leftarrow \alpha[1:n]$ and $d \leftarrow \alpha[n+1]$;

Solving the Spline Smoothing Problem for Splines on the Circle. We adapt Algorithm 3 to fit periodic splines on $[0, 1]$ using the seminorm wiggliness penalty $J_{m,\mathcal{X}}(u) = \int_0^1 (u^{(m)}(x))^2 \, \mathrm{d}x$ in Algorithm 5.

Figure 1 compares two kernels associated with the penalty (2), which differ only in their boundary conditions (the natural cubic spline conditions, or the periodic cubic spline conditions), along with their corresponding smoothing spline solutions to two related data sets. This result can be applied to

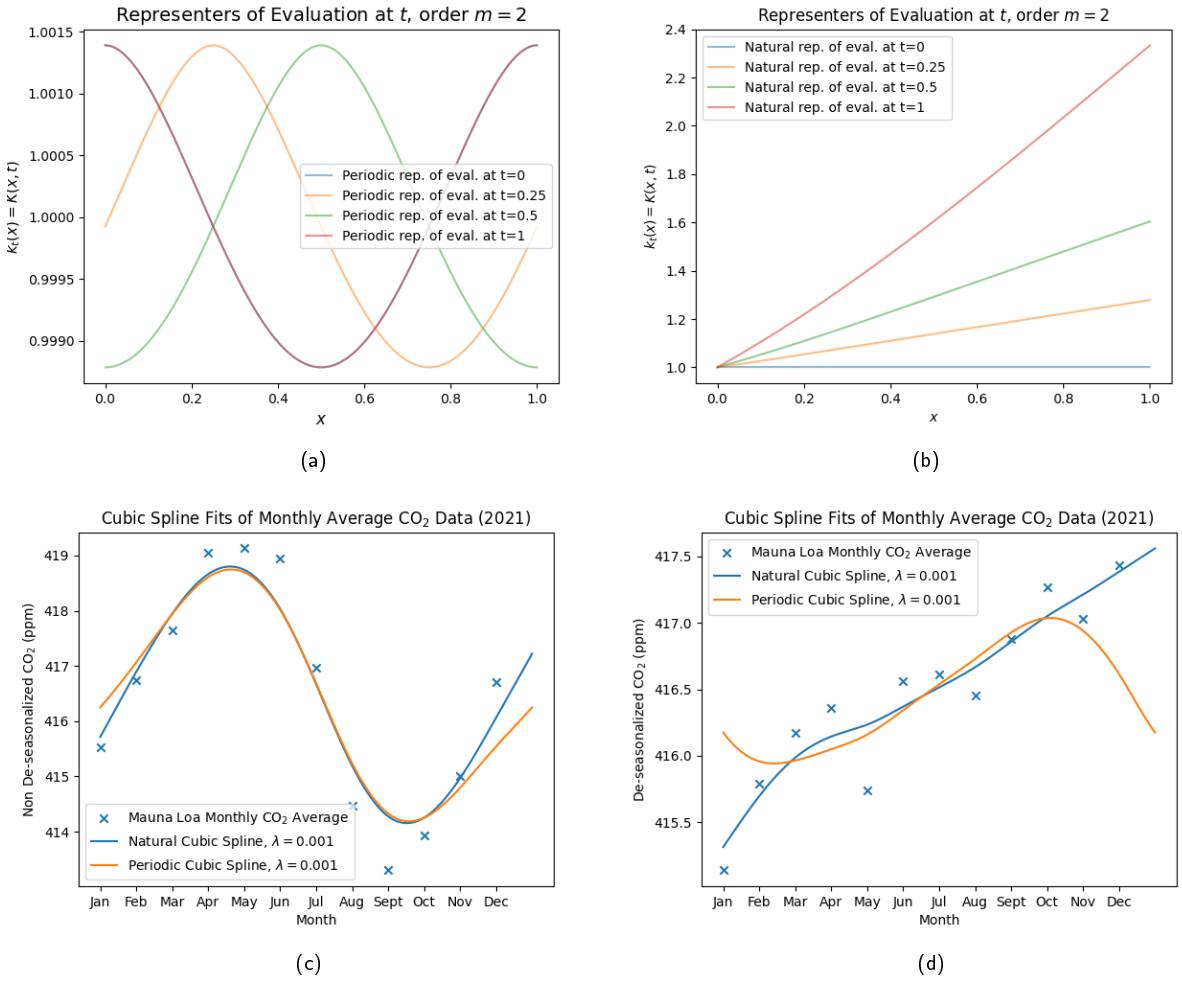


Figure 1: **(1a)** plots the representers of evaluation $k_t = k_t^0 + k_t^1 = k(\cdot, t)$ of the spline on the circle (cubic spline with periodic boundary conditions) at $t = 0, 0.25, 0.5$, and 1. In this case $k(\cdot, 0)$ and $k(\cdot, 1)$ coincide. **(1b)** does the same, but with the kernel for the natural cubic splines. While the kernels are quite different (the former is periodic, treating points with small along-the-circle distance similarly; the latter is cubic up till t, then linear), their Gram matrices' spectra are alike with few, regularly spaced data points. **(1c)-(1d)** compare the corresponding spline curves with $\lambda = 0.001$ on the twelve monthly average CO_2 measurements taken in 2021 at the Mauna Loa station of the Global Monitoring Laboratory network [30], both the untreated data (left), with a naturally occurring periodic component, and the de-seasonalized data (right), reflecting increasing atmospheric CO_2 concentrations. To avoid delving into the complexities of the signal processing chain that produces these monthly averages from discrete measurements, we take these monthly averages to be samples at times $t_i = (i-1)\frac{1}{12}$ for $i \in \{1, \ldots, 12\}$. Of course, monthly averages are bounded linear functionals, so the representer theorem (Section 2.5) can be applied to more complicated monthly averaging operators L_i than taking a single sample at each t_i.

the construction of piecewise polynomial splines at knots [32]; on equally spaced data, the minimizer of the spline empirical risk over $\mathcal{H}$ resembles the Butterworth filter [32, 158].

2.6.3 Thin-plate Splines on Graphs

Kernel methods are most often used to compare graphs, using computed features from each graph. For example, the subgraph matching kernel expresses similarity between any pair of graphs (each representing, say, a molecule) in a space of graphs by counting the subgraphs they have in common. (Computing these features is not always easy!)

In our case, however, we define kernels using the geometric structure of a single graph. Such kernels can be used to interpolate scattered observations within a graph, or any finite index set with correlated data due, for instance, to geometric proximity. The finite set can also be a graph approximation of, say, a compact manifold.

In this section, the index set $\mathcal{X}$ is finite and can be taken to be $\{1, \dots, n\}$ via a labeling isomorphism. Without imposing additional structure on $\mathcal{X}$, spline smoothing and interpolation are rather dull.

Example 2.63 ($\mathbb{R}^{\{1,\dots,n\}}$)**.** *The space $\mathbb{R}^n$ with the standard Euclidean inner product $\langle x, y\rangle_{\mathbb{R}^n} = x^T y$ is an RKHS with the standard basis functions e_i as representers of evaluation. A function $f : \{1, \dots, n\} \to \mathbb{R}$ can be represented as a vector and evaluated pointwise: $f(i) = e_i^T f$. The corresponding kernel $k(i, j) = \delta_{i,j}$, where δ is the Kronecker delta. The similarity metric from Remark 2.17 is the discrete metric*

$$d_{\mathcal{X}}(i, j) = \begin{cases} \sqrt{2}, & \text{if } i \neq j \\ 0, & \text{otherwise.} \end{cases}$$

In effect, vertex i tells us nothing about vertex j unless $i = j$. Given observations $\{y_{x_i}\}_{i=1}^m$ at $\{x_i\}_{i=1}^m$ with $x_i \in \mathcal{X} = \{1, \dots, n\}$, the minimum weight solution f^* to the associated exact-interpolation problem is easily seen to be

$$f^*(x) = \begin{cases} y_j, & \text{if } x = j \text{ for some } j \text{ already observed: } j \in \{x_i\}_{i=1}^m; \\ 0, & \text{otherwise.} \end{cases}$$

The solution to the spline smoothing problem

$$f^* = \operatorname*{arg\,min}_{f \in \mathbb{R}^n} \frac{1}{m} \sum_{i=1}^m (f(x_i) - y(x_i))^2 + \lambda ||f||^2_{\mathbb{R}^n},$$

can be found using the representer theorem (Proposition 2.60): $f^* = \sum_{i=1}^m \alpha_i k_{x_i}$. The function f^* is specified by its weights $\alpha \in \mathbb{R}^m$ on the representers of evaluation. The linear algebra problem

$$\alpha^* = \operatorname*{arg\,min}_{\alpha \in \mathbb{R}^m} (\mathbf{K}\alpha - y)^2 + \lambda \alpha^T \mathbf{K} \alpha,$$

is solved by setting the gradient with respect to α to 0: $\alpha^* = \frac{1}{1+m\lambda} y$. Thus,

$$f^* = \sum_{i=1}^m \alpha_i k_{x_i} = \sum_{i=1}^m \frac{y_i}{1 + m\lambda} k_{x_i}.$$

We can evaluate the interpolant using the reproducing property

$$f^*(j) = \left\langle \sum_{i=1}^m \alpha_i k_{x_i}, \mathbb{1}_j \right\rangle_{\mathcal{H}} = \sum_{i=1}^m \frac{y_i}{1 + \lambda} \delta_{j,x_i} = \begin{cases} \frac{y_j}{1+m\lambda}, & \text{if } j \text{ already observed;} \\ 0, & \text{otherwise.} \end{cases}$$

Graphs are a convenient mechanism of expressing the structure on $\mathcal{X}$ that is needed for more interesting notions of similarity between sample points in $\mathcal{X}$ and of "smoothness" in interpolation and smoothing problems. We define a Sobolev-like seminorm using the Laplace matrix of the graph. Not all RKHSs on $\{1, \dots, n\}$ can be represented in this way. As will become apparent soon, there is no finite simple graph associated with the above example; nevertheless, it can be be seen as a thin-plate spline of order $m = 0$ associated with any graph on $\{1, \dots, n\}$ with the usual convention that $\mathbf{L}^0 = \mathbf{I}$ for any matrix $\mathbf{L}$.

Let us first recall some facts about graphs. Let $G = (V, E)$ be an undirected graph with vertices $V = \{1, \dots, n\}$ and edges $E \subset \{(i, j) \mid i \in V, j \in V, i < j\}$. The adjacency matrix $\mathbf{A}$ of G is the $n \times n$ matrix that reposes in its ith row and jth column the value

$$(\mathbf{A})_{i,j} = \begin{cases} 1, & \text{if } (i,j) \in E \text{ or } (j,i) \in E \\ 0, & \text{otherwise.} \end{cases}$$

The Laplacian $\mathbf{L}$ of G is the $n \times n$ matrix defined as $\mathbf{L} = \mathbf{D} - \mathbf{A}$, where $\mathbf{D}$ is the diagonal matrix whose ith diagonal element contains the degree of node i. Thus its rows sum to zero $\mathbf{L}\mathbf{1}_n = 0$, and $\mathbf{L}$ always has 0 as an eigenvalue. Since $\mathbf{D}$ and $\mathbf{A}$ are symmetric, so too is $\mathbf{L}$. $\mathbf{L}$ encodes the topology of the graph via local information (the edges in the adjacency matrix) and can be used to define what "smoothness" means on a graph: not too much variation across the edges. Note that $\mathbf{A}, \mathbf{D}$, and $\mathbf{L}$ can be defined for graphs with nonnegative weighted edges in the obvious way: $\mathbf{A}$ stores the edge weights (or 0) between every pair of nodes and the degree of a node is the sum of the weights of the edges it participates in.

An analogy with resistor networks is instructive: if $f(x)$ encodes the potential at node x, and each edge represents wire with unit resistance (or, in a weighted graph, if the weight of edge e is $w(e)$, the resistance along the edge is $1/w(e)$), Δf simply gives a current balance at each node; the harmonic equation $(\Delta f)(x) = 0$ is satisfied at nodes x with no source or sink of current. By the discrete version of Liouville's theorem, the only harmonic functions on an entire finite graph are constant over the connected components. Our interpolation task with $m = 1$ consists in fixing the potential at certain nodes and determining potentials at the remaining nodes so that their current is balanced.

Lemma 2.64. *For $m = 1$, the discrete analogue of the Sobolev seminorm integration by parts*

$$\int_{\mathcal{X}} (f'(x))^2 \, \mathrm{d}x = -\int_{\mathcal{X}} f(x)(\Delta f)(x) \, \mathrm{d}x \quad \textit{(with appropriate boundary conditions)},$$

holds on graphs G (defined so as to prohibit self-loops).

Proof.

$$\begin{aligned} -\sum_{i=1}^{n} f(i)(\Delta f)(i) = f^T(\mathbf{L}f) &= f^T\mathbf{D}f - f^T\mathbf{A}f \\ &= \sum_{i=1}^{n} \mathbf{D}_{ii} f(i)^2 - \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{A}_{ij} f(i) f(j) \\ &= \sum_{(i,j)\in E} f(i)^2 + f(j)^2 - 2\sum_{(i,j)\in E} f(i)f(j) = \sum_{(i,j)\in E} (f(i) - f(j))^2. \end{aligned}$$

We made the identification $\Delta = -\mathbf{L}$. □

In terms of our physics analogy, this simply states that $f^T\mathbf{L}f$ gives the energy dissipated in the circuit, that is, the sum of the square potential drops over the resistors. For $m > 1$, significant

additional boundary conditions must be placed on a general graph for the equivalent relation to hold, though for certain graphs, such as chains, the result $\int_{\mathcal{X}} f(x)(\Delta^m f)(x)\,\mathrm{d}x = \int_{\mathcal{X}} f(x)(\Delta^{2m} f)(x)\,\mathrm{d}x$ can be enforced with constraints analogous to the splines on the circle.

Corollary 2.65. $\mathbf{L}$ *is a positive-semidefinite matrix.*

Proof. By the spectral theorem, any eigenvalue λ of the real, symmetric matrix $\mathbf{L}$ must be real; moreover, it must be nonnegative, since $\mathbf{L}u = \lambda u$ implies that

$$\lambda||u||_2^2 = \lambda u^T u = u^T \mathbf{L} u = \sum_{(i,j)\in E} (u_i - u_j)^2 \geq 0.$$

Since $||u||_2^2$ must be nonnegative, $\lambda \geq 0$; the matrix $\mathbf{L}$ is therefore positive-semidefinite. Indeed, $\mathbf{L} = \mathbf{E}\mathbf{E}^T$, where the $n \times |E|$ edge-incidence matrix $\mathbf{E}$ is defined as follows:

$$\mathbf{E}_{ij} = \begin{cases} 1, & \text{if } e_j = (i, \cdot) \\ -1, & \text{if } e_j = (\cdot, i) \\ 0, & \text{if edge } e_j \text{ does not involve node } i. \end{cases} \tag{48}$$

□

Remark 2.66. *In discrete differential geometry [31], Lemma 2.64 follows from applying Green's first identity (integration by parts with the gradient product rule) on the entire graph,*

$$\int_V \langle \nabla u, \nabla v \rangle_{\mathbb{R}^2}\,\mathrm{d}V + \int_V u\Delta v\,\mathrm{d}V = \int_{\partial V} u\nabla v \cdot \mathrm{d}\overrightarrow{S} = 0,$$

after setting $u = v, \Delta = -\mathbf{L}$, *and* $\nabla = \mathbf{E}^T$ *maps graph functions (functions defined on vertices) to 1-forms (functions defined on edges).*

$\mathbf{L}$ is never strictly positive-definite, as $\mathbf{1}_n$ is always an eigenvector with eigenvalue 0 (since a vertex's degree equals the number of edges it participates in), and any function that is constant over each of the graph's connected components is similarly an eigenvector with eigenvalue 0. Conversely, if e is an eigenvector of $\mathbf{L}$ with eigenvalue 0, we see that

$$||e||^2 = e^t \underbrace{\mathbf{L}e}_{0e} = 0 = \sum_{(i,j)\in E} (e_i - e_j)^2,$$

so that no edge can join nodes for which their values differ; e_i must equal e_j for every edge $(i,j) \in E$. Thus, the null space of $\mathbf{L}$ is precisely the space of functions that are constant on each connected component of G. If the 0 eigenvalue of $\mathbf{L}$ has multiplicity r, then G has r connected components [173]. The eigenvectors of $\mathbf{L}$ can be used as a Fourier basis useful in smooth approximation of arbitrary functions. On unweighted chains, these eigenvectors are precisely the discrete cosine transform basis functions, the specific type depending on the boundary conditions [143].

We are now ready to introduce the model space $\mathcal{H}$ in which we will find our splines.

Model space $\mathcal{H}$: The space of functions on G. By the obvious isomorphism, we represent them as vectors in $\mathbb{R}^n$ and consider $\mathcal{H} = \mathbb{R}^n$. The seminorm corresponding to our wiggliness penalty is

$$J_{m,\mathcal{H}}(f) = f^T \mathbf{L}^m f.$$

Note that, since $\mathbf{L}$ is normal, the eigenvectors of $\mathbf{L}^m$ are the same as those of $\mathbf{L}$; the eigenvalues are simply modified by their being taken to the mth power.

If G has r connected components, we can endow $\mathcal{H}$ with the inner product

$$\langle f, g\rangle_{\mathcal{H}} = \sum_{c=1}^{r} \mathrm{mean}_c\,(f) \cdot \mathrm{mean}_c\,(g) + f^T \mathbf{L}^m g, \tag{49}$$

where $\mathrm{mean}_c\,(f)$ is the mean value of f on the connected component c.

Null space $\mathcal{H}_0$: The null space of our penalty seminorm is the space of functions that are constant on each connected component. $\mathcal{H}_0$ can be made into an RKHS with inner product

$$\langle f, g\rangle_{\mathcal{H}_0} = \sum_{c=1}^{r} \mathrm{mean}_c\,(f) \cdot \mathrm{mean}_c\,(g).$$

The representer of evaluation k_v at a particular vertex v is the function $1_{c(v)}\mathcal{H}_0$ that is 1 on all nodes on v's connected component $c(v)$ and 0 otherwise, and so inhabits the null space of $J_{m,\mathcal{H}}$. Then if $f \in \mathcal{H}_0$,

$$\langle f, k_v\rangle_{\mathcal{H}_0} = \mathrm{mean}_{c(v)}\,(f) \cdot 1 = f(v),$$

since $f(v) = \mathrm{mean}_{c(v)}\,(f)$. The r distinct representers of evaluation $\{1_1, \ldots, 1_r\}$ form an orthonormal basis of $\mathcal{H}_0$, and the orthogonal projection operator $\mathrm{P}_0 : \mathcal{H} \to \mathcal{H}_0$ assigns to each vertex v the mean of f on $c(v)$, since the functions $1_c \in \mathbf{L}^m$, and so

$$\mathrm{P}_0 f = \sum_{c=1}^{r} \langle f, 1_c\rangle_{\mathcal{H}} 1_c = \sum_{c=1}^{r} \mathrm{mean}_c\,(f) 1_c.$$

In terms of the graph Laplacian, $\mathcal{H}_0 = \mathrm{null}\,\mathbf{L}^m = \mathrm{null}\,\mathbf{L}$ (with the latter equality holding because $\mathbf{L}$ is normal–diagonalize it in unitary eigenvectors). Since $\mathcal{H}_0$ is finite-dimensional, its kernel is fully specified by its rank r, $n \times n$ Gram matrix $\mathbf{K}_0$, in whose ith row and jth column rests the value $\mathbb{1}_{c(i)=c(j)}$. If G has one connected component, $\mathbf{K}_0$ is the ones matrix; more generally, $\mathbf{K}_0$ is permutation-similar to the block-diagonal matrix whose cth diagonal block is the ones matrix of size $|c| \times |c|$, where $|c|$ is the number of nodes in the cth connected component.

Space of wiggly functions $\mathcal{H}_1$: We define $\mathcal{H}_1$ to be the space of signals that have zero mean on each connected component. By construction, for all $f \in \mathcal{H}_1$, we have $||f||^2_{\mathcal{H}_0} = \sum_{c=1}^{r} \mathrm{mean}_c\,(f)^2 = 0$ and for all f and g in $\mathcal{H}_1$, $\langle f, g\rangle_{\mathcal{H}_0} = 0$. Moreover, the bilinear form

$$\langle f, g\rangle_{\mathcal{H}_1} = f^T \mathbf{L}^m g,$$

is definite on $\mathcal{H}_1$ (since the DC functions over each connected component $\{1_1, \ldots, 1_r\}$ form a basis of $\mathrm{null}\,\mathbf{L}^m$, and functions in $\mathcal{H}_1$ are zero-mean), and coincides with $\langle f, g\rangle_{\mathcal{H}}$ on its restriction to $\mathcal{H}_1$.

For any matrix $\mathbf{A}$, the operator $\mathbf{A}\mathbf{A}^\dagger$ is the orthogonal projector onto $\mathrm{range}\,\mathbf{A}$ (and $\mathbf{A}^\dagger\mathbf{A}$ the projector onto $\mathrm{range}\,\mathbf{A}^*$). Moreover, if $\mathbf{A}$ is normal, $\mathbf{A}\mathbf{A}^\dagger = \mathbf{A}^\dagger\mathbf{A} = \mathbf{A}^m(\mathbf{A}^\dagger)^m = (\mathbf{A}^\dagger)^m\mathbf{A}^m$ (diagonalize it in unitary eigenvectors). For any graph signal $g \in \mathcal{H}_1$, we have that $g \perp \{1_1, \ldots, 1_r\}$, with respect to both $\langle\cdot,\cdot\rangle_{\mathcal{H}_0}$ and $\langle\cdot,\cdot\rangle_{\mathbb{R}^n}$, and so $g \in (\mathrm{null}\,\mathbf{L})^\perp = \mathrm{range}\,\mathbf{L}$. The orthogonal projection of g onto $\mathrm{range}\,\mathbf{L}$ then recovers g: $\mathbf{L}\mathbf{L}^\dagger g = g$. Thus, since $\mathbf{L}$ is a normal matrix (it is real and symmetric), $(\mathbf{L}^\dagger)^m\mathbf{L}^m = \mathbf{L}\mathbf{L}^\dagger$ and

$$\langle (\mathbf{L}^\dagger)^m e_v, g\rangle_{\mathcal{H}_1} = e_v^T(\mathbf{L}^\dagger)^m\mathbf{L}^m g = e_v^T \underbrace{\mathbf{L}\mathbf{L}^\dagger g}_{g} = e_v^T g = g(v), \tag{50}$$

where e_v is the vth standard basis vector of $\mathbb{R}^n$. Equation (50) shows that the vth column of $(\mathbf{L}^\dagger)^m$ (and row, since $\mathbf{L}$–and therefore $(\mathbf{L}^\dagger)^m$–is symmetric) is the representer of evaluation $k_v = e_v^T(\mathbf{L}^\dagger)^m$ at vertex v. Since $\mathbf{L}$ is real and symmetric, with eigenvalues $\lambda_1 = \ldots = \lambda_r = 0$, we have that

$$\mathbf{L} = \sum_{i=1}^{n} \lambda_i u_i u_i^T = \sum_{i=r+1}^{n} \lambda_i u_i u_i^T, \text{ and so } (\mathbf{L}^\dagger)^m = \sum_{i=r+1}^{n} \lambda_i^{-m} u_i u_i^T. \tag{51}$$

Then the vth column of $(\mathbf{L}^\dagger)^m$

$$k_v = (\mathbf{L}^\dagger)^m[:, v] = \sum_{i=r+1}^{n} \frac{u_i[v]}{\lambda_i^m} u_i$$

is a weighted sum of $n - r$ linearly independent eigenvectors u_i of $\mathbf{L}$ with eigenvalue $\lambda_i > 0$. Since the $\{u_i\}_{i=1}^n$ form an orthonormal system for $\mathbb{R}^n$ and $u_i = 1_i$ for $i = 1, \ldots, r$, we have that, for $i = r+1, \ldots, n$, $u_i \perp \{1_1, \ldots, 1_r\}$ with respect to $\langle \cdot, \cdot, \rangle_{\mathbb{R}^n}$ and hence are zero-mean on each connected component. For $v = 1, \ldots, n$, then, $k_v \in \mathcal{H}_1$. Taking the inner products $\langle \cdot, \cdot \rangle_{\mathcal{H}_1}$ between the representers of evaluation, we see that the kernel over the finite-dimensional $\mathcal{H}_1$ is fully specified by its Gram matrix $\mathbf{K}_1 = (\mathbf{L}^\dagger)^m$.

Remark 2.67. *The above analysis holds with Laplacians of weighted graphs or normalized graph Laplacians. Moreover, the eigenvectors of the Laplacian (and hence the eigenvectors of the Laplacian pseudoinverse kernel) are used to define many other kernels, with different eigenvalues. For instance, the diffusion kernel, which, like the Gaussian kernel, is a solution to the diffusion equation, has the same eigenvectors [83].*

Solving the Spline Smoothing Problem on Graphs. We can now adapt Algorithm 3 to solve the spline smoothing problem on a graph with n vertices, of which n_{obs} are observed. In the pseudocode, we set the $n_{obs} \times r$ matrix $(\mathbf{T})_{\texttt{obs},c} = 1_c[\texttt{obs}]$, where `obs` is a list of indices of the n_{obs} observed nodes, and $\mathbf{K}_1$ selects the n_{obs} observed rows and columns of our $n \times n$ matrix $(\mathbf{L}^\dagger)^m$. We can also replace the evaluation functionals $x_i \mapsto u(x_i)$ with arbitrary bounded linear functionals, replacing $\mathbf{K}_1$ with the corresponding matrix $\mathbf{\Sigma}$ of the inner products of the representers of the functionals, as in (35).

In Figures 2-5, we demonstrate the use of graphs to perform spline smoothing (Equation (27)) in $\mathbb{R}^{12}$, with graphs specifying similarity of vector elements. We assign the 12 months in which we considered Mauna Loa CO_2 readings from Figures 1c-1d to nodes in the graph G or G_c (Figures 2 and 3, respectively), display their representers of evaluation with $m = 3$ (Figure 4), and show solutions of the spline smoothing problem on these 12 data points (Figure 5).

Figure 2: Weighted chain graph G associated with the sample points $t_1 < t_2 < \ldots < t_{12}$. Given our (approximately–some months are abridged!) regularly spaced months in time $\mathcal{X} = \{1, \ldots, 12\}$, we may want a similarity metric on $\mathcal{X}$ induced by the graph in which we give each edge equal weight: for instance, we can set $t_i = i - 1$ and give edges weight 1. If human activity matters more than seasonality, we may not wish to link January (1) with December (12).

Thin-plate splines on graphs are summarized in Algorithm 6.

Remark 2.68 (Graph approximations of compact Riemannian manifolds)**.** *We can use graphs to derive splines over a compact manifold. We approximate the manifold as a finite point cloud–an ϵ-net, for instance, or randomly scattered points on the manifold–and add edges between nearby points using one of many techniques from computational geometry: k-nearest neighbors, Delaunay triangulations, Gabriel graphs, etc. With the compact manifold embedded in a Euclidean space $\mathbb{R}^d$, we can set edge weights according to the Euclidean metric of the space $d_{\mathbb{R}^d}(x, y) = ||x - y||$, so that the vertices associated with two nearby points x, y in our point cloud are bridged by an edge whose weight is inversely related to $||x - y||^2$, which is a good approximation (an underestimate good to order 4) of the squared geodesic distance on our compact manifold. In a so-called Gaussian-weighted graph, the edge weight is as follows: $w(x, y; \sigma) = e^{-\frac{1}{4\sigma}||x-y||^2}$. This choice guarantees that the graph Laplacian of a random point cloud converges to the Laplace-Beltrami operator pointwise and in spectrum [13].*

Algorithm 6: Our index set $\mathcal{X}$ consists of the n vertices (in r connected components) of a graph G, whose $n \times n$ (weighted or unweighted) Laplacian matrix is $\mathbf{L}$. Define the seminorm wiggliness penalty of a graph signal $g \in \mathbb{R}^n$ as $J_{m,\mathcal{X}}(g) = ||\mathrm{P}_1 g||^2_{\mathcal{H}_1} = g^T \mathbf{L}^m g$, where $(\mathrm{P}_1 g)(x)$ is $g(x) - \mathrm{mean}_{c(x)}(g)$–that is, it subtracts from each vertex x the mean over x's connected component of g. Spline smoothing in $\mathcal{H} = \mathbb{R}^n$, given observations $\{y_i\}_{i=1}^{n_{obs}}$ at vertices $\{x_i\}_{i=1}^{n_{obs}}$, is the following empirical risk minimization problem

$$u^* = \underset{u \in \mathbb{R}^n}{\arg\min} \frac{1}{n_{obs}} \sum_{i=1}^{n_{obs}} (u(x_i) - y_i)^2 + \lambda u^T \mathbf{L}^m u.$$

The representer theorem (Proposition 2.60) allows us to write its solution as

$$u^* = \sum_{j=1}^{r} d_j 1_j + \sum_{i=1}^{n_{obs}} c_i (\mathbf{L}^\dagger)^m[:, x_i].$$

where $\phi_j(i) = 1_j = \mathbb{1}_{i \text{ in the } j\text{th connected component}}$ and $k^1(:, x_i)$ is the x_ith column of $(\mathbf{L}^\dagger)^m$. This algorithm finds the vectors of weights c and d. Be warned: for notational simplicity, in this pseudocode, we use 1-indexing.

Data: A set of n_{obs} sample locations $\{x_i\}_{i=1}^{n_{obs}}$ in $\mathcal{X} = \{1, \ldots, n\}$ and n_{obs} corresponding sample values $y_i \in \mathbb{R}$. The graph G whose Laplacian $\mathbf{L}$ defines our seminorm penalty has r connected components.

Parameters: A regularization penalty $\lambda \geq 0$ and (implicitly) a choice of model space $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$ with reproducing kernel $k = k^0 + k^1$ and seminorm wiggliness penalty $||\mathrm{P}_1 \cdot ||^2_{\mathcal{H}_1}$, whose finite-dimensional null space $\mathcal{H}_0$ has basis $\{\phi_1, \ldots, \phi_m\}$. Here $\mathcal{H}_0 = \mathrm{span}\{1_1, \ldots, 1_r\}$ and $\mathcal{H}_1 = \mathrm{span}\{(\mathbf{L}^\dagger)^m[:, 1], \ldots, (\mathbf{L}^\dagger)^m[:, n]\}$.

Result: A set of basis function weights $c \in \mathbb{R}^{n_{obs}}$ and $d \in \mathbb{R}^r$ specifying the empirical risk minimizing function u^*.

Get the indices of the observed vertices $\texttt{obs} \leftarrow [x_1, \ldots, x_{n_{obs}}]$;

Compute the $n_{obs} \times n_{obs}$ Gram matrix $\mathbf{K}_1$

$$\mathbf{K}_1 \leftarrow (\mathbf{L}^\dagger)^m[\texttt{obs}, \texttt{obs}];$$

Compute the $n_{obs} \times r$ matrix $\mathbf{T}$, whose i and column is

$$\mathbf{T}[:, i] \leftarrow 1_i[\texttt{obs}];$$

Augment the Gram matrix of k^1 on our data set with null-space basis function matrix $\mathbf{T}$ to form an $(n_{obs} + r) \times (n_{obs} + r)$ matrix $\mathbf{K}$ and set y accordingly

$$\mathbf{K} \leftarrow \begin{pmatrix} \mathbf{K}_1 + \lambda n_{obs} \mathbf{I}_{n_{obs} \times n_{obs}} & \mathbf{T} \\ \mathbf{T}^T & \mathbf{0}_{r \times r} \end{pmatrix} \text{ and } y \leftarrow \begin{pmatrix} y \\ \mathbf{0}_r \end{pmatrix};$$

Solve $\mathbf{K}\alpha = y$,

$$\alpha \leftarrow \mathbf{K}^{-1} y \text{ (or, if } \lambda = 0, \alpha \leftarrow \mathbf{K}^\dagger y);$$

Return the spline weights $c \leftarrow \alpha[1 : n]$ and $d \leftarrow \alpha[n + 1 : n + m]$;

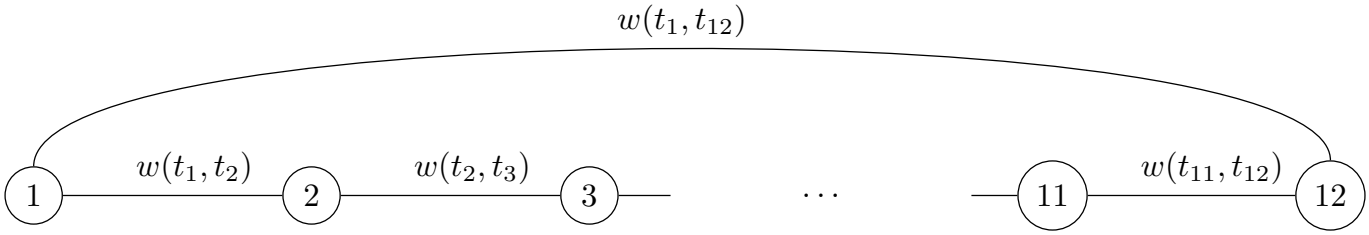
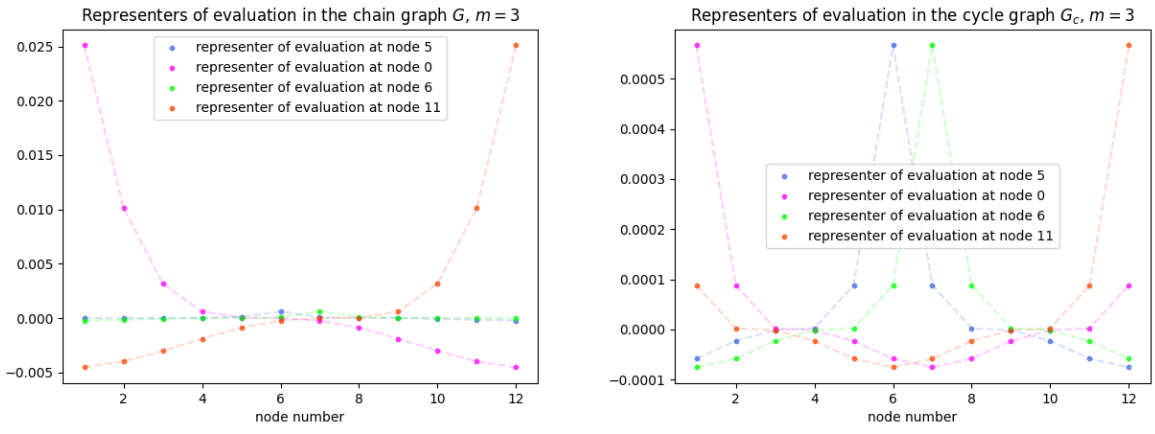


Figure 3: Weighted cycle graph G_c associated with the sample points $t_1 < t_2 < \ldots < t_{12}$. We add the edge with weight w between nodes 1 and 12, modeling similarity between the first and last months. We may choose w to be less than the other weights as a compromise between seasonality effects (such as CO_2 exchange in deciduous forests at the latitude) and the gigatons of CO_2 emitted in the interim by combusting fossil fuels. If we wish the resulting smoother to be "more periodic", we can even set w to be greater than the other weights.



(a) (b)

Figure 4: **(4a)** plots the representers of evaluation in $\mathcal{H}_1$ (with order $m = 3$), i.e. $k_t^1 = k^1(\cdot, t)$, of the kernel k^1 associated with the chain graph G (shown in Figure 2) on $\mathcal{X} = \{1, \ldots 12\}$ at $t = 1, 6, 7$, and 12. **(4b)** gives the same, but for the kernel associated with the cycle graph G_c (shown in Figure 3). In both cases, all edges assume the same weight (12). Dashed lines interpolating between the nodes are included as a visual aid but carry no meaning.

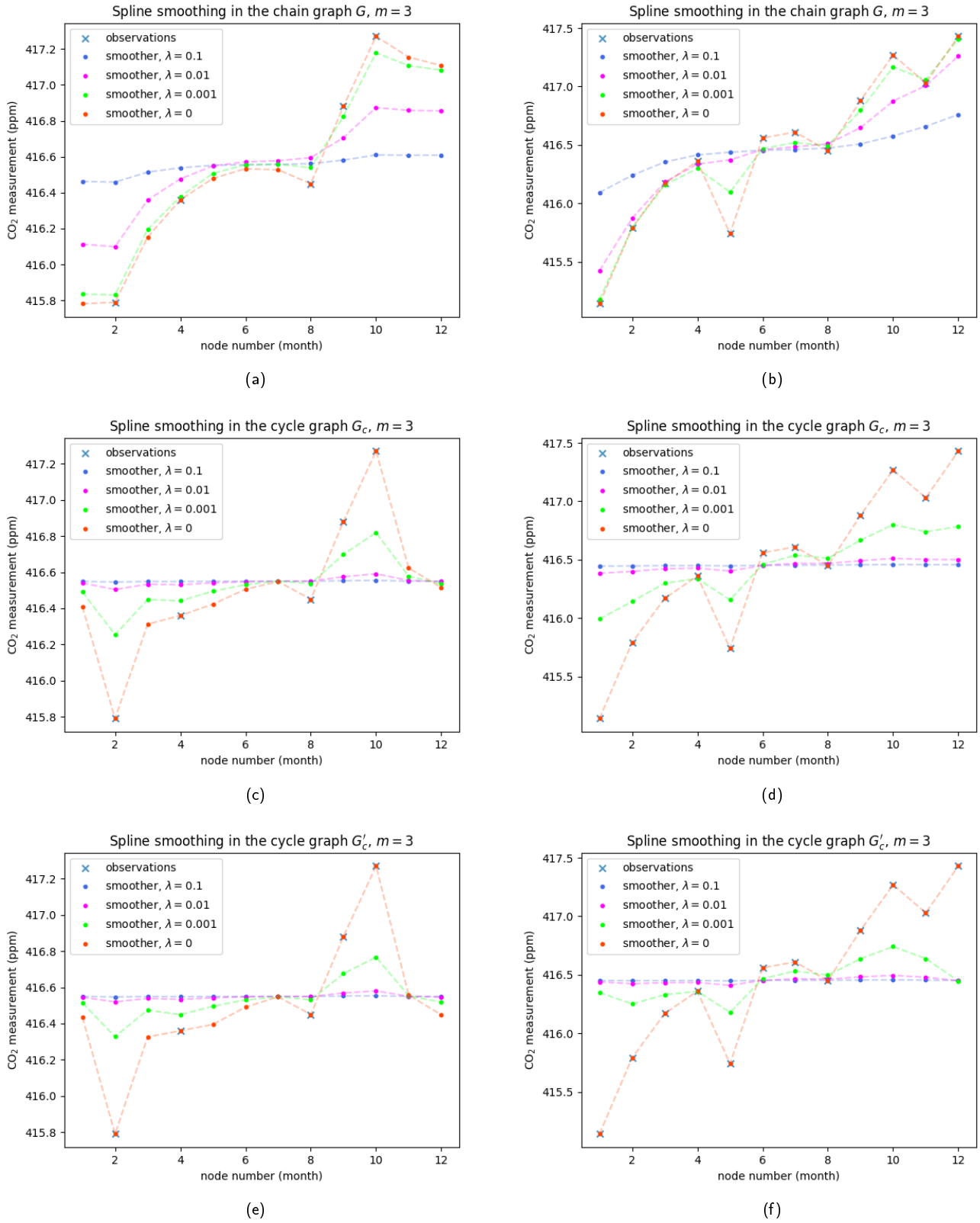


Figure 5: For four distinct choices of wiggliness penalty parameter λ, **(5a)**, **(5c)**, and **(5e)** give the smoothing splines after observing nodes $\{2, 4, 8, 9, 10\}$; **(5b)**, **(5d)**, and **(5f)**, after observing all nodes 1-12. The top row corresponds to the chain graph G (Figure 2, all edge weights 12); the middle, G_c (Figure 3, edge weights 12); and the bottom, G'_c, which is G_c but with $w(t_1, t_{12})$ set to 144 to further penalize the non-periodicity of the smoothing splines (though a constrained optimization or smooth-periodic decomposition can be applied).

2.6.4 Thin-plate Splines in Euclidean Space

Two-dimensional splines can be constructed via the product of two one-dimensional splines; the corresponding RKHS is the tensor product of the kernels' RKHSs. If $\mathcal{H}_1$ is an RKHS with reproducing kernel k_1 and $\mathcal{H}_2$ an RKHS with reproducing kernel k_2, then $\mathcal{H}_1 \otimes \mathcal{H}_2$ is an RKHS with reproducing kernel $k((x_1, y_1), (x_2, y_2)) = k_1(x_1, x_2)k_2(y_1, y_2)$ (see, e.g., [4], Part I, Section 8, Theorems I-II). Thus, two-dimensional splines in the plane, for instance, can be constructed via natural polynomial splines on each axis and two-dimensional splines on a cylinder or sphere can be constructed via a natural polynomial spline on the non-periodic axis and a periodic polynomial spline on the periodic (longitudinal) axis. More generally, one can create splines by aggregating one-dimensional splines fit on many data-aligned or random projections of the index set[37]

However, this method is unsatisfactory in certain applications. Kernels express a notion of similarity on the index set and related notions of smoothness of functions in the RKHS. One particularly sensible and well-motivated wiggliness energy on Euclidean space, which generalizes the natural cubic splines' wiggliness penalty, gives rise to the thin-plate splines.

The thin-plate splines were introduced by Harder and Desmarais [61] and by Duchon [39, 40], with early theory developed by Duchon, Meinguet [99, 100, 101], depending on results from Attéia [5, 6], Deny and Lions [37], and Matheron [97]. While the original theory was based on the integration by parts of the energy functional and not the reproducing property, the splines were found to fit nicely into the reproducing kernel Hilbert space (RKHS) framework [158], with the thin-plate spline interpolant of scattered data being a nice application of Wahba's representer theorem (Proposition 2.60).

The thin-plate splines in Euclidean 2-space minimize the energy[38]

$$J_{2,\mathbb{R}^2}(u) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\left(\frac{\partial^2 u}{\partial x_1^2}\right)^2 + 2\left(\frac{\partial^2 u}{\partial x_1 \partial x_2}\right)^2 + \left(\frac{\partial^2 u}{\partial x_2^2}\right)^2\right) \mathrm{d}x_1\,\mathrm{d}x_2\,, \tag{52}$$

which, via Hooke's law, represents the bending energy of a thin[39] interpolating sheet of an isotropic material like steel in the linear elastic regime, with in-plane deformation unpenalized[40].

[37] Given k directions $v_1, \ldots, v_k$ in $\mathcal{X} = \mathbb{R}^d$, placed in a $k \times d$ matrix $\mathbf{V}$, with $k > d$ when sparse representations of the data distribution in $\mathcal{X}$ are unavailable, one can fit a spline or Gaussian process regression to each of the k projections of the data set $\{(\langle x_i, v_1\rangle_{\mathbb{R}^d}, \ldots, \langle x_i, v_k\rangle_{\mathbb{R}^d})^T, y_i\}_{i=1}^n$. This vector of splines predicts, at any $x \in \mathcal{X}$, k values. They can be aggregated, for instance, by inverting our geometric model $\mathbf{V}^\dagger$ and applying it to the predictions. We can also take into account the local informativeness of each projection – an inverse proxy for which is the error bar on the spline or Gaussian process of direction j at $\langle x, v_j\rangle_{\mathbb{R}^d}$.

[38] Other sorts of bending energies based on local geometry can be imagined. Given a cross field on a surface, for instance, we could interpolate scattered data to minimize the local bending energy of the interpolant along the two orthogonal axes associated with each point, defined using a frame field operator [111]. We would thereby fit to scattered data *orthotropic* thin-plate splines.

[39] For sufficiently thin plates, the Kirchoff-Love hypothesis holds that points on a normal of the middle plane of the plate remain on the (surface) normal after deformation and that axial deformation, which maps vectors normal to the midplane before deformation to vectors normal to the deformed midsurface, is an isometry.

[40] If u represents the vertical displacement of a thin sheet whose mid-plane, before deformation, is placed in the xy-plane, then, with the Hessian matrix as

$$H = \begin{pmatrix} \frac{\partial^2 u}{\partial x^2} & \frac{\partial^2 u}{\partial x \partial y} \\ \frac{\partial^2 u}{\partial y \partial x} & \frac{\partial^2 u}{\partial y^2} \end{pmatrix},$$

the total deformation energy [8, 147] (ignoring the relative dilation of the plate in the lateral direction orthogonal to the bending, or, equivalently, if the material's Poisson ratio is 0) is proportional to

$$\operatorname{trace}\left(H^2\right) = \left(\frac{\partial^2 u}{\partial x_1^2}\right)^2 + 2\left(\frac{\partial^2 u}{\partial x_1 \partial x_2}\right)^2 + \left(\frac{\partial^2 u}{\partial x_2^2}\right)^2.$$

By extension, they are solutions to the more general scattered data fitting problem of functions on $\mathbb{R}^d$ involving the penalty functional of order m

$$J_{m,\mathbb{R}^d}(u) = \sum_{\alpha_1+\cdots+\alpha_d=m} \frac{m!}{\alpha_1!\dots\alpha_d!} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\frac{\partial^m u}{\partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}} \right)^2 \mathrm{d}x_1 \dots \mathrm{d}x_d \,. \tag{53}$$

Formally integrating (53) by parts (e.g., assume u is sufficiently smooth and rapidly decreasing), we can write the penalty functional in terms of the mth-iterated Laplacian Δ^m, where the Laplacian in $\mathbb{R}^d$ is

$$\Delta u \stackrel{\text{def}}{=} \frac{\partial^2 u}{\partial x_1^2} + \cdots + \frac{\partial^2 u}{\partial x_d^2};$$

namely, we can write

$$J_{m,\mathbb{R}^d}(u) = (-1)^m \int_{\mathbb{R}^d} \dots \int u(x_1, \dots, x_d) \cdot (\Delta^m u)(x_1, \dots, x_d) \, \mathrm{d}x_1 \dots \mathrm{d}x_d \,. \tag{54}$$

Notice that we penalize wiggliness throughout $\mathbb{R}^d$. Recall that, when constructing the natural polynomial splines on $\mathcal{X} = \mathbb{R}^1$, we penalized interpolant wiggliness only in the interval between the smallest and largest sample location–without loss of generality, on $(0, 1)$. The optimal solution happens to exhibit no wiggliness beyond the sample data (i.e., natural cubic splines, with $m = 2$, extrapolate beyond the data samples as a degree-$m-1$ polynomial: an affine function). As Attéia demonstrates in Section 2 of [6], when $\mathcal{X} = \mathbb{R}^d$ for $d \geq 2$, the situation is rather different. The choice of domain in which to enforce the wiggliness penalty affects the structure of the solution. We decide, therefore, to apply the penalty everywhere. For this approach to work, we need the technical constraint $m > d/2$.

This functional $J_{m,\mathbb{R}^d}$ is induced by a bilinear form with which the fundamental solution $\Phi_{d,m}$ of the mth iterated Laplacian formally satisfies the reproducing property. Since

$$\Delta^m \Phi_{d,m} \left(||x - x'||_{\mathbb{R}^d} \right) = \delta(x - x')$$

as distributions, we can see the reproducing property via integration by parts: for u in the Schwartz space $\mathcal{S}$ of rapidly decreasing functions,

$$(-1)^m \int_{\mathbb{R}^d} \Phi_{d,m} \left(||\omega||_{\mathbb{R}^d} \right) (\Delta^m u)(x - \omega) \, \mathrm{d}\omega = \int_{\mathbb{R}^d} \left(\Delta^m \Phi_{d,m}(||\omega - x||_{\mathbb{R}^d}) \right) u(\omega) \, \mathrm{d}\omega = u(x),$$

by the sifting property of the Dirac delta. Note that the inverse Fourier transform of $\Delta^m u(x - \cdot)$ is $\omega \mapsto (-1)^m e^{i\langle x, \omega \rangle_{\mathbb{R}^d}} ||\omega||_{\mathbb{R}^d}^2 \widehat{u}(\omega)$, which we can see using integration by parts and the fact that the complex exponentials $x \mapsto e^{i\langle x, \omega \rangle_{\mathbb{R}^d}}$ are (generalized) eigenfunctions of the m-iterated Laplacian with corresponding eigenvalues $(-1)^m ||\omega||_{\mathbb{R}^d}^{2m}$. Accordingly, we can show[41] that if $\widehat{\Phi}_{d,m} \left(||\omega||_{\mathbb{R}^d} \right) =$

[41] The manipulations here are permissible as the function Φ we have identified by enforcing the reproducing property has "generalized Fourier transform" $\widehat{\cdot}$ of order $l = m - \lceil \frac{d}{2} \rceil + 1$ on $\mathbb{R}^d \backslash \{0\}$ (see [162], Theorems 8.16-8.17) and $\widehat{u}(\omega) ||\omega||_{\mathbb{R}^d}^{2m}$ is in $\mathcal{S}$ and is $O(||\omega||_{\mathbb{R}^d}^{2m})$ as $||\omega||_{\mathbb{R}^d}$ approaches 0. The function Φ is then conditionally positive definite of this order, and in particular of order $m \geq l$ (see [162], Theorems 8.2 and 10.36). The generalized Fourier transform of order m coincides with the classical Fourier transform for functions in $L^1(\mathbb{R}^d)$ and the distributional Fourier transform on the subset of the Schwartz space that converges toward 0 sufficiently rapidly: $O(||\omega||_{\mathbb{R}^d}^{2m})$. However, it enables us to give a characterization of *conditionally* positive-definite radial basis functions that extends the characterization of Bochner, which is stated in terms of the Fourier-Stieltjes integral of nonnegative Borel measures, when we do not have integrability before we "project out" polynomials. The generalized Fourier transform of a polynomial of degree at most $2m - 1$ is the zero function, of order m.

$(2\pi)^{-d/2}\,\|\omega\|_{\mathbb{R}^d}^{-2m}$, the reproducing property appears to hold[42]

$$\begin{aligned}(-1)^m\int_{\mathbb{R}^d}\Phi_{d,m}\left(||\omega||_{\mathbb{R}^d}\right)\Delta^m u(x-\omega)\,\mathrm{d}\omega &= (-1)^m\int_{\mathbb{R}^d}\widehat{\Phi}_{d,m}(||\omega||_{\mathbb{R}^d})\cdot\mathcal{F}^{-1}\left(\Delta^m u(x-\cdot)\right)\mathrm{d}\omega\\ &= (-1)^m\int_{\mathbb{R}^d}(2\pi)^{-d/2}||\omega||_{\mathbb{R}^d}^{-2m}\cdot\left((-1)^m e^{i\langle x,\omega\rangle_{\mathbb{R}^d}}||\omega||_{\mathbb{R}^d}^{2m}\widehat{u}(\omega)\right)\mathrm{d}\omega\\ &= (2\pi)^{-d/2}\int_{\mathbb{R}^d}\widehat{u}(\omega)e^{i\langle x,\omega\rangle_{\mathbb{R}^d}}\,\mathrm{d}\omega = u(x).\end{aligned}$$

In the spatial domain,

$$\Phi_{d,m}(||s-t||_{\mathbb{R}^d}) = \eta_d\begin{cases}||s-t||_{\mathbb{R}^d}^{2m-d}, & \text{if } d \text{ is odd;}\\ ||s-t||_{\mathbb{R}^d}^{2m-d}\log||s-t||_{\mathbb{R}^d}, & \text{if } d \text{ is even,}\end{cases}\tag{55}$$

where $\eta_d > 0$ is a proportionality constant. Dividing out η_d from $\Phi_{d,m}$, we will write

$$E_m(s,t) = \frac{1}{\eta_d}\Phi(||s-t||_{\mathbb{R}^d}).$$

Importantly (and not surprisingly!), this function E_m (like $\Phi_{d,m}$) depends only on the Euclidean distance between its arguments, and is therefore called a *radial basis function* in the literature. But it is not a reproducing kernel. While the function $E_m(||\cdot - t||_{\mathbb{R}^d})$ reproduces evaluation at t with respect to our bilinear form, it is not positive definite[43]. Moreover, it is not of finite wiggliness.

It is, however, *conditionally* positive-definite. This means that, for all $n \in \mathbb{N}$, all pairwise distinct sets $\{x_i\}_{i=1}^n \subseteq \mathbb{R}^d$, and all $\alpha \in \mathbb{R}^n$ satisfying $\sum_{i=1}^n \alpha_i p(x_i) = 0$ for all polynomials p of degree at most $m-1$, the quadratic form $\sum_{i=1}^n\sum_{j=1}^n \alpha_i\alpha_j\Phi(||x_i - x_j||_{\mathbb{R}^d}) \geq 0$. In other words, $\Phi_{d,m}$ is positive definite on the space of functions "orthogonal to" the polynomials of degree at most $m-1$.

We can derive a positive-definite kernel through the decomposition of Section 2.4. In effect, we can do so by limiting E_m to have partial derivatives of order m of finite energy by projecting out polynomial functions of each argument. This kernel is not, however, radial (see Figure 6), as it depends on the choice of basis for the penalty null space – a space of polynomials. Fortunately, only the reproducing kernel, not the wiggliness seminorm, is affected by this choice of basis.

In past examples, our wiggliness penalty – once appropriate conditions were placed to ensure definiteness – was expressed using a positive-definite kernel and the model space was the corresponding RKHS. In this case, however, the wiggliness penalty is determined by a conditionally positive-definite function. We must take greater care in identifying the model space, which is not simply the closure of the span of the representations of evaluation on the index set, as it is with positive-definite kernels.

Model space: Suppose $m > d/2$. Let $\mathcal{H} = BL_m(L^2(\mathbb{R}^d))$ be the Beppo Levi space[44] of order m, that is, the space of distributions whose weak partial derivatives of total order m are in $L^2(\mathbb{R}^d)$[45]. Thus,

$$BL_m(L^2(\mathbb{R}^d)) = \{u \in \mathcal{S}'(\mathbb{R}^d)\,|\,D^\alpha u \in L^2(\mathbb{R}^d),\ \text{for all } |\alpha| = m\},$$

[42]This reproducing property only works if we place certain restrictions on u. For instance, if we add a harmonic function h to u, h will be annihilated, and $u+h$ cannot exhibit the reproducing property. One way out of this pickle is to not reproduce a single pointwise evaluation but rather a weighted sum of evaluations such that the weighted sum always annihilates harmonic functions, as in [162], Theorem 10.41.

[43]See Remark 2.27 and Proposition 2.26. A radial function that yields a positive-definite kernel in Euclidean space of any dimension cannot have zeros. More generally, those that are positive definite in a fixed Euclidean space $\mathbb{R}^d$ have a Hankel transform characterization.

[44]These are often called homogeneous Sobolev spaces. Beppo Levi was not fond of the appellation [106].

[45]The Beppo Levi spaces, more generally, can take any separable complete space in the place of $L^2(\mathbb{R}^d)$ [37].

where $\alpha \geq 0$ is a multi-index in $\mathbb{N}^d$ with $|\alpha| = \sum_{i=1}^d \alpha_i$, D^α the corresponding weak partial derivative, and $\mathcal{S}'(\mathbb{R}^d)$ the tempered distributions. The space $\mathcal{H}$ is endowed with the semi-inner product

$$\begin{aligned}\langle f, g\rangle_{\mathcal{H}} &= \sum_{|\alpha|=m} \frac{m!}{\alpha!} \langle D^\alpha f, D^\alpha g\rangle_{L^2(\mathbb{R}^d)} \\ &= \sum_{\alpha_1+\cdots+\alpha_d=m} \frac{m!}{\alpha_1!\ldots\alpha_d!} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{\partial^m f}{\partial x_1^{\alpha_1}\ldots\partial x_d^{\alpha_d}} \frac{\partial^m g}{\partial x_1^{\alpha_1}\ldots\partial x_d^{\alpha_d}} \,\mathrm{d}x_1\ldots\mathrm{d}x_d\,,\end{aligned}$$

with weights chosen so that the seminorm matches our Laplacian-based wiggliness penalty ($\Delta^m = \sum_{|\alpha|=m} \frac{m!}{\alpha!} D^\alpha$). In particular, any member of $BL_m(L^2(\mathbb{R}^d)$ has finite seminorm $||\cdot||_{\mathcal{H}}$. The space can also be defined on the Fourier side, using the multinomial theorem. This penalty is the final term of the usual Sobolev norm, so that we do not penalize as wiggly the polynomials of degree at most $m-1$. Since we apply the penalty throughout Euclidean space, the Beppo Levi space does not coincide algebraically with the classical Sobolev space, as was the case for the natural polynomial splines. For example, the affine functions are included in the Beppo Levi space of order 2 but not the corresponding classical (inhomogeneous) Sobolev space.

By a classic result[46], the null space $\mathcal{H}_0$ of the seminorm penalty $J_{m,\mathbb{R}^d}$ in $\mathcal{H} = BL_m(L^2(\mathbb{R}^d))$ is the $M = \binom{m+d-1}{d}$-dimensional [42] space of polynomials in the variables $x_1, \ldots, x_d$ of total degree at most $m-1$, for $m > d/2$.

Deny and Lions [37] showed that the quotient space $\mathcal{H}/\mathcal{H}_0$ is a Hilbert space in which the seminorm (54) is a definite inner product. This is due to the fact that the compactly supported test functions $C_0^\infty(\mathbb{R}^d)$ are not just elements of $BL_m(L^2(\mathbb{R}^d))$ – all derivatives of order m are continuous and compactly supported – but *dense* in $BL_m(L^2(\mathbb{R}^d))$ (see, e.g., [162], Theorem 10.40, or [37], Theorem 2.3). This density allows us to write the semi-inner product on $BL_m(L^2(\mathbb{R}^d))$ as follows[47]:

$$\langle f, g\rangle_{\mathcal{H}} = (-1)^m \int_{\mathbb{R}^d} f(x)(\Delta^m g)(x)\,\mathrm{d}x\,.$$

Using the decomposition principle, we can view the Beppo Levi space as an RKHS of slowly growing, continuous functions.

Null space $\mathcal{H}_0$: The null space $\mathcal{H}_0$ of the seminorm $J_{m,\mathbb{R}^d}$ is the $M = \binom{m+d-1}{d}$-dimensional [42] space of polynomials in the variables $x_1, \ldots, x_d$ of total degree at most $m-1$. A finite-dimensional space of continuous functions, it is an RKHS. With the most common penalty (52), $d = 2$ and $m = 2$, so we get $M = 3$; the null space is spanned by $\{\phi_1, \phi_2, \phi_3\}$, where $\phi_1(x_1, x_2) = 1$, $\phi_2(x_1, x_2) = x_1$, and $\phi_3(x_1, x_2) = x_2$ form the basis of $\mathcal{H}_0$. We endow $\mathcal{H}_0$ with the following inner product

$$\langle f, g\rangle_{\mathcal{H}_0} = \sum_{i=1}^M f(x_i)g(x_i), \text{ given a unisolvent set } U = \{x_1, \ldots, x_M\} \text{ of } M = \binom{m+d-1}{d} \text{ points.}$$

Recall, a unisolvent set in $\mathbb{R}^d$ is a set of $M = \binom{m+d-1}{d}$ points such that the only polynomial in $x_1, \ldots, x_d$ of total degree at most $m-1$ that evaluates to zero at each of the points is the zero polynomial[48]. In particular, the evaluation functionals at the points in U are linearly independent. Thus, the unique partition of unity in $\mathcal{H}_0$ of U–that is, a set of polynomials $p_1, \ldots, p_M$ such that

[46] In effect, the only slowly growing harmonic functions are polynomials. See [133], on p. 60, the corollary to Theorem VI. Alternatively, see [162], Lemma 10.38, or [37], pages 366-368.

[47] See the technical conditions on the density result of [162], Theorem 10.40, and their use in Theorem 10.41.

[48] Put another way, the least-squares regression of observations at each of the M points in the set U on the M polynomial basis functions of total degree at most $m-1$ is unique: the matrix whose ith row and jth column contains the ith such polynomial evaluated at the jth point in the set is of full rank.

$p_j(x_i) = \delta_{i,j}$–forms an orthonormal basis of the M-dimensional RKHS $\mathcal{H}_0$ (as in Lemma 2.58). With respect to this basis, we can find the representer of evaluation $k_t \in \mathcal{H}_0$ by expanding it on the basis functions $k_t = \sum_{i=1}^M \alpha_i p_i$ and setting $\alpha_i = p_i(t)$, so that, for any polynomial $f \in \mathcal{H}_0$ written $f = \sum_{i=1}^M \beta_i p_i$, we establish the reproducing property

$$\langle f, k_t \rangle_{\mathcal{H}_0} = \sum_{i=1}^M \left(\sum_{j=1}^M \beta_j p_j \right)(x_i) \left(\sum_{j=1}^M p_j(t) p_j \right)(x_i) = \sum_{i=1}^M \beta_i p_i(t) = f(t).$$

Using our orthonormal basis, define the operator $\mathrm{P}_0 : \mathcal{H} \to \mathcal{H}_0$ by $f \mapsto \sum_{i=1}^M f(x_i) p_i$. Then for any f, g in our Beppo Levi model space, some unisolvent set U of M points $x_1, \ldots, x_M$ in $\mathbb{R}^d$, and the corresponding partition of unity $p_1, \ldots, p_M$ satisfying $p_i(x_j) = \delta_{ij}$, we have that

$$\begin{aligned}\langle \mathrm{P}_0 f, \mathrm{P}_0 g \rangle_{\mathcal{H}_0} &= \sum_{i=1}^M (\mathrm{P}_0 f)(x_i)(\mathrm{P}_0 g)(x_i) \\ &= \sum_{i=1}^M \left(\sum_{j=1}^M f(x_j) p_j \right)(x_i) \cdot \left(\sum_{j=1}^M g(x_j) p_j \right)(x_i) = \sum_{i=1}^M f(x_i) g(x_i).\end{aligned}$$

Wiggly space $\mathcal{H}_1$: Define $\mathcal{H}_1$ to be the space of codimension M whose representative elements are functions that evaluate to 0 on each point in U. Thus, their projection onto $\mathcal{H}_0$ is 0; they are orthogonal to $\mathcal{H}_0$: for each $j \in \{1, \ldots, M\}$, we have that

$$0 = \langle f, p_j \rangle_{\mathcal{H}_0} = \sum_{i=1}^M f(x_i) \underbrace{p_j(x_i)}_{\delta_{ij}} = f(x_j).$$

The projection operator $\mathrm{P}_1 = \mathrm{I} - \mathrm{P}_0$ from $\mathcal{H}$ to $\mathcal{H}_1$ that satisfies $\mathrm{P}_1 f = f - \sum_{i=1}^M f(x_i) p_i$ thereby projects out the "polynomial component" of f.

We endow $\mathcal{H}_1$ with the inner product

$$\langle f, g \rangle_{\mathcal{H}_1} = (-1)^m \int_{\mathbb{R}^d} f(x) \cdot (\Delta^m g)(x) \, \mathrm{d}x \,.$$

By construction, for each $f_1 \in \mathcal{H}_1$, $\langle \cdot, f_1 \rangle_{\mathcal{H}_0} = 0$; for each $f_0 \in \mathcal{H}_0$, we have, moreover, that $\langle \cdot, f_0 \rangle_{\mathcal{H}_1} = 0$, since f_0 is a polynomial of total degree at most $m - 1$. Since the smooth, compactly supported test functions $C_0^\infty(\mathbb{R}^d)$ are dense in $BL_m(\mathbb{R}^d)$, we know a reproducing kernel for $\mathcal{H}_1$ must satisfy the following: for all $f \in C_0^\infty(\mathbb{R}^d)$,

$$(\mathrm{P}_1 f)(x) = \langle f - \mathrm{P}_0 f, k_x^1 \rangle_{\mathcal{H}_1} = \langle f, k_x^1 \rangle_{\mathcal{H}_1} = (-1)^m \int_{\mathbb{R}^d} f(y) (\Delta^m k_x^1)(y) \, \mathrm{d}y = (f, (-1)^m \Delta^m k_x^1),$$

where $(\cdot, \cdot)$ is the canonical dual pairing. But since obviously, for any $f \in C_0^\infty(\mathbb{R}^d)$,

$$(\mathrm{P}_1 f)(x) = f(x) - \sum_{j=1}^M p_i(x) f(x_i) = \left(f, \delta(\cdot - x) - \sum_{i=1}^M p_i(x) \delta(\cdot - x_i) \right),$$

our reproducing kernel k^1 must satisfy the following distributional differential equation:

$$(-1)^m \Delta k_x^1 = \delta(\cdot - x) - \sum_{i=1}^M p_i(x) \delta(\cdot - x_i).$$

By superposition, we can give a solution to the above equation in terms of the fundamental solution E_m of $(-1)^m\Delta^m f = \delta$

$$\text{candidate } k_x^1(y)\text{: } C_x(y) = E_m(x,y) - \sum_{i=1}^{M} p_i(x)E_m(x_i,y).$$

The above candidate function does not reside in $\mathcal{H}_1$ but it does reside in $\mathcal{H}$ if $m > d/2$. (This does not hold for E_m, with one argument fixed. It is an easy matter to verify that its order-m partials do not have finite energy.) Thus, for each fixed x, subtracting out a weighted sum of partial evaluations of E_m (with the weights being a polynomial in x) from $y \mapsto E_m(x,y)$, which is not in the Beppo Levi space $\mathcal{H} = BL_m(L^2(\mathbb{R}^d))$, forms $y \mapsto C_x(y)$, which is. This function $y \mapsto C_x(y)$ is thus "a polynomial away" from being in $\mathcal{H}_1$. This polynomial difference, moreover, can (using a density argument) be added to either side of the $\mathcal{H}_1$ innner product $\langle\cdot,\cdot\rangle_{\mathcal{H}_1}$ (but not both!) without changing its values. We orthogonally project the function $y \mapsto C_x(y)$ onto $\mathcal{H}_1$ to form the reproducing kernel for $\mathcal{H}_1$

$$\begin{aligned}
k_x^1(y) = (\mathrm{P}_1 C_x)(y) &= \left(E_m(x,y) - \sum_{i=1}^{M} p_i(x)E_m(x_i,y)\right) - \sum_{j=1}^{M} p_j(y)\left(E_m(x,x_j) - \sum_{i=1}^{M} p_i(x)E_m(x_i,x_j)\right) \\
&= E_m(x,y) - \sum_{i=1}^{M} p_i(x)E_m(x_i,y) - \sum_{j=1}^{M} p_j(y)E_m(x,x_j) + \sum_{i=1}^{M}\sum_{j=1}^{M} p_i(x)p_j(y)E_m(x_i,x_j).
\end{aligned}$$

This reproducing kernel preserves the reproducing property but affects the wiggliness penalty in general. However, for certain linear combinations of the conditionally positive-definite kernel evaluated on a data set $\{E_m(\cdot,x_i')\}_{i=1}^n$, projecting out polynomials does not affect the wiggliness penalty. Indeed, given n data points $\{x_i'\}_{i=1}^n$ (the $'$ distinguishes the data sample locations from our unisolvent set $\{x_l\}_{l=1}^M$), for any vector $c \in \mathbb{R}^n$ such that $\sum_{i=1}^n c_i p_j(x_i') = 0$ for $j = 1,\ldots,M$, we have that

$$\begin{aligned}
J_{m,d}\left(\sum_{i=1}^{n} c_i k_{x_i'}^1\right) &= c^T\mathbf{K}c = \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j k^1(x_i',x_j') \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \Bigg(E_m(x_i',x_j') - \sum_{l=1}^{M} E_m(x_i',x_l)p_l(x_j') - \sum_{l=1}^{M} E_m(x_l,x_j')p_l(x_i') + \\
&\qquad \sum_{l=1}^{M}\sum_{q=1}^{M} p_l(x_i')p_q(x_j')E_m(x_l,x_q)\Bigg) \\
&= \left(\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j E_m(x_i',x_j')\right) - \sum_{i=1}^{n} c_i \sum_{l=1}^{M} E_m(x_i',x_l)\underbrace{\sum_{j=1}^{n} c_j p_l(x_j')}_{0} - \\
&\qquad \sum_{j=1}^{n} c_j \sum_{l=1}^{M} E_m(x_l,x_j')\underbrace{\sum_{i=1}^{n} c_i p_l(x_i')}_{0} + \sum_{l=1}^{M}\sum_{q=1}^{M} E_m(x_l,x_q)\underbrace{\left(\sum_{i=1}^{n} c_i p_l(x_i')\right)}_{0}\underbrace{\left(\sum_{j=1}^{n} c_j p_q(x_j')\right)}_{0} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j E_m(x_i',x_j') = c^T\mathbf{E}c = J_{m,d}\left(\sum_{i=1}^{n} c_i E_m(\cdot,x_i')\right),
\end{aligned}$$

where $\mathbf{K}$ is the Gram matrix of k^1 and $\mathbf{E}$ the Gram matrix of E_m on our sample locations $\{x_i'\}_{i=1}^n$.

With a unisolvent set of $(n > M)$ sample points $\{x'_i\}_{i=1}^n$, we assumed that

$$\text{for all } j \in \{1, \dots, M\},\ \sum_{i=1}^n c_i p_j(x'_i) = 0,$$

or, equivalently,

$$c \in \operatorname{null} \mathbf{T}^T,\ \text{where } (\mathbf{T})_{i,j} = p_j(x'_i) \text{ and the } n \times M \text{ matrix } \mathbf{T} \text{ has rank } M.$$

We found that this constraint forced the wiggliness penalty of the $\sum_{i=1}^n c_i k^1(\cdot, x'_i)$ to equal the wiggliness penalty of $\sum_{i=1}^n c_i E_m(\cdot, x'_i)$. This is explained by the fact that any $c \in \operatorname{null} \mathbf{T}^T$ is called a *generalized divided difference* [69, 118, 158] of order m since it annihilates all polynomials of total degree less than m, much as first differences annihilate constant functions, second differences linear functions, and so forth. In particular, the difference between $c^T \mathbf{K} c$ and $c^T \mathbf{E} c$ is annihilated.

Thus, from our conditionally positive-definite kernel E_m – whose partial evaluations do not even reside in $\mathcal{H}$ – we have twice "projected out the null space" to compute a positive-definite kernel for $\mathcal{H}_1$

$$k^1(s,t) = E_m(s,t) - \sum_{i=1}^M p_i(t) E_m(x_i, s) - \sum_{j=1}^M p_j(s) E_m(t, x_j) + \sum_{i=1}^M \sum_{j=1}^M p_i(t) p_j(s) E_m(x_j, x_i).$$

With fixed t, then, we have that, modulo a polynomial in s of degree at most $m-1$ (i.e., the last two terms above),

$$k^1_t(s) = k^1(s,t) \equiv C_t(s) = E_m(s,t) - \sum_{i=1}^M p_i(t) E_m(x_i, s) \text{ ([100], equation 19)}, \tag{56}$$

so that $\mathrm{P}_1 k^1_t$ is given by the $M+1$ terms on the right-hand side of (56) (and $\mathrm{P}_0 k^1_t$ does not affect the wiggliness penalty). If $m > d/2$, the k^1_t so defined resides in $\mathcal{H}_1$ and serves as its representation of evaluation at t. Indeed, we can confirm that $\mathrm{P}_0 k^1_t = \phi_0(t)$ is identically zero as Equation (56) annihilates polynomials of degree at most $m-1$. As the $n = M+1$ points $w_i = (t, x_1, \dots, x_M)$ are a unisolvent set, with $c = (1, -p_1(t), \dots, -p_M(t))$, we have a generalized divided difference

$$\sum_{k=1}^{M+1} c_i p_j(w_i) = p_j(t) - \sum_{i=1}^M p_i(t) p_j(x_i) = p_j(t) - p_j(t) = 0 \text{ ([158], equation 2.4.28)}.$$

In summary, E_m is not a reproducing kernel for $\mathcal{H}_1$ because its partial evaluations $E_m(\cdot, x)$ for any $x \in \mathbb{R}^d$ do not reside in $\mathcal{H}_1$; however, it does reproduce function evaluation in $\mathcal{H}_1$. With one argument fixed, it is "a polynomial away" from being in $\mathcal{H}_1$. Removing this polynomial forms k^1, which is not radial (see Figure 6) but is a reproducing kernel for $\mathcal{H}_1$. For any nondegenerate data set $\{x'_i\}_{i=1}^n$ of size $n \geq M$, if we choose $c \in \mathbb{R}^n$ so that

$$\text{for } j = 1, \dots, M, \text{ the basis function of } \mathcal{H}_0 p_j \text{ satisfies } \sum_{i=1}^n c_i p_j(x'_i) = 0,$$

then the wiggliness of $\sum_{i=1}^n \alpha_i k_{x'_i}$ is the same as the wiggliness of $\sum_{i=1}^n \alpha_i E_m(\cdot, x'_i)$. Moreover, we have, by the generalized divided difference of the right-hand side of Equation (56), that

$$k^1(s,t) = \langle C_s, C_t \rangle_{\mathcal{H}_1}.$$

Example 2.69. *Let $d = 2$ and $m = 2$ so that $E_m(s,t) = ||s-t||^2_{\mathbb{R}^2} \log ||s-t||_{\mathbb{R}^2}$ (we have divided out a positive constant η_d from $\Phi_{m,d}$). Consider the set of points $S = \{(0,0)^T, (0,1)^T, (0,2)^T\}$. The Gram matrix $K^S_{E_m}$ of E_m on S is not positive definite*

$$K^S_{E_m} = \begin{pmatrix} 0 & 0 & 4\log 2 \\ 0 & 0 & 0 \\ 4\log 2 & 0 & 0 \end{pmatrix},$$

a nonzero matrix with zero trace, has both positive and negative eigenvalues.

We define $\mathcal{H}_0$ so that we can "project out" the polynomial contributions to E_m. The set $U = \{x_1, x_2, x_3\}$, with $x_1 = (0,0)^T, x_2 = (1,0)^T$, and $x_3 = (0,1)^T$, is obviously unisolvent (not collinear). A simple calculation[49] shows U has the following partition of unity

$$p_1(x) = \langle (-1,-1)^T, x\rangle_{\mathbb{R}^2} + 1; \qquad p_2(x) = \langle (1,0)^T, x\rangle_{\mathbb{R}^2}; \qquad p_3(x) = \langle (0,1)^T, x\rangle_{\mathbb{R}^2}.$$

In particular, $p_i(x_j) = \delta_{ij}$ (Kronecker delta) and for all $x \in \mathbb{R}^2$,

$$p_1(x) + p_2(x) + p_3(x) = \langle (-1,-1)^T + (1,0)^T + (0,1)^T, x\rangle_{\mathbb{R}^2} + (1+0+0) = 1$$

and

$$x_1 p_1(x) + x_2 p_2(x) + x_3 p_3(x) = (x^T(1,0)^T)\cdot(1,0)^T + (x^T(0,1)^T)\cdot(0,1)^T = x.$$

Then we can define an inner product on $\mathcal{H}_0$, the space of affine functions, for which $\{p_1, p_2, p_3\}$ serves as an orthonormal basis

$$\text{for all affine functions } f \text{ and } g \text{ on } \mathbb{R}^2,\ \langle f, g\rangle_{\mathcal{H}_0} = \sum_{i=1}^{3} f(x_i)g(x_i),$$

where $\{x_1, x_2, x_3\} = U$. Let $f = x \mapsto \alpha^T x + \beta$ be an arbitrary affine function; for any $x \in \mathbb{R}^2$, $k_x = t \mapsto \sum_{i=1}^3 p_i(x)p_i(t)$ reproduces evaluation at x

$$\begin{aligned}\langle f, k_x\rangle_{\mathcal{H}_0} &= \sum_{i=1}^{3} f(x_i)k_x(x_i) = \sum_{i=1}^{3}(\alpha^T x_i + \beta)\left(\sum_{j=1}^{3}\underbrace{p_j(x_i)}_{\delta_{ij}}p_j(x)\right) = \sum_{i=1}^{3}(\alpha^T x_i + \beta)p_i(x) \\ &= \alpha^T\left(\sum_{i=1}^{3} x_i p_i(x)\right) + \beta\left(\sum_{i=1}^{3} p_i(x)\right) = \alpha^T x + \beta = f(x).\end{aligned}$$

For all $t \in \mathbb{R}^2$, we can define

$$\begin{aligned}k^1_t = k^1(\cdot,t) &= E_m(\cdot,t) - \mathrm{P}_0 E_m(\cdot,t) = E_m(\cdot,t) - \sum_{i=1}^{3} p_i(\cdot)E_m(x_i,t) \\ &= \varphi(||\cdot - t||_{\mathbb{R}^2}) - \varphi(||(0,0)^T - t||_{\mathbb{R}^2})\left(1 - (1,1)^T\cdot\right) - \varphi(||(1,0)^T - t||_{\mathbb{R}^2})\left((1,0)^T\cdot\right) - \\ &\quad \varphi(||(0,1)^T - t||_{\mathbb{R}^2})\left((0,1)^T\cdot\right).\end{aligned}$$

[49] For i and j in $\{1,2\}$, let x_i^j (α_i^j) be the jth element of x_i (α_i), using 1-indexing. Set $p_1 = x \mapsto \alpha_1^T x + \beta_1; p_2 = x \mapsto \alpha_2^T x + \beta_2$; and $p_3 = x \mapsto \alpha_3^T x + \beta_3$. Solve the following system for the α and β parameters to recover the partition of unity

$$\begin{pmatrix} \alpha_1^1 & \alpha_1^2 & \beta_1 \\ \alpha_2^1 & \alpha_2^2 & \beta_2 \\ \alpha_3^1 & \alpha_3^2 & \beta_3 \end{pmatrix}\begin{pmatrix} x_1^1 & x_2^1 & x_3^1 \\ x_1^2 & x_2^2 & x_3^2 \\ 1 & 1 & 1 \end{pmatrix} = I_{3\times 3}.$$

where $\varphi(r) = r^2 \log(r)$. *Moreover, the reproducing kernel for* $\mathcal{H}_1$ *can be evaluated as follows: for all* $(s,t) \in \mathbb{R}^2 \times \mathbb{R}^2$, *with* $s = (s_1, s_2)$ *and* $t = (t_1, t_2)$, *we compute*

$$\begin{aligned}
k^1(s,t) &= \langle (\mathrm{I} - \mathrm{P}_0)k_s, (\mathrm{I} - \mathrm{P}_0)k_t \rangle_{\mathcal{H}_1} = \int_{\mathbb{R}^2} k_t^1(x) \Delta k_s^1(x)\, \mathrm{d}x \\
&= \int_{\mathbb{R}^2} \left(E_m(x,t) - \sum_{i=1}^{3} p_i(x) E_m(x_i, t) \right) \Delta \left(E_m(x,s) - \sum_{j=1}^{3} p_j(x) E_m(x_j, s) \right) \mathrm{d}x \\
&= E_m(s,t) - \sum_{i=1}^{3} p_i(s) E_m(x_i, t) - \sum_{j=1}^{3} p_j(t) E_m(x_j, s) + \sum_{i=1}^{3} \sum_{j=1}^{3} p_i(t) p_j(s) E_m(x_i, x_j) \\
&= \varphi(||s-t||_{\mathbb{R}^2}) + \log 2(t_1 s_2 + s_1 t_2) - (1 - s_1 - s_2)\varphi(||t||_{\mathbb{R}^2}) - s_1 \varphi(||t - (1,0)^T||_{\mathbb{R}^2}) - \\
&\quad s_2 \varphi(||t - (0,1)^T||_{\mathbb{R}^2}) - (1 - t_1 - t_2)\varphi(||s||_{\mathbb{R}^2}) - t_1 \varphi(||s - (1,0)^T||_{\mathbb{R}^2}) - t_2 \varphi(||s - (0,1)^T||_{\mathbb{R}^2}).
\end{aligned}$$

For each $s \in \mathbb{R}^2$, *the Beppo Levi wiggliness of* k_s^1 *is the following:*

$$\begin{aligned}
\langle k_s^1, k_s^1 \rangle_{\mathcal{H}_1} = k^1(s,s) = (2\log 2) s_1 s_2 - 2(1 - s_1 - s_2)\varphi(||s||_{\mathbb{R}^2}) - 2 s_1 \varphi(||s - (1,0)^T||_{\mathbb{R}^2})) \\
- 2 s_2 \varphi(||s - (0,1)^T||_{\mathbb{R}^2})
\end{aligned}$$

If $x = x_i$ *for some* $x_i \in U$, *the wiggliness is 0. Otherwise it is positive, and it grows quadratically with the distance to the unisolvent set.*

The Gram matrix of k^1 *(the reproducing kernel for* $\mathcal{H}_1$*) on* S *is symmetric, positive semidefinite*

$$K_{k^1}^S = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 8\log 2 \end{pmatrix}.$$

The conditionally positive-definite kernel E_m *has partial evaluations that do not reside in* $BL_m(L^2(\mathbb{R}^2))$*; the associated measure of similarity* grows *with the Euclidean distance between its arguments. On the other hand, the positive-definite kernel* k^1, *which is derived from* E_m *by "projecting out" the semi-norm null space, better expresses similarity on the index set. And it serves as a reproducing kernel for* $BL_m(L^2(\mathbb{R}^2))/\mathcal{P}_{m-1}$. *See Figure 6.*

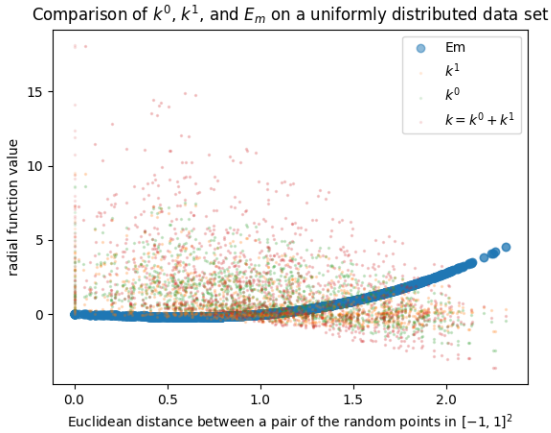
We summarize the argument as follows. The only solutions of $\Delta^m E = \delta$ in the Beppo Levi space $\mathcal{H} = BL_m(L^2(\mathbb{R}^d))$ with Beppo Levi seminorm $||u||_{BL_m(L^2(\mathbb{R}^d))} = (-1)^m \int_{\mathbb{R}^d} u(x)(\Delta u)(x)\, \mathrm{d}x$ are the polynomials of degree at most $m-1$, which span the null space of the norm. This Beppo Levi space is a semi-Hilbert space of continuous, slowly growing functions when $m > d/2$. We can make definite the semi-inner product of the Beppo Levi space by applying an inner product on its null space. We do this by creating a partition of unity for the space $\mathcal{H}_0$ of polynomials of degree at most $m-1$. A reproducing kernel for the Beppo Levi space after projecting out the polynomials of degree $m-1$ must satisfy the distributional partial differential equation

$$(-1)^m \Delta^m k_x^1 = \delta(\cdot - x) - \sum_{j=1}^{M} p_j(x) \delta(\cdot - x_i);$$

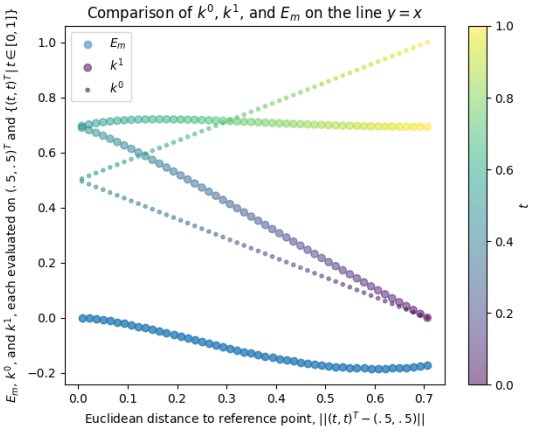
i.e. these two sides integrate against test functions identically. The fundamental solution to

$$(-1)^m \Delta^m f = \delta$$

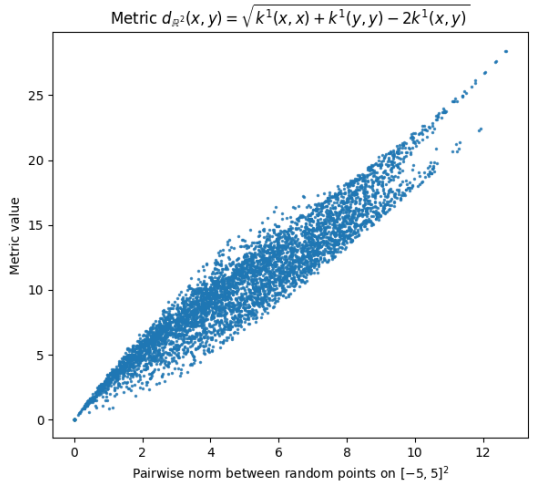
is known to be E_m, whose order-m partials do not satisfy the square-integrability criteria for inclusion in the Beppo Levi space. Projecting out the polynomial component of E_m gives a particular solution



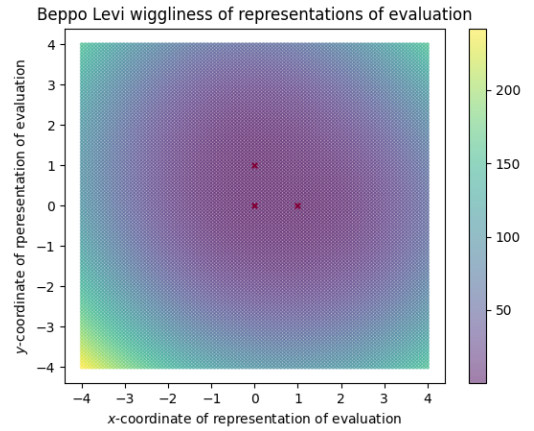
(a) E_m, k^0, and k^1 evaluated pairwise on points placed uniformly in $[-1, 1]^2$, using the unisolvent set of Example 2.69.



(b) E_m and k^1 evaluated on $y = x$, compared to a fixed reference $(0.5, 0.5)^T$. Clearly k^0 and k^1 are not radial.



(c) k^1 involves a relatively sensible notion of distance on the index set, though it is influenced by proximity to the unisolvent set.



(d) The Beppo Levi wiggliness of the representations of evaluation at a point vary substantially with the distance of the point to the unisolvent set. This is troubling as the unisolvent set is chosen independently of the data we wish to interpolate.

Figure 6: The (normalized) fundamental solution E_m of $(-1)^m \Delta^m f = \delta$ in Euclidean space is a radial function, but not a positive-definite kernel for all d: as a function of the Euclidean distance between its two inputs, it has one zero when d is odd and two zeros when d is even and thus fails Schoenberg's criterion (the second item of Proposition 2.26). Having fixed our unisolvent set as in Example 2.69, the function k^1 is positive definite and exhibits a more sensible notion of distance on the index set (though highly dependent on the choice of unisolvent set used to define $\mathcal{H}_1$!) than does E_m, according to which the similarity between points grows more than quadratically in the Euclidean distance separating them (for larger distances, with $d = m = 2$). But it is not radial, as the clouds in **(6a)** and plots with color gradient in **(6b)** attest. Values of the pseudometric (it is indeed pseudo- as there are distinct points–on the unisolvent set–that have distance zero) between points chosen uniformly at random on $[-5, 5]^2$ are plotted against the Euclidean distance in **(6c)**. The wiggliness of the representers of evaluation are plotted against their position in **(6d)**, along with the three points of the unisolvent set.

of the above differential equation

$$k_x^1 = E_m(\cdot, x) - \sum_{j=1}^{M} p_j(x) E_m(\cdot, x_j).$$

An argument based on the distributional Fourier transform and approximation by convolution can show that, when $m > d/2$, the space $\mathcal{H}_1$ is an RKHS of continuous functions that contains k_x^1 and for which k_x^1 serves as a representation of evaluation at x. Then $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$ is an RKHS with defined inner product

$$\langle f, g\rangle_{\mathcal{H}} = \langle \mathrm{P}_0 f, \mathrm{P}_0 g\rangle_{\mathcal{H}_0} + \langle (\mathrm{I} - \mathrm{P}_0) f, (\mathrm{I} - \mathrm{P}_0) g\rangle_{\mathcal{H}_1} = \sum_{i=1}^{M} f(x_i) g(x_i) + \langle (\mathrm{I} - \mathrm{P}_0) f, (\mathrm{I} - \mathrm{P}_0) g\rangle_{BL_m(L^2(\mathbb{R}^d))}.$$

Solving the Spline Smoothing Problem. Provided that $2m - d > 0$ and that c is a generalized divided difference for the n scattered data locations $\{x_i\}_{i=1}^n$, we can write the minimum-norm interpolant u^*

$$u^*(t) = \operatorname*{arg\,min}_{u \in BL_m(L^2(\mathbb{R}^d))} \sum_{i=1}^{n} (u(x_i) - y_i)^2 + \lambda J_{m,\mathbb{R}^2}(u),$$

in the form

$$u^*(t) = \sum_{j=1}^{M} d_j p_j(t) + \sum_{i=1}^{n} c_i E_m(x_i, t), \tag{57}$$

where p_j is the jth polynomial in the basis of $\mathcal{H}_0$, the space of all polynomials in $x_1, \ldots, x_d$ of maximum degree at most $m - 1$.

We give pseudocode for the thin-plate splines in Algorithm 7.

Wahba's representer theorem (Proposition 2.60) reduces this empirical risk minimization problem over an infinite-dimensional space to finite-dimensional linear algebra. The complete solution is given in [100] and summaries may be found in [41, 158, 167].

Remark 2.70 (The $m > d/2$ constraint)**.** *The technical constraint of a well-known Sobolev embedding theorem, that $2m > d$, ensures that the Sobolev space $W^{m,2}(\mathcal{X})$ embeds continuously in $C^0(\mathbb{R}^d)$, the space of continuous functions on $\mathbb{R}^d$ endowed with the L^∞ norm. This constraint does the same for the Beppo Levi space $BL_m(L^2(\mathbb{R}^d))$. This means that as functions approach each other in the Sobolev norm, they approach each other pointwise, and the pointwise evaluation operator is continuous. For the kernel of the thin-plate splines, this constraint assures the integrability of the reproducing property. Thus, in high dimensions, to get an RKHS, we need to define wiggliness in terms of high-order partial derivatives in order to implement thin-plate splines in Euclidean space. This restriction can lead to poor modeling choices in high dimensions. Strategies for working in high dimensions with a wiggliness penalty depending on partial derivatives of low total order include dimensionality-reduction techniques (random projections, t-SNE, UMAP, PCA, etc.), constructing a kernel via tensor product of lower-dimensional kernels, and approximating the space by a point cloud with a proximity graph structure.*

2.7 Thin-plate Splines on the Sphere

Motivated by meteorological applications, Wahba and Wendelberger [154, 156, 158, 161] in the late 1970s and early 1980s considered extending the thin-plate splines (and the periodic splines on the circle) to the $(d-1)$-sphere in $\mathbb{R}^d$. (In the present article, we restrict our attention to the 2-sphere $\mathbb{S}^2$

Algorithm 7: An algorithm that fits thin-plate splines on $\mathcal{X} = \mathbb{R}^d$ based on the Beppo Levi space seminorm penalty $J_{m,\mathbb{R}^d}$. By the representer theorem (Proposition 2.60), the solution to

$$u^* = \arg\min_{u \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^{n} (u(x_i) - y_i)^2 + \lambda J_{m,\mathbb{R}^d}(u)$$

takes the form

$$u^* = \sum_{j=1}^{M} d_j p_j + \sum_{i=1}^{n} c_i E_m(x_i, \cdot),$$

subject to the constraint (56), which guarantees a finite seminorm energy. This algorithm recovers $c \in \mathbb{R}^n$ and $d \in \mathbb{R}^M$ from samples $\{y_i\}_{i=1}^n$, $y_i \in \mathbb{R}$, taken at scattered values $\{x_i\}_{i=1}^n$, $x_i \in \mathbb{R}^d$.

Data: A set of n sample locations $\{x_i\}_{i=1}^n$ in $\mathbb{R}^d$ and n corresponding sample values $y_i \in \mathbb{R}$.
Parameters: A regularization penalty parameter $\lambda \geq 0$ and order m for the seminorm wiggliness penalty $J_{m,\mathcal{X}}$. We require that the penalty order $m > d/2$.
Result: A set of basis function weights $c \in \mathbb{R}^n$ and $d \in \mathbb{R}$ specifying the empirical risk minimizing function u^*.

Compute the $n \times n$ Gram matrix $\mathbf{E}$ in whose ith row and jth column reposes the value

$$(\mathbf{E})_{i,j} \leftarrow E_m(x_i, x_j),$$

where

$$E_m(x_i, x_j) = \begin{cases} ||x_i - x_j||_{\mathbb{R}^d}^{2m-d}, & \text{if } d \text{ is odd} \\ ||x_i - x_j||_{\mathbb{R}^d}^{2m-d} \log ||x_i - x_j||_{\mathbb{R}^d}, & \text{otherwise.} \end{cases} \tag{58}$$

Form the null-space basis function matrix $\mathbf{T} = \texttt{hstack}([\mathbf{1}_n, x[:,1], x[:,2] \ldots])$, where each column is one of the M basis functions on the data set $(x_i \mapsto 1, x_i \mapsto x_i[1], \ldots)$. This matrix will be used to ensure c is a generalized divided difference of order m (i.e., $c \in \operatorname{null} \mathbf{T}^T$).

$$(\mathbf{T})_{i,j} = p_j(x_i).$$

Augment the Gram matrix of E_m to form an $(n+M) \times (n+M)$ matrix $\mathbf{K}$ and set y accordingly

$$\mathbf{K} \leftarrow \begin{pmatrix} \mathbf{E} + \lambda n \mathbf{I}_{n \times n} & \mathbf{T} \\ \mathbf{T}^T & \mathbf{0}_{M \times M} \end{pmatrix} \text{ and } y \leftarrow \begin{pmatrix} y \\ \mathbf{0}_M \end{pmatrix};$$

Solve $\mathbf{K}\alpha = y$:

$$\alpha \leftarrow \mathbf{K}^{-1} y;$$

Return the spline weights $c \leftarrow \alpha[1:n]$ and $d \leftarrow \alpha[n+1:]$;

in $\mathbb{R}^3$; readers interested in thin-plate splines on the sphere $\mathbb{S}^{d-1}$ for $d > 3$ are referred to [11].) Using the analogue of Δ on the 2-sphere, the Laplace-Beltrami operator (see [105], Section 14, and [14])

$$\Delta_S u = \frac{1}{\sin^2(\theta)}\frac{\partial^2 u}{\partial \phi^2} + \frac{1}{\sin(\theta)}\frac{\partial}{\partial \theta}\left(\sin(\theta)\frac{\partial u}{\partial \theta}\right) = \csc^2(\theta)\frac{\partial^2 u}{\partial \phi^2} + \cot(\theta)\frac{\partial u}{\partial \theta} + \frac{\partial^2 u}{\partial \theta^2}, \tag{59}$$

where $\theta \in [0, \pi]$ is the colatitude[50] and $\phi \in [0, 2\pi]$ the longitude[51], they define a penalty functional analogue of (54) of order m on the 2-sphere as

$$J_{m,\mathbb{S}^2}(u) = \begin{cases} \frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi}\left(\Delta_S^{m/2}u\right)^2 \sin(\theta)\,\mathrm{d}\theta\,\mathrm{d}\phi\,, & m \text{ even} \\ \frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi}\left(\frac{\left(\frac{\partial}{\partial\phi}\left(\Delta_S^{(m-1)/2}u\right)\right)^2}{\sin^2(\theta)} + \left(\frac{\partial}{\partial\theta}\left(\Delta_S^{(m-1)/2}u\right)\right)^2\right)\sin(\theta)\,\mathrm{d}\theta\,\mathrm{d}\phi\,, & m \text{ odd.} \end{cases} \tag{60}$$

If u is sufficiently smooth, then this matches, via Green's first identity and the product rule for gradients,

$$J_{m,\mathbb{S}^2}(u) = (-1)^m \int_0^{2\pi}\int_0^{\pi} u(\theta,\phi)(\Delta u)(\theta,\phi)\sin(\theta)\,\mathrm{d}\theta\,\mathrm{d}\phi\,. \tag{61}$$

Rather than searching for a closed form of the Green's function of the mth iterated Laplace-Beltrami operator on the sphere (as we did for the natural cubic splines or thin-plate splines in Euclidean space), we use Mercer synthesis (Section 2.2.1). Because the spherical harmonics form a complete orthonormal system for $L^2(\mathbb{S}^2)$ and are eigenfunctions of the Laplace-Beltrami operator, defining the splines on the "Fourier side"–as we did with the periodic splines–simplifies the derivation. The resulting kernel exhibits the Funk-Hecke multiplicities (the eigenvalue of Δ_S on the eigenfunction Y_l^n depends on the degree l but not the order n; see Proposition 2.40) and is therefore isotropic; we can accordingly use the results from Section 2.2.3.

2.7.1 A Series Form of the Thin-plate Spline Penalty on the Sphere

By the completeness of the spherical harmonics [78], any function $u \in L^2(\mathbb{S}^2)$ can be written

$$u(\theta,\phi) \sim \sum_{l=0}^{\infty}\sum_{n=-l}^{l}(u)_{l,n}Y_l^n(\theta,\phi),$$

where the right-hand side converges to the left-hand side in $L^2(\mathbb{S}^2)$ and the Fourier expansion of u is given by

$$(u)_{l,n} = \langle u, Y_l^n\rangle_{L^2(\mathbb{S}^2)} = \frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi} u(\theta,\phi)Y_l^n(\theta,\phi)\sin(\theta)\,\mathrm{d}\theta\,\mathrm{d}\phi\,.$$

It is easy to see that the penalty (as (60) or (61)) can be formally written as an infinite series in terms of these Fourier coefficients, using (18) and the orthonormality of the spherical harmonics.

[50] According to the "physics convention" of spherical coordinates, the colatitude is $\frac{\pi}{2}$ minus the "math convention" latitude; that is, the colatitude is 0 at the "North Pole"–along the positive rectangular z-axis–and π at the "South Pole".

[51] The longitude is the azimuthal angle measured counterclockwise from the positive rectangular x-axis, through which our reference meridian passes.

Indeed, for the case (60) with m even,

$$\begin{aligned}
J_{m,\mathbb{S}^2}(u) &= \frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi}\left(\Delta_S^{m/2}u\right)^2\sin(\theta)\,\mathrm{d}\theta\,\mathrm{d}\phi \\
&= \frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi}\Delta_S^{m/2}\left(\sum_{l=0}^{\infty}\sum_{n=-l}^{l}(u)_{l,n}Y_l^n(\theta,\phi)\right)\Delta_S^{m/2}\left(\sum_{l=0}^{\infty}\sum_{n=-l}^{l}(u)_{l,n}Y_l^n(\theta,\phi)\right)\sin(\theta)\,\mathrm{d}\theta\,\mathrm{d}\phi \\
&= \left\langle\sum_{l=0}^{\infty}\sum_{n=-l}^{l}(u)_{l,n}(l(l+1))^{m/2}Y_l^n(\theta,\phi),\sum_{l=0}^{\infty}\sum_{n=-l}^{l}(u)_{l,n}(l(l+1))^{m/2}Y_l^n(\theta,\phi)\right\rangle_{L^2(\mathbb{S}^2)} \\
&= \sum_{l=0}^{\infty}\sum_{n=-l}^{l}(u)_{l,n}^2\,(l(l+1))^m\underbrace{\langle Y_l^n,Y_l^n\rangle_{L^2(\mathbb{S}^2)}}_{1} \\
&= \sum_{l=0}^{\infty}(l(l+1))^m\sum_{n=-l}^{l}(u)_{l,n}^2 = \sum_{l=1}^{\infty}(l(l+1))^m\sum_{n=-l}^{l}(u)_{l,n}^2\,,
\end{aligned}$$

where the second equality on the last line follows from the fact that Δ_S annihilates the DC component of a signal. Thus, the DC component of a signal $(u)_{0,0}$ contributes nothing to the wiggliness. Intuitively, subtracting a mean from a function should not affect the penalty.

We want to be careful to ensure that the penalty $J_{m,\mathbb{S}^2}(u)$ is finite. Proceeding as in Section 2.2.1, we let $\mathcal{H}$ be the space of functions[52] $u\in L^2(\mathbb{S}^2)$ for which $J_{m,\mathbb{S}^2}(u)$ is finite. Define

$$\langle f,g\rangle_{\mathcal{H}} = \mathrm{mean}(f)\mathrm{mean}(g) + \sum_{l=1}^{\infty}\frac{(f)_{n,l}(g)_{n,l}}{l^{-m}(l+1)^{-m}}.$$

Observe that if $u\in\mathcal{H}$ has zero mean, then

$$J_{m,\mathbb{S}^2}(u) = \sum_{l=0}^{\infty}(l(l+1))^m\sum_{n=-l}^{l}(u)_{l,n}^2 = \langle u,u\rangle_{\mathcal{H}} = ||u||_{\mathcal{H}}^2.$$

2.7.2 The Decomposition $\mathcal{H}=\mathcal{H}_0\oplus\mathcal{H}_1$ for the Thin-plate Splines of Order m

Null space $\mathcal{H}_0$: To define a penalty seminorm on the Fourier side, we must define which Fourier components do not contribute to wiggliness. The natural choice, first introduced by Wahba and Wendelberger in the early 1980s [154, 158, 161], is simply to exclude the DC component, since this is the only spherical harmonic annihilated by Δ_S. Other authors have proposed a richer wiggliness penalty seminorm null space and therefore different kernel by allowing for certain spherical polynomial trends to escape penalization [11].

Thus, $\mathcal{H}_0 = \mathrm{span}\,\{1\}$. The space $\mathcal{H}_0$ endowed with the inner product

$$\langle f,g\rangle_{\mathcal{H}_0} = \mathrm{mean}(f)\mathrm{mean}(g) = \frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi}f(\theta,\phi)\sin(\theta)\,\mathrm{d}\theta\,\mathrm{d}\phi\cdot\frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi}g(\theta,\phi)\sin(\theta)\,\mathrm{d}\theta\,\mathrm{d}\phi$$

is trivially an RKHS. The reproducing kernel for $\mathcal{H}_0$ is the constant function 1, since $1\in\mathrm{span}\{1\}$ and

$$\forall f\in\mathcal{H}_0 \text{ and all } x\in\mathbb{S}^2,\ f(x) = \mathrm{mean}(f) = \mathrm{mean}(f)\cdot 1 = \langle f,1\rangle_{\mathcal{H}_0}.$$

[52] The elements of $\mathcal{H}$ are the continuous class representers of equivalence classes of functions that coincide almost everywhere (with respect to the Lebesgue measure) with continuous functions.

The orthogonal projection P_0 onto $\mathcal{H}_0$ consists of extracting the mean, i.e., taking the coefficient $u_{0,0}$ of $u \in \mathcal{H}$ on the DC spherical harmonic Y_0^0.

Wiggly space $\mathcal{H}_1$: In direct analogy with the circular splines, $\mathcal{H}_1$ is the space of zero-mean functions with well-defined wiggliness penalty. That is, their Fourier coefficients decay sufficiently quickly so that the wiggliness penalty is finite. Letting

$$\langle f, g\rangle_{\mathcal{H}_1} = \sum_{l=1}^{\infty}\sum_{n=-l}^{l} \frac{(f)_{n,l}(g)_{n,l}}{l^{-m}(l+1)^{-m}} \text{ (no DC component)},$$

this means that, writing the Fourier coefficient of the expansion of u onto the spherical harmonic of degree l and order n as $(u)_{n,l}$,

$$u \in \mathcal{H}_1 \iff \text{mean}(u) = 0 \text{ and } ||u||_{\mathcal{H}_1}^2 = \sum_{l=1}^{\infty}\sum_{n=-l}^{l} l^m(l+1)^m((u)_{l,n})^2 < \infty.$$

The orthogonal projection P_1 onto $\mathcal{H}_1$ consists of subtracting out the mean. Thus, functions $u \in \mathcal{H}_1$ can be written with an expansion of the form

$$u = \sum_{l=1}^{\infty}\sum_{n=-l}^{l} (u)_{l,n} Y_l^n \text{ (no DC component)},$$

which converges uniformly in $\mathcal{H}_1$ by Proposition 2.24 and the results of Section 2.2.3[53]. Consequently, the penalty $J_{m,\mathbb{S}^2}$ on $\mathcal{H}_1$ exhibits definiteness

$$u \in \mathcal{H}_1 \implies (u)_{0,0}; \text{ then if } J_{m,\mathbb{S}^2}(u) = 0, (u)_{n,l} = 0\,\forall (n,l) \in \mathbb{N}^2 \text{ and } u \equiv 0,$$

since all eigenvalues of the Laplace-Beltrami operator Δ_S on spherical harmonics of nonzero degree are strictly positive.

The orthogonality of $\mathcal{H}_0$ and $\mathcal{H}_1$ is easily shown. For any $f \in \mathcal{H}_1$, $||f||_{\mathcal{H}_0}^2 = \text{mean}(f)^2 = 0$. On the other hand, observe that the constant functions in $\mathcal{H}_0$ have no wiggliness; they satisfy

$$f \in \mathcal{H}_0 \implies ||f||_{\mathcal{H}_1}^2 = 0.$$

$\mathcal{H}_0$ and $\mathcal{H}_1$ are therefore orthogonal, and $\mathcal{H}_0$ is indeed the null space of the norm $||\cdot||_{\mathcal{H}_1}$: if $f \in \mathcal{H}$ and $||f||_{\mathcal{H}_1} = 0$, then $f \in \mathcal{H}_0$. The norm $||\cdot||_{\mathcal{H}_1}$ is in fact positive-definite on $\mathcal{H}_1$: in $\mathcal{H}_1$, only the zero function has norm zero.

Noting that $l^m(l+1)^m$ are the eigenvalues associated with the spherical harmonics of order l of Δ_S^m, using Mercer synthesis (Proposition 2.38), we can obtain the reproducing kernel for $\mathcal{H}_1$ by giving it a Fourier expansion on the spherical harmonics with weights $\alpha_l = l^{-m}(l+1)^{-m}$ (see Section 2.2.3). Using Equation (22), we can identify the reproducing kernel for $\mathcal{H}_1$

$$k_{3,m}(p,p') = \sum_{l=1}^{\infty} \frac{(2l+1)\alpha_l}{4\pi} P_l^0(\cos(\sphericalangle(p,p'))) = \frac{1}{4\pi}\sum_{l=0}^{\infty} \frac{2l+1}{l^m(l+1)^m} P_l^0(\cos(\sphericalangle(p,p'))). \tag{62}$$

By the addition theorem for spherical harmonics (19), with $p = (\theta,\phi)$ and $p' = (\theta',\phi')$,

$$k_{3,m}(p,p') = \sum_{l=1}^{\infty}\sum_{n=-l}^{l} \underbrace{\left(\frac{1}{l^m(l+1)^m} Y_l^n(\theta,\phi)\right)}_{(k_{3,m}(\cdot,p))_{l,n}} Y_l^n(\theta',\phi'),$$

[53] If $m > 1$, the sequence $\{\lambda_{l,n}\}_{l=0}^{\infty}$ given by $\lambda_{n,l} = \alpha_l = l^{-m}(l+1)^{-m}$ satisfy $\{\frac{2l+1}{l^m(l+1)^m}\} \in \ell^1$, and all the results in that section hold. If $m = 1$, the sequence $\{\alpha_l(2l+1)\}_{l=0}^{\infty} = \{\frac{(l+1)^2-l^2}{l(l+1)}\}_{l=0}^{\infty}$ is in $\ell^2 \setminus \ell^1$; as a result, the resulting kernel $k_{3,1}$ has a singularity (at $p = p'$, i.e., $\cos(\sphericalangle(p,p')) = 1$).

and the reproducing property holds

$$\begin{aligned}\forall f \in \mathcal{H}_1 \, \langle f, k_{3,m}(\cdot, p)\rangle_{\mathcal{H}_1} &= \sum_{l=1}^{\infty} \sum_{n=-l}^{l} \frac{(f)_{l,n} \left(\frac{1}{l^m (l+1)^m} Y_l^n(\theta, \phi) \right)}{l^{-m}(l+1)^{-m}} \\ &= \sum_{l=1}^{\infty} \sum_{n=-l}^{l} (f)_{n,l} Y_l^n(\theta, \phi),\end{aligned}$$

which we know converges uniformly to $f(p)$ for $m > 1$ by Proposition 2.38, since f has zero mean.

Here the 3 in the index of $k_{3,m}$ indicates that we are working in $\mathbb{S}^2 \subset \mathbb{R}^3$ (so that our notation agrees with that of [11]).

Since $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$, we can sum the two orthogonal subspace kernels to find the reproducing kernel for the space $\mathcal{H}$ [3]

$$k^m(p, p') = k_0^m(p, p') + k_{3,m}(p, p') = 1 + \frac{1}{4\pi} \sum_{l=0}^{\infty} \frac{2l+1}{l^m (l+1)^m} P_l^0(\cos(\sphericalangle(p, p'))).$$

Pseudocode for solving smoothing problems on the sphere using this kernel is given in the next section.

The notorious uselessness of series expression for (62) derives from its slowness to converge. Fortunately, there are closed-form formulas (in terms of a special function) for $k_{3,m}$ for $m = 2, 3$, and in some cases the series definition (62) can be used in practice. More details on how to compute this kernel are given in the following section.

3 Implementing Thin-plate Splines on the Sphere

The chief implementation challenge of the thin-plate splines on the sphere consists in computing the kernel sum (62) for $m = 2$ and $m = 3$. Wahba suggested that we modify their numerators to yield a topologically equivalent seminorm, more willing to being manipulated into closed form and more amenable to practical use [154] (later corrected in the erratum [155]).

Keller and Borkowski found that, despite its slow convergence, just 40 terms of the sum (62) are needed in practice [77]. Indeed, this seems to be the case with the data set presented in the Example 4.1, as Figure 7g indicates. Nevertheless, closed-form expressions for the kernel simplify many calculations in practice: from the reproducing kernel of an RKHS $\mathcal{H}$, one can obtain the representer of any bounded linear functional on $\mathcal{H}$–not just the representers of evaluation at points on the index set of $\mathcal{H}$–if one can easily apply the bounded linear functional to the kernel's arguments (see Section 2.5). Additionally, even if 40 terms of a slowly converging series that defines the kernel may be sufficient to represent evaluation functionals for interpolation, many more terms may be required to recover a bounded linear function, such as a regional mean or derivative, applied to the interpolant.

While not widely known, there exist closed-form expressions (in terms of the polylogarithm) for some kernels of thin-plate splines.

3.1 Closed-form Formulas for the Reproducing Kernel of Thin-plate Splines

Closed-form expressions (in terms of special functions) for the thin-plate spline on the sphere have been found by Wendelberger, Martinez-Morales, and Beatson and zu Castell.

For the (uninteresting) case of the order-1 splines, it is easy to derive

$$k_{3,1}(p,p') = \frac{1}{4\pi}\sum_{l=1}^{\infty}\frac{2l+1}{l(l+1)}P_l^0(\cos(\sphericalangle(p,p'))) = \frac{-\ln\left(\frac{1-x}{2}\right)-1}{4\pi},$$

where we set $x = \cos(\sphericalangle(p,p'))$. In his thesis [161], Wendelberger derived formulas for $k_{3,2}$ and $k_{3,3}$ (giving them the argument $x = \cos(\sphericalangle(p,p'))$ rather than two points on the sphere)

$$k_{3,2}(x) = \frac{1}{4\pi}\sum_{l=1}^{\infty}\frac{2l+1}{l^2(l+1)^2}P_l(x) = \frac{1-\mathrm{Li}_2(1)+\mathrm{Li}_2\left(\frac{1+x}{2}\right)}{4\pi}, \text{ for } |x|\le 1 \text{ ([161], Corollary 4.4.1);}$$

$$k_{3,3}(x) = \frac{1}{4\pi}\sum_{l=1}^{\infty}\frac{2l+1}{l^3(l+1)^3}P_l(x) = \frac{1}{4\pi}\Bigg(-2+\mathrm{Li}_2(1)+2\mathrm{Li}_3(1)-\mathrm{Li}_2\left(\frac{1+x}{2}\right)+$$
$$\ln\left(\frac{1-x}{2}\right)\mathrm{Li}_2\left(\frac{1-x}{2}\right)-2\mathrm{Li}_3\left(\frac{1-x}{2}\right)\Bigg), \text{ for } |x|<1 \text{ ([161], Corollary 4.5.1).}$$

On the boundary, we take the limiting values: as $x\to+1$, $k_{3,3}(x)\to\frac{2(\zeta(3)-1)}{4\pi}$ and as $x\to-1$, $k_{3,3}(x)\to\frac{\zeta(2)-2}{4\pi}$. While we proved that $k_{3,3}$, like $k_{3,2}$, is continuous on $[-1,1]$ in Section 2.2.3, this closed-form representation of $k_{3,3}$ is not well-defined on the boundary (the product term evaluates to $\infty\cdot 0$ when $x=1$).

Here, Li_s is the polylogarithm [88, 89] of order s

$$\mathrm{Li}_s(z) = \sum_{n=1}^{\infty}\frac{z^n}{n^s},$$

which is valid for arbitrary $s\in\mathbb{C}$ and all $z\in\mathbb{C}$ for which the sum converges, though we consider only $s\in\{2,3\}$ and $z\in[-1,1]$. Its name comes from a recursive relation, which shows that the polylogarithm is a repeated integral of lower orders of itself

$$\mathrm{Li}_{s+1}(z) = \int_0^z \frac{\mathrm{Li}_s(t)}{t}\,\mathrm{d}t, \text{ with } \mathrm{Li}_1(z) = -\ln(1-z).$$

This recursive formula comes in handy when using Theorem 4.2 to estimate derivatives or gradients of a function, with access only to its scattered observations. In our Python-based IPOL demo, we use the `mpmath` package [73] to calculate the polylogarithms; the calculation may be sped up by altering the parameter `mp.dps` to reduce the number of decimal places of precision.

The Wendelberger formulas for $k_{3,2}$ and $k_{3,3}$ agree[54] with those independently derived by Beatson and zu Castell in [11], which are presented in section 6 of that work. Defining $u(x) = \frac{1-x}{2}$, these formulas are as follows:

$$(4\pi)\cdot k_{3,2}(x) = \mathrm{Li}_2\left(1-u(x)\right)+1-\frac{\pi^2}{6} = 1-\frac{\pi^2}{6}+\mathrm{Li}_2\left(\frac{1+x}{2}\right),$$

$$(4\pi)\cdot k_{3,3}(x) = -2\mathrm{Li}_3\left(u(x)\right)-\mathrm{Li}_2\left(1-u(x)\right)+\ln\left(u(x)\right)\mathrm{Li}_2\left(u(x)\right)+2\zeta(3)+\frac{\pi^2}{6}-2$$
$$= -2+\frac{\pi^2}{6}+2\zeta(3)+\ln\left(\frac{1-x}{2}\right)\mathrm{Li}_2\left(\frac{1-x}{2}\right)-\mathrm{Li}_2\left(\frac{1+x}{2}\right)-2\mathrm{Li}_3\left(\frac{1-x}{2}\right).$$

Using an operator defined by Martinez-Morales [96], Beatson and zu Castell found recurrence relations that facilitate the derivation of some thin-plate splines for higher-dimension spheres (and for penalties with different null spaces), although advances in special function theory are likely needed to derive an expression for the sum that produces the order-4 thin-plate spline on $\mathbb{S}^2$, that is, $k_{3,4}$ [11, 161].

[54] Up to the factor of 4π, which comes from the addition theorem and which is omitted in Beatson and zu Castell [11]. Simply note that $\mathrm{Li}_s(1) = \zeta(s)$ and $\zeta(2) = \frac{\pi^2}{6}$.

3.2 Pseudocode for Thin-plate Splines on the Sphere

Smoothing and interpolation thin-plate splines are found by solving the linear system (38). We provide pseudocode for thin-plate splines on the sphere in Algorithm 8.

Algorithm 8: Find the $\{\alpha_i\}_{i=1}^n$ weights on the spline basis functions (representers of evaluation at the n data points) and mean α_0.

Data: A training set consisting of n latitude and longitude values $x_i = (\theta_i, \phi_i) \in [0, \pi] \times [0, 2\pi)$ and n samples $y_i \in \mathbb{R}$, for $i = 1, \ldots, n$. The parameters consist of a regularization penalty $\lambda \geq 0$ and an order $m \in \{2, 3\}$. (Those seeking an interpolator with order $m > 3$ have recourse to the infinite series that defines $k_{3,m}(\cdot)$.)

Result: A global mean value α_0 and basis function weights $\{\alpha_i\}_{i=1}^n$.

Compute the cosine of the spherical angle $\sphericalangle$ between each pair of data points

$$\cos\left(\sphericalangle(x_i, x_j)\right) = \cos(\theta_i)\cos(\theta_j) + \sin(\theta_i)\sin(\theta_j)\cos(\phi_i - \phi_j).$$

Compute the $n \times n$ matrix $\mathbf{K}_1$ in whose ith row and jth column reposes the value

$$(\mathbf{K}_1)_{i\,j} \leftarrow k_{3,m}\left(x_i, x_j\right).$$

(Expressions, in terms of the polylogarithm, for $k_{3,1}$, $k_{3,2}$, and $k_{3,3}$ are given in Section 3.1.)

$$\mathbf{K} \leftarrow \begin{pmatrix} \mathbf{K}_1 + n\lambda\mathbf{I}_{n\times n} & \mathbf{1}_n \\ \mathbf{1}_n^T & 0 \end{pmatrix}.$$

Solve $\mathbf{K}\alpha = y$,

$$\alpha \leftarrow \mathbf{K}^{-1}y.$$

Recover the $\{\alpha_i\}_{i=1}^n$ and mean value α_0:

$\alpha_0 \leftarrow \alpha[-1]$

for i *in* $[\![1, n]\!]$ **do**

  $\alpha_i \leftarrow \alpha[i-1]$;

end

---

4 Using Splines to Compute Integrals and Derivatives and to Solve Inverse Problems

Many approaches to the problem of computing an average of a quantity over a sphere from scattered data are in use [66, 67, 150]. Some require latitude-longitude gridding [51, 103] and others more sophisticated forms of gridding[55]. Irregular meshes and multiscale approaches are put to use for global interpolations and averaging of data that are highly nonstationary over the sphere (such as topography) [54, 67, 72].

While some of these methods possess computational advantages and offer theoretical performance guarantees, they can be difficult to use with scattered data. After all, they reduce what is effectively an interpolation problem (computing a global average requires some understanding of the behavior of the unknown function between the scattered samples) to regridding–another interpolation problem,

[55]For instance, grids derived to be equidistributed and to minimize the discrepancy of the grid to the Laplace-Beltrami operator (an error term based on the Hlawka-Koksma theorem); see [67], Chapter 14.

often no easier. In practice, many real-world global averaging systems incorporate domain-specific knowledge of how the quantity under consideration varies spatially to map scattered data from the spherical surface to an interval, where the averaging problem becomes easier. In Example 4.1, we consider how global measurements of greenhouse gases are produced in practice and compare these estimates with those produced using thin-plate splines on the sphere.

Example 4.1 (Computing averages of scattered data on the sphere)**.** *In this example, we seek to estimate the global average of CO_2 from scattered measurements. On Earth, atmospheric transport via horizontal winds curtails the variance of greenhouse gases across each parallel. As a result, we can perform zonal averaging (average scattered measurements over binned latitudes) or fit a curve to the scattered data values plotted against their latitudes.*

The surface integral is thereby reduced to a single integral, which, thanks to a clever choice of latitude parameterization, can take a particularly simple form. The NOAA GML Carbon Cycle Group computes global averages of surface greenhouse gas concentrations using calibrated, weekly latitude-averaged measurements taken from marine boundary layer air [30, 146]. A curve of greenhouse gas concentration versus sine latitude is fit to weekly measurements and used to compute the global average. If the concentration T depends only on the latitude θ, or this approximation is justified by the atmospheric transport model and distribution of measurement sites relative to sources and sinks of the gas, then we can write

$$\operatorname{mean}(T) = \frac{1}{4\pi}\int_{-\pi/2}^{\pi/2}\int_{-\pi}^{\pi} T(\theta)\cos(\theta)\,\mathrm{d}\phi\,\mathrm{d}\theta = \frac{1}{2}\int_{-\pi/2}^{\pi/2} T(\theta)\cos(\theta)\,\mathrm{d}\theta = \frac{1}{2}\int_{-1}^{1} T\left(x(\theta)\right)\mathrm{d}x\,, \tag{63}$$

after making the substitution $x = \sin(\theta)$. Accordingly, the global mean can be computed by fitting an interpolant to the scattered data plotted in sin(latitude) *and estimating its 1-D integral*[56]*. In the last five rows of Table 2, we apply variations of this technique to all measurements of site monthly averages of flask-air CO_2 available from the Global Monitoring Laboratory network for the most recent month available at time of analysis. The scattered data of CO_2 concentration vs.* sin(latitude) *are used to fit natural cubic splines, Gaussian process regression curves, first-order hold curves, and second-order hold curves to the data, from which the integral (63) is computed. We also use Euclidean thin-plate splines based on the penalty (52) and thin-plate splines on the sphere to compute global averages directly over the sphere simply by calculating the global average of the interpolating surface, over $[0,\pi]\times[0,2\pi)$ and over $\mathbb{S}^2$, respectively. (In the latter case, the global average is given "for free" in the form of α_0.)*

Additional results are given in Figures 7-8.

Because the null space $\mathcal{H}_0$ of the thin-plate spline on the sphere is spanned by $\{1\}$, the parameter α_0 gives the spherical mean. However, for the planar thin-plate spline in the example above, we fit the spline surface and then calculate the mean value of that surface over $[0,\pi]\times[0,2\pi]$, not $\mathbb{R}^2$. Similarly, the Berkeley Earth Surface Temperature project uses Kriging to fit an interpolating surface from scattered data on Earth's surface and integrates the interpolant over the land surface to compute a global average land surface temperature [122]. In many Earth science applications, spatial correlations of an observed parameter between two points tend to follow known transport phenomena, such as lateral winds. Consequently, Kriging methods, using this outside knowledge,

[56] In practice, a low-pass Butterworth filter of order six is used on resampled data, rather than a spline fit to scattered data. Individual measurements are, depending on their quality, replaced with 1-10 measurements, equally spaced in $\sin(\theta)$, so that a digital Butterworth filter may be applied without having to numerically solve a difference equation (taken from the transfer function of the order-6 filter) on scattered data. This process is repeated twice with different cutoff frequencies. For further details, see [146]. In Table 2, we instead use more standard curve-fitting techniques for scattered data.

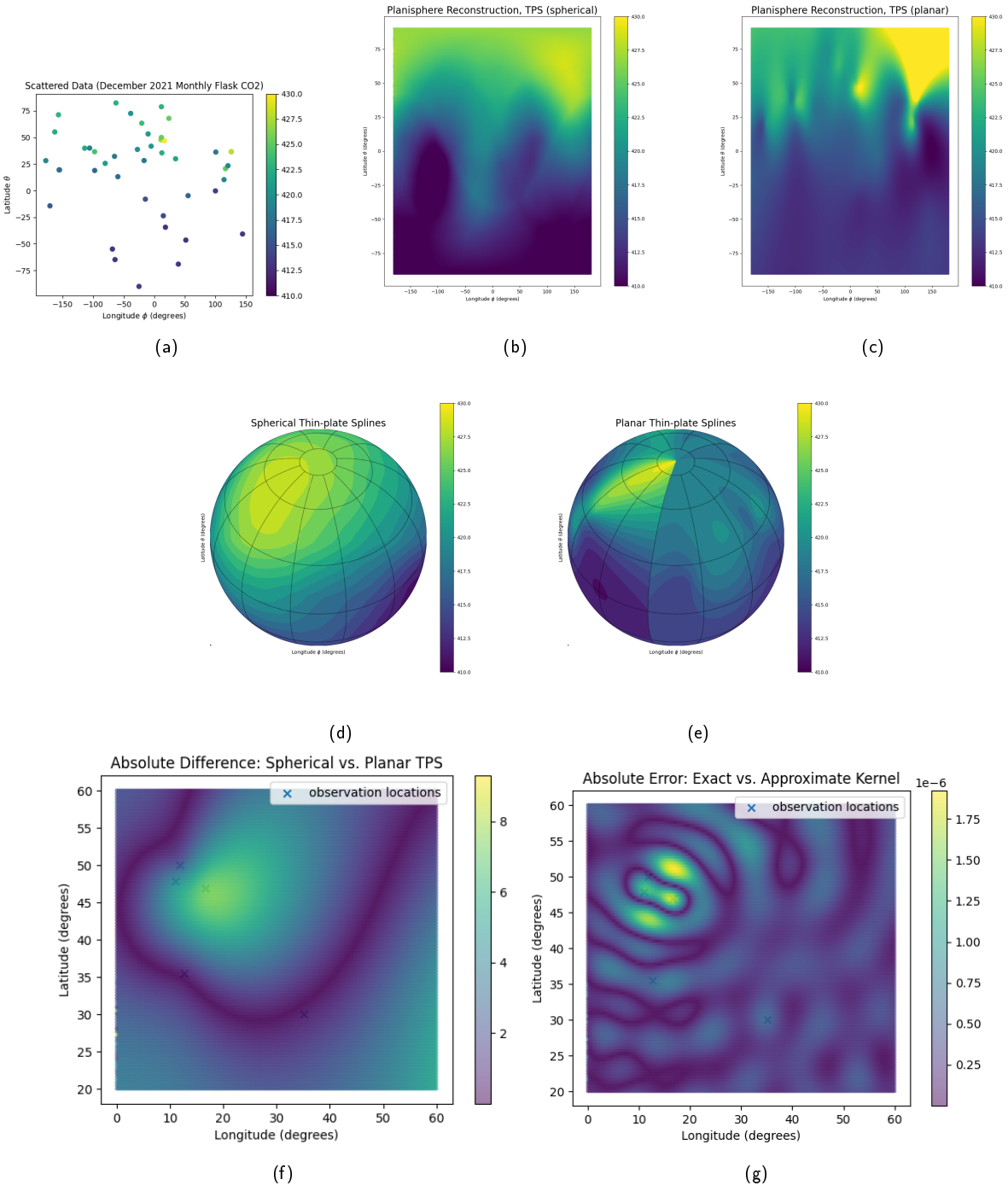
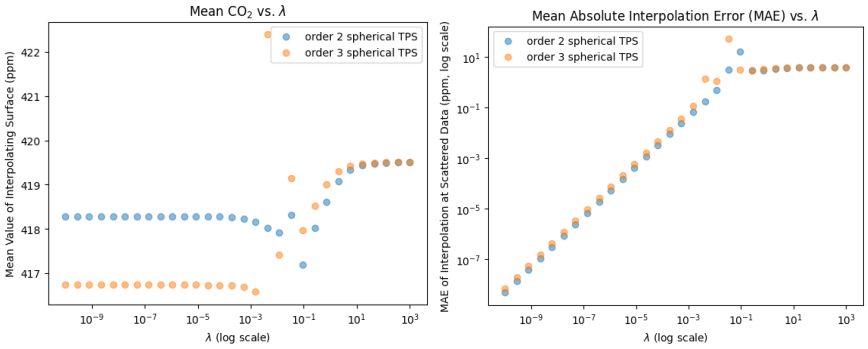


Figure 7: **(7a)** Scattered data of December 2021 monthly averages of flask-air CO_2 readings at 44 sites in the Global Monitoring Laboratory network[57]. **(7b)**-**(7c)** Spherical and planar thin-plate spline fits of these data, respectively, both of order 2. The spherical thin-plate spline interpolant is more coherent with the geometry of the sphere. The planar interpolation surface lacks longitudinal periodicity and embarks on a mean-altering but wiggliness-minimizing excursion beyond the sample points in the top-right corner of **(7c)**. **(7d)**-**(7e)** display samples of the same surfaces as in **(7b)**-**(7c)**, respectively, but these sample points, uniformly spaced in the plane under an equirectangular projection, are mapped back onto a sphere, which throws into relief the meridional discontinuity of planar thin-plate interpolating surface. In the top-right corner of **(7c)**, the interpolating surface, outside the convex hull of the planar control points, loses its curvature. **(7f)**-**(7g)** show the differences between the spherical and planar thin-plate splines, trained on all 44 points, in their reconstruction of a mesh of points between $0°$ E and $60°$ E longitude, and $20°$ N and $60°$ N longitude.

can lead to better global average estimates than thin-plate splines, which encode spatial correlation via geodesic distance.

Method	Global Average, CO_2 (ppm)
Thin-plate (spherical, $\lambda = 0$)	416.73
Thin-plate (planar, $\lambda = 0$)	414.63
Natural cubic ($\lambda = 0.001$)	417.16
Natural cubic ($\lambda = 0.1$)	417.26
Kriging ($\sigma = 1$)	417.23
Trapezoidal approximation of (63)	415.36
Simpson's approximation of (63)	415.56

Table 2: Computation of the integral in (63) using different scattered data interpolation techniques.



(a) (b)

Figure 8: **(8a)** A plot of the global mean CO_2 of the interpolating surface as a function of the regularization parameter λ used in constructing it. With low values of λ, the order 2 and order 3 spherical splines interpolate the data and, due to their different wiggliness penalties, form different interpolating surfaces with different mean concentrations of CO_2 to do so. With high values of λ, the order 2 and order 3 spherical splines agree, as they each yield what is effectively an element of $\mathcal{H}_0 = \operatorname{span}\{1\}$ due to the excessive penalty imposed on wiggliness. With high, but not excessive values λ, erroneous averages may be found. **(8b)** The mean absolute error in interpolation at the data points increases with the regularization penalty λ, until the interpolation surface is essentially constant.

A theorem from Wahba and Kimeldorf [80, 153, 158] situates this seemingly ad-hoc approach–fit a curve to data using a wiggliness penalty, then apply a functional to the curve–in Wiener-Hopf-Kolmogorov linear estimation [109] and Gaussian process [76] theory.

Theorem 4.2 (Wahba [158], Theorem 1.5.2)**.** *We can estimate the posterior mean value of a bounded linear functional* L_0 *applied to a signal given scattered data by applying the functional to the spline fit of the data if we put a Gaussian process prior on the signal.*

Specifically, suppose Y *is a zero-mean Gaussian process over an index set* $\mathcal{X}$ *with covariance* $\mathbb{E}[Y(s)Y(t)] = k^1(s,t)$ *for all* $(s,t) \in \mathcal{X}^2$*. Let*

$$F(s) = \sum_{i=1}^{m} \theta_i \phi_i(s) + b^{1/2} Y(s) \text{ for all } s \in \mathcal{X},$$

where the linearly independent, deterministic basis functions ϕ_i *are known and the parameters* θ_i *and* $b^{1/2}$ *are unknown but fixed. Suppose we have a collection of* n *noisy observations* X_i *of bounded linear functionals* $\mathrm{L}_1, \ldots, \mathrm{L}_n$ *applied to* F *(for instance, evaluation or observation through certain*

[57] The raw data were accessed July 31, 2022, from the GML Data Finder (https://gml.noaa.gov/dv/data/index.php?parameter_name=Carbon%2BDioxide&type=Flask&frequency=Monthly%2BAverages), and the positions of and CO_2 measurements taken at the 44 sites with available December 2021 monthly average data are available in one convenient spreadsheet (https://www.kaggle.com/datasets/maxdunitz/scattered-spherical-data-mean-monthly-co2-dec21/).

instruments at points in $\mathcal{X}$ or means taken over regions in $\mathcal{X}$)

$$X_i = \mathrm{L}_i F + \epsilon_i, \text{ for } i = 1, \dots, n,$$

where the measurement error is independently and identically distributed: $\epsilon_i \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$. Our goal is to estimate $\mathrm{L}_0 F$ given our observations the $X_1, \dots, X_n$. Let $\mathcal{H}_1$ be the RKHS of the kernel k^1, $\mathcal{H}_0 = \operatorname{span}\{\phi_1, \dots \phi_m\}$, $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$ and P_1 be the orthogonal projection from $\mathcal{H}$ onto $\mathcal{H}_1$. Solve the spline smoothing problem (identical to (33), with $\lambda = \sigma^2/(nb)$) using the representer theorem (Proposition 2.60)

$$f^* = \arg\min_{f \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^{n} (X_i - \mathrm{L}_i f)^2 + \frac{\sigma^2}{nb} ||\mathrm{P}_1 f||^2_{\mathcal{H}_1}.$$

Then $\mathrm{L}_0 f^*$ is the minimum variance, linear, unbiased with respect to θ estimate $\widehat{\mathrm{L}_0 F}$ of $\mathrm{L}_0 F$. That is,

$$L_0 f^* = \arg\min_{\beta \in \mathbb{R}^n} \mathbb{E}\left[\widehat{\mathrm{L}_0 F} - \mathrm{L}_0 F\right]^2 \text{ subject to } \widehat{\mathrm{L}_0 F} = \sum_{i=1}^{n} \beta_i x_i \text{ and } \mathbb{E}\left[(\widehat{\mathrm{L}_0 F} - L_0 F)|\theta\right] = 0.$$

Remark 4.3. *Using this theorem, we can recover and extend some classic quadrature rules. (See [38, 80, 81, 158].) In Example 2.6, the representer of evaluation at x was $\mathbb{1}_{\cdot \le x}$ and the kernel associated with the RKHS is $k(x, x') = \langle \mathbb{1}_{\cdot \le x}, \mathbb{1}_{\cdot \le x'} \rangle_{L^2([0,1])} = \min(x, x')$, which is the kernel of a Wiener process (Brownian motion) W. Thus, W is the stationary, zero-mean stochastic process, satisfies $W(0) = 0$ and has stationary, independent increments (see [135, 158]). Given a set of observations at $t_1, \dots, t_n$, the spline smoothing interpolant is the first-order hold, which is easy to see, since, by the representer theorem (Proposition 2.60), $f \in \operatorname{span}\{\mathbb{1}_{\cdot \le t_1}, \dots, \mathbb{1}_{\cdot \le t_n}\}$. From Theorem 4.2, we can see that if we model the prior on a signal f as Brownian motion W, that is, as a Gaussian process with kernel $k(x, x') = \int_0^1 G_1(x, u) G_1(x', u) \, \mathrm{d}u = \min(x, x')$, then our Bayesian posterior mean estimate of $\int_0^1 f(t) \, \mathrm{d}t$ given measurements $\{f(t_i)\}_{i=1}^m$ is the area under the spline fit, which corresponds to the standard trapezoidal rule applied to these measurements. If we model the prior on f as once-integrated[58] Brownian motion, with measurements subject to i.i.d. Gaussian noise, then Bayes's rule is equivalent to fitting a natural cubic spline to the observed data and integrating that [38, 60].*

This theorem, we reiterate, extends beyond Bayesian quadrature (deriving rules to approximate an integral by taking a linear combination of (noisy) function samples) to encapsulate the use of arbitrary bounded linear functional evaluations to approximate another arbitrary bounded linear functional. It associates with each spline smoothing problem a Bayesian estimation problem, where the penalized RKHS space $\mathcal{H}_1$ has the same kernel as the Gaussian process prior's covariance and the unpenalized RKHS space $\mathcal{H}_0$ corresponds to a deterministic process. The duality between RKHS and Gaussian processes–identified by Parzen [112], Wahba's thesis supervisor, using work from Loève,

[58] The $m-1$th integrated Wiener process $X_{m-1}(t) = \int_0^1 G_m(t, u) \, \mathrm{d}W(u)$, where $G_m(t, u) = \frac{(t-u)_+^{m-1}}{(m-1)!}$ as defined in Section 2.6.1. Thus, the once-integrated Wiener process is $X_1(t) = \int_0^1 G_2(t, u) \, \mathrm{d}W(u) = \int_0^1 (t-u)_+ \, \mathrm{d}W(u)$. From the stationary independent increments property of Wiener processes, it can be seen that the $m-1$th integrated process X_{m-1} has covariance

$$\mathbb{E}[X_{m-1}(s) X_{m-1}(t)] = \mathbb{E}\left[\int_0^1 G_m(s, u) \, \mathrm{d}W(u) \int_0^1 G_m(t, u) \, \mathrm{d}W(u)\right] = \int_0^1 G_m(s, u) G_m(t, u) \, \mathrm{d}W(u) = k^1(s, t),$$

where k^1 is the reproducing kernel for the space $\mathcal{H}_1$ (defined in (41)) associated with the natural polynomial splines of order m, whose squared norm is $\int_0^1 (f^{(m)}(x))^2 \, \mathrm{d}x$. For the once-integrated Wiener process X_1, $\mathbb{E}[X_1(x) X_1(y)] = xy \min(x, y) - \frac{x+y}{2} \min(x, y)^2 + \frac{1}{3} \min(x, y)^3$ (see Algorithm 4).

Parzen's supervisor–can be used as a confidence check for our modeling assumptions. In this article, we started with wiggliness penalty seminorm and identified the corresponding spaces $\mathcal{H}_0$ and $\mathcal{H}_1$ and the kernel k^1; do these choices seem reasonable when framed as probabilistic assumptions? Additionally, the theorem can aid the design or deployment of measurement instruments to survey a natural process that can profitably be modeled as a Gaussian process.

Important technical clarifications are given in the note by Angelika van der Linde-Ploumbidis [152]; the posterior variances of the Bayesian estimates of a linear functional given observations of other linear functionals are particularly influenced by modeling assumptions. For example, one can arrive at different error bars on the value of an integral even if the posterior mean is given by the same quadrature rule. A comprehensive, modern review of the connections between Gaussian processes and kernel methods is given in [76]; see also [145].

5 IPOL Demo

The IPOL demo[59] takes as input a `csv` file with three columns: one called `‘‘latitudes’’`, which stores string representations of floating-point numbers corresponding to each observation's latitude (degrees in $[-90, 90]$); one called `‘‘longitudes’’`, which gives the observations' longitudes (degrees in $[-180, 180)$); and one called `‘‘observations’’`, which stores the real-valued samples to interpolate. Alternatively, one can give a `png` image, which will be interpreted as samples of the grayscale image value at regularly spaced points under an equirectangular projection, and from which 200 samples will be selected uniformly at random[60] to construct the thin-plate spline.

Due to computational time constraints on the demo server, if a `csv` file with more than $N = 200$ samples is provided, the demo randomly selects a subset of N data points and proceeds. As implemented, the limiting computational step is not the inversion of the modified Gram matrix $\mathbf{K}$ used to solve the linear system (38) but rather the evaluations of the polylogarithm (to 15 decimal places of precision, using `mpmath`) and their conversion to floating point numbers. To speed up the demo, we precompute the values of $\mathrm{Li}_3(x)$ and $\mathrm{Li}_2(x)$ for $x \in$ `linspace(-1,1,2e6+1)` and quantize $\cos(\gamma)$ to the nearest multiple of `1e-6` before computing $\mathrm{Li}_s(\cos(\gamma))$. As with the modified Gram matrix's inverse, these polylogarithm evaluations can be precomputed in applications with a fixed set of measurement locations (or, as we have done, an arbitrary set of measurement locations quantized to fixed precision) and a fixed set of points at which to evaluate the resulting interpolating surface (or its image after applying a bounded linear operator).

5.1 Comparison with Planar Thin-plate Spline Interpolant

We also output, for the sake of comparison, the result of a thin-plate spline smoother (Algorithm 7) of order 2 (that is, using $J_{2,\mathbb{R}^2}$ in Equation (52)) for the same penalty parameter λ and the image displaying the pixelwise error between the two methods. We compute the spherical mean of the interpolating surface f

$$\frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi} f(\theta,\phi)\sin(\theta)\,\mathrm{d}\theta\,.\,\mathrm{d}\phi \tag{64}$$

Using the order-2 planar thin-plate spline expansion

$$f(\theta,\phi) = \alpha_0 + \alpha_1\theta + \alpha_2\phi + \sum_{i=1}^{N}\alpha_{i+2}\left\|(\theta,\phi)-(\theta_i,\phi_i)\right\|_{\mathbb{R}^2}^2 \ln\left\|(\theta,\phi)-(\theta_i,\phi_i)\right\|_{\mathbb{R}^2},$$

[59] https://doi.org/10.5201/ipol.2026.451

[60] We take random vectors in $\mathbb{R}^3$, project them on the sphere (if the norm is not too small), and find the corresponding pixel in which it lies.

we compute the interpolant's spherical mean (64) by noting

$$\begin{aligned}\frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi}\alpha_0\sin(\theta)\,\mathrm{d}\theta\,\mathrm{d}\phi &= \alpha_0\\ \frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi}\alpha_1\theta\sin(\theta)\,\mathrm{d}\theta\,\mathrm{d}\phi &= \frac{\pi}{2}\alpha_1\\ \frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi}\alpha_2\phi\sin(\theta)\,\mathrm{d}\theta\,\mathrm{d}\phi &= \pi\alpha_2,\end{aligned}$$

and precomputing on quantized measurement locations $(\theta_i, \phi_i) \in [0, \pi] \times [0, 2\pi]$

$$\frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi}\|(\theta,\phi)-(\theta_i,\phi_i)\|_{\mathbb{R}^2}^2\ln\|(\theta,\phi)-(\theta_i,\phi_i)\|_{\mathbb{R}^2}\sin(\theta)\,\mathrm{d}\theta\,\mathrm{d}\phi = w_i.$$

The integral (64) may be given as a linear combination of the weights α learned in fitting the spline with Algorithm 7

$$\frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi}f(\theta,\phi)\sin(\theta)\,\mathrm{d}\theta\,\mathrm{d}\phi = \alpha_0+\frac{\pi}{2}\alpha_1+\pi\alpha_2+\sum_{i=1}^{N}w_i\alpha_{i+2}. \tag{65}$$

We use this procedure to output the mean value of the interpolating surface.

5.2 Comparison with Natural Cubic Spline Interpolant of Data Versus Sine Latitude

Finally, we use a natural cubic spline fit (Algorithm 4) to fit the observations, plotted in one dimension, against sine latitude, as in Example 4.1. Note that these sine latitude data are defined on $[-1, 1]$, not $[0, 1]$, as was the case in Algorithm 4. We adapt that algorithm with a reparameterization; see Remark 2.62 for details. The basis functions of $\mathcal{H}_0$ are mapped to 1 and $1 + x$ and the kernel of the space of wiggly functions

$$k^1_{[0,1]}(x,y) = xy\min(x,y) - \frac{x+y}{2}\min(x,y)^2 + \frac{1}{3}\min(x,y),$$

becomes

$$k^1_{[-1,1]}(x,y) = xy(\min(x,y)+1) - \frac{x+y}{2}(\min(x,y)^2+1) + \frac{1}{3}(\min(x,y)^3+1).$$

Integrating this natural cubic spline fit, we estimate the global mean of the inputted data. In some use cases, this is actually quite reasonable: horizontal winds ensure roughly constant measurement values at each latitude parallel; if this assumption does not hold, this approach to estimating the global mean may still be of use if the measurements are sufficiently rich and varied that the spline fit approximates the average observation value at each (sine) latitude parallel.

We precompute the area under the representers of evaluation at x_i of the natural cubic spline on $[-1, 1]$,

$$\int_{-1}^{1}k^1(x,x_i)\,\mathrm{d}x = \frac{1}{24}x_i^4-\frac{1}{6}x_i^3+\frac{1}{4}x_i^2+\frac{7}{6}x_i+\frac{17}{24},$$

so that the integral (63) can be computed from the natural cubic spline fit u term by term

$$u(\sin(\theta)) = \alpha_0+\alpha_1(1+\sin(\theta))+\sum_{i=1}^{N}\alpha_{i+1}k^1(\sin(\theta),\sin(\theta_i)),$$

and hence

$$\text{mean}(u(\sin(\theta))) = \alpha_0 + \alpha_1 + \sum_{i=1}^{N} \frac{\alpha_{i+1}}{2} \left(\frac{1}{24}\sin(\theta_i)^4 - \frac{1}{6}\sin(\theta_i)^3 + \frac{1}{4}\sin(\theta_i)^2 + \frac{7}{6}\sin(\theta_i) + \frac{17}{24} \right). \quad (66)$$

The demo places each spherical mean estimate on the image corresponding to the interpolant used to produce the estimate.

6 Discussion

6.1 Challenges

Kernel methods for interpolation, smoothing, and solving inverse problems based on scattered data are useful only insofar as (38) can be solved in practice, which means their scalability is limited by the size and conditioning of the Gram matrices (whose size grows quadratically with the number of data points). While solving (38) clearly poses challenges for machine learning practitioners working with large data sets, those seeking, for instance, to interpolate or compute averages of sparse measurements taken over a sphere or region thereof–a common task in geosciences [67] and graphics [23]–are likely to arrive at a solution to (38) without difficulty.

Classic techniques for dealing with large ill-conditioned Gram matrices include taking care to choose an appropriate solver [59] for (38), using only a random subset of the data, and finding low-rank approximations of the Gram matrix. For the latter task, the Nyström method is a common choice and implemented in popular software packages such `scikit-learn` [53, 113]. Wood suggested using the Lanczos algorithm, though this too poses numerical stability challenges [167].

New techniques for finding low-rank approximations of $n \times n$ Gram matrices or $k \times k$ Gram matrices, with $k' < n$, that perform well are contributing to a resurgence in the popularity of RKHS interpolation methods, such as Gaussian process interpolation [9, 28, 33, 115, 165]. For instance, rather than forming a Gram matrix $(\mathbf{K})_{i,j} = k(x_i, x_j)$ from a data set $\{x_i\}_{i=1}^n$ and $\{y_i\}_{i=1}^n$, one uses automatic differentiation to choose a set of control points $\{x'_j\}_{j=1}^{k'}$ and $\{y'_j\}_{j=1}^{k'}$–not necessarily among the data one has access to–such that spline fit learned by solving (38), namely

$$\sigma(x) = \sum_{j=1}^{m} d_j \phi_j(x) + \sum_{i=1}^{k'} c_i k^1(x, x'_i),$$

to these control points minimizes some loss on the data set one has access to. The Gaussian process literature calls the sample locations found via such an approach "inducing variables" [65]. In practice, it is much easier to optimize over $\{x'_i\}_{i=1}^{k'}$, $\{y'_i\}_{i=1}^{k'}$, and λ than $\{x'_i\}_{i=1}^{k'}$, c, and d, due to the greater interpretability (unless $\mathbf{K}$ is exceptionally well-conditioned more reasonable gradients) of the former. When trying to fit splines to many examples, it can be more efficient to use neural networks to learn the inducing points from the examples[61].

Such techniques are useful not just after data acquisition but can be used prospectively, with domain knowledge or simulated data, to optimize the design of instrument arrays or the deployment of sensors in the field. When considering the problem of estimating an interpolating surface of greenhouse gas concentrations across a spherical model of Earth, knowledge of areas with significant greenhouse gas exchange or high variance can be used to identify new sensor locations.

[61] We update Isaac Jacob Schoenberg's [130] dictum that "polynomials are wonderful even after they are cut into pieces, but the cutting must be done with care". The cutting can be done with `keras`.

6.2 Extensions

Discontinuities–for instance, along coastlines, mountain ranges, or fronts–can be encoded via the decomposition principle, explained in detail in [157]. Other forms of nonlinearities–present in the observation functionals in (33), rather than the native space $\mathcal{H}$–are discussed in [110, 157, 169]. Convex constraints such as monotonicity can be incorporated into the smoothing problem (33) with techniques such as [10, 166].

Splines for vector-valued functions on the sphere have also been considered; see [138] for an application to estimating Earth's magnetic field based on scattered measurements. Moreover, splines have been adapted to sphere-like surfaces, most often using mesh methods in practice, though a wide variety of techniques are available [2, 35, 45, 92].

6.3 Other Splines on the Sphere

Of course, thin-plate splines (and Wahba's approximation thereof) are far from being the only splines from which scattered data fitting applications can profit. A veritable zoo of such functions can be found in Chapter 10.6 of [78], along with further development of RKHS theory, as well as in [43]. The characterization and study of positive-definite [129] and strictly positive-definite functions [171, 26] on the sphere remains an active research area, with many applications to machine learning (as these are the correlation functions of isotropic Gaussian processes) [12, 56, 71, 108, 170].

6.4 Other Implementations

Wood has implemented several splines on the sphere, including the thin-plate "pseudo-splines" in [154], for the `R` programming language in the library `mgcv` [168].

Acknowledgment

This work was funded in part by the Region of Île-de-France and the Fondation Mathématique Jacques Hadamard; their support of ML Briefs and my thesis are greatly appreciated. Thank you to the organizers of the ML Briefs workshop, in which I began this article (neither ML nor brief), for their dedication to reproducible and accessible research and patience with the participants' myriad software questions, and to my advisors Miguel Colom and Jean-Michel Morel for their guidance, wisdom, and confidence. Thank you also to Julien Mairal and Jean-Philippe Vert, whose MVA course sparked my interest in the material, and to Grace Wahba, whose open course materials and classic text aided my quest to better understand splines. Above all, I wish to acknowledge the contributions of Jim Wendelberger, not just for his scrupulous attention to detail and heroic devotion exhibited in computing the penalty sums ([161], corollaries 4.4.1 and 4.5.1) with extremely limited access to computational resources, but also for his sharing with me the thesis chapter in which he summarizes these efforts.

References

[1] J. Ahlberg, E. Nilson, and J. Walsh, *Complex Cubic Splines*, Transactions of the American Mathematical Society, 129 (1967), pp. 391–413, https://doi.org/10.2307/1994597.

[2] P. Alfeld, M. Neamtu, and L. L. Schumaker, *Fitting Scattered Data on Sphere-Like Surfaces Using Spherical Splines*, Journal of Computational and Applied Mathematics, 73 (1996), pp. 5–43, https://doi.org/10.1016/0377-0427(96)00034-9.

[3] N. Aronszajn, *La Théorie des Noyaux Reproduisants et ses Applications Première Partie*, in Mathematical Proceedings of the Cambridge Philosophical Society, vol. 39.3, Cambridge University Press, 1943, pp. 133–153, https://doi.org/10.1017/S0305004100017813.

[4] ——, *Theory of Reproducing Kernels*, Transactions of the American Mathematical Society, 68 (1950), pp. 337–404, https://doi.org/10.1090/S0002-9947-1950-0051437-7.

[5] M. Attéia, *Existence et Détermination des Fonctions Spline À Plusieurs Variables*, Comptes Rendus Hebdomadaires des Séances de L'Académie des Sciences : Série A, 262 (1966), p. 575.

[6] ——, *Fonctions « Spline » et Noyaux Reproduisants D'Aronszajn-Bergman*, Revue Française D'informatique et de Recherche Opérationnelle. Série Rouge, 4 (1970), pp. 31–43.

[7] M. Attéia and J. Gaches, *Approximation Hilbertienne : Splines - Ondelettes - Fractales*, Grenoble Sciences, 1999.

[8] V. Balek, *Mechanics of Plates*, ArXiv Preprint ArXiv:1110.2050, (2011), https://doi.org/10.48550/arXiv.1110.2050.

[9] S. Bartels and P. Hennig, *Conjugate Gradients for Kernel Machines.*, Journal of Machine Learning Research, 21 (2020), pp. 55–1. https://dl.acm.org/doi/abs/10.5555/3455716.3455771.

[10] X. Bay, L. Grammont, and H. Maatouk, *A New Method for Interpolating in a Convex Subset of a Hilbert Space*, Computational Optimization and Applications, 68 (2017), pp. 95–120, https://doi.org/10.1007/s10589-017-9906-9.

[11] R. K. Beatson and W. zu Castell, *Thinplate Splines on the Sphere*, SIGMA. Symmetry, Integrability and Geometry: Methods and Applications, 14 (2018), p. 083, https://doi.org/10.3842/SIGMA.2018.083.

[12] R. K. Beatson, W. zu Castell, and Y. Xu, *A Pólya Criterion for (Strict) Positive-Definiteness on the Sphere*, IMA Journal of Numerical Analysis, 34 (2014), pp. 550–568, https://doi.org/10.1093/imanum/drt008.

[13] M. Belkin and P. Niyogi, *Convergence of Laplacian Eigenmaps*, Advances in Neural Information Processing Systems, 19 (2006), https://doi.org/10.7551/mitpress/7503.003.0021.

[14] ——, *Towards a Theoretical Foundation for Laplacian-Based Manifold Methods*, Journal of Computer and System Sciences, 74 (2008), pp. 1289–1308, https://doi.org/10.1016/j.jcss.2007.08.006.

[15] S. Bergman and M. Schiffer, *Kernel Functions and Conformal Mapping*, Compositio Mathematica, 8 (1951), pp. 205–249, https://doi.org/10.1090/surv/005.

[16] M. Bergounioux, *On Poincaré-Wirtinger Inequalities in Spaces of Functions of Bounded Variation*, Control and Cybernetics, 40 (2011), pp. 921–930.

[17] A. Berlinet and C. Thomas-Agnan, *Reproducing Kernel Hilbert Spaces in Probability and Statistics*, Springer Science & Business Media, 2011, https://doi.org/10.1007/978-1-4419-9096-9.

[18] S. Bochner, *Monotone Funktionen, Stieltjessche Integrale Und Harmonische Analyse*, Math. Ann, 108 (1933), pp. 378–410.

[19] S. Bochner et al., *Lectures on Fourier Integrals*, vol. 42, Princeton University Press, 1959, https://doi.org/10.1515/9781400881994.

[20] J. Bognár, *Indefinite Inner Product Spaces*, vol. 78, Springer Science & Business Media, 2012, https://doi.org/10.1007/978-3-642-65567-8.

[21] A. Bouhamidi and A. Le Méhauté, *Radial Basis Functions Under Tension*, Journal of Approximation Theory, 127 (2004), pp. 135–154, https://doi.org/10.1016/B978-0-12-079050-0.50015-1.

[22] F. Brafman, *A Generating Function for Associated Legendre Polynomials*, The Quarterly Journal of Mathematics, 8 (1957), pp. 81–83, https://doi.org/10.1093/qmath/10.1.156.

[23] S. R. Buss and J. P. Fillmore, *Spherical Averages and Applications to Spherical Splines and Interpolation*, ACM Transactions on Graphics (TOG), 20 (2001), pp. 95–126, https://doi.org/10.1145/502122.502124.

[24] S. Canu, C. S. Ong, and X. Mary, *Splines with Non Positive Kernels*, in More Progresses in Analysis, World Scientific, 2009, pp. 163–173, https://doi.org/10.1142/9789812835635_0015.

[25] C.-T. Chen, *Linear System Theory and Design*, Saunders College Publishing, 1984.

[26] D. Chen, V. Menegatto, and X. Sun, *A Necessary and Sufficient Condition for Strictly Positive Definite Functions on Spheres*, Proceedings of the American Mathematical Society, 131 (2003), pp. 2733–2740, https://doi.org/10.1090/S0002-9939-03-06730-3.

[27] Y. Chen, G. Conforti, and T. T. Georgiou, *Measure-Valued Spline Curves: An Optimal Transport Viewpoint*, SIAM Journal on Mathematical Analysis, 50 (2018), pp. 5947–5968, https://doi.org/10.1137/18M1166249.

[28] J.-P. Chilès and N. Desassis, *Fifty Years of Kriging*, in Handbook of Mathematical Geosciences, Springer, Cham, 2018, pp. 589–612, https://doi.org/10.1007/978-3-319-78999-6_29.

[29] M. Contino, A. Maestripieri, and S. Marcantognini, *Operator Least Squares Problems and Moore-Penrose Inverses in Krein Spaces*, Integral Equations and Operator Theory, 90 (2018), pp. 1–23, https://doi.org/10.1007/s00020-018-2456-4.

[30] T. J. Conway, P. P. Tans, L. S. Waterman, K. W. Thoning, D. R. Kitzis, K. A. Masarie, and N. Zhang, *Evidence for Interannual Variability of the Carbon Cycle from the National Oceanic and Atmospheric Administration/Climate Monitoring and Diagnostics Laboratory Global Air Sampling Network*, Journal of Geophysical Research: Atmospheres, 99 (1994), pp. 22831–22855, https://doi.org/10.1029/94JD01951.

[31] K. Crane, *Discrete Differential Geometry: An Applied Introduction*, Notices of the AMS, Communication, (2018), pp. 1153–1159.

[32] P. Craven and G. Wahba, *Smoothing Noisy Data with Spline Functions*, Numerische Mathematik, 31 (1978), pp. 377–403, https://doi.org/10.1007/BF01404567.

[33] N. Cressie and G. Johannesson, *Fixed Rank Kriging for Very Large Spatial Data Sets*, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 70 (2008), pp. 209–226, https://doi.org/10.1111/j.1467-9868.2007.00633.x.

[34] F. Cucker and S. Smale, *On the Mathematical Foundations of Learning*, Bulletin of the American Mathematical Society, 39 (2002), pp. 1–49, https://doi.org/10.1090/S0273-0979-01-00923-5.

[35] L. Cui, X. Qi, C. Wen, N. Lei, X. Li, M. Zhang, and X. Gu, *Spherical Optimal Transportation*, Computer-Aided Design, 115 (2019), pp. 181–193, https://doi.org/10.1016/j.cad.2019.05.024.

[36] L. Debnath and P. Mikusinski, *Introduction to Hilbert Spaces with Applications*, Academic Press, 2005.

[37] J. Deny and J.-L. Lions, *Les Espaces du Type de Beppo Levi*, in Annales de L'institut Fourier, vol. 5, 1954, pp. 305–370, https://doi.org/10.5802/aif.55.

[38] P. Diaconis, *Bayesian Numerical Analysis*, Statistical Decision Theory and Related Topics IV, 1 (1988), pp. 163–175.

[39] J. Duchon, *Interpolation des Fonctions de Deux Variables Suivant le Principe de la Flexion des Plaques Minces*, Revue Française D'automatique, Informatique, Recherche Opérationnelle. Analyse Numérique, 10 (1976), pp. 5–12, https://doi.org/10.1051/m2an/197610R300051.

[40] ——, *Splines Minimizing Rotation-Invariant Semi-Norms in Sobolev Spaces*, in Constructive Theory of Functions of Several Variables, Springer, 1977, pp. 85–100, https://doi.org/10.1007/BFb0086566.

[41] D. Eberly, *Thin Plate Splines*, Geometric Tools Inc, 2002 (2002), p. 116.

[42] P. Ehrenfest and H. K. Onnes, *Simplified Deduction of the Formula from the Theory of Combinations which Planck Uses as the Basis of his Radiation Theory*, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 29 (1915), pp. 297–301, https://doi.org/10.1080/14786440208635308.

[43] U. Elahi, Z. Khalid, and R. A. Kennedy, *On the Choice of Kernel for Signal Interpolation on the Sphere Using Reproducing Kernel Hilbert Spaces*, in 2017 11th International Conference on Signal Processing and Communication Systems (ICSPCS), IEEE, 2017, pp. 1–7, https://doi.org/10.1109/ICSPCS.2017.8270506.

[44] L. C. Evans, *Partial Differential Equations*, vol. 19, American Mathematical Society, 2010, https://doi.org/10.1090/gsm/019.

[45] G. E. Fasshauer, *Meshfree Approximation Methods with MATLAB*, vol. 6, World Scientific, 2007, https://doi.org/10.1142/6437.

[46] J. Ferreira and V. Menegatto, *Eigenvalues of Integral Operators Defined by Smooth Positive Definite Kernels*, Integral Equations and Operator Theory, 64 (2009), pp. 61–81, https://doi.org/10.1007/s00020-009-1680-3.

[47] E. Fischer, *Applications D'un Théorème Sur la Convergence en Moyenne*, Comptes Rendus de L'Académie des Sciences, 104 (1907), pp. 1148–1151.

[48] ——, *Sur la Convergence en Moyenne*, Comptes Rendus de L'Académie des Sciences, 104 (1907), p. 1022–1024.

[49] G. B. Folland, *Introduction to Partial Differential Equations*, Princeton University Press, 1995, https://doi.org/10.2307/j.ctvzsmfgn.

[50] P. Funk, *Beiträge Zur Theorie Der Kugelfunktionen*, Mathematische Annalen, 77 (1915), pp. 136–152.

[51] M. Ganesh and H. N. Mhaskar, *Matrix-Free Interpolation on the Sphere*, SIAM Journal on Numerical Analysis, 44 (2006), pp. 1314–1331, https://doi.org/10.1137/050624005.

[52] P. Garrett, *Harmonic Analysis on Spheres*. https://www-users.cse.umn.edu/~garrett/m/mfms/notes_2013-14/09_spheres.pdf, 2014.

[53] B. Ghojogh, A. Ghodsi, F. Karray, and M. Crowley, *Multidimensional Scaling, Sammon Mapping, and Isomap: Tutorial and Survey*, ArXiv Preprint ArXiv:2009.08136, (2020), https://doi.org/10.48550/arXiv.2009.08136.

[54] Q. L. Gia, I. H. Sloan, and H. Wendland, *Multiscale Analysis in Sobolev Spaces on the Sphere*, SIAM Journal on Numerical Analysis, 48 (2010), pp. 2065–2090, https://doi.org/10.1137/090774550.

[55] J. I. Giribet, A. Maestripieri, and F. M. Pería, *Abstract Splines in Krein Spaces*, Journal of Mathematical Analysis and Applications, 369 (2010), pp. 423–436, https://doi.org/10.1016/j.jmaa.2010.03.016.

[56] T. Gneiting, *Strictly and Non-Strictly Positive Definite Functions on Spheres*, Bernoulli, 19 (2013), pp. 1327–1349, https://doi.org/10.3150/12-BEJSP06.

[57] C. Godsil and G. F. Royle, *Algebraic Graph Theory*, vol. 207, Springer Science & Business Media, 2001, https://doi.org/10.1007/978-1-4613-0163-9.

[58] P. Green and B. Silverman, *Nonparametric Regression and Linear Models: A Roughness Penalty Approach*, 1994.

[59] A. Greenbaum, *Iterative Methods for Solving Linear Systems*, SIAM, 1997, https://doi.org/10.1137/1.9781611970937.

[60] C. Gu and Y.-J. Kim, *Penalized Likelihood Regression: General Formulation and Efficient Approximation*, Canadian Journal of Statistics, 30 (2002), pp. 619–628, https://doi.org/10.2307/3316100.

[61] R. L. Harder and R. N. Desmarais, *Interpolation Using Surface Splines*, Journal of Aircraft, 9 (1972), pp. 189–191, https://doi.org/10.2514/3.44330.

[62] R. Harte and M. Mbekhta, *Generalized Inverses in C*-Algebras*, Studia Mathematica, 106 (1993), pp. 129–138.

[63] E. Hecke, *Über Orthogonal-Invariante Integralgleichungen*, Mathematische Annalen, 78 (1917), pp. 398–404, https://doi.org/10.1007/BF01457114.

[64] S. Helgason, *Groups and Geometric Analysis: Integral Geometry, Invariant Differential Operators, and Spherical Functions*, vol. 83, American Mathematical Society, 2022.

[65] J. Hensman, N. Fusi, and N. D. Lawrence, *Gaussian Processes for Big Data*, ArXiv Preprint ArXiv:1309.6835, (2013), https://doi.org/10.48550/arXiv.1309.6835.

[66] K. Hesse, I. H. Sloan, and R. S. Womersley, *Numerical Integration on the Sphere*, in Handbook of Geomathematics, Springer, 2010, https://doi.org/10.1007/978-3-642-01546-5_40.

[67] K. Hesse, I. H. Sloan, and R. S. Womersley, *Numerical Integration on the Sphere*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2015, pp. 2671–2710, https://doi.org/10.1007/978-3-642-54551-1_40.

[68] E. P. Hsu, *Stochastic Analysis on Manifolds*, American Mathematical Society, 2002, https://doi.org/10.1090/gsm/038.

[69] E. Isaacson and H. B. Keller, *Analysis of Numerical Methods*, Courier Corporation, 2012, https://doi.org/10.1063/1.3034415.

[70] J. D. Jackson, *Classical Electrodynamics*, 1999, https://doi.org/10.1063/1.3057859.

[71] J. Jäger et al., *A Note on the Derivatives of Isotropic Positive Definite Functions on the Hilbert Sphere*, SIGMA. Symmetry, Integrability and Geometry: Methods and Applications, 15 (2019), p. 081, https://doi.org/10.3842/SIGMA.2019.081.

[72] K. Jetter, J. Stöckler, and J. Ward, *Error Estimates for Scattered Data Interpolation on Spheres*, Mathematics of Computation, 68 (1999), pp. 733–747, https://doi.org/10.1090/S0025-5718-99-01080-7.

[73] F. Johansson et al., *Mpmath: a Python Library for Arbitrary-Precision Floating-Point Arithmetic (Version 0.18)*, 2013. http://code.google.com/p/mpmath.

[74] K. Jörgens, *Linear Integral Operators*, vol. 13, Pitman Advanced Publishing Program, 1982.

[75] H. Kalf, *On the Expansion of a Function in Terms of Spherical Harmonics in Arbitrary Dimensions.*, Bulletin of the Belgian Mathematical Society-Simon Stevin, 2 (1995), pp. 361–380, https://doi.org/10.36045/bbms/1103408694.

[76] M. Kanagawa, P. Hennig, D. Sejdinovic, and B. K. Sriperumbudur, *Gaussian Processes and Kernel Methods: A Review on Connections and Equivalences*, ArXiv Preprint ArXiv:1807.02582, (2018), https://doi.org/10.48550/arXiv.1807.02582.

[77] W. Keller and A. Borkowski, *Thin Plate Spline Interpolation*, Journal of Geodesy, 93 (2019), pp. 1251–1269, https://doi.org/10.1007/s00190-019-01240-2.

[78] R. A. Kennedy and P. Sadeghi, *Hilbert Space Methods in Signal Processing*, Cambridge University Press, 2013, https://doi.org/10.1017/CBO9780511844515.

[79] G. Kimeldorf and G. Wahba, *Some Results on Tchebycheffian Spline Functions*, Journal of Mathematical Analysis and Applications, 33 (1971), pp. 82–95, https://doi.org/10.1016/0022-247X(71)90184-3.

[80] G. S. Kimeldorf and G. Wahba, *A Correspondence Between Bayesian Estimation on Stochastic Processes and Smoothing by Splines*, The Annals of Mathematical Statistics, 41 (1970), pp. 495–502, https://doi.org/10.1214/aoms/1177697089.

[81] ——, *Spline Functions and Stochastic Processes*, Sankhyā: The Indian Journal of Statistics, Series A, (1970), pp. 173–180.

[82] A. Kolmogorov and S. Fomin, *Introductory Real Analysis*, Courier Corporation, 1975.

[83] R. I. Kondor and J. Lafferty, *Diffusion Kernels on Graphs and Other Discrete Structures*, in Proceedings of the 19th International Conference on Machine Learning, vol. 2002, 2002, pp. 315–322.

[84] T. W. Körner, *Fourier Analysis*, Cambridge University Press, 2022, https://doi.org/10.1017/9781009230063.

[85] S. Krein, *Linear Equations in Banach Spaces*, Springer, 1982, https://doi.org/10.1007/978-1-4684-8068-9.

[86] T. Kühn, *Eigenvalues of Integral Operators Generated by Positive Definite Hölder Continuous Kernels on Metric Compacta*, in Indagationes Mathematicae (Proceedings), vol. 90.1, Elsevier, 1987, pp. 51–61, https://doi.org/10.1016/S1385-7258(87)80006-9.

[87] L. Kuipers and B. Meulenbeld, *On a Generalization of Legendre's Associated Differential Equation*, Proceedings of the Koninklijke Nederlandse Akademie Van Wetenschappen, (1957), pp. 436–450.

[88] L. Lewin, *Polylogarithms and Associated Functions*, Elsevier Science Limited, 1981.

[89] D. W. Lozier, *NIST Digital Library of Mathematical Functions*, Annals of Mathematics and Artificial Intelligence, 38 (2003), pp. 105–119.

[90] D. G. Luenberger, *Optimization by Vector Space Methods*, John Wiley & Sons, 1997, https://doi.org/10.1137/1012072.

[91] T. M. MacRobert, *Spherical Harmonics: An Elementary Treatise on Harmonic Functions, with Applications*, vol. 98, Dover Publications, 1967.

[92] J. Maes and A. Bultheel, *Modeling Sphere-Like Manifolds with Spherical Powell-Sabin B-Splines*, Computer Aided Geometric Design, 24 (2007), pp. 79–89, https://doi.org/10.1016/j.cagd.2006.11.001.

[93] J. Mairal and J.-P. Vert, *Machine Learning with Kernel Methods*, 2017. Lecture Notes.

[94] J. C. Mairhuber, *On Haar's Theorem Concerning Chebychev Approximation Problems Having Unique Solutions*, Proceedings of the American Mathematical Society, 7 (1956), pp. 609–615, https://doi.org/10.2307/2033359.

[95] J. H. Manton, P.-O. Amblard, et al., *A Primer on Reproducing Kernel Hilbert Spaces*, Foundations and Trends in Signal Processing, 8 (2015), pp. 1–126, https://doi.org/10.1561/9781680830934.

[96] J. L. Martinez-Morales, *Generalized Legendre Series and the Fundamental Solution of the Laplacian on the N-Sphere*, Analysis Mathematica, 31 (2005), pp. 131–150, https://doi.org/10.1007/s10476-005-0009-y.

[97] G. Matheron, *The Intrinsic Random Functions and their Applications*, Advances in Applied Probability, 5 (1973), pp. 439–468, https://doi.org/10.2307/1425829.

[98] T. Maunu, T. Zhang, and G. Lerman, *A Well-Tempered Landscape for Non-Convex Robust Subspace Recovery*, Journal of Machine Learning Research, 20 (2019), pp. 1–59.

[99] J. Meinguet, *An Intrinsic Approach to Multivariate Spline Interpolation at Arbitrary Points*, in Polynomial and Spline Approximation, Springer, 1979, pp. 163–190, https://doi.org/10.1007/978-94-009-9443-0_12.

[100] ——, *Multivariate Interpolation at Arbitrary Points Made Simple*, Zeitschrift Für Angewandte Mathematik Und Physik ZAMP, 30 (1979), pp. 292–304, https://doi.org/10.1007/BF01601941.

[101] ——, *Surface Spline Interpolation: Basic Theory and Computational Aspects*, in Approximation Theory and Spline Functions, Springer, 1984, pp. 132–142, https://doi.org/10.1007/978-94-009-6466-2_5.

[102] J. Mercer, *Functions of Positive and Negative Type, and their Connection the Theory of Integral Equations*, Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character, 209 (1909), pp. 415–446.

[103] P. E. Merilees, *The Pseudospectral Approximation Applied to the Shallow Water Equations on a Sphere*, Atmosphere, 11 (1973), pp. 13–20, https://doi.org/10.1080/00046973.1973.9648342.

[104] A. Mosamam and J. Kent, *Semi-Reproducing Kernel Hilbert Spaces, Splines and Increment Kriging*, Journal of Nonparametric Statistics, 22 (2010), pp. 711–722, https://doi.org/10.1080/10485250903388886.

[105] C. Müller, *Analysis of Spherical Symmetries in Euclidean Spaces*, vol. 129, Springer Science & Business Media, 2012, https://doi.org/10.1007/978-1-4612-0581-4.

[106] J. Naumann, *Remarks on the Prehistory of Sobolev Spaces*, Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät, 2002.

[107] M. Nica, *Eigenvalues and Eigenfunctions of the Laplacian*, The Waterloo Mathematics Review, 1 (2011), pp. 23–34.

[108] Z. Nie and C. Ma, *Isotropic Positive Definite Functions on Spheres Generated from those in Euclidean Spaces*, Proceedings of the American Mathematical Society, 147 (2019), pp. 3047–3056, https://doi.org/10.1090/proc/14454.

[109] A. V. Oppenheim and G. C. Verghese, *Signals, Systems and Inference*, Pearson, 2015.

[110] F. O'Sullivan and G. Wahba, *A Cross Validated Bayesian Retrieval Algorithm for Nonlinear Remote Sensing Experiments*, Journal of Computational Physics, 59 (1985), pp. 441–455, https://doi.org/10.1016/0021-9991(85)90121-4.

[111] D. Palmer, O. Stein, and J. Solomon, *Frame Field Operators*, Computer Graphics Forum, 40 (2021), pp. 231–245, https://doi.org/10.1111/cgf.14370.

[112] E. Parzen, *Statistical Inference on Time Series by Hilbert Space Methods, I*, tech. report, Stanford University Applied Mathematics and Statistics Lab, 1959.

[113] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, *Scikit-Learn: Machine Learning in Python*, Journal of Machine Learning Research, 12 (2011), pp. 2825–2830.

[114] M. A. Pinsky, *Endpoint Convergence of Legendre Series*, Birkhäuser Boston, Boston, MA, 1999, pp. 79–85, https://doi.org/10.1007/978-1-4612-2236-1_7.

[115] G. Pleiss, M. Jankowiak, D. Eriksson, A. Damle, and J. Gardner, *Fast Matrix Square Roots with Applications to Gaussian Processes and Bayesian Optimization*, Advances in Neural Information Processing Systems, 33 (2020), pp. 22268–22281.

[116] H. Poincaré, *Sur les Équations aux Dérivées Partielles de la Physique Mathématique*, American Journal of Mathematics, (1890), pp. 211–294, https://doi.org/10.2307/2369620.

[117] H. Pollard, *The Convergence Almost Everywhere of Legendre Series*, Proceedings of the American Mathematical Society, 35 (1972), pp. 442–444, https://doi.org/10.1090/S0002-9939-1972-0302973-7.

[118] T. Popoviciu, *Sur le Reste dans Certaines Formules Linéaires D'approximation de L'analyse*, Mathematica (Cluj), 1 (1959), pp. 95–142.

[119] F. Riesz, *Sur les Systèmes Orthogonaux de Fonctions*, Comptes Rendus de L'Académie des Sciences, 104 (1907), pp. 615–619.

[120] F. Riesz, *Sur les Opérations Functionnelles Linéaires*, Gauthier-Vllars/Elsevier, 1909.

[121] F. Riesz and B. Szökefalvi-Nagy, *Leçons D'analyse Fonctionnelle*, Gauthier-Villars/Elsevier, (1972).

[122] R. Rohde, R. Muller, R. Jacobsen, S. Perlmutter, A. Rosenfeld, J. Wurtele, J. Curry, C. Wickhams, and S. Mosher, *Berkeley Earth Temperature Averaging Process*, Geoinformatics Geostatistics An Overview, 1 (2013), pp. 20–100.

[123] S. Rosenberg, *The Laplacian on a Riemannian Manifold: an Introduction to Analysis on Manifolds*, no. 31, Cambridge University Press, 1997, https://doi.org/10.1017/CBO9780511623783.

[124] A. Saha and B. Palaniappan, *Learning with Operator-Valued Kernels in Reproducing Kernel Krein Spaces*, Advances in Neural Information Processing Systems, 33 (2020), pp. 13856–13866.

[125] R. Saxena, *Expansion of Continuous Differentiable Functions in Fourier Legendre Series*, Canadian Journal of Mathematics, 19 (1967), pp. 823–827, https://doi.org/10.4153/CJM-1967-076-1.

[126] S. R. Schach, *New Identities for Legendre Associated Functions of Integral Order and Degree. I*, SIAM Journal on Mathematical Analysis, 7 (1976), pp. 59–69, https://doi.org/10.1137/0507007.

[127] E. Schmidt, *Entwickelung Willkürlicher Functionen Nach Systemen Vorgeschriebener*, Dieterich'sche Univ.-Buchdr.(WF Kaestner), 1905.

[128] I. J. Schoenberg, *Metric Spaces and Completely Monotone Functions*, Annals of Mathematics, 39 (1938), pp. 811–841, https://doi.org/10.2307/1968466.

[129] ——, *Positive Definite Functions on Spheres*, Duke Mathematical Journal, 9 (1942), pp. 96–108, https://doi.org/10.1007/978-1-4612-3948-2_13.

[130] I. J. Schoenberg, *On Interpolation by Spline Functions and its Minimal Properties*, in On Approximation Theory/Über Approximationstheorie, Springer, 1964, pp. 109–129, https://doi.org/10.1007/978-1-4899-0433-1_21.

[131] B. Schölkopf, R. Herbrich, and A. J. Smola, *A Generalized Representer Theorem*, in International Conference on Computational Learning Theory, Springer, 2001, pp. 416–426, https://doi.org/10.1007/3-540-44581-1_27.

[132] L. Schwartz, *Sous-Espaces Hilbertiens D'espaces Vectoriels Topologiques et Noyaux Associés (Noyaux Reproduisants)*, Journal D'analyse Mathématique, 13 (1964), pp. 115–256, https://doi.org/10.1007/BF02786620.

[133] L. Schwartz, *Théorie des Distributions*, Publications de l'Institut de Mathématiques de l'Université de Strasbourg, Hermann, 1966, https://doi.org/10.4236/jsea.2012.512B004.

[134] J. Shawe-Taylor, N. Cristianini, et al., *Kernel Methods for Pattern Analysis*, Cambridge University Press, 2004, https://doi.org/10.1017/CBO9780511809682.

[135] L. A. Shepp, *Radon-Nikodym Derivatives of Gaussian Measures*, The Annals of Mathematical Statistics, (1966), pp. 321–354, https://doi.org/10.1214/aoms/1177699516.

[136] G. Shilov, *Integral, Measure and Derivative: a Unified Approach*.

[137] A. Shiryayev, *Interpolation and Extrapolation of Stationary Random Sequences*, in Selected Works of AN Kolmogorov, Springer, 1992, pp. 272–280, https://doi.org/10.1007/978-94-011-2260-3_28.

[138] L. Shure, R. L. Parker, and G. E. Backus, *Harmonic Splines for Geomagnetic Modelling*, Physics of the Earth and Planetary Interiors, 28 (1982), pp. 215–229, https://doi.org/10.1016/0031-9201(82)90003-6.

[139] A. J. Smola and R. Kondor, *Kernels and Regularization on Graphs*, in Learning Theory and Kernel Machines, Springer, 2003, pp. 144–158, https://doi.org/10.1007/978-3-540-45167-9_12.

[140] E. Stein and G. Weiss, *Introduction to Fourier Analysis on Euclidean Spaces*, vol. 1, Princeton University Press, 1971, https://doi.org/10.1515/9781400883899.

[141] I. Steinwart and A. Christmann, *Support Vector Machines*, Springer Science & Business Media, 2008, https://doi.org/10.1007/978-0-387-77242-4.

[142] I. Steinwart, D. Hush, and C. Scovel, *An Explicit Description of the Reproducing Kernel Hilbert Spaces of Gaussian RBF Kernels*, IEEE Transactions on Information Theory, 52 (2006), pp. 4635–4643, https://doi.org/10.1109/TIT.2006.881713.

[143] G. Strang, *The Discrete Cosine Transform*, SIAM Review, 41 (1999), pp. 135–147, https://doi.org/10.1137/S0036144598336745.

[144] ——, *Linear Algebra and its Applications*, Belmont, CA: Thomson, Brooks/Cole, 2006, https://doi.org/10.1137/1024124.

[145] A. M. Stuart, *Inverse Problems: a Bayesian Perspective*, Acta Numerica, 19 (2010), pp. 451–559, https://doi.org/10.1017/S0962492910000061.

[146] P. P. Tans, T. J. Conway, and T. Nakazawa, *Latitudinal Distribution of the Sources and Sinks of Atmospheric Carbon Dioxide Derived from Surface Observations and an Atmospheric Transport Model*, Journal of Geophysical Research: Atmospheres, 94 (1989), pp. 5151–5172, https://doi.org/10.1029/JD094iD04p05151.

[147] R. Temam, *Problèmes Mathématiques en Plasticité*, vol. 12, Gauthier-Villars/Elsevier, 1983.

[148] C. Thomas-Agnan, *Computing a Family of Reproducing Kernels for Statistical Applications*, Numerical Algorithms, 13 (1996), pp. 21–32, https://doi.org/10.1007/BF02143124.

[149] G. P. Tolstov, *Fourier Series*, Courier Corporation, 2012, https://doi.org/10.1063/1.3050756.

[150] A. Townsend, H. Wilber, and G. B. Wright, *Computing with Functions in Spherical and Polar Geometries I. The Sphere*, SIAM Journal on Scientific Computing, 38 (2016), pp. C403–C425, https://doi.org/10.1137/15M1045855.

[151] M. Unser, J. Fageot, and J. P. Ward, *Splines Are Universal Solutions of Linear Inverse Problems with Generalized TV Regularization*, SIAM Review, 59 (2017), pp. 769–793, https://doi.org/10.1137/16M1061199.

[152] A. van der Linde, *A Note on Smoothing Splines as Bayesian Estimates*, Statistics & Risk Modelling, 11 (1993), pp. 61–68, https://doi.org/10.1524/strm.1993.11.1.61.

[153] G. Wahba, *Improper Priors, Spline Smoothing and the Problem of Guarding Against Model Errors in Regression*, Journal of the Royal Statistical Society: Series B (Methodological), 40 (1978), pp. 364–372, https://doi.org/10.1111/j.2517-6161.1978.tb01050.x.

[154] ——, *Spline Interpolation and Smoothing on the Sphere*, SIAM Journal on Scientific and Statistical Computing, 2 (1981), pp. 5–16, https://doi.org/10.1137/0902002.

[155] ——, *Erratum: Spline Interpolation and Smoothing on the Sphere*, SIAM Journal on Scientific and Statistical Computing, 3 (1982), pp. 385–3866, https://doi.org/10.1137/0903024.

[156] ——, *Surface Fitting with Scattered Noisy Data on Euclidean D-Space and on the Sphere*, The Rocky Mountain Journal of Mathematics, (1984), pp. 281–299, https://doi.org/10.1216/RMJ-1984-14-1-281.

[157] ——, *Three Topics in Ill-Posed Problems*, in Inverse and Ill-Posed Problems, Elsevier, 1987, pp. 37–51, https://doi.org/10.1016/B978-0-12-239040-1.50008-9.

[158] ——, *Spline Models for Observational Data*, SIAM, 1990, https://doi.org/10.1137/1.9781611970128.

[159] G. Wahba and Y. Wang, *Representer Theorem*, Wiley StatsRef: Statistics Reference Online, (2014), pp. 1–11, https://doi.org/10.1002/9781118445112.stat08200.

[160] E. W. Weisstein, *Spherical Harmonic*, 2004. MathWorld–A Wolfram Web Resource, https://mathworld.wolfram.com/SphericalHarmonic.html.

[161] J. G. Wendelberger, *Smoothing Noisy Data with Multidimensional Splines and Generalized Cross-Validation*, The University of Wisconsin-Madison, 1982.

[162] H. Wendland, *Scattered Data Approximation*, vol. 17, Cambridge University Press, 2004, https://doi.org/10.1017/CBO9780511617539.

[163] E. Whittaker and G. Watson, *A Course of Modern Analysis: an Introduction to the General Theory of Infinite Processes and of Analytic Functions, with an Account of the Principal Transcendental Functions*, Cambridge University Press, 1928, https://doi.org/10.2307/2972291.

[164] C. K. Williams and C. E. Rasmussen, *Gaussian Processes for Machine Learning*, vol. 2, MIT Press Cambridge, MA, 2006.

[165] A. Wilson and H. Nickisch, *Kernel Interpolation for Scalable Structured Gaussian Processes (KISS-GP)*, International Conference on Machine Learning, (2015), pp. 1775–1784.

[166] S. N. Wood, *Monotonic Smoothing Splines Fitted by Cross Validation*, SIAM Journal on Scientific Computing, 15 (1994), pp. 1126–1133, https://doi.org/10.1137/0915069.

[167] ——, *Thin Plate Regression Splines*, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 65 (2003), pp. 95–114, https://doi.org/10.1111/1467-9868.00374.

[168] S. N. Wood, *Generalized Additive Models: An Introduction with R*, Chapman and Hall/CRC, 2 ed., 2017, https://doi.org/10.1201/9781315370279.

[169] D. Xiang and G. Wahba, *A Generalized Approximate Cross Validation for Smoothing Splines with Non-Gaussian Data*, Statistica Sinica, (1996), pp. 675–692.

[170] Y. Xu, *Positive Definite Functions on the Unit Sphere and Integrals of Jacobi Polynomials*, Proceedings of the American Mathematical Society, 146 (2018), pp. 2039–2048, https://doi.org/10.1090/proc/13913.

[171] Y. Xu and E. W. Cheney, *Strictly Positive Definite Functions on Spheres*, Proceedings of the American Mathematical Society, (1992), pp. 977–981, https://doi.org/10.1090/S0002-9939-1992-1096214-6.

[172] A. I. Zayed, *On the Singularities of Gegenbauer (Ultraspherical) Expansions*, Transactions of the American Mathematical Society, 262 (1980), pp. 487–503, https://doi.org/10.1090/S0002-9947-1980-0586730-1.

[173] F. Zhang and E. R. Hancock, *Graph Spectral Image Smoothing Using the Heat Kernel*, Pattern Recognition, 41 (2008), pp. 3328–3342, https://doi.org/10.1016/j.patcog.2008.05.007.