CHAPTER 1
Background

1.1 Positive-definite functions, covariances, and reproducing kernels.

We begin with a general index set $\mathcal{T}$. Examples of $\mathcal{T}$ that are of interest follow:

$$\begin{aligned}
\mathcal{T} &= (1, 2, \ldots, N) \\
\mathcal{T} &= (\ldots, -1, 0, 1, \ldots) \\
\mathcal{T} &= [0, 1] \\
\mathcal{T} &= E^d \qquad \text{(Euclidean } d\text{-space)} \\
\mathcal{T} &= \mathcal{S} \qquad \text{(the unit sphere)} \\
\mathcal{T} &= \text{ the atmosphere (the volume between two concentric spheres),}
\end{aligned}$$

etc. The text below is generally written as though the index set were continuous, but the discrete examples are usually special cases. A symmetric, real-valued function $R(s,t)$ of two variables $s, t \in \mathcal{T}$ is said to be *positive definite* if, for any real $a_1, \ldots, a_n$, and $t_1, \ldots, t_n \in \mathcal{T}$

$$\sum_{i,j=1}^{n} a_i a_j \; R(t_i, t_j) \geq 0,$$

and *strictly positive definite* if ">" holds. If $R(\cdot,\cdot)$ is positive definite, then we can always define a family $X(t)$, $t \in \mathcal{T}$ of zero-mean Gaussian random variables with covariance function R, that is,

$$E\, X(s)X(t) = R(s,t), \qquad s, t \in \mathcal{T}. \tag{1.1.1}$$

All functions and random variables in this book will be real valued unless specifically noted otherwise.

The existence of such a well-defined family of random variables in the continuous case is guaranteed by the Kolmogorov consistency theorem (see, e.g., Cramer and Leadbetter (1967, Chap. 3)). Given a positive-definite function $R(\cdot,\cdot)$ we are going to associate with it a *reproducing kernel Hilbert space* (r.k.h.s.). A (real) r.k.h.s. is a Hilbert space of real-valued functions on $\mathcal{T}$ with the property that, for each $t \in \mathcal{T}$, the evaluation functional L_t, which associates

f with $f(t)$, $L_t f \to f(t)$, is a bounded linear functional. The boundedness means that there exists an $M = M_t$ such that

$$|L_t f| = |f(t)| \le M\|f\| \quad \text{for all } f \quad \text{in the r.k.h.s.},$$

where $\|\cdot\|$ is the norm in the Hilbert space. We remark that the familiar Hilbert space $\mathcal{L}_2[0,1]$ of square integrable functions on $[0,1]$ does not have this property, no such M exists, and, in fact, elements in $\mathcal{L}_2[0,1]$ are not even defined pointwise.

If $\mathcal{H}$ is an r.k.h.s., then for each $t \in \mathcal{T}$ there exists, by the Riesz representation theorem, an element R_t in $\mathcal{H}$ with the property

$$L_t f = < R_t, f >= f(t), \qquad f \in \mathcal{H}. \tag{1.1.2}$$

R_t is called the representer of evaluation at t. Here, and elsewhere, we will use $< \cdot, \cdot >$ for the inner product in a reproducing kernel space. This inner product will, of course, depend on what space we are talking about. This leads us to the following theorem.

THEOREM 1.1.1. *To every* r.k.h.s. *there corresponds a unique positive-definite function (called the reproducing kernel* (r.k.)) *and conversely, given a positive-definite function R on $\mathcal{T} \times \mathcal{T}$ we can construct a unique* r.k.h.s. *of real-valued functions on $\mathcal{T}$ with R as its* r.k.

The proof is simple. If $\mathcal{H}$ is an r.k.h.s., then the r.k. is $R(s,t) = < R_s, R_t >$, where for each s, t, R_s and R_t are the representers of evaluation at s and t. $R(\cdot,\cdot)$ is positive definite since, for any $t_1, \ldots, t_n \in \mathcal{T}$, $a_1, \ldots, a_n$,

$$\begin{aligned} \sum_{i,j} a_i a_j \; R(t_i, t_j) &= \sum_{i,j} a_i a_j < R_{t_i}, R_{t_j} > \\ &= \|\Sigma a_j R_{t_j}\|^2 \ge 0. \end{aligned}$$

Conversely, given R we construct $\mathcal{H} = \mathcal{H}_R$ as follows. For each fixed $t \in \mathcal{T}$, denote by R_t the real-valued function with

$$R_t(\cdot) = R(t, \cdot). \tag{1.1.3}$$

By this is meant: R_t is the function whose value at s is $R(t,s)$. Then construct a linear manifold by taking all finite linear combinations of the form

$$\sum_i a_i R_{t_i} \tag{1.1.4}$$

for all choices of n, $a_1, \ldots, a_n$, $t_1, \ldots, t_n$ with the inner product

$$< \Sigma a_i \; R_{t_i}, \; \Sigma b_j \; R_{s_j} >= \sum_{ij} a_i b_j < R_{t_i}, R_{s_j} >= \Sigma a_i b_j R(t_i, s_j).$$

This is a well-defined inner product, since R is positive definite, and it is easy to check that for any f of the form (1.1.4) $< R_t, f >= f(t)$. In this linear manifold, norm convergence implies pointwise convergence, since

$$|f_n(t) - f(t)| = | < f_n - f, R_t > | \le \|f_n - f\| \; \|R_t\|.$$

Thus, to this linear manifold we can adjoin all limits of Cauchy sequences of functions in the linear manifold, which will be well defined as pointwise limits. The resulting Hilbert space is seen to be the r.k.h.s. $\mathcal{H}_R$ with r.k. R.

R is called the reproducing kernel, since

$$< R_s, R_t >=< R(s,\cdot),\ R(t,\cdot) >= R(s,t).$$

We will frequently denote by $\mathcal{H}_R$ the r.k.h.s. with r.k. R, and its inner product by $< \cdot,\cdot >_R$ or just $< \cdot,\cdot >$ if it is clear which inner product is meant. As a positive-definite function, under some general circumstances, R has an eigenvector-eigenvalue decomposition that generalizes the eigenvector-eigenvalue decomposition of a positive-definite matrix Σ, $\Sigma = \Gamma D \Gamma'$ with Γ orthogonal and D diagonal. Below we will show why the squared norm $\|f\|^2 = \|f\|_R^2$ can be thought of as a generalization of the expression $f'\Sigma^{-1}f$ with f a vector that appears in the multivariate normal density function with covariance Σ. In particular, suppose $R(s,t)$ continuous and

$$\int_{\mathcal{T}}\int_{\mathcal{T}} R^2(s,t)\,ds\,dt < \infty. \tag{1.1.5}$$

Then there exists an orthonormal sequence of continuous eigenfunctions, $\Phi_1, \Phi_2, \ldots$ in $\mathcal{L}_2[\mathcal{T}]$ and eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq 0$, with

$$\int_{\mathcal{T}} R(s,t)\Phi_\nu(t)\,dt = \lambda_\nu \Phi_\nu(s), \quad \nu = 1,2,\ldots, \tag{1.1.6}$$

$$R(s,t) = \sum_{\nu=1}^{\infty} \lambda_\nu \Phi_\nu(s)\Phi_\nu(t), \tag{1.1.7}$$

$$\int_{\mathcal{T}}\int_{\mathcal{T}} R^2(s,t)\,ds\,dt = \sum_{\nu=1}^{\infty} \lambda_\nu^2 < \infty. \tag{1.1.8}$$

See the Mercer–Hilbert–Schmidt theorems of Riesz and Sz.-Nagy (1955, pp. 242–246) for proofs of (1.1.6)–(1.1.8). Note that if we rewrite this result for the case $\mathcal{T} = (1,2,\ldots,N)$, then (1.1.6)–(1.1.8) become

$$\begin{aligned} R\Phi_\nu &= \lambda_\nu \Phi_\nu, \\ R &= \Gamma D \Gamma', \\ \text{trace } R^2 &= \sum_{\nu=1}^{N} \lambda_\nu^2, \end{aligned}$$

where R is the $N \times N$ matrix with ijth entry $R(i,j)$, Φ_ν is the vector with jth entry $\Phi_\nu(j)$, D is the diagonal matrix with $\nu\nu$th entry λ_ν, and Γ is the orthogonal matrix with νth column Φ_ν.

We have the following lemma.

LEMMA 1.1.1. *Suppose* (1.1.5) *holds. If we let*

$$f_\nu = \int_{\mathcal{T}} f(t)\Phi_\nu(t)\,dt, \tag{1.1.9}$$

then $f \in \mathcal{H}_R$ *if and only if*

$$\sum_{\nu=1}^{\infty} \frac{f_\nu^2}{\lambda_\nu} < \infty \tag{1.1.10}$$

and

$$\|f\|_R^2 = \sum_{\nu=1}^{\infty} \frac{f_\nu^2}{\lambda_\nu}. \tag{1.1.11}$$

Proof. The collection of all functions f with $\Sigma(f_\nu^2/\lambda_\nu) < \infty$ is clearly a Hilbert space with $\|f\|^2 = \Sigma(f_\nu^2/\lambda_\nu)$. We must show that R with

$$R(s,t) = \Sigma\lambda_\nu\Phi_\nu(s)\Phi_\nu(t)$$

is its r.k. That is, we must show that $R_t \in \mathcal{H}_R$ and

$$< f, R_t >= f(t), \qquad f \in \mathcal{H}_R\,, \qquad t \in \mathcal{T}$$

for $R_t(s) = R(t,s)$. Expanding f and $R(t,\cdot)$ in Fourier series with respect to $\Phi_1, \Phi_2, \ldots,$ we have

$$\begin{aligned} f(\cdot) &\sim \sum_\nu f_\nu \Phi_\nu(\cdot), \\ R(t,\cdot) &\sim \sum_\nu \{\lambda_\nu\Phi_\nu(t)\}\Phi_\nu(\cdot)\,. \end{aligned}$$

$R_t \in \mathcal{H}_R$ since $\sum_\nu \{\lambda_\nu\Phi_\nu(t)\}^2/\lambda_\nu = \sum_\nu \lambda_\nu\Phi_\nu^2(t) = R(t,t) < \infty$ and

$$< f, R_t >=< f, R(t,\cdot) >= \sum_\nu f_\nu\{\lambda_\nu\Phi_\nu(t)\}/\lambda_\nu, \qquad t \in \mathcal{T}$$

using the inner product induced by the norm in (1.1.11). But

$$\sum_\nu f_\nu\{\lambda_\nu\Phi_\nu(t)\}/\lambda_\nu = \sum_\nu f_\nu\Phi_\nu(t) = f(t),$$

and the result is proved.

We remark that if we begin with R satisfying (1.1.5) and construct the Hilbert space of functions with $\sum(f_\nu^2/\lambda_\nu) < \infty$, it is easy to show that the evaluation functionals are bounded:

$$\begin{aligned} |f(t)| = \left|\sum_{\nu=1}^{\infty} \frac{f_\nu\{\sqrt{\lambda_\nu}\Phi_\nu(t)\}}{\sqrt{\lambda_\nu}}\right| &\le \sqrt{\sum_{\nu=1}^{\infty}\frac{f_\nu^2}{\lambda_\nu}\sum_{\nu=1}^{\infty}\lambda_\nu\Phi_\nu^2(t)} \\ &= \|f\|\sqrt{R(t,t)} = \|f\|\;\|R_t\|. \end{aligned}$$

We remind the reader of the Karhunen–Loeve expansion. Suppose R is a covariance for which (1.1.5) holds, and let $X(t),\ t \in \mathcal{T}$ be a family of zero-mean Gaussian random variables with $EX(s)X(t) = R(s,t) = \sum_{\nu=1}^{\infty} \lambda_\nu \Phi_\nu(s)\Phi_\nu(t)$. Then $X(t), t \in \mathcal{T}$ has a (quadratic mean) representation

$$X(t) \sim \sum_{\nu=1}^{\infty} X_\nu \Phi_\nu(t),$$

where $X_1, X_2, \ldots$ are independent, Gaussian random variables with

$$EX_\nu = 0, \qquad EX_\nu^2 = \lambda_\nu$$

and

$$X_\nu = \int_T X(s)\Phi_\nu(s)\,ds. \tag{1.1.12}$$

The integral in (1.1.12) is well defined in quadratic mean (see Cramer and Leadbetter (1967)). However, sample functions of $X(t),\ t \in \mathcal{T}$ are not (with probability 1) in $\mathcal{H}_R$, if R has more than a finite number of nonzero eigenvalues. We do not prove this (see Hajek (1962a)), but merely consider the following suggestion of this fact. Let

$$X_K(t) = \sum_{\nu=1}^{K} X_\nu \Phi_\nu(t), \qquad t \in \mathcal{T},$$

then for each fixed t, $X_K(t)$ tends to $X(t)$ in quadratic mean, since

$$E|X_K(t) - X(t)|^2 = E|\sum_{K+1}^{\infty} X_\nu \Phi_\nu(t)|^2 = \sum_{K+1}^{\infty} \lambda_\nu \Phi_\nu^2(t) \to 0;$$

however,

$$E\|X_K(\cdot)\|^2 = E\sum_{\nu=1}^{K} \frac{X_\nu^2}{\lambda_\nu} = K \to \infty \text{ as } K \to \infty.$$

This very important fact, namely, that the assumptions that $f \in \mathcal{H}_R$ and f a sample function from a zero-mean Gaussian stochastic process are *not* the same, will have important consequences later.

1.2 Reproducing kernel spaces on $[0,1]$ with norms involving derivatives.

We remind the reader of Taylor's theorem with remainder: If f is a real- valued function on $[0,1]$ with $m-1$ continuous derivatives and $f^{(m)} \in \mathcal{L}_2[0,1]$, then we may write

$$f(t) = \left\{ \sum_{\nu=0}^{m-1} \frac{t^\nu}{\nu!} f^{(\nu)}(0) \right\} + \left\{ \int_0^1 \frac{(t-u)_+^{m-1}}{(m-1)!} f^{(m)}(u)\,du \right\}, \tag{1.2.1}$$

where $(x)_+ = x$ for $x \geq 0$ and $(x)_+ = 0$ otherwise. Let $\mathcal{B}_m$ denote the class of functions satisfying the boundary conditions $f^{(\nu)}(0) = 0,\ \nu = 0, 1 \ldots, m-1$. If $f \in \mathcal{B}_m$ then

$$\begin{aligned} f(t) &= \int_0^1 \frac{(t-u)_+^{m-1}}{(m-1)!} f^{(m)}(u)\, du \\ &= \int_0^1 G_m(t,u) f^{(m)}(u)\, du, \quad \text{say}, \end{aligned} \tag{1.2.2}$$

where

$$G_m(t,u) = (t-u)_+^{m-1}/(m-1)!. \tag{1.2.3}$$

G_m is the *Green's function* for the problem $D^m f = g,\ f \in \mathcal{B}_m$, where D^m denotes the mth derivative. Equation (1.2.2) can be verified by interchanging the order of integration in $f(t) = \int_0^t dt_{m-1} \int_0^{t_{m-1}} dt_{m-2} \cdots \int_0^{t_1} f^{(m)}(u)\, du$. Denote by W_m^0 the collection of functions on $[0,1]$ with

$$\{f :\ f \in \mathcal{B}_m,\ f, f', \ldots, f^{(m-1)} \text{ absolutely continuous},\ f^{(m)} \in \mathcal{L}_2\}.$$

It is not hard to show that W_m^0 is a Hilbert space with square norm $\|f\|^2 = \int_0^1 (f^{(m)}(t))^2\, dt$. We claim that W_m^0 is an r.k.h.s. with r.k.

$$R(s,t) = \int_0^1 G_m(t,u) G_m(s,u)\, du. \tag{1.2.4}$$

To show that the evaluation functionals are bounded, note that for $f \in W_m^0$ we have

$$f(s) = \int_0^1 G_m(s,u) f^{(m)}(u)\, du \tag{1.2.5}$$

so that by the Cauchy–Schwarz inequality

$$\begin{aligned} |f(s)| &\leq \sqrt{\int_0^1 G_m^2(s,u)\, du} \sqrt{\int_0^1 (f^{(m)}(u))^2\, du} \\ &= \sqrt{R\,(s,s)}\, \|f\|. \end{aligned}$$

To show that $R(\cdot,\cdot)$ is the r.k. for W_m^0 we must show that $R_t(\cdot) = R(t,\cdot)$ is in W_m^0 and that $< R_t, f >= f(t)$, all $f \in W_m^o$. But

$$R_t(v) = \int_0^1 G_m(v,u) G_m(t,u)\, du$$

and hence R_t is in W_m^0,

$$\left(\frac{\partial^m}{\partial v^m} R_t\right)(v) = G_m(t,v),$$

as can be seen by letting $f = R_t$ in (1.2.5). Thus

$$< f, R_t > = < R_t, f > = \int_0^1 G_m(t, v) f^{(m)}(v)\, dv = f(t).$$

Now let $\phi_\nu(t) = t^{\nu-1}/(\nu-1)!$ for $\nu = 1, 2, \ldots, m$ and denote the m-dimensional space of polynomials of degree $m-1$ or less spanned by $\phi_1, \cdots \phi_m$ as $\mathcal{H}_0$. Note that $D^m(\mathcal{H}_0) = 0$. Since

$$\begin{aligned} (D^{\mu-1}\phi_\nu)(0) &= 1, \quad \mu = \nu \\ &= 0, \quad \mu \neq \nu, \quad \mu, \nu = 1, \cdots, m, \end{aligned}$$

$\mathcal{H}_0$ endowed with the squared norm

$$\|\phi\|^2 = \sum_{\nu=0}^{m-1} [(D^\nu \phi)(0)]^2,$$

is an m-dimensional Hilbert space with $\phi_1, \ldots, \phi_m$ as an orthonormal basis, and it is not hard to show that then the r.k. for $\mathcal{H}_0$ is

$$\sum_{\nu=1}^{m} \phi_\nu(s)\phi_\nu(t).$$

To see this, let $R_t(\cdot) = \sum_{\nu=1}^m \phi_\nu(t)\phi_\nu(\cdot)$; then

$$< R_t, \phi_\alpha > = \sum_{\nu=1}^{m} \phi_\nu(t) < \phi_\nu, \phi_\alpha > = \phi_\alpha(t), \quad \alpha = 1, 2, \ldots, m.$$

We are now ready to construct the so-called Sobolev–Hilbert space W_m;

$$W_m : W_m[0,1] = \{f : \ f, f', \ldots, f^{m-1} \text{absolutely continuous}, f^{(m)} \in \mathcal{L}_2\}.$$

There are a number of ways to construct a norm on W_m. The norm we give here is given in Kimeldorf and Wahba (1971) and has associated with it an r.k. that will be particularly useful for our purposes. Different (but topologically equivalent) norms on this space will be introduced below and in Section 10.2. "Sobolev space" is the general term given for a function space (not necessarily a Hilbert space) whose norm involves derivatives. For more on Sobolev spaces, see Adams (1975). Each element in W_m has a Taylor series expansion (1.2.1) to order m and hence a unique decomposition

$$f = f_0 + f_1$$

with $f_0 \in \mathcal{H}_0$ and $f_1 \in W_m^o$, given by the first and second terms in brackets in (1.2.1). Furthermore, $\int_0^1((D^m f_0)(u))^2 du = 0$ and $\sum_{\nu=0}^{m-1}[(D^\nu f_1)(0)]^2 = 0$. Thus, denoting W_m^0 by $\mathcal{H}_1$, we claim

$$W_m = \mathcal{H}_0 \oplus \mathcal{H}_1,$$

and, if we endow W_m with the square norm

$$\|f\|^2 = \sum_{\nu=0}^{m-1} [(D^\nu f)(0)]^2 + \int_0^1 (D^m f)^2(u)\,du$$

then $\mathcal{H}_0$ and $\mathcal{H}_1$ will be perpendicular. With this norm, it is not hard to show that the r.k. for W_m is

$$R(s,t) = \sum_{\nu=1}^{m} \phi_\nu(s)\phi_\nu(t) + \int_0^1 G_m(s,u)G_m(t,u)\,du$$

where G_m is given by (1.2.3). The reproducing kernel for the direct sum of two perpendicular subspaces is the sum of the r.k.'s (see Aronszajn (1950)). An important geometrical fact that we will use later is that the penalty functional $J_m(f) = \int_0^1 (f^{(m)}(u))^2\,du$ may be written

$$J_m(f) = \|P_1 f\|^2_{W_m}$$

where P_1 is the orthogonal projection of f onto $\mathcal{H}_1$ in W_m.

We may replace D^m in $\int_0^1 (D^m f)^2\,du$ by more general differential operators. Let $a_1, a_2, \ldots, a_m$ be strictly positive functions with $a_i(0) = 1$ and as many derivatives as needed and let

$$L_m = D\frac{1}{a_1}D\frac{1}{a_2}\cdots D\frac{1}{a_m}.$$

Also let

$$\begin{aligned}
M_0 &= I \text{ (the identity)} \\
M_1 &= D\frac{1}{a_m} \\
M_2 &= D\frac{1}{a_{m-1}}\,D\frac{1}{a_m} \\
&\vdots \\
M_{m-1} &= D\frac{1}{a_2}\,D\frac{1}{a_3}\cdots D\frac{1}{a_m}
\end{aligned}$$

and let $\omega_1, \ldots, \omega_m$ be defined by

$$\begin{aligned}
\omega_1(t) &= a_m(t) \\
\omega_2(t) &= a_m(t) \int_0^t a_{m-1}(t_{m-1})\,dt_{m-1} \\
&\vdots \\
\omega_m(t) &= a_m(t) \int_0^t a_{m-1}(t_{m-1})\,dt_{m-1} \\
&\quad \cdot \int_0^{t_{m-1}} a_{m-2}(t_{m-2})\,dt_{m-2} \cdots \int_0^{t_2} a_1(t_1)\,dt_1.
\end{aligned}$$

Note that

$$\begin{aligned}(M_{\mu-1}\omega_\nu)(0) &= 1, \qquad \mu=\nu \\ &= 0, \qquad \mu\neq\nu, \qquad \mu,\nu=1,\ldots,m.\end{aligned}$$

The $\{\omega_\nu\}$ are an "extended Tchebycheff system" and share the following property with the polynomials $\phi_1,\ldots,\phi_m$. Let $t_1,\ldots,t_n$ be distinct, with $n\geq m$; then the $n\times m$ matrix T with i,νth entry $\omega_\nu(t_i)$ is of rank m (see Karlin (1968)).

Now, let $\tilde{\mathcal{B}}_m$ denote the class of functions satisfying the boundary conditions

$$(M_\nu f)(0)=0, \qquad \nu=0,1,\ldots,m-1,$$

and let $\tilde{G}_m$ be the Green's function for the problem $L_m f=g,\ f\in\tilde{\mathcal{B}}_m$. We have $f\in\tilde{\mathcal{B}}_m\Rightarrow$

$$\begin{aligned}f(t) &= a_m(t)\int_0^t a_{m-1}(t_{m-1})\,dt_{m-1} \\ &\quad\cdot\int_0^{t_{m-1}} a_{m-2}(t_{m-2})\,dt_{m-2}\cdots\int_0^{t_1}(L_mf)(u)\,du \\ &= \int_0^t (L_mf)(u)\,du\{a_m(t)\int_u^t a_1(t_1)\,dt_1 \\ &\quad\cdot\int_{t_1}^t a_2(t_2)\,dt_2\cdots\int_{t_{m-2}}^t a_{m-1}(t_{m-1})\,dt_{m-1}\} \\ &= \int_0^t \tilde{G}_m(t,u)(L_mf)(u)\,du,\end{aligned} \tag{1.2.6}$$

where $\tilde{G}_m(t,m)$ is equal to the expression in brackets in (1.2.6).

Let $\tilde{W}_m^0$ be the collection of functions on $[0,1]$ given by

$$\{f:\ f\in\tilde{\mathcal{B}}_m,\ M_0f,\ M_1f,\cdots,M_{m-1}f\ \text{ absolutely continuous},\ L_mf\in\mathcal{L}_2\}.$$

Then by the same arguments as before, $\tilde{W}_m^0$ is an r.k.h.s. with the squared norm $\|f\|^2=\int_0^1(L_mf)^2(u)\,du$, and reproducing kernel

$$R(s,t)=\int_0^1\tilde{G}_m(s,u)\tilde{G}_m(t,u)\,du.$$

Letting $\mathcal{H}_0$ be span $\{\omega_1,\ldots,\omega_m\}$ and $\mathcal{H}_1$ be $\tilde{W}_m^0$, then letting $\tilde{W}_m$ be the Hilbert space

$$\tilde{W}_m=\mathcal{H}_0\oplus\mathcal{H}_1,$$

we have that $\tilde{W}_m$ is an r.k.h.s. with

$$\|f\|^2=\sum_{\nu=0}^{m-1}[(M_\nu f)(0)]^2+\int_0^1(L_mf)^2\,du$$

and r.k.

$$\sum_{\nu=0}^{m-1} \omega_\nu(s)\omega_\nu(t) + \int_0^1 \tilde{G}_m(s,u)\tilde{G}_m(t,u)\,du.$$

Furthermore, we have the geometrical relation

$$\int_0^1 (L_m f)^2\,du = \|P_1 f\|^2_{\tilde{W}_m},$$

where P_1 is the orthogonal projection in $\tilde{W}_m$ onto $\mathcal{H}_1$.

We have, for $f \in \tilde{W}_m$, the *generalized Taylor series expansion*

$$f(t) = \sum_{\nu=1}^{m} \omega_\nu(t)(M_{\nu-1}f)(0) + \int_0^t \tilde{G}_m(t,u)(L_m f)(u)\,du.$$

We remark that W_m and $\tilde{W}_m$ are topologically equivalent; they have the same Cauchy sequences. Another topologically equivalent norm involving boundary rather than initial values will be introduced in Section 10.2.

1.3 The special and general spline smoothing problems.

The data model associated with the special spline smoothing problem is

$$y_i = f(t_i) + \epsilon_i, \qquad i = 1, 2, \ldots, n \tag{1.3.1}$$

where $t \in \mathcal{T} = [0,1]$, $f \in W_m$, and $\epsilon = (\epsilon_1, \ldots, \epsilon_n)' \sim \mathcal{N}(0, \sigma^2 I)$. An estimate of f is obtained by finding $f \in W_m$ to minimize

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - f(t_i))^2 + \lambda \int_0^1 (f^{(m)}(u))^2\,du. \tag{1.3.2}$$

The data model associated with the general spline smoothing problem is

$$y_i = L_i f + \epsilon_i, \quad i = 1, 2, \ldots, n \tag{1.3.3}$$

where ϵ is as before. Now $\mathcal{T}$ is arbitrary, $f \in \mathcal{H}_R$, a given r.k.h.s. of functions on $\mathcal{T}$, and $L_1, \ldots, L_n$ are bounded linear functionals on $\mathcal{H}_R$. $\mathcal{H}_R$ is supposed to have a decomposition

$$\mathcal{H}_R = \mathcal{H}_0 \oplus \mathcal{H}_1$$

where $\dim \mathcal{H}_0 = M \le n$. An estimate of f is obtained by finding $f \in \mathcal{H}_R$ to minimize

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - L_i f)^2 + \lambda\|P_1 f\|^2_R, \tag{1.3.4}$$

where P_1 is the orthogonal projection of f onto $\mathcal{H}_1$, in $\mathcal{H}_R$.

One of the useful properties of reproducing kernels is that from them one can obtain the representer of any bounded linear functional. Let η_i be the representer for L_i, that is,

$$< \eta_i, f > = L_i f, \quad f \in \mathcal{H}_R.$$

Then

$$\eta_i(s) = <\eta_i, R_s> = L_i R_s = L_{i(\cdot)} R(s, \cdot) \tag{1.3.5}$$

where $L_{i(\cdot)}$ means L_i is applied to what follows as a function of $(\cdot)$. That is, one can apply L_i to $R(s,t)$ considered as a function of t, to obtain $\eta_i(s)$. For example, if $L_i f = \int w_i(u) f(u)\, du$, then $\eta_i(s) = \int w_i(u) R(s,u)\, du$, and if $L_i f = f'(t_i)$, then $\eta_i(s) = (\partial/\partial u) R(s,u)|_{u=t_i}$. On the other hand L_i is a bounded linear functional on $\mathcal{H}_R$ only if $\eta_i(\cdot)$ obtained by $\eta_i(s) = L_{i(\cdot)} R(s,\cdot)$ is a well-defined element of $\mathcal{H}_R$. To see the argument behind this note that if $L_i f = \sum_\ell a_\ell f(t_\ell)$ for any finite sum, then its representer is $\eta_i = \sum_\ell a_\ell\, R_{t_\ell}$, and any η_i in $\mathcal{H}_R$ will be a limit of sums of this form and will be the representer of the limiting bounded linear functional. As an example, $L_i f = f'(t_i)$ a bounded linear functional in $\mathcal{H}_R \Rightarrow \eta_i = \lim_{h \to 0}(1/h)(R_{t_i+h} - R_{t_i})$, where the limit is in the norm topology, which then entails that

$$\frac{\partial}{\partial t} R(t,s)|_{t=t_i} = \eta_i(s) \quad \text{with } \eta_i \in \mathcal{H}_R.$$

$f^{(k)}(t_i)$ can be shown to be a bounded linear functional in W_m for $k = 0, 1, \ldots, m-1$. More details can be found in Aronszajn (1950).

We will now find an explicit formula for the minimizer of (1.3.4), which can now be written

$$\frac{1}{n} \sum_{i=1}^{n} (y_i - <\eta_i, f>)^2 + \lambda\, \|P_1 f\|_R^2. \tag{1.3.6}$$

THEOREM 1.3.1. *Let $\phi_1, \ldots, \phi_M$ span the null space $(\mathcal{H}_0)$ of P_1 and let the $n \times M$ matrix $T_{n \times M}$ defined by*

$$T_{n \times M} = \{L_i \phi_\nu\}_{i=1}^{n}\ {}_{\nu=1}^{M} \tag{1.3.7}$$

be of full column rank. Then f_λ, the minimizer of (1.3.6), is given by

$$f_\lambda = \sum_{\nu=1}^{M} d_\nu \phi_\nu + \sum_{i=1}^{n} c_i \xi_i \tag{1.3.8}$$

where

$$\begin{aligned} \xi_i &= P_1 \eta_i, \\ d = (d_1, \ldots, d_M)' &= (T'M^{-1}T)^{-1} T' M^{-1} y, \\ c = (c_1, \ldots, c_n)' &= M^{-1}(I - T(T'M^{-1}T)^{-1} T' M^{-1}) y, \\ M &= \Sigma + n\lambda I, \\ \Sigma &= \{<\xi_i, \xi_j>\}. \end{aligned} \tag{1.3.9}$$

(Do not confuse the index M and the matrix M.)

Before giving the proof we make a few remarks concerning the ingredients of the minimizer. Letting $\mathcal{H}_R = \mathcal{H}_0 \oplus \mathcal{H}_1$, with $\mathcal{H}_0 \perp \mathcal{H}_1$, where

$$R(s,t) = R^0(s,t) + R^1(s,t)$$

and R^α is the r.k. for $\mathcal{H}_\alpha$, $\alpha = 0, 1$, we then have

$$\begin{aligned} \xi_i(t) = <\xi_i, R_t> &= <P_1\eta_i, R_t> = <\eta_i, P_1 R_t> \\ &= <\eta_i, R_t^1> \\ &= L_i R_t^1 \end{aligned} \tag{1.3.10}$$

where R_t^1 is the representer of evaluation at t in $\mathcal{H}_1$. We have used that the projection P_1 is self-adjoint. Furthermore,

$$<\xi_i, \xi_j> = <\eta_i, \xi_j>$$

since $<\eta_i - \xi_i, \xi_j> = 0$, so that

$$<\xi_i, \xi_j> = L_i \xi_j = L_{i(s)} L_{j(t)} R^1(s,t).$$

To prove the theorem, let the minimizer f_λ be of the form

$$f_\lambda = \sum_{\nu=1}^{M} d_\nu \phi_\nu + \sum_{i=1}^{n} c_i \xi_i + \rho \tag{1.3.11}$$

where ρ is some element in $\mathcal{H}_R$ perpendicular to $\phi_1, \ldots, \phi_M$, $\xi_1, \ldots, \xi_n$. Any element in $\mathcal{H}_R$ has such a representation by the property of Hilbert spaces. Then (1.3.6) becomes

$$\frac{1}{n}\|y - (\Sigma c + Td)\|^2 + \lambda(c'\Sigma c + \|\rho\|^2) \tag{1.3.12}$$

and we must find c, d, and ρ to minimize this expression. It is then obvious that $\|\rho\|^2 = 0$, and a straightforward calculation shows that the minimizing c and d of

$$\frac{1}{n}\|y - (\Sigma c + Td)\|^2 + \lambda c'\Sigma c \tag{1.3.13}$$

are given by

$$d = (T'M^{-1}T)^{-1}T'M^{-1}y, \tag{1.3.14}$$

$$c = M^{-1}(I - T(T'M^{-1}T)^{-1}T'M^{-1})y. \tag{1.3.15}$$

These formulae are quite unsuitable for numerical work, and, in fact, were quite impractical when they appeared in Kimeldorf and Wahba (1971). Utreras (1979) provided an equivalent set of equations with more favorable properties, and another improvement was given in Wahba (1978b) with the aid of an anonymous referee, who was later unmasked as Silverman. Multiplying the left and right sides of (1.3.15) by M and substituting in (1.3.14) gives (1.3.16) and multiplying (1.3.15) by T' gives (1.3.17):

$$Mc + Td = y, \tag{1.3.16}$$

$$T'c = 0, \tag{1.3.17}$$

these being equivalent to (1.3.14) and (1.3.15).

To compute c and d, let the QR decomposition (see Dongarra et al. (1979)) of T be

$$T = (Q_1 : Q_2) \begin{pmatrix} R \\ 0 \end{pmatrix} \tag{1.3.18}$$

where Q_1 is $n \times M$ and Q_2 is $n \times (n-M)$, $Q = (Q_1 : Q_2)$ is orthogonal and R is upper triangular, with $T'Q_2 = 0_{M \times (n-M)}$. Since $T'c = 0$, c must be in the column space of Q_2, giving $c = Q_2\gamma$ for some γ an $n-M$ vector. Substituting $c = Q_2\gamma$ into (1.3.16) and multiplying through by Q_2', recalling that $Q_2'T = 0$, gives

$$\begin{aligned} Q_2'MQ_2\gamma &= Q_2'y, \\ c &= Q_2\gamma = Q_2(Q_2'MQ_2)^{-1}Q_2'y, \end{aligned} \tag{1.3.19}$$

and multiplying (1.3.16) by Q_1' gives

$$Rd = Q_1'(y - Mc). \tag{1.3.20}$$

For later use the influence matrix $A(\lambda)$ will play an important role. $A(\lambda)$ is defined as the matrix satisfying

$$\begin{pmatrix} L_1 f_\lambda \\ \vdots \\ L_n f_\lambda \end{pmatrix} = A(\lambda)y. \tag{1.3.21}$$

To obtain a simple formula for $I - A(\lambda)$ we observe by substitution in (1.3.11) with $\rho = 0$ that

$$\begin{pmatrix} L_1 f_\lambda \\ \vdots \\ L_n f_\lambda \end{pmatrix} = Td + \Sigma c. \tag{1.3.22}$$

Subtracting this from (1.3.16) gives

$$(I - A(\lambda))y = n\lambda c = n\lambda Q_2(Q_2'MQ_2)^{-1}Q_2'y$$

for any y, thus

$$I - A(\lambda) = n\lambda Q_2(Q_2'MQ_2)^{-1}Q_2'. \tag{1.3.23}$$

Of course Q_2 may be replaced by any $n \times (n-M)$ matrix whose columns are any orthonormal set perpendicular to the M columns of T. We will discuss efficient numerical methods for computing c and d in conjunction with data-based methods for choosing λ later.

For the special spline smoothing problem we will demonstrate that f_λ of (1.3.8) is a natural polynomial spline. Here

$$\begin{aligned} L_i f &= f(t_i), \\ \|P_1 f\|^2 &= \int_0^1 (f^{(m)}(u))^2\,du, \\ R^1(s,t) &= \int_0^1 \frac{(s-u)_+^{m-1}(t-u)_+^{m-1}}{[(m-1)!]^2}\,du \end{aligned}$$

and

$$\xi_i(\cdot) = R^1(\cdot, t_i).$$

It is easy to check that here

$$\begin{aligned} \xi_i(\cdot) &\in \pi^{2m-1} \text{ for } s \in [0, t_i] \\ &\in \pi^{m-1} \text{ for } s \in [t_i, 1], \end{aligned}$$

and

$$\xi_i(\cdot) \in C^{2m-2}.$$

Thus

$$\begin{aligned} f_\lambda(t) = \sum_{\nu=1}^{m} d_\nu \phi_\nu(t) + \sum_{i=1}^{n} c_i \xi_i(t) &\in \pi^{m-1} \text{ for } t \in [t_n, 1] \\ &\in \pi^{2m-1} \text{ for } t \in [t_i, t_{i+1}] \\ &\in C^{2m-2}. \end{aligned}$$

We will show that the condition $T'c = 0$ guarantees that $f_\lambda \in \pi^{m-1}$ for $t \in [0, t_1]$, as follows. For $t < t_1$, we can remove the "+" in the formula for ξ_i and write

$$\sum_{i=1}^{n} c_i \xi_i(t) = \int_0^t \frac{(t-u)^{m-1}}{(m-1)!} \sum_{i=1}^{n} c_i (t_i - u)^{m-1} \, du, \ t < t_1. \tag{1.3.24}$$

But $\sum_{i=1}^{n} c_i t_i^k = 0$ for $k = 0, 1, \ldots, m-1$ since $T'c = 0$, so that (1.3.24) is 0 for $t < t_1$ and the result is proved.

We remark that it can be shown that $\lim_{\lambda \to \infty} f_\lambda$ is the least squares regression onto $\phi_1, \ldots, \phi_M$ and $\lim_{\lambda \to 0} f_\lambda$ is the interpolant to $L_i f = y_i$ in $\mathcal{H}$ that minimizes $\|P_1 f\|$. The important choice of λ from the data will be discussed later.

1.4 The duality between r.k.h.s. and stochastic processes.

Later we will show how spline estimates are also Bayes estimates, with a certain prior on f. This is no coincidence, but is a consequence of the duality between the Hilbert space spanned by a family of random variables and its associated r.k.h.s. The discussion of this duality follows Parzen (1962, 1970).

Let $X(t)$, $t \in \mathcal{T}$ be a family of zero-mean Gaussian random variables with $EX(s)X(t) = R(s,t)$. Let $\mathcal{X}$ be the *Hilbert space spanned by* $X(t)$, $t \in \mathcal{T}$. This is the collection of all random variables of the form

$$Z = \Sigma a_j X(t_j) \tag{1.4.1}$$

$t_j \in \mathcal{T}$, with inner product $< Z_1, Z_2 >= EZ_1 Z_2$, and all of their quadratic mean limits, i.e. Z is in $\mathcal{X}$ if and only if there is a sequence Z_l, $l = 1, 2, \ldots$ of random variables each of the form (1.4.1), with $\lim_{l \to \infty} E(Z - Z_l)^2 =$

$\|Z - Z_l\|^2 \to 0$. Letting $\mathcal{H}_R$ be the r.k.h.s. with r.k. R, we will see that $\mathcal{H}_R$ is isometrically isomorphic to $\mathcal{X}$, that is, there exists a 1:1 inner product preserving correspondence between the two spaces. The correspondence is given by Table 1.1. This correspondence is clearly 1:1 and preserves inner products,

TABLE 1.1
The 1:1 *correspondence between* $\mathcal{H}_R$ *and* $\mathcal{X}$.

$\mathcal{X}$		$\mathcal{H}_R$
$X(t)$	$\sim$	R_t
$\Sigma a_j X(t_j)$	$\sim$	$\Sigma a_j R_{t_j}$
$\lim \Sigma a_j X(t_j)$	$\sim$	$\lim \Sigma a_j R_{t_j}$.

since

$$< X(s), X(t) >= EX(s)X(t) = R(s,t) =< R_s, R_t > .$$

Let L be a bounded linear functional in $\mathcal{H}_R$ with representer η. Then η is the limit of a sequence of elements of the form $\Sigma a_{t_l} R_{t_l}$, by construction of $\mathcal{H}_R$. The random variable Z corresponds to η if Z is the (quadratic mean) limit of the corresponding sequence of random variables $\Sigma a_{t_l} X(t_l)$ and we can finally denote this limiting random variable by LX (although $X \notin \mathcal{H}_R$ and we do not think of L as a bounded linear functional applied to X). Then $EZX(t) = < \eta, R_t >= \eta(t) = LR_t$. Examples are $Z = \int w(t)X(t)\,dt$ and $Z = X'(t)$, if they exist.

We are now ready to give a simple example of the duality between Bayesian estimation on a family of random variables and optimization in an r.k.h.s. Consider $X(t),\ t \in \mathcal{T}$ a zero-mean Gaussian stochastic process with $EX(s)X(t) = R(s,t)$. Fix t for the moment and compute $E\{X(t)|X(t_1) = x_1, \ldots, X(t_n) = x_n\}$. The joint covariance matrix of $X(t),\ X(t_1), \ldots, X(t_n)$ is

$$\begin{pmatrix} R(t,t) & R(t,t_1), \ldots, & R(t,t_n) \\ R(t,t_1) & & \\ \vdots & R_n & \\ R(t,t_n) & & \end{pmatrix}$$

where R_n is the $n \times n$ matrix with ijth entry $R(t_i, t_j)$. We will assume for simplicity in this example that R_n is strictly positive definite. Using properties of the multivariate normal distribution, as given, e.g., in Anderson (1958), we have

$$E\{X(t) \mid X(t_i) = x_i,\ i = 1, \ldots, n\}$$
$$= (R(t,t_1), \ldots, R(t,t_n))\, R_n^{-1} x = \hat{f}(t), \tag{1.4.2}$$

say. The Gaussianness is not actually being used, except to call $\hat{f}(t)$ a conditional expectation. If $\hat{f}(t)$ were just required to be the minimum variance unbiased linear estimate of $X(t)$, given the data, the result $\hat{f}(t)$ would be the same, independent of the form of the joint distribution and depending only on the first and second moments.

Now consider the following problem. Find $f \in \mathcal{H}_R$, the r.k.h.s. with r.k. R, to minimize $\|f\|^2$ subject to $f(t_i) = x_i$, $i = 1, \ldots, n$. By a special case of the argument given before, f must be of the form

$$f = \sum_{j=1}^{n} c_j \, R_{t_j} + \rho$$

for some $\rho \perp$ to $R_{t_1}, \ldots, R_{t_n}$, that is, ρ satisfies $< R_{t_i}, \rho >= \rho(t_i) = 0$, $i = 1, \ldots, n$. $\|f\|^2 = c' R_n c + \|\rho\|^2$ and so $\|\rho\| = 0$. Setting $f(t_i) = \sum_{j=1}^{n} c_j R_{t_j}(t_i) = x_i$, $i = 1, \ldots, n$ gives $x = (x_1, \ldots, x_n)' = R_n c$, and so the minimizer is f given by

$$f = x' \, R_n^{-1} \begin{pmatrix} R_{t_1} \\ \vdots \\ R_{t_n} \end{pmatrix} = \hat{f},$$

which is exactly equal to $\hat{f}$ of (1.4.2)!

1.5 The smoothing spline and the generalized smoothing spline as Bayes estimates.

We first consider

$$W_m = \mathcal{H}_0 \oplus W_m^0,$$

where W_m^0 has the r.k.

$$R^1(s,t) = \int_0^1 \frac{(s-u)_+^{m-1}}{(m-1)!} \, \frac{(t-u)_+^{m-1}}{(m-1)!} \, du. \tag{1.5.1}$$

Let

$$X(t) = \int_0^1 \frac{(t-u)_+^{m-1}}{(m-1)!} \, dW(u) \tag{1.5.2}$$

where $W(\cdot)$ is the Wiener process. Formally, $X \subset \mathcal{B}_m$, $D^m X = dW =$ "white noise." $X(\cdot)$ is the $m-1$ fold integrated Wiener process described in Shepp (1966). We remind the reader that the Wiener process is a zero-mean Gaussian stochastic process with stationary independent increments and $W(0) = 0$. Stationary, independent increments means, for any s_1, s_2, s_3, s_4, the joint distribution of $W(s_2) - W(s_1)$ and $W(s_4) - W(s_3)$ is the same as that of $W(s_2 + h) - W(s_1 + h)$ and $W(s_4 + h) - W(s_3 + h)$, and, if the intervals $[s_1, s_2]$ and $[s_3, s_4]$ are nonoverlapping then $W(s_2) - W(s_1)$ and $W(s_4) - W(s_3)$ are independent. Integrals of the form

$$\int_0^1 g(u) \, dW(u) \tag{1.5.3}$$

are defined as quadratic-mean limits of the Riemann–Stieltjes sums

$$\sum g(u_l)\,[W(u_{l+1}) - W(u_l)] \tag{1.5.4}$$

for partitions $\{u_1, \ldots, u_L\}$ of $[0, 1]$ (see Cramer and Leadbetter (1967, Chap. 5)). It can be shown to follow from the stationary independent increments property, that $E[W(u+h) - W(u)]^2 = \text{const. } h$, for some constant, which we will take here to be 1. Using the definition of (1.5.3) as a limit of the form (1.5.4), it can be shown using the independent increments property, that if g_1 and g_2 are in $\mathcal{L}_2[0,1]$, then

$$E\int_0^1 g_1(u)\,dW(u)\int_0^1 g_2(u)\,dW(u) = \int_0^1 g_1(u)g_2(u)\,du. \tag{1.5.5}$$

Thus,

$$EX(s)X(t) = \int_0^1 \frac{(t-u)_+^{m-1}}{(m-1)!}\,\frac{(s-u)_+^{m-1}}{(m-1)!}\,du = R^1(s,t) \tag{1.5.6}$$

and the Hilbert space spanned by the $m-1$ fold integrated Wiener process is isometrically isomorphic to W_m^0.

We will consider two types of Bayes estimates, both of which lead to a smoothing spline estimate. The first was given in Kimeldorf and Wahba (1971) and might be called the "fixed effects" model, and the second might be called the "random effects model with an improper prior," and was given in Wahba (1978b).

The first model is

$$\begin{aligned} F(t) &= \sum_{\nu=1}^{M}\theta_\nu\phi_\nu(t) + b^{1/2}X(t), \quad t\in[0,1], \\ Y_i &= F(t_i) + \epsilon_i, \quad i = 1,\ldots,n. \end{aligned} \tag{1.5.7}$$

Here $\theta = (\theta_1, \ldots, \theta_M)'$ is considered to be a fixed, but unknown, vector, b is a positive constant, $X(t),\ t \in [0,1]$ is a zero-mean Gaussian stochastic process with covariance $R^1(s,t)$ of (1.5.6), and $\epsilon \sim \mathcal{N}(0, \sigma^2 I)$. We wish to construct an estimate of $F(t),\ t \in \mathcal{T}$, given $Y_i = y_i,\ i = 1, \ldots, n$.

An estimate $\hat{F}(t)$ of $F(t)$ will be called unbiased with respect to θ if

$$E(\hat{F}(t)|\theta) = E(F(t)|\theta).$$

(Here, t is considered fixed.) Let $\hat{F}(t)$ be the minimum variance, unbiased (with respect to θ) linear estimate of $F(t)$ given $Y_i = y_i,\ i = 1, \ldots, n$. That is,

$$\hat{F}(t) = \sum_{j=1}^{n}\beta_j(t)y_j$$

for some $\beta_j(t)$ (linearity), and $\hat{F}(t)$ minimizes

$$E(\hat{F}(t) - F(t))^2$$

(minimum variance) subject to

$$E(\hat{F}(t) - F(t)|\theta) = 0 \quad \text{for all } t \in [0,1].$$

We have the following theorem.

THEOREM 1.5.1. *Let f_λ be the minimizer in W_m of*

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - f(t_i))^2 + \lambda \int_0^1 (f^{(m)}(u))^2\,du.$$

Then

$$\hat{F}(t) = f_\lambda(t)$$

with $\lambda = \sigma^2/nb$.

Proof. A proof can be obtained by straightforward calculation (see Kimeldorf and Wahba (1971)).

The general version of this theorem follows. Let $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$ where $\mathcal{H}_0$ is spanned by $\phi_1, \ldots, \phi_M$, and $\mathcal{H}_1$ has r.k. $R^1(s,t)$. Let

$$F(t) = \sum_{\nu=1}^{M} \theta_\nu \phi_\nu(t) + b^{1/2}X(t), \quad t \in \mathcal{T}$$

where θ is as before and $EX(s)X(t) = R^1(s,t)$. Let $L_1, \ldots, L_n$ be bounded linear functionals on $\mathcal{H}$; then $\sum_{\nu=1}^{M} \theta_\nu L_i \phi_\nu$ is a well-defined constant and $b^{1/2}L_i X$ is a well-defined random variable in the Hilbert space spanned by $X(t),\ t \in \mathcal{T}$. Let

$$Y_i = L_i F + \epsilon_i, \quad i = 1, \ldots, n$$

where ϵ is as before. Here and elsewhere it is assumed that the $n \times M$ matrix T with $i\nu$th element $L_i\phi_\nu$ is of rank M (that is, least squares regression on $\phi_1, \ldots, \phi_M$ is uniquely defined). Let L_0 be another bounded linear functional on $\mathcal{H}$. The goal is to estimate L_0F (again a well-defined random variable), given $Y_i = y_i$, $i = 1, \ldots, n$. Call the estimate $\widehat{L_0F}$. Let $\widehat{L_0F}$ be the minimum variance, linear, unbiased with respect to θ estimate. Then

$$\widehat{L_0F} = \sum_{j=1}^{n} \beta_j y_j$$

where $\beta = (\beta_1, \ldots, \beta_n)$ is chosen to minimize

$$E(\widehat{L_0F} - L_0F)^2$$

subject to

$$E[(\widehat{L_0F} - L_0F)|\theta] = 0.$$

We have the following theorem.

THEOREM 1.5.2.

$$\widehat{L_0 F} = L_0 f_\lambda$$

where f_λ is the minimizer in $\mathcal{H}$ of

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - L_i f)^2 + \lambda\|P_1 f\|^2$$

with $\lambda = \sigma^2/nb$.

This theorem says that if you want to estimate $L_0 F$, then you can find the generalized smoothing spline f_λ for the data and take $L_0 f_\lambda$ as the estimate.

A practical application of this result that we will return to later is the estimation of $f'(t)$ given data:

$$y_i = f(t_i) + \epsilon_i, \quad i = 1, 2, \ldots, n.$$

One can take the smoothing spline for the data and use its derivative as an estimate of f'.

The second, or "random effects model with an improper prior," leads to the same smoothing spline, and goes as follows:

$$\begin{aligned} F(t) &= \sum_{\nu=1}^{M} \theta_\nu \phi_\nu(t) + b^{1/2} X(t), \\ Y_i &= L_i F + \epsilon_i, \end{aligned} \tag{1.5.8}$$

where everything is as before except θ, which is assumed to be $\mathcal{N}(0, aI)$, and we will let $a \to \infty$.

THEOREM 1.5.3. *Let*

$$\hat{F}_a(t) = E(F(t)|Y_i = y_i,\ i = 1, \ldots, n)$$

and let f_λ be the minimizer of

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - L_i f)^2 + \lambda\|P_1 f\|^2$$

with $\lambda = \sigma^2/nb$. Then, for each fixed t,

$$\lim_{a\to\infty} \hat{F}_a(t) = f_\lambda(t).$$

To prove this, by the correspondence between $\mathcal{H}_{R^1}$ and the Hilbert space spanned by $X(t), t \in \mathcal{T}$, we have

$$\begin{aligned} E(L_i X)X(t) &= L_{i(s)} R^1(s,t) = \xi_i(t), \\ EL_i X L_j X &= L_{i(s)} L_{j(t)} R^1(s,t) = <\xi_i, \xi_j> . \end{aligned}$$

Then, letting $Y = (Y_1, \dots, Y_n)'$, we have

$$EF_a(t)Y = aT \begin{pmatrix} \phi_1(t) \\ \vdots \\ \phi_M(t) \end{pmatrix} + b \begin{pmatrix} \xi_1(t) \\ \vdots \\ \xi_n(t) \end{pmatrix},$$

$$EYY' = aTT' + b\Sigma + \sigma^2 I \tag{1.5.9}$$

where T and Σ are as in (1.3.7) and (1.3.9).

Setting $\lambda = \sigma^2/nb$, $\eta = a/b$ and $M = \Sigma + n\lambda I$ gives

$$\begin{aligned} E(F_a(t)|Y = y) &= (\phi_1(t), \dots, \phi_M(t))\eta T'(\eta TT' + M)^{-1}y \\ &\quad + (\xi_1(t), \dots, \xi_n(t))(\eta TT' + M)^{-1}y. \end{aligned} \tag{1.5.10}$$

Comparing (1.3.14), (1.3.15), and (1.5.8), it only remains to show that

$$\lim_{\eta \to \infty} \eta T'(\eta TT' + M)^{-1} = (T'M^{-1}T)^{-1}T'M^{-1} \tag{1.5.11}$$

and

$$\lim_{\eta \to \infty} (\eta TT' + M)^{-1} = M^{-1}(I - T(T'M^{-1}T)^{-1}T'M^{-1}). \tag{1.5.12}$$

It can be verified that

$$\begin{aligned} &(\eta TT' + M)^{-1} \\ &\quad = M^{-1} - M^{-1}T(T'M^{-1}T)^{-1}\left\{I + \eta^{-1}(T'M^{-1}T)^{-1}\right\}^{-1} T'M^{-1}, \end{aligned}$$

expanding in powers of η and letting $\eta \to \infty$ completes the proof.