Gaussian Processes on $\mathbb{R}^n$

Robert L. Wolpert
Department of Statistical Science
Duke University, Durham, NC, USA
Version: October 27, 2016, 09:22

1 Isotropic Covariance Functions

Let $\{Z(s)\}$ be a Gaussian process on $\mathbb{R}^n$, *i.e.*, a collection of jointly normal random variables $Z(s)$ associated with n-dimensional locations $s \in \mathbb{R}^n$. The joint distribution of $\{Z(s)\}$ depends only on the means $\mu(s) = \mathsf{E}Z(s)$ and the covariances $C(s,t) = \mathsf{E}\big(Z(s) - \mu(s)\big)\big(Z(t) - \mu(t)\big)$.

The process is called *stationary* or *translation invariant* if the distribution wouldn't change under a rigid translation of the entire collection of locations, *i.e.*, if $\mu(s) = \mu(s+h)$ and $C(s+h, t+h) = C(s,t)$ for all h; in this case $\mu(s) \equiv \mu$ is constant and $C(s,t) = C(s-t, 0)$ can only depend on the difference $h = (s-t)$ between the two locations, so must be of the form $C(s,t) = C_0(s-t)$ for some function $C_0(h) = C(h, 0)$ on $\mathbb{R}^n$. Not just any function $C_0(h)$ can be a covariance function; let's see what the choices are.

It's easy to see that the function c must be *even*, *i.e.*, must satisfy $C_0(h) = C_0(-h)$, since $C_0(s-t) = \mathsf{E}\big(Z(s) - \mu(s)\big)\big(Z(t) - \mu(t)\big) = C_0(t-s)$. But more is true: if $\{s_j\}$ any collection of locations, then complex linear combinations $a^\intercal(Z-\mu) = \sum a_j(Z_j - \mu_j)$ of the centered random variables $Z_j = Z(s_j)$ (with means $\mu_j = \mu(s_j)$) must have nonnegative squared modulus $\mathsf{E}\big|\sum a_j(Z_j - \mu_j)\big|^2 = \sum a_j C_0(s_j - s_k)\bar{a}_k \geq 0$ for every set of complex numbers $\{a_j\} \subset \mathbb{C}$. A function $C_0(h)$ is called *positive semi-definite* if it always satisfies the inequality $\sum_{jk} a_j C_0(s_j - s_k)\bar{a}_k \geq 0$ for any locations s_j and complex numbers a_j; this is equivalent to asking that $C_0(h) = C_0(-h)$ for every $h \in \mathbb{R}^n$ and that $\sum a_j C_0(s_j - s_k) a_k \geq 0$ for all *real* numbers $a_j \in \mathbb{R}$. One way to get a symmetric positive semi-definite function $C_0(h)$ is by taking the Fourier transform

$$C_0(h) = \int_{\mathbb{R}^n} e^{ih\cdot\omega} G(\omega)\, d^n\omega = \int_{\mathbb{R}^n} e^{ih\cdot\omega} G(d\omega) \tag{1}$$

of any positive function $G \in L_1(\mathbb{R}^n)$ or, more generally, of any finite positive Borel measure $G(d\omega)$, because then

$$\begin{aligned}\sum_{jk} a_j C_0(s_j - s_k)\bar{a}_k &= \int_{\mathbb{R}^n} \sum_{jk} (a_j e^{is_j\cdot\omega})\overline{(a_k e^{is_k\cdot\omega})} G(d\omega) \\ &= \int_{\mathbb{R}^n} \Big|\sum_j a_j e^{is_j\cdot\omega}\Big|^2 G(d\omega) \geq 0.\end{aligned}$$

It turns out that this is the *only* way to get one— that every positive semi-definite function can be written in this form for some finite positive measure $G(d\omega)$, called the *spectral measure* (if $G(d\omega) = G(\omega)\,d\omega$ is absolutely continuous, $G(\omega)$ is called the *spectral density*). Known as "Bochner's Theorem," this result is really just the Fourier inversion formula in an unfamiliar setting:

$$G(\omega) = (2\pi)^{-n} \int_{\mathbb{R}^n} e^{-ih\cdot\omega} C_0(h)\, d^n h.$$

Since the process $\{Z(s)\}$ is real-valued, the spectral density $G(\omega) = G(-\omega)$ must be an even function and so we can write

$$\begin{aligned} C_0(h) &= \int_{\mathbb{R}^n} \cos(h\cdot\omega) G(\omega)\, d^n\omega \\ G(\omega) &= (2\pi)^{-n} \int_{\mathbb{R}^n} \cos(h\cdot\omega) C_0(h)\, d^n\omega \end{aligned}$$

If the Gaussian process is also *isotropic*, or invariant under rotations, then $G(\omega) = g(|\omega|)$ must also be invariant under rotations and depend only on the length $r = |\omega|$ of the vector $\omega \in \mathbb{R}^n$. In this case we can simplify these integrals by transforming to polar coordinates.

1.1 Polar Coordinates for Probabilists

Polar coordinates are a familiar tool in two-dimensional integrals, where the change of variables from $x \in \mathbb{R}^2$ to $r = \sqrt{x_1^2 + x_2^2}$ and $\theta = \arctan x_2/x_1$ (so $x_1 = r\cos\theta$, $x_2 = r\sin\theta$) and a change from d^2x to $r\,dr\,d\theta$ lead to simple expressions for the integrals of radial functions. Equivalently, we can let σ have a uniform probability distribution (denoted by $d\sigma$) over the unit circle $S^1 = \{x : x_1^2 + x_2^2 = 1\}$, and change variables from $x = (x_1, x_2)$ to (r, σ), with $d^2x = dx_1\,dx_2$ replaced by $2\pi r\,dr\,d\sigma$.

In three dimensions the first polar approach has its analogue in the Euler angles, while the second is simpler with uniform measure for σ on the unit sphere $S^2 \subset \mathbb{R}^3$, with $d^3x = dx_1\,dx_2\,dx_3$ replaced by $4\pi r^2\,dr\,d\sigma$. Notice that $2\pi r$ and $4\pi r^2$ are the circumference of the circle and the area of the sphere of radius r, respectively. In any number n of dimensions the sphere rS^{n-1} of radius r has area $2\pi^{n/2} r^{n-1}/\Gamma(n/2)$ (the derivative w.r.t. r of the volume $\omega_n r^n$ of the ball of radius r, where $\omega_n = \pi^{n/2}/\Gamma(1+n/2)$ is the volume of the unit ball), and we can again again evaluate integrals in polar coordinates with the uniform probability distribution $d\sigma$ for $\sigma \in S^{n-1} \subset \mathbb{R}^n$, and $d^n x = \frac{2\pi^{n/2}}{\Gamma(n/2)} r^{n-1}\,dr\,d\sigma$. This makes it easy to compute integrals of radial functions; for functions that also depend on one or more of the components x_j, it is sometimes helpful to note that the squares $\{{\sigma_j}^2\}$ have a Dirichlet $\mathsf{Di}(\frac{1}{2}, ..., \frac{1}{2})$ joint distribution, so each σ_j is distributed as the square root of a $\mathsf{Be}(\frac{1}{2}, \frac{n-1}{2})$ random variable.

1.2 Evaluating $C_0(h)$

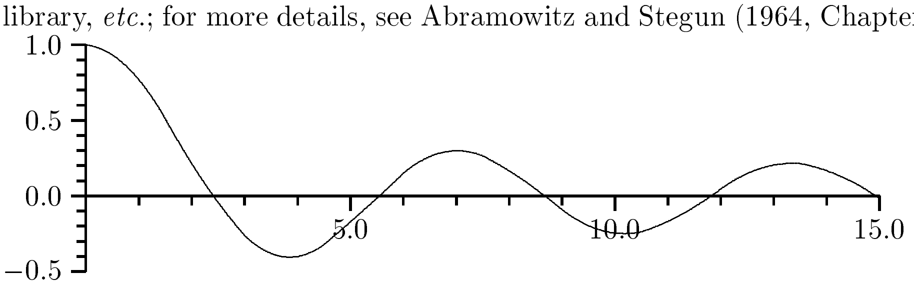
Switching to polar coordinates $r = |\omega| \geq 0$ and $\sigma = \omega/|\omega| \in S^{n-1}$ (where $d\sigma$ denotes the uniform probability measure on the unit sphere S^{n-1} in $\mathbb{R}^n$), and noting that the component $\sigma_h = \sigma \cdot h/|h|$ of $\sigma \in S^{n-1}$ in the direction h again has the same distribution as the square root of a $\mathsf{Be}(\frac{1}{2}, \frac{n-1}{2})$ random variable, writing ρ for $|h|$ and $g(|\omega|)$ for $G(\omega)$,

$$
\begin{aligned}
C_0(h) &= \int_{\mathbb{R}^n} \cos\big(h\cdot\omega\big) g(|\omega|)\, d^n\omega \\
&= \iint_{\mathbb{R}_+\times S^{n-1}} \cos\big(r\rho\sigma_h\big) g(r) \frac{2\pi^{n/2}\, r^{n-1}}{\Gamma(n/2)}\, dr\, d\sigma \\
&= \int_{\mathbb{R}_+}\int_0^1 \cos\big(r\rho\sqrt{u}\big) g(r) \frac{2\pi^{n/2}\, r^{n-1}}{\Gamma(n/2)} \frac{\Gamma(n/2)}{\Gamma(\frac{1}{2})\,\Gamma(\frac{n-1}{2})} u^{1/2-1}(1-u)^{(n-1)/2-1}\, dr\, du \\
&= \int_0^\infty \rho\,(2\pi r/\rho)^{\nu+1}\, J_\nu(r\rho)\, g(r)\, dr, \qquad \nu \equiv \frac{n}{2}-1 && (2) \\
&= \int_0^\infty (r\rho/2)^{-\nu}\Gamma(\nu+1) J_\nu(r\rho)\,\gamma(dr) && (3) \\
&= \begin{cases} \int_0^\infty 2\cos(r\rho) g(r)\, dr & \text{if } n=1 \\ \int_0^\infty 2\pi r\, J_0(r\rho)\, g(r)\, dr & \text{if } n=2 \\ \int_0^\infty \rho(2\pi r/\rho)^{3/2} J_{1/2}(r\rho)\, g(r)\, dr & \text{if } n=3 \end{cases}
\end{aligned}
$$

where

$$
J_\nu(z) = \frac{(z/2)^\nu}{\sqrt{\pi}\,\Gamma(\nu+1/2)} \int_0^\pi \cos(z\cos\theta)\,\sin(\theta)^{2\nu}\, d\theta
$$

is the Bessel function of the first kind of order ν (see Watson, 1944). Bessel functions aren't as familiar as sines and cosines, but they're common in engineering and physics and are in the standard C library, the GNU Scientific library (GSL), `R`, Maple and Mathematica, MatLab, Python's SciPy library, *etc.*; for more details, see Abramowitz and Stegun (1964, Chapter 9). Here's a plot of $J_0(z)$:



The plot of $J_0(z)$ looks a little like a sine or cosine, but falls off like $1/\sqrt{z}$ as $z\to\infty$.

The most general isotropic covariance is given in (3), with the absolutely continuous measure $g(r)\,\frac{2\pi^{n/2}}{\Gamma(n/2)}\,r^{n-1}dr$ replaced by an arbitrary positive finite measure $\gamma(dr)$ on $[0,\infty)$. Any isotropic covariance function may be approximated by one with a discrete spectral measure $\gamma(dr)=\sum\gamma_j\delta_{r_j}(dr)$ assigning mass γ_j to finitely many points r_j:

$$
\begin{aligned}
C_0(h) = C(\rho) &\approx \sum_j (2/r_j\rho)^\nu\Gamma(\nu+1) J_\nu(r_j\rho)\gamma_j && (4) \\
&= \begin{cases} \sum_j \gamma_j \cos(r_j\rho) & \text{if } n=1 \\ \sum_j \gamma_j J_0(r_j\rho) & \text{if } n=2 \\ \sum_j \gamma_j \sqrt{\pi/2r_j\rho}\, J_{1/2}(r_j\rho) & \text{if } n=3 \end{cases}
\end{aligned}
$$

but a more common approach is to choose small parametric families of densities $g^\theta(r)$ or measures $\gamma^\theta(dr)$.

We can recover a spectral density $g(r) = G(\omega)$ (for $r = |\omega|$) through the Fourier inversion formula, using polar coordinates with $\rho = |h| \in \mathbb{R}_+$ and $\sigma = h/|h| \in S^{n-1}$:

$$\begin{aligned} g(r) = G(\omega) &= \frac{1}{(2\pi)^n} \int \cos\big(-h\cdot\omega\big) C_0(h)\, d^n h \\ &= \frac{1}{(2\pi)^n} \iint_{\mathbb{R}_+ \times S^{n-1}} \cos\big(-r\rho\,\sigma_\omega\big) C(\rho) \frac{2\pi^{n/2}\,\rho^{n-1}}{\Gamma(n/2)}\, d\rho\, d\sigma \\ &= \int_0^\infty r(\rho/2\pi r)^{n/2} J_\nu(r\rho) C(\rho)\, d\rho, \qquad \nu \equiv \frac{n}{2} - 1 \qquad (5) \\ &= \begin{cases} \int_0^\infty \frac{2}{\pi} \cos(r\rho) C(\rho)\, d\rho & \text{if } n = 1 \\ \int_0^\infty (\rho/2\pi) J_0(r\rho)\, C(\rho)\, d\rho & \text{if } n = 2 \\ \int_0^\infty r(\rho/2\pi r)^{3/2} J_{1/2}(r\rho)\, C(\rho)\, d\rho & \text{if } n = 3 \end{cases} \qquad (6) \end{aligned}$$

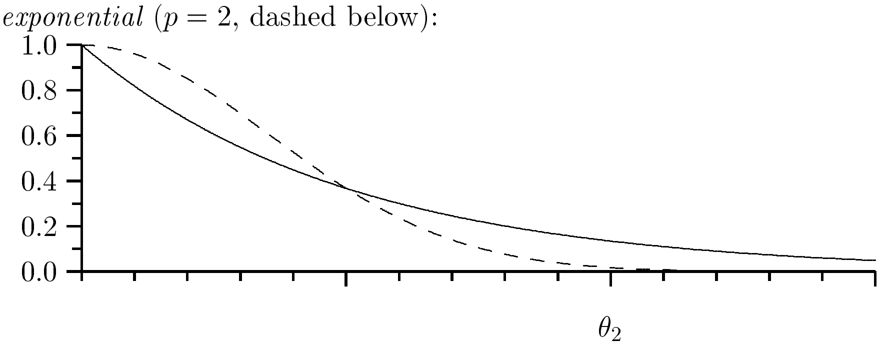
1.3 Parametric Families of Covariance Functions

It is hard to imagine what $C_0(h)$ would look like for different choices of $g(r)$; a simple approach is to take whatever symmetric functions $G(u)$ whose Fourier transforms we can find, and see what we get. Here are some commonly used covariance families, mostly in $n = 2$ dimensions. In each case $\theta_1 = C_0(0)^{1/2}$ is an overall level parameter (in the same units as $Z(s)$), the marginal standard deviation, and θ_2 is a characteristic length scale (in the same units as s):

- Power Exponential family

$$C(\rho \mid \theta, p) = {\theta_1}^2 \exp\{-|\rho/\theta_2|^p\},\ 0 < p \le 2$$

Two notable covariograms in this family are the *exponential* ($p = 1$, solid below) and the *squared exponential* ($p = 2$, dashed below):

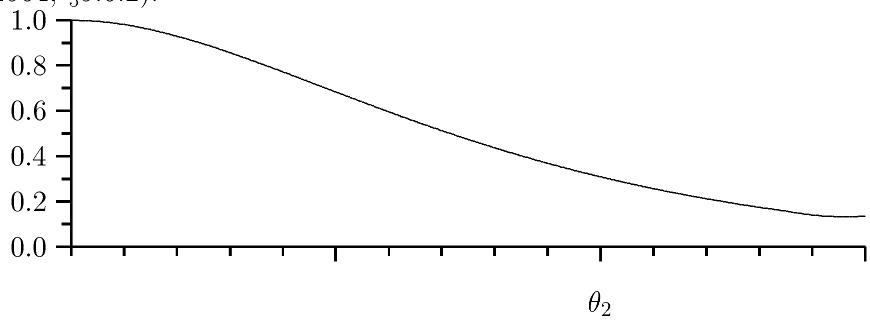


Notice that the exponential has a negative derivative at $z = 0$, so it falls off quickly at first, then slowly levels off, while the squared exponential has zero derivative near $z = 0$ then falls off very quickly. From (6) it follows that the exponential has spectral density function $g(r) = ({\theta_1}^2{\theta_2}^2/2\pi)/\big(1 + r^2{\theta_2}^2\big)^{3/2}$, proportional to a bivariate Cauchy density function, while the squared exponential has spectral density $g(r) = ({\theta_1}^2{\theta_2}^2/4\pi)\exp\big(-r^2{\theta_2}^2/4\big)$, proportional to a normal density.

- Matérn Family

$$C(\rho \mid \theta) = {\theta_1}^2 \frac{2^{1-\theta_3}}{\Gamma(\theta_3)} \Big(\sqrt{2\theta_3}\frac{\rho}{\theta_2}\Big)^{\theta_3} K_{\theta_3}\Big(\sqrt{2\theta_3}\frac{\rho}{\theta_2}\Big) \qquad (7)$$

where $K_\nu(z)$ is the modified Bessel function of the third kind of order ν (Abramowitz and Stegun, 1964, §9.6.2):



The displayed plot has shape parameter $\theta_3 = 2$. The Matérn class is quite flexible and includes the exponential family (with $\theta_3 = \frac{1}{2}$), the squared exponential family (in the limit as $\theta_3 \to \infty$), and many others. Note authors differ in how they parameterize the Matérn class: Rasmussen and Williams (2006, *p.* 84) use $\nu = \theta_3$ and $\ell = \theta_2\sqrt{8\theta_3}$, for example, while Stein (1999) uses $\nu = \theta_3$ and $\alpha = 1/2\theta_2$.

In n dimensions the Matérn spectral density function is

$$g(r) = \frac{{\theta_1}^2}{\Gamma(\theta_3)}\Big(\frac{{\theta_2}^2}{2\pi\theta_3}\Big)^{n/2}\Big(1+\frac{{\theta_2}^2}{2\theta_3}r^2\Big)^{-\theta_3-n/2},$$

proportional to the familiar n-variate Student's t density function with $2\theta_3$ degrees of freedom and scale $\sigma = \sqrt{2}/\theta_2$. This lends more insight into how the Matérn reduces to the exponential when $\theta_3 = 1/2$ and to the squared exponential when $\theta_3 \to \infty$.

Sample paths of Matérn processes have derivatives of orders less than θ_3, but not of order θ_3 or higher. Thus all three parameters have clear interpretations: ${\theta_1}^2$ is the marginal variance, θ_2 is the distance scale, and θ_3 is a smoothness parameter.

Modified Bessel functions K_ν for any $\nu > 0$ are available in most computing environments (GSL, `R`, Maple, Mathematica, MatLab, SciPy, *etc.*), but for the special case of half-odd-integers $\nu = n{+}1/2$ they have simple polynomial expressions of degree n in $(1/z)$ (Abramowitz and Stegun, 1964, §10.2.15). For $\nu > 0$,

$$K_{\pm\nu}(z) = e^{-z}\sqrt{\pi/2z}\,\sum_{j=0}^{n}\frac{(j+n)!}{j!(n-j)!}\,(2z)^{-j}.$$

For example,

$$\begin{aligned} K_{1/2}(z) &= \sqrt{\pi/2z}\ e^{-z} \\ K_{3/2}(z) &= \sqrt{\pi/2z}\ e^{-z}[1+1/z] \\ K_{5/2}(z) &= \sqrt{\pi/2z}\ e^{-z}[1+3/z+3/z^2], \end{aligned}$$

leading to simple expressions for the covariance of Eqn (7). These can also be computed from the simple expression for $K_{\pm 1/2} = \sqrt{\pi/2z}\ e^{-z}$ and the recurrence relation

$$K_{\nu+1}(z) = K_{\nu-1}(z) + \frac{2\nu}{z}K_\nu(z).$$

For this reason it is common to fix parameter $\theta_3 = 3/2$ or $5/2$ in the Matérn class imposing an intentionally selected degree of smoothness, and the estimate ${\theta_1}^2$ and θ_2 from data. For example, for $\theta_3 = 3/2$ we have

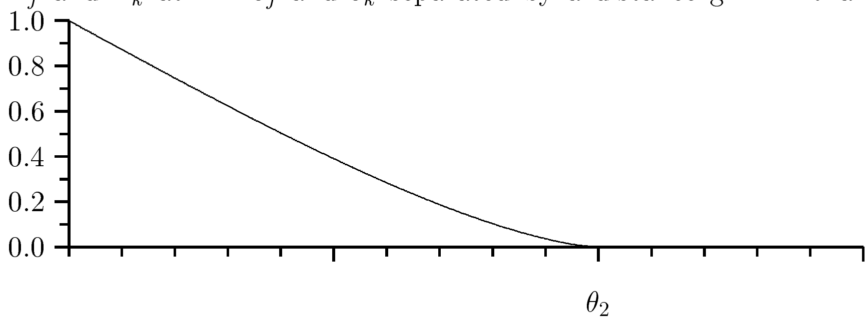
$$C(\rho \mid \theta) = {\theta_1}^2 e^{-\rho/\theta_2}\big[1 + \rho/\theta_2\big].$$

For an ardent advocate of this class see Stein (1999, §1.6,1.7,2.7,2.10).

- Spherical Family

$$C(\rho \mid \theta) = \begin{cases} {\theta_1}^2 \left[1 - \frac{2}{\pi}\left(\frac{\rho}{\theta_2}\sqrt{1 - (\frac{\rho}{\theta_2})^2} + \sin^{-1}\frac{\rho}{\theta_2}\right)\right] & \text{for } \rho < \theta_2 \\ 0 & \text{for } \rho \ge \theta_2 \end{cases}$$

The spherical covariance function is proportional to the area of intersection for two discs of diameter θ_2 with centers separated by distance ρ. In this model the Gaussian quantities Z_j and Z_k at loci s_j and s_k separated by a distance greater than θ_2 will be independent.



This is not quite linear. Like the exponential, it has a negative slope at $z = 0$ and falls off rapidly at first; like the squared exponential, it falls off rapidly later and in fact reaches zero. The spectral density, while available in closed form, isn't illuminating; it's best to think of the spherical process as a convolution or moving average of Gaussian white noise, integrated at each locus over the surrounding ball of diameter θ_2.

The same idea works in any dimension d: a constant ${\theta_1}^2/\omega_d(\theta_2/2)^d$, times the volume of intersection of two d-balls of diameter θ_2 whose centers are separated by a distance ρ, is a valid covariance function in dimensions $n \le d$:

$$\begin{aligned} C(\rho \mid \theta) &= \frac{2\,{\theta_1}^2}{\omega_d(\theta_2/2)^d}\int_{\rho/2}^{\theta_2/2}\Big\{\omega_{d-1}\big[(\theta_2/2)^2 - z^2\big]^{(d-1)/2}\Big\}\,dz \\ &= {\theta_1}^2 I_{(1-\rho^2/{\theta_2}^2)}\big(\tfrac{d+1}{2}, \tfrac{1}{2}\big), \quad 0 \le \rho \le \theta_2 \end{aligned}$$

where

$$\begin{aligned} I_z(\alpha, \beta) &:= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\int_0^z x^{\alpha-1}(1-x)^{\beta-1}\,dx \\ &= 1 - I_{1-z}(\beta, \alpha) \end{aligned}$$

is the normalized incomplete Beta function (Abramowitz and Stegun, 1964, §6.6.2), the CDF of the $\mathsf{Be}(\alpha, \beta)$ distribution. For $d = 2$ this reduces to the equation above, for $d = 3$ to $C(\rho \mid \theta) = ({\theta_1}^2/2)[(\rho/\theta_2)^3 - 3(\rho/\theta_2) + 2]\mathbf{1}_{\{\rho \le \theta_2\}}$, and for $d = 1$ to ${\theta_1}^2(1 - \rho/\theta_2)\mathbf{1}_{\{\rho \le \theta_2\}}$. It even works for non-integer real $d \ge n$.

- Cubic (Integrated Linear)
  If X_t is a stationary Gaussian process (in $n = 1$ dimension) with covariance function $\gamma(\rho)$, and $g(t)$ is an integrable function, then $Z_t := \int_{\mathbb{R}} g(t-s) X_s \, ds = g \star X(t)$ is also a stationary Gaussian process, with covariance $\gamma_g(\rho) = \iint g(s)\gamma(s-t+\rho)g(t)\, ds\, dt$, and typically with smoother sample paths than X_t. For example, if X_s has the Spherical covariance with $d = 1$ above and if $g(\rho) = (3/4\theta_2{}^2)\mathbf{1}_{\{\rho \le \theta_2\}}$, then $Z = g \star X$ has covariance

$$\begin{aligned}\gamma(\rho) &= \frac{3\theta_1{}^2}{4\theta_2{}^2}\int_0^{\theta_2}\int_{\rho}^{\rho+\theta_2}\big[1-|s-t|/\theta_2\big]\mathbf{1}_{\{|s-t|\le\theta_2\}}\,dt\,ds \\ &= \begin{cases}(\theta_1/2)^2\big[3(\rho/\theta_2)^3-6(\rho/\theta_2)^2+4\big] & 0\le\rho\le\theta_2\\ (\theta_1/2)^2(2-\rho/\theta_2)^3 & \theta_2\le\rho\le 2\theta_2\\ 0 & 2\theta_2\le\rho<\infty,\end{cases}\end{aligned}$$

  a compactly-supported piecewise-cubic function of $\rho = |s-t|$ called the *nonnegative cubic covariance* when it is used in computer modeling.

1.4 Mean-Square Continuity & Smoothness

An L_2 process Z_t is called *mean square continuous* at $t \in \mathcal{T}$ if

$$\lim_{s\to t} \mathsf{E}|Z_s - Z_t|^2 = 0.$$

Evidently this property depends only on the covariance function $C(s,t)$, since

$$\mathsf{E}|Z_s - Z_t|^2 = C(s,s) - C(s,t) - C(t,s) + C(t,t)$$

or, for stationary processes,

$$= 2[C_0(0) - C_0(s-t)].$$

Thus, a stationary L_2 process Z is mean square continuous if and only if its covariance function is continuous at zero. This implies C_0 is uniformly continuous, since

$$\begin{aligned}|C_0(s) - C_0(t)| &= |\mathsf{Cov}\big[(Z_s - Z_t), Z_0\big]| \\ &\le \{\mathsf{V}(Z_s - Z_t)\}^{1/2}\{\mathsf{V}(Z_0)\}^{1/2} \\ &= |C_0(0) - C_0(s-t)|^{1/2}\ |2C_0(0)|^{1/2}.\end{aligned}$$

Mean-square continuity doesn't imply path continuity— for example, the two-state continuous-time Markov process that alternates between plus and minus one after independent $\mathsf{Ex}(1)$ waiting times has continuous covariance $C_0(h) = e^{-2|h|}$, and paths are almost-surely continuous at each specific time t, but paths do jump with probability one.

A real-valued L_2 process Z_t indexed by $t \in \mathcal{T} \subseteq \mathbb{R}$ is called *mean square differentiable* with derivative Z_t' at $t \in \mathcal{T}$ if

$$\lim_{h\to 0} \mathsf{E}\left|\frac{Z_{t+h} - Z_t}{h} - Z_t'\right|^2 = 0.$$

This too depends only on the covariance function $C_0(s-t)$ for stationary processes, since the covariance C_h of the process $Z_h(t) := (Z_{t+h} - Z_t)/h$ is

$$\begin{aligned} C_h(t) &= \mathsf{E}\left[\frac{Z_h - Z_0}{h}\ \frac{Z_{t+h} - Z_t}{h}\right] \\ &= h^{-2}\left\{2C_0(t) - C_0(t+h) - C_0(t-h)\right\} \\ &\to -C_0''(t) \quad \text{as } h \to 0 \end{aligned}$$

if C_0 is twice differentiable at zero. If so, then Z_t is mean square differentiable, *i.e.*, Z_h converges as $h \to 0$ to a limiting process Z_t' with covariance

$$\mathsf{Cov}\big[Z_s', Z_t'\big] = -C_0''(s-t).$$

Similarly, Z_t is m-times mean square differentiable if and only if C_0 has $2m$ derivatives at zero, and in that case

$$\mathsf{Cov}\big[Z_s^{(m)}, Z_t^{(m)}\big] = (-1)^m C_0^{(2m)}(s-t).$$

The same idea works for processes Z_t indexed by vector arguments $t \in \mathcal{T} \subset \mathbb{R}^n$ but now the first derivative (or gradient) Z_t' will be an n-dimensional vector, the second derivative (or Hessian) Z_t'' will be an $n \times n$ matrix, and higher-order derivatives will be higher-order tensors.

Mean-square continuity and differentiability for isotropic random fields can also be taken from the spectral representation of the covariance. From

$$\begin{aligned} C_0(h) &= \int_{\mathbb{R}^n} \cos\big(h \cdot \omega\big) g(|\omega|)\, d\omega, \\ &= \iint_{\mathbb{R}_+ \times S^{n-1}} \cos\big(r\rho\sigma_h\big) g(r) \frac{2\pi^{n/2}\, r^{n-1}}{\Gamma(n/2)}\, dr\, d\sigma \end{aligned}$$

we see that

$$-C_0''(h) = \int_{\mathbb{R}^n} \omega\omega' \cos\big(h \cdot \omega\big) g(|\omega|)\, d\omega$$

is well-defined and finite provided that

$$\int_{\mathbb{R}^n} |\omega|^2 g(|\omega|)\, d\omega = \int_0^\infty n\, \omega_n\, r^{n+1} g(r)\, dr < \infty$$

and, similarly, Z_t with $t \in \mathcal{T} \subset \mathbb{R}^n$ is m-times mean square differentiable if and only if

$$\int_0^\infty r^{n+2m-1} g(r)\, dr < \infty.$$

I suspect but don't know that there is a Sobolev-type result that Z_t will have m-times continuously differentiable sample paths if a somewhat stricter condition holds, like $\int_0^\infty r^{n+2m} g(r)\, dr < \infty$.

In any dimension the covariance function $C(\rho \mid \theta, p)$ of the Power Exponential family of Section (1.3) is continuous but it is not twice differentiable for any $p < 2$, so sample paths are mean-square continuous but not differentiable even once. For $p = 2$ the covariance is infinitely differentiable, so sample paths are m-times mean square differentiable for all m.

The Matérn family in Section (1.3) has spectral density proportional to $(1+\theta_2^2 r^2/2\theta_3)^{-\theta_3-n/2}$, so it will be m-times mean square differentiable if and only if $r^{n+2m-1-2\theta_3-n}$ is integrable at infinity, *i.e.*, if and only if $\theta_3 > m$, in any dimension n.

Near $\rho \approx 0$ the covariance for the Spherical Family is $C(\rho \mid \theta) \asymp {\theta_1}^2(1-c\rho)$, for constant $c = d\Gamma(d/2)/\theta_2^2\sqrt{\pi}\Gamma((d+1)/2) > 0$, so it is continuous but not differentiable and the process is mean square continuous but not mean square differentiable.

2 Constructing GPs from their Spectral Measures

The integral representations of isotropic covariance functions can be exploited to *construct* random fields with specified covariance structure, in at least two different ways.

2.1 White-Noise Convolutions

Many processes may be constructed similarly as kernel integrals of standard Gaussian white noise,

$$Z(h) = \int_{\mathbb{R}^n} k(h-s)\,\zeta(ds);$$

where "standard" means that $\mathsf{E}[\zeta(ds)] = 0$ and $\mathsf{E}[\zeta(ds)^2] = ds$ (more formally, that ζ is countably additive with $\zeta(A) \sim \mathsf{No}(0, |A|)$ for Borel $A \subset \mathbb{R}^n$ of finite Lebesgue measure $|A|$). The covariance is

$$C_0(h) = \mathsf{E}\big[Z(0)\overline{Z(h)}\big] = \int_{\mathbb{R}^n} k(h-s)\,\overline{k(-s)}\,ds$$

with spectral density

$$\begin{aligned} G(\omega) &= (2\pi)^{-n} \int e^{-i\omega\cdot h} C_0(h)\,dh \\ &= (2\pi)^{-n} \iint e^{-i\omega\cdot h} k(h-s)\,\overline{k(-s)}\,ds\,dh \\ &= (2\pi)^{-n} \left| \int e^{-i\omega\cdot x} k(x)\,dx \right|^2. \end{aligned}$$

Thus an isotropic kernel may be computed from the spectral density as

$$k(x) = (2\pi)^{-n/2} \int e^{i\omega\cdot x} G(\omega)^{1/2}\,d^n\omega$$

or, in polar coordinates,

$$\begin{aligned} k(\rho) &= \int_0^\infty r^{\nu+1}\rho^{-\nu} J_\nu(r\rho) g(r)^{1/2}\,dr \\ &= \begin{cases} \int_0^\infty \sqrt{\frac{2}{\pi}}\cos(r\rho)\sqrt{g(r)}\,dr & \text{if } n=1 \\ \int_0^\infty J_0(r\rho)\,r\,\sqrt{g(r)}\,dr & \text{if } n=2 \\ \int_0^\infty J_{1/2}(r\rho)\,r^{3/2}\rho^{-1/2}\,\sqrt{g(r)}\,d\rho & \text{if } n=3 \end{cases} \end{aligned}$$

provided that the square root of the spectral density *is* the Fourier transform of a finite positive function, *i.e.*, is itself positive semidefinite. For the Matérn class, the root spectral density $\sqrt{g(r)} \propto$

$\left(1+\theta_2{}^2 r^2/2\theta_3\right)^{-(\theta_3+n/2)/2}$ will be another n-variate t density provided $\theta_3 > n/2$ and in this case, setting $\epsilon = (2\theta_3 - n)/4 > 0$, we find [check this]

$$k(\rho) = \frac{2\theta_1{}^2}{(\rho\theta_2\sqrt{2/\theta_3})^{\epsilon+n/2}}\Gamma(\epsilon+n/2)\sqrt{\Gamma(2\epsilon+n/2)}\pi^{n/4}K_\epsilon(\rho\sqrt{2\theta_3}/\theta_2)$$

leads to a moving-average kernel representation for the Matérn covariance class. In any number $n \geq 1$ of dimensions the restriction $\epsilon > 0$ entails $\theta_3 > n/2 \geq 1/2$, ruling out the exponential covariance, but the squared exponential covariance (the limiting case as $\theta_3 \to \infty$) is available in any number of dimensions, with

$$k(\rho) = \theta_1{}^2(\pi\theta_2{}^2/8\theta_3)^{-n/2}e^{-4\theta_3\rho^2/\theta_2{}^2}.$$

2.2 Spectral Representation

We generalize Hida and Hitsuda (1993, §III.2, *pp.* 42 ff). Let $\Theta \subseteq \mathbb{R}^n$ be the *spectrum*, the support of the spectral measure $G(d\omega)$ (see (1)), and let $\zeta(d\omega)$ be a Gaussian random measure on the Borel sets of Ω with control measure $G(d\omega)$ (so $\zeta(A) \sim \mathsf{No}\big(0, G(A)\big)$ for Borel sets $A \subset \Omega$ of finite G-measure). For $s \in \mathbb{R}^n$ set

$$Z_s := \int_\Omega e^{is\cdot\omega}\,\zeta(d\omega).$$

Then Z is a complex-valued Gaussian stochastic process with mean zero and autocovariance

$$\mathsf{E} Z_s \bar{Z}_t = \int_\Omega e^{i(s-t)\cdot\omega}\,G(d\omega) = C_0(s-t)$$

I think we can arrange for Z to be real-valued, but probably not ζ since if ζ were real then we would have $Z_{-t} \equiv \bar{Z}_t$.

For the special case of $n = 1$, we can start with independent real-valued Gaussian random measures $\zeta_0(d\omega)$ and $\zeta_1(d\omega)$ on $\Omega_+ = \Omega \cap \mathbb{R}_+$, each with control measure $\frac{1}{2}G(d\omega)$ restricted to Ω_+, and set

$$\zeta(A) = \left[\zeta_0(A_+) + \zeta_0(A_-)\right] + i\left[\zeta_1(A_+) - \zeta_1(A_-)\right]$$

(where of course $A_+ = A \cap \mathbb{R}_+$ and $A_- = A \cap \mathbb{R}_-$). Then ζ is a complex Gaussian measure on $\mathbb{R}$ with mean zero and covariance

$$\mathsf{E}\zeta(ds)\bar{\zeta}(dt) = \delta_s(dt)\,G(ds)$$

Set

$$Z_s = \int_\Omega e^{i\omega s}\zeta(d\omega) = \int_{\Omega_+}\cos(\omega s)\zeta_0(d\omega) - \int_{\Omega_+}\sin(\omega s)\zeta_1(d\omega) \tag{8}$$

I'm not sure how to think of this in Polar coordinates, or how to extend it to ≥ 2 dimensions, but I expect that's possible useful interesting and, probably, old. The Free Euclidean Field should be an example.

3 Green's Functions

In d dimensions the negative Laplacian is given in polar coordinates by

$$-\Delta f(r,\theta) \equiv -\left\{\frac{d-1}{r}\frac{\partial f}{\partial r} + \frac{\partial^2 f}{\partial r^2} + \frac{1}{r^2}\Delta_{S^{d-1}} f\right\}$$

where $\Delta_{S^{d-1}}$ denotes the Laplace-Beltrami operator on the $(d-1)$ sphere. The radial solutions to $-\Delta f(r) = 0$ are of the form $f(r) = a + br$ in $d = 1$ dimension, $f(r) = a + b\log r$ in $d = 2$ dimensions, and $f(r) = a + br^{2-d}$ in $d > 2$ dimensions. For $\beta > 0$ the formal solutions to the Green's Function equation $(-\Delta + \beta^2)\gamma(s-t) = \delta_t(s)$ are:

$$\begin{array}{ll} d=1 & g(s,t) = \frac{1}{2\beta}e^{-|s-t|} \\ d=2 & g(s,t) = \frac{1}{2\pi}\log|s-t| \\ d=3 & g(s,t) = \frac{1}{4\pi}|s-t|^{-1} \\ d>3 & g(s,t) = \frac{\Gamma(d/2)}{2\pi^{d/2}}|s-t|^{2-d} \end{array}$$

4 Compact Ranges

Fourier transforms and harmonic analysis are most commonly studied on Euclidean space $\mathbb{R}^d$, but many of the ideas extend to bounded spaces like the closed unit ball $B^d \subset \mathbb{R}^d$ or its boundary the unit sphere S^{d-1} (the disc and circle, for $d = 2$), or even to locally compact groups.

In this section we'll consider functions on the unit ball B^d. Because it is compact, functions in $L_2(B^d)$ will have Fourier *series* with a discrete index set, rather than the Fourier *transforms* of Section (1).

4.1 Fourier-Bessel Expansions in $d = 2$ Dimensions

The key for $d = 2$ is the orthogonality relationship for Bessel functions of the first kind (Abramowitz and Stegun, 1964, §9.1):

$$\int_0^1 J_\alpha(rj_{\alpha,n})\, J_\alpha(rj_{\alpha,n'})\, r\, dr = \tfrac{1}{2}\delta_{nn'}[J_{\alpha+1}(j_{\alpha,n})]^2,$$

where $j_{\alpha,n}$ is the n^{th} positive zero of $J_\alpha(x)$, for any $n, n' \in \mathbb{Z}$ and any $\alpha \in \mathbb{R}$. For integers α, $J_\alpha(-x) = (-1)^\alpha J_\alpha(x)$, so $[J_{\alpha+1}(j_{\alpha,n})]^2 = [J_{|\alpha|+1}(j_{\alpha,n})]^2$ and $j_{\alpha,n} = j_{-\alpha,n}$. It follows that for $m \in \mathbb{Z}$ and $n \in \mathbb{N}$ the functions

$$\phi_{mn}(r,\theta) \equiv \frac{1}{\sqrt{\pi}\, J_{|m|+1}(j_{m,n})} J_m(rj_{m,n})\, e^{im\theta} \tag{9}$$

are orthonormal in $L_2(B^2)$. In fact they form a CONS, so any function $f \in L_2(B^2)$ has a convergent expansion

$$f(r,\theta) = \sum_{\substack{m\in\mathbb{Z}\\ n\in\mathbb{N}}} a_{mn}\, \phi_{mn}(r,\theta) = \sum_{\substack{m\in\mathbb{Z}\\ n\in\mathbb{N}}} \frac{a_{mn}}{c_{mn}}\ J_m(rj_{m,n})e^{im\theta}$$

where

$$c_{mn} = \sqrt{\pi}\, J_{|m|+1}(j_{m,n}), \qquad a_{mn} = \int_{B^2} f(r,\theta)\overline{\phi_{mn}(r,\theta)}\, r\, dr\, d\theta. \tag{10}$$

Note a_{mn} from (10) can also be evaluated in *Cartesian* coordinates if, for example, f is known on a rectangular grid containing B^2— just multiply the integrand by $\mathbf{1}_{B^2}$ and replace $r\,dr\,d\theta$ with $dx\,dy$. Both the Bessel functions J_m and the zeros $j_{m,n}$ are available in R. From Bessel's differential equation

$$z^2\frac{\partial^2}{\partial z^2}J_m(z)+z\frac{\partial}{\partial z}J_m(z)+(z^2-m^2)J_m(z)=0$$

(Abramowitz and Stegun, 1964, §9.1.1), ϕ_{mn} is an eigenfunction of the Laplacian with eigenvalue $-j_{m,n}^2$ satisfying Dirichlet boundary conditions $\phi_{mn}(1,\theta)\equiv 0$ on $S^1=\partial B^2$, so the negative Laplacian of f is given in polar coordinates by

$$\begin{aligned}-\Delta f(r,\theta)&\equiv-\left\{\frac{1}{r}\frac{\partial f}{\partial r}+\frac{\partial^2 f}{\partial r^2}+\frac{1}{r^2}\frac{\partial^2 f}{\partial\theta^2}\right\}\\ &=\sum(j_{m,n})^2(a_{mn}/c_{mn})J_m(rj_{m,n})e^{im\theta}.\end{aligned}$$

The squared L_2 and Sobolev H_1 norms of f are

$$\langle f,f\rangle=\sum|a_{mn}|^2,\qquad \|f\|_1^2=\langle(-\Delta+I)f,f\rangle=\sum(1+j_{m,n}^2)\,|a_{mn}|^2,$$

so one reasonable scale-invariant measure of roughness would be

$$\frac{\langle\nabla f,\nabla f\rangle}{\langle f,f\rangle}=\frac{\langle-\Delta f,f\rangle}{\langle f,f\rangle}=\frac{\sum j_{m,n}^2\,|a_{mn}|^2}{\sum|a_{mn}|^2}$$

or, similarly but more generally, the ratio of squared Sobolev norms

$$\frac{\|f\|_s^2}{\|f\|_0^2}=\frac{\langle(-\Delta+I)^s f,f\rangle}{\langle f,f\rangle}=\frac{\sum(1+j_{m,n}^2)^s\,|a_{mn}|^2}{\sum|a_{mn}|^2}$$

for any $s>0$. Another measure depending only on angular roughness would be

$$\frac{\sum|m|^{2s}\,|a_{mn}|^2}{\sum|a_{mn}|^2}$$

for any $s>0$. Similar results hold for $d=3$ (using the spherical Bessel function j_α) or for other boundary conditions. Each of these measures is the moment of some "roughness" aspect, for the discrete probability distribution with masses proportional to the squared Fourier coefficients $|a_{mn}|^2$.

4.2 Fourier Expansions in $d=1$ Dimension

This section has a brief review of one-dimensional Sobolev spaces on an interval, for context and intuition.

Let $\mathcal{T}=[0,T]$ be a closed interval in $\mathbb{R}^1$. The functions

$$\phi_n(x)=\sin(n\pi x/T)\sqrt{2/T}$$

are a CONS in $L_2(\mathcal{T},dx)$ of eigenfunctions of $-\Delta$ (with eigenvalues $n^2\pi^2/T^2$), with Dirichlet boundary conditions at $\partial\mathcal{T}$, so every function $f\in L_2(\mathcal{T},dx)$ has a convergent expansion of the form

$$f(x)=\sum_{n\in\mathbb{N}}a_n\phi_n(x), \tag{11}$$

with coefficients

$$a_n := \int_{\mathcal{T}} f(x)\, \phi_n(x)\, dx.$$

For $s \in \mathbb{R}$ the closed linear span H_s of $\{\phi_n\}$ in the norm

$$\begin{aligned} \|f\|_s &:= \left\{ \sum |a_n|^2\, (1 + n^2\pi^2/T^2)^s \right\}^{\frac{1}{2}} \\ &= \langle (-\Delta + I)^s\, f, f \rangle^{\frac{1}{2}} \end{aligned} \tag{12}$$

is a Banach space (complete separable normed metric space) whose dual space is H_{-s}. The self-dual Hilbert space H_0 is just $L_2(\mathcal{T}, dx)$.

Differentiating (11) gives

$$\begin{aligned} \|f'\|_0^2 &= \Big\| \sum_{n\in\mathbb{N}} a_n \frac{n\pi}{T} \cos(n\pi \cdot /T) \sqrt{2/T} \Big\|_0^2 \\ &= \sum_{n,m\in\mathbb{N}} a_n a_m \frac{nm\pi^2}{T^2} \frac{2}{T} \int_0^T \cos(n\pi x/T) \cos(m\pi x/T)\, dx \\ &= \sum_{n\in\mathbb{N}} a_n^2\, \frac{n^2\, \pi^2}{T^2} = \sum_{n\in\mathbb{N}} |a_n|^2\, n^2\pi^2/T^2 \\ &= \|f\|_1^2 - \|f\|_0^2, \end{aligned}$$

so $f \in H_1$ if and only if $f \in H_0$ and $f' \in H_0$. Similarly one can show that $f \in H_s \Rightarrow f' \in H_{s-1}$ for any $s \in \mathbb{R}$. For $s > 0$, H_s consists of those functions with s derivatives in L_2, while H_{-s} consists of s^{th} derivatives of L_2 functions.

In particular, for $t \in \mathcal{T}$ the Dirac delta function $\delta_t \in H_{-1}$ but $\delta_t \notin H_0$. For example, for $T = \pi$ and $t = \pi/2$, $f \equiv \delta_t$ has coefficients

$$a_n \equiv \int_0^\pi f(x) \sin(nx)\, \sqrt{2/\pi}\, dx = \begin{cases} \sqrt{2/\pi} & n \equiv 1 \pmod 4 \\ -\sqrt{2/\pi} & n \equiv 3 \pmod 4 \end{cases},$$

otherwise zero, with eigenvalues n^2, so

$$\|\delta_t\|_s^2 = \frac{2}{\pi} \sum_{k=1}^{\infty} \left(1 + (2k+1)^2\right)^s.$$

This is only finite for $s < -\frac{1}{2}$, so $\delta_t \in H_s$ only for $s < -\frac{1}{2}$.

The boundary $\partial\mathcal{T}$ and boundary conditions play a subtle rôle in $H_s(\mathcal{T})$. For example, in the interior of $\mathcal{T} = [0, T]$ for $T > 0$ the function

$$f(x) = x(T - x)$$

satisfies $f''(x) \equiv -2$ and hence

$$(-\Delta + I) f(x) = x(T - x) + 2$$

and so, at least for non-negative integers s,

$$(-\Delta + I)^s f(x) = x(T-x) + 2s.$$

One might expect $f \in H_s$ for all $s > 0$, with

$$\begin{aligned}
\|f\|_s^2 &\equiv \langle (-\Delta + I)^s f,\, f\rangle \\
&\stackrel{?}{=} \int_0^T [x(T-x)+2s][x(T-x)]\,dx \\
&= T^5/30 + sT^3/3,
\end{aligned}$$

suggesting (wrongly) that $f \in H_s$ for all $s > 0$. In fact the coefficients are

$$a_n := \int_{\mathcal{T}} f(x)\,\phi_n(x)\,dx = 4\sqrt{2}T^{5/2}/(n\pi)^3$$

for odd n, and zero for even n, so

$$\begin{aligned}
\|f\|_s^2 &\equiv \sum_{n=1}^{\infty} |a_n|^2 \left(n^2\pi^2/T^2 + 1\right)^s \\
&= \frac{32T^5}{\pi^6} \sum_{k=1}^{\infty} \frac{\left((2k+1)^2\pi^2/T^2 + 1\right)^s}{(2k+1)^6}
\end{aligned}$$

which is only finite for $s < 5/2$. The problem is that even the first derivative f' fails to satisfy Dirichlet boundary conditions $f(\partial\mathcal{T}) = 0$. Integrating by parts,

$$\begin{aligned}
\int_{\mathcal{T}} f'(x)\,f'(x)\,dx &= f'(x)f(x)\Big|_0^T - \int_{\mathcal{T}} f''(x)\,f(x)\,dx \\
&= -\int_{\mathcal{T}} f''(x)\,f(x)\,dx = \langle(-\Delta)f, f\rangle \\
\int_{\mathcal{T}} f''(x)\,f''(x)\,dx &= f''(x)f'(x)\Big|_0^T - \int_{\mathcal{T}} f'''(x)\,f'(x)\,dx \\
&= 4T - f'''(x)\,f(x)\Big|_0^T + \int_{\mathcal{T}} f''''(x)\,f(x)\,dx \\
&= \langle(-\Delta)^2 f, f\rangle + 4T
\end{aligned}$$

so for $s \geq 2$ a non-zero boundary term enters the integration by parts and the naive calculation fails. Intuitively, what happens is that $f(x)$ takes the constant value of zero outside $\mathcal{T}$ and so derivatives beyond the first pick up something like a Dirac delta function *at the boundary.*

Something similar happens in $d = 2$ dimensions for functions f that satisfy Dirichlet boundary conditions at ∂B^2 but whose derivatives do not, such as

$$f(x, y) = (1 - x^2 - y^2) \vee 0$$

for which $-\Delta f \equiv 4$ and hence $(-\Delta + I)^s f = f + 4s$ in the interior. Still, $f \in H_s$ only for $s < 5/2$ because by circular symmetry its Fourier-Bessel coefficients are $a_{mn} = 0$ for $m \neq 0$, while

$$a_{0n} = \frac{2\pi}{\sqrt{\pi} J_1(j_{0,n})} \int_0^1 (r - r^3) J_0(r j_{0,n})\, dr = \frac{8\sqrt{\pi}}{j_{0,n}^3}, \quad \text{so}$$

$$\|f\|_s^2 = 64\pi \sum_{n=1}^{\infty} (1 + \lambda_{0n}^2)^s j_{0,n}^{-6},$$

which is only finite for $s < 5/2$ since $j_{0,n} \asymp n\pi$.

5 Stochastic Expansions

5.1 Brownian Bridge on One Dimensional Interval

As in Sec. 4.2, let

$$\phi_n(x) = \sin(n\pi x/T)\sqrt{2/T}$$

be a CONS in $L_2(\mathcal{T}, dx)$ for $\mathcal{T} = [0, T] \subset \mathbb{R}^1$, and consider the *Brownian Bridge* Gaussian stochastic process X_t with mean and covariance

$$\mathsf{E} X_t \equiv 0 \qquad \mathsf{Cov}(X_s, X_t) = \mathsf{E} X_s\, X_t = (s \wedge t)\big(1 - (s \vee t)/T\big),$$

identical to unit-rate Brownian motion conditioned to reach $X_T = 0$. The Fourier coefficients for a random path $\{X_t\}$ are Gaussian random variables

$$A_n := \int_{\mathcal{T}} Z_t \phi_n(s)\, dt$$

with mean zero and covariance

$$\begin{aligned} \mathsf{E} A_m\, A_n =& \frac{2}{T^2} \int_{\mathcal{T}^2} (s \wedge t)\big(T - (s \vee t)\big) \sin(m\pi s/T)\, \sin(n\pi t/T)\, ds\, dt \\ =& \begin{cases} 0 & m \neq n \\ \frac{T^2}{n^2\pi^2} & m = n \end{cases} \end{aligned}$$

so the $\{A_n\} \overset{\text{ind}}{\sim} \mathsf{No}\big(0, \big(\frac{T}{n\pi}\big)^2\big)$ are independent and the expected squared H_s norm of the sample path $s \rightsquigarrow Z_s$ is

$$\begin{aligned} \mathsf{E}\|Z_\bullet\|_s^2 =& \sum_{n \in \mathbb{N}} \frac{T^2(1 + n^2\pi^2/T^2)^s}{n^2\pi^2} \\ & < \infty \text{ if and only if } s < \tfrac{1}{2}, \end{aligned}$$

with $\mathsf{E}\|Z_\bullet\|_0^2 = T^2/6$. Roughly speaking, Brownian motion paths have just under one-half of a derivative.

Beginning with iid $\zeta_n \overset{\text{iid}}{\sim} \mathsf{No}(0, 1)$ one can *construct* a Brownian Bridge process as the sum

$$Z_t := \sum_{n=1}^{\infty} \frac{T}{n\pi} \zeta_n \phi_n(t)$$

or fractional derivatives of it by

$$Z_t^\sigma := \sum_{n=1}^{\infty} \left[\frac{T}{n\pi}\right]^{1-\sigma} \zeta_n \phi_n(t),$$

which will be centered Gaussian processes with about $\frac{1}{2}-\sigma$ derivatives— smooth paths, for $\sigma \ll 0$ or, for $\sigma > 0$, generalized stochastic processes that only make sense as convolutions $Z[\psi] = \int Z_t \psi(t)\, dt$ for $\psi \in H_s$ with $s > \sigma - \frac{1}{2}$.

5.2 Stochastic Expansions in Two Dimensions

As in Section (4.1) let

$$\phi_{mn}(r,\theta) \equiv \frac{1}{\sqrt{\pi}\, J_{|m|+1}(j_{m,n})} J_m(r j_{m,n})\, e^{im\theta} \tag{9}$$

be a CONS in $L_2(B^2)$, and let $\zeta_{mn} \overset{\text{iid}}{\sim} \mathsf{No}(0,1)$ for $m \in \mathbb{Z}$, $n \in \mathbb{N}$. Fix $\sigma \in \mathbb{R}$ and set

$$Z^\sigma(r,\theta) := \sum [j_{m,n}]^\sigma\, \zeta_{mn}\, \phi_{mn}(r,\theta)$$

for $(r,\theta) \in B^2$. The expected squared H_s norm of Z^σ is

$$\begin{aligned} \mathsf{E}\langle Z^\sigma, Z^\sigma\rangle_s &= \mathsf{E}\langle(-\Delta + I)^s\, Z^\sigma, Z^\sigma\rangle \\ &= \sum (j_{m,n}^2 + 1)^s j_{m,n}^\sigma \end{aligned}$$

Since $j_{m,n} \asymp \pi(n + |m|/2 - 1/4)$ (Abramowitz and Stegun, 1964, §9.5.12), this is finite by the integral test if and only if

$$\infty > \int_1^\infty (r^2+1)^s r^\sigma\, r\, dr,$$

so $Z^\sigma \in H_s$ almost-surely for $\sigma < -2(s+1)$— or $\sigma < -2$ for $s = 0$.

5.3 Reproducing Kernel Hilbert Space (RKHS)

Now let X_s be an arbitrary mean-zero Gaussian process with (positive definite) covariance kernel

$$k(s,t) = \mathsf{E} X_s X_t$$

for $s,t \in \mathcal{T}$, and consider the Hilbert space $\mathcal{H}$ consisting of limits of linear combinations of the $\{X_{s_j}\}$ for $\{s_j\} \subset \mathcal{T}$. For tame enough functions $f, g : \mathcal{T} \to \mathbb{R}$ this space will include elements $X_f \equiv \int_{\mathcal{T}} f(s) Z_s\, ds$ and $X_g \equiv \int_{\mathcal{T}} g(t) Z_t\, dt$, each Gaussian with mean zero and covariance

$$\langle X_f, X_g\rangle \equiv \int_{\mathcal{T}^2} f(s) k(s,t) g(t)\, ds\, dt,$$

the inner-product in $L_2(\mathcal{T})$ of f and Kg for the operator $K:\ L_2(\mathcal{T}) \to L_2(\mathcal{T})$ with kernel k,

$$Kg(s) \equiv \int_{\mathcal{T}} k(s,t)\, g(t)\, dt.$$

If $k \in L_2(\mathcal{T}^2)$ that operator is Hilbert-Schmidt, with a complete orthornormal set $\{\phi_n\}$ of eigenfunctions

$$K\phi_n(s) \equiv \int_{\mathcal{T}} k(s,t)\,\phi_n(t)\,dt = \lambda_n \phi_n(s)$$

whose eigenvalues are square-summable, with

$$\int_{\mathcal{T}^2} k(s,t)^2\,ds\,dt = \sum \lambda_n^2 < \infty.$$

In this case for iid $\{\zeta_n\} \overset{\text{iid}}{\sim} \mathsf{No}(0,1)$ the series

$$Z_s \equiv \sum \lambda_n\,\zeta_n\,\phi_n(s)$$

converges almost-surely in L_2 to a Gaussian process with mean zero and covariance k, the so-called *Karhunen-Loève* expansion. The example of Sec. 5.1 was an example of this. As in that example, often k is the Greens function for a differential operator (here $-\partial^2/\partial x^2$ with Dirichlet boundary conditions), and the eigenfunctions may be found as solutions to differential equations.

References

Abramowitz, M. and Stegun, I. A., eds. (1964), *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables*, *Applied Mathematics Series*, volume 55, Washington, D.C.: National Bureau of Standards, reprinted in paperback by Dover (1974); on-line at `http://www.math.sfu.ca/~cbm/aands/`.

Hida, T. and Hitsuda, M. (1993), *CGaussian Processes*, *Translations of Mathematical Monographs*, volume 120, Providence, RI: American Mathematical Society, translated by the authors from 1976 Japanese edition published by Kinokuniya.

Rasmussen, C. E. and Williams, C. K. I. (2006), *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*, Boston, MA: The MIT Press.

Stein, M. L. (1999), *Interpolation of Spatial Data: Some Theory for Kriging*, New York, NY: Springer-Verlag.

Watson, G. N. (1944), *A treatise on the theory of Bessel functions*, Cambridge, UK: Cambridge University Press, second edition, first edition 1922.