Chapter 11
Spectral Expansions

The mark of a mature, psychologically healthy mind is indeed the ability to live with uncertainty and ambiguity, but only as much as there really is.

JULIAN BAGGINI

This chapter and its sequels consider several *spectral methods* for uncertainty quantification. At their core, these are orthogonal decomposition methods in which a random variable stochastic process (usually the solution of interest) over a probability space $(\Theta, \mathscr{F}, \mu)$ is expanded with respect to an appropriate orthogonal basis of $L^2(\Theta, \mu; \mathbb{R})$. This chapter lays the foundations by considering spectral expansions in general, starting with the *Karhunen–Loève bi-orthogonal decomposition*, and continuing with orthogonal polynomial bases for $L^2(\Theta, \mu; \mathbb{R})$ and the resulting *polynomial chaos decompositions*. Chapters 12 and 13 will then treat two classes of methods for the determination of coefficients in spectral expansions, the intrusive and non-intrusive approaches respectively.

11.1 Karhunen–Loève Expansions

Fix a domain $\mathcal{X} \subseteq \mathbb{R}^d$ (which could be thought of as 'space', 'time' or a general parameter space) and a probability space $(\Theta, \mathscr{F}, \mu)$. The Karhunen–Loève expansion of a square-integrable stochastic process $U \colon \mathcal{X} \times \Theta \to \mathbb{R}$ is a particularly nice spectral decomposition, in that it decomposes U in a *bi-orthogonal* fashion, i.e. in terms of components that are both orthogonal over the spatio-temporal domain $\mathcal{X}$ and the probability space Θ.

To be more precise, consider a stochastic process $U \colon \mathcal{X} \times \Theta \to \mathbb{R}$ such that

- for all $x \in \mathcal{X}$, $U(x) \in L^2(\Theta, \mu; \mathbb{R})$;
- for all $x \in \mathcal{X}$, $\mathbb{E}_\mu[U(x)] = 0$;
- the *covariance function* $C_U(x, y) := \mathbb{E}_\mu[U(x)U(y)]$ is a well-defined continuous function of $x, y \in \mathcal{X}$.

Remark 11.1. (a) The condition that U is a zero-mean process is not a serious restriction; if U is not a zero-mean process, then simply consider $\tilde{U}$ defined by $\tilde{U}(x, \theta) := U(x, \theta) - \mathbb{E}_\mu[U(x)]$.

(b) It is common in practice to see the covariance function interpreted as providing some information on the *correlation length* of the process U. That is, $C_U(x, y)$ depends only upon $\|x - y\|$ and, for some function $g \colon [0, \infty) \to [0, \infty)$, $C_U(x, y) = g(\|x - y\|)$. A typical such g is $g(r) = \exp(-r/r_0)$, and the constant r_0 encodes how similar values of U at nearby points of $\mathcal{X}$ are expected to be; when the correlation length r_0 is small, the field U has dissimilar values near to one another, and so is rough; when r_0 is large, the field U has only similar values near to one another, and so is more smooth.

By abuse of notation, C_U will also denote the *covariance operator* of U, which the linear operator $C_U \colon L^2(\mathcal{X}, \mathrm{d}x; \mathbb{R}) \to L^2(\mathcal{X}, \mathrm{d}x; \mathbb{R})$ defined by

$$(C_U f)(x) := \int_{\mathcal{X}} C_U(x, y) f(y) \, \mathrm{d}y.$$

Now let $\{\psi_n \mid n \in \mathbb{N}\}$ be an orthonormal basis of eigenvectors of $L^2(\mathcal{X}, \mathrm{d}x; \mathbb{R})$ with corresponding eigenvalues $\{\lambda_n \mid n \in \mathbb{N}\}$, i.e.

$$\int_{\mathcal{X}} C_U(x, y) \psi_n(y) \, \mathrm{d}y = \lambda_n \psi_n(x),$$

$$\int_{\mathcal{X}} \psi_m(x) \psi_n(x) \, \mathrm{d}x = \delta_{mn}.$$

Definition 11.2. Let $\mathcal{X}$ be a first-countable topological space. A function $K \colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is called a *Mercer kernel* if

(a) K is continuous;
(b) K is symmetric, i.e. $K(x, x') = K(x', x)$ for all $x, x' \in \mathcal{X}$; and
(c) K is positive semi-definite in the sense that, for all choices of finitely many points $x_1, \ldots, x_n \in \mathcal{X}$, the *Gram matrix*

$$G := \begin{bmatrix} K(x_1, x_1) & \cdots & K(x_1, x_n) \\ \vdots & \ddots & \vdots \\ K(x_n, x_1) & \cdots & K(x_n, x_n) \end{bmatrix}$$

is positive semi-definite, i.e. satisfies $\xi \cdot G\xi \geq 0$ for all $\xi \in \mathbb{R}^n$.

Theorem 11.3 (Mercer). *Let $\mathcal{X}$ be a first-countable topological space equipped with a complete Borel measure μ. Let $K\colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a Mercer kernel. If $x \mapsto K(x,x)$ lies in $L^1(\mathcal{X}, \mu; \mathbb{R})$, then there is an orthonormal basis $\{\psi_n\}_{n\in\mathbb{N}}$ of $L^2(\mathcal{X}, \mu; \mathbb{R})$ consisting of eigenfunctions of the operator*

$$f \mapsto \int_{\mathcal{X}} K(\,\cdot\,, y)f(y)\,\mathrm{d}\mu(y)$$

with non-negative eigenvalues $\{\lambda_n\}_{n\in\mathbb{N}}$. Furthermore, the eigenfunctions corresponding to non-zero eigenvalues are continuous, and

$$K(x,y) = \sum_{n\in\mathbb{N}} \lambda_n \psi_n(x)\psi_n(y),$$

and this series converges absolutely, uniformly over compact subsets of $\mathcal{X}$.

The proof of Mercer's theorem will be omitted, since the main use of the theorem is just to inform various statements about the eigendecomposition of the covariance operator in the Karhunen–Loève theorem. However, it is worth comparing the conditions of Mercer's theorem to those of Sazonov's theorem (Theorem 2.49): together, these two theorems show which integral kernels can be associated with covariance operators of Gaussian measures.

Theorem 11.4 (Karhunen–Loève). *Let $U\colon \mathcal{X} \times \Theta \to \mathbb{R}$ be square-integrable stochastic process, with mean zero and continuous and square-integrable[1] covariance function. Then*

$$U = \sum_{n\in\mathbb{N}} Z_n \psi_n$$

in L^2, where the $\{\psi_n\}_{n\in\mathbb{N}}$ are orthonormal eigenfunctions of the covariance operator C_U, the corresponding eigenvalues $\{\lambda_n\}_{n\in\mathbb{N}}$ are non-negative, the convergence of the series is in $L^2(\Theta, \mu; \mathbb{R})$ and uniform among compact families of $x \in \mathcal{X}$, with

$$Z_n = \int_{\mathcal{X}} U(x)\psi_n(x)\,\mathrm{d}x.$$

Furthermore, the random variables Z_n are centred, uncorrelated, and have variance λ_n:

$$\mathbb{E}_\mu[Z_n] = 0, \text{ and } \mathbb{E}_\mu[Z_m Z_n] = \lambda_n \delta_{mn}.$$

Proof. By Exercise 2.1, and since the covariance function C_U is continuous and square-integrable on $\mathcal{X} \times \mathcal{X}$, it is integrable on the diagonal, and hence is a Mercer kernel. So, by Mercer's theorem, there is an orthonormal basis $\{\psi_n\}_{n\in\mathbb{N}}$ of $L^2(\mathcal{X}, \mathrm{d}x; \mathbb{R})$ consisting of eigenfunctions of the covariance operator with non-negative eigenvalues $\{\lambda_n\}_{n\in\mathbb{N}}$. In this basis, the covariance function has the representation

---

[1] In the case that $\mathcal{X}$ is compact, it is enough to assume that the covariance function is continuous, from which it follows that it is bounded and hence square-integrable on $\mathcal{X} \times \mathcal{X}$.

$$C_U(x,y) = \sum_{n \in \mathbb{N}} \lambda_n \psi_n(x) \psi_n(y).$$

Write the process U in terms of this basis as

$$U(x,\theta) = \sum_{n \in \mathbb{N}} Z_n(\theta) \psi_n(x),$$

where the coefficients $Z_n = Z_n(\theta)$ are given by orthogonal projection:

$$Z_n(\theta) := \int_{\mathcal{X}} U(x,\theta) \psi_n(x) \, dx.$$

(Note that these coefficients Z_n are real-valued random variables.) Then

$$\mathbb{E}_\mu[Z_n] = \mathbb{E}_\mu \left[\int_{\mathcal{X}} U(x) \psi_n(x) \, dx \right] = \int_{\mathcal{X}} \mathbb{E}_\mu[U(x)] \psi_n(x) \, dx = 0.$$

and

$$\begin{aligned}
\mathbb{E}_\mu[Z_m Z_n] &= \mathbb{E}_\mu \left[\int_{\mathcal{X}} U(x) \psi_m(x) \, dx \int_{\mathcal{X}} U(x) \psi_n(x) \, dx \right] \\
&= \mathbb{E}_\mu \left[\int_{\mathcal{X}} \int_{\mathcal{X}} \psi_m(x) U(x) U(y) \psi_n(y) \, dy dx \right] \\
&= \int_{\mathcal{X}} \psi_m(x) \int_{\mathcal{X}} \mathbb{E}_\mu[U(x)U(y)] \psi_n(y) \, dy dx \\
&= \int_{\mathcal{X}} \psi_m(x) \int_{\mathcal{X}} C_U(x,y) \psi_n(y) \, dy dx \\
&= \int_{\mathcal{X}} \psi_m(x) \lambda_n \psi_n(x) \, dx \\
&= \lambda_n \delta_{mn}.
\end{aligned}$$

Let $S_N := \sum_{n=1}^{N} Z_n \psi_n \colon \mathcal{X} \times \Theta \to \mathbb{R}$. Then, for any $x \in \mathcal{X}$,

$$\begin{aligned}
&\mathbb{E}_\mu \left[|U(x) - S_N(x)|^2 \right] \\
&= \mathbb{E}_\mu[U(x)^2] + \mathbb{E}_\mu[S_N(x)^2] - 2\mathbb{E}_\mu[U(x)S_N(x)] \\
&= C_U(x,x) + \mathbb{E}_\mu \left[\sum_{n=1}^{N} \sum_{m=1}^{N} Z_n Z_m \psi_m(x) \psi_n(x) \right] - 2\mathbb{E}_\mu \left[U(x) \sum_{n=1}^{N} Z_n \psi_n(x) \right] \\
&= C_U(x,x) + \sum_{n=1}^{N} \lambda_n \psi_n(x)^2 - 2\mathbb{E}_\mu \left[\sum_{n=1}^{N} \int_{\mathcal{X}} U(x)U(y) \psi_n(y) \psi_n(x) \, dy \right] \\
&= C_U(x,x) + \sum_{n=1}^{N} \lambda_n \psi_n(x)^2 - 2 \sum_{n=1}^{N} \int_{\mathcal{X}} C_U(x,y) \psi_n(y) \psi_n(x) \, dy \\
&= C_U(x,x) - \sum_{n=1}^{N} \lambda_n \psi_n(x)^2 \\
&\to 0 \text{ as } N \to \infty,
\end{aligned}$$

where the convergence with respect of x, uniformly over compact subsets of $\mathcal{X}$, follows from Mercer's theorem. $\qquad\square$

Among many possible decompositions of a random field, the Karhunen–Loève expansion is optimal in the sense that the mean-square error of any truncation of the expansion after finitely many terms is minimal. However, its utility is limited since the covariance function of the solution process is often not known a priori. Nevertheless, the Karhunen–Loève expansion provides an effective means of representing *input* random processes when their covariance structure is known, and provides a simple method for sampling Gaussian measures on Hilbert spaces, which is a necessary step in the implementation of the methods outlined in Chapter 6.

Example 11.5. Suppose that $C\colon \mathcal{H} \to \mathcal{H}$ is a self-adjoint, positive-definite, nuclear operator on a Hilbert space $\mathcal{H}$ and let $m \in \mathcal{H}$. Let $(\lambda_k, \psi_k)_{k \in \mathbb{N}}$ be a sequence of orthonormal eigenpairs for C, ordered by decreasing eigenvalue λ_k. Let $\Xi_1, \Xi_2, \ldots$ be independently distributed according to the standard Gaussian measure $\mathcal{N}(0,1)$ on $\mathbb{R}$. Then, by the Karhunen–Loève theorem,

$$U := m + \sum_{k=1}^{\infty} \lambda_k^{1/2} \Xi_k \psi_k \tag{11.1}$$

is an $\mathcal{H}$-valued random variable with distribution $\mathcal{N}(m, C)$. Therefore, a finite sum of the form $m + \sum_{k=1}^{K} \lambda_k^{1/2} \Xi_k \psi_k$ for large K is a reasonable approximation to a $\mathcal{N}(m, C)$-distributed random variable; this is the procedure used to generate the sample paths in Figure 11.1.

Note that the real-valued random variable $\lambda_k^{1/2} \Xi_k$ has Lebesgue density proportional to $\exp(-|\xi_k|^2 / 2\lambda_k)$. Therefore, although Theorem 2.38 shows that the infinite product of Lebesgue measures on $\mathrm{span}\{\psi_k \mid k \in \mathbb{N}\}$ cannot define an infinite-dimensional Lebesgue measure on $\mathcal{H}$, $U - m$ defined by (11.1) may be said to have a 'formal Lebesgue density' proportional to

$$\prod_{k \in \mathbb{N}} \exp\left(-\frac{|\xi_k|^2}{2\lambda_k}\right) = \exp\left(-\frac{1}{2} \sum_{k \in \mathbb{N}} \frac{|\xi_k|^2}{\lambda_k}\right)$$

$$= \exp\left(-\frac{1}{2} \sum_{k \in \mathbb{N}} \frac{|\langle u - m, \psi_k \rangle_{\mathcal{H}}|^2}{\lambda_k}\right)$$

$$= \exp\left(-\frac{1}{2} \left\| C^{-1/2}(u - m) \right\|_{\mathcal{H}}^2\right)$$

by Parseval's theorem and the eigenbasis representation of C. This formal derivation should make it intuitively reasonable that U is a Gaussian random variable on $\mathcal{H}$ with mean m and covariance operator C. For more general sampling schemes of this type, see the later remarks on the sampling of Besov measures.

a

$$K = 10$$

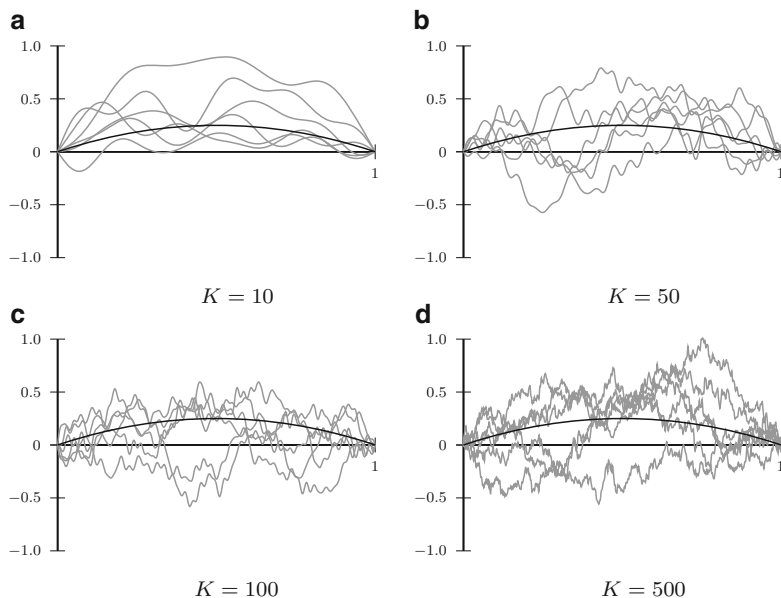**b**

$$K = 50$$

c

$$K = 100$$

d

$$K = 500$$

Fig. 11.1: Approximate sample paths of the Gaussian distribution on $H_0^1([0,1])$ that has mean path $m(x) = x(1-x)$ and covariance operator $\left(-\frac{\mathrm{d}^2}{\mathrm{d}x^2}\right)^{-1}$. Along with the mean path (black), six sample paths (grey) are shown for truncated Karhunen–Loève expansions using $K \in \mathbb{N}$ terms. Except for the non-trivial mean, these are approximate draws from the unit Brownian bridge on $[0,1]$.

Principal Component Analysis. As well as being useful for the analysis of random paths, surfaces, and so on, Karhunen–Loève expansions are also useful in the analysis of finite-dimensional random vectors and sample data:

Definition 11.6. A *principal component analysis* of an $\mathbb{R}^N$-valued random vector U is the Karhunen–Loève expansion of U seen as a stochastic process $U \colon \{1, \dots, N\} \times \mathcal{X} \to \mathbb{R}$. It is also known as the *discrete Karhunen–Loève transform*, the *Hotelling transform* and the *proper orthogonal decomposition*.

Principal component analysis is often applied to sample data, and is intimately related to the singular value decomposition:

Example 11.7. Let $X \in \mathbb{R}^{N \times M}$ be a matrix whose columns are M independent and identically distributed samples from some probability measure on $\mathbb{R}^N$, and assume without loss of generality that the samples have empirical mean zero. The empirical covariance matrix of the samples is

$$\widehat{C} := \tfrac{1}{M} X X^{\mathsf{T}}.$$

(If the samples do not have empirical mean zero, then the empirical mean should be subtracted first, and then $\frac{1}{M}$ in the definition of $\widehat{C}$ should be replaced by $\frac{1}{M-1}$ so that $\widehat{C}$ will be an unbiased estimator of the true covariance matrix C.) The eigenvalues λ_n and eigenfunctions ψ_n of the Karhunen–Loève expansion are just the eigenvalues and eigenvectors of this matrix $\widehat{C}$. Let $\Lambda \in \mathbb{R}^{N \times N}$ be the diagonal matrix of the eigenvalues λ_n (which are non-negative, and are assumed to be in decreasing order) and $\Psi \in \mathbb{R}^{N \times N}$ the matrix of corresponding orthonormal eigenvectors, so that $\widehat{C}$ diagonalizes as

$$\widehat{C} = \Psi \Lambda \Psi^{\mathsf{T}}.$$

The principal component transform of the data X is $W := \Psi^{\mathsf{T}} X$; this is an orthogonal transformation of $\mathbb{R}^N$ that transforms X to a new coordinate system in which the greatest component-wise variance comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on.

On the other hand, taking the singular value decomposition of the data (normalized by the number of samples) yields

$$\frac{1}{\sqrt{M}} X = U \Sigma V^{\mathsf{T}},$$

where $U \in \mathbb{R}^{N \times N}$ and $V \in \mathbb{R}^{M \times M}$ are orthogonal and $\Sigma \in \mathbb{R}^{N \times M}$ is diagonal with decreasing non-negative diagonal entries (the singular values of $\frac{1}{\sqrt{M}} X$). Then

$$\widehat{C} = U \Sigma V^{\mathsf{T}} (U \Sigma V^{\mathsf{T}})^{\mathsf{T}} = U \Sigma V^{\mathsf{T}} V \Sigma^{\mathsf{T}} U^{\mathsf{T}} = U \Sigma^2 U^{\mathsf{T}}.$$

from which we see that $U = \Psi$ and $\Sigma^2 = \Lambda$. This is just another instance of the well-known relation that, for any matrix A, the eigenvalues of AA^* are the singular values of A and the right eigenvectors of AA^* are the left singular vectors of A; however, in this context, it also provides an alternative way to compute the principal component transform.

In fact, performing principal component analysis via the singular value decomposition is numerically preferable to forming and then diagonalizing the covariance matrix, since the formation of XX^{T} can cause a disastrous loss of precision; the classic example of this phenomenon is the Läuchli matrix

$$\begin{bmatrix} 1 & \varepsilon & 0 & 0 \\ 1 & 0 & \varepsilon & 0 \\ 1 & 0 & 0 & \varepsilon \end{bmatrix} \quad (0 < \varepsilon \ll 1),$$

for which taking the singular value decomposition (e.g. by bidiagonalization followed by QR iteration) is stable, but forming and diagonalizing XX^{T} is unstable.

Karhunen–Loève Sampling of Non-Gaussian Besov Measures. The Karhunen–Loève approach to generating samples from Gaussian measures of known covariance operator, as in Example 11.5, can be extended to more general settings, in which a basis is prescribed a priori and (not necessarily Gaussian) random coefficients with a suitable decay rate are used. The choice of basis elements and the rate of decay of the coefficients together control the smoothness of the sample realizations; the mathematical hard work lies in showing that such random series do indeed converge to a well-defined limit, and thereby define a probability measure on the desired function space.

One method for the construction of function spaces — and hence random functions — of desired smoothness is to use wavelets. Wavelet bases are particularly attractive because they allow for the representation of sharply localized features — e.g. the interface between two media with different material properties — in a way that globally smooth basis functions such as polynomials and the Fourier basis do not. Omitting several technicalities, a wavelet basis of $L^2(\mathbb{R}^d)$ or $L^2(\mathbb{T}^d)$ can be thought of as an orthonormal basis consisting of appropriately scaled and shifted copies of a single basic element that has some self-similarity. By controlling the rate of decay of the coefficients in a wavelet expansion, we obtain a family of function spaces — the *Besov spaces* — with three scales of smoothness, here denoted p, q and s. In what follows, for any function f on $\mathbb{R}^d$ or $\mathbb{T}^d$, define the scaled and shifted version $f_{j,k}$ of f for $j, k \in \mathbb{Z}$ by

$$f_{j,k}(x) := f(2^j x - k). \tag{11.2}$$

The starting point of a wavelet construction is a *scaling function* (also known as the *averaging function* or *father wavelet*) $\widetilde{\phi}\colon \mathbb{R} \to \mathbb{R}$ and a family of closed subspaces $\mathcal{V}_j \subseteq L^2(\mathbb{R})$, $j \in \mathbb{Z}$, called a *multiresolution analysis* of $L^2(\mathbb{R})$, satisfying

(a) (nesting) for all $j \in \mathbb{Z}$, $\mathcal{V}_j \subseteq \mathcal{V}_{j+1}$;
(b) (density and zero intersection) $\overline{\bigcup_{j\in\mathbb{Z}} \mathcal{V}_j} = L^2(\mathbb{R})$ and $\bigcap_{j\in\mathbb{Z}} \mathcal{V}_j = \{0\}$;
(c) (scaling) for all $j, k \in \mathbb{Z}$, $f \in \mathcal{V}_0 \iff f_{j,k} \in \mathcal{V}_j$;
(d) (translates of $\widetilde{\phi}$ generate $\mathcal{V}_0$) $\mathcal{V}_0 = \mathrm{span}\{\widetilde{\phi}_{0,k} \mid k \in \mathbb{Z}\}$;
(e) (Riesz basis) there are finite positive constants A and B such that, for all sequences $(c_k)_{k\in\mathbb{Z}} \in \ell^2(\mathbb{Z})$,

$$A\|(c_k)\|_{\ell^2(\mathbb{Z})} \leq \left\|\sum_{k\in\mathbb{Z}} c_k \widetilde{\phi}_{0,k}\right\|_{L^2(\mathbb{R})} \leq B\|(c_k)\|_{\ell^2(\mathbb{Z})}.$$

Given such a scaling function $\widetilde{\phi}\colon \mathbb{R} \to \mathbb{R}$, the associated *mother wavelet* $\widetilde{\psi}\colon \mathbb{R} \to \mathbb{R}$ is defined as follows:

$$\text{if } \widetilde{\phi}(x) = \sum_{k \in \mathbb{Z}} c_k \widetilde{\phi}(2x - k),$$

$$\text{then } \widetilde{\psi}(x) = \sum_{k \in \mathbb{Z}} (-1)^k c_{k+1} \widetilde{\phi}(2x + k).$$

It is the scaled and shifted copies of the mother wavelet $\widetilde{\psi}$ that will form the desired orthonormal basis of L^2.

Example 11.8. (a) The indicator function $\widetilde{\phi} = \mathbb{I}_{[0,1)}$ satisfies the self-similarity relation $\widetilde{\phi}(x) = \widetilde{\phi}(2x) + \widetilde{\phi}(2x - 1)$; the associated $\widetilde{\psi}$ given by

$$\widetilde{\psi}(x) = \widetilde{\phi}(2x) - \widetilde{\phi}(2x - 1) = \begin{cases} 1, & \text{if } 0 \le x < \frac{1}{2}, \\ -1, & \text{if } \frac{1}{2} \le x < 1, \\ 0, & \text{otherwise.} \end{cases}$$

is called the *Haar wavelet*.

(b) The B-spline scaling functions σ_r, $r \in \mathbb{N}_0$, are piecewise polynomial of degree r and globally $\mathcal{C}^{r-1}$, and are defined recursively by convolution:

$$\sigma_r := \begin{cases} \mathbb{I}_{[0,1)}, & \text{for } r = 0, \\ \sigma_{r-1} \star \sigma_0, & \text{for } r \in \mathbb{N}, \end{cases} \tag{11.3}$$

where

$$(f \star g)(x) := \int_{\mathbb{R}} f(y) g(x - y) \, \mathrm{d}y.$$

Here, the presentation focusses on Besov spaces of 1-periodic functions, i.e. functions on the unit circle $\mathbb{T} := \mathbb{R}/\mathbb{Z}$, and on the d-dimensional unit torus $\mathbb{T}^d := \mathbb{R}^d/\mathbb{Z}^d$. To this end, set

$$\phi(x) := \sum_{s \in \mathbb{Z}} \widetilde{\phi}(x + s) \quad \text{and} \quad \psi(x) := \sum_{s \in \mathbb{Z}} \widetilde{\psi}(x + s).$$

Scaled and translated versions of these functions are defined as usual by (11.2). Note that in the toroidal case the spaces $\mathcal{V}_j$ for $j < 0$ consist of constant functions, and that, for each scale $j \in \mathbb{N}_0$, $\phi \in \mathcal{V}_0$ has only 2^j distinct scaled translates $\phi_{j,k} \in \mathcal{V}_j$, i.e. those with $k = 0, \ldots, 2^{j-1}$. Let

$$\mathcal{V}_j := \mathrm{span}\{\phi_{j,k} \mid k = 0, \ldots, 2^j - 1\},$$
$$\mathcal{W}_j := \mathrm{span}\{\psi_{j,k} \mid k = 0, \ldots, 2^j - 1\},$$

so that $\mathcal{W}_j$ is the orthogonal complement of $\mathcal{V}_j$ in $\mathcal{V}_{j+1}$ and

$$L^2(\mathbb{T}) = \overline{\bigcup_{j \in \mathbb{N}_0} \mathcal{V}_j} = \bigoplus_{j \in \mathbb{N}_0} \mathcal{W}_j$$

Indeed, if ψ has unit norm, then $2^{j/2}\psi_{j,k}$ also has unit norm, and

$$\{2^{j/2}\psi_{j,k} \mid k = 0, \dots, 2^j - 1\}\text{is an orthonormal basis of } \mathcal{W}_j, \text{ and}$$

$$\{2^{j/2}\psi_{j,k} \mid j \in \mathbb{N}_0, k = 0, \dots, 2^j - 1\} \text{ is an orthonormal basis of } L^2(\mathbb{T}),$$

a so-called *wavelet basis*.

To construct an analogous wavelet basis of $L^2(\mathbb{T}^d)$ for $d \geq 1$, proceed as follows: for $\nu \in \{0,1\}^d \setminus \{(0,\dots,0)\}$, $j \in \mathbb{N}_0$, and $k \in \{0,\dots,2^j-1\}^d$, define the scaled and translated wavelet $\psi_{j,k}^\nu \colon \mathbb{T}^d \to \mathbb{R}$ by

$$\psi_{j,k}^\nu(x) := 2^{dj/2}\psi^{\nu_1}(2^j x_1 - k_1)\cdots\psi^{\nu_d}(2^j x_d - k_d)$$

where $\psi^0 = \phi$ and $\psi^1 = \psi$. The system

$$\left\{\psi_{j,k}^\nu \,\middle|\, j \in \mathbb{N}_0, k \in \{0,\dots,2^j-1\}^d, \nu \in \{0,1\}^d \setminus \{(0,\dots,0)\}\right\}$$

is an orthonormal wavelet basis of $L^2(\mathbb{T}^d)$.

The Besov space $B_{pq}^s(\mathbb{T}^d)$ can be characterized in terms of the summability of wavelet coefficients at the various scales:

Definition 11.9. Let $1 \leq p, q < \infty$ and let $s > 0$. The *Besov* (p,q,s) *norm* of a function $u = \sum_{j,k,\nu} u_{j,k}^\nu \psi_{j,k}^\nu \colon \mathbb{T}^d \to \mathbb{R}$ is defined by

$$\left\|\sum_{j\in\mathbb{N}_0}\sum_{\nu,k} u_{j,k}^\nu\psi_{j,k}^\nu\right\|_{B_{pq}^s(\mathbb{T}^d)} := \left\|j \mapsto 2^{js}2^{jd(\frac{1}{2}-\frac{1}{p})}\left\|(k,\nu)\mapsto u_{j,k}^\nu\right\|_{\ell^p}\right\|_{\ell^q(\mathbb{N}_0)}$$

$$:= \left(\sum_{j\in\mathbb{N}_0} 2^{qjs}2^{qjd(\frac{1}{2}-\frac{1}{p})}\left(\sum_{\nu,k}|u_{j,k}^\nu|^p\right)^{q/p}\right)^{1/q},$$

and the *Besov space* $B_{pq}^s(\mathbb{T}^d)$ is the completion of the space of functions for which this norm is finite.

Note that at each scale j, there are $(2^d - 1)2^{jd} = 2^{(j+1)d} - 2^{jd}$ wavelet coefficients. The indices j, k and ν can be combined into a single index $\ell \in \mathbb{N}$. First, $\ell = 1$ corresponds to the scaling function $\phi(x_1)\cdots\phi(x_d)$. The remaining numbering is done scale by scale; that is, we first number wavelets with $j = 0$, then wavelets with $j = 1$, and so on. Within each scale $j \in \mathbb{N}_0$, the $2^d - 1$ indices ν are ordered by thinking them as binary representation of integers, and an ordering of the 2^{jd} translations k can be chosen arbitrarily. With this renumbering,

$$\sum_{\ell=1}^\infty c_\ell\psi_\ell \in B_{pq}^s(\mathbb{T}^d) \iff 2^{js}2^{jd(\frac{1}{2}-\frac{1}{p})}\left(\sum_{\ell=2^{jd}}^{2^{(j+1)d}-1}|c_\ell|^p\right)^{1/p} \in \ell^q(\mathbb{N}_0)$$

For $p = q$, since at scale j it holds that $2^{jd} \leq \ell < 2^{(j+1)d}$, an equivalent norm for $B_{pp}^s(\mathbb{T}^d)$ is

$$\left\|\sum_{\ell \in \mathbb{N}} u_\ell \psi_\ell\right\|_{B_{pp}^s(\mathbb{T}^d)} \simeq \left\|\sum_{\ell \in \mathbb{N}} u_\ell \psi_\ell\right\|_{X^{s,p}} := \left(\sum_{\ell=1}^\infty \ell^{(ps/d+p/2-1)}|u_\ell|^p\right)^{1/p};$$

in particular if the original scaling function and mother wavelet are r times differentiable with $r > s$, then B_{22}^s coincides with the Sobolev space H^s. This leads to a Karhunen–Loève-type sampling procedure for $B_{pp}^s(\mathbb{T}^d)$, as in Example 11.5: U defined by

$$U := \sum_{\ell \in \mathbb{N}} \ell^{-(\frac{s}{d}+\frac{1}{2}-\frac{1}{p})}\kappa^{-\frac{1}{p}}\Xi_\ell \psi_\ell, \tag{11.4}$$

where Ξ_ℓ are sampled independently and identically from the generalized Gaussian measure on $\mathbb{R}$ with Lebesgue density proportional to $\exp(-\frac{1}{2}|\xi_\ell|^p)$, can be said to have 'formal Lebesgue density' proportional to $\exp(-\frac{\kappa}{2}\|u\|_{B_{pp}^s}^p)$, and is therefore a natural candidate for a 'typical' element of the Besov space $B_{pp}^s(\mathbb{T}^d)$. More generally, given *any* orthonormal basis $\{\psi_k \mid k \in \mathbb{N}\}$ of some Hilbert space, one can define a Banach subspace $X^{s,p}$ with norm

$$\left\|\sum_{\ell \in \mathbb{N}} u_\ell \psi_\ell\right\|_{X^{s,p}} := \left(\sum_{\ell=1}^\infty \ell^{(ps/d+p/2-1)}|u_\ell|^p\right)^{1/p}$$

and define a *Besov-distributed random variable* U by (11.4).

It remains, however, to check that (11.4) not only defines a measure, but that it assigns unit probability mass to the Besov space from which it is desired to draw samples. It turns out that the question of whether or not $U \in X^{s,p}$ with probability one is closely related to having a Fernique theorem (q.v. Theorem 2.47) for Besov measures:

Theorem 11.10. *Let U be defined as in (11.4), with $1 \leq p < \infty$ and $s > 0$. Then*

$$\|U\|_{X^{t,p}} < \infty \text{ almost surely} \iff \mathbb{E}[\exp(\alpha\|U\|_{X^{t,p}}^p)] < \infty \text{ for all } \alpha \in (0, \tfrac{\kappa}{2})$$
$$\iff t < s - \tfrac{d}{p}$$

Furthermore, for $p \geq 1$, $s > \frac{d}{p}$, and $t < s - \frac{d}{p}$, there is a constant r^* depending only on p, d, s and t such that, for all $\alpha \in (0, \frac{\kappa}{2r^*})$,

$$\mathbb{E}[\exp(\alpha\|U\|_{\mathcal{C}^t})] < \infty.$$

11.2 Wiener–Hermite Polynomial Chaos

The next section will cover polynomial chaos (PC) expansions in greater generality, and this section serves as an introductory prelude. In this, the classical and notationally simplest setting, we consider expansions of a real-valued random variable U with respect to a single standard Gaussian random variable Ξ, using appropriate orthogonal polynomials of Ξ, i.e. the Hermite polynomials. This setting was pioneered by Norbert Wiener, and so it is known as the Wiener–Hermite polynomial chaos. The term 'chaos' is perhaps a bit confusing, and is not related to the use of the term in the study of dynamical systems; its original meaning, as used by Wiener (1938), was something closer to what would nowadays be called a stochastic process:

> "Of all the forms of chaos occurring in physics, there is only one class which has been studied with anything approaching completeness. This is the class of types of chaos connected with the theory of Brownian motion."

Let $\Xi \sim \gamma = \mathcal{N}(0, 1)$ be a standard Gaussian random variable, and let $\mathrm{He}_n \in \mathfrak{P}$, for $n \in \mathbb{N}_0$, be the *Hermite polynomials*, the orthogonal polynomials for the standard Gaussian measure γ with the normalization

$$\int_{\mathbb{R}} \mathrm{He}_m(\xi) \mathrm{He}_n(\xi) \, \mathrm{d}\gamma(\xi) = n! \delta_{mn}.$$

By the Weierstrass approximation theorem (Theorem 8.20) and the approximability of L^2 functions by continuous ones, the Hermite polynomials form a complete orthogonal basis of the Hilbert space $L^2(\mathbb{R}, \gamma; \mathbb{R})$ with the inner product

$$\langle U, V \rangle_{L^2(\gamma)} := \mathbb{E}[U(\Xi)V(\Xi)] \equiv \int_{\mathbb{R}} U(\xi)V(\xi) \, \mathrm{d}\gamma(\xi).$$

Definition 11.11. Let $U \in L^2(\mathbb{R}, \gamma; \mathbb{R})$ be a square-integrable real-valued random variable. The *Wiener–Hermite polynomial chaos expansion* of U with respect to the standard Gaussian Ξ is the expansion of U in the orthogonal basis $\{\mathrm{He}_n\}_{n \in \mathbb{N}_0}$, i.e.

$$U = \sum_{n \in \mathbb{N}_0} u_n \mathrm{He}_n(\Xi)$$

with scalar *Wiener–Hermite polynomial chaos coefficients* $\{u_n\}_{n \in \mathbb{N}_0} \subseteq \mathbb{R}$ given by

$$u_n = \frac{\langle U, \mathrm{He}_n \rangle_{L^2(\gamma)}}{\|\mathrm{He}_n\|_{L^2(\gamma)}^2} = \frac{1}{n! \sqrt{2\pi}} \int_{-\infty}^{\infty} U(\xi) \mathrm{He}_n(\xi) e^{-\xi^2/2} \, \mathrm{d}\xi.$$

Note that, in particular, since $\mathrm{He}_0 \equiv 1$,

$$\mathbb{E}[U] = \langle \mathrm{He}_0, U \rangle_{L^2(\gamma)} = \sum_{n \in \mathbb{N}_0} u_n \langle \mathrm{He}_0, \mathrm{He}_n \rangle_{L^2(\gamma)} = u_0,$$

so the expected value of U is simply its 0th PC coefficient. Similarly, its variance is a weighted sum of the squares of its PC coefficients:

$$
\begin{aligned}
\mathbb{V}[U] &= \mathbb{E}\left[|U - \mathbb{E}[U]|^2\right] \\
&= \mathbb{E}\left[\left|\sum_{n \in \mathbb{N}} u_n \mathrm{He}_n\right|^2\right] && \text{since } \mathbb{E}[U] = u_0 \\
&= \sum_{m,n \in \mathbb{N}} u_m u_n \langle \mathrm{He}_m, \mathrm{He}_n \rangle_{L^2(\gamma)} \\
&= \sum_{n \in \mathbb{N}} u_n^2 \|\mathrm{He}_n\|^2_{L^2(\gamma)} && \text{by Hermitian orthogonality} \\
&= \sum_{n \in \mathbb{N}} u_n^2 n!.
\end{aligned}
$$

Example 11.12. Let $X \sim \mathcal{N}(m, \sigma^2)$ be a real-valued Gaussian random variable with mean $m \in \mathbb{R}$ and variance $\sigma^2 \geq 0$. Let $Y := e^X$; since $\log Y$ is normally distributed, the non-negative-valued random variable Y is said to be a *log-normal random variable*. As usual, let $\Xi \sim \mathcal{N}(0, 1)$ be the standard Gaussian random variable; clearly X has the same distribution as $m + \sigma\Xi$, and Y has the same distribution as $e^m e^{\sigma\Xi}$. The Wiener–Hermite expansion of Y as $\sum_{k \in \mathbb{N}_0} y_k \mathrm{He}_k(\Xi)$ has coefficients

$$
\begin{aligned}
y_k &= \frac{\langle e^{m + \sigma\Xi}, \mathrm{He}_k(\Xi) \rangle}{\|\mathrm{He}_k(\Xi)\|^2} \\
&= \frac{e^m}{k!} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{\sigma\xi} \mathrm{He}_k(\xi) e^{-\xi^2/2} \, \mathrm{d}\xi \\
&= \frac{e^{m + \sigma^2/2}}{k!} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{He}_k(\xi) e^{-(\xi - \sigma)^2/2} \, \mathrm{d}\xi \\
&= \frac{e^{m + \sigma^2/2}}{k!} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{He}_k(w + \sigma) e^{-w^2/2} \, \mathrm{d}w.
\end{aligned}
$$

This Gaussian integral can be evaluated directly using the Cameron–Martin formula (Lemma 2.40), or else using the formula

$$
\mathrm{He}_n(x + y) = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} \mathrm{He}_k(y),
$$

which follows from the derivative property $\mathrm{He}_n' = n\mathrm{He}_{n-1}$, with $x = \sigma$ and $y = w$: this formula yields that

$$
y_k = \frac{e^{m + \sigma^2/2}}{k!} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \sum_{j=0}^{k} \binom{k}{j} \sigma^{k-j} \mathrm{He}_j(w) e^{-w^2/2} \, \mathrm{d}w = \frac{e^{m + \sigma^2/2} \sigma^k}{k!}
$$

since the orthogonality relation $\langle \mathrm{He}_m, \mathrm{He}_n \rangle_{L^2(\gamma)} = n! \delta_{mn}$ with $n = 0$ implies that every Hermite polynomial other than He_0 has mean 0 under standard Gaussian measure. That is,

$$Y = e^{m+\sigma^2/2} \sum_{k \in \mathbb{N}_0} \frac{\sigma^k}{k!} \mathrm{He}_k(\Xi). \tag{11.5}$$

The Wiener–Hermite expansion (11.5) reveals that $\mathbb{E}[Y] = e^{m+\sigma^2/2}$ and

$$\mathbb{V}[Y] = e^{2m+\sigma^2} \sum_{k \in \mathbb{N}} \left(\frac{\sigma^k}{k!} \right)^2 \|\mathrm{He}_k\|_{L^2(\gamma)}^2 = e^{2m+\sigma^2} \left(e^{\sigma^2} - 1 \right).$$

Truncation of Wiener–Hermite Expansions. Of course, in practice, the series expansion $U = \sum_{k \in \mathbb{N}_0} u_k \mathrm{He}_k(\Xi)$ must be truncated after finitely many terms, and so it is natural to ask about the quality of the approximation

$$U \approx U^K := \sum_{k=0}^{K} u_k \mathrm{He}_k(\Xi).$$

Since the Hermite polynomials $\{\mathrm{He}_k\}_{k \in \mathbb{N}_0}$ form a complete orthogonal basis for $L^2(\mathbb{R}, \gamma; \mathbb{R})$, the standard results about orthogonal approximations in Hilbert spaces apply. In particular, by Corollary 3.26, the truncation error $U - U^K$ is orthogonal to the space from which U^K was chosen, i.e.

$$\mathrm{span}\{\mathrm{He}_0, \mathrm{He}_1, \ldots, \mathrm{He}_K\},$$

and tends to zero in mean square; in the stochastic context, this observation was first made by Cameron and Martin (1947, Section 2).

Lemma 11.13. *The truncation error $U - U^K$ is orthogonal to the subspace*

$$\mathrm{span}\{\mathrm{He}_0, \mathrm{He}_1, \ldots, \mathrm{He}_K\}$$

of $L^2(\mathbb{R}, \mathrm{d}\gamma; \mathbb{R})$. Furthermore, $\lim_{K \to \infty} U^K = U$ in $L^2(\mathbb{R}, \gamma; \mathbb{R})$.

Proof. Let $V := \sum_{m=0}^{K} v_m \mathrm{He}_m$ be any element of the subspace of $L^2(\mathbb{R}, \gamma; \mathbb{R})$ spanned by the Hermite polynomials of degree at most K. Then

$$
\begin{aligned}
\langle U - U^K, V \rangle_{L^2(\gamma)} &= \left\langle \left(\sum_{n > K} u_n \mathrm{He}_n \right), \left(\sum_{m=0}^{K} v_m \mathrm{He}_m \right) \right\rangle \\
&= \sum_{\substack{n > K \\ m \in \{0, \ldots, K\}}} u_n v_m \langle \mathrm{He}_n, \mathrm{He}_m \rangle \\
&= 0.
\end{aligned}
$$

Hence, by Pythagoras' theorem,

$$\|U\|^2_{L^2(\gamma)} = \|U^K\|^2_{L^2(\gamma)} + \|U - U^K\|^2_{L^2(\gamma)},$$

and hence $\|U - U^K\|_{L^2(\gamma)} \to 0$ as $K \to \infty$. $\qquad\qquad\square$

11.3 Generalized Polynomial Chaos Expansions

The ideas of polynomial chaos can be generalized well beyond the setting in which the elementary random variable Ξ used to generate the orthogonal decomposition is a standard Gaussian random variable, or even a vector $\Xi = (\Xi_1, \dots, \Xi_d)$ of mutually orthogonal Gaussian random variables. Such expansions are referred to as *generalized polynomial chaos* (gPC) expansions.

Let $\Xi = (\Xi_1, \dots, \Xi_d)$ be an $\mathbb{R}^d$-valued random variable with independent (and hence L^2-orthogonal) components, called the *stochastic germ*. Let the measurable rectangle $\Theta = \Theta_1 \times \cdots \times \Theta_d \subseteq \mathbb{R}^d$ be the support (i.e. range) of Ξ. Denote by $\mu = \mu_1 \otimes \cdots \otimes \mu_d$ the distribution of Ξ on Θ. The objective is to express any function (random variable, random vector, or even random field) $U \in L^2(\Theta, \mu)$ in terms of elementary μ-orthogonal functions of the stochastic germ Ξ.

As usual, let $\mathfrak{P}^d$ denote the ring of all d-variate polynomials with real coefficients, and let $\mathfrak{P}^d_{\leq p}$ denote those polynomials of total degree at most $p \in \mathbb{N}_0$. Let $\Gamma_p \subseteq \mathfrak{P}^d_{\leq p}$ be a collection of polynomials that are mutually orthogonal, orthogonal to $\mathfrak{P}^d_{\leq p-1}$, and span $\mathfrak{P}^d_{=p}$. Assuming for convenience, as usual, the completeness of the resulting system of orthogonal polynomials, this yields the orthogonal decomposition

$$L^2(\Theta, \mu; \mathbb{R}) = \bigoplus_{p \in \mathbb{N}_0} \operatorname{span} \Gamma_p.$$

It is important to note that there is a lack of uniqueness in these basis polynomials whenever $d \geq 2$: each choice of ordering of multi-indices $\alpha \in \mathbb{N}_0^d$ can yield a different orthogonal basis of $L^2(\Theta, \mu)$ when the Gram–Schmidt procedure is applied to the monomials ξ^α.

Note that (as usual, assuming separability) the L^2 space over the product probability space $(\Theta, \mathscr{F}, \mu)$ is isomorphic to the Hilbert space tensor product of the L^2 spaces over the marginal probability spaces:

$$L^2(\Theta_1 \times \cdots \times \Theta_d, \mu_1 \otimes \cdots \otimes \mu_d; \mathbb{R}) = \bigotimes_{i=1}^d L^2(\Theta_i, \mu_i; \mathbb{R});$$

hence, as in Theorem 8.25, an orthogonal system of multivariate polynomials for $L^2(\Theta, \mu; \mathbb{R})$ can be found by taking products of univariate orthogonal polynomials for the marginal spaces $L^2(\Theta_i, \mu_i; \mathbb{R})$. A *generalized polynomial chaos* (gPC) expansion of a random variable or stochastic process U is simply the expansion of U with respect to such a complete orthogonal polynomial basis of $L^2(\Theta, \mu)$.

Example 11.14. Let $\Xi = (\Xi_1, \Xi_2)$ be such that Ξ_1 and Ξ_2 are independent (and hence orthogonal) and such that Ξ_1 is a standard Gaussian random variable and Ξ_2 is uniformly distributed on $[-1, 1]$. Hence, the univariate orthogonal polynomials for Ξ_1 are the Hermite polynomials He_n and the univariate orthogonal polynomials for Ξ_2 are the Legendre polynomials Le_n. Thus, by Theorem 8.25, a system of orthogonal polynomials for Ξ up to total degree 3 is

$$
\begin{aligned}
\Gamma_0 &= \{1\}, \\
\Gamma_1 &= \{\mathrm{He}_1(\xi_1), \mathrm{Le}_1(\xi_2)\} \\
&= \{\xi_1, \xi_2\}, \\
\Gamma_2 &= \{\mathrm{He}_2(\xi_1), \mathrm{He}_1(\xi_1)\mathrm{Le}_1(\xi_2), \mathrm{Le}_2(\xi_2)\} \\
&= \{\xi_1^2 - 1, \xi_1\xi_2, \tfrac{1}{2}(3\xi_2^2 - 1)\}, \\
\Gamma_3 &= \{\mathrm{He}_3(\xi_1), \mathrm{He}_2(\xi_1)\mathrm{Le}_1(\xi_2), \mathrm{He}_1(\xi_1)\mathrm{Le}_2(\xi_2), \mathrm{Le}_3(\xi_2)\} \\
&= \{\xi_1^3 - 3\xi_1, \xi_1^2\xi_2 - \xi_2, \tfrac{1}{2}(3\xi_1\xi_2^2 - \xi_1), \tfrac{1}{2}(5\xi_2^3 - 3\xi_2)\}.
\end{aligned}
$$

Remark 11.15. To simplify the notation in what follows, the following conventions will be observed:

(a) To simplify expectations, inner products and norms, $\langle \cdot \rangle_\mu$ or simply $\langle \cdot \rangle$ will denote integration (i.e. expectation) with respect to the probability measure μ, so that the $L^2(\mu)$ inner product is simply $\langle X, Y \rangle_{L^2(\mu)} = \langle XY \rangle_\mu$.

(b) Rather than have the orthogonal basis polynomials be indexed by multi-indices $\alpha \in \mathbb{N}_0^d$, or have two scalar indices, one for the degree p and one within each set Γ_p, it is convenient to order the basis polynomials using a single scalar index $k \in \mathbb{N}_0$. It is common in practice to take $\Psi_0 = 1$ and to have the polynomial degree be (weakly) increasing with respect to the new index k. So, to continue Example 11.14, one could use the graded lexicographic ordering on $\alpha \in \mathbb{N}_0^2$ so that $\Psi_0(\xi) = 1$ and

$$
\begin{aligned}
\Psi_1(\xi) &= \xi_1, & \Psi_2(\xi) &= \xi_2, & \Psi_3(\xi) &= \xi_1^2 - 1, \\
\Psi_4(\xi) &= \xi_1\xi_2, & \Psi_5(\xi) &= \tfrac{1}{2}(3\xi_2^2 - 1), & \Psi_6(\xi) &= \xi_1^3 - 3\xi_1, \\
\Psi_7(\xi) &= \xi_1^2\xi_2 - \xi_2, & \Psi_8(\xi) &= \tfrac{1}{2}(3\xi_1\xi_2^2 - \xi_1), & \Psi_9(\xi) &= \tfrac{1}{2}(5\xi_2^3 - 3\xi_2).
\end{aligned}
$$

(c) By abuse of notation, Ψ_k will stand for both a polynomial function (which is a deterministic function from $\mathbb{R}^d$ to $\mathbb{R}$) and for the real-valued random variable that is the composition of that polynomial with the stochastic germ Ξ (which is a function from an abstract probability space to $\mathbb{R}$).

Truncation of gPC Expansions. Suppose that a gPC expansion of the form $U = \sum_{k \in \mathbb{N}_0} u_k \Psi_k$ is truncated, i.e. we consider

$$U^K = \sum_{k=0}^{K} u_k \Psi_k.$$

It is an easy exercise to show that the truncation error $U - U^K$ is orthogonal to $\operatorname{span}\{\Psi_0, \dots, \Psi_K\}$. It is also worth considering how many terms there are in such a truncated gPC expansion. Suppose that the stochastic germ Ξ has dimension d (i.e. has d independent components), and we work only with polynomials of total degree at most p. The total number of coefficients in the truncated expansion U^K is

$$K + 1 = \frac{(d + p)!}{d! p!}.$$

That is, the total number of gPC coefficients that must be calculated grows combinatorially as a function of the number of input random variables and the degree of polynomial approximation. Such rapid growth limits the usefulness of gPC expansions for practical applications where d and p are much greater than the order of 10 or so.

Expansions of Random Variables. Consider a real-valued random variable U, which we expand in terms of a stochastic germ Ξ as

$$U^K(\Xi) = \sum_{k \in \mathbb{N}_0} u_k \Psi_k(\Xi),$$

where the basis functions Ψ_k are orthogonal with respect to the law of Ξ, and with the usual convention that $\Psi_0 = 1$. A first, easy, observation is that

$$\mathbb{E}[U] = \langle \Psi_0 U \rangle = \sum_{k \in \mathbb{N}_0} u_k \langle \Psi_0 \Psi_k \rangle = u_0,$$

so the expected value of U is simply its 0th gPC coefficient. Similarly, its variance is a weighted sum of the squares of its gPC coefficients:

$$\mathbb{E}\left[|U - \mathbb{E}[U]|^2\right] = \mathbb{E}\left[\left|\sum_{k \in \mathbb{N}_0} u_k \Psi_k\right|^2\right]$$

$$= \sum_{k,\ell \in \mathbb{N}} u_k u_\ell \langle \Psi_k \Psi_\ell \rangle$$

$$= \sum_{k \in \mathbb{N}} u_k^2 \langle \Psi_k^2 \rangle.$$

Similar remarks apply to any truncation $U^K = \sum_{k=1}^{K} u_k \Psi_k$ of the gPC expansion of U. In view of the expression for the variance, the gPC coefficients can be used as sensitivity indices. That is, a natural measure of how strongly U depends upon $\Psi_k(\Xi)$ is

$$\frac{u_k^2 \langle \Psi_k^2 \rangle}{\sum_{\ell \geq 1} u_\ell^2 \langle \Psi_\ell^2 \rangle}.$$

Expansions of Random Vectors. Similarly, if $U_1, \ldots, U_n$ are (not necessarily independent) real-valued random variables, then the $\mathbb{R}^n$-valued random variable $\boldsymbol{U} = [U_1, \ldots, U_n]^\mathsf{T}$ with the U_i as its components can be given a (possibly truncated) expansion

$$\boldsymbol{U}(\xi) = \sum_{k \in \mathbb{N}_0} \boldsymbol{u}_k \Psi_k(\xi),$$

with vector-valued gPC coefficients $\boldsymbol{u}_k = [u_{1,k}, \ldots, u_{n,k}]^\mathsf{T} \in \mathbb{R}^n$ for each $k \in \mathbb{N}_0$. As before,

$$\mathbb{E}[\boldsymbol{U}] = \langle \Psi_0 \boldsymbol{U} \rangle = \sum_{k \in \mathbb{N}_0} \boldsymbol{u}_k \langle \Psi_0 \Psi_k \rangle = \boldsymbol{u}_0 \in \mathbb{R}^n$$

and the covariance matrix $C \in \mathbb{R}^{n \times n}$ of U is given by

$$C = \sum_{k \in \mathbb{N}} \boldsymbol{u}_k \boldsymbol{u}_k^\mathsf{T} \langle \Psi_k^2 \rangle$$

i.e. its components are $C_{ij} = \sum_{k \in \mathbb{N}} u_{i,k} u_{j,k} \langle \Psi_k^2 \rangle$.

Expansions of Stochastic Processes. Consider now a stochastic process U, i.e. a function $U \colon \Theta \times \mathcal{X} \to \mathbb{R}$. Suppose that U is square integrable in the sense that, for each $x \in \mathcal{X}$, $U(\cdot, x) \in L^2(\Theta, \mu)$ is a real-valued random variable, and, for each $\theta \in \Theta$, $U(\theta, \cdot) \in L^2(\mathcal{X}, \mathrm{d}x)$ is a scalar field on the domain $\mathcal{X}$. Recall that

$$L^2(\Theta, \mu; \mathbb{R}) \otimes L^2(\mathcal{X}, \mathrm{d}x; \mathbb{R}) \cong L^2(\Theta \times \mathcal{X}, \mu \otimes \mathrm{d}x; \mathbb{R}) \cong L^2\big(\Theta, \mu; L^2(\mathcal{X}, \mathrm{d}x)\big),$$

so U can be equivalently viewed as a linear combination of products of $\mathbb{R}$-valued random variables with deterministic scalar fields, or as a function on $\Theta \times \mathcal{X}$, or as a field-valued random variable. As usual, take $\{\Psi_k \mid k \in \mathbb{N}_0\}$ to be an orthogonal polynomial basis of $L^2(\Theta, \mu; \mathbb{R})$, ordered (weakly) by total degree, with $\Psi_0 = 1$. A gPC expansion of the random field U is an L^2-convergent expansion of the form

$$U(x, \xi) = \sum_{k \in \mathbb{N}_0} u_k(x) \Psi_k(\xi).$$

The functions $u_k \colon \mathcal{X} \to \mathbb{R}$ are called the *stochastic modes* of the process U. The stochastic mode $u_0 \colon \mathcal{X} \to \mathbb{R}$ is the *mean field* of U:

$$\mathbb{E}[U(x)] = u_0(x).$$

The variance of the field at $x \in \mathcal{X}$ is

$$\mathbb{V}[U(x)] = \sum_{k \in \mathbb{N}} u_k(x)^2 \langle \Psi_k^2 \rangle,$$

whereas, for two points $x, y \in \mathcal{X}$,

$$\mathbb{E}[U(x)U(y)] = \left\langle \sum_{k \in \mathbb{N}_0} u_k(x) \Psi_k(\xi) \sum_{\ell \in \mathbb{N}_0} u_\ell(y) \Psi_\ell(\xi) \right\rangle$$
$$= \sum_{k \in \mathbb{N}_0} u_k(x) u_k(y) \langle \Psi_k^2 \rangle$$

and so the covariance function of U is given by

$$C_U(x, y) = \sum_{k \in \mathbb{N}} u_k(x) u_k(y) \langle \Psi_k^2 \rangle.$$

The previous remarks about gPC expansions of vector-valued random variables are a special case of these remarks about stochastic processe, namely $\mathcal{X} = \{1, \ldots, n\}$. At least when $\dim \mathcal{X}$ is low, it is very common to see the behaviour of a stochastic field U (or its truncation U^K) summarized by plots of the mean field and the variance field, as well as a few 'typical' sample realizations. The visualization of high-dimensional data is a subject unto itself, with many ingenious uses of shading, colour, transparency, videos and user interaction tools.

Changes of gPC Basis. It is possible to change between representations of a stochastic quantity U with respect to gPC bases $\{\Psi_k \mid k \in \mathbb{N}_0\}$ and $\{\Phi_k \mid k \in \mathbb{N}_0\}$ generated by measures μ and ν respectively. Obviously, for such changes of basis to work in both directions, μ and ν must at least have the same support. Suppose that

$$U = \sum_{k \in \mathbb{N}_0} u_k \Psi_k = \sum_{k \in \mathbb{N}_0} v_k \Phi_k.$$

Then, taking the $L^2(\nu)$-inner product of this equation with Φ_ℓ,

$$\langle U \Phi_\ell \rangle_\nu = \sum_{k \in \mathbb{N}_0} u_k \langle \Psi_k \Phi_\ell \rangle_\nu = v_\ell \langle \Psi_\ell^2 \rangle_\nu,$$

provided that $\Psi_k \Phi_\ell \in L^2(\nu)$ for all $k \in \mathbb{N}_0$, i.e.

$$v_\ell = \sum_{k \in \mathbb{N}_0} \frac{u_k \langle \Psi_k \Phi_\ell \rangle_\nu}{\langle \Psi_\ell^2 \rangle_\nu}.$$

Similarly, taking the $L^2(\mu)$-inner product of this equation with Ψ_ℓ yields that, provided that $\Phi_k \Psi_\ell \in L^2(\mu)$ for all $k \in \mathbb{N}_0$,

$$u_\ell = \sum_{k \in \mathbb{N}_0} \frac{v_k \langle \Phi_k \Psi_\ell \rangle_\mu}{\langle \Psi_\ell^2 \rangle_\mu}.$$

Remark 11.16. It is possible to adapt the notion of a gPC expansion to the situation of a stochastic germ Ξ with arbitrary dependencies among its components, but there are some complications. In summary, suppose that $\Xi = (\Xi_1, \ldots, \Xi_d)$, taking values in $\Theta = \Theta_1 \times \cdots \times \Theta_d$, has joint law μ, which is not necessarily a product measure. Nevertheless, let μ_i denote the marginal law of Ξ_i, i.e.

$$\mu_i(E_i) := \mu(\Theta_1 \times \cdots \times \Theta_{i-1} \times E_i \times \Theta_{i+1} \times \cdots \times \Theta_d).$$

To simplify matters further, assume that μ (resp. μ_i) has Lebesgue density ρ (resp. ρ_i). Now let $\phi_p^{(i)} \in \mathfrak{P}$, $p \in \mathbb{N}_0$, be univariate orthogonal polynomials for μ_i. The *chaos function* associated with a multi-index $\alpha \in \mathbb{N}_0^d$ is defined to be

$$\Psi_\alpha(\xi) := \sqrt{\frac{\rho_1(\xi_1) \ldots \rho_d(\xi_d)}{\rho(\xi)}} \phi_{\alpha_1}^{(1)}(\xi_1) \ldots \phi_{\alpha_d}^{(d)}(\xi_d).$$

It can be shown that the family $\{\Psi_\alpha \mid \alpha \in \mathbb{N}_0^d\}$ is a complete orthonormal basis for $L^2(\Theta, \mu; \mathbb{R})$, so we have the usual series expansion $U = \sum_\alpha u_\alpha \Psi_\alpha$. Note, however, that with the exception of $\Psi_0 = 1$, the functions Ψ_α are not polynomials. Nevertheless, we still have the usual properties that truncation error is orthogonal to the approximation subspace, and

$$\mathbb{E}_\mu[U] = u_0, \quad \mathbb{V}_\mu[U] = \sum_{\alpha \neq 0} u_\alpha^2 \langle \Psi_\alpha^2 \rangle_\mu.$$

Remark 11.17. Polynomial chaos expansions were originally introduced in stochastic analysis, and in that setting the stochastic germ Ξ typically has countably infinite dimension, i.e. $\Xi = (\Xi_1, \ldots, \Xi_d, \ldots)$. Again, for simplicity, suppose that the components of Ξ are independent, and hence orthogonal; let Θ denote the range of Ξ, which is an infinite product domain, and let $\mu = \bigotimes_{d \in \mathbb{N}} \mu_d$ denote the law of Ξ. For each $d \in \mathbb{N}$, let $\{\psi_{\alpha_d}^{(d)} \mid \alpha_d \in \mathbb{N}_0\}$ be a system of univariate orthogonal polynomials for $\Xi_d \sim \mu_d$, again with the usual convention that $\psi_0^{(d)} \equiv 1$. Products of the form

$$\psi_\alpha(\xi) := \prod_{d \in \mathbb{N}} \psi_{\alpha_d}^{(d)}(\xi_d)$$

are again polynomials when only finitely many $\alpha_d \neq 0$, and form an orthogonal system of polynomials in $L^2(\Theta, \mu; \mathbb{R})$.

As in the finite-dimensional case, there are many choices of ordering for the basis polynomials, some of which may lend themselves to particular problems. One possible orthogonal PC decomposition of $u(\Xi)$ for $u \in L^2(\Theta, \mu; \mathbb{R})$, in which summands are arranged in order of increasing 'complexity', is

$$
\begin{aligned}
u(\Xi) = f_0 &+ \sum_{d \in \mathbb{N}} u_{\alpha_d} \psi_{\alpha_d}^{(d)}(\Xi_d) \\
&+ \sum_{d_1, d_2 \in \mathbb{N}} u_{\alpha_{d_1} \alpha_{d_2}} \psi_{\alpha_{d_1}}^{(d_1)}(\Xi_{d_1}) \psi_{\alpha_{d_2}}^{(d_2)}(\Xi_{d_2}) \\
&\cdots \\
&+ \sum_{d_1, d_2, \ldots, d_k \in \mathbb{N}} u_{\alpha_{d_1} \alpha_{d_2} \ldots \alpha_{d_k}} \psi_{\alpha_{d_1}}^{(d_1)}(\Xi_{d_1}) \psi_{\alpha_{d_2}}^{(d_2)}(\Xi_{d_2}) \cdots \psi_{\alpha_{d_k}}^{(d_k)}(\Xi_{d_k}) \\
&\cdots ;
\end{aligned}
$$

i.e., writing $\Psi_{\alpha_d}^{(d)}$ for the image random variable $\psi_{\alpha_d}^{(d)}(\Xi_d)$,

$$
\begin{aligned}
U = u_0 &+ \sum_{d \in \mathbb{N}} u_{\alpha_d} \Psi_{\alpha_d}^{(d)} \\
&+ \sum_{d_1, d_2 \in \mathbb{N}} u_{\alpha_{d_1} \alpha_{d_2}} \Psi_{\alpha_{d_1}}^{(d_1)} \Psi_{\alpha_{d_2}}^{(d_2)}(\Xi_{d_2}) \\
&\cdots \\
&+ \sum_{d_1, d_2, \ldots, d_k \in \mathbb{N}} u_{\alpha_{d_1} \alpha_{d_2} \ldots \alpha_{d_k}} \Psi_{\alpha_{d_1}}^{(d_1)} \Psi_{\alpha_{d_2}}^{(d_2)} \cdots \Psi_{\alpha_{d_k}}^{(d_k)} \\
&\cdots .
\end{aligned}
$$

The PC coefficients $u_{\alpha_d} \in \mathbb{R}$, etc. are determined by the usual orthogonal projection relation. In practice, this expansion must be terminated at finite k, and provided that u is square-integrable, the L^2 truncation error decays to 0 as $k \to \infty$, with more rapid decay for smoother u, as in, e.g., Theorem 8.23.

11.4 Wavelet Expansions

Recall from the earlier discussion of Gibbs' phenomenon in Chapter 8 that expansions of non-smooth functions in terms of smooth basis functions such as polynomials, while guaranteed to be convergent in the L^2 sense, can have poor pointwise convergence properties. However, to remedy such problems, one can consider spectral expansions in terms of orthogonal bases of functions

in $L^2(\Theta, \mu; \mathbb{R})$ that are no longer polynomials: a classic example of such a construction is the use of *wavelets*, which were developed to resolve the same problem in harmonic analysis and its applications. This section considers, by way of example, orthogonal decomposition of random variables using Haar wavelets, the so-called *Wiener–Haar expansion*.

Definition 11.18. The *Haar scaling function* is $\phi(x) := \mathbb{I}_{[0,1)}(x)$. For $j \in \mathbb{N}_0$ and $k \in \{0, \ldots, 2^j - 1\}$, let $\phi_{j,k}(x) := 2^{j/2}\phi(2^j x - k)$ and

$$\mathcal{V}_j := \operatorname{span}\{\phi_{j,0}, \ldots, \phi_{j,2^j-1}\}.$$

The *Haar function* (or *Haar mother wavelet*) $\psi \colon [0,1] \to \mathbb{R}$ is defined by

$$\psi(x) := \begin{cases} 1, & \text{if } 0 \le x < \frac{1}{2}, \\ -1, & \text{if } \frac{1}{2} \le x < 1, \\ 0, & \text{otherwise.} \end{cases}$$

The *Haar wavelet family* is the collection of scaled and shifted versions $\psi_{j,k}$ of the mother wavelet ψ defined by

$$\psi_{j,k}(x) := 2^{j/2}\psi(2^j x - k) \quad \text{for } j \in \mathbb{N}_0 \text{ and } k \in \{0, \ldots, 2^j - 1\}.$$

The spaces $\mathcal{V}_j$ form an increasing family of subspaces of $L^2([0,1], \mathrm{d}x; \mathbb{R})$, with the index j representing the level of 'detail' permissible in a function $f \in \mathcal{V}_j$: more concretely, $\mathcal{V}_j$ is the set of functions on $[0,1]$ that are constant on each half-open interval $[2^{-j}k, 2^{-j}(k+1))$. A straightforward calculation from the above definition yields the following:

Lemma 11.19. *For all $j, j' \in \mathbb{N}_0$, $k \in \{0, \ldots, 2^j-1\}$ and $k' \in \{0, \ldots, 2^{j'}-1\}$,*

$$\int_0^1 \psi_{j,k}(x)\,\mathrm{d}x = 0, \quad \text{and}$$

$$\int_0^1 \psi_{j,k}(x)\psi_{j',k'}(x)\,\mathrm{d}x = \delta_{jj'}\delta_{kk'}.$$

Hence, $\{1\} \cup \{\psi_{j,k} \mid j \in \mathbb{N}_0, k \in \{0, 1, \ldots, 2^j - 1\}\}$ is a complete orthonormal basis of $L^2([0,1], \mathrm{d}x; \mathbb{R})$. If $\mathcal{W}_j$ denotes the orthogonal complement of $\mathcal{V}_j$ in $\mathcal{V}_{j+1}$, then

$$\mathcal{W}_j = \operatorname{span}\{\psi_{j,0}, \ldots, \psi_{j,2^j-1}\}, \quad \text{and}$$
$$L^2([0,1], \mathrm{d}x; \mathbb{R}) = \bigoplus_{j \in \mathbb{N}_0} \mathcal{W}_j.$$

Consider a stochastic germ $\Xi \sim \mu \in \mathcal{M}_1(\mathbb{R})$ with cumulative distribution function $F_\Xi \colon \mathbb{R} \to [0,1]$. For simplicity, suppose that F_Ξ is continuous and strictly increasing, so that F_Ξ is differentiable (with $F_\Xi' = \frac{\mathrm{d}\mu}{\mathrm{d}x} = \rho_\Xi$) almost

everywhere, and also invertible. We wish to write a random variable $U \in L^2(\mathbb{R}, \mu; \mathbb{R})$, in particular one that may be a non-smooth function of Ξ, as

$$U(\xi) = u_0 + \sum_{j \in \mathbb{N}_0} \sum_{k=0}^{2^j - 1} u_{j,k} \psi_{j,k}(F_\Xi(\xi))$$

$$= u_0 + \sum_{j \in \mathbb{N}_0} \sum_{k=0}^{2^j - 1} u_{j,k} W_{j,k}(\xi);$$

such an expansion will be called a *Wiener–Haar* expansion of U. See Figure 11.2 for an illustration comparing the cumulative distribution function of a truncated Wiener–Haar expansion to that of a standard Gaussian, showing the 'clumping' of probability mass that is to be expected of Wiener–Haar wavelet expansions but not of Wiener–Hermite polynomial chaos expansions. Indeed, the (sample) law of a Wiener–Haar expansion even has regions of zero probability mass.

Note that, by a straightforward change of variables $x = F_\Xi(\xi)$:

$$\int_{\mathbb{R}} W_{j,k}(\xi) W_{j',k'}(\xi) \, \mathrm{d}\mu(\xi) = \int_{\mathbb{R}} W_{j,k}(\xi) W_{j',k'}(\xi) \rho_\Xi(\xi) \, \mathrm{d}\xi$$

$$= \int_0^1 \psi_{j,k}(x) \psi_{j',k'}(x) \, \mathrm{d}x$$

$$= \delta_{jj'} \delta_{kk'},$$

so the family $\{W_{j,k} \mid j \in \mathbb{N}_0, k \in \{0, \dots, 2^j - 1\}\}$ forms a complete orthonormal basis for $L^2(\mathbb{R}, \mu; \mathbb{R})$. Hence, the Wiener–Haar coefficients are determined by

$$u_{j,k} = \langle U W_{j,k} \rangle = \int_{\mathbb{R}} U(\xi) W_{j,k}(\xi) \rho_\Xi(\xi) \, \mathrm{d}\xi$$

$$= \int_0^1 U(F_\Xi^{-1}(x)) \psi_{j,k}(x) \, \mathrm{d}x.$$

As in the case of a gPC expansion, the usual expressions for the mean and variance of U hold:

$$\mathbb{E}[U] = u_0 \quad \text{and} \quad \mathbb{V}[U] = \sum_{j \in \mathbb{N}_0} \sum_{k=0}^{2^j - 1} |u_{j,k}|^2.$$

Comparison of Wavelet and gPC Expansions. Despite the formal similarities of the corresponding expansions, there are differences between wavelet and gPC spectral expansions. For gPC expansions, the globally smooth orthogonal polynomials used as the basis elements have the property that expansions of smooth functions/random variables enjoy a fast convergence
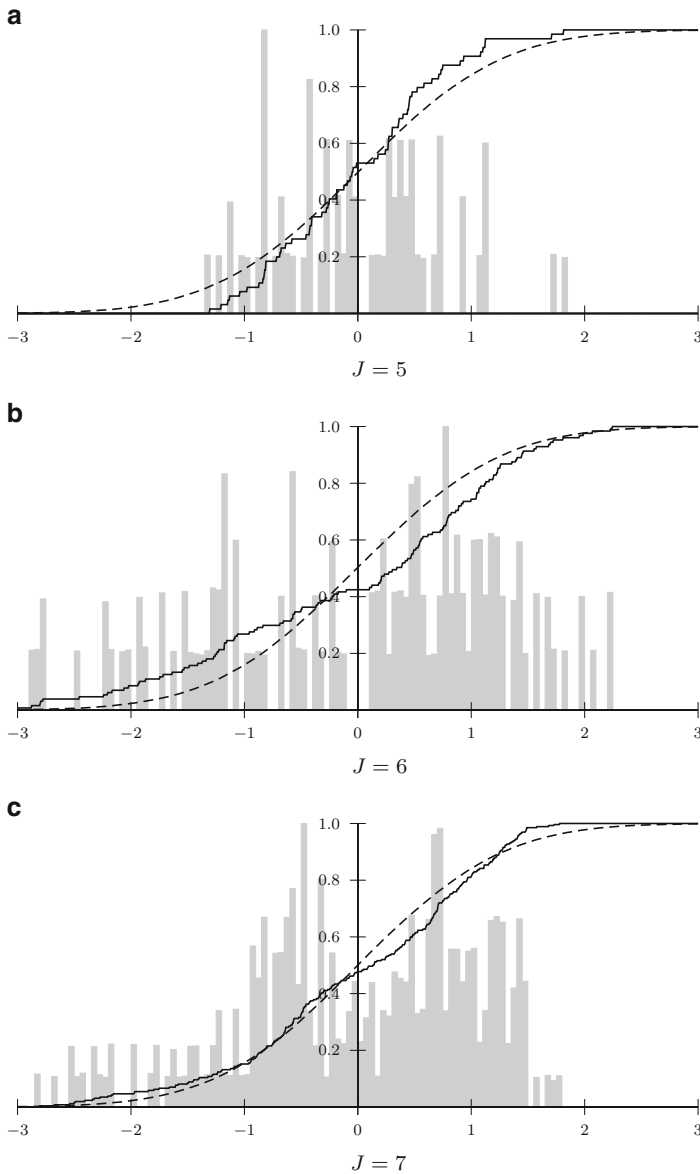
a

b

c

Fig. 11.2: The cumulative distribution function and binned peak-normalized probability density function of 10^5 i.i.d. samples of a random variable U with truncated Wiener–Haar expansion $U = \sum_{j=0}^{J} \sum_{k=0}^{2^j-1} u_{j,k} W_{j,k}(\Xi)$, where $\Xi \sim \mathcal{N}(0,1)$. The coefficients $u_{j,k}$ were sampled independently from $u_{j,k} \sim 2^{-j}\mathcal{N}(0,1)$. The cumulative distribution function of a standard Gaussian is shown dashed for comparison.

rate, as in Theorem 8.23; no such connection between smoothness and convergence rate is to be expected for Wiener–Haar expansions, in which the basis functions are non-smooth. However, in cases in which U shows a localized sharp variation or a discontinuity, a Wiener–Haar expansion may be more efficient than a gPC expansion, since the convergence rate of the latter would be impaired by Gibbs-type phenomena. Another distinctive feature of the Wiener–Haar expansion concerns products of piecewise constant processes. For instance, for $f, g \in \mathcal{V}_j$ the product fg is again an element of $\mathcal{V}_j$; it is not true that the product of two polynomials of degree at most n is again a polynomial of degree at most n. Therefore, for problems with strong dependence upon high-degree/high-detail features, or with multiplicative structure, Wiener–Haar expansions may be more appropriate than gPC expansions.

11.5 Bibliography

Mercer (1909) proved Theorem 11.3 for positive semi-definite kernels on $[a, b] \times [a, b]$; the general theorem as used here can be found in many standard works on linear functional analysis, e.g. that of Dunford and Schwartz (1963, pp. 1087–1088); Steinwart and Scovel (2012) consider Mercer-type theorems for non-compact domains. The Karhunen–Loève expansion bears the names of Karhunen (1947) and Loève (1978), but Karhunen–Loève-type series expansions of stochastic processes were considered earlier by Kosambi (1943). Jolliffe (2002) gives a general introduction to principal component analysis, and de Leeuw (2013) gives a survey of the history of PCA and its nonlinear generalizations.

The application of Wiener–Hermite PC expansions to engineering systems was popularized by Ghanem and Spanos (1991); the extension to gPC and the connection with the Askey scheme is due to Xiu and Karniadakis (2002). The extension of gPC expansions to arbitrary dependency among the components of the stochastic germ, as in Remark 11.16, is due to Soize and Ghanem (2004). For a more pedagogical approach, see the discussion of the UQ applications of spectral expansions in the books of Le Maître and Knio (2010, Chapter 2), Smith (2014, Chapter 10), and Xiu (2010, Chapter 5).

The orthogonal decomposition properties of the Haar basis were first noted by Haar (1910). Meyer (1992) provides a thorough introduction to wavelets in general. Wavelet bases for UQ, which can better resolve locally non-smooth features of random fields, are discussed by Le Maître and Knio (2010, Chapter 8) and in articles of Le Maître et al. (2004a,b, 2007). Wavelets are also used in the construction and sampling of Besov measures, as in the articles of Dashti et al. (2012) and Lassas et al. (2009), and Theorem 11.10 is synthesized from results in those two papers. A thorough treatment of Besov spaces from a Fourier-analytic perspective can be found in Bahouri et al. (2011, Chapter 2).

11.6 Exercises

Exercise 11.1. Consider the negative Laplacian operator $\mathcal{L} := -\frac{d^2}{dx^2}$ acting on real-valued functions on the interval $[0, 1]$, with zero boundary conditions. Show that the eigenvalues μ_n and normalized eigenfunctions ψ_n of $\mathcal{L}$ are

$$\mu_n = (\pi n)^2,$$
$$\psi_n(x) = \sqrt{2}\sin(\pi n x).$$

Hence show that $C := \mathcal{L}^{-1}$ has the same eigenfunctions with eigenvalues $\lambda_n = (\pi n)^{-2}$. Hence, using the Karhunen–Loève theorem, generate figures similar to Figure 11.1 for your choice of mean field $m\colon [0, 1] \to \mathbb{R}$.

Exercise 11.2. Do the analogue of Exercise 11.1 for $\mathcal{L} = (-\Delta)^\alpha$ acting on real-valued functions on the square $[0, 1]^2$, again with zero boundary conditions. Try $\alpha = 2$ first, then try $\alpha = 1$, and try coarser and finer meshes in each case. You should see that your numerical draws from the Gaussian field with $\alpha = 1$ fail to converge, whereas they converge nicely for $\alpha > 1$. Loosely speaking, the reason for this is that a Gaussian random variable with covariance $(-\Delta)^\alpha$ is almost surely in the Sobolev space H^s or the Hölder space $\mathcal{C}^s$ for $s < \alpha - \frac{d}{2}$, where d is the spatial dimension; thus, $\alpha = 1$ on the two-dimensional square is exactly on the borderline of divergence.

Exercise 11.3. Show that the eigenvalues λ_n and eigenfunctions e_n of the exponential covariance function $C(x, y) = \exp(-|x - y|/a)$ on $[-b, b]$ are given by

$$\lambda_n = \begin{cases} \frac{2a}{1+a^2 w_n^2}, & \text{if } n \in 2\mathbb{Z}, \\ \frac{2a}{1+a^2 v_n^2}, & \text{if } n \in 2\mathbb{Z}+1, \end{cases}$$

$$e_n(x) = \begin{cases} \sin(w_n x)/\sqrt{b - \frac{\sin(2w_n b)}{2w_n}}, & \text{if } n \in 2\mathbb{Z}, \\ \cos(v_n x)/\sqrt{b + \frac{\sin(2v_n b)}{2v_n}}, & \text{if } n \in 2\mathbb{Z}+1, \end{cases}$$

where w_n and v_n solve the transcendental equations

$$\begin{cases} aw_n + \tan(w_n b) = 0, & \text{for } n \in 2\mathbb{Z}, \\ 1 - av_n \tan(v_n b) = 0, & \text{for } n \in 2\mathbb{Z}+1. \end{cases}$$

Hence, using the Karhunen–Loève theorem, generate sample paths from the Gaussian measure with covariance kernel C and your choice of mean path. Note that you will need to use a numerical method such as Newton's method to find approximate values for w_n and v_n.

Exercise 11.4 (Karhunen–Loève-type sampling of Besov measures). Let $\mathbb{T}^d := \mathbb{R}^d/\mathbb{Z}^d$ denote the d-dimensional unit torus. Let $\{\psi_\ell \mid \ell \in \mathbb{N}\}$ be an orthonormal basis for $L^2(\mathbb{T}^d, dx; \mathbb{R})$. Let $q \in [1, \infty)$ and $s \in (0, \infty)$, and define a new norm $\|\cdot\|_{X^{s,q}}$ on series $u = \sum_{\ell \in \mathbb{N}} u_\ell \psi_\ell$ by

$$\left\| \sum_{\ell \in \mathbb{N}} u_\ell \psi_\ell \right\|_{X^{s,q}} := \left(\sum_{\ell \in \mathbb{N}} \ell^{\frac{sq}{d} + \frac{q}{2} - 1} |u_\ell|^q \right)^{1/q}.$$

Show that $\|\cdot\|_{X^{s,q}}$ is indeed a norm and that the set of u with $\|u\|_{X^{s,q}}$ finite forms a Banach space. Now, for $q \in [1, \infty)$, $s > 0$, and $\kappa > 0$, define a random function U by

$$U(x) := \sum_{\ell \in \mathbb{N}} \ell^{-(\frac{s}{d} + \frac{1}{2} - \frac{1}{q})} \kappa^{-\frac{1}{q}} \Xi_\ell \psi_\ell(x)$$

where Ξ_ℓ are sampled independently and identically from the generalized Gaussian measure on $\mathbb{R}$ with Lebesgue density proportional to $\exp(-\frac{1}{2}|\xi|^q)$. By treating the above construction as an infinite product measure and considering the product of the densities $\exp(-\frac{1}{2}|\xi_\ell|^q)$, show formally that U has 'Lebesgue density' proportional to $\exp(-\frac{\kappa}{2}\|u\|_{X^{s,q}}^q)$.

Generate sample realizations of U and investigate the effect of the various parameters q, s and κ. It may be useful to know that samples from the probability measure $\frac{\beta^{1/2}}{2\Gamma(1+\frac{1}{q})} \exp(-\beta^{q/2}|x - m|^q) \, \mathrm{d}x$ can be generated as $m + \beta^{-1/2} S |Y|^{1/q}$ where S is uniformly distributed in $\{-1, +1\}$ and Y is distributed according to the gamma distribution on $[0, \infty)$ with parameter q, which has Lebesgue density $q e^{-qx} \mathbb{I}_{[0,\infty)}(x)$.