Spectral Representations of Stochastic Processes

BY STEPHEN CROWLEY

April 13, 2024

Abstract

This is a nearly verbatim reproduction of most of the first two chapters of Roger G. Ghanem and Pol D. Spanos. *Stochastic finite elements: a spectral approach*. Springer-Verlag, Berlin, Heidelberg, 1991.

Table of contents

1 The Spectral Approach to Stochastic Processes

1.1 The Gelfand-Vilenkin Spectral Representation

The central concepts of the theory of representation for random processes can be formulated where the majority of the results have been derived for the class of second order processes.[1] A very important result is the `Gelfand-Vilenkin spectral representation of random processes` which, in its most general form, can be stated as

$$w(x,\theta) = \int g(x)\, d\,\mu(\theta) \tag{1}$$

where $w\,(x,\theta)$ is a stochastic process whose covariance function $C_{ww}(x_1,x_2)$ can be expressed as

$$C_{ww}(x_1,x_2)=\int g(x_1)\,g(x_2)\,d\,\mu_1(\theta)\,d\,\mu_2(\theta) \tag{2}$$

In equation (1), $g(x)$ is a deterministic function. Further, $d\,\mu(\theta)$ is an orthogonal set function, also termed orthogonal stochastic measure, defined on the σ-field Ψ of random events. An important specialization of the spectral decomposition occurs if the process $w\,(x,\theta)$ is wide stationary. In this case, equation (1) can be shown to reduce to the Wiener-Khintchine relation (Yaglom, 1962) and the following equations hold

$$w\,(x,\theta)=\int_{-\infty}^{\infty} e^{i\omega^T x}\,d\,\mu(\omega,\theta) \tag{3}$$

and

$$C_{ww}(x_1,x_2)=\int_{-\infty}^{\infty} e^{i(\omega_1 x_1-\omega_2 x_2)^T}\,S(\omega)\,d\,\omega \tag{4}$$

Here, the symbol T denotes vector transposition, $S(\omega)$ is the usual `spectral density` of the stationary process, and ω is the `wave number vector`. In the equation for $C_{ww}(x_1,x_2)$:

- $e^{i(\omega_1 x_1-\omega_2 x_2)^T}$ represents the complex exponential function used in Fourier transforms, indicating oscillations at frequencies ω_1 and ω_2 along dimensions x_1 and x_2, respectively.

- ω_1 and ω_2 are spatial frequency components corresponding to different axes/dimensions (e.g., x_1 and x_2). They indicate how the wave's frequency components vary along these dimensions.

- The term $S(\omega)$ represents the spectral density of the process at frequency ω, detailing how energy is distributed across different frequencies.

- The integral sums these contributions across all frequencies to calculate the covariance between two points x_1 and x_2 in the process, incorporating the spatial structure and dependence of the process.

The impact of these spectral representations on the theory of random processes is significant. However, their use has been limited to deterministic systems subjected to random excitations. This limitation arises because these representations require differentials of random functions, placing them in an infinite-dimensional space, which complicates the development of computational algorithms. Similarly,[2, Ch.1 Bibliogaphical notes, p.18] highlights a challenge in extending von Neumann's zero-sum two-person games to stochastic processes, hindered by the same issue of infinite-dimensionality.

1.2 The Karhunen-Loeve Expansion

Another spectral representation theorem, and one which is extensively used in the sequel, is the Karhunen-Loeve expansion whereby a random process $w(x,\theta)$ can be expanded in terms of a denumerable set of orthogonal random variables in the form

$$w(x,\theta) = \sum_{i=1}^{\infty} \mu_i(\theta)\, g_i(x) \tag{5}$$

where $\{\mu_i(\theta)\}$ is a set of orthogonal random variables and $\{g_i(x)\}$ are deterministic functions, which can be related to the covariance kernel of $w(x,\theta)$. Note that since equation (?) constitutes a representation of the random process in terms of a denumerable set of random variables, it may be regarded a quantizationd of the random process. It is important to note that this equation can be viewed as a representation of the process $w(x,\theta)$ as a curve in the Hilbert space spanned by the set $\{g_i(x)\}$ expressed as a direct sum of orthogonal projections in this Hilbert space whereby the magnitudes of the projections on successive basis vectors are proportional to the corresponding eigenvalues of the covariance function associated with the process $w(x,\theta)$. Collectively, the representations discussed up to [1, 2.2] can be thought of as linear operators or filters acting on processes with independent increments.

1.2.1 Derivation

One of the major difficulties associated with the numerical incorporation of random processes in finite element analyses, is the necessity of dealing with infinite-dimensional spaces. The major conceptual difficulty from the viewpoint of the class of problems considered involves the treatment of functions defined on abstract spaces of random variables defined on σ-fields of random events. The most widely used method, the Monte Carlo simulation, consists of sampling these functions at randomly chosen elements of this σ-field, in a random, collocation-like, scheme. Obviously, a quite large number of points needs to be sampled if a good approximation is to be achieved. I've always thought the method was a terrible hack and led to no good. The Karhunen-Loeve expansion is a muc theoretically elegant and computationally appealing way to deal with abstract measures spaces can sometimes have limited physical intuitive support. It works by expanding functions in Fourier-like series as

$$w(x,\theta) = \sum_{n=0}^{\infty} \sqrt{\lambda_n}\, \xi_n(\theta)\, f_n(x) \tag{6}$$

where $\{\xi_n(\theta)\}$ is a usually said to be a set of random variables to be determined, λ_n is some constant, and $\{f_n(x)\}$ is an orthonormal set of deterministic functions. I find it a bit of a misnomer to refer to the sequence $\xi_n(\theta)$ as random variables because none of the mathematics is dependent upon the stochastic nature to which they are usually applied or regarded. For instance, the realization of a well known complex analytic function with a very elaborate behaviour can be shown to be the principal realization of a Gaussian process having a certain kernel. In this instance, even though we use the mathematics of randomness and stochastics the function to which it applies can be computed to any degree of precision at any given point and in that sense there is nothing random about it, other than the fact that I have yet to point out to everyone how it comes together.

1.2.2 Spectral Covariance Representations

Let $w\,(x,\theta)$ be a random process, function of the position vector x defined over the domain D, with θ belonging to the space of random events Ω. Let $\bar{w}(x)$ denote the expected value of $w\,(x,\theta)$ over all possible realizations of the process, and $C(x_1,x_2)$ denote its covariance function. By definition of the covariance function, it is bounded, symmetric and positive definite. Thus, it has the spectral decomposition [1, 2.3.1]

$$C(x_1,x_2)=\sum_{n=0}^{\infty}\lambda_n\,f_n(x_1)\,f_n(x_2) \tag{7}$$

where λ_n and $f_n(x)$ are the eigenvalue and the eigenfunction of the covariance kernel. And, specifically, that they are the solution to the integral equation

$$\int_D C(x_1,x_2)\,f_n(x_1)\,d\,x_1=\lambda_n\,f_n(x_2) \tag{8}$$

Due to the symmetry and the positive definiteness of the covariance kernel, its eigenfunctions are orthogonal and form a complete set. They can be normalized according to the following criterion

$$\int_D f_n(x)\,f_m(x)\,d\,x=\delta_{nm} \tag{9}$$

where δ_{nm} is the Kronecker delta. Clearly, $w(x,\theta)$ can be written as

$$w\,(x,\theta)=\bar{w}(x)+\alpha(x,\theta) \tag{10}$$

where $\alpha(x,\theta)$ is a process with zero mean and covariance function $C(x_1,x_2)$. The process $\alpha(x,\theta)$ can be expanded in terms of the eigenfunctions $f_n(x)$ as

$$\alpha(x,\theta)=\sum_{n=0}^{\infty}\xi_n(\theta)\sqrt{\lambda_n}\,f_n(x) \tag{11}$$

Second order properties of the random variables ξ_n can be determined by multiplying both sides of equation (11) by $\alpha(x_2,\theta)$ and taking the expectation on both sides. Specifically, it is found that

$$\begin{aligned} C(x_1,x_2) &= \langle\alpha(x_1,\theta)\,\alpha(x_2,\theta)\rangle \\ &= \sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\langle\xi_n(\theta)\,\xi_m(\theta)\rangle\sqrt{\lambda_n\,\lambda_m}\,f_n(x_1)\,f_m(x_2) \end{aligned} \tag{12}$$

Then, multiplying both sides of equation (12) by $f_k(x_2)$, integrating over the domain D, and making use of the orthogonality of the eigenfunctions, yields

$$\begin{aligned}\int_D C(x_1,x_2)\, f_k(x_2)\, d\,x_2 \; &= \lambda_k\, f_k(x_1) \\ &= \sum_{n=0}^{\infty} \langle \xi_n(\theta)\, \xi_k(\theta)\rangle \sqrt{\lambda_n\, \lambda_k}\; f_n(x_1)\end{aligned} \tag{13}$$

Multiplying once more by $f_l(x_1)$ and integrating over D gives

$$\int_D\int_D f_l(x_1)\, f_k(x_1)\, d\,x_1 = \sum_{n=0}^{\infty} E\, \langle \xi_n(\theta)\, \xi_k(\theta)\rangle \sqrt{\lambda_n\, \lambda_k}\, \delta_{nl} \tag{14}$$

Then, using equation (9) leads to

$$\lambda_k\, \delta_{kl} = \sqrt{\lambda_k\, \lambda_l}\, \langle \xi_k(\theta)\, \xi_l(\theta)\rangle \tag{15}$$

Equation (15) can be rearranged to give

$$\langle \xi_k(\theta)\, \xi_l(\theta)\rangle = \delta_{kl} \tag{16}$$

Thus, the random process $w\,(x,\theta)$ can be written as

$$w\,(x,\theta) = \bar{w}(x) + \sum_{n=0}^{\infty} \xi_n(\theta) \sqrt{\lambda_n}\; f_n(x) \tag{17}$$

where

$$\langle \xi_n(\theta)\rangle = 0 \tag{18}$$

$$\langle \xi_n(\theta)\, \xi_m(\theta)\rangle = \delta_{nm} \tag{19}$$

and λ_n, $f_n(x)$ are solution to equation (8). Truncating the series in equation (17) at the M^{th} term, gives

$$w\,(x,\theta) = \bar{w}(x) + \sum_{n=0}^{M} \xi_n(\theta) \sqrt{\lambda_n}\; f_n(x) \tag{20}$$

An explicit expression for $\xi_n(\theta)$ can be obtained by multiplying equation (11) by $f_n(x)$ and integrating over the domain D. That is,

$$\xi_n(\theta) = \frac{\int_D \alpha(x,\theta)\, f_n(x)\, d\,x}{\sqrt{\lambda_n}} \tag{21}$$

1.2.3 Reproducing Kernel Hilbert Spaces (RKHS)

The essential thing that the Reproducing Kernel Hilbert Space (RKHS) (Aronszajn, 1950; Parzen, 1959), either of equations (11) or (20), is an expression for the congruence that maps the Hilbert space spanned by the functions $f_n(x)$ to the Hilbert space spanned by the random process, or equivalently, the space spanned by the set of random variables $\{\xi_n(\theta)\}$. It is this congruence along with the covariance function of the process that determines uniquely the random process $w\,(x,\theta)$. Observe the similarity of equations (11) and (20) with equations (7) and (8), respectively. Indeed, it can be shown (Parzen, 1959) that if a function can be represented in terms of linear operations on the family $\{C(\,\cdot,x_2)\}$, then f belongs to the RKHS corresponding to the kernel $C(x_1,x_2)$, and the congruence between the two Hilbert spaces means of the same linear operations used to represent f in terms of $\{C(\,\cdot,x_2)\}$, $x_2 \in D$.

Another point of practical importance is that the expansion given by equation (19) can be used in a numerical simulation scheme to obtain numerical realizations of the random process. It is optimal in the Fourier sense, as it minimizes the mean square error resulting from truncation after a finite number of terms.It is well known from functional analysis that the steeper a bilinear form decays to zero as a function of one of its arguments, the more terms are needed in its spectral representation in order to reach a preset accuracy. Noting that the Fourier transform operator is a spectral representation, it may be concluded that the faster the autocorrelation function tends to zero, the broader is the corresponding spectral density, and the greater the number of requisite terms to represent the underlying random process by the Karhunen-Loeve expansion.

1.3 Properties

1.3.1 Error Minimization

Lemma 1

Error Minimization: *The generalized coordinate system defined by the eigenfunctions of the covariance kernel is optimal in the sense that the mean-square error resulting from a finite representation of the process $w\,(x,\theta)$ is minimized.*

Proof. Given a complete orthonormal set of functions $h_n(x)$, the process $w\,(x,\theta)$ can be approximated in a convergent series of the form

$$w\,(x,\theta) = \sum_{n=0}^{\infty} \lambda_n\, \xi_n(\theta)\, h_n(x) \tag{22}$$

Truncating equation (21) at the M^{th} term results in an error ϵ_M equal to

$$\epsilon_M = \sum_{n=M+1}^{\infty} \lambda_n\, \xi_n(\theta)\, h_n(x) \tag{23}$$

Multiplying equation (21) by $h_m(x)$ and integrating throughout gives

$$\xi_m(\theta) = \frac{\int_D w\,(x,\theta)\, h_m(x)\, d\,x}{\sqrt{\lambda_m}} \tag{24}$$

where use is made of the orthogonality property of the set $h_n(x)$. Substituting equation (23) for $\xi_m(\theta)$ back into equation (22), the mean-square error ϵ_M^2 can be written as

$$\epsilon_M^2 = \left[\sum_{m=M+1}^{\infty} \sum_{n=M+1}^{\infty} \int_D \int_D \langle \xi_m(\theta)\, \xi_n(\theta) \rangle\, h_m(x_1)\, h_n(x_2)\, d\,x_1\, d\,x_2 \right]^2 \tag{25}$$

Integrating equation (24) over D and using the orthonormality of the set $\{h_i(x)\}$ yields

$$\int_D \epsilon_M^2\, d\,x = \sum_{m=M+1}^{\infty} \int_D \int_D R_{ww}(x_1, x_2)\, h_m(x_1)\, h_m(x_2)\, d\,x_1\, d\,x_2 \tag{26}$$

The problem, then, is to minimize $\int_D \epsilon_M^2$ subject to the orthonormality of the functions $h_n(x)$. In other words, the solution minimizes the functional given by the equation

$$\mathcal{F}[h(x)] = \sum_{m=M+1}^{\infty} \int_D \int_D R_{ww}(x_1, x_2)\, h_m(x_1)\, h_m(x_2)\, d\,x_1\, d\,x_2 - \lambda_m \left[\int_D h_m(x)\, h_m(x)\, d\,x - 1 \right] \tag{27}$$

Differentiating equation (26) with respect to $h_i(x)$ and setting the result equal to zero, gives

$$\frac{\partial \mathcal{F}[h(x)]}{\partial h_i(x)} = \int_D \int_D R_{ww}(x_1, x_2)\, h_i(x_1)\, d\,x_1 - \lambda_i\, h_i(x_2)\, d\,x_2 = 0 \tag{28}$$

which is satisfied when

$$\int_D R_{ww}(x_1, x_2)\, h_i(x_2)\, d\,x_2 = \lambda_i\, h_i(x_1) \tag{29}$$ □

1.3.2 Uniqueness of the Expansion

Lemma 2

Uniqueness: *The random variables appearing in an expansion of the kind given by equation (10) are orthonormal if and only if the orthonormal functions $\{f_n(x)\}$ and the constants $\{\lambda_n\}$ are respectively the eigenfunctions and the eigenvalues of the covariance kernel as given by equation (8).*

Proof. The "if" part is an immediate consequence of equation (11). To show the "only if" part, equation (12) can be used with

$$\langle \xi_n(\theta)\, \xi_m(\theta) \rangle = \delta_{nm} \tag{30}$$

to obtain

$$C(x_1, x_2) = \sum_{n=0}^{\infty} \lambda_n \, f_n(x_1) \, f_n(x_2) \tag{31}$$

Multiplying both sides by $f_m(x_2)$ and integrating over D gives

$$\int_D C(x_1, x_2) \, f_m(x_2) \, d\,x_2 = \sum_{n=0}^{\infty} \lambda_n \, f_n(x_1) \, \delta_{nm} = \lambda_m \, f_m(x_1) \tag{32}$$ □

1.3.3 Expansion of Gaussian Processes

Let $w\,(x, \theta)$ be a Gaussian process with covariance function $C(x_1, x_2)$. Then $w\,(x, \theta)$ has the Karhunen-Loeve decomposition given by equation (17) with the random variables $\xi_i(\theta)$ forming a Gaussian vector. That is, any subset of $\{\xi_i(\theta)\}$ is jointly Gaussian. Since these random variables are uncorrelated, their Gaussian property implies their independence. Some important consequences derive from this property. Specifically,

$$\langle \xi_1(\theta), \ldots, \xi_{2n+1}(\theta) \rangle = 0 \tag{33}$$

and

$$\langle \xi_1(\theta), \ldots, \xi_{2n}(\theta) \rangle = \sum \prod \langle \xi_i(\theta) \, \xi_j(\theta) \rangle \tag{34}$$

where the summation extends over all the partitions of the set $\{\xi_i(\theta)\}_{i=1}^{2n}$ into sets of two elements, and the product is over all such sets in a given partition. Furthermore, it can be shown (Loeve, 1977) that for Gaussian processes, the Karhunen-Loeve expansion is almost surely convergent.

1.3.4 Minimum Representation Entropy

The minimum representation of entropy property is worth mentioning even though it only references and no details were given in [1].

1.4 Solution of the Integral Equation

The usefulness of the Karhunen-Loeve expansion hinges on the ability to solve the integral equation of the form

$$\int_D C(x_1, x_2) \, f(x_2) \, d\,x_2 = \lambda \, f(x_1) \tag{35}$$

where $C(x_1, x_2)$ is an autocovariance function. Equation (35) is a homogeneous Fredholm integral equation of the second kind. Being an autocovariance function, by definition the kernel $C(x_1, x_2)$ is bounded, symmetric, and positive definite. This fact simplifies the ensuing analysis considerably by virtue of the fact that positive definite kernels implies the existence of an RKHS for the kernel and its spectral density given by its Fourier transform is non-negative and non-decreasing over its domain of definition by Aronszajn's Theorem and Bochner's theorem respectively.

1.5 Irrational Spectra

There is no general method for the solution of the integral equation (35) corresponding to irrational spectra.

1.5.1 Wiener process

Another kernel that may be treated by the same method is the kernel of the Wiener process. It is given by the equation

$$C(x_1, x_2) = \min(x_1, x_2) \forall (x_1, x_2) \in [0, T] \times [0, T] \tag{36}$$

The resulting normalized eigenfunctions and eigenvalues are

$$f_n(x_1) = \sqrt{2} \sin\left(\frac{n \pi x_1}{T}\right) \tag{37}$$

and

$$\lambda_n = \frac{4 T^2}{\pi^2 (2n+1)^2} \forall n = 0, 1, \ldots \tag{38}$$

Note that the Wiener process is an example of a nonstationary process, a fact that emphasizes the generality of the Karhunen-Loeve expansion and its applicability to such processes.

Remark 3. A final remark is in order concerning the choice of the domain D of definition of the random process being investigated. Taking D to be the finite domain over which the process is being observed may often be the most obvious choice. Clearly, this choice does not induce the ergodic assumption for the process, which involves observing infinite length records. This is by no means a handicap of this approach since the ergodic assumption is usually introduced for convenience and is not necessary for the present study. If ergodicity is needed for some particular problem, then it may be recovered by extending the limits of integration in equation (35) to infinity.[1, 2.3.2]

1.6 Numerical Solution

In this section, a Galerkin type procedure is described for the solution of the Fredholm equation (35). Let $h_i(x)$ be a complete set of functions in the Hilbert space H. Each eigenfunction of the kernel $C(x_1, x_2)$ may be represented as

$$f_k(x) = \sum_{i=1}^{N} d_i^{(k)} h_i(x) \tag{39}$$

with an error e_N resulting from truncating the summation after the Nth term. This error is equal to the difference between the left hand side and the right hand side of equation (35). Substituting equation (39) into equation (35) yields the following expression for the error

$$e_N = \sum_{i=1}^{N} d_i^{(k)} \left[\int_D \int_D C(x_1, x_2) h_i(x_2) \, d x_2 - \lambda_n h_i(x_1) \right] \tag{40}$$

Requiring the error to be orthogonal to the approximating space yields equations of the following form $\forall j = 1, \ldots, N$

$$(e_N, h_j(x)) = 0 \tag{41}$$

Equivalently,

$$\sum_{i=1}^{N} d_i^{(k)} \left[\int_D \int_D C(x_1, x_2) h_i(x_2) \, d x_2 \, h_j(x_1) \, d x_1 \right] - \lambda_n \int_D h_i(x) h_j(x) \, d x = 0 \tag{42}$$

Denoting

$$C_{ij} = \int_D \int_D C(x_1, x_2) h_i(x_2) h_j(x_1) \, d x_1 \, d x_2 \tag{43}$$

$$B_{ij} = \int_D h_i(x) h_j(x) \, d x \tag{44}$$

$$D_{ij} = d_i^{(j)} \tag{45}$$

$$\Lambda_{ij} = \delta_{ij} \lambda_j \tag{46}$$

equation (42) becomes

$$C D = \Lambda B D \tag{47}$$

where C, B and D are three N-dimensional matrices whose elements are given by equations (43)-(44). Equation (47) represents a generalized algebraic eigenvalue problem which may be solved for the matrix D and the eigenvalues λ_k. Backsubstituting into equation (39) yields the eigenfunctions of the covariance kernel. The preceding procedure can be implemented using piecewise polynomials as the basis for the expansion. With this choice of basis functions, the columns of the matrix $\mathbf{D}$ become the eigenvectors computed at the respective nodal points of the induced mesh, and the $i\,j^{th}$ element of the matrix $\mathbf{C}$ becomes the weighted correlation between the process at nodes i and j. Note that both matrices $\mathbf{C}$ and $\mathbf{B}$ are symmetric positive definite, a fact that substantially simplifies the numerical solution. The Galerkin scheme described above can be shown to be equivalent to a variational treatment of the problem. This property ensures that the computed eigenvalues are a lower bound of the correspondingly numbered exact eigenvalues. This implies that the convergence of each eigenvalue is monotonic in N. Further, note that the accuracy in estimating the eigenvalues is better than that achieved for the eigenfunctions (Delves and Mohamed, 1985).

Bibliography

[1] Roger G. Ghanem and Pol D. Spanos. *Stochastic finite elements: a spectral approach*. Springer-Verlag, Berlin, Heidelberg, 1991.

[2] Malempati M. Rao. *Stochastic Processes: Inference Theory*. Springer Monographs in Mathematics. Springer, 2nd edition, 2014.