Polynomial chaos

The theory of polynomial chaos dates back to Wiener [18, 1]. It was originally developed in a rather general/abstract setting, i.e., to represent L^2 functionals of the Browninan motion (Wiener) process $W(t;\omega)$. What is a functional of the Brownian motion process? Think about the solution to the following SDE driven by $W(t;\omega)$.

$$dX(t;\omega) = m(X(t;\omega))dt + dW(t;\omega). \tag{1}$$

The solution of the ODE at final time $t = T$, $X(T;\omega)$ is a functional of the Brownian motion $W(t;\omega)$, i.e., it depends on the whole path of $W(t;\omega)$ from $t = 0$ to $t = T$. The main result of Wiener's theory is that *any* functional of the Brownian motion can be expanded in the so-called Wiener-Hermite series involving orthogonal[1] polynomial functionals of the Brownian motion, and the expansion converges in the mean square sense. This has important applications in nonlinear system identification. In fact, by measuring the response of a nonlinear system to Gaussian white noise we can in principle "identify" the system, i.e., determine a functional power series expansion that allows us to compute the output of the system corresponding to arbitrary inputs. Denote by $C_0([0,T])$ the set of continuous functions on the interval $[0,T]$ vanishing at zero and

$$F : C_0([0,T]) \to \mathbb{R}$$

a real-valued functional mapping functions on $C_0([0,T])$ onto the real line, then we have the following series expansion [1, 2]

$$F([W(t;\omega)]) = \lim_{N\to\infty} \sum_{n=0}^{N} G_n([W(t;\omega)]), \tag{2}$$

where G_0 is a constant, and $G_n([W(t;\omega)])$ are Wiener-Hermite polynomials functionals. The first two of such functionals are[2] [19, p. 32]

$$G_1([W]) = \int_0^T \kappa_1(t_1) dW(t_1;\omega), \tag{4}$$

$$G_2([W]) = \int_0^T \int_0^T \kappa_2(t_1,t_2) dW(t_1;\omega) dW(t_2;\omega) - \int_0^T \kappa_2(t_1,t_1) dt_1, \tag{5}$$

where

$$\int_0^T \kappa_1(t_1)^2 dt_1 = 1, \qquad 2\int_0^T \int_0^T \kappa_2(t_1,t_2)^2 dt_1 dt_2 = 1. \tag{6}$$

Note that G_1 and G_2 are random variables. The kernel functions κ_1, κ_2, etc., satisfy a number of conditions that follow from the orthogonality requirements

$$\mathbb{E}\left\{G_0, G_1\right\} = G_0 \mathbb{E}\left\{G_1\right\} = 0, \qquad \mathbb{E}\left\{G_0, G_2\right\} = G_0 \mathbb{E}\left\{G_2\right\} = 0, \qquad \mathbb{E}\left\{G_1, G_2\right\} = 0 \tag{7}$$

and the normalization conditions (see [19, Lecture 3])

$$\mathbb{E}\left\{G_0^2\right\} = \mathbb{E}\left\{G_1^2\right\} = \mathbb{E}\left\{G_2^2\right\} = 1. \tag{8}$$

Note that (4)-(6) already include (7) and (8).

[1] Wiener-Hermite polynomial functionals are orthogonal with respect to the Gaussian measure.

[2] Note that the integrals in (4)-(5) do not exist in the ordinary Stieltjes sense because $W(t;\omega)$ is nowhere differentiable. However, we can get around this by defining the integrals such as (4) using integration by parts as

$$G_1([W]) = \int_0^T \kappa_1(t_1) dW(t_1;\omega) = \kappa_1(T)W(T;\omega) - \kappa_1(0)W(0;\omega) - \int_0^T \kappa_1'(t_1) W(t_1;\omega) dt_1 = \kappa_1(T)W(T;\omega) - \int_0^T \kappa_1'(t_1) W(t_1;\omega) dt_1. \tag{3}$$

Computation of Wiener-Hermite kernels. The series expansion (2) allows us to identify a nonlinear system by simply recording its response to Gaussian white noise [11, 13]. To this end, first average (2) to obtain

$$G_0 = \mathbb{E}\left\{F([W])\right\}. \tag{9}$$

We then multiply (2) left and right by the following first-order Hermite functional (same form as $G_1([W])$ but with different kernel)

$$Q_1([W]) = \int_0^T g_1(s) dW(s;\omega) \tag{10}$$

where $g_1(s)$ is to be chosen. We then average to obtain

$$\mathbb{E}\left\{F([W]) \int_0^T g(s) dW(s;\omega)\right\} = \int_0^T \kappa_1(s) g_1(s) ds. \tag{11}$$

In fact, by construction, Q_1 is orthogonal to all G_k (for all $k \neq 1$). Next we choose

$$g_1(s) = \begin{cases} 1 & \text{for } s \in [t_1, t_1 + dt] \text{ (t_1 arbitrary in $[0,T]$, except $t_1 = T$)} \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

With this choice we have

$$\kappa_1(s) = \frac{1}{dt}\mathbb{E}\left\{F([W]) dW(t_1;\omega)\right\}, \tag{13}$$

or, equivalently,

$$\kappa_1(t_1) = \mathbb{E}\left\{F([W])\dot{W}(t_1;\omega))\right\} \tag{14}$$

where $\dot{W}(t;\omega)$ is Gaussian white noise (derivative of Wiener process). Other kernels can be determined in a similar way, i.e., by using the orthogonality between $Q_n([W])$ and $G_j([W])$ (for $n \neq j$), and then choosing the kernel of $Q_n([W])$ appropriately (see [19, p. 42]).

Example: Consider the linear functional

$$F([W] = \int_0^T \sin(t)\dot{W}(t;\omega) dt \tag{15}$$

Applying (9) yields

$$G_0 = \int_0^T \sin(t)\mathbb{E}\{\dot{W}(t;\omega)\} dt = 0 \tag{16}$$

On the other hand, using (14) we obtain the kernel

$$\kappa_1(t_1) = \int_0^T \sin(t) \underbrace{\mathbb{E}\left\{\dot{W}(t;\omega)\dot{W}(t_1;\omega))\right\}}_{\delta(t-t_1)} dt = \sin(t_1). \tag{17}$$

Generalized Wiener-Hermite expansion. There were attempts to *generalize* the Wiener-Hermite functional expansion to input processes other than the Brownian motion, e.g., more general independent increment processes [12, 9, 2]. The reason for such generalization is obvious. For instance, such expansions can be used to represent the solution of an ODE driven by random noise other than Brownian motion. However, it was found in [12, 9] that expanding a given functional in terms of series of orthogonal polynomial functionals of processes other than Brownian motion can yield non-convergent expansions. As we shall see hereafter, this is also true in the much simpler case of systems driven by a finite number of random variables, or even just one random variable.

Generalized polynomial chaos expansion for systems driven by one random variable.

While theoretically sound, the Wiener-Hermite expansion in terms of orthogonal polynomial functionals does not have a great deal of practical applicability, mostly because it is an expansion relative to an infinite-dimensional stochastic process, i.e., the Brownian motion process. However, the theory can be simplified substantially for systems driven by a finite number of random variables [20, 2]. The simplest case is a system driven by only one random variable, i.e., a mapping of the form

$$\eta(\omega) = g(\xi(\omega)). \tag{18}$$

The *generalized polynomial chaos (gPC)* expansion of $\eta(\omega)$ is a series expansion of $g(\xi(\omega))$ in terms of polynomials of $\xi(\omega)$ orthogonal with respect to the PDF of $\xi(\omega)$. Let us write such gPC expansion as

$$g(\xi(\omega)) = \sum_{k=0}^{\infty} a_k P_k(\xi(\omega)), \tag{19}$$

where a_k are real numbers (they play the same role as the kernels κ_n in (4)-(5)), and $P_k(\xi(\omega))$ are polynomials of the random variable $\xi(\omega)$ satisfying the orthogonality conditions

$$\mathbb{E}\left\{P_k(\xi)P_j(\xi)\right\} = \int_{-\infty}^{\infty} P_k(x)P_j(x)p_\xi(x)dx = \mathbb{E}\left\{P_k^2(\xi)\right\}\delta_{kj}. \tag{20}$$

A substitution of (20) into (19) yields

$$a_k = \frac{\mathbb{E}\left\{P_k(\xi)g(\xi)\right\}}{\mathbb{E}\left\{P_k^2(\xi)\right\}}. \tag{21}$$

Remark: The theory of orthogonal polynomials is summarized in [20, Ch. 3] and in Appendix A of this note. One of the key elements is that there exists a one-to-one correspondence between the PDF $p_\xi(x)$ and a set of (monic) orthogonal polynomials. In other words, the function $p_\xi(x)$ defines uniquely a set of orthogonal polynomials, e.g., through the Stieltjes algorithm [4, 3] (see Appendix A).

In Table 1 we summarize the generalized polynomial chaos corresponding to continuous random variables $\xi(\omega)$ with known probability distribution.

Theorem 1 (Convergence of gPC expansion)**.** The set of orthogonal polynomials associated with the random variable $\xi(\omega)$ is dense in $L^2(\Omega, \mathcal{F}, P)$ if and only if the moment problem for $\xi(\omega)$ is uniquely solvable[3].

The proof of is given in [2]. Stated differently, Theorem 1 says that the sequence of random variables

$$g_n(\xi) = \sum_{k=0}^{n} a_k P_k(\xi) \qquad a_k = \frac{\mathbb{E}\left\{P_k(\xi)g(\xi)\right\}}{\mathbb{E}\left\{P_k^2(\xi)\right\}}, \tag{22}$$

where $P_k(\xi)$ are orthogonal polynomials relative to the PDF of ξ, converges to the random variable $\eta(\omega) = g(\xi(\omega))$ in $L^2(\Omega, \mathcal{F}, P)$, i.e., in the mean square sense (see Appendix B), if and only if the moment problem for $\xi(\omega)$ is uniquely solvable. In other words,

$$\lim_{n\to\infty} \mathbb{E}\left\{|g(\xi) - g_n(\xi)|^2\right\} = 0. \tag{23}$$

[3]The "moment problem" here refers to the question of whether the PDF of $\xi(\omega)$ can be uniquely identified by the sequence of its moments.

PDF of $\xi(\omega)$	gPC	support
Gaussian	Hermite	$(-\infty, \infty)$
Uniform	Legendre	$[-1, 1]$
Gamma	Laguerre	$[0, \infty)$
Arbitrary PDF	Stieltjes algorithm	$[a, b]$

Table 1: Correspondence between the PDF of the continuous random variable $\xi(\omega)$ and the gPC basis.

In Appendix B we show that mean square convergence implies convergence in probability and therefore convergence in distribution. This means that the if the random variables are continuous then the PDF of $g_n(\xi)$ converges to the PDF of $g(\xi)$ pointwise.

An important question at this point is: under which conditions is the moment problem for a random variable uniquely solvable?

Theorem 2 (Uniqueness of the solution to the moment problem)**.** The moment problem for the distribution function of a random variable $\xi(\omega)$ is uniquely solvable if one of the following conditions is satisfied:

1. The PDF $p_\xi(x)$ is compactly supported;
2. The moment generating function $m(a) = \mathbb{E}\{e^{a\xi(\omega)}\}$ exists and it is finite in a neighborhood of $a = 0$;
3. $\xi(\omega)$ is exponentially integrable, i.e.,
$$\mathbb{E}\{e^{a|\xi(\omega)|}\} < \infty \quad \text{for some } a > 0; \tag{24}$$
4. The sequence of moments $m_n = \mathbb{E}\{\xi^n\}$ satisfies
$$\sum_{n=0}^{\infty} \left(\frac{1}{m_{2n}}\right)^{\frac{1}{2n}} = \infty. \tag{25}$$

The proof is given in [2].

Example: The moment problem is uniquely solvable for Gaussian PDFs, uniform PDFs, and gamma PDFs
$$p_\xi(x) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-x/\theta}, \qquad x > 0, \qquad k, \theta > 0. \tag{26}$$

Example: A log-normal random variable is defined as
$$\xi(\omega) = \log(X(\omega)), \tag{27}$$
where $X(\omega)$ is normal. It is straightforward to show that
$$p_\xi(x) = \frac{1}{x\sqrt{2\pi}} e^{-\log(x)^2/2} \qquad x > 0. \tag{28}$$
The moments of ξ are
$$\mathbb{E}\{\xi^n\} = e^{n^2/2}, \tag{29}$$

and clearly exist for all $n \geq 1$. However, the moment problem is not uniquely solvable in this case. Indeed, there are multiple PDFs with exactly the same sequence of moments. For example, for any $v \in (0,1)$ and any $k > 0$ the PDF

$$p_\eta(x) = \frac{1}{x\sqrt{2\pi}} e^{-\log(x)^2/2} \left[1 + v \sin(2k\pi \log(x)\right] \qquad x > 0 \tag{30}$$

has exactly the same moments as (28) (see [2, §4.1]). In other words,

$$\int_0^\infty x^n p_\xi(x) dx = \int_0^\infty x^n p_\eta(x) dx \quad \text{for all } n \geq 1. \tag{31}$$

Note that condition 4 in Theorem 2 does not hold for lognormal variables. Indeed, for lognormal variables we have $m_n = e^{n^2/2}$ (see Eq. (29)) and therefore

$$\sum_{n=0}^\infty \left(\frac{1}{e^{2n^2}} \right)^{\frac{1}{2n}} = \sum_{n=0}^\infty \frac{1}{e^n} = \frac{e}{e-1} \neq \infty \tag{32}$$

gPC expansion for systems driven by multiple random variables.

Consider the random variable $\eta(\omega)$ defined as a scalar function of M *independent* random variables $\{\xi(\omega), \ldots, \xi_M(\omega)\}$

$$\eta = g(\xi_1, \ldots, \xi_M) \tag{33}$$

We have seen that the PDF η can be represented as a multidimensional convolution of the PDFs of $\{\xi_j\}$. Denote by $\{P_{j_i}^{(i)}(\xi_i)\}$ the gPC expansion associated with the random variable $\xi_i(\omega)$. By leveraging the separability of $L^2_{p(\boldsymbol{\xi})}$ following from the independence assumption on $\{\xi_n(\omega)\}$, we have the following multivariate gPC expansion

$$g(\xi_1, \ldots, \xi_M) = \sum_{j_1=0}^\infty \cdots \sum_{j_M=0}^\infty a_{j_1 \ldots j_M} P_{j_1}^{(1)}(\xi_1) \cdots P_{j_M}^{(M)}(\xi_M), \tag{34}$$

where

$$a_{j_1 \ldots j_M} = \frac{\mathbb{E}\{g(\xi_1, \ldots, \xi_M) P_{j_1}^{(1)}(\xi_1) \cdots P_{j_M}^{(M)}(\xi_M)\}}{\mathbb{E}\left\{P_{j_1}^{(1)}(\xi_1)^2\right\} \cdots \mathbb{E}\left\{P_{j_M}^{(M)}(\xi_M)^2\right\}} \tag{35}$$

If the moment problem for each random variable $\xi_i(\omega)$ is uniquely solvable, then the *tensor product* gPC expansion (34)-(35) converges in the mean square sense (see [2]), i.e., in the $L^2(\Omega, \mathcal{F}, P)$ sense

$$\lim_{n_1 \to \infty} \cdots \lim_{n_M \to \infty} \mathbb{E}\left\{ g(\xi_1, \ldots, \xi_M) - \sum_{j_1}^{n_1} \cdots \sum_{j_M}^{n_M} a_{j_1 \ldots j_M} P_{j_1}^{(1)}(\xi_1) \cdots P_{j_M}^{(M)}(\xi_M) \right\} = 0 \tag{36}$$

It convenient to write the expansion (34) more compactly. Upon definition of $\boldsymbol{\xi} = (\xi_1, \ldots, xi_M)$ we have

$$g(\boldsymbol{\xi}) = \sum_{k=0}^\infty a_k \Phi_k(\boldsymbol{\xi}), \tag{37}$$

where $\Phi_k(\boldsymbol{\xi})$ are multivariate polynomials constructed by taking products of one-dimensional polynomials $P_{j_i}^i(\xi_i)$. A convenient way to arrange the polynomials $\Phi_k(\boldsymbol{\xi})$ is to sort the tensor product in a degree lexicographic order. In Table 2) we summarize such ordering for the three-dimensional polynomial chaos

$$\Phi_k(\boldsymbol{\xi}) = P_{j_1}^{(1)}(\xi_1) P_{j_2}^{(2)}(\xi_2) P_{j_3}^{(3)}(\xi_3), \tag{38}$$

j_1	j_2	j_3	k	Total degree	gPC basis
0	0	0	0	0	$\Phi_0(\boldsymbol{\xi}) = P_0^{(1)}(\xi_1)P_0^{(2)}(\xi_2)P_0^{(3)}(\xi_3)$
0	0	1	1	1	$\Phi_1(\boldsymbol{\xi}) = P_0^{(1)}(\xi_1)P_0^{(2)}(\xi_2)P_1^{(3)}(\xi_3)$
0	1	0	2	1	$\Phi_2(\boldsymbol{\xi}) = P_0^{(1)}(\xi_1)P_1^{(2)}(\xi_2)P_0^{(3)}(\xi_3)$
1	0	0	3	1	$\Phi_3(\boldsymbol{\xi}) = P_1^{(1)}(\xi_1)P_0^{(2)}(\xi_2)P_0^{(3)}(\xi_3)$
0	0	2	4	2	$\Phi_4(\boldsymbol{\xi}) = P_0^{(1)}(\xi_1)P_0^{(2)}(\xi_2)P_2^{(3)}(\xi_3)$
0	1	1	5	2	$\Phi_5(\boldsymbol{\xi}) = P_0^{(1)}(\xi_1)P_1^{(2)}(\xi_2)P_1^{(3)}(\xi_3)$
0	2	0	6	2	$\Phi_6(\boldsymbol{\xi}) = P_0^{(1)}(\xi_1)P_2^{(2)}(\xi_2)P_0^{(3)}(\xi_3)$
1	0	1	7	2	$\Phi_7(\boldsymbol{\xi}) = P_1^{(1)}(\xi_1)P_0^{(2)}(\xi_2)P_1^{(3)}(\xi_3)$
1	1	0	8	2	$\Phi_8(\boldsymbol{\xi}) = P_1^{(1)}(\xi_1)P_1^{(2)}(\xi_2)P_0^{(3)}(\xi_3)$
2	0	0	9	2	$\Phi_9(\boldsymbol{\xi}) = P_2^{(1)}(\xi_1)P_0^{(2)}(\xi_2)P_0^{(3)}(\xi_3)$

Table 2: Degree lexicographic order of the multivariate polynomial chaos $\Phi_k(\boldsymbol{\xi}) = P_{j_1}^{(1)}(\xi_1)P_{j_2}^{(2)}(\xi_2)P_{j_3}^{(3)}(\xi_3)$. Shown are polynomials up to total degree 2.

		p					
		1	2	3	4	5	6
M	1	2	3	4	5	6	7
	2	3	6	10	15	21	28
	3	4	10	20	35	56	84
	4	5	15	35	70	126	210
	5	6	21	56	126	252	462
	6	7	28	84	210	462	924

Table 3: Dimensionality of multivariate gPC for different values of p (max polynomial degree in each 1D gPC expansion) and M (number of random variables).

It is clear that the number of terms grows with the dimension M and maximum polynomial degree in each variable quite fast (exponentially fast as a matter of fact). For instance, gPC of degree $p = 2$ in $M = 3$ random variables yields 10 basis elements $\{\Phi_0, \ldots, \Phi_9\}$ (see Table 2 and Table 3). The combinatorial nature of the tensor product basis allows us to calculate the number of terms for each fixed M and polynomial degree p exactly. If we denote by $K+1$ the total number terms, i.e., a truncation of the (37) to K terms, then we have

$$K + 1 = \binom{p+M}{p} = \frac{(M+p)!}{M!p!}. \tag{39}$$

Note that $K+1$ chosen in this way allows a full development of total degree terms up to (and including) p in the gPC basis $\{\Phi_0, \ldots, \Phi_K\}$. In Table 3 we summarize the dimensionality of gPC, i.e., the number of terms $K+1$ for different p (max polynomial degree in each 1D polynomial expansion) and M (number of random variables).

Statistical properties. Once the gPC series expansion of a mapping between random variables is available it is rather straightforward to compute statistical properties such as moments, cumulants, and even the PDF of η (using sampling). To this end, let

$$\eta = g(\boldsymbol{\xi}) \simeq \sum_{j=0}^{K} a_j \Phi_j(\boldsymbol{\xi}), \qquad a_j = \frac{\mathbb{E}\left\{\eta \Phi_j(\boldsymbol{\xi})\right\}}{\mathbb{E}\{\Phi_j^2\}}, \tag{40}$$

be the gPC expansion of $\eta = g(\boldsymbol{\xi})$. It is straightforward to show that,

$$\mathbb{E}\left\{g(\boldsymbol{\xi})\right\} = a_0, \tag{41}$$

$$\mathbb{E}\left\{g(\boldsymbol{\xi})^2\right\} = \sum_{k=0}^{K} a_k^2 \mathbb{E}\{\Phi_k^2\}. \tag{42}$$

In fact, by construction, all 1D orthogonal polynomials of degree larger or equal to one defining $\Phi_k(\boldsymbol{\xi})$ average to zero as a consequence of orthogonality[4]. Also, the constant polynomial in each 1D expansion is, by construction always equal to one and therefore $\Phi_0(\boldsymbol{\xi}) = 1$, which implies $\mathbb{E}\{\Phi_0\} = 1$. Equations (41)-(42) allow us to express the variance of $g(\boldsymbol{\xi})$ (second cumulant) as

$$\text{var}\{g(\boldsymbol{\xi})\} = \sum_{k=1}^{K} a_k^2 \mathbb{E}\{\Phi_k^2\}. \tag{44}$$

Regarding the PDF of η, we recall that the gPC expansion (40) converges in the mean square sense, and therefore in distribution (see Appendix B). This means that if we sample each random variable ξ_i according to its PDF and substitute such samples into the the gPC expansion then we obtain samples of $\boldsymbol{\eta}$.

Multi-element generalized polynomial chaos (ME-gPC) expansion. The ME-gPC expansion was originally developed in [17, 16] to address the loss of accuracy of gPC simulations of certain time-dependent problems. One of the reasons that leads to a loss accuracy in gPC simulations is related to the complexity of the mapping being approximated by gPC, which eventually requires more and more terms as time evolves. As a simple example consider an harmonic oscillator with random frequency $\xi(\omega)$, uniformly distributed in $[0, 1]$,

$$\ddot{x} + \xi^2(\omega)x = 0, \qquad \dot{x}(0) = 1, \qquad x(0) = 0. \tag{45}$$

As is well known, the solution to (45) is

$$x(t, \xi) = \sin(\xi(\omega)t). \tag{46}$$

It is clear that the gPC representation of the solution (46) requires Legendre polynomials with an increasing degree as t increases. The reason is clearly explained in Figure 1, where we see that as t increases the function $\xi \to \sin(\xi t)$ has more and more zeroes in $[0, 1]$. Another example in which gPC fails miserably is the approximation of the solution to the simple decay problem

$$\dot{x} = -\xi(\omega)x, \qquad x(0) = 1, \qquad \xi \sim U([0, 1]), \tag{47}$$

i.e.,

$$x(t; \omega) = e^{-\xi t}. \tag{48}$$

The basic idea of ME-gPC is to partition the support of the joint PDF of the random input variables, i.e., the range of the random input variables, into non-overlapping elements and construct a *local gPC* series expansion corresponding to each element. To describe ME-gPC we consider, for simplicity, only one random input variable $\xi(\omega)$, continuous and with bounded range $[a, b]$.

First, we partition the range of ξ into two non-overlapping elements (see Figure 2)

$$E_1 = \{x \in \mathbb{R} : a \leq x \leq c\}, \qquad E_2 = \{x \in \mathbb{R} : c \leq x \leq b\}. \tag{49}$$

[4]In fact, since $P_0^{(j)}(\xi_j)$ are constants we have

$$\mathbb{E}\{P_0^{(j)}(\xi_j)P_q^{(j)}(\xi_j)\} = 0 \quad \text{for all} \quad q \neq 0 \quad \Rightarrow \quad P_0^{(j)}(\xi_j)\mathbb{E}\{P_q^{(j)}(\xi_j)\} = 0 \quad \Rightarrow \quad \mathbb{E}\{P_q^{(j)}(\xi_j)\} = 0 \quad \text{for all} \quad q \neq 0. \tag{43}$$

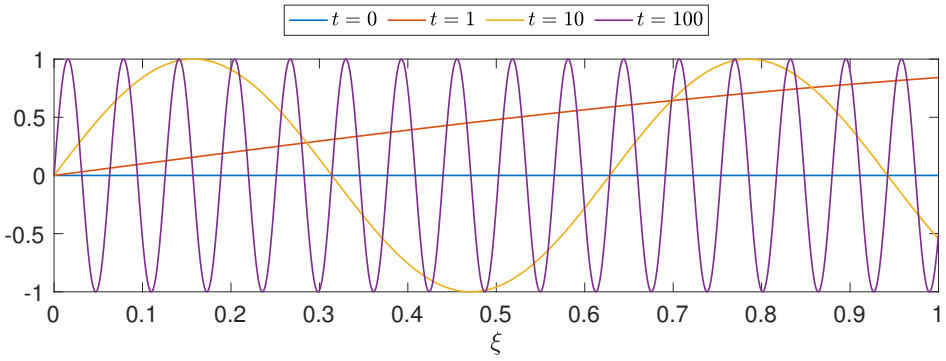

Figure 1: Random frequency problem. A gPC expansion of the time-dependent function $\sin(\xi t)$ requires polynomials of higher and higher degrees as t increases. To see this, simply note that $\sin(\xi t)$ has 4 zeros in ξ at $t = 10$ and 32 zeros in ξ at $t = 100$. Therefore, at $t = 4$ we need a gPC expansion of degree of at least 4 while at $t = 100$ we a gPC expansion of degree at least of 32. Such estimates are of course a lower bound for the gPC degree that actually guarantees a specified accuracy.

Then we define two indicator functions

$$I_{E_i} = \begin{cases} 1 & \text{if } \xi(\omega) \in E_i \\ 0 & \text{otherwise} \end{cases} \qquad i = 1, 2. \tag{50}$$

Clearly, $A_i = I_{E_i}^{-1}(1) \subset \Omega$ (pre-image of 1 under I_{E_i}) represents the subset of the sample space Ω such that $\xi(\omega) \in E_i$, i.e.,

$$A_1 = \{\omega \in \Omega : \xi(\omega) \in E_1\} \quad \text{and} \quad A_2 = \{\omega \in \Omega : \xi(\omega) \in E_2\}. \tag{51}$$

Clearly A_1 and A_2 are non-intersecting subsets of Ω such that

$$\Omega = A_1 \cup A_2. \tag{52}$$

At this point, consider the input-output map

$$\eta(\omega) = g(\xi(\omega)). \tag{53}$$

We know that the statistical properties of η are fully described by the distribution function

$$F_\eta(y) = P(\underbrace{\{\omega \in \Omega : g(\xi(\omega)) \leq y\}}_{\text{set } B_y}). \tag{54}$$

The set B_y can be written as union of two non-intersecting[5] sets

$$\begin{aligned} B_y =& B_y \cap \Omega \\ =& B_y \cap (A_1 \cup A_2) \\ =& (B_y \cap A_1) \cup (B_y \cap A_2). \end{aligned} \tag{55}$$

Since $(B_y \cap A_1)$ and $(B_y \cap A_2)$ are disjoint we have

$$P(B_y) = P(B_y \cap A_1) + P(B_y \cap A_2). \tag{56}$$

[5]To be more precise A_1 and A_2 do intersect, but the intersection set has zero measure.

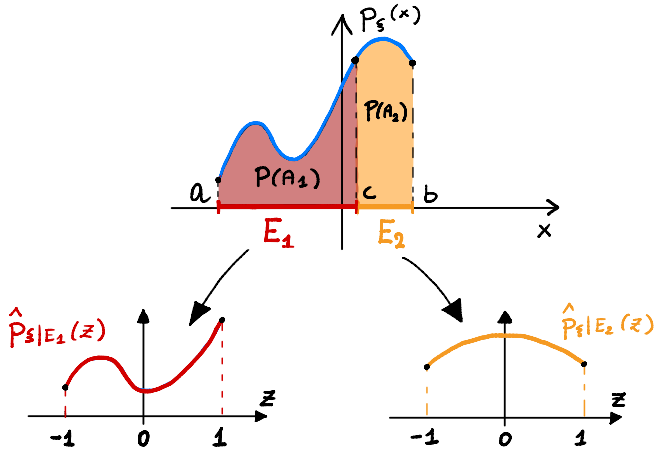

Figure 2: Basic idea of Multi-Element generalized Polynomial Chaos (ME-gPC). The range of the random input variable $\xi(\omega)$, i.e., the support of the PDF $p_\xi(x)$ is partitioned into non-overlapping elements, say E_1 and E_2. A local gPC expansion is then constructed relative to the conditional PDF of $\xi(\omega)$ in E_1 and E_2. Such conditional PDF is obtained by simply rescaling the PDF $p_\xi(x)$ restricted to each element E_1 and E_2, and eventually remapping it to the standard element $[-1, 1]$. The latter step allows standardization of the Stieltjes algorithm to construct the set of orthogonal polynomials corresponding to the PDFs $\hat{p}_{\xi|E_i}(z)$.

In terms of conditional probabilities this can be written as

$$P(B_y) = P(B_y|A_1)P(A_1) + P(B_y|A_2)P(A_2). \tag{57}$$

Recall that $P(A_1)$ represents the probability that $\xi(\omega)$ is in the element E_1, while $P(A_2)$ represents the probability that $\xi(\omega)$ is in the element E_2. Such probabilities can be expressed in terms of the PDF of ξ as (See Figure 2)

$$P(A_1) = \int_{E_1} p_\xi(x)dx \qquad P(A_2) = \int_{E_2} p_\xi(x)dx. \tag{58}$$

By combining (54), (57) and (58) we finally obtain

$$F_\eta(y) = F_{\eta|\xi\in E_1}(y) \int_{E_1} p_\xi(x)dx + F_{\eta|\xi\in E_2}(y) \int_{E_2} p_\xi(x)dx. \tag{59}$$

By differentiating this expression with respect to y we obtain the corresponding expression for the PDF of η

$$p_\eta(y) = p_{\eta|\xi\in E_1}(y) \int_{E_1} p_\xi(x)dx + p_{\eta|\xi\in E_2}(y) \int_{E_2} p_\xi(x)dx. \tag{60}$$

Based on this formula, we see that the PDF of the output η is represented as a weighted mean of two conditional PDFs, i.e.,

$$p_{\eta|\xi\in E_1}(y) \quad \text{and} \quad p_{\eta|\xi\in E_2}(y). \tag{61}$$

Such conditional PDFs represent the *response* of the system to two conditionally independent random variables $\xi|E_1$ and $\xi|E_2$ with PDF that coincide with the conditionals $p_\xi(x|\xi \in E_1)$ and $p_\xi(x|\xi \in E_2)$ (suitably normalized). Hence, if we compute two different gPC expansions of the response $\eta = g(\xi)$ corresponding to the conditionally independent random variables to $\xi|E_1$ and $\xi|E_2$ and combine the results as in (60) then we can compute any statistical properties of η, including the PDF of η

ME-gPC algorithm:

1. Partition the range of $\xi(\omega)$, i.e., the support of $p_\xi(x)$ as in Figure 2, i.e., as a covering of non-overlapping elements.
2. Determine the conditionals $p_\xi(x|\xi \in E_i)$ and map them onto $[-1,1]$ using the transformation (104). The mapping is not strictly necessary but it enhances stability especially if the measure of E_i is very small.
3. Generate a gPC expansion relative to each mapped PDF with the Stieltjes algorihtm (Appendix B).
4. Compute the polynomial chaos coefficients relative to each local gPC expansion.

This allows us to compute an element-by-element representation of the response. For example, to generate samples of $p_\eta(y)$ we can generate independent samples of $p_{\eta|\xi\in E_i}(y)$ using the gPC expansion in terms of the random variable $\xi|E_i$ with PDF $p_\xi(x|\xi \in E_1)$.

Remark: Another approach to address the random frequency problem associated with the solution of certain random dynamical systems was proposed in [8]. The key idea is to approximate with gPC the mapping that pushes forward the solution of the random ODE in time rather than the solution itself. With such flow map approximation and composition it is demonstrated that gPC retains accuracy as t increases. At the same time the gPC polynomial degree naturally increases in the scheme, which makes it essentially inapplicable for system driven by multiple random variables.

The stochastic Galerkin method

The stochastic Galerkin method is a projection operator method to solve a wide variety of UQ problems ranging from random eigenvalue problems, to system of ordinary or partial differential equations evolving from random initial states, with random boundary conditions, random parameters, or random forcing terms. The basic idea it to represent the solution of the UQ problem in a polynomial chaos expansion with unknown coefficients, substitute the expansion into the equations defining the problem, and the project (in the sense of $L^2(\Omega,\mathcal{F},P)$) the resulting equation onto the gPC basis to obtain a system of *deterministic* equations for the gPC coefficients. The number of such equations depends on the number of random input and the polynomial chaos order as summarized in Table 3.

Decay problem (linear ODE). Consider the simple linear ODE

$$\frac{dx}{dt} = -\xi(\omega)^2 x, \qquad x(0;\omega) = 1, \tag{62}$$

where $\xi(\omega)$ is a uniform random variable in $[-1,1]$, and the initial condition is deterministic. We expand the solution in a Legendre polynomial chaos expansion (see Table 1)

$$x(t,\omega) = \sum_{k=0}^{n} a_k(t) L_k(\xi(\omega)), \tag{63}$$

where $L_k(\xi)$ are Legendre polynomials[6] of the uniform random variable ξ. Note that here the polynomial chaos modes are function of time and defined through projection

$$a_k(t) = \frac{\mathbb{E}\{x(t;\omega)L_k(\xi)\}}{\mathbb{E}\{L_k^2(\xi)\}}. \tag{64}$$

[6]Legendre polynomials are defined are orthogonal with respect to the recursively in Eq. (100).

A substitution of (63) into (62) yields

$$\sum_{k=0}^{n}\frac{da_k(t)}{dt}L_k(\xi)=-\xi(\omega)^2\sum_{k=0}^{n}a_k(t)L_k(\xi)+R_n(t;\xi),\qquad \sum_{k=0}^{n}a_k(0)L_k(\xi)=x_0, \tag{65}$$

where $R_n(t,\xi)$ is the residual arising from the fact that (63) does not satisfy the ODE (62) exactly. In the stochastic Galerkin method we impose that the residual is orthogonal to the gPC space $B_n = \text{span}\{L_0,\ldots,L_n\}$ relative to the $L^2(\Omega,\mathcal{F},P)$ inner product[7]. In practice, we multiply (65) by $L_j(\xi)$ and then integrate relative to the PDF of $\xi(\omega)$, i.e., take the expectation, to obtain

$$\begin{cases}\displaystyle\sum_{k=0}^{n}\frac{da_k(t)}{dt}\mathbb{E}\{L_kL_j\}=-\sum_{k=0}^{n}a_k(t)\mathbb{E}\{\xi^2L_kL_j\}\\ \displaystyle\sum_{k=0}^{n}a_k(0)\mathbb{E}\{L_kL_j\}=x_0\end{cases} \tag{66}$$

Using the orthogonality of $\{L_k\}$ relative to the uniform PDF of ξ and recalling that

$$L_0(\xi)=1\qquad L_1(\xi)=\xi\qquad L_2(\xi)=\frac{3}{2}\xi^2-\frac{1}{2}, \tag{67}$$

i.e.,

$$\xi^2=\frac{2L_2(\xi)+1}{3}. \tag{68}$$

we can write (66) as

$$\begin{cases}\displaystyle\frac{da_j(t)}{dt}=-\frac{a_j(t)}{3}-\frac{2}{3\mathbb{E}\{L_j^2\}}\sum_{k=0}^{n}\mathbb{E}\{L_2L_kL_j\}a_k(t)\qquad j=0,\ldots,n\\ a_0(0)=1\\ a_j(0)=0\qquad j=1,\ldots,n\end{cases} \tag{69}$$

This is a system of $n+1$ linear ODEs that can solved numerically with any discretization scheme. Once the gPC modes $\{a_0(t),\ldots,a_n(t)\}$ are available, we can substitute them back into the gPC expansion of the solution (63), and compute statistical properties such as the mean,

$$\mathbb{E}\{x(t;\omega)\}=a_0(t), \tag{70}$$

the variance

$$\text{var}\{x(t;\omega)\}=\sum_{k=1}^{n}a_k^2(t)\mathbb{E}\{L_k^2\}, \tag{71}$$

or the PDF of $x(t;\omega)$ by sampling or transforming the polynomial chaos expansion (63).

Remark: Recall that the PDF of $x(t;\omega)$ can be also computed by solving the Liouville equation for the joint PDF of $x(t;\omega)$ and ξ and then marginalizing out ξ. Alternatively, we can try to solve the BBGKY equation for the PDF of $x(t;\omega)$ alone, e.g., by computing a data-driven closure for the conditional expectations appearing in the reduced-order PDF equation.

Remark: What happens if $\xi(\omega)$ is a uniform random variable in $[a,b]$ instead of $[-1,1]$? Not much of a difference. We simply need to generate a gPC expansion for a uniform random variable defined in $[a,b]$.

[7] In numerical methods for deterministic PDEs this procedure is also known as Galerkin projection method [6, **?**].

How do we do that? We first change the coordinate system and map the support $[a,b]$ to $[-1,1]$. In such new coordinates we generate the orthogonal polynomial basis, which is made of Legendre polynomials. Once the polynomials are available in $[-1,1]$ we map them back then we map them back to $[a,b]$. As easily seen, these are still orthogonal polynomials. What changes is simply that there is a scaling factor $(b-a)/2$ appearing when computing $\mathbb{E}\{L_kL_j\}$.

The coefficients

$$\mathbb{E}\{L_iL_jL_k\} = \int_{-1}^{1} L_i(x)L_j(x)L_k(x)dx \tag{72}$$

appearing in the the *gPC propagator* (66) can be pre-computed offline using Gauss quadrature, or can be computed analytically using the so-called *linearization formulas* [15, Appendix] for orthogonal polynomials. Such formulas basically express the product of two orthogonal polynomials in terms of polynomials that belong to same family as

$$L_k(\xi)L_j(\xi) = \sum_{m=0}^{k+j} \beta_m(k,j)L_m(\xi). \tag{73}$$

Indeed a substitution of (73) (with β_m known) into (72) yields

$$\mathbb{E}\{L_iL_jL_k\} = \sum_{m=0}^{k+j} \beta_m(k,j)\mathbb{E}\{L_mL_i\} = \beta_i(k,j)\mathbb{E}\{L_i^2\}. \tag{74}$$

Heat equation with random boundary condition. Consider the following initial/boundary value problem

$$\begin{cases} \dfrac{\partial u}{\partial t} = \dfrac{\partial^2 u}{\partial x^2} \qquad x\in[0,L] \\ u(x,0) = 0 \\ u(0,t) = u_0(t) \\ u(L,t) = A + \sigma\xi(\omega)\sin(t) \end{cases} \tag{75}$$

where A, σ are a positive constant, and $\xi(\omega)$ is a random variable with known distribution supported in $[-1,1]$. To solve this problem we first compute the gPC expansion corresponding to the PDF of ξ. To this end, we can use the Stieltjes algorithm summarized in Appendix B. Such algorithm produces a set of polynomials $\{P_0,P_1,\ldots\}$ orthogonal in $[-1,1]$ with respect to the PDF of ξ. We expand the solution of (75) relative to the (monic) gPC basis $\{P_0,P_1,\ldots\}$ as

$$u(x,t;\omega) = \sum_{k=0}^{n} a_k(x,t)P_k(\xi). \tag{76}$$

Substituting (76) into (75) and imposing that the residual is orthogonal to $B_n = \text{span}\{P_0,\ldots,P_n\}$ relative to the $L^2(\Omega,\mathcal{F},P)$ yields the gPC propagator

$$\begin{cases} \dfrac{\partial a_k(x,t)}{\partial t} = \dfrac{\partial^2 a_k}{\partial x^2} \qquad k=0,\ldots,K \qquad x\in[0,L] \\ a_k(x,0) = 0 \\ a_0(0,t) = u_0(t) \\ a_k(0,t) = 0 \qquad k=1,\ldots,n \\ a_0(L,t) = A \\ a_1(L,t) = \sigma\sin(t)\mathbb{E}\{P_1^2\} \\ a_k(L,t) = 0 \qquad k=2,\ldots,n \end{cases} \tag{77}$$

This is a system of $n+1$ uncoupled initial/boundary value problems for the polynomial chaos modes $a_k(x,t)$. Note that these modes are functions of space and time in the case of PDEs.

Burgers equation with random initial condition. Consider the following initial/boundary value problem

$$\begin{cases} \dfrac{\partial u}{\partial t} + u\dfrac{\partial u}{\partial x} = \dfrac{\partial^2 u}{\partial x^2} \qquad x \in [0, 2\pi] \\ u(x,0) = u_0(x) + \sigma \displaystyle\sum_{j=1}^{M} \xi_j(\omega)\psi_j(x) \\ \text{Periodic B.C.} \end{cases} \tag{78}$$

The random initial condition here assumed to be a correlated Gaussian random process represented in terms of a Karhunen-Loéve expansion with M independent Gaussian random variables $\boldsymbol{\xi} = \{x_1, \ldots, \xi_M\}$. To construct a gPC basis, we first build the gPC basis for one Gaussian random variable, which is known to be made of Hermite polynomials (see Table 1), and then build a tensor product basis using the degree lexicographic ordering summarized in Table 2. Once the multivariate gPC basis $\{\Phi_0, \ldots, \Phi_K\}$ is available, we expand the solution of (78) as

$$u(x,t;\omega) = \sum_{k=0}^{K} a_k(x,t)\Phi_k(\boldsymbol{\xi}). \tag{79}$$

A substitution of (79) into (78) and subsequent projection onto the gPC basis $\{\Phi_j\}$ yields

$$\begin{cases} \dfrac{\partial a_k}{\partial t} + \dfrac{1}{\mathbb{E}\{\Phi_k^2\}} \displaystyle\sum_{i,j} \mathbb{E}\{\Phi_k\Phi_i\Phi_j\} a_i(x,t)\frac{\partial a_j(x,t)}{\partial x} = \frac{\partial^2 a_k(x,t)}{\partial x^2} \qquad k = 0, \ldots K \qquad x \in [0, 2\pi] \\ a_0(x,0) = u_0(x) \\ a_1(x,0) = \sigma\psi_M(x) \\ a_2(x,0) = \sigma\psi_{M-1}(x) \\ \vdots \\ a_M(x,0) = \sigma\psi_1(x) \\ a_k(x,0) = 0 \qquad k = M+1, \ldots, K \\ \text{Periodic B.C. for each } a_k(x,t) \end{cases} \tag{80}$$

Note that if the KL expansion of the random initial condition in (78) involves just 6 random variables, and we use a gPC expansion of degree 5 then $K+1 = 462$ (see Table 3). This means that the number of coupled PDEs in the gPC propagator (80) is 462!

Stochastic thermal convection. Consider the system of PDEs system

$$\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u}\cdot\nabla)\boldsymbol{u} = -\nabla p + Pr\nabla^2 u + RaPr\, T\boldsymbol{j} \tag{81}$$

$$\frac{\partial T}{\partial t} + \boldsymbol{u}\cdot\nabla T = \nabla^2 T \tag{82}$$

$$\nabla\cdot\boldsymbol{u} = 0 \tag{83}$$

describing the motion of an incompressible fluid within the square cavity shown in Figure 3. The fluid motion is sustained by buoyancy forces (natural convection) induced by the the temperature difference between the horizontal sides of the cavity. In (81)-(83) $\boldsymbol{u}(\boldsymbol{x},t)$ is the (dimensionless) velocity field, $T(\boldsymbol{x},t)$ is the (dimensionless) temperature field, $\boldsymbol{j}$ is the upward unit vector, $Pr = \nu/\alpha^2$ is the Prandtl number,

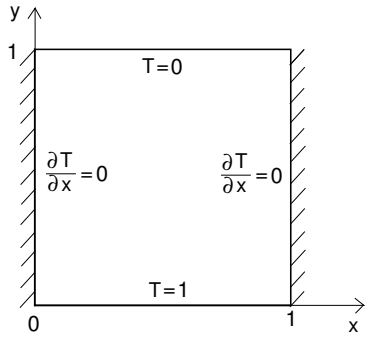

Figure 3: Schematic of the geometry and dimensionless temperature boundary conditions. The velocity boundary conditions are of no-slip type, i.e. $\boldsymbol{u} = 0$ at the walls.

and $Ra = g\beta L^3 \Delta\tau/(\nu\alpha^2)$ is the Rayleigh number. The bifurcation analysis of the PDE system near the onset of convective instability is shown in Figure 4.

Next, we assume that the Rayleigh number in (81) is a uniform random variable centered at $Ra_c = 2585$ (onset of convective instability), i.e.,

$$Ra = Ra_c\,(1 + \sigma\xi)\,, \quad \xi \sim U([-1,1]) \qquad \sigma = 0.05. \tag{84}$$

We are interested in computing the velocity, pressure and temperature fields corresponding to such random Rayleigh number (see [14]). To this end, consider the gPC expansions

$$\boldsymbol{u}\,(\boldsymbol{x},t;\xi) = \sum_{i=0}^{n} \widehat{\boldsymbol{u}}_i\,(\boldsymbol{x},t)\,\Phi_i\,(\xi)\,, \tag{85}$$

$$p\,(\boldsymbol{x},t;\xi) = \sum_{i=0}^{n} \widehat{p}_i\,(\boldsymbol{x},t)\,\Phi_i\,(\xi)\,, \tag{86}$$

$$T\,(\boldsymbol{x},t;\xi) = \sum_{i=0}^{n} \widehat{T}_i\,(\boldsymbol{x},t)\,\Phi_i\,(\xi)\,. \tag{87}$$

where, $\Phi_i\,(\xi)$ are Legendre polynomials of the uniform random variable ξ. A substitution of (85)-(87)) into the system (81)-(83) and subsequent projection onto the basis $\{\Phi_i\}$ yields the gPC propagator

$$\frac{\partial\widehat{\boldsymbol{u}}_k}{\partial t} + \sum_{i,j=0}^{n} \frac{\mathbb{E}\{\Phi_i\Phi_j\Phi_k\}}{\mathbb{E}\{\Phi_k^2\}}\,(\widehat{\boldsymbol{u}}_i\cdot\nabla)\,\widehat{\boldsymbol{u}}_j = -\nabla\widehat{p}_k + Pr\nabla^2\widehat{\boldsymbol{u}}_k + Ra_cPr\left(\widehat{T}_k + \sigma\sum_{i,j=0}^{n}\frac{\mathbb{E}\{\Phi_1\Phi_j\Phi_k\}}{\mathbb{E}\{\Phi_k^2\}}\widehat{T}_j\right)\widehat{\boldsymbol{j}}, \tag{88}$$

$$\frac{\partial\widehat{T}_k}{\partial t} + \sum_{i,j=0}^{n} \frac{\mathbb{E}\{\Phi_i\Phi_j\Phi_k\}}{\mathbb{E}\{\Phi_k^2\}}\widehat{\boldsymbol{u}}_i\cdot\nabla\widehat{T}_j = \nabla^2\widehat{T}_k, \tag{89}$$

$$\nabla\cdot\widehat{\boldsymbol{u}}_k = 0. \tag{90}$$

This is a system in $3(n+1)$ coupled PDEs of the form (81)-(83), where n is the gPC order. In Figure 5 we compare the performance of gPC and ME-gPC in predicting the mean and standard deviation of the velocity and temperature fields at $y = 0.5$.

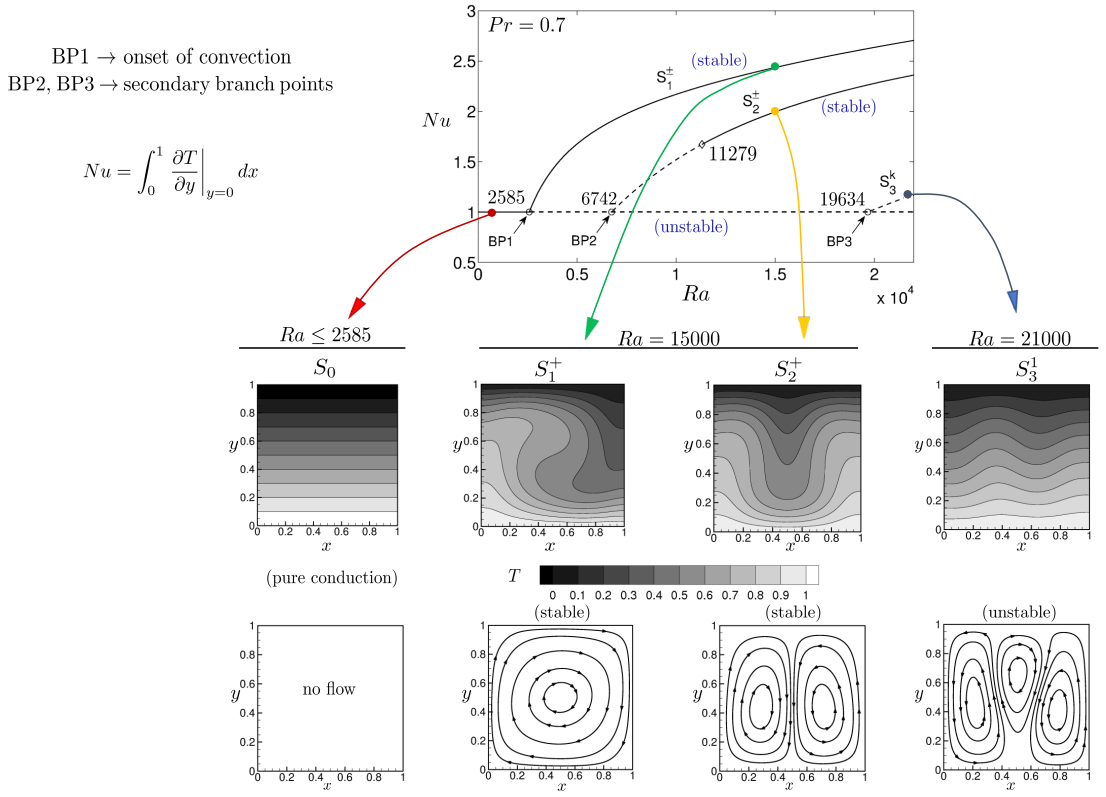

Figure 4: Bifurcation analysis of the cavity flow near the onset of convective instability.

Random eigenvalue problems

Consider a random matrix $\boldsymbol{A}(\xi)$ whose entries depend on the components of a random variable ξ. The case where the entries of $\boldsymbol{A}$ depend on a random vector $\boldsymbol{\xi}$ can be treated similarly. Suppose we are interested in computing the (random) eigenvalues and (random) eigenvectors of $\boldsymbol{A}$, i.e., solve the random eigenvalue problem

$$\boldsymbol{A}(\xi)\boldsymbol{v}_k(\xi) = \lambda_k(\xi)\boldsymbol{v}_k(\xi). \tag{91}$$

The eigenvalues $\lambda_j(\xi)$ are known to be at least continuous functions of coefficients of the characteristic polynmomial, which in turn are functions of the matrix entries. Hence, if $\boldsymbol{A}$ depends continuously on ξ then λ is at least continuous in ξ the entries – see [7] for a thorough characterization of the eigenvalues as a function of perturbation parameters. Let us expand the matrix, the eigenvalues and eigenvectors in (91) in a gPC series expansion as

$$\boldsymbol{A}(\xi) = \sum_{j=0}^{M} \boldsymbol{A}_j P_j(\xi), \tag{92}$$

$$\lambda_k(\xi) = \sum_{j=0}^{M} \lambda_{kj} P_j(\xi), \tag{93}$$

$$\boldsymbol{v}_k(\xi) = \sum_{j=0}^{M} \boldsymbol{v}_{kj} P_j(\xi). \tag{94}$$

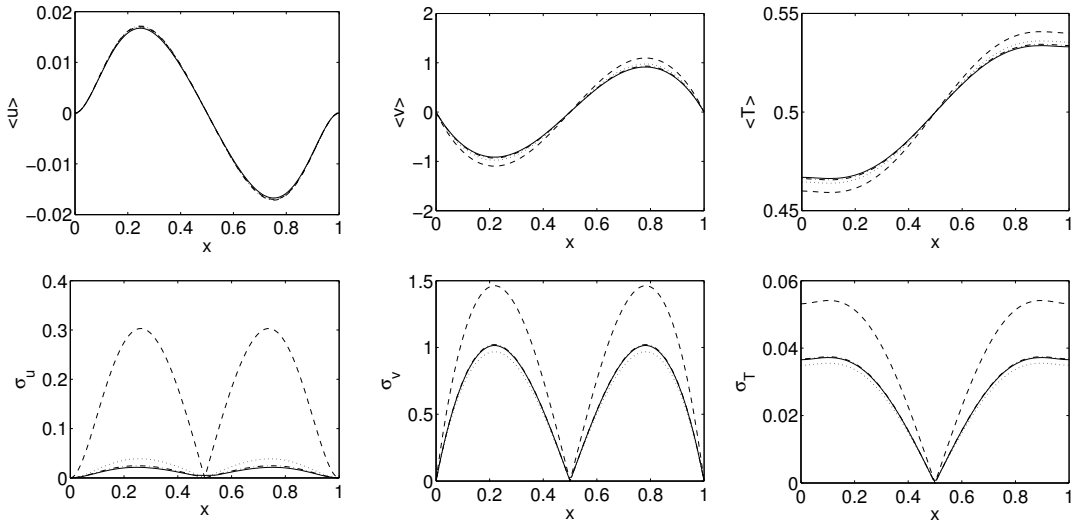

Figure 5: Stochastic convection near the onset. Shown are means (first row) and standard deviations (second row) of velocity and temperature fields along the crossline $y = 0.5$. We plot different results: MC benchmark $(-)$, gPC order 3 $(--)$, ME-gPC 2 elements of order 3 $(\cdots)$, ME-gPC 8 elements of order 3 $(-\cdot)$.

A substitution of (92)-(94) and subsequent projection onto $\{P_q\}$ yields

$$\sum_{j,l=0}^{M} \mathbb{E}\{P_q P_j P_l\} \boldsymbol{A}_l \boldsymbol{v}_{kj} = \sum_{j,l=0}^{M} \lambda_{kj} \boldsymbol{v}_{kq} \mathbb{E}\{P_q P_j P_l\} \tag{95}$$

This system can be written as

$$\boldsymbol{K}\boldsymbol{V}_k = \boldsymbol{\Lambda}_k \boldsymbol{V}_k, \tag{96}$$

where $\boldsymbol{V}_k$ is a $(M+1) \times n$-dimensional vector (vertical stack of $\{\boldsymbol{v}_{k0} \ldots, \boldsymbol{v}_{kM}\}$ and $\boldsymbol{K}$, $\boldsymbol{\Lambda}_k$ are matrices depending on $\boldsymbol{A}_l$, λ_{jk} and $\mathbb{E}\{P_q P_j P_l\}$. For the full definition of such matrices see [5, Eqs (10)-(11)].

With the solution of the (deterministic) eigenvalue problem (96) available, we derive the gPC modes $\boldsymbol{v}_{kj}$ and λ_{kj}, and compute the gPC expansions (93)-(94). This allows us to compute statistical properties of the random eigenvalues and random eigenvectors we may be interested in.

Appendix A: Orthogonal polynomials

A polynomial of degree n can be written as

$$Q_n(x) = b_n x^n + \cdots + b_1 x + b_0, \qquad b_n \neq 0. \tag{97}$$

We denote by $\pi_n(x) = Q_n(x)/b_n$ the *monic* version of $Q_n(x)$, i.e., a polynomial with leading coefficient equal to one. A system of polynomials $\{Q_n(x)\}$ is said to be orthogonal in L^2_μ with respect to a real positive weight function $\mu(x)$ if

$$\int_{\text{supp}(\mu)} Q_n(x) Q_m(x) \mu(x) dx = \delta_{nm}\gamma_n \qquad \text{where} \qquad \gamma_n = \int_{\text{supp}(\mu)} Q_n(x)^2 \mu(x) dx, \tag{98}$$

where δ_{nm} is the Kronecker delta. The weight function $\mu(x)$ defines the set of orthogonal polynomials uniquely. It is well-known that all orthogonal polynomials $\{Q_n(x)\}$ satisfy a *three-term recurrence relation* (see, [4, 6])

$$\begin{cases} Q_{n+1}(x) = (A_n x + B_n) Q_n(x) - C_n Q_{n-1}(x) \\ Q_0(x) = 1 \\ Q_{-1}(x) = 0 \end{cases} \tag{99}$$

where $A_n \neq 0$, $C_n \neq 0$ and $C_n A_n A_{n-1} > 0$ for all n (Favard's theorem [20, p. 26]).

Legendre polynomials: Legendre polynomials are orthogonal in $[-1, 1]$ with respect to the weight function $\mu(x) = 1$, and they satisfy the three-term recurrence relation

$$L_{n+1}(x) = \frac{2n+1}{n+1} x L_n(x) - \frac{n}{n-1} L_{n-1}(x). \tag{100}$$

Hermite polynomials: Hermite polynomials are orthogonal in $(-\infty, \infty)$ with respect to the weight function

$$\mu(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \tag{101}$$

and they satisfy the three-term recurrence relation

$$H_{n+1}(x) = x H_n(x) - n H_{n-1}(x) \tag{102}$$

For monic orthogonal polynomials the three-term recurrence relation simplifies to

$$\begin{cases} \pi_{n+1}(x) = (x - \alpha_n)\pi_n(x) - \beta_n \pi_{n-1}(x) \\ \pi_0(x) = 1 \\ \pi_{-1}(x) = 0 \end{cases} \tag{103}$$

The coefficients α_n and β_n are *uniquely determined* by the weight function $\mu(x)$.

Stieltjes algorithm. The coefficients α_n and β_n in (103), which define monic orthogonal polynomials corresponding to a given weight function, can be computed numerically using a simple algorithm known as *Stieltjes algorithm*. To this end, suppose that $\mu(x) \geq 0$ is continuous and supported[8] on $[-1, 1]$. Define

[8] If the measure $\mu(x)$ is supported on a general interval $[a, b]$ then we can map it to the standard interval $[-1, 1]$ by using the transformation

$$x = \frac{b-a}{2} z + \frac{b+a}{2} \qquad z \in [-1, 1]. \tag{104}$$

the inner product

$$(p,q)=\int_{-1}^{1}p(x)q(x)\mu(x)dx. \tag{105}$$

Multiplying (103) by $\pi_n(x)$ and imposing orthogonality yields

$$\alpha_n=\frac{(x\pi_n,\pi_n)}{(\pi_n,\pi_n)} \qquad n=0,1,2,\dots \tag{106}$$

$$\beta_n=\frac{(\pi_n,\pi_n)}{(\pi_{n-1},\pi_{n-1})} \qquad n=1,2,3,\dots \tag{107}$$

This allows us to derive the following algorithm (known as Stieltjes algorithm) to compute the recurrence coefficients α_k and β_k in (103):

1. Set $n=0$ and $\pi_0(x)=1$ in (106). Compute α_0.
2. With α_0 and $\pi_0(x)=1$ available compute

$$\pi_1(x)=(x-\alpha_0)\pi_0(x)-\beta_0\underbrace{\pi_{-1}(x)}_{=0}=(x-\alpha_0). \tag{108}$$

3. With $\pi_1(x)$ and $\pi_0(x)$ available compute β_1 form (107).
4. Compute α_1 from (106), $\pi_2(x)$ from (103), β_2 from (107), and so on so forth.

In practice, we can compute α_n and β_n to machine precision by replacing the inner product (105) with, e.g., a Gaussian quadrature rule [10]

$$(p,q)\simeq\sum_{j=0}^{M}w_jp(x_j)q(x_j)\mu(x_j), \tag{109}$$

w_j being the Gaussian quadrature weights. This yields a numerical approximation for the coefficients α_n and β_n.

Polynomial approximation theory. Denote by

$$\mathbb{P}_n([a,b])=\text{span}\{1,x,\dots,x^n\} \tag{110}$$

the space of polynomial of degree at most n defined on the interval $[a,b]$. It is well-known that any continuous function $f(x)$ defined on $[a,b]$ can be approximated by a polynomial $p_n(x)\in\mathbb{P}_n([a,b])$ as close as we like, where "close" here means in the uniform (i.e., $L^\infty([a,b])$) norm. This is summarized in the following theorem.

Theorem 3 (Weierstrass)**.** Let $f\in C_0([a,b])$. Then for any $\epsilon>0$ there exists $n_\epsilon\in\mathbb{N}$ and a polynomial $p_{n_\epsilon}(x)\in\mathbb{P}_n([a,b])$ such that

$$\|f-p_{n_\epsilon}\|_{L^\infty([a,b])}=\sup_{x\in[a,b]}|f(x)-p_{n_\epsilon}(x)|\leq\epsilon. \tag{111}$$

This theorem does not provide a constructive way to determine $p_{n_\epsilon}(x)$. It just states the existence of such a polynomial.

However, if we consider the polynomial approximation problem of a function $f(x)$ in the function space $L^2_\mu([a,b])$ (which is a Hilbert space) rather than the Banach space $C_0([a,b])$ then it is rather straightforward to develop a constructive approximation theory, i.e., a systematic way to build the approximating polynomial with estimated on the convergence rate of the approximation. To this end, let $\{Q_0(x), Q_1(x), \ldots, Q_n(x)\}$ be a set of polynomials orthogonal with respect to the inner product

$$(Q_i, Q_j) = \int_a^b Q_i(x) Q_j(x) \mu(x) dx, \tag{112}$$

i.e., $(Q_i, Q_j) = \delta_{ij} \left\| Q_j \right\|^2_{L^2_\mu}$. For each function $f(x) \in L^2_\mu([a,b])$ we define the *orthogonal projection operator* onto the span of $\{Q_0(x), Q_1(x), \ldots, Q_n(x)\}$

$$\mathcal{P}_n : L^2_\mu([a,b]) \to \mathbb{P}_n([a,b]) \tag{113}$$

as

$$\mathcal{P}_n f(x) = \sum_{k=0}^{n} a_k Q_k(x), \qquad a_k = \frac{(f, Q_k)}{(Q_k, Q_k)}. \tag{114}$$

It is straightforward to show that $\mathcal{P}_n f(x)$ is the best polynomial of degree n approximating $f(x)$ in the sense of $L^2_\mu([a,b])$, i.e.,

$$\left\| f - \mathcal{P}_n f \right\|^2_{L^2_\mu} = \inf_{p \in \mathbb{P}_n([a,b])} \left\| f - p \right\|^2_{L^2_\mu}. \tag{115}$$

It can be shown that polynomials are *dense in* $L^2_\mu([a,b])$, meaning that every function $f \in L^2_\mu([a,b])$ can be approximated as a limit of a convergent sequence of polynomials (the limit being in L^2_μ). Since every polynomial of degree n is in the span of $\{Q_0(x), Q_1(x), \ldots, Q_n(x)\}$ this implies that

$$\lim_{n \to \infty} \left\| f - \mathcal{P}_n f \right\|^2_{L^2_\mu} = 0. \tag{116}$$

An important question is how fast $\mathcal{P}_n f$ converges to f. This depends on the *smoothness* of f, and on the specific class orthogonal polynomials. In particular, for Legendre polynomials (100) we have the following approximation result (see [6, p. 109] or [20, p. 33]).

Theorem 4. Let $H^s([-1,1])$ be the Sobolev space of degree s, and $f(x) \in H^s([-1,1])$. Then there exists a constant C, independent of n, such that

$$\left\| f - \mathcal{P}_n f \right\|^2_{L^2([-1,1])} \leq C n^{-s} \left\| f \right\|_{H^s([-1,1]} \tag{117}$$

where $\mathcal{P}_n f$ is the orthogonal projection of f onto the space of Legendre polynomials (Eq. (114)).

This theorem demonstrates that the error, as measured in the $L^2([-1,1])$ norm, decays *spectrally*, i.e., as n^{-s}. Moreover, the rate of decay (the exponent s), is defined by how smooth f is. Indeed, the statement $f \in H^s$ means that f is differentiable s times, and that all derivatives up to the order s are in $L^2([-1,1])$. If f is of class C^∞, i.e., infinitely differentiable in $[-1,1]$ then the convergence rate becomes *exponential*

$$\left\| f - \mathcal{P}_n f \right\|^2_{L^2([-1,1])} \sim e^{-\beta n}. \tag{118}$$

Appendix B: Modes of convergence of sequences of random variables

In this appendix we briefly review the basic modes of convergence of sequences of random variables.

Convergence in distribution. Let $\{X_j(\omega)\}_{j=1,2,\dots}$ be a sequence of random variables defined on the probability space $(\Omega, \mathcal{F}, P)$. We say that the sequence $\{X_j(\omega)\}$ converges to the random variable $X(\omega)$ *in distribution* if for all bounded continuous functions $h : \mathbb{R} \to \mathbb{R}$ we have that

$$\lim_{j\to\infty} \mathbb{E}\left\{h(X_j)\right\} = \mathbb{E}\left\{h(X)\right\}. \tag{119}$$

This equation can be equivalently written as

$$\lim_{j\to\infty} \int_{-\infty}^{\infty} h(x) dF_{X_j}(x) = \int_{-\infty}^{\infty} h(x) dF_X(x) \quad \text{for all bounded continuous functions } h(x), \tag{120}$$

where $F_{X_j}(x)$ and $F_X(x)$ are the distribution functions of $X_j(\omega)$ and $X(\omega)$, respectively. For continuous random variables we know that $F_{X_j}(x)$ and $F_X(x)$ are continuous. In this case, it follows from (120) that $F_{X_j}(x)$ converges to $F_X(x)$ pointwise, i.e.,

$$\sup_x \left|F_{X_j}(x) - F_X(x)\right| \xrightarrow[j\to\infty]{} 0. \tag{121}$$

Moreover, if $F_{X_j}(x)$ and $F_X(x)$ admit PDFs $p_{X_j}(x)$ and $p_X(x)$, i.e.,

$$dF_{X_j}(x) = p_{X_j}(x)dx, \qquad dF_X(x) = p_X(x)dx, \tag{122}$$

then (120) implies that

$$\sup_x \left|p_{X_j}(x) - p_X(x)\right| \xrightarrow[j\to\infty]{} 0, \tag{123}$$

i.e., the PDF of X_j converges to the PDF of X pointwise as we increase j.

Convergence in probability. Let $\{X_j(\omega)\}_{j=1,2,\dots}$ be a sequence of random variables defined on the probability space $(\Omega, \mathcal{F}, P)$. We say that the sequence $\{X_j(\omega)\}$ converges to the random variable $X(\omega)$ *in probability* if for every $\epsilon \geq 0$

$$P\left(\{\omega \in \Omega : |X_j(\omega) - X(\omega)| > \epsilon\}\right) \xrightarrow[j\to\infty]{} 0. \tag{124}$$

Theorem 5. Let $\{X_j(\omega)\}_{j=1,2,\dots}$ be a sequence of random variables defined on the probability space $(\Omega, \mathcal{F}, P)$. If $\{X_j(\omega)\}$ converges to $X(\omega)$ in probability then $\{X_j(\omega)\}$ converges to $X(\omega)$ in distribution.

Proof. We first notice that for every pair of random variables X_j and X, every $a \in \mathbb{R}$ and every $\epsilon \geq 0$ we have (see Figure 6)

$$\{\omega : X_j(\omega) \leq a\} \subseteq \{\omega : X(\omega) \leq a + \epsilon\} \cup \{\omega : |X_j(\omega) - X(\omega)| > \epsilon\}. \tag{125}$$

Since the two set at the right hand side of (125) do intersect, we have[9]

$$\underbrace{P(\{\omega : X_j(\omega) \leq a\})}_{F_{X_j}(a)} \leq \underbrace{P(\{\omega : X(\omega) \leq a + \epsilon\})}_{F_X(a+\epsilon)} + P(\{\omega : |X_j(\omega) - X(\omega)| > \epsilon\}). \tag{127}$$

[9]Recall that for every pair of events A and B in the σ-algebra we have:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \leq P(A) + P(B). \tag{126}$$

Moreover, recall that if $A \subseteq B$ then $P(A) \leq P(B)$.

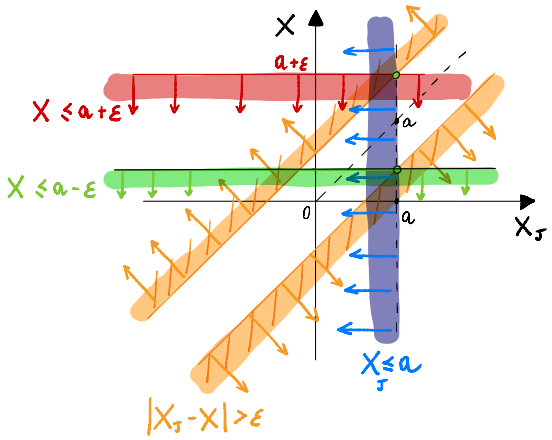

Figure 6: Sketch of the sets in equation (125). Clearly, $\{X_j \leq a\}$ is a subset of $\{X \leq a+\epsilon\} \cup \{|X_j - X| > \epsilon\}$, and $\{X \leq a-\epsilon\}$ is a subset of $\{X_j \leq a\} \cup \{|X_j - X| > \epsilon\}$.

Similarly,

$$\{\omega : X(\omega) \leq a - \epsilon\} \subseteq \{\omega : X_j(\omega) \leq a\} \cup \{\omega : |X_j(\omega) - X(\omega)| > \epsilon\} \tag{128}$$

(see Figure 6), and therefore

$$P(\{\omega : X(\omega) \leq a - \epsilon\}) \leq P(\{\omega : X_j(\omega) \leq a\}) + P(\{\omega : |X_j(\omega) - X(\omega)| > \epsilon\}). \tag{129}$$

Combining (127)-(129) yields

$$F_X(a-\epsilon) - P(\{\omega : |X_j(\omega) - X(\omega)| > \epsilon\}) \leq F_{X_j}(a) \leq F_X(a+\epsilon) + P(\{\omega : |X_j(\omega) - X(\omega)| > \epsilon\}). \tag{130}$$

If $\{X_j(\omega)\}$ converges to $X(\omega)$ in probability then for every $\epsilon \geq 0$

$$\lim_{j\to\infty} P(\{\omega : |X_j(\omega) - X(\omega)| > \epsilon\}) = 0. \tag{131}$$

This implies that in the limit $j \to \infty$

$$F_X(a-\epsilon) \leq F_{X_j}(a) \leq F_X(a+\epsilon). \tag{132}$$

If we send ϵ to zero we obtain (under continuity assumptions for F_X and F_{X_j}) that $F_{X_j}(a)$ converges to $F_X(a)$ for every $a \in \mathbb{R}$, i.e., $\{X_j(\omega)\}$ converges to $X(\omega)$ in distribution.

□

Mean square convergence. Let $\{X_j(\omega)\}_{j=1,2,\ldots}$ be a sequence of random variables defined on the probability space $(\Omega, \mathcal{F}, P)$. We say that the sequence $\{X_j(\omega)\}$ converges to the random variable $X(\omega)$ in the *mean square* sense (or in $L^2(\Omega, \mathcal{F}, P)$) if

$$\lim_{j\to\infty} \mathbb{E}\left\{|X_j(\omega) - X(\omega)|^2\right\} = 0. \tag{133}$$

By using the Markov inequality

$$P(\{\omega : |X_j(\omega) - X(\omega)| > \epsilon\}) \leq \frac{1}{\epsilon^2}\mathbb{E}\left\{|X_j(\omega) - X(\omega)|^2\right\}, \tag{134}$$

we see that if $\{X_j(\omega)\}$ converges to the random variable $X(\omega)$ in L^2 then it converges in probability, and therefore in distribution.

Hence, *mean square convergence implies convergence in distribution.* In other words, X and $\{X_j\}$ have PDFs then (134) implies that the PDF of X_j converges to the PDF of X pointwise (see (123)).

References

[1] R. H. Cameron and W. T. Martin. The orthogonal development of non-linear functionals in series of Fourier-Hermite functionals. *Annals of Mathematics*, 48(2):385–392, 1947.

[2] O. G. Ernst, A. Mugler, H.-J. Starkloff, and E. Ullmann. On the convergence of generalized polynomial chaos expansions. *ESAIM: Math. Model. Numer. Anal.*, 46(2):317–339, 2012.

[3] W. Gautschi. On generating orthognal polynomials. *SIAM J. Sci. and Stat. Comput.*, 3(3):289–317, 1982.

[4] W. Gautschi. *Orthogonal polynomials: computation and approximation.* Oxford University Press, 2004.

[5] R. Ghanem and D. Ghosh. Efficient characterization of the random eigenvalue problem. *Int. J. Numer. Methods in Engineering*, 72:486–504, 2007.

[6] J. S. Hesthaven, S. Gottlieb, and D. Gottlieb. *Spectral methods for time-dependent problems*, volume 21 of *Cambridge Monographs on Applied and Computational Mathematics.* Cambridge University Press, Cambridge, 2007.

[7] T. Kato. *Perturbation theory for linear operators.* Classics in Mathematics. Springer-Verlag, Berlin, 1995. Reprint of the 1980 edition.

[8] D. M. Luchtenburg, S. L. Brunton, and C. W. Rowley. Long-time uncertainty propagation using generalized polynomial chaos and flow map composition. *J. Comput. Phys.*, 274:783–802, 2014.

[9] H. Ogura. Orthogonal functionals of the Poisson process. *IEEE Trans. Inf. Theory*, 4:473–481, 1972.

[10] A. Quarteroni, R. Sacco, and F. Salieri. *Numerical mathematics.* Springer, 2007.

[11] W. J. Rugh. *Nonlinear system theory: the Volterra/Wiener approach.* Johns Hopkins University Press, 1981.

[12] A. Segall and T. Kailath. Orthogonal functionals of independent-increment processes. *IEEE Trans. Inf. Theory*, 22(3):287–298, 1976.

[13] M. Shetzen. *The Volterra and Wiener theories of nonlinear systems.* Wiley, New York, 1980.

[14] D. Venturi, X. Wan, and G. E. Karniadakis. Stochastic bifurcation analysis of rayleigh-bénard convection. *Journal of Fluid Mechanics*, 650:391–413, 2010.

[15] D. Venturi, X. Wan, R. Mikulevicius, B. L Rozovskii, and G. E. Karniadakis. Wick-Malliavin approximation to nonlinear stochastic partial differential equations: analysis and simulations. *Proc. R. Soc. A*, 469(2158):1–20, 2013.

[16] X. Wan and G. E. Karniadakis. An adaptive multi-element generalized polynomial chaos method for stochastic differential equations. *J. Comput. Phys.*, 209(2):617–642, 2005.

[17] X. Wan and G. E. Karniadakis. Multi-element generalized polynomial chaos for arbitrary probability measures. *SIAM J. Sci. Comput.*, 28(3):901–928, 2006.

[18] N. Wiener. The homogeneous chaos. *American Journal of Mathematics*, 60:897–936, 1938.

[19] N. Wiener. *Nonlinear problems in random theory.* MIT Press, 1966.

[20] D. Xiu. *Numerical Methods for Stochastic Computations: A Spectral Method Approach.* Princeton University Press, 2010.