Review of probability theory

Can you predict where a leaf falling from a tree will land? Will there be clouds above Santa Cruz tomorrow at noon? Being scientists, we know that there are physical laws and models we could integrate in time which may provide an answer to such questions. For the falling leaf, we have the equations of fluid mechanics coupled with the equations describing the leaf mechanics (fluid-structure interaction). For the weather forecast in Santa Cruz, we have a quite complicated (usually data-driven) model for dynamics of the atmosphere. However, even if we firmly believe that our equations truthfully represent physical reality, i.e., that there is no *model uncertainty*, we still have some problem when making inferences on the two systems mentioned above. In the case of the leaf falling from the tree, we do not know the exact shape of the leaf, nor the distribution of mass within the leaf, nor whether there is a tiny wind gust pushing the leaf in a direction we did not expect, or having it flipping in a way we did not anticipate. We can of course try to control some of these uncertainties, e.g., by designing a *sterilized experiment* in which we are reasonably sure that there is no wind gust and we know “exactly” the geometry, mechanics, and mass distribution of the leaf. Would the result of such an experiment be useful to make inferences about the behavior of the falling leaf in real world? Perhaps not.

Alternatively, we could study the system for which we have available equations and physical laws using techniques that allow us to account for uncertainties in the initial condition, boundary conditions, forcing terms, geometry,and physical parameters.

The most common approach to study uncertainty propagation, and perhaps the first one that was ever developed, is *random sampling*. In this approach we basically study the response of the system, e.g., the trajectory of the leaf and where it lands, corresponding to randomly sampled realizations of the uncertain parameters and stochastic processes driving the system. Such parameters and processes can be modeled as random variables, random functions or random fields. Computing the solution to such *stochastic models* by sampling involves solving the ODE/PDE system many times, so that a sufficiently large ensemble of solutions is available to compute statistics such as mean, standard deviations, and even probability distribution functions. There are many different types of sampling methods that were developed for this purpose. For instance, Monte Carlo methods and their variants (quasi-MC, multi-level MC, etc.), sparse grids, probabilistic collocation methods, etc. Sampling method are often classified as *non-intrusive*. This means that we do not need to modify the equations of our model to perform uncertainty analysis, but simply sample them many times for different conditions.

Another approach to compute the statistics of a given model problem (set of equations describing a physical system, neural network, etc) in the presence of uncertainties is to represent to output of the model relative to a set of stochastic basis functions, e.g., multivariate polynomials of random variables with given probability distributions. This approach is known as *polynomial chaos* (PC) [23], and has many different variants (generalized PC, multi-element PC). The method allows us to compute the solution to a model problem with a small number of random variables and often exhibits exponential convergence rate. Polynomial chaos and related methods based on series expansions of the model problem are often classified as *intrusive methods*. The adjective “intrusive” emphasizes the fact that the equations of motion to propagate uncertainty are problem-dependent and require an ad-hoc derivation and corresponding coding.

A third class of methods relies on transforming the model problem from the state space to probability space and solve for the probability density function of the solution. An example of such transformation is the Liouville equation (linear hyperbolic PDE) for the probability density function of the solution of a nonlinear dynamical system evolving from a random initial state. Another example is the Fokker-Plank equation governing the PDF of the state of a nonlinear dynamical system driven by random (white) noise. In the context of partial differential equations (infinite-dimensional dynamical systems), the PDF equations corresponding to the solution of nonlinear PDEs evolving from random initial states are *functional differ-*

ential equations [20, 18, 2]. Probability density function methods can also be used also to model and study neural nets (neural nets are essentially discrete dynamical systems [21]). Moreover, all Bayesian inference approaches, e.g., Gaussian regression, probabilistic graphical models, and data assimilation techniques, heavily rely probability density function methods.

PDF methods are also very attractive for systems with *unknown governing equations* (if such equations even exist!), or systems for which governing equations can be discovered only "locally" and in an approximate form. Examples of such systems are mathematical models of brain, large random networks of interacting individuals, the mechanical behavior random heterogeneous materials, disease propagation, stock market models. In all these cases it not straightforward to derive a computational model that accurately describes the system in all its features, and can be used to accurately forecast quantities of interest. Recent advances in data-driven modeling and artificial intelligence, open the possibility to discover model and equations from data. Of course, dealing with model uncertainty on the top of uncertainty in operating conditions, parameters, forcing terms, etc., opens a whole new dimension to the problem of modeling and prediction.

It also raises deep philosophical questions regarding the appropriateness of the mathematics we are using to build our models, and therefore the validity of our computations.

Probability space

There is a well-developed mathematical theory that allows us to describe randomness in the world we live in, or at least the way we perceive it. Such theory is known as probability theory [15]. The proper mathematical foundations of probability theory are quite abstract and technical, as they involve rather advanced concepts of *measure theory* [9]. However, for our purposes it possible to avoid most technicalities and have a version of probability theory that allows for computation, and can be digested by the most (including myself). Let me describe hereafter the basic ingredients of such theory.

To formally describe the outcome of an "experiment" from a mathematical viewpoint it is convenient to define the *probability space* $(\Omega, \mathcal{F}, P)$ which consists of the following items:

- Ω (sample space): the set of all possible outcomes of the experiment
- $\mathcal{F}$ (event space): set of events, en event being a set defined as union or intersection of elements in the sample space.
- P (probability measure): this function assigns each event in $\mathcal{F}$ a probability, which is a number between 0 and 1.

Example 1: Suppose the experiment is rolling a fair dice with 6 faces once. In this case we can define the sample space as

$$\Omega = \{1, 2, 3, 4, 5, 6\} \qquad \text{(sample space).} \tag{1}$$

The definition of the event space depends on what we are interested in. In particular, we may be interested in the following events:

$$\mathcal{F} = \{\emptyset, \Omega, \underbrace{\{1, 3, 5\}}_{\text{odd}}, \underbrace{\{2, 4, 6\}}_{\text{even}}\}. \tag{2}$$

These events can be phrased as: "rolling the dice produces no number" (event $\emptyset$); "rolling the dice returns any number between 1 and 6" (event Ω); "rolling the dice gives an even number" (event $\{2, 4, 6\}$), "rolling the dice returns an odd number" (event $\{1, 3, 5\}$). Clearly, we can assign probabilities to these events as:

$$P(\emptyset) = 0, \qquad P(\Omega) = 1, \qquad P(\{1, 3, 5\}) = \frac{1}{2}, \qquad P(\{2, 4, 6\}) = \frac{1}{2}. \tag{3}$$

Note that in this case assigning probabilities is rather straightforward as we can imagine the process of rolling a dice, and its outcome quite easily. In a similar way, we can assign, e.g., the probability of winning various prizes in the Powerball or the Mega-Millions (assuming the lottery is fair). A rather different story is when we are asked to assign probabilities to complex processes influenced by many variables, e.g., where the leaf falling from a tree is going to land.

Example 2: Let $(\theta(\omega), r(\omega))$ be the polar coordinated identifying where the leaf falling off a tree is going to land. Suppose that $r = 0$ identifies the center of the tree. Clearly $(\theta(\omega), r(\omega))$ is a vector with two random components. In this case, the outcome of the experiment are realizations of two real random variables (coordinates $(\theta(\omega), r(\omega))$ of the leaf after it lands). We can define the following set of events (distance from the tree):

$$\mathcal{F} = \{\emptyset, \Omega, \underbrace{\{\omega : r(\omega) \leq 1\}}_{\text{event 1}}, \underbrace{\{\omega : 1 < r(\omega) \leq 2\}}_{\text{event 2}}, \underbrace{\{\omega : r(\omega) > 2\}}_{\text{event 3}}\}. \tag{4}$$

At this point we need to assign probabilities[1] to each event in (4), which can be done, e.g., by running a very complicated fluid dynamics model (repeated simulations), or by observing many many leaves falling off a tree.

Example 3: Consider an infinite (uncountable) collection of continuous functions $X(t; \omega)$ (stochastic process) defined in the temporal interval $[0, T]$. Let the sample space Ω be the collection of such functions and consider the event space $\mathcal{F}$

$$\mathcal{F} = \{\emptyset, \Omega, \underbrace{\{\omega : X(t; \omega) < 1\}}_{\text{event 1}}, \underbrace{\{\omega : X(t; \omega) \geq 1\}}_{\text{event 2}}\}. \tag{5}$$

In other words, here we are interested in two events only, namely whether the stochastic process $X(t, \omega)$ is (for all $t \in [0, T]$) strictly smaller than one, or larger or equal to one. We can assign a probability to each event in $\mathcal{F}$, e.g., as

$$P(\emptyset) = 0, \qquad P(\Omega) = 1, \qquad P(\{\omega : X(t; \omega) < 1\}) = a, \qquad P(\{\omega : X(t; \omega) \geq 1\}) = 1 - a, \tag{6}$$

where $a \in [0, 1]$. Again, the way a is computed depends on the statistical characterization of the process $X(t; \omega)$. In other words, the calculation leading to $P(\{\omega : X(t; \omega) < 1\})$ may involve quite a lot of operations. Alternatively, the probability of an event $E \in \mathcal{F}$ can be estimated using a frequency approach, i.e., $P(E) \simeq n_E/n$, where n_E is the number E occurs over n trials.

σ-algebra. As we shall see hereafter, in order to perform set operations and corresponding operations on probabilities we need to make sure that $\mathcal{F}$ has the structure of a σ-algebra on Ω. A σ-algebra on Ω is a collection of subsets of Ω that is closed under complement, countable unions, and countable intersections. In other words,

$$A, B \in \mathcal{F} \quad \Rightarrow \quad \begin{cases} A \cap B \in \mathcal{F} \\ A \cup B \in \mathcal{F} \\ A^c, B^c \in \mathcal{F} \quad \text{(complement of } A \text{ and } B\text{, i.e., } A^c = \Omega \setminus A) \end{cases} \tag{7}$$

From this conditions it also follows that $\emptyset, \Omega \in \mathcal{F}$. Moreover, if $\{A_i\}_{i=1}^{\infty} \in \mathcal{F}$ then

$$\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}, \qquad \bigcap_{i=1}^{\infty} A_i \in \mathcal{F} \qquad \text{(countable union and intersection).} \tag{8}$$

[1] For a thorough discussion on the meaning of probability and how to assign probabilities see [15, Chapters 1-3].

Examples of σ-algebras:

- Consider the sample space $\Omega = \{a, b, c\}$. The power set of Ω, i.e., the combination of all possible elements of Ω (including the empty set), is a σ-algebra.

$$2^{\Omega} = \{\emptyset, a, b, c, \{a, b\}, \{a, c\}, \{b, c\}, \underbrace{\{a, b, c\}}_{\Omega}\} \qquad \text{(power set).} \tag{9}$$

  The cardinality of the power set, i.e., the number of elements of the set 2^{Ω} is equal to $2^{\#\Omega}$ (where # denotes the number of elements of a set). In the specific case of (9) we have $\#\Omega = 3$, and therefore $\#2^{\Omega} = 2^3 = 8$.

- If the sample space Ω is countably infinite (i.e., the elements of Ω can be put in a correspondence with $\mathbb{N}$) then the power set 2^{Ω} is isomorphic to $\mathbb{R}$, i.e., it is an uncountable set.

- If the sample space Ω is uncountably infinite, e.g., $\Omega = [0, 1]$ then any σ-algebra $\mathcal{F}$ on Ω can be represented as a sub-algebra of the power set 2^{Ω} (Stone's representation theorem [9]). This is why the σ-algebra $\mathcal{F}$ on an uncountably infinite sample space Ω is often written as a subset of the power set 2^{Ω}, i.e., $\mathcal{F} \subseteq 2^{\Omega}$.

- The σ-algebra on $\Omega = \mathbb{R}$ is the σ-algebra of the collection of all open subsets of $\mathbb{R}$. Such σ-algebra necessarily contains all open sets, all closes sets, and all (countable) unions and intersections of open and closed sets. Such σ-algebra is a sub-algebra of the power set 2^{Ω}.

Probability measure. The probability function

$$P : \mathcal{F} \to [0, 1] \tag{10}$$

assigns to each event A in the σ-algebra F a number $P(A) \in [0, 1]$. In other words, $P(A)$ measures the likelihood that A occurs. The probability function P satisfies the properties of a *measure* (hence the name probability measure[2]):

1. $P(\emptyset) = 0$.

2. $P(\Omega) = 1$.

3. For all countable collections of *disjoint* sets $A_i \in \mathcal{F}$

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i). \tag{11}$$

4. For all $A, B \in \mathcal{F}$,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \tag{12}$$

From these properties it follows that of the event $B \in \mathcal{F}$ is a subset of $A \in \mathcal{F}$ then $A = B \cup (B^c \cap A)$, which implies that (note that B and $B^c \cap A$ are disjoint)

$$P(A) = P(B) + P(B^c \cap A) \geq P(B). \tag{13}$$

[2] In real analysis, the pair $(\Omega, \mathcal{F})$ is called *measurable space* [9]. The elements of $\mathcal{F}$, i.e., the events, are called *measurable sets*. The triple $(\Omega, \mathcal{F}, P)$ is called *probability space*, which is essentially a measurable space in which we define a probability measure.

Frequency interpretation of the probability measure: Suppose that in an experiment the event A shows up n_A times out of n trials. If we define the empirical distribution

$$\mu_n(A) = \frac{n_A}{n} \tag{14}$$

then

$$P(A) = \lim_{n\to\infty} \mu_n(A). \tag{15}$$

Random variables

Let $(\Omega, \mathcal{F}, P)$ be a probability space. A real-valued random variable $X(\omega)$ is a measurable map from the sample space Ω into $\mathbb{R}$, i.e.,

$$X : \Omega \to \mathbb{R}. \tag{16}$$

The *distribution function* of the random variable $X(\omega)$ is defined as

$$F(x) = P(\underbrace{\{\omega : X(\omega) \leq x\}}_{\text{event}}) \qquad x \in \mathbb{R}. \tag{17}$$

The distribution function represents the measure of the set (event) $\{\omega \in \Omega : X(\omega) \leq x\}$, i.e., the probability that $X(\omega)$ is smaller than a given real number x. By using the properties of the probability measure P it is straightforward to conclude that:

1. $F(-\infty) = 0$,
2. $F(\infty) = 1$,
3. $F(x)$ is non-decreasing, i.e., $x_1 < x_2 \quad \Rightarrow F(x_1) \leq F(x_2)$,
4. $P\left(\{\omega : X(\omega) > x\}\right) = 1 - F(x)$,
5. $F(x)$ is continuous from the right, i.e.,

$$\lim_{\epsilon\to 0^+} F(x+\epsilon) = F(x), \tag{18}$$

6. $F(x)$ is not continuous from the left (for discrete random variables),
7. $P\left(\{\omega : a < X(\omega) \leq b\}\right) = F(b) - F(a)$
8. $P\left(\{\omega : a \leq X(\omega) \leq b\}\right) = F(b) - \lim_{\epsilon\to 0^+} F(a+\epsilon)$.

The proof of 1.-8. can be found in [15, Chapter 4].

If $F(x)$ is continuous in x then we say that the random variable $X(\omega)$ is *continuous*. If $F(x)$ is a staircase function then the random variable $X(\omega)$ is *discrete*. $F(x)$ is discontinuous and not staircase, then we say that $X(omega)$ is *mixed*.

Frequency interpretation of the distribution function $F(x)$: Suppose we perform an experiment n-times and observe n realization of the random variable $X(\omega)$, say $\{X(\omega_1), \dots, X(\omega_n)\}$. Let us place all these numbers on the x axis of a Cartesian plane, and form a staircase function, where each step at $X(\omega_i)$ has height $1/n$. Then the staircase function $F_n(x)$ converges to $F(x)$ in the limit $n \to \infty$.

Probability density function. The probability density function (PDF) $p(x)$ of the random variable $X(\omega)$ is (technically speaking) the Radon–Nikodym derivative[3] (assuming it exists) of the probability measure P. The existence of the Radon–Nikodym derivative allows us to write the cumulative distribution function (17) as

$$F(x) = \int_{-\infty}^{x} p(y)dy. \tag{21}$$

Equivalently $p(x)$ can be interpreted as the (weak) derivative of $F(x)$, i.e.,

$$p(x) = \frac{dF(x)}{dx}. \tag{22}$$

By taking the limit of Lebesgue-integrable Dirac delta sequences, we can make sense of Radon–Nikodym PDFs converging to Dirac deltas. This is useful when dealing with the PDF of deterministic (non-random) variables, or discrete random variables. For example,

$$p(x) = \delta(x-a) \qquad \text{(PDF of the random variable } X(\omega) = a \text{ for all } \omega \in \Omega), \tag{23}$$

and

$$p(x) = \sum_{i=1}^{N} p_i \delta(x - x_i) \qquad \text{(PDF of a discrete random variable with range } \{x_1, \ldots, x_n\}). \tag{24}$$

In particular, the PDF of a fair dice with 6 faces is

$$p(x) = \frac{1}{6}\sum_{i=1}^{6} \delta(x-i). \tag{25}$$

By using the properties of the cumulative distribution function $F(x)$ it is straightforward to derive the following properties for the PDF

$$p(x) \geq 0 \quad \text{(positivity)}, \qquad \int_{-\infty}^{\infty} p(x)dx = 1 \quad \text{(normalization)}. \tag{26}$$

Other properties are

$$P\left(\{\omega : x_1 < X(\omega) \leq x_2\}\right) = \int_{x_1}^{x_2} p(x)dx, \qquad P\left(\{\omega : x < X(\omega) \leq x + dx\}\right) = p(x)dx \tag{27}$$

Frequency interpretation of PDFs: Suppose we sample the random variable $X(\omega)$ n times and find that $n_{\Delta x}$ samples fall between x and $x + \Delta x$. By using equation (27), and the frequency interpretation of probability we conclude that

$$p(x)\Delta x \simeq \frac{n_{\Delta x}}{n} \quad \Rightarrow \quad p(x) \simeq \frac{1}{\Delta x}\frac{n_{\Delta x}}{n}. \tag{28}$$

[3] A probability measure P on the measurable space $(\Omega, \mathcal{F})$ is said to be *absolutely continuous* with respect to another measure ν if for all events $E \in \mathcal{F}$ such that $P(E) = 0$ we have $\nu(E) = 0$. In other words, P is absolutely continuous with respect to ν if all impossible events (measured relative to P) are also impossible relative to ν. This is denoted as $\nu \ll P$. Consider, in particular, the Lebesgue measure $d\nu = dx$. The Radon-Nikodym theorem says that if P is absolutely continuous with respect to the Lebesgue measure, then there exists a unique function $p(x)$ such that

$$P(E) = \int_E p(x)dx. \tag{19}$$

Setting the event E in (19) as

$$E = \{w : X(\omega) \leq x\} \in \mathcal{F} \tag{20}$$

yields equation (21).

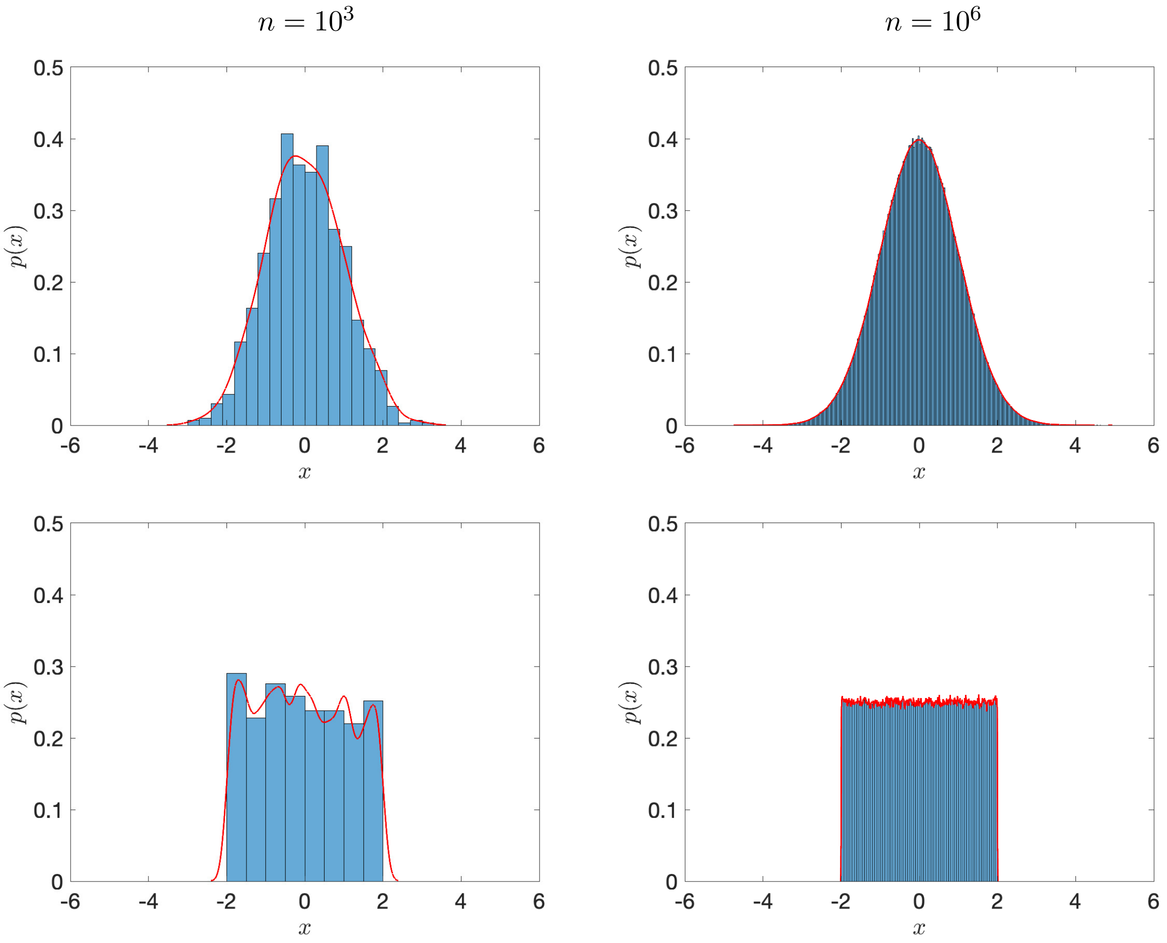


Figure 1: Estimation of the PDF of a Gaussian random variable (first row) and a uniform random variable (second row) using the frequency approach, i.e., formula (28), and the kernel density estimate discussed in [4] (red line) . We plot results for a different number of samples n.

Hence, by dividing the support of the random variable $X(\omega)$ into bins and counting the number of samples within each bin allows us to estimate the PDF of $X(\omega)$ in a rather straightforward way. This is at the basis of the Monte-Carlo estimation method for random variables. There are of course more effective methods to estimate the PDF of one random variable from data (see, e.g., [4]). In figure 1 we estimate the PDF of a Gaussian random variable using frequency approach, i.e., equation (28), and the kernel density estimation method discussed in [4].

Examples of one-dimensional PDFs:

- Gaussian (continuous):

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad x \in \mathbb{R}. \tag{29}$$

- Uniform (continuous):

$$p(x) = \frac{1}{b-a}, \qquad x \in [a,b]. \tag{30}$$

- Binomial (discrete):

$$p(x) = \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} \delta(x-i), \qquad p \in]0,1[, \qquad x \geq 0. \tag{31}$$

- Poisson (discrete):

$$p(x) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \delta(x-k), \qquad \lambda \in]0,\infty[, \qquad x \geq 0. \tag{32}$$

Functions of one random variable. In this section we discuss how to compute the probability density function of a random variable $Y(\omega)$ defined as a deterministic nonlinear function of another random variable $X(\omega)$. To this end, let $(\Omega, \mathcal{F}, P)$ be a probability space,

$$g : \mathbb{R} \to \mathbb{R} \tag{33}$$

a deterministic function,

$$X, Y : \Omega \to \mathbb{R} \tag{34}$$

random variables. Suppose we are given the PDF $p_X(x)$ of $X(\omega)$, and that

$$Y(\omega) = g(X(\omega)) \tag{35}$$

for all $\omega \in \Omega$. What is PDF $Y(\omega)$? Since X and Y are defined on the same probability space we have

$$F_Y(y) = P\left(\{\omega : Y(\omega) \leq y\}\right) = P\left(\{\omega : g(X(\omega)) \leq y\}\right). \tag{36}$$

Therefore, to determine the distribution function $F_Y(y)$ we just need to measure the set

$$B_y = \{\omega : g(X(\omega)) \leq y\} \tag{37}$$

for each y in the set of $g(\mathscr{R}(X))$ (where $\mathscr{R}(X)$ denotes the range of the random variable X). The set B_y is shown in Figure 2 (in yellow) for a prototype function $g(x)$ and a specific value of y. Clearly, the distribution function $F_Y(y)$ must be defined case-by-case. With reference to Figure 2 we have

$$F_Y(y) = F_X(x_1(y)) + 1 - F_X(x_2(y)), \tag{38}$$

where $x_1(y)$ and $x_2(y)$ are the branches of the inverse function $g^-1(y)$. The function (38) represents the distribution function of Y in terms of cumulative distribution function of X, which we know.

With the cumulative distribution function of Y available, it is straightforward to compute the PDF of Y, by taking the (weak) derivative of $F_Y(y)$. This is formalized in the following theorem

Theorem 1. Let X be a random variable with PDF $p_X(x)$, $g \in C^1(\mathbb{R})$ a continuously differentiable function. Then the PDF of $Y = g(X)$ is given by

$$p_Y(y) = \sum_{i=1}^{r} \frac{p_X(x_i(y))}{|g'(x_i(y))|}, \tag{39}$$

where $x_i(y)$ $(i = 1, \ldots, r)$ are the real roots of the equation $g(x) = y$, and $g'(x_i(y))$ is assumed to be non-zero[4].

[4] If $g'(x_i(y)) = 0$ then formula (39) does not apply, and we need to resort to a different method. For example we can use the distribution function approach outlined in Figure 2, i.e., we could measure sets depending on y with the probability measure P and connect such set to the distribution function of X.

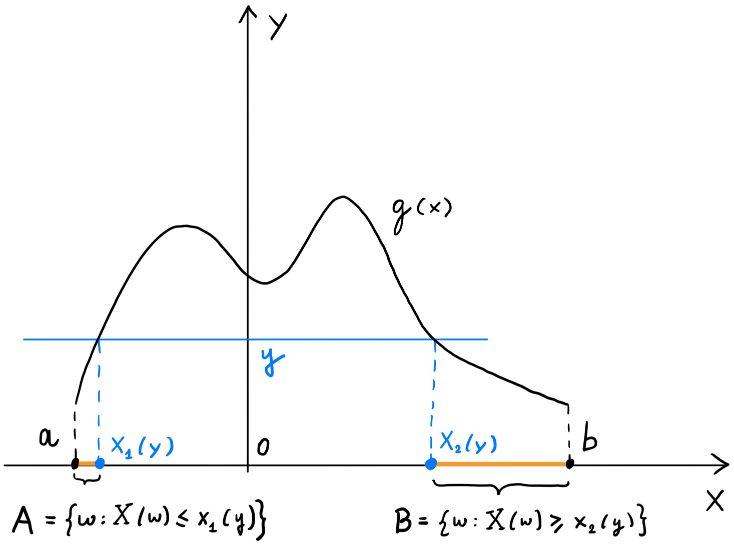


Figure 2: Sketch of the set B_y defined in equation (37) (yellow lines). The random variable X is compactly supported in $[a, b]$. The distribution function of the random variable $Y(\omega) = g(X(\omega))$ evaluated at y is the measure of the set $B_y = A \cup B$ (union of the two yellow lines), i.e., $F_Y(y) = F_X(x_1(y)) + 1 - F_X(x_2(y))$.

Proof. We prove the theorem using Fourier transforms[5]. Let

$$\phi_Y(a) = \int_{-\infty}^{\infty} e^{iay} p_Y(y)dy = \int_{-\infty}^{\infty} e^{iag(x)} p_X(x)dx. \tag{40}$$

Taking the inverse Fourier transform yields

$$p_Y(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{ia(g(x)-y)} p_X(x)dxda. \tag{41}$$

Next, recall that

$$\delta\left(g(x) - y\right) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ia(g(x)-y)} da. \tag{42}$$

Substituting (42) into (41) yields (see also [10])

$$p_Y(y) = \int_{-\infty}^{\infty} \delta\left(g(x) - y\right) p_X(x)dx. \tag{43}$$

At this point we use the well-known identity[6]

$$\delta\left(g(x) - y\right) = \sum_{i=1}^{r} \frac{\delta\left(x - x_i(y)\right)}{\left|g'\left(x_i(y)\right)\right|}, \tag{44}$$

where $x_i(y)$ are the real roots of the $y = g(x)$ for each $y \in \mathbb{R}$. A substitution of (44) into (43) yields (39). This completes the proof.

□

[5] The Fourier transform of a the probability density function $p_X(x)$ is known as *characteristic function* of the random variable $X(\omega)$ (see Eq. (90)).

[6] The identity (44) if and only if $g'(x_i(y)) \neq 0$.

Examples of probability density mappings: Let X be a random variable with probability density function $p_X(x)$. In the following examples we derive the PDF of $Y = g(X)$ for a few prototype $g(x)$.

- Consider the random variable $Y(\omega) = X(\omega)^2$. The mapping $y = g(x) = x^2$ between the random variables X and Y can be inverted (with real roots) for all $y \geq 0$. This yields

  $$x_1(y) = \sqrt{y}, \qquad x_2(y) = -\sqrt{y} \qquad y \geq 0. \tag{45}$$

  By using Theorem 1 we immediately obtain

  $$p_Y(y) = \frac{1}{2\sqrt{y}} \left[p_X(\sqrt{y}) + p_X(-\sqrt{y})\right]. \tag{46}$$

  For instance, if $p_X(x)$ is Gaussian , i.e.,

  $$p_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \tag{47}$$

  then

  $$p_Y(y) = \frac{e^{-y/2}}{\sqrt{2\pi y}} \qquad (\chi^2\text{-distribution}). \tag{48}$$

  Similarly, if X is uniformly distributed in $[-1, 1]$ then[7]

  $$p_Y(y) = \frac{1}{2\sqrt{y}} \quad \text{for all } 0 \leq y \leq 1. \tag{49}$$

- Consider the random variable $Y(\omega) = e^{tX(\omega)}$, where $t \geq 0$ is a real parameter. The mapping $y = g(x) = e^{tx}$ can be inverted (with unique solution) for all $y > 0$ as

  $$x = \frac{\log(y)}{t} \qquad y > 0. \tag{50}$$

  The derivative of $g(x)$ is $g'(x) = te^{tx}$. Therefore

  $$p_Y(y) = \frac{1}{ty} p_X\left(\frac{\log(y)}{t}\right) \qquad y > 0. \tag{51}$$

Application to dynamical systems. Let us briefly discuss two applications of the PDF mapping technique to simple one-dimensional dynamical systems.

- Consider the following Cauchy problem for one ODE evolving from a random initial state

  $$\begin{cases} \dfrac{dx}{dt} = f(x) \\ x(0) = X(\omega) \end{cases} \tag{52}$$

  We know from AM 214 that if f is continuously differentiable in x then the system generates a smooth flow map $x(t) = x(t, X(\omega))$ (differentiable in X) that takes any initial state $X(\omega)$ (at $t = 0$) and maps it to the corresponding solution at time t. Given the PDF of the initial condition $p_X(x)$ we can compute the PDF of $x(t)$ as

  $$p(x, t) = \int_{-\infty}^{\infty} \delta\left(x - x(t, y)\right) p_X(y) dy. \tag{53}$$

[7] Recall that a uniform PDF in $[-1, 1]$ is $p_X(x) = 1/2$ for all $x \in [-1, 1]$.

A convenient way to actually compute such PDF is by sampling, i.e., compute sample paths of $x(t)$ corresponding to samples of $X(\omega)$. However, if the flow map is available analytically then we can also compute (53) analytically. To this end, consider the system

$$\begin{cases} \dfrac{dx}{dt} = x^2 \\ x(0) = X(\omega) \end{cases} \tag{54}$$

We known that the analytical solution (flow map) is

$$x(t, X) = \frac{X(\omega)}{1 - tX(\omega)}. \tag{55}$$

Suppose that $X(\omega)$ is uniformly distributed in $[-1, 0]$ so that the flow map exists for all $t \geq 0$ (no blow-up). What is then the PDF of $x(t, X)$ at each fixed time t? Clearly, we can invert

$$g(x) = \frac{x}{1 - tx} = y \tag{56}$$

uniquely for each $t \geq 0$ ($x \leq 0$) as

$$x(1 + ty) = y \quad \Rightarrow \quad x(y) = \frac{y}{1 + ty}. \tag{57}$$

The first derivative of (56) with respect to x evaluated at the unique root $x(y) = y/(1 + ty)$ is

$$g'(x) = \frac{1}{(1 - tx(y))^2} = (1 + ty)^2. \tag{58}$$

At this point we use Theorem 1 to conclude that the PDF of the solution to the ODE (54) at each fixed time t is

$$p(x, t) = \frac{1}{(1 + tx)^2} p_X\left(\frac{x}{1 + tx}\right). \tag{59}$$

In particular, if p_X is the PDF of a uniform random variable in $[-1, 0]$ then the support of $p(x, t)$ is defined by the condition

$$-1 \leq \frac{x}{1 + tx} \leq 0 \quad \Rightarrow -\frac{1}{(1 + t)} \leq x \leq 0. \tag{60}$$

Hence, as t goes to infinity the support of $p(x, t)$ shrinks to 0 and $p(x, t)$ converges to a Dirac delta function at $x = 0$. Note that for each fixed t we have that the normalization condition of the PDF $p(x, t)$ is satisfied. In fact,

$$\int_{-\infty}^{\infty} p(x, t) dx = \int_{-1/(1+t)}^{0} \frac{1}{(1 + tx)^2} \underbrace{1}_{p_X\left(\frac{x}{1+tx}\right)} dx = 1. \tag{61}$$

- Next, consider the linear decay problem

$$\begin{cases} \dfrac{dx}{dt} = \xi(\omega) x \\ x(0) = 1 \end{cases} \tag{62}$$

where $\xi(\omega)$ is a random variable with known probability density $p_\xi(x)$. The analytical solution to (62) is

$$x(t, \xi(\omega)) = e^{t\xi(\omega)}. \tag{63}$$

By using equation (51), we immediately conclude that the probability density of the solution to (62) is

$$p(x,t) = \frac{1}{tx} p_X\left(\frac{\log(x)}{t}\right) \qquad x > 0. \tag{64}$$

For instance, if p_X is a uniform PDF in $[-2, 0]$ then the support of $p(x,t)$ is defined by

$$-2t \leq \log(x) \leq 0 \quad \Rightarrow \quad e^{-2t} \leq x \leq 1. \tag{65}$$

At $t = 0$ the PDF of the solution is supported only at one point, i.e., $x = 1$. Indeed,

$$p(x,0) = \delta(x-1) \quad \text{(deterministic initial condition)}. \tag{66}$$

For $t > 0$ the PDF of the solution to (62) is[8]

$$p(x,t) = \frac{1}{2tx} \quad \text{for} \quad e^{-2t} \leq x \leq 1. \tag{68}$$

Data-driven identification of the PDF of the initial state. What is the probability density of the initial state $p(x,0)$ that generates an envelope of trajectories that is as close as possible to a measured quantity of interest $h(x(t))$? This is an inverse problem that can be solved by minimizing a performance metric, i.e., a dissimilarity measure between the measurements at various times and the envelope of trajectories, over the parameters representing the initial probability density function.

Liouville equation. The PDF of the solution to the Cauchy problem (52) satisfies the following linear hyperbolic conservation law (see Appendix A of the present course note)

$$\frac{\partial p(x,t)}{\partial t} + \frac{\partial}{\partial x}\left(f(x)p(x,t)\right) = 0, \qquad p(x,0) = p_X(x). \tag{69}$$

This equation is known as Liouville equation. It is straightforward to show by using the method of characteristics that (59) is the solution of Liouville equation (69) for $f(x) = x^2$, i.e., for the dynamical system (54). In Appendix A, we prove that the joint probability density function of the phase space variables of any n-dimensional nonlinear dynamical system

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{f}(\boldsymbol{x}), \qquad \boldsymbol{x}(0) = \boldsymbol{X}(\omega) \tag{70}$$

evolving from a random initial state satisfies $\boldsymbol{X}(\omega)$ satisfies the Liouville equation

$$\frac{\partial p(\boldsymbol{x},t)}{\partial t} + \nabla \cdot \left(\boldsymbol{f}(\boldsymbol{x})p(\boldsymbol{x},t)\right) = 0. \tag{71}$$

To solve (71) one could propagate characteristic curves from the support of random initial state $p(\boldsymbol{x},0)$, or use more sophisticated methods, e.g., numerical tensor methods [7, 8] or physics-informed neural network techniques [16].

Sampling from arbitrary one-dimensional PDFs. Let $X(\omega)$ be a uniform random variable in $[0,1]$. We would like to find a mapping $g(X)$ such that the (continuous) random variable $Y(\omega) = g(X)$ has a

[8] Note that the PDF (68) integrates to one. In fact

$$\int_{-\infty}^{\infty} p(x,t)dx = \int_{e^{-2t}}^{1} \frac{1}{tx} \underbrace{\frac{1}{2}}_{p_X\left(\frac{\log(x)}{t}\right)} dx = 1. \tag{67}$$

desired probability density $p_Y(y)$. With such mapping g available we can transform each sample of $X(\omega)$ to a sample of the PDF p_Y, hence constructing a sampler for $Y(\omega)$. As we shall see hereafter, if we denote by $F_Y(y)$ the cumulative distribution of the continuous random variable Y (the random variable we are interested in sampling) then the mapping g is simply the inverse of F_Y, i.e., $Y(\omega) = F_Y^{-1}(X(\omega))$.

Lemma 1. Let $X(\omega)$ be a uniform random variable in $[0, 1]$. Consider a second random variable $Y(\omega)$ with PDF p_Y and cumulative distribution function

$$F_Y(y) = \int_{-\infty}^{y} p_Y(x)dx \tag{72}$$

The random variable $Y = F_Y^{-1}(X)$ has cumulative distribution function $F_Y(y)$.

Proof. Suppose that F_Y is invertible. Let us show that the random variable $F_Y^{-1}(X)$ has indeed cumulative distribution function $F_Y(y)$. By definition,

$$\begin{aligned} F_Y(y) &= P\left(\{\omega : Y(\omega) \leq y\}\right) \\ &= P\left(\{\omega : F_Y^{-1}(X(\omega)) \leq y\}\right) \\ &= P\left(\{\omega : X(\omega) \leq F_Y(y)\}\right) \qquad (F_Y \text{ invertible and nondecreasing}) \\ &= F_X(F_Y(y)) \\ &= F_Y(y). \end{aligned} \tag{73}$$

In fact, since $X(\omega)$ is uniform in $[0, 1]$ we have $F_X(x) = x$ for all $x \in [0, 1]$.

□

Expectation, moments and cumulants. Let $(\Omega, \mathcal{F}, P)$ be a probability space, $X : \Omega \to \mathbb{R}$ a random variable with cumulative distribution function $F_X(x)$ and PDF $p_X(x)$. For any function $g(X)$ we define the *expectation* of $g(X)$ as [9]

$$\mathbb{E}\{g(X)\} = \int_{-\infty}^{\infty} g(x)dF_X(x) = \int_{-\infty}^{\infty} g(x)p_X(x)dx. \tag{75}$$

Clearly, if $Y(\omega) = g(X(\omega))$ is a random variable with PDF $p_Y(y)$, we can equivalently express the expectation as

$$\mathbb{E}\{g(X)\} = \mathbb{E}\{Y\} = \int_{-\infty}^{\infty} ydF_Y(y) = \int_{-\infty}^{\infty} yp_Y(y)dy. \tag{76}$$

In particular, if we set $g(X) = X^k$ then $\mathbb{E}\left\{X^k\right\}$ are called *moments*[10] of the random variable X

$$\mathbb{E}\left\{X^k\right\} = \int_{-\infty}^{\infty} x^k dF_X(x) = \int_{-\infty}^{\infty} x^k p_X(x)dx. \tag{77}$$

[9] We do not need to assume the existence of the PDF to define the expectation operator. In fact, a more general expression for (75) is

$$\mathbb{E}\{g(X(\omega))\} = \int_{\Omega} g(X(\omega))dP(\omega). \tag{74}$$

[10] There are random variables for which moments do not exist. An example is the Cauchy random variable. Random variables with compactly supported range have all moments. For such compactly supported random variables it is always possible to reconstruct the PDF p_X from the knowledge of its moments or cumulants. In other words, the so-called *moment problem* has a unique solution for compactly supported PDFs.

The first few moments of a random variable X are

$$\mathbb{E}\{X\} = \int_{-\infty}^{\infty} x p_X(x) dx \qquad \text{(mean)}, \tag{78}$$

$$\mathbb{E}\left\{X^2\right\} = \int_{-\infty}^{\infty} x^2 p_X(x) dx \qquad \text{(second-order moment)}, \tag{79}$$

$$\mathbb{E}\left\{X^3\right\} = \int_{-\infty}^{\infty} x^3 p_X(x) dx \qquad \text{(third-order moment)}. \tag{80}$$

The moments of random variable are the coefficients of the power series expansion of the so-called *moment generating function*

$$M(a) = \mathbb{E}\left\{e^{aX(\omega)}\right\} \tag{81}$$

In fact,

$$M(a) = M(0) + \underbrace{\frac{dM(0)}{da}}_{\mathbb{E}\{X\}} a + \frac{1}{2}\underbrace{\frac{d^2M(0)}{da^2}}_{\mathbb{E}\{X^2\}} a^2 + \cdots . \tag{82}$$

In general,

$$\mathbb{E}\left\{X^k\right\} = \frac{d^kM(0)}{da^k}. \tag{83}$$

A function related to the moment generating function is the *cumulant generating function*

$$\Psi(a) = \log(M(a)). \tag{84}$$

The coefficients of the power series expansion of $\Psi(a)$ are called *cumulants* of the random variable $X(\omega)$

$$\Psi(a) = \Psi(0) + \underbrace{\frac{d\Psi(0)}{da}}_{\mathbb{E}\{X\}} a + \frac{1}{2}\underbrace{\frac{d^2\Psi(0)}{da^2}}_{\mathbb{E}\{X^2\}-\mathbb{E}\{X\}^2} a^2 + \cdots \tag{85}$$

The cumulants of a random variable X are often denotes as $\left\langle X^k\right\rangle_c$. For example, we have

$$\langle X\rangle_c = \mathbb{E}\{X\}, \tag{86}$$

$$\left\langle X^2\right\rangle_c = \mathbb{E}\left\{X^2\right\} - \mathbb{E}\{X\}^2, \tag{87}$$

$$\left\langle X^3\right\rangle_c = \mathbb{E}\left\{X^3\right\} - 3\mathbb{E}\{X\}\mathbb{E}\left\{X^2\right\} + 2\mathbb{E}\{X\}^3, \tag{88}$$

$$\cdots$$

The quantity

$$\left\langle X^2\right\rangle_c = \mathbb{E}\left\{X^2\right\} - \mathbb{E}\{X\}^2, \tag{89}$$

is the *variance* of the random variable X. Finally, we define the the *characteristic function* of the random variable $X(\omega)$ as

$$\phi(a) = \mathbb{E}\left\{e^{iaX(\omega)}\right\} \tag{90}$$

where i is the imaginary unit. We have seen this function already, i.e., in the proof of Theorem 1. The characteristic function is the Fourier transform of the probability density function $p(x)$. It is straightforward to show that

$$\mathbb{E}\left\{X^k\right\} = \frac{1}{i^k}\frac{d^k\phi(0)}{da^k}. \tag{91}$$

By expanding the complex exponential function in a power series, and using the definition of cumulants we obtain the following *cumulant expansion* of $\phi(a)$ (see, e.g., [12])

$$\phi(a) = \exp\left[\sum_{j=1}^{\infty} \left\langle X^j \right\rangle_c \frac{(ia)^j}{j!}\right]. \tag{92}$$

Example: The characteristic function of a Gaussian random variable with mean μ and variance σ^2 is

$$\phi(a) = e^{i\mu a - \sigma^2 a^2/2}. \tag{93}$$

This expression can be derived by the taking the Fourier transform of (29), or by using (92). In fact, for Gaussian random variables we have that only the first two cumulants are non-zero, i.e.,

$$\langle X \rangle_c = \mathbb{E}\{X\} = \mu, \tag{94}$$

$$\left\langle X^2 \right\rangle_c = \mathbb{E}\left\{X^2\right\} - \mathbb{E}\{X\}^2 = \sigma^2, \tag{95}$$

$$\left\langle X^k \right\rangle_c = 0 \quad \text{for all } k \geq 3. \tag{96}$$

Substituting these expressions into (92) yields (93).

Random vectors

Let $(\Omega, \mathcal{F}, P)$ be a probability space. A real-valued random vector $\boldsymbol{X}(\omega) = (X_1(\omega), \dots, X_n(\omega))$ is a measurable map from Ω into $\mathbb{R}^n$, i.e.,

$$\boldsymbol{X} : \Omega \to \mathbb{R}^n. \tag{97}$$

Each component $X_i(\omega)$ of the random vector $\boldsymbol{X}(\omega)$ is a real-valued random variable. The *distribution function* of the random vector $\boldsymbol{X}(\omega)$ is defined as

$$F(x_1, \dots, x_n) = P(\underbrace{\{\omega : X_1(\omega) \leq x_1\} \cap \dots \cap \{\omega : X_n(\omega) \leq x_n\}}_{\text{element of } \mathcal{F} \text{ (event) defined as intersection of events}}). \tag{98}$$

As before, if P is absolutely continuous with respect to the Lebesgue measure $dx_1 \cdots dx_n$ then there exists a (Lebesgue integrable) probability density function[11] $p(x_1, \dots, x_n)$ such that

$$F(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} p(y_1, \dots, y_n) dy_1 \cdots dy_n. \tag{99}$$

Equivalently, we can express $p(x_1, \dots, x_n)$ as a (weak) derivative of $F(x_1, \dots, x_n)$ as

$$p(x_1, \dots, x_n) = \frac{\partial^n F(x_1, \dots, x_n)}{\partial x_1 \cdots \partial x_n}. \tag{100}$$

The multivariate distribution function F and associated probability density function p satisfy similar properties as the properties we have seen for one one random variable (see [15] for details).

Frequency interpretation of the joint PDF: Suppose we observe realizations of a random vector $\boldsymbol{X}(\omega)$ with only two components, i.e., $X_1(\omega)$ and $X_2(\omega)$. By using (98)-(99), we have

$$P(\{\omega : x_1 \leq X_1(\omega) \leq x_1 + \Delta x_1\} \cap \{\omega : x_2 \leq X_2(\omega) \leq x_2 + \Delta x_2\}) \simeq p(x_1, x_2) \Delta x_1 \Delta x_2. \tag{101}$$

[11] As before, the probability density function $p(x_1, \dots, x_n)$ is the Radon-Nikodym derivative of the probability measure P relative to the Lebesgue measure $dx_1 \cdots dx_n$.

Let us partition the tensor product space $\mathbb{R}^2$ with an evenly-spaced grid of width Δx_1 (along x_1) and Δx_2 (along x_2). Suppose we observe n realizations of the random vector $\boldsymbol{X}(\omega) = (X_1(\omega), X_2(\omega))$, and suppose that $n_A < n$ instances satisfy the condition

$$\{x_1 \leq X_1(\omega) \leq x_1 + \Delta x_1\} \quad \text{and} \quad \{x_2 \leq X_2(\omega) \leq x_2 + \Delta x_2\}. \tag{102}$$

Then from (101) we obtain the PDF estimate

$$p(x_1, x_2) \simeq \frac{1}{\Delta x_1 \Delta x_2} \frac{n_A}{n}. \tag{103}$$

More efficient and accurate methods to estimate the PDF from data are based on kernels [4] (see Figure 4)

Marginal probability density and marginal distribution. Let $\boldsymbol{X}(\omega) = (X_1(\omega), X_2(\omega))$ be a random vector with joint distribution function function $F(x_1, x_2)$. The distribution of the random variable $X_1(\omega)$ can be obtained from $F(x_1, x_2)$ simply by sending x_2 to infinity, i.e.,

$$F(x_1) = \lim_{x_2 \to \infty} F(x_1, x_2). \tag{104}$$

In fact,

$$\lim_{x_2 \to \infty} F(x_1, x_2) = P(\{\omega : X_1(\omega) \leq x_1\} \cap \{\omega : X_2(\omega) \leq \infty\}) = P(\{\omega : X_1(\omega) \leq x_1\}) = F(x_1). \tag{105}$$

We can write the last equation in terms of PDFs as

$$\lim_{x_2 \to \infty} \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} p(y_1, y_2) dy_1 dy_2 = \int_{-\infty}^{x_1} p(y_1) dy_1. \tag{106}$$

Since x_1 is arbitrary, it follows from (106) that

$$p(x_1) = \int_{-\infty}^{\infty} p(x_1, x_2) dx_2 \qquad \text{(marginalization rule)}. \tag{107}$$

Moreover, we have $F(\infty, \infty) = 1$, i.e.,

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x_1, x_2) dx_1 dx_2 = 1 \qquad \text{(normalization condition)}. \tag{108}$$

It is straightforward to extend these formulas to distribution functions and PDFs in more than two variables. For example, if $\boldsymbol{X}(\omega) = (X_1(\omega), X_2(\omega), X_3(\omega), X_4(\omega))$ is a four-dimensional random vector with distribution function $F(x_1, \ldots, x_4)$ and PDF $p(x_1, \ldots, x_4)$, then we can obtain the joint distribution function and the joint PDF of X_2 and X_3, respectively, as

$$F(x_2, x_3) = F(\infty, x_2, x_3, \infty), \qquad p(x_2, x_3) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x_1, x_2, x_3, x_4) dx_1 dx_4. \tag{109}$$

Example (Gaussian distribution): Consider the multivariate Gaussian PDF

$$p(x_1, \ldots, x_n) = \frac{1}{\sqrt{(2\pi)^n \det(\boldsymbol{\Sigma})}} e^{-(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}-\boldsymbol{\mu})/2}, \tag{110}$$

where

$$\boldsymbol{x}^T = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}, \tag{111}$$

$$\boldsymbol{\mu}^T = \begin{bmatrix} \mathbb{E}\{X_1\} & \dots & \mathbb{E}\{X_n\} \end{bmatrix} \qquad \text{(mean)}, \tag{112}$$

$$\Sigma_{ij} = \mathbb{E}\{X_i X_j\} - \mathbb{E}\{X_i\}\mathbb{E}\{X_j\} \qquad \text{(covariance matrix)}. \tag{113}$$

It is straightforward to show that all marginal PDF and distribution functions are still Gaussians of the form (110).

Independence. Let $(\Omega, \mathcal{F}, P)$ be a probability space. Two events $A \in F$ and $B \in \mathcal{F}$ are said to be *independent* if the probability of their intersection (that means the probability that both events A and B happen) equals the product of their probabilities, i.e.,

$$A, B \in \mathcal{F} \quad \text{independent} \quad \Leftrightarrow \quad P(A \cap B) = P(A)P(B). \tag{114}$$

Consider now a random vector $\boldsymbol{X}(\omega) = (X_1(\omega), X_2(\omega))$ with components $X_1(\omega)$ and $X_2(\omega)$. We say that the random variables $X_1(\omega)$ and $X_2(\omega)$ are statistically independent if

$$P(\underbrace{\{\omega : X_1(\omega) \leq x_1\}}_{\text{event } A} \cap \underbrace{\{\omega : X_2(\omega) \leq x_2\}}_{\text{event } B}) = P(\{\omega : X_1(\omega) \leq x_1\})P(\{\omega : X_2(\omega) \leq x_2\}), \tag{115}$$

for all $x_1, x_2 \in \mathbb{R}$. Equation (115) can be written in terms of the cumulative distribution function as

$$F(x_1, x_2) = F(x_1)F(x_2). \tag{116}$$

This also implies that the joint PDF of X_1 and X_2 (if it exists) is simply the product of the PDF of X_1 and the PDF of X_2, i.e.,

$$p(x_1, x_2) = p(x_1)p(x_2). \tag{117}$$

These formulas can be generalized to n independent random variables as

$$F(x_1, \dots, x_n) = F(x_1) \cdots F(x_n), \qquad p(x_1, \dots, x_n) = p(x_1) \cdots p(x_n). \tag{118}$$

Examples:

- Jointly uniform random vector. Let $\boldsymbol{X}$ be a n-dimensional random vector with zero-mean i.i.d. (independent identically distributed) uniform components in [-1,1]. The joint PDF of $\boldsymbol{X}$ is

$$p(x_1, \dots, x_n) = \begin{cases} \dfrac{1}{2^n} & (x_1, \dots, x_n) \in [-1, 1]^n \\ 0 & \text{otherwise} \end{cases} \tag{119}$$

- Jointly normal random vector. Let $\boldsymbol{X}$ be a n-dimensional random vector with zero-mean i.i.d. Gaussian components with variance equal to one. The joint PDF of $\boldsymbol{X}$ is

$$p(x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2}} e^{-\boldsymbol{x}^T \boldsymbol{x}/2} \qquad \boldsymbol{x} \in \mathbb{R}^n. \tag{120}$$

  Clearly, from equation (110) we see that Gaussian random variables are independent if and only if

$$\mathbb{E}\{X_i X_j\} = \mathbb{E}\{X_i\}\mathbb{E}\{X_j\} \quad \text{for } i \neq j. \tag{121}$$

  In general, if (121) is satisfied then we say that X_i and X_j are *uncorrelated.* Lack of correlation is a much weaker statement than independence, yet sufficient to claim independence for Gaussian random variables.

Conditional distribution function and conditional PDF. Conditional probability is a measure of the probability of an event A occurring, given that another event B has already occurred. Suppose that the two aforementioned events belong to the σ-algebra $\mathcal{F}$ of a probability space $(\Omega, \mathcal{F}, P)$. Then the probability of A under the condition B is defined as[12]

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \tag{122}$$

Note that the conditional probability is non-zero if A and B are intersecting. Also note that if B is a subset of A then $P(A|B) = 1$.

Clearly, if A and B are independent events then by equation (114) we have that $P(A \cap B) = P(A)P(B)$. This implies that if A and B are independent then $P(A|B) = P(A)$. In other words, B has no effect whatsoever on the probability of A occurring. Moreover, $P(A \cap B) \leq P(B)$ and therefore we always have that $P(A|B) \leq 1$.

In the context of random vectors with multiple components, we may be interested in determining the conditional probability of an event involving one component, given that another event involving another component has already occurred. This yields the concept of conditional distribution function and conditional probability density. Let us first clarify these concepts for a random vector with only two components $\boldsymbol{X}(\omega) = (X_1(\omega), X_2(\omega))$. By using the definition of the cumulative distribution function (98) we obtain (see [15, Ch. 7])

$$F(x_1|x_2) = \frac{F(x_1, x_2)}{F(x_2)} \quad \Leftrightarrow \quad F(x_1, x_2) = F(x_1|x_2)F(x_2). \tag{123}$$

The determination of the conditional density of $X_1(\omega)$ assuming $X_2(\omega) = x_2$, i.e., a specific value of $X_2(\omega)$ is of particular interest. This density cannot be derived directly from (122) because, in general, the event $X_2(\omega) = x_2$ has zero probability. However, one can make sense of such conditional probability by taking a suitable limit. Specifically, consider

$$P(\{X_1(\omega) \leq x_1\} \cap \{x_2 < X_2(\omega) \leq x_2 + \Delta x_2\}) = F(x_1, x_2 + \Delta x_2) - F(x_1, x_2) \tag{124}$$

and

$$P(\{x_2 < X_2(\omega) \leq x_2 + \Delta x_2\}) = F(x_2 + \Delta x_2) - F(x_2). \tag{125}$$

In (124) it is understood that $F(x_1, x_2)$ is the joint distribution function of (X_1, X_2), while in (125) $F(x_2)$ denotes the distribution function of X_2 alone. Clearly, for small Δx_2

$$F(x_1, x_2 + \Delta x_2) - F(x_1, x_2) \simeq \Delta x_2 \int_{-\infty}^{x_1} p(y_1, x_2) dy_1, \tag{126}$$

and

$$F(x_2 + \Delta x_2) - F(x_2) \simeq p(x_2)\Delta x_2. \tag{127}$$

By differentiating (123) with respect to x_1, and taking into account (126)-(127) yields the conditional PDF

$$p(x_1|X_2 = x_2) = \frac{\Delta x_2 \dfrac{\partial}{\partial x_1} \displaystyle\int_{-\infty}^{x_1} p(y_1, x_2) dy_1}{\Delta x_2 p(x_2)}, \tag{128}$$

[12] An example of conditional probability could be the following:

- Event A: "Daniele's team scores a goal".
- Event B: "Daniele takes a shot at the goal".

The conditional probability $P(A|B)$, i.e., the probability that Daniele's team scores a goal, conditional to Daniele taking a shot equals the probability that Daniele takes a shot and scores a goal, divided by the probability that Daniele takes a shot.

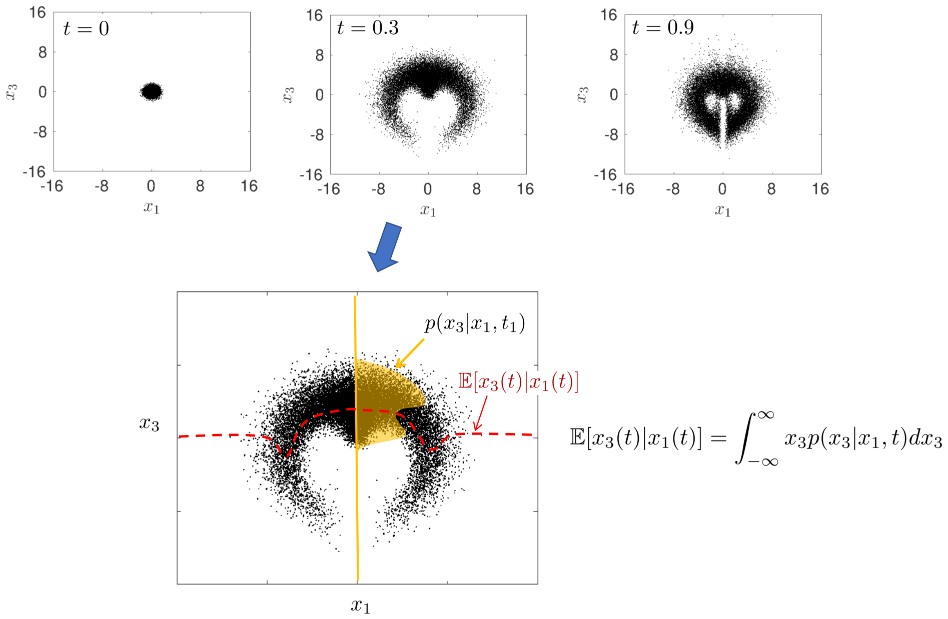


Figure 3: Point clouds representing the joint PDF of the phase variables $x_1(t)$ and $x_3(t)$ of the Kraichnan-Orzag system at different times, i.e., $p(x_3, x_1, t)$. Shown is the procedure to compute the conditional PDF $p(x_3|x_1, t)$ and the corresponding conditional mean $\mathbb{E}\{X_3|X_1 = x_1\}$.

i.e.,

$$p(x_1|X_2 = x_2) = \frac{p(x_1, x_2)}{p(x_2)} \quad \text{(conditional PDF)}. \tag{129}$$

In summary, to compute the conditional PDF, $p(x_1|X_2 = x_2)$ we literally take a section of the joint $p(x_1, x_2)$ for some fixed value of x_2 and then rescale the function we obtain by the number $p(x_2)$, i.e., the one-dimensional PDF of $p(x)$ of $X_2(\omega)$ evaluated at $x = x_2$. This procedure is illustrated in Figure 3 for a PDF represented in terms of a point cloud.

Equation (129) can be written as

$$p(x_1, x_2) = p(x_1|x_2)p(x_2) = p(x_2|x_1)p(x_1) \tag{130}$$

which yields the identities

$$p(x_2) = \int_{-\infty}^{\infty} p(x_2|x_1)p(x_1)dx_1, \qquad p(x_1) = \int_{-\infty}^{\infty} p(x_1|x_2)p(x_2)dx_2. \tag{131}$$

The conditional probability density rule can be generalized to multiple random variables. For instance, if $p(x_1, x_2, x_3, x_4)$ denotes the joint PDF of four random variables then

$$p(x_1, x_2, x_3, x_4) = p(x_1|x_2, x_3, x_4)p(x_2, x_3, x_4) = p(x_1|x_2, x_3, x_4)p(x_2|x_3, x_4)p(x_3|x_4)p(x_4). \tag{132}$$

Moreover, conditional probability densities satisfy the *marginalization rule*. For instance

$$p(x_1, x_3|x_4, x_5) = \int_{-\infty}^{\infty} p(x_1, x_2, x_3|x_4, x_5)dx_2. \tag{133}$$

This property follows directly from the definition of conditional probability density (129).

Expectation, joint moments, and joint cumulants. Let $\boldsymbol{X}(\omega) = (X_1(\omega), \dots, X_n(\omega))$ be a random vector defined on the probability space $(\Omega, \mathcal{F}, P)$. For any measurable function $g(X_1, \dots, X_n)$ we define the expectation[13] as

$$\mathbb{E}\left\{g(X_1, \dots, X_n)\right\} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \dots, x_n) p(x_1, \dots, x_n) dx_1 \cdots dx_n. \tag{135}$$

In particular, if $g(X_1, \dots, X_n) = X_1^{k_1} \cdots X_n^{k_n}$ then

$$\mathbb{E}\left\{X_1^{k_1} \cdots X_n^{k_n}\right\} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_1^{k_1} \cdots x_n^{k_n} p(x_1, \dots, x_n) dx_1 \cdots dx_n \qquad \text{(joint moments)} \tag{136}$$

The correlation matrix[14] and the covariance matrix are defined as (see, e.g., (110))

$$\mathbb{E}\left\{X_i X_j\right\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_i x_j p(x_i, x_j) dx_i dx_j \qquad \text{(correlation matrix)}, \tag{138}$$

$$\mathbb{E}\left\{(X_i - \mu_i)(X_j - \mu_j)\right\} = \mathbb{E}\left\{X_i X_j\right\} - \mu_i \mu_j \qquad \text{(covariance matrix)}. \tag{139}$$

where $\mu_i = \mathbb{E}\left\{X_i\right\}$ (mean of X_i).

Remark: We say that two random variables $X_i(\omega)$ and $X_j(\omega)$ are *uncorrelated* if

$$\mathbb{E}\left\{X_i X_j\right\} = \mathbb{E}\left\{X_i\right\} \mathbb{E}\left\{X_j\right\}. \tag{140}$$

Independent random variables are always uncorrelated. In fact, let $p(x_i, x_j)$ be the joint PDF of X_i and X_j. We know that if X_i and X_j are independent then $p(x_i, x_j)$ can be factorized as

$$p(x_i, x_j) = p(x_i) p(x_j). \tag{141}$$

A substitution of (141) into (138) immediately yields (140).

We define the *moment generating function* of the random vector $\boldsymbol{X}(\omega) = (X_1(\omega), \dots, X_n(\omega))$ as

$$m(a_1, \dots, a_n) = \mathbb{E}\left\{e^{a_1 X_1 + \cdots + a_n X_n}\right\}. \tag{142}$$

[13] Note that the expectation $\mathbb{E}\{\cdot\}$ is a linear operator from a space of functions, e.g., the space of real-valued functions that are measurable with respect $p(x_1, \dots, x_n)$. Also, we do not need to assume the existence of the PDF to define the expectation operator. In fact, a more general expression for (135) is

$$\mathbb{E}\left\{g(X_1, \dots, X_n)\right\} = \int_{\Omega} g(X_1(\omega), \dots, X_n(\omega)) dP(\omega). \tag{134}$$

[14] Note that (138) follows from (136) using the marginalization property of the PDF. For instance

$$\begin{aligned}\mathbb{E}\left\{X_1 X_2\right\} &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_1 x_2 p(x_1, \dots, x_n) dx_1 \cdots dx_n \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 \left(\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(x_1, \dots, x_n) dx_3 \cdots dx_n \right) dx_1 dx_2 \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 p(x_1, x_2) dx_1 dx_2. \end{aligned} \tag{137}$$

.

It is straightforward to show that

$$\begin{aligned}\frac{\partial m(0,\ldots,0)}{\partial a_i} =&\mathbb{E}\{X_i\}, &(143)\\ \frac{\partial^2 m(0,\ldots,0)}{\partial a_j\partial a_i} =&\mathbb{E}\{X_iX_j\} &(144)\\ \frac{\partial^3 m(0,\ldots,0)}{\partial a_j\partial a_i\partial a_k} =&\mathbb{E}\{X_iX_jX_k\}, \\ &\cdots &(145)\end{aligned}$$

Hence, the partial derivatives of the moment generating function evaluated at zero represent the joint moments of the components of random vector $\boldsymbol{X}$. Clearly, if $m(a_1,\ldots,a_n)$ admits a convergent power series expansion at 0 then all joint moments exist.

The *cumulant generating function* of the random vector $\boldsymbol{X}(\omega)=(X_1(\omega),\ldots,X_n(\omega))$ is defined as

$$\Psi(a_1,\ldots,a_n)=\log(m((a_1,\ldots,a_n))). \tag{146}$$

It is straightfoward to show that

$$\begin{aligned}\frac{\partial \Psi(0,\ldots,0)}{\partial a_i} =&\mathbb{E}\{X_i\}, &(147)\\ \frac{\partial^2 \Psi(0,\ldots,0)}{\partial a_j\partial a_i} =&\mathbb{E}\{X_iX_j\}-\mathbb{E}\{X_i\}\mathbb{E}\{X_j\}, &(148)\\ \frac{\partial^3 \Psi(0,\ldots,0)}{\partial a_j\partial a_i\partial a_k} =&\mathbb{E}\{X_iX_jX_k\}-\mathbb{E}\{X_i\}\mathbb{E}\{X_jX_k\}-\mathbb{E}\{X_j\}\mathbb{E}\{X_iX_k\}-\mathbb{E}\{X_k\}\mathbb{E}\{X_iX_j\}\\ &+2\mathbb{E}\{X_i\}\mathbb{E}\{X_j\}\mathbb{E}\{X_k\},\\ &\cdots\end{aligned}$$

The quantities at the right hand side are known as *joint cumulants* of the random variables $(X_1,\ldots,X_n)$. The cumulants are often denoted as $\langle X_iX_j\cdots\rangle_c$ (see, e.g., [12])

$$\begin{aligned}\langle X_i\cdots\rangle_c =&\mathbb{E}\{X_i\},\\ \langle X_iX_j\cdots\rangle_c =&\mathbb{E}\{X_iX_j\}-\mathbb{E}\{X_i\}\mathbb{E}\{X_j\}, &(149)\\ &\cdots.\end{aligned}$$

The *characteristic function* of the random vector $\boldsymbol{X}(\omega)=(X_1(\omega),\ldots,X_n(\omega))$ is defined as

$$\phi(a_1,\ldots,a_n)=\mathbb{E}\left\{e^{i(a_1X_1+\cdots+a_nX_n)}\right\}. \tag{150}$$

Note that the characteristic function is the Fourier transform of the joint probability density function $p(x_1,\ldots,x_n)$ and therefore it essentially carries the same information. The joint moments of $\boldsymbol{X}$ can be computed as

$$\mathbb{E}\left\{X_1^{k_1}\cdots X_n^{k_n}\right\}=\frac{1}{i^{k_1+\cdots+k_n}}\frac{\partial^{k_1+\cdots+k_n}\phi(0,\ldots,0)}{\partial^{k_1}a_1\cdots\partial^{k_n}a_n}. \tag{151}$$

It is interesting to notice that the marginalization operation we have seen for the PDF, e.g.,

$$p(x_1)=\int_{-\infty}^{\infty}p(x_1,x_2,\ldots,x_n)dx_2\cdots dx_n \tag{152}$$

turns out to be simplified quite substantially in Fourier space. Indeed

$$\phi(a_1) = \phi(a_1, 0, \dots, 0) = \mathbb{E}\left\{e^{ia_1X_1+i0X_2\cdots+i0X_n}\right\}. \tag{153}$$

By using the well known series expansion of the complex exponential, it is possible to show that (see, e.g., [12])

$$\phi(a_1, a_2, \dots, a_n) = \exp\left[\sum_{\nu_1,\dots,\nu_n=0}^{\infty} \langle X_1^{\nu_1}\cdots X_n^{\nu_n}\rangle_c \prod_{k=1}^{n}\frac{(ia_k)^{\nu_k}}{\nu_k!}\right] \tag{154}$$

where the series at the exponent excludes the case $\nu_1 = \cdots = \nu_n = 0$. For example,

$$\phi(a_1, a_2) = \phi(a_1)\phi(a_2)\exp\left[\sum_{j,k=1}^{\infty}\left\langle X_1^jX_2^k\right\rangle_c \frac{(ia_1)^j(ia_2)^k}{j!k!}\right], \tag{155}$$

where we used (92). Clearly, if X_1 and X_2 are independent we have $\left\langle X_1^jX_2^k\right\rangle_c = 0$ for all i and j and therefore (155) reduces to

$$\phi(a_1, a_2) = \phi(a_1)\phi(a_2). \tag{156}$$

Clearly, this equation is the Fourier transform of the PDF $p(x_1, x_2) = p(x_1)p(x_2)$, and shows that if X_1 and X_2 are independent both the joint PDF and the joint characteristic function can be factorized as a product of one-dimensional functions.

Conditional expectation. Let $\boldsymbol{X}(\omega)$ and $\boldsymbol{Y}(\omega)$ be two random vectors defined on the probability space $(\Omega, \mathcal{F}, P)$. The *conditional mean* of $u(\boldsymbol{X}(\omega))$ (u is an arbitrary measurable function) assuming $\boldsymbol{Y}(\omega) = \boldsymbol{y}$ is defined as[15]

$$\mathbb{E}\{g(\boldsymbol{X})|\boldsymbol{Y} = \boldsymbol{y}\} = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g(\boldsymbol{x})p(\boldsymbol{x}|\boldsymbol{y})d\boldsymbol{x}, \tag{157}$$

where

$$p(\boldsymbol{x}|\boldsymbol{y}) = \frac{p(\boldsymbol{x}, \boldsymbol{y})}{p(\boldsymbol{y})} \tag{158}$$

is the conditional probability density of $\boldsymbol{X}(\omega)$ given $\boldsymbol{Y}(\omega) = \boldsymbol{y}$. Note that the $\mathbb{E}\{g(\boldsymbol{X})|\boldsymbol{Y} = \boldsymbol{y}\}$ is a function of $\boldsymbol{y}$. The conditional mean defined in equation (157) allows us to write the conditional moments of a random variable or a random vector, given information on another random vector. For example, the conditional mean and conditional correlation of $\boldsymbol{X}$ given $\boldsymbol{Y}(\omega) = \boldsymbol{y}$ are defined as

$$\mathbb{E}\{X_i|\boldsymbol{Y} = \boldsymbol{y}\} = \int_{-\infty}^{\infty} x_ip(x_i|\boldsymbol{y})dx_i, \tag{159}$$

$$\mathbb{E}\{X_iX_j|\boldsymbol{Y} = \boldsymbol{y}\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_ix_jp(x_i, x_j|\boldsymbol{y})dx_idx_j. \tag{160}$$

The conditional mean of a system with two random variables is visualized in Figure 3.

By combining (158), (157) and (135) we see that

$$\mathbb{E}\{g(\boldsymbol{X})\} = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\mathbb{E}\{g(\boldsymbol{X})|\boldsymbol{Y} = \boldsymbol{y}\}p(\boldsymbol{y})d\boldsymbol{y}. \tag{161}$$

In this sense, $\mathbb{E}\{g(\boldsymbol{X})|\boldsymbol{Y} = \boldsymbol{y}\}$ can be interpreted as a random variable, i.e., a scalar function of the random variable $\boldsymbol{Y}$ which, if averaged over $p(\boldsymbol{y})$, yuelds exactly $\mathbb{E}\{g(\boldsymbol{X})\}$.

[15] The conditional mean in equation (157) is often written as $\mathbb{E}\{g(\boldsymbol{X})|\boldsymbol{Y}\}$.

Joint PDF of m functions of n random variables. Let $\boldsymbol{X}(\omega) = (X_1(\omega), \dots, X_n(\omega))$ be a random vector with joint probability density function $p(x_1, \dots, x_n)$. Define

$$\begin{cases} Y_1 = g_1(X_1, \dots, X_n) \\ \quad \vdots \\ Y_m = g_m(X_1, \dots, X_n) \end{cases} \tag{162}$$

What is the joint probability density function of the random vector $\boldsymbol{Y} = (Y_1, \dots, Y_m)$? Note that m can be smaller, equal or larger than n. These cases need to be handled differently.

- If $n = m$ and $\{g_1, \dots, g_m\}$ are distinct functions we proceed as in Theorem 2 below.
- If $m < n$ and $\{g_1, \dots, g_m\}$ are distinct functions we can add $m - n$ equations to complement the system so that we have n independent equations in n variables:

$$\begin{cases} Y_1 = g_1(X_1, \dots, X_n) \\ \quad \vdots \\ Y_m = g_m(X_1, \dots, X_n) \\ Y_{m+1} = X_{m+1} \\ \quad \vdots \\ Y_n = X_n \end{cases} \tag{163}$$

  Once the joint PDF of $Y_1, \dots, Y_n$ is known (using Theorem 2 below) then we can marginalize it with respect to $(y_{m+1}, \dots, y_n)$ to obtain $p(y_1, \dots, y_m)$ as

$$p(y_1, \dots, y_m) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(y_1, \dots, y_m, y_{m+1}, \dots y_n) dy_{m+1} \cdots dy_n. \tag{164}$$

- If we have more equations than variables (i.e. $m > n$) then the computation of the joint PDF of $(Y_1, \dots, Y_m)$ is not as straightforward as above. Consider for example the mapping $Y_1(\omega) = X(\omega)$ and $Y_2(\omega) = X^2(\omega)$. Here we have two functions of the same random variable. Note also that $Y_2 = Y_1^2$. It can be shown that the joint PDF of $Y_1 = X$ and $Y_2 = X^2$ is

$$p(y_1, y_2) = p_X(y_1)\delta(y_2 - y_1^2), \tag{165}$$

  where p_X is the PDF of X and $\delta(\cdot)$ is the Dirac delta function.

Theorem 2. Let $\boldsymbol{x}_k(\boldsymbol{y})$ $(k = 1, \dots, r)$ be the zeros of the nonlinear system of equations $\boldsymbol{y} = \boldsymbol{g}(\boldsymbol{x})$ defined in (162) (for $n = m$) or in (163) (for $m < n$). The joint PDF of $Y_1, \dots, Y_n$ is given by

$$p_{\boldsymbol{Y}}(\boldsymbol{y}) = \sum_{i=1}^{r} \frac{p_{\boldsymbol{X}}(\boldsymbol{x}_i(\boldsymbol{y}))}{|J(\boldsymbol{x}_i(\boldsymbol{y}))|}, \tag{166}$$

where J is the Jacobian determinant[16] associated with the mapping $\boldsymbol{g}(\boldsymbol{x})$ evaluated at $\boldsymbol{x}_i(\boldsymbol{y})$ (assumed non-zero).

[16] In (166) it is assumed that

$$J(\boldsymbol{x}_i(\boldsymbol{y})) = \det\left[\frac{\partial \boldsymbol{g}(\boldsymbol{x})}{\partial \boldsymbol{x}}\right]_{\boldsymbol{x}=\boldsymbol{x}_i(\boldsymbol{y})} \neq 0. \tag{167}$$

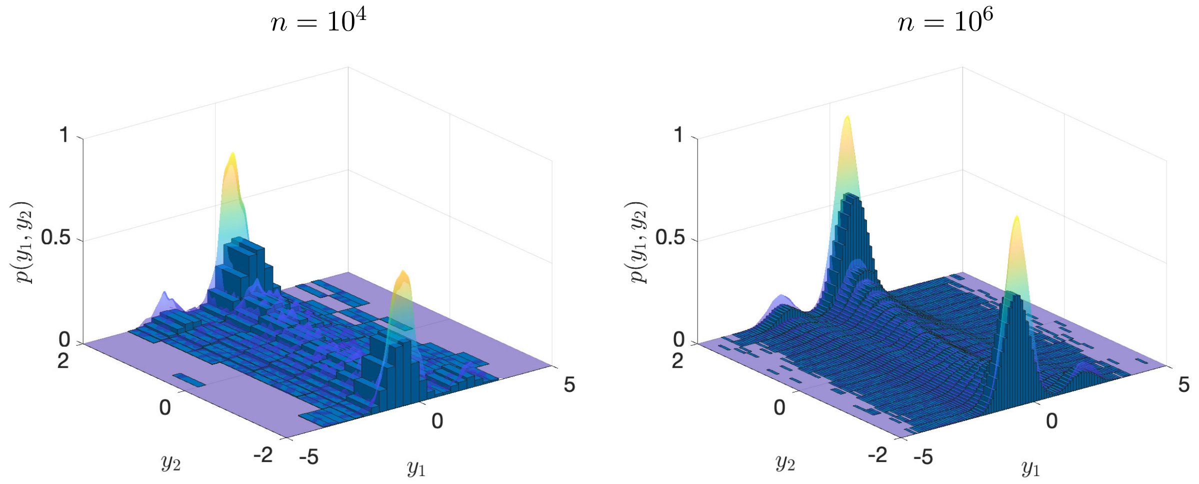


Figure 4: Estimation of the joint PDF of the random variables $Y_1 = X_1$ and $Y_2 = 2\sin(2X_1 + X_2)$ where X_1 and X_2 and independent Gaussians with zero mean and variance one. We show the results we obtain with the frequency approach, i.e., formula (103) and the 2D kernel density estimation method discussed in [4] (transparent surface plot). We plot results for a different number of samples n.

The proof of this theorem is provided in [15, Chapter 8].

Example: Consider the mapping

$$Y_1 = X_1^2 \qquad Y_2 = X_1 + X_2. \tag{168}$$

Suppose we know the joint PDF of X_1 and X_2. What's the joint PDF of Y_1 and Y_2? The following mapping from (X_1, X_2) to (Y_1, Y_2) can be inverted as

$$\begin{cases} y_1 = x_1^2 \\ y_2 = x_1 + x_2 \end{cases} \Rightarrow \quad \begin{cases} x_1 = \pm\sqrt{y_1} \\ x_2 = y_2 \mp \sqrt{y_1} \end{cases} . \tag{169}$$

The Jacobian determinant of (169) is easily obtained as

$$J(x_1, x_2) = \det \begin{bmatrix} 2x_1 & 0 \\ 1 & 1 \end{bmatrix} = 2x_1. \tag{170}$$

Hence, by applying Theorem 2, we obtain the following joint PDF of Y_1 and Y_2 is

$$p_{\boldsymbol{Y}}(y_1, y_2) = \frac{1}{2\sqrt{y}} \left[p_{\boldsymbol{X}}(\sqrt{y_1}, y_2 - \sqrt{y_1}) + p_{\boldsymbol{X}}(-\sqrt{y_1}, y_2 + \sqrt{y_1}) \right] \qquad y_1 \geq 0. \tag{171}$$

Example: Consider the mapping

$$Y_1(\omega) = X_1 \qquad Y_2(\omega) = 2\sin\left(2X_1(\omega) + X_2(\omega)\right), \tag{172}$$

where X_1 and X_2 and independent Gaussians with zero mean and variance one. In Figure 4 we estimate the joint PDF of Y_1 and Y_2 using the frequency approach approach, i.e., formula (103), and the 2D kernel density estimation method discussed in [4].

Alternative methods to compute the joint PDF of functions of random vectors. There are alternative equivalent methods to compute the joint PDF $(Y_1, \ldots, Y_m)$, given the joint PDF $(Y_1, \ldots, Y_n)$, e.g., methods based on the Dirac delta function [10] or methods based on the joint characteristic function. With reference to the previous example we have the joint characteristic function

$$\phi_{\boldsymbol{Y}}(a_1,a_2)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{ia_1x_1^2+ia_2(x_1+x_2)}p(x_1,x_2)dx_1dx_2. \tag{173}$$

Clearly, if $\phi_{\boldsymbol{Y}}(a_1,a_2)$ can be computed then we can simply inverse Fourier transform it to obtain the joint PDF of (Y_1,Y_2). By using Dirac delta functions we can represent directly the joint PDF of the random variable

$$Y(\omega)=g(X_1(\omega),\dots,X_n(\omega)), \tag{174}$$

as

$$\begin{aligned}p(y)=&\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\delta(y-g(x_1,\dots,x_n))p(x_1,\dots,x_n)dx_1\cdots dx_n & (175)\\ =&\frac{1}{2\pi}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}e^{ia(y-g(x_1,\dots,x_n))}p(x_1,\dots,x_n)dx_1\cdots dx_nda. & (176)\end{aligned}$$

Example: Let $Y_1=X$ and $Y_2=X^2$ (two functions of one random variable). What is the joint PDF of Y_1 and Y_2? The mapping (162) yields a Jacobian determinant that is zero, and therefore the mapping it is not invertible. This implies that theorem (2) cannot be applied. However, using the characteristic function approach we obtain

$$\phi(a_1,a_2)=\int_{-\infty}^{\infty}e^{ia_1x+ia_2x^2}p_X(x)dx. \tag{177}$$

Taking the inverse Fourier transform yields,

$$\begin{aligned}p(y_1,y_2)=&\frac{1}{(2\pi)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{ia_1(x-y_1)+ia_2(x^2-y_2)}p_X(x)dxda_1da_2\\ =&\frac{1}{2\pi}\int_{-\infty}^{\infty}\delta(x-y_1)e^{ia_1(x^2-y_2)}p_X(x)dxda_2\\ =&\delta(y_1^2-y_2)p_X(y_1). & (178)\end{aligned}$$

Remark: If $(X_1,\dots,X_n)$ are independent random variables and $(g_1,\dots,g_n)$ are n functions from $\mathbb{R}$ into $\mathbb{R}$, then $Y_1=g_1(X_1)$, …, $Y_n=g_n(X_n)$ are independent random variables. It is straightforward to prove this statement using the Dirac delta function representation (or the characteristic function) of PDF mapping [10]. To this end, let

$$Y_i(\omega)=g_i(X_i(\omega)). \tag{179}$$

We have

$$\begin{aligned}p(y_1,\dots,y_n)=&\int_{-\infty}^{\infty}\prod_{j=1}^{n}\delta(y_j-g_j(x_j))p(x_1,\dots,x_n)dx_1\cdots dx_n\\ =&\prod_{j=1}^{n}\int_{-\infty}^{\infty}\delta(y_j-g_j(x_j))p(x_j)dx_j\\ =&p(y_1)\cdots p(y_n). & (180)\end{aligned}$$

Remark: The PDF of the sum of independent random variables is the *convolution* the PDF of each variable. For example, let

$$Y = X_1 + X_2 + X_3 \tag{181}$$

be the sum of three independent random variables X_1, X_2 and X_3, with PDFs $p_1(x_1)$, $p_2(x_2)$ and $p_3(x_3)$ respectively. By using (175) we obtain

$$\begin{aligned}
p(y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \delta(y - x_1 - x_2 - x_3)p(x_1, x_2, x_3)dx_1dx_2dx_3 \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \delta(x_1 - y + x_2 + x_3)p_1(x_1)p_2(x_2)p_3(x_3)dx_1dx_2dx_3 \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p_1(x_2 + x_3 - y)p_2(x_2)p_3(x_3)dx_2dx_3 \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p_1(x_1 - y)p_2(x_1 - x_3)p_3(x_3)dx_1dx_3.
\end{aligned} \tag{182}$$

In the last equality we considered the mapping $x_1 = x_2 + x_3$ as a coordinate change from x_1 to x_2 with parameter x_3. Note that the process of computing the PDF of the sum of independent random variables can be also seen as a hierarchical process in which we proceed with two variables at a time To this end, we first compute the PDF of $Z = X_2 + X_3$ as

$$p_Z(z) = \int_{-\infty}^{\infty} p_2(z - x_3)p_3(x_3)dx_3. \tag{183}$$

Clearly, Z is independent of X_1 and therefore the PDF of $Y = Z + X_1$ is

$$p_Y(y) = \int_{-\infty}^{\infty} p_1(y - x_1)p_Z(x_1)dx_1. \tag{184}$$

A substitution of (183) into (184) yields (182).

Lebesgue spaces of random variables. The expectation operator $\mathbb{E}\{\cdot\}$ is a linear integral operator over a probability measure. Such an operator can be used to define norms and eventually inner products in spaces of random variables. For example,

$$\mathbb{E}\left\{|X|^q\right\} = \int_\Omega |X(\omega)|^q dP(\omega) \qquad q \in \mathbb{N} \tag{185}$$

is essentially a weighted q norm. The space of random variables satisfying $\mathbb{E}\left\{|X|^q\right\} < \infty$ is denoted as $L^q(\Omega, \mathcal{F}, P)$, in analogy with the classical Lebesgue space for functions. The case $q = 2$ is of particular importance as it has the structure of a Hilbert space. Specifically, for any two random variables in $L^2(\Omega, \mathcal{F}, P)$ we have the inner product

$$\mathbb{E}\left\{XY\right\} = \int_\Omega X(\omega)Y(\omega)dP(\omega) \tag{186}$$

and the norm

$$\mathbb{E}\left\{X^2\right\} = \int_\Omega X(\omega)^2 dP(\omega). \tag{187}$$

The inner product (186) allows us to define orthogonal random variables. Specifically, $X(\omega)$ and $Y(\omega)$ are orthogonal in $L^2(\Omega, \mathcal{F}, P)$ if they are uncorrelated, i.e., $\mathbb{E}\left\{XY\right\} = 0$. Also, $X(\omega)$ and $Y(\omega)$ are orthonormal if they are orthogonal and have norm equal to one, i.e., $\mathbb{E}\left\{X^2\right\} = \mathbb{E}\left\{Y^2\right\} = 1$.

Application to dynamical systems

Consider the following linear dynamical system

$$\begin{cases} \dot{x}(t) + \xi(\omega)x(t) = 0 \\ x(0) = x_0(\omega) \end{cases} \tag{188}$$

where $\xi(\omega)$ and $x_0(\omega)$ are independent random variables. Specifically $\xi(\omega)$ is uniformly distributed in $[0,1]$, while $x_0(\omega)$ is Gaussian random variable with mean zero and variance one. As is well-known, the analytical solution of (188) is

$$x(t;\omega) = x_0(\omega)e^{-t\xi(\omega)}. \tag{189}$$

Let us compute the mean, the second-order moment and the auto-correlation function of the solution $x(t;\omega)$, i.e., $\mathbb{E}\{x(t;\omega)\}$, $\mathbb{E}\{x(t;\omega)^2\}$, and $\mathbb{E}\{x(t;\omega)x(t'\omega)\}$ versus time. We have

$$\mathbb{E}\left\{x(t;\omega)\right\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x_0 e^{-x_0^2/2}dx_0 \int_0^1 e^{-t\xi}d\xi = 0, \tag{190}$$

$$\mathbb{E}\left\{x(t;\omega)^2\right\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x_0^2 e^{-x_0^2/2}dx_0 \int_0^1 e^{-2t\xi}d\xi = \frac{1}{2t}\left(1-e^{-2t}\right), \tag{191}$$

$$\mathbb{E}\left\{x(t;\omega)x(t';\omega)\right\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x_0^2 e^{-x_0^2/2}dx_0 \int_0^1 e^{-(t+t')\xi}d\xi = \frac{1}{t+t'}\left(1-e^{-(t+t')}\right). \tag{192}$$

The one-time probability density function of $x(t;\omega)$ can be easily computed by using the Dirac delta function approach [10]. Indeed,

$$\begin{aligned} p(x,t) =& \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\int_0^1 \delta\left(x - x_0e^{-\xi t}\right)e^{-x_0^2/2}dx_0 d\xi \\ =& \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\int_0^1 \frac{\delta\left(x_0 - xe^{\xi t}\right)}{e^{-\xi t}}e^{-x_0^2/2}dx_0 d\xi \qquad (193) \\ =& \frac{1}{\sqrt{2\pi}}\int_0^1 e^{\xi t-(xe^{\xi t})^2/2}d\xi. \qquad (194) \end{aligned}$$

Now consider the change of variables

$$u = \frac{xe^{\xi t}}{\sqrt{2}} \qquad \Rightarrow \qquad d\xi = \frac{\sqrt{2}}{\xi t}e^{-\xi t}du. \tag{195}$$

A substitution of (195) into (194) yields

$$\begin{aligned} p(x,t) =& \frac{1}{\xi t\sqrt{\pi}}\int_{x/\sqrt{2}}^{xe^t/\sqrt{2}} e^{-u^2}du \qquad (196) \\ =& \frac{1}{\xi t\sqrt{\pi}}\left[\text{erf}\left(\frac{xe^t}{\sqrt{2}}\right) - \text{erf}\left(\frac{x}{\sqrt{2}}\right)\right]. \qquad (197) \end{aligned}$$

Liouville equation approach: We can transform the linear system (188) involving one random variable at the right hand side to an equivalent 2D linear system evolving from a random initial state (an no random variables at the right hand side). To this end, we notice that

$$\begin{cases} \dot{x}(t) + yx(t) = 0 \\ \dot{y}(t) = 0 \\ x(0) = x_0(\omega) \\ y(0) = \xi(\omega) \end{cases} \tag{198}$$

is completely equivalent to (188). In this setting, we can derive a linear transport equation for the joint PDF of $x(t;\omega)$ and $y(t;\omega)$, i.e., $x(t;\omega)$ and $\xi(\omega)$. Such PDF equation takes the form

$$\begin{cases} \dfrac{\partial p(x,y,t)}{\partial t} + \dfrac{\partial}{\partial x}\left(xyp(x,y,t)\right) + \dfrac{\partial}{\partial y}\left(xyp(x,y,t)\right) = 0 \\ p(x,y,0) = p_{x_0}(x)p_{\xi}(y) \end{cases} \tag{199}$$

It can be verified by a direct substitution that the solution the initial value problem (199) is

$$p(x,y,t) = \frac{1}{\sqrt{2\pi}} e^{yt-(xe^{yt})^2/2}. \qquad y \in [0,1], \qquad x \in \mathbb{R}. \tag{200}$$

Note that the joint PDF (200) was already obtained in equation (194), right before marginalizing with respect to ξ.

Data-driven identification of random dynamical systems. A system with random parameters and/or random initial states generates an envelope of trajectories that depends on the joint PDF of the random variables driving the system. It is possible to identify such joint PDF from data, e.g., by minimizing a performance metric, i.e., a dissimilarity measure (e.g., a Wasserstein norm) between the measurements of a quantity of interest at various times and the envelope of trajectories, over the degrees of freedom representing the joint probability density function. of the random variables. In this way, we are essentially trying to reduce *model uncertainty* by shrinking a continuous trajectory tube generated by a random dynamical system of the form

$$\begin{cases} \dfrac{d\boldsymbol{x}(t)}{dt} = \boldsymbol{f}(\boldsymbol{x}(t);\omega) \\ \boldsymbol{x}(0) = \boldsymbol{x}_0(\omega) \end{cases} \tag{201}$$

around measurements of some phase space function $\boldsymbol{h}(\boldsymbol{x}(t))$. Note that $\boldsymbol{f}(\boldsymbol{x}(t);\omega)$ is a *random vector field* and $\boldsymbol{x}_0(\omega)$ is a random initial state. We will shall see hereafter that $\boldsymbol{f}(\boldsymbol{x}(t);\omega)$ can be represented in a Karhunen-Loève expansion

$$f_i(\boldsymbol{x}(t);\omega) = \sum_{k=1}^{\infty} \xi_i^k(\omega)\theta_i^k(\boldsymbol{x}) \qquad i = 1,\dots,n \tag{202}$$

where $\xi_i^k(\omega)$ are uncorrelated random variables and $\theta_i^k(\boldsymbol{x})$ are orthonormal basis functions. Other representations of $f_i(\boldsymbol{x}(t);\boldsymbol{\xi}(\omega))$ can be built, e.g., using tensor expansions in weighted L^2 spaces, e.g., functional tensor train [7, 3, 14]. The minimization procedure discussed above essentially identifies the degrees of freedom of the joint probability density function of $\boldsymbol{x}_0(\omega)$ and $\boldsymbol{\xi}(\omega)$, i.e., $p(\boldsymbol{x}_0, \boldsymbol{\xi})$, either in the form of a *sampler*, e.g., using Wasserstein generative neural networks [1], or the actual multivariate function $p(\boldsymbol{x}_0, \boldsymbol{\xi})$.

Random processes and random fields

Let $\Omega, \mathcal{F}, P)$ be a probability space. A real valued stochastic process in the time interval $[0,T]$ is a mapping

$$X : \Omega \times [0,T] \to \mathbb{R}. \tag{203}$$

The process can be continuous in time (e.g., Brownian motion) discontinuous in time (e.g., telegrapher's random process), or time-discrete, e.g., represented by a sequence of random variables $X(t_j;\omega)$ $j = 1,\dots,n$.

Remark: The notion of continuity we know for real valued functions can be generalized substantially when dealing with stochastic processes. We have, for example,

- Continuity in probability:

$$\lim_{s\to t} P(\{\omega : |X(t;\omega) - X(s;\omega)| > \epsilon\}) = 0 \quad \text{for all } \epsilon > 0. \tag{204}$$

- Mean-square continuity:

$$\lim_{s\to t} \mathbb{E}\{|X(t;\omega) - X(s;\omega)|^2\} = 0. \tag{205}$$

- Continuity in distribution:

$$\lim_{s\to t} F(x,s) = F(x,t) \qquad (F(x,t) \text{ distribution function of } X(t;\omega). \tag{206}$$

Continuity in mean-square ⇒ continuity in probability ⇒ continuity in distribution.

Continuity in probability follows from mean-square continuity[17] thanks to the Markov's inequality

$$P(\{\omega : |X(t;\omega) - X(s;\omega)| > \epsilon\}) \leq \frac{1}{\epsilon^2}\mathbb{E}\{|X(t;\omega) - X(s;\omega)|^2\} \qquad \forall t, s \in [0,T]. \tag{208}$$

Other properties of $X(t;\omega)$ very much depend on the way we characterize the process, i.e., the set of rules and specifications that allow us to fully characterize the process. Clearly, $X(t;\omega)$ is a random variable for each fixed t. This means that $X(t;\omega)$ admits a distribution function

$$F(x,t) = P\left(\{\omega : X(t;\omega) \leq x\}\right), \tag{209}$$

and eventually a probability density function

$$p(x,t) = \frac{dF(x,t)}{dx}. \tag{210}$$

With $F(x,t)$ or $p(x,t)$ available we can compute the statistical moments at time t, e.g.,

$$\mathbb{E}\left\{X(t;\omega)^k\right\} = \int_{-\infty}^{\infty} x^k p(x,t)dx, \quad k \in \mathbb{N}. \tag{211}$$

The PDF $p(x,t)$, however, provides very limited statistical information about the process $X(t;\omega)$. In fact, it does not allow us to compute any joint statistics at different times, for example the autocorrelation function

$$\mathbb{E}\{X(t;\omega)X(s;\omega)\} = \int_{-\infty}^{\infty} x_1 x_2 p(x_1,x_2,t_1,t_2)dx_1 dx_2, \tag{212}$$

where $p(x_1,x_2,t_1,t_2)$ is the joint probability density function of the random variables $X(t_1;\omega)$ and $X(t_2;\omega)$ (t_1 and t_2 here can vary in $[0,T]$). A straightforward generalization of this line of thinking leads us to construct the joint PDF of $\{X(t_1;\omega),\ldots,X(t_n;\omega)\}$ for an increasing number of distinct time instants

[17] Mean square continuity also implies that the mean process $\mathbb{E}\{X(t;\omega)\}$ is continuous in t. In fact, using the inequality $|\mathbb{E}\{X\}|^2 \leq \mathbb{E}\{X^2\}$ we obtain

$$|\mathbb{E}\{X(t;\omega) - X(s;\omega)\}| \leq \sqrt{\mathbb{E}\{|X(t;\omega) - X(s;\omega)|^2\}}. \tag{207}$$

Similarly, if the process is mean-square continuous then the auto-correlation function $\mathbb{E}\{X(t;\omega)X(s;\omega)\}$ continuous in both s and t.

$t_i \in [0, T]$. Similarly, we can construct the joint characteristic function of the random process $X(t;\omega)$ at distinct time instants $(t_1, \dots, t_n)$ as

$$\phi(a_1, \dots, a_n; t_1, \dots, t_n) = \mathbb{E}\left\{ e^{ia_1 X(t_1;\omega) + \cdots + ia_n X(t_n;\omega)} \right\}. \tag{213}$$

This expression can be obtained (at least formally) from the so-called Hopf characteristic functional [20, 11] associated with the stochastic process $X(t;\omega)$, i.e.,

$$\Phi([\theta(t)]) = \mathbb{E}\left\{ \exp\left(\int_0^T X(\tau;\omega)\theta(\tau)d\tau \right) \right\}, \tag{214}$$

where $\theta(t)$ is a deterministic test function which we are free to choose. For example, if we pick

$$\theta(t) = \sum_{i=1}^{n} a_i \delta(t - t_i), \tag{215}$$

and substitute it into (214) then we obtain (213). The Hopf functional[18] (214) provides full statistical information about the stochastic process $X(t;\omega)$, including all joint statistical moments, all multi-time PDFs, etc. For instance, the functional derivatives of Φ evaluated at $\theta = 0$ coincide with the statistical moments (see, e.g. [18])

$$\left.\frac{\delta^q \Phi([\theta])}{\delta\theta(t)^q}\right|_{\theta=0} = \frac{1}{i^q}\mathbb{E}\{X(t;\omega)^q\}, \qquad \left.\frac{\delta^{q+p} \Phi([\theta])}{\delta\theta(t)^q \delta\theta(s)^p}\right|_{\theta=0} = \frac{1}{i^{q+p}}\mathbb{E}\{X(t;\omega)^q X(s;\omega)^p\}. \tag{216}$$

In [11] the Hopf functional is determined for various types of stochastic processes.

Remark: To fully characterize a stochastic process it is not necessary to identify or provide the Hopf functional. A stochastic process can be defined in many different ways, some of which are not even explicit. However, if the Hopf characteristic functional is available, then the process is fully specified, perhaps in the most compact possible way (see [13] for applications of Hopf functional methods to turbulence).

Gaussian processes. The Hopf characteristic functional for a Gaussian process is (see, e.g., [11])

$$\Phi([\theta(t)]) = \mathbb{E}\left\{ \exp\left(i\int_0^T \mu(\tau)\theta(\tau) - \int_0^T \int_0^T C(\tau, s)\theta(\tau)\theta(s)d\tau ds \right) \right\}, \tag{217}$$

where

$$\mu(t) = \mathbb{E}\{X(t;\omega)\} \qquad \text{(mean)}, \tag{218}$$

$$C(t, s) = \mathbb{E}\{X(t;\omega)X(s;\omega)\} - \mu(t)\mu(s) \qquad \text{(covariance function)}. \tag{219}$$

Higher order moments can be computed using functional differentiation (e.g., (216)), or by noticing that the joint characteristic function the random process $X(t;\omega)$ at an arbitrary number of distinct time instants is

$$\phi(a_1, \dots, a_n; t_1, \dots, t_n) = \exp\left(i\sum_{k=1}^{n} a_k \mu(t_k) - \sum_{k,j=1}^{n} C(t_k, t_j) a_k a_j \right). \tag{220}$$

[18] Recall that a functional is a mapping from a certain space of functions (or distributions) into the real line or the complex plane. The Hopf functional is a complex-valued nonlinear functional into $\mathbb{C}$.

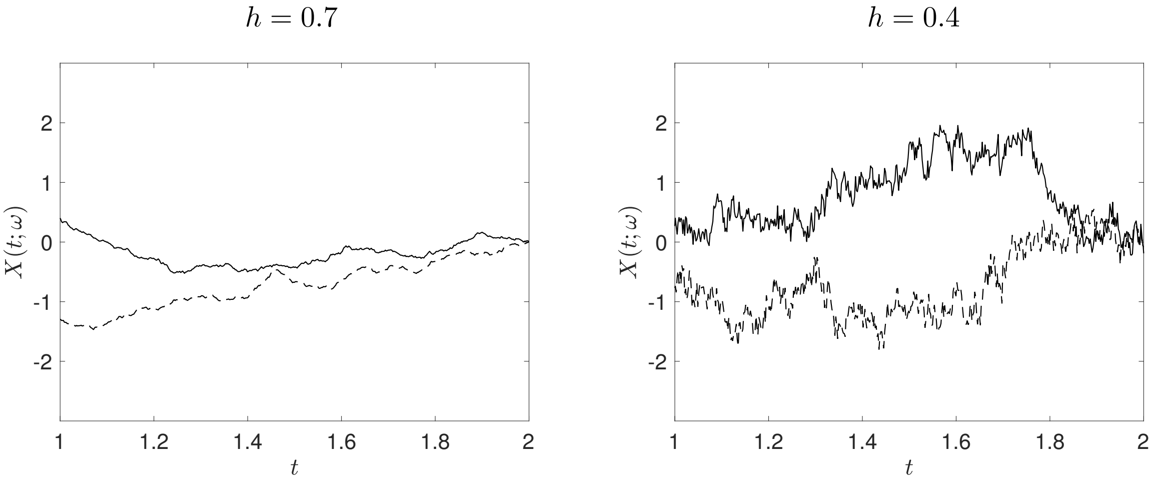


Figure 5: Samples of zero-mean Gaussian process with covariance function (222) and $\sigma = 1$. We show samples corresponding to different values of the Hurst parameter h.

Sampling Gaussian processes: To sample a Gaussian process with mean $\mu(t)$ and covariance function $C(s,t)$ it is sufficient to construct a temporal grid in $[0,T]$ and then sample a Gaussian random vector with mean $\mu_i = \mu(t_i)$, and covariance matrix with entries $C(t_k, t_j)$. To this end, it is sufficient to recall that if $\boldsymbol{X}(\omega)$ is a zero-mean Gaussian random vector (column vector) with independent entries of variance one, and $\boldsymbol{C} = \boldsymbol{R}\boldsymbol{R}^T$ is the Cholesky decomposition of the covariance matrix[19] $\boldsymbol{C}$, then $\boldsymbol{Y} = \boldsymbol{R}\boldsymbol{X}$ is a zero-mean Gaussian random vector with covariance $\boldsymbol{C}$. In fact,

$$\mathbb{E}\{\boldsymbol{Y}(\omega)\boldsymbol{Y}^T(\omega)\} = \mathbb{E}\{\boldsymbol{R}\boldsymbol{X}(\omega)\boldsymbol{X}^T(\omega)\boldsymbol{R}^T\} = \boldsymbol{R}\underbrace{\mathbb{E}\{\boldsymbol{X}(\omega)\boldsymbol{X}^T(\omega)\}}_{\text{(identity matrix)}}\boldsymbol{R}^T = \boldsymbol{C}. \tag{221}$$

In figure 5 we plot a few samples of a Gaussian random process with zero mean and covariance function

$$C(s,t) = \frac{\sigma}{2}\left(|s|^{2h} + |t|^{2h} - |s-t|^{2h}\right), \tag{222}$$

where $0 < h < 1$ is the so-called Hurst parameter. A Gaussian process with covariance function (222) is called fractional Brownian motion.

Gaussian random fields: The procedure we used to sample of Gaussian stochastic process with covariance $C(s,t)$ (e.g., (222)) can be extended to *Gaussian random fields* [17], i.e., random functions defined of a domain $V \subseteq \mathbb{R}^d$. For example, we could sample a zero-mean Gaussian random field $X(\boldsymbol{x};\omega)$ defined on the square domain $V = [0,1] \times [0,1]$ with covariance function

$$C(\boldsymbol{x},\boldsymbol{y}) = \frac{\sigma}{2}\left(\|\boldsymbol{x}\|_2^{2h} + \|\boldsymbol{y}\|_2^{2h} - \|\boldsymbol{x}-\boldsymbol{y}\|_2^{2h}\right), \tag{223}$$

to this end we first construct the covariance matrix $C(\boldsymbol{x}_i, \boldsymbol{x}_j)$ and then use the procedure we used before, i.e.: i) sample a zero-mean i.i.d. Gaussian random variable with variance one at each spatial location $\boldsymbol{x}_i$, and ii) multiply the sample of the random vector constructed in this way by the matrix $\boldsymbol{R}$ obtained by the Cholesky decomposition of the autocovariance function (223). In Figure 6 we provide a few samples of a zero mean Gaussian random field with covariance (223).

[19] The entries of the covariance matrix $\boldsymbol{C}$ are $C(t_i, t_j)$, where $C(t,s)$ is the covariance function of the random process.

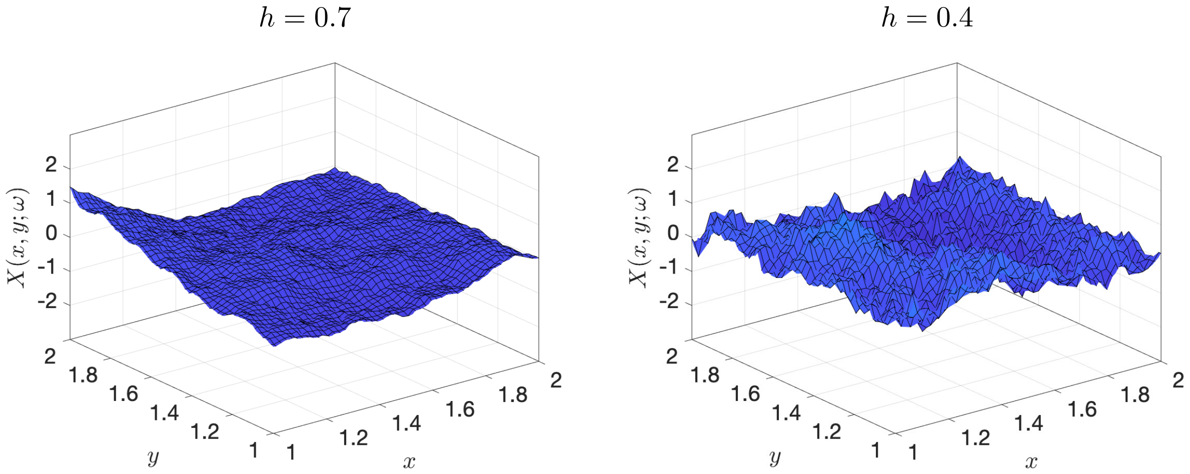


Figure 6: Samples of zero-mean Gaussian random field with covariance function (223) and $\sigma = 1$. We show samples corresponding to different values of the Hurst parameter h.

Discrete Markov processes. Consider a discrete set of distinct temporal time instant, say $\{t_1, \ldots, t_n\}$ and a time-discrete random process which is essentially a collection of random variables

$$X_i(\omega) = X(t_i; \omega), \tag{224}$$

or a collection of random vectors

$$\boldsymbol{X}_i(\omega) = \boldsymbol{X}(t_i; \omega). \tag{225}$$

The random process (224) can be defined in many different ways, for example as a recurrence relation[20]

$$X_{i+1}(\omega) = h(X_i(\omega)) + \xi_i(\omega), \tag{226}$$

where $\xi_i(\omega)$ are random variables and $X_0(\omega)$ is random as well. Similarly, we can define a vector valued discrete process as

$$\boldsymbol{X}_{i+1}(\omega) = \boldsymbol{h}(\boldsymbol{X}_i(\omega)) + \boldsymbol{\xi}_i(\omega), \tag{227}$$

Note that the structure of (227) is the same as a recurrent neural network perturbed by noise [21].

Disregarding how we generate the sequence of random variables $X_0, \ldots, X_n$, in (226), we can characterize the statistics of the process X_i in terms of the joint PDF (assuming it exists) $p(x_n, \ldots, x_0)$. By using the definition of conditional probability density we have

$$p(x_n, \ldots, x_1, x_0) = p(x_n | x_{n-1}, \ldots, x_1, x_0) p(x_{n-1}, \ldots, x_0). \tag{228}$$

If the system is memoryless (or Markovian), we have that the conditional PDF of X_n given the entire history of X_i equals $p(x_n | x_{n-1})$, i.e.,

$$p(x_n | x_{n-1}, \ldots, x_1, x_0) = p(x_n | x_{n-1}). \tag{229}$$

In other words, the PDF of $X_n(\omega)$ conditional to any set of variables $\{X_j(\omega)\}$ with $j < n$ equals to $p(x_n | x_{n-1})$, i.e., it depends only on the value of $X_{n-1}(\omega)$. By applying (229) recursively we obtain

$$p(x_n, x_{n-1}, \ldots, x_1, x_0) = p(x_n | x_{n-1}) p(x_{n-1} | x_{n-2}) \cdots p(x_1 | x_0) p(x_0). \tag{230}$$

[20] The discrete process (226) is also called *autoregressive process*.

Hence, the process is fully specified by the transition density $p(x_{k+1}|x_k)$. Denoting by p_{ξ_k} the PDF of ξ_k in (226), and assuming that $\{\xi_1, \ldots, \xi_{n-1}\}$ are statistically independent we have that the transition probability defined by the Markov chain (226) is

$$p(x_{k+1}|x_k) = p_{\xi_k}(x_{x+1} - F(x_k)). \tag{231}$$

Remark: More general auto-regressive random vector processes of the form (227) are the discussed in the book [5]. For example, vector auto-regressive moving-average (VARMA) processes, integrated VARMA (VARIMA) processes, etc.

Markov Chain Monte Carlo (MCMC). Markov Chain Monte Carlo (MCMC) refers to a class of methods that allow us to sample high-dimensional probability density functions [6]. In MCMC we construct a discrete Markov process that has a stationary PDF that coincides with the distribution of interest, i.e., the PDF we'd like to sample from. Hence, simulations of the Markov chain[21] provide samples of the high-dimensional PDF we are interested in, once a transient, i.e., the so-called *burn-in* phase of the chain, is completed. There are several MCMC algorithms to sample from high-dimensional PDFs. Perhaps the simplest ones are the Gibbs sampling and the Metropolis-Hastings algorithms. Let us briefly describe the Gibbs sampling method. To this end, suppose you are given a three-dimensional PDF $p(x_1, x_2, x_3)$ and that the conditional PDFs $p(x_1|x_2, x_3)$, $p(x_2|x_1, x_3)$ and $p(x_3|x_1, x_2)$ are all available[22]. To sample from $p(x_1, x_2, x_3)$ we proceed as follows:

1. Initialize $x_2 = x_2^{(i)}$ and $x_3 = x_3^{(i)}$. Here $x_2^{(i)}$ and $x_3^{(i)}$ are two real numbers. The superscript "i" is an integer number that labels the discrete Markov process
$$\boldsymbol{X}_i(\omega) = \begin{bmatrix} x_1^{(i)}(\omega) & x_2^{(i)}(\omega) & x_3^{(i)}(\omega) \end{bmatrix} \qquad i \in \mathbb{N}. \tag{232}$$
2. Sample a new $x_1^{(i+1)}$ from the one-dimensional conditional PDF $p\left(x_1|x_2^{(i)}, x_3^{(i)}\right)$.
3. With the sample $x_1^{(i+1)}$ available, sample a new $x_2^{(i+1)}$ from the one-dimensional conditional PDF $p\left(x_2|x_1^{(i+1)}, x_3^{(i)}\right)$.
4. With the sample $x_2^{(i+1)}$ available, sample a new $x_3^{(i+1)}$ from the one-dimensional conditional PDF $p\left(x_3|x_1^{(i+1)}, x_2^{(i+1)}\right)$.
5. Update $x_j^{(i)} \leftarrow x_j^{(i+1)}$ for $j = 1, 2, 3$ and go back to point 2.

This algorithm allows us to compute $\boldsymbol{X}_{i+1}$ from $\boldsymbol{X}_i$ by sampling known one-dimensional conditional transition densities. To sample from such arbitrary one-dimensional transition densities we can use different methods. If the inverse cumulative distribution of each conditional PDF is known, then we have seen that it is sufficient to sample a uniform PDF and then map such sample using the inverse cumulative distribution function. Alternatively, we can determine the mapping between uniform random variables and conditionally distributed random variables using *polynomial chaos expansions*. The mapping $\boldsymbol{X}_i \to \boldsymbol{X}_{i+1}$ defines a *random walk* in $\mathbb{R}^3$. The stationary distribution of such random walk coincides with $p(x_1, x_2, x_3)$.

[21] Simulations of a Markov chain are usually performed with the Monte Carlo method, hence the name Markov Chain Monte Carlo.

[22] Recall that to compute the conditional PDF $p(x_1|x_2, x_3)$ we literally set x_2 and x_3 in $p(x_1, x_2, x_3)$ equal to some number, say $x_2 = x_2^*$ and $x_3 = x_3^*$ and then normalize the one-dimensional function $p(x_1, x_2^*, x_3^*)$ so that the integral with respect to x_1 equals one.

In other words, after the burn-in phase is completed, i.e., for sufficiently large i, we have that $X_i(\omega)$ are samples of the joint PDF $p(x_1, x_2, x_3)$.

Karhunen-Loève expansion. Let $X(t;\omega)$ be a zero-mean square-integrable stochastic process defined on the probability space $(\Omega, \mathcal{F}, P)$. "Square-integrable" means that $X(t;\omega)$ has finite second order moment, i.e.,

$$\mathbb{E}\left\{\int_0^T X(t;\omega)^2 dt\right\} < \infty. \tag{233}$$

By using the properties of $L^2(\Omega, \mathcal{F}, P)$ spaces (probability spaces of square integrable random variables), it can be shown that $X(t;\omega)$ admits a series expansion

$$X(t;\omega) = \sum_{k=1}^{\infty} \sqrt{\lambda_k}\xi_k(\omega)\psi_k(t), \tag{234}$$

where $\{\xi_1(\omega), \xi_2(\omega), \ldots\}$ is a set of uncorrelated, i.e., orthonormal, random variables satisfying

$$\mathbb{E}\left\{\xi_i(\omega)\xi_j(\omega)\right\} = \delta_{ij}, \tag{235}$$

and $\{\psi_1(t), \psi_2(t), \ldots\}$ are orthonormal (in $L^2([0,T])$) temporal modes

$$\int_0^T \psi_i(t)\psi_j(t)dt = \delta_{ij}. \tag{236}$$

By using the orthogonality properties (235)-(236), we obtain the so-called dispersion relations[23]

$$\xi_k(\omega) = \frac{1}{\sqrt{\lambda_k}}\int_0^T X(t;\omega)\psi_k(t)dt, \tag{238}$$

$$\psi_k(t) = \frac{1}{\sqrt{\lambda_k}}\mathbb{E}\{\xi_k(\omega)X(t;\omega)\}. \tag{239}$$

A substitution of (238) into (239) yields the eigenvalue problem

$$\tag{240}$$

$$\int_0^T C(t,s)\psi_k(s)ds = \lambda_k^2\psi_k(t). \tag{241}$$

where

$$C(t,s) = \mathbb{E}\left\{X(t;\omega)X(s;\omega)\right\} \tag{242}$$

is the autocorrelation function of the process. In other words, the KL temporal modes are are eigenfunctions of the the auto-correlation function of the process. Since $C(t,s)$ is a Mercer's kernel (continuous symmetric non-negative definite kernel) we have that $\{\psi_k(t)\}$ is a complete orthonormal basis of $L^2([0,T])$.

Example: Let us compute the KL expansion of a stochastic process with *exponential* auto-correlation function

$$C(t,s) = \frac{\sigma^2}{2\tau}e^{-|t-s|/\tau}, \tag{243}$$

[23] It is straightforward to show that (239) follows form the variational principle

$$\min_{\psi_k} E([\psi_1, \psi_2, \ldots]) = \min_{\psi_k}\int_0^T \mathbb{E}\left\{\left|X(t;\omega) - \sum_{k=1}^{\infty}\sqrt{\lambda_k}\xi_k(\omega)\psi_k(t)\right|^2\right\} dt. \tag{237}$$

where τ denotes the correlation time. Note that (243) is an element of a Dirac delta sequence. This implies that

$$\lim_{\tau\to 0}\frac{\sigma^2}{2\tau}e^{-|t-s|/\tau}=\sigma^2\delta(t-s). \tag{244}$$

The eigenvalue problem (243) with $C(t,s)$ defined in(243) admits the analytical solution[24] (see [**?**])

$$\psi_k(t)=\frac{\tau z_k\cos(z_k t)+\sin(z_k t)}{\sqrt{\frac{1}{2}\left(\tau^2 z_k^2+1\right)T+\left(\tau^2 z_k^2-1\right)\frac{\sin(2z_kT)}{4z_k}+\frac{\tau}{2}\left(1-\cos(2z_kT)\right)}}, \tag{248}$$

where z_k are solution of the transcendental equation

$$\left(z_k^2-\frac{1}{\tau}\right)\tan(z_kT)-\frac{2z_k}{\tau}=0, \tag{249}$$

and

$$\lambda_k=\frac{\sigma^2}{\left(z_k^2\tau^2+1\right)} \tag{250}$$

are the KL eigenvalues. The KL eigenvalues become smaller and smaller as z_k increases. The eigenvalue decay is more pronounced for larger correlation lengths τ, while for very small correlation lengths the eigenvalue decay rate is very small, eventually zero for zero correlation length.

For practical purposes, the KL series expansion (234) is usually truncated to a finite number of terms. As we just discussed, the number of terms is inversely proportional to τ: the smaller τ the larger the number of terms. The number of terms M in the KL series expansion (234) is usually chosen by thresholding the relative "energy" of the process as

$$\frac{\displaystyle\sum_{k=1}^{M}\lambda_k}{\displaystyle\sum_{k=1}^{\infty}\lambda_k}\simeq 0.95. \tag{251}$$

This implies that the modes we retain in the series capture about 95% of the process "energy". In Figure 7 we plot samples of the exponentially correlated Gaussian random process

$$X(t;\omega)=\sin(t)+\frac{\sigma}{2\tau}\sum_{k=1}^{M}\sqrt{\lambda_k}\xi_k(\omega)\psi_k(t)\qquad t\in[0,20], \tag{252}$$

for $\tau=1$ and $\tau=0.1$.

[24] To compute the analytical solution of the KL eigenvalue problem (241) with exponential covariance (243) let us first rewrite it as

$$\int_0^T e^{-c|t-s|}\psi_k(s)ds=\hat{\lambda}_k\psi_k(t),\qquad c=\frac{1}{\tau},\qquad \hat{\lambda}_k=\frac{2\tau}{\sigma^2}\lambda_k. \tag{245}$$

Differentiating with respect to t the equivalent expression

$$\int_0^t e^{-c(t-s)}\psi_k(s)ds+\int_t^T e^{c(t-s)}\psi_k(s)ds=\hat{\lambda}_k\psi_k(t) \tag{246}$$

yields the second-order boundary value problem

$$\begin{cases}\dfrac{d^2\psi_k}{dt^2}=\dfrac{c^2\hat{\lambda}_k-2c}{\hat{\lambda}_k}\psi_k(t)\\[2mm] \dfrac{d\psi_k(t)}{dt}=c\psi(0)\\[2mm] \dfrac{d\psi_k(T)}{dt}=c\psi(T)\end{cases} \tag{247}$$

The solution of the BVP (247) is (248)-(250).

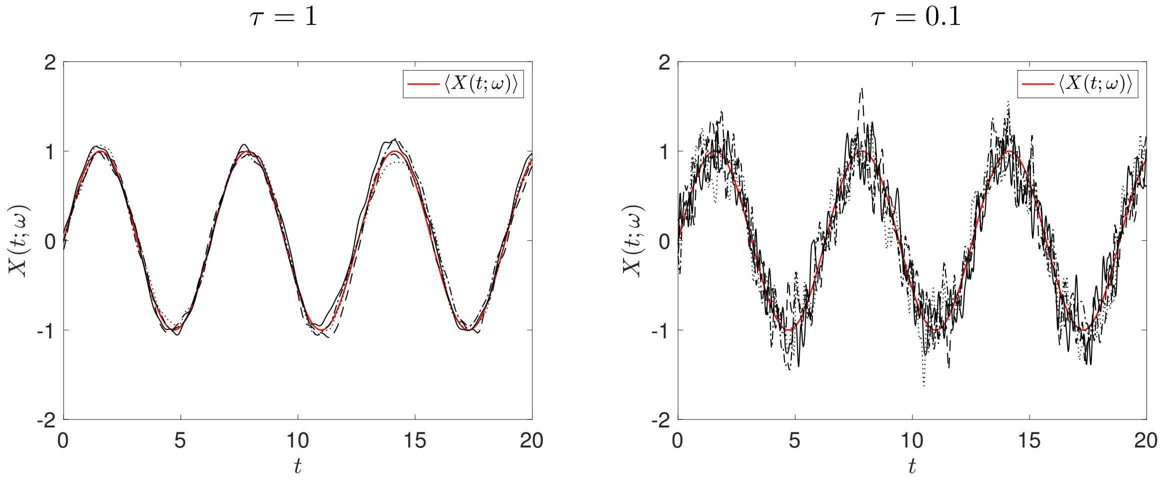


Figure 7: Samples of the exponentially correlated Gaussian random process (252) for different correlation times τ. The KL The mean of the process is shown in red. The truncation threshold for the number of terms M is set at 95% of the energy of the process (see Eq. (251)).

Remark: In the case where (241) cannot be solved analytically, we can resort to numerical method for Fredholm eigenvalue problems, e.g., Finite-difference methods, spectral methods, or Galerkin methods (see e.g., [19]). Of course it is also possible to define KL expansions of *random fields* by simply generalizing the bi-orthogonal series (234) as

$$X(\boldsymbol{x};\omega) = \sum_{k=1}^{\infty} \sqrt{\lambda}_k \xi_k(\omega)\psi_k(\boldsymbol{x}). \tag{253}$$

The computation of the KL expansion follows exactly the same steps as before, i.e., $\psi_k(\boldsymbol{x})$ are solutions to the eigenvalue problem

$$\int_V C(\boldsymbol{x},\boldsymbol{y})\psi_k(\boldsymbol{y})d\boldsymbol{y} = \lambda_k\psi_k(\boldsymbol{x}), \tag{254}$$

where V is some spatial domain.

Remark: To sample realizations of the random process (234) we need to sample the random variables $\{\xi_1,\ldots,xi_M\}$. Such random variables are (by construction) orthonormal (see (235)), i.e., they are uncorrelated and have variance equal to one. Clearly, if $\{\xi_1,\ldots,xi_M\}$ are jointly Gaussian then we know that the condition (235) is necessary and sufficient for independence. Hence, in the Gaussian case, sampling the joint PDF of $\{\xi_1,\ldots,xi_M\}$ reduces to sampling the PDF of an independent set of one-dimensional Gaussian random variables with zero mean and variance one. More generally, if we have available the joint PDF $p(\xi_1,\ldots,xi_M)$, e.g., by computing (238), then we can sample it using Markov Chain Monte Carlo (MCMC) methods, e.g., the Metropolis-Hastings algorithm or Gibbs sampling.

Wiener process. The Wiener process is a zero-mean continuous-time random process satisfying the following conditions:

- The increment $X(t+\tau;\omega) - X(t;\omega)$ is a Gaussian random variable with zero mean and variance τ. In other words, the conditional probability density of $X(t+\tau;\omega)$ given $X(t;\omega)$, is Gaussian with zero mean and variance τ.
- The random variables (increments)

$$X(t_1;\omega) - X(t_0;\omega) \qquad \text{and} \qquad X(t_3;\omega) - X(t_2;\omega) \tag{255}$$

are statistically independent for $t_0 < t_1 \leq t_2 < t_3$. In other words, the Wiener process is an *independent increment* process.

- The process $X(t;\omega)$ is continuous with probability one, i.e.,

$$P\left(\{\omega : \lim_{s\to t} |X(s;\omega) - X(t;\omega)|\}\right) = 1 \qquad \text{for all } t \geq 0. \tag{256}$$

  This means that almost all (except sets of measure zero) sample paths are continuous in the classical sense, but the process $X(t;\omega)$ is nowhere differentiable. Continuity with probability one implies continuity in probability, and therefore mean square continuity and continuity in distribution.

An very clear description of the Wiener process is provided by Wiener himself in [22, Lecture 1]. The simplest algorithm to sample a Wiener process leverages the fact that the process has Gaussian distributed independent increments. Let $\{t_k\}_{k=1,\dots,n}$ be n distinct time instants

$$0 = t_0 < t_1 < \ldots < t_n. \tag{257}$$

Then

$$X(t_k;\omega) = \sum_{j=1}^{k} \sqrt{\Delta t_j}\xi_j(\omega) \qquad \Delta t_j = t_j - t_{j-1}, \tag{258}$$

where $\{\xi_j(\omega)\}$ are independent random variables with mean zero and variance 1. A closer look at (258), reveals

$$X(t_1;\omega) =\sqrt{\Delta t_1}\xi_1(\omega), \tag{259}$$

$$X(t_2;\omega) =X(t_1;\omega) + \sqrt{\Delta t_2}\xi_2(\omega) = \sqrt{\Delta t_1}\xi_1(\omega) + \sqrt{\Delta t_2}\xi_2(\omega) \tag{260}$$

$$\cdots$$

Since $X(t_k;\omega)$ is a superimposition of essentially an infinite number of independent random vairable, it is rather straightforward to show that the one time PDF of $X(t;\omega)$ is

$$p(x,t) = \frac{1}{\sqrt{2\pi t}} e^{-x^2/(2t)}, \tag{261}$$

i.e., Gaussian. This equation also follows from the conditional PDF identity

$$p(x,t) = \int_{\infty}^{\infty} p(x,t|y,s)p(y,s)dy \qquad t > s, \tag{262}$$

where $p(x,t|y,s)$ is the transition density[25], and $p(y,s)$ is the PDF of $X(s;\omega)$. If we set $s = 0$ then $p(y,0) = \delta(y)$ and, of course, this yields (261). The auto-correlation function of the Wiener process is

$$C(t,s) = \min(t,s) \tag{264}$$

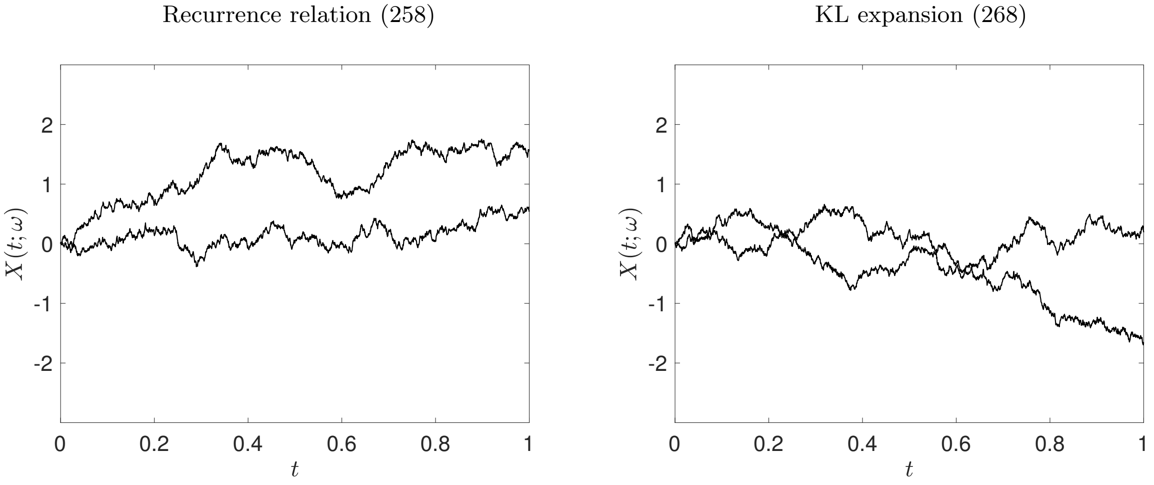
With the autocorrelation function available we can compute a KL expansion of the Wiener process following the procedure outlined in the previous section. If we consider the time interval $[0,1]$ this yields the eigenvalue problem

$$\int_0^1 C(t,s)\psi_k(s) = \lambda_k \psi_k(t), \tag{265}$$

[25] From the recurrence relation

$$X(t_{k+1};\omega) = X(t_k;\omega) + \sqrt{\Delta t_{k+1}}\xi_{k+1}(\omega) \tag{263}$$

with $\xi_{k+1}(\omega)$ Gaussian with zero mean and variance one we see that the conditional PDF $p(x,t|y,s)$ is Gaussian with zero mean and variance $t-s$.


Recurrence relation (258)

KL expansion (268)

Figure 8: Wiener processes obtained by sampling the Karhunen-Loève expansion (268) with 10^5 terms on a temporal grid with 2000 points in $[0, 1]$, and by iterating (258) on the same temporal grid.

the solution of which is

$$\psi_k(t) = \sqrt{2}\sin\left(\left[k - \frac{1}{2}\right]\pi t\right) \qquad k = 1, 2, \ldots \tag{266}$$

and

$$\lambda_k = \frac{4}{\pi^2 (2k-1)^2}. \tag{267}$$

Substituting (266) and (267) into (234) yields

$$X(t;\omega) = \sum_{k=1}^{\infty} \frac{2\sqrt{2}}{\pi (2k-1)} \xi_k(\omega) \sin\left(\left[k - \frac{1}{2}\right]\pi t\right), \tag{268}$$

where $\xi_k(\omega)$ are independent Gaussian random variables with zero mean and variance one (they satisfy (235)). The series expansion in (268) can be eventually truncated to a finite number of terms, depending on the threshold set on the eigenvalues (267) (which decay as $1/k$). In Figure 8 we plot a few samples of the Wiener process we obtain by sampling (268) with 10^5 terms on a temporal grid with 2000 points in $[0, 1]$, and the Wiener process we obtain by iterating (258) on the same temporal grid. Note that if $X(t;\omega)$ is a Wiener process in $t \in [0, 1]$ then

$$\sqrt{T} X\left(\frac{t}{T};\omega\right) \qquad t \in [0, T] \tag{269}$$

is a Wiener process in $[0, T]$. This expression is obtained by simply changing the variables in the integral equation (265). The expression (269) shows that features of a Wiener process do not change while zooming in or out. In other words, the Wiener process is self-similar.

Appendix A: Derivation of the Liouville equation

Consider the nonlinear dynamical system

$$\begin{cases} \dfrac{d\boldsymbol{x}(t)}{dt} = \boldsymbol{f}(\boldsymbol{x}(t)) \\ \boldsymbol{x}(0) = \boldsymbol{x}_0(\omega) \end{cases} \tag{270}$$

where $\boldsymbol{x}_0(\omega)$ is a random vector with known joint probability density function $p_0(\boldsymbol{x})$. We know that if $\boldsymbol{f}(\boldsymbol{x})$ is continuously differentiable in $\boldsymbol{x}$ then (270) admits a smooth flow $\boldsymbol{x}(t, \boldsymbol{x}_0(\omega))$, which is at least continuously differentiable in $\boldsymbol{x}_0$. The flow is also continuously differentiable in t, i.e., $x(t, \boldsymbol{x}_0(\omega))$ is a diffeomorphism in t. We are interested in determining an evolution equation for $p(\boldsymbol{x}, t)$, i.e., the probability density function of $\boldsymbol{x}(t, \boldsymbol{x}_0)$ at time t. To this end, consider the characteristic function representation of the PDF $p(\boldsymbol{x}, t)$

$$\phi(\boldsymbol{a}, t) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{i\boldsymbol{a}\cdot\boldsymbol{x}(t;\boldsymbol{x}_0)} p(\boldsymbol{x}_0) d\boldsymbol{x}_0 = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{i\boldsymbol{a}\cdot\boldsymbol{x}} p(\boldsymbol{x}, t) d\boldsymbol{x} \tag{271}$$

Differentiating with respect to t yields

$$\begin{aligned}
\frac{\partial \phi(\boldsymbol{a}, t)}{\partial t} &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} i\boldsymbol{a} \cdot \frac{\partial \boldsymbol{x}(t, \boldsymbol{x}_0)}{\partial t} e^{i\boldsymbol{a}\cdot\boldsymbol{x}(t;\boldsymbol{x}_0)} p(\boldsymbol{x}_0) d\boldsymbol{x}_0 \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} i\boldsymbol{a} \cdot \boldsymbol{f}\left(\boldsymbol{x}(t, \boldsymbol{x}_0)\right) e^{i\boldsymbol{a}\cdot\boldsymbol{x}(t;\boldsymbol{x}_0)} p(\boldsymbol{x}_0) d\boldsymbol{x}_0 \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} i\boldsymbol{a} \cdot \boldsymbol{f}\left(\boldsymbol{x}\right) e^{i\boldsymbol{a}\cdot\boldsymbol{x}} p(\boldsymbol{x}, t) d\boldsymbol{x} \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{\partial}{\partial \boldsymbol{x}} \left(e^{i\boldsymbol{a}\cdot\boldsymbol{x}}\right) \cdot \boldsymbol{f}\left(\boldsymbol{x}\right) p(\boldsymbol{x}, t) d\boldsymbol{x} \\
&= -\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{i\boldsymbol{a}\cdot\boldsymbol{x}} \nabla \cdot \left(\boldsymbol{f}\left(\boldsymbol{x}\right) p(\boldsymbol{x}, t)\right) d\boldsymbol{x}. \quad \text{(integrating by parts)}
\end{aligned} \tag{272}$$

By using (271) and (272) we obtain

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{i\boldsymbol{a}\cdot\boldsymbol{x}} \left[\frac{\partial p(\boldsymbol{x}, t)}{\partial t} + \nabla \cdot \left(\boldsymbol{f}\left(\boldsymbol{x}\right) p(\boldsymbol{x}, t)\right)\right] d\boldsymbol{x} = 0, \quad \text{for all } \boldsymbol{a} \in \mathbb{R}^n, \tag{273}$$

which implies that the function between square bracket must be equal to zero for all $\boldsymbol{x}$ and all t, i.e.,

$$\frac{\partial p(\boldsymbol{x}, t)}{\partial t} + \nabla \cdot \left(\boldsymbol{f}\left(\boldsymbol{x}\right) p(\boldsymbol{x}, t)\right) = 0 \qquad \text{(Liouville equation).} \tag{274}$$

References

[1] M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein GAN. *arXiv:1701.07875*, pages 1–30, 2017.

[2] M. J. Beran. *Statistical continuum theories.* New York: Interscience Publishers, 1968.

[3] D. Bigoni, A. P. Engsig-Karup, and Y. M. Marzouk. Spectral tensor-train decomposition. *SIAM J. Sci. Comput.*, 38(4):A2405–A2439, 2016.

[4] Z. I. Botev, J. F. Grotowski, and D. P. Kroese. Kernel density estimation via diffusion. *Annals of Statistics*, 38(5):2916–2957, 2010.

[5] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel. *Time Series Analysis: Forecasting and Control.* Wiley, 2008.

[6] S. Brooks, A. Gelman, G. Jones, and X.-L. Meng. *Handbook of Markov Chain Monte Carlo.* Chapman & Hall/CRC, 2011.

[7] A. Dektor, A. Rodgers, and D. Venturi. Rank-adaptive tensor methods for high-dimensional nonlinear pdes. *Journal of Scientific Computing*, 88(36):1–27, 2021.

[8] A. Dektor and D. Venturi. Dynamically orthogonal tensor methods for high-dimensional nonlinear PDEs. *J. Comput. Phys.*, 404:109125, 2020.

[9] G. B. Folland. *Real Analysis: Modern Techniques and Their Applications.* Wiley, second edition, 2007.

[10] A. I. Khuri. Applications of Dirac's delta function in statistics. *Int. J. Math. Educ. Sci. Technol.*, 35(2):185–195, 2004.

[11] V. I. Klyatskin. *Dynamics of stochastic systems.* Elsevier Publishing Company, 2005.

[12] R. Kubo. Generalized cumulant expansion method. *Journal of the Physical Society of Japan*, 17(7):1100–1120, 1962.

[13] A. S. Monin and A. M. Yaglom. *Statistical Fluid Mechanics, Volume II: Mechanics of Turbulence.* Dover, 2007.

[14] I. V. Oseledets. Tensor-train decomposition. *SIAM J. Sci. Comput.*, 33(5):2295—-2317, 2011.

[15] A. Papoulis. *Probability, random variables and stochastic processes.* McGraw-Hill, third edition, 1991.

[16] M. Raissi, P. Perdikaris, and G. E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.*, 378:606–707, 2019.

[17] H. Rue and L. Held. *Gaussian Markov random fields.* Chapman & Hall/CRC, 2005.

[18] D. Venturi. The numerical approximation of nonlinear functionals and functional differential equations. *Physics Reports*, 732:1–102, 2018.

[19] D. Venturi, M. Choi, and G. E. Karniadakis. Supercritical quasi-conduction states in stochastic Rayleigh-Bénard convection. *Int. J. Heat and Mass Transfer*, 55(13-14):3732–3743, 2012.

[20] D. Venturi and A. Dektor. Spectral methods for nonlinear functionals and functional differential equations. *Res. Math. Sci.*, 8(27):1–39, 2021.

[21] D. Venturi and X. Li. The Mori-Zwanzig formulation of deep learning. *ArXiv*, (2209.05544):1–40, 2022.

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

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

PDF equations for random dynamical systems

Consider the following n-dimensional dynamical system

$$\begin{cases} \dfrac{d\boldsymbol{x}}{dt} = \boldsymbol{f}(\boldsymbol{x}) \\ \boldsymbol{x}(0;\omega) = \boldsymbol{x}_0(\omega) \end{cases} \tag{1}$$

where $\boldsymbol{x}_0(\omega)$ is a random initial state with joint PDF $p_0(\boldsymbol{x})$. We are interested in studying the statistical properties of the solution to (1) using probability density function (PDF) methods. As we shall see hereafter, systems of the form (1) include systems in which we have random variables at appearing the right hand side of the ODE, i.e., systems with random parameters.

Systems with random parameters. It is straightforward to show that a non-autonomous system of the form

$$\begin{cases} \dfrac{d\boldsymbol{x}}{dt} = \boldsymbol{G}(\boldsymbol{x}, \boldsymbol{\xi}(\omega), t) \\ \boldsymbol{x}(0;\omega) = \boldsymbol{x}_0(\omega) \end{cases} \tag{2}$$

can be transformed into an autonomous system evolving form a random initial state. To this end, we define the phase variables of $z(t) = t$ and $\boldsymbol{y}(t) = \boldsymbol{\xi}(\omega)$ rewrite (2) as

$$\begin{cases} \dfrac{d\boldsymbol{x}}{dt} = \boldsymbol{G}(\boldsymbol{x}, \boldsymbol{y}, z) \\ \dfrac{d\boldsymbol{y}}{dt} = \boldsymbol{0} \\ \dfrac{dz}{dt} = 1 \\ \boldsymbol{x}(0;\omega) = \boldsymbol{x}_0(\omega), \qquad \boldsymbol{y}(0;\omega) = \boldsymbol{\xi}(\omega), \qquad z(0,\omega) = 0. \end{cases} \tag{3}$$

Remarkably, system of the form (2) include also dynamical systems driven by finite-dimensional random processes, i.e., random processes that can be represented in terms of series expansions involving a finite number of random variables.

Example: An simple example of a system of the form (2) is

$$\frac{dx}{dt} = f(x) + \eta(t;\omega) \tag{4}$$

i.e., a scalar ODE driven by *colored random noise* $\eta(t;\omega)$ [12, 21, 17]. Let us represent $\eta(t;\omega)$ as a truncated Karhunen-Loève series expansion (see [21] for an application to cancer modeling)

$$\eta(t;\omega) \simeq \sum_{k=1}^{M} \sqrt{\lambda}_k \xi_k(\omega)\psi_k(t) \tag{5}$$

involving a finite number of uncorrelated random variables $\{\xi_1, \ldots, \xi_M\}$. We shall call M the *dimensionality* of the noise process[1] The adjective "colored" refers to the fact that the Fourier power spectral density of the random noise $\eta(t;\omega)$ is in general not flat as in the case of white noise[2]. The power spectral density is the inverse Fourier transform of the temporal auto-correlation function of the noise, i.e.,

$$\mathbb{E}\left\{f(t;\omega)f(t';\omega)\right\}=\sum_{k=1}^{M}\lambda_k\psi_k(t)\psi_k(t'). \tag{7}$$

Remark: A random dynamical systems is a systems driven by a finite number of random variables. An example is the system (4)-(5), in which the the random input process is finite-dimensional (M finite). On the other hand, a "stochastic dynamical system" is usually driven by infinite-dimensional random processes, i.e., processes that can be represented in terms of an infinite (countable or uncountable) number of random variables. An example is the ODE (4) if we choose $\eta(t;\omega)$ to be, e.g., Gaussian white noise process (derivative of a Wiener process). In this case, it is more appropriate to write the ODE as

$$dx=f(x)dt+d\zeta(t), \tag{8}$$

where $d\zeta(t)$ denotes the increment of a Wiener process.

Liouville equation. Let $\boldsymbol{x}(t;\boldsymbol{x}_0)$ be the flow generated by (1). The PDF of $\boldsymbol{x}(t;\boldsymbol{x}_0)$, i.e., the solution of (1) at time t, satisfies the Liouville equation

$$\frac{\partial p(\boldsymbol{x},t)}{\partial t}+\nabla\cdot(\boldsymbol{f}(\boldsymbol{x})p(\boldsymbol{x},t))=0, \qquad p(\boldsymbol{x},0)=p_0(\boldsymbol{x}), \tag{9}$$

where $p_0(x)$ is the PDF of the random initial state $\boldsymbol{x}_0(\omega)$. To derive equation (9), consider the characteristic function representation of $p(\boldsymbol{x},t)$

$$\phi(\boldsymbol{a},t)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}e^{i\boldsymbol{a}\cdot\boldsymbol{x}}p(\boldsymbol{x},t)d\boldsymbol{x}=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}e^{i\boldsymbol{a}\cdot\boldsymbol{x}(t;\boldsymbol{x}_0)}p(\boldsymbol{x}_0)d\boldsymbol{x}_0. \tag{10}$$

Differentiating (10) with respect to t yields

$$\begin{aligned}\frac{\partial\phi(\boldsymbol{a},t)}{\partial t}&=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}i\boldsymbol{a}\cdot\frac{\partial\boldsymbol{x}(t,\boldsymbol{x}_0)}{\partial t}e^{i\boldsymbol{a}\cdot\boldsymbol{x}(t;\boldsymbol{x}_0)}p(\boldsymbol{x}_0)d\boldsymbol{x}_0\\&=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}i\boldsymbol{a}\cdot\boldsymbol{f}\left(\boldsymbol{x}(t,\boldsymbol{x}_0)\right)e^{i\boldsymbol{a}\cdot\boldsymbol{x}(t;\boldsymbol{x}_0)}p(\boldsymbol{x}_0)d\boldsymbol{x}_0\\&=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}i\boldsymbol{a}\cdot\boldsymbol{f}\left(\boldsymbol{x}\right)e^{i\boldsymbol{a}\cdot\boldsymbol{x}}p(\boldsymbol{x},t)d\boldsymbol{x}\\&=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\frac{\partial}{\partial\boldsymbol{x}}\left(e^{i\boldsymbol{a}\cdot\boldsymbol{x}}\right)\cdot\boldsymbol{f}\left(\boldsymbol{x}\right)p(\boldsymbol{x},t)d\boldsymbol{x}\\&=-\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}e^{i\boldsymbol{a}\cdot\boldsymbol{x}}\nabla\cdot\left(\boldsymbol{f}\left(\boldsymbol{x}\right)p(\boldsymbol{x},t)\right)d\boldsymbol{x}.\end{aligned} \tag{11}$$

[1] To sample realizations of the random process (5) we need to sample the random variables $\{\xi_1,\ldots,xi_M\}$. Such random variables are (by construction) orthonormal, i.e., they are uncorrelated and have variance one:

$$\mathbb{E}\{\xi_i(\omega)\xi_j(\omega)\}=\delta_{ij}. \tag{6}$$

If $\{\xi_1,\ldots,xi_M\}$ are jointly Gaussian then we know that (6) is necessary and sufficient for independence. Hence, in this case sampling the joint PDF $\{\xi_1,\ldots,xi_M\}$ reduces to sampling the PDF of an independent set of one-dimensional Gaussian PDFs with zero mean and variance one. More generally, the joint PDF of $\{\xi_1,\ldots,xi_M\}$ can be sampled using Markov Chain Monte Carlo methods, e.g., the Metropolis-Hastings algorithm or the Gibbs sampling algorithm.

[2] A flat power spectral density implies that all frequencies contribute equally to the signal. The adjective "white" follows from an analogy the power spectrum of visible colors, in which the color white has all visible frequencies contributing equally. Stochastic ODEs driven by Gaussian white noise, and corresponding models are discussed extensively in the course AM216. If the power spectral density of a random signal decays with the frequency ν as $1/\nu^{\alpha}$ ($\alpha\in[1,2]$) then the noise is called "pink".

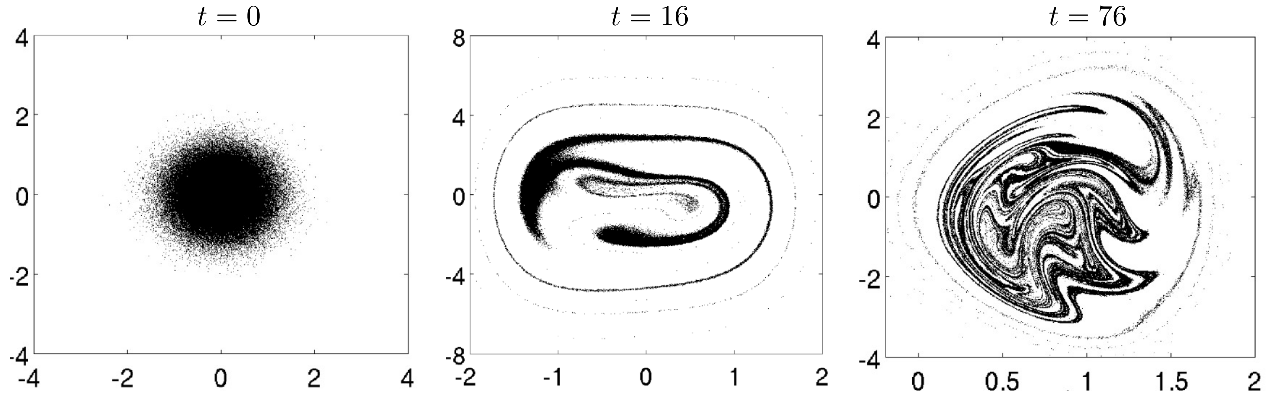


Figure 1: Point clouds corresponding to a jointly Gaussian initial PDF advected by the flow map generated by the Duffing equation. The x-axis corresponds to x while the y-axis represents $\dot{x}$. We plot the joint PDF of $x(t)$ and $\dot{x}(t)$ at different times.

In the last step we used integration by parts and the fact that the PDF $p(\boldsymbol{x}, t)$ decays to zero at infinity sufficiently fast. By combining (10) and (11) we obtain

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{i\boldsymbol{a}\cdot\boldsymbol{x}} \left[\frac{\partial p(\boldsymbol{x}, t)}{\partial t} + \nabla \cdot (\boldsymbol{f}(\boldsymbol{x})\, p(\boldsymbol{x}, t))\right] d\boldsymbol{x} = 0, \quad \text{for all } \boldsymbol{a} \in \mathbb{R}^n, \tag{12}$$

which implies that the function between square bracket must be equal to zero for all $\boldsymbol{x}$ and all t. This proves the Liouville equation (9).

Note that from a mathematical viewpoint, the Liouville equation (9) is a linear hyperbolic conservation law in as many variables as the dimension of the system (1). Therefore, computing its solution can be challenging due to high-dimensionality (PDE in n independent variables), normalization and positivity constraints of the solution (the solution is a PDF), as well as potential multiple scales (The PDF is a hyperbolic conservation law). Related to the last point, in Figure 1 we show what happens to a jointly Gaussian initial state when samples from such PDF are evolved forward in time by the flow map generated by the 2D Duffing oscillator [1]

$$\frac{d^2x}{dt^2} = -x - \frac{1}{50}\frac{dx}{dt} - 5x^3 + 8\cos\left(\frac{t}{2}\right). \tag{13}$$

By using the method of characteristics, it is straightforward to obtain the following formal solution to the Liouville equation (9)

$$p(\boldsymbol{x}, t) = p_0\left(\boldsymbol{x}_0(\boldsymbol{x}, t)\right) \exp\left(-\int_0^t \nabla \cdot \boldsymbol{f}\left(\boldsymbol{x}(\tau, \boldsymbol{x}_0)\right) d\tau\right), \tag{14}$$

where $\boldsymbol{x}_0(\boldsymbol{x}, t)$ denotes the inverse flow map generated by (1). Equation (14) follows from the well-known characteristic system[3]

$$\begin{cases} \dfrac{d\boldsymbol{x}(t,\boldsymbol{x}_0)}{dt} = \boldsymbol{f}\left(\boldsymbol{x}(t,\boldsymbol{x}_0)\right) \\ \boldsymbol{x}(0,\boldsymbol{x}_0) = \boldsymbol{x}_0 \\ \dfrac{dp(\boldsymbol{x}(t,\boldsymbol{x}_0),t)}{dt} = -p(\boldsymbol{x}(t,\boldsymbol{x}_0),t)\nabla\cdot\boldsymbol{f}\left(\boldsymbol{x}(t,\boldsymbol{x}_0)\right) \\ p(\boldsymbol{x}(0,\boldsymbol{x}_0),0) = p_0(\boldsymbol{x}_0) \end{cases} \tag{16}$$

Example: The Liouville equation corresponding to the system (4)-(5) is

$$\frac{\partial p(x,\boldsymbol{y},t)}{\partial t} + \frac{\partial}{\partial x}\left(f(x)p(x,\boldsymbol{y},t)\right) + \frac{\partial p(x,\boldsymbol{y},t)}{\partial x}\sum_{k=1}^{M}\sqrt{\lambda_k}y_k\psi_k(t) = 0. \tag{17}$$

If $x_0(\omega)$ and $\boldsymbol{\xi}$ are statistically independent then the initial PDF can be factored as $p(x_0,\boldsymbol{y},0) = p_{x_0}(x_0)p_{\boldsymbol{\xi}}(\boldsymbol{y})$. It is important to emphasize that the joint PDF equation involves both the state variable $x(t,\omega)$ and the variables y_k representing the variables ξ_k in the noise process (5).

Example: The Liouville equation corresponding to the three-dimensional dynamical system

$$\dot{x}_1 = x_1x_3. \qquad \dot{x}_2 = -x_2x_3, \qquad \dot{x}_3 = -x_1^2 + x_2^2. \tag{18}$$

is

$$\frac{\partial p(\boldsymbol{x},t)}{\partial t} = -\frac{\partial}{\partial x_1}\left(x_1x_3p(\boldsymbol{x},t)\right) + \frac{\partial}{\partial x_2}\left(x_2x_3p(\boldsymbol{x},t)\right) + \frac{\partial}{\partial x_3}\left((x_1^2 - x_2^2)p(\boldsymbol{x},t)\right). \tag{19}$$

Reduced-order PDF equations for dynamical systems. The Liouville equation (9) describes the exact dynamics of the joint PDF of state variables $\boldsymbol{x}(t)$. In most cases, however, we are only interested in a smaller subset of such variables, e.g., in the scalar quantity of interest

$$z(t,\omega) = u(\boldsymbol{x}(t,\boldsymbol{x}_0(\omega))) \qquad \text{(phase space function)}. \tag{20}$$

We have seen that the probability density function of such phase space function can be written as

$$p(z,t) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\delta\left(z - u(\boldsymbol{x})\right)p(\boldsymbol{x},t)d\boldsymbol{x} = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\delta\left(z - u(\boldsymbol{x}(t,\boldsymbol{x}_0))\right)p(\boldsymbol{x}_0)d\boldsymbol{x}_0, \tag{21}$$

where $\delta(\cdot)$ is the Dirac's delta function (see [8, 20, 14]) and z is the phase space variable representing $u(\boldsymbol{x}(t))$. Multiplying the Liouville equation (9) by $\delta\left(z - u(\boldsymbol{x})\right)$ and integrating over all phase variables yields

$$\frac{\partial p(z,t)}{\partial t} + \frac{1}{2\pi}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}e^{ia(z-u(\boldsymbol{x}))}\nabla\cdot\left(\boldsymbol{f}(\boldsymbol{x})p(\boldsymbol{x},t)\right)d\boldsymbol{x}da = 0, \tag{22}$$

[3] Note that (9) can be written as

$$\frac{\partial p(\boldsymbol{x},t)}{\partial t} + \boldsymbol{f}(\boldsymbol{x})\cdot\nabla p(\boldsymbol{x},t) = -p(\boldsymbol{x},t)\nabla\cdot\boldsymbol{f}(\boldsymbol{x}) \qquad p(\boldsymbol{x},0) = p_0(\boldsymbol{x}). \tag{15}$$

Applying the method of characteristics to (15) yields the ODE system (16). In practice, the PDF $p(\boldsymbol{x},t)$ is computed using (16) along each characteristic curve. Clearly, this is computationally challenging in high-dimensions.

$$p(x_1,t) = \int_{-\infty}^{\infty} p(x_1,x_2,t)dx_2$$

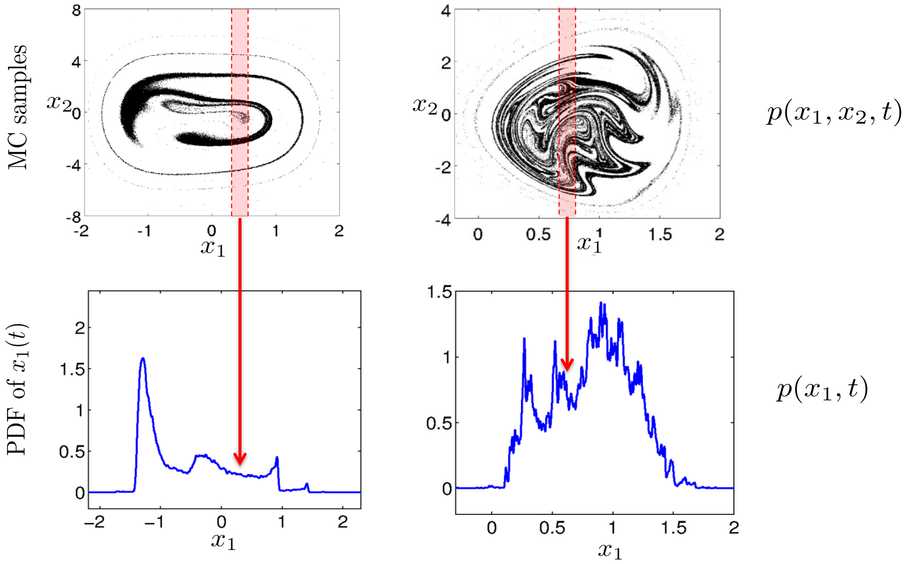


Figure 2: Regularization of PDFs by integration/marginalization. The PDF of $x_1(t)$ at one specific location is obtained by summing up the probability mass within the strips highlighted in red. The figures at the top represent the joint PDF of x_1 and x_2, i.e., x and $\dot{x}$ in the Duffing equation (13) at different times (see also Figure 1).

where we used the Fourier representation of the Dirac delta function $\delta\left(z-u(\boldsymbol{x})\right)$. In general, equation (22) is *unclosed* in the sense that there are terms at the right hand side that cannot be represented or computed based on $p(z,t)$ alone. If we set $u(\boldsymbol{x}(t)) = x_k(t)$, i.e., the quantity of interest is the k-th component of the dynamical system (1), then (22) reduces to

$$\frac{\partial p(x_k,t)}{\partial t} + \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\frac{\partial}{\partial x_k}\left(f_k(\boldsymbol{x})p(\boldsymbol{x},t)\right)dx_1\ldots dx_{k-1}dx_{k+1}\ldots dx_N = 0. \tag{23}$$

The specific form of this equation depends on the vector field $\boldsymbol{f}(\boldsymbol{x})$.

Remark: Low-dimensional marginals of high-dimensional PDF are usually smoother functions than the original PDF. This is illustrated in Figure 2 with reference to the Duffing equation (13). Hence, deriving and solving low-dimensional PDF equations for quantities of interest, has advantages relative to full Liouville equation. In particular: 1) the PDF equation for the quantity of interest is low dimensional, 2) we expect the solution to a reduced-order PDF equations to relatively smooth because of the “regularization by integration” effect.

BBGKY hierarchy. By integrating the Liouville equation (9) with respect to different phase variables it is possible to derive a hierarchy of PDEs known as *Bogoliubov-Born-Green-Kirkwood-Yvon (BBGKY) hierarchy* involving PDFs with an increasing number of phase variables. The first set of PDEs is (23), and it clearly depends on PDFs with a larger number of variables, unless $f_k(\boldsymbol{x})$ depends only on x_k (in which case the system (1) is uncoupled). Hereafter we provide specific examples of BBGKY hierarchies.

Example: Consider the Kraichnan-Orszag three-mode problem [13, 24]

$$\dot{x}_1 = x_1 x_3. \qquad \dot{x}_2 = -x_2 x_3, \qquad \dot{x}_3 = -x_1^2 + x_2^2. \tag{24}$$

The associated Liouville equation is

$$\frac{\partial p(\boldsymbol{x},t)}{\partial t} = -\frac{\partial}{\partial x_1}\left(x_1 x_3 p(\boldsymbol{x},t)\right) + \frac{\partial}{\partial x_2}\left(x_2 x_3 p(\boldsymbol{x},t)\right) + \frac{\partial}{\partial x_3}\left((x_1^2 - x_2^2) p(\boldsymbol{x},t)\right). \tag{25}$$

Suppose we are interested in the PDF of the first component of the system, i.e., set $u(\boldsymbol{x}(t)) = x_1(t)$ in equation (20). By integrating (25) with respect to x_2 and x_3, and assuming that $p(\boldsymbol{x},t)$ decays fast enough at infinity, we obtain

$$\frac{\partial p(x_1,t)}{\partial t} = -\frac{\partial}{\partial x_1}\int_{-\infty}^{\infty} x_1 x_3 p(x_1,x_3,t) dx_3. \tag{26}$$

From this equation we see that the evolution of $p(x_1,t)$ depends on an integral involving $p(x_1,x_3,t)$. Hence, to compute $p(x_1,t)$ we need to know what $p(x_1,x_3,t)$ is. The evolution equation for $p(x_1,x_3,t)$ can be obtained by integrating (25) with respect to x_2, i.e.,

$$\frac{\partial p(x_1,x_3,t)}{\partial t} = -\frac{\partial}{\partial x_1}\left(x_1 x_3 p(x_1,x_3,t)\right) + x_1^2 \frac{\partial}{\partial x_3}\left(x_3 p(x_1,x_3,t)\right) - \frac{\partial}{\partial x_3}\int_{-\infty}^{\infty} x_2^2 p(x_1,x_2,x_3,t) dx_2. \tag{27}$$

The PDE system (26)-(27) represents the first two levels of a BBGKY hierarchy. Note that the hierarchy be closed only at the level of the Liouville equation (25). Indeed, the integral at the right hand side of (27) involves $p(x_1,x_2,x_3,t)$, which is unknown unless we solve (25).

At this point we notice that we can represent the term involving $p(x_1,x_3,t)$ in (26) in a different way. Specifically, we can write the joint PDF of $x_1(t)$ and $x_3(t)$ at time t as

$$p(x_1,x_3,t) = p(x_1,t) p(x_3|x_1,t), \tag{28}$$

where $p(x_3|x_1,t)$ is the conditional probability density of $x_3(t)$ given $x_1(t)$. A substitution of (28) into (26) yields

$$\frac{\partial p(x_1,t)}{\partial t} = -\frac{\partial}{\partial x_1}\left(x_1 p(x_1,t)\mathbb{E}[x_3(t)|x_1(t)]\right), \tag{29}$$

where

$$\mathbb{E}[x_3(t)|x_1(t)] = \int_{-\infty}^{\infty} x_3 p(x_3|x_1,t) dx_3 \qquad \text{(conditional expectation of } x_3(t) \text{ given } x_1(t)). \tag{30}$$

As we shall see hereafter, $\mathbb{E}[x_3(t)|x_1(t)]$ can be estimated from sample trajectories of (24).

Note that the reduced-order PDF equation (29) is a scalar conservation law where the (compressible) advection velocity field is equal to $x_1\mathbb{E}[x_3(t)|x_1(t)]$. It is important to emphasize that the "innocent-looking" PDE (26) is actually a PDE involving derivatives of $p(x_1,t)$ up to order *infinity* in the phase variable x_1. In fact, by using Kubo's cumulant expansion [9] of the joint characteristic function of $x_3(t)$ and $x_1(t)$ (i.e. Eq. (156) in lecture notes 1)

$$\phi(a_1,a_3,t) = \phi(a_1,t)\phi(a_3,t)\exp\left[\sum_{j,k=1}^{\infty}\left\langle x_1^j(t) x_3^k(t)\right\rangle_c \frac{(ia_1)^j (ia_3)^k}{j!k!}\right], \tag{31}$$

and the correspondence

$$p(x_1,x_3,t) = \frac{1}{(2\pi)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-i(a_1x_1+a_3x_3)}\phi(a_1,a_3,t) da_1 da_3 \tag{32}$$

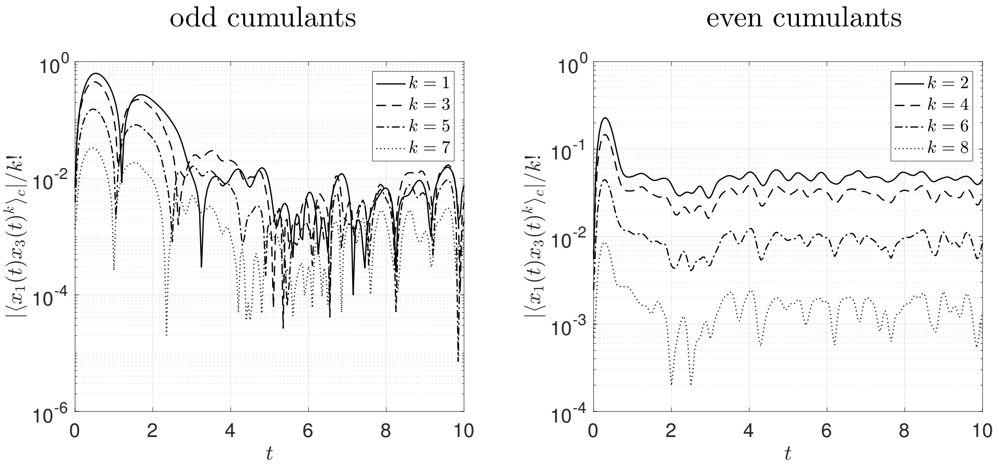


Figure 3: Kraichnan-Orszag three mode problem. Absolute values of the first 8 rescaled cumulants $\langle x_1(t)x_3(t)^k\rangle_c/k!$. The initial condition $x_i(0)$ $(i = 1, 2, 3)$ in (24) is set to be i.i.d. Gaussian with mean and variance 1. We estimated the cumulants numerically by using Monte Carlo (50000 sample paths) and ensemble averages. It is seen that the odd cumulants decay slowly with k, suggesting that the cumulant expansion (33) cannot be truncated at low order. This implies that any reasonably accurate approximation of the reduced-order equation (35) involves high-order derivatives of $p(x_1, t)$ with respect to x_1.

we can prove that

$$\int_{-\infty}^{\infty} x_3 p(x_3, x_1, t)dx_3 = \mathbb{E}[x_3(t)]p(x_1, t) + \sum_{k=1}^{\infty}(-1)^{k+1}\frac{\langle x_1(t)x_3(t)^k\rangle_c}{k!}\frac{\partial^k p(x_1, t)}{\partial x_1^k}, \tag{33}$$

where $\langle x_1(t)x_3(t)^k\rangle_c$ are cumulant averages[4]. A substitution of (33) into (26) yields the infinite-order PDE

$$\frac{\partial p(x_1, t)}{\partial t} = -\mathbb{E}[x_3(t)]\frac{\partial\left(x_1 p(x_1, t)\right)}{\partial x_1} + \sum_{k=1}^{\infty}(-1)^{k+1}\frac{\langle x_1(t)x_3(t)^k\rangle_c}{k!}\frac{\partial^{k+1}\left(x_1 p(x_1, t)\right)}{\partial x_1^{k+1}}. \tag{35}$$

As shown in Figure 3, the rescaled cumulants $\langle x_1(t)x_3(t)^k\rangle_c/k!$ decay slowly with k, suggesting that the cumulant expansion (33) cannot be truncated at low-order. This implies that any reasonably accurate approximation of the reduced-order PDF equation (35) involves high-order derivatives of $p(x_1, t)$ with respect to x_1. The data-driven cumulant expansion approach we just described relies on computing sample paths of (24), estimating the cumulant averages $\langle x_1(t)x_3(t)^k\rangle_c$ using ensemble averaging, and then solving the PDE (35). Clearly this is not practical since such PDE potentially involves high-order derivatives of $p(x_1, t)$ with respect to x_1. Another approach relies on estimating the conditional expectation (30) directly from data and then solving the hyperbolic conservation law (29), which is a first-order linear PDE.

Example: Consider the following N-dimensional nonlinear dynamical system

$$\frac{dx_i}{dt} = -\sin(x_{i+1})x_i - Ax_i + F, \qquad i = 1, ..., N, \tag{36}$$

where $x_{N+1}(t) = x_1(t)$ (periodic boundary conditions). Depending on the value of F, A and on the number of phase variables N, this system can exhibit different behaviors. In Figure 4 we plot a 2D section of the

[4] The cumulant averages appearing in equation (33) are defined as

$$\langle x_1(t)x_3(t)^k\rangle_c = \mathbb{E}[x_1(t)x_3(t)^k] - \mathbb{E}\left[x_1(t)\right]\mathbb{E}[x_3(t)^k]. \tag{34}$$

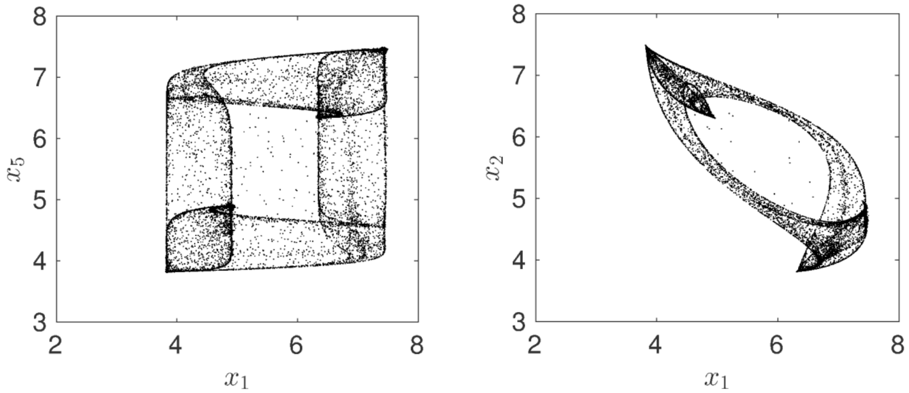


Figure 4: 2D sections of the point clouds generated by the dynamical system (36) at $t = 14$. The random initial condition samples are taken from a Gaussian distribution.

point cloud we obtain at $t = 14$ by sampling the initial condition from a Gaussian random vector. Here we set $F = 10$, $A = 0.2$ and $N = 1000$. The Liouville transport equation associated with (36) is

$$\frac{\partial p(\boldsymbol{x},t)}{\partial t} = -\sum_{i=1}^{N} \frac{\partial}{\partial x_i} \left[(F - \sin(x_{i+1})x_i - Ax_i)\, p(\boldsymbol{x},t) \right]. \tag{37}$$

This PDF is very hard to solve numerically because of the very high number of phase variables. The evolution equation for the PDF of each phase variable $x_i(t)$ can be obtained by integrating (37) with respect to all other variables. This yields the unclosed equation

$$\frac{\partial p(x_i,t)}{\partial t} = -\frac{\partial}{\partial x_i} \int_{-\infty}^{\infty} \left[(F - \sin(x_{i+1})x_i - Ax_i)\, p(x_i, x_{i+1}, t) \right] dx_{i+1}. \tag{38}$$

We can write (38) equivalently as

$$\frac{\partial p(x_i,t)}{\partial t} = -F\frac{\partial p(x_i,t)}{\partial x_i} + A\frac{\partial (x_i p(x_i,t))}{\partial x_i} - \frac{\partial}{\partial x_i} x_i \int_{-\infty}^{\infty} \sin(x_{i+1}) p(x_i, x_{i+1}, t) dx_{i+1}. \tag{39}$$

Note that all equations for $p(x_i,t)$ have the same structure, independently of the index i. This means that if the random initial state $\boldsymbol{x}_0$ has i.i.d. components, then the evolution of each $p(x_i,t)$ does not depend on i, i.e., it is the same for all $i = 1, ..., N$. A similar conclusion holds for the joint distributions $p(x_i, x_{i+1}, t)$, which satisfy the equations

$$\begin{aligned} \frac{\partial p(x_i, x_{i+1}, t)}{\partial t} = & -\frac{\partial}{\partial x_i} \left[(F - \sin(x_{i+1})x_i - Ax_i)\, p(x_i, x_{i+1}, t) \right] - \\ & \frac{\partial}{\partial x_{i+1}} \int_{-\infty}^{\infty} \left[(F - \sin(x_{i+2})x_{i+1} - Ax_{i+1})\, p(x_i, x_{i+1}, x_{i+2}, t) \right] dx_{i+2}. \end{aligned} \tag{40}$$

The PDE system (39)-(40) represents the first two levels of the BBGKY hierarchy corresponding to (36).

Let us set $i = 1$ in equation (39) and express the integral in terms of the conditional expectation of $\sin(x_2(t))$ given $x_1(t)$. This yields

$$\frac{\partial p(x_1,t)}{\partial t} = \frac{\partial}{\partial x_1} \left(x_1 p(x_1,t) \mathbb{E}\left[\sin(x_2(t)) | x_1(t) \right] \right) + \frac{\partial}{\partial x_1} \left[(Ax_1 - F) p(x_1,t) \right], \tag{41}$$

where

$$\mathbb{E}\left[\sin(x_2(t)) | x_1(t) \right] = \int_{-\infty}^{\infty} \sin(x_2) p(x_2 | x_1, t) dx_2. \tag{42}$$

Example: Consider the Liouville equation (17) corresponding to (4)-(5). Is it possible to derive an evolution equation for $p(x,t)$, i.e., integrate the variables $\boldsymbol{y}$ representing $(\xi_1,\ldots,\xi_M)$ in the KL expansion of the noise (5)? By applying the marginalization rule

$$p(x,t)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}p(x,\boldsymbol{y},t)d\boldsymbol{y} \tag{43}$$

to the Liouville equation (17) we obtain

$$\frac{\partial p(x,t)}{\partial t}+\frac{\partial}{\partial x}\left(f(x)p(x,t)\right)+\sum_{k=1}^{M}\sqrt{\lambda_k}\psi_k(t)\frac{\partial}{\partial x}\int_{-\infty}^{\infty}y_k p(x,y_k,t)dy_k=0. \tag{44}$$

Note that the PDF $p(x,t)$ depends on M joint PDFs $p(x,y_k,t)$. Therefore (44) is an unclosed PDF equation. We can of course derive an evolution equation for each $p(x,y_k,t)$ as

$$\frac{\partial p(x,y_k,t)}{\partial t}+\frac{\partial}{\partial x}\left(f(x)p(x,y_k,t)\right)+\frac{\partial p(x,y_k,t)}{\partial x}\sqrt{\lambda_k}y_k\psi_k(t)+\sum_{\substack{j=1\\ j\neq k}}^{M}\sqrt{\lambda_j}\psi_j(t)\frac{\partial}{\partial x}\int_{-\infty}^{\infty}y_j p(x,y_j,y_k,t)dy_j=0. \tag{45}$$

These are additional M unclosed PDEs involving $p(x,y_j,y_k,t)$. At this point we could derive the evolution equation for the joint PDF $p(x,y_j,y_k,t)$, and go on and on. The BBGKY hierarchy is formally closed only at the level of the Liouville equation, unless the system has a special structure, or a *closure approximation* is introduced. For instance, if $p(x,y_j,y_k,t)$ can be factored in terms of lower-order PDFs as

$$p(x,y_j,y_k,t)\simeq p(x,y_k,t)p(y_j) \tag{46}$$

then (44)-(45) is a closed system of PDEs.

A substitution of

$$p(x,y_k,t)=p(y_k|x,t)p(x,t), \tag{47}$$

where $p(y_k|x,t)$ is the conditional PDF of y_k given $x(t;\omega)$, into (44) yields the low dimensional PDE

$$\frac{\partial p(x,t)}{\partial t}+\frac{\partial}{\partial x}\left(f(x)p(x,t)\right)+\sum_{k=1}^{M}\psi_k(t)\frac{\partial}{\partial x}\left(p(x,t)\mathbb{E}\{\xi_k(\omega)|x(t;\omega)=x\}\right). \tag{48}$$

Here,

$$\mathbb{E}\{\xi_k(\omega)|x(t;\omega)=x\}=\int_{-\infty}^{\infty}y_k p(y_k|x,t)dy_k \tag{49}$$

is the conditional expectation of $\xi_k(\omega)$ given $x(t;\omega)=x$. We shall see hereafter that such conditional expectation can be estimated from sample trajectories of (4)-(5).

Data-driven closure approximation of BBGKY hierarchies. Computing conditional expectations from data or sample trajectories is a key step in determining accurate closure approximations of reduced-order PDF equations. A major challenge to fitting a conditional expectation is ensuring accuracy and stability. More importantly, the estimator must be flexible and effective for a wide range of numerical applications. Let us briefly recall what conditional expectations are and, more importantly, how to compute them based on sample paths of (1). To this end, consider the random processes $x_1(t)$ and $x_3(t)$ defined by the dynamical system (24) evolving from a random initial state. The conditional expectation of $x_3(t)$ given $x_1(t)$ is defined mathematically in equation (30). The geometric meaning of such conditional expectation is illustrated in Figure 5. We first compute sample trajectories of (24) – see Figure 5(a) – by sampling the initial condition and evolving it forward in time. We then project the solution samples we obtain at time t

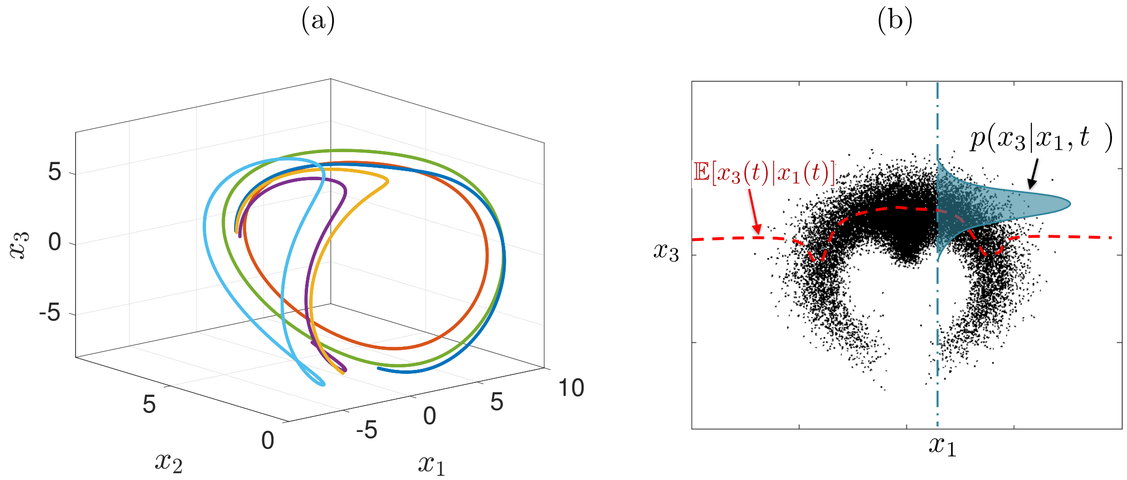


Figure 5: Kraichnan-Orszag three mode problem. (a) Sample trajectories of (24) corresponding to random samples projected on the plane (x_1, x_3). For each value of x_1, the conditional PDF $p(x_3|x_1,t)$ can be estimated based on samples sitting on or lying nearby the vertical dashed line. The conditional expectation $\mathbb{E}[x_3(t)|x_1(t)]$ is the mean of such conditional PDF.

on the plane (x_1, x_3), to obtain the scatter plot in Figure 5(b). For each value of x_1, the conditional PDF $p(x_3|x_1,t)$ can be estimated based on all samples sitting on or lying nearby the vertical dashed line. The conditional expectation $\mathbb{E}[x_3(t)|x_1(t)]$ is the mean of such conditional PDF.

In this section, we present two different approaches to estimate conditional expectations from data based on *moving averages* and *smoothing splines*. The moving average estimate is obtained by first sorting the data into bins and then computing the average within each bin. With such averages available, we can construct a smooth interpolant using the average value within each bin. Important factors affecting the bin average are the bin size (the number of samples in each bin) and the interpolation method used in the final step. Another approach to estimate conditional expectations uses smoothing splines. This approach seeks to minimize a penalized sum of squares. A smoothing parameter determines the balance between smoothness and goodness-of-fit in the least-squares sense [3]. The choice of smoothing parameter is critical to the accuracy of the results. Specifying the smoothing parameter a priori generally yields poor estimates [15]. Instead, cross-validation and maximum likelihood estimators can guide the choice the optimal smoothing value for the data set [23]. Such methods can be computationally intensive, and is not recommended when the spline estimate is performed at each time step. Other techniques to estimate conditional expectations can be built upon deep-neural nets.

In Figure 6 we compare the performance of the moving average and smoothing splines approaches in approximating the conditional expectation of two jointly Gaussian random variables. Specifically, we consider the joint distribution

$$p(x_1,x_2)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)}{\sigma_1^2}\frac{(x_2-\mu_2)}{\sigma_2^2}-\frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2)}\right]\right) \tag{50}$$

with parameters $\rho = 3/4$, $\mu_1 = 0$, $\mu_2 = 2$, $\sigma_1 = 1$ $\sigma_2 = 2$. As is well known [14], the conditional expectation

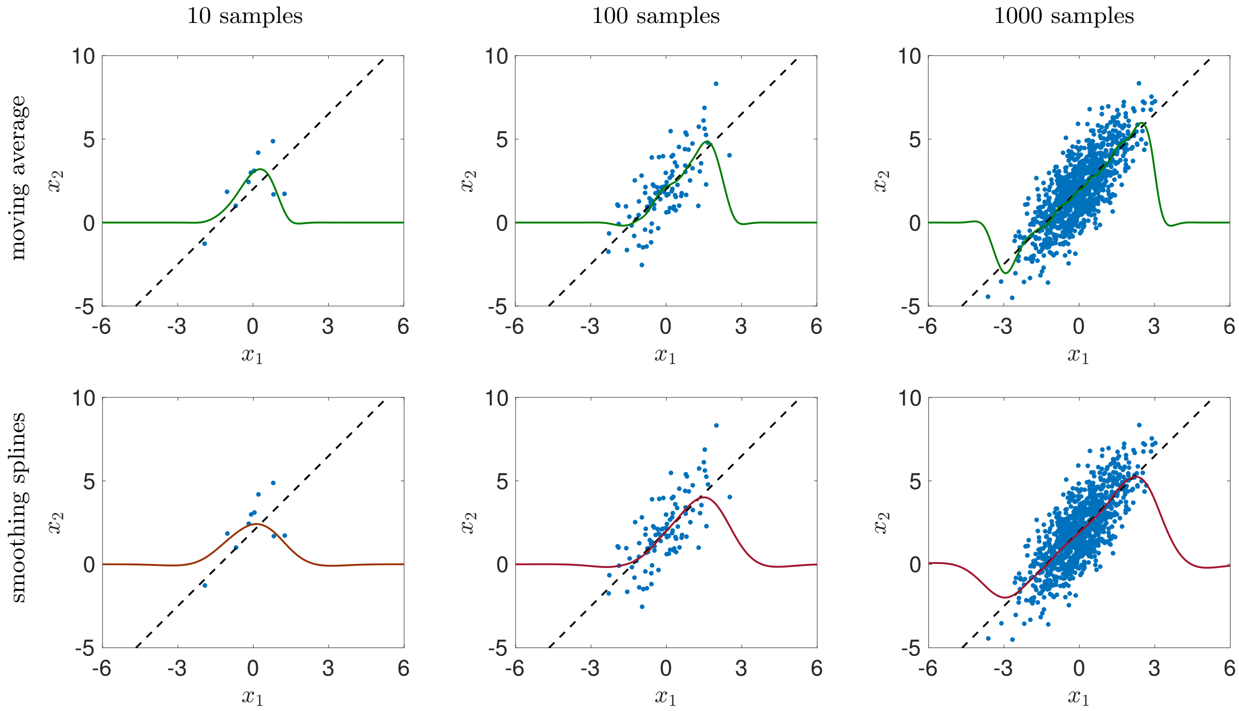


Figure 6: Numerical estimation of the conditional expectation (52) for different number of samples of (50). Shown are results obtained with moving averages (first row) and cubic smoothing splines (second row). It is seen that both methods converge to the correct conditional expectation in the active region as we increase the number of samples.

of x_2 given x_1 can be expressed as[5]

$$\mathbb{E}[x_2|x_1] = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1 - \mu_1) = 2 + \frac{3}{2}x_1. \tag{52}$$

Such expectation is plotted in Figure 6 (dashed line), together with the plots of the estimates we obtain with the moving average and the smoothing spline approaches for different numbers of samples. It is seen that both methods converge to the correct conditional expectation as we increase the number of samples. Both estimators are parametric, i.e., they require setting suitable parameters to compute the expectation, e.g., the width of the moving average window in the moving average approach, or the smoothing parameter in the cubic spline.

If the joint PDF of x_1 and x_2 is not compactly supported, then the conditional expectation is defined on the whole real line. It is computationally challenging to estimate (52) in regions where the PDF is very small. At the same time, if we are not interested in rare events (i.e., tails of probability densities), then resolving the dynamics in such regions of small probability is not really needed. This means that if we have available a sufficient number of sample trajectories, then we can identify the regions of the phase space where dynamics is happening with high probability, and approximate the conditional expectation only within such regions. Outside the active regions, we can set the expectation equal to zero. However, keep in mind that if the joint PDF of x_1 and x_2 is compactly supported, e.g. uniform on the square $[0, 1]^2$, then conditional expectation is undefined outside the support of the joint PDF.

[5] Given two random variables with joint PDF $p(x_1, x_2)$, the conditional expectation of x_2 given x_1 is defined as

$$\mathbb{E}[x_2|x_1] = \int_{-\infty}^{\infty} x_2 p(x_2|x_1)dx_2 = \frac{1}{p(x_1)}\int_{-\infty}^{\infty} x_2 p(x_1, x_2)dx_2, \tag{51}$$

where $p(x_1)$ is the marginal of $p(x_1, x_2)$ with respect to x_2.

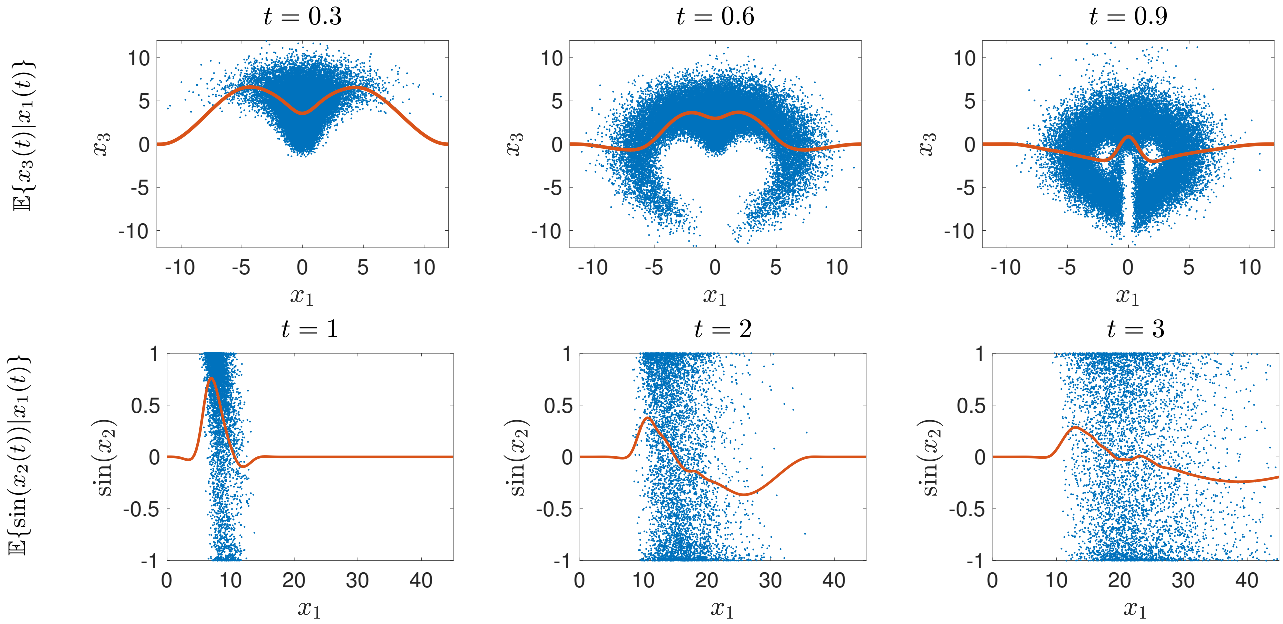


Figure 7: Data-driven estimates of the conditional expectations (30) and (42) defining the reduced-order PDF models (29) and (41).

In Figure 7, we summarize the results we obtain by applying the smoothing spline conditional expectation estimator to the dynamical systems (24) and (36). In figure 8 and figure 9 we provide numerical simulation result for (29) and (41), respectively.

PDF equations for nonlinear PDEs evolving from random initial conditions

The procedure we used to derive reduced-order PDF equations for dynamical systems can be extended to PDEs evolving from random initial states, or PDEs with random forcing (see, e.g., [10, 7]). To describe the method, consider the prototype problem of a one dimensional heat equation evolving from a random initial state

$$\frac{\partial u(x,t;\omega)}{\partial t} = \kappa^2 \frac{\partial^2 u(x,t;\omega)}{\partial x^2}, \qquad u(x,0;\omega) = u_0(x;\omega). \tag{53}$$

We have seen in Chapter 1 that the Hopf functional[6]

$$\Phi([\theta],t) = \mathbb{E}\left\{\exp\left[i\int_{-\infty}^{\infty} u(x,t;\omega)\theta(x)dx\right]\right\} \tag{54}$$

provides full statistical information on $u(x,t;\omega)$ at each time t. This includes, e.g., multi-point statistical moments such as

$$\mathbb{E}\left\{u(x_i,t;\omega)u(x_j,t;\omega)\right\} \quad \text{and} \quad \mathbb{E}\left\{u(x_i,t;\omega)u(x_j,t;\omega)u(x_k,t;\omega)\right\}, \tag{55}$$

or multi-point probability density functions. It is possible to derive an evolution equation for the Hopf functional corresponding to the solution of (53). To this end, let us differentiate (54) with respect to time

[6] In (54) we assumed that the spatial domain for the heat equation (53) is $\mathbb{R}$. If the spatial domain is a compact subset $\mathbb{R}$, say $[0, 2\pi]$, then the domain on which the integral in (54) is evaluated changes accordingly.

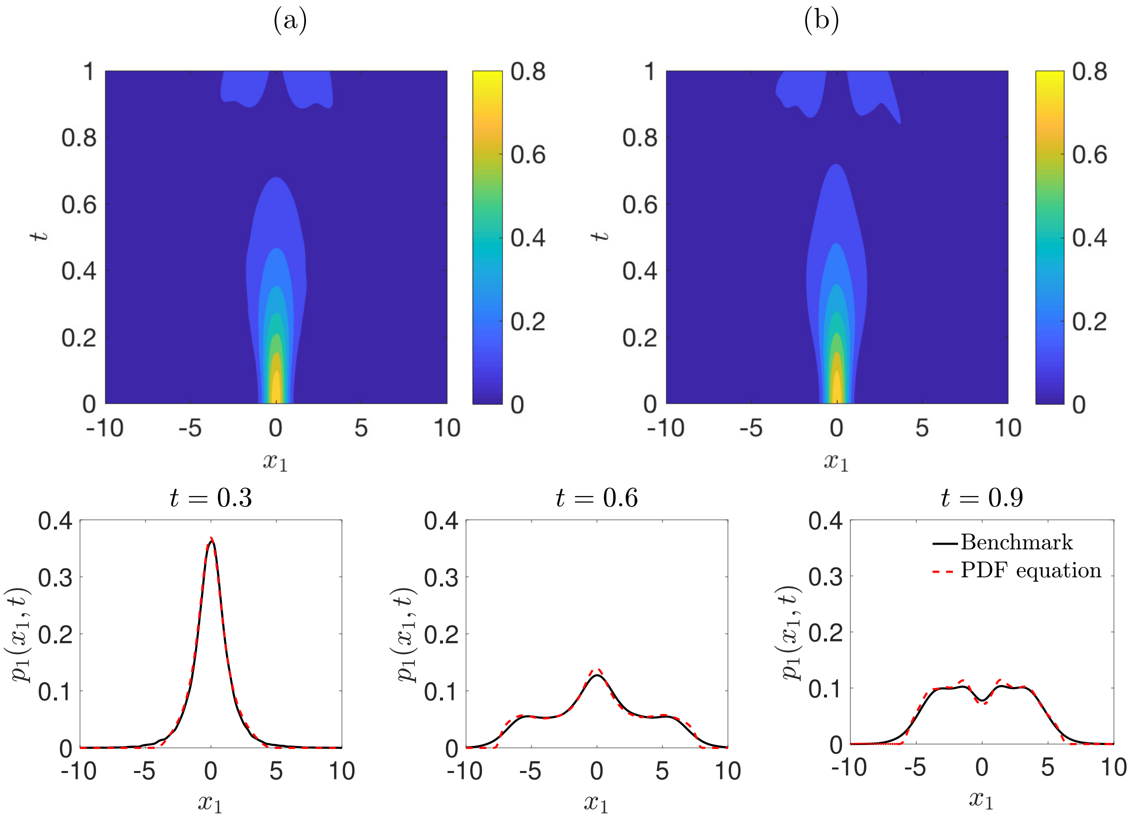


Figure 8: Kraichnan-Orszag three-mode problem. (a) Accurate kernel density estimate of $p_1(x_1, t)$ based on 30000 sample trajectories. (b) Numerical solution of (29) obtained by estimating $\mathbb{E}[x_3(t)|x_1(t)]$ with 5000 sample trajectories.

to obtain

$$\begin{aligned}
\frac{\partial \Phi([\theta], t)}{\partial t} =& \mathbb{E}\left\{\exp\left[i\int_{-\infty}^{\infty} u(x,t;\omega)\theta(x)dx\right] i\int_{-\infty}^{\infty} \frac{\partial u(x,t;\omega)}{\partial t}\theta(x)dx\right\}\\
=& i\kappa^2 \int_{-\infty}^{\infty} \mathbb{E}\left\{\exp\left[i\int_{-\infty}^{\infty} u(x,t;\omega)\theta(x)dx\right] \frac{\partial^2 u(x,t;\omega)}{\partial x^2}\right\}\theta(x)dx\\
=& i\kappa^2 \int_{-\infty}^{\infty} \frac{\partial^2}{\partial x^2}\left(\mathbb{E}\left\{\exp\left[i\int_{0}^{2\pi} u(x,t;\omega)\theta(x)dx\right] u(x,t;\omega)\right\}\right)\theta(x)dx\\
=& i\kappa^2 \int_{-\infty}^{\infty} \frac{\partial^2}{\partial x^2}\left(\frac{\delta\Phi([\theta],t)}{\delta\theta(x)}\right)\theta(x)dx, \qquad (56)
\end{aligned}$$

where $\delta\Phi([\theta], t)/\delta\theta(x)$ denotes the first-order functional derivative of the nonlinear functional (54) (see [19] or [5, p. 309]). Technically speaking, equation (56) is a *functional-differential* equation (FDE) as it involves derivatives with respect to functions and derivatives with respect to independent variables x and t. The solution to (56) is a time-dependent nonlinear functional, i.e., a nonlinear operator from a space of functions into $\mathbb{C}$. The functional differential equation (56) is essentially an infinite-dimensional PDE, i.e., a PDE in an infinite number of independent variables which may be approximated by a PDE in a finite (though very large) number of variables using the functional methods described in [19].

The Hopf equation (56) plays the same role for the heat equation (53) as the Fourier transform of the Liouville equation (9) does for finite-dimensional dynamical systems (1).

Example: Another well-known example of a FDE involves the characteristic functional of the solution to

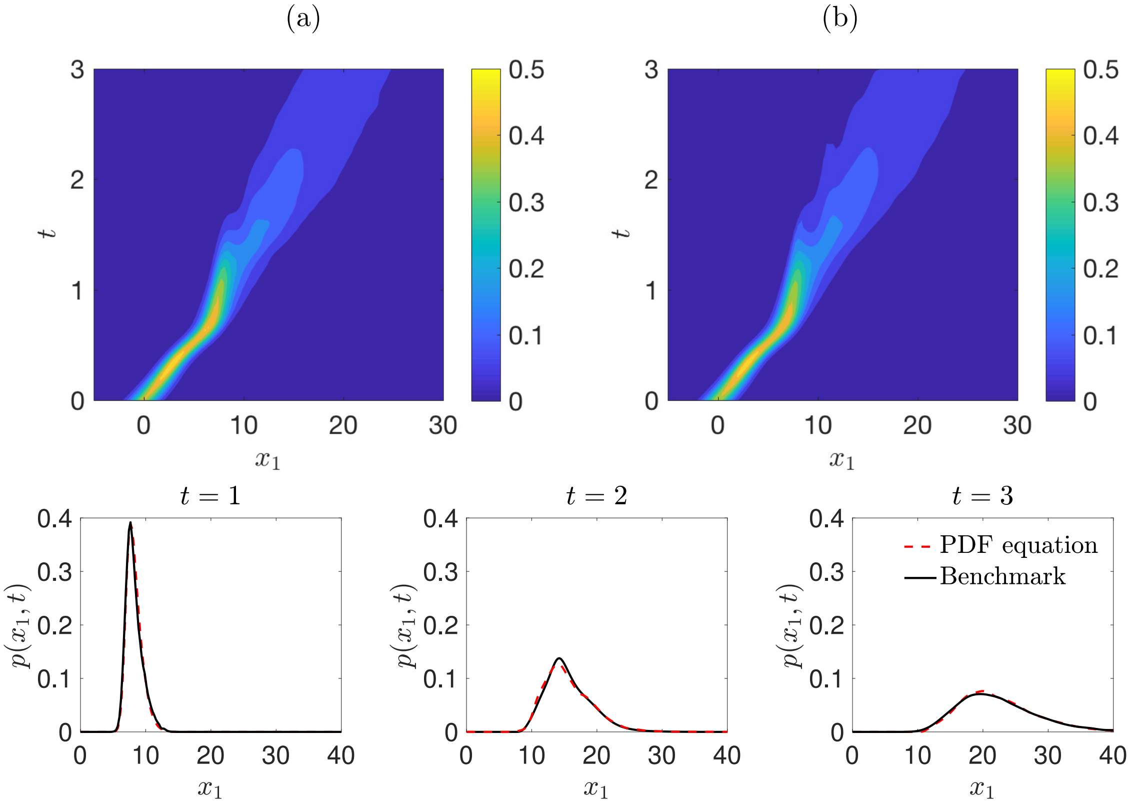


Figure 9: Nonlinear dynamical system (36). (a) Accurate kernel density estimate [2] of $p(x_1,t)$ based on 20000 sample trajectories. (b) Data-driven solution of the transport equation (41). We estimated the conditional expectation $\mathbb{E}\left[\sin(x_2(t))|x_1(t)\right]$ based on 5000 sample trajectories of (36) (see Figure 7).

the Navier-Stokes equations

$$\frac{\partial \boldsymbol{u}}{\partial t}+(\boldsymbol{u}\cdot\nabla)\boldsymbol{u}=-\nabla p+\nu\nabla^2\boldsymbol{u}\qquad \nabla\cdot\boldsymbol{u}=0. \tag{57}$$

Such a FDE can be written as [6, 11, 19]

$$\frac{\partial\Phi([\boldsymbol{\theta}],t)}{\partial t}=\sum_{k=1}^{3}\int_V\theta_k(\boldsymbol{x})\left(i\sum_{j=1}^{3}\frac{\partial}{\partial x_j}\frac{\delta^2\Phi([\boldsymbol{\theta}],t)}{\delta\theta_k(\boldsymbol{x})\delta\theta_j(\boldsymbol{x})}+\nu\nabla^2\frac{\delta\Phi([\boldsymbol{\theta}],t)}{\delta\theta_k(\boldsymbol{x})}\right)d\boldsymbol{x}, \tag{58}$$

where

$$\Phi([\theta],t)=\mathbb{E}\left\{\exp\left[i\int_V\boldsymbol{u}(\boldsymbol{x},t;\omega)\cdot\boldsymbol{\theta}(\boldsymbol{x})d\boldsymbol{x}\right]\right\}. \tag{59}$$

Here, $\boldsymbol{u}(x,t;\omega)$ represents a stochastic solution to the Navier-Stokes equation (57) corresponding to a random initial state, and $\mathbb{E}\{\cdot\}$ is the expectation over the probability measure of such random initial state. Equation (58) was deemed by Monin and Yaglom ([11, Ch. 10]) to be *"the most compact formulation of the general turbulence problem"*, which is the problem of determining the statistical properties of the velocity field generated by the Navier-Stokes equations given statistical information on the initial state[7].

Remark: Clearly, if we discretize the PDE (53) or (57) in the spatial domain, e.g., with finite-differences, then we obtain a system of ODEs which can be handled with the mathematical tools we discussed in the

[7] In equations (58)-(59), $V\subseteq\mathbb{R}^3$ is a periodic box, $\boldsymbol{\theta}(\boldsymbol{x})=(\theta_1(\boldsymbol{x}),\theta_2(\boldsymbol{x}),\theta_3(\boldsymbol{x}))$ is a vector-valued (divergence-free) function, and $\delta/\delta\theta_j(\boldsymbol{x})$ denotes the first-order functional derivative.

previous section. In particular, it is possible to derive a Liouville equation for such finite-dimensional ODE system and correspondingly a BBGKY hierarchy for the solution evaluated e.g., at the spatial grid points. An interesting question is how to compute statistical properties at spatial locations that do not coincide with the grid points. For example, is it possible to "interpolate" the joint characteristic function of the solution $u(x,t;\omega)$ at n spatial nodes $\{x_k\}$ and obtain an approximation of the joint characteristic at a different set of m nodes? To answer this question, consider the 2-point characteristic function

$$\phi_2(a_1,a_2,t) = \mathbb{E}\left\{e^{ia_1 u(x_1,t;\omega)+ia_2 u(x_2,t;\omega)}\right\}. \tag{60}$$

Let x^* be a point in between x_1 and x_2. Assuming that $u(x,t;\omega)$ is a smooth solution to a PDE, we can construct an interpolant for $u(x^*,t,\omega)$, e.g., a linear interpolant as

$$u(x^*,t;\omega) = u(x_1,t;\omega)\ell_1(x^*) + u(x_2,t;\omega)\ell_2(x^*) \tag{61}$$

where $\ell_1(x) = (x-x_2)/(x_1-x_2)$ and $\ell_2(x) = (x-x_1)/(x_2-x_1)$ are Lagrange characteristic polynomials. This representation allows us to represent the three-point joint characteristic function of $u(x,t;\omega)$ at x_1, x_2 and x^* as

$$\begin{aligned}\phi_3(a_1,a_2,a_3,t) =&\mathbb{E}\left\{e^{ia_1u(x_1,t;\omega)+ia_2u(x_2,t;\omega)+ia_3u(x^*,t;\omega)}\right\}\\ =&\mathbb{E}\left\{e^{ia_1u(x_1,t;\omega)+ia_2u(x_2,t;\omega)+ia_3(u(x_1,t;\omega)\ell_1(x^*)+u(x_2,t;\omega)\ell_2(x^*))}\right\}\\ =&\mathbb{E}\left\{e^{i(a_1+a_3\ell_1(x^*))u(x_1,t;\omega)+i(a_2+a_3\ell_2(x^*))u(x_2,t;\omega)}\right\}\\ =&\phi_2(a_1+a_3\ell_1(x^*),a_2+a_3\ell_2(x^*),t).\end{aligned} \tag{62}$$

This expression provides an approximation of the three-point characteristic function in terms of the two point characteristic function. Of course the method can be generalized to n point characteristic functions. If the spatial discretization is sufficiently fine, and the interpolants are accurate, we can represent the $2n$, $3n$, etc., characteristic functions in terms of one *core* characteristic function e.g., involving the solution at n spatial points.

Lundgren-Monin-Novikov (LMN) hierarchy. We are interested in deriving the PDF equation governing the PDF of $u(x,t)$. To this end, consider the characteristic function

$$\phi(a,x,t) = \mathbb{E}\left\{e^{iau(x,t;\omega)}\right\}, \tag{63}$$

and differentiate it with respect to time to obtain

$$\begin{aligned}\frac{\partial\phi(a,x,t)}{\partial t} =&ia\mathbb{E}\left\{\frac{\partial u(x,t;\omega)}{\partial t}e^{iau(x,t;\omega)}\right\}\\ =&ia\kappa^2\mathbb{E}\left\{\frac{\partial^2 u(x,t;\omega)}{\partial x^2}e^{iau(x,t;\omega)}\right\}\\ =&ia\kappa^2\lim_{y\to x}\mathbb{E}\left\{\frac{\partial^2 u(y,t;\omega)}{\partial y^2}e^{iau(x,t;\omega)}\right\}\\ =&ia\kappa^2\lim_{y\to x}\frac{\partial^2}{\partial y^2}\mathbb{E}\left\{u(y,t;\omega)e^{iau(x,t;\omega)}\right\}.\end{aligned} \tag{64}$$

Recalling that the two-point characteristic function is defined as

$$\phi(a,b,x,y,t) = \mathbb{E}\left\{e^{iau(x,t;\omega)+ibu(y,t;\omega)}\right\} \tag{65}$$

we see that we can write the term at the right hand side of (64) as

$$i\mathbb{E}\left\{u(y,t;\omega)e^{iau(x,t;\omega)}\right\} = \lim_{b\to 0}\frac{\partial}{\partial b}\phi(a,b,x,y,t). \tag{66}$$

Substituting into (66) into (64) yields

$$\frac{\partial\phi(a,x,t)}{\partial t} = \kappa^2 \lim_{b\to 0}\lim_{y\to x}\frac{\partial^2}{\partial y^2}a\frac{\partial\phi(a,b,x,y,t)}{\partial b}. \tag{67}$$

Next, we transform this equation for the characteristic function to an equation for the PDF. To this end, we first recall that

$$\phi(a,x,t) = \int_{-\infty}^{\infty} e^{iau}p(u,x,t)du, \qquad \phi(a,b,x,y,t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{iau+ibv}p(u,v,x,y,t)dudv. \tag{68}$$

This allows us to write the right hand side of (67) as

$$a\frac{\partial\phi(a,b,x,y,t)}{\partial b} = ia\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{iau+ibv}vp(u,v,x,y,t)dudv. \tag{69}$$

Taking the limit

$$\begin{aligned}
\lim_{b\to 0} a\frac{\partial\phi(a,b,x,y,t)}{\partial b} =& ia\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} ve^{iau}p(u,v,x,y,t)dudv\\
=& \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} v\frac{\partial\left(e^{iau}\right)}{\partial u}p(u,v,x,y,t)dudv\\
=& -\int_{-\infty}^{\infty} e^{iau}\int_{-\infty}^{\infty} v\frac{\partial p(u,v,x,y,t)}{\partial u}dudv.
\end{aligned} \tag{70}$$

Hence,

$$\frac{\partial p(u,x,t)}{\partial t} = -\kappa^2\lim_{y\to x}\frac{\partial^2}{\partial y^2}\int_{-\infty}^{\infty} v\frac{\partial p(u,v,x,y,t)}{\partial u}dv. \tag{71}$$

In other words, the dynamics of the one-point PDF $p(u,x,t)$ (PDF of the solution at location x and time t) depends on the joint PDF of $u(x,t;\omega)$ and $u(y,t,\omega)$ through some quite unusual limit. Equation (71) is the first equation PDF of an infinite hierarchy known as *Lundgren-Monin-Novikov (LMN) hierarchy* [10, 4, 22], first developed by Thomas Lundgren to study the statistical properties of turbulence. The second equation of the LMN hierarchy is an equation for the time derivative of the two point PDF $p(u,v,x,y,t)$. It was shown in [7] that the Hopf functional equation is completely equivalent to the LMN hierarchy.

Remark: Why do we get a closure problem for the one-point one-time PDF equation of the solution to the heat equation? The reason is rather simple, and can be understood by recalling that the analytical solution of (53) in an infinite domain is

$$u(x,t;\omega) = \int_{-\infty}^{\infty}\mathscr{G}\left(x,t|x',t'\right)u(x',t')dx' \qquad t\geq t', \tag{72}$$

where

$$\mathscr{G}\left(x,t|x',t'\right) = \frac{1}{[4\pi\kappa^2(t-t')]^{1/2}}\exp\left(-\frac{(x-x')^2}{4\kappa^2(t-t')}\right) \tag{73}$$

is the heat kernel, i.e., the Green function of the diffusion equation on the real line. For an infinitesimal time increment Δt we have that the random variable $u(x,t+\Delta t;\omega)$ (for fixed x) depends on all random variables at the previous time step $u(x',t;\omega)$ (arbitrary $x' \in \mathbb{R}$). To see this more clearly, consider the following quadrature approximation of the integral in (72) (e.g., Hermite quadrature)

$$u(x_p, t+\Delta t;\omega) = \sum_{j=1}^{M} \frac{w_j}{[4\pi\kappa^2\Delta t]^{1/2}} \exp\left(-\frac{(x_p-x_j)^2}{4\kappa^2\Delta t}\right) u(x_j,t;\omega), \tag{74}$$

where x_j are Gauss-Hermite nodes, and w_j are quadrature weights. Clearly, (74) represents a mapping from M random variables $\{u(x_1,t;\omega),\dots,u(x_M,t;\omega)\}$ into one random variable $u(x_p,t+\Delta t;\omega)$. We know that the PDF of $u(x_p,t+\Delta t;\omega)$ can be computed if and only if the joint PDF of the random vector $\{u(x_1,t;\omega),\dots,u(x_M,t;\omega)\}$ is available. In other words, the fact that the solution (72) is *non-local* in space implies that the statistical properties at some fixed spatial point x and time $t+\Delta t$ are determined by the joint statistics at all points x' at a previous time instant. Hence, a closed equation for the one-point PDF cannot exist.

By using similar methods, it is possible to derive LMN PDF hierarchies corresponding to rather general nonlinear PDEs, e.g., the Navier-Stokes equation (57), evolving from random initial states (see [10]).

Data-driven closure approximation of LMN hierarchies. The integral at the right hand side of (71) can be written in terms of a conditional expectation of $u(y,t;\omega)$ given $u(x,t;\omega)$. A substitution of the identity

$$p(u,v,x,y,t) = p(v,y,t|u,x,t)p(u,x,t) \tag{75}$$

into (71) yields

$$\begin{aligned}\frac{\partial p(u,x,t)}{\partial t} =& -\kappa^2 \lim_{y\to x} \frac{\partial^2}{\partial y^2}\int_{-\infty}^{\infty} v p(v,y,t|u,x,t)\frac{\partial p(u,x,t)}{\partial u} dv\\ =& -\kappa^2 \frac{\partial p(u,x,t)}{\partial u} \lim_{y\to x}\frac{\partial^2}{\partial y^2}\mathbb{E}\left\{u(y,t;\omega)|u(x,t;\omega)\right\}.\end{aligned} \tag{76}$$

As before, if we estimate the conditional expectation $\mathbb{E}\left\{u(y,t;\omega)|u(x,t;\omega)\right\}$ from sample paths of (53) then we can solve (76) as we did in the case of data-driven closures for BBGKY hierarchies. The development of efficient methods for data-driven estimation of conditional expectations such as $\mathbb{E}\left\{u(y,t;\omega)|u(x,t;\omega)\right\}$ nearby $x=y$ is (to my knowledge) an open problem.

Example: Consider the Kuramoto-Sivashinsky equation

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + \frac{\partial^2 u}{\partial x^2} + \nu\frac{\partial^4 u}{\partial x^4} = 0 \tag{77}$$

By using the methods we just outlined it can be shown that the first equation of the LMN hierarchy is

$$\begin{aligned}\frac{\partial p(u,x,t)}{\partial t} + \int_{-\infty}^{u} \frac{\partial p(u',x,t)}{\partial x} du' + u\frac{\partial p(u,x,t)}{\partial x} =& \\ -\lim_{y\to x}\left[\frac{\partial^2}{\partial y^2}\mathbb{E}\left\{u(y,t;\omega)|u(x,t;\omega)\right\} + \nu\frac{\partial^4}{\partial y^4}\mathbb{E}\left\{u(y,t;\omega)|u(x,t;\omega)\right\}\right].\end{aligned} \tag{78}$$

Hence, once again, the PDF equation can be closed by estimating $\mathbb{E}\left\{u(y,t;\omega)|u(x,t;\omega)\right\}$ from data.

Appendix A: Fokker-Planck equation and generalized Fokker-Planck equations

Consider the following stochastic ODE

$$d\boldsymbol{X} = \boldsymbol{G}(\boldsymbol{X}, t)dt + \boldsymbol{M}(\boldsymbol{X}, t)d\boldsymbol{\zeta}(t;\omega), \qquad \boldsymbol{X}(0;\omega) = \boldsymbol{X}_0(\omega). \tag{79}$$

where $\boldsymbol{\zeta}(t)$ is a vector-valued m-dimensional random process with known statistical properties, and $\boldsymbol{M}(\boldsymbol{X}, t)$ is a $n\times m$ matrix of functions. We have seen at the beginning of this Chapter that if $\boldsymbol{\zeta}(t)$ is finite-dimensional (i.e., it can be represented in terms of a finite-number of random variables) then it is possible to derive an exact transport equation (i.e., (9)) for the joint PDF of $\boldsymbol{X}(t;\omega)$ and all random variables representing $\boldsymbol{\zeta}(t;\omega)$. By integrating out the phase variables corresponding to the noise, i.e., by marginalizing the Liouville equation with respect to the phase variables representing the noise, it is straightforward to obtain an evolution equation for the PDF of $\boldsymbol{X}(t;\omega)$ alone. Such an equation represents the first equation of a BBGKY hierarchy and it is usually not closed, meaning that it involves quantities that cannot be computed just based on the PDF of $\boldsymbol{X}(t;\omega)$. However, there are cases in which the integration of the noise can be carried out exactly, and a closed equation for the PDF of $\boldsymbol{X}(t;\omega)$ can be derived. Perhaps the most famous example is the case where $\boldsymbol{xi}(t)$ is a Wiener process. In this case it was shown in [16] that $p(\boldsymbol{x}, t)$ satisfies the Fokker-Plank equation

$$\frac{\partial p(\boldsymbol{x}, t)}{\partial t} = -\sum_{k=1}^{n} \frac{\partial}{\partial x_k}\left(G_k(\boldsymbol{x}, t)p(\boldsymbol{x})\right) + \frac{1}{2}\sum_{i,k=1}^{n} \frac{\partial^2}{\partial x_i \partial x_k}\left(\sum_{j=1}^{m} M_{ij}(\boldsymbol{x}, t)M_{kj}(\boldsymbol{x}, t)p(\boldsymbol{x}, t)\right). \tag{80}$$

Let us denote by $\mathscr{K}(\boldsymbol{x}, t)$ the Kolmogorov operator defining the right hand side of (80), i.e.,

$$\mathscr{K}(\boldsymbol{x}, t)p(\boldsymbol{x}, t) = -\sum_{k=1}^{n} \frac{\partial}{\partial x_k}\left(G_k(\boldsymbol{x}, t)p(\boldsymbol{x})\right) + \frac{1}{2}\sum_{i,k=1}^{n} \frac{\partial^2}{\partial x_i \partial x_k}\left(\sum_{j=1}^{m} M_{ij}(\boldsymbol{x}, t)M_{kj}(\boldsymbol{x}, t)p(\boldsymbol{x}, t)\right). \tag{81}$$

It is shown in [16, p. 86] that $\mathscr{K}$ represents the first-order term in the short-time expansion of the transition density

$$p_{t+dt|t}(\boldsymbol{x}, t+dt|\boldsymbol{y}, t) = \left[I + \mathscr{K}(\boldsymbol{x}, t)dt + \mathcal{O}(dt^2)\right]\delta(\boldsymbol{x} - \boldsymbol{y}), \tag{82}$$

where $\delta(\cdot)$ is the multivariate Dirac delta function. The transition density (82) allows us to compute the PDF $p(\boldsymbol{x}, t)$ of the random vector $\boldsymbol{X}(t;\omega)$ appearing in (79) given the PDF $p(\boldsymbol{x}, s)$ of $\boldsymbol{X}(s;\omega)$ at any time $s \leq t$

$$p(\boldsymbol{x}, t) = \int p_{t|s}(\boldsymbol{x}, t|\boldsymbol{y}, s)p(\boldsymbol{y}, s)d\boldsymbol{y}. \tag{83}$$

We emphasize that the PDE governing the PDF of the solution to (4) depends substantially on the statistical properties of the noise $\boldsymbol{\zeta}(t)$. For instance, if we replace the Wiener process $\boldsymbol{\zeta}(t;\omega)$ in (4) with a Lévy random walk then the PDF equation for $\boldsymbol{X}(t;\omega)$ comes with a *fractional Laplace operator* [18], i.e., it is a fractional PDE.

Similarly, for weakly colored random noise, i.e., noise with short temporal correlation, it is possible to leverage the quasi-Markovian nature of the system, and integrate out the noise, e.g., using functional integration [12, 21].

References

[1] C. Bonatto, J. A. C. Gallas, and Y. Ueda. Chaotic phase similarities and recurrences in a damped-driven Duffing oscillator. *Phys. Rev. E*, 77:026217(1–5), 2008.

[2] Z. I. Botev, J. F. Grotowski, and D. P. Kroese. Kernel density estimation via diffusion. *Annals of Statistics*, 38(5):2916–2957, 2010.

[3] P. Craven and G. Wahba. Smoothing noisy data with spline functions. *Numerische Mathematik*, 31(4):377–403, 1979.

[4] R. Friedrich, A. Daitche, O. Kamps, J. Lülff, M. Voβkuhle, and M. Wilczek. The Lundgren-Monin-Novikov hierarchy: kinetic equations for turbulence. *Comptes Rendus Physique*, 13(9-10):929–953, 2012.

[5] P. Hänggi. Colored noise in continuous dynamical system. In F. Moss and P. V. E. McClintock, editors, *Noise in nonlinear dynamical systems (Vol. 1)*, pages 307–347. Cambridge Univ. Press, 1989.

[6] E. Hopf. Statistical hydromechanics and functional calculus. *J. Rat. Mech. Anal.*, 1(1):87–123, 1952.

[7] I. Hosokawa. Monin-Lundgren hierarchy versus the Hopf equation in the statistical theory of turbulence. *Phys. Rev. E*, 73:067301(1–4), 2006.

[8] A. I. Khuri. Applications of Dirac's delta function in statistics. *Int. J. Math. Educ. Sci. Technol.*, 35(2):185–195, 2004.

[9] R. Kubo. Generalized cumulant expansion method. *Journal of the Physical Society of Japan*, 17(7):1100–1120, 1962.

[10] T. S. Lundgren. Distribution functions in the statistical theory of turbulence. *Phys. Fluids*, 10(5):969–975, 1967.

[11] A. S. Monin and A. M. Yaglom. *Statistical Fluid Mechanics, Volume II: Mechanics of Turbulence.* Dover, 2007.

[12] F. Moss and P. V. E. McClintock, editors. *Noise in nonlinear dynamical systems. Volume 1: theory of continuous Fokker-Planck systems.* Cambridge Univ. Press, 1995.

[13] S. A. Orszag and L. R. Bissonnette. Dynamical properties of truncated Wiener-Hermite expansions. *Physics of Fluids*, 10(12):2603–2613, 1967.

[14] A. Papoulis. *Probability, random variables and stochastic processes.* McGraw-Hill, third edition, 1991.

[15] S. B. Pope and R. Gadh. Fitting noisy data using cross-validated cubic smoothing splines. *Communications in Statistics - Simulation and Computation*, pages 349–376, 1988.

[16] H. Risken. *The Fokker-Planck equation: methods of solution and applications.* Springer-Verlag, second edition, 1989. Mathematics in science and engineering, vol. 60.

[17] K. Sobczyk. *Stochastic differential equations: with applications to physics and engineering.* Springer, 2001.

[18] J. P. Taylor-King, R. Klages, S. Fedotov, and R. A. Van Gorder. Fractional diffusion equation for an n-dimensional correlated Lévy walk. *Phys. Rev. E*, 94:012104, 2016.

[19] D. Venturi and A. Dektor. Spectral methods for nonlinear functionals and functional differential equations. *Res. Math. Sci.*, 8(27):1–39, 2021.

[20] D. Venturi and G. E. Karniadakis. Convolutionless Nakajima-Zwanzig equations for stochastic analysis in nonlinear dynamical systems. *Proc. R. Soc. A*, 470(2166):1–20, 2014.

[21] D. Venturi, T. P. Sapsis, H. Cho, and G. E. Karniadakis. A computable evolution equation for the joint response-excitation probability density function of stochastic dynamical systems. *Proc. R. Soc. A*, 468(2139):759–783, 2012.

[22] M. Waclawczyk, N. Staffolani, M. Oberlack, A. Rosteck, M. Wilczek, and R. Friedrich. Statistical symmetries of the Lundgren-Monin-Novikov hierarchy. *Phys. Rev. E*, 90:013022(1–11), 2014.

[23] G. Wahba. A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem. *Annals of Statistics*, 13(4):1378–1402, 1985.

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

Deep learning with stochastic neural networks

It has been recently shown that new insights on deep learning can be obtained by regarding the process of training a deep neural network as a discretization of an optimal control problem involving nonlinear differential equations [5, 4, 8]. One attractive feature of this formulation is that it allows to use tools from dynamical system theory to study deep learning from a rigorous mathematical perspective [12, 9, 14]. For instance, it has been recently shown that by idealizing deep residual networks (ResNet) as continuous-time dynamical systems it is possible to derive sufficient conditions for universal approximation in L^p, which can be understood as an approximation theory built on flow maps generated by dynamical systems [13].

In this note we present a formulation deep neural networks obtained by applying simple probabilistic tools to discrete dynamical systems. Specifically, we consider two types of neural network models:

- Neural networks perturbed by additive random noise;
- Neural networks with random weights and biases.

Modeling neural networks as discrete stochastic dynamical systems. Let us begin by modeling the input-output map of a neural network as a discrete dynamical system (see Figure 1)

$$\boldsymbol{X}_1 = \boldsymbol{F}_0(\boldsymbol{X}_0, \boldsymbol{w}_0) + \boldsymbol{\xi}_0 \qquad \boldsymbol{X}_{n+1} = \boldsymbol{F}(\boldsymbol{X}_n, \boldsymbol{w}_n) + \boldsymbol{\xi}_n, \tag{1}$$

Here the index n labels a specific layer in the network, $\boldsymbol{X}_0 \in \mathbb{R}^d$ is the input, $\boldsymbol{X}_n \in \mathbb{R}^N$ $(n = 1, \ldots, L$ represents output of the n-th layer, and $\{\boldsymbol{\xi}_0, \ldots, \boldsymbol{\xi}_{L-1}\}$ is set of statistically independent random vectors, or more generally a vector-valued Markov process. We allow the initial state $\boldsymbol{X}_0$ to be random as well, which can be directly connected to a data set in a training algorithm. A neural networks of the form (1) is called *recurrent*, to emphasize the fact that the mapping

$$\boldsymbol{F}(\boldsymbol{X}_n, \boldsymbol{w}_n) = \boldsymbol{\varphi}(\boldsymbol{W}_n \boldsymbol{X}_n + \boldsymbol{b}_n) \qquad \boldsymbol{w}_n = \{\boldsymbol{W}_n, \boldsymbol{b}_n\}, \tag{2}$$

between one layer and the next has the same functional form. In (2) $\boldsymbol{\varphi} : \mathbb{R}^N \mapsto \mathbb{R}^N$ is the *activation function* of the network, $\boldsymbol{W}_n$ is a $N \times N$ weight matrix and $\boldsymbol{b}_n \in \mathbb{R}^N$ is a bias vector.

In a supervised learning setting, the degrees of freedom

$$\boldsymbol{w} = \{\boldsymbol{w}_0, \ldots, \boldsymbol{w}_{L-1}\}, \tag{3}$$

are determined by optimizing a suitable performance metric depending on the network output. For instance, if we are interested in using the network depicted in Figure 1 to approximate a multivariate function $g(\boldsymbol{x}) \in L^2([0,1]^d)$ then we can identify the degrees of freedom (3) by minimizing, e.g., the non-convex functional

$$\{\boldsymbol{\alpha}, \boldsymbol{w}\} = \underset{\boldsymbol{\alpha}, \boldsymbol{w}}{\operatorname{argmin}} \left\| g(\boldsymbol{x}) - \boldsymbol{\alpha} \cdot \mathbb{E}\left[\boldsymbol{X}_L | \boldsymbol{X}_0 = \boldsymbol{x}\right] \right\|^2_{L^2([0,1]^d)}, \tag{4}$$

where $\boldsymbol{\alpha}$ are the output weights, and $\mathbb{E}\left[\boldsymbol{X}_L | \boldsymbol{X}_0 = \boldsymbol{x}\right]$ conditional expectation of $\boldsymbol{X}_L$ given $\boldsymbol{X}_0 = \boldsymbol{x}$. In this setting, it is clear that the process of training a neural network is basically an optimal control problem (the controls being the weights and biases) of a discrete stochastic differential equation.

A different stochastic neural network model can be defined by randomizing weights and biases [6, 21]. In this setting we have

$$\boldsymbol{X}_{n+1} = \boldsymbol{\varphi}(\boldsymbol{W}_n(\omega) \boldsymbol{X}_n + \boldsymbol{b}_n(\omega)), \tag{5}$$

where $\boldsymbol{W}_n(\omega)$ are random weight matrices and $\boldsymbol{b}_n(\omega)$ are random bias vectors. We shall assume that $\boldsymbol{W}_n$ and $\boldsymbol{b}_n$ corresponding to different layers are statistically independent.

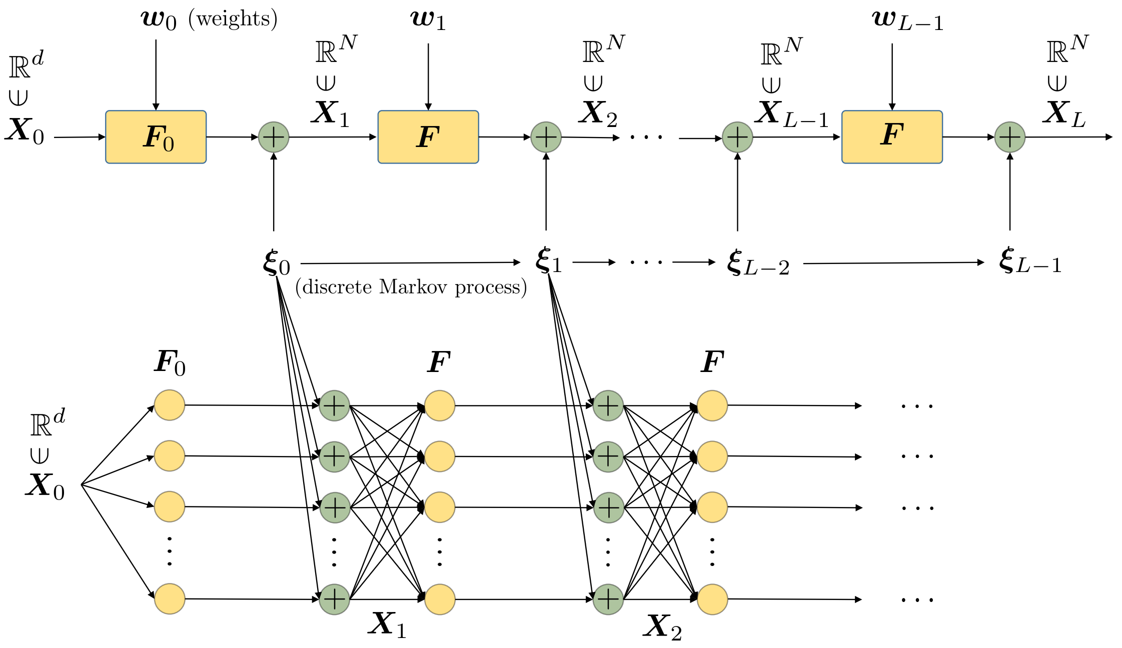

$\boldsymbol{w}_0$ (weights)
$\boldsymbol{w}_1$
$\boldsymbol{w}_{L-1}$
$\mathbb{R}^d \ni \boldsymbol{X}_0$
$\boldsymbol{F}_0$
$\boldsymbol{F}$
$\mathbb{R}^N \ni \boldsymbol{X}_1$
$\mathbb{R}^N \ni \boldsymbol{X}_2$
$\mathbb{R}^N \ni \boldsymbol{X}_{L-1}$
$\mathbb{R}^N \ni \boldsymbol{X}_L$
$\boldsymbol{\xi}_0$
$\boldsymbol{\xi}_1$
$\boldsymbol{\xi}_{L-2}$
$\boldsymbol{\xi}_{L-1}$
(discrete Markov process)
$\boldsymbol{F}_0$
$\boldsymbol{F}$
$\boldsymbol{F}$
$\mathbb{R}^d \ni \boldsymbol{X}_0$
$\boldsymbol{X}_1$
$\boldsymbol{X}_2$

Figure 1: Sketch of the stochastic neural network model (1). Note that the transfer function $\boldsymbol{F}$ is the same in every layer (except the first one). This implies that the random vectors $\{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_L\}$ all have the same dimension.

By adding random noise to the output of each neural network layer, or by randomizing weights and biases, we are essentially adding an infinite number of degrees of freedom to our system. This allows us to rethink the process of training the neural network from a probabilistic perspective. For instance, random noise allows us to approximately encode secret messages in fully trained deterministic neural networks by selecting an appropriate transition probability for the noise process.

Composition and transfer operators

Let us now derive the composition and transfer operators associated with the neural network models (1) and (5), which map, respectively, the conditional expectation $\mathbb{E}\{\boldsymbol{u}(\boldsymbol{X}_L)|\boldsymbol{X}_n = \boldsymbol{x}\}$ and $p_n(\boldsymbol{x})$ (the probability density of $\boldsymbol{X}_n$) forward and backward across the network. To this end, we assume that $\{\boldsymbol{\xi}_0, \ldots, \boldsymbol{\xi}_{L-1}\}$ in (1) are independent random vectors. Similarly, we assume that the random matrices $\{\boldsymbol{W}_0, \ldots, \boldsymbol{W}_{L-1}\}$ and the random vectors $\{\boldsymbol{b}_0, \ldots, \boldsymbol{b}_{L-1}\}$ in (5) are statistically independent. These assumptions imply that the sequences of vectors $\{\boldsymbol{X}_0, \boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_L\}$ generated by either (1) or (5) are *discrete Markov processes*[1]. Therefore, the joint probability density function (PDF) of the random vectors $\{\boldsymbol{X}_0, \ldots, \boldsymbol{X}_L\}$, i.e., joint PDF of the state of the entire neural network, can be factored[2] as

$$p(\boldsymbol{x}_0, \ldots, \boldsymbol{x}_L) = p_{L|L-1}(\boldsymbol{x}_L|\boldsymbol{x}_{L-1}) p_{L-1|L-2}(\boldsymbol{x}_{L-1}|\boldsymbol{x}_{L-2}) \cdots p_{1|0}(\boldsymbol{x}_1|\boldsymbol{x}_0) p_0(\boldsymbol{x}_0). \tag{6}$$

By using the identity

$$p(\boldsymbol{x}_{k+1}, \boldsymbol{x}_k) = p_{k+1|k}(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k) p_k(\boldsymbol{x}_k) = p_{k|k+1}(\boldsymbol{x}_k|\boldsymbol{x}_{k+1}) p_{k+1}(\boldsymbol{x}_{k+1}) \tag{7}$$

[1] The independence assumption assumption of the random noise vector $\{\boldsymbol{\xi}_0, \ldots, \boldsymbol{\xi}_{L-1}\}$ or the random weights and biases $\{\boldsymbol{W}_1, \ldots, \boldsymbol{W}_1\}$ and $\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_m\}$ is not necessary for the process $\{\boldsymbol{X}_0, \boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_L\}$ to be Markov. Such assumption just simplifies the expression of the transition density $p(\boldsymbol{x}_{i+1}|\boldsymbol{x}_i)$.

[2] In equation (6) we used the shorthand notation $p_{i|j}(\boldsymbol{x}|\boldsymbol{y})$ to denote the conditional probability density function of the random vector $\boldsymbol{X}_i$ given $\boldsymbol{X}_j = \boldsymbol{x}_j$. With this notation we have that the conditional probability density of $\boldsymbol{X}_i$ given $\boldsymbol{X}_i = \boldsymbol{y}$ is $p_{i|i}(\boldsymbol{x}|\boldsymbol{y}) = \delta(\boldsymbol{x} - \boldsymbol{y})$, where $\delta(\cdot)$ is the Dirac delta function.

we see that the chain of transition probabilities (6) can be reverted, yielding

$$p(\boldsymbol{x}_0,\ldots,\boldsymbol{x}_L) = p_{0|1}(\boldsymbol{x}_0|\boldsymbol{x}_1)p_{1|2}(\boldsymbol{x}_1|\boldsymbol{x}_2)\cdots p_{L-1|L}(\boldsymbol{x}_{L-1}|\boldsymbol{x}_L)p_L(\boldsymbol{x}_L). \tag{8}$$

From these expression, it follows that

$$p_{n|q}(\boldsymbol{x}|\boldsymbol{y}) = \int p_{n|j}(\boldsymbol{x}|\boldsymbol{z})p_{j|q}(\boldsymbol{z}|\boldsymbol{y})d\boldsymbol{z}, \tag{9}$$

for all indices n, j and q in $\{0,\ldots,L\}$, excluding $n=j=q$ (see footnote 2). The transition probability equation (9) is known as *discrete Chapman-Kolmogorov equation* and it allows us to define the transfer operator mapping the PDF $p_n(\boldsymbol{x}_n)$ into $p_{n+1}(\boldsymbol{x}_{n+1})$, together with the composition operator for the conditional expectation $\mathbb{E}\{\boldsymbol{u}(\boldsymbol{x}_L)|\boldsymbol{X}_n=\boldsymbol{x}_n\}$. As we shall will see hereafter, the discrete composition and transfer operators are adjoint to one another.

Transfer operator. Let us denote by $p_q(\boldsymbol{x})$ the PDF of $\boldsymbol{X}_q$, i.e., the output of the q-th neural network layer. We first define the operator that maps $p_q(\boldsymbol{x})$ into $p_n(\boldsymbol{x})$. By integrating the joint probability density of $\boldsymbol{X}_n$ and $\boldsymbol{X}_q$, i.e., $p_{n|q}(\boldsymbol{x}|\boldsymbol{y})p_q(\boldsymbol{y})$ with respect to $\boldsymbol{y}$ we immediately obtain

$$p_n(\boldsymbol{x}) = \int p_{n|q}(\boldsymbol{x}|\boldsymbol{y})p_q(\boldsymbol{y})d\boldsymbol{y}. \tag{10}$$

At this point, it is convenient to define the operator

$$\mathcal{N}(n,q)f(\boldsymbol{x}) = \int p_{n|q}(\boldsymbol{x}|\boldsymbol{y})f(\boldsymbol{y})d\boldsymbol{y}. \tag{11}$$

$\mathcal{N}(n,q)$ is known as *transfer operator* [3]. From a mathematical viewpoint $\mathcal{N}(n,q)$ is an integral operator with kernel $p_{n|q}(\boldsymbol{x},\boldsymbol{y})$, i.e., the transition density integrated “from the right”. It follows from the Chapman-Kolmogorov identity (9) that the set of integral operators $\{\mathcal{N}(n,q)\}$ forms a group. Namely,

$$\mathcal{N}(n,q) = \mathcal{N}(n,j)\mathcal{N}(j,q), \qquad \mathcal{N}(j,j)=\mathcal{I}, \qquad \forall n,j,q\in\{0,\ldots,L\}. \tag{12}$$

The operator $\mathcal{N}$ allows us to map the one-layer PDF, e.g., the PDF of $\boldsymbol{X}_q$, either forward or backward across the neural network (see Figure 2). As an example, consider a network with four layers with states $\boldsymbol{X}_0$ (input), $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, $\boldsymbol{X}_3$, and $\boldsymbol{X}_4$ (output). Then Eq. (11) implies that,

$$p_2(\boldsymbol{x}) = \underbrace{\mathcal{N}(2,1)\mathcal{N}(1,0)}_{\mathcal{N}(2,0)}p_0(\boldsymbol{x}) = \underbrace{\mathcal{N}(2,3)\mathcal{N}(3,4)}_{\mathcal{N}(2,4)}p_4(\boldsymbol{x}).$$

In summary, we have

$$p_n(\boldsymbol{x}) = \mathcal{N}(n,q)p_q(\boldsymbol{x}) \qquad \forall n,q\in\{0,\ldots,L\}, \tag{13}$$

where

$$\mathcal{N}(n,q)p_q(\boldsymbol{x}) = \int p_{n|q}(\boldsymbol{x}|\boldsymbol{y})p_q(\boldsymbol{y})d\boldsymbol{y}. \tag{14}$$

We emphasize that modeling PDF dynamics via neural networks has been studied extensively in machine learning, e.g., in the theory of normalizing flows for density estimation or variational inference [17, 10, 18].

Composition operator For any measurable deterministic function $\boldsymbol{u}(\boldsymbol{x})$, the conditional expectation of $\boldsymbol{u}(\boldsymbol{X}_j)$ given $\boldsymbol{X}_n=\boldsymbol{x}$ is defined as

$$\mathbb{E}\left\{\boldsymbol{u}(\boldsymbol{X}_j)|\boldsymbol{X}_n=\boldsymbol{x}\right\} = \int \boldsymbol{u}(\boldsymbol{y})p_{j|n}(\boldsymbol{y}|\boldsymbol{x})d\boldsymbol{y}. \tag{15}$$

A substitution of (9) into (15) yields

$$\mathbb{E}\left\{\boldsymbol{u}(\boldsymbol{X}_j)|\boldsymbol{X}_n=\boldsymbol{x}\right\}=\int\mathbb{E}\left\{\boldsymbol{u}(\boldsymbol{X}_j)|\boldsymbol{X}_q=\boldsymbol{y}\right\}p_{q|n}(\boldsymbol{y}|\boldsymbol{x})d\boldsymbol{y},\tag{16}$$

which holds for all $j,n,q\in\{0,\dots,L-1\}$. At this point we define the integral operator

$$\mathcal{M}(n,q)f(\boldsymbol{x})=\int f(\boldsymbol{y})p_{q|n}(\boldsymbol{y}|\boldsymbol{x})d\boldsymbol{y},\tag{17}$$

which is known as *composition* [3] or "stochastic Koopman" [19, 23] operator. Thanks to the Chapman-Kolmogorov identity (9), the set of operators $\{\mathcal{M}(q,j)\}$ forms a group, i.e.,

$$\mathcal{M}(n,q)=\mathcal{M}(n,j)\mathcal{M}(j,q),\qquad\mathcal{M}(j,j)=\mathcal{I},\qquad\forall n,j,q\in\{0,\dots,L\}.\tag{18}$$

Equation (18) allows us to map the conditional expectation (15) of any measurable phase space function $\boldsymbol{u}(\boldsymbol{X}_j)$ forward or backward through the network. As an example, consider again a neural network with four layers and states $\{\boldsymbol{X}_0,\dots,\boldsymbol{X}_4\}$. We have

$$\begin{aligned}\mathbb{E}\{\boldsymbol{u}(\boldsymbol{X}_j)|\boldsymbol{X}_2=\boldsymbol{x}\}=&\mathcal{M}(2,3)\mathcal{M}(3,4)\mathbb{E}\{\boldsymbol{u}(\boldsymbol{X}_j)|\boldsymbol{X}_4=\boldsymbol{x}\}\\=&\mathcal{M}(2,1)\mathcal{M}(1,0)\mathbb{E}\{\boldsymbol{u}(\boldsymbol{X}_j)|\boldsymbol{X}_0=\boldsymbol{x}\}.\end{aligned}\tag{19}$$

Equation (19) holds for every $j\in\{0,..,4\}$. Of particular interest in machine-learning context is the conditional expectation of $\boldsymbol{u}(\boldsymbol{X}_L)$ (network output) given $\boldsymbol{X}_0=\boldsymbol{x}$, which can be computed as

$$\begin{aligned}\mathbb{E}\{\boldsymbol{u}(\boldsymbol{X}_L)|\boldsymbol{X}_0=\boldsymbol{x}\}&=\mathcal{M}(0,L)\boldsymbol{u}(\boldsymbol{x}),\\&=\mathcal{M}(0,1)\mathcal{M}(1,2)\cdots\mathcal{M}(L-1,L)\boldsymbol{u}(\boldsymbol{x}),\end{aligned}\tag{20}$$

i.e., by propagating $\boldsymbol{u}(\boldsymbol{x})=\mathbb{E}\{\boldsymbol{u}(\boldsymbol{X}_L)|\boldsymbol{X}_L=\boldsymbol{x}\}$ *backward* through the neural network using single layer operators $\mathcal{M}(i-1,i)$. Similarly, we can compute, e.g., $\mathbb{E}\{\boldsymbol{u}(\boldsymbol{X}_0)|\boldsymbol{X}_L=\boldsymbol{x}\}$ as

$$\mathbb{E}\{\boldsymbol{u}(\boldsymbol{X}_0)|\boldsymbol{X}_L=\boldsymbol{x}\}=\mathcal{M}(L,0)\boldsymbol{u}(\boldsymbol{x}).\tag{21}$$

For subsequent analysis, it is convenient to define

$$\boldsymbol{q}_n(\boldsymbol{x})=\mathbb{E}\{\boldsymbol{u}(\boldsymbol{X}_L)|\boldsymbol{X}_{L-n}=\boldsymbol{x}\}.\tag{22}$$

In this way, if $\mathbb{E}\{\boldsymbol{u}(\boldsymbol{X}_L)|\boldsymbol{X}_n=\boldsymbol{x}\}$ is propagated *backward* through the network by $\mathcal{M}(n-1,n)$, then $\boldsymbol{q}_n(x)$ is propagated *forward* by the operator

$$\mathcal{G}(n,q)=\mathcal{M}(L-n,L-q).\tag{23}$$

In fact, equations (22)-(23) allow us to write (20) in the equivalent form

$$\begin{aligned}\boldsymbol{q}_L(\boldsymbol{x})=&\mathcal{G}(L,L-1)\boldsymbol{q}_{L-1}(\boldsymbol{x})\\=&\mathcal{G}(L,L-1)\cdots\mathcal{G}(1,0)\boldsymbol{q}_0(\boldsymbol{x}),\end{aligned}\tag{24}$$

i.e., as a forward propagation problem (see Figure 2). Note that we can write (24) (or (20)) explicitly in terms of iterated integrals involving single layer transition densities as

$$\begin{aligned}\boldsymbol{q}_L(\boldsymbol{x})&=\int\boldsymbol{u}(\boldsymbol{y})p_{0|L}(\boldsymbol{y}|\boldsymbol{x})d\boldsymbol{y}\\&=\int\boldsymbol{u}(\boldsymbol{y})\left(\int\cdots\int p_{L|L-1}(\boldsymbol{y}|\boldsymbol{x}_{L-1})\cdots p_{2|1}(\boldsymbol{x}_2|\boldsymbol{x}_1)p_{1|0}(\boldsymbol{x}_1|\boldsymbol{x})d\boldsymbol{x}_{L-1}\cdots d\boldsymbol{x}_1\right)d\boldsymbol{y}.\end{aligned}\tag{25}$$

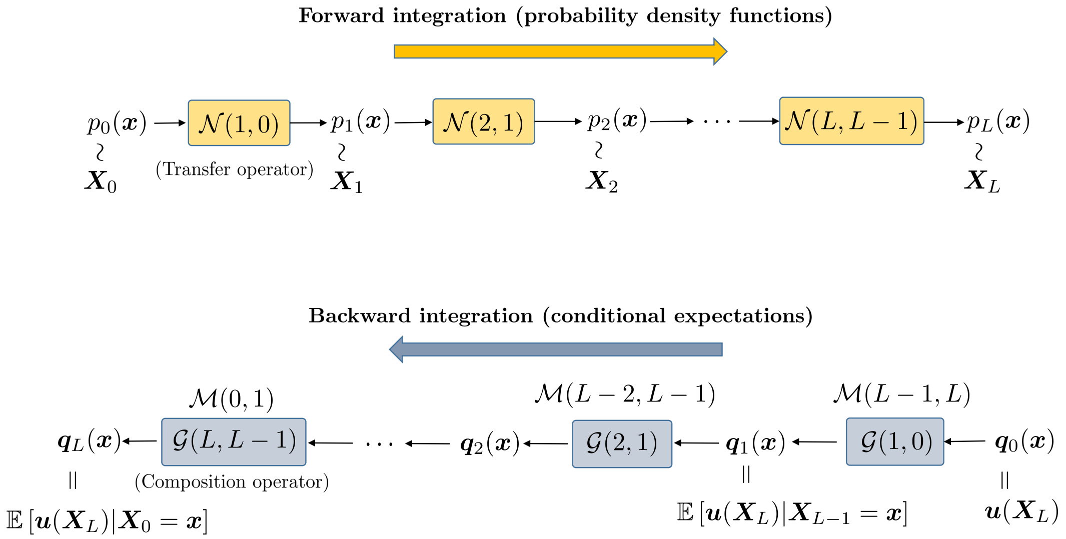

Forward integration (probability density functions)
$p_0(\boldsymbol{x})$
$\mathcal{N}(1,0)$
$p_1(\boldsymbol{x})$
$\mathcal{N}(2,1)$
$p_2(\boldsymbol{x})$
$\cdots$
$\mathcal{N}(L,L-1)$
$p_L(\boldsymbol{x})$
$\boldsymbol{X}_0$
(Transfer operator)
$\boldsymbol{X}_1$
$\boldsymbol{X}_2$
$\boldsymbol{X}_L$
Backward integration (conditional expectations)
$\mathcal{M}(0,1)$
$\mathcal{M}(L-2,L-1)$
$\mathcal{M}(L-1,L)$
$\boldsymbol{q}_L(\boldsymbol{x})$
$\mathcal{G}(L,L-1)$
$\cdots$
$\boldsymbol{q}_2(\boldsymbol{x})$
$\mathcal{G}(2,1)$
$\boldsymbol{q}_1(\boldsymbol{x})$
$\mathcal{G}(1,0)$
$\boldsymbol{q}_0(\boldsymbol{x})$
(Composition operator)
$\mathbb{E}\left[\boldsymbol{u}(\boldsymbol{X}_L)|\boldsymbol{X}_0=\boldsymbol{x}\right]$
$\mathbb{E}\left[\boldsymbol{u}(\boldsymbol{X}_L)|\boldsymbol{X}_{L-1}=\boldsymbol{x}\right]$
$\boldsymbol{u}(\boldsymbol{X}_L)$

Figure 2: Sketch of the forward/backward integration process for probability density functions (PDFs) and conditional expectations. The transfer operator $\mathcal{N}(n+1,n)$ maps the PDF $p_n(\boldsymbol{x})$ of the state $\boldsymbol{X}_n$ into $p_{n+1}(\boldsymbol{x})$ forward through the neural network. On the other hand, the composition operator $\mathcal{M}$ maps the conditional expectation $\mathbb{E}\left[\boldsymbol{u}(\boldsymbol{X}_L)|\boldsymbol{X}_{n+1}=\boldsymbol{x}\right]$ backwards to $\mathbb{E}\left[\boldsymbol{u}(\boldsymbol{X}_L)|\boldsymbol{X}_n=\boldsymbol{x}\right]$. By defining the operator $\mathcal{G}(n,m)=\mathcal{M}(L-n,L-m)$ we can transform the backward propagation problem for $\mathbb{E}\left[\boldsymbol{u}(\boldsymbol{X}_L)|\boldsymbol{X}_n=\boldsymbol{x}\right]$ into a forward propagation problem for $\boldsymbol{q}_n(\boldsymbol{x})=\mathbb{E}\left[\boldsymbol{u}(\boldsymbol{X}_L)|\boldsymbol{X}_{L-n}=\boldsymbol{x}\right]$.

Relation between composition and transfer operators. The integral operators $\mathcal{M}$ and $\mathcal{N}$ defined in (17) and (11) involve the same kernel function, i.e., the multi-layer transition probability density $p_{q|n}(\boldsymbol{x},\boldsymbol{y})$. In particular, we noticed that $\mathcal{M}(n,q)$ integrates $p_{q|n}$ "from the left", while $\mathcal{N}(q,n)$ integrates it "from the right". It is easy to show that $\mathcal{M}(n,q)$ and $\mathcal{N}(q,n)$ are adjoint to each other relative to the standard inner product in L^2 (see [3] for the continuous-time case). In fact,

$$\begin{aligned}\mathbb{E}\{\boldsymbol{u}(\boldsymbol{X}_k)\} &= \int \mathbb{E}\{\boldsymbol{u}(\boldsymbol{X}_k)|\boldsymbol{X}_q=\boldsymbol{x}\}p_q(\boldsymbol{x})d\boldsymbol{x}\\ &= \int \left[\mathcal{M}(q,j)\mathbb{E}\{\boldsymbol{u}(\boldsymbol{X}_k)|\boldsymbol{X}_j=\boldsymbol{x}\}\right]p_q(\boldsymbol{x})d\boldsymbol{x}\\ &= \int \mathbb{E}\{\boldsymbol{u}(\boldsymbol{X}_k)|\boldsymbol{X}_j=\boldsymbol{x}\}\mathcal{N}(j,q)p_q(\boldsymbol{x})d\boldsymbol{x}. \end{aligned} \tag{26}$$

Therefore

$$\mathcal{M}(q,j)^* = \mathcal{N}(j,q) \qquad \forall q,j\in\{0,\ldots,L\}, \tag{27}$$

where $\mathcal{M}(q,j)^*$ denotes the operator adjoint of $\mathcal{M}(q,j)$ with respect to the L^2 inner product. By invoking the definition (23), we can also write (27) as

$$\mathcal{G}(L-q,L-j)^* = \mathcal{N}(j,q), \qquad \forall j,q\in\{0,\ldots,L\}. \tag{28}$$

In Appendix A we show that if the cumulative distribution function of each random vector $\boldsymbol{\xi}_n$ in the noise process has partial derivatives that are Lipschitz in $\mathscr{R}(\boldsymbol{\xi}_n)$ (range of $\boldsymbol{\xi}_n$), then the composition and transfer operators defined in Eqs. (17) and 11 are *bounded* in L^2 (see Proposition 0.9 and Proposition 0.10). Moreover, is possible to choose the probability density of $\boldsymbol{\xi}_n$ such that the single layer composition and transfer operators become strict *contractions*.

Conditional transition density

We have seen that the composition and the transfer operators $\mathcal{M}$ and $\mathcal{N}$ defined in Eqs. (17) and (11), allow us to push forward and backward conditional expectations and probability densities across the entire neural network. Moreover such operators are adjoint to one another (see equation (27)) [3, 22, 2], and also have the same kernel, i.e., the transition density $p_{n|q}(\boldsymbol{x}_n|\boldsymbol{x}_q)$. In this section, we determine an explicit expression for such transition density. To this end, we fist derive analytical formulas for the one-layer transition density $p_{n+1|n}(\boldsymbol{x}_{n+1}|\boldsymbol{x}_n)$ for various types of neural network models. The multi-layer transition density $p_{n|q}(\boldsymbol{x}_n|\boldsymbol{x}_q)$ is then obtained by composing one-layer transition densities as follows

$$p_{n|q}(\boldsymbol{x}_n|\boldsymbol{x}_q) = \int \cdots \int p_{n|n-1}(\boldsymbol{x}_n|\boldsymbol{x}_{n-1}) \cdots p_{q+1|q}(\boldsymbol{x}_{q+1}|\boldsymbol{x}_q) d\boldsymbol{x}_{n-1} \cdots d\boldsymbol{x}_{q+1}. \tag{29}$$

Neural network with additive noise. Let us first consider the neural network model

$$\boldsymbol{X}_{n+1} = \boldsymbol{F}(\boldsymbol{X}_n, \boldsymbol{w}_n) + \boldsymbol{\xi}_n, \tag{30}$$

where $\{\boldsymbol{\xi}_0, \boldsymbol{\xi}_1, \ldots\}$ is a discrete (vector-valued) Markov process indexed by "n". By (30), $\boldsymbol{X}_{n+1}$ is the sum of two *independent* random vectors[3], i.e., $\boldsymbol{F}(\boldsymbol{X}_n, \boldsymbol{w}_n)$ and $\boldsymbol{\xi}_n$. Given any measurable function $h(\boldsymbol{x})$ we clearly have

$$\begin{aligned}\mathbb{E}\left\{h(\boldsymbol{X}_{n+1})\right\} &= \int h(\boldsymbol{x}) p_{n+1}(\boldsymbol{x}) d\boldsymbol{x}\\ &= \int\int h(\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{w}_n) + \boldsymbol{\xi}) p_n(\boldsymbol{x}) \rho_n(\boldsymbol{\xi}) d\boldsymbol{x} d\boldsymbol{\xi}\\ &= \int\int h(\boldsymbol{y})) \underbrace{\rho_n(\boldsymbol{x} - \boldsymbol{F}(\boldsymbol{y}, \boldsymbol{w}_n))}_{p_{n+1|n}(\boldsymbol{x}|\boldsymbol{y})} p_n(\boldsymbol{x}) d\boldsymbol{y} d\boldsymbol{x},\end{aligned} \tag{31}$$

where $\rho_n(\boldsymbol{x})$ denotes the probability density of the random vector $\boldsymbol{\xi}_n$. Therefore, the one-layer transition density for the neural network model (30) is[4]

$$p_{n+1|n}(\boldsymbol{x}_{n+1}|\boldsymbol{x}_n) = \rho_n(\boldsymbol{x}_{n+1} - \boldsymbol{F}(\boldsymbol{x}_{\boldsymbol{n}}, \boldsymbol{w}_n)). \tag{34}$$

Note that such transition density depends on the PDF of the random noise ρ_n, the neural activation function $\boldsymbol{F}$, and the neural network weights $\boldsymbol{w}_n$.

Neural network with random weights and random biases. Next, consider the recurrent neural network model

$$\boldsymbol{X}_{n+1} = \boldsymbol{\varphi}(\boldsymbol{W}_n(\omega)\boldsymbol{X}_n + \boldsymbol{b}_n(\omega)), \tag{35}$$

[3] Recall that $\boldsymbol{X}_n$ and $\boldsymbol{\xi}_n$ are statistically independent random vectors. Hence, $\boldsymbol{F}(\boldsymbol{X}_n, \boldsymbol{w}_n)$ and $\boldsymbol{\xi}_n$ are statistically independent random vectors.

[4] Equation (34) can be derived in a more general setting by recalling the conditional probability identity

$$p_{n+1|n}(\boldsymbol{x}_{n+1}|\boldsymbol{x}_n) = \int p_{\boldsymbol{X}_{n+1}|\boldsymbol{X}_n,\boldsymbol{\xi}_n}(\boldsymbol{x}_{n+1}|\boldsymbol{x}_n, \boldsymbol{z}) \rho_n(\boldsymbol{z}) d\boldsymbol{z}. \tag{32}$$

The conditional density of $\boldsymbol{X}_{n+1}$ given $\boldsymbol{X}_n = \boldsymbol{x}_n$ and $\boldsymbol{\xi}_n = \boldsymbol{z}$, i.e., $p_{\boldsymbol{X}_{n+1}|\boldsymbol{X}_n,\boldsymbol{\xi}_n}(\boldsymbol{x}_{n+1}|\boldsymbol{x}_n, \boldsymbol{z})$, can be immediately computed by using (30) as

$$p_{\boldsymbol{X}_{n+1}|\boldsymbol{X}_n,\boldsymbol{\xi}_n}(\boldsymbol{x}_{n+1}|\boldsymbol{x}_n, \boldsymbol{z}) = \delta(\boldsymbol{x}_{n+1} - \boldsymbol{F}(\boldsymbol{x}_n, \boldsymbol{w}_n) - \boldsymbol{z}), \tag{33}$$

where $\delta(\boldsymbol{x})$ is the multivariate Dirac delta function. Substituting (33) into (32), and integrating over $\boldsymbol{z}$ yields (34).

where $\boldsymbol{W}_n(\omega)$ are random weight matrices and $\boldsymbol{b}_n(\omega)$ are random bias vectors. The PDF of $\boldsymbol{X}_{n+1}$ given $\boldsymbol{X}_n = \boldsymbol{x}_n$, i.e., the conditional density we are interested in, can be obtained first by computing the PDF of the random vector

$$\boldsymbol{Z}_n(\omega) = \boldsymbol{W}_n(\omega)\boldsymbol{X}_n + \boldsymbol{b}_n(\omega), \tag{36}$$

i.e., a linear mapping between independent random variables, and then computing the PDF of $\boldsymbol{X}_{n+1} = \boldsymbol{\varphi}(\boldsymbol{Z}_n)$, where $\boldsymbol{\varphi}$ is the (invertible) activation function. By using the methods we have seen in chapter 1, it is rather straightforward to obtain and expression for the conditional density of $\boldsymbol{X}_{n+1}$ given $\boldsymbol{X}_n = \boldsymbol{x}_n$ for specific probability distributions of $\boldsymbol{W}_n(\omega)$ (random matrix ensembles) and $\boldsymbol{b}_n$.

General neural networks. The transition density of a general neural network of the form

$$\boldsymbol{X}_{n+1} = \boldsymbol{H}(\boldsymbol{X}_n, \boldsymbol{w}_n, \boldsymbol{\xi}_n), \tag{37}$$

where $\{\boldsymbol{\xi}_0, \ldots, \boldsymbol{\xi}_{L-1}\}$ are statistically independent and do not depend on $\{\boldsymbol{X}_j\}$ can be written as (see, e.g., [7])

$$p(\boldsymbol{x}_{n+1}|\boldsymbol{x}_n) = \int \underbrace{\delta\left(\boldsymbol{x}_{n+1} - \boldsymbol{H}(\boldsymbol{x}_n, \boldsymbol{w}_n, \boldsymbol{\xi}_n)\right)}_{p(\boldsymbol{x}_{n+1}|\boldsymbol{x}_n,\boldsymbol{\xi}_n)} p(\boldsymbol{\xi}_n) d\boldsymbol{\xi}_n. \tag{38}$$

Remark: The transition density (34) associated with the neural network model (30) can be computed explicitly once we choose a probability model for $\boldsymbol{\xi}_n \in \mathbb{R}^N$. For instance, if we assume that $\{\boldsymbol{\xi}_0, \boldsymbol{\xi}_1, \ldots, \}$ are i.i.d. Gaussian random vectors with PDF

$$\rho_n(\boldsymbol{\xi}) = \frac{1}{(2\pi)^{N/2}} e^{-\boldsymbol{\xi}^T\boldsymbol{\xi}/2} \qquad \text{for all } n = 0, \ldots, L \tag{39}$$

then we can explicitly write the one-layer transition density (34) as

$$p_{n+1|n}(\boldsymbol{x}_{n+1}|\boldsymbol{x}_n) = \frac{1}{(2\pi)^{N/2}} \exp\left[-\frac{\left[\boldsymbol{x}_{n+1} - \boldsymbol{F}(\boldsymbol{x}_n, \boldsymbol{w}_n)\right]^T \left[\boldsymbol{x}_{n+1} - \boldsymbol{F}(\boldsymbol{x}_n, \boldsymbol{w}_n)\right]}{2}\right]. \tag{40}$$

In Appendix A we provide an analytical example of transition density for a neural network with two layers (one neuron per layer), $\tanh(\cdot)$ activation function, and uniformly distributed random noise.

The zero noise limit An important question is what happens to the neural network as we send the amplitude of the noise to zero. To answer this question consider the system (30) and the introduce the parameter $\epsilon \geq 0$, i.e.,

$$\boldsymbol{X}_{n+1} = \boldsymbol{F}(\boldsymbol{X}_n, \boldsymbol{w}_n) + \epsilon\boldsymbol{\xi}_n, \tag{41}$$

We are interested in studying the orbits of this system as $\epsilon \to 0$. To this end, we assume the $\boldsymbol{\xi}_n$ to be independent random vectors each having the same density $\rho(\boldsymbol{x})$. This implies that for all $n = 0, \ldots, L-1$, the PDF of $\epsilon\boldsymbol{\xi}_n$ is

$$\epsilon\boldsymbol{\xi}_n \sim \frac{1}{\epsilon}\rho_n\left(\frac{\boldsymbol{x}}{\epsilon}\right). \tag{42}$$

It is shown in [11, Proposition 10.6.1] that the operator $\mathcal{N}(n+1, n)$ defined in (11)

$$\begin{aligned} p_{n+1}(\boldsymbol{x}) =& \mathcal{N}(n+1, n)p_n(\boldsymbol{x}) \\ =& \int \frac{1}{\epsilon}\rho_n\left(\frac{\boldsymbol{x} - \boldsymbol{F}(\boldsymbol{z}, \boldsymbol{w}_n)}{\epsilon}\right) p_n(\boldsymbol{z}) d\boldsymbol{z} \end{aligned} \tag{43}$$

converges in norm to the Frobenious-Perron operator corresponding to $\boldsymbol{F}(\boldsymbol{X}_n, \boldsymbol{w}_n)$ as $\epsilon \to 0$. Indeed, in the limit $\epsilon \to 0$ we have, formally

$$\lim_{\epsilon\to 0} p_{n|n+1}\left(\boldsymbol{x}_{n+1}|\boldsymbol{x}_n\right) = \lim_{\epsilon\to 0}\int \frac{1}{\epsilon}\rho_n\left(\frac{\boldsymbol{x}_{n+1}-\boldsymbol{F}(\boldsymbol{x}_n,\boldsymbol{w}_n)}{\epsilon}\right) = \delta\left(\boldsymbol{x}_{n+1}-\boldsymbol{F}(\boldsymbol{x}_n,\boldsymbol{w}_n)\right). \tag{44}$$

Substituting this expression into (11), one gets,

$$p_{n+1}(\boldsymbol{x}) = \mathcal{N}(n+1,n)p_n(\boldsymbol{x}) = \int \delta\big(\boldsymbol{x}-\boldsymbol{F}(\boldsymbol{z},\boldsymbol{w}_n)\big)p_n(\boldsymbol{z})d\boldsymbol{z}. \tag{45}$$

Similarly, a substitution into equation (24) yields

$$\boldsymbol{q}_{n+1}(\boldsymbol{x}) = \mathcal{G}(n+1,n)\boldsymbol{q}_n(\boldsymbol{x}) = q_n\left(\boldsymbol{F}(\boldsymbol{x},\boldsymbol{w}_{L-n-1})\right), \tag{46}$$

i.e, the familiar function composition representation of neural network mappings

$$\boldsymbol{q}_{n+1} = \boldsymbol{q}_0\Big(\boldsymbol{F}\big(\boldsymbol{F}(\cdots\boldsymbol{F}(\boldsymbol{x},\boldsymbol{w}_{L-n})\cdots,\boldsymbol{w}_{L-1}),\boldsymbol{w}_L\big)\Big). \tag{47}$$

Training over weights versus training over noise

By adding random noise to the output of each layer in a neural network we are essentially adding an infinite number of degrees of freedom to our system. This allows us to rethink the process of training the neural network from a probabilistic perspective. In particular, instead of optimizing a performance metric[5] relative to the neural network weights $\boldsymbol{w} = \{\boldsymbol{w}_0, \boldsymbol{w}_1, \ldots, \boldsymbol{w}_{L-1}\}$ for fixed noise, we can now optimize the transition density[6] $p_{n+1|n}(\boldsymbol{x}_{n+1}|\boldsymbol{x}_n)$. Clearly, such transition density is connected to the neural network weights, e.g., by equation (34). Hence, if we prescribe the PDF of the random noise, i.e., $\rho_n(\cdot)$ in (34), then the transition density $p_{n+1|n}$ is uniquely determined by the functional form of the activation function $\boldsymbol{F}$, and by the weights $\boldsymbol{w}_n$. On the other hand, we can *optimize* ρ_n (probability density of the random noise $\boldsymbol{\xi}_n$) while keeping the weights $\boldsymbol{w}_n$ *fixed*. As we shall see hereafter, this process opens the possibility approximately encode and decode secret messages in a fully trained neural network using random noise.

Encoding secret messages in neural networks using random noise. An interesting question is whether random noise added to the output of each layer in the neural network can enhance features of the output, or allow us to encode/decode secret signals in the network. The interaction between random noise and the nonlinear dynamics modeled by the network can yield indeed many surprising results. For example, in stochastic resonance [16, 20] it is well known that random noise added to a properly tuned bi-stable system can induce a peak in the Fourier power spectrum of the output, hence effectively amplifying the signal. Similarly, random noise added to a neural network can have remarkable effects. In particular, it allows us to re-purpose (to some extent) a previously trained network by hiding a secret signal in it, which can be approximately encoded and decoded by using random noise. To this end, it is sufficient to optimize

[5] In a supervised learning setting the neural network weights are usually determined by minimizing a dissimilarity measure between the output of the network and a target function. Such measure may be an entropy measure, the Wasserstein distance, the Kullback–Leibler divergence, or other measures defined by classical L^p norms.

[6] In a deterministic setting, the transition density for a neural network model of the form $\boldsymbol{X}_{n+1} = \boldsymbol{F}(\boldsymbol{X}_n, \boldsymbol{w}_n)$ is simply

$$p_{n+1|n}(\boldsymbol{x}_{n+1}|\boldsymbol{x}_n) = \delta\left(\boldsymbol{x}_{n+1}-\boldsymbol{F}(\boldsymbol{x}_n,\boldsymbol{w}_n)\right), \tag{48}$$

where $\delta(\cdot)$ is the Dirac delta function. Such density does not have any degree of freedom other than $\boldsymbol{w}_n$. On the other hand, in a stochastic setting we are free to *choose* the PDF of $\boldsymbol{\xi}_n$. For a neural network model of the form $\boldsymbol{X}_{n+1} = \boldsymbol{F}(\boldsymbol{X}_n, \boldsymbol{w}_n) + \boldsymbol{\xi}_n$ the transition density has the form

$$p_{n+1|n}(\boldsymbol{x}_{n+1}|\boldsymbol{x}_n) = \rho_n\left(\boldsymbol{x}_{n+1}-\boldsymbol{F}(\boldsymbol{x}_n,\boldsymbol{w}_n)\right), \tag{49}$$

where $\rho_n(\boldsymbol{\xi})$ is the PDF of $\boldsymbol{\xi}_n$. We are clearly free to choose the functional form of ρ_n.

the PDF of the noise appropriately, and then train the conditional expectation of the output over such PDF.

To describe the method, suppose that we are given a fully trained deterministic neural network with only two layers, and weights chosen to represent an input-output map defined on some domain $\Omega \subseteq \mathbb{R}^d$. In the absence of noise we can write the output of the neural network as

$$q_2(\boldsymbol{x}) = \boldsymbol{\alpha}^T \boldsymbol{F}(\boldsymbol{F}_0(\boldsymbol{x}, \boldsymbol{w}_0), \boldsymbol{w}_1) \tag{50}$$

where $\{\boldsymbol{\alpha}, \boldsymbol{w}_0, \boldsymbol{w}_1\}$ can be optimized to minimize the distance between $q_2(\boldsymbol{x})$ and a given target function $f(\boldsymbol{x})$ $(\boldsymbol{x} \in \Omega)$. Injecting noise $\boldsymbol{\xi}_0$ in the output of the first layer yields the input-output map

$$h_2(\boldsymbol{x}) = \boldsymbol{\alpha}^T \int \boldsymbol{F}(\boldsymbol{y} + \boldsymbol{F}_0(\boldsymbol{x}, \boldsymbol{w}_0), \boldsymbol{w}_1)\rho_0\left(\boldsymbol{y}\right) d\boldsymbol{y}, \tag{51}$$

where $\{\boldsymbol{\alpha}, \boldsymbol{w}_1, \boldsymbol{w}_0\}$ here are fixed, and ρ_0 is the PDF of $\boldsymbol{\xi}_0$. Equation (51) resembles a Fredholm integral equation of the first kind. In fact, it can be written as

$$h_2(\boldsymbol{x}) = \int \kappa_2(\boldsymbol{x}, \boldsymbol{y})\rho_0\left(\boldsymbol{y}\right) d\boldsymbol{y}, \tag{52}$$

where

$$\kappa_2(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{\alpha}^T \boldsymbol{F}(\boldsymbol{y} + \boldsymbol{F}_0(\boldsymbol{x}, \boldsymbol{w}_0), \boldsymbol{w}_1). \tag{53}$$

However, differently from standard Fredholm equations of the first kind, here we have $\boldsymbol{x} \in \Omega \subseteq \mathbb{R}^d$ while $\boldsymbol{y} \in \mathbb{R}^N$, i.e., the integral operator with kernel κ_2 maps functions in N variables into functions in d variables. We are interested in finding a PDF $\rho_0(\boldsymbol{y})$ that solves (51) for a given function $h_2(\boldsymbol{x})$. In other words, we are re-purposing the neural network (50) with output $q_2(\boldsymbol{x}) \simeq f(\boldsymbol{x})$ to approximate now a different function $h_2(\boldsymbol{x}) \simeq v(\boldsymbol{x})$, without modifying the weights $\{\boldsymbol{\alpha}, \boldsymbol{w}_1, \boldsymbol{w}_0\}$ but rather simply adding noise $\boldsymbol{\xi}_0$ and averaging the output over the PDF ρ_0 of the noise (Eq. (52)). Equation (52) is unfortunately ill-posed in the space of probability distributions. In other words, for a given kernel κ_2 and a given target q_2 there is (in general) no PDF ρ_0 that satisfies (52) exactly. However, one can proceed by optimization. For instance, ρ_0 can be determined by solving the constrained least squares problem[7]

$$\rho_0 = \underset{\rho}{\operatorname{argmin}} \left\| h_2(\boldsymbol{x}) - \int \kappa_2(\boldsymbol{x}, \boldsymbol{y})\rho(\boldsymbol{y}) d\boldsymbol{y} \right\|_{L^2(\Omega)} \qquad \text{subject to} \qquad \|\rho\|_{L^1(\mathbb{R}^N)} = 1 \qquad \rho \geq 0. \tag{54}$$

[7] The optimization problem (54) is a quadratic program with linear constraints if we represent ρ_0 in the span of a basis made of positive functions, e.g., Gaussian kernels [1].

Appendix A: Functional setting

Let $(\mathcal{S}, \mathcal{F}, \mathscr{P})$ be a probability space. Consider the neural network model (see Figure 1)

$$\boldsymbol{X}_1 = \boldsymbol{F}_0(\boldsymbol{X}_0, \boldsymbol{w}_0) + \boldsymbol{\xi}_0 \qquad \boldsymbol{X}_{n+1} = \boldsymbol{F}(\boldsymbol{X}_n, \boldsymbol{w}_n) + \boldsymbol{\xi}_n \qquad n = 1, \dots, L-1, \tag{55}$$

where $\{\boldsymbol{\xi}_0, \dots, \boldsymbol{\xi}_{L-1}\}$ is a discrete (vector-valued) Markov process. Suppose we are interested in using the model (55) to approximate a multivariate function $f(\boldsymbol{x})$. This is usually done by taking a linear combination of the network output, e.g., Eq. (4). In this setting, the neural network can be thought of as a process of constructing an adaptive basis by function composition. Here we consider the case where the function we are approximating is defined on a compact subset Ω_0 of $\mathbb{R}^d$. This means that the input vector of the neural network, i.e. $\boldsymbol{X}_0$, is and element of Ω_0. We assume that the following conditions are satisfied

1. $\boldsymbol{X}_0 \in \Omega_0 \subseteq \mathbb{R}^d$ (Ω compact), $\boldsymbol{X}_n \in \mathbb{R}^N$ for $n = 1, \dots, L-1$;
2. The image of $\boldsymbol{F}_0$ and $\boldsymbol{F}$ is the hyper-cube $[-1, 1]^N$.

For example, if $\boldsymbol{F}$ in (55) is of the form

$$\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{w}) = \tanh(\boldsymbol{W}\boldsymbol{x} + \boldsymbol{b}) \qquad \boldsymbol{w} = \{\boldsymbol{W}, \boldsymbol{b}\}, \tag{56}$$

then conditions 1. and 2. imply that $\boldsymbol{W}_0 \in M_{N\times d}(\mathbb{R})$ and $\boldsymbol{W}_n \in M_{N\times N}(\mathbb{R})$ for $n = 1, \dots L-1$, while the biases are $\boldsymbol{b}_n \in M_{N\times 1}(\mathbb{R})$ for $n = 0, \dots L-1$. The random vectors $\{\boldsymbol{\xi}_0, \dots, \boldsymbol{\xi}_{L-1}\}$ added to the output of each layer make $\{\boldsymbol{X}_1, \dots, \boldsymbol{X}_K\}$ a discrete Markov process (each $\boldsymbol{X}_i$ is a random vector). The range of $\boldsymbol{X}_{n+1}$ depends essentially on the range of $\boldsymbol{\xi}_n$, as the image of $\boldsymbol{F}$ is the hyper-cube $[-1, 1]^N$ (see condition 2. above). Let us define[8]

$$\begin{aligned}\Omega_{n+1} =& [-1, 1]^N + \mathscr{R}(\boldsymbol{\xi}_n)\\ =& \{\boldsymbol{c} \in \mathbb{R}^N \,:\, \boldsymbol{c} = \boldsymbol{a} + \boldsymbol{b} \quad \boldsymbol{a} \in [-1, 1]^N, \boldsymbol{b} \in \mathscr{R}(\boldsymbol{\xi}_n)\},\end{aligned} \tag{58}$$

where $\mathscr{R}(\boldsymbol{\xi}_n)$ denotes the range of the random vector $\boldsymbol{\xi}_n$, i.e.,

$$\mathscr{R}(\boldsymbol{\xi}_n) = \{\boldsymbol{\xi}_n(\omega) \in \mathbb{R}^N : \omega \in \mathcal{S}\}. \tag{59}$$

Clearly, the range of the random vector $\boldsymbol{X}_{n+1}$ is a subset[9] of Ω_{n+1}, i.e., $\mathscr{R}(\boldsymbol{X}_{n+1}) \subseteq \Omega_{n+1}$. This implies the following lemma.

Lemma 0.1. *Let* $\lambda(\Omega_{n+1})$ *the Lebesgue measure of the set* (58). *The Lebesgue measure of the range of* $\boldsymbol{X}_{n+1}$ *satisfies*

$$\lambda(\mathscr{R}(\boldsymbol{X}_{n+1})) \leq \lambda(\Omega_{n+1}). \tag{60}$$

Proof. The proof follows from the inclusion $\mathscr{R}(\boldsymbol{X}_{n+1}) \subseteq \Omega_{n+1}$. □

[8] The notation $[-1, 1]^N$ denotes a Cartesian product of N one-dimensional domains $[-1, 1]$, i.e.,

$$[-1, 1]^N = \bigtimes_{k=1}^{N} [-1, 1] = \underbrace{[-1, 1] \times [-1, 1] \times \cdots \times [-1, 1]}_{N \text{ times}}. \tag{57}$$

[9] We emphasize that if we are given a specific form of the activation function $\boldsymbol{F}$ together with suitable bounds on the neural network weights and biases $\{\boldsymbol{W}, \boldsymbol{b}\}$ then we can easily identify a domain that is smaller than Ω_n, and that still contains $\mathscr{R}(\boldsymbol{X}_n)$. This allows us to construct a tighter bound for $\lambda(\mathscr{R}(\boldsymbol{X}_{n+1})$ in Lemma 0.1, which depends on the activation function and on the bounds we set on neural network weights and biases.

The L^∞ norm of the random vector $\boldsymbol{\xi}$ is defined as the largest value of $r \geq 0$ that yields a nonzero probability on the event $\{\omega \in \mathcal{S} : \|\boldsymbol{\xi}(\omega)\|_\infty > r\} \in \mathcal{F}$, i.e.,

$$\|\boldsymbol{\xi}\|_\infty = \sup_{r\in\mathbb{R}}\{ \mathscr{P}(\{\omega \in \mathcal{S} : \|\boldsymbol{\xi}(\omega)\|_\infty > r\}) > 0\}. \tag{61}$$

This definition allows us to bound the Lebesgue measure of Ω_{n+1} as follows.
Proposition 0.2. *The Lebesgue measure of the set Ω_{n+1} defined in (58) can be bounded as*

$$\lambda(\Omega_{n+1}) \leq \left(\sqrt{N} + \|\boldsymbol{\xi}_n\|_\infty\right)^N \frac{\pi^{N/2}}{\Gamma(1+N/2)}, \tag{62}$$

where N is the number of neurons and $\Gamma(\cdot)$ is the Gamma function.

Proof. As is well known, the length of the diagonal of the hypercube $[-1,1]^N$ is $\sqrt{N}$. Hence, $\sqrt{N} + \|\boldsymbol{\xi}_n\|_\infty$ is the radius of a ball that encloses all elements of Ω_{n+1}. The Lebesgue measure of such ball is obtained by multiplying the Lebesgue measure of the unit ball in $\mathbb{R}^N$, i.e., $\pi^{N/2}/\Gamma(1+N/2)$ by the scaling factor $\left(\sqrt{N} + \|\boldsymbol{\xi}_n\|_\infty\right)^N$.

□

Lemma 0.3. *If $\mathscr{R}(\boldsymbol{\xi}_n)$ is bounded then $\mathscr{R}(\boldsymbol{X}_{n+1})$ is bounded.*

Proof. The image of the activation function $\boldsymbol{F}$ is a bounded set. If $\mathscr{R}(\boldsymbol{\xi}_n)$ is bounded then Ω_{n+1} in (58) is bounded. $\mathscr{R}(\boldsymbol{X}_{n+1}) \subseteq \Omega_{n+1}$ and therefore $\mathscr{R}(\boldsymbol{X}_{n+1})$ is bounded.

□

Clearly, if $\{\boldsymbol{\xi}_0, \ldots, \boldsymbol{\xi}_{L-1}\}$ are i.i.d. random variables then there exists a domain $V = \Omega_1 = \cdots = \Omega_L$ such that

$$\mathscr{R}(\boldsymbol{\xi}_n) \subseteq \mathscr{R}(\boldsymbol{X}_{n+1}) \subseteq V \qquad \forall n = 0, \ldots, L-1. \tag{63}$$

In fact, if $\{\boldsymbol{\xi}_0, \ldots, \boldsymbol{\xi}_{L-1}\}$ are i.i.d. random variables then we have

$$\mathscr{R}(\boldsymbol{\xi}_0) = \mathscr{R}(\boldsymbol{\xi}_1) = \cdots = \mathscr{R}(\boldsymbol{\xi}_{L-1}), \tag{64}$$

which implies that all of Ω_i defined in (58) are the same. If the range of each random vector $\boldsymbol{\xi}_n$ is a tensor product of one-dimensional domain, e.g., if the components of $\boldsymbol{\xi}_n$ are statistically independent, then $V = \Omega_1 = \cdots = \Omega_L$ becomes particularly simple, i.e., a hypercube.

Lemma 0.4. *Let $\{\boldsymbol{\xi}_0, \ldots, \boldsymbol{\xi}_{L-1}\}$ be i.i.d. random variables with bounded range and suppose that each $\boldsymbol{\xi}_k$ has statistically independent components with range $[a,b]$. Then all domains $\{\Omega_1, \ldots, \Omega_L\}$ defined in equation (58) are the same, and they are equivalent to*

$$V = \bigtimes_{k=1}^{N} [-1+a, 1+b]. \tag{65}$$

V includes the range of all random vectors $\boldsymbol{X}_n$ $(n = 1, \ldots, L)$ and has Lebesgue measure

$$\lambda(V) = (2+b-a)^N. \tag{66}$$

Proof. The proof is trivial and therefore omitted.

□

Remark: It is worth noticing that if each $\boldsymbol{\xi}_k$ is a uniformly distributed random vector with statistically independent components in $[-1,1]$, then for $N=10$ (number of neurons) the upper bound in (62) is 3.98×10^6 while the exact result (66) gives 1.05×10^6. Hence the estimate (62) is quite sharp in the case of uniform random vectors.

Boundedness of composition and transfer operators

Lemma 0.3 states that if we perturb the output of the n-th layer of a neural network by a random vector $\boldsymbol{\xi}_n$ with finite range then we obtain a random vector $\boldsymbol{X}_{n+1}$ with finite range. In this hypothesis it is straightforward to show that that the composition and transfer operators defined in (17) and (11) are bounded. We have seen that these operators can be written as

$$\mathcal{M}(n,n+1)v=\int_{\mathscr{R}(\boldsymbol{X}_{n+1})}v(\boldsymbol{y})p_{n+1|n}(\boldsymbol{y}|\boldsymbol{x})d\boldsymbol{y},\qquad \mathcal{N}(n+1,n)v=\int_{\mathscr{R}(\boldsymbol{X}_{n})}p_{n+1|n}(\boldsymbol{x}|\boldsymbol{y})v(\boldsymbol{y})d\boldsymbol{y},\tag{67}$$

where $p_{n+1|n}(\boldsymbol{y}|\boldsymbol{x})=\rho_n(\boldsymbol{y}-\boldsymbol{F}(\boldsymbol{x},\boldsymbol{w}_n))$ is the conditional transition density of $\boldsymbol{X}_{n+1}$ given $\boldsymbol{X}_n$, and ρ_n is the joint PDF of the random vector $\boldsymbol{\xi}_n$. The conditional transition density $p_{n+1|n}(\boldsymbol{y}|\boldsymbol{x})$ is always non-negative, i.e.,

$$p_{n+1|n}(\boldsymbol{y}|\boldsymbol{x})\geq 0\qquad \forall\boldsymbol{y}\in\mathscr{R}(\boldsymbol{X}_{n+1}),\quad \forall\boldsymbol{x}\in\mathscr{R}(\boldsymbol{X}_{n}).\tag{68}$$

Moreover, the conditional density $p_{n+1|n}$ is defined on the set

$$\mathscr{B}_n=\{(\boldsymbol{x},\boldsymbol{y})\in\mathscr{R}(\boldsymbol{X}_n)\times\mathscr{R}(\boldsymbol{X}_{n+1}):\,(\boldsymbol{y}-\boldsymbol{F}(\boldsymbol{x},\boldsymbol{w}_n))\in\mathscr{R}(\boldsymbol{\xi}_n)\}.\tag{69}$$

It is also important to emphasize that $\boldsymbol{y}\in\mathscr{R}(\boldsymbol{X}_{n+1})$ and $\boldsymbol{x}\in\mathscr{R}(\boldsymbol{X}_n)$. Both $\mathscr{R}(\boldsymbol{X}_{n+1})$ and $\mathscr{R}(\boldsymbol{X}_n)$ depend on Ω_0 (domain of the input), the neural network weights, and the noise amplitude. Thanks to Lemma 0.1, we have that

$$\mathscr{B}_n\subseteq\Omega_n\times\Omega_{n+1}.\tag{70}$$

The Lebesgue measure of $\mathscr{B}_n$ can be calculated as follows.

Lemma 0.5. *The Lebesgue measure of the set $\mathscr{B}_n$ defined in* (69) *is equal to the product of the measure of $\lambda(\mathscr{R}(\boldsymbol{X}_n))$ and the measure of $\mathscr{R}(\boldsymbol{\xi}_n)$, i.e.,*

$$\lambda(\mathscr{B}_n)=\lambda(\mathscr{R}(\boldsymbol{X}_n))\lambda(\mathscr{R}(\boldsymbol{\xi}_n)).\tag{71}$$

Moreover, $\lambda(\mathscr{B}_n)$ is bounded by $\lambda(\mathscr{R}(\Omega_n))\lambda(\mathscr{R}(\boldsymbol{\xi}_n))$, which is independent of the neural network weights.

Proof. Let χ_n be the indicator function of the set $\mathscr{R}(\boldsymbol{\xi}_n)$, $\boldsymbol{y}\in\mathscr{R}(\boldsymbol{X}_{n+1})$ and $\boldsymbol{x}\in\mathscr{R}(\boldsymbol{X}_n)$. Then

$$\begin{aligned}\lambda(\mathscr{B}_n)=&\int_{\mathscr{R}(\boldsymbol{X}_{n+1})}\int_{\mathscr{R}(\boldsymbol{X}_{n})}\chi_n(\boldsymbol{y}-\boldsymbol{F}(\boldsymbol{x},\boldsymbol{w}_n))d\boldsymbol{x}d\boldsymbol{y}\\ =&\lambda(\mathscr{R}(\boldsymbol{\xi}_n))\int_{\mathscr{R}(\boldsymbol{X}_{n})}d\boldsymbol{x}\\ =&\lambda(\mathscr{R}(\boldsymbol{X}_n))\lambda(\mathscr{R}(\boldsymbol{\xi}_n)).\end{aligned}\tag{72}$$

By using Lemma 0.1 we conclude that $\lambda(\mathscr{B}_n)$ is bounded from above by $\lambda(\mathscr{R}(\Omega_{L-m}))\lambda(\mathscr{R}(\boldsymbol{\xi}_n))$, which is independent of the neural network weights.

□

Remark: The result (71) has a straightforward geometrical interpretation in two dimensions. Pick a ruler of length $r = \lambda(\mathscr{R}(\xi_n))$ with endpoints that can leave markings if we slide it on a rectangular table with side lengths $s_b = \lambda(\mathscr{R}(X_{n+1}))$ (horizontal side) $s_h = \lambda(\mathscr{R}(X_n))$ (vertical side). Then slide the ruler from the top to the bottom of the table, while keeping it horizontal, i.e., parallel to the horizontal sides of the table (see Figure 3). The area of the domain defined by the two curves drawn by the endpoints of the ruler is always $r \times s_h$ independently of the way we slide the ruler laterally – provided the ruler never gets out of the table.

Lemma 0.6. *If the range of $\boldsymbol{\xi}_{n-1}$ is a bounded subset of $\mathbb{R}^N$ then the transition density $p_{n+1|n}(\boldsymbol{y}|\boldsymbol{x})$ is an element of $L^1\left(\mathscr{R}(\boldsymbol{X}_{n+1}) \times \mathscr{R}(\boldsymbol{X}_n)\right)$.*

Proof. Note that

$$\int_{\mathscr{R}(\boldsymbol{X}_{n+1})}\int_{\mathscr{R}(\boldsymbol{X}_n)} p_{n+1|n}(\boldsymbol{y}|\boldsymbol{x})d\boldsymbol{y}d\boldsymbol{x} = \lambda\left(\mathscr{R}(\boldsymbol{X}_n)\right) \leq \lambda(\Omega_n). \tag{73}$$

The Lebesgue measure $\lambda(\Omega_n)$ can be bounded as (see Proposition 0.2)

$$\lambda(\Omega_n) \leq \left(\sqrt{N} + \|\boldsymbol{\xi}_{n-1}\|_\infty\right)^N \frac{\pi^{N/2}}{\Gamma(1+N/2)}. \tag{74}$$

Since the range of $\boldsymbol{\xi}_{n-1}$ is bounded by hypothesis we have that there exists a finite real number $M > 0$ such that $\|\boldsymbol{\xi}_{n-1}\|_\infty \leq M$. This implies that the integral in (73) is finite, i.e., that the transition kernel $p_{n+1|n}(\boldsymbol{y}|\boldsymbol{x})$ is in $L^1\left(\mathscr{R}(\boldsymbol{X}_{n+1}) \times \mathscr{R}(\boldsymbol{X}_n)\right)$.

□

Theorem 0.7. *Let $C_{\boldsymbol{\xi}_n}(\boldsymbol{x})$ be the cumulative distribution function $\boldsymbol{\xi}_n$. If $C_{\boldsymbol{\xi}_n}(\boldsymbol{x})$ is Lipschitz continuous on $\mathscr{R}(\boldsymbol{\xi}_n)$ and the partial derivatives $\partial C_{\boldsymbol{\xi}_n}/\partial x_k$ ($k = 1, \ldots, N$) are Lipschitz continuous in x_1, x_2, ..., x_N, respectively, then the joint probability density function of $\boldsymbol{\xi}_n$ is bounded on $\mathscr{R}(\boldsymbol{\xi}_n)$.*

Proof. By using Rademacher's theorem we have that if $C_{\boldsymbol{\xi}_n}(\boldsymbol{x})$ is Lipschitz on $\mathscr{R}(\boldsymbol{\xi}_n)$ then it is differentiable almost everywhere on $\mathscr{R}(\boldsymbol{\xi}_n)$ (except on a set with zero Lebesgue measure). Therefore the partial derivatives $\partial C_{\boldsymbol{\xi}_n}/\partial x_k$ exist almost everywhere on $\mathscr{R}(\boldsymbol{\xi}_n)$. If, in addition, we assume that $\partial C_{\boldsymbol{\xi}_n}/\partial x_k$ are Lipschitz continuous with respect to x_k (for all $k = 1, \ldots, N$) then by applying [15, Theorem 9] recursively we conclude that the joint probability density function of $\boldsymbol{\xi}_n$ is bounded.

□

Lemma 0.8. *Under the same assumptions of Theorem 0.7 we have that the conditional PDF $p_{n+1|n}(\boldsymbol{y}|\boldsymbol{x}) = \rho_n(\boldsymbol{y} - \boldsymbol{F}(\boldsymbol{x}, \boldsymbol{w}))$ is bounded on $\mathscr{R}(\boldsymbol{X}_{n+1}) \times \mathscr{R}(\boldsymbol{X}_n)$.*

Proof. Theorem 0.7 states that ρ_n is a bounded function. This implies that the conditional density $p_{n+1|n}(\boldsymbol{y}|\boldsymbol{x}) = \rho_n(\boldsymbol{y} - \boldsymbol{F}(\boldsymbol{x}, \boldsymbol{w}))$ is bounded on $\mathscr{R}(\boldsymbol{X}_{n+1}) \times \mathscr{R}(\boldsymbol{X}_n)$.

□

Proposition 0.9. *Let $\mathscr{R}(\boldsymbol{\xi}_n)$ and $\mathscr{R}(\boldsymbol{\xi}_{n-1})$ be bounded subsets of $\mathbb{R}^N$. Then, under the same assumptions of Theorem 0.7, we have that the composition and the transfer operators defined in* (67) *are bounded in L^2.*

Proof. Let us first prove that $\mathcal{M}(n, n+1)$ is a bounded linear operator from $L^2(\mathscr{R}(\boldsymbol{X}_{n+1}))$ into $L^2(\mathscr{R}(\boldsymbol{X}_n))$. To this end, note that

$$\begin{aligned}
\|\mathcal{M}(n,n+1)v\|^2_{L^2(\mathscr{R}(\boldsymbol{X}_n))} =& \int_{\mathscr{R}(\boldsymbol{X}_n)} \left| \int_{\mathscr{R}(\boldsymbol{X}_{n+1})} v(\boldsymbol{y}) p_{n+1|n}(\boldsymbol{y}|\boldsymbol{x}) d\boldsymbol{y} \right|^2 d\boldsymbol{x} \\
\leq & \|v\|^2_{L^2(\mathscr{R}(\boldsymbol{X}_{n+1}))} \underbrace{\int_{\mathscr{R}(\boldsymbol{X}_n)} \int_{\mathscr{R}(\boldsymbol{X}_{n+1})} p_{n+1|n}(\boldsymbol{y}|\boldsymbol{x})^2 d\boldsymbol{y} d\boldsymbol{x}}_{K_n} \\
=& K_n \|v\|^2_{L^2(\mathscr{R}(\boldsymbol{X}_{n+1}))}.
\end{aligned} \tag{75}$$

Clearly, $K_n < \infty$. In fact, if $\mathscr{R}(\boldsymbol{\xi}_n)$ and $\mathscr{R}(\boldsymbol{\xi}_{n-1})$ are bounded then $\mathscr{R}(\boldsymbol{X}_{n+1})$ and $\mathscr{R}(\boldsymbol{X}_n)$ are bounded. Moreover, thanks to Lemma 0.8 we have that $p_{n+1|n}(\boldsymbol{y}|\boldsymbol{x})$ is bounded on $\mathscr{R}(\boldsymbol{X}_{n+1}) \times \mathscr{R}(\boldsymbol{X}_n)$. Hence, K_n is the integral of the square of a bounded function defined on a bounded domain, and therefore it is finite. By following the same steps it is straightforward to show that the transfer operator $\mathcal{N}$ is a bounded linear operator. Alternatively, simply recall that $\mathcal{N}$ is the adjoint of $\mathcal{M}$, and the adjoint of a bounded linear operator is bounded. Specifically we have,

$$\|\mathcal{N}(n+1,n)p\|^2_{L^2(\mathscr{R}(\boldsymbol{X}_{n+1}))} \leq K_n \|p\|^2_{L^2(\mathscr{R}(\boldsymbol{X}_n))}. \tag{76}$$

□

Remark: The integrals

$$K_n = \int_{\mathscr{R}(\boldsymbol{X}_n)} \int_{\mathscr{R}(\boldsymbol{X}_{n+1})} p_{n+1|n}(\boldsymbol{y}|\boldsymbol{x})^2 d\boldsymbol{y} d\boldsymbol{x} \tag{77}$$

can be computed by noting that

$$p_{n+1|n}(\boldsymbol{y}|\boldsymbol{x}) = \rho_n(\boldsymbol{y} - \boldsymbol{F}(\boldsymbol{x}, \boldsymbol{w})) \tag{78}$$

is essentially a *shift* of the PDF ρ_n by a quantity $\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{w})$ that depends on $\boldsymbol{x}$ and $\boldsymbol{w}$ (see, e.g., Figure 3). Such a shift does not influence the integral with respect to $\boldsymbol{y}$, meaning that the integral of $p_{n+1|n}(\boldsymbol{y}|\boldsymbol{x})$ or $p_{n+1|n}(\boldsymbol{y}|\boldsymbol{x})^2$ with respect to $\boldsymbol{y}$ is the same for all $\boldsymbol{x}$. Hence, by changing variables we have that the integral (77) is equivalent to

$$K_n = \lambda(\mathscr{R}(\boldsymbol{X}_n)) \int_{\mathscr{R}(\boldsymbol{\xi}_n)} \rho_n(\boldsymbol{x})^2 d\boldsymbol{x}, \tag{79}$$

where $\lambda(\mathscr{R}(\boldsymbol{X}_n))$ is the Lebesgue measure of $\mathscr{R}(\boldsymbol{X}_n)$, and $\mathscr{R}(\boldsymbol{\xi}_n)$ is the range of $\boldsymbol{\xi}_n$. Note that K_n depends on the neural net weights only through the Lebesgue measure of $\mathscr{R}(\boldsymbol{X}_n)$. Clearly, since the set Ω_n includes $\mathscr{R}(\boldsymbol{X}_n)$ we have by Lemma 0.1 that $\lambda(\mathscr{R}(\boldsymbol{X}_n)) \leq \lambda(\Omega_n)$. This implies that

$$K_n \leq \lambda(\Omega_n) \int_{\mathscr{R}(\boldsymbol{\xi}_n)} \rho_n(\boldsymbol{x})^2 d\boldsymbol{x}. \tag{80}$$

The upper bound here does not depend on the neural network weights. The following lemma summarizes all these remarks.

Proposition 0.10. *Under the same assumptions of Theorem 0.7, we have that the composition and the transfer operators defined in* (67) *can be bounded as*

$$\|\mathcal{M}(n,n+1)\|^2 \leq K_n, \qquad \|\mathcal{N}(n+1,n)\|^2 \leq K_n, \tag{81}$$

where

$$K_n = \lambda(\mathscr{R}(\boldsymbol{X}_n)) \int_{\mathscr{R}(\boldsymbol{\xi}_n)} \rho_n(\boldsymbol{x})^2 d\boldsymbol{x}. \tag{82}$$

Moreover, K_n can be bounded as

$$K_n \leq \lambda(\Omega_n) \int_{\mathscr{R}(\boldsymbol{\xi}_n)} \rho_n(\boldsymbol{x})^2 d\boldsymbol{x}, \tag{83}$$

where Ω_n is defined in (58) and ρ_n is the PDF of $\boldsymbol{\xi}_n$. The upper bound in (83) does not depend on the neural network weights and biases.

Under additional assumptions on the PDF $\rho_n(\boldsymbol{x})$ it is also possible to bound the integrals at the right hand side of (82) and (83). Specifically we have the following sharp bound.

Lemma 0.11. *Let*

$$s_n = \inf_{\boldsymbol{x}\in\mathscr{R}(\boldsymbol{\xi}_n)} \rho_n(\boldsymbol{x}), \qquad S_n = \sup_{\boldsymbol{x}\in\mathscr{R}(\boldsymbol{\xi}_n)} \rho_n(\boldsymbol{x}). \tag{84}$$

If $s_n > 0$ then under the same assumptions of Theorem 0.7 we have that

$$\int_{\mathscr{R}(\boldsymbol{\xi}_n)} \rho_n(\boldsymbol{x})^2 d\boldsymbol{x} \leq \frac{1}{\lambda(\mathscr{R}(\boldsymbol{\xi}_n))} \frac{(S_n + s_n)^2}{4S_n s_n}. \tag{85}$$

Proof. First we notice that if the random vector $\boldsymbol{\xi}_n$ satisfies the assumptions of Theorem 0.7 then the upper bound S_n is finite. By using the definition (84) we have

$$(\rho_n(\boldsymbol{x}) - s_n)(S_n - \rho_n(\boldsymbol{x})) \geq 0 \quad \text{for all} \quad \boldsymbol{x} \in \mathscr{R}(\boldsymbol{\xi}_n). \tag{86}$$

This implies

$$\int_{\mathscr{R}(\boldsymbol{\xi}_n)} \rho_n(\boldsymbol{x})^2 d\boldsymbol{x} \leq (S_n + s_n) - S_n s_n \lambda(\mathscr{R}(\boldsymbol{\xi}_n)), \tag{87}$$

where we used the fact that the PDF ρ_n integrates to one over $\mathscr{R}(\boldsymbol{\xi}_n)$. Next, define

$$R_n = \frac{1}{\lambda(\mathscr{R}(\boldsymbol{\xi}_n))} \frac{(S_n + s_n)^2}{4S_n s_n}. \tag{88}$$

Clearly,

$$R_n \left(1 - \frac{2S_n s_n}{s_n + S_n} \lambda(\mathscr{R}(\boldsymbol{\xi}_n))\right)^2 = R_n - (S_n + s_n) + S_n s_n \lambda(\mathscr{R}(\boldsymbol{\xi}_n)) \geq 0 \tag{89}$$

which implies that

$$(S_n + s_n) - S_n s_n \lambda(\mathscr{R}(\boldsymbol{\xi}_n)) \leq R_n. \tag{90}$$

A substitution of (90) into (87) yields (85).

□

An example. Let $X_0 \in \Omega_0 = [-1, 1]$ and consider

$$X_1 = \tanh(X_0 + 3) + \xi_0, \qquad X_2 = \tanh(2X_1 - 1) + \xi_1, \tag{91}$$

where ξ_0 and ξ_1 are uniform random variables with range $\mathscr{R}(\xi_0) = \mathscr{R}(\xi_1) = [-2, 2]$. In this setting,

$$\begin{aligned} \mathscr{R}(X_1) =& [\tanh(2) - 2, \tanh(4) + 2], \\ \mathscr{R}(X_2) =& [\tanh(2\tanh(2) - 5) - 2, \tanh(2\tanh(4) + 3) + 2]. \end{aligned}$$

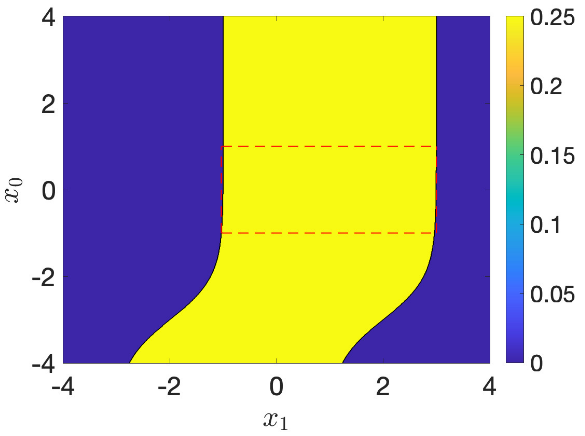


Figure 3: Conditional probability density function $p_{1|0}(x_1|x_0)$ defined in equation (92). The domain $\mathscr{R}(X_1) \times \mathscr{R}(X_0)$ is the interior of the rectangle delimited by dashed red lines.

The conditional density of X_1 given X_0 is given by

$$p_{1|0}(x_1|x_0) = \begin{cases} \dfrac{1}{4} & \text{if } |x_1 - \tanh(x_0+3)| \leq 2 \\ 0 & \text{otherwise} \end{cases} \tag{92}$$

This function is plotted in Figure 3 together with the domain $\mathscr{R}(X_1) \times \mathscr{R}(X_0)$ (interior of the rectangle delimited by dashed red lines). Clearly, the integral of the conditional PDF (92) is

$$\int_{\mathscr{R}(X_0)} \int_{\mathscr{R}(X_1)} p_{1|0}(x_1|x_0) dx_1 dx_0 = \lambda(\mathscr{R}(X_0)) = 2, \tag{93}$$

where $\lambda(\mathscr{R}(X_0))$ is the Lebesgue measure of $\mathscr{R}(X_0) = [-1,1]$. The L^2 norm of the operators $\mathcal{N}$ and $\mathcal{M}$ is bounded by[10]

$$K_0 = \int_{\mathscr{R}(X_0)} \int_{\mathscr{R}(X_1)} p_{1|0}(x_1|x_0)^2 dx_1 dx_0 = \frac{\lambda(\mathscr{R}(X_0))}{\lambda(\mathscr{R}(\xi_0))} = \frac{1}{2}. \tag{96}$$

Hence, both operators $\mathcal{N}(1,0)$ and $\mathcal{M}(0,1)$ are contractions (Proposition 0.10). On the other hand,

$$K_1 = \frac{\lambda(\mathscr{R}(X_1))}{\lambda(\mathscr{R}(\xi_1))} = 1 + \frac{\tan(4) - \tan(2)}{4} > 1. \tag{97}$$

Next, define V as in Lemma 0.4, i.e., $V = [-3,3]$. Clearly, both $\mathscr{R}(X_0)$ and $\mathscr{R}(X_1)$ are subsets of V. If we integrate the conditional PDF shown in Figure 3 in $V \times V$ we obtain

$$\int_V \int_V p_{1|0}(x_1|x_0)^2 dx_1 dx_0 = \frac{\lambda(V)}{\lambda(\mathscr{R}(\xi_0))} = \frac{3}{2}. \tag{98}$$

[10] For uniformly distributed random variables we have that

$$\int_{\mathscr{R}(\boldsymbol{\xi}_n)} \rho_n(\boldsymbol{x})^2 d\boldsymbol{x} = \frac{1}{\lambda(\mathscr{R}(\boldsymbol{\xi}_n))}. \tag{94}$$

Therefore equation (82) yields

$$K_n = \frac{\lambda(\mathscr{R}(\boldsymbol{X}_n))}{\lambda(\mathscr{R}(\boldsymbol{\xi}_n))} \leq \frac{\lambda(\Omega_n)}{\lambda(\mathscr{R}(\boldsymbol{\xi}_n))}. \tag{95}$$

Depending on the ratio between the Lebesgue measure of $\mathscr{R}(\boldsymbol{X}_n)$ and $\mathscr{R}(\boldsymbol{\xi}_n)$ one can have K_n smaller or larger than 1.

Random noise can induce contractions

In this section we prove a result on neural networks perturbed by random noise of increasing amplitude which states that it is possible to make both operators $\mathcal{N}$ and $\mathcal{M}$ in (67) contractions[11] if the noise is properly chosen. To this end, we begin with the following lemma.

Lemma 0.12. *Let*

$$\|\rho_n\|^2_{L^2(\mathscr{R}(\boldsymbol{\xi}_n))} = \int_{\mathscr{R}(\boldsymbol{\xi}_n)} \rho_n(\boldsymbol{x})^2 d\boldsymbol{x}. \tag{99}$$

If

$$\|\rho_n\|^2_{L^2(\mathscr{R}(\boldsymbol{\xi}_n))} \leq \frac{\kappa}{\lambda(\Omega_n)} \qquad 0 \leq \kappa < 1 \tag{100}$$

then $\mathcal{M}(n, n+1)$ and $\mathcal{N}(n+1, n)$ are contractions. The condition (100) *is independent of the neural network weights.*

Proof. The proof follows from equation (83).

□

Hereafter we specialize Lemma 0.12 to neural network perturbed by uniformly distributed random noise.

Proposition 0.13. *Let $\{\boldsymbol{\xi}_0, \ldots, \boldsymbol{\xi}_{L-1}\}$ be independent random vectors. Suppose that the components of each $\boldsymbol{\xi}_n$ are zero-mean i.i.d. uniform random variables with range $[-b_n, b_n]$ $(b_n > 0)$. If*

$$b_0 \geq \frac{1}{2}\left(\frac{\lambda(\Omega_0)}{\kappa}\right)^{1/N} \qquad \textit{and} \qquad b_n \geq \frac{b_{n-1}+1}{\kappa^{1/N}} \qquad n = 1, \ldots, L-1, \tag{101}$$

where Ω_0 is the domain of the neural network input, $0 \leq \kappa < 1$, and N is the number of neurons in each layer, then both operators $\mathcal{M}(n, n+1)$ and $\mathcal{N}(n+1, n)$ defined in (67) *are contractions for all $n = 0, \ldots, L-1$, i.e., their norm can be bounded by a constant $K_n \leq \kappa$, independently of the weights of the neural network.*

Proof. If $\boldsymbol{\xi}_n$ is uniformly distributed then from (82) we have that

$$K_n = \frac{\lambda\left(\mathscr{R}(\boldsymbol{X}_n)\right)}{\lambda\left(\mathscr{R}(\boldsymbol{\xi}_n)\right)}. \tag{102}$$

By using Lemma 0.4 we can bound K_n as

$$K_n \leq \left(\frac{1+b_{n-1}}{b_n}\right)^N, \tag{103}$$

where N is the number of neurons in each layer of the neural network. Therefore, if $b_n \geq (b_{n-1}+1)/\kappa^{1/N}$ $(n = 1, \ldots, L-1)$ we have that K_n is bounded by a quantity κ smaller than one. Regarding b_0, we notice that

$$K_0 = \frac{\lambda\left(\mathscr{R}(\boldsymbol{X}_0)\right)}{\lambda\left(\mathscr{R}(\boldsymbol{\xi}_0)\right)} = \frac{\lambda(\Omega_0)}{(2b_0)^N}, \tag{104}$$

where Ω_0 is the domain of the neural network input. Hence, if b_0 satisfies (102) then $K_0 \leq \kappa$.

□

[11] An linear operator is called a contraction if its operator norm is smaller than one.

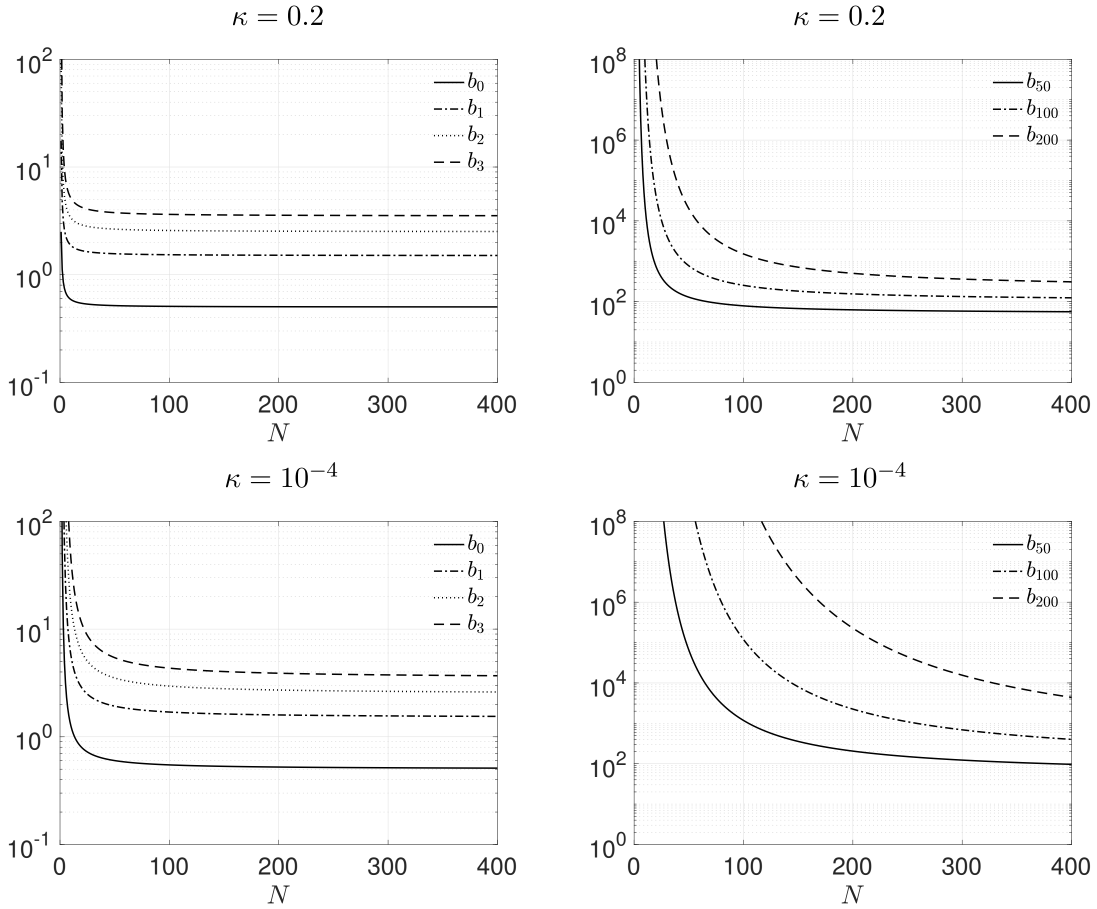


Figure 4: Lower bound on the coefficients b_n defined in (106) for $\lambda(\Omega_0) = 1$ as a function of the number of neurons N and number of layers of the neural network. With such values of b_n the operator $\mathcal{G}(L-n+1, L-n)$ is a contraction satisfying $\|\mathcal{G}(L-n+1, L-n)\|^2 \leq \kappa$. Shown are results for $\kappa = 0.2$ and $\kappa = 10^{-4}$ (contraction index).

One consequence of Proposition 0.13 is that the L^2 norm of the neural network output decays with both the number of layers and the number of neurons if the noise amplitude from one layer to the next increases as in (101). For example, if we represent the input-output map as a sequence of conditional expectations (see (20)), and set $u(\boldsymbol{x}) = \boldsymbol{\alpha}^T \boldsymbol{x}$ (linear output) then we have

$$q_0(\boldsymbol{x}) = \mathcal{M}(0,1)\mathcal{M}(1,2)\cdots\mathcal{M}(L-1,L)(\boldsymbol{\alpha}^T\boldsymbol{x}). \tag{105}$$

By iterating the inequalities (101) in Proposition 0.13 we find that

$$b_n \geq \frac{1}{2\kappa^{n/N}}\left(\frac{\lambda(\Omega_0)}{\kappa}\right)^{1/N} + \sum_{k=1}^{n}\frac{1}{\kappa^{k/N}} \qquad n = 0,\ldots,L-1, \tag{106}$$

In Figure 4 we plot the lower bound at the right hand side of (106) for $\kappa = 0.2$ and $\kappa = 10^{-4}$ as a function of the number of neurons (N). With b_n given in (106) we have that the operator norms of $\mathcal{M}(n, n+1)$ and $\mathcal{N}(n+1, n)$ ($n = 0, \ldots, L-1$) are bounded exactly by κ (see Lemma 0.12). Hence, by taking the L^2 norm of (105), and recalling that $\|\mathcal{M}(n, n+1)\|^2 \leq \kappa$ we obtain

$$\|q_0\|^2_{L^2(\Omega_0)} \leq Z^2\, \|\boldsymbol{\alpha}\|_2^2\, \kappa^L, \tag{107}$$

where[12]

$$Z^2 = \sum_{k=1}^{N} \int_{\mathscr{R}(\boldsymbol{X}_L)} x_k^2 d\boldsymbol{x} \qquad \text{and} \qquad \|\boldsymbol{\alpha}\|_2^2 = \sum_{k=1}^{N} \alpha_k^2. \tag{109}$$

The inequality (107) shows that the 2-norm of the vector of weights $\boldsymbol{\alpha}$ must increase exponentially fast with the number of layers L if we chose the noise amplitude as in (106). As shown in the following Lemma, the growth rate of b_n that guarantees that both $\mathcal{M}$ and $\mathcal{N}$ are contractions is linear (asymptotically with the number of neurons).

Lemma 0.14. *Consider a neural network satisfying the hypotheses of Proposition 0.13. Then, in the limit of an infinite number of neurons ($N \to \infty$), the noise amplitude* (106) *satisfies*

$$\lim_{N\to\infty} b_n = \frac{1}{2} + n, \tag{110}$$

independently of the contraction factor κ and the domain Ω_0. This means that for a finite number of neurons the noise amplitude b_n that guarantees that $\|\mathcal{M}(n, n+1)\| \leq \kappa$ is bounded from below ($\kappa < 1$) or from above ($\kappa > 1$) by a function that increases linearly with the number of layers.

Proof. The proof follows by taking the limit of (106) for $N \to \infty$. □

An example: Set $k = 10^{-4}$, $L = 4$ (four layers) and $N = 10$ neurons per layer. The factor κ^L in (107) is 10^{-16}. If we are interested in representing a d-dimensional function in the unit cube $\Omega_0 = [0, 1]^d$ then we have[13]

$$\lambda(\Omega_0) = 1 \qquad Z^2 \leq N \frac{2^N (1 + b_3)^{N+2}}{3}, \tag{112}$$

where $b_3 = 3.684$ (see Figure 4 for $N = 10$). Hence, the norm of the output (107) is bounded by

$$\|q_L\|_{L^2(\Omega_0)} \leq C \|\boldsymbol{\alpha}\|_2 \qquad C = 10^{-16} \sqrt{10 \frac{2^{10}(1 + b_3)^{12}}{3}} \simeq 6.17 \times 10^{-11}. \tag{113}$$

This means that the 2-norm of the output coefficients $\boldsymbol{\alpha}$ has to be of the order of 10^{11} to represent, e.g., a two-dimensional function of norm about one on the square $\Omega_0 = [0, 1]^2$.

[12] In equation (107) we used the Cauchy-Schwarz inequality

$$\left\| \boldsymbol{\alpha}^T \boldsymbol{x} \right\|_{L^2(\mathscr{R}(\boldsymbol{X}_L))}^2 \leq Z^2 \|\boldsymbol{\alpha}\|_2^2. \tag{108}$$

[13] In fact,

$$\sum_{i=1}^{N} \int_{\mathscr{R}(\boldsymbol{X}_4)} x_i^2 d\boldsymbol{x} \leq \sum_{i=1}^{N} \int_{\Omega_4} x_i^2 d\boldsymbol{x} = \sum_{i=1}^{N} \underbrace{\int_{-1-b_3}^{1+b_3} \cdots \int_{-1-b_3}^{1+b_3}}_{N \text{ times}} x_i^2 d\boldsymbol{x} = \frac{2^N N (1 + b_3)^{N+2}}{3}. \tag{111}$$

References

[1] Z. I. Botev, J. F. Grotowski, and D. P. Kroese. Kernel density estimation via diffusion. *Ann. Stat.*, 38(5):2916–2957, 2010.

[2] C. Brennan and D. Venturi. Data-driven closures for stochastic dynamical systems. *J. Comp. Phys.*, 372:281–298, 2018.

[3] J. M. Dominy and D. Venturi. Duality and conditional expectations in the Nakajima-Mori-Zwanzig formulation. *J. Math. Phys.*, 58(8):082701, 2017.

[4] W. E. A proposal on machine learning via dynamical systems. *Commun. Math. Stat.*, 5:1–10, 2017.

[5] W. E, J. Han, and Q. Li. A mean-field optimal control formulation of deep learning. *Res. Math. Sci.*, 6:1–41, 2019.

[6] L. Gonon, L. Grigoryeva, and J.-P. Ortega. Risk bounds for reservoir computing. *JMLR*, 21(240):1–61, 2020.

[7] N. J. Gordon, D. J. Salmond, and A. F. M. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings F. Radar and Signal Processing*, 140(2):107–113, 1993.

[8] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pages 770–778, 2016.

[9] K. He, X. Zhang, S. Ren, and J. Sun. Identity mappings in deep residual networks. In *ECCV*, pages 630–645. Springer, 2016.

[10] I. Kobyzev, S. J. D. Prince, and M. A. Brubaker. Normalizing flows: An introduction and review of current methods. *IEEE transactions on pattern analysis and machine intelligence*, 43(11):3964–3979, 2020.

[11] A. Lasota and M. C. Mackey. *Chaos, fractals and noise: stochastic aspects of dynamics.* Springer–Verlag, second edition, 1994.

[12] Q. Li, L. Chen, C. Tai, and W. E. Maximum principle based algorithms for deep learning. *JMLR*, 18:1–29, 2018.

[13] Q. Li, T. Lin, and Z. Shen. Deep learning via dynamical systems: An approximation perspective. *J. Eur. Math. Soc.*, (published online first), 2022.

[14] Y. Lu, Zhong A, Q. Li, and B. Dong. Beyond finite layer neural networks: Bridging deep architectures and numerical differential equations. *arXiv:1710.10121*, 2017.

[15] E. Minguzzi. The equality of mixed partial derivatives under weak differentiability conditions. *eal Anal. Exch.*, 40(1):81–98, 2014/2015.

[16] D. Nozaki, D. J. Mar, P. Grigg, and J. J. Collins. Effects of colored noise on stochastic resonance in sensory neurons. *Phys. Rev. Lett.*, 82(11):2402–2405, 1999.

[17] D. Rezende and S. Mohamed. Variational inference with normalizing flows. In *International conference on machine learning*, pages 1530–1538. PMLR, 2015.

[18] E. G. Tabak and E. Vanden-Eijnden. Density estimation by dual ascent of the log-likelihood. *Comm. Math. Sci.*, 8(1):217–233, 2010.

[19] N. Črnjarić-Žic, S. Maćešić, and I. Mezić. Koopman operator spectrum for random dynamical systems. *J. Nonlinear Sci.*, 30:2007–2056, 2020.

[20] D. Venturi and G. E. Karniadakis. Convolutionless Nakajima-Zwanzig equations for stochastic analysis in nonlinear dynamical systems. *Proc. R. Soc. A*, 470(2166):1–20, 2014.

[21] T. Yu, Y. Yang, D. Li, T. Hospedales, and T. Xiang. Simple and effective stochastic neural networks. In *Proc. Innov. Appl. Artif. Intell. Conf.*, volume 35, pages 3252–3260, 2021.

[22] Y. Zhu, J. M. Dominy, and D. Venturi. On the estimation of the Mori-Zwanzig memory integral. *J. Math. Phys*, 59(10):103501, 2018.

[23] Y. Zhu and D. Venturi. Hypoellipticity and the Mori-Zwanzig formulation of stochastic differential equations. *J. Math. Phys.*, 62:1035051, 2021.

Polynomial chaos

The theory of polynomial chaos dates back to Wiener [16, 1]. It was originally developed in a rather general/abstract setting, i.e., to represent L^2 functionals of the Browninan motion process $X(t;\omega)$ (see course note 1). What is a functional of the Brownian motion process? Think about the solution to an ODE driven by Brownian motion. The solution of the ODE at final time is a functional of the Brownian motion (forcing term). The main result is that any L^2 functional of the Brownian motion can be expanded in the so-called Wiener-Hermite series involving orthogonal[1] polynomial functionals of the Brownian motion. Long story short, if we denote by $C_0([0,1])$ the set of continuous functions on the interval $[0,1]$ vanishing at zero and

$$F : C_0([0,1]) \to \mathbb{R}$$

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

$$F([X(t;\omega)]) = \lim_{N\to\infty} \sum_{n=0}^{N} G_n([X(t;\omega)]), \tag{1}$$

where G_0 is a constant, and $G_n([X(t;\omega)])$ are Wiener-Hermite polynomials functionals. The first two of such functionals are[2] [17, p. 32]

$$G_1([X]) = \int_0^1 \kappa_1(t_1)dX(t_1;\omega), \tag{3}$$

$$G_2([X]) = \int_0^1\int_0^1 \kappa_2(t_1,t_2)dX(t_1;\omega)dX(t_2;\omega) - \int_0^1 \kappa_2(t_1,t_1)dt_1, \tag{4}$$

The kernel functions κ_1, κ_2, etc., satisfy a certain number of conditions that follow from the orthogonality requirements

$$\mathbb{E}\{G_0, G_1\} = 0, \qquad \mathbb{E}\{G_0, G_2\} = \mathbb{E}\{G_1, G_2\} = 0 \tag{5}$$

and the normalization conditions (see [17, Lecture 3])

$$\mathbb{E}\left\{G_0^2\right\} = \mathbb{E}\left\{G_1^2\right\} = \mathbb{E}\left\{G_2^2\right\} = 1. \tag{6}$$

Roughly speaking, the series expansion (1) says that it is possible to identify a nonlinear system by simply recording its response to Gaussian white noise [9, 11]. Another usage of (1) is to represent the solution to a problem, e.g. an ODE or PDE, driven by Gaussian white noise. The solution to such problems is a functional of the Browninan motion, and therefore it admits a Wiener-Hermite expansion.

There were attempts to *generalize* the Wiener-Hermite functional expansion to processes other than the Brownian motion (see, e.g., [10, 7, 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 [10, 7] 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 one random variable.

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

[2] Note that the integrals in (3)-(4) do not exist in the ordinary Stieltjes sense because $X(t;\omega)$ is nowhere differentiable. However, we can get around this by defining the integrals such as (3) using integration by parts as

$$G_1([X]) = \int_0^1 \kappa_1(t_1)dX(t_1;\omega) = \kappa_1(1)X(1;\omega) - \kappa_1(0)X(0;\omega) - \int_0^1 \kappa_1'(t_1)X(t_1;\omega)dt_1 = \kappa_1(1)X(1;\omega) - \int_0^1 \kappa_1'(t_1)X(t_1;\omega)dt_1. \tag{2}$$

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 [18, 2]. The simplest case is a system driven by only one random variable, i.e., a mapping of the form

$$\eta(\omega) = g(\xi(\omega)). \tag{7}$$

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

$$g(\xi(\omega)) = \sum_{k=0}^{\infty} a_k P_k(\xi(\omega)), \tag{8}$$

where a_k are real numbers, and $P_k(\xi(\omega))$ are polynomials of the random variable $\xi(\omega)$ satisfying the orthogonality conditions

$$\mathbb{E}\left\{P_k(\xi)P_j(\xi)\right\} = \int_{-\infty}^{\infty} P_k(x)P_j(x)p_\xi(x)dx = \mathbb{E}\left\{P_k^2(\xi)\right\}\delta_{kj}. \tag{9}$$

A substitution of (9) into (8) yields

$$a_k = \frac{\mathbb{E}\left\{P_k(\xi)g(\xi)\right\}}{\mathbb{E}\left\{P_k^2(\xi)\right\}}. \tag{10}$$

Remark: The theory of orthogonal polynomials is summarized in [18, 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).

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.

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

$$g_n(\xi) = \sum_{k=0}^{n} a_k P_k(\xi) \qquad a_k = \frac{\mathbb{E}\left\{P_k(\xi)g(\xi)\right\}}{\mathbb{E}\left\{P_k^2(\xi)\right\}}, \tag{11}$$

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). In other words

$$\lim_{n\to\infty} \mathbb{E}\left\{\left|g(\xi) - g_n(\xi)\right|^2\right\} = 0. \tag{12}$$

We have shown in Appendix B 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?

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.

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.,

$$\mathbb{E}\{e^{a|\xi(\omega)|}\} < \infty \quad \text{for some } a > 0; \tag{13}$$

4. The sequence of moments $m_n = \mathbb{E}\{\xi^n\}$ satisfies

$$\sum_{n=0}^{\infty} \left(\frac{1}{m_{2n}}\right)^{\frac{1}{2n}} = \infty. \tag{14}$$

The proof of the theorem is provided in [2] and therefore omitted here.

Example: The moment problem is uniquely solvable for Gaussian PDFs, uniform PDFs, and gamma PDFs

$$p_\xi(x) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-x/\theta}, \qquad x > 0, \qquad k, \theta > 0. \tag{15}$$

Example: A log-normal random variable is defined as

$$\xi(\omega) = \log(X(\omega)), \tag{16}$$

where $X(\omega)$ is normal. It is straightforward to show that

$$p_\xi(x) = \frac{1}{x\sqrt{2\pi}} e^{-\log(x)^2/2} x > 0. \tag{17}$$

The moments of ξ are

$$\mathbb{E}\{\xi^n\} = e^{n^2/2}, \tag{18}$$

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

$$p_\eta(x) = \frac{1}{x\sqrt{2\pi}} e^{-\log(x)^2/2} \left[1 + v \sin(2k\pi \log(x)\right] \qquad x > 0 \tag{19}$$

has exactly the same moments as (17) (see [2, §4.1]). In other words,

$$\int_0^\infty x^n p_\xi(x)dx = \int_0^\infty x^n p_\eta(x)dx \quad \text{for all } n \geq 1. \tag{20}$$

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. (18)) and therefore

$$\sum_{n=0}^{\infty} \left(\frac{1}{e^{2n^2}}\right)^{\frac{1}{2n}} = \sum_{n=0}^{\infty} \frac{1}{e^n} = \frac{e}{e-1}. \tag{21}$$

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

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 $\{\xi(\omega), \ldots, \xi_M(\omega)\}$

$$\eta = g(\xi_1, \ldots, \xi_M) \tag{22}$$

We have seen in Chapter 1 that the PDF η can be represented as a multidimensional convolution of the PDFs of $\{\xi_j\}$. Denote by $\{P^{(i)}_{j_i}(\xi_i)\}$ 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

$$g(\xi_1, \ldots, \xi_M) = \sum_{j_1=0}^{\infty} \cdots \sum_{j_M=0}^{\infty} a_{j_1 \ldots j_M} P^{(1)}_{j_1}(\xi_1) \cdots P^{(M)}_{j_M}(\xi_M), \tag{23}$$

where

$$a_{j_1 \ldots j_M} = \frac{\mathbb{E}\{g(\xi_1, \ldots, \xi_M) P^{(1)}_{j_1}(\xi_1) \cdots P^{(M)}_{j_M}(\xi_M)\}}{\mathbb{E}\left\{P^{(1)}_{j_1}(\xi_1)^2\right\} \cdots \mathbb{E}\left\{P^{(M)}_{j_M}(\xi_M)^2\right\}} \tag{24}$$

If the moment problem for each random variable $\xi_i(\omega)$ is uniquely solvable, then the *tensor product* gPC expansion (23)-(24) converges in the mean square sense (see [2]), i.e., in the $L^2(\Omega, \mathcal{F}, P)$ sense

$$\lim_{n_1 \to \infty} \cdots \lim_{n_M \to \infty} \mathbb{E}\left\{ g(\xi_1, \ldots, \xi_M) - \sum_{j_1}^{n_1} \cdots \sum_{j_M}^{n_M} a_{j_1 \ldots j_M} P^{(1)}_{j_1}(\xi_1) \cdots P^{(M)}_{j_M}(\xi_M) \right\} = 0 \tag{25}$$

It convenient to write the expansion (23) more compactly. Upon definition of $\boldsymbol{\xi} = (\xi_1, \ldots, xi_M)$ we have

$$g(\boldsymbol{\xi}) = \sum_{k=0}^{\infty} a_k \Phi_k(\boldsymbol{\xi}), \tag{26}$$

where $\Phi_k(\boldsymbol{\xi})$ are multivariate polynomials constructed by taking products of one-dimensional polynomials $P^i_{j_i}(\xi_i)$. 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

$$\Phi_k(\boldsymbol{\xi}) = P^{(1)}_{j_1}(\xi_1) P^{(2)}_{j_2}(\xi_2) P^{(3)}_{j_3}(\xi_3), \tag{27}$$

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 $M = 3$

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).

random variables yields 10 basis elements $\{\Phi_0, \dots, \Phi_9\}$ (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 (26) to K terms, then we have

$$K + 1 = \binom{p+M}{p} = \frac{(M+p)!}{M!p!}. \tag{28}$$

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 $\{\Phi_0, \dots, \Phi_K\}$. 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

$$\eta = g(\boldsymbol{\xi}) \simeq \sum_{j=0}^{K} a_j \Phi_j(\boldsymbol{\xi}), \qquad a_j = \frac{\mathbb{E}\left\{\eta \Phi_j(\boldsymbol{\xi})\right\}}{\mathbb{E}\{\Phi_j^2\}}, \tag{29}$$

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

$$\mathbb{E}\left\{g(\boldsymbol{\xi})\right\} = a_0, \tag{30}$$

$$\mathbb{E}\left\{g(\boldsymbol{\xi})^2\right\} = \sum_{k=0}^{K} a_k^2 \mathbb{E}\{\Phi_k^2\}. \tag{31}$$

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[3]. 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 (30)-(31) allow us to express the variance of $g(\boldsymbol{\xi})$ (second cumulant) as

$$\mathrm{var}\{g(\boldsymbol{\xi})\} = \sum_{k=1}^{K} a_k^2 \mathbb{E}\{\Phi_k^2\}. \tag{33}$$

Regarding the PDF of η, we recall that the gPC expansion (29) 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 [15, 14] 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 $[0, 1]$,

$$\ddot{x} + \xi^2(\omega)x = 0, \qquad \dot{x}(0) = 1, \qquad x(0) = 0. \tag{34}$$

As is well known, the solution to (34) is

$$x(t, \xi) = \sin(\xi(\omega)t). \tag{35}$$

It is clear that the gPC representation of the solution (35) requires polynomials of increasing order 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

$$\dot{x} = -\xi(\omega)x, \qquad x(0) = 1, \qquad \xi \sim U([0, 1]), \tag{36}$$

i.e.,

$$x(t; \omega) = e^{-\xi t}. \tag{37}$$

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)

$$E_1 = \{x \in \mathbb{R} : a \le x \le c\}, \qquad E_2 = \{x \in \mathbb{R} : c \le x \le b\}. \tag{38}$$

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

$$\mathbb{E}\{P_0^{(j)}(\xi_j)P_q^{(j)}(\xi_j)\} = 0 \quad \text{for all} \quad q \neq 0 \quad \Rightarrow \quad P_0^{(j)}(\xi_j)\mathbb{E}\{P_q^{(j)}(\xi_j)\} = 0 \quad \Rightarrow \quad \mathbb{E}\{P_q^{(j)}(\xi_j)\} = 0 \quad \text{for all} \quad q \neq 0. \tag{32}$$

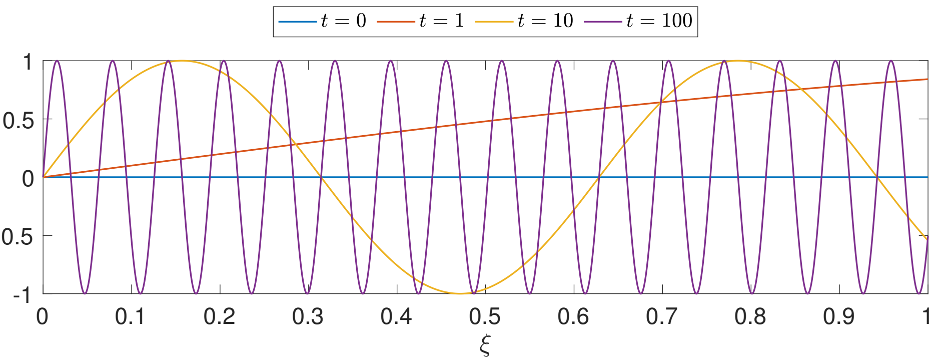


Figure 1: Random frequency problem. A gPC expansion of the time-dependent function $\sin(\xi t)$ requires polynomials of higher and higher degrees as t increases. To see this, simply note that $\sin(\xi t)$ has 4 zeros in ξ at $t=10$ and 32 zeros in ξ at $t=100$. Therefore, at $t=4$ we need a gPC expansion 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

$$I_{E_i} = \begin{cases} 1 & \text{if } \xi(\omega) \in E_i \\ 0 & \text{otherwise} \end{cases} \qquad i = 1, 2. \tag{39}$$

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

$$A_1 = \{\omega \in \Omega : \xi(\omega) \in E_1\} \quad \text{and} \quad A_2 = \{\omega \in \Omega : \xi(\omega) \in E_2\} \tag{40}$$

are non-intersecting subsets of Ω such that

$$\Omega = A_1 \cup A_2. \tag{41}$$

At this point, consider the input-output map

$$\eta(\omega) = g(\xi(\omega)). \tag{42}$$

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

$$F_\eta(y) = P(\underbrace{\{\omega \in \Omega : g(\xi(\omega)) \leq y\}}_{\text{set } B_y}). \tag{43}$$

The set B_y can be written as union of two non-intersecting[4] sets

$$\begin{aligned} B_y =& B_y \cap \Omega \\ =& B_y \cap (A_1 \cup A_2) \\ =& (B_y \cap A_1) \cup (B_y \cap A_2). \end{aligned} \tag{44}$$

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

$$P(B_y) = P(B_y \cap A_1) + P(B_y \cap A_2). \tag{45}$$

In terms of conditional probabilities this can be written as

$$P(B_y) = P(B_y|A_1)P(A_1) + P(B_y|A_2)P(A_2). \tag{46}$$

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

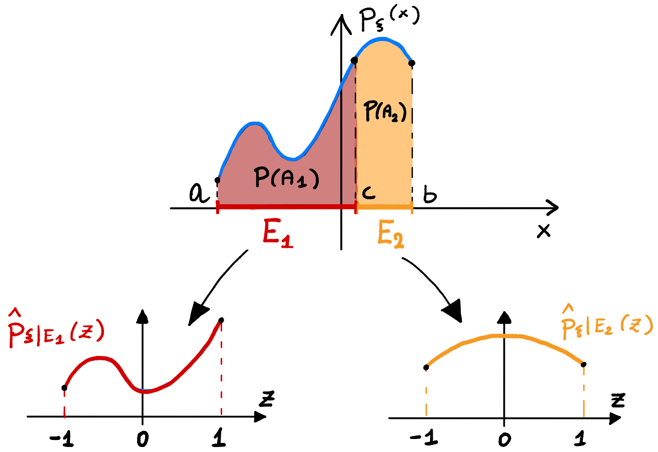


Figure 2: Basic idea of Multi-Element generalized Polynomial Chaos (ME-gPC). The range of the random input variable $\xi(\omega)$, i.e., the support of the PDF $p_\xi(x)$ is partitioned into non-overlapping elements, say E_1 and E_2. A local gPC expansion is then constructed relative to the conditional 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 allow standardization of the Stieltjes algorithm to construct the set of orthogonal polynomials corresponding to the PDFs $\hat{p}_{\xi|E_i}(z)$.

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)

$$P(A_1) = \int_{E_1} p_\xi(x)dx \qquad P(A_2) = \int_{E_2} p_\xi(x)dx. \tag{47}$$

By combining (43), (46) and (47) we finally obtain

$$F_\eta(y) = F_{\eta|\xi\in E_1}(y)\int_{E_1} p_\xi(x)dx + F_{\eta|\xi\in E_2}(y)\int_{E_2} p_\xi(x)dx. \tag{48}$$

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

$$p_\eta(y) = p_{\eta|\xi\in E_1}(y)\int_{E_1} p_\xi(x)dx + p_{\eta|\xi\in E_2}(y)\int_{E_2} p_\xi(x)dx. \tag{49}$$

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.,

$$p_{\eta|\xi\in E_1}(y) \quad \text{and} \quad p_{\eta|\xi\in E_2}(y). \tag{50}$$

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 (49) then we can compute any statistical properties of η, including the PDF of η

The procedure for ME-gPC is as follows:

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 (87).

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 [6]. 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

$$\frac{dx}{dt} = -\xi(\omega)^2 x, \qquad x(0;\omega) = 1, \tag{51}$$

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)

$$x(t,\omega) = \sum_{k=0}^{n} a_k(t) L_k(\xi(\omega)), \tag{52}$$

where $L_k(\xi)$ are Legendre polynomials[5] of the uniform random variable ξ. Note that here the polynomial chaos modes are function of time and defined through projection

$$a_k(t) = \frac{\mathbb{E}\{x(t;\omega)L_k(\xi)\}}{\mathbb{E}\{L_k^2(\xi)\}}. \tag{53}$$

A substitution of (52) into (51) yields

$$\sum_{k=0}^{n} \frac{da_k(t)}{dt} L_k(\xi) = -\xi(\omega)^2 \sum_{k=0}^{n} a_k(t) L_k(\xi) + R_n(t;\xi), \qquad \sum_{k=0}^{n} a_k(0) L_k(\xi) = x_0, \tag{54}$$

where $R_n(t,\xi)$ is the residual arising from the fact that (52) does not satisfy the ODE (51) exactly. In the stochastic Galerkin method we impose that the residual is orthogonal to the gPC space $B_n =$

[5] Legendre polynomials are defined are orthogonal with respect to the recursively in Eq. (83).

span$\{L_0,\ldots,L_n\}$ relative to the $L^2(\Omega,\mathcal{F},P)$ inner product[6]. In practice, we multiply (54) by $L_j(\xi)$ and then integrate relative to the PDF of $\xi(\omega)$, i.e., take the expectation, to obtain

$$\begin{cases}\displaystyle\sum_{k=0}^{n}\frac{da_k(t)}{dt}\mathbb{E}\{L_kL_j\}=-\sum_{k=0}^{n}a_k(t)\mathbb{E}\{\xi^2L_kL_j\}\\ \displaystyle\sum_{k=0}^{n}a_k(0)\mathbb{E}\{L_kL_j\}=x_0\end{cases}\tag{55}$$

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

$$L_0(\xi)=1\qquad L_1(\xi)=\xi\qquad L_2(\xi)=\frac{3}{2}\xi^2-\frac{1}{2},\tag{56}$$

i.e.,

$$\xi^2=\frac{2L_2(\xi)+1}{3}.\tag{57}$$

we can write (55) as

$$\begin{cases}\displaystyle\frac{da_j(t)}{dt}=-\frac{a_j(t)}{3}-\frac{2}{3\mathbb{E}\{L_j^2\}}\sum_{k=0}^{n}\mathbb{E}\{L_2L_kL_j\}a_k(t)\qquad j=0,\ldots,n\\ a_0(0)=1\\ a_j(0)=0\qquad j=1,\ldots,n\end{cases}\tag{58}$$

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 (52), and compute statistical properties such as the mean,

$$\mathbb{E}\{x(t;\omega)\}=a_0(t),\tag{59}$$

the variance

$$\mathrm{var}\{x(t;\omega)\}=\sum_{k=1}^{n}a_k^2(t)\mathbb{E}\{L_k^2\},\tag{60}$$

or the PDF of $x(t;\omega)$ by sampling or transforming the polynomial chaos expansion (52).

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 $[a,b]$. 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 $\mathbb{E}\{L_kL_j\}$.

[6] In numerical methods for deterministic PDEs this procedure is also known as Galerkin projection method [5, ?].

The coefficients

$$\mathbb{E}\{L_iL_jL_k\} = \int_{-1}^{1} L_i(x)L_j(x)L_k(x)dx \tag{61}$$

appearing in the the *gPC propagator* (55) can be pre-computed offline using Gauss quadrature, or can be computed analytically using the so-called *linearization formulas* [13, Appendix] for orthogonal polynomials. Such formulas basically express the product of two orthogonal polynomials in terms of polynomials that belong to same family as

$$L_k(\xi)L_j(\xi) = \sum_{m=0}^{k+j} \beta_m L_m(\xi). \tag{62}$$

Indeed a substitution of (62) (with β_m known) into (61) yields

$$\mathbb{E}\{L_iL_jL_k\} = \sum_{m=0}^{k+j} \beta_m \mathbb{E}\{L_mL_i\} = \beta_i\mathbb{E}\{L_i^2\}. \tag{63}$$

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

$$\begin{cases} \dfrac{\partial u}{\partial t} = \dfrac{\partial^2 u}{\partial x^2} \qquad x \in [0, L] \\ u(x, 0) = 0 \\ u(0, t) = u_0(t) \\ u(L, t) = A + \sigma\xi(\omega)\sin(t) \end{cases} \tag{64}$$

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 (64) relative to the (monic) gPC basis $\{P_0, P_1, \ldots\}$ as

$$u(x, t; \omega) = \sum_{k=0}^{n} a_k(x, t)P_k(\xi). \tag{65}$$

Substituting (65) into (64) 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

$$\begin{cases} \dfrac{\partial a_k(x,t)}{\partial t} = \dfrac{\partial^2 a_k}{\partial x^2} \qquad k = 0, \ldots, K \qquad x \in [0, L] \\ a_k(x, 0) = 0 \\ a_0(0, t) = u_0(t) \\ a_k(0, t) = 0 \qquad k = 1, \ldots, n \\ a_0(L, t) = A \\ a_1(L, t) = \sigma \sin(t)\mathbb{E}\{P_1^2\} \\ a_k(L, t) = 0 \qquad k = 2, \ldots, n \end{cases} \tag{66}$$

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.

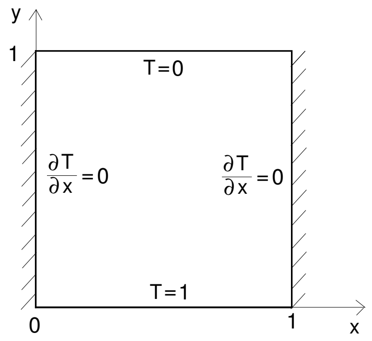


Figure 3: Schematic of the geometry and dimensionless temperature boundary conditions. The velocity boundary conditions are of no-slip type, i.e. $\boldsymbol{u} = 0$ at the walls.

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

$$\begin{cases} \dfrac{\partial u}{\partial t} + u\dfrac{\partial u}{\partial x} = \dfrac{\partial^2 u}{\partial x^2} \qquad x \in [0, 2\pi] \\ u(x,0) = u_0(x) + \sigma \displaystyle\sum_{j=1}^{M} \xi_j(\omega)\psi_j(x) \\ \text{Periodic B.C.} \end{cases} \tag{67}$$

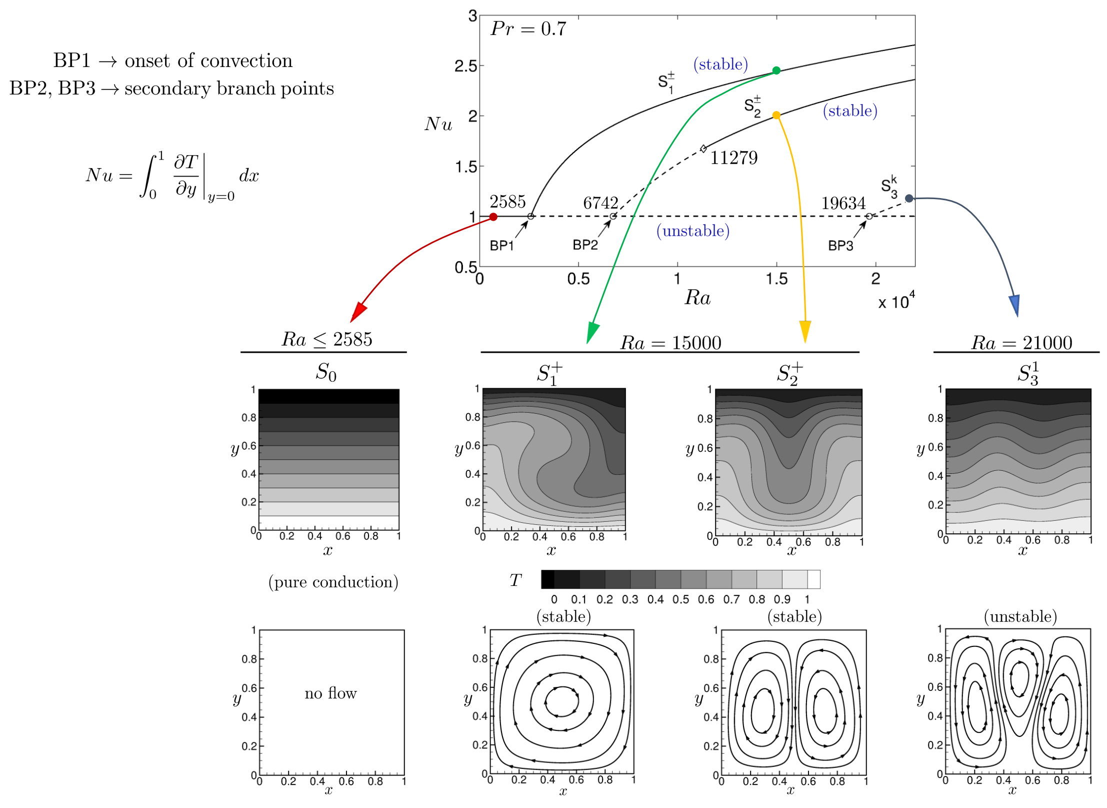
The random initial condition here assumed to be a correlated Gaussian random process represented in terms of a Karhunen-Loéve expansion with M independent Gaussian random variables $\boldsymbol{\xi} = \{x_1, \ldots, \xi_M\}$. To construct a gPC basis, we first build the gPC basis for one Gaussian random 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 (67) as

$$u(x,t;\omega) = \sum_{k=0}^{K} a_k(x,t)\Phi_k(\boldsymbol{\xi}). \tag{68}$$

A substitution of (68) into (67) and subsequent projection onto the gPC basis $\{\Phi_j\}$ yields

$$\begin{cases} \dfrac{\partial a_k}{\partial t} + \dfrac{1}{\mathbb{E}\{\Phi_k^2\}} \displaystyle\sum_{i,j} \mathbb{E}\{\Phi_k\Phi_i\Phi_j\} a_i(x,t) \dfrac{\partial a_j(x,t)}{\partial x} = \dfrac{\partial^2 a_k(x,t)}{\partial x^2} \qquad k = 0, \ldots K \qquad x \in [0, 2\pi] \\ a_0(x,0) = u_0(x) \\ a_1(x,0) = \sigma\psi_M(x) \\ a_2(x,0) = \sigma\psi_{M-1}(x) \\ \vdots \\ a_M(x,0) = \sigma\psi_1(x) \\ a_k(x,0) = 0 \qquad k = M+1, \ldots, K \\ \text{Periodic B.C. for each } a_k(x,t) \end{cases} \tag{69}$$

Note that if the KL expansion of the random initial condition in (67) 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 (69) is 462!



BP1 → onset of convection
BP2, BP3 → secondary branch points

$$Nu = \int_0^1 \frac{\partial T}{\partial y}\bigg|_{y=0} dx$$

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

Stochastic thermal convection. Consider the system of PDEs system

$$\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u}\cdot\nabla)\boldsymbol{u} = -\nabla p + Pr\nabla^2 u + Ra\, Pr\, T\boldsymbol{j} \tag{70}$$

$$\frac{\partial T}{\partial t} + \boldsymbol{u}\cdot\nabla T = \nabla^2 T \tag{71}$$

$$\nabla\cdot\boldsymbol{u} = 0 \tag{72}$$

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 (70)-(72) $\boldsymbol{u}(\boldsymbol{x},t)$ 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, 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 (70) is a uniform random variable centered at $Ra_c = 2585$ (onset of convective instability), i.e.,

$$Ra = Ra_c\,(1+\sigma\xi)\,, \quad \xi \sim U([-1,1]) \qquad \sigma = 0.05. \tag{73}$$

We are interested in computing the velocity, pressure and temperature fields corresponding to such random

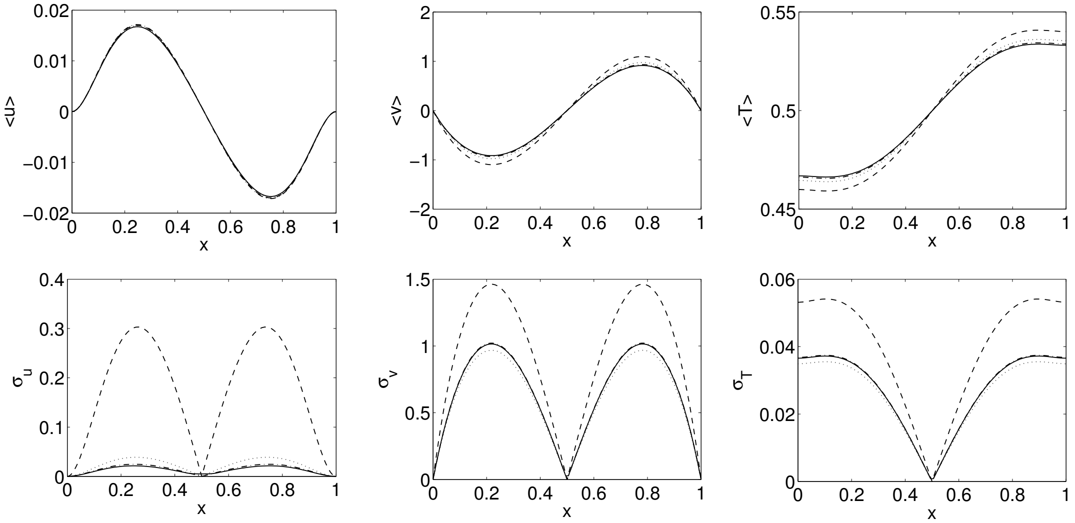


Figure 5: Stochastic convection near the onset. Shown are means (first row) and standard deviations (second row) of velocity and temperature fields along the crossline $y = 0.5$. We plot different results: MC benchmark (−), gPC order 3 (−−), ME-gPC 2 elements of order 3 (· · ·), ME-gPC 8 elements of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

Rayleigh number (see [12]). To this end, consider the gPC expansions

$$\boldsymbol{u}\left(\boldsymbol{x},t;\xi\right) = \sum_{i=0}^{n} \widehat{\boldsymbol{u}}_i\left(\boldsymbol{x},t\right)\Phi_i\left(\xi\right), \tag{74}$$

$$p\left(\boldsymbol{x},t;\xi\right) = \sum_{i=0}^{n} \widehat{p}_i\left(\boldsymbol{x},t\right)\Phi_i\left(\xi\right), \tag{75}$$

$$T\left(\boldsymbol{x},t;\xi\right) = \sum_{i=0}^{n} \widehat{T}_i\left(\boldsymbol{x},t\right)\Phi_i\left(\xi\right). \tag{76}$$

where, $\Phi_i\left(\xi\right)$ are Legendre polynomials of the uniform random variable ξ. A substitution of (74)-(76)) into the system (70)-(72) and subsequent projection onto the basis $\{\Phi_i\}$ yields the gPC propagator

$$\frac{\partial \widehat{\boldsymbol{u}}_k}{\partial t} + \sum_{i,j=0}^{n} \frac{\mathbb{E}\{\Phi_i\Phi_j\Phi_k\}}{\mathbb{E}\{\Phi_k^2\}}\left(\widehat{\boldsymbol{u}}_i\cdot\nabla\right)\widehat{\boldsymbol{u}}_j = -\nabla\widehat{p}_k + Pr\nabla^2\widehat{\boldsymbol{u}}_k + Ra_cPr\left(\widehat{T}_k + \sigma\sum_{i,j=0}^{n}\frac{\mathbb{E}\{\Phi_1\Phi_j\Phi_k\}}{\mathbb{E}\{\Phi_k^2\}}\widehat{T}_j\right)\widehat{\boldsymbol{j}}, \tag{77}$$

$$\frac{\partial \widehat{T}_k}{\partial t} + \sum_{i,j=0}^{n} \frac{\mathbb{E}\{\Phi_i\Phi_j\Phi_k\}}{\mathbb{E}\{\Phi_k^2\}}\widehat{\boldsymbol{u}}_i\cdot\nabla\widehat{T}_j = \nabla^2\widehat{T}_k, \tag{78}$$

$$\nabla\cdot\widehat{\boldsymbol{u}}_k = 0. \tag{79}$$

This is a system in $3(n+1)$ coupled PDEs of the form (70)-(72), 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 $y = 0.5$.

Appendix A: Orthogonal polynomials

A polynomial of degree n can be written as

$$Q_n(x) = b_n x^n + \cdots + b_1 x + b_0, \qquad b_n \neq 0. \tag{80}$$

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

$$\int_{\text{supp}(\mu)} Q_n(x)Q_m(x)\mu(x)dx = \delta_{nm}\gamma_n \qquad \text{where} \qquad \gamma_n = \int_{\text{supp}(\mu)} Q_n(x)^2\mu(x)dx, \tag{81}$$

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, 5])

$$\begin{cases} Q_{n+1}(x) = (A_n x + B_n)Q_n(x) - C_n Q_{n-1}(x) \\ Q_0(x) = 1 \\ Q_{-1}(x) = 0 \end{cases} \tag{82}$$

where $A_n \neq 0$, $C_n \neq 0$ and $C_n A_n A_{n-1} > 0$ for all n (Favard's theorem [18, 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

$$L_{n+1}(x) = \frac{2n+1}{n+1} x L_n(x) - \frac{n}{n-1} L_{n-1}(x). \tag{83}$$

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

$$\mu(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \tag{84}$$

and they satisfy the three-term recurrence relation

$$H_{n+1}(x) = xH_n(x) - nH_{n-1}(x) \tag{85}$$

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

$$\begin{cases} \pi_{n+1}(x) = (x - \alpha_n)\pi_n(x) - \beta_n \pi_{n-1}(x) \\ \pi_0(x) = 1 \\ \pi_{-1}(x) = 0 \end{cases} \tag{86}$$

The coefficients α_n and β_n are *uniquely determined* by the weight function. Let us show show how.

Stieltjes algorithm. The coefficients α_n and β_n in (86), which define monic orthogonal polynomials corresponding to a given measure, can be computed numerically using a simple algorithm known as *Stieltjes algorithm*. To this end, suppose that the weight function $\mu(x) \geq 0$ is continuous and supported[7] on $[-1, 1]$.

[7] 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

$$x = \frac{b-a}{2} z + \frac{b+a}{2} \qquad z \in [-1, 1]. \tag{87}$$

Define the inner product

$$(p,q)=\int_{-1}^{1}p(x)q(x)\mu(x)dx. \tag{88}$$

Multiplying (86) by $\pi_n(x)$ and imposing orthogonality yields

$$\alpha_n =\frac{(x\pi_n,\pi_n)}{(\pi_n,\pi_n)} \qquad n=0,1,2,\ldots \tag{89}$$

$$\beta_n =\frac{(\pi_n,\pi_n)}{(\pi_{n-1},\pi_{n-1})} \qquad n=1,2,3,\ldots \tag{90}$$

This allows us to derive the following algorithm (known as Stieltjes algorithm) to compute the recurrence coefficients α_k and β_k in (86):

1. Set $n=0$ and $\pi_0(x)=1$ in (89). Compute α_0.
2. With α_0 and $\pi_0(x)=1$ available compute

$$\pi_1(x)=(x-\alpha_0)\pi_0(x)-\beta_0\underbrace{\pi_{-1}(x)}_{=0}=(x-\alpha_0). \tag{91}$$

3. With $\pi_1(x)$ and $\pi_0(x)$ available compute β_1 form (90).
4. At this point we can compute α_1 from (89), $\pi_2(x)$ from (86), β_2 from (90), and so on so forth.

In practice, we can compute α_n and β_n to machine precision by replacing the inner product (88) with, e.g., a Gaussian quadrature rule [8]

$$(p,q)\simeq\sum_{j=0}^{M}w_jp(x_j)q(x_j)\mu(x_j), \tag{92}$$

w_j being the Gaussian quadrature weights.

Polynomial approximation theory. Denote by

$$\mathbb{P}_n([a,b])=\text{span}\{1,x,\ldots,x^n\} \tag{93}$$

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

$$\|f-p_{n_\epsilon}\|_{L^\infty([a,b])}=\sup_{x\in[a,b]}|f(x)-p_{n_\epsilon}(x)|\leq\epsilon. \tag{94}$$

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

$$(Q_i, Q_j) = \int_a^b Q_i(x)Q_j(x)\mu(x)dx, \tag{95}$$

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 $\{Q_0(x), Q_1(x), \ldots, Q_n(x)\}$

$$\mathcal{P}_n : L^2_\mu([a,b]) \to \mathbb{P}_n([a,b]) \tag{96}$$

as

$$\mathcal{P}_n f(x) = \sum_{k=0}^{n} a_k Q_k(x), \qquad a_k = \frac{(f, Q_k)}{(Q_k, Q_k)}. \tag{97}$$

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.,

$$\left\|f - \mathcal{P}_n f\right\|^2_{L^2_\mu} = \inf_{p \in \mathbb{P}_n([a,b])} \left\|f - p\right\|^2_{L^2_\mu}. \tag{98}$$

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

$$\lim_{n\to\infty} \left\|f - \mathcal{P}_n f\right\|^2_{L^2_\mu} = 0. \tag{99}$$

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 (83) we have the following approximation result (see [5, p. 109] or [18, 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

$$\left\|f - \mathcal{P}_n f\right\|^2_{L^2([-1,1])} \leq C n^{-s} \left\|f\right\|_{H^s([-1,1]} \tag{100}$$

where $\mathcal{P}_n f$ is the orthogonal projection of f onto the space of Legendre polynomials (Eq. (97)).

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*

$$\left\|f - \mathcal{P}_n f\right\|^2_{L^2([-1,1])} \sim e^{-\beta n}. \tag{101}$$

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 $\{X_j(\omega)\}_{j=1,2,\ldots}$ 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 distribution* if for all bounded continuous functions $h : \mathbb{R} \to \mathbb{R}$ we have that

$$\lim_{j\to\infty} \mathbb{E}\left\{h(X_j)\right\} = \mathbb{E}\left\{h(X)\right\}. \tag{102}$$

This equation can be equivalently written as

$$\lim_{j\to\infty}\int_{-\infty}^{\infty} h(x)dF_{X_j}(x) = \int_{-\infty}^{\infty} h(x)dF_X(x) \quad \text{for all bounded continuous functions } h(x), \tag{103}$$

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 (103) that $F_{X_j}(x)$ converges to $F_X(x)$ pointwise, i.e.,

$$\sup_x \left|F_{X_j}(x) - F_X(x)\right| \xrightarrow[j\to\infty]{} 0. \tag{104}$$

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

$$dF_{X_j}(x) = p_{X_j}(x)dx, \qquad dF_X(x) = p_X(x)dx, \tag{105}$$

then (103) implies that

$$\sup_x \left|p_{X_j}(x) - p_X(x)\right| \xrightarrow[j\to\infty]{} 0, \tag{106}$$

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 $\epsilon \geq 0$

$$P\left(\{\omega\in\Omega : |X_j(\omega) - X(\omega)| > \epsilon\}\right) \xrightarrow[j\to\infty]{} 0. \tag{107}$$

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)

$$\{\omega : X_j(\omega)\leq a\} = \{\omega : X(\omega)\leq a+\epsilon\}\cup\{\omega : |X_j(\omega)-X(\omega)|>\epsilon\}. \tag{108}$$

Since the two set at the right hand side of (108) do intersect, we have[8]

$$\underbrace{P(\{\omega : X_j(\omega)\leq a\})}_{F_{X_j}(a)} \leq \underbrace{P(\{\omega : X(\omega)\leq a+\epsilon\})}_{F_X(a+\epsilon)} + P(\{\omega : |X_j(\omega)-X(\omega)|>\epsilon\}). \tag{110}$$

Similarly (see Figure 6),

$$P(\{\omega : X(\omega)\leq a-\epsilon\}) \leq P(\{\omega : X_j(\omega)\leq a\}) + P(\{\omega : |X_j(\omega)-X(\omega)|>\epsilon\}). \tag{111}$$

Combining (110)-(111) yields

$$F_X(a-\epsilon) - P(\{\omega : |X_j(\omega)-X(\omega)|>\epsilon\}) \leq F_{X_j}(a) \leq F_X(a+\epsilon) + P(\{\omega : |X_j(\omega)-X(\omega)|>\epsilon\}). \tag{112}$$

[8] Recall that for every pair of events A and B in the σ-algebra $\mathcal{F}$ we have:

$$P(A\cup B) = P(A) + P(B) - P(A\cap B) \leq P(A) + P(B). \tag{109}$$

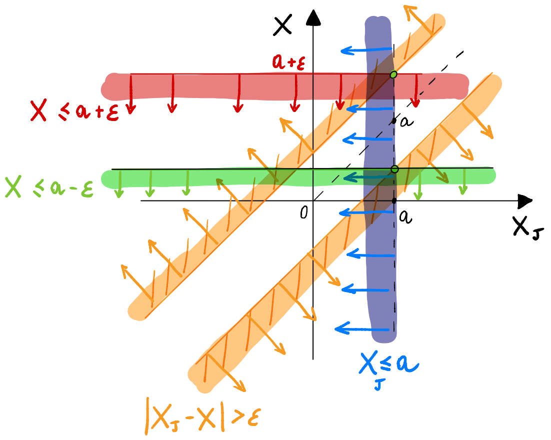


Figure 6: Sketch of the sets in equation (108). Clearly, $\{X_j \leq a\}$ is a subset of $\{X \leq a+\epsilon\} \cup \{|X_j - X| > \epsilon\}$, and $\{X \leq a - \epsilon\}$ is a subset of $\{X_j \leq a\} \cup \{|X_j - X| > \epsilon\}$.

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

$$\lim_{j\to\infty} P(\{\omega : |X_j(\omega) - X(\omega)| > \epsilon\}) = 0. \tag{113}$$

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

$$F_X(a-\epsilon) \leq F_{X_j}(a) \leq F_X(a+\epsilon). \tag{114}$$

If we send ϵ to zero we obtain (under continuity assumptions) 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 $\{X_j(\omega)\}_{j=1,2,\dots}$ 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 the *mean square* sense (or in $L^2(\Omega, \mathcal{F}, P)$) if

$$\lim_{j\to\infty} \mathbb{E}\left\{|X_j(\omega) - X(\omega)|^2\right\} = 0. \tag{115}$$

By using the Markov inequality

$$P(\{\omega : |X_j(\omega) - X(\omega)| > \epsilon\}) \leq \frac{1}{\epsilon^2}\mathbb{E}\left\{|X_j(\omega) - X(\omega)|^2\right\}, \tag{116}$$

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 (116) implies that the PDF of X_j converges to the PDF of X pointwise (see (106)).

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] 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.

[6] 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.

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

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

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

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

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

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

[13] 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.

[14] 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.

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

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

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

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

Sampling Methods

In this lecture note we discuss several sampling methods commonly used to propagate uncertainty in numerical simulations of nonlinear systems. The basic idea is very simple: compute samples of the solution to the model equations, corresponding to suitable samples of the random input variables. Such samples can be randomly generated (using pseudo-random number generators) as in the Monte Carlo method [7], or can be part of deterministic sequences as in the quasi Monte Carlo method [7, 4], sparse grids [3], or in the probabilistic collocation method [10, 5].

The most appropriate sampling scheme depends on the application, in particular on the number of random variables and the quantity of interest. For instance, if we are interested in approximating the expectation of a quantity of interest in a system that depends only on one random variable $\xi(\omega)$ with bounded range then perhaps Monte-Carlo is not the most efficient method. Indeed, in this case, it is rather straightforward to derive a highly accurate Gauss quadrature rule to approximate the expectation as

$$\mathbb{E}\left\{h(\xi)\right\} = \int_a^b h(x)p_\xi(x)dx \simeq \sum_{k=1}^N h(\xi_k)w_k. \tag{1}$$

As we will see, if the function h of class C^∞, then the Gauss quadrature rule (1) can converge exponentially fast with N (number of samples). On the other hand, the convergence rate of the Monte Carlo method to approximate (1) is $1/\sqrt{N}$. On the other hand, if the system is driven by high-dimensional random input vector then Gauss quadrature becomes impractical, and oftentimes we are left with no other choice than random sampling.

A distinctive advantage of sampling methods over polynomial chaos or PDF methods is that they are *non-intrusive*. This means that they do not require devising equations or writing new codes and algorithms from scratch perform UQ analyses, but rather simply run existing deterministic algorithms and codes many times, eventually in a massively parallel way.

Monte Carlo (MC)

Monte Carlo methods are a broad class of computational algorithms that rely on repeated *random sampling* to obtain numerical results of various types, e.g., estimation of high-dimensional PDFs or approximation of high-dimensional integrals representing, e.g., expectation operators.

Example (PDF estimation): Suppose we are interested in estimating the PDF of a random variable Y depending on three random variables X_1, X_2 and X_3. We are given joint PDF of (X_1, X_2, X_3), i.e., $p(x_1, x_2, x_3)$ and the mapping

$$Y = g(\boldsymbol{X}). \tag{2}$$

In a (Markov-Chain) Monte-Carlo setting the estimation of the PDF Y proceeds as follows:

1. Determine N samples of $p(x_1, x_2, x_3)$, e.g., using Gibbs sampling (see Chapter 1). This yields $\left\{\boldsymbol{X}^{[1]}, \ldots, \boldsymbol{X}^{[N]}\right\}$;
2. Compute N samples of Y using (2), i.e., $Y^{[j]} = g\left(\boldsymbol{X}^{[j]}\right)$;
3. Estimate the joint PDF of $Y(\omega)$ using relative frequencies, or kernel density estimation [2].

Example (Expectation operator): Suppose we are interested in approximating the mean of scalar phase space function of interest $h(\boldsymbol{x})$ depending on the solution of the ODE system

$$\begin{cases} \dfrac{d\boldsymbol{x}}{dt} = \boldsymbol{G}(\boldsymbol{x}, \boldsymbol{\xi}(\omega), t) \\ \boldsymbol{x}(0;\omega) = \boldsymbol{x}_0 \end{cases} \tag{3}$$

where $\boldsymbol{x}_0$ is deterministic and $\boldsymbol{\xi}$ is a random vector. The expectation of $h(\boldsymbol{x}(t;\omega))$ can be written as

$$\mathbb{E}\left\{h(\boldsymbol{x}(t;\omega))\right\} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h\left(\boldsymbol{x}\left(t;\boldsymbol{y}\right)\right) p_{\boldsymbol{\xi}}(\boldsymbol{y}) d\boldsymbol{y}, \tag{4}$$

i.e., as a high-dimensional integral over the PDF of $\boldsymbol{\xi}$. In a Monte-Carlo setting, such an integral is approximated by an *equal-weight* quadrature formula of the form

$$\mathbb{E}\left\{h(\boldsymbol{x}(t;\omega))\right\} \simeq \frac{1}{N}\sum_{k=1}^{N} h\left(\boldsymbol{x}\left(t;\boldsymbol{\xi}^{[k]}\right)\right), \tag{5}$$

where $\{\boldsymbol{\xi}^{[1]}, \ldots, \boldsymbol{\xi}^{[N]}\}$ are independent random samples obtained from $p(\boldsymbol{\xi})$ using, e.g., Gibbs sampling.

Similarly, Monte Carlo can be used to obtain response samples of random eigenvalue problems (random eigenvalues and random eigenvectors), solution to PDE, etc.

Monte Carlo integration. Consider the following mapping between an n-dimensional random vector $\boldsymbol{X}$ and a random variable Y

$$Y = g(\boldsymbol{X}). \tag{6}$$

We are interested in computing

$$\mathbb{E}\{Y\} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(\boldsymbol{x}) p_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x}. \tag{7}$$

To this end, we draw N independent random samples $\left\{\boldsymbol{X}^{[1]}, \ldots, \boldsymbol{X}^{[N]}\right\}$, and approximate the integral at the right hand side of (7) as

$$\mathbb{E}\{Y\} \simeq \frac{1}{N}\sum_{k=1}^{N} g\left(\boldsymbol{X}^{[k]}\right). \tag{8}$$

The following error bound then holds true.

Theorem 1. For all functions $g \in L^2_{p_{\boldsymbol{X}}}(\mathbb{R}^n)$, i.e., for all random variables $Y = G(\boldsymbol{X})$ with finite second-order moment we have

$$\widehat{\mathbb{E}}\left\{\left|\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(\boldsymbol{x}) p_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x} - \frac{1}{N}\sum_{k=1}^{N} g\left(\boldsymbol{X}^{[k]}\right)\right|^2\right\} = \frac{\sigma^2(g)}{N}, \tag{9}$$

where $\widehat{\mathbb{E}}$ is an expectation of the joint PDF of $\left\{\boldsymbol{X}^{[1]}, \ldots, \boldsymbol{X}^{[N]}\right\}$ (treated as independent random vectors), and

$$\sigma^2(g) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g^2(\boldsymbol{x}) p_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x} - \left(\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(\boldsymbol{x}) p_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x}\right)^2 \tag{10}$$

is the variance of $g(\boldsymbol{X})$, i.e., the variance of Y in (6).

The proof of this theorem is available in [4, 7] and therefore omitted here. Note that if we approximate (7) using (8) and different sets of samples $\{\boldsymbol{X}^{[1]}, \ldots, \boldsymbol{X}^{[N]}\}$ then we obtain different results. Hence, we should really think of (8) as a sum of independent random variables $\boldsymbol{X}^{[k]}$, each one of which is distributed as $p_{\boldsymbol{X}}(\boldsymbol{x})$. This means that the right hand side of (8) can be thought of as a sum of *independent random variables.*

By using the central limit theorem[1] it is straightforward to obtain the following probabilistic error bound on the MC approximation (8)

$$\lim_{N\to\infty} P\left(\left\{\omega : \left|\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g(\boldsymbol{x})p_{\boldsymbol{X}}(\boldsymbol{x})d\boldsymbol{x} - \frac{1}{N}\sum_{k=1}^{N} g\left(\boldsymbol{X}^{[k]}(\omega)\right)\right| \leq c\frac{\sigma(g)}{\sqrt{N}}\right\}\right) = \frac{1}{\sqrt{2\pi}}\int_{-c}^{c} e^{-y^2/2}dy. \quad (14)$$

To this end, we simply substitute $Y^{[k]} = g\left(\boldsymbol{X}^{[k]}\right)$, $\sigma^2 = \sigma^2(g)$ and

$$m = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g(\boldsymbol{x})p_{\boldsymbol{X}}(\boldsymbol{x})d\boldsymbol{x}. \quad (15)$$

into (13).

Equation (14) is an asymptotic probabilistic error bound stating that as we increase the number of samples the MC approximation goes to zero as[2] $1/\sqrt{N}$. Note that (14) is *independent of the dimension of the integral* (dimension of the vector $\boldsymbol{x}$), which is a great deal that makes MC suitable for high-dimensional integration. Similarly, by using the Markov inequality, it follows from (9) that

$$P\left(\left\{\omega : \left|\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g(\boldsymbol{x})p_{\boldsymbol{X}}(\boldsymbol{x})d\boldsymbol{x} - \frac{1}{N}\sum_{k=1}^{N} g\left(\boldsymbol{X}^{[k]}(\omega)\right)\right| \geq \epsilon\right\}\right) \leq \frac{\sigma(g)}{\epsilon\sqrt{N}}. \quad (16)$$

Remark: While independent of the dimension of the integral, the convergence rate $O(N^{-1/2})$ of the Monte Carlo approximation (8) is not too great. Roughly speaking, to obtain a one digit increase in accuracy we need 100 times more samples! To show this Let E_1 be the integration error. We know that E_1 is proportional to $N_1^{-1/2}$, i.e.,

$$E_1 = CN_1^{-1/2}. \quad (17)$$

To obtain an error $E_2 = E_1/10$, i.e., gain one digit accuracy, we need

$$CN_2^{-1/2} = CN_1^{-1/2}/10 \quad \Rightarrow \quad N_2 = 100N_1. \quad (18)$$

Hence, if we get the first two digits of our integral right with an MC formula involving 5000 random samples, then we would need roughly 500000 samples to get third digit right!

[1] The central limit theorem can be stated as follows: let $\{Y^{[1]}, \ldots, Y^{[N]}\}$ be a sequence of i.i.d. random variables with mean m and variance σ^2. Define

$$Z_N = \sqrt{N}\left(\frac{1}{N}\sum_{k=1}^{N} Y^{[k]} - m\right). \quad (11)$$

Then the PDF of Z_N converges to a normal distribution with zero mean and variance σ^2, i.e.,

$$\lim_{N\to\infty} p_{Z_N}(x) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-x^2/(2\sigma^2)}. \quad (12)$$

This means that

$$\lim_{N\to\infty} P\left(\{\omega : |Z_N(\omega)| \leq c\sigma\}\right) = \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-c\sigma}^{c\sigma} e^{-x^2/(2\sigma^2)}dx = \frac{1}{\sqrt{2\pi}}\int_{-c}^{c} e^{-y^2/2}dy. \quad (13)$$

[2] Simply set $c = 3$ in (14) to obtain the right hand side approximately equal to 1.

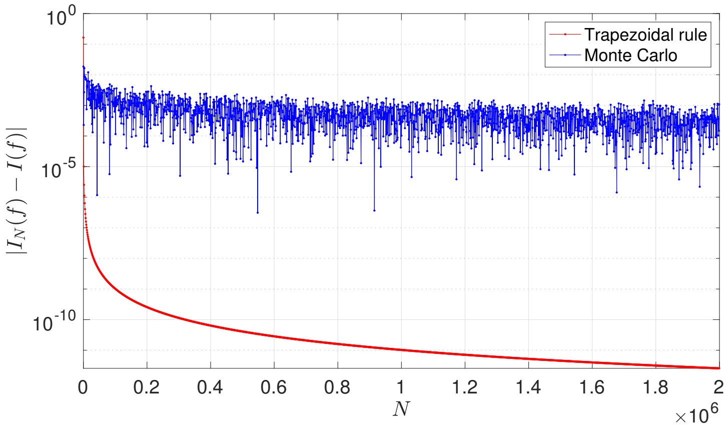


Figure 1: Error in the numerical approximation of the integral (19) using the Monte Carlo rule (8) and the trapezoidal rule versus the number of points N.

As an example, in Figure 1 we compare the error in the numerical approximation of the integral

$$I(f) = \int_{-1}^{1} g(x)dx, \qquad g(x) = e^{-x}x^2 \sin(10x)^2 \tag{19}$$

using Monte Carlo and the trapezoidal rule. For Monte Carlo, we simply compute N independent samples of a uniform random variable X in $[-1,1]$ and compute the sum

$$I_N(f) = \frac{2}{N}\sum_{k=1}^{N} g(X^{[k]}). \tag{20}$$

The factor 2 accounts for the fact that the PDF of a uniform variable in $[-1,1]$ is $1/2$. Note that on average the convergence rate of MC is $1/\sqrt{N}$ while the converge rate of the trapezoidal rule is $1/N^2$.

Quasi Monte Carlo (QMC)

Just like Monte Carlo, quasi-Monte Carlo methods [7, 4] aim at representing multidimensional integrals of the form (7) as equal-weights quadrature rules (8). However, in QMC the sequence of points $\boldsymbol{X}^{[k]}$ are not realizations of a random vector, but rather elements of a *deterministic sequence* called low-discrepancy sequence. The whole point such low-discrepancy sequences is to improve the (very) slow convergence rate of MC (i.e., $O(N^{-1/2})$) when evaluating multidimensional integrals of the form

$$I(g) = \int_{[0,1]^n} g(\boldsymbol{x})d\boldsymbol{x}. \tag{21}$$

QMC methods are usually classified based on how the points in the low-discrepancy sequences are computed. In particular, we can have sequences of points that can be increased without recomputing the first few points (open QMC formulas), and sequences of points that require recalculation of all points if N changes (closed QMC formulas) [4]. Hereafter we provide a two examples of QMC rules leveraging the radical inverse function.

Radical inverse function. Let $b \geq 2$ be a natural number. As is well known, any integer number can be represented relative to the base b as

$$i = \sum_{k=1}^{\infty} i_k b^{k-1} \tag{22}$$

where i_k can take values in $\{0, 1, \dots, b-1\}$. For example, the number 11 can be written in base 2 and base 3 as c

$$11 = 1 \times 2^0 + 1 \times 2^1 + 0 \times 2^2 + 1 \times 2^3 = [\cdots 01011]_2, \tag{23}$$

$$= 2 \times 3^0 + 0 \times 3^1 + 1 \times 3^2 = [\cdots 0102]_3. \tag{24}$$

We define the radical inverse function corresponding to an integer number $i \in \mathbb{N}_0$

$$\phi_b(i) = \sum_{k=1}^{\infty} \frac{i_k}{b^k}. \tag{25}$$

The function (25) operates as follows:

$$i = [\cdots i_3 i_2 i_1]_b \qquad \Rightarrow \qquad \phi_b(i) = [0.i_1 i_2 i_3 \cdots]_b. \tag{26}$$

With reference to (23)-(24) we have, for example,

$$\phi_2(11) = 1 \times 2^{-1} + 1 \times 2^{-2} + 1 \times 2^{-4} = \frac{1}{2} + \frac{1}{4} + \frac{1}{16} = \frac{13}{16}, \tag{27}$$

$$\phi_3(11) = 2 \times 3^{-1} + 1 \times 3^{-3} = \frac{2}{3} + 1\frac{1}{27} = \frac{19}{27}. \tag{28}$$

Halton's sequence. The Halton's sequence is a point set in the hypercube $[0,1]^n$ defined as

$$\boldsymbol{X}_{Hl}^{[i]} = (\phi_{p_1}(i), \dots, \phi_{p_n}(i)) \qquad i = 1, 2, \dots, N, \tag{29}$$

where $\{p_1, \dots, p_n\}$ are the first n prime numbers, and $\phi_{p_j}(i)$ is the radical inverse function (25). For example, in dimension $n = 5$ we have

$$\boldsymbol{X}_{Hl}^{[i]} = (\phi_2(i), \phi_3(i), \phi_5(i), \phi_7(i), \phi_{11}(i)) \qquad i = 1, 2, \dots, N. \tag{30}$$

By using the Halton's sequence we can approximate integrals relative to uniform PDFs in $[0,1]^n$ as

$$\int_{[0,1]^n} g(\boldsymbol{x}) d\boldsymbol{x} \simeq \frac{1}{N} \sum_{k=1}^{N} g\left(\boldsymbol{X}_{Hl}^{[k]}\right). \tag{31}$$

It can be shown that (see [4])

$$\left| \int_{[0,1]^n} g(\boldsymbol{x}) d\boldsymbol{x} - \frac{1}{N} \sum_{k=1}^{N} g\left(\boldsymbol{X}_{Hl}^{[k]}\right) \right| \leq C_n \frac{(\log(N))^n}{N} V_{HK}(g), \tag{32}$$

where $V_{HK}(g)$ is the variation of $g(\boldsymbol{x})$ in the sense of Hardy and Krause (see [4]). For fixed g we have that $V_{HK}(g)$ is a number depending only on g. The function $(\log(N))^n / N$ defining the upper bound in (32) has an asymptote at $N = 0$, a minimum at $N = 1$ (equal to zero), a maximum at $N = e^n$ (equal to $(n/e)^n$), and goes to zero faster than $N^{-1/2}$ as N goes to infinity (see Figure 2. In dimension $n = 10$ we have $e^n = 22026$. Hence, to go past the "hump" in dimension $n = 10$ we need $N > 22026$ samples.

Hammersley's sequence. The Hammersley's sequence is a point set in the hypercube $[0,1]^n$ defined as

$$\boldsymbol{X}_{Hm}^{[i]} = \left(\frac{i}{N}\phi_{p_1}(i), \dots, \phi_{p_n}(i)\right) \qquad i = 1, 2, \dots, N-1 \tag{33}$$

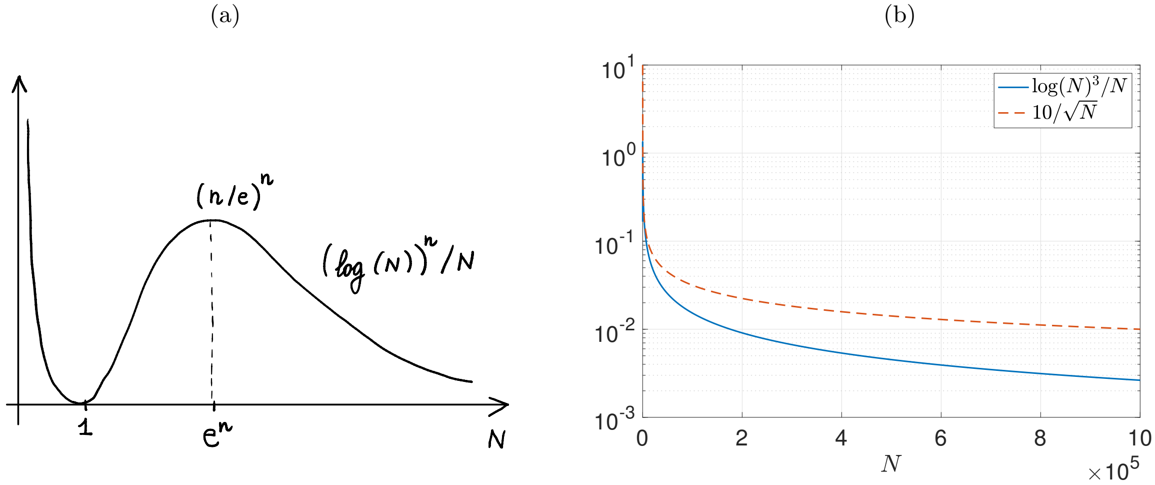


Figure 2: (a) Sketch of the upper bound of the approximation error of the quasi-Monte Carlo quadradure rule based on the Halton's sequence applied to an n-dimensional integral versus the number of samples N. (b) Comparison between the decay rate for of Halton QMC and MC for $n = 3$.

where $\{p_1, \ldots, p_n\}$ are the first n prime numbers, and $\phi_{p_j}(i)$ is the radical inverse function (25). Note that the first column in (33) needs to be recomputed if we change N. It can be shown (e.g., [4]) that

$$\left| \int_{[0,1]^n} g(\boldsymbol{x}) d\boldsymbol{x} - \frac{1}{N} \sum_{k=1}^{N} g\left(\boldsymbol{X}_{Hm}^{[i]} \right) \right| \leq C_n \frac{(\log(N))^{n-1}}{N} V_{HK}(g), \tag{34}$$

which represents a slight improvement over (32).

Remark: To further improve the convergence rate of quasi-Monte Carlo one can introduce *randomizations* of the QMC point sets, e.g., in the form of *random shifts* or *point scrambling*. The randomization allows us derive probabilistic error bounds similar to MC, and at the same time can improve the convergence rate of QMC (see [7, 4]).

Probabilistic collocation method (PCM)

The probabilistic collocation method is a high-order method based on deterministic point sets that allows us to compute expectation operators involving low-dimensional integrals. The method leverages high-order interpolatory quadrature rules [9], in particular, Gaussian quadrature.

We have seen in previous lecture that orthogonal polynomials play a fundamental role in the approximation of smooth functions. As we shall see hereafter, orthogonal polynomials play also a crucial role in devising interpolants and quadrature formulae with maximal *degrees of exactness*[3]. These formulae are known as Gaussian quadrature formulae [9, §10.2].

To introduce Gaussian quadrature in the context of UQ, suppose we are given a random variable X with range $[a, b]$ and PDF $p_X(x)$. We have seen that for every measurable function $g : [a, b] \to \mathbb{R}$ the expectation of $g(X)$ is defined as

$$\mathbb{E}\{g(X)\} = \int_a^b g(x) p_X(x) dx. \tag{35}$$

[3] The degree of exactness of a quadrature formula is the maximum degree of the polynomial that can be integrated exactly by the formula. In other words, we say that a quadrature formula has degree of exactness p if it can integrate exactly polynomials of degree p or less.

By using the coordinate transformation

$$x = \frac{b-a}{2}z + \frac{b+a}{2} \qquad z = \frac{2}{b-a}\left(x - \frac{b+a}{2}\right) \qquad z \in [-1,1] \tag{36}$$

we can rewrite the expectation in (35) as

$$\int_a^b g(x)p_X(x)dx = \frac{b-a}{2}\int_{-1}^{1} f(z)\mu(z)dz, \tag{37}$$

where

$$f(z) = g\left(\frac{b-a}{2}z + \frac{b+a}{2}\right), \qquad \mu(z) = p_X\left(\frac{b-a}{2}z + \frac{b+a}{2}\right). \tag{38}$$

For the approximation of the weighted integral at the right hand side of (37), we consider the quadrature rule

$$\int_{-1}^{1} f(z)\mu(z)dz \simeq \sum_{k=0}^{M} f(z_k)w_k, \tag{39}$$

where $\{z_0, \dots, z_M\}$ are quadrature points in $[-1,1]$ while $\{w_0, \dots, w_M\}$ are quadrature weights.

If we approximate $f(z)$ by the Lagrange interpolation polynomial $\Pi_M f(z)$ at the $M+1$ nodes $\{z_0, \dots, z_M\}$ then (39) is a quadrature formula that has degrees of exactness at least equal to M, and explicit expression for the quadrature weights w_k. This follows from

$$\int_{-1}^{1} f(z)\mu(z)dz \simeq \int_{-1}^{1} \Pi_M f(z)\mu(z)dz = \sum_{k=0}^{M} f(z_k)\underbrace{\int_{-1}^{1} l_k(z)\mu(z)dz}_{w_k}, \tag{40}$$

where

$$l_k(z) = \prod_{\substack{j=0\\ j\neq k}}^{M} \frac{z-z_j}{z_k - z_j} \tag{41}$$

are the Lagrange characteristic polynomial associated with the grid $\{z_0, \dots, z_M\}$.

At this point the question is whether suitable choices of the nodes exist such that the degree of exactness is greater than M, say, equal to $r = M+m$ for some $m > 0$. The answer is given by the following theorem

Theorem 2 (Gaussian quadrature - Jacobi's theorem)**.** For any given $m > 0$ the interpolatory quadrature rule (39) has degree of exactness $M+m$ if and only if the polynomial

$$q_{M+1}(z) = \prod_{j=0}^{M}(z - z_j) \tag{42}$$

associated with the nodes $\{z_0, \dots, z_M\}$ satisfies the orthogonality conditions

$$\int_{-1}^{1} q_{M+1}(z)b(z)\mu(z)dz = 0 \tag{43}$$

for all polynomial $b(z)$ of degree at most $m-1$.

In other words, if we can find a set of nodes $\{z_0, \dots, z_M\}$ such that $q_{M+1}(z)$ is orthogonal in $L^2_\mu([-1,1])$ to any polynomial of degree $m-1$ then the quadrature rule (39) has degree of exactness $M+m$.

Proof. Suppose that $f(z)$ in (39) is a polynomial of degree $m + M$. Divide $f(z)$ by (42) to obtain[4]

$$f(z) = \underbrace{q_{M+1}(z)}_{\text{divisor}}\underbrace{d_{m-1}(z)}_{\text{quotient}} + \underbrace{r_M(z)}_{\text{reminder}} \tag{45}$$

Note that the degree of the quotient is $(m + M) - (M + 1) = m - 1$ while the degree of the remainder is $(M + 1) - 1 = M$ (i.e., a polynomial that cannot be divided by $q_{M+1}(z)$). Since $r_M(z)$ is a polynomial of degree M it can be integrated exactly by the quadrature rule with $M + 1$ nodes. This yields,

$$\sum_{k=1}^{M} w_k r_M(z_k) = \int_{-1}^{1} r_M(z)\mu(z)dz = \int_{-1}^{1} f(z)\mu(z)dz - \int_{-1}^{1} q_{M+1}(z)d_{m-1}(z)\mu(z)dz. \tag{46}$$

If hypothesis (43) holds true then the last term at the right hand side vanishes. This allows us to conclude that

$$\int_{-1}^{1} f(z)\mu(z)dz = \sum_{k=1}^{M} w_k r_M(z_k), \tag{47}$$

i.e., that the polynomial $f(z)$ of degree $M + m$ can be integrated exactly on the grid with $M + 1$ points $\{z_0, \ldots, z_M\}$ satisfying the condition (43).

□

Example (Gauss-Legendre quadrature): Let $\{z_0, \ldots, z_M\}$ the zeros of the Legendre orthogonal polynomial $L_{M+1}(z)$, i.e., $L_{M+1}(z_j) = 0$. Clearly, the nodal polynomial $q_{M+1}(z)$ in theorem (2) coincides (modulus sign) with $L_{M+1}(z)$. In fact, $q_{M+1}(z)$ and $L_{M+1}(z)$ have the same zeros. Setting $\mu(z) = 1$ in (43) (Legendre polynomials are orthogonal in $[-1, 1]$ with respect to $\mu(z) = 1$) yields

$$\int_{-1}^{1} L_{M+1}(z)b(z)dz = 0. \tag{48}$$

At this point we write the polynomial $b(z)$ (of degree $m - 1$) in terms of a linear combination of Legendre polynomials

$$b(z) = \sum_{j=0}^{m-1} b_j L_j(z). \tag{49}$$

Next, substitute (49) into (48) to obtain

$$\sum_{j=0}^{m-1} b_j \int_{-1}^{1} L_{M+1}(z)L_j(z)dz = 0. \tag{50}$$

By using orthogonality of the Legendre polynomials we see that the maximum degree $m - 1$ of the polynomial $b(z)$ that satisfies equation (50) is $m - 1 = M$ (i.e., $m = M + 1$). Hence, the degree of exactness of Gauss-Legendre quadrature is $M + m = 2M + 1$. This means that with $M + 1$ points we can integrate

[4] As an example, consider the polynomial division of $f(z) = z^3 + z^2 - 3z + 4$ by $q_2(z) = z^2 - 3z + 2$. To this end, we firt multiply $q_2(z)$ by z to obtain $z^3 - 3z^2 + 2z$. Subtracting this from $f(z)$ yields the reminder $4z^2 - 5z + 4$. At this point we multiply $q_2(z)$ by 4, i.e., $4q_2(z) = 4z^2 - 12z + 8$ and subtract it from $4z^2 - 5z + 4$ to obtain the final remainder $7z - 4$. Hence, we obtained the factorization

$$z^3 + z^2 - 3z + 4 = \underbrace{(z^2 - 3z + 2)}_{\text{divisor}}\underbrace{(z + 4)}_{\text{quotient}} + \underbrace{(7z - 4)}_{\text{reminder}}. \tag{44}$$

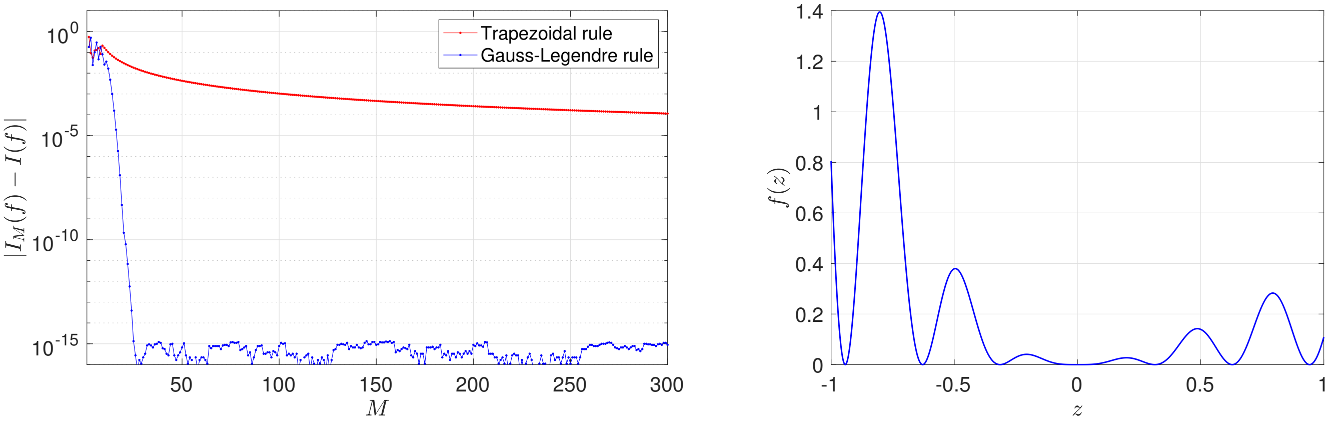


Figure 3: Error in the numerical approximation of the integral (53) using the Gauss-Legendre quadrature rule and the trapezoidal rule versus the number of collocation points M. Note that the Gauss-Legendre rule converges exponentially fast. In particular, with only 25 points the Gauss-Legendre rule achieves error 1.3×10^{-15}. On the other hand, the trapezoidal rule with 300 points achieves error 10^{-4}.

exactly polynomials up to degree $2M + 1$! Not really intuitive, huh? Regarding the integration weights for the Gauss-Legendre quadrature, it can be shown that

$$w_j = \frac{2}{(1 - z_j^2)\left[L'_{M+1}(z_j)\right]^2} \qquad j = 0, \ldots, n. \tag{51}$$

Moreover, for every $f \in H^s([-1,1])$ we have the following spectral convergence result[5] [9, p. 437]

$$\left| \int_{-1}^{1} f(z)dz - \sum_{k=0}^{M} f(z_k)w_k \right| \leq CM^{-s} \left\| f \right\|_{H^s([-1,1])}. \tag{52}$$

In Appendix A, we discuss similar results for Chebyshev-Gauss-Lobatto quadrature. If f is infinitely differentiable then convergence is exponential. As an example, in Figure 3 we compare the error in the numerical approximation of the integral

$$I(f) = \int_{-1}^{1} f(z)dz \qquad f(z) = e^{-z}z^2\sin(10z)^2 \tag{53}$$

using the Gauss-Lobatto rule and the trapezoidal rule.

Lemma 1. The *maximum degree of exactness* of the interpolatory quadrature formula (39) is $2M + 1$.

Proof. The proof is very simple. Suppose we could choose the max degree of $b(z)$ to be $m = M + 2$. Following what we just said for the Gauss-Legendre quadrature this would imply

$$\int_{-1}^{1} q_{M+1}^2(z)\mu(z)dz = 0, \tag{54}$$

i.e., $q_{M+1}^2(z) = 0$ which is impossible (see [9, Corollary 10.2]).

□

[5] The error estimate holds for Gauss-Legendre-Lobatto quadrature (see Table 1), which has degree of exactness $2M - 1$.

	Gauss-Legendre	Gauss-Legendre-Lobatto
nodes $\{z_0,\ldots,z_M\}$	$L_{M+1}(z)=0$	$(1-x^2)L'_M(z)=0$
Lagrange polynomials	$l_i(z)=\dfrac{L_{M+1}(z)}{(z-z_i)L'_{M+1}(z)}$	$l_i(z)=-\dfrac{1}{M(M+1)}\dfrac{(1-z^2)}{(z-z_i)}\dfrac{L'_M(z)}{L'_M(z_i)}$
Integration weights	$w_i(z)=\dfrac{2}{(1-z_i^2)\left[L'_{M+1}(z_i)\right]^2}$	$w_i(z)=\dfrac{1}{M(M+1)L_M(z_i)^2}$

Table 1: Gauss-Legendre (GL) and Gauss-Lobatto-Legendre (GLL) quadrature and interpolation rules. The GL rule has degree of exactness $2M+1$, while the GLL rule has degree of exactness $2M-1$.

Gauss and Gauss-Lobatto points have also excellent properties when used for polynomial interpolation (see Appendix B).

Computation of Gaussian quadrature points and weights. With the exception of a few special cases, like the Chebyshev polynomials, no closed form expressions for the quadrature nodes and weights are known (see, e.g., Table 1). Nevertheless, there is a simple and elegant way of computing these nodes as well as the corresponding weights based on the eigenvalues suitable tridiagonal matrices [6, §11.2]. The method relies on the three-term recurrence relation for orthogonal polynomials, written hereafter for monic orthogonal polynomials

$$\pi_{n+1}(z)=(z-\alpha_n)\pi_n(x)-\beta_n\pi_{n-1}(x). \tag{55}$$

We have seen that the coefficients α_n and β_n can be computed for every measure $\mu(z)$ using the Stieltjes algorithm (see Appendix B of Chapter 4). Equation (55) can be rewritten as

$$z\pi_n(z)=\pi_{n+1}(z)+\alpha_n\pi_n(x)+\beta_n\pi_{n-1}(x), \tag{56}$$

or in the following convenient matrix-vector form

$$z\underbrace{\begin{bmatrix}\pi_0(z)\\ \pi_1(z)\\ \pi_2(z)\\ \vdots\\ \pi_{n-1}(z)\\ \pi_n(z)\end{bmatrix}}_{\boldsymbol{\pi}(z)}=\underbrace{\begin{bmatrix}\alpha_0 & 1 & 0 & 0 & \cdots & 0\\ \beta_1 & \alpha_1 & 1 & 0 & \cdots & 0\\ 0 & \beta_2 & \alpha_2 & 1 & \cdots & 0\\ \vdots & & \ddots & \ddots & \ddots & \vdots\\ 0 & \cdots & 0 & \beta_{n-1} & \alpha_{n-1} & 1\\ 0 & \cdots & 0 & 0 & \beta_n & \alpha_n\end{bmatrix}}_{\text{Jacobi matrix } \boldsymbol{J}}\underbrace{\begin{bmatrix}\pi_0(z)\\ \pi_1(z)\\ \pi_2(z)\\ \vdots\\ \pi_{n-1}(z)\\ \pi_n(z)\end{bmatrix}}_{\boldsymbol{\pi}(z)}+\begin{bmatrix}0\\ 0\\ 0\\ \vdots\\ 0\\ \pi_{n+1}(z)\end{bmatrix}. \tag{57}$$

At this point, it is clear that the zeros of $\pi_{n+1}(z)$ are eigenvalues of the Jacobi matrix $\boldsymbol{J}$. In fact, if z_j is such that $\pi_{n+1}(z_j)=0$ then

$$\boldsymbol{J}\boldsymbol{\pi}(z_j)=z_j\boldsymbol{\pi}(z_j). \tag{58}$$

This eigenvalue problem may be solved using the QR algorithm. This yields the Gauss quadrature points $\{z_0,\ldots,z_n\}$. The corresponding quadrature weights can be computed by expanding each Lagrange polynomial $l_j(z)$ in (40) in terms of $\pi_j(z)$, and using orthogonality of $\pi_j(z)$ relative $\mu(z)$ to we obtain

$$w_k=\int_{-1}^{1}l_k(z)\mu(z)dz=\sum_{j=0}^{M}a_{kj}\int_{-1}^{1}\pi_j(z)\mu(z)dz=a_{k0}\int_{-1}^{1}\mu(z)dz \qquad \text{(integration weights)}. \tag{59}$$

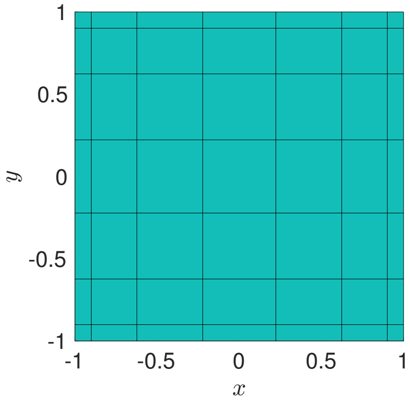


Figure 4: Tensor product of two one-dimensional Chebyshev grids (97) of 8 points.

Quadrature and interpolation on tensor product grids. Suppose we are interested in approximating the integral of a two-dimensional function $g(x, y)$ in $[-1, 1]^2$ relative to the separable integration weight

$$\mu(x, y) = \mu_1(x)\mu_2(y), \tag{60}$$

i.e.,

$$\int_{[-1,1]^2} g(x, y)\mu_1(x)\mu_2(y)dxdy. \tag{61}$$

By leveraging the isomorphism

$$L^2_\mu([-1, 1]^2) = L^2_{\mu_1}([-1, 1]) \otimes L^2_{\mu_2}([-1, 1]) \tag{62}$$

we see that we can represent $g(x, y)$ in terms of a tensor product of one-dimensional bases involving functions of x alone and y alone. In particular, such bases could be Lagrange characteristic polynomials corresponding to appropriate one-dimensional grids in $x \in [-1, 1]$ and $y \in [-1, 1]$. Let us denote by

$$\{x_0, \ldots, x_M\} \quad \text{and} \quad \{y_0, \ldots, y_N\} \tag{63}$$

the aforementioned one-dimensional grids, and by $\{l_j(x)\}$ and $h_i(y)$ the corresponding Lagrange polynomials. Then 2D polynomial interpolant of the dataset $\{g(x_i, y_j)\}$ can be written as

$$\Pi g(x, y) = \sum_{i=0}^{M}\sum_{j=0}^{N} g(x_i, y_j)\, l_i(x)h_j(y). \tag{64}$$

Clearly, $\Pi g(x, y)$ is a polynomial of total degree $M + N$. In Figure 4 we show a tensor product of two Gauss-Chebyshev-Lobatto one-dimensional grids (97). In Figure 5 we plot a few 2D Lagrange characteristic polynomials $l_i(x)h_j(y)$ associated with the Chebyshev grid shown in Figure 4. The convergence rate of the interpolant (64) is determined by the tensor product interpolation grid. In particular, for each fixed $x = x_j$ or $y = y_k$ it is clear that the spectral convergence results summarized in Appendix B hold. With the 2D interpolant (64) available, it is straightforward to derive a 2D interpolatory quadrature rule. In

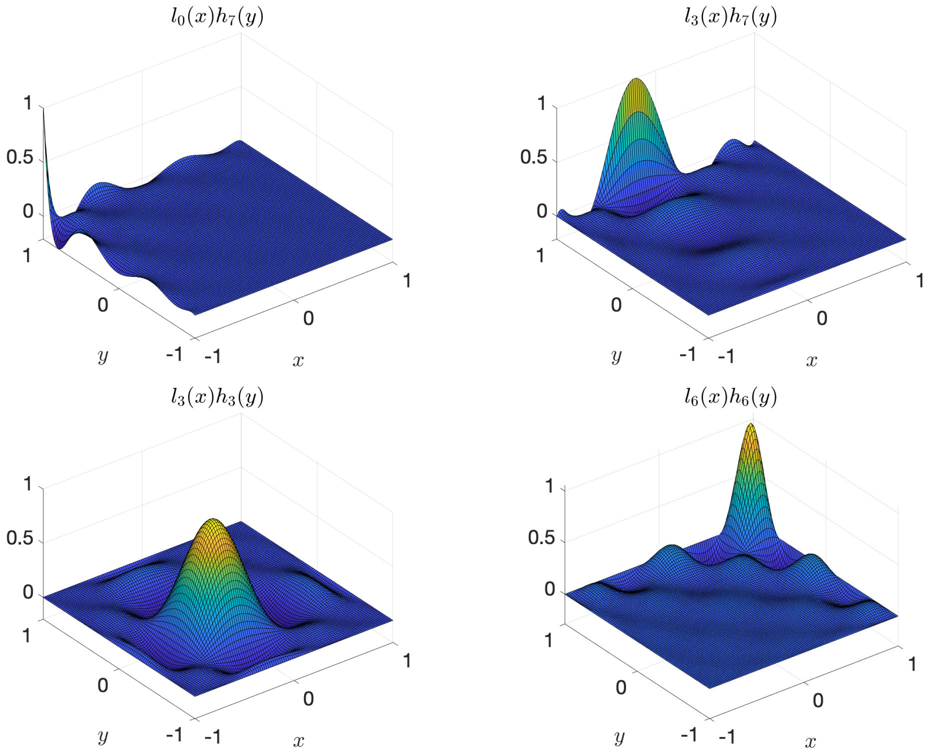


Figure 5: 2D Lagrange characteristic polynomials $l_i(x)h_j(y)$ associated with the 2D Chebyshev grid shown in Figure 4.

fact,

$$\begin{aligned}\int_{-1}^{1}\int_{-1}^{1} g(x,y)\mu_1(x)\mu_2(y)dxdy &\simeq \int_{-1}^{1}\int_{-1}^{1} \Pi g(x,y)\mu_1(x)\mu_2(y)dxdy\\ &=\sum_{i=0}^{M}\sum_{j=0}^{N} g\left(x_i,y_j\right)\underbrace{\int_{-1}^{1} l_i(x)\mu_1(x)dx}_{w_i}\underbrace{\int_{-1}^{1} h_j(y)\mu_2(y)dy}_{q_j}\\ &=\sum_{i=0}^{M}\sum_{j=0}^{N} g\left(x_i,y_j\right)w_i q_j. \end{aligned}\tag{65}$$

Next, consider the random variable

$$\eta(\omega)=g(\xi_1,\ldots,\xi_n)\tag{66}$$

and assume that all random variables $\{\xi_1,\ldots,\xi_n\}$ are i.i.d. with PDF $p_\xi(x)$ supported in $[-1,1]$. We have

$$\mathbb{E}\{\eta(\omega)\}=\int_{-1}^{1}\cdots\int_{-1}^{1} g(x_1,\ldots,x_n)p_\xi(x_1)\cdots p_\xi(x_n)dx_1\cdots dx_n.\tag{67}$$

We approximate this integral using tensor product PCM. To this end, we first construct a one-dimensional quadrature rule with high-degree of exactness (i.e., Gauss or Gauss-Lobatto) using the methods described in previous sections. With the one-dimensional quadrature points $\{x_0,\ldots,x_M\}$ and quadrature weights $\{w_0,\ldots,w_M\}$ available we approximate the integral in (67) as

$$\mathbb{E}\{\eta(\omega)\}=\sum_{j_1=0}^{N}\cdots\sum_{j_n=0}^{N} g\left(x_{j_1},\ldots,x_{j_n}\right)w_{j_1}\cdots w_{j_n}.\tag{68}$$

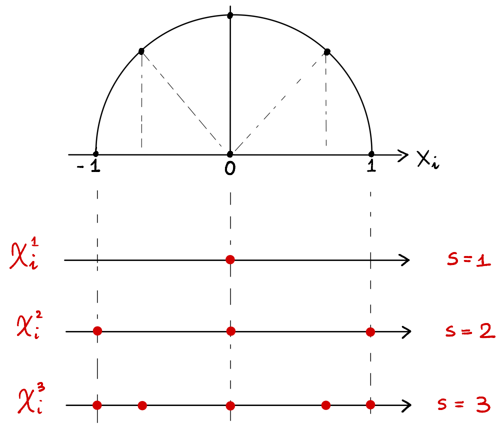


Figure 6: Chebyshev-Gauss-Lobatto (GCL) nested point sets χ_i^1, χ_i^2, and χ_i^3.

Computational cost: To compute all the sums we essentially need to evaluate all $g(\xi_1, \dots, x_n)$ at a number of points that grows as $(N+1)^n$, i.e., polynomially in N (number of points) and *exponentially* in n (dimension). In dimension $n = 10$ with just $N+1 = 10$ points per dimension this yields 10^{10} collocation points! Each point has n coordinates. Hence, to store the grid in double precision floating point (64 bits, 8 Bytes per floating point number) we need

$$10^{10} \times 10 \times 8 = 800\,\text{GB}. \tag{69}$$

To store $g\left(x_{j_1}, \dots, x_{j_n}\right)$ we need an extra 80 GB, and 1.25 Bytes for the vector of weights. Hence, similarly to polynomial chaos, tensor product PCM undergoes an exponential growth of degrees of freedom with the dimension of the problem. However, differently than polynomial chaos, PCM is a *non-intrusive* method that allows us to perform UQ calculations on legacy codes in a straightforward way, without coding polynomial chaos propagators or PDF equation solvers from scratch.

Sparse grids

Sparse grids are numerical techniques to represent, integrate or interpolate high dimensional functions. They were originally developed by the Russian mathematician Sergey A. Smolyak, and are based on a *sparse tensor product* construction [3]. The fundamental building block of sparse grids is a one-dimensional nested points set, e.g., the Gauss-Chebyshev-Lobatto grid (97) for $M = 2, 4, 8, \dots, 2^s$, or any other nested point set. To describe how sparse grids are constructed, let

$$\chi_i^s = \{x_i^1 \dots, x_i^{n_s}\} \tag{70}$$

be the nested points set in the variable x^i where

$$n_1 = 1, \qquad n_s = 2^{s-1} + 1 \quad (s \geq 2), \tag{71}$$

is the total number of points in the nested point set, e.g., in the Gauss-Chebyshev-Lobatto grid. In Figure 6 we provide a graphical representation of the GCL nested point set.

The *level* q sparse grids in d dimensions is defined as the multidimensional point set

$$H_q^d = \bigcup_{q+1-d \leq i_1 + \cdots + i_d \leq q} \chi_1^{i_1} \times \cdots \times \chi_d^{i_d}, \tag{72}$$

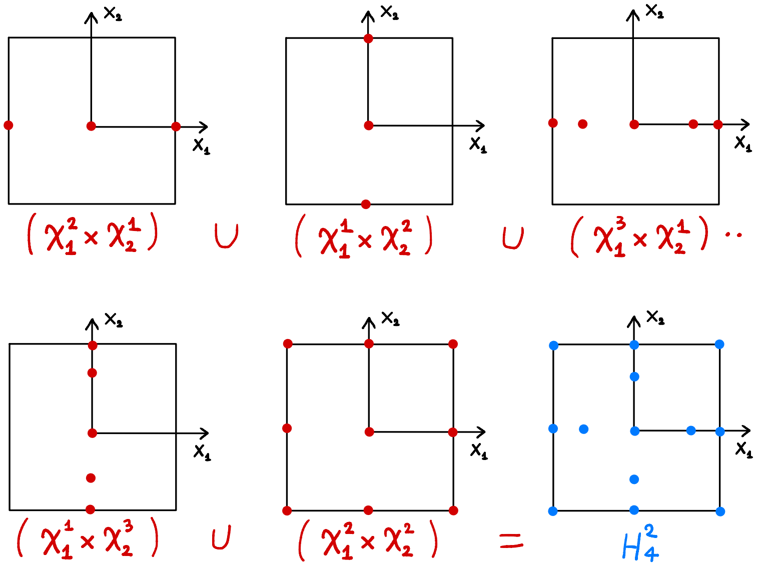

x_2
x_1
$(\chi_1^2 \times \chi_2^1) \cup (\chi_1^1 \times \chi_2^2) \cup (\chi_1^3 \times \chi_2^1) \cdots$
$(\chi_1^1 \times \chi_2^3) \cup (\chi_1^2 \times \chi_2^2) = H_4^2$

Figure 7: Construction of two-dimensional Chebyshev-Gauss-Lobatto (GCL) sparse grids of level 4 (see Eq. (74)). The final point set is denoted by H_4^2.

i.e., the union of suitable Cartesian products of one-dimensional grids. As we will see sparse grids follow naturally from the definition of Smolyak interpolant of a multivartiate function, which we will discuss in detail in the next section. For now, we simply notice that the point set (72) is nested in the sense that

$$H_{q-1}^d \subset H_q^d \tag{73}$$

if χ_i^s is a nested point set.

Example (Level 4 Gauss-Chebyshev-Lobatto (GCL) sparse grids in two-dimensions): To derive the level 4 GCL sparse grids in two dimensions we set $q = 4$ and $d = 2$ in (72). This yields

$$\begin{aligned} H_4^2 &= \bigcup_{3\leq i_1+i_2\leq 4} \chi_1^{i_1} \times \chi_2^{i_2} \\ &= \left(\chi_1^2 \times \chi_2^1\right) \cup \left(\chi_1^1 \times \chi_2^2\right) \cup \left(\chi_1^3 \times \chi_2^1\right) \cup \left(\chi_1^1 \times \chi_2^3\right) \cup \left(\chi_1^2 \times \chi_2^2\right). \end{aligned} \tag{74}$$

The Cartesian product grids appearing in this expression can be easily derived by taking Cartesian products of the elementary 1D grids shown in Figure 6. Such product grids are shown in Figure 7 In Figure 8 we plot CGL sparse grids of level 5 and 6 in 2D and 3D.

Interpolation on sparse grids. Let Π_i^s be the interpolation operator in the variable x_i corresponding to the 1D point set (70). Note that for $s = 1$ we have that χ_i^1 has only one point (see Figure 6). Therefore that Π_i^1 is an interpolant at one point only (for polynomial this is the constant function). Define the difference between two interpolation operators as[6]

$$\Delta_i^0 = 0 \qquad \Delta_i^s = \Pi_i^s - \Pi_i^{s-1}. \tag{75}$$

The *Smolyak interpolant* of a multivariate function $f(x_1, \ldots, x_d)$ is defined as

$$S_q^d(f) = \sum_{i_1+\cdots+i_d\leq q} \Delta_1^{i_1} \otimes \cdots \otimes \Delta_d^{i_d}, \tag{76}$$

[6] In equation (75) Π_i^s denotes the one-dimensional interpolant of a function $f(\boldsymbol{x})$ on a grid in the variable x_i.

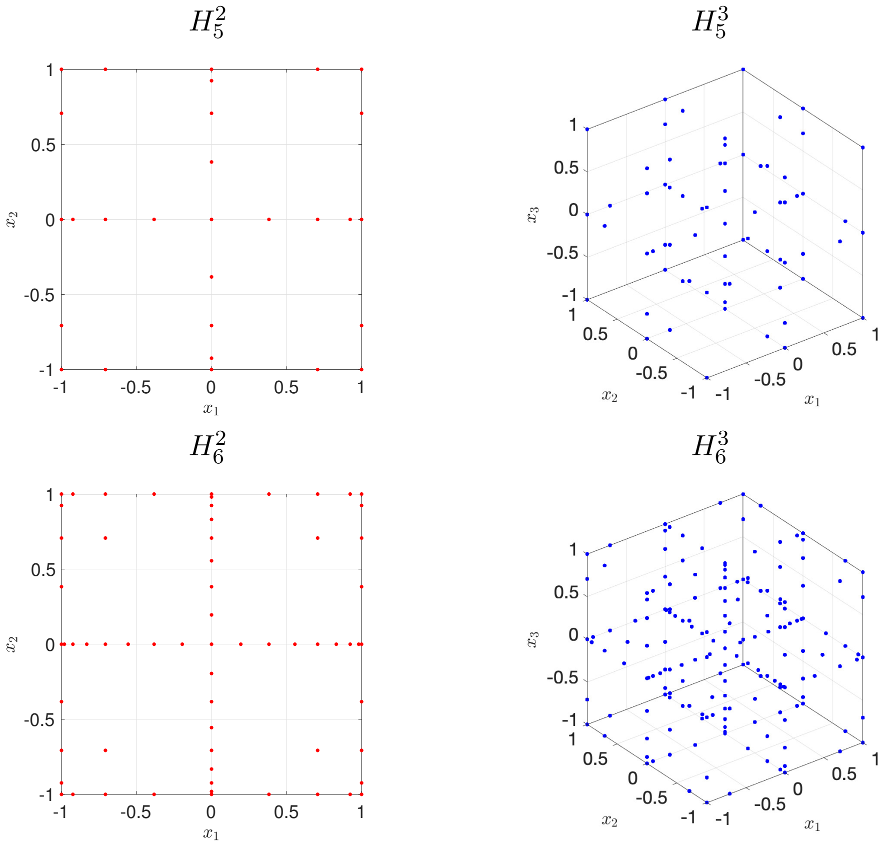


Figure 8: Chebyshev-Gauss-Lobatto (GCL) sparse grids of level 5 (first row) and 6 (second row) in dimension 2 and 3.

where $i_j \geq 1$, and $q \geq d$ is a parameter called sparse grids level. To clarify the meaning of (76), let us compute the two-dimensional Smolyak interpolant of level 3 of a two-dimensional function $f(x_1, x_2)$. By definition,

$$\begin{aligned} S_3^2(f) = &\sum_{i_1+i_2\leq 3} \Delta_1^{i_1} \otimes \Delta_2^{i_2} \\ =&\Delta_1^1 \otimes \Delta_2^1 + \Delta_1^2 \otimes \Delta_2^1 + \Delta_1^1 \otimes \Delta_2^2 \\ =&\Pi_1^1 \otimes \Pi_2^1 + \left(\Pi_1^2 - \Pi_1^1\right) \otimes \Pi_2^1 + \Pi_1^1 \otimes \left(\Pi_2^2 - \Pi_2^1\right) \\ =&\Pi_1^2 \otimes \Pi_2^1 + \Pi_1^1 \otimes \Pi_2^2 - \Pi_1^1 \otimes \Pi_2^1. \end{aligned} \tag{77}$$

The interpolant (77) is built upon the sparse grid H_3^2 as shown in Figure 9. Note that each point is accounted for only once in the final grid H_3^2 (the origin is summed up twice and subtracted once). Specifically, we have

$$\begin{aligned} S_3^2(f) = &\left[f(-1,0)l_1^{(1)}(x_1) + f(0,0)l_2^{(1)}(x_1) + f(1,0)l_3^{(1)}(x_1)\right] l_1^{(2)}(x_2) + \\ &\left[f(0,-1)l_1^{(1)}(x_2) + f(0,0)l_2^{(1)}(x_2) + f(0,1)l_3^{(1)}(x_2)\right] l_1^{(2)}(x_1) - \\ &f(0,0)l_1^{(2)}(x_1)l_1^{(2)}(x_2), \end{aligned} \tag{78}$$

where $l_j^{(i)}$ are the Lagrange polynomials shown in Figure 9. By substituting (75) into (76) we can rewrite the Smolyak interpolant in terms of elementary one-dimensional interpolants as

$$S_q^d(f) = \sum_{q+1-d\leq i_1+\cdots+i_d\leq q} (-1)^{q-i_1-\cdots-i_d} \binom{d-1}{q-i_1-\cdots-i_d} \Pi_1^{i_1} \otimes \cdots \otimes \Pi_d^{i_d}. \tag{79}$$

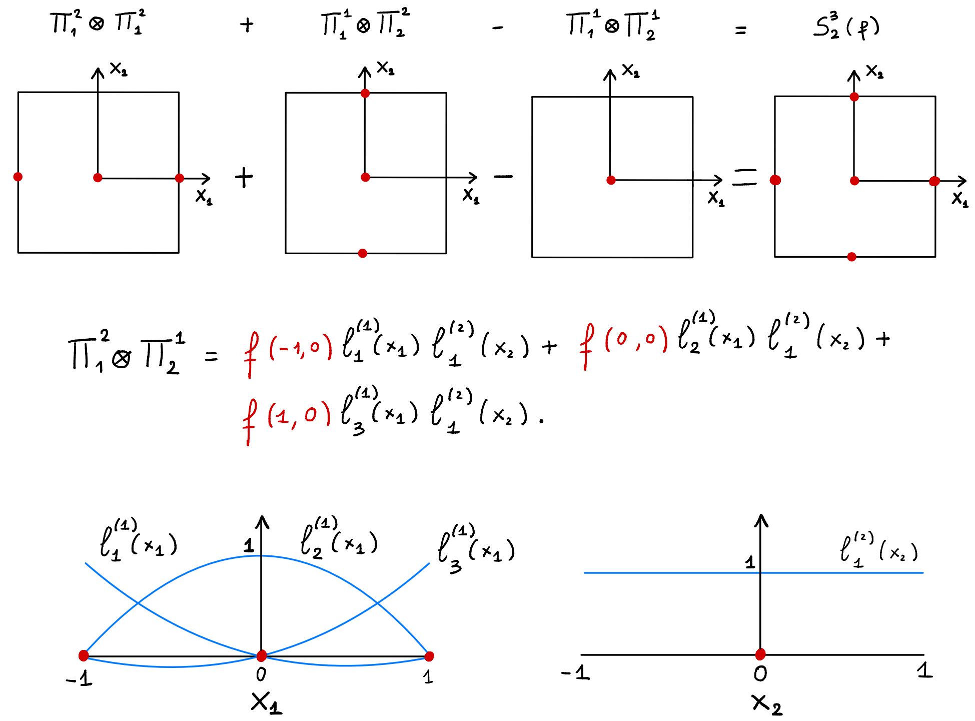

$\Pi_1^2 \otimes \Pi_1^2$ + $\Pi_1^1 \otimes \Pi_2^2$ - $\Pi_1^1 \otimes \Pi_2^1$ = $S_2^3(f)$
x_2
x_1
+
-
=
$\Pi_1^2 \otimes \Pi_2^1 = f(-1,0)\, \ell_1^{(1)}(x_1)\, \ell_1^{(2)}(x_2) + f(0,0)\, \ell_2^{(1)}(x_1)\, \ell_1^{(2)}(x_2) + f(1,0)\, \ell_3^{(1)}(x_1)\, \ell_1^{(2)}(x_2).$
$\ell_1^{(1)}(x_1)$
1
$\ell_2^{(1)}(x_1)$
$\ell_3^{(1)}(x_1)$
-1
0
1
x_1
1
$\ell_1^{(2)}(x_2)$
-1
0
1
x_2

Figure 9: Construction of the Smolyak interpolant (77) and corresponding grids. Similar expression can be derived for $\Pi_1^1 \otimes \Pi_2^2$ and $\Pi_1^1 \otimes \Pi_2^1$.

Hereafter we summarize an error estimate obtained in [1, Remark 11].

Theorem 3. Let $H_\mu^s([-1,1]^d)$ be the weighted Sobolev space of order s, with weight[7]

$$\mu(x_1, \dots, x_d) = \left[(1-x_1^2)\cdots(1-x_d^2)\right]^{-1/2}. \tag{80}$$

Then,

$$\left\| f - S_q^d(f) \right\|_{L^2_\mu([-1,1]^d)} \leq C(s,d) n^{-s} \log(n)^{(s+1)(d-1)} \|f\|_{H^s_\mu([-1,1]^d)}, \tag{81}$$

where $C(s,d)$ is a constant that depends on s and d, and $n = n$ is the total number of sparse grids points (which depends on d and q).

As easily seen, convergence is no longer spectral (unless $d = 1$) because of the factor $\log(n)^{(s+1)(d-1)}$.

Integration on sparse grids. The Smolyak algorithm can be used to construct cubature formulas to integrate high-dimensional functions. The key idea is very simple: replace the function with the Smolyak interpolant on a sparse grid and then integrate. Assuming that the integration weight is separable as in Theorem 3 we

$$\int_{[-1,1]^d} f(\boldsymbol{x}) \prod_{j=1}^d \mu_j(x_j) d\boldsymbol{x} \simeq \underbrace{\int_{[-1,1]^d} S_q^d(f)(\boldsymbol{x}) \prod_{j=1}^d \mu_j(x_j) d\boldsymbol{x}}_{I_q^d(f)}. \tag{82}$$

[7] Note that the weight (80) corresponds to a tensor product of Chebyshev polynomials

This yields an interpolatory quadrature rule with high degree of exactness [8]. In particular, by substituting (79) into (82) we obtain

$$\begin{aligned} I_q^d(f) &= \int_{[-1,1]^d} S_q^d(f)(\boldsymbol{x}) \prod_{j=1}^{d} \mu_j(x_j) d\boldsymbol{x} \\ &= \sum_{q+1-d\leq i_1+\cdots+i_d\leq q} (-1)^{q-i_1-\cdots-i_d} \binom{d-1}{q-i_1-\cdots-i_d} U_1^{i_1}\otimes\cdots\otimes U_d^{i_d}, \end{aligned} \tag{83}$$

where

$$U_j^{i_j} = \int_{-1}^{1} \Pi_j^{i_j} \mu_j(x_j) dx_j. \tag{84}$$

Example: To illustrate (83), let us integrate the two-dimensional interpolant $S_3^2(f)$ defined in (78) on $[-1,1]^2$. This yields the interpolatory quadrature formula

$$\begin{aligned} I_3^2(f) =& f(-1,0)w_{11}^1 w_{12}^2 + f(0,0)\left[w_{21}^1 w_{12}^2 + w_{22}^1 w_{11}^2 - w_{11}^1 w_{12}^1\right] + \\ & f(0,-1)w_{12}^1 w_{11}^2 + f(1,0)w_{31}^1 w_{12}^2 + f(0,1)w_{32}^1 w_{11}^2, \end{aligned} \tag{85}$$

where

$$w_{ip}^j = \int_{-1}^{1} l_i^{(j)}(x_p)\mu(x_p)dx_p. \tag{86}$$

Note that the integration weights in sparse grids are not necessarily all positive. For example, the weight multiplying $f(0,0)$, i.e., $w_{21}^1 w_{12}^2 + w_{22}^1 w_{11}^2 - w_{11}^1 w_{12}^1$, could be negative. Regarding the degree of exactness, for CGL sparse grids we have the following result (see [8, Corollary 3])

Theorem 4. Let $q = \sigma d + \tau$, $\sigma \in \mathbb{N}$, $\tau \in \{0,\ldots,d-1\}$ Then $I_q^d(f)$ has degree of exactness

$$\begin{cases} 2(q-d)+1 & \text{if} \quad q < 4d \\ 2^{\sigma-2}(d+1+\tau)+2d-1 & \text{if} \quad q \geq 4d \end{cases} \tag{87}$$

Other convergence estimates for interpolatory quadrature rules on sparse grids can be derived based on convergence estimates of one-dimensional quadrature (see [1, Remark 11]).

Appendix A: Chebyshev-Gauss-Lobatto quadrature

Let briefly review the main ingredients of the Gauss-Lobatto Chebyshev expansion. For more details we refer to [6]. We first recall that the Chebyshev polynomials of the first kind are defined as[8]

$$T_k(x) = \cos(k \arccos(x)) \qquad x \in [-1, 1] \qquad \text{(trigonometric representation)}. \tag{91}$$

It can be shown that $T_k(x)$ (like any other orthogonal polynomial) satisfy the three-term recurrence relation

$$\begin{aligned} T_0(x) &= 1 \\ T_1(x) &= x \\ T_{n+1}(x) &= 2x\, T_n(x) - T_{n-1}(x). \end{aligned} \tag{92}$$

and the orthogonality conditions

$$\int_{-1}^{1} T_k(x) T_j(x) \underbrace{\frac{1}{\sqrt{1-x^2}}}_{\mu(x)} dx = \delta_{kj} \left\| T_k \right\|^2_{L^2_\mu}. \tag{93}$$

Note that the first polynomials which gives

$$T_2(x) = 2x^2 - 1, \qquad T_3(x) = 4x^3 - 3x \qquad T_4(x) = 8x^4 - 8x^2 + 1, \ldots. \tag{94}$$

The Chebyshev-Gauss-Lobatto nodes are zeros of the polynomial

$$Q_{M+1}(x) = (1 - x^2) \frac{dT_M(x)}{dx}, \tag{95}$$

i.e., $x_0 = -1$, $x_M = 1$ and all maxima and minima of $T_M(x)$. By differentiating (91) with respect to x we obtain

$$\frac{dT_M(x)}{dx} = \frac{\sin(M \arccos(x))}{\sqrt{1-x^2}}. \tag{96}$$

Hence $Q_{M+1}(x) = 0$ implies that

$$x_j = -\cos\left(\frac{k\pi}{M}\right) \qquad j = 0, \ldots, M \qquad \text{(Chebyshev-Gauss-Lobatto points)}. \tag{97}$$

These points are obtained by dividing half unit circle in evenly-spaced parts and projecting them onto the x-axis. Note also that Chebyshev grid points are *nested* for $M = 2, 4, 8, \ldots, 2^s$.

It can be shown that the Lagrange characteristic polynomials associated with the Gauss-Chebyshev-Lobatto nodes are

$$l_j(x) = \frac{(-1)^{M+j+1}(1-x^2)}{d_j M^2 (x - x_j)} \frac{dT_M(x)}{dx} = \frac{(-1)^{M+j+1}\sqrt{(1-x^2)}}{d_j M^2 (x - x_j)} \sin(M \arccos(x)), \tag{98}$$

where x_j is given in (97) and

$$d_0 = d_M = 2 \qquad d_1 = d_2 = \cdots = d_{M-1} = 1. \tag{99}$$

[8] Note that (91) are indeed polynomials. For example,

$$\begin{aligned} \cos(\arccos(x)) =& x, & (88) \\ \cos(2\arccos(x)) =& 2\left(\cos(\arccos(x))\right)^2 - 1 = 2x^2 - 1, & (89) \\ \cos(3\arccos(x)) =& 4\left(\cos(\arccos(x))\right)^3 - 3\cos(\arccos(x)) = 4x^3 - 3x. & (90) \end{aligned}$$

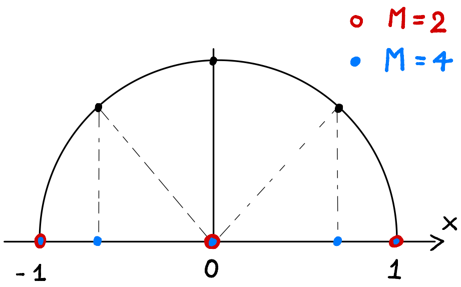


Figure 10: Nested property of Chebyshev grids for $M = 2^s$ $(s = 1, 2, 3, \ldots)$.

For any function $f(x)$ defined in $[-1, 1]$ we have the following Lagrangian interpolant

$$\Pi_M f(x) = \sum_{k=0}^{M} f(x_k) l_k(x), \qquad x \in [-1, 1]. \tag{100}$$

At this point we integrate (100) to obtain the quadrature formula

$$\int_{-1}^{1} f(x) \frac{1}{\sqrt{1 - x^2}} dx \simeq \sum_{k=0}^{M} f(x_k) w_k, \tag{101}$$

where

$$w_k = \int_{-1}^{1} \frac{l_k(x)}{\sqrt{1 - x^2}} dx = \frac{\pi}{M d_j} \tag{102}$$

and d_j is defined in (99).

Appendix B: Lagrangian interpolation at Gauss points

The quadrature rule (39) induces a discrete inner product that establishes a correspondence between series expansions in terms of orthogonal polynomials[9] and Lagrangian interpolation formulas. To show this, let

$$f(z) \simeq \sum_{k=0}^{M} a_k P_k(z) \qquad a_k = \frac{(f, P_k)_{L^2_\mu([-1,1])}}{(P_k, P_k)_{L^2_\mu([-1,1])}} \tag{103}$$

be a polynomial expansion of $f(z)$ in $[-1,1]$, where $\{P_0, \ldots, P_M\}$ is a set of polynomials orthogonal relative to the weight function $\mu(z)$. Consider the following Gauss approximation the inner product

$$\begin{aligned}(f, P_k)_{L^2_\mu([-1,1])} &= \int_{-1}^{1} f(z) P_k(z) \mu(z) dz \\ &\simeq \sum_{j=0}^{M} f(z_j) P_k(z_j) w_j, \qquad \text{(discrete inner product)}\end{aligned} \tag{104}$$

where

$$\{z_0, \ldots, z_M\} \quad \text{and} \quad \{w_0, \ldots, w_M\} \tag{105}$$

are $M+1$ Gauss quadrature points and quadrature weights, respectively. Recall that the Gauss rule (104) has degree of exactness $2M+1$ and therefore it can be used to compute

$$\gamma_k = (P_k, P_k)_{L^2_\mu} \tag{106}$$

exactly up to $k = M$. A substitution of (104) into (103) yields

$$f(z) \simeq \sum_{j=0}^{M} f(z_j) \underbrace{\sum_{k=0}^{M} \frac{w_j}{\gamma_k} P_k(z_j) P_k(z)}_{l_j(z)}. \tag{107}$$

In this form, we recognize that the Lagrangian interpolation formula, where

$$l_j(z) = \sum_{k=0}^{M} \frac{w_j}{\gamma_k} P_k(z_j) P_k(z) \tag{108}$$

are Lagrange characteristic polynomials associated with the Gauss nodes (105).

The identification of the approximation (103) with the Lagrangian interpolant (107) at Gauss nodes (105) suggests a mathematically equivalent but computationally different way of representing the function $f(z)$. Regarding the approximation error in (103), the following general estimate in terms of the uniform norm holds true.

Theorem 5. Let $f \in C^0([-1,1])$ and $\Pi_M f(z)$ the polynomial of degree M interpolating $f(z)$ at $\{z_0, \ldots, z_M\}$. Then

$$\|f(z) - \Pi_M f(z)\|_\infty \leq (1 + \Lambda_M) \inf_{\Psi \in \mathbb{P}_M} \|f(z) - \Psi(z)\|_\infty \tag{109}$$

where

$$\Lambda_M = \max_{z \in [-1,1]} \lambda_M(z) \qquad \text{(Lebesgue constant)}, \tag{110}$$

$$\lambda_M(z) = \sum_{j=0}^{M} |l_j(z)| \qquad \text{(Lebesgue function)}. \tag{111}$$

[9] We have seen in Chapter 4 that orthogonal polynomial expansions exhibit spectral convergence.

Proof. The proof for the upper bound (109) is very simple. Let $\Psi \in \mathbb{P}_M$ be the best approximating polynomial

$$\left\|f(z) - \Pi_M f(z)\right\|_\infty \leq \left\|f(z) - \Psi(z)\right\|_\infty + \left\|\Psi(z) - \Pi_M f(z)\right\|_\infty . \tag{112}$$

At this point, we represent $\Psi(z)$ and $\Pi_f(z)$ in terms of the same set of Lagrange polynomials associated with the grid $\{z_0, \ldots, z_M\}$ to obtain

$$\begin{aligned}\left\|\Psi(z) - \Pi_M f(z)\right\|_\infty &= \left\| \sum_{j=0}^{M} \left[\Psi(z_j) - f(z_j)\right] l_j(z) \right\|_\infty \\ &\leq \left\|\Psi(z) - f(z)\right\|_\infty \underbrace{\max_{z\in[-1,1]} \sum_{j=0}^{M} |l_j(z)|}_{\Lambda_M} .\end{aligned} \tag{113}$$

A substitution of (113) into (112) yields (109).

□

Note that the Lebesgue constant depends only on the set of grid points. Clearly, the smaller the Lebesgue constant, the smaller the interpolation error in the uniform norm. It can be shown that, no matter how we choose the points, the Lebesgue constant grows at least logarithmically with M, i.e. (see [6, p.102]),

$$\Lambda_M \geq \frac{2}{\pi} \log(1+M) + C \quad \text{as} \quad M \to \infty. \tag{114}$$

Note that this does not mean that the interpolation error necessarily grows with M. It just means that the upper bound in (109) diverges as $M \to \infty$, i.e., that we cannot grant uniform convergence of Lagrangian interpolation using (109). In other words, for any given set of grid points there exist continuous functions for which the polynomial interpolant will exhibit non-uniform convergence. On the other hand, one can also show that for any given continuous function one can always construct a set of grid points that will result in a uniformly convergent polynomial representation.

It is possible to bound the Lebesgue constant corresponding to various types grids. For instance, for evenly-spaced grids of $M+1$ points in $[-1,1]$ we have

$$\frac{2^{M-2}}{M^2} \leq \Lambda_M \leq \frac{2^{M+3}}{M}. \tag{115}$$

Similarly, for the Gauss-Chebyshev-Lobatto (GCL) grid (97) we have (e.g., [6, p. 105])

$$\Lambda_M \leq \frac{2}{\pi} \log(M) + B \qquad \text{(finite } M\text{)}, \tag{116}$$

where B is a suitable constant independent of M.

Example: In Figure 11 we plot the Lagrangian interpolant of

$$f(z) = \frac{1}{1+10z^2} \tag{117}$$

computed at 17 evenly-spaced nodes or 17 Gauss-Chebyshev-Lobatto (GCL) nodes ($M = 16$). In the same Figure we plot the Lebesgue functions of both interpolation problems. The Lebesgue constants for the evenly-spaced grid and the GCL grid are obtained, respectively, as

$$\Lambda_M^{eq} = 934.532 \qquad \Lambda_M^{GCL} = 2.468. \tag{118}$$

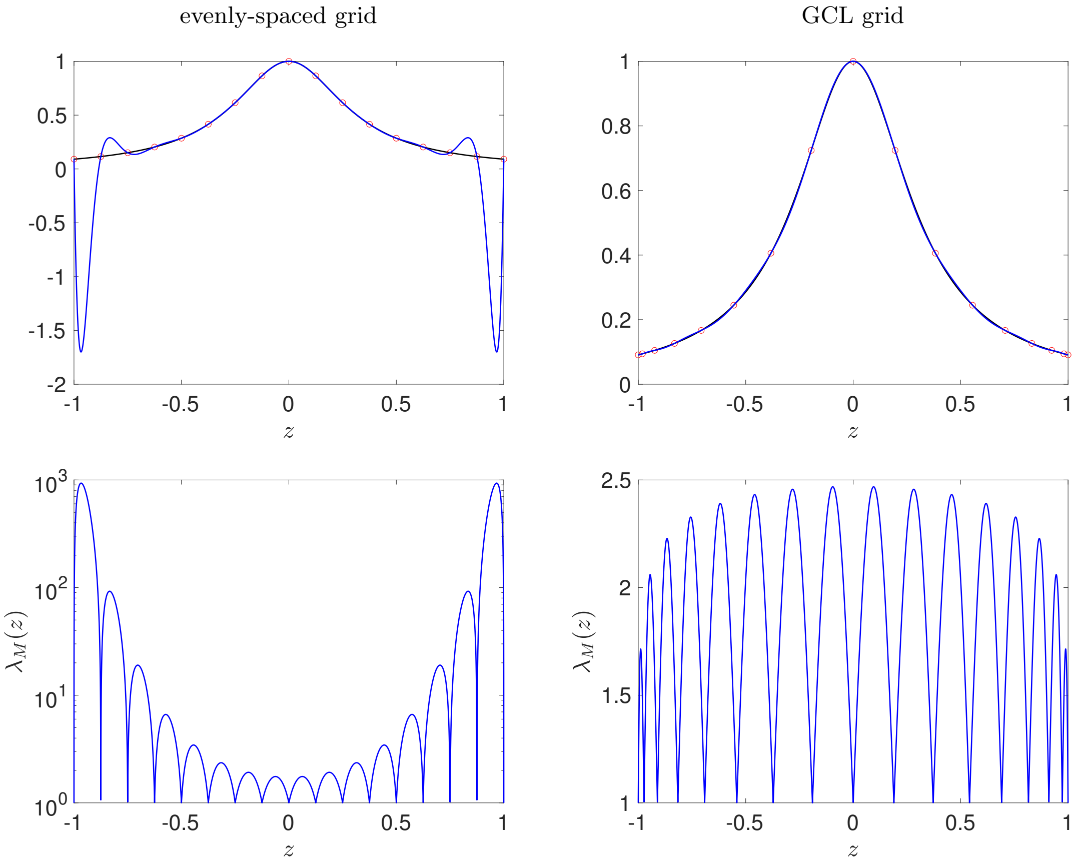


Figure 11: Lagrangian interpolation of $f(z) = (1 + 10z^2)^{-1}$ using 17 evenly-spaced nodes (left), and 17 GCL nodes (right). The Lebesgue functions $\lambda_M(z)$ associated with the evenly-spaced grid and the GCL grid have maxima $\Lambda_M^{eq} = 934.532$ and $\Lambda_M^{GCL} = 2.468$, respectively.

If we measure the interpolation error in terms of the $L^2_\mu([-1,1])$ (instead of the uniform norm we used in Theorem (5)) then by leveraging the correspondence between orthogonal polynomial expansions and the Lagrangian interpolant in Eq. (107) is possible to obtain spectral convergence results. For example, the following convergence result holds for Gauss-Legendre and Gauss-Legendre-Lobatto interpolation (see Table 1 and [6, p. 114]).

Theorem 6. Let $f(z) \in H^s([-1,1])$, $p \geq 1$. Then

$$\left\| f(z) - \sum_{k=0}^{M} f(z_k) l_k(z) \right\|_{L^2([-1,1])} \leq C M^{-s} \left\| f(z) \right\|_{H^s([-1,1])} \tag{119}$$

where $\{z_0, \ldots, z_M\}$ are either Gauss-Legendre points or Gauss-Legendre-Lobatto points (see Table 1).

References

[1] V. Barthelmann, E. Novak, and K. Ritter. High dimensional polynomial interpolation on sparse grids. *Advances in Computational Mechanics*, 12:273–288, 2000.

[2] Z. I. Botev, J. F. Grotowski, and D. P. Kroese. Kernel density estimation via diffusion. *Annals of Statistics*, 38(5):2916–2957, 2010.

[3] H. J. Bungartz and M. Griebel. Sparse grids. *Acta Numerica*, 13:147–269, 2004.

[4] J. Dick, F. Y. Kuo, and I. H. Sloan. High-dimensional integration: the quasi-Monte Carlo way. *Acta Numer.*, 22:133–288, 2013.

[5] J. Foo and G. E. Karniadakis. Multi-element probabilistic collocation method in high dimensions. *J. Comput. Phys.*, 229:1536–1557, 2010.

[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] C. Lemieux. *Monte Carlo and Quasi-Monte Carlo Sampling*. Springer, 2009.

[8] E. Novak and K. Ritter. Simple cubature formulas with high polynomial exactness. *Constr. Approx.*, 15:499–522, 1999.

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

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

Tensor methods for deterministic and stochastic PDEs

We have seen how to propagate uncertainty in PDE models involving random initial conditions, random parameters, random boundary conditions or random forcing terms. Specifically, we disussed PDF methods (Hopf equation and Lundgren-Monin-Novikov hierarchies), polynomial chaos methods (gPC, ME-gPC), and sampling methods (MC, qMC, PCM, ME-PCM, and sparse grids). In this lecture note we discuss another method that relies upon orthogonal tensor expansions to compute the solution of stochastic PDEs. The same theory can be used to compute the numerical solution of high-dimensional PDEs such as the Liouville equation or the Fokker-Plank equation.

Dynamically orthogonal (DO) tensor methods for stochastic PDEs

The dynamically orthogonal field equation method for SPDEs was pioneered by Sapsis and Lermusiaux in [14], and it is essentially a tensor method for linear or nonlinear PDEs in a separable Hilbert space [7, 6]. To describe DO, suppose we are interested in computing the solution to an initial/boundary value problem for a stochastic PDE of the form

$$\begin{cases} \dfrac{\partial u(\boldsymbol{x},t;\omega)}{\partial t} = G_\omega(u(\boldsymbol{x},t;\omega)), \\ u(\boldsymbol{x},0;\omega) = u_0(\boldsymbol{x};\omega), \end{cases} \tag{1}$$

where $\boldsymbol{x} \in V \subseteq \mathbb{R}^d$ (V is the spatial domain $d \geq 1$), and G_ω is a *random nonlinear operator* which may take into account random forcing terms, random parameters or random boundary conditions. A simple example of $G_\omega(u(\boldsymbol{x},t;\omega))$ could be[1]

$$G_\omega\left(u(\boldsymbol{x},t;\omega) = \nabla \cdot \left[\kappa(\boldsymbol{x};\omega)\nabla u(\boldsymbol{x},t;\omega)\right], \qquad \kappa(\boldsymbol{x};\omega) > 0, \tag{2}$$

in $\mathbb{R}^d$. We look for a representation of the solution to (1) of the form

$$u(\boldsymbol{x},t;\omega) = \mathbb{E}\{u(\boldsymbol{x},t;\omega)\} + \sum_{k=1}^{\infty} \widehat{u}_k(\boldsymbol{x},t)Y_k(t;\omega), \tag{3}$$

where $\{\widehat{u}_1(\boldsymbol{x},t), \widehat{u}_2(\boldsymbol{x},t), \cdots\}$ are deterministic spatio-temporal modes, while $\{Y_1(t;\omega), Y_2(t;\omega), \ldots\}$ are random temporal modes. Note the time redundancy in both the space-time modes $\widehat{u}_k(\boldsymbol{x},t)$ and the random modes $Y_k(t;\omega)$. The theoretical justification of the series expansion (3) relies on a tensor product representation of the Hilbert space $L^2(V \times T \times \Omega)$ (T is the temporal domain and Ω is the sample space) as

$$L^2(V \times T \times \Omega) = L^2(V \times T) \otimes L^2(T \times \Omega). \tag{4}$$

The expansion (3) includes time-dependent gPC [9] as a sub-case.

Properties of the modes $\widehat{u}_k(\boldsymbol{x},t)$ and $Y_k(t;\omega)$. The random temporal modes $Y_k(t;\omega)$ are clearly zero mean. In fact, by applying the expectation operator to (3) we obtain

$$\sum_{k=1}^{\infty} \widehat{u}_k(\boldsymbol{x},t)\mathbb{E}\{Y_k(t;\omega)\} = 0 \quad \Rightarrow \quad \mathbb{E}\{Y_k(t;\omega)\} = 0. \tag{5}$$

[1] The PDE (1)-(2) describes heat conduction in a heterogeneous medium with random thermal conductivity $\kappa(\boldsymbol{x};\omega)$.

We also assume that the space-time modes $\widehat{u}_k(\boldsymbol{x},t)$ satisfy the *gauge*[2] conditions

$$\left\langle \widehat{u}_k(\boldsymbol{x},t), \frac{\partial \widehat{u}_j(\boldsymbol{x},t)}{\partial t}\right\rangle_{L^2(V)} = \int_V \widehat{u}_k(\boldsymbol{x},t)\frac{\partial \widehat{u}_j(\boldsymbol{x},t)}{\partial t} d\boldsymbol{x} = 0 \qquad \text{for all } t \geq 0 \qquad \text{and all } j,k \geq 1. \tag{6}$$

These conditions are called *dynamically orthogonal (DO) conditions.* The reason is that if the set of modes $\{\widehat{u}_k(\boldsymbol{x},t)\}$ is initially orthonormal, i.e.,

$$\langle \widehat{u}_k(\boldsymbol{x},0), \widehat{u}_j(\boldsymbol{x},0)\rangle_{L^2(V)} = \delta_{kj}. \tag{7}$$

then it stays orthonormal in time. In fact, for all $t \geq 0$ we have

$$\frac{\partial}{\partial t}\langle \widehat{u}_k(\boldsymbol{x},t), \widehat{u}_j(\boldsymbol{x},t)\rangle_{L^2(V)} = \left\langle \frac{\partial \widehat{u}_k(\boldsymbol{x},t)}{\partial t}, \widehat{u}_j(\boldsymbol{x},t)\right\rangle_{L^2(V)} + \left\langle \widehat{u}_k(\boldsymbol{x},t), \frac{\partial \widehat{u}_j(\boldsymbol{x},t)}{\partial t}\right\rangle_{L^2(V)} = 0 \quad \text{for all } i,j \geq 1. \tag{8}$$

This implies that

$$\langle \widehat{u}_k(\boldsymbol{x},t), \widehat{u}_j(\boldsymbol{x},t)\rangle_{L^2(V)} = \langle \widehat{u}_k(\boldsymbol{x},0), \widehat{u}_j(\boldsymbol{x},0)\rangle_{L^2(V)} = \delta_{kj}, \tag{9}$$

i.e., space time modes $\widehat{u}_k(\boldsymbol{x},t)$ that are orthogonal at $t = 0$ remain orthogonal at later times. For this reason we shall call $\widehat{u}_k(\boldsymbol{x},t)$ *dynamically orthogonal* modes.

DO propagator. At this point we have all elements to derive a coupled system of equations for the DO modes $\widehat{u}_j(\boldsymbol{x},t)$, the stochastic modes $Y_t(t;\omega)$ and the mean field

$$\overline{u}(\boldsymbol{x},t) = \mathbb{E}\{u(\boldsymbol{x},t;\omega\} \tag{10}$$

appearing in (3). To this end, we first substitute a truncated expansion of the form (3), i.e.,

$$u_M(\boldsymbol{x},t;\omega) = \overline{u}(\boldsymbol{x},t) + \sum_{k=1}^{M} \widehat{u}_k(\boldsymbol{x},t) Y_k(t;\omega), \tag{11}$$

into the SPDE (1) to obtain

$$\frac{\partial \overline{u}(\boldsymbol{x},t)}{\partial t} + \sum_{k=1}^{M}\left(\frac{\partial \widehat{u}_k(\boldsymbol{x},t)}{\partial t} Y_k(t;\omega) + \widehat{u}_k(\boldsymbol{x},t)\frac{dY_k(t;\omega)}{dt}\right) = G_\omega\left(u_M(\boldsymbol{x},t;\omega)\right) + R_M(\boldsymbol{x},t;\omega). \tag{12}$$

Then we impose that the residual $R_M(\boldsymbol{x},t;\omega)$ is orthogonal to

$$S_M = \text{span}\left\{\widehat{u}_1(\boldsymbol{x},t),\dots,\widehat{u}_M(\boldsymbol{x},t)\right\} \quad \text{and} \quad Z_M = \text{span}\left\{Y_1(t;\omega),\dots,Y_M(t;\omega)\right\} \tag{13}$$

relative to the inner products $\langle\cdot\rangle_{L^2(V)}$ (see Eq. (6)) and $\mathbb{E}\{\cdot\}$. This gives the $2M + 1$ conditions

$$0 = \mathbb{E}\left\{R_M(\boldsymbol{x},t;\omega)\right\}, \tag{14}$$

$$0 = \mathbb{E}\left\{R_M(\boldsymbol{x},t;\omega)Y_k(t;\omega)\right\} \qquad k = 1,\dots,M, \tag{15}$$

$$0 = \langle R_M(\boldsymbol{x},t;\omega)\widehat{u}_k(\boldsymbol{x},t)\rangle_{L^2(V)} \qquad k = 1,\dots,M \tag{16}$$

which are sufficient to identify a set of equation for the mean field $\boldsymbol{u}(\boldsymbol{x},t)$, the DO modes $\{\widehat{u}_k(\boldsymbol{x},t)\}$, and the stochastic modes $\{Y_k(t;\omega)\}$. By taking the expectation of (12) and taking into account (14) we obtain

$$\frac{\partial \overline{u}}{\partial t} = \mathbb{E}\left\{G_\omega\left(u_M\right)\right\} \quad \text{(evolution equation for the mean field).} \tag{17}$$

[2] In physics, choosing a gauge denotes a mathematical procedure for coping with redundant degrees of freedom in field variables. In the case of the series expansion (3), t is the redundant degree of freedom. We also emphasize that the inner product (6) can be generalized to include, e.g., a weight function $\mu(\boldsymbol{x})$ (weighted $L^2_\mu(V)$ space), or spatial derivatives of $\widehat{u}_k(\boldsymbol{x},t)$ (Sobolev space $H^s(V)$).

Next, we project (12) onto $\widehat{u}_p(\boldsymbol{x},t)$ and take (16) into account to obtain

$$\left\langle \frac{\partial \overline{u}}{\partial t}, \widehat{u}_p \right\rangle_{L^2(V)} + \sum_{k=1}^{M} \underbrace{\left\langle \frac{\partial \widehat{u}_k}{\partial t}, \widehat{u}_p \right\rangle_{L^2(V)}}_{=0} Y_k + \sum_{k=1}^{M} \underbrace{\langle \widehat{u}_k \widehat{u}_p \rangle_{L^2(V)}}_{=\delta_{kp}} \frac{dY_k}{dt} = \langle G_\omega \left(u_M\right) \widehat{u}_p \rangle_{L^2(V)}\,, \tag{18}$$

where we assumed that the DO modes $\{\widehat{u}_k(\boldsymbol{x},t)\}$ are orthonormal at $t=0$ and therefore at all t (see Eq. (9)). Equation (18) can be written as

$$\frac{dY_p}{dt} = \langle [G_\omega \left(u_M\right) - \mathbb{E}\{G_\omega \left(u_M\right)\}]\, \widehat{u}_p \rangle_{L^2(V)}\,. \tag{19}$$

Finally we project (12) onto $Y_p(t;\omega)$ and take (16) into account to obtain

$$\underbrace{\mathbb{E}\left\{ \frac{\partial \overline{u}}{\partial t} Y_p \right\}}_{=0} + \sum_{k=1}^{M} \frac{\partial \widehat{u}_k}{\partial t} \underbrace{\mathbb{E}\left\{Y_k Y_p\right\}}_{\Sigma_{kp}(t)} + \sum_{k=1}^{M} \widehat{u}_k \mathbb{E}\left\{ \frac{dY_k}{dt} Y_p \right\} = \mathbb{E}\left\{G_\omega \left(u_M\right) Y_p\right\}. \tag{20}$$

Note that

$$\Sigma_{kp}(t) = \mathbb{E}\left\{Y_k(t;\omega) Y_p(t;\omega)\right\} \tag{21}$$

is the *covariance function* of the random process $Y_k(t;\omega)$ and $Y_p(t;\omega)$. By using (19) we can write the last terms at the right hand side of (20) as[3]

$$\mathbb{E}\left\{ \frac{dY_k}{dt} Y_p \right\} = \mathbb{E}\left\{ \langle G_\omega(u_M) \widehat{u}_k \rangle_{L^2(V)} Y_p \right\}. \tag{23}$$

A substitution of (23) into (20) yields

$$\sum_{k=1}^{M} \frac{\partial \widehat{u}_k}{\partial t} \Sigma_{kp}(t) = \mathbb{E}\left\{G_\omega \left(u_M\right) Y_p\right\} - \sum_{k=1}^{M} \langle \mathbb{E}\left\{G_\omega \left(u_M\right) Y_p\right\} \widehat{u}_k \rangle_{L^2(V)} \widehat{u}_k. \tag{24}$$

In summary, the *DO propagator* can be written as [14, 5] (for $p = 1, \ldots, M$)

$$\begin{cases} \dfrac{\partial \overline{u}}{\partial t} = \mathbb{E}\left\{G_\omega \left(u_M\right)\right\}, \\ \dfrac{dY_p}{dt} = \langle [G_\omega \left(u_M\right) - \mathbb{E}\{G_\omega \left(u_M\right)\}]\, \widehat{u}_p \rangle_{L^2(V)}\,, \\ \displaystyle\sum_{k=1}^{M} \frac{\partial \widehat{u}_k}{\partial t} \Sigma_{kp}(t) = \mathbb{E}\left\{G_\omega \left(u_M\right) Y_p\right\} - \sum_{k=1}^{M} \langle \mathbb{E}\left\{G_\omega \left(u_M\right) Y_p\right\} \widehat{u}_k \rangle_{L^2(V)} \widehat{u}_k. \end{cases} \tag{25}$$

The initial and boundary conditions for this PDE system are obtained by projection (see [14]). Clearly, the evolution equations for the DO modes $\widehat{u}_k$ in (25) have some issues if the covariance matrix Σ_{kp} of the stochastic modes is singular. This happens, for example when a random mode Y_k has zero energy, e.g., when we add a mode during integration to increase accuracy. In this case, the system (25) becomes algebraic-differential (covariance matrix singular). This requires special numerical techniques for temporal integration. One can overcome this problem by considering pseudo-inverse matrix operations [1]. More rigorously, it can be shown that is possible to rewrite the system (25) in fully equivalent form that does not require covariance matrix inversion, and solve such a system using *operator splitting* (see, e.g., [6]).

[3] Note that

$$\mathbb{E}\left\{\mathbb{E}\left\{G_\omega(u_M)\right\} Y_p\right\} = \mathbb{E}\left\{G_\omega(u_M)\right\} \mathbb{E}\left\{Y_p\right\} = 0. \tag{22}$$

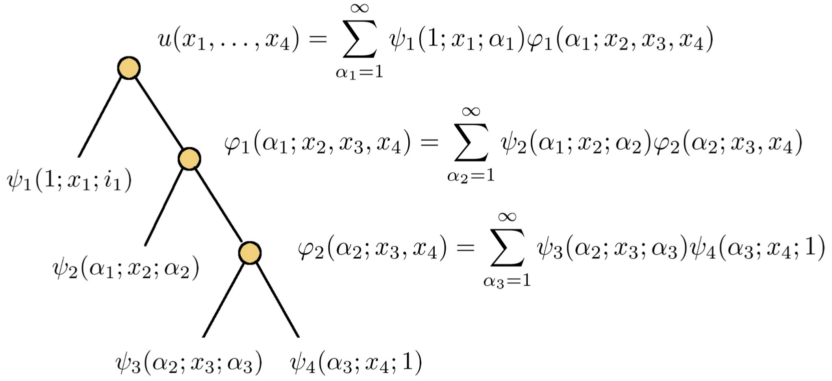

$u(x_1,\dots,x_4)=\sum_{\alpha_1=1}^{\infty}\psi_1(1;x_1;\alpha_1)\varphi_1(\alpha_1;x_2,x_3,x_4)$
$\varphi_1(\alpha_1;x_2,x_3,x_4)=\sum_{\alpha_2=1}^{\infty}\psi_2(\alpha_1;x_2;\alpha_2)\varphi_2(\alpha_2;x_3,x_4)$
$\psi_1(1;x_1;i_1)$
$\varphi_2(\alpha_2;x_3,x_4)=\sum_{\alpha_3=1}^{\infty}\psi_3(\alpha_2;x_3;\alpha_3)\psi_4(\alpha_3;x_4;1)$
$\psi_2(\alpha_1;x_2;\alpha_2)$
$\psi_3(\alpha_2;x_3;\alpha_3)$
$\psi_4(\alpha_3;x_4;1)$

Figure 1: Construction of functional tensor train (FTT). Shown is the sequence of hierarchical Schmidt decomposition for a function in four variables.

An advantage of (25) over, e.g., polynomial chaos is that the stochastic modes are evolving with time in a way that depends on the PDE. Moreover, it can be shown that the DO equations (25) satisfy an optimality (variational) principle similar to the one satisfied by the Karhunen-Loève expansion (see [7]), which implies that we can obtain accurate stochastic solutions of (1) using a series expansion (11) with a relatively small number of modes M.

In a parallel research effort, T. Hou and collaborators developed an alternative version of DO on bi-orthogonal (BO) expansions [3, 4]. Bi-orthogonality essentially represents a different *gauge* condition, which yields a propagator, i.e., a coupled system of equations for the modes $\widehat{u}_k$ and Y_k that differs from (25)(see [3, 4] for details). The correspondence between DO and BO was investigated in [5, 7].

Dynamically orthogonal tensor methods for high-dimensional deterministic PDE

In this section we generalize the series expansion (3) to compute the numerical solution of a high-dimensional deterministic PDE of the form

$$\frac{\partial u(\boldsymbol{x},t)}{\partial t}=G(u(\boldsymbol{x},t)),\qquad u(\boldsymbol{x},0)=u_0(\boldsymbol{x}),\tag{26}$$

where $u:V\times[0,T]\to\mathbb{R}$ is a (time-dependent) scalar field in d variables defined on the domain $V\subseteq\mathbb{R}^d$ and G is a nonlinear operator which may depend on the spatial variables, and may incorporate boundary conditions.

The PDE (26) may be a Liouville equation, a Fokker-Planck equation, or an approximation of the Hopf characteristic functional equation we have seen in Chapter 2.

Functional tensor train (FTT). Let $V\subseteq\mathbb{R}^d$ be a Cartesian product of d real intervals $V_i=[a_i,b_i]$

$$V=\bigtimes_{i=1}^{d}V_i,\tag{27}$$

μ a finite product measure on V

$$\mu(\boldsymbol{x})=\prod_{i=1}^{d}\mu_i(x_i),\tag{28}$$

and

$$H=L^2_\mu(V)\tag{29}$$

the standard weighted Hilbert space[4] of square–integrable functions on V. It was shown in [12, 2, 8] that any function $u(\boldsymbol{x}) \in H$ can be represented as

$$u(\boldsymbol{x}) = \sum_{\alpha_1,\ldots,\alpha_{d-1}=1}^{\infty} \psi_1(1;x_1;\alpha_1)\psi_2(\alpha_1;x_2;\alpha_2)\cdots\psi_d(\alpha_{d-1};x_d;1), \tag{30}$$

where $\psi_i(\alpha_{i-1};x_i;\alpha_i)$ are matrices of functions depending only on the variable x_i. Such functions are computed by solving a hierarchical sequence of eigenvalue problems that is similar to the Karhunen-Loève eigenvalue problem.

Computation of FTT. In Figure 1 we show the sequence of hierarchical (Schmidt) decompositions to compute the functional tensor train expansion for a four-dimensional function. The first step is to solve the eigenvalue problem

$$\lambda_1\psi_1(1;x_1;\alpha_1) = \int_{V_1} K(x_1,x_1')\psi_1(1;x_1';\alpha_1)dx_1', \tag{31}$$

where

$$K_1(x_1,x_1') = \int_{V_2\times V_3\times V_4} u(x_1,x_2,x_3,x_4)u(x_1',x_2,x_3,x_4)dx_2dx_3dx_4. \tag{32}$$

The (not-normalized) modes $\varphi_1(\alpha_1;x_2,x_3,x_4)$ are obtained by projection of u onto the orthonormal modes ψ_1 as

$$\varphi_1(\alpha_1;x_2,x_3,x_4) = \int_{V_1} u(x_1,x_2,x_3,x_4)\psi_1(1;x_1;\alpha_1)dx_1. \tag{33}$$

At this point we perform another Schmidt decomposition by solving the eigenvalue problem

$$\lambda_2\psi_2(\alpha_1;x_2;\alpha_2) = \int_{V_2} K_2(x_2,x_2';\alpha_1)\psi_2(\alpha_1;x_2';\alpha_2)dx_2', \tag{34}$$

where

$$K_2(x_2,x_2';\alpha_1) = \int_{V_3\times V_4} \varphi_1(\alpha_1;x_2,x_3,x_4)\varphi_1(\alpha_1;x_2',x_3,x_4)dx_3dx_4. \tag{35}$$

Note that the kernel K_2 is defined by the orthogonal modes φ_1 we obtained from the previous decomposition. We project $\varphi_1(\alpha_1;x_2,x_3,x_4)$ onto the orthonormal modes $\psi_2(\alpha_1;x_2';\alpha_2)$ to obtain

$$\varphi_2(\alpha_2;x_3,x_4) = \sum_{\alpha_1=1}^{\infty}\int_{V_2} \varphi_1(\alpha_1;x_2,x_3,x_4)\psi_2(\alpha_1;x_2;\alpha_2)dx_2. \tag{36}$$

Lastly we perform a decomposition the $\varphi_2(\alpha_2;x_3,x_4)$, which yields the modes $\psi_3(\alpha_2;x_3;\alpha_3)$ and $\psi_4(\alpha_3;x_4;1)$ (see Figure 1). The final expansion corresponds to the following sequence of function space decompositions

$$\begin{aligned} H(V_1\times V_2\times V_3\times V_4) =&[H(V_1)\otimes H(V_2\times V_3\times V_4)]\\ &[H(V_1)\otimes[H(V_2)\otimes H(V_3\times V_4)]]\\ &[H(V_1)\otimes[H(V_2)\otimes[H(V_3)\otimes H(V_4)]]], \end{aligned} \tag{37}$$

where the notation $[H(V_1)\otimes H(V_2\times V_3\times V_4)]$ emphasizes the fact that that we diagonalized the expansion involving the function spaces within the bracket.

In a finite-dimensional setting, such decomposition are essentially generated by a hierarchical sequence of singular value decompositions corresponding to various flattening of a multi-dimensional array (see Figure

[4] Note that the Hilbert space H in equation (29) can be equivalently chosen to be a Sobolev space $W^{2,p}$ (see [7] for details).

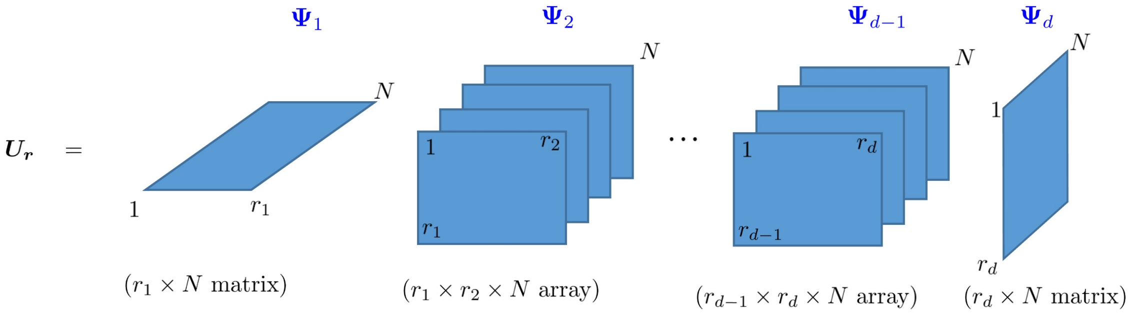


Figure 2: Construction of functional tensor train (FTT). Shown is the sequence of hierarchical Schmidt decomposition for a function in four variables.

2). By truncating the expansion (30) so that the terms corresponding to the largest eigenvalues in (31), (34), etc., are retained yields

$$u_{\boldsymbol{r}}(\boldsymbol{x}) = \sum_{\alpha_1,\ldots,\alpha_{d-1}=1}^{\boldsymbol{r}} \psi_1(1;x_1;\alpha_1)\psi_2(\alpha_1;x_2;\alpha_2)\cdots\psi_d(\alpha_{d-1};x_d;1), \tag{38}$$

where $\boldsymbol{r} = (1, r_1, \ldots, r_{d-1}, 1)$ is the FTT-rank.

It was shown by Bigoni *et al.* [2] that the truncated FTT expansion expansion (38) converges optimally (in $\mathbf{r}$) with respect to the $L^2_\mu(V)$ norm. More precisely, for any given function $u \in L^2_\mu(V)$ the FTT approximant (38) minimizes the residual $\|u - u_{\boldsymbol{r}}\|_{L^2_\mu(V)}$ relative to independent variations of the functions $\psi_i(\alpha_{i-1};x_i;\alpha_i)$ on a tensor manifold with constant rank $\boldsymbol{r}$. It is convenient to write (38) in a more compact form as

$$u_{\boldsymbol{r}}(\boldsymbol{x}) = \boldsymbol{\Psi}_1(x_1)\boldsymbol{\Psi}_2(x_2)\cdots\boldsymbol{\Psi}_d(x_d), \tag{39}$$

where $\boldsymbol{\Psi}_i(x_i)$ is a $r_{i-1}\times r_i$ matrix with entries $[\boldsymbol{\Psi}_i(x_i)]_{jk} = \psi_i(j;x_i;k)$. The matrix-valued functions $\boldsymbol{\Psi}_i(x_i)$ are known as FTT *tensor cores*. To simplify notation even more we can suppress explicit tensor core dependence on the spatial variable x_i, allowing us to simply write $\boldsymbol{\Psi}_i = \boldsymbol{\Psi}_i(x_i)$ as the spatial dependence is indicated by the tensor core subscript. If we discretize the domain V in terms of a grid with N points in each variable then we can represent (39) as a product of 2D and 3D matrices (see Figure 2).

FTT tensor manifold. It was shown in [15, 8] that the set of truncated tensors (38) (with invertible covariance matrices of each tensor modes) belongs to a *smooth manifold*[5] $\mathcal{M}_{\boldsymbol{r}}$, i.e., a manifold in which we can define a tangent space $T_{u_{\boldsymbol{r}}}\mathcal{M}_{\boldsymbol{r}}$ at a point $u_{\boldsymbol{r}} \in \mathcal{M}_{\boldsymbol{r}}$. Specifically, let us denote by $H^{(i)}_{r_{i-1}\times r_i}$ the set of all tensor cores $\boldsymbol{\Psi}_i \in M_{r_{i-1}\times r_i}(L^2_{\mu_i}(V_i))$ with the property that the autocovariance matrices $\left\langle \boldsymbol{\Psi}_i^{\mathrm{T}}\boldsymbol{\Psi}_i\right\rangle_i \in M_{r_i\times r_i}(\mathbb{R})$ and $\left\langle \boldsymbol{\Psi}_i\boldsymbol{\Psi}_i^{\mathrm{T}}\right\rangle_i \in M_{r_{i-1}\times r_{i-1}}(\mathbb{R})$ are invertible for $i = 1, \ldots, d$. The set

$$\mathcal{M}_{\boldsymbol{r}} = \{u_{\boldsymbol{r}} \in L^2_\mu(V) : \quad u_{\boldsymbol{r}} = \boldsymbol{\Psi}_1\boldsymbol{\Psi}_2\cdots\boldsymbol{\Psi}_d, \quad \boldsymbol{\Psi}_i \in H^{(i)}_{r_{i-1}\times r_i}, \quad \forall i = 1, 2, \ldots, d\}, \tag{40}$$

consisting of fixed-rank FTT tensors, is a smooth sub-manifold of $L^2_\mu(V)$. We represent elements in the tangent space, $T_{u_{\boldsymbol{r}}}\mathcal{M}_{\boldsymbol{r}}$, of $\mathcal{M}_{\boldsymbol{r}}$ at the point $u_{\boldsymbol{r}} \in \mathcal{M}_{\boldsymbol{r}}$ as equivalence classes of velocities of continuously differentiable curves on $\mathcal{M}_{\boldsymbol{r}}$ passing through $u_{\boldsymbol{r}}$

$$T_{u_{\boldsymbol{r}}}\mathcal{M}_{\boldsymbol{r}} = \left\{\gamma'(s)|_{s=0} : \quad \gamma \in \mathcal{C}^1\left((-\delta,\delta), \mathcal{M}_{\boldsymbol{r}}\right), \quad \gamma(0) = u_{\boldsymbol{r}}\right\}. \tag{41}$$

[5] A manifold is a generalization and abstraction of the notion of a curved surface. In particular, the manifold of the FTT tensors with fixed rank is a topological function space that admits a tangent space at each point, an inner product defined on the tangent space, etc.

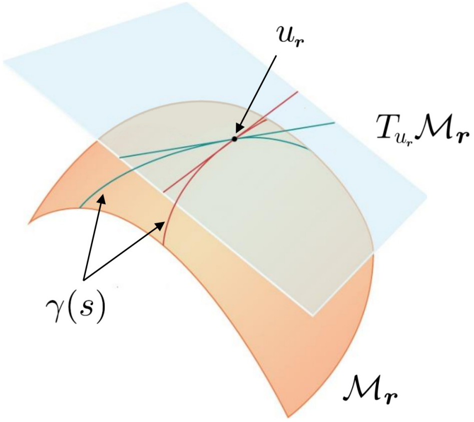

$u_{\boldsymbol{r}}$
$T_{u_r}\mathcal{M}_{\boldsymbol{r}}$
$\gamma(s)$
$\mathcal{M}_{\boldsymbol{r}}$

Figure 3: Sketch of the tensor manifold $\mathcal{M}_{\boldsymbol{r}}$ and the tangent space $T_{u_{\boldsymbol{r}}}\mathcal{M}_{\boldsymbol{r}}$ at $u_{\boldsymbol{r}} \in \mathcal{M}_{\boldsymbol{r}}$. The tangent space is defined as equivalence classes of velocities of continuously differentiable curves $\gamma(s)$ on $\mathcal{M}_{\boldsymbol{r}}$ passing through $u_{\boldsymbol{r}}$.

A sketch of $\mathcal{M}_{\boldsymbol{r}}$ and $T_{u_{\boldsymbol{r}}}\mathcal{M}_{\boldsymbol{r}}$ is provided in Figure 3. Since $L^2_\mu(V)$ is an inner product space, for each $u \in L^2_\mu(V)$ the tangent space $T_u L^2_\mu(V)$ is canonically isomorphic to $L^2_\mu(V)$. Moreover, for each $u_{\boldsymbol{r}} \in \mathcal{M}_{\boldsymbol{r}}$ the normal space to $\mathcal{M}_{\boldsymbol{r}}$ at the point $u_{\boldsymbol{r}}$, denoted by $N_{u_{\boldsymbol{r}}}\mathcal{M}_{\boldsymbol{r}}$, consists of all vectors in $L^2_\mu(V)$ that are orthogonal to $T_{u_{\boldsymbol{r}}}\mathcal{M}_{\boldsymbol{r}}$ with respect to the inner product in $L^2_\mu(V)$

$$N_{u_{\boldsymbol{r}}}\mathcal{M}_{\boldsymbol{r}} = \{w \in L^2_\mu(V) : \langle w, v\rangle_{L^2_\mu(V)} = 0, \quad \forall v \in T_{u_{\boldsymbol{r}}}\mathcal{M}_{\boldsymbol{r}}\}. \tag{42}$$

Since the tangent space $T_{u_{\boldsymbol{r}}}\mathcal{M}_{\boldsymbol{r}}$ is closed, for each point $u_{\boldsymbol{r}} \in \mathcal{M}_{\boldsymbol{r}}$ the space $L^2_\mu(V)$ admits a decomposition into tangential and normal components

$$L^2_\mu(V) = T_{u_{\boldsymbol{r}}}\mathcal{M}_{\boldsymbol{r}} \oplus N_{u_{\boldsymbol{r}}}\mathcal{M}_{\boldsymbol{r}}. \tag{43}$$

We represent elements of the tangent space $T_{u_{\boldsymbol{r}}}\mathcal{M}_{\boldsymbol{r}}$ as equivalence classes of velocities of curves passing through the point $u_{\boldsymbol{r}}$

$$T_{u_{\boldsymbol{r}}}\mathcal{M}_{\boldsymbol{r}} = \left\{y'(s)|_{s=0} : \quad y \in \mathcal{C}^1\left((-\delta,\delta), \mathcal{M}_{\boldsymbol{r}}\right), \quad y(0) = u_{\boldsymbol{r}}\right\}. \tag{44}$$

Here $\mathcal{C}^1\left((-\delta,\delta), \mathcal{M}_{\boldsymbol{r}}\right)$ is the space of continuously differentiable functions from the interval $(-\delta,\delta)$ to the space of constant rank FTT tensors $\mathcal{M}_{\boldsymbol{r}}$.

Next, we can now define a projection onto the tangent space of $\mathcal{M}_{\boldsymbol{r}}$ at $u_{\boldsymbol{r}}$ by

$$\begin{aligned} P_{u_{\boldsymbol{r}}} &: L^2_\mu(V) \to T_{u_{\boldsymbol{r}}}\mathcal{M}_{\boldsymbol{r}} \\ & P_{u_{\boldsymbol{r}}} v = \underset{v_{\boldsymbol{r}} \in T_{u_{\boldsymbol{r}}}\mathcal{M}_{\boldsymbol{r}}}{\operatorname{argmin}} \|v - v_{\boldsymbol{r}}\|_{L^2_\mu(V)}. \end{aligned} \tag{45}$$

For fixed $u_{\boldsymbol{r}}$, the map $P_{u_{\boldsymbol{r}}}$ is linear and bounded. Each $v \in L^2_\mu(V)$ admits a unique representation as $v = v_t + v_n$ where $v_t \in T_{u_{\boldsymbol{r}}}\mathcal{M}_{\boldsymbol{r}}$ and $v_n \in N_{u_{\boldsymbol{r}}}\mathcal{M}_{\boldsymbol{r}}$ (see equation (43)). From this representation it is clear that $P_{u_{\boldsymbol{r}}}$ is an orthogonal projection onto the tangent space $T_{u_{\boldsymbol{r}}}\mathcal{M}_{\boldsymbol{r}}$.

An arbitrary element of the tangent space $T_{u_{\boldsymbol{r}}}\mathcal{M}_{\boldsymbol{r}}$ can be expressed as

$$\dot{u}_{\boldsymbol{r}} = \dot{\boldsymbol{\Psi}}_1\boldsymbol{\Psi}_{\geq 2} + \cdots + \boldsymbol{\Psi}_{\leq i-1}\dot{\boldsymbol{\Psi}}_i\boldsymbol{\Psi}_{\geq i+1} + \cdots + \boldsymbol{\Psi}_{\leq d-1}\dot{\boldsymbol{\Psi}}_d, \tag{46}$$

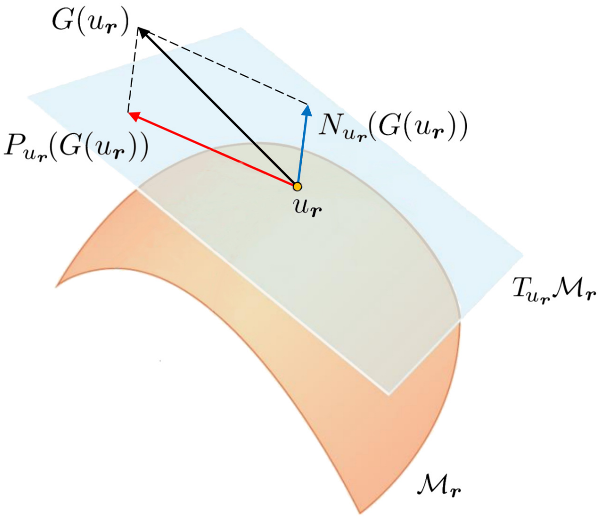

$G(u_{\boldsymbol{r}})$
$N_{u_{\boldsymbol{r}}}(G(u_{\boldsymbol{r}}))$
$P_{u_{\boldsymbol{r}}}(G(u_{\boldsymbol{r}}))$
$u_{\boldsymbol{r}}$
$T_{u_{\boldsymbol{r}}}\mathcal{M}_{\boldsymbol{r}}$
$\mathcal{M}_{\boldsymbol{r}}$

Figure 4: Tangent and normal components of $G\left(u_{\boldsymbol{r}}\right)=\partial u_{\boldsymbol{r}}/\partial t$ at $u_{\boldsymbol{r}}$. The tensor rank of the solution is increased at time t_i if the norm of the normal component $N_{u_{\boldsymbol{r}}}(G(u_{\boldsymbol{r}}))$ is larger than a specified threshold ϵ_{inc}.

where $\dot{u}_{\boldsymbol{r}}=\partial u_{\boldsymbol{r}}/\partial t$ and $\dot{\boldsymbol{\Psi}}_i=\partial\boldsymbol{\Psi}_i/\partial t$.

Dynamic tensor approximation of high-dimensional nonlinear PDEs. With the machinery on FTT tensors available, we can now approximate the solution of (26) on the tensor manifold $\mathbb{M}_{\boldsymbol{r}}$. To this end, suppose that the initial condition $u_0(\boldsymbol{x})$ is on the manifold $\mathcal{M}_{\boldsymbol{r}}$. Clearly, the solution to the initial/boundary value problem (see Figure 4)

$$\begin{cases}\dfrac{\partial u_{\boldsymbol{r}}}{\partial t}=P_{u_{\boldsymbol{r}}}G(u_{\boldsymbol{r}}),\\ u(\boldsymbol{x},0)=u_0(\boldsymbol{x}),\end{cases}\tag{47}$$

remains on the manifold $\mathcal{M}_{\boldsymbol{r}}$ for all $t\geq 0$. Here G is the nonlinear operator on the right hand side of equation (1). The solution to (47) is known as a dynamic approximation to the solution of (1). To compute the tangent space projection of the PDE (48) we solve the *convex optimization problem*

$$\min_{v(\boldsymbol{x},t)\in T_{u(\boldsymbol{x},t)}\mathcal{M}_{\boldsymbol{r}}}\left\|v(\boldsymbol{x},t)-G(u_{\boldsymbol{r}}(\boldsymbol{x},t))\right\|^2_{L^2_\mu(V)}.\tag{48}$$

subject to the DO constraints

$$\left\langle\dot{\boldsymbol{\Psi}}_i^{\mathrm{T}}\boldsymbol{\Psi}_i\right\rangle_i=\mathbf{0}_{r_i\times r_i},\qquad i=1,\ldots,d-1,\tag{49}$$

which ensures that $\left\langle\boldsymbol{\Psi}_i^{\mathrm{T}}(t)\boldsymbol{\Psi}_i(t)\right\rangle_i=\boldsymbol{I}_{r_i\times r_i}$ for all $i=1,\ldots,d-1$ and for all $t\geq 0$.

DO-TT propagator. It was shown in [8] that under these constraints, the convex minimization problem (48) admits a unique minimum for vectors in the tangent space (46) satisfying the PDE system

$$\begin{cases}\dot{\boldsymbol{\Psi}}_1=\left[\left\langle G(u_{\boldsymbol{r}})\boldsymbol{\Psi}_{\geq 2}^{\mathrm{T}}\right\rangle_{\geq 2}-\boldsymbol{\Psi}_1\left\langle\boldsymbol{\Psi}_1^{\mathrm{T}}G(u_{\boldsymbol{r}})\boldsymbol{\Psi}_{\geq 2}^{\mathrm{T}}\right\rangle_{\geq 1}\right]\left\langle\boldsymbol{\Psi}_{\geq 2}\boldsymbol{\Psi}_{\geq 2}^{\mathrm{T}}\right\rangle_{\geq 2}^{-1},\\ \dot{\boldsymbol{\Psi}}_k=\left[\left\langle\boldsymbol{\Psi}_{\leq k-1}^{\mathrm{T}}G(u_{\boldsymbol{r}})\boldsymbol{\Psi}_{\geq k+1}^{\mathrm{T}}\right\rangle_{\leq k-1,\geq k+1}-\right.\\ \qquad\left.\boldsymbol{\Psi}_k\left\langle\boldsymbol{\Psi}_{\leq k}^{\mathrm{T}}G(u_{\boldsymbol{r}})\boldsymbol{\Psi}_{\geq k+1}^{\mathrm{T}}\right\rangle_{\geq 1}\right]\left\langle\boldsymbol{\Psi}_{\geq k+1}\boldsymbol{\Psi}_{\geq k+1}^{\mathrm{T}}\right\rangle_{\geq k+1}^{-1},\qquad k=2,3,\ldots,d-1,\\ \dot{\boldsymbol{\Psi}}_d=\left\langle\boldsymbol{\Psi}_{\leq d-1}^{\mathrm{T}}G(u_{\boldsymbol{r}})\right\rangle_{\leq d-1}.\end{cases}\tag{50}$$

Here, $u_{\boldsymbol{r}}(\boldsymbol{x},t) = \boldsymbol{\Psi}_1(t)\boldsymbol{\Psi}_2(t)\cdots\boldsymbol{\Psi}_d(t) \in \mathcal{M}_r$ and we have introduced the notation

$$\begin{aligned}
\langle \boldsymbol{\Psi} \rangle_{\leq k} &= \int_{V_1\times\cdots\times V_k} \boldsymbol{\Psi}(\boldsymbol{x}) d\mu_1(x_1)\cdots\mu_k(x_k),\\
\langle \boldsymbol{\Psi} \rangle_{\geq k} &= \int_{V_k\times\cdots\times V_d} \boldsymbol{\Psi}(\boldsymbol{x}) d\mu_k(x_k)\cdots\mu_d(x_d),\\
\langle \boldsymbol{\Psi} \rangle_{\leq k-1,\geq k+1} &= \int_{V_1\times\cdots\times V_{k-1}\times V_{k+1}\times\cdots\times V_d} \boldsymbol{\Psi}(\boldsymbol{x}) d\mu_1(x_1)\cdots\mu_{k-1}(x_{k-1})\mu_{k+1}(x_{k+1})\cdots\mu_d(x_d),
\end{aligned} \tag{51}$$

for any matrix $\boldsymbol{\Psi}(\boldsymbol{x}) \in M_{r\times s}\left(L^2_\mu(V)\right)$. The DO-FTT system (50) involves several inverse covariance matrices $\left\langle \boldsymbol{\Psi}_{\geq k}\boldsymbol{\Psi}^{\mathrm{T}}_{\geq k}\right\rangle^{-1}_{\geq k}$, which can become poorly conditioned in the presence of tensor modes with small energy (i.e. autocovariance matrices with small singular values). This phenomenon has been shown to be a result of the fact that the curvature of the tensor manifold at a tensor is inversely proportional to the smallest singular value present in the tensor [10, section 4]. To overcome the problem of inverting potentially ill-conditioned covariance matrices a rank-adaptive operator splitting method was proposed in [6].

Numerical application of DO-TT to the Fokker-Planck equation. We have seen in Chapter 2 that the Fokker–Planck equation describes the evolution of the probability density function (PDF) of the state vector solving the Itô stochastic differential equation (SDE) [13]

$$d\boldsymbol{X}_t = \boldsymbol{\mu}(\boldsymbol{X}_t,t)dt + \boldsymbol{\sigma}(\boldsymbol{X}_t,t)d\boldsymbol{W}_t. \tag{52}$$

Here, $\boldsymbol{X}_t$ is the d-dimensional state vector, $\boldsymbol{\mu}(\boldsymbol{X}_t,t)$ is the d-dimensional drift, $\boldsymbol{\sigma}(\boldsymbol{X}_t,t)$ is an $d\times m$ matrix and $\boldsymbol{W}_t$ is an m-dimensional standard Wiener process. The Fokker–Planck equation that corresponds to (52) has the form

$$\begin{cases} \dfrac{\partial p(\boldsymbol{x},t)}{\partial t} = \mathcal{L}(\boldsymbol{x},t)p(\boldsymbol{x},t),\\[2ex] p(\boldsymbol{x},0) = p_0(\boldsymbol{x}), \end{cases} \tag{53}$$

where $p_0(\boldsymbol{x})$ is the PDF of the initial state $\boldsymbol{X}_0$, $\mathcal{L}$ is a second-order linear differential operator defined as

$$\mathcal{L}(\boldsymbol{x},t)p(\boldsymbol{x},t) = -\sum_{k=1}^{d} \frac{\partial}{\partial x_k}\left(\mu_k(x,t)p(\boldsymbol{x},t)\right) + \sum_{k,j=1}^{d} \frac{\partial^2}{\partial x_k \partial x_j}\left(D_{ij}(\boldsymbol{x},t)p(\boldsymbol{x},t)\right), \tag{54}$$

and $\boldsymbol{D}(\boldsymbol{x},t) = \boldsymbol{\sigma}(\boldsymbol{x},t)\boldsymbol{\sigma}(\boldsymbol{x},t)^{\mathrm{T}}/2$ is the diffusion tensor. For our numerical demonstration we set

$$\boldsymbol{\mu}(\boldsymbol{x}) = \alpha \begin{bmatrix} \sin(x_1)\\ \sin(x_3)\\ \sin(x_4)\\ \sin(x_1) \end{bmatrix}, \qquad \boldsymbol{\sigma}(\boldsymbol{x}) = \sqrt{2\beta} \begin{bmatrix} g(x_2) & 0 & 0 & 0\\ 0 & g(x_3) & 0 & 0\\ 0 & 0 & g(x_4) & 0\\ 0 & 0 & 0 & g(x_1) \end{bmatrix}, \tag{55}$$

where $g(x) = \sqrt{1 + k\sin(x)}$. With the drift and diffusion matrices chosen in (55) the operator (54) takes the form

$$\begin{aligned}
\mathcal{L} =& -\alpha\left(\cos(x_1) + \sin(x_1)\frac{\partial}{\partial x_1} + \sin(x_3)\frac{\partial}{\partial x_2} + \sin(x_4)\frac{\partial}{\partial x_3} + \sin(x_1)\frac{\partial}{\partial x_4}\right)\\
&+ \beta\left((1 + k\sin(x_2))\frac{\partial^2}{\partial x_1^2} + (1 + k\sin(x_3))\frac{\partial^2}{\partial x_2^2} + (1 + k\sin(x_4))\frac{\partial^2}{\partial x_3^2} + (1 + k\sin(x_1))\frac{\partial^2}{\partial x_4^2}\right).
\end{aligned} \tag{56}$$

Clearly $\mathcal{L}$ is a linear, time-independent separable operator of rank 9, since it can be written as

$$\mathcal{L} = \sum_{i=1}^{9} L_i^{(1)} \otimes L_i^{(2)} \otimes L_i^{(3)} \otimes L_i^{(4)}, \tag{57}$$

where each $L_i^{(j)}$ operates on x_j only. Specifically, we have

$$\begin{aligned}
&L_1^{(1)} = -\alpha\cos(x_1), && L_2^{(1)} = -\alpha\sin(x_1)\frac{\partial}{\partial x_1}, && L_3^{(2)} = -\alpha\frac{\partial}{\partial x_2}, && L_3^{(3)} = \sin(x_3),\\
&L_4^{(3)} = -\alpha\frac{\partial}{\partial x_3}, && L_4^{(4)} = \sin(x_4), && L_5^{(1)} = -\alpha\sin(x_1), && L_5^{(4)} = \frac{\partial}{\partial x_4},\\
&L_6^{(1)} = \beta\frac{\partial^2}{\partial x_1^2}, && L_6^{(2)} = 1 + k\sin(x_2), && L_7^{(2)} = \beta\frac{\partial^2}{\partial x_2^2}, && L_7^{(3)} = 1 + k\sin(x_3),\\
&L_8^{(3)} = \beta\frac{\partial^2}{\partial x_3^2}, && L_8^{(2)} = 1 + k\sin(x_4), && L_9^{(4)} = \beta\frac{\partial^2}{\partial x_4^2}, && L_9^{(1)} = 1 + k\sin(x_1),
\end{aligned} \tag{58}$$

and all other unspecified $L_i^{(j)}$ are identity operators. We set the parameters in (55) as $\alpha = 0.1$, $\beta = 2.0$, $k = 1.0$ and solve (53) on the four-dimensional flat torus $\mathbb{T}^4$. The initial PDF is set as

$$p_0(\boldsymbol{x}) = \frac{\sin(x_1)\sin(x_2)\sin(x_3)\sin(x_4) + 1}{16\pi^4}. \tag{59}$$

Note that (59) is a four-dimensional FTT tensor with multilinear rank $\boldsymbol{r} = \begin{bmatrix} 1 & 2 & 2 & 2 & 1 \end{bmatrix}$. Upon normalizing the modes appropriately we obtain the left orthogonalized initial condition required to begin integration

$$\begin{aligned}
p_0(\boldsymbol{x}) = &\psi_1(1; x_1; 1)\psi_2(1; x_2; 1)\psi_3(1; x_3; 1)\psi_4(1; x_4; 1)\sqrt{\lambda(1)}\\
&+ \psi_1(1; x_1; 2)\psi_2(2; x_2; 2)\psi_3(2; x_3; 2)\psi_4(2; x_4; 1)\sqrt{\lambda(2)},
\end{aligned} \tag{60}$$

where

$$\psi_i(1; x_i; 1) = \frac{\sin(x_i)}{\sqrt{\pi}}, \qquad \sqrt{\lambda(1)} = \frac{1}{16\pi^2}. \tag{61}$$

All other tensor modes are equal to $1/\sqrt{2\pi}$, and $\sqrt{\lambda(2)} = 1/(2\pi^2)$. To obtain a benchmark solution with which to compare the rank-adaptive FTT solution, we solve the PDE (53) using a Fourier pseudo-spectral method on the flat torus $\mathbb{T}^4$ with $21^4 = 194481$ evenly-spaced points. As before, the operator $\mathcal{L}$ is represented in terms of pseudo-spectral differentiation matrices, and the resulting semi-discrete approximation (ODE system) is integrated with an explicit fourth-order Runge Kutta method using time step $\Delta t = 10^{-4}$. The numerical solution we obtained in this way is denoted by $p_{\text{ref}}(\boldsymbol{x}, t)$. We also solve the Fokker-Planck using the proposed rank-adaptive FTT method with first-order Lie-Trotter time integrator and normal vector thresholding. We run three simulations all with time step $\Delta t = 10^{-4}$: one with no rank adaption, and two with rank-adaptation and normal component thresholds set to $\epsilon_{\text{inc}} = 10^{-3}$ and $\epsilon_{\text{inc}} = 10^{-4}$. In Figure 5 we plot three time snapshots of the two-dimensional solution marginal

$$p(x_1, x_2, t) = \int_0^{2\pi}\int_0^{2\pi} p(x_1, x_2, x_3, x_4, t)dx_3dx_4 \tag{62}$$

computed with the rank-adaptive FTT integrator ($\epsilon_{\text{inc}} = 10^{-4}$) and the full tensor product pseudo-spectral method (reference solution). In Figure 6(a) we compare the $L^2(\Omega)$ errors of the rank-adaptive method relative to the reference solution. It is seen that as we decrease the threshold the solution becomes more accurate. In Figure 6(b) we plot the component of $\mathcal{L}p_{\boldsymbol{r}}$ normal to the tensor manifold. Note that in the rank-adaptive FTT solution with thresholds $\epsilon_{\text{inc}} = 10^{-3}$ and $\epsilon_{\text{inc}} = 10^{-4}$ the solver performs both mode addition as well as mode removal. This is documented in Figure 7. The abrupt change in rank

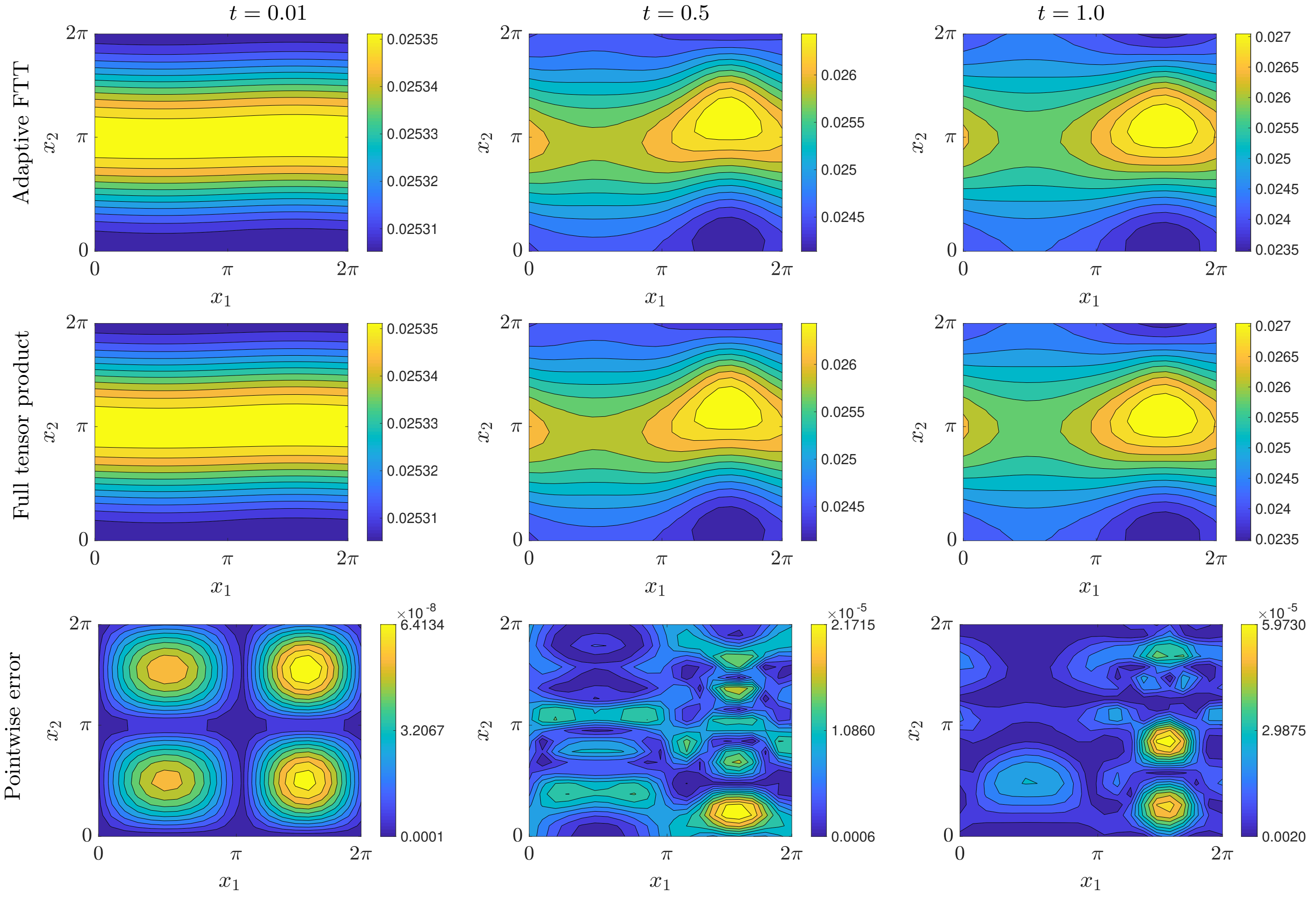


Figure 5: Time snapshots of marginal PDF $p_{\boldsymbol{r}}(x_1, x_2, t)$ corresponding to the solution to the Fokker-Planck equation (53). We plot marginals computed with the rank-adaptive FTT integrator using $\epsilon_{\text{inc}} = 10^{-4}$ (top row) and with the full tensor product Fourier pseudo-spectral method (middle row). We also plot the pointwise error between the two numerical solutions (bottom row). The initial condition is the FTT tensor (59).

observed in Figure 7(a)-(c) near time $t = 0.4$ corresponding to the rank-adaptive solution with threshold $\epsilon\text{inc} = 10^{-4}$ is due to the time step size Δt being equal to ϵ_{inc}. This can be justified as follows. Recall that the solution is first order accurate in Δt and therefore the approximation of the component of $\mathcal{L}p_{\boldsymbol{r}}$ normal to the tensor manifold $\mathcal{M}_{\boldsymbol{r}}$ is first-order accurate in Δt. If we set $\epsilon_{\text{inc}} \leq \Delta t$, then the rank-adaptive scheme may overestimate the number of modes needed to achieve accuracy on the order of Δt. This does not affect the accuracy of the numerical solution due to the robustness of the Lie-Trotter integrator to over-approximation [11]. Moreover we notice that the rank-adaptive scheme removes the unnecessary modes ensure that the tensor rank is not unnecessarily large. In fact, the diffusive nature of the Fokker-Plank equation on the flat torus $\mathbb{T}^4$ yields relaxation to a statistical equilibrium state that depends on the drift and diffusion coefficients in (53). Such an equilibrium state may be well-approximated by a low-rank FTT tensor.

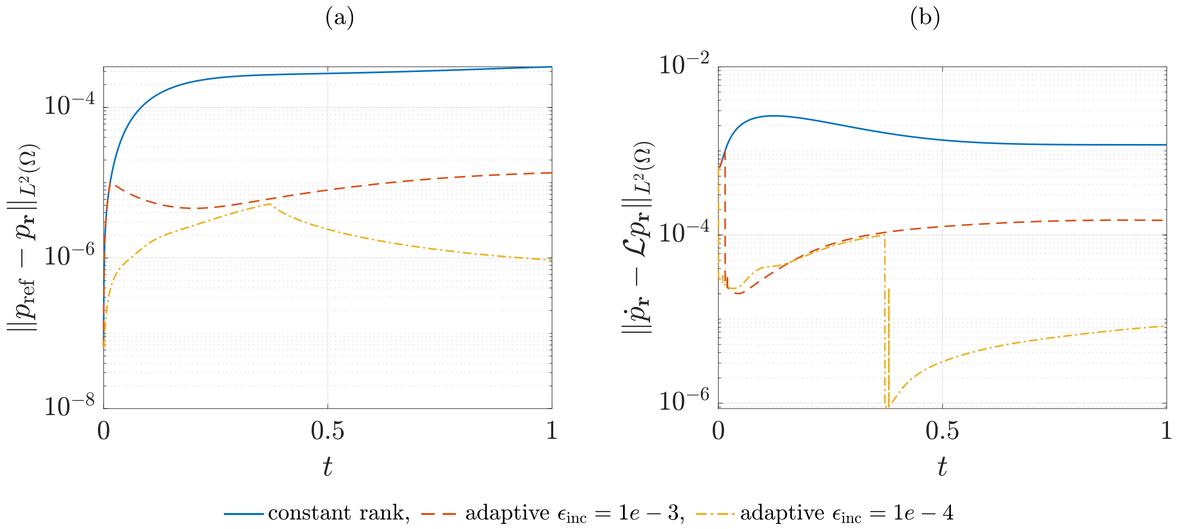


Figure 6: (a) The $L^2(\Omega)$ error of the FTT solution $p_{\boldsymbol{r}}(\boldsymbol{x}, t)$ relative to the benchmark solution $p_{\text{ref}}(\boldsymbol{x}, t)$ computed with a Fourier pseudo-spectral method on a tensor product grid. (b) Norm of the component of $\mathcal{L}p_{\boldsymbol{r}}$ normal to the tensor manifold (see Figure 4). Such component is approximated a two-point BDF formula at each time step.

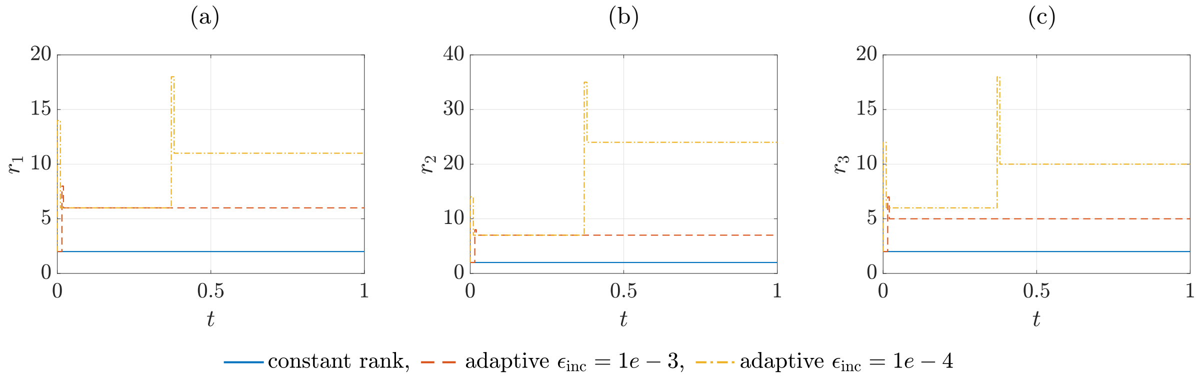


Figure 7: Tensor rank $\boldsymbol{r} = [1\, r_1\, r_2\, r_3\, 1]$ of adaptive FTT solution to the four dimensional Fokker-Planck equation (53).

References

[1] H. Babaee, M. Choi, T. P. Sapsis, and G. E. Karniadakis. A robust bi-orthogonal/dynamically-orthogonal method using the covariance pseudo-inverse with application to stochastic flow problems. *J. Comput. Phys.*, 344:303–319, 2017.

[2] D. Bigoni, A. P. Engsig-Karup, and Y. M. Marzouk. Spectral tensor-train decomposition. *SIAM J. Sci. Comput.*, 38(4):A2405–A2439, 2016.

[3] M. Cheng, T. Y. Hou, and Z. Zhang. A dynamically bi-orthogonal method for time-dependent stochastic partial differential equations I: derivation and algorithms. *J. Comput. Phys.*, 242:843–868, 2013.

[4] M. Cheng, T. Y. Hou, and Z. Zhang. A dynamically bi-orthogonal method for time-dependent stochastic partial differential equations II: adaptivity and generalizations. *J. Comput. Phys.*, 242:753–776, 2013.

[5] M. Choi, T. Sapsis, and G. E. Karniadakis. On the equivalence of dynamically orthogonal and bi-orthogonal methods: Theory and numerical simulations. *J. Comput. Phys.*, 270:1–20, 2014.

[6] A. Dektor, A. Rodgers, and D. Venturi. Rank-adaptive tensor methods for high-dimensional nonlinear pdes. *Journal of Scientific Computing*, 88(36):1–27, 2021.

[7] A. Dektor and D. Venturi. Dynamically orthogonal tensor methods for high-dimensional nonlinear PDEs. *J. Comput. Phys.*, 404:109125, 2020.

[8] A. Dektor and D. Venturi. Dynamic tensor approximation of high-dimensional nonlinear pdes. *J. Comput. Phys.*, 437:110295, 2021.

[9] M. Gerritsma, J.-B. van der Steen, P. Vos, and G. E. Karniadakis. Time-dependent generalized polynomial chaos. *J. Comput. Phys.*, 229(22):8333–8363, 2010.

[10] O. Koch and C. Lubich. Dynamical low-rank approximation. *SIAM J. Matrix Anal. Appl.*, 29(2):434–454, 2007.

[11] C. Lubich, I. V. Oseledets, and B. Vandereycken. Time integration of tensor trains. *SIAM J. Numer. Anal.*, 53(2):917–941, 2015.

[12] I. V. Oseledets. Constructive representation of functions in low-rank tensor formats. *Constructive Approximation*, 37:1–18, 2013.

[13] H. Risken. *The Fokker-Planck equation: methods of solution and applications*. Springer-Verlag, second edition, 1989. Mathematics in science and engineering, vol. 60.

[14] T. P. Sapsis and P. F .J. Lermusiaux. Dynamically orthogonal field equations for continuous stochastic dynamical systems. *Physica D*, 238(23-24):2347–2360, 2009.

[15] A. Uschmajew and B. Vandereycken. The geometry of algorithms using hierarchical tensors. *Linear Algebra Appl.*, 439(1):133–166, 2013.