SANDIA REPORT

SAND2008-0963
Unlimited Release
Printed February 2008

Elements Of A Function Analytic Approach to Probability

John R. Red-Horse, Roger G. Ghanem

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation,
a Lockheed Martin Company, for the United States Department of Energy's
National Nuclear Security Administration under Contract DE-AC04-94-AL85000.

Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: http://www.osti.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd.
Springfield, VA 22161

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.fedworld.gov
Online order: http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND2008-0963
Unlimited Release
Printed February 2008

Elements Of A Function Analytic Approach to Probability

John R. Red-Horse Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87185-0828
jrredho@sandia.gov

Roger G. Ghanem
University of Southern California
Department of Aerospace and Mechanical Engineering
Los Angeles, California 90089-2531
ghanem@usc.edu

Abstract

We first provide a detailed motivation for using probability theory as a mathematical context in which to analyze engineering and scientific systems that possess uncertainties. We then present introductory notes on the function analytic approach to probabilistic analysis, emphasizing the connections to various classical deterministic mathematical analysis elements. Lastly, we describe how to use the approach as a means to augment deterministic analysis methods in a particular Hilbert space context, and thus enable a rigorous framework for commingling deterministic and probabilistic analysis tools in an application setting.

Contents

Figures

Summary

Probability theory is a mature mathematical discipline, which has a long history as a toolset for analyzing systems possessing uncertainty. One enduring drawback to full acceptance in some areas of science and engineering is that most popular probabilistic approaches utilize specialized probabilistic descriptors, such as so-called probability density functions (PDFs), or cumulative probability functions (CDFs) that infer more information into an analysis than is justified by the available observed data. In these approaches, unbalanced probabilistic inferences can actually obscure vital information present in the accompanying process models. Moreover, for historical reasons, these same traditional probabilistic methods often employ obscure definitions, and analytical tools so as to render them less appealing to general process analysts.

We are proposing to overcome these, and other, roadblocks to acceptance by returning to a relatively old idea first described by Kolmogorov and Weiner in the early part of the 20th century. Namely, that probabilistic systems can be described in an equivalent, alternative manner to specialized probability descriptors by using a functional approach.

Since, functional analysis provides the foundation for developing approximate solutions in many applications in science and engineering, it is not surprising that this same function analytic approach, applied in a probabilistic setting, provides a more satisfying path towards the development of a rigorous framework for augmenting traditional approximation-based deterministic analysis methods to accommodate uncertainty quantification (UQ).

The key is to recognize that the so-called random mappings, that is, random variables (RVs) and random fields (RFs), are, in fact, not random but rather deterministic mappings where at least a portion of the domain consists of sets in a probability sample space. The randomness enters the picture via the probability associated with the occurrance of a given domain set.

The path one follows is first to establish the building blocks for the case of RVs. Then one migrates from RVs to RFs by building on them to develop a product space-style set expansion, similar to what one uses to increase the domain in a deterministic setting when migrating from an ODE to a PDE, and to build a new measure space in this expanded domain.

The key distingishing feature between the two is that in the deterministic migration, one has to deal only with one measure in each of the constituent subdomains, specifically Lebesgue measure. In the UQ-enabled setting one has to generalize to product measures where a wide class of general probability measures are a constituent. Naturally, this fact complicates the story somewhat.

In this document, we provide a narrative in which we do these things using, primarily, a particular application setting: the classical Hilbert space, L_2. In the

course of this narrative, we demonstrate our ideas through simple examples, where, in the RF case, decompose our function mappings into products, just as one encounters in classical separation of variables solutions of PDEs, and discuss the relationships that exist in the UQ setting that tie the two decomposed function spaces together. Where appropriate, we provide a discussion on how one might seek to achieve results on more general function space settings.

1 Introduction

The bases of science and engineering can be thought of as evidence-driven discovery processes tasked with advancing the state of knowledge. One important goal of these efforts is to develop models that approximate reality, and that are, at best, tailored to the current state of knowledge. As scientific discovery evolves, along with knowledge and technology, the level of detail that a model can capture increases. However, regardless of the detail present in a given model, the presence of uncertainty is ubiquitous in physical scenarios of interest. Regardless of this fact, there has historically been persistent disagreement among scientists and engineers regarding the very concept. We give our perspective here.

1.1 General Thoughts on Uncertainty

Uncertainty can be thought of as having two broad classifications. First, there is inherent uncertainty; that is, uncertainty that cannot be reduced by gathering more information, at least at the scale at which available observations have been made. The reference to scale is intentional: The origins of this class of uncertainty typically entail phenomena that occur at scales that are smaller than either the model or the observations that are being effected in a given study. We mention in passing that this scale-limitation effect is known in quantum mechanics as the Heisenberg Uncertainty Principal.

Sub-scale information can be thought as being packaged for use at the analysis scale; this packaging thus consists of a scale-bridging step, either explicitly or implicitly, and typically results in a truncated view of the sub-scale information. This class of uncertainty is often termed irreducible, aleatoric, and variability, to name a few of the more common labels.

Second, there is reducible uncertainty, or ignorance. This class finds its origins in lack of information, and, typically, is associated with limitations in models or available experimental data.

There is a subtle point here in that any models that are present in a given analysis, including those tasked with modeling uncertainty, are idealizations of the phenomena they address. These approximations always involve some level of ignorance; thus, it is imperative for the analyst to understand and to explore the effects of ignorance of each model a given analysis incorporates to determine the combined effects of the approximations, and to address efficient ways to gather information necessary to reduce these effects, when possible.

1.1.1 An Example: Coin Tossing

Let's try to illustrate the above points with the classic example of coin tossing.

Coin tossing, while very often taken for granted as the prototypical "random event," is, at least conceptually, a deterministic procedure. That is, in theory, if one were able to make every observation necessary with sufficient detail to assess all of the factors affecting the tossing event, then, with modeling capabilities that are appropriate for the mechanics involved, one could make a prediction on the outcome. Yet even for this seemingly simple event, assessing these factors is surprisingly complicated. These include the physical properties of the coin itself, the externally applied forces, its environment while airborne, and the conditions that it will encounter upon striking the landing surface.

Given the breadth of information required, perhaps it's not surprising that this single, simple affair has been ordained as the arbiter for many mainstream affairs, including those in the political arena. Nonetheless, it has continued to this day to be a focus of contemporary research [10, 26, 9].

The inherent variability in the above factors have been used to justify departing totally from a physics-based model of the event. Most often, the coin-tossing outcomes are represented by a discrete random variable with two possible values. A toss of a so-called fair coin is presumed to have a 50% chance of resulting in a head or a tail, a consistent cumulative probability function is assigned to the random variable, and this function is utilized in subsequent analyses. Doing the above means that the underlying physics are replaced entirely with an uncertainty-based model; no cause and effect understanding is presumed or utilized. For the case of the coin, this is a decision based on practicality; making such a decision for other physical scenarios is assuredly non-trivial.

While we acknowledge that the above description appeals to a frequentist interpretation of probability, this interpretation is, in general, transcended by probability theory. For more on this, the underlying philosophical characteristics of coin tossing, and to uncertainty modeling in general, we refer the reader to [25].

1.1.2 Probabilistic Approach to Uncertainty Analysis

Implicitly in the above discussion, we have demonstrated the prime motivation for addressing uncertainty: It is always present; and, even if it's presence can be safely neglected for a given application, it is often necessary to perform a coarse, uncertainty-enabled analysis to establish this.

For those cases where uncertainty cannot neglected, the analyst must go through steps to account for it properly for the application at hand. This includes characterizing it, developing an appropriate mathematical context, and producing models for

it within this context.

1.1.3 A Brief History of Probability

Based on its long history, and strong association, albeit relatively recent, with traditional mathematics disciplines, probability theory is a strong candidate for providing such a context. Here we provide some background on this history and association.

The formal birth of probability theory can be traced to two French mathematicians, Blaise Pascal and Pierre de Fermat, in the mid-17th century as a means to analyze games of chance. The popularity of these games at that time, instigated keen interest in probability theory, and applications were almost exclusively composed of them.

It was in the early 19th century when Pierre de Laplace, in his book Théorie Analytique des Probabilités, extended the application base to many scientific and practical problems. It was during the 19th century, that many important applications, including the theory of errors, actuarial mathematics, and statistical mechanics, were broached using probability theory.

In the early 20th century, largely due to the development of Lebesgue integration, probability theory began a path towards ever-stronger association with rigorous mathematical disciplines. This burgeoning nexus was made explicit by the Russian mathematician, A. N. Kolmogorov, who, in a 1933 monograph in Russian, for the first time posited an axiomatic approach to probability theory. This monograph was translated into English in 1950 as *Foundations of Probability Theory* [28]. In his book, Kolmogorov took the first sophisticated steps towards aligning probability theory with what is the general discipline known as measure theory.

But, as was the case from the early days, the driving forces behind the development of probability theory were applications. The development of theoretical and experimental physics and chemistry, involving atomic and other sub-microscopic entities, was symbiotic with the theoretical development of probabilistic methods.

It was for such an application that N. Wiener developed his theory of polynomial chaos expansions [52] as a means to filter inputs to systems containing Brownian motion paths [14]. This was perhaps the first foray into applications of functional analysis in probability.

The state of the art in probability, bolstered by its strong association with mathematics disciplines such as functional analysis, numerical analysis, and topology, has continued to evolve over the the past 50 years. This evolution has kept probability relevant in the face of ever-changing application analysis requirements.

1.2 Functional Analytic Approach to Probability: An Overview

We now place our attention on one analysis path in particular; we do so through the use of a simple case, which we expand on in some detail later.

In the field of probability, there are two primary means of analysis. The first of these is the more traditional approach, in which one is concerned with properties of certain probabilistic entities, such as cumulative distribution functions, probability density functions, or a statistical moment, and their behavior under transformation or limit operations.

There is, however, an alternative approach, which we will refer to as function analytic probability. The basis of this approach is the recognition that probabilistically-defined random functions, such as random variables (RVs) and random fields (RFs), are well-defined functional mappings with, at a minimum, a subset of their domain of definition being a sample space, Ω, of elementary events. This concept implicitly assumes many things. For example, the sample space is assumed to be understood in the context of a probability space, which is composed of the measure triple $(\Omega, \mathbf{S}, \mathbf{P})$. This measure triple, and thus the corresponding probability space, consists of a sigma algebra, $\mathbf{S}$, of subsets of Ω called events, and a probability measure, P, which itself is just a measure constrained to $P(\Omega) = 1$. Each of these entities has a well-established and precise mathematical structure and property set associated with it [16, 44, 12].

Using the above structure, it is easy to observe that a random function is, in fact, not random at all, and that whatever uncertainty that appears in the range space of these mappings originates in the uncertainty for events in $\mathbf{S}$. These are completely specified in the probability space by P. This fact, combined with the long history of function approximation in traditional engineering analyses, means that casting these random functions similarly makes perfect sense as a context for the examination of same classes of application in the presence of uncertainty.

Within this mathematical setting, there are many analysis alternatives that depend on the complexity of the mathematical structure warranted by the physical problem: algebraic, semi-group, topological, etc. In this document, we describe one of these alternatives, Hilbert space, and pay particular attention to using it in applications relevant to models one encounters in a variety of engineering disciplines. Since, our main thrust will be to enrich the mathematical structure typically associated with deterministic approaches by augmenting the problem domains with suitable probabilistic subdomains, we will generalize the deterministic mathematical elements, such as measures, norms, and inner products where appropriate, using product space techniques. The establishment of analysis fundamentals, such as the augmented measure triple and the properties of the various classes of generalized mappings, will readily accommodate the use of well-established functional analysis methods in our uncertainty-enabled context.

The motivation for this approach is that these new building blocks, which leverage analysis techniques developed throughout the rich history of deterministic functional analysis, will be useful as the basis for selecting problem schema suitable for addressing such diverse applications as prediction via partial differential equations (PDEs), optimization under uncertainty (OUU), and model calibration and validation, including experiment design and regularization.

In closing, we note that under identical assumptions, the probabilistic solutions that result from either of the two main analytical paths are identical. They offer competing approaches to package information for subsequent analysis; the specific approach taken should be dictated by the particulars associated with a given application class. Finally, a consequence is that the above mappings are also subject to the same approximation criteria as one might encounter, say, in a deterministic finite element analysis, only in this case we have generalized the distance metrics to allow for the use of probability measures in associated norms defining the approximations. Naturally, the deterministic solutions are fully recovered when the probabilistic subdomains are eliminated from the analysis specifications.

2 Mathematical Framework

Probability theory has evolved greatly from the aforementioned axiomatic formulation statement of Kolmogorov. This evolution has had the effect of aligning it rigorously with sub-elements of traditional classical mathematical analysis, including set theory, topology, measure theory and integration, and functional analysis.

Our goal is to cite the necessary connections and to exploit them to construct, ultimately, our random process approximations in a Hilbert Space context, and to use these approximations in our analyses as a means to expand upon deterministic results to accommodate uncertainty. To achieve this, we will rely upon the concepts of norm (distance) and inner product (geometry), both of which require integration. First, we assemble a collection of mathematical building blocks.

2.1 Measure Spaces and Probability

A probability space is a measure space, which is completely specified by a measure triple $(\Omega, \mathcal{S}, P)$ [12]. Here Ω is the non-empty universal set of elementary events, or sample space; $\mathcal{S}$ is a σ-algebra of subsets of Ω; and, P is a measure with $P(\Omega) = 1$. This last constraint is what distinguishes a probability measure versus more general measures. One will often see a probability space referred to as a probability triple.

Let us describe what these terms mean in greater detail.

The sample space, Ω, can be any abstractly defined set. The fact that it is the universal set means that complements of subsets of Ω are assumed to be taken relative to it.

Members of the σ-algebra in a probability space, $\mathcal{S}$, which are referred to as events, are simply the P-measurable sets. By definition $\mathcal{S}$ is a collection of subsets of the sample space that contains Ω and its complement, the empty set, $\emptyset$, and is closed under countable unions and intersections of sets in the collection. We explicitly state here that countable can include countably infinite should Ω be an uncountable set. Events are more generally referred to as measurable sets, and, as the name implies, the restriction of subsets of the domain of a measure is always necessary to ensure the consistency necessary for measures, probability or otherwise, to make sense.

Finally, we note that the process of creation of a measure means specifying all of the elements of its triple; it is quite a detailed and complicated affair to do this. For example, Lebesgue measure, $\boldsymbol{m}$, on the real line, $\boldsymbol{R}^1$, which is the typical measure space for deterministic problems on $\boldsymbol{R}^1$, is developed via an extremely complicated process despite being based on a simple distance metric for any open or closed interval. Specifically, for any such interval $I \in \boldsymbol{R}^1$,

$$m(I) = b - a, \text{ where } b > a. \tag{1}$$

Thus the phrase Lebesgue measure really refers to the measure space, $(\boldsymbol{R}^1, \mathcal{L}, \boldsymbol{m})$, implicitly. Here $\mathcal{L}$ is the σ-algebra of Lebesgue measurable sets; we will have more to say on this later. Naturally, when we restrict Lebesgue measure to any unit interval, or to any finite interval with $\hat{\boldsymbol{m}} = (\frac{1}{(b-a)})\boldsymbol{m}$ replacing $\boldsymbol{m}$, in the measure triple, and adjusting its corresponding σ-algebra appropriately, the result is uniform probability measure. We refer the interested reader to [42] for additional information on this topic.

One last subtle issue relating to measure spaces is that of completion of the measure space. We mentioned previously that the process of developing the Lebesgue measure was complicated. One of these complicating factors is that we have to enrich the smallest σ-algebra achieved through countable unions and intersections of open intervals in $\boldsymbol{R}^1$ to arrive at $\mathcal{L}$.

For this case, the smallest σ-algebra, denoted by $\mathcal{B}$, is the so-called Borel σ-algebra generated by the open intervals in $\boldsymbol{R}^1$. Elements of $\mathcal{B}$ are called the Borel sets on the real line, and this measure triple, also derived from interval length, is called Borel measure. Generally, Borel sets are an important mathematical entity; they are defined for any topology, not just those on $\boldsymbol{R}^1$, as the smallest σ-algebra containing that topology. Such a close relationship between Borel σ-algebras and measure spaces enables us to connect the concepts of measurability and continuity for functional mappings on these spaces. We mention, however, that even with this connection, we still have to take care to restrict our underlying topological spaces to be locally compact, σ-compact Hausdorff spaces, and our measures to be *sigma*-finite Borel measures. Fortunately, we can always guarantee this for Euclidian spaces composed entirely of copies of $\boldsymbol{R}^1$; we direct the interested reader to [44] for additional information on this.

Regardless, for our case on the real line, we evolve from $\mathcal{B}$ to $\mathcal{L}$, that is, from Borel to Lebesgue measure through a process known as completion of measures [44]. This process entails enlarging $\mathcal{B}$ to include all subsets, say N_i, of Borel sets of measure zero, $m(B) = 0$, that cannot be achieved through countable unions and intersections of sets in $\mathcal{B}$. Sets of measure zero play an important role in measure theory, so understanding this seemingly subtle point is important as well.

With all of the basic definitions, we are now in a position to state the Axioms of Probability explicitly:

- The probability of an event is a non-negative real number:

$$P(E) \geq 0 \qquad \forall E \in \Omega. \tag{2}$$

- The probability for the entire sample space is 1:

$$P(\Omega) = 1. \tag{3}$$

- We require that P be countably additive. That is, given a countable collection of a mutually disjoint set of events, that is $A_i \in \mathcal{S}, i = 1, \ldots, \infty$ such that $A_i \bigcap A_j = \emptyset$ for all $i \neq j$, then:

$$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i) \tag{4}$$

2.2 Random Variables

The cornerstone of the function analytic probability approach is the fact that mappings exist from a sample space to an analytical range space that is amenable to conventional analysis in a manner that is consistent with deterministic methods. Random variables (RVs) and random processes or fields (RFs) will be the mechanism for accomplishing this. Since we will assume these mappings take range values in the real line, $\boldsymbol{R}^1$, we will refer to them as real RVs. Naturally, RVs can be generalized to the complex plane, vectors, and processes with the same foundational elements.

Consider a real RV, X. It is by definition a measurable mapping from a probability space to a measure space on $\boldsymbol{R}^1$, which can be thought of as a function on a sample space Ω

$$X : \Omega \longrightarrow \boldsymbol{R}^1 \tag{5}$$

or as a set function on the corresponding set of events $\mathcal{S}$

$$X : \mathcal{S} \longrightarrow \mathcal{B} \tag{6}$$

where $\mathcal{B}$ is an appropriate σ-algebra that is part of the measure triple $(\boldsymbol{R}^1, \mathcal{B}, \mu)$. By definition, a function is measurable if the inverse mappings of sets in $\mathcal{B}$ are sets in $\mathcal{S}$. So both of the perspectives on the mapping definition guarantee that X is constrained so that both of the measure spaces that it connects are well-defined, and, depending on the range measure space, X we say that it is Borel or Lebesgue measurable.

Since our mappings result in an explicit connection between the domain and range σ-algebras, we can guarantee that $P(\boldsymbol{X}^{-1}(B))(\in \mathcal{S}), B \in \mathcal{B}$ is well-defined. Thus, we can now state that μ is also measure, and we explicitly define it using P via X by $\mu(B) = P(X^{-1}(B)$ for any set $B \in \mathcal{B}$.

Our μ above is known as the distribution of X, and is denoted by dist_X. Several commonly encountered probabilistic entities derive from the distribution. These including the cumulative distribution function (CDF),

$$F_X(x) = \mu\{(-\inf, x]\}, \tag{7}$$

and the probability density function (PDF), when it exists,

$$p_X(x) = \left. \frac{dF_X}{dx} \right|_x. \tag{8}$$

This concept of a PDF in Eq 8 is quite general. We can represent the measure on $\boldsymbol{R}^1$, $\mu = \text{dist}_X$ of random variable X defined above using the Radon-Nikodym Theorem [16]. Namely, the Lebesgue decomposition of μ is

$$\mu(A) = \int_A f(x)d\lambda + \mu^s(A), \quad A \in \mathcal{B}, \tag{9}$$

where μ^s is a discrete measure; when $\mu^s = 0$, the measure μ is said to be absolutely continuous. The function $f(x)$ is defined to be the Radon-Nikodym derivative of μ with respect to λ, and is often referred to as the probability density of the absolutely continuous part random variable X with respect to the measure λ. A common, and important, situation occurs when λ is Borel or Lebesgue measure on $\boldsymbol{R}^1$, was the case for Eq 8; we note, however, that λ could have been any σ-finite measure on a given measure space.

There is a final noteworthy characteristic of mappings on measure spaces: Generally, any two can differ from one another on a domain set of measure zero, and thus be equal on any set of P-measure > 0. This property is referred to as equality almost everywhere relative to the measure P. Two RVs X and Y that have this relationship, which we denote them $X = Y \text{a.e.} P$, are members of an equivalence class of RVs that also share the same relationship. This means that families of RVs can be partitioned into groups of equivalence classes; in each class any two members are equal almost everywhere P. It is important to be aware of this property when attempting to operate on measure spaces that are subspaces of others, as the equivalence classes tend not to be identical since these subspaces do not share all the same sets of measure zero.

2.3 Expectation

With the preliminaries now established, the stage is set for us to define expectation, which we do in terms of the Lebesgue integral. We do this for two reasons: To ensure closure of the various function spaces, in particular Banach and Hilbert spaces, a necessary evil in any function analytic setting; we note that this is not the case for the Riemann definition of integration. Secondly, Lebesgue integration provides a uniform integral regardless of whether or not a given RV is discrete or continuous, that is whether or not the RV takes range values that are countable in R^1. Of course we can only define this for measurable functions. When the integrands are RVs and the measures probability measures, the Lebesgue integral defines expectation. We note here alternate definitions of the integral obtained by extending continuous linear functionals defined on the range of these functions [8, 37]. This definition is well-adapted to integrals of vector-valued functions and we will used it when we introduce stochastic processes.

We outline the classical approach to defining the integral. We begin with a definition for non-negative simple functions; then extend the definition to include more

general non-negative measurable functions; and, finally, extend it a final time to allow any real-valued measurable function in the integrand.

Simple functions are those functions that can only take a finite number of values in their range space. Consider the so-called indicator, or characteristic, function is defined as follows

$$I_A(x) = \begin{cases} 1 & , \quad x \in A \\ 0 & , \quad \text{Otherwise} \end{cases} . \tag{10}$$

Whenever A is measurable, it is easy to see that I_A can be shown to be a measurable mapping.

We can now define simple functions in terms of indicator functions. If $\alpha_1, \ldots, \alpha_n$ are the distinct values in the range of s, and $A_i = \{x : s(x) = \alpha_i\}$, then

$$s(x) = \sum_{i=1}^{n} \alpha_i I_{A_i}(x) \tag{11}$$

where $I_{A_i}(x)$ is as defined in Eq (10). Note that it is always possible to construct monotonically increasing simple functions that converge to any positive measurable function.

Now, let s be a simple function and $C \in \mathcal{S}$, then the integral of s is defined by

$$\int_C s dP = \sum_{i=1}^{n} \alpha_i P(A_i \cap C). \tag{12}$$

For X a positive measurable set function with domain Ω, and $C \in \mathcal{S}$, the integral is defined as

$$\int_C X dP = \sup \int_C s dP \tag{13}$$

with the supremum being taken over all simple functions, s, such that $0 \leq s \leq X$. We note that the operation of taking a supremum can also be shown to be measurable.

For general measurable X, one first decomposes X into its positive and negative parts, $X^+ = \max\{X, 0\}$ and $X^- = \max\{-X, 0\}$, so that $X = X^+ - X^-$ and defines the integral as

$$\int_C X dP = \int_C X^+ dP - \int_C X^- dP. \tag{14}$$

So, as we mentioned, the expectation of a RV is defined in terms of the Lebesgue integral on the probability measure, P. Specifically,

$$E[X] = \int_\Omega X dP, \tag{15}$$

Eq (15) and, more generally, for measurable functions, g, of X by

$$E[g(X)] = \int_\Omega g(X) dP. \tag{16}$$

We've mentioned two advantages of Lebesgue integration over Riemann integration; these result from the behavior of the former under limit operations. We note that under the more restrictive Riemann integration criteria, the two integrals coincide.

Another artifact of Lebesgue integration is the so-called Change of Variables Theorem, which is motivated by the desire to cast the integration problem onto $\boldsymbol{R}^1$ rather than the more abstract underlying probability space. We state the result for the simplest form here; the interested reader can find the more general result in [4].

Theorem 1 (Change of Variable)**.** *Given a RV X and a measurable mapping g from $\boldsymbol{R}^1$ to $\boldsymbol{R}^1$, then*

$$\int_{\Omega} g(X)dP = \int_{-\infty}^{\infty} g(t)d\mu, \tag{17}$$

where $\mu = dist_X$ is the distribution of X. That is, the integration has been shifted from the probability space to the corresponding induced measure space on real line.

The theorem states that random variables can be integrated as "usual" functions on the real line. There is no technical reason why the range measure space must be defined on R^1, however.

Another noteworthy entity identified with $\mu = \text{dist}_X$ is the characteristic functional of the random variable X, defined as

$$\hat{\mu}(t) = E\left[e^{iXt}\right] = \int_{-\infty}^{\infty} e^{ixt} d\mu(x) \ . \tag{18}$$

One can readily observe that the characteristic functional is the Fourier transform of a positive function. It therefore has a number of very useful properties among which: (1) it uniquely determines the distribution function of X and (2) pointwise convergence follows from weak convergence of the corresponding distributions of approximations to X.

2.4 Spaces of Functions

In functional analysis, we consider each function to be one element in a set of functions. Sets of functions are often defined by function properties. For example, we define the integrable functions on a given measure space, L_1, to be the set of functions for which the integral of each member function is finite, and the set of square integrable functions on a measure space, L_2 to be the collection of those functions whose square is integrable. These spaces need not be defined through integration; for example, we could consider those functions that are continuous on the R^1, and designate them by C_0, or those functions with continuous first derivatives, C_1, etc.

Often, all of the members of a set of functions will have properties that can be exploited to infer various other desirable properties. For example, a Banach space is linear function space that is a complete, normed space.

An important subclass of Banach space is one for which we can define an inner product. These spaces are called Hilbert spaces, and they're important in analysis since, although they are infinite dimensional, they extend several of the concepts of Euclidian geometry in an intuitive way. The space L_2 cited above is an example of a Hilbert space. We will have more on this space later. We should note here that an inner product space can be defined without the requirement for completeness, and thus need not be a Banach space. This is the structure of the so-called pre-Hilbert space which when completed with respect to the norm induced by the inner product, becomes a Hilbert space and hence a Banach space.

In Section 2.2, we described the concept of almost everywhere (a.e.) P equality for RVs. Of course this concept also holds true for general measurable functions defined on a measure space. This means that function spaces can be thought of as spaces comprised of subgroups of member functions where each subgroup contains functions that differ only on a set of measure zero relative to the measure on the domain. In fact, equality a.e. μ, μ a measure on the range space, is known as an equivalence relation, and each of the subgroups of functions described above represent an equivalence class. Measure and integration of any member of a given equivalence class yields an identical result, so this fine point is often not discussed in functional analysis. However, one must be aware that work with subspaces of function spaces will often be complicated by the fact that the corresponding equivalence classes aren't identical due to discrepancies in the collections of sets of measure zero.

There are other forms of equivalence, such as equivalence in distribution, or mean square, etc, where the distance measures, and the function topologies they induce, are clearly defined. One should always be aware of the fact that, in general, uniqueness properties of these functions applies to the equivalence classes, and not to individual member functions themselves.

The relationship between RVs and measurable functions is the prime motive for exploring a function analytic approach to probability. We give a specific example of this in the following section.

2.5 Putting Things Together, Part One: Scalar Polynomial Chaos Expansions In $L_2(\Omega)$

As we mentioned in Section 1.1.3 above, the polynomial chaos expansion (PCE) method was first conceived by N. Wiener as a means to integrate differential equation-type operators where differential Brownian motion, which at the time was viewed as chaotic, was an external forcing influence. While the random process he described

is quite general, for this early exposition we will take a simpler route, and note that the transition to the more general case, first progressing to vectors containing RVs as components, then to more general random processes, is possible. We will address the particulars of this progression later in this text.

Now, consider two real-valued, scalar, measurable RVs, X and Y, each with finite variance. Assume that there exists a functional transformation, T, between X and Y; that is, that $X = T(Y)$ is well-defined.

Since we've constrained X to have finite variance by assumption, it is well known that it is a member of the class of square-integrable functions on its domain, the sample space; thus $X \in L_2(\Omega)$. It is also known that members of L_2 constitute a Hilbert space [44] of functions. Specifically, L_2 is a space of functions that comprise a complete normed inner-product space. As one might expect, there are a number of mathematical properties that this membership entails. Most important to our purposes is that we know that it is possible to construct a generalized Fourier expansion [29] in $L_2(\Omega)$ to approximate X, and that this can be done to arbitrary accuracy [50].

For the case of PCEs, and for reasons that we'll discuss shortly, we choose to construct this generalized Fourier expansion in terms of a set of orthogonal polynomials in a standard normal RV, ξ, such that $Y = \xi$; we denote these polynomials as $\Gamma_i(\xi)$. The mathematical machinery that we've described above ensures that we can construct approximations, $T^{(n)}$, such that the norm of the difference $\left\|T - T^{(n)}\right\|$, consistent with L_2 approximation for the target RV, can be made arbitrarily small.

The Γ_i are referred to as Hermite polynomials, and their properties can be found in any of a number of references documenting orthogonal polynomials such as [1]. In one dimension, the following formula yields the one-dimensional polynomials

$$\Gamma_i(\xi) = \sum_{j=0}^{[i/2]} (-1)^j \frac{i!}{(i-2j)!j!2^j} \xi^{i-2j} \tag{19}$$

where the expression $[r]$ evaluates to the largest integer less than or equal to r and $0! = 1$. For example, the first four of these are

$$\begin{aligned} \Gamma_0(\xi) &= 1 \\ \Gamma_1(\xi) &= \xi \\ \Gamma_2(\xi) &= \xi^2 - 1 \\ \Gamma_3(\xi) &= \xi^3 - 3\xi. \end{aligned} \tag{20}$$

Many of these so-called complete orthogonal polynomial systems derive from particular ordinary differential equations defined on Hilbert spaces [55] known as Sturm-Liouville systems. These systems induce an associated inner product weighting function that defines their orthogonality properties. For the Hermite polynomials this

weighting function is, to within a constant, identical to the probability density function of a standard normal RV. This explains our choice of polynomial system in the preceding paragraphs. Now, with these specifics in hand, we can exploit the orthogonality and other properties of the inner product to build our PCE-based approximations:

$$X^{(n)} = T^{(n)}(\xi) = \sum_{i=0}^{n} g_i \Gamma_i(\xi), \tag{21}$$

where the generalized Fourier coefficients,

$$g_i = \frac{E[X\Gamma_i(\xi)]}{\Gamma_i^2(\xi)}. \tag{22}$$

Perhaps not surprisingly, there are a number of other known probability density functions that also can be identified, simply by inspection, as weighting functions for inner products from different Sturm-Liouville systems. For example, an exponentially-distributed RV defined on the interval $[0, \infty)$, is affiliated with the Laguerre polynomials. Note that none of the underlying theory, nor the operations necessary for constructing a generalized Fourier expansion, rely on a particular RV/orthogonal polynomial pairing. Thus, theoretically, it is possible that PCEs can be generalized to any appropriate pairing. These generalizations, often termed Askey Expansions in the literature [3], are the subject of active research [53, 54].

Less obvious is the recent use of non-orthogonal, but still complete, RV/function pairings as the expansion bases for a given analysis [56]. Researchers are actively pursuing such approaches to achieve explicit segregation in analysis domains; these are typically obviated by the need to confine analyses to stable solution regimes. Here again, we emphasize the flexibility of the function analytic approach as an attractive feature for accommodating specific classes of application.

2.6 A Simple One-Dimensional Example

In the previous section, we outlined a Hilbert space formulation to building approximations for RVs. Specifically, we offered Eq (21) as an approximation in L_2 for a RV X. The issue at hand is to demonstrate how to compute the g_i given by Eq (22).

Here, we present a simple, one-dimensional example to illustrate how one might use the method to achieve an approximate result in a problem setting. Consider a cantilevered beam, 1 m in length, width equal to 1 cm and height equal to 2 cm, and composed of 6061-T6 aluminum . Let the beam be loaded by a 100 N transverse point load at its free end. The formula for the deflection at the free end of the beam is given by $PL^3/(3EI)$. We assume that the Young's modulus, E, is a Gaussian-distributed random variable of mean equal to 69 GPa and standard deviation equal to 6.9 GPa. We also assumed that the applied load is a Gaussian distributed with mean 100 N

and standard deviation 5 N, and that the length of the beam itself is Gaussian with mean 1m and standard deviation 1cm.

One-dimensional fourth order PC expansions for P, L and E obtained by using the Rosenblatt transformation, specifically,

$$E = P_E^{-1}\Phi(\xi_1), \quad L = P_L^{-1}\Phi(\xi_2), \quad P = P_P^{-1}\Phi(\xi_3). \tag{23}$$

This results in a 3-dimensional fourth order PC expansion for the deflection δ Here, P_E, P_L and P_P are the distributions of E, L and P, respectively, and are given each by an independent truncated Gaussian distribution [23].

Each of the chaos coefficients is evaluated using equation (22) where the expected value in the numerator has been approximated by a quadrature formula based on Monte Carlo sampling. Thus $N_q = 1000$ realizations of the variables ξ_i, $(i = 1, 2, 3)$ are sampled from a Gaussian distribution, for each one of them, the polynomials, $\Gamma_i(\xi)$ are evaluated, so are realizations of the variables E, L, and P, using equation (23). With these sets of realizations, the integrals in the numerator of equation (22) are evaluated for each of E, P, and L. Once the polynomial chaos expansions for these random variables have been constructed, realizations of these variables can be obtained from the resulting representation in order to construct a smooth approximation of the density function. These realizations are computationally inexpensive as they do not involve the solution of any physics problem, but rather sampling from a response surface.

The figures below show the results from a Monte Carlo analysis (1000 samples) and a histogram sampled from a PCE expansion representation of the solution. In order to provide a consistent comparison, the histogram generated by sampling from the chaos expansion has been constrained to have the same number of bins as the histogram obtained from a Monte Carlo analysis of the problem. A two-sample Kolmogorov-Smirnov goodness-of-fit two-sided hypothesis test at $\alpha = 1\%$ significance level was carried out for these two histograms. The conclusion is that the null hypothesis, to the effect that both PC and MC samples are drawn from the same distribution, should not be rejected at the 1% significance level.

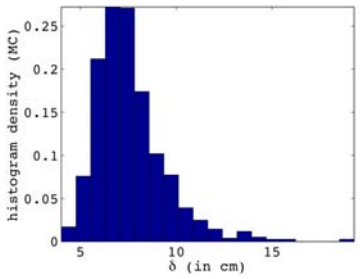

Figure 1. Results from Monte Carlo Analysis

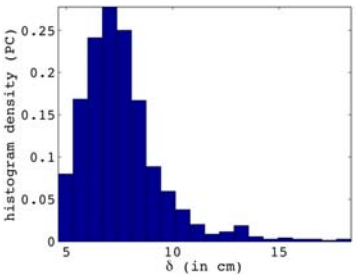

Figure 2. Results from a One-Dimensional Polynomial Chaos Representation.

2.7 More on the Hilbert Space Approach for RVs

We wrap up our introduction of the functional analytic approach to probability using our simple PCE description with a brief discussion of some relevant issues. Some of these are resolved; some, like the Askey Expansions discussed above, are the subject of current research.

First, implicit in the above discussion, is the assumption that we can justify the necessary changes of variables. For example, for PCEs, we assume that the transformation from a standard normal RV to an arbitrarily distributed RV exists; among other things, this connection allows us to evaluate the numerator of Eq 22. Our rationale for presuming that this is possible is based upon the following theorem, which is due to Paul Lévy [42, 7, 43]:

Theorem 2. *Let U be a uniform RV over $[0,1]$, and let $F : \boldsymbol{R}^1 \rightarrow [0,1]$ be any cumulative distribution function (CDF). Define Ψ as*

$$\Psi(u) = \inf_{x \in \boldsymbol{R}^1} \{x : F(x) \geq u\} \quad \forall u \in (0,1).$$

Then $Z = \Psi(U)$ is a RV whose CDF is F.

Again, specific to the case for PCEs one uses the uniformly distributed RV as the intermediary between the original and standard normal RVs.

Second is the issue of computational feasibility. A large number of underlying random variables, which is labeled as the stochastic dimension, can require one to compute a large number of Fourier coefficients to achieve a given approximation accuracy. In fact, the growth can become factorial in nature.

Finally, taking a function analytic approach in Hilbert Space has many beneficial features that deserve mention; there are also some caveats, to be aware of. We briefly describe several of these here.

- With each n, the expansion $T^{(n)}$ has a probability distribution that is implicitly defined. For PCEs, samples from this distribution, or realizations, are almost trivial to generate since they are known functional transformations of a standard normal RV.

- While the convergence of a given expansion is guaranteed in the appropriate space, one should be aware that transformations among absolutely continuous random variables, say, are only guaranteed to make sense in distribution [5]. This property, means that it is possible for more than one random variable to be defined by a given distribution. Thus, special care must be taken when reducing information, such as statistical moments from an expansion result, to ensure that information that is only indirectly related to a distribution is utilized appropriately.

- The use of generalized Fourier expansions, and PCEs specifically, in Hilbert Spaces can accommodate extension to random fields, as Wiener demonstrated many years ago. However, there are constraints that must be satisfied. For our case, we are considering the function spaces to be separable Hilbert spaces, and, further, we assume that the function spaces satisfy the Minlos-Sazonov [49] conditions that are the criteria for guaranteeing the existence of the continuum measure that our finite-dimensional approximations are presumed to converge to.

- Since it is possible to generate large number of transformations, $T^{(n)}$, and to use them to construct tests against experimentally-derived statistical constraints, it is possible to use the approach as a tool for addressing epistemic uncertainty. Successful applications of this have been performed for random variables, vectors, and fields [41, 18].

- Finally, a consequence approach is that the mappings are subject to the same approximation criteria as one might encounter, say, in a deterministic finite element analysis, only in this case the distance metrics will have to be generalized to allow for the use of probability measures in associated norms defining the approximations. Using this approach, the deterministic solutions are fully recovered when the probabilistic subdomains are eliminated from the analysis specifications.

2.8 Measures on Product Spaces

We have stated previously that the key idea is that the uncertainty sources are assumed to be fully constrained to the probabilistic subdomains, and that the associated mappings themselves are deterministic. This suggests that our approach will be to define random fields as (constrained) mappings on domains that are extensions of deterministic domains to those that allow a mechanism for including probabilistic characteristics. We provide the details to this approach here.

2.8.1 Direct Products of Sets and σ-Algebras

Let S and T be sets. The collection of all ordered pairs of the form (s,t), where $s \in S$ and $t \in T$, is the Cartesian product $S \times T$. For sets $A \subset S$ and $B \subset T$ the set $A \times B$ is a subset of $S \times T$ called a generalized rectangle with sides A and B.

If $\mathcal{S}$ and $\mathcal{T}$ are σ-algebras of subsets of S and T, respectively, then a measurable rectangle is a set of the form $A \times B$, where $A \in \mathcal{S}$ and $B \in \mathcal{T}$.

An elementary set is defined by $Q = C_1 \cup \cdots \cup C_n$, where each of the C_i are measurable rectangles and $C_i \cap C_j = \emptyset$ for all $i \neq j$.

Finally, we define $\mathcal{S} \times \mathcal{T}$ to be smallest σ-algebra of subsets of $S \times T$ containing every measurable rectangle. It is also the smallest σ-algebra containing the class of all elementary sets.

2.8.2 Product Measure

Consider two measure triples, $(S, \mathcal{S}, \mu)$ and $(T, \mathcal{T}, \theta)$. We define the product measure triple to be $(S \times T, \mathcal{S} \times \mathcal{T}, \mu \times \theta)$, where $\mu \times \theta$, the product measure, is defined on the σ-algebra, $\mathcal{S} \times \mathcal{T}$, that we described in the previous section.

When the constituent σ-algebras are both of the Borel category, then the product measure can be given as

$$\mu \times \theta(A \times B) = \mu(A)\theta(B) \tag{24}$$

where $\mathcal{S}$ and $\mathcal{T}$ are the Borel σ-algebras, $A \in \mathcal{S}$, and $B \in \mathcal{T}$.

Product measures can require σ-algebras that are larger than the collections of sets that are assembled simply by taking combinations of sets from their constituent σ-algebras. More precisely, even if both $(S, \mathcal{S}, \mu)$ and $(T, \mathcal{T}, \theta)$ are complete measure spaces, it is not the case that $(S \times T, \mathcal{S} \times \mathcal{T}, \mu \times \theta)$ will be complete. This is an artifact of the process necessary for completion of measures that was discussed in Section 2.1, and entails expanding $\mathcal{S} \times \mathcal{T}$ to include all subsets of sets of measure zero in $\mathcal{S} \times \mathcal{T}$ that can be constructed through the product of sets of measure zero in the constituents. Since the sets included in the larger σ-algebra, which we will denote as M, differ from Borel sets only by a subset of measure zero, they can be replaced, for the purpose of the measure, by a bracketing Borel set; this latter set is guaranteed to be measurable relative to the constituent measures. To take care of this process automatically, we have to generalize our definition of product measure, Eq (24), for complete product spaces to be

$$\mu \times \theta(Q) = \int_S d\theta \int_T I_Q(s,t) d\mu = \int_T d\mu \int_S I_Q(s,t) d\theta \tag{25}$$

where $Q \in (\mathcal{S} \times \mathcal{T})^*$ and $I_Q(s,t)$ is the indicator function defined in Eq (10).

Our construction, with the constraints that have been assumed on the constituent measure spaces, results in a product measure space that is σ-finite.

2.8.3 Definition of Measurability

Now we consider function mappings on product spaces.

Scalar, real-valued functions with domain $S \times T$ are termed measurable if inverse mappings of Borel, or where applicable, Lebesgue, sets are elements of $\mathcal{S} \times \mathcal{T}$, or $(\mathcal{S} \times \mathcal{T})^*$, respectively.

2.8.4 Integration on Product Space

Our primary motive for our function analytic approach to probability is to expand the domains of our deterministic functions to product space mappings with one sub-domain a probabilistic sample space, to expand the notion of measure to product measure, and to define integration for these functions relative to product measure. Finally, we will use these building blocks as means to define norms and inner products for these new functions.

We state the following important theorem without proof:

Theorem 3 (Fubini)**.** *Let $(S, \mathcal{S}, \mu)$ and $(T, \mathcal{T}, \theta)$ be σ-finite measure spaces, and let f be an $(\mathcal{S} \times \mathcal{T})$-measurable real-valued function on $S \times T$. If either:*

(a) $f \geq 0$, or

(b)

$$\int_s d\mu \left(\int_T |f| d\theta \right) < \infty \tag{26}$$

then

$$\int_{S \times T} f d(\mu \times \theta) = \int_s d\mu \left(\int_T f d\theta \right) = \int_T d\theta \left(\int_S f d\mu \right). \tag{27}$$

When the product space has been completed, which means, in our notation, that f is assumed to be $(\mathcal{S} \times \mathcal{T})^*$-measurable, then we understand the inner integrals in Eq (27) to only be defined almost everywhere relative to the outer integration measure. That is, $\int_T f d\theta$ is understood to be defined a.e.-μ, and similarly for $\int_S f d\mu$.

2.9 Random Fields

With this in mind, we define scalar, real-valued random fields (RFs) to be measurable functions with one of the subdomains assumed to be a probability space in the sense defined above. Specifically,

$$\alpha : D \times \Omega \rightarrow \boldsymbol{R}^1 \tag{28}$$

where, for the sake of discussion, we assume D to be a deterministic subdomain, and Ω to be a probabilistic sample space. Finally, we assume that α is measurable with respect to the measure triple associated with the domain, $(D \times \Omega, \mathcal{D} \times \mathcal{S}, \mu \times P)$

A common statistical entity of interest is the correlation function defined as

$$R(d_1, d_2) = E[\alpha(d_1, \omega)\alpha(d_2, \omega)] \tag{29}$$

where measurability of α ensures that the entity on the right hand side of the equation is well-defined.

This construction of random fields does not take into consideration the particular structure of the functional space in which each realization of the process claims membership, be it the space of square-integrable functions of a Sobolev space, for instance. An example of the significance of these functional space constraints on the efficiency of canonical approximation is well documented [30, 13]. This product measure definition forms the building block for our functional analysis construction and will be further refined, below, to account for additional more specific character of the functions being described by the random process.

As was the case for single constituent mappings, the product space mappings that we have just described can themselves be thought of as existing in the framework of a collection. For example, it is possible to consider the collection of all such (real-valued) mappings, say, $\mathcal{F}$, that map our product space domain into the real line. The next section details an analysis of random functions as elements of these function spaces.

3 Analysis of Random Processes

As we can see from the previous section, RFs can be thought of as generalizations either to deterministic functions, or to RVs. Fortunately, a function analytic approach provides a consistent framework in which to handle this generalization.

In many problems in science and engineering, realizations of RFs, $\alpha_{\omega_j}(d) = \alpha(d, \omega_j), \omega_j \in \Omega$, are constrained to membership in a well-structured functional space. Fortunately, the restriction of a probability measure from the space of all functions to specific functional spaces can also be readily carried out, and the machinery discussed above for constructing measures on product spaces remains valid.

Also, as with RVs, the construction of measures for RFs is quite complicated; it is further complicated by the fact that the so-called indexing set, D in Eq (28), is often an uncountable set. One common approach is to construct this measure from finite-dimensional measures of RVs achieved by considering finite collections of RVs, $\{\alpha_{d_i}(\omega) = \alpha(d_i, \omega), d_i \in D, i = 1, \cdots, n\}$, that one can compute or observe from the RF.

In this section, we build on our development in Section 2 by taking advantage of the structure on the range space of these random variables. A number of references already exist to guide us in this endeavor [17, 36, 21, 40].

Before we begin the analysis, we will provide a succinct description of duality, a concept of fundamental importance and that is particularly useful in our approach to functional analytic representations. Most physical quantities appear in pairs. From the point of view of one item in this pair, the second item, its dual pair, is a quantity that can help it achieve its value. Thus strain is the dual of stress (achieving energy), velocity is the dual of force (achieving power), price is the dual of a commodity (achieving value). We can thus talk about a certain space of objects and its dual. The importance of the duality concept in functional analysis relies chiefly in the possibility of relating properties of representations in one space to those in its dual. The algebraic dual of a linear vector space is defined as the set of all linear functionals defined on the space. We will be mostly interested in the topological dual of a given linear vector space. This is the set of all continuous linear functionals defined on the original space. Given the continuity of these functionals, a topology is naturally induced in the dual space, and the concept of neighborhoods and proximity thus follows naturally. In addition to the topology induced in the dual space itself, these continuous linear functionals induce a topology in the original space where neighborhoods are defined in terms of their images under the duality pairing. This is the so-called weak topology in the original space and is weaker, hence the name, than the original topology. We note that the dual of a normed space is a Banach space, and that the dual of a Hilbert space is itself a Hilbert which may be identified with the original space.

A most significant concept in the functional analytic description of probability

theory, is the ability to further abstract the concept of integration. Specifically, the Lebesgue integral has been generalized to vector-valued functions along two distinct paths [11]. The weak integral of a random variable with values in a topological vector space B can be defined as the regular integral of the numerical function $\langle f, X\rangle$, for f in the dual space of B, and where $\langle , \rangle$ denotes duality pairing. Alternatively, the strong integral, is defined as a natural extension of the Lebesgue integral with convergence interpreted as being with respect to the norm in B.

The characteristic functional and the covariance operator of a stochastic process turn out to be key for a functional analysis approach to RFs. We construct each using the topological dual of the space in which the RF assumes its values; we will refer to this as their phase space. The properties of the characteristic functional and the covariance operator will therefore reflect the functional analytic structure of their phase space. While the covariance operator provides a more standard machinery for analysis in function space, the characteristic functional enables the definition of a probability measure on spaces with a milder structure, and in the case of a Hilbert space, it provides the means for characterizing the measures on subspaces obtained through orthogonal projections. In what follows, we will describe the evolution of the probabilistic structure of a RF as the functional structure of its phase space changes; we will start by assuming that it is a metric space, and end with the assumption that it is a Hilbert space.

3.1 Generalizing the Concept of Random Variable

In Section 2.1, we described the Borel σ-algebra on the real line in terms of its topological origins, and at that time also mentioned that this could be generalized to incorporate other topological spaces. We discuss this here.

Consider the probability space $(\Omega, \mathcal{S}, P)$ and the measurable space $(H, \mathcal{H})$, where by measurable space, we mean a set/σ-algebra pairing absent a measure. We will call an $(\mathcal{S}, \mathcal{H})$-measurable mapping from Ω into H an H-valued random variable. Also, we denote a realization of the random variable X at any $\omega \in \Omega$, by $X(\omega)$. In Section 2.2, we defined the distribution of a real-valued RV; we generalize this now for the case of an H-valued RV. The $(\mathcal{S}, \mathcal{H})$-measurability of this generalized random variable guarantees that the probability of X belonging to a subset $A \in \mathcal{H}$ is is well-defined. This is given as

$$\text{dist}_X \equiv P(X \in A) = P(X^{-1}(A)) \quad A \in \mathcal{H}, \tag{30}$$

where $X^{-1}(A)$ is in $\mathcal{S}$. The probability measure, $\eta = \text{dist}_X$, defined on $(H, \mathcal{H})$, yields a well-defined measure space; specifically a probability triple, $(H, \mathcal{H}, \eta)$, which is induced by the random variables X on the space of its values.

Significantly, the inverse mapping of all $A \in \mathcal{A}$ is generally only a subset of $\mathcal{S}$, which we denote by $\mathcal{S}_X$, and is called the σ-algebra induced in Ω by the random variable X. This property also means that the topology of open sets associated with

$\mathcal{S}_X$ is weaker than that of $\mathcal{S}$.

If we then consider collections of random variables, $\{X_i\}$, each with values in the measurable space $(G_i, \mathcal{G}_i)$, they each induce a σ-algebra $\mathcal{S}_{X_i}$ in Ω with a similar property. Taking the union of all of these induced σ-algebras yields a σ-algebra that is induced by that family. Note that this σ-algebra is still only a subset of $\mathcal{S}$. Properties of such families are, in fact, precisely what we will exploit in our development.

We will denote by $\mathcal{M}(H)$ the space of all probability measures or distributions on H. It can be shown that $\mathcal{M}(H)$ can be metrized as a separable metric space if and only if H is a separable metric space [36]. A weak topology is usually defined on $\mathcal{M}(H)$, which is of some significance in statistical inference problems as it is required for investigating the proximity between two elements of $\mathcal{M}(H)$.

Clearly, the structure of the space H, be it a topological vector space, a metric space, a Banach space, or a Hilbert space, etc, will have significant ramifications on the properties of the mappings X and the associated measure η_x. We will next explore how the properties of this measure evolve as the functional analysis structure of H is gradually enriched. The exposition is meant to be accurate but brief and is far from exhaustive. The interested reader is referred to the references.

3.1.1 Measure on Metric Spaces

As above, let H be a metric space and $\mathcal{H}$ the Borel σ-algebra on H induced by its metric. Then any measure on H is regular in the sense that the measure of any Borel subset of H can be approached as the limit from below and above using, respectively, a sequence of closed and open sets.

Let $C(H)$ denote the space of all bounded real-valued continuous functions on H. Under the supremum norm, $C(H)$ becomes a Banach space [44]. It can be shown that if H is compact then any positive linear functional on $C(H)$ can be identified, uniquely, with the expectation operator relative to some probability measure on $\mathcal{H}$ [36]. If H is not compact, the expectation operator is merely relative to a finitely additive regular measure.

By endowing H with a separable metric group structure, the space of measures becomes a topological semigroup under the binary operation of convolution. Thus on a metric space, the concepts of topological completeness and separability are significant. It is should be noted here that a semi-group is a set over which an associative closed operation has been defined. A group is a semi-group in which an identity element exists and where each element has an inverse defined with respect to the group operation [24].

3.1.2 Measure on Topological Groups

A locally compact abelian group is an abelian group which is given the structure of a locally compact topological space compatible with the group operation [38]. That is, the group operation is continuous. Locally compact groups have a natural measure that is invariant under under group operation, the so-called Haar measure, thus paving the way for defining the integral on members of the group. A space is said to be second countable if its topology has a countable base. Every locally compact group that is second countable is separable, metrizable as a topological group and complete [38]. Moreover, these groups have a metric that is invariant under the group operation (i.e. translation invariant).

By endowing H further with the structure of a *locally compact second countable abelian group*, it becomes topologically complete and the space $\mathcal{M}(H)$ becomes a topological semigroup. With this structure, Fourier analysis on the set of measures is possible and infinitely divisible distributions can be introduced [24, 22]. Any such distribution can be represented as the composition of a Gaussian and a Poisson distributions.

Letting the character group associated with H be denoted by G, and the value of the character g at $h \in H$ by $\langle h, g\rangle$, the characteristic function, $\hat{\mu}(g)$ for each $\mu \in \mathcal{M}(H)$ can be defined as

$$\hat{\mu}(g) = \int_H \langle h, g\rangle d\mu(h) \ , \tag{31}$$

for all continuous $g \in G$. The notation $\langle h, g\rangle$ refers to duality pairing, which can be written as

$$\langle h, g\rangle = e^{il(h,g)} \tag{32}$$

for some continuous mapping l: $H \times G \mapsto R$. It can be shown that a function ϕ defined on G is the characteristic function of a measure $\mu \in \mathcal{M}(H)$ if and only if 1) it is equal to 1 at the zero element of H, 2) it is continuous, and 3) it is positive definite. Clearly these properties of the characteristic function carry over as additional structure is bestowed on the space H. Moreover, with this topological group structure, Gaussian measure can be defined on H.

A generalization of characteristic functionals, defined for one-dimension in Eq (18), was introduced on n-dimensional Euclidian space [6] and was subsequently furhter generalized to other functional spaces, namely Hilbert and Banach spaces, and topological groups [46, 39, 47].

3.1.3 Measure in Weakly Complete Topological Linear Space

Until now, the characterization of H-valued random variables has been accomplished in terms of the topological structure of the space H. This has led us to expressions of the corresponding characteristic functional, through which we then obtained a characterization of the measure of a stochastic process.

By adding a linear space structure, the characteristic functional introduced previously is still a valuable tool, defined in an analogous manner, with the character group replaced by the topological dual. With this topological linear structure, however, additional analysis becomes possible. To explore this new structure, we let H be a topological linear space with the weak topology determined by the linear functionals. Assume that the space H is weakly complete, and denote its dual by H'. The element $m_x \in H$ obtained as the weak integral with respect to the measure η_x of $X \in H$, is called the expectation of the random variable X.

Canonical Expansions: Generalized Karhunen-Loève Expansion Next consider a zero-mean random variable X and the H-valued random variable $X\overline{\langle f, X\rangle}$ where $f \in H'$ and an overbar denotes complex conjugation. The expectation of this random variable,

$$S_X f \equiv E\left\{X\overline{\langle f, X\rangle}\right\} = \int_\Omega X(\omega)\overline{\langle f, X(\omega)\rangle} dP(\omega) \tag{33}$$

defines an operator $S_X f$ of random variable X, from a subspace of H' into H. This is the covariance of random variable X. Similarly, the cross-covariance operator of two random variables, $X \in H_1$ and $Y \in H_2$ is defined as the mapping from H_2' into H_1 given by

$$S_{XY}(f) \equiv E\left\{X\overline{\langle f, Y\rangle}\right\} = \int_\Omega X(\omega)\overline{\langle f, Y(\omega)\rangle} dP(\omega)\ . \tag{34}$$

The covariance operator is symmetric and positive definite and we will use it to define a scalar product.

While this definition is quite distinct from the usual definition of a covariance of a random variable, it reduces to the standard definition for the well-known cases. For instance, if X is an element in $\boldsymbol{R}^n$, then the element f of its dual space can be represented as the transpose of an $\boldsymbol{R}^n$ vector. The quantity $S_X f$ then becomes the matrix-vector product of the $n \times n$ covariance matrix of vector X with f. As another example, considering $X = X(t)$ with realizations in some linear vector space H, the action of an element f of its dual space on X can be represented as $\langle f, X\rangle = \sum_k f_k X(t_k)$. The quantity $S_X f$ can then be written as $S_X f = \sum_k \overline{f_k} E\left\{X(t)\overline{X}(t_k)\right\}$.

Let S_X, a linear operator from H' to H, denote the covariance operator of H-

valued random variable which can be shown to be continuous in the weak topology of the space H. Denoting by N_0 the null space of S_X, the factor space H'/N_0 defines equivalence classes on H'. In the sequel, while working in H'/N_0, reference will be made to H' for notational convenience. Next, we define on $H' \times H'$ the functional,

$$(f, g) = \langle f, S_X g \rangle \ . \tag{35}$$

This functional defines a scalar product which turns H' into a Hilbert space, which we denote by H_X, associated with random variable X. All functionals $f \in H'$ are continuous and hence the expression $V = \langle f, X \rangle$ determines an isometric mapping from H_X on a Hilbert space H_V, with scalar product given by

$$(U, V)_{H_V} = E\left\{U\bar{V}\right\} \ . \tag{36}$$

It can readily be shown that if $V = \langle f, X \rangle$ and $W = \langle g, X \rangle$ for $f, g \in H_X$, then

$$(V, W)_{H_V} = (f, g)_{H_X}, \tag{37}$$

where the subscript each inner product specifies the corresponding Hilbert space. Note in particular that for $V = \langle f, X \rangle$,

$$\|V\|_{H_V} = \|f\|_{H_X}. \tag{38}$$

Suppose for now that the space H_X is separable, and let f_ν denote any orthonormal basis in H_X. Then any element $f \in H_X$ can be represented as

$$f = \sum_{\nu=1}^{\infty} (f, f_\nu)_{H_X} f_\nu \ , \tag{39}$$

so that as $n \to \infty$,

$$\|f - \sum_{\nu=1}^{n} (f, f_\nu)_{H_X} f_\nu\|^2_{H_X} = E\left\{\langle f, X \rangle - \sum_{\nu=1}^{n} (f, f_\nu)_{H_X} \langle f_\nu, X \rangle\right\}^2 \to 0 \ , \tag{40}$$

where we have used the inner product to infer the norm in H_X, namely that

$$\|f\|_{H_X} = (f, f)_{H_X} = \langle f, S_X f \rangle \ . \tag{41}$$

We note that $(f, f_\nu)_{H_X} = \langle f, S_X f_\nu \rangle$, and denote $X_\nu = S_X f_\nu$ where $X_\nu \in H$. Then letting $V_\nu = \langle f_\nu, X \rangle$, results in

$$\langle f, X\rangle = \sum_{\nu=1}^{\infty} V_\nu \langle f, X_\nu\rangle, \quad \forall f \in H_X \ . \tag{42}$$

This results in the following expression for X,

$$X = \sum_{\nu=1}^{\infty} V_\nu X_\nu, \tag{43}$$

which is mean-square convergent to X in the weak topology of the space H. It also follows from equation (37) that,

$$E\{V_\nu V_\mu\} = \delta_{\nu\mu} \ , \tag{44}$$

where $\delta_{\nu\mu}$ denotes the kronecker delta function. Noting that $(f_\nu, f_\mu) = \langle f_\nu, S_X f_\mu\rangle = \langle f_\nu, X_\mu\rangle$, we deduce a biorthogonality relation between $\{f_\nu\}$ and $\{X_\nu\}$. Clearly, both the orthogonal random variables $\{V_\nu\}$ and the corresponding coordinate functions X_ν are dependent on the choice of the orthonormal basis set $\{f_\nu\}$.

It should be emphasized that the treatment in this section, is equally valid for Hilbert space-valued random variables. This is particularly useful in situations where the dual of the Hilbert space is not identified with the space itself, such as when dealing with Sobolev spaces. In this context, we note that the above treatment sets on firmer ground previous treatments of generalized Karhunen-Loève expansions [27, 48, 30, 20, 19] where the added structure in the Sobolev space was addressed either through regularization [27, 48, 30] or through an orthogonal expansion in the Sobolev space itself [20, 19]. The present treatment to this problem is more versatile, as it applies to any complete topological vector space, highlighting the role played by its dual.

3.1.4 Measure in Separable Hilbert Spaces

With the Hilbert space structure on the space H, the characteristic functional is written as,

$$\hat{\eta}(g) = \int_H e^{i\langle h, g\rangle} d\eta(h) \ . \tag{45}$$

where the bracket denotes the inner product, or more generally, duality pairing in H.

The Hilbert space structure brings additional tools to bear on the analysis of H-valued random variables. We are mainly interested in the canonical representation of these variables.

Nuclear Spaces But before we delve into defining a probability measure on a Hilbert space, let us introduce a particular space of functions, of significance in

stochastic analysis, the so-called nuclear space [17]. To set the stage for their definition, we start by defining a countably Hilbert space as a complete linear topological space in which the topology is given by a countable set of compatible norms that are derived from an associated set of scalar product. Thus if Φ denotes a linear space, and $(\phi, \phi)_k,\ k \in \mathcal{K}$ a system of associated scalar products, then $\|\phi\|_k = \sqrt{(\phi, \phi)_k}$ denotes a corresponding system of norms. Next we introduce in Φ the topology taking as the complete neighborhood basis of the origin the sets associated with the inequality $\|\phi\|_n \leq \epsilon$. If we then let Φ_k denote the completion of Φ with respect to $\|\phi\|_k$, it follows that $\Phi = \bigcap_{n=1}^{\infty} \Phi_n$. This last condition could also be used as the defining criterion for countably Hilbert spaces. It is worth noting that the compatibility requirement on the norms associated with a countable Hilbert space implies an ordering in these norms, so that if $m \leq n$ then $\|\phi\|_m \leq \|\phi\|_n$. We also note in passing that if Φ'_n denotes the dual of Φ_n, then the dual of Φ is given as $\Phi' = \bigcup_{n=1}^{\infty} \Phi'_n$.

We introduce one last ingredient before defining a nuclear space, namely the concept of a nuclear operator. Specifically, an operator T from a Hilbert space Φ_n into a Hilbert space Φ_m is said to be nuclear if it can be represented in the form,

$$T\phi = \sum_{k=1}^{\infty} \lambda_k (\phi, \phi_k) \psi_k \ , \tag{46}$$

where $\{\phi_k\}$ and $\{\psi_k\}$ are orthonormal systems in Φ_n and Φ_m, respectively, $\lambda_k > 0$ and $\sum_{k=1}^{\infty} \lambda_k$ converges. We note that this last requirement ensures that every nuclear operator is also of Hilbert-Schmidt type. We now define a nuclear space as a countable Hilbert space Φ for which the canonical embedding, T_m^n, of Φ_n into Φ_m is nuclear.

We note that similar definitions can be made for nuclear normed spaces and nuclear linear topological vector spaces.

A motivation for the significance of nuclear operators and countable Hilbert spaces in the analysis of stochastic processes can be attributed to the Kolmogorov consistency requirements. Interpreting these constraints in the context of an infinite-dimensional measure can be more readily carried out in terms of projections on subspaces. This association of nuclear operators with Kolmogorov consistency is best highlighted by the Minlos-Sazanov theorem [49]. A second motivation is associated with the observation that both the covariance operator and the characteristic functional of a stochastic process are nuclear operators.

Canonical Expansions: Karhunen-Loève Expansion The construction of the canonical expansion in the previous section did not require a Hilbert space structure on H. It merely used the Hilbert space structure induced on H' by the scalar product (35). If the space H is indeed a Hilbert space such that $E\{\|X\|^2\} < \infty,\ \ \forall X \in H$, then by Riesz theorem it could be identified with its own dual and consequently, the orthogonal basis in Eq (39) becomes a basis in H_X.

It should be noted, however, that the identification of H with its own dual, while convenient for some problems, is not natural for many problems of great interest in engineering applications involving PDEs. These problems involve Sobolev spaces, the duals of which are constructed as distributions [35].

A particular set of basis functions is of singular interest when H is a Hilbert space. In this case, the covariance operator, S_X, is nuclear, self-adjoint, positive, and completely continuous, and therefore has a finite or countable set of eigenvalues, $\{\lambda_n\}$ and corresponding orthonormal eigenfunction, $\{\phi_n\}$, such that,

$$S_X\phi_n = \lambda_n\phi_n, \qquad (\phi_n, \phi_m) = \delta_{nm} \ . \tag{47}$$

An expansion in terms of the eigenfunctions of its covariance operator can thus be affected. This expansion is convergent almost surely and in mean-square. This is the standard Karhunen-Loève expansion for H-valued random variables.

It should be noted in this case, that the space H_X becomes a reproducing kernel Hilbert space (RKHS) [2, 34, 31, 32, 33, 45] associated with the vector-valued random variable X. Also in this case, the canonical expansion is bi-orthogonal, with both the set of random variables V_ν and coordinate functions X_ν orthogonal in their respective spaces.

3.1.5 Transition from Series of Random Elements to Stochastic Process: Consistent Measures

A particularly fine point in the theory of stochastic processes addresses the compatibility of finite-dimensional measures of these processes. Specifically, under what conditions is a collection of finite-dimensional distributions compatible with a distribution on the whole, possibly infinite-dimensional, space. These conditions can be written down in terms of the characteristic functional, which is conveniently expressed for random variables with values in Hilbert spaces. In a Hilbert space, the covariance operator can be redefined as

$$(SX, X) = \int (y, X)^2 \, d\eta(y) \ . \tag{48}$$

This covariance operator is nuclear and hermitian and can be used to construct the so-called S-topology [36].

Then conditions for the compatibility of a characteristic function in Hilbert space H can be stated as,

- The necessary and sufficient conditions for a function $\phi(y)$, $y \in H$ to be the characteristic function of a probability measure η, it should be equal to 1 at the origin; it should be positive definite, and it should be continuous in the S-topology.

If in addition, $\int \|x\|^2 d\eta < \infty$, then the conditions become,

- The necessary and sufficient conditions for a function $\phi(y)$, $y \in H$ to be the characteristic function of a probability measure η, it should be equal to 1 at the origin; it should be positive definite; it should be continuous in norm; and $\sum_{n=1}^{\infty} \int_{-\infty}^{+\infty} x^2 dF_n(x) < \infty$ where F_n is the measure on the real line defined in Eq (7) with characteristic function $\phi_n(t) = \phi(te_n)$ and $\{e_i\}$ a fixed orthonormal basis in H.

One of the few compatibility results in terms of the finite-dimensional distribution functions is available on the space $C[0,1]$ and can be stated as follows:

- Let $\{\xi_t : 0 \leq t \leq 1\}$ be a stochastic process and $P^{t_1,\cdots,t_k}$ the probability distribution in R^k of the vector $(\xi_{t_1}, \cdots, \xi_{t_k})$. If there are constants $\alpha, \delta, K > 0$ such that $E\,|\xi_{t_1} - \xi_{t_2}|^{\alpha} = \int \int_{R^2} |u - v|^{\alpha} dP^{t_1,t_2}(u,v) \leq K|t_1 - t_2|^{1+\delta}$, for all $t_1, t_2 \in [0,1]$, then there exists a unique measure η on C such that $P^{t_1,\cdots,t_k} = \eta^{t_1,\cdots,t_k}$ for all k and all $t_1, \cdots, t_k \in [0,1]$.

4 Putting Things Together, Part Two: Multi-Dimensional Polynomial Chaos Expansions In $L_2(\Omega)$

The canonical expansions developed in the previous section provide a discretization of random process in terms of a denumerable set of random variables, $\{V_\nu\}$, orthogonal with respect to the infinite-dimensional measure on the space $(H, \mathcal{H})$. The phase space character of the random process is inherited by the deterministic coordinate functions in the canonical expansion. As a corollary to the Kolmogorov consistency requirement on multi-dimensional distribution functions, this orthogonality can be construed to be in terms of the two-dimensional joint distribution between two random variables V_ν and V_μ. This consistency requirement also ensures that the finite-dimensional random variable $\{V_\nu\}_{\nu=1}^N$ can be legitimately characterized by their N-dimensional joint distribution function. We note at this point that, although orthogonal, these random variables are generally neither identically distributed nor statistically independent. Hence their joint measure cannot be characterized by the product of their respective one-dimensional marginal distributions. In addition to canonical representations of stochastic processes, random variables can appear in a given problem through the direct characterization of a physical phenomenon or model parameter. The collection of these variables, say n of them, can then be regarded as a vector-valued random variable with phase space in n-dimensional Euclidean space or, if they are independent, then they can be represented as n separate real-valued random variables.

Considering then n jointly-distributed random variables, or equivalently, a random variable, $\boldsymbol{V}$, with values in $\mathbb{R}^n$, we will next be concerned with canonical representations of functions, $\boldsymbol{f}(\boldsymbol{V})$, of these random variables. We encounter this situation in two broad classes of problems. First, the mapping in question may be explicitly provided by the physical constraints in the problem, such as conservation laws or geometrical constraints. Typically in this case, the random variable $\boldsymbol{X}$ provides a probabilistic representation of model parameters. Alternatively, in many other cases of interest, $\boldsymbol{Y} = \boldsymbol{f}(\boldsymbol{X})$ defines an equivalence between random variable $\boldsymbol{X}$ and random variable $\boldsymbol{Y}$, usually in the sense that their respective probability distributions are identified.

The first step in developing a canonical representation for $\boldsymbol{f}(\boldsymbol{X})$ consists in constructing a basis set in a suitable functional space. We first assume that the distribution function of random variable $\boldsymbol{X}$ is continuous with respect to Lebesgue measure, and using the Radon-Nikodym theorem, we can define the probability density function of random variable $\boldsymbol{X}$ as follows,

$$P_{\boldsymbol{X}}(d\boldsymbol{x}) = p_{\boldsymbol{X}}(\boldsymbol{x})\ d\boldsymbol{x}. \tag{49}$$

Let $p_{X_1}(x_1), \ldots, p_{X_n}(x_n)$ be the marginal probability density functions of order 1 given by

$$p_{X_k}(x_k) = \int_{\mathbb{R}^{n-1}} p_{\boldsymbol{X}}(x_1) \ldots, x_{k-1}, x_k, x_{k+1}, \ldots, x_n) \; dx_1 \ldots dx_{k-1} dx_{k+1} \ldots dx_n \; . \quad (50)$$

First we define $\mathbb{H}$, as

$$\mathbb{H} = L^2_{P_{\boldsymbol{X}}} \left(\mathbb{R}^n, \mathbb{R}\right) \; , \quad (51)$$

where $L^2_{P_{\boldsymbol{X}}}$ is the space of real-valued functions on $\mathbb{R}^n$ that are square-integrable with respect to the measure $P_{\boldsymbol{X}}$. We then introduce the real Hilbert space associated with marginal distribution $p_{X_k}(x_k) \; dx_k$,

$$\mathbb{H}_k = L^2_{P_{X_k}}(\mathbb{R}, \mathbb{R}) \quad (52)$$

equipped with the inner product

$$\langle r, s \rangle_{\mathbb{H}_k} = \int_{\mathbb{R}} r(q) \; s(q) \; p_{X_k}(q) \; dq = E\left\{r(X_k) \; s(X_k)\right\} \; . \quad (53)$$

Letting $\left\{\psi^k_\ell, \; \ell \in \mathbb{N}\right\}$ denote a Hilbertian basis of the real Hilbert space $\mathbb{H}_k$, it can be shown [51] that,

Lemma 1. *For all $\boldsymbol{x} = (x_1, \ldots, x_n)$ belonging to the support of $p_{\boldsymbol{X}}(\boldsymbol{x})$, Hilbertian basis $\{\phi_{\boldsymbol{\alpha}}, \quad \boldsymbol{\alpha} \in \mathbb{N}^m\}$ of real Hilbert space $\mathbb{H}$ is given by*

$$\phi_{\boldsymbol{\alpha}}(\boldsymbol{x}) = \left(\frac{p_{X_1}(x_1) \times \ldots \times p_{X_n}(x_n)}{p_{\boldsymbol{X}}(\boldsymbol{x})}\right)^{1/2} \psi^1_{\alpha_1}(x_1) \times \ldots \times \psi^n_{\alpha_n}(x_n) \; , \quad (54)$$

where $\{\psi^k_{\alpha_k}(x_k)\}_{\alpha_k}$ is a Hilbertian basis of real Hilbert space $\mathbb{H}_k$ and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m)$, $p_{\boldsymbol{X}}(\boldsymbol{x})$ is the joint probability density function of random variable $\boldsymbol{X}$ and p_{X_m} is the marginal density of the real random variable X_m.

We note that in case the components of random variable $\boldsymbol{X}$ are statistically independent, the basis functions with respect to the joint probability measure is obtained as the tensor product of the basis functions for all the one-dimensional marginal density functions.

Next we turn our attention back to the task of representing $\boldsymbol{Y} = \boldsymbol{f}(\boldsymbol{X})$ under condition that $\boldsymbol{Y} \in \mathbb{H}$. A representation of the solution is sought with respect to the Chaos basis in the form,

$$\boldsymbol{Y} = \sum_{\boldsymbol{\alpha} \in \mathbb{N}^n} \boldsymbol{Y}_{\boldsymbol{\alpha}} \; \phi_{\boldsymbol{\alpha}}(\boldsymbol{X}) \quad (55)$$

with the $\mathbb{R}^n$-valued coefficient $\boldsymbol{Y}_{\boldsymbol{\alpha}}$ which can be estimated using an identical procedure to that constructed in section 2.5.

5 Statistical Inference

A most interesting application of the theory of stochastic processes, and one of great relevance in statistical model validation, is the treatment of the n-sample distribution function, $F_n^*(x)$, as a stochastic process. A very useful theoretical result ensues [21].

Given n independent observations, $\xi_1, \cdots, \xi_n$ on a statistical population with distribution function $F(x)$, the Glivenko-Cantelli theorem asserts that

$$\lim_{n\to\infty} \sup_x \left[F_n^*(x) - F(x)\right] = 0 \tag{56}$$

where $F_n^*(x)$ denotes the n-sample distribution. The asymptotic distribution of $\sup_x \left[F_n^*(x) - F(x)\right]$ can be obtained by considering the following stochastic process

$$\eta_n(x) = \sqrt{n}\left(F_n^*(x) - F(x)\right), \quad -\infty < x < \infty \ . \tag{57}$$

It can then be shown, using results from measures on Hilbert spaces (namely the space $D[0,1]$) that the finite dimensional distributions of the process $\eta_n(x)$ converge weakly to the finite dimensional distributions of a zero-mean Gaussian process, $\eta(x)$ with covariance function $E(\eta(x)\eta(y)) = F(x)(1 - F(y))$ for $-\infty < x \leq y < \infty$.

6 Concluding Remarks

Functional analysis provides the foundation for developing approximate solutions in many applications in science and engineering.

Thus, it is not surprising that this same function analytic approach, applied in a probabilistic setting, provides one path to developing a similar rigorous framework for augmenting traditional deterministic analysis methods to accommodate uncertainty quantification (UQ).

The main idea is to recognize that the so-called random mappings, that is, random variables (RVs) and random fields (RFs), are, in fact, not random but rather deterministic mappings where at least a part of the domain consists of sets in a probability sample space. The randomness itself lies in the probabilility associated with the occurrance of a given domain set.

One uses this to build the necessary analytical components via direct ties to existing deterministic analysis, measure theoretical, and function analysis tools. For example, one defines expectation in terms of Lebesgue integration, a probability density function using Radon-Nikodym, etc.

The path one follows is first to establish the building blocks for the case of RVs. Then one migrates from RVs to RFs by building on them to develop a product space-style set expansion, similar to what one uses to increase the domain in a deterministic setting when migrating from an ODE to a PDE, and to build a new measure space in this expanded domain.

One key difference between the two is that in the deterministic case, one has to deal only with one measure, specifically Lebesgue measure. In the UQ-enabled setting one has to deal with product measures where a wide class of general probability measures are a constituent. Naturally, this fact imposes constraints.

In this document, we have provided a narrative in which we do these things using, primarily, a particular application setting: the classical Hilbert space, L_2. In the course of this narrative, we demonstrate our ideas through simple examples, where, in the RF case, decompose our function mappings into products, just as one encounters in classical separation of variables solutions of PDEs, and discuss the relationships that exist in the UQ setting that tie the two decomposed function spaces together.

Where appropriate, we provide a discussion on how one might seek to achieve results on more general function space settings.

References

[1] M. Abramowitz and I. A. Stegun, editors. *Handbook of Mathematical Functions*. 9th Printing, Dover Publications, Inc., New York, 1970.

[2] N. Aronszajn. Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68(3):337–403, May 1950.

[3] Richard Askey and James Wilson. *Some Basic Hypergeometric Orthogonal Polynomials That Generalize Jacobi Polynomials*, volume 54. Memoirs of the American Mathematical Society, Providence, RI, March 1985.

[4] Patrick Billingsley. *Probability and Measure*. J. Wiley and Sons, New York, 1995.

[5] Patrick Billingsley. *Convergence of Probability Measures*. J. Wiley and Sons, New York, second edition, 1999.

[6] S. Bochner. Stochastic processes. *Annals of Mathematics*, 48:1014, 1947.

[7] G. E. P. Box and M. E. Muller. A note on the generation of random normal deviates. *Annals. Math. Stat.*, 29:610–611, 1958.

[8] P.J. Daniell. A general form of integral. *Annals of Mathematics*, 19:27994, 1918.

[9] Persi Diaconis, Susan Holmes, and Richard Montgomery. Dynamical bias in the coin toss. *SIAM Review*, 49(2):211–235, 2007.

[10] Persi Diaconis and Charles Stein. Some tauberian theorems related to coin tossing. *The Annals of Probability*, 6(3):483–490, June 1978.

[11] J. Diestel and J.J. Uhl. *Vector Measures*. American Mathematical Society, 1977.

[12] J. L. Doob. *Measure Theory*. Springer-Verlag, 1994.

[13] A. Doostan, R. Ghanem, and J. Red-Horse. Stochastic model reduction for chaos representations. *Computer Methods in Applied Mechanics and Engineering*, 196:3951–3966, 2007.

[14] A. Einstein. On the movement of small particles suspended in a stationary liquid demanded by the molecular kinetic theory of heat. *Annalen der Physik*, 17, 1905. (Reprinted in Einstein, 1956).

[15] A. Einstein. *Investigations on the Theory of Brownian Movement*. Dover Publications, 1956.

[16] Avner Friedman. *Foundations of Modern Analysis*. Dover Publications, Inc., New York, NY, 1982.

[17] I.M. Gel'fand and N. Ya. Vilenkin. *Generalized Functions, Vol 4: Applications of Harmonic Analysis*. Academic Press, 1964.

[18] R. Ghanem and A. Doostan. On the construction and analysis of stochastic models: characterization and propagation of the errors associated wit h limited data. *J. Comp. Phys.*, 217(1):63–81, 2006.

[19] R. Ghanem and J. Red-Horse. Optimal representations of stochastic processes. In *9th ASCE EMD-SEI-GI-AD Joint Specialty Conference on Probabilistic Mechanics and Structural Reliability*, Albuquerque, NM, July 2004.

[20] R. Ghanem and J. Red-Horse. Orthogonal representations of stochastic processes and their propagation in mechanics. In *43rd IEEE Conference on Decision and Control*, Bahamas, December 2004.

[21] I. Gikhman and A. Skorohod. *The Theory of Stochastic Processes I.* Springer-Verlag, Berlin, 1974.

[22] U. Grenander. *Probabilities on Algebraic Structures.* John Wiley, New York, 1963.

[23] R.L. Hall. Inverse moments for a class of truncated normal distributions. *Sankhya, Series B*, 41:66–76, 1979.

[24] E. Hille and R.S. Phillips. Functional analysis and semi-groups. *American Mathematical Society colloquium Publication*, 31, 1957.

[25] E.T. Jaynes. *Probability Theory.* Cambridge, 2003.

[26] J. B. Keller. The probability of heads. *American Mathematical Monthly*, 93:191–197, March 1986.

[27] M. Kirby. Minimal dynamical systems from pdes using sobolev eigenfunctions. *Physica D*, 57:466–475, 1992.

[28] A. N. Kolmogorov. *Foundations of the Theory of Probability.* Chelsea Publishing Co, New York, NY, 2nd english edition edition, 1956.

[29] A. N. Kolmogorov and S. V. Fomin. *Elements of the Theory of Functions and Functional Analysis.* Dover Publications, Mineola, NY, 1957.

[30] A. Levy and J. Rubinstein. Some properties of smoothed principal component analysis for functional data. *Journal of The Optical Society of America*, 16(1):28–35, 1999.

[31] Reuven Meidan. A generalized karhunen-loève expansion. *Advances in Applied Probability*, 9(2):231–232, June 1977.

[32] Reuven Meidan. Reproducing-kernel hilbert spaces of distributions and generalized stochastic processes. *SIAM J. Math Anal*, 10(1):62–70, January 1979.

[33] Reuven Meidan. On the connection between ordinary and generalized stochastic processes. *Journal of Mathematical Analysis and Applications*, 76:124–133, 1980.

[34] M. Z. Nashed and G. Wahba. Generalized inverses in reproducing kernel spaces: An approach to regularization of linear operator equations. *SIAM J. Math Anal*, 5(6):974–987, November 1974.

[35] T.J. Oden and L.F. Demkowicz. *Applied Functional Analysis*. CRC Press, 1996.

[36] K.R. Parthasarathy. *Probability Measures on Metric Spaces*. AMS Chelsea Publishing, 1967.

[37] B.J. Pettis. On integration in vector spaces. *Transactions of the American Mathematical Society*, 44(2):277–304, 1938.

[38] L. Pontrjagin. The theory of topological commutative groups. *Annals of Mathematics*, 35(2):361–388, 1934.

[39] Y. Prokhorov and V. Sazanov. Some results associated with bochner's theorem. *Theory of Probability and its Applications*, 6(1):82–87, 1961.

[40] V.S. Pugachev and I.N. Sinitsyn. *Stochastic Systems: Theory and Applications*. World Scientific, New Jersey, 2001.

[41] J.R. Red-Horse and A. Benjamin. A probabilistic approach to uncertainty quantification with limited information. *Reliability Engineering & System Safety*, 85(1-3):183–190, 2004.

[42] J. S. Rosenthal. *A First Look at Rigorous Probability Theory*. World Scientific, 2000.

[43] R.Y. Rubinstein. *Simulation and the Monte Carlo Method*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, NY, 1981.

[44] Walter Rudin. *Real and Complex Analysis*. McGraw-Hill, New York, third edition, 1987.

[45] S. Saitoh. *Integral Transforms, Reproducing Kernels and their Applications*. Addison Wesley Longman Limited, 1997.

[46] V. Sazanov. A remark on characteristic functionals. *Theory of Probability and its Applications*, 3:188–192, 1958.

[47] V.V. Sazanov and V.N. Tutubalin. Probability distributions on topological groups. *Theory of Probability and its Applications*, 11(1):1–45, 1966.

[48] B.W. Silverman. Smoothed functional principal component analysis by choice of norm. *Annals of Statistics*, 24(1):1–24, 1996.

[49] A. Skorohod. *Integration in Hilbert Space*. Springer-Verlag, 1972.

[50] A. V. Skorokhod. *Integration in Hilbert Spaces*. Springer-Verlag, 1974.

[51] C. Soize and R. Ghanem. Physical systems with random uncertainties: Chaos representations with arbitrary probability measure. *SIAM Journal of Scientific Computing*, 26(2):395–410, 2004.

[52] N. Wiener. The homogeneous chaos. *Amer. J. Math*, 60:897–936, 1938.

[53] Dongbin Xiu and George Em Karniadakis. The wiener-askey polynomial chaos for stochastic differential equations. *SIAM Journal on Scientific Computing*, 24(2):619–644, 2002.

[54] Dongbin Xiu, Didier Lucor, C.-H. Su, and George Em Karniadakis. Stochastic modeling of flow-structure interactions using generalized polynomial chaos. *Journal of Fluids Engineering*, 124:51–59, 2002.

[55] N. Young. *An Introduction to Hilbert Space*. Cambridge University Press, Cambridge, UK, 1997.

[56] D.X. Zhang and Z.M. Lu. An efficient, high-order perturbation approach for flow in random porous media via karhunen-loeve and polynomial expansions. *Journal of Computational Physics*, 194(2):773–794, 2004.

DISTRIBUTION:

1 Ivo M. Babuska
Department of Aerospace Engineering and Engineering Mechanics
1 University Station, C0200
201 E. 21st Street, ACE 6.416
Austin, TX 78712

1 Roger Ghanem
Department of Aerospace and Mechanical Engineering
University of Southern California, 210 KAP Hall
Los Angeles, CA 90089

1 Pol D. Spanos
Dept of Mechanical Engineering, 233 MEB Rice University
6100 Main
Houston, TX 77005-1827

1 Raul Tempone
School of Computational Science
Florida State University
Dirac Science Library
Tallahassee, FL 32306-4120

1 MS 1318 J. R. Stewart, 1411

1 MS 1318 B. M. Adams, 1411

1 MS 1318 R. Bartlett, 1411

1 MS 1318 M. S. Eldred, 1411

1 MS 1318 E. T. Phipps, 1411

1 MS 1318 D. Ridzal, 1411

1 MS 1318 L. P. Swiler, 1411

1 MS 0370 T. G. Trucano, 1411

1 MS 1318 C. G. Webster, 1411

1 MS 1318 K. F. Alvin, 1414

1 MS 1318 P. B. Bochev, 1414

1 MS 1318 R. B. Lehoucq, 1414

1	MS 1318	B. van Bloemen Waanders, 1414
1	MS 1318	S. S. Collis, 1416
1	MS 1318	M. A. Heroux, 1416
1	MS 0825	B. Hassan, 1511
1	MS 0346	R. V. Field, 1526
1	MS 0380	H. S. Morgan, 1540
1	MS 0382	E. S. Hertel, 1541
1	MS 0382	M. W. Glass, 1541
1	MS 0380	J. Jung, 1542
1	MS 0380	M. K. Bhardwaj, 1542
1	MS 0380	M. W. Heinstein, 1542
1	MS 0380	S. W. Key, 1542
1	MS 0828	A. A. Giunta, 1544
1	MS 0828	A. R. Black, 1544
1	MS 0828	K. J. Dowding, 1544
1	MS 0557	T. L. Paez, 1544
1	MS 0828	V. J. Romero, 1544
10	MS 0828	J. R. Red-Horse, 1544
1	MS 0828	W. R. Witkowski, 1544
1	MS 9051	Y. Marzouk, 8351
1	MS 9051	P. P. Pebay, 8351
1	MS 9159	P. T. Boggs, 8961
1	MS 9159	K. R. Long, 8961
1	MS 0899	Technical Library, 9536 (electronic only)