Analysis of Probabilistic and Parametric Reduced Order Models*

Hermann G. Matthies

Institute of Scientific Computing
Technische Universität Braunschweig
38092 Braunschweig, Germany
e-mail: wire@tu-bs.de

5th September 2018

Abstract

Stochastic models share many characteristics with generic parametric models. In some ways they can be regarded as a special case. But for stochastic models there is a notion of weak distribution or generalised random variable, and the same arguments can be used to analyse parametric models. Such models in vector spaces are connected to a linear map, and in infinite dimensional spaces are a true generalisation. Reproducing kernel Hilbert space and affine- / linear- representations in terms of tensor products are directly related to this linear operator. This linear map leads to a generalised correlation operator, and representations are connected with factorisations of the correlation operator. The fitting counterpart in the stochastic domain to make this point of view as simple as possible are algebras of random variables with a distinguished linear functional, the state, which is interpreted as expectation. The connections of factorisations of the generalised correlation to the spectral decomposition, as well as the associated Karhunen-Loève- or proper orthogonal decomposition will be sketched. The purpose of this short note is to show the common theoretical background and pull some lose ends together.

Keywords: stochastic models, parametric models, correlation, factorisation, spectral decomposition

MSC Classification: 35B30, 37M99, 41A05, 41A45, 41A63, 60G20, 60G60, 65J99, 93A30

*Partly supported by the Deutsche Forschungsgemeinschaft (DFG) through SPP 1886 and SFB 880.
1 Introduction

Probabilistic and parametric models, used in many areas of science, engineering, and economics, share many similarities. Probabilistic models are used to describe uncertainties or random phenomena, whereas parametric models describe variations or changes of some system as some parameters are changed. Typically these are part of some larger mathematical model describing some system with such characteristics. A parameter can of course be a random variable, and this is the connection between these two kinds of models. Here the interest is mainly in system models with an infinite dimensional state space, e.g. systems described by ordinary or partial differential equations. This often also makes it necessary to theoretically consider infinitely many parameters. In an actual numerical computation this has of course to be reduced through some kind of discretisation to a finite number. And obviously one would like to have this number as small as possible while still retaining acceptable accuracy. This is the realm of reduced order models.

These reduced order models lessen the possibly high computational demand, and are hence probabilistic or parametrised reduced order models. The survey [1] and the recent collection [2], as well as the references therein, provide a good account of parametric reduced order models and some of the areas where they appear. The interested reader may find there further information on parametrised reduced order models and how to generate them.

Here we build on our recent work [20, 19] analysing parametrised reduced order systems, which itself is a continuation of [18]. In these publications the theoretical background of such parametrised models is treated in a functional analysis setting. The purpose of the present note is to use the same kind of techniques for stochastic or probabilistic models, where some generalisations are required due to the wish to cover infinite dimensional state spaces, and combine this with the description of parametric reduced order models.

As an example, assume that some physical system is investigated, which is modelled by an evolution equation for its state \( v(t) \in \mathcal{V} \) at time \( t \in [0, T] \), where \( \mathcal{V} \) is assumed to be a Hilbert space for the sake of simplicity: \( \dot{v}(t) = A(\varsigma, \mu; v(t)) + f(\varsigma, \mu; t) \); \( v(0) = v_0 \), where the superimposed dot signifies the time derivative, \( A \) is an operator modelling the physics of the system, and \( f \) is some external excitation. Here \( \varsigma \) is a random variable (RV) defined on an event space \( \Omega \) with values in some Hilbert space \( \mathcal{S} \) (again for simplicity), and \( \mu \in \mathcal{M} \) are parameters that can be controlled, and can be used to evaluate the design of the system, control its behaviour, or optimise the performance in some way. No specific structure is assumed for the set \( \mathcal{M} \). We assume that for all possible values of \( \varsigma \) and for all \( \mu \) of interest the system is well-posed. This will make the system state \( v(\varsigma, \mu; t) \) a random variable as well, depending on the value of the parameters \( \mu \).

One may be interested in the state of the system \( v(\varsigma, \mu; t) \) and its statistics, or some functional of it, say \( \Psi(\mu) = \mathbb{E}(\psi(v(\varsigma, \mu))) \), where \( \mathbb{E} \) is an expectation operator. While evaluating \( A(\varsigma, \mu) \) or \( f(\varsigma, \mu) \) for a certain \( \mu \) may be straightforward, evaluating \( v(\varsigma, \mu; t) \) or \( \Psi(\mu) \) may be very costly. This is why one wants representations of \( v(\varsigma, \mu; t) \) or \( \Psi(\mu) \) which allow a cheaper evaluation. This is achieved through reduced order models, which are often also called proxy- or surrogate-models. It turns out that such random and parametric objects can be analysed by associated linear maps [20, 19], which renders them much more accessible to the techniques of linear functional analysis, a well understood subject. This association with linear mappings has probably been known for a long time,
see [15] for an exposition in the context of stochastic models. In Section 2 the association of parametric and stochastic models with linear maps will be explained, in passing touching on reproducing kernel Hilbert spaces. The classical probabilistic framework (cf. [26]), starting from measurable spaces and \( \sigma \)-algebras, can be used to define algebras of random variables (RVs) as measurable functions on these measure spaces, and the expectation operator as integral of these RVs w.r.t. the probability measure. These algebras of RVs can be used in the case of probabilistic models to build the range or image space for these linear maps as spaces of classical RVs. But alternatively one may also start by using as fundamental concepts algebras of objects that we want to call RVs together with the expectation operator (cf. [24]) as a linear functional, and if this algebra of RVs is Abelian or commutative one essentially recovers equivalence with classical probability. This approach allows for non-commuting algebras of RVs, which is important (cf. [21]) in order to deal with e.g. random matrices, random fields of tensors, quantum theory and quantum fields. More important for our immediate purposes here, this view greatly facilitates the specification of stochastic models on infinite dimensional spaces. Such an algebra of RVs, whether generated classically as derived concept as an algebra of measurable functions, or used as a primary model of possibly non-commuting of RVs, seems to be a natural object to use in the case of stochastic models on infinite dimensional vector spaces, as it allows to generalise such stochastic models to so-called weak distributions or generalised processes (cf. [22] [14] [8] [23]), and thereby elegantly circumvent many problems which arise when one tries to define \( \sigma \)-additive set functions for example on Hilbert spaces. This algebraic and analytic view on probability will be explained in Section 3. Everything is tied together in Section 4 in the analysis of the generalised correlation operator, its factorisations, as well as its spectral decomposition, and the last Section 5 concludes by pointing out once more the connection between functions in high-dimensional spaces and the associated linear maps and correlation operators, where well-known methods can be used to analyse their structure.

2 Parametric and stochastic models

We start with a short recap of [20] [19], where the interested reader may find more detail. Let \( r : \mathcal{M} \to U \) be a generic substitute for any one of the parametric objects alluded to in the introduction, e.g. things like \( \mu \mapsto v(\varsigma, \mu, t) \in \mathcal{V} \) or \( \mu \mapsto \dot{v}(\varsigma, \mu, \cdot) \in L^2([0, T]) \otimes \mathcal{V} \); \( \omega \mapsto \varsigma(\omega) \in \mathcal{S} \)—with \( \Omega \) taking the rôle of \( \mathcal{M} \); \( (\mu, \omega) \mapsto v(\varsigma(\omega), \mu, t) \in \mathcal{V} \)—with \( \mathcal{M} \times \Omega \) taking the rôle of \( \mathcal{M} \); \( \omega \mapsto f(\varsigma(\omega), \mu, t) \in \mathcal{V} \)—with \( \Omega \) taking the rôle of \( \mathcal{M} \), or \( \mu \mapsto A(\varsigma, \mu, \cdot) \in (\mathcal{V} \to \mathcal{V}) \)—the space of maps from \( \mathcal{V} \) to \( \mathcal{V} \), etc.

The space \( U \) is assumed for the sake of simplicity as a separable Hilbert space. The function \( r \) can thus be either a parametric input, or a random input—i.e. a random variable (RV), in which case \( \mathcal{M} \) would be a measure space—to a model like that described in Section 1 or the operator of that model, or the state (solution) of that system. Assuming—without significant loss of generality—that the image \( \text{span} r(\mathcal{M}) = \text{span} r \subseteq U \) is dense in \( U \), one may to each such function \( r \) associate a linear map \( R : U \ni u \mapsto \langle r(\cdot) | u \rangle_U \in \mathbb{R}^\mathcal{M} \) into the space \( (\mathcal{M} \to \mathbb{R}) \) of all real-valued functions on \( \mathcal{M} \). By construction, \( R \) restricted to \( \text{span} r = \text{span} r(\mathcal{M}) \) is injective. In Section 3 it will be explained how—in the case of a probabilistic or random model—the Hilbert space can be generated from an algebra of RVs.
As an aside, note that on its restricted range \( \tilde{R} := R(\text{span im } r) \subseteq \mathbb{R}^M \) one may define an inner product as \( \langle \phi|\psi \rangle_R := \langle R^{-1}\phi|R^{-1}\psi \rangle_U \) for all \( \phi, \psi \in \tilde{R} \). Denote the completion with this inner product by \( \mathcal{R} \). This makes \( R \) and \( R^{-1} \) into bijective isometries, hence *unitary* maps between \( \mathcal{U} \) and \( \mathcal{R} \). It may easily be shown \([20, 19]\) that \( \mathcal{R} \) is a *reproducing kernel Hilbert space* (RKHS) \([3, 12]\) with reproducing kernel with this inner product by \( R \).

Maps between unitary such that the reproducing property \( \langle R(\mu)\cdot \rangle_R = \phi(\mu) \) holds for all \( \phi \in \tilde{R} \). In this note the RKHS \( \mathcal{R} \) will not be used, but the important thing to keep in mind is that the map \( R \) and the space \( \mathcal{R} \) of scalar functions on the set \( \mathcal{M} \)—one might view them as problem oriented co-ordinates—carry the same information as the parametric object \( r(\mu) \).

Often some information of what is important in the set \( \mathcal{M} \) is also available, here it is assumed to be given by a Hilbert subspace \( \mathcal{Q} \subseteq \mathbb{R}^M \), usually different from \( \mathcal{R} \). From now on we shall by slight abuse of notation view the map \( R \) as mapping into \( \mathcal{Q} \) and still assume that it is injective as well as closed, for the sake of simplicity. Details like the assumption that the subspace \( R^{-1}(\mathcal{Q}) \) is dense in \( \mathcal{U} \) will not always be spelt out in detail for the sake of brevity. The idea is that with \( u \in \mathcal{U} \) of unit length the vectors \( Ru \in \mathcal{Q} \) with large norm are more important, and this will be considered in building reduced order models. As will be shown \([20, 19]\) in Section 4, the map \( C : \mathcal{U} \to \mathcal{U} \) defined by \( C = R^*R \), where \( R^* \) is the adjoint of \( R \), is central to the analysis. More precisely, with the above assumptions on \( R \) the adjoint \( R^* \) is surjective, and \( C \) is a densely defined self-adjoint positive definite operator, which we shall call the ‘correlation’ of the model \( r(\mu) \).

A random variable or stochastic model as exemplified by the RV \( \varsigma \) in Section 1 is usually formulated as a measurable map \( \varsigma : \Omega \to \mathcal{I} \), where \( (\Omega, \mathfrak{A}, \mathbb{P}) \) is a probability space with \( \sigma \)-algebra \( \mathfrak{A} \) and probability measure \( \mathbb{P} \). One may view the set \( \Omega \) as a parameter set like \( \mathcal{M} \) above, and one can construct a linear map into the just defined space \( \mathbb{R}^\Omega \), i.e. the scalar random variables. Without loss of generality, we assume that \( \text{span } \varsigma(\Omega) = \text{span im } \varsigma \subseteq \mathcal{I} \) is dense in the separable Hilbert space \( \mathcal{I} \), and define \([15]\)

\[
S : \mathcal{I} \ni \xi \mapsto \langle \varsigma(\cdot)|\xi \rangle_{\mathcal{I}} \in \mathbb{R}^\Omega. \tag{1}
\]

It remains to define an inner product on \( \mathbb{R}^\Omega \) and a subspace corresponding to \( \mathcal{Q} \) for the parametric case above. This will be done in Section 3. For the time being assume that this has been defined, i.e. there is an inner product \( \langle \cdot|\cdot \rangle_{\mathcal{I}} \) and a corresponding Hilbert space of (equivalence classes) of RVs \( \mathcal{Y} \subseteq \mathbb{R}^\Omega \), and we regard \( S \) as a map \( S : \mathcal{I} \to \mathcal{Y} \) with the same properties as assumed for \( R \) above. Obviously the densely defined self-adjoint positive definite operator \( C_\varsigma = S^*S : \mathcal{I} \to \mathcal{I} \) corresponding to \( C = R^*R \) above is indeed the correlation operator of the RV \( \varsigma \).

In case \( \varsigma \) is an input to a dynamical system like the one alluded to in Section 4 the state of the system \( v(\varsigma, \mu; t) \) also becomes a stochastic quantity, and inner product with a vector \( w \in \mathcal{V} \) leads for fixed \( \mu \) and \( t \) automatically to a linear mapping

\[
P : \mathcal{V} \ni w \mapsto \langle v(\varsigma|\cdot), \mu; t|w \rangle_{\mathcal{V}} \in \mathbb{R}^\Omega, \tag{2}
\]

which we shall regard again as a map \( P : \mathcal{V} \to \mathcal{V} \) into the just defined space \( \mathcal{V} \). This defines a third correlation operator \( C_\nu = P^*P : \mathcal{V} \to \mathcal{V} \).

It may be seen that with the correspondences

\[
R : \mathcal{U} \to \mathcal{Q} \quad — \quad S : \mathcal{I} \to \mathcal{Y} \quad — \quad P : \mathcal{V} \to \mathcal{V} \tag{3}
\]
all three situations are completely analogous, and may in the simplest case be dealt with in the same formalism. The idea on how to obtain representations of \( r(\mu) \) resp. \( \varsigma(\omega) \) resp. \( v(\omega) \) is the following [20][19], which we shall mainly demonstrate for \( r(\mu) \): choose a complete basis \( \{q_j\}_j \subset \mathcal{Q} \), and represent \( r(\mu) \) as

\[
r(\mu) = \sum_j \alpha_j R^* q_j(\mu). \tag{4}
\]

A good reduced order model is one where

\[
r_{\text{ROM}}(\mu) = \sum_{j=1}^J \alpha_j R^* q_j(\mu) \tag{5}
\]
is a good approximation to \( r(\mu) \approx r_{\text{ROM}}(\mu) \) with a small \( J \), i.e. with not too many terms. In Section 4 some other possibilities for the choice of basis \( \{q_j\}_j \) will be discussed, where the \( \mu \)-dependence is encoded in the scalar functions from \( \mathcal{Q} \), but where a basis of \( \mu \)-independent vectors is picked from \( \mathcal{U} \), and where again for the sake of brevity and simplicity we shall confine ourselves to complete orthonormal systems (CONS). The important message here is that with \( R \) one has a factorisation of \( C = R^* R \), and that the adjoint is the map which carries a representation on the function space to the space \( \mathcal{U} \). Later we shall indicate [20][10] how every representation leads to a factorisation of \( C \), and that — with some additional assumptions on \( C \) — every factorisation leads to a representation. But the description and analysis via factorisations is more general [22][9][8][23][15], and this is needed in the formulation of probabilistic models where \( \mathcal{U} \) resp. \( \mathcal{F} \) is an infinite dimensional Hilbert space.

## 3 Algebras of random variables

Here we shall take a closer look at the stochastic or probabilistic model \( \varsigma : \Omega \to \mathcal{F} \) and the associated linear map \( S : \mathcal{F} \to \mathcal{V} \), as well as the space of RVs \( \mathcal{V} \) and how it is generated. Although there are classical ways of specifying the space \( \mathcal{V} \), the most natural one seems to be the algebraic approach to probability. These ideas are certainly also used in the classical approach, but the algebraic probability approach distills the essential components in an abstract setting and allows at the same time generalisations. Historically, when looking back as how in the beginnings of probability theory the Bernoullis treated random variables (RVs), it is clear that they added them and took multiples—hence they form a vector space—and that they multiplied them with each other—so they form an algebra. Although the formalisation of probability as formulated by Kolmogorov used the concept of measure and this algebraic background was largely ignored, it was revived with the advent of quantum theory. It turns out that here this view is essential, as not all observables can be observed simultaneously, and this is reflected in the fact that they do not commute in the algebra. Another topic where this view is very advantageous are random matrices and more generally random fields of even-order tensors.

We are mainly interested in ‘real’ or self-adjoint RVs as they will later be called. But for analytical convenience we shall treat complex RVs, following Paul Painlevé’s and Jacques Hadamard’s adage that the shortest path between two truths in the real domain passes through the complex domain — “le plus court chemin entre deux vérités dans le
domaine réel passe par le domaine complexe”. Some algebraic language is needed, but most of the terms will be familiar from complex numbers and from matrices, which are indeed two simple but prime examples of algebras. Let us start right away with a simple and mostly familiar example from probability theory, which will at the same time serve as motivation, concrete example, and explanation of the abstract setting.

3.1 Specifying the algebra

Consider a probability space \((\Omega, \mathcal{A}, \mathbb{P})\) with a set of elementary events \(\Omega\), \(\sigma\)-algebra \(\mathcal{A}\) of measurable subset of \(\Omega\), and probability measure \(\mathbb{P}\). In the vector space \(L_0(\Omega, \mathcal{A}, \mathbb{P}; \mathbb{C})\) of complex-valued measurable functions / classical random variables on \(\Omega\)—which for the sake of brevity shall be denoted just by \(L_0(\Omega)\)—let \(\mathcal{A}_s := L_{0s}(\Omega) \subset L_0(\Omega)\) be the vector subspace of complex-valued simple measurable functions, i.e. complex linear combinations of functions \(1_E\), which for \(E \in \mathcal{A}\) are defined to be \(1_E(\omega) = 1\) if \(\omega \in E \subseteq \Omega\), and zero otherwise. Hence \(\mathcal{A}_s\) are the RVs where each one of them can only take finitely many different values.

On this vector space we may define a multiplication by just pointwise multiplication of two such RVs, and the product is obviously again a simple function; in fact for \(E, F \in \mathcal{A}\) one has \(1_E 1_F = 1_{E \cap F}\), i.e. the multiplication in \(\mathcal{A}_s\) reflects the intersection in the \(\sigma\)-algebra \(\mathcal{A}\). This means that the space \(\mathcal{A}_s\) is closed under multiplication and hence thanks to the properties of the multiplication on \(\mathbb{C}\) is a complex, associative, and commutative or Abelian algebra, with the familiar distributive law from \(\mathbb{C}\) coupling addition and multiplication also on \(\mathcal{A}_s\). Another way of saying this is to state that the multiplication is a bilinear map from \(\mathcal{A}_s \times \mathcal{A}_s\) to \(\mathcal{A}_s\). Let us note in passing that with the same definition of pointwise multiplication also \(L_0(\Omega)\) is an associative and commutative algebra—with \(\mathcal{A}\), a sub-algebra—as the pointwise product of two measurable functions is again measurable, but we shall see later that for our purposes \(L_0(\Omega)\) is in general too big. The element \(1_\Omega \in \mathcal{A}_s \subset L_0(\Omega)\) which is constant equal to unity is obviously a neutral element or unit for the multiplication, and hence \(\mathcal{A}_s\) and \(L_0(\Omega)\) are called unital algebras. For \(\psi \in \mathcal{A}_s\) one can now compute powers \(\psi^n = \psi \psi^{n-1}\) for any integer \(n \geq 1\), and if we define \(\psi^0 = 1_\Omega\) in a unital algebra even for any \(n \geq 0\). Given a polynomial \(Q(X) = \sum_{k=0}^n \alpha_k X^k \in \Pi_1\) in one unknown \(X\) with complex co-efficients \(\alpha_k \in \mathbb{C}\), it is now possible to evaluate \(Q(\psi)\) in \(\mathcal{A}_s\) for any \(\psi \in \mathcal{A}_s\). For some \(\phi \in \mathcal{A}_s\), there is a \(\psi \in \mathcal{A}_s\) such that \(\phi \psi = 1_\Omega\). This is then called the (multiplicative) inverse \(\psi = \phi^{-1}\), such that \(\phi \phi^{-1} = 1_\Omega\).

For a complex number \(\zeta \in \mathbb{C}\) its complex conjugate is denoted by \(\zeta^* \in \mathbb{C}\), and this operation is an involution, as \((\zeta^*)^* = \zeta\). One may extend this involution from \(\mathbb{C}\) to the algebra \(L_0(\Omega)\) through a pointwise definition of complex conjugation, and hence also to its sub-algebra \(\mathcal{A}_s\). For \(\phi, \psi \in L_0(\Omega)\) and \(\zeta \in \mathbb{C}\) this involution obviously satisfies 
\[(\phi + \zeta \psi)^* = \phi^* + \zeta^* \psi^*\] and is thus anti-linear. As regards the product of two RVs, it satisfies \((\phi \psi)^* = \psi^* \phi^*\), and it is easy to verify that both \(\mathcal{A}_s\) and \(L_0(\Omega)\) are closed under this involution. Associative algebras with such an anti-linear involution and the indicated behaviour on products are called \(*\)-algebras—the element \(\psi^*\) is usually called in algebraic terms the adjoint of \(\psi\)—and both \(L_0(\Omega)\) and its sub-algebra \(\mathcal{A}_s = L_{0s}(\Omega)\) are thus \(*\)-algebras.

Let \(\Pi_2^\mathbb{C}\) denote the set of all polynomials \(Q(X, Y)\) with complex co-efficients in two commuting variables \(X, Y\). For \(\phi \in \mathcal{A}_s\) the unital sub-\(*\)-algebra \(\mathbb{C}[\phi, \phi^*] := \{Q(\phi, \phi^*) \mid Q \in \Pi_2^\mathbb{C}\} \subset \mathcal{A}_s\) is called the sub-algebra generated by \(\phi \in \mathcal{A}_s\). Observe that if \(\psi \in L_0(\Omega)\)}
is self-adjoint, i.e. $\psi = \psi^*$, then $\psi$ has only real values, and if $\psi = \phi^*\phi$ for some $\phi \in \mathcal{L}_0(\Omega)$, then $\psi$ is self-adjoint (real) and is called positive as it can not take negative values, i.e. $0 \leq \psi = \phi^*\phi$—in case $0 < \psi$ it is usually called strictly positive. One says that for self-adjoint $\phi, \psi \in \mathcal{L}_0(\Omega)$ one has $\psi \leq \phi$ iff $\phi - \psi$ is positive, and thus one can define a partial order on $\mathcal{A}_s$ and $\mathcal{L}_0(\Omega)$. Positive self-adjoint elements $\psi \in \mathcal{L}_0(\Omega)$ which are idempotent, i.e. satisfy $\psi^2 = \psi$, are called projections. Observe that each $1_E$ is a projection, and that the unit $1_\Omega$ is a maximal projection in the order mentioned. In fact all projections in $\mathcal{L}_0(\Omega)$ and $\mathcal{A}_s$ have the form $1_E$ for some $E \in \mathfrak{A}$. Ultimately, one is only interested in the self-adjoint elements of the algebra $\mathcal{A}_s$, as they take real values; they are therefore often also called observables. The other elements of the algebra may be regarded as merely a kind of analytical completion to make the theory nice. It may be remarked that the self-adjoint elements of $\mathcal{A}_s$ form a real subspace of $\mathcal{A}_s$. Obviously an arbitrary $\phi \in \mathcal{A}_s$ may be decomposed into real and imaginary parts: $\phi = \Re\phi + i\Im\phi$ with real resp. self-adjoint $\Re\phi = (\phi + \phi^*)/2$ and $\Im\phi = (\phi - \phi^*)/(2i)$, so that the whole algebra is the complex span of the self-adjoint elements or observables.

To extract the essential point from this example and generalise, we start with an associative algebra $\mathcal{A}$ of what we want to call random variables (RVs) $a, b, \cdots \in \mathcal{A}$, i.e. a vector space $[23]$ equipped with an associative and bi-linear multiplication which will be denoted just by juxtaposition: $\mathcal{A} \times \mathcal{A} \ni (a, b) \mapsto ab \in \mathcal{A}$. As was noted before, it is advantageous to assume the algebra to be a complex algebra, which is no loss of generality as any real algebra may be embedded into a complex one. For $a \in \mathcal{A}$ the powers $a^n$ are defined for any integer $n \geq 1$ in the natural recursive fashion. Additionally assume that the algebra is unital, i.e. has a multiplicative unit $e$ such that $ae = ea = a$ for any $a \in \mathcal{A}$, and one defines the power $a^n$ for $n = 0$ by $a^0 = e$. Hence for a polynomial $Q(X) \in \Pi_1$ it is now possible to evaluate $Q(a)$ for any $a \in \mathcal{A}$. Also assume that there is an anti-linear involution defined, called the ‘adjoint’, denoted as $a^*$, such that $(a^*)^* = a$ and $(ab)^* = b^*a^*$.

Let $\Pi_2^a$ be a set of all polynomials $Q(X, Y)$ with complex co-efficients in two non-commuting variables $X, Y$, then for $a \in \mathcal{A}$ the unital sub*-algebra $\mathbb{C}\{ a, a^* \} := \{Q(a, a^*) \mid Q \in \Pi_2^a \} \subset \mathcal{A}$ is called the sub-algebra generated by $a \in \mathcal{A}$. Elements $a \in \mathcal{A}$ such that $a = a^*$ are called self-adjoint, and self-adjoint elements which may be factored as $a = b^*b$ are called positive. Positive elements form a salient pointed cone which defines an order relation on $\mathcal{A}$. Positive elements $p$ which are idempotent $p = pp = p^2 = p^*p = p^*$ are called projections. Observe that $e$ is a projection, and that it is maximal w.r.t. the order mentioned. Succinctly stated, we assume that $\mathcal{A}$ is a complex associative unital *-algebra, not necessarily commutative. As was shown, both $\mathcal{L}_0(\Omega)$ and $\mathcal{A}_s$ considered above are commutative examples of such algebras. Again, one is later ultimately interested in the self-adjoint elements of $\mathcal{A}$—the observables. Also in the general abstract case they form a real subspace of $\mathcal{A}$, and an arbitrary $a \in \mathcal{A}$ may be decomposed into two parts $a = a_s + ia_w$ with self-adjoint $a_s = (a + a^*)/2$ and $a_w = (a - a^*)/(2i)$—also called the symmetric and skew parts—so that the whole algebra is the complex span of the self-adjoint elements, the observables. And naturally, if for some $a \in \mathcal{A}$ there is a $c \in \mathcal{A}$ such that $ac = ca = e$, then $c = a^{-1}$ is the unique multiplicative inverse of $a$. 


3.2 States and the expectation functional

To continue, we return to the example $\mathcal{A}_s$ above. Just as classical probability builds on the measurable space $(\Omega, \mathcal{A})$ on one hand and the probability measure $\mathbb{P}$ on the other hand, in the algebraic framework the second entity needed is the linear expectation functional $E : \mathcal{A}_s \to \mathbb{C}$. To define the expected value for a RV $\phi \in \mathcal{A}_s$ one only has to look at the generating elements $1_{\mathcal{E}}$ with $\mathcal{E} \in \mathcal{A}$. Here one defines $E(1_{\mathcal{E}}) := \int_{\Omega} 1_{\mathcal{E}}(\omega) \mathbb{P}(d\omega) = \mathbb{P}(\mathcal{E})$ and extends this by linearity to all of $\mathcal{A}_s$. Thus the probability of an event $\mathcal{E} \in \mathcal{A}$ is given in terms of the expected value of the associated projection $1_{\mathcal{E}}$. For a typical $\phi(\omega) = \sum_k \alpha_k 1_{\mathcal{E}_k}(\omega) \in \mathcal{A}_s$ with $\alpha_k \in \mathbb{C}$ this gives $E(\phi) = \int_{\Omega} \phi(\omega) \mathbb{P}(d\omega) = \sum_k \alpha_k \mathbb{P}(\mathcal{E}_k) \in \mathbb{C}$. Obviously, as $\mathbb{P}(\Omega) = 1$, the expected value of the unit is $E(1_{\Omega}) = 1$, a kind of normalisation of the expectation functional.

This linear functional $E$ additionally satisfies $E(\phi^*) = (E(\phi))^*$ and thus carries the adjoint to its complex conjugate and hence is real on self-adjoint elements. Such a linear functional is itself called self-adjoint. In addition, $E(\phi^* \phi) = \sum_k (\alpha_k^\ast \alpha_k) E(1_{\mathcal{E}_k}) = \sum_k |\alpha_k|^2 \mathbb{P}(\mathcal{E}_k) \geq 0$, i.e. the functional is non-negative on positive $\psi = \phi^* \phi \in \mathcal{A}_s$. Such a self-adjoint linear functional is itself called positive. If $\rho \in \mathcal{A}_s$ is positive with unit expected value $E(\rho) = 1$, one may define a new expectation functional—corresponding to a change of probability measure—via $E_\rho(\phi) := E(\rho \phi) = \int_{\Omega} \rho(\omega) \phi(\omega) \mathbb{P}(d\omega)$. It is easily checked that $E_\rho$ is linear, self-adjoint, positive, and normalised. Such linear functionals which can serve as expectation are called states, an element of the dual space $\mathcal{A}_s^\ast$.

The element $\tilde{\phi} := E(\phi) 1_{\Omega} \in \mathcal{A}_s$ is called the mean of $\phi \in \mathcal{A}_s$ and the additive rest $\check{\phi} = \phi - \tilde{\phi} \in \mathcal{A}_s$ is its zero-mean or centred or fluctuating part. The one-dimensional unital $*$-algebra $\mathcal{A}_{sc} := \mathbb{C}[1_{\Omega}] = \text{span}\{1_{\Omega}\} \subset \mathcal{A}_s$—isomorphic to $\mathbb{C}$—are the constants, whereas the subspace $\mathcal{A}_{s0} := \ker E$ are the zero-mean or centred RVs, such that $\mathcal{A}_s = \mathcal{A}_{sc} \oplus \mathcal{A}_{s0} = \mathbb{C}[1_{\Omega}] \oplus \ker E$ as a direct sum.

One may observe that in general not every measurable $\phi \in L_0(\Omega)$ has a finite integral. Thus the algebra of all classical RVs $L_0(\Omega)$ is too big for our purpose as one would like $E(\cdot)$ to be defined on the whole algebra. This is the reason to start with the ‘smaller’ algebra $\mathcal{A}_s = L_{0s}(\Omega)$. It is a building block from which more complicated RVs can be built via limiting processes.

In the general abstract case one also wants a linear, self-adjoint, positive, and normalised functional—a state—$E : \mathcal{A} \to \mathbb{C}$ with $E(\alpha^*) = E(\alpha)^*$. Such a state is called faithful if $E((a^*a) = 0$ implies $a = 0$. If a state is not faithful, then one can start to work with an algebra of equivalence classes, where two elements $a,b \in \mathcal{A}$ are considered equivalent if $E((a-b)^*(a-b)) = 0$. It is therefore no loss of generality to assume that the state is faithful. The projections $p \in \mathcal{A}$ are also identified with events, and the probability of the event $p \in \mathcal{A}$ may be defined as $\mathbb{P}(p) := E(p)$. As $E$ is positive, one has $\mathbb{P}(p) \geq 0$, and as $\varepsilon$ is a maximal projection, $\mathbb{P}(\varepsilon) = E(\varepsilon) = 1$. One defines the mean part of a RV as a multiple of the identity $\bar{a} := E(a) 1_{\Omega}$ and the fluctuating zero-mean or centred part as $\tilde{a} := a - \bar{a}$ with $E(\tilde{a}) = 0$. The one dimensional sub-$*$-algebra $\mathcal{A}_c = \mathbb{C}[\varepsilon] = \text{span}\{\varepsilon\}$ of constants—$\mathcal{A}_c$—are multiples of the identity, and the subspace of zero-mean fluctuating parts $\mathcal{A}_0 = \ker E$ is the kernel of the state, and the whole algebra is the direct sum of both parts $\mathcal{A} = \mathcal{A}_c \oplus \mathcal{A}_0 = \mathbb{C}[\varepsilon] \oplus \ker E$. An abstract algebra which satisfies all these requirements together with a distinguished faithful state as expectation is called a probability algebra. If $\varrho \in \mathcal{A}$ is positive with unit expectation $E(\varrho) = 1$, then one may define a new weighted state by $E_\varrho(a) := E(\varrho \bar{a})$ for $a \in \mathcal{A}$.
A faithful state may be used to define an inner product on $A$, via a positive definite sesqui-linear form:

$$A^2 \ni (a,b) \mapsto \langle a|b \rangle_2 := \mathbb{E}(b^*a) \in \mathbb{C}. \quad (6)$$

As usual, one may define the square of a norm via $\|a\|^2 := \langle a|a \rangle_2$. The completion of $A$ in the uniform topology generated by this norm is a Hilbert space denoted by $L_2(A)$, which is one candidate for $\mathcal{F} := L_2(A)$. Later we shall see more possible ways of generating a Hilbert space of RVs. With this inner product the above direct sum of the mean is one candidate for $V := \mathbb{E}_N$.

As the expectation or state is normally also continuous in the topology of the associated Hilbert space $\mathcal{F}$, it can be defined also on $\mathcal{F}$ giving an orthogonal decomposition $\mathcal{F} = \ker \mathbb{E} \oplus (\ker \mathbb{E})^\perp =: \mathcal{F}_0 \oplus \mathbb{C}[e]$. For the probabilistic model $S : \mathcal{F} \to \mathcal{F}$ this means that it can be extended to $\xi \in \mathcal{F}$ as $\mathbb{E}_\xi(\xi) := \mathbb{E}(S\xi)$, and with it an orthogonal decomposition of $\mathcal{F} = \mathcal{F}_0 \oplus \mathcal{F}_1 := \ker \mathbb{E}_\xi \oplus (\ker \mathbb{E}_\xi)^\perp$, where $(\ker \mathbb{E}_\xi)^\perp = \text{span}(S^*e)$ are multiples of the mean $\xi := S^*e \in \mathcal{F}$ of the RV $\xi$. Instead of looking at the correlation operator $C_\xi = S^*S$, one is usually only interested in the correlation $\hat{C}_\xi = S\hat{S}$ of $\hat{S}$, where $\hat{S} : \mathcal{F} \ni \xi \mapsto S\xi - \mathbb{E}_\xi(\xi)e \in \mathcal{F}_0 - \hat{C}_\xi$ is called the covariance operator. Completely analogous statements can be made for the map $P : \mathcal{F} \ni w \mapsto \langle w|\xi \rangle_{\hat{P}} \in \mathcal{F}$, the associated expectation $\mathbb{E}_P(w) := \mathbb{E}(P\xi)$, the orthogonal split $\mathcal{F} = \mathcal{F}_0 \oplus \mathcal{F}_1 := \ker \mathbb{E}_P \oplus \text{span}\{P* e\}$, and the associated covariance operator.

In the example algebra $A_s = L_{0s}(\Omega)$ from above, identifying $1_\mathcal{E}$ and $1_\mathcal{F}$ if $\mathcal{E}, \mathcal{F} \in \mathfrak{A}$ differ only by a null-set $\mathcal{N} \in \mathfrak{A}$ with $\mathbb{P}(\mathcal{N}) = 0$, the integral or expected value becomes a faithful state. As is well known [24], the construction in Eq. (6) defines the $L_2$ inner product $\langle \phi|\psi \rangle_2 = \mathbb{E}(\psi^*\phi) = \int_\Omega \psi(\omega)^*\phi(\omega)\mathbb{P}(d\omega)$ for $\phi, \psi \in A_s = L_{0s}(\Omega)$, and the completion is the familiar Hilbert space $L_2(\Omega) = L_2(A_s)$. The inner product $\langle \phi|\psi \rangle_2$ of two RVs $\phi, \psi \in A_s$ is also called their correlation, and one may continue and define the covariance in the usual way by $\text{cov}(\phi, \psi) := \mathbb{E}\left(\psi^*\phi - \langle \phi|\psi \rangle_2\right)$, i.e. the inner product or correlation of the fluctuating parts. The variance of a RV $\phi \in A_s$ is then $\text{var}(\phi) := \text{cov}(\phi, \phi)$, and one has from Pythagoras’s theorem $\|\phi\|^2 = \|\phi\|^2_2 + \|\phi\|^2_2 = \mathbb{E}(\phi^2) + \text{var}(\phi)$. Two RVs $\phi, \psi \in A_s$ are uncorrelated iff their covariance vanishes: $\text{cov}(\phi, \psi) = 0$, i.e. if $\mathbb{P}(\mathcal{N}) = 0$, the integral or expected value becomes a faithful state. As is well known [24], the construction in Eq. (6) defines the $L_2$ inner product $\langle a|b \rangle_2$, the covariance as the inner product of the fluctuating parts $\text{cov}(a, b) := \langle a|b \rangle_2$, and the variance as $\text{var}(a) := \text{cov}(a, a)$. Pythagoras’s theorem can be applied here as well to give $\|a\|^2 = \|\hat{a}\|^2 + \|\hat{a}\|^2 = \mathbb{E}(a^2) + \text{var}(a)$. Two RVs $a, b \in A$ are uncorrelated iff their covariance vanishes: $\text{cov}(a, b) = 0$, i.e. if their fluctuating parts are orthogonal $\langle a|a \rangle_2 = 0$. The two RVs $a, b \in A$ are independent iff $\text{cov}(Q_1(a, a^*), Q_2(b, b^*)) = 0$ for all $Q_1, Q_2 \in \mathbb{P}_2(\Omega)$ with $\mathbb{E}(Q_1(a, a^*)) = \mathbb{E}(Q_2(b, b^*)) = 0$, i.e. if the centred subspaces of the algebras generated by them are orthogonal, i.e. $(\mathbb{C}\{a, a^*\} \cap \ker \mathbb{E}) \perp (\mathbb{C}\{b, b^*\} \cap \ker \mathbb{E})$. In the non-commutative case, the concept of freeness and free independence becomes more important, cf. [27, 10, 21, 25], but we shall not further pursue this topic here.

We have seen that the example algebra $A_s = L_{0s}(\Omega)$ satisfies all the requirements and is thus a concrete example of a probability algebra, and generates the Hilbert space.
3.3 More examples

For the example algebra \( \mathcal{A}_s = L_{0s}(\Omega) \) it is also well known that one may define the \( \ell_p \)-norms for any \( 1 \leq p < \infty \) via 
\[
\|\phi\|_p := \mathrm{E} \left( (\phi \ast \phi)^{p/2} \right) = \int_{\Omega} |\phi(\omega)|^p \mathrm{d}\mathbb{P}(\omega).
\]
For \( p = \infty \) one sets \( \|\phi\|_\infty := \sup_{\Omega} |\phi| \). The completion of \( \mathcal{A}_s = L_{0s}(\Omega) \) in any of the norms \( \|\cdot\|_p \) for \( 1 \leq p \leq \infty \) gives the familiar Banach spaces \( L_p(\Omega) \). This gives two more concrete examples of probability algebras, namely \( L_\infty(\Omega) \) and \( L_\infty^{-}(\Omega) := \bigcap_{1 \leq p < \infty} L_p(\Omega) \).

The last example contains unbounded RVs, e.g. all the Gaussian RVs. Obviously one has \( \mathcal{A}_s = L_{0s}(\Omega) \subset L_\infty(\Omega) \subset L_\infty^{-}(\Omega) \subset L_0(\Omega) \), i.e. the classical simple RVs in \( \mathcal{A}_s \) are a probability sub-algebra of the classical bounded RVs \( L_\infty(\Omega) \), which is a probability sub-algebra of the algebra \( L_\infty^{-}(\Omega) \) of unbounded RVs which have finite moments of any order, which in turn is a sub-*-algebra of the *-algebra of all RVs, which is not a probability algebra as not every element has a finite expected value.

One more classical example which should be mentioned is the case when \( \Omega \) is in addition a compact Hausdorff topological space, the \( \sigma \)-algebra \( \mathfrak{F} \) is the Borel algebra \( \mathfrak{B}(\Omega) \), and the probability measure a Radon measure. Then the RVs given by the continuous complex-valued functions \( C(\Omega; \mathbb{C}) \)—for brevity only \( C(\Omega) \)—are a sub-probability algebra of \( L_\infty(\Omega) \), in fact a \( C^* \)-algebra— a Banach space in the \( \|\cdot\|_\infty \) norm such that \( \|\phi \psi\|_\infty \leq \|\phi\|_\infty \|\psi\|_\infty \) and \( \|\phi \ast \phi^*\|_\infty = \|\phi\|_\infty \|\phi^*\|_\infty = \|\phi\|_\infty^2 \) such that the product and adjoint are continuous—called the uniform algebra on \( \Omega \).

These are all examples of classical commutative resp. Abelian algebras of RVs with the state the usual Lebesgue integral (i.e. the usual expected value) w.r.t the measure \( \mathbb{P} \). The bounded RVs \( L_\infty(\Omega) \) are a maximal Abelian \( W^* \)-algebra \([24]\)—a \( W^* \)-algebra is in simplest terms defined as a \( C^* \)-algebra which as Banach space is the dual of another Banach space. It may be shown conversely that any complex maximal Abelian \( W^* \)-probability algebra \( \mathcal{A} \) is isomorphic to an \( L_\infty \)-algebra on a probability space, a result that will be used in the sequel—this is the Segal representation. Thus the algebraic approach to probability can completely recover the classical approach due to Kolmogorov which starts from measure spaces and defines RVs as measurable functions. Similarly it can be shown that unital Abelian \( C^* \)-algebras are isomorphic to the uniform algebra on a compact space—the Gel’fand representation. Abelian algebras of this kind are therefore often called ‘function algebras’.

Let us now consider some non-commutative examples. A simple one is \( M_n(\mathbb{C}, n) = \mathbb{C}^{n \times n} \), the algebra of complex \( n \times n \) matrices with complex conjugate transposition as involution. The language of the algebra is completely the same, except that projections in the abstract setting—which are self-adjoint—are called orthogonal projections here. This kind of algebra corresponds to RVs which can take no more than \( n \) different values. Let \( \varrho \in M_n(\mathbb{C}, n) \) be a self-adjoint positive definite matrix with \( \text{tr} \varrho = 1 \), called a density matrix. Then \( \mathbb{E}_\varrho(A) := \text{tr}(\varrho A) \) is a faithful state. Of course any sub-algebra of \( M_n(\mathbb{C}, n) \) which contains the identity matrix is another example, and the diagonal matrices are an example of a commutative sub-algebra. More powerful is the algebra \( M(L_\infty(\Omega), n) \) of \( n \times n \) random matrices with entries from \( L_\infty(\Omega) \), and the expectation is the expected value of a matrix state, i.e. for \( A \in M(L_\infty(\Omega), n) \) one may set \( \mathbb{E}(A) := \int_\Omega \mathbb{E}_\varrho(A(\omega)) \mathrm{d}\mathbb{P}(\omega) \).

An example generalising the previous case is \( \mathcal{L}(\mathcal{H}) \), the algebra of bounded linear
maps on a complex Hilbert space \( \mathcal{H} \) with the adjoint taking the rôle of the involution, or any unital sub-algebra thereof. \( \mathcal{L}(\mathcal{H}) \) is a \( W^* \)-algebra, non-commutative if \( \dim \mathcal{H} > 1 \). If \( \varrho \in \mathcal{L}(\mathcal{H}) \) is a nuclear resp. trace-class positive definite operator with unit trace \( \text{tr} \varrho = 1 \)—called again a density matrix—then a state may be defined for \( A \in \mathcal{L}(\mathcal{H}) \) as \( \mathbb{E}_\varrho(A) := \text{tr}(\varrho A) \). The example is in some way universal, as with the Gel’fand-Naimark-Segal (GNS) construction any algebra with faithful state may be embedded (faithfully represented) into an algebra of operators on a complex Hilbert space \([24, 22, 23, 27]\); namely \( a \in A \) is represented as \( L_a : A \ni b \mapsto ab \in A \) in \( \mathcal{L}(L_2(A)) \).

When the Hilbert space \( \mathcal{H} \) in question is a Lebesgue space \( L_2(\Omega) \), then any \( \kappa \in L_\infty(\Omega) \) can be represented as a linear map \( M_\kappa : L_2(\Omega) \ni \varphi \mapsto \kappa \varphi \in L_2(\Omega) \). Thus the Abelian algebra \( L_\infty(\Omega) \) is represented as a maximal Abelian \( W^* \)-sub-algebra of \( \mathcal{L}(L_2(\Omega)) \), it is called the multiplication algebra of \( L_2(\Omega) \).

### 3.4 Weights, spectrum, and spectral calculus

In this abstract setting we have now seen RVs and their expectation and what can be deduced from these concepts. The question arises now as to what an actual observation or sample of such an RV really is. To this end a bit more theory is needed. First it turns out that with non-commuting observables, in an experiment or other observation, only commuting observables (self-adjoint elements) can be observed simultaneously \([28]\). This is implied by the uncertainty relation. Let \( a, b \in A \) be two self-adjoint elements resp. observables, and \( [a, b] = ab - ba \) be their commutator. The Cauchy-Bunyakovsky-Schwarz inequality for non-commutative variables easily gives the uncertainty relation

\[
\text{var}(a)\text{var}(b) \geq \mathbb{E}(i[a, b])^2 / 4 \geq 0;
\]

where the the expected value on the right hand side is real, it is easy to see that \( i[a, b] \) is self-adjoint. Once say \( a \) has been observed, it is known and its variance vanishes. This shows that it is not possible to observe \( a \) and \( b \) simultaneously, unless they commute.

Therefore the way to approach this is to consider for some observation or experiment all relevant commuting RVs which can be observed simultaneously, say \( a_1, \ldots, a_k \in A \). They, and hence any powers or polynomials in commuting variables of them can be observed simultaneously, in fact any element of the Abelian sub-probability algebra \( \mathcal{A}_x := \mathbb{C}[a_1, \ldots, a_k] \subseteq A \) generated by them. We shall shortly add more functions beyond polynomials to this list.

As \( a_1, \ldots, a_k \) commute, so do the linear operators \( L_{a_1}, \ldots, L_{a_k} \) in the GNS-representation, and the algebra \( \mathcal{L}_x := \mathbb{C}[L_{a_1}, \ldots, L_{a_k}] \subseteq \mathcal{L}(L_2(A)) \) generated by them is an Abelian algebra isomorphic to \( \mathcal{A}_x \). It is worthwhile at this point to remember that for linear operators the fact that they commute means that they have the same spectral resolution, and the Gel’fand representation of Abelian \( C^* \)-algebras and the Segal representation of maximal Abelian \( W^* \)-algebras can now be used \([8, 24, 17]\). This can in fact be employed to obtain a version of the spectral theorem for linear operators. We defer this for a moment in order to point out the importance of spectral theory to the subject.

The concept of a state as a self-adjoint positive normalised linear functional was already introduced. The set of all possible states \( S(\mathcal{A}_x) \) is clearly a subset of the dual \( \mathcal{A}_x^* \), and due to the normalisation they are actually on the unit ball of \( \mathcal{A}_x^* \). One can easily show that \( S(\mathcal{A}_x) \) is a closed, convex, and hence weak-* compact subset of the unit ball of the dual. The extreme points of \( S(\mathcal{A}_x) \) are called pure states, and their convex combinations are weak-* dense in \( S(\mathcal{A}_x) \). In the case of classical RVs, the states are naturally represen-
ted by probability measures, which are known to form a convex weak-* compact subset of the unit ball in the space of all measures of bounded total variation. The extreme points in that case are well known to be Dirac-δ-measures.

A weight, or more specifically a representational weight, also called a multiplicative character, \( \alpha \in S(\mathcal{A}_x) \) is a special kind of state, namely one that is also an algebra *-homomorphism \( \mathcal{A}_x \to \mathbb{C} \). This means that for \( b, c \in \mathcal{A}_x \) and \( \eta, \zeta \in \mathbb{C} \) it holds not only that \( \alpha(\eta b + \zeta c) = \langle \alpha, \eta b + \zeta c \rangle = \eta\alpha(b) + \zeta\alpha(c) \) (linearity), but also that \( \alpha(b^*) = (\alpha(a))^* \) and \( \alpha(bc) = \alpha(b)\alpha(c) \). The set of all weights —one-dimensional representations of \( \mathcal{A}_x \)— is denoted by \( \hat{\mathcal{A}}_x \) and is called the spectrum of \( \mathcal{A}_x \); it is also a weak-* compact subset \( \hat{\mathcal{A}}_x \subset S(\mathcal{A}_x) \subset B_1(0) \subset \mathcal{A}_x^* \) of the unit ball of the dual. In the case of classical algebras of RVs the Dirac-δ-measures are a good example of weights.

The best known meaning of the term spectrum is certainly when used with regard to a linear map or an element \( c \in \mathcal{A} \) as the set \( \sigma(c) = \{ \lambda \in \mathbb{C} \mid c - \lambda \mathbf{1} \text{ is not invertible} \} \). Now let \( \alpha \in \hat{\mathcal{A}}_x \) be any weight, and \( b \in \mathcal{A}_x \). If \( b \) is invertible with inverse \( b^{-1} \), then \( e = bb^{-1} \) implies \( 1 = \alpha(e) = \alpha(bb^{-1}) = \alpha(b)\alpha(b^{-1}) \), and hence \( \alpha(b) \neq 0 \). Invertible elements can thus not be mapped to 0 by any weight, i.e. any element in the spectrum \( \hat{\mathcal{A}}_x \).

Looking at \( b = c - \alpha(c)e \), one sees that \( \alpha(b) = \alpha(c - \alpha(c)e) = \alpha(c) - \alpha(c)\alpha(e) = 0 \), hence \( b = c - \alpha(c)e \) can not be invertible and therefore \( \alpha(c) \in \sigma(c) \) for any weight \( \alpha \in \hat{\mathcal{A}}_x \). This explains the name spectrum for the set of weights \( \hat{\mathcal{A}}_x \), i.e. each \( \alpha(c) \) is in the spectrum of \( c \). In fact, for any \( \lambda \in \sigma(c) \) there is a \( \alpha \in \hat{\mathcal{A}}_x \) such that \( \alpha(c) = \lambda \).

The interpretation now is that when one observes a RV, i.e. sees a sample, then one sees the action of some weight on the RV. Hence the possible values (sample observations) of an abstract RV \( a \in \mathcal{A} \) are given by the action of all weights on the RV, \( \{ \alpha(a) = \langle \alpha, a \rangle \mid \alpha \in \hat{\mathcal{A}}_x \} \). Therefore one concludes that all possible observations of a RV \( a \) are given by its spectrum \( \sigma(a) \); and as the observables are self-adjoint the spectrum is real, \( \sigma(a) \subset \mathbb{R} \).

Considering general non-commutative probability algebras, the spectrum of the algebra is often empty as there are no non-zero one-dimensional representations—another sign that these observables cannot be observed simultaneously—but in the case of Abelian algebras like \( L_x \), the ones we are considering when examining a concrete experiment or observation, the Gel’fand and Segal representations tell us that the spectrum is rich enough. One may hence use spectral theory of linear operators to determine the set of possible values, as \( a \in \mathcal{A}_x \) and \( L_x \in \mathcal{L}_x \) in the GNS-construction have the same spectrum.

The representation theorems state [24] that an Abelian probability algebra is isomorphic to a sub-algebra of \( L_\infty(\mathcal{X}) \) on a compact Hausdorff space \( \mathcal{X} \). In fact, the compact space may be chosen as \( \mathcal{X} := \hat{\mathcal{A}}_x \). The version of the spectral theorem for linear operators which is most useful here—and will be used again for a different purpose in Section 4—is that an Abelian algebra of operators like \( \mathcal{L}_x \) is not only isomorphic but unitarily equivalent to a sub-algebra of the multiplication algebra on some measure space \( \mathcal{Y} \) [24] with total measure equal to unity, i.e. a classical probability space. The spectrum of such a multiplication operator \( M_\kappa \) with the function or RV \( \kappa \in L_\infty(\mathcal{Y}) \) [24] is the essential range of the function \( \kappa \). Hence any of the commuting RVs \( a_\ell \) resp. \( L_\ell \), is represented by a multiplication operator \( M_\kappa_\ell \), and hence as algebra by an RV \( \kappa_\ell \in L_\infty(\mathcal{Y}) \). We may thus say that \( \sigma(a_\ell) = \sigma(L_\ell) = \sigma(M_\kappa_\ell) = \sigma(\kappa_\ell) = \text{ess range } \kappa_\ell \).

In the classical framework where RVs are measurable maps on a probability space, one important and relevant fact is that the composition of measurable functions is again a measurable function, and one can form new RVs by applying a measurable function to an existing RV. In the algebraic framework presented so far only polynomials—which
are kind of natural when dealing with algebras—have appeared. Now if \( f : \mathbb{R} \to \mathbb{R} \)—or more generally \( f : \sigma(a) \subseteq \mathbb{R} \to \mathbb{R} \)—is an essentially bounded measurable function, so is \( \gamma = f \circ \kappa \in \mathcal{L}(\mathcal{Y}) \). Hence there is a corresponding \( M_\gamma := f(M_\kappa) \) in the multiplication algebra, and a \( L_\gamma := f(L_\kappa) \in \mathcal{L}(L_2(\mathcal{A})) \), and a \( g := f(a) \) in the weak-* closure of \( \mathcal{A}_x \).

This defines the function \( f \) now on the algebra \( \mathcal{L}_x \) or \( \mathcal{A}_x \); and is the essence of spectral calculus, used here to obtain new RVs by applying a measurable function \( f \).

### 3.5 Extensions

With the spectral calculus in place, one may define non-commutative analogues of the classical \( \mathcal{L}_p \)-spaces for all \( 1 \leq p \leq \infty \) by extending any probability algebra \( \mathcal{A} \) through completion in a certain uniform topology, and not just for \( p = 2 \) as above. First note that for a positive element \( a = b^*b \in \mathcal{A} \) one can always find a unique positive \( c \in \mathcal{A} \) such that \( a = cc = c^2 \) via spectral calculus, as this \( c = a^{1/2} \in \mathcal{A} \) is the square root. This allows one to define for any \( a \in \mathcal{A} \) the absolute value as the positive element \( |a| := (a^*a)^{1/2} \in \mathcal{A} \). Similarly one may compute the \( p \)-th power for real \( p > 0 \). For \( 1 \leq p < \infty \) the expression \( \|a\|_p^p := \mathbb{E}(|a|^p) \) defines the \( p \)-th power of a norm. Completion of \( \mathcal{A} \) w.r.t. any of those norms gives non-commutative Banach spaces \( \mathcal{L}_p(\mathcal{A}) \), and this agrees for \( p = 2 \) with the previous definition. It also immediately gives a new algebra \( \mathcal{L}_\infty(\mathcal{A}) := \bigcap_{1 \leq p < \infty} \mathcal{L}_p(\mathcal{A}) \).

Recalling the spectral calculus from the end of the previous Subsection 3.4, one may now state that \( \mathcal{L}_p(\mathcal{A}) \) contains elements \( f(a) \) for \( a \in \mathcal{A} \) and certain measurable functions \( f \in \mathcal{L}_0(\sigma(\hat{a})) \). These measurable functions have to be such that in the representation of the Abelian probability sub-algebra \( \mathbb{C}[a] \), where \( a \) is represented by the multiplication operator \( M_\kappa \) on \( L_2(\mathcal{Y}) \) with \( \kappa \in \mathcal{L}_\infty(\mathcal{Y}) \), and where \( \sigma(a) = \sigma(\kappa) = \text{ess range } \kappa \), the composite function satisfies \( f \circ \kappa \in \mathcal{L}_p(\mathcal{A}) \).

For \( p = \infty \) one has to look at the representation of \( a \in \mathcal{A} \) through the linear map \( L_a \) in the GNS-construction above and define the \( \|a\|_\infty := \|L_a\|_{op} \) as the operator norm of \( L_a \), effectively \( \|a\|_\infty := \sup_{\|\cdot\|_2=1} \|ba\|_2 \). One may also define a topology corresponding to the weak operator topology through the semi-norms \( q_{b,c}(a) := \|L_a(b)c\|_2 = [\mathbb{E}(c^*ab)] \). Completion of the sub-algebra \( \mathcal{A}_\infty := \{a \mid \|a\|_\infty < \infty \} \subseteq \mathcal{A} \) with finite \( \infty \)-norm w.r.t. the uniform locally convex topology generated by the semi-norms \( q_{b,c}(\cdot) \) gives the probability \( W^* \)-algebra \( \mathcal{L}_\infty(\mathcal{A}) \). This shows that the \( \mathcal{L}_p \)-spaces of non-commutative RVs can be generated just as in the classical Abelian case.

As already mentioned, the space \( L_2(\mathcal{A}) \) is a possible candidate for the space \( \mathcal{Y} \) appearing in the probabilistic model \( S : \mathcal{Y} \to \mathcal{Y} \). Other candidates may be generated by the following very general construction: if \( \mathcal{H} \) is a Hilbert space with inner product \( \langle \cdot | \cdot \rangle_0 \), and \( \mathcal{A} \) a possibly unbounded self-adjoint positive operator in \( \mathcal{H} \) with dense domain \( dom \mathcal{A} \), one may via spectral calculus define \( A^s \) for any \( s > 0 \) with dense domain \( dom A^s \). The positive definite sesqui-linear form given by \( \langle f|g \rangle_s := \langle f|g \rangle_0 + \langle A^s f|g \rangle_0 \) for \( f, g \in dom A^s \) defines an inner product on \( dom A^s \), the completion of which in the associated topology defines the densely embedded Hilbert space \( \mathcal{H}_s \hookrightarrow \mathcal{H} \) Obviously one also has dense embeddings \( \mathcal{H}_s \hookrightarrow \mathcal{H}_t \) for \( s > t > 0 \). Identifying \( \mathcal{H} \) with its dual and denoting the dual of \( \mathcal{H}_s \) by \( \mathcal{H}_{-s} \), one obtains Gel’fand triplets \([8, 17]\) or ‘sandwiched’ dense embeddings \( \mathcal{H}_s \hookrightarrow \mathcal{H} \hookrightarrow \mathcal{H}_{-s} \) of Hilbert spaces. One may even go a step further and introduce the projective limit \( S = \lim_{s \to 0} \mathcal{H}_s \), depending on \( \mathcal{A} \) often a nuclear space, which in our case usually will be a new probability algebra. The dual construction of inductive limit \( S^* = \lim_{s \to 0} \mathcal{H}_{-s} \) then generates the dual space of generalised objects, like the distributions in the sense of...
Sobolev and Schwartz.

It is worthwhile to recall that the familiar Sobolev-Hilbert spaces $H^s(\mathbb{R}^n)$ are generated in this way by taking $\mathcal{H} = H^0(\mathbb{R}^n) = L_2(\mathbb{R}^n)$ and $A = -\Delta + M_{1/2}$, essentially the negative Laplacian added to a multiplication operator. Then the Schwartz space of rapidly decaying smooth functions $\mathcal{S}(\mathbb{R}^n)$ is the projective limit and additionally an Abelian algebra, and its dual $\mathcal{S}'(\mathbb{R}^n)$, the inductive limit, is the Schwartz space of tempered distributions.

The same device can be used here by choosing $\mathcal{H} = L_2(\mathcal{A})$—a space which is naturally given by the expectation state—and an appropriate operator $A$; then all the spaces $\mathcal{H}_t$, $t \in \mathbb{R}$, are possible candidates for $\mathcal{V}$, and the ‘regularity’ of the RVs in $\mathcal{V} := \mathcal{H}_t$ can be controlled by the parameter $t \in \mathbb{R}$. For $t < 0$ these are spaces of ‘generalised’ RVs, only defined via the duality, similar to the Sobolev-Hilbert spaces with negative exponent.

One possible classical choice for the linear operator $A$ for $\mathcal{H} = L_2(\Omega) = L_2(L_\infty(\Omega))$ is the following: denote by $H^m_n$, $n \in \mathbb{N}_0$, the $n$-th homogeneous chaos in Wiener’s polynomial chaos decomposition $\mathcal{H} = \bigoplus_{n=0}^\infty H^m_n$; and define $A$ by $Ah := nh$ for any $h \in H^m_n$; a self-adjoint operator with spectrum $\sigma(A) = \mathbb{N}_0$, called the number operator. More examples of Hilbert spaces of RVs which can be generated in this way may be found in [11, 12], they are all practically defined with the help of the Wiener-Itô polynomial chaos expansion and are all possible candidates for the space $\mathcal{V}$.

### 3.6 Weak or generalised distributions

In any case, this construction of a unital algebra with involution and faithful state leads to an inner product and Hilbert space $\mathcal{V}$, and the state $\mathbb{E}$ may be extended as continuous functional onto the whole space $\mathcal{V}$. This may be used in the mapping $S : \mathcal{F} \to \mathcal{V}$ in Section 2. With the possibility of also using non-commutative algebras, this approach also allows to deal with objects such as random matrices, or more generally random fields of tensors of even order [20, 19], which is much more cumbersome in the traditional measure space approach. Our first example $\mathcal{A}_s = L_{0s}(\Omega)$ also indicates that the algebraic approach is more general and can completely recover the measure space approach [24, 27, 10, 25, 21]. The state takes the place of the usual expectation operator, and it has all its usual properties.

Nevertheless, even in the general abstract setting of a probability algebra, it is possible to define a distribution probability measure or ‘law’ on $\mathbb{R}$ for any non-commutative self-adjoint RV, i.e. an observable. Classically, for a real-valued or self-adjoint RV $\phi \in L_0(\Omega)$ the law of $\phi$ is the push-forward $\phi_* \mathbb{P}$ of the probability measure $\mathbb{P}$, given for an element $B$ of the Borel-$\sigma$-algebra $\mathcal{B}(\mathbb{R})$ by $\phi_* \mathbb{P}(B) := \mathbb{P}(\phi^{-1}(B))$.

In the abstract setting, for any $a \in \mathcal{A}$ one may define the law of $a$ as a map $\tau_a : \Pi_1 \to \mathcal{C}$ which assigns to any polynomial $Q \in \Pi_1$ the number $\tau_a(Q) := \mathbb{E}(Q(a))$. With $a \in \mathcal{A}$ self-adjoint, we know that the spectrum is real: $\sigma(a) \subseteq \mathbb{R}$. Let $J \subset \mathbb{R}$ be a compact interval which contains the spectrum $\sigma(a)$. The polynomials $\Pi_1^J$ with real co-efficients are known to be dense in $C(J)$ due to the Stone-Weierstrass theorem, and $\tau_a$ can be shown to be a continuous map, hence may be extended to all of $C(J)$. From the Riesz-Markov representation theorem it now follows that there is a Radon probability measure $\mathbb{P}_a$ such that $\int_J Q(t) \mathbb{P}_a(dt) = \tau_a(Q)$ for any $Q \in \Pi_1^J$, called the distribution measure or law of the self-adjoint RV $a \in \mathcal{A}$.

This more general approach via a mapping like $S : \mathcal{F} \to \mathcal{V}$ and abstract probability algebras $\mathcal{A}$ related to $\mathcal{V}$ is also needed in many concrete analytic situations. As a simple
example, consider, as in Section 1 and Section 2, a RV $\varsigma$ with values in an infinite dimensional Hilbert space $\mathcal{H}$. For this to be an ‘honest’ RV, the push-forward distribution $\varsigma_{\mathcal{P}} = \mathcal{P} \circ \varsigma^{-1}$ of the probability measure $\mathcal{P}$ should be a $\sigma$-additive measure on the Borel sets $\mathcal{B}(\mathcal{H})$ of $\mathcal{H}$. It is well known that on a Hilbert space this is only possible (Sazonov’s theorem, cf. e.g. [4, 20]) if the correlation $C_\varsigma$ already mentioned in Section 2 is a nuclear or trace-class operator. In particular, there is no iso-Gaussian measure—i.e. where $C_\varsigma = I$ is the identity, invariant under unitaries—on an infinite-dimensional Hilbert space; one has to resort to so-called cylindrical pro-measures (which are not $\sigma$-additive) or enlargements of the Hilbert space.

The formulations such as with the mapping $S$ or $P$ from above or Section 2 circumvent all the difficulties mentioned in the previous paragraph with non-nuclear correlation or covariance operators, and such an assignment is called a weak distribution or generalised RV [22, 23, 9, 24] resp. a generalised process [8]. For example the aforementioned iso-Gaussian weak distribution resp. generalised process—this is also called white noise on the Hilbert space $\mathcal{H}$—is very simply defined: Pick any complete orthonormal system $\{s_n\}_n$ in $\mathcal{H}$ and an infinite sequence of independent identically distributed (iid) standard Gaussian RVs $\{s_n\}$ (zero mean, unit variance) as CONS, and let $\mathcal{H}$ be the Hilbert space generated by them. Define a linear map $W : \mathcal{H} \ni s_n \mapsto s_n \in \mathcal{H}$, and it is clear that its covariance is $C_W = W^*W = I$, as $W$ is by construction unitary. Hence $W$ defines a weak white noise distribution on $\mathcal{H}$. Other extensions covered by this use of weak distributions are the cases when the covariance has continuous spectrum, as often happens for translation invariant covariance kernels [14] which are diagonalised by the Fourier transform [5].

From all this we conclude that one may define a stochastic model as a weak distribution on $\mathcal{H}$ via a linear map $S : \mathcal{H} \to \mathcal{V}$, where $\mathcal{V}$ was generated by a probability algebra $\mathcal{A}$ as described above, and similarly for $P : \mathcal{V} \to \mathcal{V}$. For a conventional probability model we assume that the algebra is Abelian, but the non-commutative case is useful to model e.g. random matrices or tensor fields [20, 19]. For a dynamical system like the one mentioned in Section 1 the equality in the equation is to be understood in a probabilistically weak sense as just described: both sides of the equation are mapped into the space $\mathcal{V}$, and have to be equal as elements of that space, i.e. in a $\mathcal{V}$-weak sense. First we spell out the meaning of the map $P$:

\[
P(\dot{v}(t)) = P(A(\varsigma, \mu; v(t))) + P(f(\varsigma, \mu; t)) \iff \forall w \in \mathcal{V} : \langle \dot{v}(t) | w \rangle_\mathcal{V} = \langle A(\varsigma, \mu; v(t)) | w \rangle_\mathcal{V} + \langle f(\varsigma, \mu; t) | w \rangle_\mathcal{V}, \tag{7}\]

as an element of $\mathcal{V}$, which in detail in $\mathcal{V}$ means

\[
\forall \varphi \in \mathcal{V} : \langle P(\dot{v}(t)) | \varphi \rangle_\mathcal{V} = \langle P(A(\varsigma, \mu; v(t))) | \varphi \rangle_\mathcal{V} + \langle P(f(\varsigma, \mu; t)) | \varphi \rangle_\mathcal{V}. \tag{8}\]

This allows one to deal with a much wider range of probabilistic situations, including white noise as already alluded to, as well as white noise or a Wiener process in time, as the Itô-integral can be understood as a weak stochastic distribution [14]. The way Eq. (7) and Eq. (8) are formulated also immediately suggests numerical approximations by Galerkin’s method—called the stochastic Galerkin method [17]—using finite dimensional subspaces $\mathcal{V}_n \subseteq \mathcal{V}$ and $\mathcal{V}_m \subseteq \mathcal{V}$.

It may be noted that this whole development is analogous on how generalised functions or distributions are introduced in the Sobolev-Schwartz framework. There they are linear
In general, one may specify [15, 20, 19] a densely defined map $C$ in $\mathcal{U}$ through the bilinear form
\[ \forall u, v \in \mathcal{U} : \langle C u | v \rangle_{\mathcal{U}} := \langle Ru | Rv \rangle_{\mathcal{Q}}. \quad (9) \]

The map $C = R^* R$, may be called the ‘correlation’ operator and is by construction self-adjoint and positive, and if $R$ is continuous so is $C$. In case the inner product $\langle \cdot | \cdot \rangle_{\mathcal{Q}}$ comes from a measure $\varpi$ on $\mathcal{M}$, so that for two functions $\phi$ and $\psi$ on $\mathcal{M}$, one has
\[ \langle \phi | \psi \rangle_{\mathcal{Q}} := \int_{\mathcal{M}} \phi(\mu) \psi(\mu) \, \varpi(d\mu), \]
the usual formula for the correlation. The space $\mathcal{Q}$ may then be taken as $\mathcal{Q} := L^2(\mathcal{M}, \varpi)$. A special case is when $\varpi$ is a probability measure, $\varpi(\mathcal{M}) = 1$, as for $\mathcal{M} = \mathcal{O}$ and $\varpi = \mathcal{P}$, this inspired the term ‘correlation’ operator. In terms of the developments in Section 3, the Hilbert space $\mathcal{Q}$ would be replaced by any of the candidates for $\mathcal{V}$ and instead of $C = R^* R$ we would be investigating $C_\varsigma = S^* S$ or $C_v = P^* P$.

The spectral theorem for operators in a Hilbert space was already used in Section 3, but here we start in a gentler way. To make everything as simple as possible to explain the main underlying idea, assume first that $C$ is a non-singular trace class or nuclear operator. This means that it is compact, the spectrum $\sigma(C)$ is a point spectrum, has a CONS $\{v_m\}_m \subset \mathcal{U}$ consisting of eigenvectors, with each eigenvalue $\lambda_m \geq \lambda_{m+1} \cdots \geq 0$ positive and counted decreasingly according to their finite multiplicity, and has finite trace $\text{tr} C = \sum_m \lambda_m < \infty$. Then a version of the spectral decomposition of $C$ is
\[ C = \sum_m \lambda_m (v_m \otimes v_m). \quad (10) \]

Use this CONS to define a new CONS $\{s_m\}_m$ in $\mathcal{Q}$: $\lambda_m^{1/2} s_m := R v_m$, to obtain the corresponding singular value decomposition (SVD) of $R$ and $R^*$:
\[ R = \sum_m \sqrt{\lambda_m} (s_m \otimes v_m); \quad R^* = \sum_m \sqrt{\lambda_m} (v_m \otimes s_m); \]
\[ r(\mu) = \sum_m \sqrt{\lambda_m} s_m(\mu) v_m = \sum_m (R^* s_m)(\mu), \quad (11) \]

The set $\varsigma(R) = \{\sqrt{\lambda_m}\}_m = \sqrt{\sigma(C)} \subset \mathbb{R}_+$ are the singular values of $R$ and $R^*$. The last relation is the so-called Karhunen-Loève expansion or proper orthogonal decomposition.
(POD). The finite trace condition of \( C \) translates into the fact that \( r \) is in \( \mathcal{U} \otimes \mathcal{Q} \). If in that relation the sum is truncated at \( n \in \mathbb{N} \), i.e.

\[
r(\mu) \approx r_{\text{ROM}}(\mu) = \sum_{m=1}^{n} \sqrt{\lambda_m} s_m(\mu) v_m = \sum_{m=1}^{n} (R^* s_m)(\mu),
\]

we obtain the best \( n \)-term approximation to \( r(\mu) \) in the norm of \( \mathcal{U} \). Observe that \( r \) is linear in the \( s_m \). This means that by choosing the ‘co-ordinate transformation’ \( \mathcal{M} \ni \mu \mapsto (s_1(\mu), \ldots, s_m(\mu), \ldots) \in \mathbb{R}^n \) one obtains a linear / affine representation where the first co-ordinates are the most important ones. For the stochastic cases \( C \gamma = S^* S \) and \( C_v = P^* P \) we point out again as in Section 3 that the nuclearity of \( C \) resp. \( C_v \) is necessary for the existence of a measurable map \( \gamma : \Omega \to \mathcal{F} \) resp. \( v(\cdot, \mu; t) : \Omega \to \mathcal{V} \).

Equivalently this means that \( S \) resp. \( P \) has to be a Hilbert-Schmidt operator, e.g. \[\Omega\], a condition which severely restricts stochastic models. There is a practical need to consider more general classes of correlation operators, as already evidenced in the seminal paper by Karhunen \[13, 14\], where integral transforms for representations as in Eq. (12) were investigated. This more general view is for example necessary to consider homogeneous or stationary random fields or stochastic processes, cf. e.g. \[16\].

One formulation of the spectral decomposition extending Eq. (10), already used implicitly in Section 3, which does not require \( C \) to be nuclear \[\Omega\] [24], nor do \( C \) or \( R \) have to be continuous, which was used already in Section 3 and has to be applied here to the Abelian algebra \( \mathbb{C}[L] \), as is follows. The densely defined self-adjoint and positive operator \( C : \mathcal{U} \to \mathcal{U} \) is unitarily equivalent with a multiplication operator \( M_\gamma \) on an appropriate measure space \( \mathcal{T} \),

\[
C = VM_\gamma V^*,
\]

where the unitary map is \( V : L_2(\mathcal{T}) \to \mathcal{U} \), and \( M_\gamma \) multiplies a \( \psi \in L_2(\mathcal{T}) \) with a real-valued function \( \gamma ; M_\gamma : \psi \mapsto \gamma \psi \). In case \( C \) is bounded, so is \( \gamma \in L_\infty(\mathcal{T}) \). As \( C \) is positive, \( \gamma(t) \geq 0 \) for \( t \in \mathcal{T} \), and the essential range of \( \gamma \) is the spectrum of \( C \). In Section 3 this was already used for the Abelian algebra \( \mathbb{C}[a] \) resp. \( \mathbb{C}[L_\gamma] \), which says then that any member of that algebra is unitarily equivalent to a multiplication operator.

As already indicated, via spectral calculus one may define the square root \( M_\gamma^{1/2} := M_\gamma^{1/2} \), and a factorisation similar to \( C = R^* R \) is obtained via \( C = (VM_\gamma V^*)(VM_\gamma V^*)^* =: G^* G \). From this factorisation and the spectral decomposition Eq. (13) follows another singular value decomposition (SVD) of \( R \) and \( R^* \), which is

\[
R = UM_\gamma V^*, \quad R^* = VM_\gamma U^*;
\]

where \( U : L_2(\mathcal{T}) \to \mathcal{Q} \) is a unitary operator. Having \( M_\gamma^{1/2} \) allows us to compute the square root of \( C \): \( C^{1/2} = VM_\gamma^{1/2} V^* \), and from it the self-adjoint positive definite factorisation \( C = C^{1/2} C^{1/2} \).

Consider now an arbitrary factorisation \( C = B^* B \), where \( B : \mathcal{U} \to \mathcal{H} \) is a map to a Hilbert space \( \mathcal{H} \). Any two such factorisations \( B_1 : \mathcal{U} \to \mathcal{H}_1 \) and \( B_2 : \mathcal{U} \to \mathcal{H}_2 \) with \( C = B_1^* B_1 = B_2^* B_2 \) are \[20\] unitarily equivalent in that there is a unitary map \( X_{21} : \mathcal{H}_1 \to \mathcal{H}_2 \) such that \( B_2 = X_{21} B_1 \). Each such factorisation is also unitarily equivalent to \( R \), i.e. there is a unitary \( X : \mathcal{H} \to \mathcal{Q} \) such that \( R = XB \). For finite dimensional spaces, a favourite choice for such a decomposition of \( C \) is the Cholesky factorisation \( C = LL^* \), where \( B = L^* \) is represented by an upper triangular matrix.
Let us go back to the situation of Eq. (10) and how the SVD of the factors \( R \) Eq. (11) in the factorisation \( C = R^*R \) was generated. In the same way a SVD of any of the factorisations just considered may be generated with left-singular vectors \( h_m := BC^{-1}R^*s_m = BC^{-1/2}v_m \), plus the analogue of Eq. (11), i.e.

\[
B = \sum_m \sqrt{\lambda_m}(h_m \otimes v_m); \quad B^* = \sum_m \sqrt{\lambda_m}(v_m \otimes h_m); \quad r = \sum_m \sqrt{\lambda_m}h_mv_m = \sum_mB^*h_m,
\]

and with \( W = X^*U \):

\[
B = WM\sqrt{\pi}V^*, \quad R^* = VM\sqrt{\pi}W^*.
\]

The left-singular vectors \( h_m \) can now be thought of living on any of the spaces which appeared in the factorisation, i.e. generically \( \mathcal{H} \), for which we have just seen the examples \( \mathcal{H} = L_2(\mathcal{T}) \) and \( \mathcal{H} = \mathcal{U} \) (not necessarily very useful) [20].

Instead of \( C = B^*B \), one may of course consider

\[
C_\mathcal{H} = BB^* = WM_\mathcal{H}W^*
\]

(15)
on \( \mathcal{H} \), which has the same spectrum as \( C \)—with \( C \) nuclear, \( C_\mathcal{H} \) is also nuclear—and the whole game can be repeated by looking at the spectral decompositions of \( C_\mathcal{H} \).

When one takes the special case \( \mathcal{H} = \mathcal{Q} \) with \( C_\mathcal{Q} = RR^* \), we see that \( C_\mathcal{Q}s_m = \lambda_m s_m \), and \( s_m = UV^*v_m \), as well as \( C_\mathcal{Q} = UV^*CVU^* \). This abstract equation can be spelt out in more analytical detail for the special case when the inner product on \( \mathcal{Q} \) is given by a measure \( \varpi \) on \( \mathcal{P} \), as it then becomes

\[
\langle C_\mathcal{Q}\phi|\psi\rangle_\mathcal{Q} = \langle R^*\varphi|R^*\psi\rangle_\mathcal{U} = \int\int_{\mathcal{M} \times \mathcal{M}} \varphi(\mu_1)\varphi(\mu_2)\psi(\mu_2)\varpi(\mu_1)\varpi(\mu_2), \quad (16)
\]
i.e. \( C_\mathcal{Q} \) is a Fredholm integral operator with kernel \( \varphi \)—on \( \mathcal{Q} \) the kernel is in general not reproducing—and its spectral decomposition \( C_\mathcal{Q} = \sum_m \lambda_m s_m \otimes s_m \) is nothing but the familiar theorem of Mercer [6]. Factorisations of \( C_\mathcal{Q} \) are then factorisations of the kernel \( \varphi(\mu_1,\mu_2) \) and the corresponding representations of \( r(\mu) \) are obtained by integral transforms [20] [19], as already indicated by Karhunen in [13] [14]. The abstract setting outlined in this section can now be applied to the analysis of a great number of different situations, see [20] for more detail.

As already indicated, the spectral decomposition Eq. (13) allows one to go beyond the requirement that \( C \) be nuclear, but in the case of a probability assignment the push-forward is not a measure any more on \( \mathcal{U} \), but it can still be useful in the computation considering weak distributions. Another formulation of the spectral decomposition in the same vein as Eq. (10) allows also to cover the general case [7] [8]. The space \( \mathcal{U} = \bigoplus_j \mathcal{U}_j \) can be decomposed into a orthogonal direct sum of invariant subspaces \( \mathcal{U}_j \) on each of which the operator has a simple spectrum. So we may assume for this that the operator has a simple spectrum, otherwise consider each subspace \( \mathcal{U}_j \) in turn. It turns out that one can find a so-called rigged Hilbert space or Gel’fand triplet: \( \mathcal{N} \hookrightarrow \mathcal{U} \hookrightarrow \mathcal{N}^\ast \) with \( \mathcal{N} \) nuclear and a densely embedded in \( \mathcal{U} \). The eigenvalue equation for a self-adjoint operator \( C \) can be written in weak form: for \( \lambda \in \sigma(C) \) find \( v_\lambda \in \mathcal{U} \), s.t. for all \( w \in \mathcal{U} \), \( \langle w|Cv_\lambda \rangle = \lambda \langle w|v_\lambda \rangle \), but there may be no \( v_\lambda \in \mathcal{U} \) if \( \lambda \) is merely in the spectrum and not also an eigenvalue.

Using duality, this is now weakened to: for \( \lambda \in \sigma(C) \) find \( v_\lambda \in \mathcal{N}^\ast \), s.t. for all \( w \in \mathcal{N} \), \( \langle Cv_\lambda\rangle = \lambda \langle w, v_\lambda \rangle \), and it turns out that one can find such \( v_\lambda \in \mathcal{N}^\ast \), in the larger space.
With this the Eq. (10) may be generalised, where, as the spectrum $\sigma(C)$ may be continuous, the sum in general has to be replaced by an integral w.r.t. a measure $\rho$ on $\sigma(C) \subseteq \mathbb{R}$. As $C = R^*R$, the operator $C_Q = RR^*$ has the same spectrum, and can be decomposed in a Gel’fand triplet or rigged Hilbert space $\mathcal{P} \hookrightarrow Q \hookrightarrow \mathcal{P}^*$ with $s_\lambda \in \mathcal{P}^*$:

$$C = \int_{\sigma(C)} \lambda \, v_\lambda \otimes v_\lambda \, \rho(d\lambda); \quad C_Q = \int_{\sigma(C)} \lambda \, s_\lambda \otimes s_\lambda \, \rho(d\lambda). \quad (17)$$

The $s_\lambda \in \mathcal{P}^*$ may be seen as generalised functions, and both decompositions together in Eq. (17) allow to write a SVD-like decomposition of $R$ and $R^*$, corresponding to Eq. (11), and have a representation of $r(\mu)$ in a weak sense as a Karhunen-Loève integral over $\mathcal{P}^*$-generalised functions:

$$R = \int_{\sigma(C)} \sqrt{\lambda} \, (s_\lambda \otimes v_\lambda) \, \rho(d\lambda); \quad R^* = \int_{\sigma(C)} \sqrt{\lambda} \, (v_\lambda \otimes s_\lambda) \, \rho(d\lambda);$$

$$r(\mu) = \int_{\sigma(C)} \sqrt{\lambda} \, s_\lambda(\mu) v_\lambda \, \rho(d\lambda) = \int_{\sigma(C)} (R^* s_\lambda)(\mu) \, \rho(d\lambda). \quad (18)$$

One familiar and frequent place where this occurs (e.g. [16]) is the classical spectral representation of a stationary stochastic process

$$q(t) = \int_{\mathbb{R}} \sqrt{S(\omega)} \exp(i\omega t) \, Z(d\omega),$$

where $\sqrt{S(\omega)}$ is the square root of the spectral density—corresponding to $\sqrt{\lambda}$—and $Z(d\omega)$ is a random measure with orthogonal increments and unit variance. This random measure corresponds to $v_\lambda \, \rho(d\lambda)$ in Eq. (13), the space $Q$ corresponds to $L_2(\mathbb{R})$, the space of generalised functions $\mathcal{P}^*$ corresponds to the Schwartz space of tempered distributions $\mathcal{S}'(\mathbb{R})$, and the generalised eigenfunction $s_\lambda(\mu)$ corresponds to $\exp(i\omega t)$, a generalised eigenfunction of a stationary covariance kernel which is in $\mathcal{S}'(\mathbb{R})$ but not in $L_2(\mathbb{R})$ [5].

5 Conclusion

Parametric mappings have been analysed together with random variables with values in infinite dimensional spaces and their generalisations via an associated linear map, enabling the analysis by using well known techniques for the analysis of linear mappings. In the case of stochastic elements this leads to what is called weak distributions, a generalisation of the usual concept of a random variable.

In this connection algebras of random variables, the so-called algebraic approach to probability, leads to a concise description of the generation of appropriate spaces of random variables, and can naturally be used to specify randomness on infinite dimensional spaces via weak distributions. This has as a fundamental building block, next to the algebra of random variables, a distinguished self-adjoint, positive, and normalised linear functional called the state, which may be interpreted as an expectation operator. It is this setting that turns out to be conceptually much simpler than the measure-theoretic point of view, especially in the infinite dimensional setting. In particular this allows a natural approach to random matrices and tensor fields, where the random variables do not necessarily have to commute, and the interesting object is the behaviour of their spectra,
a distinctly analytic and algebraic concept which is much more complicated to treat with the usual measure-theoretic background.

The associated linear map leads to the self-adjoint and positive definite so-called ‘correlation operator’, as well as its different factorisations. Different representations generate different factorisations and thus allow a uniform analysis of their behaviour via an analysis of linear maps. It is in particular the different factorisations, and especially the spectral decomposition, which lead to suggestions for reduced order models and their analysis.

Not only does each separated representation define an associated linear map, but conversely under the restrictive conditions of a nuclear or trace-class correlation operator each factorisation induces a Karhunen-Loève- or proper orthogonal decomposition (POD)-like separated representation. The extension of this idea to arbitrary non-nuclear correlations operators is indicated through integral transforms, exemplified through the use of appropriate spectral decompositions, either via multiplication operators or as spectral integrals with rigged Hilbert spaces. These representations must be classed as generalised maps or generalised random variables, they can only be considered in a duality framework in a weak sense. This can be seen as an analogy to how normal generalised functions or distributions in the Sobolev-Schwartz sense are treated as a dual space of very smooth functions, and in fact the theoretical treatment follows along similar lines.

As this is a very short note touching on many diverse subjects to show their interconnection, it can naturally only be brief and in many cases just provides hints which have to be followed further with the references indicated. The analytic techniques used are ‘classical’ and have been developed along with the growth of quantum theory in the 1940s. It is their combination and uniform view from the point of linear functional analysis which is novel here.

References

[1] P. Benner, S. Gugercin, and K. Willcox, A survey of projection-based model reduction methods for parametric dynamical systems, SIAM Review 57 (2015), 483–531, doi:10.1137/130932715.

[2] P. Benner, M. Ohlberger, A. T. Patera, G. Rozza, and K. Urban (eds.), Model Reduction of Parametrized Systems, MS&A — Modeling, Simulation & Applications, vol. 17, Berlin, Springer, 2017, doi:10.1007/978-3-319-58786-8.

[3] A. Berlinet and C. Thomas-Agnan, Reproducing kernel Hilbert spaces in probability and statistics, Kluwer, Dordrecht, 2004.

[4] V. I. Bogachev and O. G. Smolyanov, Topological vector spaces and their applications, Springer, Berlin, 2017, doi:10.1007/978-3-319-57117-1.

[5] R. N. Bracewell, The Fourier transform and its applications, McGraw-Hill, New York, NY, 1978.

[6] R. Courant and D. Hilbert, Methods of mathematical physics, John Wiley & Sons, Chichester, 1989.

[7] R. Dautray and J.-L. Lions, Spectral theory and applications, Mathematical Analysis and Numerical Methods for Science and Technology, vol. 3, Springer, Berlin, 1990.
[8] I. M. Gel’fand and N. Y. Vilenkin, *Applications of harmonic analysis*, Generalized Functions, vol. 4, Academic Press, New York, NY, 1964.

[9] L. Gross, *Measurable functions on Hilbert space*, Transactions of the American Mathematical Society 105 (1962), no. 3, 372–390, doi:10.2307/1993726

[10] F. Hiai and D. Petz, *The semicircle law, free random variables and entropy*, Mathematical Surveys and Monographs, vol. 77, American Mathematical Society, Providence, RI, 2000.

[11] H. Holden, B. Øksendal, J. Uböe, and T.-S. Zhang, *Stochastic partial differential equations*, Birkhäuser, Basel, 1996.

[12] S. Janson, *Gaussian Hilbert spaces*, Cambridge Tracts in Mathematics, 129, Cambridge University Press, Cambridge, 1997.

[13] K. Karhunen, *Über lineare Methoden in der Wahrscheinlichkeitsrechnung*, Ann. Acad. Sci. Fennicae. Ser. A. I. Math.-Phys. 37 (1947), 1–79.

[14] K. Karhunen and I. Selin (transl.), *On linear methods in probability theory — Über lineare Methoden in der Wahrscheinlichkeitsrechnung — 1947, U.S. Air Force — Project RAND T-131, The RAND Corporation, St Monica, CA, USA, August 1960, English Translation, Available from: https://www.rand.org/pubs/translations/T131.html

[15] P. Krée and C. Soize, *Mathematics of random phenomena—random vibrations of mechanical structures*, D. Reidel, Dordrecht, 1986.

[16] H. G. Matthies, *Uncertainty quantification with stochastic finite elements*, Encyclopaedia of Computational Mechanics (E. Stein, R. de Borst, and T. J. R. Hughes, eds.), vol. 1, John Wiley & Sons, Chichester, 2007, Part 1. Fundamentals. Encyclopaedia of Computational Mechanics, doi:10.1002/0470091355.ecm071

[17] H. G. Matthies and A. Keese, *Galerkin methods for linear and nonlinear elliptic stochastic partial differential equations*, Computer Methods in Applied Mechanics and Engineering 194 (2005), no. 12-16, 1295–1331, doi:10.1016/j.cma.2004.05.027 MR MR2121216 (2005j:65146)

[18] H. G. Matthies, A. Litvinenko, O. Pajonk, B. V. Rosić, and E. Zander, *Parametric and uncertainty computations with tensor product representations*, Uncertainty Quantification in Scientific Computing (A. Dienstfrey and R. Boisvert, eds.), IFIP Advances in Information and Communication Technology, vol. 377, Springer, Boulder, CO, 2012, pp. 139–150, doi:10.1007/978-3-642-32677-6

[19] H. G. Matthies and R. Ohayon, *Analysis of parametric models for coupled systems* [online], arXiv: 1806.07255 [math.NA], 2018, Available from: http://arxiv.org/1806.07255.

[20] ______, *Analysis of parametric models — linear methods and approximations* [online], arXiv: 1806.01101 [math.NA], 2018, Available from: http://arxiv.org/1806.01101
[21] J. A. Mingo and R. Speicher, *Free probability and random matrices*, Fields Institute Monographs, vol. 35, Springer, Berlin, 2017, [doi:10.1007/978-1-4939-6942-5](https://doi.org/10.1007/978-1-4939-6942-5).

[22] I. E. Segal, *Distributions in Hilbert space and canonical systems of operators*, Transactions of the American Mathematical Society 88 (1958), no. 1, 12–41, [doi:10.2307/1993234](https://doi.org/10.2307/1993234).

[23] ______, *Nonlinear functions of weak processes. I*, Journal of Functional Analysis 4 (1969), no. 3, 404–456, [doi:10.1016/0022-1236(69)90007-X](https://doi.org/10.1016/0022-1236(69)90007-X).

[24] I. E. Segal and R. A. Kunze, *Integrals and operators*, Springer, Berlin, 1978.

[25] R. Speicher, *Free probability theory*, Jahresber Dtsch Math-Ver 119 (2017), 3–30, [doi:10.1365/s13291-016-0150-5](https://doi.org/10.1365/s13291-016-0150-5).

[26] T. J. Sullivan, *Introduction to uncertainty quantification*, Texts in Applied Mathematics, vol. 63, Springer, Berlin, 2015, [doi:10.1007/978-3-319-23395-6](https://doi.org/10.1007/978-3-319-23395-6).

[27] D. V. Voiculescu, K. J. Dykema, and A. Nica, *Free random variables*, CRM Monograph Series, vol. 1, American Mathematical Society, Providence, RI, 1992.

[28] P. Whittle, *Probability via expectation*, 4th ed., Springer Texts in Statistics, Springer, Berlin, 2000.