HAL open science

QUANTUM MECHANICS REVISITED (v.3)

Jean Claude Dutailly

To cite this version:

Jean Claude Dutailly. QUANTUM MECHANICS REVISITED (v.3). 2015. hal-00770220v3

HAL Id: hal-00770220

https://hal.science/hal-00770220v3

Preprint submitted on 1 Jul 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Quantum Mechanics Revisited

Jean Claude Dutailly
Paris (France)

July 1, 2015

Abstract

The paper proposes a new approach in the foundations of Quantum Mechanics. It does not make any assumption about the physical world, but looks at the consequences of the formalism used in models. Whenever a system is represented by variables which meet precise, but common, mathematical properties, one can prove theorems which are very close to the axioms of Quantum Mechanics (Hilbert spaces, observables, eigen values,...). It is then possible to explore the conditions of the validity of these axioms and to give a firm ground to the usual computations. Moreover this approach sheds a new ligth on the issues of determism, and interacting systems.

In the third edition of this paper developments have been added about the statistical procedures used to detect anomalies in Physical Laws.

1 Introduction

Quantum Physics encompasses several theories, with three distinct areas:

i) Quantum Mechanics (QM) proper, which can be expressed as a collection of axioms, such as summarized by Weinberg :

- Physical states of a system are represented by vectors ψ in a Hilbert space H, defined up to a complex number (a ray in a projective Hilbert space)

- Observables are represented by Hermitian operators

- The only values that can be observed for an operator are one of its eigen values λ_k corresponding to the eigen vector ψ_k

- The probability to observe λ_k if the system is in the state ψ is proportional to $|\langle\psi, \psi_k\rangle|^2$

- If two systems with Hilbert space H_1, H_2 interact, the states of the total system are represented in $H_1 \otimes H_2$

and, depending on the authors, the Schrödinger's equation.

ii) Wave Mechanics, which states that particles can behave like fields which propagate, and conversely force fields can behave like pointwise particles. Moreover particles are endowed with a spin. In itself it constitutes a new theory, with the introduction of new concepts, for which QM is the natural formalism.

Actually this is essentially a theory of electromagnetism, and is formalized in Quantum Electrodynamics (QED).

iii) The Quantum Theory of Fields (QTF) is a theory which encompasses theoretically all the phenomena at the atomic or subatomic scale, but has been set up mainly to deal with the other forces (weak and strong interactions) and the organization of elementary particles. It uses additional concepts (such as gauge fields) and formalism and computation rules (Feynman diagrams, path integrals).

I will address in this paper *QM only*.

The status of the axioms listed above is special. They are not Physical laws, they do not define any physical object, or physical property (if we except the Schrödinger's equation which is or not part of the corpus). they are deemed valid for any system at a scale which is not even mentioned but they are not falsifiable (how could we check that an observable is a Hermitian operator ?). This strange status, quite unique in Science, is at the origin of the search for interpretations, and for the same reason, makes so difficult any sensible discussion on the topic. Actually these axioms have emerged slowly from the practices of great physicists, kept without any change in the last decenniums, and endorsed by the majority, mostly because it is part of their environment.

Most of these interpretations (there are hundreds of them) look for what one can tell "physical interpretations" : the axioms reflect the physical world, so we must explain their statements through the properties of natural objects or phenomena, how bizarre they could seem. I will not enter into this debate : my proposals do not assume anything about the physical properties of the real world.

Others have proposed a different direction : these axioms come from foundamental limitations in our capability to know the real world. The main endeavour has been, since seminal books and articles of von Neumann and Birkhoff, to set up a formal system in which the assertions done in Physics can be expressed and used in the predictions of experiments, and so granting to Physics a status which would be less speculative and more respecting of the facts as they can actually be established. This is actually similar to what is done in Mathematics for Arithmetic or Sets Theory. The work has been pursued, notably by Jauch, Haag, Varadarajan and Francis in the recent years. An extension which accounts for Relativity has been proposed by Wightman and has been developed as an Axiomatic Quantum Field Theory (Haag, Araki, Halvorson, Borchers, Doplicher, Roberts, Schroer, Fredenhagen, Buchholz, Summers, Longo,...). It assumes the existence of the formalism of Hilbert space itself, so the validity of most of the axioms, and emphasizes the role to the algebra of operators. Since all the information which can be extracted from a system goes through operators, it can be conceived to define the system itself as the set of these operators. This is a more comfortable venue, as it is essentially mathematical, which has been studied by several authors (Bratelli and others). Recently this approach has been completed by attempts to link QM with Information Theory, either in the framework of Quantum Computing, or through the use of the Categories

Theory.

These works share some philosophical convictions, supported with a strength depending on the authors, but which are nonetheless present :

i) A deep mistrust with regard to realism, the idea that there is a real world, which can be understood and described through physical concepts such as particles, location,...At best they are useless, at worst they are misleading.

ii) A great faith in the mathematical formalism, which should ultimately replace the concepts.

iii) The preeminence of experimentation over theories : experimental facts are seen as the unique source of innovation, physical laws are essentially the repeated occurrences of events whose correlation must be studied by statistical methods, the imperative necessity to consider the conditions in which the experiments can or cannot be made.

As any formal system, the axiomatic QM defines its own objects, which are basically the assertions that a physicist can make from the results of experiments ("the yes-no experiments" of Jauch), and set up a system of rules of inference according to which other assertions can be made, with a special attention given to the possibility to make simultaneous measures, and the fact that any measure is the product of a statistical estimation. With the addition of some axioms, which obviously cannot reflect any experimental work (it is necessary to introduce infinity), the formal system is then identified, by a kind of structural isomorphism, with the usual Hilbert space and its operators of Mathematics. And from there the axioms of QM are deemed to be safely grounded.

One can be satisfied or not by this approach. But some remarks can be done.

In many ways this attempt is similar to the one by which mathematicians tried to give an ultimate, consistent and logical basis to Mathematics. Their attempt has not failed, but have shown the limits of what can be achieved : the necessity to detach the objects of the formal system from any idealization of physical objects, the non unicity of the axioms, and the fact that they are justified by experience and efficiency and not by a logical necessity. The same limits are obvious in axiomatic QM. If to acknowledge the role of experience and efficiency in the foundations of the system should not be disturbing, the pretense to enshrine them in axioms, not refutable and not subject to verification, places a great risk to the possibility of any evolution. And indeed the axioms have not changed for more than 50 years, without stopping the controversies about their meaning. The unavoidable replacement of physical concepts, identification of physical objects and their properties, by formal and abstract objects, which is consistent with the philosophical premises, is specially damaging in Physics. Because there is always a doubt about the meaning of the objects (for instance it is quite impossible to find the definition of a "state") the implementation of the system sums up practically to a set of "generally accepted computations", it makes its learning and teaching perilous (the Feynmann's affirmation that it cannot be understood), and eventually to the recurring apparitions of "unidentified physical objects" whose existence is supposed to fill the gap. In many ways the formal system has replaced the Physical Theories, that is a set of

objects, properties and behaviors, which can be intuitively identified and understood. The Newton's laws of motion are successful, not only because they can be checked, but also because it is easy to understand them. This is not the case for the decoherence of the wave function...

Nevertheless, this attempt is right in looking for the origin of these axioms in the critique (in the Kantian meaning) of the method specific to Physics. But it is aimed at the wrong target : the concepts are not the source of the problems, they are and will stay necessary because they make the link between formalism and real world, and are the field in which new ideas can germinate. And the solution is not in a sanctification of the experiments, which are too diverse to be submitted to any analytical method. Actually these attempts have missed a step, which always exists between the concepts and the collection of data : the mathematical formalization itself, in models. Models, because they use a precise formalism, can be easily analyzed and it is possible to show that, indeed, they have specific properties of their own, which do not come from the reality they represent, but from their mathematical properties and the way they are used. The objects of an axiomatic QM, if one wishes to establish one, are then clearly identified, without disturbing the elaboration or the implementation of theories. The axioms can then be proven, they can also be safely used, as we will show in this paper.

QM is about the representation of physical phenomena, and not a representation of these phenomena (as can be Wave Mechanics, QED or QTF). It expresses properties of the data which can be extracted from measures of physical phenomena but not properties of physical objects. To sum up : QM is not about how the physical world works, it is about how it looks.

The results presented here are theorems : they are proven, as consequence of some basic assumptions about the mathematical properties of the models used commonly in Physics. They state precise conditions, and use common Mathematics Theorems, which can be be found in my book "Mathematics for Theoretical Physics") which is freely available. They will be referred to as (Maths.XXX).

In the first section we will introduce the Hilbert space and its mandatory tool, the Hilbertian basis.

In the second section we will introduce Observables, and their main properties.

In the third section we will see how we can understand the introduction of probability in QM.

In the fourth section we will prove a theorem, similar to the Wigner's theorem, about the change of variables and the use of QM with the representation of groups.

In the fifth section we will introduce two theorems, similar to the Schrödinger's equation, for the evolution of systems.

In the sixth section we will consider interacting systems.

The comparison between the results and the usual axioms is done in the last section.

2 HILBERT SPACE

2.1 Representation of a system

To implement a scientific law, either to check it or to use it for practical purpose (to predict an outcome), scientists and engineers use models. A **model** can be seen as the general representation of the law. It comprises :

- a system : the area in which the system is located and the time frame during which it is observed, the list of the objects and of their properties which are considered

- the circumstances if they are specific (temperature, interference with the exterior of the system,...)

- the variables representing the properties, associated each to a mathematical object with more specific mathematical properties if necessary (a scalar can be positive, a function can be continuous,...)

- the procedures used to collect and analyze the data, notably if statistical methods are used.

Building and using models are a crucial part of the scientific work. Any engineer or theoretical physicist use them, either to compute solutions of a problem from well established laws, or to explore the consequences of more general hypotheses. A model is a representation, usually simplified, of part of the reality, built from concepts, assumptions and accepted laws. The simplification helps to focus on the purpose, trading accuracy for efficiency. Models provide both a framework in which to make the computations, using some formalism in an ideal representation, and a practical procedure to organize the collection and analysis of the data. They are the embodiment of scientific laws, implemented in more specific circumstances, but still with a large degree of generality which enables to transpose the results from one realization to another. Actually most, if not all, scientific laws can be expressed in the framework of a model.

Models use a formalism, that is a way to represent the properties in terms of variables, which can take different values according to the specific realizations of the model, and which are used to make computations to predict a result. The main purpose of the formalism is efficiency, because it enables to use rules and theorems well established in a more specific field. In Physics the formalism is mathematical, but other formalisms exist (for instance the atomic representation used in Chemistry).

The most elaborate models are in Analytic Mechanics and, indeed, they stand at the heart of QM. A system, meaning a delimited area of space comprising material bodies, is represented by scalar generalized coordinates $q = (q_1, ..., q_N)$ its evolution by the derivatives $q' = (q'_1, ..., q'_N)$. By extension q can be the coordinates of a point Q of some manifold M to account for additional constraints, and then the state of the system at a given time is fully represented by a point of the vector bundle $TM : W = (Q, V_Q)$. By mathematical transformations the derivatives q' can be exchanged with conjugate momenta, and the state of the system is then represented in the phase space, with a symplectic structure. But we will not use this addition and stay at the very first step, that

is the representation of the system by (q, q') .

Trouble arises when one considers the other fundamental objects of Physics : force fields. By definition their value is defined all over the space x time. So in the previous representation one should account, at a given time, for the value of the fields at each point, and introduce unaccountably infinitely many coordinates. This issue has been at the core of many attempts to improve Analytic Mechanics.

But let us consider two facts :

- Analytic Mechanics, as it is usually used, is aimed at representing the evolution of the system over a whole period of time $[0, T]$, as it is clear in the Lagrangian formalism : the variable are accounted, together, for the duration of the experiment;

- the state of the system is represented by a map $W : [0, T] \to (Q, V_Q)$: the knowledge of this map sums up all that can be said on the system, the map itself represents the state of the system.

Almost all the problems in Physics involve a model which comprises the following :

i) a set of physical objects (material bodies or particles, force fields) in a delimited area Ω of space x time (it can be in the classical or the relativist framework) called the system;

ii) the state of the system is represented by a fixed finite number N of variables $X = (X_k)_{k=1}^N$ which can be maps defined on Ω , with their derivatives;

so that the state of the system is defined by a finite number of maps, which usually belong themselves to infinite dimensional vector spaces.

And it is legitimate to substitute the maps to the coordinates in Ω. We still have infinite dimensional vector spaces, but by proceeding first to an aggregation by maps, the vector space is more manageable, and we have some mathematical tools to deal with it. But we need to remind the definition of a manifold (more in Maths.15.1.1).

2.2 Manifold

Let M be a set, E a topological vector space, an atlas, denoted $A = (O_i, \varphi_i, E)_{i \in I}$ is a collection of :

subsets $(O_i)_{i \in I}$ of M such that $\cup_{i \in I} O_i = M$ (this is a cover of M)

maps $(\varphi_i)_{i \in I}$ called **charts**, such that :

i) $\varphi_i : O_i \to U_i :: \xi = \varphi_i(m)$ is bijective and ξ are the coordinates of M in the chart

ii) U_i is an open subset of E

iii) $\forall i, j \in I : O_i \cap O_j \neq \varnothing$:

$\varphi_i(O_i \cap O_j), \varphi_j(O_i \cap O_j)$ are open subsets of E, and there is a bijective, continuous map, called a transition map :

$\varphi_{ij} : \varphi_i(O_i \cap O_j) \to \varphi_j(O_i \cap O_j)$

Notice that no mathematical structure of any kind is required on M. A topological structure can be imported on M, by telling that all the charts are continuous, and conversely if there is a topological structure on M the charts must

be compatible with it. But the set M has no algebraic structure : a combination such as $am + bm'$ has no meaning.

Two atlas $A = (O_i, \varphi_i, E)_{i\in I}, A' = \left(O'_j, \varphi'_j, E\right)_{j\in J}$ of M are said to be compatible if their union is still an atlas. Which implies that :

$\forall i \in I, j \in J : O_i \cap O'_j \neq \varnothing : \exists \varphi_{ij} : \varphi_i \left(O_i \cap O'_j\right) \to \varphi'_j \left(O_i \cap O'_j\right)$ which is a homeomorphism

The relation $A, A'$are compatible atlas of M, is a relation of equivalence. A class of equivalence is a **structure of manifold** on the set M.

The key points are :

- there can be different structures of manifold on the same set. On $\mathbb{R}^4$ there are unaccountably many non equivalent structures of smooth manifolds (this is special to $\mathbb{R}^4$: on $\mathbb{R}^n, n \neq 4$ all the smooth structures are equivalent !).

- all the interesting properties on M come from E : the dimension of M is the dimension of E (possibly infinite); if E is a Fréchet space we have a Fréchet manifold, if E is a Banach space we have a Banach manifold and then we can have differentials, if E is a Hilbert space we have a Hilbert manifold, but these additional properties require that the transition maps φ_{ij} meet additional properties.

- for many sets several charts are required (a sphere requires at least two charts) but an atlas can have only one chart, then the manifold structure is understood as the same point M will be defined by a set of compatible charts.

The usual, euclidean, 3 dimensional space of Physics is an affine space. It has a structure of manifold, which can use an atlas with orthonormal frames, or with curved coordinates (spherical or cylindrical). Passing from one system of coordinates to another is a change of charts, and represented by transition maps φ_{ij}.

2.3 Fundamental theorem

We will consider models which meet the following conditions:

Condition 1 *i) The system is represented by a fixed finite number N of variables $(X_k)_{k=1}^N$*

ii) Each variable belongs to an open subset O_k of a separable Fréchet real vector space V_k

iii) At least one of the vector spaces $(V_k)_{k=1}^N$ is infinite dimensional

iv) For any other model of the system using N variables $(X'_k)_{k=1}^N$ belonging to open subset O'_k of V_k, and for $X_k, X'_k \in O_k \cap O'_k$ there is a continuous map : $X'_k = \digamma_k (X_k)$

Remarks :

i) The *variables must be vectorial*. This condition is similar to the superposition principle which is assumed in QM. This is one of the most important condition. By this we mean that the associated physical phenomena can be represented as vectors (or tensors, or scalars). The criterion, to check if this is the case, is : if the physical phenomenon can be represented by X and X',

does the phenomenon corresponding to any linear combination $\alpha X + \beta X'$ has a physical meaning ?

Are usually vectorial variables : the speed of a material point, the electric or magnetic field, a force, a moment,...and the derivatives, which are, by definition, vectors.

Are not usually vectorial variables : qualitative variables (which take discrete values), a point in the euclidean space or on a circle, or any surface. The point can be represented by coordinates, but these coordinates are not the physical object, which is the material point. For instance in Analytic Mechanics the coordinates $q = (q_1, ..., q_N)$ are not a geometric quantity : usually a linear combination $\alpha q + \beta q'$ has no physical meaning. The issue arises because physicists are used to think in terms of coordinates (in euclidean or relativist Lorentz frame) which leads to forget that the coordinates are just a representation of an object which, even in its mathematical form (a point in an affine space) is not vectorial.

So this condition, which has a simple mathematical expression, has a deep physical meaning : it requires to understand clearly why the properties of the physical phenomena can be represented by a vectorial variable, and reaches the most basic assumptions of the theory. The status, vectorial or not, of a quantity is not something which can be decided at will by the Physicist : it is part of the Theory which he uses to build his model.

However we will see that the addition of a variable which is not a vector can be very useful (Theorem 24).

ii) The variables are assumed to be independent, in the meaning that there is no given relation such that $\sum_k X_k = 1$. Of course usually the model is used with the purpose to compute or check relations between the variables, but these relations do not matter here. Actually to check the validity of a model one considers all the variables, those which are given and those which can be computed, they are all subject to measures and this is the comparison, after the experiment, between computed values and measured values which provides the validation. So in this initial stage of specification of the model there is no distinction between the variables, which are on the same footing.

Similarly there is no distinction between variables internal and external to the system : if the evolution of a variable is determined by the observer or by phenomena out of the system (it is external) its value must be measured to be accounted for in the model, so it is on the same footing as any other variable. And it is assumed that the value of all variables can be measured (we will come back on this point in the next section).

The derivative $\frac{dX_k}{dt}$ (or partial derivative at any order) of a variable X_k is considered as an independent variable, as it is usually done in Analytic Mechanics and in the mathematical formalism of r-jets.

iii) The variables can be restricted to take only some range (for instance it must be positive). The vector spaces are infinite dimensional whenever the variables are functions. The usual case is when they represent the evolution of the system with the time t : then X_k is the function itself : $X_k : \mathbb{R} \to O_k :: X_k(t)$. What we consider here are variables which cover the whole evolution of

the system over the time, and not only just a snapshot $X_k(t)$ at a given time. But the condition encompasses other cases, notably fields F which are defined over a domain Ω. *The variables are the maps* $F_k : \Omega \to O_k$ *and not their values* $F_k(\xi)$ *at a given point* $\xi \in \Omega$.

iv) A Fréchet space is a Hausdorff, complete, topological space endowed with a countable family of semi-norms (Maths.971). It is locally convex and metric. This, quite complicated, mathematical definition is required because we will prove a theorem, and as usual in Mathematics we need to be precise in stating the conditions of its validity.

Are Fréchet spaces :

- any Banach vector space : the spaces of bounded functions, the spaces $L^p(E, \mu, \mathbb{C})$ of integrable functions on a measured space (E, μ) (Maths.2270), the spaces $L^p(M, \mu, E)$ of integrable sections of a vector bundle (valued in a Banach E) (Maths.2276)

- the spaces of continuously differentiable sections on a vector bundle (Maths.2310), the spaces of differentiable functions on a manifold (Maths.2314).

A topological vector space is separable if it has a dense countable subset (Maths.590) which, for a Fréchet space, is equivalent to be second countable (Maths.698). A totally bounded ($\forall r > 0$ there is a finite number of balls which cover V), or a connected locally compact Fréchet space, is separable (Maths.702, 703). The spaces $L^p(\mathbb{R}^n, dx, \mathbb{C})$ of integrable functions for $1 \leq p < \infty$, the spaces of continuous functions on a compact domain, are separable (Lieb).

Thus this somewhat complicated specification encompasses most of the usual cases.

A case which appears quite often in Physics is the following : maps : $X : \Omega \to E$ from a relatively compact subset Ω of a manifold M to a finite dimensional vector space, endowed with a norm (it is important to notice that a definite positive scalar product is not required, a norm suffices). Then the space of maps such that $\int_\Omega \|X(m)\| \varpi(m) < \infty$ where ϖ is a measure on M (a volume measure) is an infinite dimensional, separable, Fréchet space.

v) The condition iv addresses the case when the variables are defined over connected domains. But it implicitly tells that any other set of variables which represent the same phenomena are deemed compatible with the model. This point is addressed more precisely in another section, with the change of variables.

The set of all potential states of the system is then given by the set $S = \left\{ (X_k)_{k=1}^N, X_k \in O_k \right\}$. If there is some relation between the variables, stated by a physical law or theory, its consequence is to restrict the domain in which the state of the system will be found, but as said before we stay at the step before any experiment, so O_k represents the set of all possible values of X_k.

Theorem 2 *For any system represented by a model meeting the conditions 1, there is a separable, infinite dimensional, Hilbert space H, defined up to isomorphism, such that* $\mathcal{S}$ *can be embedded as an open subset* $\Omega \subset H$ *which contains 0 and a convex subset.*

Proof. i) Each value of the set $\mathcal{S}$ of variables defines a state of the system, denoted X, belonging to the product $O = \prod_{k=1}^{N} O_k \subset V = \prod_{k=1}^{N} V_k$. The couple (O, X), together with the property iv) defines the structure of a Fréchet manifold M on the set $\mathcal{S}$, modelled on the Fréchet space $V = \prod_{k1}^{N} V_k$. The coordinates are the values $(x_k)_{k=1}^{N}$ of the functions X_k. This manifold is infinite dimensional. Any Fréchet space is metric, so V is a metric space, and M is metrizable.

ii) As M is a metrizable manifold, modelled on an infinite dimensional separable Fréchet space, the Henderson's theorem (Henderson - corollary 5, Maths.1386) states that it can be embedded as a open subset Ω of an infinite dimensional separable Hilbert space H, defined up to isomorphism. Moreover this structure is smooth, the set $H - \Omega$ is homeomorphic to H, the border $\partial\Omega$ is homeomorphic to Ω and its closure $\overline{\Omega}$.

iii) Translations by a field vector are isometries. Let us denote $\langle\rangle_H$ the scalar product on H (this is a bilinear symmetric positive definite form). The map : $\Omega \to \mathbb{R} :: \langle\psi, \psi\rangle_H$ is bounded from below and continuous, so it has a minimum (possibly not unique) ψ_0 in Ω. By translation of H with ψ_0 we can define an isomorphic structure, and then assume that 0 belongs to Ω. There is a largest convex subset of H contained in Ω, defined as the intersection of all the convex subset contained in Ω. Its interior is an open convex subset C. It is not empty : because 0 belongs to Ω which is open in H, there is an open ball $B_0 = (0, r)$ contained in Ω. ■

So the state of the system can be represented by a single vector ψ in a Hilbert space.

From a practical point of view, often V itself can be taken as the product of Hilbert spaces, notably of square summable functions such as $L^2(\mathbb{R}, dt)$ which are separable Hilbert spaces and then the proposition is obvious.

If the variables belong to an open O' such that $O \subset O'$ we would have the same Hilbert space, and an open Ω' such that $\Omega \subset \Omega'$. V is open so we have a largest open $\Omega_V \subset H$ which contains all the Ω.

Notice that this is a real vector space.

The interest of Hilbert spaces lies with Hilbertian basis, and we now see how to relate such basis of H with a basis of the vector space V. It will enable us to show a linear chart of the manifold M.

2.4 Basis

Theorem 3 *For any basis $(e_i)_{i\in I}$ of V contained in O, there are unique families $(\varepsilon_i)_{i\in I}, (\phi_i)_{i\in I}$ of independent vectors of H, a linear isometry $\Upsilon : V \to H$ such that :*

$\forall X \in O : \Upsilon(X) = \sum_{i\in I} \langle\phi_i, \Upsilon(X)\rangle_H \varepsilon_i \in \Omega$

$\forall i \in I : \varepsilon_i = \Upsilon(e_i)$

$\forall i,j \in I : \langle \phi_i, \varepsilon_j \rangle_H = \delta_{ij}$
and Υ is a compatible chart of M.

Proof. i) Let $(e_i)_{i\in I}$ be a basis of V such that $e_i \in O$ and $V_0 = Span\,(e_i)_{i\in I}$. Thus $O \subset V_0$.

Any vector of V_0 reads : $X = \sum_{i\in I} x_i e_i$ where only a finite number of x_i are non null. Or equivalently the following map is bijective :

$\pi_V : V_0 \to \mathbb{R}_0^I :: \pi_V \left(\sum_{i\in I} x_i e_i\right) = x = (x_i)_{i\in I}$

where the set $\mathbb{R}_0^I \subset \mathbb{R}^I$ is the subset of maps $I \to \mathbb{R}$ such that only a finite number of components x_i are non null.

(O, X) is an atlas of the manifold M and M is embedded in H, let us denote $\Xi : O \to \Omega$ a homeomorphism accounting for this embedding.

The inner product on H defines a positive kernel :

$K : H \times H \to \mathbb{R} :: K(\psi_1, \psi_2) = \langle \psi_1, \psi_2 \rangle_H$

Then $K_V : O \times O \to \mathbb{R} :: K_V(X, Y) = K(\Xi(X), \Xi(Y))$ defines a positive kernel on O (Math.1196).

K_V defines a definite positive symmetric bilinear form on V_0, denoted $\langle\rangle_V$, by :

$\left\langle \sum_{i\in I} x_i e_i, \sum_{i\in I} y_i e_i \right\rangle_V = \sum_{i,j\in I} x_i y_j K_{ij}$ with $K_{ij} = K_V(e_i, e_j)$

which is well defined because only a finite number of monomials $x_i y_j$ are non null. It defines a norm on V_0.

ii) Let : $\varepsilon_i = \Xi(e_i) \in \Omega$ and $H_0 = Span\,(\varepsilon_i)_{i\in I}$ the set of finite linear combinations of vectors $(\varepsilon_i)_{i\in I}$. It is a vector subspace (Math.901) of H. The family $(\varepsilon_i)_{i\in I}$ is linearly independent, because, for any finite subset J of I, the determinant

$\det\left[\langle \varepsilon_i, \varepsilon_j \rangle_H\right]_{i,j\in J} = \det\left[K_V(e_i, e_j)\right]_{i,j\in J} \neq 0.$

Thus $(\varepsilon_i)_{i\in I}$ is a non Hilbertian basis of H_0.

H_0 can be defined similarly by the bijective map :

$\pi_H : H_0 \to \mathbb{R}_0^I :: \pi_H \left(\sum_{i\in I} y_i \varepsilon_i\right) = y = (y_i)_{i\in I}$

iii) By the Gram-Schmidt procedure (which works for infinite sets of vectors) it is always possible to built an orthonormal basis $(\widetilde{\varepsilon}_i)_{i\in I}$ of H_0 starting with the vectors $(\varepsilon_i)_{i\in I}$ indexed on the same set I (as H is separable I can be assimilated to $\mathbb{N}$).

$\ell^2(I) \subset \mathbb{R}^I$ is the set of families $y = (y_i)_{i\in I} \subset \mathbb{R}^I$ such that :

$\sup\left(\sum_{i\in J} (y_i)^2\right) < \infty$ for any countable subset J of I.

$\mathbb{R}_0^I \subset \ell^2(I)$

The map : $\chi : \ell^2(I) \to H_1 :: \chi(y) = \sum_{i\in I} y_i \widetilde{\varepsilon}_i$ is an isomorphism to the closure $H_1 = \overline{Span\,(\widetilde{\varepsilon}_i)_{i\in I}} = \overline{H_0}$ of H_0 in H (Math.1121). H_1 is a closed vector subspace of H, so it is a Hilbert space. The linear span of $(\widetilde{\varepsilon}_i)_{i\in I}$ is dense in H_1, so it is a Hilbertian basis of H_1 (Math.1122).

Let $\pi : H \to H_1$ be the orthogonal projection on H_1 : $\|\psi - \pi(\psi)\|_H = \min_{u\in H_1} \|\psi - u\|_H$ then :

$\psi = \pi(\psi) + o(\psi)$ with $o(\psi) \in H_1^\perp$ which implies : $\|\psi\|^2 = \|\pi(\psi)\|^2 + \|o(\psi)\|^2$

There is a open convex subset, containing 0, which is contained in Ω so there is $r > 0$ such that :

$\|\psi\| < r \Rightarrow \psi \in \Omega$ and as $\|\psi\|^2 = \|\pi(\psi)\|^2 + \|o(\psi)\|^2 < r^2$
then $\|\psi\| < r \Rightarrow \pi(\psi), o(\psi) \in \Omega$
$o(\psi) \in H_1^\perp, H_0 \subset H_1 \Rightarrow o(\psi) \in H_0^\perp$
$\Rightarrow \forall i \in I : \langle \varepsilon_i, o(\psi)\rangle_H = 0 = K_V\left(\Xi^{-1}(\varepsilon_i), \Xi^{-1}(o(\psi))\right) = K_V\left(e_i, \Xi^{-1}(o(\psi))\right)$
$\Rightarrow \Xi^{-1}(o(\psi)) = 0 \Rightarrow o(\psi) = 0$
$H_1^\perp = 0$ thus H_1 is dense in H (Math.1115), and as it is closed : $H_1 = H$
$(\widetilde{\varepsilon}_i)_{i\in I}$ is a Hilbertian basis of H and
$\forall \psi \in H : \psi = \sum_{i\in I} \langle \widetilde{\varepsilon}_i, \psi\rangle_H \widetilde{\varepsilon}_i$ with $\sum_{i\in I} |\langle \widetilde{\varepsilon}_i, \psi\rangle_H|^2 < \infty$
$\Leftrightarrow (\langle \widetilde{\varepsilon}_i, \psi\rangle_H)_{i\in I} \in \ell^2(I)$

H_0 is the interior of H, it is the union of all open subsets contained in H, so $\Omega \subset H_0$

$H_0 = Span\left((\widetilde{\varepsilon}_i)_{i\in I}\right)$ thus the map :
$\widetilde{\pi}_H : H_0 \to \mathbb{R}_0^I :: \widetilde{\pi}_H\left(\sum_{i\in I} \widetilde{y}_i \widetilde{\varepsilon}_i\right) = \widetilde{y} = (\widetilde{y}_i)_{i\in I}$
is bijective and : $\widetilde{\pi}_H(H_0) = \widetilde{R}_0 \subset \mathbb{R}_0^I \subset \ell^2(I)$
Moreover : $\forall \psi \in H_0 : \widetilde{\pi}_H(\psi) = (\langle \widetilde{\varepsilon}_i, \psi\rangle_H)_{i\in I} \in \mathbb{R}_0^I$
Thus :
$\forall X \in O : \Xi(X) = \sum_{i\in I} \langle \widetilde{\varepsilon}_i, \Xi(X)\rangle_H \widetilde{\varepsilon}_i \in \Omega$
and $\widetilde{\pi}_H(\Xi(X)) = (\langle \widetilde{\varepsilon}_i, \Xi(X)\rangle_H)_{i\in I} \in \widetilde{R}_0$
$\forall i \in I, e_i \in O \Rightarrow \Xi(e_i) = \varepsilon_i = \sum_{j\in I} \langle \widetilde{\varepsilon}_j, \varepsilon_i\rangle_H \widetilde{\varepsilon}_j$
and $\widetilde{\pi}_H(\varepsilon_i) = \left(\langle \widetilde{\varepsilon}_j, \varepsilon_i\rangle_H\right)_{j\in I} \in \widetilde{R}_0$
iv) Let be : $\widetilde{e}_i = \Xi^{-1}(\widetilde{\varepsilon}_i) \in V_0$ and $\mathcal{L}_V \in GL(V_0; V_0) :: \mathcal{L}_V(e_i) = \widetilde{e}_i$
We have the following diagram :

$$\begin{array}{ccccccc}
 & \Xi & & \mathcal{L}_H^{-1} & & & \\
e_i & \rightarrow & \varepsilon_i & \rightarrow & \widetilde{\varepsilon}_i & & \\
 & \searrow & & & \downarrow & & \\
 & \mathcal{L}_V & \searrow & & \downarrow & \Xi^{-1} & \\
 & & & \searrow & \downarrow & & \\
 & & & & \widetilde{e}_i & &
\end{array}$$

$\langle \widetilde{e}_i, \widetilde{e}_j\rangle_V = \langle \Xi(\widetilde{e}_i), \Xi(\widetilde{e}_j)\rangle_H = \langle \widetilde{\varepsilon}_i, \widetilde{\varepsilon}_j\rangle_H = \delta_{ij}$
So $(\widetilde{e}_i)_{i\in I}$is an orthonormal basis of V_0 for the scalar product K_V
$\forall X \in V_0 : X = \sum_{i\in I} \widetilde{x}_i \widetilde{e}_i = \sum_{i\in I} \langle \widetilde{e}_i, X\rangle_V \widetilde{e}_i$ and $(\langle \widetilde{e}_i, X\rangle_V)_{i\in I} \in \mathbb{R}_0^I$
The coordinates of $X \in O$ in the basis $(\widetilde{e}_i)_{i\in I}$ are $(\langle \widetilde{e}_i, X\rangle_V)_{i\in I} \in \mathbb{R}_0^I$
The coordinates of $\Xi(X) \in H_0$ in the basis $(\widetilde{\varepsilon}_i)_{i\in I}$ are $(\langle \widetilde{\varepsilon}_i, \Xi(X)\rangle_H)_{i\in I} \in \mathbb{R}_0^I$
$\langle \widetilde{\varepsilon}_i, \Xi(X)\rangle_H = \langle \Xi(\widetilde{e}_i), \Xi(X)\rangle_H = \langle \widetilde{e}_i, X\rangle_V$
Define the maps :
$\widetilde{\pi}_V : V_0 \to \mathbb{R}_0^I :: \widetilde{\pi}_V\left(\sum_{i\in I} \widetilde{x}_i \widetilde{e}_i\right) = \widetilde{x} = (\widetilde{x}_i)_{i\in I}$
$\Upsilon : V_0 \to H_0 :: \Upsilon = \widetilde{\pi}_H^{-1} \circ \widetilde{\pi}_V^{-1}$

which associates to each vector of V the vector of H with the same components in the orthonormal bases, then :

$\forall X \in O : \Upsilon(X) = \Xi(X)$

and Υ is a bijective, linear map, which preserves the scalar product, so it is continuous and is an isometry.

v) There is a bijective linear map : $\mathcal{L}_H \in GL(H_0; H_0)$ such that : $\forall i \in I$: $\varepsilon_i = \mathcal{L}_H(\widetilde{\varepsilon}_i)$.

$(\widetilde{\varepsilon}_i)_{i\in I}$ is a basis of H_0 thus $\varepsilon_i = \sum_{j\in I} [\mathcal{L}_H]_i^j \widetilde{\varepsilon}_j$ where only a finite number of coefficients $[\mathcal{L}_H]_i^j$ is non null.

Let us define : $\varpi_i : H_0 \to \mathbb{R} :: \varpi_i \left(\sum_{j\in I} \psi_j \varepsilon_j\right) = \psi_i$

This map is continuous at $\psi = 0$ on H_0 :

take $\psi \in H_0, \|\psi\| \to 0$

then $\psi = \sum_{i\in I} \langle \widetilde{\varepsilon}_i, \psi\rangle_H \widetilde{\varepsilon}_i$ and $\widetilde{\psi}_j = \langle \widetilde{\varepsilon}_i, \psi\rangle_H \to 0$

so if $\|\psi\| < r$ then $\|\psi\|^2 = \sum_{j\in I} \left|\widetilde{\psi}_j\right|^2 < r^2$ and $\forall j \in I : \left|\widetilde{\psi}_j\right| < r$

$\psi_i = \sum_{j\in J} [\mathcal{L}_H]_i^j \widetilde{\psi}_j \Rightarrow |\psi_i| < \varepsilon \sum_{j\in I} \max\left|[\mathcal{L}_H]_i^j\right|$ and $\left(\left|[\mathcal{L}_H]_i^j\right|\right)_{j\in I}$ is bounded $\Rightarrow |\psi_i| \to 0$

Thus ϖ_i is continuous and belongs to the topological dual H_0' of H_0. It can be extended as a continuous map $\overline{\varpi}_i \in H'$ according to the Hahn-Banach theorem (Maths.958). Because H is a Hilbert space, there is a vector $\phi_i \in H$ such that : $\forall \psi \in H : \overline{\varpi}_i(\psi) = \langle \phi_i, \psi\rangle_H$ so that :

$\forall X \in O : \Upsilon(X) = \Xi(X) = \sum_{i\in I} \psi_i \varepsilon_i$

$= \sum_{i\in I} \langle \phi_i, \psi\rangle_H \varepsilon_i = \sum_{i\in I} \langle \phi_i, \Xi(X)\rangle_H \varepsilon_i$

$\forall i \in I$:

$\Xi(e_i) = \varepsilon_i = \Upsilon(e_i) = \sum_{j\in I} \langle \phi_j, \varepsilon_i\rangle_H \varepsilon_j \Rightarrow \langle \phi_j, \varepsilon_i\rangle_H = \delta_{ij}$

$\Xi(\widetilde{e}_i) = \sum_{j\in I} \langle \phi_j, \Xi(\widetilde{e}_i)\rangle_H \varepsilon_j = \widetilde{\varepsilon}_i = \sum_{j\in I} \langle \phi_j, \widetilde{\varepsilon}_i\rangle_H \varepsilon_j$

vi) The map $\Upsilon : O \to \Omega$ is a linear chart of M, using two orthonormal bases : it is continuous, bijective so it is an homeomorphism, and is obviously compatible with the chart Ξ. ■

2.4.1 Remarks

i) Because $(\widetilde{\varepsilon}_i)_{i\in I}$ is a Hilbertian basis of the separable infinite dimensional Hilbert space H, I is a countable set which can be identified to $\mathbb{N}$. The assumption about $(e_i)_{i\in I}$ is that it is a Hamel basis, which is the most general because any vector space has one. From the proposition above we see that this basis must be of cardinality $\aleph_0$. Hamel bases of infinite dimensional normed vector spaces must be uncountable, however our assumption about V is that it is a Fréchet space, which is a metrizable but not a normed space. If V is a Banach vector space then, according to the Mazur theorem, it implies that there it has an infinite dimensional vector subspace W which has a Shauder basis : $\forall X \in W : X = \sum_{i\in I} x_i e_i$ where the sum is understood in the topological limit. Then the same reasoning as above shows that the closure of W is itself a Hilbert space. Moreover it has been proven that any separable Banach space is homeomorphic to a Hilbert space, and most of the applications will concern spaces of integrable functions (or sections of vector bundle endowed with a norm) which are separable Fréchet spaces.

One interesting fact is that we assume that the variables belong to an open subset O of V. The main concern is to allow for variables which can take values

only in some bounded domain. But this assumption addresses also the case of a Banach vector space which is "hollowed out" : O can be itself a vector subspace (in an infinite dimensional vector space a vector subspace can be open), for instance generated by a countable subbasis of a Hamel basis, and we assume explicitly that the basis $(e_i)_{i\in I}$ belongs to O.

ii) For $O = V$ we have a largest open Ω_V and a linear map $\Upsilon : V \to \Omega_V$ with domain V.

iii) To each (Hamel) basis on V is associated a linear chart Υ of the manifold, such that a point of M has the same coordinates both in V and H. So Υ depends on the choice of the basis, and similarly the positive kernel K_V depends on the basis.

iv) In the proof we have introduced a map : $K_V : O \times O \to \mathbb{R} :: K_V(X, Y)$ which is not bilinear, but is definite positive in a precise way. It plays an important role in several following demonstrations. From a physical point of view it can be seen as related to the probability of transition between two states X, Y often used in QM.

2.5 Complex structure

The variables X and vector space V are real and H is a real Hilbert space. The condition that the vector space V is real is required only in Theoorem 2 to prove the existence of a Hilbert space, because the Henderson's theorem holds only for real structures. However, as it is easily checked, if H exists, all the following theorems hold even if H is a complex Hilbert space. This is specially useful when the space V over which the maps X are defined is itself a complex Hilbert space, as this is often the case.

Moreover it can be useful to endow H with the structure of a complex Hilbert space : the set does not change but one distinguishes real and imaginary components, and the scalar product is given by a Hermitian form. Notice that this is a convenience, not a necessity.

Theorem 4 *Any real separable infinite dimensional Hilbert space can be endowed with the structure of a complex separable Hilbert space*

Proof. H has a infinite countable Hilbertian basis $(\varepsilon_\alpha)_{\alpha\in\mathbb{N}}$ because it is separable.

A complex structure is defined by a linear map : $J \in \mathcal{L}(H; H)$ such that $J^2 = -Id$. Then the operation : $i \times \psi$ is defined by : $i\psi = J(\psi)$.

Define :

$J(\varepsilon_{2\alpha}) = \varepsilon_{2\alpha+1}; J(\varepsilon_{2\alpha+1}) = -\varepsilon_{2\alpha}$

$\forall\psi \in H : i\psi = J(\psi)$

So : $i(\varepsilon_{2\alpha}) = \varepsilon_{2\alpha+1}; i(\varepsilon_{2\alpha+1}) = -\varepsilon_{2\alpha}$

The bases $\varepsilon_{2\alpha}$ or $\varepsilon_{2\alpha+1}$ are complex bases of H :

$\psi = \sum_\alpha \psi^{2\alpha}\varepsilon_{2\alpha} + \psi^{2\alpha+1}\varepsilon_{2\alpha+1} = \sum_\alpha \left(\psi^{2\alpha} - i\psi^{2\alpha+1}\right)\varepsilon_{2\alpha}$

$= \sum_\alpha \left(-i\psi^{2\alpha} + \psi^{2\alpha+1}\right)\varepsilon_{2\alpha+1}$

$\|\psi\|^2 = \sum_\alpha \left|\psi^{2\alpha} - i\psi^{2\alpha+1}\right|^2$
$= \sum_\alpha \left|\psi^{2\alpha}\right|^2 + \left|\psi^{2\alpha+1}\right|^2 + i\left(-\overline{\psi}^{2\alpha}\psi^{2\alpha+1} + \psi^{2\alpha}\overline{\psi}^{2\alpha+1}\right)$
$= \sum_\alpha \left|\psi^{2\alpha}\right|^2 + \left|\psi^{2\alpha+1}\right|^2 + i\left(-\psi^{2\alpha}\psi^{2\alpha+1} + \psi^{2\alpha}\psi^{2\alpha+1}\right)$
Thus $\varepsilon_{2\alpha}$ is a Hilbertian complex basis
H has a structure of complex vector space that we denote $H_\mathbb{C}$
The map : $T : H \to H_\mathbb{C} : T(\psi) = \sum_\alpha \left(\psi^{2\alpha} - i\psi^{2\alpha+1}\right)\varepsilon_{2\alpha}$ is linear and continuous
The map : $\overline{T} : H \to H_\mathbb{C} : \overline{T}(\psi) = \sum_\alpha \left(\psi^{2\alpha} + i\psi^{2\alpha+1}\right)\varepsilon_{2\alpha}$ is antilinear and continuous
Define : $\gamma(\psi, \psi') = \left\langle \overline{T}(\psi), T(\psi')\right\rangle_H$
γ is sesquilinear
$\gamma(\psi, \psi') = \left\langle \sum_\alpha \left(\psi^{2\alpha} + i\psi^{2\alpha+1}\right)\varepsilon_{2\alpha}, \sum_\alpha \left(\psi'^{2\alpha} - i\psi'^{2\alpha+1}\right)\varepsilon_{2\alpha}\right\rangle_H$
$= \sum_\alpha \left(\psi^{2\alpha} + i\psi^{2\alpha+1}\right)\left(\psi'^{2\alpha} - i\psi'^{2\alpha+1}\right)$
$= \sum_\alpha \psi^{2\alpha}\psi'^{2\alpha} + \psi^{2\alpha+1}\psi'^{2\alpha+1} + i\left(\psi^{2\alpha+1}\psi'^{2\alpha} - \psi^{2\alpha}\psi'^{2\alpha+1}\right)$
$\gamma(\psi, \psi) = 0 \Rightarrow \langle\psi, \psi\rangle_H = 0 \Rightarrow \psi = 0$
Thus γ is definite positive ■

2.6 Decomposition of the Hilbert space

V is the product $V = V_1 \times V_2 ... \times V_N$ of vector spaces, thus the proposition implies that the Hilbert space H is also the direct product of Hilbert spaces $H_1 \times H_2 ... \times H_N$ or equivalently $H = \oplus_{k=1}^N H_k$ where H_k are Hilbert vector subspaces of H. More precisely :

Theorem 5 *If the model is comprised of N continuous variables $(X_k)_{k=1}^N$, each belonging to a separable Fréchet vector space V_k, then the real Hilbert space H of states of the system is the Hilbert sum of N Hilbert space $H = \oplus_{k=1}^N H_k$ and any vector ψ representing a state of the system is uniquely the sum of N vectors ψ_k, each image of the value of one variable X_k in the state ψ*

Proof. By definition $V = \prod_{k=1}^N V_k$ .The set $V_k^0 = \{0, .., V_k, ...0\} \subset V$ is a vector subspace of V. A basis of V_k^0 is a subfamily $(e_i)_{i\in I_k}$ of a basis $(e_i)_{i\in I}$ of V. V_k^0 has for image by the continuous linear map Υ a closed vector subspace H_k of H. Any vector X of V reads : $X \in \prod_{k=1}^N V_k : X = \sum_{k=1}^N \sum_{i\in I_k} x^i e_i$ and it has for image by $\Upsilon : \psi = \Upsilon(X) = \sum_{k=1}^N \sum_{i\in I_k} x^i \varepsilon_i = \sum_{k=1}^N \psi_k$ with $\psi_k \in H_k$ .This decomposition of $\Upsilon(X)$ is unique.

Conversely, the family $(e_i)_{i\in I_k}$ has for image by Υ the set $(\varepsilon_i)_{i\in I_k}$ which are linearly independent vectors of H_k.It is always possible to build an orthonormal basis $(\widetilde{\varepsilon}_i)_{i\in I_k}$ from these vectors as done previously. H_k is a closed subspace of H, so it is a Hilbert space. The map : $\widehat{\pi}_k : \ell^2(I_k) \to H_k :: \widehat{\pi}_k(x) = \sum_{i\in I_k} x^i \widetilde{\varepsilon}_i$ is an isomorphism of Hilbert spaces and :$\forall\psi \in H_k : \psi = \sum_{i\in I_k} \langle\widetilde{\varepsilon}_i, \psi\rangle_H \widetilde{\varepsilon}_i$.

$\forall \psi_k \in H_k, \psi_l \in H_l, k \neq l : \langle \psi_k, \psi_l \rangle_H = \left\langle \Upsilon^{-1}(\psi_k), \Upsilon^{-1}(\psi_l) \right\rangle_E = 0$

Any vector $\psi \in H$ reads : $\psi = \sum_{k=1}^{N} \pi_k(\psi)$ with the orthogonal projection $\pi_k : H \to H_k :: \pi_k(\psi) = \sum_{i \in I_k} \langle \widetilde{\varepsilon}_i, \psi \rangle_H \widetilde{\varepsilon}_i$ so H is the Hilbert sum of the H_k ■

As a consequence the definite positive kernel of (V, Υ) decomposes as :
$K((X_1, ...X_N), (X'_1, ...X'_N))$
$= \sum_{k=1}^{N} K_k(X_k, X'_k)$
$= \sum_{k=1}^{N} \langle \Upsilon(X_k), \Upsilon(X'_k) \rangle_{H_k}$

This decomposition comes handy when we have to translate relations between variables into relations between vector states, notably it they are linear. But it requires that we keep the real Hilbert space structure.

2.7 Discrete variables

It is common in a model to have discrete variables $(D_k)_{k=1}^{K}$, taking values in a finite discrete set. They correspond to different cases:

i) the discrete variables identify different elementary systems (such as different populations of particles) which coexist simultaneously in the same global system, follow different rules of behavior, but interact together. We will see later how to deal with these cases (tensorial product).

ii) the discrete variables identify different populations, whose interactions are not relevant. Actually one could consider as many different systems but, by putting them together, one increases the size of the samples of data and improve the statistical estimations. They are not of great interest here, in a study of formal models.

iii) the discrete variables represent different kinds of behaviors, which cannot be strictly identified with specific populations. Usually a discrete variable is then used as a proxy for a quantitative parameter which tells how close the system is from a specific situation.

We will focus on this third case. The system is represented as before by quantitative variables X, whose possible values belong to some set M, which has the structure of an infinite dimensional manifold. The general idea in the third case is that the possible states of the system can be regrouped in two distinct subsets. That we formalize in the following assumption : the set O of possible states of the system has two connected components O_1, O_2

Theorem 6 *If the condition of the theorem 2 are met, and the set O of possible states of the system has two connected components O_1, O_2 then there is a continuous function $f : H \to [0,1]$ such that $f(\Upsilon(X)) = 1$ in O_1 and $f(\Upsilon(X)) = 0$ in O_2*

Proof. The connected components O_1, O_2 of a topological space are closed, so O_1, O_2 are disjoint and both open and closed in V (Maths.624). Using a linear continuous map Υ then Ω has itself two connected components, $\Omega_1 = \Upsilon^{-1}(O_1), \Omega_2 = \Upsilon^{-1}(O_2)$ both open and closed, and disjoint. H is metric, so

it is normal (Maths.705). Ω_1, Ω_2 are disjoint and closed in H. Then, by the Urysohn's Theorem (Maths.596) there is a continuous function f on H valued in [0,1] such that $f(\psi) = 1$ in H_1 and $f(\psi) = 0$ in H_2. ■

The set of continuous, bounded functions is a Banach vector space, so it is always possible, in these conditions, to replace a discrete variable by a quantitative variable with the same features.

3 OBSERVABLES

The key point in the conditions 1 above is that the variables are maps, which take an infinite number of values (usually non countable). So the variables would require the same number of data to be totally known, which is impossible. The physicist estimates the variable by statistical methods. But any practical method involves a first step : the scope of all maps is reduced from V to a smaller subset W, so that any map of W can be characterized by a finite number of parameters. The procedure sums up to replace X by another variable $\Phi(X)$ that we will call an observable, which is then estimated from a finite batch of data. The mechanism of estimating the variables $X \subset V$ is then the following :

- the observer collects data, as a set $Y = \{x_p\}_{p=1}^N$ of values assumed to be taken by the variable X, in the mathematical format fitted to X (scalars, vectors,..for different values of the arguments)

- he proceeds to the estimation $\widehat{X}$ of the map $\Phi(X)$ by statistical adjustment to the data $\{x_p\}_{p=1}^N$. Because there are a finite number of parameters (the coordinates of $\Phi(X)$ in W) this is possible

- the estimation is : $\widehat{X} = \varphi(Y) \in W$: this is a map which is a simplified version of X.

The procedure of the replacement of X by $\Phi(X)$, called the choice of a **specification**, is done by the physicist, and an observable is not unique. However we make three general assumptions about Φ :

Definition 7 *i) an* ***observable*** *is a linear map :* $\Phi \in L(V;V)$

ii) the range of an observable is a finite dimensional vector W subspace of V : $W \subset V, \dim \Phi(W) < \infty$

iii) $\forall X \in O, \Phi(X)$ *is an admissible value, that is* $\Phi(O) \subset O$.

Using the linear chart Υ given by any basis, to Φ one can associate a map :

$$\widehat{\Phi} : H \to H :: \widehat{\Phi} = \Upsilon \circ \Phi \circ \Upsilon^{-1} \tag{1}$$

and $\widehat{\Phi}$ is an operator on H. And conversely.

The image of W by Υ is a finite dimensional vector subspace $H_\Phi = \Upsilon(W)$ of H, so it is closed and a Hilbert space : $\widehat{\Phi} \in \mathcal{L}(H; H_\Phi)$

$$\begin{array}{ccccccc}
 & & & \Phi & & & \\
 & V & \to & \to & \to & W & \\
 & \downarrow & & & & \downarrow & \\
\Upsilon & \downarrow & & & & \downarrow & \Upsilon \\
 & \downarrow & & \widehat{\Phi} & & \downarrow & \\
 & H & \to & \to & \to & H_\Phi &
\end{array}$$

3.1 Primary observables

The simplest specification for an observable is, given a basis $(e_i)_{i\in I}$, to define Φ as the projection on the subspace spanned by a finite number of vectors of the basis. For instance if X is a function $X(t)$ belonging to some space such as : $X(t) = \sum_{n\in\mathbb{N}} a_n e_n(t)$ where $e_n(t)$ are fixed functions, then a primary observable would be $Y_J(X(t)) = \sum_{n=0}^{N} a_n e_n(t)$ meaning that the components $(a_n)_{n>N}$ are discarded and the data are used to compute $(a_n)_{n=0}^{N}$. To stay at the most general level, we define :

Definition 8 *A **primary observable** $\Phi = Y_J$ is the projection of $X = \{X_k, k = 1...N\}$ on the vector subspace V_J spanned by the vectors $(e_i)_{i\in J} \equiv \left(e_i^k\right)_{i\in J_k}$ where $J = \prod_{k=1}^{N} J_k \subset I = \prod_{k=1}^{N} I_k$ is a finite subset of I and $(\varepsilon_i)_{i\in I} = \prod_{k=1}^{N} \left(e_i^k\right)_{i\in I_k}$ is a basis of V.*

So the procedure can involve simultaneously several variables. It requires *the choice of a finite set of independent vectors of V.*

Theorem 9 *To any primary observable Y_J is associated uniquely a self-adjoint, compact, trace-class operator $\widehat{Y}_J$ on $H : Y_J = \Upsilon^{-1} \circ \widehat{Y}_J \circ \Upsilon$ such that the measure $Y_J(X)$ of the primary observable Y_J, if the system is in the state $X \in O$, is*

$Y_J(X) = \sum_{i\in I} \left\langle \phi_i, \widehat{Y}_J(\Upsilon(X)) \right\rangle_H e_i$

Proof. i) We use the notations and definitions of the previous section. The family of variables $X = (X_k)_{k=1}^{N}$ define the charts : $\Xi : O \to \Omega$ and the basis $(e_i)_{i\in I}$ defines the bijection $\Upsilon : V \to H$

$\forall X = \sum_{i\in I} x_i e_i \in O :$

$\Upsilon(X) = \sum_{i\in I} x_i \Upsilon(e_i) = \sum_{i\in I} x_i \varepsilon_i = \sum_{i\in I} \langle \phi_i, \Upsilon(X)\rangle_H \varepsilon_i$

$\Leftrightarrow x_i = \langle \phi_i, \Upsilon(X)\rangle_H$

$\forall i, j \in I : \langle \phi_i, \varepsilon_j\rangle_H = \delta_{ij}$

ii) The primary observable Y_J is the map :

$Y_J : V \to V_J :: Y_J(X) = \sum_{j\in J} x_j e_j$

This is a projection : $Y_J^2 = Y_J$

$Y_J(X) \in O$ so it is associated to a vector of H :

$\Upsilon(Y_J(X)) = \Upsilon\left(\sum_{j\in J} x_j e_j\right) = \sum_{j\in J} \langle \phi_j, \Upsilon(Y_J(X))\rangle_H \varepsilon_j$

$= \sum_{j\in J} \langle \phi_j, \Upsilon(X)\rangle_H \varepsilon_j$

iii) $\forall X \in O : \Upsilon(Y_J(X)) \in H_J$ where H_J is the vector subspace of H spanned by $(\varepsilon_j)_{j\in J}$. It is finite dimensional, thus it is closed in H. There is a unique map (Math.1111) :

$\widehat{Y}_J \in \mathcal{L}(H; H) :: \widehat{Y}_J^2 = \widehat{Y}_J, \widehat{Y}_J = \widehat{Y}_J^*$

$\widehat{Y}_J$ is the orthogonal projection from H onto H_J. It is linear, self-adjoint, and compact because its range is a finite dimensional vector subspace. As a projection : $\left\|\widehat{Y}_J\right\| = 1$.

$\widehat{Y}_J$ is a Hilbert-Schmidt operator (Maths.1143) : take the Hilbertian basis $\widetilde{\varepsilon}_i$ in H:

$\sum_{i\in I}\left\|\widehat{Y}_J\left(\widetilde{\varepsilon}_i\right)\right\|^2 = \sum_{ij\in J}\left|\langle\phi_j,\widetilde{\varepsilon}_i\rangle\right|^2\left\|\varepsilon_j\right\|^2 = \sum_{j\in J}\left\|\phi_j\right\|^2\left\|\varepsilon_j\right\|^2 < \infty$

$\widehat{Y}_J$ is a trace class operator (Maths.1147) with trace $\dim H_J$

$\sum_{i\in I}\left\langle\widehat{Y}_J\left(\widetilde{\varepsilon}_i\right),\widetilde{\varepsilon}_i\right\rangle = \sum_{ij\in J}\left\langle\phi_j,\widetilde{\varepsilon}_i\right\rangle\left\langle\varepsilon_j,\widetilde{\varepsilon}_i\right\rangle = \sum_{j\in J}\left\langle\phi_j,\varepsilon_j\right\rangle = \sum_{j\in J}\delta_{jj} = \dim H_J$

iv) $\forall\psi\in H_J : \widehat{Y}_J\left(\psi\right) = \psi$

$\forall X\in O : \Upsilon\left(Y_J\left(X\right)\right)\in H_J$

$\forall X\in O : \Upsilon\left(Y_J\left(X\right)\right) = \widehat{Y}_J\left(\Upsilon\left(X\right)\right) \Leftrightarrow Y_J\left(X\right) = \Upsilon^{-1}\circ\widehat{Y}_J\left(\Upsilon\left(X\right)\right) \Leftrightarrow Y_J = \Upsilon^{-1}\circ\widehat{Y}_J\circ\Upsilon$

v) The value of the observable reads : $Y_J\left(X\right) = \sum_{i\in I}\left\langle\phi_i,\widehat{Y}_J\left(\Upsilon\left(X\right)\right)\right\rangle_H e_i$

■

3.2 von Neumann algebras

There is a bijective correspondence between the projections, meaning the maps $P\in\mathcal{L}\left(H;H\right) : P^2 = P, P = P^*$ and the closed vector subspaces of H (Maths.1111). Then P is the orthogonal projection on the vector subspace. So the operators $\widehat{Y}_J$ for any finite subset J of I are the orthogonal projections on the finite dimensional, and thus closed, vector subspace H_J spanned by $\left(\varepsilon_j\right)_{j\in J}$.

We will enlarge the family of primary observables in several steps, in keeping the same basis $\left(e_i\right)_{i\in I}$ of V.

1. For any given basis $\left(e_i\right)_{i\in I}$ of V, we extend the definition of these operators $\widehat{Y}_J$ to any finite or infinite, subset of I by taking $\widehat{Y}_J$ as the orthogonal projection on the closure $\overline{H_J}$ in H of the vector subspace H_J spanned by $\left(\varepsilon_j\right)_{j\in J}$: $\overline{H_J} = \overline{Span\left(\varepsilon_j\right)_{j\in J}}$.

Theorem 10 *The operators* $\left\{\widehat{Y}_J\right\}_{J\subset I}$ *are self-adjoint and commute*

Proof. Because they are projections the operators $\widehat{Y}_J$ are such that : $\widehat{Y}_J^2 = \widehat{Y}_J, \widehat{Y}_J^* = \widehat{Y}_J$

$\widehat{Y}_J$ has for eigen values :

1 for $\psi\in\overline{H_J}$

0 for $\psi\in\left(\overline{H_J}\right)^{\perp}$

For any subset J of I, by the Gram-Schmidt procedure one can built an orthonormal basis $\left(\widetilde{\varepsilon}_i\right)_{i\in J}$ of H_J starting with the vectors $\left(\varepsilon_i\right)_{i\in J}$ and an orthonormal basis $\left(\widetilde{\varepsilon}_i\right)_{i\in J^c}$ of H_{J^c} starting with the vectors $\left(\varepsilon_i\right)_{i\in J^c}$

Any vector $\psi \in H$ can be written :

$\psi = \sum_{j\in I} x_j \widetilde{\varepsilon}_j = \sum_{j\in J} x_j \widetilde{\varepsilon}_j + \sum_{j\in J^c} x_j \widetilde{\varepsilon}_j$ with $(x_j)_{j\in I} \in \ell^2(I)$

$\overline{H_J}$ is defined as $\sum_{j\in J} x_j \widetilde{\varepsilon}_j$ with $(x_j)_{j\in J} \in \ell^2(J)$ and similarly $\overline{H_{J^c}}$ is defined as $\sum_{j\in J^c} x_j \widetilde{\varepsilon}_j$ with $(x_j)_{j\in J^c} \in \ell^2(J^c)$

So $\widehat{Y}_J$ can be defined as : $\widehat{Y}_J\left(\sum_{j\in I} x_j \widetilde{\varepsilon}_j\right) = \sum_{j\in J} x_j \widetilde{\varepsilon}_j$

For any subsets $J_1, J_2 \subset I$:

$\widehat{Y}_{J_1} \circ \widehat{Y}_{J_2} = \widehat{Y}_{J_1\cap J_2} = \widehat{Y}_{J_2} \circ \widehat{Y}_{J_1}$

$\widehat{Y}_{J_1\cup J_2} = \widehat{Y}_{J_1} + \widehat{Y}_{J_2} - \widehat{Y}_{J_1\cap J_2} = \widehat{Y}_{J_1} + \widehat{Y}_{J_2} - \widehat{Y}_{J_1} \circ \widehat{Y}_{J_2}$

So the operators commute. ■

2. Let us define $W = Span\left\{\widehat{Y}_i\right\}_{i\in I}$ the vector subspace of $\mathcal{L}(H;H)$ comprised of finite linear combinations of $\widehat{Y}_i$ (as defined in 1 above). The elements $\left\{\widehat{Y}_i\right\}_{i\in I}$ are linearly independent and constitute a basis of W.

The operators $\widehat{Y}_j, \widehat{Y}_k$ are mutually orthogonal for $j \neq k$:

$\widehat{Y}_j \circ \widehat{Y}_k(\psi) = \langle \phi_k, \psi\rangle \langle \phi_j, \varepsilon_k\rangle \varepsilon_j = \langle \phi_k, \psi\rangle \delta_{jk} = \delta_{jk}\widehat{Y}_j(\psi)$

Let us define the scalar product on W :

$\left\langle \sum_{i\in I} a_i \widehat{Y}_i, \sum_{i\in I} b_i \widehat{Y}_i \right\rangle_W = \sum_{i\in I} a_i b_i$

$\left\| \sum_{i\in I} a_i \widehat{Y}_i \right\|_W^2 = \sum_{i\in I} a_i^2 \left\| \widehat{Y}_i \right\|_W^2 = \sum_{i\in I} a_i^2$

W is isomorphic to $\mathbb{R}_0^I$ and its closure in $\mathcal{L}(H;H)$: $\overline{W} = \overline{Span\left\{\widehat{Y}_i\right\}_{i\in I}}$ is isomorphic to $\ell^2(I)$, and has the structure of a Hilbert space with :

$\overline{W} = \left\{\sum_{i\in I} a_i \widehat{Y}_i, (a_i)_{i\in I} \in \ell^2(I)\right\}$

3. Let us define A as the algebra generated by any finite linear combination or products of elements $\widehat{Y}_J, J$ finite or infinite, and $\overline{A}$ as the closure of A in $\mathcal{L}(H;H)$: $\overline{A} = \overline{Span\left\{\widehat{Y}_J\right\}_{J\subset I}}$ with respect to the strong topology, that is in norm.

Theorem 11 *$\overline{A}$ is a commutative von Neumann algebra of $\mathcal{L}(H,H)$*

Proof. It is obvious that A is a *subalgebra of $\mathcal{L}(H,H)$ with unit element $Id = \widehat{Y}_I$.

Because its generators are projections, $\overline{A}$ is a von Neumann algebra (Maths.1190).

The elements of $A = Span\left\{\widehat{Y}_J\right\}_{J\subset I}$ that is of finite linear combination of $\widehat{Y}_J$ commute

$Y, Z \in \overline{A} \Rightarrow \exists (Y_n)_{n\in\mathbb{N}}, (Z_n)_{n\in\mathbb{N}} \in A^{\mathbb{N}} : Y_n \to_{n\to\infty} Y, Z_n \to_{n\to\infty} Z$

The composition is a continuous operation.

$Y_n \circ Z_n = Z_n \circ Y_n \Rightarrow \lim(Y_n \circ Z_n) = \lim(Z_n \circ Y_n) = \lim Y_n \circ \lim Z_n = \lim Z_n \circ \lim Y_n = Z \circ Y = Y \circ Z$

So $\overline{A}$ is commutative.

$\overline{A}$ is identical to the bicommutant of its projections, that is to $\overline{A}$" (Maths.1189)

■

This result is of interest because commutative von Neumann algebras are classified : they are isomorphic to the space of functions $f \in L^{\infty}(E,\mu)$ acting by pointwise multiplication $\varphi \to f\varphi$ on functions $\varphi \in L^2(E,\mu)$ for some set E and measure μ (not necessarily absolutely continuous). They are the topic of many studies, notably in ergodic theory. The algebra $\overline{A}$ depends on the choice of a basis $(e_i)_{i\in I}$ and, as can be seen in the formulation through $(\widetilde{\varepsilon}_i)_{i\in I}$, is defined up to a unitary transformation.

In the axiomatisation of QM, it is usual to define a system by a von Neumann algebra of operators on a Hilbert space. We see here how such algebras appear naturally. However the algebra $\overline{A}$ is commutative, and this property is the consequence of the choice of a unique basis $(e_i)_{i\in I}$. It would not hold for primary observables defined through different bases : they do not even constitute an algebra. Any von Neumann algebra is the closure of the linear span of its projections (Maths.1190), and any projection can be defined through a basis, thus one can say that the "observables" (with their usual definition) of a system are the collection of all primary observables (as defined here) for all bases of V. This is a crucial issue in the axiomatic interpretation of QM, but the concept of observables introduced here allows us to deal with this issue and understand how probabilities enter the picture. But, before that, we need to see what can be said about more general observables, which are not just primary.

3.3 Secondary observables

Beyond primary observables, general observables Φ can be studied using spectral theory.

1. A spectral measure defined on a measurable space E with $\sigma-$algebra σ_E and acting on the Hilbert space H is a map : $P : \sigma_E \to \mathcal{L}(H;H)$ such that (Maths.1240) :

i) $P(\varpi)$ is a projection

ii) $P(E) = Id$

iii) $\forall \psi \in H$ the map: $\varpi \to \langle P(\varpi)\psi, \psi\rangle_H = \|P(\varpi)\psi\|^2$ is a finite positive measure on (E,σ_E).

One can show (Maths.1242) that there is a bijective correspondence between the spectral measures on H and the maps : $\chi : \sigma_E \to H$ such that :

i) $\chi(\varpi)$ is a closed vector subspace of H

ii) $\chi(E) = H$

iii) $\forall \varpi, \varpi' \in \sigma_E, \varpi \cap \varpi' = \varnothing : \chi(\varpi) \cap \chi(\varpi') = \{0\}$

then $P(\varpi)$ is the orthogonal projection on $\chi(\varpi)$, denoted : $\widehat{\pi}_{\chi(\varpi)}$

Thus, for any fixed $\psi \neq 0 \in H$ the function $\widehat{\chi}_\psi : \sigma_E \to \mathbb{R} :: \widehat{\chi}_\psi(\varpi) = \frac{\langle \widehat{\pi}_{\chi(\varpi)}\psi, \psi\rangle}{\|\psi\|^2} = \frac{\|\widehat{\pi}_{\chi(\varpi)}\psi\|^2}{\|\psi\|^2}$ is a probability law on (E,σ_E).

2. An application of standard theorems on spectral measures (Maths.1243, 1245) tells that, for any bounded measurable function $f : E \to \mathbb{R}$, the spectral integral : $\int_E f(\xi)\, \widehat{\pi}_{\chi(\xi)}$ defines a continuous operator $\widehat{\Phi}_f$ on H. $\widehat{\Phi}_f$ is such that :

$\forall \psi, \psi' \in H : \left\langle \widehat{\Phi}_f(\psi), \psi' \right\rangle = \int_E f(\xi) \left\langle \widehat{\pi}_{\chi(\xi)}(\psi), \psi' \right\rangle$

And conversely (Math.1252), for any continuous normal operator $\widehat{\Phi}$ on H, that is such that :

$\widehat{\Phi} \in \mathcal{L}(H; H) : \widehat{\Phi} \circ \widehat{\Phi}^* = \widehat{\Phi}^* \circ \widehat{\Phi}$ with the adjoint $\widehat{\Phi}^*$

there is a unique spectral measure P on $(\mathbb{R}, \sigma_{\mathbb{R}})$ such that $\widehat{\Phi} = \int_{Sp(\widehat{\Phi})} sP(s)$ where $Sp(\widehat{\Phi}) \subset \mathbb{R}$ is the spectrum of $\widehat{\Phi}$.

So there is a map $\chi : \sigma_{\mathbb{R}} \to H$ where $\sigma_{\mathbb{R}}$ is the Borel algebra of $\mathbb{R}$ such that :

$\chi(\varpi)$ is a closed vector subspace of H

$\chi(\mathbb{R}) = Id$

$\forall \varpi, \varpi' \in \sigma_{\mathbb{R}}, \varpi \cap \varpi' = \varnothing \Rightarrow \chi(\varpi) \cap \chi(\varpi') = \{0\}$

and $\widehat{\Phi} = \int_{Sp(\widehat{\Phi})} s\widehat{\pi}_{\chi(s)}$

The spectrum $Sp(\widehat{\Phi})$ is a non empty compact subset of $\mathbb{R}$. If $\widehat{\Phi}$ is normal then $\lambda \in Sp(\widehat{\Phi}) \Leftrightarrow \overline{\lambda} \in Sp(\widehat{\Phi}^*)$.

For any fixed $\psi \neq 0 \in H$ the function $\widehat{\mu}_\psi : \sigma_{\mathbb{R}} \to \mathbb{R} :: \widehat{\mu}_\psi(\varpi) = \frac{\left\langle \widehat{\pi}_{\chi(\varpi)}\psi, \psi \right\rangle}{\|\psi\|^2} = \frac{\left\| \widehat{\pi}_{\chi(\varpi)}\psi \right\|^2}{\|\psi\|^2}$ is a probability law on $(\mathbb{R}, \sigma_{\mathbb{R}})$.

3. We will define :

Definition 12 *A* ***secondary observable*** *is a linear map $\Phi \in L(V; V)$ valued in a finite dimensional vector subspace of V, such that $\widehat{\Phi} = \Upsilon \circ \Phi \circ \Upsilon^{-1}$ is a normal operator : $\widehat{\Phi} \circ \widehat{\Phi}^* = \widehat{\Phi}^* \circ \widehat{\Phi}$ with the adjoint $\widehat{\Phi}^*$*

Theorem 13 *Any secondary observable Φ is a compact, continuous map, its associated map $\widehat{\Phi} = \Upsilon \circ \Phi \circ \Upsilon^{-1}$ is a compact, self-adjoint, Hilbert-Schmidt and trace class operator.*

$\Phi = \sum_{p=1}^{n} \lambda_p Y_{J_p}$ where $\left(Y_{J_p}\right)_{p=1}^{N}$ are primary observables associated to a basis $(e_i)_{i \in I}$ of V and $(J_p)_{p=1}^{n}$ are disjoint finite subsets of I

Proof. i) $\widehat{\Phi}(H)$ is a finite dimensional vector subspace of H. So :

$\widehat{\Phi}$ has 0 for eigen value, with an infinite dimensional eigen space H_c.

$\Phi, \widehat{\Phi}$ are compact and thus continuous (Maths.912).

ii) As $\widehat{\Phi}$ is continuous and normal, there is a unique spectral measure P on $(\mathbb{R}, \sigma_{\mathbb{R}})$ such that $\widehat{\Phi} = \int_{Sp(\widehat{\Phi})} sP(s)$ where $Sp(\widehat{\Phi}) \subset \mathbb{R}$ is the spectrum of $\widehat{\Phi}$. As $\widehat{\Phi}$ is compact, by the Riesz theorem (Maths.1142) its spectrum is either finite or is a countable sequence converging to 0 (which may or not be an eigen value)

and, except possibly for 0, is identical to the set $(\lambda_p)_{p\in\mathbb{N}}$ of its eigen values (Maths.1020). For each distinct eigen value the eigen spaces H_p are orthogonal and H is the direct sum $H = \oplus_{p\in\mathbb{N}} H_p$. For each non null eigen value λ_p the eigen space H_p is finite dimensional.

Let λ_0 be the eigen value 0 of $\widehat{\Phi}$. So : $\widehat{\Phi} = \sum_{p\in\mathbb{N}} \lambda_p \widehat{\pi}_{H_p}$ and any vector of H reads : $\psi = \sum_{p\in\mathbb{N}} \psi_p$ with $\psi_p = \widehat{\pi}_{H_p}(\psi)$

Because $\widehat{\Phi}(H)$ is finite dimensional, the spectrum is finite and the non null eigen values are $(\lambda_p)_{p=1}^n$, the eigen space corresponding to 0 is $H_c = \left(\oplus_{p=1}^n H_p\right)^\perp$

$\forall \psi \in H : \psi = \psi_c + \sum_{p=1}^n \psi_p$ with $\psi_p = \widehat{\pi}_{H_p}(\psi), \psi_c = \widehat{\pi}_{H_c}(\psi)$

$\widehat{\Phi} = \sum_{p=1}^n \lambda_p \widehat{\pi}_{H_p}$

Its adjoint reads : $\widehat{\Phi}^* = \sum_{p\in\mathbb{N}} \overline{\lambda}_p \widehat{\pi}_{H_p} = \sum_{p\in\mathbb{N}} \lambda_p \widehat{\pi}_{H_p}$ because H is a real Hilbert space

$\widehat{\Phi}$ is then self-adjoint, Hilbert-Schmidt and trace class, as the sum of the trace class operators $\widehat{\pi}_{H_p}$.

iii) The observable reads :

$\Phi = \sum_{p=1}^n \lambda_p \pi_p$ where $\pi_p = \Upsilon^{-1} \circ \widehat{\pi}_{H_p} \circ \Upsilon$ is the projection on a finite dimensional vector subspace of V :

$\pi_p \circ \pi_q = \Upsilon^{-1} \circ \widehat{\pi}_{H_p} \circ \Upsilon \circ \Upsilon^{-1} \circ \widehat{\pi}_{H_q} \circ \Upsilon = \Upsilon^{-1} \circ \widehat{\pi}_{H_p} \circ \widehat{\pi}_{H_q} \circ \Upsilon = \delta_{pq} \Upsilon^{-1} \circ \widehat{\pi}_{H_p} \circ \Upsilon = \delta_{pq} \pi_p$

$\Phi \circ \pi_p = \lambda_p \pi_p$ so $\pi_p(V) = V_p$ is the eigen space of Φ for the eigen value λ_p and the subspaces $(V_p)_{p=1}^n$ are linearly independent.

By choosing any basis $(e_i)_{i\in J_p}$ of V_p, and $(e_i)_{i\in J^c}$ with $J^c = \complement_I \left(\oplus_{p=1}^n J_n\right)$ for the basis of $V_c = Span\left((e_i)_{i\in J^c}\right)$

$X = Y_{J^c}(X) + \sum_{p=1}^n Y_{J_p}(X)$

the observable Φ reads : $\Phi = \sum_{p=1}^n \lambda_p Y_{J_p}$ ■

We have :

$Y_{J_p}(X) = \sum_{i\in J_p} \left\langle \phi_i, \widehat{Y}_{Jp}(\Upsilon(X)) \right\rangle_H e_i$

$\Phi(X) = \sum_{p=1}^n \lambda_p \sum_{i\in J_p} \left\langle \phi_i, \widehat{Y}_{Jp}(\Upsilon(X)) \right\rangle_H e_i$

$= \sum_{i\in I} \left\langle \phi_i, \sum_{p=1}^n \lambda_p \widehat{Y}_{Jp}(\Upsilon(X)) \right\rangle_H e_i$

$= \sum_{i\in I} \left\langle \phi_i, \widehat{\Phi}(\Upsilon(X)) \right\rangle_H e_i$

$\Phi, \widehat{\Phi}$ have invariant vector spaces, which correspond to the direct sum of the eigen spaces.

The probability law $\widehat{\mu}_\psi : \sigma_{\mathbb{R}} \to \mathbb{R}$ reads :

$\widehat{\mu}_\psi(\varpi) = \Pr(\lambda_p \in \varpi) = \frac{\left\|\widehat{\pi}_{H_p}(\psi)\right\|^2}{\|\psi\|^2}$

To sum up :

Theorem 14 *For any primary or secondary observable Φ, there is a basis $(e_i)_{i\in I}$ of V, a compact, self-adjoint, Hilbert-Schmidt and trace class operator $\widehat{\Phi}$ on the associated Hilbert space H such that :*

$\widehat{\Phi} = \Upsilon \circ \Phi \circ \Upsilon^{-1}$

if the system is in the state $X = \sum_{i \in I} \langle \phi_i, \Upsilon(X) \rangle_H e_i$ *the value of the observable is :* $\Phi(X) = \sum_{i \in I} \left\langle \phi_i, \widehat{\Phi}(\Upsilon(X)) \right\rangle_H e_i$

$\widehat{\Phi}$ has a finite set of eigen values, whose eigen spaces (except possibly for 0) are finite dimensional and orthogonal. The vectors corresponding to the eigen value 0 are never observed, so it is convenient to represent the Hilbert space H through a basis of eigen vectors, each of them corresponding to a definite state, which usually can be identified. This is a method commonly used in Quantum Mechanics, however the vector has also a component in the eigen space corresponding to the null eigen value, which is not observed but exists. Conversely any observable (on V) can be defined through an operator on H with the required properties (compact, normal, it is then self-adjoint). We will come back on this point in the following, when a group is involved.

3.4 Efficiency of an observable

A crucial factor for the quality and the cost of the estimation procedure is the number of parameters to be estimated, which is closely related to the dimension of the vector space $\Phi(V)$, which is finite. The error made by the choice of $\Phi(X)$ when the system is in the state X is : $o_\Phi(X) = X - \Phi(X)$. If two observables Φ, Φ' are such that $\Phi(V), \Phi'(V)$ have the same dimension, one can say that Φ is more efficient than Φ' if : $\forall X : \|o_\Phi(X)\|_V \leq \|o_{\Phi'}(X)\|_V$

To assess the efficiency of a secondary observable Φ it is legitimate to compare Φ to the primary observable Y_J with a set J which has the same cardinality as the dimension of $\oplus_{p=1}^n H_p$.

The error with the choice of Φ is :

$o_\Phi(X) = X - \Phi(X) = Y_c(\psi) + \sum_{p=1}^n (1 - \lambda_p) Y_p(\psi)$

$\|o_\Phi(X)\|_V^2 = \|Y_c(\psi)\|_V^2 + \sum_{p=1}^n (1 - \lambda_p)^2 \|Y_p(\psi)\|^2$

$\widehat{o}_\Phi(\Upsilon(X)) = \Upsilon(X) - \widehat{\Phi}(\Upsilon(X)) = \widehat{\pi}_{H_c}(\psi) + \sum_{p=1}^n (1 - \lambda_p) \widehat{\pi}_{H_p}(\psi)$

$\|\widehat{o}_\Phi(\Upsilon(X))\|^2 = \|\widehat{\pi}_{H_c}(\psi)\|^2 + \sum_{p=1}^n (1 - \lambda_p)^2 \left\|\widehat{\pi}_{H_p}(\psi)\right\|^2 = \|o_\Phi(X)\|_V^2$

And for Y_J : $\|\widehat{o}_{Y_J}(\Upsilon(X))\|^2 = \|\widehat{\pi}_{H_c}(\psi)\|^2$ because $\lambda_p = 1$

So :

Theorem 15 *For any secondary observable there is always a primary observable which is at least as efficient.*

This result justifies the restriction, in the usual formalism, of observables to operators belonging to a von Neumann algebra.

3.5 Statistical estimation and primary observables

At first the definition of a primary observable seems naive, and the previous results will seem obvious to the specialists of Axiomatic QM. After all the definition of a primary observable requires only the choice of a finite number of orthonormal vectors of V. We have already seen that a primary observable is always better than a, more sophisticated, secondary observable. But we have also to compare a primary observable to what is practically done in an experiment, where we have to estimate a map from a batch of data.

Consider a model with variables X, maps, belonging to a Hilbert space H (to keep it simple), from a set M to a normed vector space E, endowed with a scalar product $\langle\rangle_E$. The physicist has a batch of data, that is a finite set $\{x_p \in E, p = 1...N\}$ of N measures of X done at different points $\Omega = \{m_p \in M, p = 1...N\}$: of M : $x_p = X(m_p)$. The estimated map $\widehat{X}$ should be a solution of the collection of equations : $x_p = X(m_p)$ where x_p, m_p are known.

The **evaluation maps**, that we will encounter several times, is the collection of maps $\mathcal{E}(m)$ on H :

$\mathcal{E}(m) : H \to E :: \mathcal{E}(m) Y = Y(m)$

Because H and E are vector spaces $\mathcal{E}(m)$ is a linear map : $\mathcal{E}(m) \in L(H; E)$, depending on both H and E. It can be continuous or not.

The set of solutions of the equations, that is of maps Y of H such that $\forall m_p \in \Omega : Y(m_p) = x_p$ is :

$A = \cap_{m_p \in \Omega} \mathcal{E}(m_p)^{-1}(x_p)$

$Y \in A \Leftrightarrow \forall m \in \Omega : Y(m) = X(m)$

It is not empty because it contains at least X. Its closed convex hull is the set B in H (Maths.361) :

$\forall Z \in B : \exists \alpha \in [0.1], Y, Y' \in A : Z = \alpha Y + (1 - \alpha) Y'$

$\Rightarrow \forall m \in \Omega : Z(m) = x_p$

B is the smallest closed set of H such that all its elements Z are solutions of the equations : $\forall p = 1..N : Z(m_p) = x_p$.

If we specify an observable, we restrict X to a finite dimensional subspace $H_J \subset H$.With the evaluation map $\mathcal{E}_J$ on H_J we can consider the same procedure, but then usually $A_J = \varnothing$. The simplification of the map to be estimated as for consequence that there is no solution to the equations. So the physicist uses a statistical method, that is a map which associates to each batch of data $X(\Omega)$ a map $\varphi(X(\Omega)) = \widehat{X} \in H_J$. Usually $\widehat{X}$ is such that it minimizes the sum of the distance between points in E : $\sum_{m \in \Omega} \left\| \widehat{X}(m) - x_p \right\|_E$ (other additional conditions can be required).

The primary observable Φ gives another solution : $\Phi(X)$ is the orthogonal projection of X on the Hilbert space H_J, it is such that it minimizes the distance between maps :

$\forall Z \in H_J : \|X - Z\|_H \geq \|X - \Phi(X)\|_H$.

$\Phi(X)$ always exist, and does not depend on the choice of an estimation procedure φ. $\Phi(X)$ minimizes the distance between maps in H, meanwhile $\varphi(X(\Omega))$

minimizes distance between points in E. Usually $\varphi(X(\Omega))$ is different from $\Phi(X)$ and $\Phi(X)$ is a better estimate than $\widehat{X}$: *a primary observable is actually the best statistical estimator* for a given size of the sample. But it requires the explicit knowledge of the scalar product and H_J. This can be practically done in some significant cases (see for an example J.C.Dutailly *Estimation of the probability of transitions between phases).*

Knowing the estimate $\widehat{X}$ provided by a statistical method φ, we can implement the previous procedure to the set $\widehat{X}(\Omega)$ and compute the set of solutions : $\widehat{A} = \cap_{m_p \in \Omega} \mathcal{E}_J(m_p)^{-1}\left(\widehat{X}(m)\right)$. It is not empty. Its closed convex hull $\widehat{B}$ in H_J can be considered as the domain of confidence of $\widehat{X}$: they are maps which take the same values as $\widehat{X}$ in Ω and as a consequence give the same value to $\sum_{m\in\Omega}\left\|\widehat{X}(m) - x_p\right\|_E$.

Because $\widehat{B}$ is closed and convex there is a unique orthogonal projection Y of X on $\widehat{B}$ (Maths.1107) and :

$\forall Z \in \widehat{B} : \|X - Z\|_H \geq \|X - Y\|_H \Rightarrow \left\|X - \widehat{X}\right\|_H \geq \|X - Y\|_H$

so Y is a better estimate than $\varphi(X(\Omega))$, and can be computed if we know the scalar product on H.

We see clearly the crucial role played by the choice of a specification. But it leads to a more surprising result, of deep physical meaning.

3.5.1 Quantization of singularities

A classic problem in Physics is to prove the existence of a singular phenomenon, appearing only for some values of the parameters m. To study this problem we use a model similar to the previous one, with the same notations. But here the variable X is comprised of two maps, X_1, X_2 with unknown, disconnected, domains $M_1, M_2 : M = M_1 + M_2$. The first problem is to estimate X_1, X_2.

With a statistical process $\varphi(X(\Omega))$ it is always possible to find estimations $\widehat{X}_1, \widehat{X}_2$ of X_1, X_2. The key point is to distinguish in the set Ω the points which belong to M_1 and M_2. There are $\frac{1}{2}\left(2^N - 2\right) = 2^{N-1} - 1$ distinct partitions of Ω in two subsets $\Omega_1 + \Omega_2$, on each subset the statistical method φ gives the estimates :

$\widehat{Y}_1 = \varphi(X(\Omega_1)), \widehat{Y}_2 = \varphi(X(\Omega_2))$

Denote : $\rho(\Omega_1, \Omega_2)$

$= \sum_{m_p\in\Omega_1} \|X(m_p) - \varphi(X(\Omega_1))(m_p)\| + \sum_{m_p\in\Omega_2} \|X(m_p) - \varphi(X(\Omega_2))(m_p)\|$

A partition (Ω_1, Ω_2) is said to be a better fit than (Ω'_1, Ω'_2) if :

$\rho(\Omega_1, \Omega_2) \leq \rho(\Omega'_1, \Omega'_2)$

Then $\widehat{X}_1 = \varphi(X(\Omega_1)), \widehat{X}_2 = \varphi(X(\Omega_2))$ is the solution for the best partition.

So there is a procedure, which provides always the best solution given the data and φ, but it does not give M_1, M_2 precisely, their estimation depends on the structure of M.

However it is a bit frustrating, if we want to test a law, because the procedure provides always a solution, even if actually there is no such partition of X. And this can happen. If we define the sets as above with the evaluation map : $\mathcal{E}_J(m) : H_J \to E :: \mathcal{E}(m) Y = Y(m)$

$A_k = \cap_{m_p \in \Omega_k} \mathcal{E}(m_p)^{-1} \left(\widehat{X}_k(m_p)\right) \subset H_J$ for $k = 1, 2$. It is not empty because it contains at least $\widehat{X}_k$.

B_k the closed convex hull of A_k in H_J

Then : $\forall Y \in B_k, m \in \Omega_k : Y(m) = \widehat{X}_k(m)$

If $B_1 \cap B_2 \neq \varnothing$ there is at least one map, which can be defined uniquely on M, belongs to H_J and is equivalent to $\widehat{X}_1, \widehat{X}_2$.

This issue is of importance because many experiments aim at proving the existence of a special behavior. We need, in addition, a test of the hypothesis (denoted H_0) : there is a partition (and then the best solution would be $\widehat{X}_1, \widehat{X}_2$) against the hypothesis (denoted H_1) there is no partition : there is a unique map $\widehat{X} \in H_J$ for the domain Ω. The simplest test is to compare $\sum_{m_p \in \Omega} \|X(m_p) - \varphi(\Omega)(m_p)\|$ to $\rho(\Omega_1, \Omega_2)$. If $\varphi(\Omega)$ gives results as good as $\widehat{X}_1, \widehat{X}_2$ we can reject the hypothesis. Notice that it accounts for the properties assumed for the maps in H_J. For instance if H_J is comprised uniquely of continuous maps, then $\varphi(X(\Omega))$ is continuous, and clearly distinct from the maps $\widehat{X}_1, \widehat{X}_2$ continuous only on M_1, M_2.

It is quite obvious that the efficiency of this test decreases with N : the smaller N, the greater the chance to accept H_0. Is there a way to control the validity of an experiment ? The Theory of Tests, a branch of Statistics, studies this kind of problems.

The problem is, given a sample of points $\Omega = (m_p)_{p=1}^N$ and the corresponding values $x = (x_p)_{p=1}^N$, decide if they obey to a simple (X, Hypothesis H_1) or a double (X_1, X_2, Hypothesis H_0) distribution law.

The choice of the points $(m_p)_{p=1}^N$ in a sample is assumed to be random : all the points m of M have the same probability to be in Ω, but the size of M_1, M_2 can be different, so it could give a different chance for a point of M_1 or M_2 to be in the sample. Let us say that :

$\Pr(m \in M_1 | H_0) = 1 - \lambda, \Pr(m \in M_2 | H_0) = \lambda, \Pr(m \in M | H_1) = 1$

(all the probabilities are for a sample of a given size N)

Then the probability for any vector of E to have a given value x depends only on the map X : this is the number of points m of M for which $X(m) = x$. For instance if there are two points m with $X(m) = x$ then x has two times the probability to appear, and if X is more concentrated in an area of E, this area has more probability to appear. Let us denote this value $\rho(x) \in [0, 1]$.

Rigorously (Maths.869), with a measure dx on E, μ on M, $\rho(x)\, dx$ is the pull-back of the measure μ on M. For any ϖ belonging to the Borel algebra σE of E :

$\int_\varpi \rho(x)\, dx = \int_{\mathcal{E}(m)^{-1}(\varpi)} \mu(m) \Leftrightarrow \rho(x)\, dx = X^* \mu$

If H_1 is true, the probability $\Pr(x|H_1) = \rho(x)$ depends only on the value x, that is of the map X.

If H_0 is true the probability depends on the maps and if $m \in M_1$ or $m \in M_2$ $(M = M_1 + M_2)$

$\Pr(x|H_0 \wedge m \in M_1) = \rho_1(x)$

$\Pr(x|H_0 \wedge m \in M_2) = \rho_2(x)$

$\Rightarrow \Pr(x|H_0) = (1-\lambda)\rho_1(x) + \lambda\rho_2(x)$

Moreover we have with some measure dx on E :

$\int_E \rho(x)\,dx = \int_E \rho_1(x)\,dx = \int_E \rho_2(x)\,dx = 1$

The likehood function is the probability of a given batch of data. It depends on the hypothesis :

$$L(x|H_0) = \Pr(x_1, x_2, ...x_N|H_0) = \prod_{p=1}^{N} ((1-\lambda)\rho_1(x_p) + \lambda\rho_2(x_p))$$

$$L(x|H_1) = \Pr(x_1, x_2, ...x_N|H_1) = \prod_{p=1}^{N} \rho(x_p)$$

The Theory of Tests gives us some rules (see Kendall t.II). A critical region is an area $w \subset E^N$ such that H_0 is rejected if $x \in w$.

One considers two risks :

- the risk of type I is to wrongly reject H_0. It has the probability : $\alpha = \Pr(x \in w|H_0)$

- the risk of type II is to wrongly accept H_0. It has the probability : $1-\beta = \Pr(x \in E^N - w|H_0)$ called the power of the test thus :

$\beta = \Pr(x \in w|H_1)$

A simple rule, proved by Neyman and Pearson, says that the best critical region w is defined by :

$w = \left\{x : \frac{L(x|H_0)}{L(x|H_1)} \leq k\right\}$

the scalar k being defined by : $\alpha = \Pr(x \in w|H_0)$. So we are left with a single parameter α, which can be seen as the rigor of the test.

The critical area $w \subset E^N$ is then :

$$w = \left\{x \in E^N : \prod_{p=1}^{N} \frac{((1-\lambda)\rho_1(x_p) + \lambda\rho_2(x_p))}{\rho(x_p)} \leq k\right\}$$

with :

$$\alpha = \int_w \prod_{p=1}^{N} ((1-\lambda)\rho_1(\xi_p) + \lambda\rho_2(\xi_p))\,(d\xi)^N$$

It provides a reliable method to build a test, but requires to know, or to estimate, $\rho, \rho_1, \rho_2, \lambda$.

In most of the cases encountered, actually one looks for an anomaly.

H_1 is unchanged, there is only one map X,defined over M. Then : $\Pr(x|H_1) = \rho(x)$

H_0 becomes :

$M = M_1 + M_2$

$\Pr(m \in M_1|H_0) = 1-\lambda, \Pr(m \in M_2|H_0) = \lambda$

On M_1 the variable is X :
$\Pr(x_p|H_0 \wedge m_p \in M_1) = \rho(x) \Rightarrow \Pr(x_p|H_0) = (1-\lambda)\rho(x)$
On M_2 the variable becomes X_2
$\Pr(x_p|H_0 \wedge m_p \in M_2) = \rho_2(x) \Rightarrow \Pr(x_p|H_0) = \lambda\rho_2(x)$
And w is :

$$w = \left\{ x \in E^N : \prod_{p=1}^{N} \frac{((1-\lambda)\rho(x_p)+\lambda\rho_2(x_p))}{\rho(x_p)} \leq k \right\}$$

$$w = \left\{ x \in E^N : \prod_{p=1}^{N} \left(1-\lambda+\lambda\frac{\rho_2(x_p)}{\rho(x_p)}\right) \leq k \right\}$$

$$\alpha = \int_w \prod_{p=1}^{N} ((1-\lambda)\rho(x_p)+\lambda\rho_2(x_p))(dx)^N$$

$$\beta = \Pr(x \in w|H_1) = \int_w \left(\prod_{p=1}^{N} \rho(x_p)\right)(dx)^N$$

If there is one observed value such that $\rho(x_p) = 0$ then H_0 should be accepted. But, because ρ, ρ_2 are not well known, and the imprecision of the experiments, H_0 would be proven if $\frac{L(x|H_0)}{L(x|H_1)} > k$ for a great number of experiments. So we can say that H_0 is scientifically proven if :

$$\forall (x_1, x_2, ...x_N) : \prod_{p=1}^{N} \left((1-\lambda)+\lambda\frac{\rho_2(x_p)}{\rho(x_p)}\right) > k$$

By taking $x_1 = x_2 = ... = x_N = x$:
$\forall x : (1-\lambda)+\lambda\frac{\rho_2(x)}{\rho(x)} > k^{1/N}$
$\frac{\rho_2(x)}{\rho(x)} > \left(k^{1/N}+\lambda-1\right)/\lambda$
When $N \to \infty : k^{1/N} \to 1 \Rightarrow \frac{\rho_2(x)}{\rho(x)} > 1$

So a necessary condition to have a chance to say that a singularity has been reliably proven is that : $\forall x : \frac{\rho_2(x)}{\rho(x)} > 1$.

The function $\frac{\rho_2(x)}{\rho(x)}$ can be called the Signal to Noise Ratio, by similarity with the Signal Theory. Notice that we have used very few assumptions about the variables. And we can state :

Theorem 16 *In a system represented by variables X which are maps defined on a set M and valued in a vector space E, a necessary condition for a singularity to be detected is that the Signal to Noise Ratio is greater than 1 for all values of the variables in E.*

This result can be seen in another way : if a signal is acknowledged, then necessarily it is such that $\frac{\rho_2(x)}{\rho(x)} > 1$. Any other signal would be interpreted as related to the imprecision of the measure. So there is a threshold under which phenomena are not acknowledged, and their value is necessarily above this threshold. The singular phenomena are quantized. One application is the Planck's law (see JC.Dutailly "Mathematics in Physics").

4 PROBABILITY

One of the main purposes of the model is to know the state X, represented by some vector $\psi \in H$. The model is fully determinist, in that the values of the variables X are not assumed to depend on a specific event : there is no probability law involved in its definition. However the value of X which will be acknowledged at the end of the experiment, when all the data have been collected and analyzed, differs from its actual value. The discrepancy stems from the usual imprecision of any measure, but also more fundamentally from the fact that we estimate a vector in an infinite dimensional vector space from a batch of data, which is necessarily finite. We will focus on this later aspect, that is on the discrepancy between an observable $\Phi(X)$ and X.

In any practical physical experiment the estimation of X requires the choice of an observable. We have seen that the most efficient solution is to choose a primary observable which, furthermore, provides the best statistical estimator. However usually neither the map Φ nor the basis $(e_i)_{i\in I}$ are explicit, even if they do exist. So we can look at the discrepancy $X - \Phi(X)$ from a different point of view : for a given, fixed, value of the state X, what is the uncertainty which stems from the choice of Φ among a large class of observables ? This sums up to assess the risk linked to the choice of a specification for the estimation of X.

4.1 Primary observables

Let us start with primary observables : the observable Φ is some projection on a finite dimensional vector subspace of V.

The bases of the vector space V_0 (such that $O \subset V_0$) have the same cardinality, so we can consider that the set I does not depend on a choice of a basis (actually one can take $I = \mathbb{N}$). The set 2^I is the largest $\sigma-$algebra on I. The set $\left(I, 2^I\right)$ is measurable (Maths.802).

For any fixed $\psi \neq 0 \in H$ the function

$\widehat{\mu}_\psi : 2^I \to \mathbb{R} :: \widehat{\mu}_\psi(J) = \frac{\langle \widehat{Y}_J\psi, \psi\rangle}{\|\psi\|^2} = \frac{\left\|\widehat{Y}_J\psi\right\|^2}{\|\psi\|^2}$

is a probability law on $\left(I, 2^I\right)$: it is positive, countably additive and $\widehat{\mu}_\psi(I) = 1$ (Maths.11.4.1).

If we see the choice of a finite subset $J \in 2^I$ as an event in a probabilist point of view, for a given $\psi \neq 0 \in H$ the quantity $\widehat{Y}_J(\psi)$ is a random variable, with a distribution law $\widehat{\mu}_\psi$

The operator $\widehat{Y}_J$ has two eigen values : 1 with eigen space $\widehat{Y}_J(H)$ and 0 with eigen space $\widehat{Y}_{J^c}(H)$. Whatever the primary observable, the value of $\Phi(X)$ will be $Y_J(X)$ for some J, that is an eigen vector of the operator $\Phi = Y_J$, and the probability to observe $\Phi(X)$, if the system is in the state X, is :

$\Pr(\Phi(X) = Y_J(X)) = \Pr(J|\psi) = \widehat{\mu}_\psi(J) = \frac{\left\|\widehat{Y}_J\psi\right\|^2}{\|\psi\|^2} = \frac{\left\|\widehat{\Phi}(\Upsilon(X))\right\|_H^2}{\|\Upsilon(X)\|_H^2}$

This result still holds if another basis had been chosen : $\Phi(X)$ will be $Y_J(X)$ for some J, expressed in the new basis, but with a set J of same cardinality. And some specification must always be chosen. So we have :

Theorem 17 *For any primary observable Φ, the value $\Phi(X)$ which is measured is an eigen vector of the operator Φ, and the probability to measure a value $\Phi(X)$ if the system is in the state X is :*

$\Pr(\Phi(X)|X) = \frac{\left\|\widehat{\Phi}(\Upsilon(X))\right\|_H^2}{\|\Upsilon(X)\|_H^2}$

4.2 Secondary observables

For a secondary observable, as defined previously :

$\Phi = \sum_{p=1}^{n} \lambda_p Y_{J_p}$

$\widehat{\Phi} = \sum_{p=1}^{n} \lambda_p \widehat{\pi}_{H_p}$

The vectors decompose as :

$X = Y_{J^c}(X) + \sum_{p=1}^{n} X_p$

with $X_p = Y_{J_p}(X) = \sum_{i \in J_p} \left\langle \phi_i, \widehat{Y}_{Jp}(\Upsilon(X)) \right\rangle_H e_i \in V_p$

$\Upsilon(X) = \psi = \psi_c + \sum_{p=1}^{n} \psi_p$ with $\psi_p = \widehat{\pi}_{H_p}(\psi), \psi_c = \widehat{\pi}_{H_c}(\psi)$

where $\psi_p$is an eigen vector of $\widehat{\Phi}$, X_p is an eigen vector of Φ both for the eigen value λ_p

and

$\Phi(X) = \sum_{p=1}^{n} \lambda_p X_p$

$\widehat{\Phi}(\psi) = \sum_{p=1}^{n} \lambda_p \psi_p$

If, as above, we see the choice of a finite subset $J \in 2^I$ as an event in a probabilist point of view then the probability that $\Phi(X) = \lambda_p X_p$ if the system is in the state X, is given by $\Pr(J_p|X) = \frac{\left\|\widehat{Y}_p \psi\right\|^2}{\|\psi\|^2} = \frac{\|\psi_p\|^2}{\|\psi\|^2}$

And we have :

Theorem 18 *For any secondary observable Φ, the value $\Phi(X)$ which is observed if the system is in the state X is a linear combination of eigen vectors X_p of Φ for the eigen value λ_p: $\Phi(X) = \sum_{p=1}^{n} \lambda_p X_p$*

The probability that $\Phi(X) = \lambda_p X_p$ is:

$\Pr(\Phi(X) = \lambda_p X_p|X) = \frac{\|\Upsilon(X_p)\|^2}{\|\Upsilon(X)\|^2}$

Which can also be stated as : $\Phi(X)$ can take the values $\lambda_p X_p$, each with the probability $\frac{\|\psi_p\|^2}{\|\psi\|^2}$, then $\Phi(X)$ reads as an expected value. This is the usual way it is expressed in QM.

The interest of these results comes from the fact that we do not need to explicit any basis, or even the set I. And we do not involve any specific property of the estimator of X, other than Φ is an observable. The operator $\widehat{\Phi}$ sums up the probability law.

Of course this result can be seen in another way : as only $\Phi(X)$ can be accessed, one can say that the system takes only the states $\Phi(\lambda_p X_p)$, with a probability $\frac{\|\psi_p\|^2}{\|\psi\|^2}$. This gives a probabilistic behavior to the system (X becoming

a random variable) which is not present in its definition, but is closer to the usual interpretation of QM.

This result can be illustrated by a simple example. Let us take a single continuous variable x, which takes its values in $\mathbb{R}$. It is clear that any physical measure will at best give a rational number $Y(x) \in \mathbb{Q}$ up to some scale. There are only countably many rational numbers for unaccountably many real scalars. So the probability to get $Y(x) \in \mathbb{Q}$ should be zero. The simple fact of the measure gives the paradox that rational numbers have an incommensurable weight, implying that each of them has some small, but non null, probability to appear. In this case I can be assimilated to $\mathbb{Q}$, the subsets J are any finite collection of rational numbers.

4.3 Wave function

The wave function is a central object in QM, but it has no general definition and is deemed non physical (except in the Bohm's interpretation). Usually this is a complex valued function, defined over the space of configuration of the system : the set of all possible values of the variables representing the system. If it is square integrable, then it belongs to a Hilbert space, and can be assimilated to the vector representing the state. Because its arguments comprise the coordinates of objects such as particles, it has a value at each point, and the square of the module of the function is proportional to the probability that the measure of the variable takes the values of the arguments at this point. Its meaning is relatively clear for systems comprised of particles, but less so for systems which include force fields, because the space of configuration is not defined. But we will see now how it can be precisely defined in our framework.

Theorem 19 *In a system modelled by N variables, collectively denoted X, which are maps : $X : M \to F$ from a common measured set M to a finite dimensional normed vector space F and belonging to an open subset of an infinite dimensional, separable, real Fréchet vector space V, such that the evaluation map : $\mathcal{E}(m) : V \to F :: \mathcal{E}(m)(X) = X(m)$ which assigns at any X its value in a fixed point m in M is measurable : then for any state X of the system there is a function : $W : M \times F \to \mathbb{R}$ such that $W(m, y) = \Pr(\Phi(X)(m) = y|X)$ is the probability that the measure of the value of any primary observable $\Phi(X)$ at m is y.*

Proof. The conditions 1 apply, there is a Hilbert space H and an isometry $\Upsilon : V \to H$.

To the primary observable $\Phi : V \to V_J$ is associated the self-adjoint operator $\widehat{\Phi} = \Upsilon \circ \Phi \circ \Upsilon^{-1}$

We can apply the theorem 17: the probability to measure a value $\Phi(X) = Y$ if the system is in the state X is :

$\Pr(\Phi(X) = Y|X) = \frac{\left\|\widehat{\Phi}(\Upsilon(Y))\right\|_H^2}{\|\Upsilon(X)\|_H^2} = \pi(Y)$

Because only the maps belonging to V_J are observed it provides a probability law π on the set $V_J : \pi : V_\sigma \to [0,1]$ where V_σ is the Borel algebra of V_J.

The evaluation map : $\mathcal{E}_J(m) : V_J \to F :: \mathcal{E}_J(m)(Y) = Y(m)$ assigns at any $Y \in V_J$ its value in the fixed point m in M.

If $y \in F$ is a given vector of F, the set of maps in V_J which gives the value y in m is : $\varpi(m,y) = \mathcal{E}_J(m)^{-1}(y) \subset V_J$.

The probability that the observable takes the value y at m $\Phi(X)(m) = y$ is

$\pi(\varpi(m,y)) = \pi\left(\mathcal{E}_J(m)^{-1}(x)\right)$

$= \frac{1}{\|\Upsilon(X)\|_H^2} \int_{Y \in \varpi(m,y)} \left\|\widehat{\Phi}(\Upsilon(Y))\right\|_H^2 \pi(Y) = W(m,y)$ ■

If M is endowed with a positive measure μ and X is a scalar function, the space V of square integrable maps $\int_\Omega |X(m)|^2 \mu(m) < \infty$ is a separable Hilbert space H, then the conditions 1 are met and H can be identified with the space of the states.

$W(m,y) = \frac{1}{\|X\|_H^2} \int_{Y \in \varpi(m,y)} |Y|_H^2 = \left(\int_\Omega |X|^2 \mu\right)^{-1} \mu\left(Y^{-1}(m,y)\right)$

No structure, other than the existence of the measure μ, is required on M. But of course if the variables X include derivatives M must be at least a differentiable manifold.

W can be identified with the square of the wave function of QM.

5 CHANGE OF VARIABLES

In the conditions 1 we have noticed that, in the model, the variables could be defined over different connected domains. Actually one can go further and consider the change of variables, which leads to a theorem similar to the well known Wigner's theorem. The problem appears in Physics in two different ways, which reflect the interpretations of Scientific laws.

5.1 Two ways to define the same state of a system

5.1.1 The first way : from a theoretical model

In the first way the scientist has built a theoretical model, using known concepts and their usual representation by mathematical objects. A change of variables appears notably when :

i) The variables are the components of a geometric quantity (a vector, a tensor,...) expressed in some basis. According to the general Principle of Relativity, the state of the system shall not depend on the observers (those measuring the coordinates). For instance it should not matter if the state of a system is measured in different units. The data change, but according to rules which depend on the mathematical representation which is used, and not on the system itself. In a change of basis coordinates change but they represent the same vectorial quantity. We will see another example with interacting, indistinguishable systems.

ii) The variables are maps, depending on arguments which are themselves coordinates of some event : $X_k = X_k(\xi_1, ...\xi_{p_k})$. Similarly these coordinates ξ can change according to some rules, while the variable X_k represents the same event. A simple example that we will develop later on is a simple function of the time $X_k(t)$ such that the time t can be expressed in different units, or with different origin : $X_k(t)$ and $X'_k(t) = X_k(t+\theta)$ represent the same state.

By definition in both cases there is a continuous bijective map $U : V \to V'$ such that X and $X' = U(X)$ represent the same state of the system. This is the way mathematicians see a change of variables, and is usually called the passive way by physicists.

Any primary or secondary observable Φ is a linear map $\Phi \in L(V;W)$ into a finite dimensional vector subspace W. For the new variable the observable is $\Phi' \in L(V;W')$. Both $W, W' \subset V$ but W' is not necessarily identical to W. However the assumption that $X' = U(X)$ and X represents the same state of the system implies that for any measure of the state we have a similar relation : $\Phi' \circ U(X) = U \circ \Phi(X) \Leftrightarrow \Phi' \circ U = U \circ \Phi$. This is actually the true meaning of "represent the same state". This means that actually one makes the measures according to a fixed procedure, given by Φ, on variables which vary with U. Because U is a bijection on V : $\Phi' = U \circ \Phi \circ U^{-1}$.

5.1.2 The second way : from experimental measures

In the second way the scientist makes measures with a device that can be adjusted according to different values of a parameter, say θ : the simplest example is using different units, but often it is the orientation of the device which can be changed. And the measures $Y(\theta)$ which are taken are related to the choice of parameter for the device. If the results of experiments show that $Y(\theta) = Q(\theta) Y(\theta_0)$ with a bijective map $Q(\theta)$ and θ_0 some fixed value of the parameter one can assume that this experimental relation is a feature of the system itself.

Physicists distinguish a passive transformation, when only the device changes, and an active transformation, when actually the experiment involves a physical change on the system. In a passive transformation we come back to the first way and it is legitimate to assume that we have actually the same state, represented by different data, reflecting some mathematical change in their expression, even if the observable, which is valued in a finite dimensional space, does not account for all the possible values of the variables. In an active transformation (for instance in the Stern-Gerlach experiment one changes the orientation of a magnetic field to which the particles are submitted) one can say that there is some map U acting on the space V of the states of the system, such that the measure is done by a unique procedure $\widetilde{\Phi}$ on a state X which is changed by a map $U(\theta)$. So that the measures are $Y(\theta) = \widetilde{\Phi} \circ U(\theta) X$ and the relation $Y(\theta) = Q(\theta) Y(\theta_0)$ reads : $\widetilde{\Phi} \circ U(\theta)(X) = U(\theta) \circ \widetilde{\Phi}(X)$. So this is very similar to the first case, where θ represents the choice of a frame.

In both cases there is the general idea that the state of the system is represented by some fixed quantity, which can be measured in different procedures, so that there is a relation, given by the way one goes from one procedure to the others, between the measures of the state. In the first way the conclusion comes from the mathematical definition in a theoretical model : this is a simple mathematical deduction using the Principle of Relativity. In the second way there is an assumption : that one can extend the experimental facts, necessarily limited to a finite number of data, to the whole set of possible values of the variable.

The Theorem 2 is based on the existence of a Fréchet manifold structure on the set of possible values of the maps X. The same manifold structure can be defined by different, compatible, atlas. So the choice of other variables can lead to the same manifold, and the fixed quantity that we identify with a state is just a point on the manifold, and the change of variables is a change of charts between compatible atlas. The variables must be related by transition maps, that is continuous bijections, but additional conditions are required, depending on the manifold structure considered. For instance for differentiable manifolds the transition maps must be differentiable. We will request that the transition maps preserve the positive kernel, which plays a crucial role in Fréchet manifolds.

5.2 Fundamental theorem for a change of variables

We will summarize these features in the following :

Condition 20

i) The same system is represented by the variables $X = (X_1, ...X_N)$ *and* $X' = (X'_1, ...X'_{N'})$ *which belong to open subsets* O, O' *of the infinite dimensional, separable, Fréchet vector space* V.

ii) There is a continuous map $U : V \to V$, *bijective on* (O, O'), *such that* X *and* $X' = U(X)$ *represent the same state of the system*

iii) U *preserves the positive kernel on* V[1]

iv) For any observable Φ *of* X, *and* Φ' *of* X' : $\Phi' \circ U = U \circ \Phi$

The map U shall be considered as part of the model, as it is directly related to the definition of the variables, and is assumed to be known. There is no hypothesis that it is linear.

Theorem 21 *Whenever a change of variables on a system meets the conditions 20 above,*

i) there is a unitary, linear, bijective map $\widehat{U} \in \mathcal{L}(H; H)$ *such that :* $\forall X \in O : \widehat{U}(\Upsilon(X)) = \Upsilon(U(X))$ *where* H *is the Hilbert space and* Υ *is the linear map :* $\Upsilon : V \to H$ *associated to* X, X'

ii) U *is necessarily a bijective linear map.*

For any observables Φ, Φ'*:*

iii) $W' = \Phi'(V)$ *is a finite dimensional vector subspace of* V, *isomorphic to* $W = \Phi(V) : W' = U(W)$

iv) the associated operators $\widehat{\Phi} = \Upsilon \circ \Phi \circ \Upsilon^{-1}, \widehat{\Phi}' = \Upsilon \circ \Phi' \circ \Upsilon^{-1}$ *are such that :* $\widehat{\Phi}' = \widehat{U} \circ \widehat{\Phi} \circ \widehat{U}^{-1}$ *and* $H'_{\Phi'} = \widehat{\Phi}'(H)$ *is a vector subspace of* H *isomorphic to* $H_\Phi = \widehat{\Phi}(H)$

Proof. i) Let $V_0 = O \cup O'$. This is an open set and we can apply the theorem 2. There is a homeomorphism $\Xi : V_0 \to H_0$ where H_0 is an open subset of a Hilbert space H. For a basis $(e_i)_{i \in I}$ of Span(V_0) there is an isometry Υ such that :

$\Upsilon : V_0 \to H_0 :: \Upsilon(Y) = \sum_{i \in I} \langle \phi_i, \Upsilon(Y) \rangle_H \varepsilon_i$

such that :

$\forall i \in I : \varepsilon_i = \Upsilon(e_i)$;

$\forall i, j \in I : \langle \phi_i, \varepsilon_j \rangle_H = \delta_{ij}$;

ii) Υdefines a positive kernel on $V_0 : K_V(Y_1, Y_2) = \langle \Upsilon Y_1, \Upsilon Y_2 \rangle_H$

The sets (V_0, Υ, H) and $(V_0, \Upsilon U, H)$ are two realizations triple of K_V. Then there is an isometry φ such that :

$\Upsilon U = \varphi \circ \Upsilon$ (Maths.1200).

[1] The positive kernel plays a role similar to the probability of transition between states of the Wigner's Theorem.

$\langle UX_1, UX_2\rangle_V = \langle \Upsilon UX_1, \Upsilon UX_2\rangle_H = \langle \varphi\circ\Upsilon X_1, \varphi\circ\Upsilon X_2\rangle_H$
$= \langle \Upsilon X_1, \Upsilon X_2\rangle_H = \langle X_1, X_2\rangle$
So U preserves the scalar product on V
Let be : $\widehat{U} = \Upsilon\circ U\circ\Upsilon^{-1}$
$\left\langle \widehat{U}\psi_1, \widehat{U}\psi_2\right\rangle_H = \left\langle \Upsilon\circ U\circ\left(\Upsilon^{-1}\psi_1\right), \Upsilon\circ U\circ\left(\Upsilon^{-1}\psi_2\right)\right\rangle_H$
$= \left\langle U\circ\left(\Upsilon^{-1}\psi_1\right), U\circ\left(\Upsilon^{-1}\psi_2\right)\right\rangle_V = \left\langle\left(\Upsilon^{-1}\psi_1\right), \left(\Upsilon^{-1}\psi_2\right)\right\rangle_V$
$= \langle\psi_1, \psi_2\rangle_H$
So $\widehat{U}$ preserves the scalar product on H

iii) As seen in Theorem 2 starting from the basis $(\varepsilon_i)_{i\in I}$ of H one can define a Hermitian basis $(\widetilde{\varepsilon}_i)_{i\in I}$ of H, an orthonormal basis $(\widetilde{e}_i)_{i\in I}$ of V for the scalar product $K_V = \langle\rangle_V$ with $\widetilde{e}_i = \Upsilon^{-1}(\widetilde{\varepsilon}_i)$

U is defined for any vector of V, so for $(\widetilde{e}_i)_{i\in I}$ of V.

Define : $\widehat{U}(\widetilde{\varepsilon}_i) = \widehat{U}(\Upsilon(\widetilde{e}_i)) = \Upsilon(U(\widetilde{e}_i)) = \widetilde{\varepsilon}'_i$

The set of vectors $(\widetilde{\varepsilon}'_i)_{i\in I}$ is an orthonormal basis of H:

$\langle\widetilde{\varepsilon}'_i, \widetilde{\varepsilon}'_j\rangle_H = \left\langle\widehat{U}(\Upsilon(\widetilde{e}_i)), \widehat{U}(\Upsilon(\widetilde{e}_j))\right\rangle_H = \langle\widetilde{e}_i, \widetilde{e}_j\rangle_V = \delta_{ij}$

The map : $\chi : \ell^2(I)\to H :: \chi(y) = \sum_{i\in I} y_i\widetilde{\varepsilon}'_i$ is an isomorphism (same as in Theorem 2) and $(\widetilde{\varepsilon}'_i)_{i\in I}$ is a Hilbertian basis of H. So we can write :

$\forall\psi\in H : \psi = \sum_{i\in I}\psi^i\widetilde{\varepsilon}_i, \widehat{U}(\psi) = \sum_{i\in I}\psi'^i\widetilde{\varepsilon}'_i$

and : $\psi^i = \langle\widetilde{\varepsilon}_i, \psi\rangle_H = \left\langle\widehat{U}(\widetilde{\varepsilon}_i), \widehat{U}(\psi)\right\rangle_H = \left\langle\widetilde{\varepsilon}'_i, \sum_{j\in I}\psi'^j\widetilde{\varepsilon}'_j\right\rangle_H = \psi'^i$

Thus the map $\widehat{U}$ reads : $\widehat{U} : H\to H :: \widehat{U}\left(\sum_{i\in I}\psi^i\widetilde{\varepsilon}_i\right) = \sum_{i\in I}\psi^i\widetilde{\varepsilon}'_i$

It is linear, continuous and unitary : $\left\langle\widehat{U}(\psi_1), \widehat{U}(\psi_2)\right\rangle = \langle\psi_1, \psi_2\rangle$ and $\widehat{U}$ is invertible

$U = \Upsilon^{-1}\circ\widehat{U}\circ\Upsilon$ is linear and bijective

iv) For any primary or secondary observable Φ there is a self-adjoint, Hilbert-Schmidt and trace class operator $\widehat{\Phi}$ on the associated Hilbert space H such that : $\widehat{\Phi} = \Upsilon\circ\Phi\circ\Upsilon^{-1}$. For the new variable the observable is $\Phi'\in L(V; W')$ and $W\prime\subset V$ is not necessarily identical to W. It is associated to the operator : $\widehat{\Phi}' = \Upsilon\circ\Phi'\circ\Upsilon^{-1}$. W and W' are finite dimensional vector subspaces of V.

$$\begin{array}{ccccccccccccc}
 & & \Phi & & & & U & & & & \Phi' & & \\
W & \leftarrow & \leftarrow & \leftarrow & V & \rightarrow & \rightarrow & \rightarrow & V & \rightarrow & \rightarrow & \rightarrow & W' \\
\downarrow & & & & \downarrow & & & & \downarrow & & & & \downarrow \\
\downarrow & \Upsilon & & & \downarrow & \Upsilon & & \Upsilon & \downarrow & & & \Upsilon & \downarrow \\
\downarrow & & \widehat{\Phi} & & \downarrow & & \widehat{U} & & \downarrow & & \widehat{\Phi'} & & \downarrow \\
H_\Phi & \leftarrow & \leftarrow & \leftarrow & H & \rightarrow & \rightarrow & \rightarrow & H & \rightarrow & \rightarrow & \rightarrow & H_{\Phi'}
\end{array}$$

Because U is a bijection on V : $\Phi'\circ U = U\circ\Phi \Rightarrow \Phi' = U\circ\Phi\circ U^{-1}$ and V is globally invariant by U

$\Phi'(V) = W' = U\circ\Phi\circ U^{-1}(V) = U\circ\Phi(V) = U(W)$

thus W' is a vector subspace of V isomorphic to W

$\widehat{\Phi}' = \Upsilon\circ\Phi'\circ\Upsilon^{-1} = \Upsilon\circ U\circ\Phi\circ U^{-1}\circ\Upsilon^{-1} = \widehat{U}\circ\Upsilon\circ\Phi\circ\Upsilon^{-1}\circ\widehat{U}^{-1} = \widehat{U}\circ\widehat{\Phi}\circ\widehat{U}^{-1}$

Let us denote : $\widehat{\Phi}(H) = H_{\Phi}, \widehat{\Phi}'(H) = H_{\Phi'}$
$\widehat{U}(H) = H$ because it is a unitary map
$\widehat{\Phi}'(H) = \widehat{U} \circ \widehat{\Phi} \circ \widehat{U}^{-1}(H) = \widehat{U} \circ \widehat{\Phi}(H) = \widehat{U}(H_{\Phi}) = H_{\Phi'}$
thus $H_{\Phi'}$ is a vector subspace of H isomorphic to H_{Φ} ■

As a consequence the map U is necessarily linear, even if this was not assumed in the conditions 20 : variables which are not linearly related (in the conditions 20) cannot represent the same state.

As $\widehat{U}$ is unitary, it cannot be self adjoint or trace class (except if $U = Id$). So it differs from an observable.

5.2.1 Change of units

A special case of this theorem is the choice of units to measure the variables. A change of units is a map : $X'_k = \alpha_k X_k$ with fixed scalars $(\alpha_k)_{k=1}^N$. As we must have :

$\langle U(X_1), U(X_2)\rangle_V = \langle X_1, X_2\rangle_V = \sum_{k=1}^N \alpha_k^2 \langle X_1, X_2\rangle_V = \langle X_1, X_2\rangle_V \Rightarrow \sum_{k=1}^N \alpha_k^2 = 1$

which implies for any single variable $X_k : \alpha_k = 1$. So the variables in the model should be dimensionless quantities. This is in agreement with the elementary rule that any formal theory should not depend on the units which are used.

More generally whenever one has a law which relates quantities which are not expressed in the same units, there should be some fundamental constant involved, to absorb the discrepancy between the units. For instance some Physicals laws involve an exponential, such as the wave equation for a plane wave :

$\psi = \exp i \left(\left\langle \overrightarrow{k}, \overrightarrow{r} \right\rangle - \varpi t \right)$

They require that the argument in the exponential is dimensionless, and because $\overrightarrow{r}$ is a length and t a time we should have a fundamental constant with the dimension of a speed (in this case c).

But also it implies that there should be some "universal system of units" (based on a single quantity) in which all quantities of the theory can be measured. In Physics this is the Planck's system which relate the units of different quantities through the values of the fundamental constants c, G (gravity), R (Boltzmann constant), $\hbar$, and the charge of the electron (see Wikipedia for more).

Usually the variables are defined with respect to some frame, then the rules for a change of frame have a special importance and are a defining feature of the model. When the rules involve a group, the previous theorem can help to precise the nature of the abstract Hilbert space H and from there the choice of the maps X.

5.3 Group representation

5.3.1 Summary of representation of groups

The theory of group representation is a key tool in Physics. We will remind some basic results here, see Maths.23 for a comprehensive study of this topic.

The left action of a group G on a set E is a map :$\lambda : G \times E \to E :: \lambda(g, x)$ such that $\lambda(gg', x) = \lambda(g, \lambda(g', x)), \lambda(1, x) = x$. And similarly for a right action $\rho(x, g)$.

The representation of a group G is a couple (E, f) of a vector space E and a continuous map $f : G \to G\mathcal{L}(E; E)$ (the set of linear invertible maps from E to E) such that :

$\forall g, g' \in G : f(g \cdot g') = f(g) \circ f(g'); f(1) = Id \Rightarrow f\left(g^{-1}\right) = f(g)^{-1}$

A representation is **faithful** if f is bijective.

A vector subspace F is **invariant** if $\forall u \in F, g \in G : f(g) u \in F$

A representation is **irreducible** if there is no other invariant subspace than $E, 0$.

A representation is not unique : from a given representation one can build many others. The sum of the representations $(E_1, f_1), (E_2, f_2)$ is $(E_1 \oplus E_2, f_1 + f_2)$.

A representation is **unitary** if there is a scalar product on E and $f(g)$ is unitary : $\forall u, v \in F, g \in G : \langle f(g) u, f(g) v\rangle = \langle u, v\rangle$

If two groups G,G' are isomorphic by ϕ, then a representation (E, f) of G provides a representation of G':

$\phi : G' \to G :: \forall g, g' \in G' : \phi(g \cdot g') = \phi(g) \cdot \phi(g'); \phi(1_{G'}) = 1_G \Rightarrow \phi\left(g^{-1}\right) = \phi(g)^{-1}$

$f : G \to G\mathcal{L}(E; E)$

Define $f' : G' \to G\mathcal{L}(E; E) :: f'(g') = f(\phi(g'))$

$f'(g_1' \cdot g_2') = f(\phi(g_1' \cdot g_2')) = f(\phi(g_2')) \circ f(\phi(g_1')) = f'(g_1') \circ f'(g_2')$

A **Lie group** is a group endowed with the structure of a manifold. On the tangent space T_1G at its unity (that we will denote 1) there is an algebraic structure of **Lie algebra**, that we will also denote generally T_1G, endowed with a bracket [] which is a bilinear antisymmetric map on T_1G.

If G is a Lie group with Lie algebra T_1G and (E, f) a representation of G, then $(E, f'(1))$ is a representation of the Lie algebra T_1G :

$f'(1) \in \mathcal{L}(T_1G; \mathcal{L}(E; E))$

$\forall X, Y \in T_1G : f'(1)([X, Y]) = f'(1)(X) \circ f'(1)(Y) - f'(1)(Y) \circ f'(1)(X)$

The converse, from the Lie algebra to the group, holds if G is simply connected, otherwise a representation of the Lie algebra provides usually multiple valued representations of the group (we will see important examples later).

Any Lie group G has the **adjoint representation** (T_1G, Ad) over its Lie algebra.

Any irreducible representation of a commutative (abelian) group is unidimensional.

Any unitary representation of a compact or finite group is reducible in the sum of orthogonal, finite dimensional, irreducible unitary representations.

Any representation of a group on a finite dimensional vector space becomes a representation on a set of matrices by choosing a basis. The representations of the common groups of matrices are tabulated. In the standard representation $(K^n, \imath)$ of a group G of $n \times n$ matrices on a field K the map $\imath$ is the usual action of matrices on column vectors in the space K^n. If G is a Lie group then the standard representation of its Lie algebra is the representation $(K^n, \imath)$ by matrices, deduced by derivation.

Two representations $(E, f), (F, \rho)$ of the same group G are **equivalent** if there is an isomorphism : $\phi : E \to F$ such that :

$\forall g \in G : f(g) = \phi^{-1} \circ \rho(g) \circ \phi$

Then from a basis $(e_i)_{i \in I}$ of E one deduces a basis $|e_i >$ of F by : $|e_i >= \phi(e_i)$. Because ϕ is an isomorphism $|e_i >$ is a basis of F. Moreover the matrix of the action of G is in this basis the same as for (E, f) :

$\rho(g)|e_i >= \sum_{j \in J} [\rho(g)]_j^i |e_j >= \rho(g)\phi(e_i) = \phi \circ f(g)(e_i)$

$= \phi\left(\sum_{j \in I} [f(g)]_i^j e_j\right) = \sum_{p \in I} [f(g)]_i^j \phi(e_j) = \sum_{p \in I} [f(g)]_i^j |e_j >$

$[\rho(g)] = [f(g)]$

If K is a subgroup of G, and (E, f) a representation of G, then (E, f) is a subrepresentation of K.

The vector subspaces F of E which are invariant by K provide representations (F, f) of K.

5.3.2 Change of variable parametrized by a group

This is the usual case in Physics. The second point of view that we have noticed above is clear when U is defined by a group. The system is represented by fixed variables, and the measures are taken according to procedures which change with g and we have :

$\Phi(g)(X) = U(g) \circ \Phi(1)(X)$

$\Phi \in L(V; W)$ and $U(g)$ is a bijection so X and $\Phi(1)(X)$ are in bijective correspondence and X must belong to $W \subset V$: we reduce the definition of the states at what can be observed. And to assume that this is true for any observable leads to redefine X as in the first way, but this requires and additional assumption.

Theorem 22 *If the conditions 20 are met, and (V, U) is a representation of the group G, then:*

i) $\left(H, \widehat{U}\right)$ is a unitary representation of the group G with $\widehat{U}(g) = \Upsilon \circ U(g) \circ \Upsilon^{-1}$

ii) For any observable $\Phi \in L(V; W)$ the vector space $W \subset V$ is invariant by U and (W, U) is a representation of G, and for the associated operator $\widehat{\Phi} = \widehat{U}(g) \circ \widehat{\Phi} \circ \widehat{U}(g)^{-1} \in L(H; H_\Phi)$, $\left(H_\Phi, \widehat{U}\right)$ is a finite dimensional unitary representation of the group G.

If G is a Lie group, and U continuous, then :

iii) U is smooth, $\widehat{U}$ is differentiable and $\left(\widehat{U}'(1), H\right)$ is an anti-symmetric representation of the Lie algebra T_1G of G

iv) For any observable $\Phi \in L(V;W)$ $\left(H_\Phi, \widehat{U}'(1)\right)$ is an anti-symmetric representation of the Lie algebra T_1G of G

If (F,f) is a unitary representation of G, equivalent to $\left(H_\Phi, \widehat{U}\right)$, and Φ a primary or secondary observable, then :

v) The results of measures of Φ for two values $1, g$ and the same state of the system are related by :

$\Phi \circ U(1)(X) = \sum_{j\in J} X^j(1) e_j, \Phi \circ U(g)(X) = \sum_{j\in J} X^j(g) e_j$ for some basis $(e_i)_{i\in I}$ of V

$X^j(g) = \sum_{k\in J} [f(g)]_k^j X^k(1)$ where $[f(g)]$ is the matrix of $f(g)$ in orthonormal bases of F

vi) If moreover G is a Lie group and U, f continuous, then the action $U'(1)(\kappa_a)$ of $U'(1)$ for vectors κ_a of T_1G are expressed by the same matrices $[K_a]$ of the action $f'(1)(\kappa_a)$:

$f'(1)(\kappa_a)(f_j) = \sum_{k\in J} [K_a]_j^k f_k \to U'(1)(\kappa_a)(e_j) = \sum_{k\in J} [K_a]_j^k e_k$

and similarly for the observable $\Phi : \Phi \circ U'(1)(\kappa_a)(e_j) = \sum_{k\in J} [K_a]_j^k e_k$

Proof. i) The map : $U : G \to G\mathcal{L}(V;V)$ is such that : $U(g \cdot g') = U(g) \circ U(g'); U(1) = Id$ where G is a group and 1 is the unit in G.

Then $U(g)$ is necessarily invertible, because $U\left(g^{-1}\right) = U(g)^{-1}$

$\widehat{U} : G \to \mathcal{L}(H;H) :: \widehat{U} = \Upsilon \circ U \circ \Upsilon^{-1}$ is such that :

$\widehat{U}(g \cdot g') = \Upsilon \circ U(g \cdot g') \circ \Upsilon^{-1} = \Upsilon \circ U(g) \circ U(g') \circ \Upsilon^{-1} = \Upsilon \circ U(g) \circ \Upsilon^{-1} \circ \Upsilon \circ U(g') \circ \Upsilon^{-1} = \widehat{U}(g) \circ \widehat{U}(g')$

$\widehat{U}(1) = \Upsilon \circ U(1) \circ \Upsilon^{-1} = Id$

So $\left(H, \widehat{U}\right)$ is a unitary representation of the group G ($\widehat{U}(g)$ is bijective, thus invertible).

ii) For any observable : $\Phi \circ U(g) = U(g) \circ \Phi, \widehat{\Phi} = \widehat{U}(g) \circ \widehat{\Phi} \circ \widehat{U}(g)^{-1}$

Let us take $Y \in W = \Phi(V) : \exists X \in V : Y = \Phi(X)$

$U(g)Y = U(g)(\Phi(X)) = \Phi(U(g)X) \in \Phi(V)$

And similarly

$\widehat{Y} \in \widehat{\Phi}(H) : \exists \psi \in H : \widehat{Y} = \widehat{\Phi}(\psi)$

$\widehat{U}(g)\widehat{Y} = \widehat{U}(g)\left(\widehat{\Phi}(\psi)\right) = \widehat{\Phi}\left(\widehat{U}(g)\psi\right) \in \widehat{\Phi}(H)$

thus $W, H_\Phi = \widehat{\Phi}(H)$ are invariant by U,$\widehat{U}$

The scalar product on H holds on the finite dimensional subspace $\widehat{\Phi}(H)$, which is a Hilbert space.

iii) If G is a Lie group and the map $U : G \to \mathcal{L}(V;V)$ continuous, then it is smooth (Maths.1789), $\widehat{U}$ is differentiable and $\left(\widehat{U}'(1), H\right)$ is an anti-symmetric representation of the Lie algebra T_1G of G :

$\forall \kappa \in T_1G : \left(\widehat{U}'(1)\kappa\right)^* = -\left(\widehat{U}'(1)\kappa\right)$

$\widehat{U}(\exp\kappa) = \exp\widehat{U}'(1)\kappa$ where the first exponential is taken on T_1G and the second on $\mathcal{L}$(H;H) (Maths.1886).

iv) Φ is a primary or secondary observable, and so is $\Phi \circ U(g)$, then $\widehat{\Phi} \circ \widehat{U}(g) = \widehat{U}(g) \circ \widehat{\Phi}$ is a self-adjoint, compact operator, and by the Riesz theorem (Math.1142) its spectrum is either finite or is a countable sequence converging to 0 (which may or not be an eigen value) and, except possibly for 0, is identical to the set $(\lambda_p(g))_{p\in\mathbb{N}}$ of its eigen values (Maths.1020). For each distinct eigen value the eigen spaces $H_p(g)$ are orthogonal and H is the direct sum $H = \oplus_{p\in\mathbb{N}} H_p(g)$. For each non null eigen value $\lambda_p(g)$ the eigen space $H_p(g)$ is finite dimensional. For a primary observable the eigen values are either 1 or 0.

Because H_Φ is finite dimensional, for each value of g there is an orthonormal basis $(\widetilde{\varepsilon}_i(g))_{i\in J}$ of H_Φ comprised of a finite number of vectors which are eigen vectors of $\widehat{\Phi} \circ \widehat{U}(g) : \widehat{\Phi} \circ \widehat{U}(g)(\widetilde{\varepsilon}_j(g)) = \lambda_j(g)\widetilde{\varepsilon}_j(g)$

Any vector of H_Φ reads :

$\psi = \sum_{j\in J} \psi^j(g)\widetilde{\varepsilon}_j(g)$ and

$\widehat{\Phi} \circ \widehat{U}(g) = \sum_{p\in\mathbb{N}} \lambda_p(g)\widehat{\pi}_{H_p(g)}$with the orthogonal projection $\widehat{\pi}_{H_p(g)}$ on $H_p(g)$.

And, because any measure belongs to H_Φ it is a linear combination of eigen vectors

$\Phi \circ U(g)(X) = \Upsilon^{-1} \circ \widehat{\Phi} \circ \widehat{U}(g) \circ \Upsilon(X) = \Upsilon^{-1}\left(\sum_{j\in J} \lambda_j(g)\psi^j(g)\widetilde{\varepsilon}_j(g)\right)$

$= \sum_{j\in J} \lambda_j(g)\psi^j \Upsilon^{-1}(\widetilde{\varepsilon}_j(g)) = \sum_{j\in J} \lambda_j(g)\psi^j e_j(g)$

for some basis $(e_i)_{i\in I}$ of V : $e_j(g) = \Upsilon^{-1}(\widetilde{\varepsilon}_j(g))$ and $\Phi \circ U(g)(e_j(g)) = \lambda_j e_j(g)$

That we can write :

$\Phi \circ U(g)(X) = \sum_{j\in J} \lambda_j \psi^j(g) e_j(g) = \sum_{j\in J} X^j(g) e_j(g) = U(g) \circ \Phi(X)$

$\Phi(X) = U(g^{-1})\left(\sum_{j\in J} X^j(g) e_j(g)\right)$

v) If the representations $\left(H_\Phi, \widehat{U}\right), (F, f)$ are equivalent (which happens if they have the same finite dimension) there is an isomorphism $\phi : H_\Phi \to F$ which can be defined by taking an orthonormal basis $(\widetilde{\varepsilon}_i(g_0))_{i\in J}, (f_j(g_0))_{j\in J}$ in each vector space, for some fixed $g_0 \in G$ that we can take $g_0 = 1 : \phi\left(\sum_{i\in J} \psi^j \widetilde{\varepsilon}_j(1)\right) = \sum_{i\in J} \psi^j f_j(1) \Leftrightarrow \phi(\widetilde{\varepsilon}_j(1)) = f_j(1)$

To a change of g corresponds a change of orthonormal basis, both in H_Φ and F, given by the known unitary map $f(g)$: $f_j(g) = f(g)(f_j(1)) = \sum_{k\in J} [f(g)]_j^k f_k(1)$ and thus we have the same matrix for $\widehat{U}(g)$:

$\widetilde{\varepsilon}_j(g) = \widehat{U}(g)(\widetilde{\varepsilon}_j(1)) = \phi^{-1} \circ f(g) \circ \phi(\widetilde{\varepsilon}_j(1)) = \phi^{-1} \circ f(g)(f_j(1)) = \sum_{k\in J} [f(g)]_j^k \widetilde{\varepsilon}_k(1)$

$$
\begin{array}{ccccccccccccc}
 & & U(g) & & & & \Phi & & & & & & \\
V & \to & \to & \to & V & \to & \to & \to & W & & & & \\
\downarrow & & & & \downarrow & & & & \downarrow & & & & \\
\downarrow & \Upsilon & & \Upsilon & \downarrow & & & \Upsilon & \downarrow & & & & \\
\downarrow & & \widehat{U}(g) & & \downarrow & & \widehat{\Phi} & & \downarrow & & \widehat{U}(g) & & \\
H & \to & \to & \to & H & \to & \to & \to & H_{\Phi} & \to & \to & \to & H_{\Phi} \\
 & & & & & & & & \downarrow & & & & \downarrow \\
 & & & & & & & \phi & \downarrow & & & \phi & \downarrow \\
 & & & & & & & & \downarrow & & f(g) & & \downarrow \\
 & & & & & & & & F & \to & \to & \to & F
\end{array}
$$

$\widetilde{\varepsilon}_j(g) = \widehat{U}(g)(\widetilde{\varepsilon}_j(1)) = \sum_{k\in J}[f(g)]_j^k \widetilde{\varepsilon}_k(1)$

$e_j(g) = \Upsilon^{-1}(\widetilde{\varepsilon}_j(g)) = \Upsilon^{-1}\left(\sum_{k\in J}[f(g)]_j^k \widetilde{\varepsilon}_k(1)\right)$

$= \sum_{k\in J}[f(g)]_j^k \Upsilon^{-1}(\widetilde{\varepsilon}_k(1)) = \sum_{k\in J}[f(g)]_j^k e_k(1)$

$e_j(g) = \Upsilon^{-1}\circ\widehat{U}(g)\circ\Upsilon(e_j(1)) = U(g)(e_j(1))$

Thus the matrix of $U(g)$ to go from 1 to g is $[f(g)]$

$\Phi(X) = U\left(g^{-1}\right)\left(\sum_{j\in J}X^j(g)\,e_j(g)\right)$

$\Phi\circ U(g)(X) = \sum_{j\in J}X^j(g)\,e_j(g) = \sum_{j\in J}X^j(g)\sum_{k\in J}\left[f\left(g^{-1}\right)\right]_j^k e_k(1)$

$\Phi\circ U(g_0)(X) = \sum_{k\in J}X^k(1)\,e_k(1) \Rightarrow \sum_{j\in J}X^j(g)\left[f\left(g^{-1}\right)\right]_j^k = X^k(1)$

$X^j(g) = \sum_{k\in J}[f(g)]_j^k X^j(1)$

The measures $\Phi\circ U(g)(X)$ transform with the known matrix $f(g)$.

vi) $\left(H_{\Phi},\widehat{U}'(1)\right),(F,f'(1))$ are equivalent, anti-symmetric (or anti-hermitian for complex vector spaces) representations of the Lie algebra T_1G. If $(\kappa_a)_{a=1}^m$ is a basis of T_1G then $f'(1)$, which is a linear map, is defined by the values of $f'(1)(\kappa_a)\in L(F;F)$.

$$
\begin{array}{ccccc}
 & & \widehat{U}'(1)(\kappa) & & \\
 & H_{\Phi} & \to \quad \to & \to & H_{\Phi} \\
 & \downarrow & & & \downarrow \\
\phi & \downarrow & & \phi & \downarrow \\
 & \downarrow & f'(1)(\kappa) & & \downarrow \\
 & F & \to \quad \to & \to & F
\end{array}
$$

$\widehat{U}'(1)(\kappa)(\psi) = \phi^{-1}\circ f'(1)(\kappa)\circ\phi(\psi)$

If we know the values of the action of $f'(1)(\kappa_a)$ on any orthonormal basis $(f_j)_{j\in J}$ of F :

$f'(1)(\kappa_a)(f_j) = \sum_{k\in J}[K_a]_j^k f_k$

we have the value of $\widehat{U}'(1)(\kappa_a)$ for the corresponding orthonormal basis $(\widehat{\varepsilon}_j)_{j\in J}$ of H_{Φ}

$\widehat{U}'(1)(\kappa_a)(\widehat{\varepsilon}_j) = \widehat{U}'(1)(\kappa_a)\,\phi^{-1}(f_j) = \phi^{-1}\circ f'(1)(\kappa_a)(f_j)$

$= \phi^{-1}\left(\sum_{k\in J}[K_a]_j^k f_k\right) = \sum_{k\in J}[K_a]_j^k \widehat{\varepsilon}_k$

So $\widehat{U}'(1)$ is represented in an orthonormal basis of H_{Φ} by the same matrices $[K_a]$

And similarly :

$\widehat{U}(g) = \Upsilon \circ U(g) \circ \Upsilon^{-1} \Rightarrow \widehat{U}'(1)(\kappa) = \Upsilon \circ U'(1)(\kappa) \circ \Upsilon^{-1}$

$U'(1)(\kappa_a)(e_j) = \Upsilon \circ U'(1)(\kappa_a) \circ \Upsilon^{-1}(e_j) = \Upsilon \circ U'(1)(\kappa_a)(\widehat{\varepsilon}_j) = \Upsilon\left(\sum_{k\in J}[K_a]_j^k \widehat{\varepsilon}_k\right) = \sum_{k\in J}[K_a]_j^k e_k$

vii) Because $\Phi \circ U(g) = U(g) \circ \Phi \Rightarrow \Phi \circ U'(1)(\kappa_a) = U'(1)(\kappa_a) \circ \Phi$:

$\Phi \circ U'(1)(\kappa_a)(e_j) = \sum_{k\in J}[K_a]_j^k \Phi(e_k)$ ■

This result is specially important in Physics. Any unitary representation of a compact or finite group is reducible in the sum of orthogonal, finite dimensional, irreducible unitary representations. As a consequence the space V of the variables X has the same structure. If, as it can be assumed, the state of the system stays in the same irreducible representation, it can belong only to some specific finite dimensional spaces, defined through the representation or an equivalent representation of G. X depends only on a finite number of parameters, This is the starting point of quantization.

Notice that the nature of the space E does not matter, only the matrices $[f(g)], [K]$.

Usually in Physics the changes are not parametrized by the group, but by a vector of the Lie algebra (for instance rotations are not parametrized by a matrix but by a vector representing the rotation), which gives a special interest to the two last results.

The usual geometric representations, based on frames defined through a point and a set of vectors, such as in Galilean Geometry and Special Relativity, have been generalized by the formalism of fiber bundles, which encompasses also General Relativity, and is the foundation of gauge theories. Gauge theories use abundantly group transformations, so they are a domain of choice to implement the previous results.

5.3.3 Fourier transform

If G is an abelian group we have more. Irreducible representations of abelian groups are unidimensional, and any unitary representation of an abelian group is the sum of projections on unidimensional vector subspaces which, for infinite dimensional representations, takes the form of spectral integrals. More precisely, there is a bijective correspondence between the unitary representation of an abelian group G and the spectral measures on the Pontryagin dual $\widehat{G}$, which is the space of continuous maps : $\vartheta : G \to T$ where T is the set of complex numbers of module 1 (Maths.1932). This can be made less abstract if G is a topological, locally compact group. Then it has a Haar measure μ and the representation $\left(H, \widehat{U}\right)$ is equivalent to $\left(L^2(G, \mu, \mathbb{C}), \mathcal{F}\right)$ that is to the Fourier transform $\mathcal{F}$ on complex valued, square integrable, functions on G (Maths.2421).

If $\varphi \in L^2(G, \mu, \mathbb{C}) \cap L^1(G, \mu, \mathbb{C})$:

$\mathcal{F}(\varphi)(\vartheta) = \int_G \varphi(g)\overline{\vartheta(g)}\mu(g)$

$\mathcal{F}^*(h)(g) = \int_{\widehat{G}} h(\vartheta)\vartheta(g)\nu(\vartheta)$ for a unique Haar measure ν on $\widehat{G}$ and $\mathcal{F}^* = \mathcal{F}^{-1}$

If G is a compact group then we have Fourier series on a space of periodic functions, and if G is a non compact, finite dimensional Lie group, G is isomorphic to some vector space E and we have the usual Fourier transform on functions on E.

These cases are important from a practical point of view as it is possible to replace the abstract Hilbert space H by more familiar spaces of functions, and usually one can assume that the space V is itself some Hilbert space. The previous tools (observables,...) are then directly available.

The most usual application is about periodic phenomena : whenever a system is inclosed in some box, it can be usually assumed that they are periodic (and null out of the box). Then the representation is naturally through Fourier series and we have convenient Hilbert bases.

5.3.4 One parameter groups

An important case, related to the previous one, is when the variables X depend on a scalar real argument, and the model is such that $X(t), X'(t') = X(t+\theta)$, with any fixed θ, represent the same state. The associated operator is parametrized by a scalar and we have a map :

$\widehat{U} : \mathbb{R}_+ \to G\mathcal{L}(H,H)$ such that :

$\widehat{U}(t+t') = \widehat{U}(t) \circ \widehat{U}(t')$

$\widehat{U}(0) = Id$

Then we have a one parameter semi-group. If moreover the map $\widehat{U}$ is strongly continuous (that is $\lim_{\theta\to 0} \left\|\widehat{U}(\theta) - Id\right\| = 0$), it can be extended to $\mathbb{R}$. $\left(\widehat{U}, H\right)$ is a unitary representation of the abelian group $(\mathbb{R},+)$. We have a one parameter group, and because $\widehat{U}$ is a continuous Lie group morphism it is differentiable with respect to θ (Maths.1784).

Any strongly continuous one parameter group of operators on a Banach vector space admits an infinitesimal generator $S \in \mathcal{L}(H;H)$ such that : $\widehat{U}(t) = \sum_{n=0}^{\infty} \frac{t^n}{n!} S^n = \exp tS$ (Maths.1033). By derivation with respect to t we get : $\frac{d}{ds}\widehat{U}(s)|_{t=s} = (\exp tS) \circ S \Rightarrow S = \frac{d}{ds}\widehat{U}(s)|_{t=0}$

Because $\widehat{U}(t)$ is unitary S is anti-hermitian :

$\left\langle \widehat{U}(t)\psi, \widehat{U}(t)\psi' \right\rangle_H = \langle \psi, \psi' \rangle_H$

$\Rightarrow \left\langle \frac{d}{dt}\widehat{U}(t)\psi, \widehat{U}(t)\psi' \right\rangle_H + \left\langle \widehat{U}(t)\psi, \frac{d}{dt}\widehat{U}(t)\psi' \right\rangle_H = 0 \Rightarrow S = -S^*$

S is normal and has a spectral resolution P :

$S = \int_{Sp(S)} sP(s)$

S is anti-hermitian so its eigen-values are pure imaginary : $\lambda = -\overline{\lambda}$. $\widehat{U}(t)$ is not compact and S is not compact, usually its spectrum is continuous, so it is not associated to any observable.

We will see in the next Chapter a striking application of this case.

5.4 Extension to manifolds

Several extensions of the theorem 2 can be considered. One problem that we will meet in the next chapters is the following. In a model variables X are maps defined on a manifold M, valued in a fixed vector space, and belong to a space V of maps with the required properties. But a variable Y is defined through X : $Y(m) = f(X(m))$ and belongs to a manifold $N(X)$ depending on X. So the conditions 1 do not apply.

To address this kind of problem we need to adapt our point of view. We have seen the full mathematical definition of a manifold in the first section. A manifold M is a class of equivalence : the same point m of M can be defined by several charts, maps $\varphi : E \to M$ from a vector space E to M, with different coordinates : $m = \varphi_a(\xi_a) = \varphi_b(\xi_b)$ so that it defines classes of equivalence between sets of coordinates : $\xi_a \sim \xi_b \Leftrightarrow \varphi_a(\xi_a) = \varphi_b(\xi_b)$. These classes of equivalence are made clear by the transitions maps $\chi_{ba} : E \to E$, which are bijective : $\xi_a \sim \xi_b \Leftrightarrow \xi_b = \chi_{ba}(\xi_a)$. And these transitions maps are the key characteristic of the manifold. To a point m of M corresponds a class of equivalence of coordinates.

So let us consider a system represented by a model which meets the following :

Condition 23 *The model is comprised of :*

i) A finite number of variables, collectively denoted X, which are maps valued in a vector space E and meeting the conditions 1 : they belong to an open subset O of a separable, infinite dimensional Fréchet space V.

ii) A variable Y, valued in a set F, defined by a map : $f : O \to F :: Y = f(X)$

iii) A collection of linear continuous bijective maps $\mathfrak{U} = (U_a \in G\mathcal{L}(V;V))_{a\in A}$, comprising the identity, closed under composition : $\forall a, b \in A : U_a \circ U_b \in \mathfrak{U}$

iv) On V and F the equivalence relation :

$R : X \sim X' \Leftrightarrow \exists a \in A : X' = U_a(X) : f(X) = f(X')$

Denote the set $N = \{Y = f(X), X \in O\}$. The quotient set : N/R is comprised of classes of equivalence of points Y which can be defined by related coordinates. This is a manifold, which can be discrete and comprising only a finite number of points. One can also see the classes of equivalence of N/R as representing states of the system, defined equivalently by the variable $X, X' = U_a(X)$.

Notice that f is unique, no condition is required on E other than to be a vector space, and nothing on F. Usually the maps U_a are defined by : $U_a(X) = \chi_a \circ X$ where the maps $\chi_a \in GL(E;E)$ are bijective on E (not F or V) but only the continuity of U_a can be defined.

We have the following result :

Theorem 24 *For a system represented by a model meeting the conditions 23 :*

i) V can be embedded as an open of a Hilbert space H with a linear isometry $\Upsilon : V \to H$, to each U_a is associated the unitary operator $\widehat{U}_a = \Upsilon \circ U_a \circ \Upsilon^{-1}$ on

H, each class of equivalence $[V]_y$ *of R on V is associated to a class of equivalence* $[H]_y$ *in H of :*

$\widehat{R} : \psi \sim \psi' \Leftrightarrow \exists a \in A : \psi' = \widehat{U}_a(\psi)$. $[V]_y$ *is a partition of V and* $[H]_y$ *of H.*

ii) If (V, U) *is a representation of a Lie group G, then* $\left(H, \widehat{U}\right)$ *is a unitary representation of G and each* $[H]_y$ *is invariant by the action of G.*

Proof. i) R defines a partition of V, we can label each class of equivalence by the value of Y, and pick one element X_y in each class :

$[V]_y = \{X \in O : f(X) \sim f(X_y) = y\} \equiv \{X \in O : \exists a \in A : X = U_a(X_y)\}$
$\equiv \{X \in O : X = U_a(X_y), a \in A\}$

The variables X meet the conditions 1, O can be embedded as an open of a Hilbert space H and there is linear isomorphism : $\Upsilon : V \to H$

In $[V]_y$ the variables $X, X' = U_a(X)$ define the same state and we can implement the theorem 21. $\widehat{U}_a = \Upsilon \circ U_a \circ \Upsilon^{-1}$ is an unitary operator on H

$\forall X \in [V]_y : \widehat{U}_a \circ \Upsilon(X_y) = \Upsilon \circ U_a(X_y) = \Upsilon(X)$

The set $[H]_y = \Upsilon\left([V]_y\right) = \left\{\psi \in H : \psi = \widehat{U}_a(\Upsilon(X_y)), a \in A\right\}$ is the class of equivalence of :

$\widehat{R} : \psi \sim \psi' \Leftrightarrow \exists a \in A : \psi' = \widehat{U}_a(\psi)$

R defines a partition of V : $V = \cup_y [V]_y$ and $\widehat{R}$ defines a partition of H : $H = \cup_y [H]_y$

ii) If (V, U) is a representation of a Lie group G then $[V]_y$ is the orbit of X_y, $\left(H, \widehat{U}\right)$ is a unitary representation of G

Each $[H]_y$ is invariant by G. The vector subspace $[F]_y$ spanned by $[H]_y$ is invariant by G, so $\left([F]_y, \widehat{U}\right)$ is a representation of G. ■

As a consequence of the last result : if U is a compact group, then the representation $\left(H, \widehat{U}\right)$ is the sum of irreducible, orthogonal, finite dimensional representations. For each value of Y the subset $[H]_y$ is invariant by the action of G, so it must belong to one of the irreducible representations, as well as $[F]_y$. The maps X, for a given value of Y, belong to a finite dimensional vector space, and depend on a finite number of parameters. This is the usual meaning of the quantization of X.

6 THE EVOLUTION OF THE SYSTEM

In many models involving maps, the variables X_k are functions of the time t, which represents the evolution of the system. So this is a privileged argument of the functions. So far we have not made any additional assumption about the model : the open Ω of the Hilbert space contains all the possible values but, due to the laws to which it is subject, only some solutions will emerge, depending on the initial conditions. They are fixed by the value $X(0)$ of the variables at some origin 0 of time. They are specific to each realization of the system, but we should expect that the model and the laws provide a general solution, that is a map : $X(0) \to X$ which determines X for each specific occurrence of $X(0)$. It will happen if the laws are determinist. One says that the problem is well posed if for any initial conditions there is a unique solution X, and that X depends continuously on $X(0)$. We will give a more precise meaning of determinism by enlarging the conditions 1 as follows :

Condition 25 *: The model representing the system meets the conditions 1. Moreover :*

i) V is an infinite dimensional separable Fréchet space V of maps : $X = (X_k)_{k=1}^N :: R \to E$ where R is an open subset of $\mathbb{R}$ and E a normed vector space

ii) $\forall t \in R$ the evaluation map : $\mathcal{E}(t) : V \to E : \mathcal{E}(t) X = X(t)$ is continuous

The laws for the evolution of the system are such that the variables $(X_k)_{k=1}^N$, which define the possible states considered for the system (that we call the admissible states) meet the conditions :

iii) The initial state of the system, defined at $t = 0 \in R$, belongs to an open subset A of E

iv) For any solutions X, X' belonging to O if the set $\varpi = \{t, X(t) = X(t')\}$ has a non null Lebesgue measure then $X = X'$.

The last condition iv) means that the system is semi determinist : to the same initial conditions can correspond several different solutions, but if two solutions are equal on some interval then they are equal almost everywhere.

The condition ii) is rather technical and should be usually met. Practically it involves some relation between the semi-norms on V and the norm on E (this is why we need a norm on E) : when two variables X, X' are close in V, then their values $X(t), X'(t)$ must be close for almost all t. More precisely, because $\mathcal{E}(t)$ is linear, the continuity can be checked at $X = 0$ and reads:

$\forall t \in R, \forall X \in O : \forall \varepsilon > 0, \exists \eta : d(X, 0)_V < \eta \Rightarrow \|X(t)\|_E < \varepsilon$ where d is the metric on V

In all usual cases (such as L^p spaces or spaces of differentiable functions) $d(X, 0)_V \to 0 \Rightarrow \forall t \in R : \|X(t)\|_E \to 0$ and the condition ii) is met, but this is not a general result.

Notice that :

- the variables X can depend on any other arguments besides t as previously
- E can be infinite dimensional but must be normed
- no continuity condition is imposed on X.

6.1 Fundamental theorems for the evolution of a system

If the model meets the conditions 25 then it meets the conditions 1 : there is a separable, infinite dimensional, Hilbert space H, defined up to isomorphism, such that the states (admissible or not) $\mathcal{S}$ belonging to O can be embedded as an open subset $\Omega \subset H$ which contains 0 and a convex subset. Moreover to any basis of V is associated a bijective linear map $\Upsilon : V \to H$.

Theorem 26 *If the conditions 25 are met, then there are :*

i) a Hilbert space F, an open subset $\widetilde{A} \subset F$

ii) a map : $\Theta : R \to \mathcal{L}(F;F)$ such that $\Theta(t)$ is unitary and, for the admissible states $X \in O \subset V$:

$X(0) \in \widetilde{A} \subset F$

$\forall t : X(t) = \Theta(t)(X(0)) \in F$

iii) for each value of t an isometry : $\widehat{\mathcal{E}}(t) \in \mathcal{L}(H;F)$ such that for the admissible states $X \in O \subset V$:

$\forall X \in O : \widehat{\mathcal{E}}(t)\Upsilon(X) = X(t)$

where H is the Hilbert space and Υ is the linear chart associated to X and any basis of V

Proof. i) Define the equivalence relation on V :

$\mathcal{R} : X \sim X' \Leftrightarrow X(t) = X'(t)$ for almost every $t \in R$

and take the quotient space $V/\mathcal{R}$, then the set of admissible states is a set $\widetilde{O}$ such that :

$\widetilde{O} \in O \subset V$

$\forall X \in \widetilde{O} : X(0) \in A$

$\forall X, X' \in \widetilde{O}, \forall t \in R : X(t) = X'(t) \Rightarrow X = X'$

ii) Define :

$\forall t \in R : \widetilde{F}(t) = \left\{X(t), X \in \widetilde{O}\right\}$ thus $\widetilde{F}(0) = A$

A is a subset of E. There are families of independent vectors belonging to A, and a largest family $(f_j)_{j\in J}$ of independent vectors. It generates a vector space $F(0)$ which is a vector subspace of E, containing A.

$\forall u \in F(0) : \exists (x_j)_{j\in J} \in \mathbb{R}_0^J : u = \sum_{j\in J} x_j f_j$

The map :

$\widetilde{\Theta}(t) : \widetilde{F}(0) \to \widetilde{F}(t) :: \widetilde{\Theta}(t)u = \mathcal{E}(t) \circ \mathcal{E}(0)^{-1} u$

is bijective and continuous

The set $F(t) = \widetilde{\Theta}(t)F(0) \subset E$ is well defined by linearity :

$\widetilde{\Theta}(t)\left(\sum_{j\in J} x_j f_j\right) = \sum_{j\in J} x_j \widetilde{\Theta}(t)(f_j)$

The map : $\widetilde{\Theta}(t) : F(0) \to F(t)$ is linear, bijective, continuous on an open subset A, thus continuous, and the spaces $F(t)$ are isomorphic, vector subspaces of E, containing $\widetilde{F}(t)$.

Define : $(\varphi_j)_{j\in J}$ the largest family of independent vectors of

$\left\{\widetilde{\Theta}(t)(f_j), t \in R\right\}$. This is a family of independent vectors of E, which generates a subspace $\widetilde{F}$ of E, containing each of the $F(t)$ and thus each of the

$\widetilde{F}(t)$. Moreover each of the φ_j is the image of a unique vector f_j for some $t_j \in R$.

The map $\widetilde{\Theta}(t)$ is then a continuous linear map $\widetilde{\Theta}(t) \in \mathcal{L}\left(\widetilde{F};\widetilde{F}\right)$

iii) The conditions of proposition 1 are met for O and V, so there are a Hilbert space H and a linear map : $\Upsilon : O \to \Omega$

Each of the φ_j is the image of a unique vector f_j for some $t \in R$,and thus there is a uniquely defined family $(X_j)_{j\in J}$ of $\widetilde{O}$ such that $X_j(t_j) = \varphi_j$.

Define on $\widetilde{F}$ the bilinear symmetric definite positive form with coefficients :

$\langle \varphi_j, \varphi_k \rangle_{\widetilde{F}} = K_V\left(\mathcal{E}(t_j)^{-1}\varphi_j, \mathcal{E}(t_k)^{-1}\varphi_k\right)$

$= \left\langle \Upsilon\mathcal{E}(t_j)^{-1}\varphi_j, \Upsilon\mathcal{E}(t_k)^{-1}\varphi_k\right\rangle_H = \langle X_j, X_k\rangle_H$

By the Gram-Schmidt procedure we can build an orthonormal basis $(\widetilde{\varphi}_j)_{j\in J}$ of $\widetilde{F}$: $\widetilde{F} = Span(\widetilde{\varphi}_j)_{j\in J}$ and the Hilbert vector space : $F = \left\{\sum_{j\in J}\widetilde{x}_j\widetilde{\varphi}_j, (\widetilde{x}_j)_{j\in J} \in \ell^2(J)\right\}$ which is a vector space containing $\widetilde{F}$ (but is not necessarily contained in E).

iv) The map : $\widetilde{\Theta}(t) \in \mathcal{L}\left(\widetilde{F};\widetilde{F}\right)$ is a linear homomorphism, $\widetilde{F}$ is dense in F, thus $\widetilde{\Theta}(t)$ can be extended to a continuous operator $\Theta(t) \in \mathcal{L}(F;F)$ (Math.1003).

$\widetilde{\Theta}(t)$ is unitary on $\widetilde{F}$: $\langle u, v\rangle_{\widetilde{F}} = K_V\left(\mathcal{E}(0)^{-1}u, \mathcal{E}(0)^{-1}v\right)$ so $\Theta(t)$ is unitary on F.

iv) Define the map :

$\widehat{\mathcal{E}}(t) : \Omega \to F :: \widehat{\mathcal{E}}(t)\Upsilon(X) = X(t)$

where $\Omega \subset H$ is the open associated to V and O.

For $X \in \widetilde{O}$:

$\widehat{\mathcal{E}}(t)\Upsilon(X) = X(t) = \widetilde{\Theta}(t)X = \mathcal{E}(t)\circ\mathcal{E}(0)^{-1}X$

$\widehat{\mathcal{E}}(t) = \mathcal{E}(t)\circ\mathcal{E}(0)^{-1}\circ\Upsilon^{-1}$

$\widehat{\mathcal{E}}(t)$ is linear, continuous, bijective on Ω, it is an isometry :

$\left\langle \widehat{\mathcal{E}}(t)\psi, \widehat{\mathcal{E}}(t)\psi'\right\rangle_F = \langle X(t), X'(t)\rangle_F = \langle \Upsilon X, \Upsilon X'\rangle_H = \langle \psi, \psi'\rangle_H$

v) $A = \widetilde{F}(0)$ is an open subset of $F(0)$, which is itself an open vector subspace of F. Thus A can be embedded as an open subset $\widetilde{A}$ of F. ■

When X depends on other arguments ξ, the result reads :

$\forall t, \forall \xi : X(t,\xi) = \Theta(t)(X(0,\xi)) \in F$

Indeed the basic feature which is used is :

$\forall X, X' \in \widetilde{O}, \forall t \in R : X(t) = X'(t) \Rightarrow X = X'$

which means : $\forall t, \forall \xi : X(t,\xi) = X'(t,\xi) \Leftrightarrow X = X'$

As a consequence the model is determinist, up to the equivalence between maps almost everywhere equal. But the operator $\Theta(t)$ depends on t and not necessarily continuously, so the problem is not necessarily well posed. Notice that each solution $X(t)$ belong to E, but the Hilbert space F can be larger than E. Moreover the result holds if the conditions apply to some variables only.

But we have a stronger result.

Theorem 27 *If the model representing the system meets the conditions 1 and moreover :*

i) V is an infinite dimensional separable Fréchet space V of maps : $X = (X_k)_{k=1}^N :: R \to E$ *where E is a normed vector space*

ii) $\forall t \in \mathbb{R}$ *the evaluation map :* $\mathcal{E}(t) : V \to E : \mathcal{E}(t) X = X(t)$ *is continuous*

iii) the variables $X'_k(t) = X_k(t+\theta)$ *and* $X_k(t)$ *represent the same state of the system, for any* $t' = t + \theta$ *with a fixed* $\theta \in \mathbb{R}$

then :

i) there is a continuous map $S \in \mathcal{L}(V;V)$ *such that :*

$\mathcal{E}(t) = \mathcal{E}(0) \circ \exp tS$

$\forall t \in \mathbb{R} : X(t) = (\exp tS \circ X)(0) = \left(\sum_{n=0}^{\infty} \frac{t^n}{n!} S^n X\right)(0)$

and the operator $\widehat{S} = \Upsilon \circ S \circ \Upsilon^{-1}$ *associated to S is anti-hermitian*

ii) there are a Hilbert space F, an open $\widetilde{A} \subset F$*, a continuous anti-hermitian map* $\widetilde{S} \in \mathcal{L}(F;F)$ *such that :*

$\forall X \in O \subset V : X(0) \in \widetilde{A} \subset F$

$\forall t : X(t) = \left(\exp t\widetilde{S}\right)(X(0)) \in F$

iii) The maps X are smooth and :

$\frac{d}{ds} X(s)|_{s=t} = \widetilde{S} X(t)$

Proof. i) We have a change of variables U depending on a parameter $\theta \in \mathbb{R}$ which reads with the evaluation map : $\mathcal{E} : \mathbb{R} \times V \to F :: \mathcal{E}(t) X = X(t)$:

$\forall t, \theta \in \mathbb{R} : \mathcal{E}(t)(U(\theta) X) = \mathcal{E}(t+\theta)(X) \Leftrightarrow \mathcal{E}(t) U(\theta) = \mathcal{E}(t+\theta) = \mathcal{E}(\theta) U(t)$:

U defines a one parameter group of linear operators:

$U(\theta+\theta') X(t) = X(t+\theta+\theta') = U(\theta) \circ U(\theta') X(t)$

$U(0) X(t) = X(t)$

It is obviously continuous at $\theta = 0$ so it is continuous.

ii) The conditions 1 are met, so there are a Hilbert space H, a linear chart Υ, and $\widehat{U} : \mathbb{R} \to \mathcal{L}(H;H)$ such that $\widehat{U}(\theta)$ is linear, bijective, unitary :

$\forall X \in O : \widehat{U}(\theta)(\Upsilon(X)) = \Upsilon(U(\theta)(X))$

$\widehat{U}(\theta+\theta') = \Upsilon \circ U(\theta+\theta') \circ \Upsilon^{-1} = \Upsilon \circ U(\theta) \circ U(\theta') \circ \Upsilon^{-1} = \Upsilon \circ U(\theta) \circ \Upsilon^{-1} \circ \Upsilon \circ U(\theta') \circ \Upsilon^{-1} = \widehat{U}(\theta) \circ \widehat{U}(\theta')$

$\widehat{U}(0) = \Upsilon \circ U(0) \circ \Upsilon^{-1} = Id$

The map : $\widehat{U} : \mathbb{R} \to \mathcal{L}(H;H)$ is uniformly continuous with respect to θ, it defines a one parameter group of unitary operators. So there is an anti-hermitian operator $\widehat{S}$ with spectral resolution P such that :

$\widehat{U}(\theta) = \sum_{n=0}^{\infty} \frac{\theta^n}{n!} \widehat{S}^n = \exp \theta \widehat{S}$

$\frac{d}{ds} \widehat{U}(s)|_{\theta=s} = \left(\exp \theta \widehat{S}\right) \circ \widehat{S}$

$\widehat{S} = \int_{Sp(S)} sP(s)$

$\left\|\widehat{U}(\theta)\right\| = 1 \leq \exp \left\|\theta \widehat{S}\right\|$

iii) $S = \Upsilon^{-1} \circ \widehat{S} \circ \Upsilon$ is a continuous map on the largest vector subspace V_0 of V which contains O, which is a normed vector space with the norm induced by the positive kernel.

$\|S\| \leq \left\|\Upsilon^{-1}\right\| \left\|\widehat{S}\right\| \|\Upsilon\| = \left\|\widehat{S}\right\|$ because Υ is an isometry.

So the series $\sum_{n=0}^{\infty} \frac{\theta^n}{n!} S^n$ converges in V_0 and :

$U(\theta) = \Upsilon^{-1} \circ \widehat{U}(\theta) \circ \Upsilon = \sum_{n=0}^{\infty} \frac{\theta^n}{n!} S^n = \exp \theta S$

$\forall \theta, t \in \mathbb{R} : U(\theta) X(t) = X(t+\theta) = (\exp \theta S) X(t)$

$\mathcal{E}(t) \exp \theta S = \mathcal{E}(t+\theta)$

Exchange θ, t and take $\theta = 0$:

$\mathcal{E}(\theta) \exp tS = \mathcal{E}(t+\theta)$

$\mathcal{E}(0) \exp tS = \mathcal{E}(t) \in \mathcal{L}(V; E)$

which reads :

$\forall t \in \mathbb{R} : U(t) X(0) = X(t) = (\exp tS) X(0)$

(U, V_0) is a continuous representation of $(\mathbb{R}, +)$, U is smooth and X is smooth :

$\frac{d}{ds} U(s) X(0) |_{s=t} = \frac{d}{ds} X(s) |_{s=t} = SX(t)$

$\Leftrightarrow \frac{d}{ds} \mathcal{E}(s) |_{s=t} = S\mathcal{E}(t)$

The same result holds whatever the size of O in V, so S is defined over V.

iv) The set : $F(t) = \{X(t), X \in V\}$ is a vector subspace of E.

Each map is fully defined by its value at one point :

$\forall t \in \mathbb{R} : X(t) = (\exp tS \circ X)(0)$

$X(t) = X'(t) \Rightarrow \forall \theta : X(t+\theta) = X'(t+\theta) \Leftrightarrow X = X'$

So the conditions 25 are met.

$\Theta(t) : F(0) \to F(t) :: \Theta(t) u = \mathcal{E}(t) \circ \mathcal{E}(0)^{-1} u = \mathcal{E}(0) \circ \exp tS \circ \mathcal{E}(0)^{-1} u$

The map $\Theta(\theta) : F \to F$ defines a one parameter group, so it has an infinitesimal generator $\widetilde{S} \in \mathcal{L}(F; F) : \Theta(\theta) = \exp \theta \widetilde{S}$ and because $\Theta(\theta)$ is unitary $\widetilde{S}$ is anti-hermitian.

$\frac{d}{ds} \Theta(s) X(0) |_{s=t} = \frac{d}{ds} X(s) |_{s=t} = \widetilde{S} X(t)$ ■

As a consequence such a model is necessarily determinist, and the system is represented by smooth maps whose evolution is given by a unique operator. It is clear that the conditions 25 are then met, so this case is actually a special case of the previous one. Notice that, even if X was not assumed to be continuous, smoothness is a necessary result. This result can seem surprising, but actually the basic assumption about a translation in time means that the laws of evolution are smooth, and as a consequence the variables depend smoothly on the time. And conversely this implies that, whenever there is some discontinuity in the evolution of the system, the conditions above cannot hold : time has a specific meaning, related to a change in the environment.

6.2 Comments

The conditions above depend deeply on how the time is understood in the model. We have roughly two cases :

A) t is a parameter used only to identify a temporal location. In Galilean Geometry the time is independent from the spatial coordinates for any observer and one can consider a change of coordinates such as : $t' = t + \theta$ with any constant θ. The variables X, X' such that $X'(t') = X(t+\theta)$ represent the

same system. Similarly in Relativist Geometry the universe can be modelled as a manifold, and a change of coordinates with affine parameters, $\xi' = \xi + \theta$ with a fixed 4 vector θ, is a change of charts. The components of any quantity defined on the tensorial tangent bundle change according to the jacobian $\left[\frac{\partial \xi'}{\partial \xi}\right]$ which is the identity, so the corresponding variables represent the same system. Then we are usually in the conditions of the Theorem 27, and this is the basis of the Schrödinger equation.

B) t is a parameter used to measure the duration of a phenomenon, usually the time elapsed since some specific event, and it is clear that the origin of time matters and the variables X, X' such that $X'(t') = X(t+\theta)$ do not represent the same system. This is the case in more specific models, such as in Engineering. The proposition 27 does not hold, but the proposition 26 holds if the model is determinist.

The conditions 25 require at least that all the variables which are deemed significant are accounted for. As it as been discussed in the previous chapter, usually probabilist laws appear because some of them are missing. The Theorem 26 precises this issue : by denoting the missing variables Y, one needs to enlarge the vector space E, and similarly F. The map $\Theta(t)$ still exists, but it encompasses the couples $(X(t), Y(t))$. The dispersion of the observed values of $X(t)$ are then imputed to the distribution of the unknown values $Y(t)$.

6.3 Observables

When a system is studied through its evolution, the observables can be considered from two different points of view :

- in the movie way : the estimation of the parameters is done at the end of the period considered, from a batch of data corresponding to several times (which are not necessarily the same for all variables). So this is the map X which is estimated through an observable $X \to \Phi(X)$.

- in the picture way : the estimation is done at different times (the same for all the variables which are measured). So there are the values $X(t)$ which are estimated. Then the estimation of $X(t)$ is given by $\varphi(X(t)) = \varphi(\mathcal{E}(t) X)$, with φ a linear map from E to a finite dimensional vector space, which usually does not depend on t (the specification stays the same).

In the best scenario the two methods should give the same result, which reads :

$\varphi(\mathcal{E}(t) X) = \mathcal{E}(t)(\Phi X) \Leftrightarrow \varphi = \mathcal{E}(t) \circ \Phi \circ \mathcal{E}(t)^{-1}$

But usually, when it is possible, the first way gives a better statistical estimation.

6.4 Phases Transitions

There is a large class of problems which involve transitions in the evolution of a system. They do not involve the maps X, which belong to the same family as above, but the values $X(t)$ which are taken over a period of time in some vector

space E. There are distinct subsets of E, that we will call **phases** (to avoid any confusion with states which involves the map X), between which the state of the system goes during its evolution, such as the transition solid / gas or between magnetic states. The questions which arise are then : what are the conditions, about the initial conditions or the maps X, for the occurrence of such an event ? Can we forecast the time at which such event takes place ?

Staying in the general model meeting the conditions 25, the first issue is the definition of the phases. The general idea is that they are significantly different states, and it can be formalized by : the set $\{X(t), t \in R, X \in O\}$ is disconnected, it comprises two disjoint subsets E_1, E_2 closed in E.

If the maps $X : R \to F$ are continuous and R is an interval of $\mathbb{R}$ (as we will assume) then the image $X(R)$ is connected, the maps X cannot be continuous, and we cannot be in the conditions of proposition 27 (a fact which is interesting in itself), but we can be in the case of proposition 26. This is a difficult but also very common issue : in the real life such discontinuous evolutions are the rule. However, as we have seen, in the physical world discontinuities happen only at isolated points : the existence of a singularity is what makes interesting a change of phase. If the transition points are isolated, there is an open subset of R which contains each of them, a finite number of them in each compact subset of R, and at most a countable number of transition points. A given map X is then continuous (with respect to t) except in a set of points $(\theta_\alpha)_{\alpha \in A}, A \subset \mathbb{N}$. If $X(0) \in E_1$ then the odd transition points $\theta_{2\alpha+1}$ mark a transition $E_1 \to E_2$ and the opposite for the even points $\theta_{2\alpha}$.

If the conditions 25 are met then Θ is continuous except in $(\theta_\alpha)_{\alpha \in A}$, the transition points do not depend on the initial state $X(0)$, but the phase on each segment does. Then it is legitimate to assume that there is some probability law which rules the occurrence of a transition. We will consider two cases.

The simplest assumption is that the probability of the occurrence of a transition at any time t is constant. Then it depends only on the cumulated lengths of the periods $T_1 = \sum_{\alpha=0} [\theta_{2\alpha}, \theta_{2\alpha+1}], T_2 = \sum_{\alpha=0} [\theta_{2\alpha+1}, \theta_{2\alpha+2}]$ respectively.

Let us assume that $X(0) \in E_1$ then the changes $E_1 \to E_2$ occur for $t = \theta_{2\alpha+1}$, the probability of transitions read :

$\Pr(X(t+\varepsilon) \in E_2 | X(t) \in E_1) = \Pr(\exists \alpha \in \mathbb{N} : t + \varepsilon \in [\theta_{2\alpha+1}, \theta_{2\alpha+2}])$
$= T_2 / (T_1 + T_2)$
$\Pr(X(t+\varepsilon) \in E_1 | X(t) \in E_2) = \Pr(\exists \alpha \in \mathbb{N} : t + \varepsilon \in [\theta_{2\alpha}, \theta_{2\alpha+1}])$
$= T_1 / (T_1 + T_2)$
$\Pr(X(t) \in E_1) = T_1 / [R]; \Pr(X(t) \in E_2) = T_2 / [R]$

The probability of a transition at t is : $T_2 / (T_1 + T_2) \times T_1 / (T_1 + T_2) + T_1 / (T_1 + T_2) \times T_2 / (T_1 + T_2) = 2T_1T_2 / (T_1 + T_2)^2$. It does not depend of the initial phase, and depends only on Θ. This probability law can be checked from a batch of data about the values of T_1, T_2 for each observed transition.

However usually the probability of a transition depends on the values of the variables. The phases are themselves characterized by the value of $X(t)$,

so a sensible assumption is that the probability of a transition increases with the proximity of the other phase . Using the Hilbert space structure of F it is possible to address practically this case.

If E_1, E_2 are *closed convex subsets* of F, which is a Hilbert space, there is a unique map : $\pi_1 : F \to E_1$. The vector $\pi_1(x)$ is the unique $y \in E_1$ such that $\|x - y\|_F$ is minimum. The map π_1 is continuous and $\pi_1^2 = \pi_1$. And similarly for E_2.

The quantity $r = \|X(t) - \pi_1(X(t))\|_F + \|X(t) - \pi_2(X(t))\|_F =$ the distance to the other subset than where $X(t)$ lies, so one can assume that the probability of a transition at t is : $f(r)$ where $f : \mathbb{R} \to [0,1]$ is a probability density. The probability of a transition depends only on the state at t, but one cannot assume that the transitions points θ_α do not depend on X.

The result holds if E_1, E_2 are closed *vector subspaces* of F such that $E_1 \cap E_2 = \{0\}$. Then

$X(t) = \pi_1(X(t)) + \pi_2(X(t))$

and $\|X(t)\|^2 = \|\pi_1(X(t))\|^2 + \|\pi_2(X(t))\|^2$

$\frac{\|\pi_1(X(t))\|^2}{\|X(t)\|^2}$ can be interpreted as the probability that the system at t is in the phase E_1.

One important application is forecasting a transition for a given map X. From the measure of $X(t)$ one can compute for each t the quantity $r(t) = \|X(t) - \pi_1(X(t))\|_F + \|X(t) - \pi_2(X(t))\|_F$ and, if we know f, we have the probability of a transition at t. The practical problem is then to estimate f from the measure of r over a past period $[0, T]$. A very simple, non parametric, estimator can be built when X are maps depending only of t (see J.C.Dutailly *Estimation of the probability of transitions between phases*). It can be used to forecast the occurrence of events such as earth quakes.

7 INTERACTING SYSTEMS

7.1 Representation of interacting systems

In the propositions above no assumption has been done about the interaction with exterior variables. If the values of some variables are given (for instance to study the impact of external factors with the system) then they shall be fully integrated into the set of variables, at the same footing as the others.

A special case occurs when one considers two systems S_1, S_2, which are similarly represented, meaning that that we have the same kind of variables, defined as identical mathematical objects and related significance. To account for the interactions between the two systems the models are of the form :

$$\begin{array}{ccc} \ulcorner & S_1 & \urcorner \\ X_1 & & Z_1 \\ V_1 & \times & W_1 \\ & \downarrow \Upsilon_1 & \\ & \psi_1 & \\ & H_1 & \end{array} \qquad \begin{array}{ccc} \ulcorner & S_2 & \urcorner \\ X_2 & & Z_2 \\ V_2 & \times & W_2 \\ & \downarrow \Upsilon_2 & \\ & \psi_2 & \\ & H_2 & \end{array}$$

$$\begin{array}{ccc} \ulcorner & S_{1+2} & \urcorner \\ X_1 & & X_2 \\ V_1 & \times & V_2 \\ & & \\ \psi_1 & & \psi_2 \\ H_1 & \times & H_2 \end{array}$$

X_1, X_2 are the variables (as above X denotes collectively a set of variables) characteristic of the systems S_1, S_2,and Z_1, Z_2 are variables representing the interactions. Usually these variables are difficult to measure and to handle. One can consider the system S_{1+2} with the direct product $X_1 \times X_2$, but doing so we obviously miss the interactions Z_1, Z_2.

We see now how it is possible to build a simpler model which keeps the features of S_1, S_2 and accounts for their interactions.

We consider the models without interactions (so with only X_1, X_2) and we assume that they meet the conditions 1. For each model $S_k, k = 1, 2$ there are

a linear map : $\Upsilon_k : V_k \to H_k :: \Upsilon_k (X_k) = \psi_k = \sum_{i \in I_k} \langle \phi_{ki}, \psi_k \rangle e_{ki}$

a positive kernel : $K_k : V_k \times V_k \to \mathbb{R}$

Let us denote S the new model. Its variables will be collectively denoted Y, valued in a Fréchet vector space V'. There will be another Hilbert space H', and a linear map $\Upsilon' : V' \to H'$ similarly defined. As we have the choice of the model, we will impose some properties to Y and V' in order to underline both that they come from S_1, S_2 and that they are interacting.

Condition 28 *i) The variable Y can be deduced from the value of X_1, X_2 : there must be a bilinear map : $\Phi : V_1 \times V_2 \to V'$*

ii) Φ must be such that whenever the systems S_1, S_2 are in the states ψ_1, ψ_2 then S is in the state ψ' and

$\Upsilon'^{-1}(\psi') = \Phi\left(\Upsilon_1^{-1}(\psi_1), \Upsilon_2^{-1}(\psi_2)\right)$

iii) The positive kernel is a defining feature of the models, so we want a positive kernel K' of (V', Υ') such that :

$\forall X_1, X_1' \in V_1, \forall X_2, X_2' \in V_2$:

$K'(\Phi(X_1, X_2), \Phi(X_1', X_2')) = K_1(X_1, X_1') \times K_2(X_2, X_2')$

We will prove the following :

Theorem 29 *Whenever two systems S_1, S_2 interact, there is a model S encompassing the two systems and meeting the conditions 28 above. It is obtained by taking the tensor product of the variables specific to S_1, S_2. Then the Hilbert space of S is the tensorial product of the Hilbert spaces associated to each system.*

Proof. First let us see the consequences of the conditions if they are met.

The map : $\varphi : H_1 \times H_2 \to H' :: \varphi(\psi_1, \psi_2) = \Phi\left(\Upsilon_1^{-1}(\psi_1), \Upsilon_2^{-1}(\psi_2)\right)$ is bilinear. So, by the universal property of the tensorial product, there is a unique map $\widehat{\varphi} : H_1 \otimes H_2 \to H'$ such that : $\varphi = \widehat{\varphi} \circ \imath$ where $\imath : H_1 \times H_2 \to H_1 \otimes H_2$ is the tensorial product (Maths.369).

The condition iii) reads :

$\langle \Upsilon_1(X_1), \Upsilon_1(X_1') \rangle_{H_1} \times \langle \Upsilon_2(X_2), \Upsilon_2(X_2') \rangle_{H_2}$

$= \langle (\Upsilon' \circ \Phi(\Upsilon_1(X_1), \Upsilon_2(X_2)), \Upsilon' \circ \Phi(\Upsilon_1(X_1'), \Upsilon_2(X_2'))) \rangle_{H'}$

$\langle \psi_1, \psi_1' \rangle_{H_1} \times \langle \psi_2, \psi_2' \rangle_{H_2} = \langle \varphi(\psi_1, \psi_2), \varphi(\psi_1', \psi_2') \rangle_{H'}$

$= \langle \widehat{\varphi}(\psi_1 \otimes \psi_2), \widehat{\varphi}(\psi_1' \otimes \psi_2') \rangle_{H'}$

The scalar products on H_1, H_2 extend in a scalar product on $H_1 \otimes H_2$, endowing the latter with the structure of a Hilbert space with :

$\langle (\psi_1 \otimes \psi_2), (\psi_1' \otimes \psi_2') \rangle_{H_1 \otimes H_2} = \langle \psi_1, \psi_1' \rangle_{H_1} \langle \psi_2, \psi_2' \rangle_{H_2}$

and then the reproducing kernel is the product of the reproducing kernels (Maths.1208).

So we must have : $\langle \widehat{\varphi}(\psi_1 \otimes \psi_2), \widehat{\varphi}(\psi_1' \otimes \psi_2') \rangle_{H'} = \langle \psi_1 \otimes \psi_2, \psi_1' \otimes \psi_2' \rangle_{H_1 \otimes H_2}$ and $\widehat{\varphi}$ must be an isometry : $H_1 \otimes H_2 \to H'$

So by taking $H' = H_1 \otimes H_2$ and $V' = V_1 \otimes V_2$ we meet the conditions. ■

The conditions above are a bit abstract, but are logical and legitimate in the view of the Hilbert spaces. They lead to a natural solution, which is not unique and makes sense only if the systems are defined by similar variables. The measure of the tensor S can be addressed as before, the observables being linear maps defined in the tensorial products $V_1 \otimes V_2$, $H_1 \otimes H_2$ and valued in finite dimensional vector subspaces of these tensor products.

7.2 Comments

A key point in this representation is the difference between the simple direct product : $V_1 \times V_2$ and the tensorial product $V_1 \otimes V_2$, an issue about which there is much confusion.

The knowledge of the states (X_1, X_2) of both systems requires two vectors of I components each, that is $2 \times I$ scalars, and the knowledge of the state S requires a vector of I^2 components. So the measure of S requires more data, and brings more information, because it encompasses all the interactions. Moreover *a tensor is not necessarily the tensorial product of vectors* (if it is so it is said to be **decomposable**), it is the sum of such tensors. There is no canonical map : $V_1 \otimes V_2 \to V_1 \times V_2$. So *there is no simple and unique way to associate two vectors* (X_1, X_2) *to one tensor S*. This seems paradoxical, as one could imagine that both systems can always be studied, and their states measured, even if they are interacting. But the simple fact that we consider interactions means that the measure of the state of one of the system shall account for the conditions in which the measure is done, so it shall precise the value of the state of the other system and of the interactions Z_1, Z_2.

If a model is arbitrary, its use must be consistent : if the scientist assumes that there are interactions, they must be present somewhere in the model, as variables for the computations as well as data to be collected. They can be dealt with in two ways. Either we opt for the two systems model, and we have to introduce the variables Z_1, Z_2 representing the interactions, then we have two separate models as in the first section. The study of their interactions can be a topic of the models, but this is done in another picture and requires additional hypotheses about the laws of the interactions. Or, if we intend to account for both systems and their interactions in a single model, we need a representation which supports more information that can bring $V_1 \times V_2$. The tensorial product is one way to enrich the model, this is the most economical and, as far as one follows the guidelines i),ii),iii) above, the only one. The complication in introducing general tensors is the price that we have to pay to account for the interactions. This representation does not, in any way, imply anything about *how* the systems interact, or even if they interact at all (in this case S is always decomposable). As usual the choice is up to the scientist, based upon how he envisions the problem at hand. But he has to live with his choice.

This issue is at the root of the paradoxes of entanglement. With many variants it is an experiment which involves two objects, which interact at the beginning, then are kept separated and non interacting, and eventually one measures the state of one of the two objects, from which the state of the other can be deduced with some probability. If we have two objects which interact at some point, with a significant result because it defines a new state, and we compare their states, then we must either incorporate the interactions, or consider that they constitute a single system and use the tensorial product. The fact that the objects cease to interact at some point does not matter : they are considered together if we compare their states. The interactions must be accounted for, one way or another and, when an evolution is considered, this is the map which represents the whole of the evolution which is significant, not its value at some time.

A common interpretation of this representation is to single out decomposable tensors $\Psi = \psi_1 \otimes \psi_2$, called "pure states", so that actual states would be a superposition of pure states (a concept popularized by the famous Schrödinger's

cat). It is clear that in an interacting system the pure states are an abstraction, which actually would represent two non interacting systems, so their superposition is an artificial construction. It can be convenient in simple cases, where the states of each system can be clearly identified, or in complicated models to represent quantities which are defined over the whole system as we will see later. But it does not imply any mysterious feature, notably any probabilist behavior, for the real systems. A state of the two interacting systems is represented by a single tensor, and a tensor is not necessarily decomposable, but it is a sum of decomposable tensors.

7.3 Homogeneous systems

The previous result can be extended to N (a number that we will assumed to be fixed) similar systems (that we will call **microsystems**), represented by the same model, interacting together. For each microsystem, identified by a label s, the Hilbert space H and the linear map Υ are the same, the state S of the total system can be represented as a vector belonging to the tensorial product $\mathbf{V}_N = \otimes_{s=1}^N V$, associated to a tensor Ψ belonging to the tensorial product $\mathbf{H}_N = \otimes_{s=1}^N H$. The linear maps $\Upsilon \in \mathcal{L}(V;H)$ can be uniquely extended as maps $\Upsilon_N \in \mathcal{L}(\mathbf{V}_N;\mathbf{H}_N)$ such that (Maths.423) :

$\Upsilon_N(X_1 \otimes ... \otimes X_N) = \Upsilon(X_1) \otimes ... \otimes \Upsilon(X_N)$

The state of the system is then totally defined by the value of tensors S, Ψ, with I^N components.

We have general properties on these tensorial products (Maths.1208).

If $(\widetilde{\varepsilon}_i)_{i\in I}$ is a Hilbertian basis of H then $E_{i_1...i_N} = \widetilde{\varepsilon}_{i_1} \otimes ... \otimes \widetilde{\varepsilon}_{i_N}$ is a Hilbertian basis of $\otimes_{s=1}^N H$. The scalar product is defined by linear extension of

$\langle \Psi, \Psi' \rangle_{\mathbf{H}_N} = \langle \psi_1, \psi_1' \rangle_H \times ... \times \langle \psi_N, \psi_N' \rangle_H$

for decomposable tensors : $\Psi = \psi_1 \otimes ... \otimes \psi_N, \Psi' = \psi_1' \otimes ... \otimes \psi_N'$.

The subspaces $\otimes_{s=1}^p H \otimes \widetilde{\varepsilon}_i \otimes_{s=p+2}^N H$ are orthogonal and $\otimes_{s=1}^N H \simeq \ell^2\left(I^N\right)$

Any operator on H can be extended on $\otimes_{s=1}^N H$ with similar properties : a self adjoint, unitary or compact operator extends uniquely as a self adjoint, unitary or compact operator (Maths.1211).

In the general case the label matters : the state $S = X_1 \otimes ... \otimes X_N$ is deemed different from $S = X_{\sigma(1)} \otimes ... \otimes X_{\sigma(N)}$ where $\left(X_{\sigma(p)}\right)_{p=1}^N$ is a permutation of $(X_s)_{s=1}^N$. If the microsystems have all the same behavior they are, for the observer, indistinguishable. Usually the behavior is related to a parameter analogous to a size, so in such cases the microsystems are assumed to have the same size. We will say that these interacting systems are homogeneous :

Definition 30 *A **homogeneous system** is a system comprised of a fixed number N of microsystems, represented in the same model, such that any permutation of the N microsystems gives the same state of the total system.*

We have the following result :

Proposition 31 *The states Ψ of homogeneous systems belong to an open subset of a subspace $\boldsymbol{h}$ of the Hilbert space $\otimes_{s=1}^{N} H$, defined by :*

i) a class of conjugacy $\mathfrak{S}(\lambda)$ of the group of permutations $\mathfrak{S}(N)$,defined itself by a decomposition of N in p parts :

$\lambda = \{0 \leq n_p \leq ... \leq n_1 \leq N, n_1 + ...n_p = N\}$.

ii) p distinct vectors $(\widetilde{\varepsilon}_j)_{j=1}^{p}$ of a Hermitian basis of H which together define a subspace H_J

iii) The space $\boldsymbol{h}$ of tensors representing the states of the system is then :

either the symmetric tensors belonging to : $\odot_{n_1} H_J \otimes \odot_{n_2} H_J ... \otimes \odot_{n_p} H_J$

or the antisymmetric tensors belonging to : $\wedge_{n_1} H_J \otimes \wedge_{n_2} H_J ... \otimes \wedge_{n_p} H_J$

Proof. i) In the representation of the general system the microsystems are identified by some label s = 1 ... N. An exchange of labels $U(\sigma)$ is a change of variables, represented by an action of the group of permutations $\mathfrak{S}(N)$: U is defined uniquely by linear extension of $U(\sigma)(X_1 \otimes ... \otimes X_N) = X_{\sigma(1)} \otimes ... \otimes X_{\sigma(N)}$ on decomposable tensors.

We can implement the Theorem 22 proven previously. The tensors ψ representing the states of the system belong to a Hilbert space $\mathbf{H}_N \subset \otimes_{s=1}^{N} H$ such that $\left(\mathbf{H}_N, \widehat{U}\right)$ is a unitary representation of $\mathfrak{S}(N)$. Which implies that $\mathbf{H}_N$ is invariant by $\widehat{U}$. The action of $\widehat{U}$ on $\otimes_{s=1}^{N} H$ is defined uniquely by linear extension of

$\widehat{U}(\sigma)(\psi_1 \otimes ... \otimes \psi_N) = \psi_{\sigma(1)} \otimes ... \otimes \psi_{\sigma(N)}$ on decomposable tensors.

$\Psi \in \otimes_{s=1}^{N} H$ reads in a Hilbert basis $(\widetilde{\varepsilon}_i)_{i \in I}$ of H :

$\Psi = \sum_{i_1 ... i_N \in I} \Psi^{i_1 ... i_N} \widetilde{\varepsilon}_{i_1} \otimes ... \widetilde{\varepsilon}_{i_N}$ and :

$\widehat{U}(\sigma)\Psi = \sum_{i_1 ... i_N \in I} \Psi^{i_1 ... i_N} \widehat{U}(\sigma)(\widetilde{\varepsilon}_{i_1} \otimes ... \widetilde{\varepsilon}_{i_N}) = \sum_{i_1 ... i_N \in I} \Psi^{i_1 ... i_N} \widetilde{\varepsilon}_{\sigma(i_1)} \otimes ... \widetilde{\varepsilon}_{\sigma(i_N)}$

$= \sum_{i_1 ... i_N \in I} \Psi^{\sigma(i_1) ... \sigma(i_N)} \widetilde{\varepsilon}_{i_1} \otimes ... \widetilde{\varepsilon}_{i_N}$

$\left\langle \widehat{U}(\sigma)\Psi, \widehat{U}(\sigma)\Psi' \right\rangle = \langle \Psi, \Psi' \rangle$

$\Leftrightarrow \sum_{i_1 ... i_N \in I} \Psi^{\sigma(i_1) ... \sigma(i_N)} \Psi'^{\sigma(i_1) ... \sigma(i_N)} = \sum_{i_1 ... i_N \in I} \Psi^{i_1 ... i_N} \Psi'^{i_1 ... i_N}$

The only vector subspaces of $\otimes_{s=1}^{N} H$ which are invariant by $\widehat{U}$ and on which $\widehat{U}$ is unitary are spaces of symmetric or antisymmetric tensors :

symmetric : $\Psi^{\sigma(i_1) ... \sigma(i_N)} = \Psi^{i_1 ... i_N}$

antisymmetric : $\Psi^{\sigma(i_1) ... \sigma(i_N)} = \epsilon(\sigma) \Psi^{i_1 ... i_N}$

ii) $\mathfrak{S}(N)$ is a finite, compact group. Its unitary representations are the sum of orthogonal, finite dimensional, unitary, irreducible representations (Maths.1948). Let $\mathbf{h} \subset \otimes_{s=1}^{N} H$ be an irreducible, finite dimensional, representation of $\widehat{U}$. Then $\forall \sigma \in \mathfrak{S}(N) : \widehat{U}(\sigma)\mathbf{h} \subset \mathbf{h}$

iii) Let J a finite subset of I with $card(J) \geq N$, H_J the associated Hilbert space, $\widehat{Y}_J : H \to H_J$ the projection, and $\widehat{Y}_{J_N} = \otimes_N \widehat{Y}_J$ be the extension of $\widehat{Y}_J$ to $\otimes_{s=1}^{N} H$:

$\widehat{Y}_{J_N}\left(\sum_{i_1 ... i_N \in I} \Psi^{i_1 ... i_N} \widetilde{\varepsilon}_{i_1} \otimes ... \widetilde{\varepsilon}_{i_N}\right) = \sum_{i_1 ... i_N \in J} \Psi^{i_1 ... i_N} \widetilde{\varepsilon}_{i_1} \otimes ... \widetilde{\varepsilon}_{i_N}$

Then :

$\forall \sigma \in \mathfrak{S}(N) : \widehat{U}(\sigma)\widehat{Y}_{J_N}\left(\sum_{i_1 ... i_N \in I} \Psi^{i_1 ... i_N} \widetilde{\varepsilon}_{i_1} \otimes ... \widetilde{\varepsilon}_{i_N}\right)$

$= \sum_{i_1 \ldots i_N \in J} \Psi^{\sigma(i_1)\ldots\sigma(i_N)} \widetilde{\varepsilon}_{i_1} \otimes \ldots \widetilde{\varepsilon}_{i_N} = \widehat{Y}_{J_N} \widehat{U}(\sigma)\Psi$

So if $\mathbf{h}$ is invariant by $\widehat{U}$ then $\widehat{Y}_{J_N}\mathbf{h}$ is invariant by $\widehat{U}$. If $\left(\mathbf{h},\widehat{U}\right)$ is an irreducible representation then the only invariant subspace are 0 and $\mathbf{h}$ itself, so necessarily $\mathbf{h} \subset \widehat{Y}_{J_N}\left(\otimes_{s=1}^{N} H\right)$ for $card(J) = N$. Which implies : $\mathbf{h} \subset \otimes_N H_J$ with $H_J = \widehat{Y}_J H$ and $card(J) = N$.

iv) There is a partition of $\mathfrak{S}(N)$ in conjugacy classes $\mathfrak{S}(\lambda)$ which are subgroups defined by a decomposition of N in p parts :

$\lambda = \{0 \leq n_p \leq \ldots \leq n_1 \leq N, n_1 + \ldots n_p = N\}$. Notice that there is an order on the sets $\{\lambda\}$. Each element of a conjugacy class is then defined by a repartition of the integers $\{1, 2, \ldots N\}$ in p subsets of n_k items (this is a Young Tableau) (Maths. 5.2.2). A class of conjugacy is an abelian subgroup of $\mathfrak{S}(N)$: its irreducible representations are unidimensional.

The irreducible representations of $\mathfrak{S}(N)$ are then defined by a class of conjugacy, and the choice of a vector.

$\mathbf{h}$ is a Hilbert space, thus it has a Hilbertian basis, composed of decomposable tensors which are of the kind $\widetilde{\varepsilon}_{j_1} \otimes \ldots \otimes \widetilde{\varepsilon}_{j_N}$ where $\widetilde{\varepsilon}_{j_k}$ are chosen among the vectors of a Hermitian basis $(\widetilde{\varepsilon}_j)_{j \in J}$ of H_J

If $\widetilde{\varepsilon}_{j_1} \otimes \ldots \otimes \widetilde{\varepsilon}_{j_N} \in H, \forall \sigma \in \mathfrak{S}(N) : \widehat{U}(\sigma)\widetilde{\varepsilon}_{j_1} \otimes \ldots \otimes \widetilde{\varepsilon}_{j_N} = \widetilde{\varepsilon}_{j_{\sigma(1)}} \otimes \ldots \otimes \widetilde{\varepsilon}_{j_{\sigma(N)}} \in \mathbf{h}$

and because the representation is irreducible the basis of $\mathbf{h}$ is necessarily composed from a set of $p \leq N$ vectors $\widetilde{\varepsilon}_j$ by action of $\widehat{U}(\sigma)$

Conversely : for any Hermitian basis $(\widetilde{\varepsilon}_i)_{i \in I}$ of H, any subset J of cardinality N of I, any conjugacy class λ, any family of vectors $(\widetilde{\varepsilon}_{j_k})_{k=1}^{p}$ chosen in $(\widetilde{\varepsilon}_i)_{i \in J}$, the action of $\widehat{U}$ on the tensor :

$\Psi_\lambda = \otimes_{n_1} \widetilde{\varepsilon}_{j_1} \otimes_{n_2} \widetilde{\varepsilon}_{j_2} \ldots \otimes_{n_p} \widetilde{\varepsilon}_{j_p}, j_1 \leq j_2 .. \leq j_p$

gives the same tensor if $\sigma \in \mathfrak{S}(\lambda) : \widehat{U}(\sigma)\Psi_\lambda = \Psi_\lambda$

gives a different tensor if $\sigma \in \mathfrak{S}(\lambda^c)$ the conjugacy class complementary to $\mathfrak{S}(\lambda) : \mathfrak{S}(\lambda^c) = \complement_{\mathfrak{S}(N)}^{\mathfrak{S}(\lambda)}$

so it provides an irreducible representation by :

$\forall \Psi \in \mathbf{h} : \Psi = \sum_{\sigma \in \mathfrak{S}(\lambda^c)} \Psi^\sigma \widehat{U}(\sigma)\left(\otimes_{n_1} \widetilde{\varepsilon}_{j_1} \otimes_{n_2} \widetilde{\varepsilon}_{j_2} \ldots \otimes_{n_p} \widetilde{\varepsilon}_{j_p}\right)$

where the components Ψ^σ are labeled by the vectors of a basis of $\mathbf{h}$. The dimension of $\mathbf{h}$ his given by the cardinality of $\mathfrak{S}(\lambda^c)$ that is : $\frac{N!}{n_1! \ldots n_p!}$. All the vector spaces $\mathbf{h}$ of the same conjugacy class (but different vectors $\widetilde{\varepsilon}_i$) have the same dimension, thus they are isomorphic.

v) A basis of $\mathbf{h}$ is comprised of tensorial products of N vectors of a Hilbert basis of H. So we can give the components of the tensors of $\mathbf{h}$ with respect to $\otimes_{s=1}^{N} H$. We have two non equivalent representation :

By symmetric tensors : $\mathbf{h}$ is then isomorphic to $\odot_{n_1} H_J \otimes \odot_{n_2} H_J \ldots \otimes \odot_{n_p} H_J$ where the symmetric tensorial product $\odot$ and the space of n order symmetric tensor on H_J is $\odot_n H_J$

By antisymmetric tensors : $\mathbf{h}$ is then isomorphic to $\wedge_{n_1} H_J \otimes \wedge_{n_2} H_J \ldots \otimes \wedge_{n_p} H_J$ and the space of n order antisymmetric tensor on H_J is $\wedge_n H_J$

The result extends to V_N by : $S = \Upsilon_N^{-1}(\Psi)$ ■

7.4 Remarks

i) For each choice of a class of conjugacy, and each choice of the vectors $(\widetilde{\varepsilon}_j)_{j=1}^p$ which defines H_J, we have a different irreducible representation with vector space **h.** Different classes of conjugacy gives non equivalent representations. But different choices of the Hermitian basis $(\widetilde{\varepsilon}_j)_{j\in I}$ and the subset J of I, for a given class of conjugacy, give equivalent representations, and they can be arbitrary. So, for a given system, the set of states is characterized by a subset J of N elements in any basis of H, and by a class of conjugacy.

A change of the state of the system can occur either inside the same vector space **h**, or between irreducible representations: **h→h'**. As we will see in the next chapters usually the irreducible representation is fixed by other variables (such that energy) and a change of irreducible representation implies a discontinuous process. The states of the total system are quantized by the interactions.

ii) $\otimes_{n_1}\widetilde{\varepsilon}_{j_1}\otimes_{n_2}\widetilde{\varepsilon}_{j_2}...\otimes_{n_p}\widetilde{\varepsilon}_{j_p}$ can be seen as representing a configuration where n_k microsystems are in the same state $\widetilde{\varepsilon}_{j_k}$.The class of conjugacy, characterized by the integers n_p, correspond to the distribution of the microsystems between fixed states.

iii) If O is a convex subset then S belongs to a convex subset, and the basis can be chosen such that $\forall\Psi\in\mathbf{h}$ is a linear combination $(y_k)_{k=1}^q$ of the generating tensors with $y_k\in[0,1]\,,\sum_{k=1}^q y_k=1$. S can then be identified to the expected value of a random variable which would take one of the value $\otimes_{n_1}X_1\otimes_{n_2}X_2...\otimes_{n_p}X_p$, which corresponds to n_k microsystems having the state X_k. As exposed above the identification with a probabilist model is formal : there is no random behavior assumed for the physical system.

iv) In the probabilist picture one can assume that each microsystem behaves independently, and has a probability π_j to be in the state represented by $\widetilde{\varepsilon}_j$ and $\sum_{j=1}^N\pi_j=1$. Then the probability that we have $(n_k)_{k=1}^p$ microstates in the states $(\widetilde{\varepsilon}_k)_{k=1}^p$ is $\frac{N!}{n_1!...n_p!}\left(\pi_{j_1}\right)^{n_1}...\left(\pi_{j_p}\right)^{n_p}$.

v) The set of symmetric tensor $\odot_n H_J$ is a closed vector subspace of $\otimes_n H_J$, this is a Hilbert space, $\dim\otimes_n H_J=C_{p+n-1}^{p-1}$ with Hilbertian basis $\frac{1}{\sqrt{n!}}\odot_{j\in J}\widetilde{\varepsilon}_j=\frac{1}{\sqrt{n!}}S_n\left(\otimes_{j\in J}\widetilde{\varepsilon}_j\right)$ where the symmetrizer is :

$S_n\left(\sum_{(i_1...i_n)}\psi^{i_1..i_n}\widetilde{\varepsilon}_{i_1}\otimes..\otimes\widetilde{\varepsilon}_{i_n}\right)=\sum_{(i_1...i_n)}\psi^{i_1..i_n}\sum_{\sigma\in\mathfrak{S}(n)}\widetilde{\varepsilon}_{\sigma(1)}\otimes....\widetilde{\varepsilon}_{\sigma(k)}$

A tensor is symmetric iff : $\Psi\in\odot_n H_J\Leftrightarrow S_n\left(\Psi\right)=n!\Psi$ (Maths. 7.2.1,13.5.2).

The set of antisymmetric tensor $\wedge_n H_J$ is a closed vector subspace of $\otimes_n H_J$, this is a Hilbert space, $\dim\wedge_n H_J=C_p^n$ with Hilbertian basis $\frac{1}{\sqrt{n!}}\wedge_{j\in J}\widetilde{\varepsilon}_j=\frac{1}{\sqrt{n!}}A_n\left(\otimes_{j\in J}\widetilde{\varepsilon}_j\right)$ with the antisymmetrizer :

$A_n\left(\sum_{(i_1...i_n)}\psi^{i_1..i_n}\widetilde{\varepsilon}_{i_1}\otimes..\otimes\widetilde{\varepsilon}_{i_n}\right)=\sum_{(i_1...i_n)}\psi^{i_1..i_n}\sum_{\sigma\in\mathfrak{S}(n)}\epsilon\left(\sigma\right)\widetilde{\varepsilon}_{\sigma(1)}\otimes....\widetilde{\varepsilon}_{\sigma(k)}$

A tensor is antisymmetric iff : $\Psi\in\wedge_n H_J\Leftrightarrow A_n\left(\Psi\right)=n!\Psi$ (Maths. 7.2.2,13.5.2)

v) for $\theta\in\mathfrak{S}\left(N\right):\widehat{U}(\theta)\Psi$ is usually different from Ψ

7.5 Global observables of homogeneous systems

The previous definitions of observables can be extended to homogeneous systems. An observable is defined on the total system, this is a map : $\Phi : \mathbf{V}_N \to W$ where W is a finite dimensional vector subspace of $\mathbf{V}_N$, but not necessarily a tensorial vector product of spaces. To Φ is associated the self-adjoint operator $\widehat{\Phi} = \Upsilon \circ \Phi \circ \Upsilon^{-1}$ and $H_\Phi = \widehat{\Phi}\left(\otimes_{s=1}^N H\right) \subset \otimes_{s=1}^N H$.

Theorem 32 *Any observable of a homogeneous system is of the form :*

$\Phi : \mathbf{V}_N \to W$ *where W is generated by vectors* Φ_λ *associated to each class of conjugacy of* $\mathfrak{S}(N)$

The value of $\Phi(X_1 \otimes ... \otimes X_N) = \varphi(X_1, ..., X_N)\Phi_\lambda$ *where* φ *is a scalar linear symmetric map, if the system is in a state corresponding to* λ

Proof. The space W must be invariant by U and H_Φ invariant by $\widehat{U}$. If the system is in a state belonging to $\mathbf{h}$ for a class of conjugacy λ, then $H_\Phi = \widehat{\Phi}\mathbf{h}$ and $\left(\widehat{\Phi}\mathbf{h}, \widehat{U}\right)$ is an irreducible representation of the abelian subgroup $\mathfrak{S}(\lambda)$ corresponding to λ. It is necessarily unidimensional and $\Phi(X_1 \otimes ... \otimes X_N)$ is proportional to a unique vector. The observable being a linear map, the function φ is a linear map of the components of the tensor. ■

There is no way to estimate the state of each microsystem. From a practical point of view, this is a vector $\gamma = \widehat{\Phi}\left(\otimes_{n_1}\widetilde{\varepsilon}_{j_1} \otimes_{n_2} \widetilde{\varepsilon}_{j_2} ... \otimes_{n_p} \widetilde{\varepsilon}_{j_p}\right)$ which is measured, and from it $\lambda, (\widetilde{\varepsilon}_{j_k})_{k=1}^p$ are estimated.

In the probabilist picture the expected value of γ is :

$\langle\gamma\rangle = z(\pi_1, ..., \pi_N)$

with

$z(\pi_1, ..., \pi_N)$

$= \sum_\lambda \frac{N!}{n_1!...n_p!} \sum_{1 \leq j_1 \leq .. \leq j_p \leq N} (\pi_{j_1})^{n_1} ... (\pi_{j_p})^{n_p} \widehat{\Phi}\left(\otimes_{n_1}\varepsilon_{j_1} ... \otimes_{n_p} \varepsilon_{j_p}\right)$

We have a classic statistical problem : estimate the π_i from a statistic given by the measure of γ. If the statistic $\widehat{\Phi}$ is sufficient, meaning that π_i depends only on γ, as F is finite dimensional whatever the number of microsystems, the Pitman-Koopman-Darmois theorem tells us that the probability law is exponential, then an estimation by the maximum likehood gives the principle of Maximum Entropy with entropy :

$E = -\sum_{j=1}^N \pi_j \ln \pi_j$

In the usual interpretation of the probabilist picture, it is assumed that the state of each microsystem can be measured independently. Then the entropy $E = -\sum_{j=1}^N \pi_j \ln \pi_j$ can be seen as a measure of the heterogeneity of the system. And, contrary to a usual idea, the interactions between the micro-systems do not lead to the homogenization of their states, but to their quantization : the states are organized according to the classes of conjugacy.

7.6 Evolution of homogeneous systems

The evolution of homogeneous systems raises many interesting issues. The assumptions are a combination of the previous conditions.

Theorem 33 *For a model representing the evolution of a homogeneous system comprised of a fixed number N of microsystems s = 1 ...N which are represented by the same model, with variables* $(X_s)_{s=1}^N$ *such that, for each microsystem :*

i) the variables X_s *are maps :* $X_s :: R \to E$ *where* R *is an open subset of* $\mathbb{R}$ *and* E *a normed vector space, belonging to an open subset* O *of an infinite dimensional Fréchet space* V

ii) $\forall t \in R$ *the evaluation map :* $\mathcal{E}(t) : O \to E : \mathcal{E}(t) X_s = X_s(t)$ *is continuous*

iii) $\forall t \in R : X_s(t) = X'_s(t) \Rightarrow X_s = X'_s$

There is a map : $S : R \to \otimes_N F$ *such that* $S(t)$ *represents the state of the system at t.* $S(t)$ *takes its value in a vector space* $f(t)$ *such that* $\left(\mathbf{f}(t), \widehat{U}_F\right)$ *, where* $\widehat{U}_F$ *is the permutation on* $\otimes_N F$*, is an irreducible representation of* $\mathfrak{S}(N)$

The crucial point is that the homogeneity is understood as the microsystems follow the same laws, but at a given time they do not have necessarily the same state.

Proof. i) Implement the Theorem 2 for each microsystem : there is a common Hilbert space H associated to V and a continuous linear map $\Upsilon : V \to H ::$ $\psi_s = \Upsilon(X_s)$

ii) Implement the Theorem 31 on the homogeneous system, that is for the whole of its evolution. The state of the system is associated to a tensor $\Psi \in \mathbf{h}$ where $\mathbf{h}$ is defined by a Hilbertian basis $(\widetilde{\varepsilon}_i)_{i \in I}$ of H, a finite subset J of I, a conjugacy class λ and a family of p vectors $(\widetilde{\varepsilon}_{j_k})_{k=1}^p$ belonging to $(\widetilde{\varepsilon}_i)_{i \in J}$. The vector space $\mathbf{h}$ stays the same whatever t.

iii) Implement the Theorem 26 on the evolution of each microsystem : there is a common Hilbert space F, a map : $\widehat{\mathcal{E}} : R \to \mathcal{L}(H; F)$ such that : $\forall X_s \in O$: $\widehat{\mathcal{E}}(t) \Upsilon(X_s) = X_s(t)$ and $\forall t \in R$, $\widehat{\mathcal{E}}(t)$ is an isometry

Define $\forall i \in I : \varphi_i : R \to F :: \varphi_i(t) = \widehat{\mathcal{E}}(t) \widetilde{\varepsilon}_i$

iv) $\widehat{\mathcal{E}}(t)$ can be uniquely extended in a continuous linear map :

$\widehat{\mathcal{E}}_N(t) : \otimes_N H \to \otimes_N F$ such that : $\widehat{\mathcal{E}}_N(t)(\otimes_N \psi_s) = \otimes_N X_s(t)$

$\widehat{\mathcal{E}}_N(t)\left(\otimes_{s=1}^N \widetilde{\varepsilon}_{i_s}\right) = \otimes_{s=1}^N \varphi_{i_s}(t)$

$\widehat{\mathcal{E}}_N(t)$ is an isometry, so $\forall t \in R : \left\{\otimes_{s=1}^N \varphi_{i_s}(t), i_s \in I\right\}$ is a Hilbertian basis of $\otimes_N F$

v) Define as the state of the system at t : $S(t) = \widehat{\mathcal{E}}_N(t)(\Psi) \in \otimes_N F$

Define : $\forall \sigma \in \mathfrak{S}(N) : \widehat{U}_F(\sigma) \in \mathcal{L}(\otimes_N F; \otimes_N F)$ by linear extension of : $\widehat{U}_F(\sigma)\left(\otimes_{s=1}^N f_s\right) = \otimes_{s=1}^N f_{\sigma(s)}$

$\widehat{U}_F(\sigma)\left(\otimes_{s=1}^N \varphi_{i_s}(t)\right) = \otimes_{s=1}^N \varphi_{\sigma(i_s)}(t) = \widehat{\mathcal{E}}_N(t) \widehat{U}(\sigma)\left(\otimes_{s=1}^N \widetilde{\varepsilon}_{i_s}\right)$

$\forall \Psi \in \mathbf{h} : \Psi = \sum_{\sigma \in \mathfrak{S}(\lambda^c)} \Psi^\sigma \widehat{U}(\sigma)\left(\otimes_{n_1} \widetilde{\varepsilon}_{j_1} \otimes_{n_2} \widetilde{\varepsilon}_{j_2} ... \otimes_{n_p} \widetilde{\varepsilon}_{j_p}\right)$

$S(t) = \sum_{\sigma \in \mathfrak{S}(\lambda^c)} \Psi^\sigma \widehat{\mathcal{E}}_N(t) \circ \widehat{U}(\sigma)\left(\otimes_{n_1} \widetilde{\varepsilon}_{j_1} \otimes_{n_2} \widetilde{\varepsilon}_{j_2} ... \otimes_{n_p} \widetilde{\varepsilon}_{j_p}\right)$

$S(t) = \sum_{\sigma \in \mathfrak{S}(\lambda^c)} \Psi^\sigma \widehat{U}_F(\sigma) \otimes_{n_1} \varphi_{j_1}(t) \otimes_{n_2} \varphi_{j_2}(t) ... \otimes_{n_p} \varphi_{j_p}(t)$

$\forall \theta \in \mathfrak{S}(\lambda) : \widehat{U}_F(\theta)\left(\otimes_{n_1} \varphi_{j_1}(t) \otimes_{n_2} \varphi_{j_2}(t) ... \otimes_{n_p} \varphi_{j_p}(t)\right)$

$= \otimes_{n_1} \varphi_{j_1}(t) \otimes_{n_2} \varphi_{j_2}(t) ... \otimes_{n_p} \varphi_{j_p}(t)$

$\forall \theta \in \mathfrak{S}(\lambda^c) : \widehat{U}_F(\theta)\left(\otimes_{n_1} \varphi_{j_1}(t) \otimes_{n_2} \varphi_{j_2}(t) ... \otimes_{n_p} \varphi_{j_p}(t)\right)$

$\neq \left(\otimes_{n_1}\varphi_{j_1}(t)\otimes_{n_2}\varphi_{j_2}(t)...\otimes_{n_p}\varphi_{j_p}(t)\right)$
and the tensors are linearly independent

So $\left\{\widehat{U}_F(\sigma)\left(\otimes_{n_1}\varphi_{j_1}(t)\otimes_{n_2}\varphi_{j_2}(t)...\otimes_{n_p}\varphi_{j_p}(t)\right),\sigma\in\mathfrak{S}(\lambda^c)\right\}$ is an orthonormal basis of

$\mathbf{f}(t)=Span\left\{\widehat{U}_F(\sigma)\left(\otimes_{n_1}\varphi_{j_1}(t)\otimes_{n_2}\varphi_{j_2}(t)...\otimes_{n_p}\varphi_{j_p}(t)\right),\sigma\in\mathfrak{S}(\lambda^c)\right\}$

$\mathbf{f}(t)=\widehat{\mathcal{E}}_N(t)(\mathbf{h})$

Let $\widetilde{f}(t)\subset\mathbf{f}(t)$ be any subspace globally invariant by $\left\{\widehat{U}_F(\theta),\theta\in\mathfrak{S}(N)\right\}$: $\widehat{U}_F(\theta)\widetilde{f}(t)\in\widetilde{f}(t)$

$\widehat{\mathcal{E}}_N(t)$ is an isometry, thus a bijective map

$\widetilde{h}=\widehat{\mathcal{E}}_N(t)^{-1}\widetilde{f}(t)\Leftrightarrow\widetilde{f}(t)=\widehat{\mathcal{E}}_N(t)\widetilde{h}$

$\widehat{U}_F(\theta)\widehat{\mathcal{E}}_N(t)\widetilde{h}\in\widehat{\mathcal{E}}_N(t)\widetilde{h}$

$\forall\Psi\in\mathbf{h}:\widehat{U}_F(\theta)\widehat{\mathcal{E}}_N(t)\Psi=\widehat{\mathcal{E}}_N(t)\widehat{U}(\theta)\Psi$

$\Rightarrow\widehat{\mathcal{E}}_N(t)\widehat{U}(\theta)\widetilde{h}\in\widehat{\mathcal{E}}_N(t)\widetilde{h}$

$\Rightarrow\widehat{U}(\theta)\widetilde{h}\in\widetilde{h}$

So $\left(\mathbf{f}(t),\widehat{U}_F\right)$ is an irreducible representation of $\mathfrak{S}(N)$ ■

For each t the space $\mathbf{f}(t)$ is defined by a Hilbertian basis $(f_i)_{i\in I}$ of F, a finite subset J of I, a conjugacy class $\lambda(t)$ and a family of p vectors $(f_{j_k}(t))_{k=1}^p$ belonging to $(f_i)_{i\in J}$. The set J is arbitrary but defined by $\mathbf{h}$, so it does not depend on t**.** For a given class of conjugacy different families of vectors $(f_{j_k}(t))_{k=1}^p$ generate equivalent representations and isomorphic spaces, by symmetrization or antisymmetrization. So for a given system one can pick up a fixed ordered family $(f_j)_{j=1}^N$ of vectors in $(f_i)_{i\in I}$ such that for each class of conjugacy $\lambda=\{0\leq n_p\leq...\leq n_1\leq N,n_1+...n_p=N\}$ there is a unique vector space $\mathbf{f}_\lambda$ defined by $\otimes_{n_1}f_1\otimes_{n_2}f_2...\otimes_{n_p}f_p$. Then if $S(t)\in\mathbf{f}_\lambda$:

$S(t)=\sum_{\sigma\in\mathfrak{S}(\lambda^c)}S^\sigma(t)\widehat{U}_F(\sigma)\left(\otimes_{n_1}f_1\otimes_{n_2}f_2...\otimes_{n_p}f_p\right)$

and at all time $S(t)\in\otimes_N F_J$.

The vector spaces $\mathbf{f}_\lambda$ are orthogonal. With the orthogonal projection π_λ on $\mathbf{f}_\lambda$:

$\forall t\in R:S(t)=\sum_\lambda\pi_\lambda S(t)$

$\|S(t)\|^2=\sum_\lambda\|\pi_\lambda S(t)\|^2$

The distance between S(t) and a given $\mathbf{f}_\lambda$ is well defined and :

$\|S(t)-\pi_\lambda S(t)\|^2=\|S(t)\|^2-\|\pi_\lambda S(Ut)\|^2$

Whenever S, and thus Θ, is continuous, the space $\mathbf{f}_\lambda$ stays the same. As we have seen previously one can assume that, in all practical cases, Θ is continuous but for a countable set $\{t_k,k=1,2..\}$ of isolated points. Then the different spaces $\mathbf{f}_\lambda$ can be seen as phases, each of them associated with a class of conjugacy λ. And there are as many possible phases as classes of conjugacy. So, in a probabilist picture, one can assume that the probability for the system to be in a phase $\lambda:\Pr(S(t)\in\mathbf{f}_\lambda)$ is a function of $\frac{\|\pi_\lambda S(t)\|^2}{\|S(t)\|^2}$. It can be estimated as seen previously from data on a past period, with the knowledge of both λ and $\frac{\|\pi_\lambda S(t)\|^2}{\|S(t)\|^2}$.

8 CORRESPONDENCE WITH QUANTUM MECHANICS

It is useful to compare the results proven in the present paper to the axioms of QM as they are usually expressed.

8.1 Hilbert space

QM : 1. *The states of a physical system can be represented by rays in a complex Hilbert space H. Rays meaning that two vectors which differ by the product by a complex number of module 1 shall be considered as representing the same state.*

In Theorem 2 we have proven that in a model meeting precise conditions the states of the system can be represented as vectors in an infinite dimensional, separable, real Hilbert space. We have seen that it is always possible to endow the Hilbert space with a complex structure, but this is not a necessity. Moreover the Hilbert space is defined up to an isometry, so notably up to the product by a fixed complex scalar of module 1.

The state ψ (motion, kinematic and EM charge characteristics) of a particle can be represented in a fiber bundle, with fiber a complex vector space E. The gauge group for the EM field is $U(1)$ and the vectors ψ are defined up to a complex scalar of module 1. And this this the origin of rays (see more in J.C.Dutailly "Mathematics in Physics").

In Quantum Physics a great attention is given to the Principle of Superposition. This Principle is equivalent to the condition that the variables of the system (and then its state) belong to a vector space. There is a distinction between pure states, which correspond to actual measures, and mixed states which are linear combination of pure states, usually not actually observed. There has been a great effort to give a physical meaning to these mixed states. Here the concept of pure states appears only in the tensors representing interacting systems, with the usual, but clear, explanation. In Quantum Mechanics some states of a system cannot be achieved (through a preparation for instance) as a combination of other states, and then super-selection rules are required to sort out these specific states. Here there is a simple explanation : because the set H_0 is not the whole of H it can happen that a linear combination of states is not inside H_0. The remedy is to enlarge the model to account for other physical phenomena, if it appears that these states have a physical meaning.

Actually the main difference comes from the precise conditions of the Theorem 2. The variables must be maps, but also belong to a vector space. Thus for instance it does not apply to the model of a solid body represented by its trajectory $x(t)$ and its speed $v(t)$: the variable $x(t)$ is a map : $x : \mathbb{R} \to M$ valued in a manifold (an affine space in Galilean geometry). So it is necessary to adapt the model, using the fiber bundle formalism, and this leads to a deep redefinition of the concept of motion (including rotation) and to the spinors. And as it has

been abundantly said, the state is defined by maps over the evolution of the system, and not pointwise.

8.2 Observables

QM : 2. *To any physical measure Φ, called an observable, which can be done on the system, is associated a continuous, linear, self-adjoint operator $\widehat{\Phi}$ on H.*

We have proven that this operator is also compact and trace-class. The main result is that we have here a clear understanding of the concept of observable, rooted in the practical way the data are analyzed and assigned to the value of the variables, with the emphasize given to the procedure of specification, an essential step in any statistical analysis and which is usually overlooked. From primary observables it is possible to define von Neumann algebras of operators, which are necessarily commutative when a fixed basis has been chosen. As the choice of a privileged basis can always be done, one can say that there is always a commutative von Neumann algebra associated to a system. But, as it can be seen, these von Neumann algebras do not play any role in the proofs of the theorems. Their introduction can be useful, but they are not a keystone in our framework.

This is the opposite in the axiomatic interpretations of QM which define the system itself from the existence of a von Neumann algebra. However such interpretation is, eventually, based on the same assumption as any other interpretation of QM : the postulate that for any system there is a quantity, the state, which has a physical meaning and can be represented in a Hilbert space. Nothing preclude the choice of a privileged Hilbertian basis for this Hilbert space (as it is always done in any practical computation in QM), with respect to which the operators can be defined, and then the algebra is commutative. Which nullifies the emphasize given to the commutation of operators. Or at least it should be given another interpretation than the simultaneity of measures.

In QM a great emphasize if given to the commutation of observables, linked to the physical possibility to measure simultaneously two variables. This concept does not play any role here, for the strong reason that we consider maps with a domain over the whole extension, spatial and temporal, of the system, there is no assumption about how the measures are done, so the simultaneity of measures is not considered. In our picture the variables and their properties are the model, they are listed explicitly and it is assumed that there is some way to estimate their value, without any consideration of the time at which the measures are done. So the question of simultaneous measures does not arise, and the product of observables itself has no clear meaning and no use. If a variable is added, we have another model, the variable gets the same status as the others, and it is assumed that it can be measured.

8.3 Measure

QM : 3. *The result of any physical measure is one of the eigen-values λ of the associated operator $\widehat{\Phi}$. After the measure the system is in the state represented by the corresponding eigen vector ψ_λ*

This is one of the most puzzling axiom. We have here a clear interpretation of this result, with primary observables, and there is always a primary observable which is at least as efficient than a secondary observable.

In our picture there is no assumption about how the measures are done, and particularly if they have or not an impact on the state of the system. If it is assumed that this is the case, a specific variable should be added to the model. Its value can be measured directly or estimated from the value of the other variables, but this does not make a difference : it is a variable as the others. We will see an example in the following chapters.

There is no assumption about the times at which the measures are taken, when the model represents a process the measures can be taken at the beginning, during the process, or at the end. The variables which are estimated are maps, and the estimation of maps requires more than one value of the arguments. The estimation is done by a statistical method which uses all the available data. From this point of view our picture is closer to what is done in the laboratories, than to the idealized vision of simultaneous measures, which should be taken all together at each time, and would be impossible because of the perturbation caused by the measure.

Actually the importance granted to the simultaneity of measures, magnified by Dirac, is somewhat strange. It is also problematic in the Relativist picture. It is clear that some measures cannot be done, at the atomic scale, without disturbing the state of the system that is studied, but this does not preclude to use the corresponding variables in a model, or give them a special status. Before the invention of radar the artillerymen used efficient models even if they were not able to measure the speed of their shells. And in a collider it is assumed that the speed and the location of particles are known when they collide.

8.4 Probability

QM : 4. *The probability that the measure is λ is equal to $|\langle\psi_\lambda,\psi\rangle|^2$ (with normalized eigen vectors). If a system is in a state represented by a normalized vector ψ , and an experiment is done to test whether it is in one of the states $(\psi_n)_{n=1}^N$ which constitutes an orthonormal set of vectors, then the probability of finding the system in the state ψ_n is $|\langle\psi_n,\psi\rangle|^2$.*

The first part is addressed by the theorem 17. The second part has no direct equivalent in our picture but can be interpreted as follows : a measure of the primary observable has shown that $\psi \in H_J$, then the probability that it belongs

to $H_{J'}$ for any subset $J' \subset J$ is $\left\|\widehat{Y}_{J'}(\psi)\right\|^2$. It is a computation of conditional probabilities :

Proof. The probability that $\psi \in H_K$ for any susbset $K \subset I$ is $\left\|\widehat{Y}_K(\psi)\right\|^2$. The probability that $\psi \in H_{J'}$ knowing that $\psi \in H_J$ is :

$\Pr(\psi \in H_{J'} | \psi \in H_J) = \frac{\Pr(\psi \in H_{J'} \wedge \psi \in H_J)}{\Pr(\psi \in H_{J'} | \psi \in H_J)} = \frac{\Pr(\psi \in H_{J'})}{\Pr(\psi \in H_{J'} | \psi \in H_J)} = \frac{\left\|\widehat{Y}_{J'}(\psi)\right\|^2}{\left\|\widehat{Y}_J(\psi)\right\|^2} = \left\|\widehat{Y}_{J'}(\psi)\right\|^2$ because $\widehat{Y}_{J'}(\psi) = \psi$ and $\|\psi\| = 1$ ■

Moreover we have seen how the concept of wave functions can be introduced, and its meaning, for models where the variables are maps defined on the same set. Of course the possibility to define such a function does not imply that it is related to a physical phenomenon.

8.5 Interacting systems

QM : 5. *When two systems interacts, the vectors representing the states belong to the tensorial product of the Hilbert states.*

This is the topic of the theorem 28. We have seen how it can be extended to N systems, and the consequences that entails for homogeneous systems. If the number of microsystems is not fixed, the formalism of Fock spaces can be used but would require a mathematical apparatus that is beyond the scope of this book.

There is a fierce debate about the issue of locality in physics, mainly related to the entanglement of states for interacting particles. It should be clear that the formal system that we have built is global : more so, it is its main asset. While most of the physical theories are local, with the tools which have been presented we can deal with variables which are global, and get some strong results without many assumptions regarding the local laws.

8.6 Wigner's theorem

QM : 6. *If the same state is represented by two rays R, R', then there is an operator $\widehat{U}$, unitary or antiunitary, on the Hilbert space H such that if the state ψ is in the ray R then $\widehat{U}\psi$ is in the ray R'.*

This the topic of the theorem 21. The issue unitary / antiunitary exists in the usual presentation of QM because of the rays. In our picture the operator is necessarily unitary, which is actually usually the case.

8.7 Schrödinger equation

QM : 7. *The vector representing the state of a system which evolves with time follows the equation : $i\hbar\frac{\partial\psi}{\partial t} = \widehat{H}\psi$* where *$\widehat{H}$ is the Hamiltonian of the system.*

This is actually the topic of the theorem 27 and the result holds for the variables X in specific conditions, including in the General Relativity context. The imaginary i does not appear because the Hilbert space is real. As for Planck's constant of course it cannot appear in a formal model. However as said before all quantities must be dimensionless, as it is obvious in the equivalent expression $\psi(t) = \exp \frac{t}{i\hbar} \widehat{H} \psi(0)$. Thus it is necessary either to involve some constant, or that all quantities (including the time t) are expressed in a universal system of units. This is commonly done by using the Planck's system of units. Which is more important is that the theorems (and notably the second) precise fairly strong conditions for their validity. In many cases the Schrödinger's equation, because of its linearity, seems "to good to be true". We can see why.

8.8 The scale issue

The results presented here hold whenever the model meets the conditions 1. So it is valid whatever the scale. But it is clear that the conditions are not met in many models used in classic physics, notably in Analytic Mechanics (the variables q are not vectorial quantities). Moreover actually in the other cases it can often be assumed that the variables belong themselves to Hilbert spaces. The results about observables and eigen values are then obvious, and those about the evolution of the system, for interacting systems or for gauge theories keep all their interest.

The "Quantic World", with its strange properties does not come from specific physical laws, which would appear below some scale, but from the physical properties of the atomic world themselves. And of course these cannot be addressed in the simple study of formal models.

So the results presented here, which are purely mathematical, give a consistent and satisfying explanation of the basic axioms of Quantum Mechanics, without the need for any exotic assumptions. They validate, and in many ways make simpler and safer, the use of techniques used for many years. Moreover, as it is easy to check, most of these results do not involve any physics at all : they hold for any scientific theory which is expressed in a mathematical formalism. From my point of view they bring a definitive answer to the issue of the interpretation of QM : the interpretations were sought in the physical world, but actually there is no such interpretation to be found. There is no physical interpretation because QM is not a physical theory.

The results presented go beyond the usual axioms of QM : on the conditions to detect an anomaly, on the quantization of a variable $Y = f(X)$, on the phases transitions. And other results can probably be found. So the method should give a fresh view of the foundations of QM in Physics.

jc.dutailly@free.fr

9 BIBLIOGRAPHY

R.D.Anderson *Some open questions in infinite dimensional topology* Proceeding of the 3d Prague symposium Praha (1972)

H.Araki *Mathematical theory of quantum fields* Oxford Science Publications (2000)

J.C. Baez, M.Stay *Physics, Topology, Logic and Computation: A Rosetta Stone* arXiv 0903.0340 (2009)

H.Baumgartel *Operator Algebraic methods in Quantum Fields* Berlin Akademie verl. (1995)

M.Le Bellac *Physique quantique* CNRS (2003)

A.Berlinet, C.Thomas-Agnan *Reproducing kernel, Hilbert spaces in probability and statistics* Springer (2004)

A.Bird *Philosophy of science* Rootledge (1998)

J.D.Bjorken,S.D.Drell *Relativistic quantum fields* Mc Graw Hill (1965)

N.Bogolubov,A.A.Logunov, A.I.Ossak, I.T.Todorov *General principles of quantum fields theory* Kluwer (1990)

O.Bratelli, D.W.Robinson *Operators algebras and quantum statistical mechanics* Springer (2002)

B.Coecke, E.O.Paquette *Categories for the practising physicist* arXiv:0905-3010v1 [quant-ph] (16 may 2009)

B.Coecke *New Structures for Physics* Lecture Notes in Physics vol. 813, Springer, Berlin, 2011

B.d'Espagnat *Reality and the physicist* Cambridge University Press (1989)

P.A.M.Dirac *The principles of Quantum Mechanics* Oxford Science Publications (1958)

J.C.Dutailly *Mathematics for theoretical physics* arXiv:1209-5665v2 [math-ph] (4 feb 2014)

J.C.Dutailly *Estimation of the probability of transition between phases* CNRS (http://hal.archives-ouvertes.fr/hal-01075940, 20 october 2014)

J.C.Dutailly *Mathematics in Physics* CNRS (http://hal.archives-ouvertes.fr/hal-01169985, 29 juin 2015)

R.P.Feynman, A.R.Hibbs *Quantum Mechanics and Path Integrals* Dover (2005)

J.Finne *Asymptotic study of canonical correlation analysis: from matrix and analytic approach to operator and tensor approach* SORT 27 (2) July-December 2003, 165-174

Francis C.E.H *A construction of full QED using finite dimensional Hilbert space* EJTP 10 N°28 (2013)

Francis C.E.H *The Hilbert space of conditional clauses* arXiv:1205-4607 (2013)

Tepper L.Gill, G.R.Pantsulaia, W.W.Zachary *Constructive analysis in infinitely many variables* arXiv 1206-1764v2 [math-FA] (26 june 2012)

H.Halvorson *Algebraic quantum fields theory* arXiv:math-ph/0602036v1 14 feb 2006

N.R. Hansen *Non-parametric likelihood based estimation of linear filters for point processes* arXiv:1304-0503v3 [stat:CO] (12 feb 2014)

D.W.Henderson *Infinite dimensional manifolds are open subsets of Hilbert spaces* (1969) Internet paper

S.S.Horuzhy *Introduction to algebraic quantum field theory* Riedel (1989)

J.M.Jauch *Foundations of Quantum Mechanics* AddisonWesley (1968)

Sir M.Kendall, A.Stuart *The advanced theory of statistics* Charles Griffin & Co (1977)

A.W.Knapp *Lie groups beyond an introduction* Birkhäuser (2005)

F.Laloë *Comprenons-nous vraiment la mécanique quantique ?* CNRS Editions (2011)

E.H.Lieb, M.Loss *Analysis* American Mathematical Society (2000)

R.Haag *Local quantum physics* 2nd Ed.Springer (1991)

G.Mackey *The mathematical fundations of Quantum Mechanics* W.A.Benjamin (1963)

J.von Neumann *Mathematical Foundations of Quantum Mechanics*, Beyer, R. T., trans., Princeton Univ. Press. (1996 edition)

R.Omnès *The interpretation of quantum mechanics* Princeton (1994)

J. E. Palomar Tarancon *Conceptual systems, conceptual convergence theory and and algebras of analogies*

K.Popper *Quantum theory and the schism in physics* Routledge (1982)

K.Popper *The logic of scientific discovery* Rootledge (1959)

I.Schnaid *Wave function perturbations propagation in multi particles system* arXiv 1307.2510v1 [physics-gen.ph] (9 july 2013)

A.Smola, A.Gretton, L.Song, B.Schölkop *A Hilbert Space Embedding for Distributions*

H.Torunczyk *Characterizing Hilbert spaces topology* Fundamental mathematica (1981)

A. Vourdas *The complete Heyting algebra of subsystems and contextuality* arXiv:1310-3604v1 [quant-ph] (14 oct 2013)

S.Weinberg *The quantum theory of fields* Cambridge University Press (1995)

S.Weinberg *Dreams of a Final Theory* Pantheon Books (1992)

H.Weyl *The theory of groups and quantum mechanics* Dover (1931 / 1950)