

# Putting perspectives in context: combining measurements in quantum mechanics and data pooling in classical probability

Sven Aerts and Diederik Aerts

*Center Leo Apostel for Interdisciplinary Studies (CLEA)  
and Foundations of the Exact Sciences (FUND)*

*Department of Mathematics, Vrije Universiteit Brussel  
1160 Brussels, Belgium*

E-Mails: [saerts@vub.ac.be](mailto:saerts@vub.ac.be), [diraerts@vub.ac.be](mailto:diraerts@vub.ac.be)

Franklin E. Schroeck

*Department of Mathematics*

*John Greene Hall, 2360 S. Gaylord St.*

*University of Denver, Denver, USA*

E-Mail: [fschroec@du.edu](mailto:fschroec@du.edu)

Jürgen Sachs

*Electronic Measurement Lab*

*Technische Universität Ilmenau*

*PSF 100 565, D-98684 Ilmenau, Germany*

E-Mail: [Juergen.Sachs@tu-ilmenau.de](mailto:Juergen.Sachs@tu-ilmenau.de)

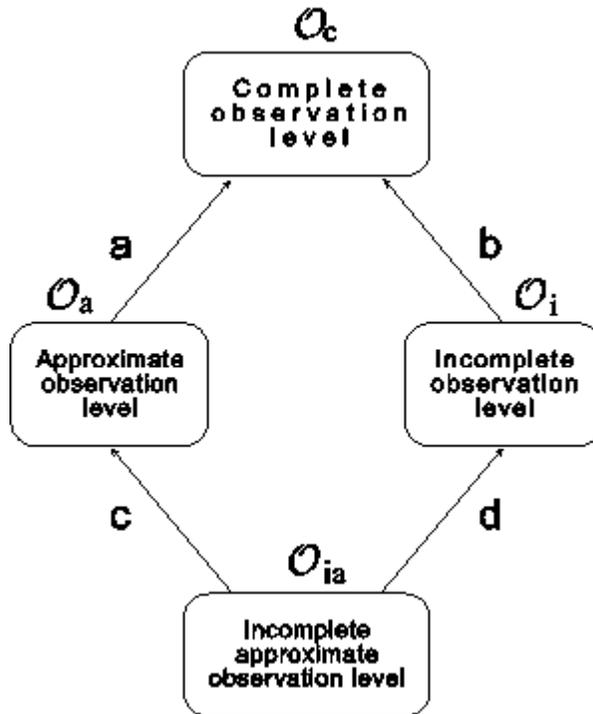
## Abstract

In a former contribution [1] we proposed the idea that both in quantum mechanics as well as in signal analysis, the mathematical framework is best understood as arising from the necessity to combine various, sometimes incompatible, measurements into a single state of the entity under investigation. In this paper we investigate the physical basic settings and the corresponding various procedures that can be applied to reconstruct the state from measurements. We deal with different levels of information, in which limited measurement accuracy and limited numbers of measurable observable quantities restrict the possibility of perfect state reconstruction.

Keywords: data pooling, state tomography, quantum probability, signal analysis, incompatible observables, observation, global view

## 1 Introduction

Knowledge of any given phenomenon starts with observation, but even a single phenomenon may be observed from various perspectives and various aspects of the same phenomenon can be brought to the foreground. The phenomenon can be measured in varying degrees of precision and accuracy. The problem is then: how do we unite these various perspectives on the same phenomenon? Suppose that two observers have measured the time of the setting of the sun, and observer  $A$  obtains the result is 5:53 PM, while observer  $B$ 's result is 6 PM. To avoid unnecessary complications, we assume throughout this article that both observers are sincere, observe in a scientifically legitimate way, and use the same system of units to express their results. One generally expects an increase of information about the setting of the sun, as a result of having access to two observations. So how are we to understand the different outcomes they obtain? There are basically two situations in which this might have happened. In a first scenario, the measurement outcome was really only an estimate. The fact that the outcome was only an estimate, means that there is uncertainty associated with it, which might well be bigger than the difference between the outcomes. In a second scenario, the uncertainty associated with the outcome may be very low, but the outcomes may deal with different perspectives on the same phenomenon. For example, the two observations of the setting of the sun may have occurred at different locations, or on different days. Despite the apparent differences in the result, both observations can claim legitimacy. In most instance single measurements will not suffice to characterize the state of the system under study. What observable quantities do we need to make sure that we have a complete description of the system? The answer that question, one can use Prugovečki's concept of *informational completeness* [2], which was originally proposed within the setting of quantum mechanics, but is valuable for any system with a generalization of the concept of state. A set of observable is called *informational complete* if there is only *one* state compatible with the probabilities obtained from it. Another name for an informational complete set of observables, is a *quorum* [3]. This means that for an informational complete set of measurements, the state can be uniquely derived from the probability distributions of these observables.



Four levels of observation.  $\mathcal{O}_c$  represents the ideal situation. Fundamental physical theories deal with this level.  $\mathcal{O}_{ia}$ , on the other hand, represents the incomplete and approximate observation level, and is the situation that prevails in most experiments. In between we find the approximate ( $\mathcal{O}_a$ ) and incomplete ( $\mathcal{O}_i$ ) observation levels. The issue we are concerned with here, is how to get from the bottom level to the top level.

Moreover, the expectation value of *any* observable (*i.e.* any point of view) can be calculated as a function of the outcome probabilities of the informational complete set [4]. We will examine how to combine diverse forms of information in a sound way within a classical and quantum setting. To facilitate the discussion we will on a few occasions differentiate between four different levels of observational information. A first level is such that we can reconstruct the complete state of the system under study. Following [5], we will call this the *complete observation level*. It implies infinitely precise information about all relevant observable quantities that pertain to the system. A second level of observation, is when we possess the probability distributions, and with arbitrary precision, for some of the observables, but not for all the necessary ones to determine the state. This level will be called the *incomplete observation level*. A third level of observation, is when

we have information about all observables that characterize the system, but we have only approximations of the corresponding probability distributions. We will call this the *approximate observation level*. Lastly, we can consider the realistic case, in which we have only approximations of a restricted set of observables. This last case will be called the *incomplete, approximate observation level*. To go from  $\mathcal{O}_{ia}$  to  $\mathcal{O}_c$  directly is not feasible. In practice, to go from  $\mathcal{O}_{ia}$  to  $\mathcal{O}_i$ , (arrow d), we need to perform more measurements of the *same* kind. To go from  $\mathcal{O}_{ia}$  to  $\mathcal{O}_a$ , (arrow c) we need to perform more measurements of a *different* kind. Of course, to go from  $\mathcal{O}_a$  to  $\mathcal{O}_c$ , we can also perform more measurements of the *same* kind (or of a different kind for going from  $\mathcal{O}_i$  to  $\mathcal{O}_c$ .) However in these last two cases there also exist theoretical alternatives that seek to obtain an approximation to  $\mathcal{O}_c$ , some of which will be treated in this paper. However, we first need to explain what different kinds of states there are and how they relate to physical situations.

## 2 Systems, states and properties

Before we can understand what it takes to combine measurements, we need a general framework to describe the system we seek to observe. Let us first abstract the system that we wish to study and call it  $S$ . If possible at all, we would like to know everything there is to know about  $S$ . We will take the metaphysical stance that a system is characterized fully by its state. This means that  $S$  will always be in a state  $q$  (states will be denoted by small roman script  $p, q, r, \dots$ ) taken from a set of states  $\Sigma_S$  of the system  $S$ . This state, in turn, is determined by the properties that are actual for that particular state. Loosely speaking, a property of a system is something that can we attribute to that system with a certain persistency. Birkhoff and von Neumann formalized the characterization of a system through its properties, starting with quantities which yield only one of two possible results, and which are testable in a reproducible sense. They called these quantities ‘experimental propositions’. This setting also forms the operational basis of the Geneva-Brussels approach, in which a property is introduced as an equivalence class of tests (also called questions or experimental projects), and a state is identified with the set of actual properties. We will introduce these notions through the state-property spaces as in [6]. Consider a system  $S$  with its (non-empty) set of states  $\Sigma_S$  and (non-empty) set of properties  $\mathcal{L}_S$ . Properties shall be denoted by small bold script  $\mathbf{a}, \mathbf{b}, \mathbf{c}, \dots$ . Depending on the state of the system, some of the properties may be actual, and some not. When we say an entity has an actual property  $\mathbf{a}$ , we tacitly assume

that for every property  $\mathbf{a}$  in  $\mathcal{L}_S$  there exists a test, let us call it  $\alpha$ , that tests property  $\mathbf{a}$ . The result of the test is a simple ‘yes’ or ‘no’. If the test yields ‘yes’ with *certainty*, we say property  $\mathbf{a}$  is *actual*.

**definition:** An entity  $S$  in the state  $p$ , is said to have *an actual property*  $\mathbf{a}$  iff the test  $\alpha$  that tests property  $\mathbf{a}$  yields ‘yes’ with certainty.

Two tests  $\alpha$  and  $\beta$  will be called *equivalent* if for any state  $p$  where  $\alpha$  gives with certainty ‘yes’, we also have that  $\beta$  gives with certainty ‘yes’, and vice versa. We will say equivalent tests, test the same property. Vice versa, a property is identified with an equivalence class of tests. A property is actual when any test in the equivalence class yields ‘yes’ with certainty, because all tests will then yield ‘yes’ with certainty. A given property  $\mathbf{a} \in \mathcal{L}_S$  may be actual for some of the states  $p \in \Sigma_S$  of the entity, but not necessarily for all. With the usual notation  $\mathcal{P}(\Sigma_S)$  for the set of all subsets of  $\Sigma_S$ , we postulate the map  $\kappa_S : \mathcal{L}_S \rightarrow \mathcal{P}(\Sigma_S)$ , called the *Cartan map*, such that  $\kappa_S(\mathbf{a})$  is the set of states in  $\Sigma_S$  for which the property  $\mathbf{a}$  is actual. The triple  $(\Sigma_S, \mathcal{L}_S, \kappa_S)$  is called a *state property space* [6] and fully characterizes what can be known about the entity with certainty.

**definition:** The triple  $(\Sigma_S, \mathcal{L}_S, \kappa_S)$ , called a *state property space*, consists of two sets  $\Sigma_S$  and  $\mathcal{L}_S$  (where  $\Sigma_S$  is the set of states of a physical entity  $S$ , and  $\mathcal{L}_S$  its set of properties), and a function  $\kappa_S : \mathcal{L}_S \rightarrow \mathcal{P}(\Sigma_S)$ , such that  $\mathbf{a} \in \mathcal{L}_S$  is actual for the entity in a state  $p$  iff  $p \in \kappa_S(\mathbf{a})$ .

We note that in the development of the Geneva-Brussels approach, further axioms are imposed on  $\mathcal{L}_S$  so that the properties of a general physical system form a complete, atomistic and orthomodular lattice [6]. The state determination axiom in this approach says that the state of the system is then the collection of all properties that are actual. The main incentive to introduce this notion of a state via the Geneva-Brussels formalism, is that it gives us a general operational characterization of what it means “to be in a state”, and how this fact relates to actual properties. We emphasize that the definitions apply equally well in the quantum and the classical domain. We will later on argue that actual experiments do not measure the properties directly, and take recourse to positive operator valued measures to represent measurements. Nevertheless, the idealized concept of property can still be regarded as a limiting case of such an approach.

## 2.1 Properties and Observables

We are often interested in assigning a numerical value to a quantity of interest. We will call such a situation the measurement of an observable (quantity)  $A$ . Obviously, measuring an observable is more general than measuring whether a property holds, because a property can be regarded as a physical quantity that takes only two possible values: ‘yes’ or ‘no’. However, the range of values the observable  $A$  can take in a measurement, *the outcome set*  $X_A$  associated with  $A$ , can be divided into subsets  $X^i$ . We associate with each of these subsets  $X_A^i \subset X_A$  the property:

$$\mathbf{a}_i : A \text{ has a value in the subset } X^i$$

With the help of such a construction, it is easily recognized that an arbitrary observable  $A$  can be approximated by a set of properties  $\{\mathbf{a}_i\}$ .

## 2.2 Classical states and properties

First we note that with ‘classical case’, we do not necessarily mean a point particle governed by Newtonian mechanics, but rather the characterization of a system by means of measurements which, in principle, have an error that only depends on the number of measurements we care to perform. In essence this means that all observables have simultaneously well-defined values at each instant, whether this is a thermodynamic system, or the kicked classical top. In classical mechanics, the state space of a system, is represented by the phase space  $\Omega$ . The set of properties for a classical system, is then the set of all subsets of  $\Omega$ , which we denote as  $\mathcal{P}(\Omega)$ . The Cartan map is in this case simply the identity, because a classical system either has the property (if its state is in a given subset  $\Delta \in \mathcal{P}(\Omega)$ ) or it does not have that property (its state is in  $\Omega \setminus \Delta \in \mathcal{P}(\Omega)$ ). The state property system for a classical system, is then

$$(\Omega, \mathcal{P}(\Omega), \text{id})$$

## 2.3 Quantum states and properties

In quantum mechanics, a single system is generally assumed to be in a *pure state*, represented by a square integrable (unit norm) element of a Hilbert space over the field of complex numbers:

$$\Sigma_S^{QM} = \{\psi \in \mathcal{H}(\mathbb{C}) : \|\psi\| = 1\}$$

Here  $\|\cdot\|$  denotes the usual Hilbert space norm, defined as the sesquilinear product of an element of  $\mathcal{H}(\mathbb{C})$  with itself. A yes/no experiment to test a

given property on the system  $S$  in a state  $\psi$ , is represented by a projection operator. The range of the projection operator is a closed subspace of  $\mathcal{H}(\mathbb{C})$ . If we denote the set of all closed subspaces of  $\mathcal{H}(\mathbb{C})$  by  $\mathcal{P}(\mathcal{H})$ , we have:

$$\mathcal{L}_S^{QM} = \mathcal{P}(\mathcal{H})$$

A property of a system, is then given by the closed subspace that is the range of the projection operator, and the Cartan map for a property  $\mathbf{a}$ , is the set of states that actualize this property

$$\kappa_S(\mathbf{a}) = \{\psi \in \Sigma_S^{QM} : \psi \in \mathbf{a}\}$$

So for a single system, we have that the state property system  $(\Sigma_S, \mathcal{L}_S, \kappa_S)$  is given by  $(\Sigma_S^{QM}, \mathcal{P}(\mathcal{H}), \kappa_S)$ . As an example, one can think of a spin 1/2 system, with state space  $\mathcal{H}(\mathbb{C}^2)$ , and with a spin state prepared in the  $z$  direction. The “spin property” is then defined for each direction that we choose to place our Stern-Gerlach apparatus to measure the spin. The test that corresponds to that property, is effected by projecting the state onto the relevant direction. Of all these properties (directions), only one is actual, and this coincides with the one dimensional closed subspace in which the state  $\psi$  lies.

### 3 Uncertainty

#### 3.1 Lack of knowledge about the state, classical case

In many instances of practical interest, we do not have precise knowledge about the state of the system, or we have an ensemble of systems with more or less similar states. This situation of lack of knowledge can be handled by probability theory. The state space is then equipped with a probability density, such that to each point in phase space, the probability density indicates the probability that the state of the system is in fact that point of phase space. Suppose for simplicity, that we have an ensemble consisting of two systems of the same type, but with different states  $\varphi_{S_1}$  and  $\phi_{S_2}$ , in relative proportions  $c$  and  $1 - c$ . A perfect measurement of property  $\mathbf{a}$  reveals outcome  $x_1$  and for system 2,  $x_2$ . Repeated application of a measurement procedure designed to measure observable  $A$ , will then yield outcome  $x_1$  with a relative frequency approaching  $c$  and outcome  $x_2$  with a relative frequency converging to  $1 - c$ . In this way we can obtain perfect knowledge about the relative fractions of  $\varphi_{S_1}$  and  $\phi_{S_2}$  in the ensemble. This point may seem trivial, but we will see no such perfect decomposition is possible in the

quantum case, unless we already know in advance of just which two types of states the mixture consists.

### 3.2 Lack of knowledge about the measurement, classical case

If the measurements are 100% faithful, we can perfectly decompose the ensemble into its constituents. Data pooling in this setting is trivial. If two measurements pertain to the same property, measured on the same system, they a fortiori have to yield the same outcome. If two measurements pertain to the same system, but to different properties  $\mathbf{a} \in \mathcal{P}(\Omega)$ ,  $\mathbf{b} \in \mathcal{P}(\Omega)$ , then the true state of the system is somewhere in  $\mathbf{a} \cap \mathbf{b} \in \mathcal{P}(\Omega)$ . However, in practice, measurements have a limited accuracy and precision. A repeated application of a measurement, designed to measure an observable quantity  $A$ , on a system  $S$  in the state  $\varphi$  yields a discrete set of outcomes  $X_A$ . This set is partitioned into outcome bins  $X_A^i$ . For each bin we count the number of measurements that result in an outcome within that bin. In the limit of an infinite number of measurements, this converges to the probability

$$P(x \in X_A^i | \varphi)$$

One can think of the measurement of the length of a given piece of very thin, very inelastic rope. Because the rope is thin and inelastic, its length is well-defined. Because we measure the same piece of rope repeatedly, there is no change in the state of the rope. Nevertheless, if we attempt to measure this length a number of times, we will find that the outcome fluctuates at least in the smallest digit the measuring rod affords. We see that even the classical case, and with measurements performed on the same entity in the same state, it can make sense to associate probability distributions to the state of the system. Obviously, these distributions refer to the process we used to obtain knowledge rather than to the system itself. In fact, this type of lack of knowledge was conjectured to be the source of quantum indeterminism in the so-called hidden-measurement approach [7], [8].

### 3.3 Lack of knowledge about the state, quantum case

If we cannot create an ensemble of identical system states, the best we can do, is to create an ensemble of which the members have pure states not too far from each other. In the discrete case, we have an ensemble of states  $\{\psi_i\}$ , each member occurring in a relative fraction  $c_i$  in the ensemble. This is usually represented by means of a density operator  $\rho$  which is a member

of the linear operators  $L(\mathcal{H})$  that act on elements of the Hilbert space  $\mathcal{H}$ . Moreover,  $\rho$  needs to be normalized, self-adjoint and positive:

$$\Sigma_S^{QM} = \{\rho \in L(\mathcal{H}) : Tr(\rho) = 1, \rho = \rho^\dagger, \langle \varphi | \rho | \varphi \rangle \geq 0\} \quad (1)$$

In this definition, we have used the standard Dirac notation, in which vectors in  $\mathcal{H}$  are denoted as  $|\cdot\rangle$  and called *ket* vectors. The space dual to  $\mathcal{H}$  contains the linear functionals on  $\mathcal{H}$  and its elements are called *bra* vectors and denoted by  $\langle \cdot |$ . The inner product of a *bra* with a *ket* (a ‘bracket’ written as  $\langle \varphi | \phi \rangle$ ) yields a complex number, the outer product, written as  $|\phi\rangle\langle \varphi|$  forms an operator. The outer product of an element with its own dual,  $|\varphi\rangle\langle \varphi|$ , is the projection operator onto the one dimensional subspace that coincides with the original vector  $\varphi$ . With these basics, one can show that  $\rho$ , being a self-adjoint operator, has a spectral decomposition

$$\rho = \sum_i c_i |\varphi_i\rangle\langle \varphi_i| \quad (2)$$

The conditions imposed on  $\rho$  can also be translated into conditions on the  $c_i$  :  $\sum c_i = 1$ ,  $c = c_i^*$  and  $c_i \geq 0$ . In the most simple case, there is only one  $c_i$  different from zero, and hence, from the condition  $\sum c_i = 1$ , we have that  $c_i = 1$  and

$$\rho = |\varphi\rangle\langle \varphi|$$

A density operator that can be written in this form, with  $|\varphi\rangle$  unit norm, is said to represent a pure state. The reason for the terminology *pure*, is that one cannot further decompose such a state. A necessary and sufficient condition for this to be the case, is idempotency of the density operator  $\rho^2 = \rho$ . In this case  $\rho$  is a projection operator onto the one dimensional subspace containing  $|\varphi\rangle$ . If in the spectral decomposition of a density operator we have two  $c_i$ ’s different from zero, we can write:

$$\rho = c|\varphi_0\rangle\langle \varphi_0| + (1 - c)|\varphi_1\rangle\langle \varphi_1| \quad (3)$$

If we form a statistical mixture of two density operators which have support on the same Hilbert space

$$\rho = c\rho_0 + (1 - c)\rho_1 \quad (4)$$

then we obtain a new quantity that satisfies all requirements to classify as a density operator. More generally, a convex combination of an arbitrary number of density operators with the same supporting Hilbert space, is a

new density operator on that same Hilbert space. It is tempting to interpret such a  $\rho$  as a classical mixture of two pure states,  $|\varphi_0\rangle$  with fraction  $c$  and  $|\varphi_1\rangle$  with fraction  $1 - c$ . In the more general case, a set of pure states  $\{|\varphi_i\rangle\}$  can be thought to generate the general state (2) by forming the appropriate convex combination (with the  $c_i$ ) of the projectors associated with each member  $|\varphi_i\rangle$ . The ensemble formed in this way, can indeed be represented by a density operator of the form (2) and vice versa, a system described by  $\rho$  with spectral density (2), cannot be distinguished from the members of an ensemble formed by such a mixing procedure. It should however be kept firmly in mind that there are many convex combinations of sets of pure vectors that yield identical density operators [9]. An observable quantity  $\mathcal{A}$  is represented by a linear self-adjoint operator  $A$ . For simplicity we consider only operators with a discrete spectrum. Then the spectral decomposition of the operator  $A$  reads:

$$A = \sum a_n |a_n\rangle\langle a_n|$$

We have, as is customary, given the same symbol to the eigenvalue and the eigenvector, the real nature of the quantity revealing itself by having brackets or not. The set of possible outcomes for the measurement of observable  $A$ , is the set of eigenvalues of  $A$ . The average value of the observable is then given by

$$\langle A \rangle = Tr(A\rho)$$

where  $Tr$  denotes the trace of the operator  $A\rho$ .

## 4 State reconstruction with complete knowledge

### 4.1 Classical state reconstruction in the ideal case.

Suppose we can perfectly isolate an ensemble of a large number of identical states and perform perfect measurements on this ensemble. By perfect measurement, we mean a measurement without error associated with it. How would we proceed to infer the state of the members in the ensemble from the outcomes of such measurements? In the classical case, the scenario for such a reconstruction is rather trivial. Because the measurement is assumed perfect, two measurements of identical observables, performed on identical states, will yield identical results. Suppose we have determined the value of observable  $A$  and determined it to be  $a$ . This represents our first piece of information and is called, in accordance with [5], the first observational

level. Then we know that the state of the system is in that part of the state space that is compatible with this observation:

$$\Sigma_S^{\{A=a\}} = \{\phi \in \Sigma_S : A = a\}$$

If we have measured a second observable, say  $B$  with the value  $b$ , then we have

$$\Sigma_S^{\{B=b\}} = \{\phi \in \Sigma_S : B = b\}$$

The data pooling of  $a$  and  $b$ , is then completely determined by the subset of the state space compatible with both observations:

$$\Sigma_S^{\{A=a, B=b\}} = \{\phi \in \Sigma_S : A = a, B = b\}$$

This state space  $\Sigma_S^{\{A=a, B=b\}}$  is defined by the set-theoretic intersection of  $\Sigma_S^{\{A=a\}}$  and  $\Sigma_S^{\{B=b\}}$ :

$$\Sigma_S^{\{A=a, B=b\}} = \Sigma_S^{\{A=a\}} \cap \Sigma_S^{\{B=b\}}$$

In case the system is a point particle, we need to observe 7 observable quantities: its three position coordinates, its three momentum coordinates, and its mass. Each additional observable will refine the state space, until we reach a singleton, which coincides with the state:

$$\phi = \Sigma_S^{\{X=x, Y=y, Z=z, P_x=p_x, P_y=p_y, P_z=p_z, M=m\}}$$

## 4.2 Quantum state reconstruction in the ideal case

In a quantum mechanical experiment, the occurrence of a single outcome has very little meaning. In fact, quantum theory does not provide us with a mechanism for the production of an outcome. The theory, as used in its standard ‘forward’ sense, allows one to derive probability distributions when the state is known. We are dealing here with the inverse problem, because we seek to estimate the state from sets of outcomes. In the ideal case, we assume that sufficient measurements have been performed so that we have arbitrarily good approximations to the probability distributions for every relevant observable. This then corresponds to the complete observation level  $\mathcal{O}_c$ . It is a basic postulate of quantum mechanics that for each observable  $A_i$  with possible outcomes  $a_j$  and corresponding eigenvectors, the probability of obtaining the outcome  $a_k$  if the system is in the state  $\psi$ , is given by:

$$P(A_i = a_j | \psi) = |\langle a_j | \psi \rangle|^2$$

Suppose we have a pure state in an  $n$  dimensional Hilbert space written in some basis  $\{|c_k\rangle, k = 1, \dots, n\}$

$$|\psi\rangle = \sum_{k=1}^n c_k |c_k\rangle$$

The  $c_k$  are complex numbers, giving us  $2n$  free parameters. Unit norm and phase invariance of the probabilities, give us two  $2(n-1)$  free parameters to determine the state. A repeated measurement of  $P(A_i = a_j|\psi)$ , can estimate the  $n$  parameters  $|\langle a_j|\psi\rangle|^2$ ,  $j = 1, \dots, n$ . Since these  $n$  parameters sum to one, only  $n - 1$  are independent. We then seek to establish the remaining  $n - 1$  parameters by means of another measurement. This can be accomplished by measuring another observable. However, two arbitrary observables can be more or less related, depending on how the eigenvectors of one observable can be written as linear functions of the eigenvectors of the other observable. Let observable  $A$  have a set of eigenvectors  $|a_k\rangle$  and observable  $B$  a set of eigenvectors  $|b_k\rangle$ . Assuming these observables act on the same Hilbert space, we can write:

$$|a_k\rangle = \sum_{j=1}^n r_{kj} |b_j\rangle$$

If the coefficients in this expansion are all zero except for the  $r_{jj}$  which are equal to 1, the two bases coincide and deliver exactly the same information. The other extreme case is when the coefficients are such that for all pairs  $\langle b_j|$  and  $|a_k\rangle$ , we have:

$$|\langle b_j|a_k\rangle|^2 = \frac{1}{n} \tag{5}$$

In this case we call the bases  $\{|a_k\rangle\}$  and  $\{|b_k\rangle\}$  *mutually unbiased*. The reason for this name, is that complete knowledge of the probabilities  $P(A_i = a_j|\psi)$  does not increase one's power of inference with respect to the probabilities  $P(B_i = b_j|\psi)$ . It is well-known that for a Hilbert space of dimension  $N$ , with  $N$  a prime number, there exist  $N + 1$  such mutually unbiased bases and there exist explicit procedures to construct the full  $N+1$  dimensional set of mutually unbiased bases. As indicated, two observables with eigenbases that are mutually unbiased, are statistically independent and knowledge of the  $2(n-1)$  parameters  $P(A_i = a_j|\psi)$  and  $P(B_i = b_j|\psi)$  suffices in principle to characterize the state completely. For non-pure states, the situation is more complex [10], [11], but mutually unbiased bases also play a crucial role. Suppose we have an  $n$ -level quantum system, described by a density operator

$\rho_n$ . How many free parameters do we need to obtain by experiment? There are  $n^2$  complex numbers in this  $n \times n$  matrix, leading to  $2n^2$  real parameters to specify. Because of the fact that a density operator is Hermitian, only  $n^2$  parameters are necessary. Lastly, the normalization condition imposes a single constraint, reducing the number free independent real parameters to  $n^2 - 1$ . Each measurement in single basis (i.e., a single observable), can at most provide us with the probabilities of  $n$  possible outcomes, giving us  $n - 1$  parameters. It follows we should have access to  $n + 1$  such observables, provided they are independent! Mutually unbiased bases have this property by construction, so can we find  $n + 1$  such bases for a  $n$  level system? The answer to this question, although mathematically well-defined, has turned out to be surprisingly difficult. It is known that for  $n$  prime, and even for  $n$  a power of a prime, there indeed exists  $n + 1$  mutually unbiased bases, and no more. However, even for low numbers  $n$ , the answer is unknown for any  $n$  that is not the power of a prime number.

## 5 State reconstruction with incomplete knowledge

### 5.1 Classical sensor pooling using the Maximum Entropy method.

A completely successful solution to the state determination problem is one for which the derived solution exists, is unique, and is stable (does not change much when the input data is slightly perturbed). In practice measurements are limited in precision and number. We often find that either the solution is not unique, is not stable, or does not even exist. Such a problem is called *ill-posed* in the sense of Haddamard and we can no longer determine the precise state of the system. The question we then face is: what is the most likely state of the system, given the outcomes we have obtained from these limited measurements. This is in essence a problem of inverse statistics, because we seek to determine which probability distribution is the most likely one, given our finite number of measurements. There are several ways to regularize an ill-posed problem. A particularly interesting method for the treatment of probabilistic data, is called the maximum entropy method, introduced by Jaynes [12] in the field of statistical mechanics in 1957. It has since then found wide spread application in spectral analysis, image restoration, quantum state determination and many other areas of interest [13]. The principle is usually stated in approximately the following formulation for the discrete case. Given an information source that can take any (but only one at a time) value  $x_j$  for a discrete random variable  $X$  and let  $p(x_j)$  denote

the probability that  $X$  takes the value  $x_j$ . We define the entropy associated with  $p$  as

$$H(p(x_1), \dots, p(x_n)) = - \sum_{i=1}^n p(x_i) \log_2(p(x_i)) \quad (6)$$

That the base of the logarithm in Eq. (6) is two, comes from the fact that the unit of information is the *bit*. However, for the purpose of the maximum entropy method, it is of no consequence what base is actually taken. The maximum entropy method can now be stated in very simple terms:

*Of all solutions compatible with the data, choose the one for which the entropy of the probability distribution derived from that data, is maximal.*

There are many arguments why the maximum entropy method should be so successful, but the most important reason is that the principle allows one to derive the most conservative density operator, in that it will not allow inferences not warranted by the actual data. We refer the interested reader to [13], [14], [15] for a discussion. A typical classical setting for the use of the maximum entropy method is the following.

Suppose we have a set of  $N$  sensors, each one assumed to measure the average value of some function  $\phi_k(x)$ . That is, we have

$$\begin{aligned} e_k &= E[\phi_k(x)] \\ &= \int \phi_k(x)p(x)dx, \quad k = 1, \dots, N \end{aligned}$$

How are we to determine the most likely distribution  $p(x)$ , when we are given the values  $e_k$  and the functions  $\phi_k(x)$ ? In most instances, this will not yield a unique answer, and the maximum entropy method can be used to select a conservative distribution. The idea is to maximize the entropy

$$S(p) = - \int p(x) \ln p(x) dx$$

with additional constraint that the averages match the measured values:

$$e_k = \int \phi_k(x)p(x)dx, \quad k = 1, \dots, N$$

The explicit solution is given by

$$p(x) = \frac{1}{Z(\theta)} \exp\left[- \sum_k^N \theta_k \phi_k(x)\right]$$

with

$$Z(\theta) = \int \exp[-\sum_k^N \theta_k \phi_k(x)] dx$$

and the  $\theta_k$  obtained through

$$-\frac{\partial Z(\theta)}{\partial \theta_k} = e_k, \quad k = 1, \dots, N$$

For more details, we refer to [16].

The maximum entropy method is sufficiently general that it can equally well be applied to the density operator estimation.

## 5.2 Quantum density operator reconstruction in $O_i$

As we have explained before, in quantum mechanics we very often describe the state of a system, not by means of a pure state, but through the density (or state) operator. The physical situation that often accompanies the description with a density operator, is one in which the preparation of systems under investigation is less than perfect. In other words: in case we fail to prepare an ensemble of identical system states, we will describe the members of the ensemble through its associated density operator. The rule to obtain the expectation value for an observable  $E$  represented by the operator  $E$  for a density operator  $\rho$  is the so-called trace rule:

$$\langle E \rangle = Tr(E\rho)$$

If the density operator is given in closed form, or measured to very good approximation, then a good measure for the purity of the density operator, is given by the entropic uncertainty measure  $\eta$ , also called the von Neumann entropy:

$$\eta(\rho) = -Tr(\rho \ln \rho) \tag{7}$$

If the state is pure, then  $\eta(\rho_{pure}) = 0$ . If the state is completely mixed, so that  $\rho_m = \frac{1}{N}1$ , then its uncertainty measure  $\eta(\rho_m)$  is maximal and equal to  $\ln(N)$ . Suppose however, that we do not possess knowledge of the density operator, but we do have the precise expectation values  $\langle E_\alpha \rangle$  of a set of observables  $E_\alpha$  with  $\alpha$  a natural number between 1 and  $n$ . We then have

$$\begin{aligned} Tr(\rho) &= 1 \\ \langle E_\alpha \rangle &= Tr(E_\alpha \rho) \end{aligned} \tag{8}$$

If we have the expectation value of an observable, this clearly puts a constraint on the set of density operators that is compatible with it. If we know nothing about the system, then the density operator can be any element of  $L$ , the set of all density operators acting on the state space of the entity:

$$\rho \in L$$

Suppose we have measured the expectation value of  $E_1$  and found it to be  $e_1$ . In that case we have

$$\begin{aligned} \rho &\in L_{E_1} \subset L \\ L_{E_1} &= \{\rho \in L : Tr(E_1\rho) = e_1\} \end{aligned}$$

Suppose we have a set of observables  $F = \{E_\alpha | \alpha \in I\}$ , and know of each observable in the set its expectation value  $e_\alpha$ . In that case the density operator belongs to smaller subset of  $L$ :

$$\begin{aligned} \rho &\in L_{\mathcal{F}} = \{\rho \in L : Tr(E_\alpha\rho) = e_\alpha, \forall \alpha \in I\} \\ L_{\mathcal{F}} &= \bigcap_{\alpha \in I} L_{E_\alpha} \end{aligned}$$

Incomplete knowledge then implies the data do not allow us to get a unique answer from this procedure, and other more or less general principles have to be invoked. Again, one of the most powerful of such techniques is the maximum entropy (MaxEnt) method [17]. The translation of the MaxEnt principle to this context now reads:

*Of all density matrices compatible with the observed expectation values as in Eq. (8) for a set of observables, one seeks the one with the largest von Neumann entropy ( $\gamma$ ).*

According to this principle, one has to select the density operator  $\rho^{MaxEnt}$  for which

$$\rho^{MaxEnt} = \arg \max_{\rho \in L_{\mathcal{F}}} [\eta(\rho)]$$

Following Jaynes, [12], [15], the actual reconstruction process can (iteratively) be performed in three steps:

- 1 Define the generalized partition function with unknown Lagrange multipliers  $\lambda_\alpha$

$$Z(\lambda_1, \dots, \lambda_n) = Tr(\exp - \sum_{\alpha} \lambda_{\alpha} E_{\alpha})$$

- 2 Solve the system of algebraic equations for the unknown Lagrange multipliers

$$\begin{aligned}\langle E_\alpha \rangle &= \text{Tr}(E_\alpha \rho) \\ &= -\frac{\partial}{\partial \lambda_\alpha} Z(\lambda_1, \dots, \lambda_n)\end{aligned}\tag{9}$$

Sometimes physically undesirable results are obtained using this last formula, which is an indication that the data do not permit even approximate reconstruction of the state. In case the operators  $E_\alpha$  do not commute, the right hand side of Eq. (9) should be evaluated using

$$\frac{\partial}{\partial a} \exp(-X(a)) = \exp(-X(a)) \int_0^1 \exp(\mu X(a)) \frac{\partial X(a)}{\partial a} \exp(-\mu X(a)) d\mu$$

- 3 The MaxEnt density operator, also known as the generalized canonical density operator, is then given by:

$$\rho^{MaxEnt} = \frac{1}{Z_0} \exp\left(-\sum_{\alpha} \lambda_{\alpha} E_{\alpha}\right)$$

Again, the principle allows one to derive the most conservative density operator, in that it will not allow inferences not warranted by the actual data. Many interesting examples of constructions with this method can be found in [5]. Because we assumed incomplete but accurate knowledge of the expectation values, the MaxEnt method yields an optimal choice for the density operator given access to observational level  $O_i$ .

### 5.3 Quantum density operator reconstruction in $O_a$

Suppose now we have access only to a limited number of measurements so that our averages for the observables can be considered only estimates for the expectation values. This corresponds to a reconstruction starting from observational level  $O_a$ . In this case a Bayesian inference procedure can be used. We start with a few simplifications to make the problem more palatable. We assume the number of outcomes for each observable is finite (discrete spectrum), and the state operator corresponds to a pure state. Please note that in the end, the reconstruction does not necessarily yield a pure state operator, and hence the von Neumann entropy can in this case be regarded as a measure of the faithfulness of the reconstruction. As measurements, we consider von Neumann projections and associate the set

of mutually orthogonal projectors  $P_E^{e_i}$  with the discrete spectrum of the observable with corresponding eigenvalues  $e_i$ . The Bayesian reconstruction then consists of performing three steps iteratively.

- 1 We form the conditional probability  $p(E, e_i|\rho)$  of finding result  $e_i$  for the measurement of  $E$ , if the state was  $\rho$

$$p(E, e_i|\rho) = \text{Tr}(P_E^{e_i}\rho)$$

- 2 The second step consist of specifying the a priori  $\rho$ , denoted as  $p_0(\rho)$ . This is to summarize our knowledge about the state before measurements were applied. If no such information exists, we assume a uniform distribution on the set of system states  $\Sigma_S$ . With the help of this a priori distribution, we define the joint probability  $p(E, e_i \wedge \rho)$  on  $\Sigma_S \times \Sigma_M$  as

$$p(E, e_i \wedge \rho) = p(E, e_i|\rho)p_0(\rho)$$

- 3 According to the well-known Bayes rule [18]:

$$p(x|y)p(y) = p(x \wedge y) = p(y|x)p(x) \quad (10)$$

we then have that

$$p(\rho|E, e_i) = \frac{p(E, e_i \wedge \rho)}{\int p(E, e_i \wedge \rho)d\Sigma_S} \quad (11)$$

The resulting density is then given by inversion of Eq. (11) using Eq. (10):

$$\rho = \frac{\int_{\Sigma_S} p(\rho|E, e_i)\rho d\Sigma_S}{\int_{\Sigma_S} p(\rho|E, e_i)d\Sigma_S}$$

After  $N$  iterations of the above three steps, we use as a priori distribution the output of the  $(N - 1)^{th}$  measurement. As mentioned before, we initially assume pureness of the density operator. The reason is that this is a necessary requirement for the uniqueness of the unitarily invariant measure  $d\Sigma_S$ . If this assumption is dropped, uniqueness is no longer guaranteed, but a purification ansatz can help to save the scheme [5]. An optimal algorithm for density operator estimation for finite ensemble exists and can be found in [19]. For a classical exposition of the Bayesian approach, we refer to [16].

## 6 Concluding remarks

If one measurement tells us something about a system, and another measurement gives us different information on that same system, then how are we to combine that knowledge in a consistent way? This is the central problem that we have addressed here. The answer turns out to be complex. Throughout the paper, we have assumed that the way to combine this information, is by reconstruction of the state of the system by means of various measurements. But even if we suppose that all measurements involved are arbitrarily precise and executed on the same state of the system under investigation, we need the proper relations between measurements to be able to reconstruct the state of the system. In the classical case, this rôle is fulfilled by the functions  $\phi_k(x)$ , in the quantum case by the commutation relations  $[X, Y]$  between the operators  $X$  and  $Y$  that represent the observables. In the ideal classical case, the measurements are assumed completely unrelated for each and every single outcome produced by the measurements. State reconstruction is then trivial. In the ideal quantum case, the measurements are assumed statistically independent (mutually unbiased bases) and state reconstruction, although it can be technically involved, is fundamentally a solved problem. In practice, however, we do not always have access to all the observables that we need to specify completely the state of the system (an informational complete set in the sense of Prugovečki). Moreover, we hardly ever have precise estimates for the expectation values of these limited sets of observables. These two limitations on data impose severe restrictions on the orthodox reconstruction schemes mentioned above. What to do then? We have considered several possible schemes for the reconstruction of the reality of the system from such limited observations. Both in the classical and the quantum regime, we find the MaxEnt and the Bayesian paradigm to be the most valuable. Both schemes have their limitations, which is only natural because one cannot produce more information by theoretical means than was already in the data. Yet both enjoy methods enjoy widespread use and seem to have a firm foundation: the Bayesian method in the standard way of handling conditional probabilities, as set out almost 250 years ago [18], and the MaxEnt method because it produces a stable answer that is conservative in the sense that it does not allow to infer something not warranted by the data [14]. If these methods still fail to produce good results, it is often an indication of poor quality of data. It should come as no surprise that the quality of observation is of tantamount importance as a starting point in the reconstruction of physical reality.

## 7 Acknowledgments

This research was part of FWO project G.0362.03N, and FWO-research community: “The construction of integrating worldviews”.

## References

- [1] S. Aerts, D. Aerts, F. E. Schroeck, The Necessity of Combining Mutually Incompatible Perspectives in the Construction of a Global View: Quantum Probability and Signal Analysis, In D. Aerts, B. D’Hooghe and N. Note (Eds.), *Worldviews, Science and Us: Redemarcating Knowledge and Its Social and Ethical Implications*. Singapore: World Scientific (2005).
- [2] E. Prugovečki, Information-theoretical aspects of quantum measurement. *Int. J. Theor. Phys.*, 16, 321–331 (1977).
- [3] W. Band, J. Park, *Found. Phys.*, 1, No 2, pp 133-144 (1970).
- [4] F. E. Schroeck, *Quantum Mechanics on Phase space*. Kluwer Academic Publishers, Dordrecht (1996).
- [5] V. Bužek, G. Drobny, R. Derka, G. Adam, H. Wiedemann, Quantum state reconstruction from incomplete data, [quant-ph/9805020](#) (1998).
- [6] D. Aerts, Foundations of quantum physics: a general realistic and operational approach, *Int. J. Theor. Phys.*, 38, 289-358 (1999).
- [7] D. Aerts, A possible explanation for the probabilities of quantum mechanics, *J. Math. Phys.*, 27, 202-210 (1986).
- [8] D. Aerts, S. Aerts, T. Durt and O. Lévêque, Classical and Quantum Probability in the  $\epsilon$ -model, *Int. J. Theor. Phys.*, 38, 407 (1999).
- [9] L.P. Hughston, R. Josza, W.K. Wootters, A complete characterization of quantum ensembles having a given density matrix, *Phys. Lett. A*, Vol. 183, No. 1, 14-18 (1993).
- [10] S. Bandyopadhyay, P. O. Boykin, V. Roychowdhury, F. Vatan, A new proof for the existence of mutually unbiased bases, [quant-ph/0103162](#) (2001).
- [11] J. Lawrence, Č. Brukner, and A. Zeilinger, Mutually unbiased binary observable sets on N qubits, *Phys. Rev. A* 65, [quant-ph/0104012](#) (2001).

- [12] E.T. Jaynes. Information theory and statistical mechanics, *Phys. Rev.*, 106, 620-630, *ibid.* 108, 171-190 (1957).
- [13] N. Wu, *The Maximum Entropy Method*, Springer Series in Information Science, Springer-Verlag, Berlin (1997).
- [14] E.T. Jaynes, *Probability theory, the logic of science*, Cambridge University Press (2003).
- [15] E.T. Jaynes, *Am. Jour. Phys.* vol 31, pp66 (1963).
- [16] M. Djafari, Probabilistic methods for data pooling, *ArXiv: physics/0111118v1* (2001).
- [17] V. Bužek, *Quantum Tomography from Incomplete Data via MaxEnt Principle*, *Lect. Notes Phys.*, 649, 189–234 (2004).
- [18] T. Bayes, *An essay towards solving a problem in the doctrine of chances*. *Phil. Trans. R. Soc.*, 53, pp 370-418 (1763).
- [19] R. Derka, V. Buzek, A. Ekert, *Universal Algorithm for Optimal Estimation of Quantum States from Finite Ensembles*, *Phys. Rev. Lett.* 80 1571 (1998).