UNIVERSITÀ DEGLI STUDI DI TRIESTE
Sede amministrativa del Dottorato di Ricerca
DIPARTIMENTO DI FISICA

XXXI CICLO DEL
DOTTORATO DI RICERCA IN FISICA

A possible origin of the quantum mechanical behavior

Settore scientifico-disciplinare FIS/02

DOTTORANDO:
Luca Curcuraci

RESPONSABILE DOTTORATO DI RICERCA:
Chiar.mo prof. Livio Lanceri (Univ. Trieste)

FIRMA: _____

RELATORE:
Prof. Angelo Bassi (Univ. Trieste)

FIRMA: _____

ANNO ACCADEMICO 2018/2019

Ai miei genitori, a mia sorella
e ai miei nonni (presenti e non).

Preface

This dissertation is submitted for the degree of Doctor of Philosophy at the university of Trieste. The research described herein was accomplished under the supervision of professor Angelo Bassi in the department of physics of the the university of Trieste, between November 2015 and November 2018. During this period I also wrote the following papers:

1. *Why do we need Hilbert spaces? - L. Curcuraci* (arXiv:1708.08326v1);
2. *On non-commutativity in quantum theory (I): from classical to quantum probability - L. Curcuraci* (arXiv:1803.04913v2);
3. *On non-commutativity in quantum theory (II): toy models for non-commutative kinematics - L. Curcuraci* (arXiv:1803.04916v2);
4. *On non-commutativity in quantum theory (III): determinantal point processes and non-relativistic quantum mechanics - L. Curcuraci* (arXiv:1803.04921v2);
5. *The notion of trajectory in quantum mechanics - A. Bassi, L. Curcuraci*;
6. *Unitary time-evolution in stochastic time-dependent Hilbert spaces - S. Bacchi, A. Bassi, L. Curcuraci*;
7. *The inverse Born problem in contextual probability theories: quantum spin and continuous random variables - S. Bacchi, L. Curcuraci, G. Gasbarri* (arXiv:1810.13443);
8. *A Thermodynamical derivation of the quantum potential and the temperature of the wave function - L.Curcuraci, M. Ramezani* (arXiv:1810.09892);
9. *Can time coarse-grain get rid-off measurement postulate? (Temporary title) - A. Bassi, L. Curcuraci, G. Gasbarri* (in preparation).

This thesis is about the papers 2,3 and 4. During the exposition of this work, we will refer to works above using the notation [CurX], where $X = 1, \cdots, 9$ is the number of the paper we want to cite.

This thesis represent an original work, except where acknowledgments and references are made to previous works.

Contents

Chapter 1

Introduction

Explaining what quantum mechanics is about is an hard task. One could start saying that quantum mechanics is the theory used to describe the quantum behavior of physical systems. However we need to explain what we mean by "quantum behavior". A naïve attempt could be the following: the quantum behavior is the typical behavior of a *microscopic* physical system. But, microscopic respect to what? One of the biggest object exhibiting quantum behavior is the C_{60} molecule [1] whose radius is of the order 7×10^{-10} m. The typical length of an Escherichia Coli (a very "famous" bacterium) is around 1×10^{-6} m. Ten orders of magnitude separate us from the C_{60} molecule, but there are only four orders of magnitude between the E.Coli and the C_{60} molecule. That said, a funny question may be: "Is the C_{60} molecule microscopic enough to be quantum even for E. Coli?". This and other misleading questions can be formulated if we base our definition of quantum behavior on the notion of microscopicity. Moreover there are macroscopic systems exhibiting quantum behavior in certain situations (typically difficult to achieve), for example Bose-Einstein condensate. A more operational definition can be given by specifying under which conditions a physical system behaves in a quantum way. It turns out that this happens when the action S of the system under study is of the order or less of the Plank constant $h = 6.62607004 \times 10^{-34}$ $\mathrm{m^2Kg/s}$, namely $|S| \lesssim h$. In addition, one must also specify that quantum mechanics describes the quantum behavior of physical systems when they can be considered non-relativistic.

The whole theoretical apparatus can be derived from a list of axioms, from which the various predictions of quantum mechanics can be derived using mathematical identities compatible with given physical situations. There are probably as many systems of axioms as there are books dealing with the subject [2, 3, 4, 5, 6, 7]. Choosing a particular set of axioms over another can already be a signature of the interpretation one gives to elements of the theory. Interpreting quantum mechanics is a notoriously controversial question, hence to be completely honest with the reader, it is better to declare from the beginning that a *statistical interpretation of quantum mechanics* will be used in this thesis. No ontological value to the wave function or other elements of the theory

is assumed: quantum theory is considered as a theory of probability. Once that this choice has been made, the central entity for the description of a physical system will be the one of *proposition*. A proposition is nothing but of a statement about a physical system, for example «*The particle position x is observed in the set A*» is a proposition. It is clear that a description in these terms can be done for any physical system (not necessarily quantum), and the theoretical prediction should tell us the number of times a given proposition holds true. Within the interpretation chosen, the system of axioms we adopt in this thesis is the one used in [3]. Below we list a summarized version of them.

Ax. 1 - To each quantum system one may associate a separable complex Hilbert space $\mathcal{H}$. Call $\mathcal{P}(\mathcal{H})$ the set of all the orthogonal projectors on $\mathcal{H}$, the experimentally testable propositions are represented by projectors belonging to a subset of $\mathcal{P}(\mathcal{H})$. Given two projectors $\hat{P}_1, \hat{P}_2$ associated to two propositions, when $[\hat{P}_1, \hat{P}_2] = 0$, the orthogonal projectors $\hat{P}_1\hat{P}_2$ and $\hat{P}_1+\hat{P}_2-\hat{P}_1\hat{P}_2$ correspond to the composition of the two propositions with the logical connectives "*and*" and "*or*", respectively. The *negation* of a proposition represented by $\hat{P}$ is the proposition associated to the orthogonal projector $\hat{\mathbb{I}} - \hat{P}$.

Ax. 2 - To each quantum system at a given time t, one may associate a positive trace-class operator $\hat{\rho}$ with $\mathrm{Tr}\,[\hat{\rho}] = 1$ called *state*. The probability that at time t a given proposition $\hat{P} \in \mathcal{P}(\mathcal{H})$ holds when the quantum system is prepared in the state $\hat{\rho}$ (at the time t) is given by $\mathrm{Tr}\left[\hat{P}\hat{\rho}\right]$.

Ax. 3 - Physical quantities used to describe a given quantum system are represented by self-adjoint operators on $\mathcal{H}$. The expectation value of the physical quantity associated to the self-adjoint operator $\hat{A}$, when the quantum system is in the state $\hat{\rho}$, is given by $\mathbb{E}_{\hat{\rho}}[\hat{A}] = \mathrm{Tr}\left[\hat{A}\hat{\rho}\right]$.

Ax. 4 - The Hilbert space associated to a single quantum particle in $\mathbb{R}^n$ with m internal degrees of freedom is $L_2(\mathbb{R}^n) \otimes \mathbb{C}^{2m+1}$.

Ax. 5 - The self adjoint operator associated to the position observable of a single quantum particle in $\mathbb{R}^n$ is the multiplicative operator $\hat{\mathbf{X}} := (\hat{X}_1, \cdots, \hat{X}_n)$, whose components are defined as

$$\hat{X}_i\psi(x_1, \cdots, x_n) = x_i\psi(x_1, \cdots, x_n),$$

where $\psi(x_1, \cdots, x_n) \in L_2(\mathbb{R}^n)$. The self adjoint operator associated to the momentum observable of a single quantum particle in $\mathbb{R}^n$ is the differential operator $\hat{\mathbf{P}} := (\hat{P}_1, \cdots, \hat{P}_n)$, whose components are defined as

$$\hat{P}_i\psi(x_1, \cdots, x_n) = -i\hbar\frac{\partial}{\partial x_i}\psi(x_1, \cdots, x_n),$$

where $\psi(x_1, \cdots, x_n) \in L_2(\mathbb{R}^n)$.

Ax. 6 - To each quantum system described by an Hilbert space $\mathcal{H}$, there exists a self-adjoint operator $\hat{H}$ with bounded-from-below spectrum, called *Hamiltonian*, and corresponding to the mechanical energy of the system, such that the state of the system evolves in time according to the equation

$$i\hbar\frac{d}{dt}\hat{\rho}(t) = [\hat{H}, \hat{\rho}(t)],$$

provided that the quantum system is isolated.

Ax. 7 - Given a quantum system composed by N (finite) quantum particles, its Hilbert space can be constructed from the Hilbert spaces associated to each quantum particle, say $\{\mathcal{H}_{i,1p}\}_{i=1,\cdots,N}$, as

$$\mathcal{H} = \bigotimes_{i=1}^{N} \mathcal{H}_{i,1p}.$$

Ax. 8 For quantum systems with Hilbert space $\mathcal{H}$ composed by N indistinguishable subsystems, the only physically admissible propositions are the ones invariant under permutations. This means that if $\hat{U}_\sigma$ is the (unitary) operator representing a permutation σ of $\{1, \cdots, N\}$ on $\mathcal{H}$, the physically admissible propositions are the ones associated to the orthogonal projectors $\hat{P} \in \mathcal{P}(\mathcal{H})$ such that $\hat{U}_\sigma^{-1}\hat{P}\hat{U}_\sigma = \hat{P}$ for every permutation σ of $\{1, \cdots, N\}$.

Ax. 9 Given a quantum system prepared in a state $\hat{\rho}$, if at time t performing a measurement, a particular proposition represented by the orthogonal projector $\hat{P} \in \mathcal{P}(\mathcal{H})$ is found to hold, the state of the quantum system right after the measurement is given by

$$\hat{\rho}' = \frac{\hat{P}\hat{\rho}\hat{P}}{Tr(\hat{P}\hat{\rho})}.$$

These axioms are not independent. In particular, a careful analysis of the mathematical structures behind them, suggests that only three are the basic underlying principles. The first axiom simply tell us how to deal with the experimentally testable propositions. It is the arrival point of *quantum logic* [8, 9], a field of research founded by G. Birkhoff and J. Von Neumann with their pioneering paper [10]. Although quantum logic is interesting in its own right, it cannot be considered a satisfactory basis for this axiom: there is no real justification (at least for the Author) for the whole mathematical structure needed to derive it. Hence, no real justification for this axiom seems to be available at the moment. This is not the case for the second axiom. Indeed, once one looks for a probability measure over the set of projectors, the *Gleason's theorem* comes into help [11]. Essentially, this theorem states that any probability measure on projectors defined on a separable Hilbert space $\mathcal{H}$, can be always written in the form used in axiom 2 provided that $\dim \mathcal{H} > 2$. In light of the axiom 4, we can see that the application

of this theorem is justified. The third axiom can be seen as consequence of the first two. Indeed, an observable of a physical system (in general, not necessarily quantum) can be seen as the collection of possible observed outcomes, together with the corresponding collection of propositions stating that a given outcome of the experiment is observed. From this one can prove that to every observable a self adjoint operator can be associated, through the use of the *spectral theorem* [3, 12, 13, 14]. In light of axiom 2, and using the notion of *spectral measure* [13, 14], the definition of expectation value given in axiom 3 follows. Axiom 4 is strictly related to the first and no actual justification seem to be available. One may argue that it should be incorporated into the first, however this axiom is valid only for a single quantum particle while the first holds for any system. For this reason they must be kept separate. Axiom 5 and 6 are both consequences of another basic fact: the law of physics of non-relativistic systems are covariant under *Galilean transformations*. In fact, once one wants to represent the Galilean group preserving the probabilistic structure derived from the first four axioms (in particular, preserve the transition probabilities), one is naturally lead to look for a *strongly-continuous unitary representation* of the group over the Hilbert space of the axiom 4 (this cannot be done directly but one has to consider the central extension of the Galilean group)[3, 13]. The generators of a Galilean transformation give rise to the operator representations of various physical observables like *position*, *momentum*, *angular momentum* and *mechanical energy*. In particular, the fundamental commutation relation $[\hat{X}, \hat{P}] = i\hat{\mathbb{I}}$ is obtained: then the *Stone - Von Neumann - Mackey theorem* [15, 16, 17, 18] implies that the position and momentum operator must be the one of axiom 5. Finally applying the *Stone theorem* [15] to the operator representing the time-translation, one can naturally derive the Schrödinger equation from which axiom 6 follows. The seventh axiom can be derived from the previous ones by means of quantum logic tools, using substantially the same arguments used to describe a composite classical system [19]. Axiom 8 is nothing but the definition of indistinguishable objects. It is worth note that in this axiom, bosons and fermions are two special cases. In general the so-called para-statistics are allowed but they are not believed to be associated to any fundamental particle (despite this, there have very interesting applications [20]). To conclude axiom 9, the controversial *measurement postulate*, within the statistical interpretation does not represent a problem. It can be seen as the Bayesian update one has to perform after a measurement: in this sense it cannot be derived from the previous but can be naturally added to them. In this respect one can see how the measurement postulate is derived in Bohmian mechanics [5] [Cur5] via Bayes theorem. A similar idea for the introduction of the measurement postulate can be found in [21] where within the framework of quantum probabilities a quantum Bayes theorem is derived. Finally a description of a (weak) measurement process within the ordinary formalism of quantum mechanics, done entirely at the level of the wave function and without invoking the measurement postulate, can be found in [Cur9].

From the arguments briefly summarized above, one can see that the system of postulates presented is essentially based on three principles:

i) the use of Hilbert spaces in order to have a consistent statistical description for a quantum system;

ii) the specific Hilbert space used for the description of a single quantum particle;

iii) the use of the Galilean group as symmetry group for a non-relativistic system.

This thesis is about the first two principles. In particular, a possible reason for the two principles will be presented starting from an assumption about the geometry of space. It will be shown that a point-like particle jumping at random on a stochastic space exhibits quantum behavior once that the stochastic space is "removed" [a]. The strategy used to prove that is based on information theoretical tools, used to detect non-commutativity at the level of the algebraic description of the system considered (particle + space). Once that this non-commutativity is detected, ordinary algebraic theorems implies the such a description can be done on Hilbert spaces.

This thesis is organized as follows. In chapter 2 two axiomatic models of probability are presented: classical (commutative) and algebraic (non-commutative) ones. Both of them are presented using a common language: the one of Hilbert spaces. In chapter 3, an useful tool to detect non-commutativity is presented and the construction of an algebraic probabilistic model from a classical model is discussed. In chapter 4 the tools developed are applied to the description of a point-like particle in a random 1-D space. It is proved that position and momentum (velocity) of this particle are representable on a common algebraic probability space only using non-commuting operators, once that the random space is eliminated from the description. Limitations of this description are discussed. In chapter 5 we pause for a moment to introduce the notion of point process. In particular the class of determinantal point process is briefly reviewed. In chapter 6, using the notion of determinantal point process and the idea developed in chapter 4, a model capable of recovering non-relativistic quantum mechanics is discussed. To conclude, limitations and outlooks are discussed.

[a] The explicit meaning of this operation will be explained in details later. For the moment it is enough to say that after this operation, the stochastic space can no longer be described in the probabilistic model associated to the particle.

Chapter 2

Probability theories on Hilbert spaces.

In this chapter we will describe two possible models of probability: the Kolmogorov's measure-theoretic model and the algebraic model[a]. In section 2.1 the measure theoretic approach is briefly reviewed. The representation of this probability model on a Hilbert space is studied, with particular attention to the link between random variables and self-adjoint operator. This enables us to represent a single random variable with an operator on some Hilbert space. In section 2.2 the algebraic approach will be presented. A possible representation on Hilbert spaces and the impossibility of the use of the measure-theoretic language due to non-commutativity is explained.

2.1 Classical probability in Hilbert space language.

In this section, we collect a series of results about measure-theoretic probability theory and its algebraic formulation. After the introduction of the necessary mathematical tools (a good reference is [3]), we will show how standard measure-theoretic probability look like in Hilbert spaces [25].

2.1.1 Measure-theoretic probability, i.e. classical probability.

With the term *classical probability* we will refer to Kolmogorov's formulation of probability theory based on measure space. According to this framework, the description of a random phenomenon is made by using the triple $(\Omega, \mathcal{E}, P)$, called *probability space*, where

[a] In some sense, von Neumann can be considered as the father of the algebraic approach, despite we need to wait till 80's in order to have a proper formalization [22, 23, 24].

i) Ω is the *sample space*, and it represents the set of all possibile elementary outcomes of a random experiment;

ii) $\mathcal{E}$ is a *σ-algebra on* Ω, namely a collection of subsets of Ω which is closed under complementation and countable union. It can be understood as the set of all propositions (also called *events*) about the random phenomenon whose truth value can be tested with an experiment;

iii) P is a *probability measure*, namely a map $P : \mathcal{E} \to [0, 1]$, which is normalised ($P(\Omega) = 1$) and σ-additive. If $A \in \mathcal{E}$ is an event, then we will interpret $P(A)$ as the degree of belief that the event A really happens.

Note that in Kolmogorov's formulation, a probability space is nothing but a measure space where the measure is normalised. A random variable X in this context, i.e. a feature of a random phenomenon, is simply any P-measurable map from the probability space to another measurable space[b] $(M, \mathcal{M})$, hence $X : (\Omega, \mathcal{E}, P) \to (M, \mathcal{M})$. The image of the probability measure P, under the map X, induces a (probability) measure on $(M, \mathcal{M})$, $\mu_X := P \circ X^{-1}$, which is called *probability distribution of the random variable* X. Statistical information about a random variable X can be obtained from the *expectation value*, defined as

$$\mathbb{E}[X] := \int_\Omega X(\omega) P(d\omega) = \int_M x \mu_X(dx)$$

Sometimes, to emphasise the probability measure which we are using to compute the expectation we write $\mathbb{E}_P$.

The definition given for P is rather obscure since we should explain the meaning of "degree of belief". This is a signature of the fact that the notion of probability is a primitive concept in the measure-theoretic formulation. A method we can use to measure P is explained by the *(weak) law of large numbers* [25].

Theorem 1. *Let* $\{A_i\}_{i \in \mathbb{N}}$ *be a collection of* independent events*, namely* $P(\cap_i A_i) = \Pi_i P(A_i)$*. If* $P(A_i) = p$ *for all* i*, namely they have all the same probability, for any* $\epsilon > 0$*, we have*

$$\lim_{n \to \infty} P\left(\left\{ \omega \in \Omega \middle| \left| \frac{K_n(\omega)}{n} - p \right| < \epsilon \right\} \right) = 1$$

where $K_n(\omega) = \sum_{j=1}^n \chi_{A_j}(\omega)$ *where* $\chi_{A_j}(\omega)$ *is the indicator function for the set* A_j*.*

Let us explain the meaning of this theorem, and what it tells us about the measurement of probability. First of all, we should accept that $P(A) = 1$ means to be sure that the event A is true. Assumed this, the meaning of this theorem is hidden in the function $K_n(\omega)$. Consider the following collection of events

$$A_j := \{\text{in the trial } j \text{ we found } \omega \in A\} \qquad \forall j \in \mathbb{N}.$$

[b] A measurable space is the couple $(M, \mathcal{M})$ where M is a set and $\mathcal{M}$ is a σ-algebra on this set. When equipped with a measure it becomes a measure space.

For this collection of events, the function $K_n(\omega)$ is just the number of times we observe $\omega \in A$ by repeating the observation n times. In addition the independence hypothesis in the theorem ensures that the observation in the i-th trial does not influence the j-th trial. Finally the requirement $P(A_i) = p, \forall i \in \mathbb{N}$, is a quite natural requirement: the probability that $\omega \in A$ is the same independently on the trial. At this point the meaning of the theorem is clear: for sufficiently many trials, the number of times we find $\omega \in A$ normalised to the total number of trials, tend to be the number p with high probability. Notice that, despite this theorem tells us how to *measure* P (via frequencies), we cannot use it to define the meaning of P: it would be a recursive definition. This argument explains why probability in the measure-theoretic framework is a primitive notion.

2.1.2 The algebra of functions $L_\infty(\Omega, \mathcal{E}, P)$.

Let $(\Omega, \mathcal{E}, P)$ be a probability space and consider the functions $f : \Omega \to \mathbb{C}$ which are measurable with respect to the σ-algebra $\mathcal{E}$. We also require f to be bounded, namely $\|f\| := \sup_{\omega\in\Omega} |f(\omega)| < \infty$. Then we can define the following class of functions

$$\mathscr{L}_\infty(\Omega, \mathcal{E}, P) := \{f : \Omega \to \mathbb{C} | f \text{ is } \mathcal{E}\text{-measurable}, \|f\| < \infty\}$$

With the equivalence relation $f \sim g$ whenever $P(\{\omega \in \Omega | f(\omega) = g(\omega)\}) = 1$, we can define the following object $L_\infty(\Omega, \mathcal{E}, P) = \mathscr{L}_\infty(\Omega, \mathcal{E}, P)/\sim$. Defining the operation of sum, multiplication by a scalar and multiplication between functions in the usual way, $L_\infty(\Omega, \mathcal{E}, P)$ becomes an algebra of functions. Finally, using the *essential supremum norm* $\|f\|_\infty := \text{ess sup} f = \inf\{\alpha \in \mathbb{R} \mid P(\{\omega \in \Omega | |f(\omega)| \leqslant \alpha\}) = 1\}$, $L_\infty(\Omega, \mathcal{E}, P)$ is an abelian C^*-algebra of functions (C^* means that $\|f^* f\|_\infty = \|f\|^2_\infty$ which is true for the complex conjugation * and the essential supremum norm $\|\cdot\|_\infty$, see section 2.1.3 for more details). Note that ess sup$f \leqslant \sup f$.

This object encodes, in an algebraic way, all the information encoded in the underlying probability space. Clearly $(\Omega, \mathcal{E}, P)$ determines uniquely $L_\infty(\Omega, \mathcal{E}, P)$, but the opposite is not exactly true. Indeed, given $L_\infty(\Omega, \mathcal{E}, P)$ we may construct a σ-algebra by setting

$$\tilde{\mathcal{E}} := \{p \in L_\infty(\Omega, \mathcal{E}, P) | p = p^* = p^2\}$$

but this is not isomorphic to the original $\mathcal{E}$, since we identified everywhere P-equal functions in the construction of $L_\infty(\Omega, \mathcal{E}, P)$. $\tilde{\mathcal{E}}$ is a *measure algebra*. Nevertheless this is an advantage instead of a limitation. Indeed, measure algebra is a coherent way to exclude set of zero measure from the probability space describing the random phenomenon (see Sec. 1.7 in [26]). Over this σ-algebra, we can define a probability measure $\tilde{P} : \tilde{\mathcal{E}} \to [0, 1]$ as $\tilde{P}(f) := \phi(f)$ for any f which is $\tilde{\mathcal{E}}$-measurable, where ϕ is a positive normalised linear functional defined to be $\phi(f) = \int_\Omega f(\omega) P(d\omega)$. Summarising, starting from the algebra $L_\infty(\Omega, \mathcal{E}, P)$ we can construct a probability space $(\Omega, \tilde{\mathcal{E}}, \tilde{P})$ which is equivalent, up to zero measure set, to the probability space $(\Omega, \mathcal{E}, P)$.

Consider an ordinary random variable $X : (\Omega, \mathcal{E}, P) \to (M, \mathcal{M})$. Since $(M, \mathcal{M}, \mu_X)$ is a probability space as well, we can associate to it an abelian C^*-algebra of functions. In this picture, X can be seen as a linear map between algebras which respects the multiplication, namely a *C^*-algebra homomorphism*. Thus we can say that the algebraic analogous of the random variable X is the C^*-algebra homomorphism $x_X : L_\infty(\Omega, \mathcal{E}, P) \to L_\infty(M, \mathcal{M}, \mu_X)$. We can see that a random phenomenon described in measure theoretic language, can be equivalently described using (abelian) algebras: this is part of the algebraic approach to probability theory. We conclude this section by observing that, in the algebraic approach, the role of the random variables is central: the elements of $L_\infty(\Omega, \mathcal{E}, P)$ are functions on $(\Omega, \mathcal{E}, P)$, i.e random variables.

2.1.3 From probability spaces to abelian von Neumann algebras.

We have seen that the information encoded in $(\Omega, \mathcal{E}, P)$, can be encoded in an equivalent manner in the algebra $L_\infty(\Omega, \mathcal{E}, P)$. Now, we will establish a link between the algebra L_∞ and a suitable von Neumann algebra of operators over some Hilbert space.

In general, an algebra $\mathcal{A}$ is a vector space equipped with a product operation. Typically such product is assumed to be associative and in some case, it can be commutative. An algebra can have or not the unit element with respect to this multiplication but in what follows we will always consider algebras with unit. We will always consider algebras having a *norm* defined on it, labeled by $\|\cdot\|$. It is also useful to consider algebras equipped with an additional map $^* : \mathcal{A} \to \mathcal{A}$, such that $(a^*)^* = a$, which is called *involution* (examples of involutions are the complex conjugation for functions or the adjoint operation for operators). At this point, we may define what is a C^*-algebra.

Definition 1. *Let $\mathcal{A}$ be an algebra with a norm $\|\cdot\|$ and an involution *. If $\mathcal{A}$ is complete with respect to the norm $\|\cdot\|$ we call this algebra* *-algebra. *If in addition,*

$$\|a^* a\| = \|a\|^2$$

we say that $\mathcal{A}$ is a C^*-algebra.

Completeness of $\mathcal{A}$ is understood in the usual way: all the Cauchy sequences in $\mathcal{A}$ with respect to a given norm are also convergent sequences. Given a *-algebra $\mathcal{A}$, a generic element $a \in \mathcal{A}$ is said to be *self-adjoint* if $a = a^*$, while it is said to be positive (and we will write $a \geqslant 0$) if we can write $a = b^* b$, for some $b \in A$.

Definition 2. *Let $\mathcal{A}$ be a *-algebra ,* a state over $\mathcal{A}$ *is a linear functional $\phi : \mathcal{A} \to \mathbb{C}$ which is positive ($\phi(aa^*) \geqslant 0$ for any $a \in \mathcal{A}$) and normalised ($\phi(\mathsf{I}) = 1$, where I is the unit of $\mathcal{A}$).*

Note that the definitions above are very abstract in the sense that we do not need to define explicitly the sum, the product, the norm or the involution.

For this reason $\mathcal{A}$ is called *abstract algebra* if such information are not declared. When all the features of the algebra are explicited, we speak of *concrete algebra*. Let us restrict our attention to the case of algebras of operators in some Hilbert space $\mathcal{H}$, and in particular to $\mathcal{A} = \mathcal{B}(\mathcal{H})$ (the bounded operators over an Hilbert space $\mathcal{H}$) which is a concrete algebra. Thanks to the notion of positivity, we have a natural ordering operation $\geqslant$ between the elements of the algebra, i.e. given two operators $\hat{A}_1$ and $\hat{A}_2$, the writing $\hat{A}_1 \geqslant \hat{A}_2$ means $\hat{A}_1 - \hat{A}_2 \geqslant 0$.

Definition 3. *Let $\mathcal{A}$ be an operator algebra and $\{\hat{A}_i\}_{i=1,2,\dots}$ be an increasing sequence of operators in $\mathcal{A}$* with strong limit $s - \lim_{n\to\infty} \hat{A}_n = \hat{A}$, *namely $\hat{A}_1 \leqslant \hat{A}_2 \leqslant \cdots$ and $\lim_{n\to\infty} \|\hat{A}_n - \hat{A}\| = 0$ for some $\hat{A} \in \mathcal{A}$. A state ϕ is said to be* normal *if $\lim_{n\to\infty} \phi(\hat{A}_n) = \phi(\hat{A})$.*

A normal state ϕ on $\mathcal{B}(\mathcal{H})$ can be written as $\phi(\cdot) = \mathrm{Tr}\,[\hat{\rho}\ \cdot\ ]$ for some $\hat{\rho} \in \mathcal{B}_1(\mathcal{H})$ (see Th. 7.1.12 in [27]).

Definition 4. *Let $\mathcal{A}$ be an algebra of operators and $\phi : \mathcal{A} \to \mathbb{C}$ a state on it. Take some $\hat{A} \in \mathcal{A}$, if $\phi(\hat{A}^*\hat{A}) = 0$ implies $\hat{A} = 0$, then ϕ is said* faithful.

Among algebras of operators a very important class is the one of von Neumann algebras.

Definition 5. *Let $\mathcal{H}$ be an Hilbert space, a* von Neumann algebra $\mathcal{V}(\mathcal{H})$ *is a *-sub-algebra of $\mathcal{B}(\mathcal{H})$ which is strongly closed (i.e. the strong limit of any sequence of operators in $\mathcal{V}(\mathcal{H})$ converge to some operator which is still in $\mathcal{V}(\mathcal{H})$).*

In general, any von Neumann algebra is a C^*-algebra, but the opposite is not true. Von Neumann algebras are concrete algebras, however in general one should consider more abstract algebras, not necessarily composed of operators, hence it is useful to introduce also the notion of representation.

Definition 6. *Let $\mathcal{A}$ be an algebra with involution and $\mathcal{H}$ an Hilbert space. An homomorphism $\pi : \mathcal{A} \to \mathcal{B}(\mathcal{H})$ preserving the involution is called* representation of $\mathcal{A}$ on $\mathcal{H}$. *A representation is said* faithful *if it is one-to-one.*

We now have all the notions needed to state the main theorem of this section. Consider the algebra $L_\infty(\Omega, \mathcal{E}, P)$ and for any $f \in L_\infty(\Omega, \mathcal{E}, P)$ define the operator $\hat{M}_f$ on the Hilbert space $L_2(\Omega, \mathcal{E}, P)$ as

$$\hat{M}_f \psi(\omega) = f(\omega)\psi(\omega) \qquad \psi(\omega) \in L_2(\Omega, \mathcal{E}, P)$$

Clearly, such a representation is faithful and represents $L_\infty(\Omega, \mathcal{E}, P)$ as multiplicative operators on $L_2(\Omega, \mathcal{E}, P)$. This is the link mentioned in the beginning. More formally we have the following theorem.

Theorem 2. *Let $(\Omega, \mathcal{E}, P)$ be a probability space. Then the algebra $\mathcal{V}_c(L_2(\Omega, \mathcal{E}, P)) := \{\hat{M}_f | f \in L_\infty(\Omega, \mathcal{E}, P)\}$ is an abelian von Neumann algebra on the Hilbert space $L_2(\Omega, \mathcal{E}, P)$ and*

$$\phi_P : \hat{M}_f \mapsto \int_\Omega f(\omega) P(d\omega)$$

is a faithful normal state on $\mathcal{V}_c(L_2(\Omega, \mathcal{E}, P))$.

Proof. See Appendix. □

More generally, the results obtained till now can be reversed: starting from a generic abelian von Neumann algebra we may construct a probability space [25].

Theorem 3. *Let $\mathcal{A}$ be an abelian von Neumann algebra of operators and ϕ a faithful normal state on it. Then there exist a probability space $(\Omega, \mathcal{E}, P)$ and a linear correspondence between $\mathcal{A}$ and $L_\infty(\Omega, \mathcal{E}, P)$, $\hat{A} \mapsto f_{\hat{A}}$, such that*

$$f_{\hat{A}\hat{B}} = f_{\hat{A}} f_{\hat{B}} \qquad f_{\hat{A}^*} = (f_{\hat{A}})^*$$
$$\|f_{\hat{A}}\|_\infty = \|\hat{A}\| \qquad \mathbb{E}[f_{\hat{A}}] = \phi(\hat{A}).$$

Summarising, the theorem above tells that any abelian C^*-algebra of functions, which is constructed from a probability space, can be described in an equivalent way by using multiplicative operators over a suitable Hilbert space that one can construct from the probability space itself. It is important to observe that the state ϕ_P is not constructed from the vectors of $L_2(\Omega, \mathcal{E}, P)$. Finally, despite we are describing a classical probability space using an Hilbert space, this Hilbert space changes when we change the probability measure P.

2.1.4 Essentials of spectral theory for bounded operators.

Here we introduce the basic notions and theorems about the spectral theory of bounded operators which will be used later. The central object of the spectral theory is the notion of PVM [3]. In order to define them in the whole generality, we recall that a *second-countable topological space* $(\mathsf{X}, \mathscr{T})$ is a set X with a topology $\mathscr{T}$ (collection of open sets) whose elements can be seen as the countable union of basis sets (i.e. elements of $\mathscr{T}$ which cannot be seen as unions of other sets).

Definition 7. *Let $\mathcal{H}$ be an Hilbert space, $(\mathsf{X}, \mathscr{T})$ a second-countable topological space and $\mathscr{B}(\mathsf{X})$ the borel σ-algebra on X. The map $\hat{P} : \mathscr{B}(\mathsf{X}) \to \mathcal{B}(\mathcal{H})$ is called* projector-valued measure *(PVM) on X, if the following conditions holds*

i) $\hat{P}(B) \geqslant 0$ *for any* $B \in \mathscr{B}(\mathsf{X})$;

ii) $\hat{P}(B)\hat{P}(B') = \hat{P}(B \cap B')$ *for any* $B, B' \in \mathscr{B}(\mathsf{X})$;

iii) $\hat{P}(\mathsf{X}) = \hat{\mathbb{I}}$;

iv) if $\{B_n\}_{n\in\mathbb{N}} \subset \mathscr{B}(\mathsf{X})$ *with* $B_n \cap B_m = \{\varnothing\}$ *for* $n \neq m$, *then*

$$\sum_{n=0}^{\infty} \hat{P}(B_n) = \hat{P}\left(\bigcup_{n=0}^{\infty} B_n\right)$$

The support of the PVM is the closed set defined as $supp(\hat{P}) := \mathsf{X}/\{\cup A | A \in \mathscr{T}, \hat{P}(A) = \hat{\mathbb{O}}\}$. *When $\mathsf{X} = \mathbb{R}^n$, $\hat{P}$ is said bounded if $supp(\hat{P})$ is a bounded set.*

In the above definition $\hat{\mathbb{O}}$ is simply the null operator. Because a $(\mathbb{Z}, \mathcal{P}(\mathbb{Z}))$, where $\mathcal{P}(A)$ means the power set of A, and $(\mathbb{R}^n, \mathcal{T}_o)$, where $\mathcal{T}_o$ is the ordinary euclidean topology, are second-countable topological spaces, with the above definition we may treat at the same time the continuous and discrete cases. PVMs are useful because they allow to define operator-valued integrals with respect to them. In fact, if we consider a bounded function $g : \mathsf{X} \to \mathbb{C}$ which is measurable, we can define

$$\hat{F}(g) := \int_{\mathsf{X}} g(x)\hat{P}(dx)$$

which is called *integral operator in* $\hat{P}$ and it is a (bounded) operator on $\mathcal{H}$. We observe that

$$\int_{\mathsf{X}} g(x)\hat{P}(dx) = \int_{\mathrm{supp}(\hat{P})} g(x)\hat{P}(dx)$$

for any measurable bounded function g, because the PVM vanishes for all $A \notin \mathrm{supp}(\hat{P})$. One can prove that, if g is measurable and bounded, then $\| \int_{\mathsf{X}} g(x)\hat{P}(dx)\| \leqslant \|g|_{\mathrm{supp}(\hat{P})}\|_\infty$ and also that the integral operator is positive for g positive. Related to PVM, another important quantity is the *spectral measure*.

Definition 8. *Let $\psi \in \mathcal{H}$, the map $\mu_\psi : \mathscr{B}(\mathsf{X}) \to \mathbb{R}$ defined as*

$$\mu_\psi(E) := \langle\psi| \int_{\mathsf{X}} \chi_E(x)\hat{P}(dx)\psi\rangle \qquad E \in \mathscr{B}(\mathsf{X}),$$

where $\langle\cdot|\cdot\rangle$ is the scalar product of $\mathcal{H}$, is a real and positive measure on $\mathbb{R}$ called spectral measure associated to ψ.

Note that if $\psi \in \mathcal{H}$ is normalised, then also μ_ψ is. An important property for any bounded and measurable function g on X is the following:

$$\langle\psi| \int_{\mathsf{X}} g(x)\hat{P}(dx)\psi\rangle = \int_{\mathsf{X}} g(x)\mu_\psi(dx)$$

which is simply a consequence of the fact that we may always write $g(x)$ as limit of a sum of indicator functions. This last equality is very important in quantum physics. Since $\int_{\mathsf{X}} f(x)\hat{P}(dx)$ is an operator on $\mathcal{H}$, i.e. $\hat{F}(g)$, the above equality tells that the quantum mechanical expectation $\langle\psi|\hat{F}(g)\psi\rangle$ coincides with the ordinary expectation value $\mathbb{E}[g]$ when it is computed with the spectral measure. As we will see, this is a very general feature of self-adjoint operators.

At this point we can state (without proof) the two central theorems of the spectral theory for bounded operators. The first important theorem is the *spectral decomposition theorem for self-adjoint operators in $\mathcal{B}(\mathcal{H})$* which tells that every self-adjoint operator in $\mathcal{B}(\mathcal{H})$ can be constructed integrating some function with respect to a specific PVM, and it is completely determined by it.

Theorem 4 (Th. 8.54 in [3])**.** *Let $\mathcal{H}$ be an Hilbert space and $\hat{A} \in \mathcal{B}(\mathcal{H})$ a self-adjoint operator.*

a) There exists a unique and bounded PVM $\hat{P}^{(\hat{A})}$ *on* $\mathbb{R}$ *such that*

$$\hat{A} = \int_{supp(\hat{P}^{(\hat{A})})} x\hat{P}^{(\hat{A})}(dx);$$

b) $\sigma(\hat{A}) = supp(\hat{P}^{(\hat{A})})$*, where* $\sigma(\hat{A})$ *is the spectrum of the operator* $\hat{A}$*;*

c) If f *is a bounded measurable function on* $\sigma(\hat{A})$*, the operator* $f(\hat{A}) := \int_{\sigma(\hat{A})} f(x)\hat{P}^{(\hat{A})}(dx)$ *commutes with every operator in* $\mathcal{B}(\mathcal{H})$ *which commutes with* $\hat{A}$*.*

The second important theorem is the so called *spectral representation theorem of self-adjoint operators in* $\mathcal{B}(\mathcal{H})$. This theorem tells that every bounded self-adjoint operator on $\mathcal{H}$ can be represented as a multiplicative operator on some L_2 Hilbert space, which is basically constructed from its spectrum.

Theorem 5 (Th. 8.56 in [3])**.** *Let* $\mathcal{H}$ *be an Hilbert space,* $\hat{A} \in \mathcal{B}(\mathcal{H})$ *a self-adjoint operator and* $\hat{P}^{(\hat{A})}$ *the associated PVM. Then*

a) $\mathcal{H}$ *splits as Hilbert sum* $\mathcal{H} = \oplus_{i\in I}\mathcal{H}_i$ *(with* I *at most countable if* $\mathcal{H}$ *is separable), where* $\mathcal{H}_i$ *are closed and mutually orthogonal subspaces such that*

i) $\forall i \in I$*, then* $\hat{A}\mathcal{H}_i \subset \mathcal{H}_i$*;*

ii) $\forall i \in I$ *there exist a positive finite borel measure* μ_i *on the Borel sets of* $\sigma(\hat{A}) \subset \mathbb{R}$*, and a surjective isometry* $\hat{U}_i : \mathcal{H}_i \to L_2(\sigma(\hat{T}), \mu_i)$ *such that*

$$\hat{U}_i \left(\int_{\sigma(\hat{A})} f(x)\hat{P}^{(\hat{A})}(dx) \right) \Big|_{\mathcal{H}_i} \hat{U}_i^{-1} = f\cdot$$

for any bounded measurable f*, where* $f\cdot$ *means multiplication by* f *in* $L_2(\sigma(\hat{A}), \mu_i)$*.*

b) $\sigma(\hat{A}) = supp(\{\mu_i\}_{i\in I})$ *where* $supp(\{\mu_i\}_{i\in I})$ *is the complement to the set of* $\lambda \in \mathbb{R}$ *for which there is an open set* $B_\lambda \subset \mathbb{R}$ *such that* $\lambda \in B_\lambda$ *and* $\mu_i(A_\lambda) = 0$ *for all* $i \in I$*.*

c) If $\mathcal{H}$ *is separable there exist a measure space* (M_A, Σ_A, μ_A) *with* $\mu_A(M_A) < +\infty$*, a bounded map* $F_A : M_A \to \mathbb{R}$ *and a unitary operator* $\hat{U}_A : \mathcal{H} \to L_2(M_A, \mu_A)$ *satisfying*

$$\left(\hat{U}_A\hat{A}\hat{U}_A^{-1}g\right)(x) = F_A(x)g(x)$$

for any $g \in \mathcal{H}$*.*

Note that in *c*) the measure is not uniquely determined by $\hat{A}$. This theorem is a more general version of the well known result about the splitting of an Hilbert space as direct sum of eigenspaces associated to a self adjoint operator.

Let us conclude this section observing that the spectral decomposition theorem tells that any self-adjoint operator (i.e. a possible quantum observable) can always be seen as an integral operator and that this decomposition is unique. The spectral measure allows to compute the quantum expectation as an ordinary expectation and, finally, the spectral representation theorem tells that the whole algebraic structure described in section 2.1.3 is present. This means that the complete probabilistic description of a *single quantum observable* is possible by using measure-theoretic probability.

2.1.5 Ordinary probability in Hilbert spaces.

We concluded the previous section observing that for a single quantum observable we can use measure-theoretic probability without problems. In this section we want to see how we can do the opposite: describe a measure-theoretic random variable with operators over an Hilbert space. In section 2.1.2 we have seen that to any measure-theoretic probability space, $(\Omega, \mathcal{E}, P)$, we may associate an abelian C^*-algebra of functions, $L_\infty(\Omega, \mathcal{E}, P)$, which can always be represented by using multiplicative operators over the Hilbert space $L_2(\Omega, \mathcal{E}, P)$, i.e. the commutative von Neumann algebra $\mathcal{V}_c(L_2(\Omega, \mathcal{E}, P))$, as shown in theorem 2. Such a theorem also tells that expectations with respect to a probability measure are nothing but states over $\mathcal{V}_c(L_2(\Omega, \mathcal{E}, P))$. We also observed that the Hilbert space $L_2(\Omega, \mathcal{E}, P)$ strongly depends on the probability measure of the underlying probability space, and so a change of the probability measure would change the Hilbert space. However, the spectral representation theorem suggests that we may find a "bigger Hilbert space" (namely $\mathcal{H} = \oplus_i \mathcal{H}_i$, as defined in the theorem) where this dependence on the probability measure seems to disappear. Finally, the spectral measure, introduced in section 2.1.4, seems to allow us to move the probabilistic content from the original probability measure to (functional of) function of this "bigger Hilbert space". In this section we want to study better this mechanism. More precisely, we want to discuss the following problem: how it is possible to construct explicitly an Hilbert space (independent on the probability measure), an operator and a *state* (defined as in definition 2) on a suitable algebra of operators on $\mathcal{H}$, which are capable to give the same statistical prediction about a random variable described in ordinary measure-theoretic setting.

Consider a probability space $(\Omega, \mathcal{E}, P)$, a measurable space $(M, \mathcal{M})$ and a random variable $X : \Omega \to M$ on it. As usual X induces a distribution ν_X such that $(M, \mathcal{M}, \nu_X)$ is a probability space. Algebraically, the random variable X can be seen as the map $x : L_\infty(\Omega, \mathcal{E}, P) \to L_\infty(M, \mathcal{M}, \nu_X)$. Clearly the random variable X can be seen also as the identity map on $L_\infty(M, \mathcal{M}, \nu_X)$, and any expectation $\mathbb{E}_P$ can be computed using a suitable state ϕ_{ν_X} over this algebra, i.e. $\mathbb{E}_P[f(X)] = \phi_{\nu_X}(f(X))$. This fact does not change if we represent the element $x \in L_\infty(M, \mathcal{M}, \nu_X)$, corresponding to the original random variable X, as a multiplicative operator $\hat{M}_x$ acting on $L_2(M, \mathcal{M}, \nu_X)$, i.e. if we consider the abelian von Neumann algebra of operators $\mathcal{A} := \{\hat{M}_f | f \in L_2(M, \mathcal{M}, \nu_X)\}$ on this Hilbert space. Clearly $L_2(M, \mathcal{M}, \nu_X)$ changes as we change the initial

probability measure P. Consider now the Hilbert space $\mathcal{H} = \oplus_i \mathcal{H}_i$ of the spectral representation theorem and a bounded operator $\hat{T} \in \mathcal{B}(\mathcal{H})$ on it with spectrum $\sigma(\hat{T})$. Then take the surjective isometry of the theorem, i.e. $\hat{U}_i : \mathcal{H}_i \to L_2(\sigma(\hat{T}), \mu_i)$. The idea is to use $\hat{U}_i$ to map $L_2(M, \mathcal{M}, \nu_X)$ in some $\mathcal{H}_i$ and to construct $\mathcal{H}$ from it. If we want to do that we can set:

a) $\sigma(\hat{T}) = M$,

b) $\mu_i = \nu_X$.

This allows to write that $\hat{U}_i : \mathcal{H}_i \to L_2(M, \nu_X)$ (we omit the σ-algebra $\mathcal{M}$ for simplicity). These requirements can be explained as follows. Since we want to represent with $\hat{T}$ the random variable X (note that $\hat{T}$ is not the operator $\hat{M}_x$ seen before) and encode the probabilistic content of $(M, \mathcal{M}, \nu_X)$ (and so of $(\Omega, \mathcal{E}, P)$) in some suitable object defined on $\mathcal{H}$, the requirement a) simply means that the set of eigenvalues of the operator coincides with the set of outcomes of the random variable. This tells us how to construct the operator $\hat{T}$ since the spectrum uniquely identifies the operator. The requirement b) is needed in order to encode the statistical information in functionals of elements of $\mathcal{H}$, allowing the Hilbert space, on which $\hat{T}$ is defined, to be capable to contain information about P. Note that at this level it is not clear what the meaning of the index i is (which is important for the construction of $\mathcal{H} = \oplus_i \mathcal{H}_i$) in the original probability space. Observing that this index determines the dimension and separability property of the Hilbert space, let us try to attach it to some feature of the random variable we want to represent. In particular, *we assume that i labels the outcome of X, i.e. $i = x \in M$.* This immediately implies that

$$\mathcal{H} = \bigoplus_{x \in M} \mathcal{H}_x$$

where $\oplus$ means direct sum or direct integral according to the cardinality of M, while the operator representing X is simply

$$\hat{T} := \int_M x \hat{P}^{(\hat{T})}(dx)$$

where $\hat{P}^{(\hat{T})}(dx)$ is the PVM having M as support. Note that in this way $\hat{T}\mathcal{H}_x \subset \mathcal{H}_x$, i.e. $\hat{P}^{(\hat{T})}(dx)|x\rangle = |x\rangle$ for any $|x\rangle \in \mathcal{H}_x$, as required by the spectral decomposition theorem. By construction the operator $\hat{T}$ has a non-degenerate spectrum and if M is a bounded subset of $\mathbb{R}$, the spectrum of $\hat{T}$ is bounded, implying that $\hat{T}$ is a bounded operator. Let us assume this for the rest of this section. The only thing that we miss is how to represent the probability distribution ν_X. At this point, we assume that the random variable X is discrete hence $\nu(x)$ can be interpreted as the probability to have $X = x$. In general on $(M, \mathcal{M}, \nu_X)$ we can describe, together with X, all random variables $f(X)$, where $f : M \to M$ are measurable and bounded functions, and they correspond to the operators $f(\hat{T})$. Hence, since $\nu(x)$ is a bounded and measurable function,

it can be represented as

$$\hat{\rho}_\nu := \nu(\hat{T}) = \int_M \nu(x)\hat{P}^{(\hat{T})}(dx).$$

Note that $\hat{\rho} \in \mathcal{B}_1(\mathcal{H})$, because $\nu(x)$ is a probability. In section 2.1.3 we have seen that a normal state $\phi(\cdot)$ on $\mathcal{B}(\mathcal{H})$ can be always written as $\mathrm{Tr}\,[\hat{\rho}\ \cdot\ ]$ for some trace class operator $\hat{\rho}$. The set of all the operators $f(\hat{T})$, equipped with the operation of sum and product of operators, forms a sub-algebra of $\mathcal{B}(\mathcal{H})$ which is in one-to-one correspondence (via the surjective isometry $\hat{U}_x$) with an abelian von Neumann algebra. Thus this set of operators form an abelian von Neumann algebra, which we label by $\mathcal{V}_T$. Then if we impose that states on $\mathcal{V}_T$ coincide with states of $L_\infty(M, \mathcal{M}, \nu_X)$ (inheriting all their properties), we must have

$$\mathbb{E}_{\nu_X}[f(X)] = \mathrm{Tr}\left[\hat{\rho} f(\hat{T})\right] \tag{2.1}$$

for any f measurable and bounded function on M. This implies that $\hat{\rho} = \hat{\rho}_\nu$. Note that, this time given ν_X (i.e. P) we can determine a unique object which encodes all the probabilistic information of the random phenomenon under study.

Heuristically, it seems that we can write the following formal "correspondence"

$$P(d\omega) \leftrightarrow \hat{\rho}_P$$
$$\int \cdots \leftrightarrow \mathrm{Tr}\,[\cdots]$$

which anyhow should be taken with care. First, additional difficulties are added if one drops the assumption that M is a bounded subset of $\mathbb{R}$. Another difficulty arises if we want to describe continuous randoms variable taking value on $\mathbb{R}$. These difficulties may be overcome, from a practical point of view, by seeing continuous unbounded operators as the limit of bounded operators with discrete spectrum: this is the solution that we will adopt in chapter 4 to deal with continuous unbounded random variables. Rigorous approaches to treat algebraically unbounded operators are available [27], while the notion of (generalized) eigenvalues for continuous unbounded operators can be formalized, from the mathematical point of view, using the *Gel'fand triples* [28].

2.2 Algebraic probability spaces.

We have seen that the usual measure-theoretic formulation of probability theory can be encoded in a satisfactory way in an abelian von Neumann algebra of functions. This suggests that a more general formulation of probability theory is possible in an algebraic context, allowing to obtain a non-commutative probability theory [29, 30, 31].

2.2.1 Basic definitions.

Some of the basic definitions we need in order to describe the algebraic approach to probability, have been already introduced in section 2.1.3. For notions like algebra, involution, state (and its classification) and representation, we will refer to that section.

Definition 9. *The pair $(\mathcal{A}, \omega)$ where $\mathcal{A}$ is a *-algebra with unit, and $\omega : \mathcal{A} \to \mathbb{C}$ a state on it, is called* algebraic probability space.

We will restrict our attention to the case where $\mathcal{A}$ is also C^* quickly. Note that the commutativity of $\mathcal{A}$ is not required in the definition. In section 2.1.3 we saw that in the abelian case, if $a \in \mathcal{A}$, $\omega(a) \in \mathbb{C}$ is its expectation value. If $a = a^*$ (i.e. a is self-adjoint), then one can prove that $\omega(a) \in \mathbb{R}$, thus self-adjoint elements of the algebra correspond to real-valued random variables. More generally, the elements of a generic algebra can be interpreted as random variables, as the following definition suggests.

Definition 10. *Given an algebraic probability space $(\mathcal{A}, \omega)$ and another *-algebra with unit $\mathcal{B}$, then an homomorphism $j : \mathcal{B} \to \mathcal{A}$ preserving the unit and the involution, is called an* algebraic random variable.

This definition is just an extension of the notion of random variable, used in the abelian case, to general algebras. As in ordinary probability theory, the algebraic random variable j, induces a state, $\omega_j = \omega \circ j$ called *distribution*, such that $(\mathcal{B}, \omega_j)$ is another algebraic probability space.

2.2.2 Two possible representations of an algebra.

Algebras are very abstract objects. For this reason, the notion of representation is very important. Here we will review the two basic representation theorems that we have at disposal, in order to pass from an abstract algebra to some concrete algebra.

A general result which allows to represent a generic abstract C^*-algebra with a concrete C^*-algebra of operators is the celebrated GNS theorem. First we need to introduce some terminology: a representation is called a **-representation* if it preserves the involution, while a vector $\psi \in \mathcal{H}$ is said to be *cyclic* for a representation π, if $\mathrm{span}\{\pi(a)\psi | a \in \mathcal{A}\}$ is a dense subspace of $\mathcal{H}$ [3, 12].

Theorem 6. *Let $\mathcal{A}$ be a C^*-algebra with unit and $\omega : \mathcal{A} \to \mathbb{C}$ a state. Then:*

i) there exists a triple $(\mathcal{H}_\omega, \pi_\omega, \Psi_\omega)$ where $\mathcal{H}_\omega$ is an Hilbert space, $\pi_\omega : \mathcal{A} \to \mathcal{B}(\mathcal{H}_\omega)$ is a *-representation of $\mathcal{A}$ on the C^*-algebra of bounded operators on $\mathcal{H}_\omega$, and $\Psi_\omega \in \mathcal{H}_\omega$ is a vector, such that:

   a) Ψ_ω is a unit vector, cyclic for π_ω;

   b) $\langle \Psi_\omega | \pi_\omega(a) \Psi_\omega \rangle = \omega(a)$ for any $a \in \mathcal{A}$.

ii) If $(\mathcal{H}, \pi, \Psi)$ is a triple such that:

a) $\mathcal{H}$ *is an Hilbert space,* $\pi : \mathcal{A} \to \mathcal{B}(\mathcal{H})$ *is a* **-representation and* $\Psi \in \mathcal{H}$ *is a unit vector cyclic for* π*;*

b) $\omega(a) = \langle \Psi | \pi(a) \Psi \rangle$*;*

then there exits a unitary operator $\hat{U} : \mathcal{H} \to \mathcal{H}_\omega$ *such that* $\Psi = \hat{U}\Psi_\omega$ *and* $\pi_\omega(a) = \hat{U}\pi(a)\hat{U}^{-1}$ *for any* $a \in \mathcal{A}$*.*

Note that in general, $\mathcal{H}_\omega \neq \mathcal{H}_{\omega'}$ for $\omega \neq \omega'$. If $\mathcal{H}_\omega$ is finite dimensional, then $\mathcal{A}$ is also a von Neumann algebra; in the infinite dimensional case, this is not true anymore. Because of this theorem, we will always use algebras of operators over some Hilbert space instead of abstract objects. For completeness, we mention that a GNS theorem for *-algebras with unit is also available (see Th. 14.20 in [3]). The contents of such a theorem are more or less the same of the GNS theorem presented here. However, it allows to represent elements of a *-algebra with unbounded operators which are closable over a state dependent domain $\mathcal{D}_\omega$. This version of the GNS theorem allows to threat in a more rigorous way unbounded random variables using unbounded operators over some Hilbert space, as mentioned at the end of section 2.1.5. If $\mathcal{A}$ is commutative, we have another result which allows to represent abstract C^*-algebras with continuous functions over some space: the commutative Gel'fand-Naimark theorem[3].

Theorem 7. *Any commutative* C^**-algebra with unit* $\mathcal{A}$ *is* **-isomorphic (i.e. the involution is preserved under the isomorphism) to the commutative* C^**-algebra with unit of continuous functions on* $\Delta(\mathcal{A})$*,* $C(\Delta(\mathcal{A}))$ *(which is* C^* *with respect to the norm* $\| \cdot \|_\infty$*), where*

$$\Delta(\mathcal{A}) := \{\phi : \mathcal{A} \to \mathbb{C} \mid \phi(ab) = \phi(a)\phi(b)\ \forall a, b \in A,\ \phi \text{ non trivial}\}.$$

Such **-isomorphism (called Gelfand's transform) is isometric.*

Note that the GNS theorem holds also for the commutative case but only in the abelian case we can construct the measure-theoretic probability space. This fact has important consequences on the concrete interpretation of algebraic probability spaces.

2.2.3 Some effect of non-commutativity.

Let us discuss some differences between the commutative and the non-commutative case, which are relevant for quantum theory, but the list of differences does not end here.

i) *The lattice of projectors.* Given a *-algebra $\mathcal{A}$, we call $p \in \mathcal{A}$ orthogonal projector if $p = p^* = p^2$ and the set of projectors on $\mathcal{A}$ will be labeled by $\mathcal{P}(\mathcal{A})$. From the abelian case, we have seen that the σ-algebra of the associated probability space can be constructed from this structure ($\mathcal{P}(\mathcal{A}) = \tilde{\mathcal{E}}$, in section 2.1.2). From the mathematical logic point of view, this means that in the abelian case $\mathcal{P}(\mathcal{A})$ has the structure of a distributive

lattice (i.e. a Boolean lattice which is always isomorphic to a Boolean σ-algebra). In the non-commutative case, this structure changes: $\mathcal{P}(\mathcal{A})$ has, in general, the structure of an orthomodular lattice (modularity depends on the type of factor of $\mathcal{A}$). The practical consequence is that we cannot interpret the propositions about "non-commutative random phenomena" using ordinary propositional calculus (the logical connectivities AND and OR are problematic) which is exactly what happens in quantum logic.

ii) *The CHSH inequality.* Consider two von Neumann algebras $\mathcal{A}$ and $\mathcal{B}$ (which are automatically C^*) that are mutually commuting and $\mathcal{A}, \mathcal{B} \subset \mathcal{B}(\mathcal{H})$. Let ω be a normal state for both algebras (hence a positive normalized linear functional from $\mathcal{B}(\mathcal{H})$ to $\mathbb{C}$) and define

$$\beta(\omega, \mathcal{A}, \mathcal{B}) := \sup \omega \left(a_1[b_1 + b_2] + a_2[b_1 - b_2] \right)$$

where the sup is taken over all $a_1, a_2 \in \mathcal{A}$ and $b_1, b_2 \in \mathcal{B}$ having norm less than 1. Then if at least one of these two algebras is abelian one can prove that $\beta(\omega, \mathcal{A}, \mathcal{B}) \leqslant 2$ for all states ω. When both $\mathcal{A}$ and $\mathcal{B}$ are non-abelian, then this bound can be violated: it is known that the maximal violation is $\beta(\omega, \mathcal{A}, \mathcal{B}) = 2\sqrt{2}$ [32]. The degree of violation depends on the type of algebra: for two mutually commuting, non-abelian von Neumann algebra, if the Schlieder property holds [29] (i.e. $ab = 0$ for $a \in \mathcal{A}$ and $b \in \mathcal{B}$ implies either $a = 0$ or $b = 0$) then there exists a normal state which maximally violate the inequality. This is nothing but the well known CHSH inequality of quantum mechanics [33].

iii) *Dispersion free state.* Let $\mathcal{A}$ be a von Neumann algebra, we say that a state ω is *dispersion-free* if $\omega((a - \omega(a))^2) = 0$ for all $a \in \mathcal{A}$. In the abelian case, a pure state can be characterised as the states for which $\omega(ab) = \omega(a)\omega(b)$ holds for any $a, b \in \mathcal{A}$. Cleary, the pure states in abelian case are dispersion-free. In the non-abelian case dispersion-free states do not exist. In quantum mechanics this is a well known fact, and it is called Heisenberg uncertainty principle. We will use this fact in chapter 3 to study a possible characterization of non-commutativity.

Other differences which are relevant from the physical point of view, between commutative and non-commutative case are, for example, the way one composes two algebras, or the algebraic generalization of the notions of independence and conditional expectation [29].

Chapter 3

Signature of non-commutativity

In the previous chapter, we have seen how it is possible to represent random variables on some probability space with operators on a Hilbert space. Different random variables on the same probability space corresponds to different operators, but all these operators are diagonal on the same basis (i.e. they all commutes). If we consider two random variables defined on two *different* probability spaces, this is not a priori true. In section 3.1 we present a tool, the entropic uncertainty relations, which can be used to test if two random variables are diagonal on the same basis or not. In section 3.2 a method, to map two random variables fulfilling an entropic uncertainty relation with operators on the same Hilbert space, is described.

3.1 Entropic uncertainty relations

The non-existence of dispersion-free states in a non-commutative probability space suggests that we cannot have delta-like marginals (of some joint probability distribution) for all the random variables of our algebra. Following this intuitive idea, we introduce a natural measure of the "spread" of a given probability distribution and then we discuss how this measure behaves in presence of non-commuting random variables.

3.1.1 Entropy in information theory

A natural measure we can use to quantify the spread of a given probability distribution is the Shannon entropy. Such entropy is the basic notion of classical information theory and for this reason it is sometimes claimed (especially in quantum physics [34]) that it cannot be used for non-commutative probability spaces. From the mathematical point of view, this claim is not true, simply because any non-commutative algebra always admits a commutative sub-algebra

where classical information theory can be applied. In addition, the Shannon entropy is not sensitive to the origin of probability [35]: it is associated to a *single* random variable.

We will introduce the Shannon entropy as done in [36], which is different to Shannon's original approach. Naïvely speaking, information quantifies the number of things we do not know about a given random phenomenon. In other words, information quantifies the unexpectedness of an event E relative to a random variable X. Let $I_X(E)$ be a measure of this unexpectedness; it is reasonable to require that

i) $I_X(E)$ is a function of the probability of E to occur, and not directly a function of the event E;

ii) $I_X(E)$ is a smooth function of the probability;

iii) if E and F are two disjoint events (hence independent), then $I_X(E,F) = I_X(E) + I_X(F)$.

It is not difficult to see that $I_X(E) = k\log_b(P(E))$ fulfils the three requirements. Typically $k = -1$ and $b = e$ are chosen, and this function is called *information content* of the event E. The Shannon entropy can be thought as the expectation value of the information content of the elementary events ($E = \{\omega\}$), i.e. $H(X) := \mathbb{E}[I_X]$. Consider a discrete random variable taking values over a discrete set $\{x_1, \cdots, x_N\}$, then $H(X)$ is just

$$H(X) = -\sum_{i=1}^{N} p_i \log p_i \tag{3.1}$$

where $p_i := P[X = x_i]$ where x_i is one of the possible outcomes of the random variable X. To better understand how $H(X)$ quantifies the spread of a distribution, let us consider the case of a certain event (determinism). Suppose we know that the event $E := \{X = k\}$ is always true. Then clearly $p_i = \delta_{ik}$, which gives $H(X) = 0$: the event is certain so our unexpectedness is zero (note we assumed $0\log 0 = 0$, as typically done in information theory). Since $-x\log x$ is always positive for $x \in [0,1]$ it is not difficult to understand that $H(X) = 0$ only for delta-like distributions. In addition *iii*) suggests that the more elementary events contribute to $H(X)$, namely the more elementary events have non zero probability, the lager its value will be. In this sense we can use $H(X)$ to quantify the spread of a probability distribution.

As already observed at the beginning of this section, the only requirement needed on $\{p_i\}$ in order to define $H(X)$ is that they come from a σ-additive, normalised measure, which happens in any algebraic probability space (σ-additive means that the measure remains finite even for countable unions of events). In the non-commutative case some usual properties of H do not hold: as a rule of thumb, all properties which depend on vectors of random variables (like (X,Y)) should be checked with care.

3.1.2 What is an entropic uncertainty relation?

Entropic uncertainty relations are a way to introduce an uncertainty principle for generic observables in quantum mechanics. Here we will review the known bounds which are interesting for our discussion. The results presented here can be found in [37, 38, 39, 40, 41, 42].

Entropic uncertainty relations are relevant relations between self-adjoint operators in an Hilbert space. Let us start with a "preliminary definition".

Definition 11. *Consider a Hilbert space $\mathcal{H}$ and two self-adjoint operators on it, $\hat{A}$ and $\hat{B}$. Then if*

$$H_{\hat{\rho}}(\hat{A}) + H_{\hat{\rho}}(\hat{B}) \geqslant C \qquad \forall \hat{\rho} \in \mathcal{B}_1(\mathcal{H}),$$

where C is a fixed positive number independent on $\hat{\rho}$, we say that $\hat{A}$ and $\hat{B}$ fulfil an entropic uncertainty relation.

In the definition above, $H_\rho(\hat{A})$ is the Shannon entropy computed with the probability distribution $\mu_{\hat{\rho}}(\cdot) = \mathrm{Tr}\left[\hat{\rho}\hat{P}^{(\hat{A})}(\cdot)\right]$, where $\hat{P}^{(\hat{A})}(\cdot)$ is the PVM associated to $\hat{A}$. The same holds for $\hat{B}$. We can clearly see why this definition should be taken with care: according to definition 3.1, what is the Shannon entropy if the spectrum of the operator is continuous? We will provide a more rigorous definition in the next section, for the moment we just observe that if the Hilbert space is finite dimensional this definition works (because all operators are compact). Typically, in quantum information, one is interested in finding the *bound* C (i.e. in the minimisation problem $\min_{\hat{\rho}\in\mathcal{B}_1(\mathcal{H})}[H_{\hat{\rho}}(\hat{A}) + H_{\hat{\rho}}(\hat{B})]$).

As a first example of the aforementioned bound (i.e. of entropic uncertainty relation), let us consider the following theorem [41].

Theorem 8. *Let $H_{\hat{\rho}}(\hat{A})$ and $H_{\hat{\rho}}(\hat{B})$ be the Shannon entropies associated to two non-degenerate self-adjoint operators $\hat{A}$ and $\hat{B}$ over a finite dimensional Hilbert space $\mathcal{H}$. Assume that $\{|\phi_a\rangle\}_{a\in\sigma(\hat{A})}$ and $\{|\psi_b\rangle\}_{b\in\sigma(\hat{B})}$ are basis of eigenvectors of $\hat{A}$ and $\hat{B}$ respectively. Then $\forall\hat{\rho} \in \mathcal{B}_1(\mathcal{H})$*

$$H_{\hat{\rho}}(\hat{A}) + H_{\hat{\rho}}(\hat{B}) \geqslant -2\log(\max_{a,b} |\langle\phi_a|\psi_b\rangle|).$$

We can see that, if the scalar product between eigenvectors is less than 1 (i.e. $\hat{A}$ and $\hat{B}$ cannot be diagonalised at the same time, $[\hat{A}, \hat{B}] \neq 0$) the bound is non-zero. This result can be generalized to the case of POVMs (which are defined as PVMs except that *ii*) in definition 7 is not required to hold), which encodes, as special case, that of degenerate operators [42]. Because we do not need all this generality, we consider the PVM case only, which can be obtained from theorem 8 by replacing the argument of the logarithm with $\max_{ab}(\|\hat{P}_a^{(\hat{A})}\hat{P}_b^{(\hat{B})}\|)$ (here $\hat{P}_a^{(\hat{A})}$ is the projector on the eigenspace associated to $a \in \sigma(\hat{A})$; same for $\hat{P}_b^{(\hat{B})}$). It is worth to say that theorem 8, and its generalisations, is a consequence of the Riesz-Thorin interpolation theorem for L_p-spaces [43]. From the physical

point of view this means that no physical assumption is needed to derive this theorem: in this sense it does not depend on the physical interpretation of the mathematical objects. Also in infinte dimensional Hilbert spaces we have a similar theorem. Nevertheless this time we need to face the problem that operators do not admit in general only a point spectrum. To include also the continuos-spectrum case, avoiding to introduce the "continuous version" of the Shannon entropy (i.e. the differential entropy [44], which is not properly a generalisation) the idea is simply to partition the spectrum. Given a generic operator $\hat{A}$ on an infinite dimensional Hilbert space $\mathcal{H}$, a *partition of the spectrum* is a collection of set $\{E_i\}_{i\in I}$ such that $\sigma(\hat{A}) = \cup_{i\in I} E_i$. Given this partition of the spectrum and $\hat{\rho} \in \mathcal{B}_1(\mathcal{H})$, we can associate to it a set of probabilities $\{p_i^{(\hat{\rho})}\}_{i\in I}$ computed via the formula $p_i^{(\hat{\rho})} = \mathrm{Tr}\left[\hat{\rho}\hat{P}^{(\hat{A})}(E_i)\right]$. Using this distribution we can compute $H_{\hat{\rho}}(\hat{A})$. Then we have the following theorem [40].

Theorem 9. *Let $H_{\hat{\rho}}(\hat{A})$ and $H_{\hat{\rho}}(\hat{B})$ be the Shannon entropies associated to self-adjoint operators $\hat{A}$ and $\hat{B}$ over (possibly infinite dimensional) Hilbert space $\mathcal{H}$. Assume that $\{E_i\}_{i\in I}$ and $\{F_j\}_{j\in J}$ are two different partitions of the spectrum of $\hat{A}$ and $\hat{B}$, respectively. Then $\forall \hat{\rho} \in \mathcal{B}_1(\mathcal{H})$*

$$H_{\hat{\rho}}(\hat{A}) + H_{\hat{\rho}}(\hat{B}) \geqslant 2\log\left(\frac{2}{\sup_{i,j} \|\hat{P}^{(\hat{A})}(E_i) + \hat{P}^{(\hat{B})}(F_j)\|}\right).$$

Again, we can see that if $\hat{A}$ and $\hat{B}$ commute, the RHS vanishes (since $1 \leqslant \sup_{i,j} \|\hat{P}^{(\hat{A})}(E_i) + \hat{P}^{(\hat{B})}(F_j)\| \leqslant 2$ and the upper bound is reached if and only if $\hat{P}^{(\hat{A})}(\cdot)$ and $\hat{P}^{(\hat{B})}(\cdot)$ have common eigenvectors).

Using the two theorems presented here above, we are able to relate non-commutativity between operators and the probability measures associated with them (i.e. the spectral measure) using entropic uncertainty relations. In the next section, we will formalize these facts in a C^*-probability space, proving that there is a link between the non-commutativity of the algebra and the properties of the probability measures associated with states on it, which can be characterized using entropic uncertainty relations.

3.1.3 Algebraic generalisation

In this section we will extend the definition of the Shannon entropy to a generic C^*-algebra. Consider a C^*- probability space $(\mathcal{A}, \omega)$. Using the GNS theorem we may equivalently consider the triple $(\mathcal{H}_\omega, \pi_\omega, \Psi_\omega)$. For any self-adjoint element $a \in \mathcal{A}$, we may consider the bounded operator $\hat{A}_\omega := \hat{\pi}_\omega(a)$ acting on $\mathcal{B}(\mathcal{H}_\omega)$. The spectral theorem ensures that there exist a PVM $\{\hat{P}^{(\hat{A}_\omega)}(E)\}_{E\subset\sigma(\hat{A}_\omega)}$ associated to $\hat{A}_\omega$, thus the probability that a takes value in E is $\langle\Psi_\omega|\hat{P}^{(\hat{A}_\omega)}(E)\Psi_\omega\rangle$. Nevertheless we cannot use this probability directly in the definition of the Shannon entropy because in general the spectrum may have a continuous part. It is a known fact that if $\hat{A}$ is a bounded self-adjoint operator its spectrum can be split

as $\sigma(\hat{A}) = \sigma_p(\hat{A}) \cup \sigma_c(\hat{A})$, where $\sigma_p(\hat{A})$ and $\sigma_c(\hat{A})$ are respectively the point and the continuous part of the spectrum. Note that at the algebraic level the classification of the spectrum may depend on the state ω. To introduce a well defined notion of entropy at the algebraic level, we have to find a way to deal with the continuous part of the spectrum. Mimicking what we did in section 3.1.2, we introduce a partition of the continuous part of the spectrum $\{E_i\}_{i\in I_\omega}$ (we always assume I_ω at most countable). Let us label with ε a generic partition, then given ε we can always construct a probability distribution $\{p_i^{(\omega,\varepsilon)}\}_{i\in\sigma_p(\hat{A}_\omega)\cup I_\omega}$ for $a \in \mathcal{A}$ as

$$p_i^{(\omega,\varepsilon)} := \begin{cases} \langle\Psi_\omega|\hat{P}^{(\hat{A}_\omega)}(\{i\})\Psi_\omega\rangle & \text{if } i \in \sigma_p(\hat{A}_\omega) \\ \langle\Psi_\omega|\hat{P}^{(\hat{A}_\omega)}(E_i)\Psi_\omega\rangle & \text{if } i \in I_\omega. \end{cases}$$

Note that these probabilities clearly depend on the partition chosen, as well as on the state. Using the probability distribution constructed in this way, we can apply without problems the definition of the Shannon entropy to any self-adjoint element of $\mathcal{A}$.

Definition 12. *Let $(\mathcal{A},\omega)$ be a C^*-probability space. Fix a partition ε and constructs for some self-adjoint $a \in \mathcal{A}$ the probability distribution $\{p_i^{(\omega,\varepsilon)}\}_{i\in\sigma_p(\hat{A}_\omega)\cup I}$, where $\hat{A}_\omega = \hat{\pi}_\omega(a)$. Then the* ε-Shannon entropy *of $a \in \mathcal{A}$ is given by*

$$H_\omega(a;\varepsilon) := -\sum_{i\in\sigma_p(\hat{A}_\omega)\cup I_\omega} p_i^{(\omega,\varepsilon)} \log p_i^{(\omega,\varepsilon)}$$

Since the probabilities depend on the partition, the entropy depends also on the partition of the spectrum as well. Thanks to this definition we can define in a proper manner an entropic uncertainty relation in an algebraic contest.

Definition 13. *Let $\mathcal{A}$ be a C^*-algebra and consider two random variables $a, b \in \mathcal{A}$ on it. Choose two partitions (different in general) ε and δ for a and b, respectively. If for any ω,*

$$H_\omega(a;\varepsilon) + H_\omega(b;\delta) \geqslant C(\varepsilon,\delta),$$

where $C(\varepsilon,\delta) \in \mathbb{R}^+/\{0\}$ is a constant which may depend on the partitions but not on the state, we say that a and b fulfil an (ε,δ)-entropic uncertainty relation.

Note that the important part of this definition is the independence on the state of the constant $C(\varepsilon,\delta)$: the LHS is bigger than this constant *for any possible state*. We need to introduce an ordering relation between partitions of the spectrum, which is nothing but the notion of "finer partition".

Definition 14. *Let $\varepsilon = \{E_i\}_{i\in I}$ and $\varepsilon' = \{E'_j\}_{j\in J}$ be two partitions. We say that ε* is finer than *ε', written $\varepsilon \subset \varepsilon'$, if*

i) $E'_j = \cup_{i\in I_j} E_i$ for some $I_j \subset I$;

ii) $I = \cup_{j\in J} I_j$.

Intuitively, a partition ϵ is finer than a partition ϵ', if combining in a suitable way the sets of ϵ, we can construct all the sets of ϵ'. The requirements i) and ii) are simply the conditions under which this combination is possible. In what follows, if we need to talk repeatedly of two partitions, say ε and δ, we will use the symbol (ε, δ). The writing $(\varepsilon, \delta) \subset (\varepsilon', \delta')$ means $\varepsilon \subset \varepsilon'$ and $\delta \subset \delta'$. At this point we may state the following theorem, which relates the non-commutativity of the C^*-algebra and the presence of entropic uncertainty relations.

Theorem 10. *Let $\mathcal{A}$ be a C^*-algebra and take two self-adjoint elements $a, b \in \mathcal{A}$. If for two partitions (ε, δ) an entropic uncertainty relation holds, namely*

$$H_\omega(a;\varepsilon) + H_\omega(b;\delta) \geqslant C(\varepsilon, \delta),$$

with $C(\varepsilon, \delta) > 0$, and this happens for any possible state ω over $\mathcal{A}$, then $[a, b] \neq 0$.

Idea of Proof. Here we give an intuitive proof of this theorem. A detailed mathematical proof can be found in appendix.

We have already seen that dispersion-free states do not exist in a non-commutative probability space. Consider a non-abelian C^*-algebra $\mathcal{A}$, and $a, b \in \mathcal{A}$, such that $[a, b] \neq 0$. Given a state ω, let $\hat{A}_\omega = \hat{\pi}_\omega(a)$ and $\hat{B}_\omega = \hat{\pi}_\omega(b)$ be the two associated GNS representations acting on $\mathcal{H}_\omega$. Assume that the spectrum is purely continuous for such representations, the discrete case can be thought as a sub-case of this. Since $[a, b] \neq 0$, we cannot find a state ω, which has a delta-like probability distributions (i.e. spectral measures) for both $\hat{A}_\omega$ and $\hat{B}_\omega$, for any possible partitions of the two spectra we can consider. The best we can do, is to choose ω which has a delta-like probability distribution for *only one* of the two random variables: hence, we are in the situation of Figure 3.1. Now, take (ε, δ) and (ε', δ') such that $(\varepsilon', \delta') \subset (\varepsilon, \delta)$. If ω induces the two probability distributions in the picture, we can see that:

i) There are partitions where the two probability distributions $\{p_i^{(\omega,\varepsilon)}(a)\}_{i\in I}$ and $\{p_i^{(\omega,\delta)}(b)\}_{i\in I}$, have a delta-like shape (i.e. all the probabilities are 0 except for one set of the partition). This is the case of the partitions (ε, δ) in the Figure 3.1. For these partitions, we have no entropic uncertainty relations.

ii) There are partitions where only one of the two probability distributions still have a delta-like shape. This is the case of the partitions (ε', δ') in the Figure 3.1. In this case, we have an entropic uncertainty relation.

Hence, if dispersion-free states do not exist (i.e. the algebra is non-commutative) we can find a partition for which an entropic uncertainty relation holds. On the other hand, it is not difficult to see that if an entropic uncertainty relation is found for a partition (ε, δ), automatically it holds for all the finer partitions $(\varepsilon', \delta') \subset (\varepsilon, \delta)$. Thus there are no dispersion-free states on the algebra, which means it is not abelian. An alternative proof of more algebraic flavor can be found in [45] □

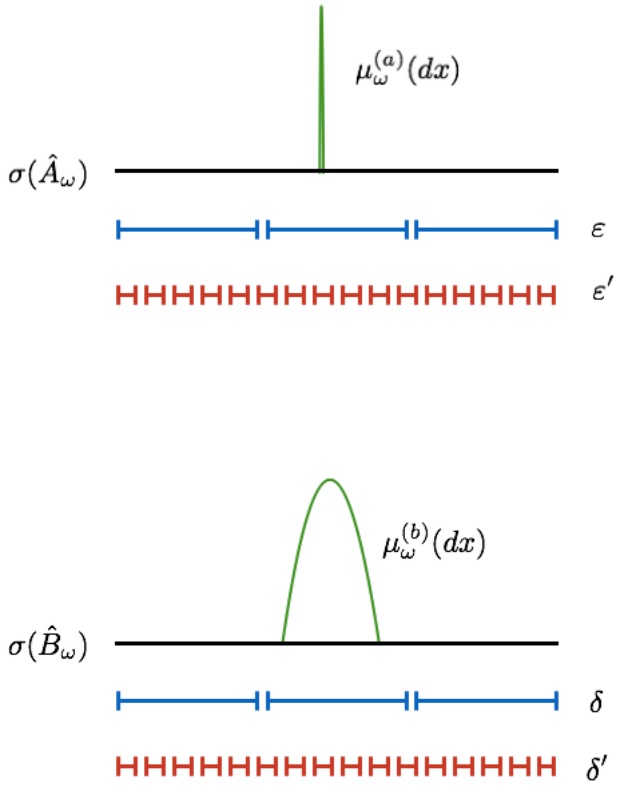

Figure 3.1: The case of a state which is not dispersion-free. As one can see we can have for all possible partitions, a delta-like probability distribution for the algebraic random variable a. This cannot happen for b, when $[a, b] \neq 0$: there are partitions for which the probability distribution of b cannot be delta-like.

Note that this theorem gives a way to test if two algebraic random variables commute or not, using purely probabilistic concepts. This result generalizes in the algebraic contest the content of the theorems 8 and 9 seen in the previous section. Note that this theorem does not say anything about the bound (contrary to theorems 8 and 9), but it asserts only that if it exists for all states then the algebra is non-commutative. A similar result, where variances are used instead of entropies can be found in [12]. As we will see, entropies turns out to be more easy to use for the problem concerning this thesis, mainly because variances depends explicitly on the possible values of the spectrum while entropies depend only on the probability distributions. The main difficulty in use theorem 10 to test non-commutativity, is that the LHS of the inequality must be varied over all the possible states. Fortunately, a further simplification can be done. Consider a state $\omega : \mathcal{A} \to \mathbb{C}$ and let $\mathcal{S}(\mathcal{A})$ be the set of all states on $\mathcal{A}$. Then we

say that the state is *pure* if it cannot be written as convex combination of other states (i.e. $\nexists \omega_1, \omega_2$ such that $\omega(\cdot) = \lambda\omega_1(\cdot) + (1-\lambda)\omega_2(\cdot)$ for some $\lambda \in [0,1]$), otherwise it is said to be *mixed*. Let $\mathcal{S}_p(\mathcal{A})$ denote the set of all the pure states on $\mathcal{A}$.

Corollary 1. *Consider a C^*-algebra $\mathcal{A}$ and take two self-adjoint elements $a, b \in \mathcal{A}$. If an entropic uncertainty relation between a and b holds for all $\omega \in \mathcal{S}_p(\mathcal{A})$, then it also holds for any state in $\mathcal{S}(\mathcal{A})$.*

Proof. See appendix. □

3.2 Non-commutativity from ordinary measure-theoretic probability

If we want to model random phenomena, we can use two (apparently) different mathematical structures: a measure space, $(\Omega, \mathcal{E}, P)$, or an algebra with a state $(\mathcal{A}, \omega)$. Here we want to discuss a possible method to obtain a non-commutative behavior of probability starting from a collection of probability spaces, i.e. how obtain $(\mathcal{A}, \omega)$ starting from $(\Omega, \mathcal{E}, P)$.

3.2.1 The general method in the algebraic setting

Before we explain the method in the Hilbert space setting, let us discuss the idea from the algebraic point of view. Suppose we have two real random variables a and b. Instead of describing them in the measure-theoretic language, we describe them using the abelian algebras that they generate, say $\mathcal{A}_a$ and $\mathcal{A}_b$ respectively. This means that the algebra $\mathcal{A}_a$ is the abelian algebra generated by the identity, a and all its polynomial $p(a)$. The same for $\mathcal{A}_b$. On these two algebras, we can define states: we label by ω_a states on $\mathcal{A}_a$ and by ω_b states on $\mathcal{A}_b$. Now we assume the following: *there exist a 1-1 map between states on $\mathcal{A}_a$ and states on $\mathcal{A}_b$*. This means that to a given state ω_a on $\mathcal{A}_a$, we can associate in a unique way a state ω_b on $\mathcal{A}_b$. This map allows neglecting the labels a and b in the symbol of the state ω. Let us now set $\mathcal{A}$ as the smallest C^*-algebra containing both $\mathcal{A}_a$ and $\mathcal{A}_b$ as subalgebras (i.e. the algebra generated by the identity, a, b and polynomials $p(a,b)$). By theorem 10, if we can prove that $H_\omega(a)+H_\omega(b) \geqslant D$ for all ω (the dependence on the partitions is omitted for simplicity), we know that $\mathcal{A}$ is a non-abelian algebra. Since $\mathcal{A}$ is non-abelian, the GNS theorem allows to represent it as an algebra of bounded operators on a suitable Hilbert space. We cannot represent $\mathcal{A}$ as an algebra of functions. Thus, starting from two ordinary random variables defined in two *different* probability spaces, we end up with an Hilbert space description where both random variables are present as operators, but they do not commute.

3.2.2 The Hilbert space structure from the entropic uncertainty relation

Previously we presented in algebraic setting a method to obtain a non-commutative probability space starting from two ordinary measure-theoretic probability spaces. In this section we will explain how to construct a concrete algebraic probability space (i.e. already represented on an Hilbert space) starting from the probability spaces of two random variables, assuming that they fulfill an entropic uncertainty relation. To keep the discussion simple, we restrict ourself to finite discrete random variables.

Let $X : (\Omega, \mathcal{E}, P) \to (M, \mathcal{M})$ and $Y : (\Omega', \mathcal{E}', P') \to (N, \mathcal{N})$ be two discrete random variables and as usual $\mu_X := P \circ X^{-1}$ and $\nu_Y := P' \circ Y^{-1}$ label their probability distributions. We assume the following conditions:

i) we have a 1-1 map between P and P', i.e. to each probability distribution μ_X for X we can associate a corresponding probability distribution ν_Y for Y and viceversa;

ii) M and N have the same cardinality, i.e. X and Y have the same number of possible distinct outcomes;

iii) X and Y fulfil an entropic uncertainty relation, namely for any μ_X and ν_Y

$$H(X) + H(Y) \geqslant D$$

with $D > 0$.

In section 2.1.5, we have seen that a consistent way to represent a random variable on an Hilbert space is obtained by using the spectral representation theorem and the spectral decomposition theorem. Thus, given the random variable X, we can construct the operator

$$\hat{T}_X := \sum_{x \in M} x |x\rangle\langle x|$$

defined on the Hilbert space

$$\mathcal{H}_X := \bigoplus_{x \in M} \mathcal{H}_x.$$

By construction $\sigma(\hat{T}_X) = M$ and $\{|x\rangle\}_{x \in M}$ is a basis of $\mathcal{H}_X$. The assumption $i)$ ensures that, in general, the operator representing the random variable X cannot be used to describe also the random variable Y. More precisely, as we have seen in section 2.1.5, if $\{|x\rangle\}_{x \in M}$ is the basis on which $\hat{X}$ is diagonal, we can represent over this basis all the random variables that are functions of X. Hence, thanks to the assumption $i)$, we can go beyond the simple case of $X = f(Y)$ (or $Y = g(X)$), where the map between P and P' is given by a simple change of variables. The random variable Y, being defined on a different probability space, cannot be seen in general as a function of X. Repeating the

whole construction for the random variable Y, also in this case we can define an operator

$$\hat{S}_Y = \sum_{y \in N} y|y\rangle\langle y|$$

on the Hilbert space

$$\mathcal{H}_Y := \bigoplus_{y \in N} \mathcal{H}_y.$$

Note that this Hilbert space is not in general $\mathcal{H}_X$. Again $\sigma(\hat{S}_Y) = N$ and $\{|y\rangle\}_{y \in N}$ is a basis of $\mathcal{H}_Y$ by construction. The assumption *ii*) ensures that the two Hilbert spaces have equal dimension, and so there exists a unitary map $\hat{U} : \mathcal{H}_X \to \mathcal{H}_Y$. This means that we can map the operator $\hat{S}_Y$ on $\mathcal{H}_X$ and $\hat{T}_X$ on $\mathcal{H}_Y$. Let us consider the first case, since the second is equivalent. The operator representing Y on $\mathcal{H}_X$ is

$$\begin{aligned} \hat{T}_Y :&= \hat{U}\hat{S}_Y\hat{U}^* \\ &= \hat{U}\sum_{y \in N} y|y\rangle\langle y|\hat{U}^* \\ &= \sum_{y \in N} y\hat{U}|y\rangle\langle y|\hat{U}^*. \end{aligned}$$

Let us set $|Uy\rangle := \hat{U}|y\rangle$ and note that the operator $\hat{T}_Y$ is diagonal in this basis. Since unitary transformation maps a basis into a basis, also $\{|Uy\rangle\}_{y \in N}$ is a basis and in particular it is the image under $\hat{U}$ of the basis in which $\hat{S}_Y$ is diagonal. At this point, the key observation is that if the assumption *iii*) is true, then the basis $\{|x\rangle\}_{x \in M}$ and the basis $\{|Uy\rangle\}_{y \in N}$ do not coincide. Indeed, the entropic uncertainty relation assumed, together the theorem 8, allows us to write that

$$-2\log(\max_{x,y} |\langle x|Uy\rangle|) \geqslant D$$

(with the equality only if one can prove that the bound is optimal) so $\max_{x,y} |\langle x|Uy\rangle| \leqslant e^{-D/2} < 1$, since D is never zero. Another way to say this is that $\hat{U}$ is not the identity transformation. Note that we can reach this conclusion only because we assumed the existence of an entropic uncertainty relation: if $D = 0$, then we cannot exclude that $|\langle x|Uy\rangle| = 1$ for some x, y (i.e. they are the same basis).

The conclusion is that, given the entropic uncertainty relations, the two operators $\hat{T}_X$ and $\hat{T}_Y$ do not commute, thus we can describe *both* random variables only on a common non-commutative algebraic probability space (i.e. with operators on an Hilbert space). How on this structure is represented the map between P and P', i.e. the state, will be discussed in the next section.

3.2.3 Conditional probabilities and representation of states

In the previous section, we have seen that starting from two random variables defined on two different probability spaces, if an entropic uncertainty relation

holds, we can construct a non-commutative algebraic probability space where both the random variables are represented by non-commuting operators. Essential for this construction is the presence of two distinct probability space, one for each random variable. Here we want to discuss how this condition can be met in a rather simple way and the consequences of this on the map between P and P'.

Given a probability space $(\Omega, \mathcal{E}, P)$ and a collection of events, conditioning with respect to each of these events, generates a collection of probability spaces. More precisely, conditional probability in measure-theoretic setting is defined via the Bayes formula

$$P_C(A) := P(A|C) = \frac{P(A \cap C)}{P(C)} \qquad A, C \in \mathcal{E}$$

P_C is again a probability measure on Ω, but this time it depends on the event C also. Given a family of events $\mathcal{C} := \{C_i\}_{i \in I}$, then by conditioning we obtain the collection of probability spaces $(\Omega, \mathcal{E}(\Omega)_{C_i}, P_{C_i})_{C_i \in \mathcal{C}}$. The trivial case $\mathcal{C} = \{C\}$ coincides with the usual measure-theoretic description, however in the more general case, this collection is called *contextual probability space* [46, 47, 48] while the C_is are called *context*. In the general contextual probability theory, not all the context are elements of a σ-algebra (i.e. events, as in this case) despite it is always possible to do so [49]. This means that it is not assumed that all the contextual probability spaces are generated by conditioning. Similar notions were introduced also in [50], where very general results are presented, and in [51].

Consider now two random variables X and Y on $(\Omega, \mathcal{E}, P)$ with distributions $\mu_X = P \circ X^{-1}$ and $\nu_Y = P \circ Y^{-1}$. Assume for simplicity that they are discrete. Conditioning alone is not sufficient to ensure that they are described in two different probability spaces. Indeed, since they are functions on the same probability space, after conditioning they can always be described on a probability space $(\Omega, \mathcal{E}(\Omega)_{C_i}, P_{C_i})$ where, from a (conditional) joint probability distribution, $\eta_{X,Y|C_i} = P_{C_i} \circ (X^{-1}, Y^{-1})$, we can derive the two marginals $\mu_{X|C_i}$ and $\nu_{Y|C_i}$ describing X and Y (after conditioning). However, we may proceed in a different manner. Suppose that X and Y are two random variables on $(\Omega, \mathcal{E}, P)$ with fixed transition probabilities $\alpha(x, y) := P[X = x|Y = y]$ and $\tilde{\alpha}(y, x) := P[Y = y|X = x]$. The random variables X and Y after conditioning are described by the conditional probability distributions $\mu_{X|C_i}$ and $\nu_{Y|C_i}$. It is not difficult to see that, if we use these fixed transition probabilities, in general

$$\alpha(x, y)\nu_{Y|C_i}(y) \neq \tilde{\alpha}(y, x)\mu_{X|C_i}(x). \tag{3.2}$$

In an ordinary measure-theoretic model of probability the product of the transition probability, times the marginal gives the joint probability distribution, which is symmetric under the exchange of its arguments (this is the *Bayes theorem* and it is a consequence of the fact that events are subsets of the *same* sample space). In our case fixing the transition probabilities and using the conditional probabilities for the two random variables, makes impossible to define a

joint probability distribution. More precisely, it does not exist a joint probability distributions which has $\mu_{X|C_i}$ and $\nu_{Y|C_i}$ as marginals, and such that $\alpha(y,x)$ and $\tilde{\alpha}(y,x)$ are the two transition probabilities which can be derived from it. Hence if we fix the transition probabilities in advance, the random variables X and Y after conditioning must be considered to be defined on two different probability spaces in general. Another way to see this is via the *Law of total probability*. From (3.2), one can conclude that

$$\delta(x|Y,C_i) = \mu_{X|C_i}(x) - \sum_y \alpha(x,y)\mu_{Y|C_i}(y) \neq 0$$

which can also be seen as a violation of the Bayes theorem. This has big consequences on the representation with a single mathematical object of the two probability distributions $\mu_{X|C_i}$ and $\mu_{Y|C_i}$. Since we are not working on a single probability space, the procedure explained in section 2.1.5 no longer work. In fact, if we follow this procedure we can associate to $\mu_{Y|C_i}(y)$ the trace class operator $\hat{\rho}_Y = \sum_y \mu_{Y|C_i}(y)|y\rangle\langle y|$, from which we have to conclude that

$$\begin{aligned} \mu_{X|C_i}(x) &= \mathrm{Tr}\left[\hat{\rho}_Y |x\rangle\langle x|\right] \\ &= \sum_y |\langle x|y\rangle|^2 \mu_{Y|C_i}(y). \end{aligned} \tag{3.3}$$

Interpreting $\alpha(x,y) = |\langle x|y\rangle|^2$, we can see that only if $\delta(x|Y,C_i) = 0$ the map between $\mu_{X|C_i}$ and $\mu_{Y|C_i}$ can be described in this way. When $\delta(x|Y,C_i) \neq 0$ we have to proceed in a different way. As explained in [47], the term $\delta(x|Y,C_i)$ play the role of the interference. Under suitable conditions on the probability distributions an algorithm, for the construction of the vector $|\psi\rangle \in \mathcal{H}$ and the representation of the two random variables by means of operators on $\mathcal{H}$, is available [46, 47, 48]. It is called *Quantum-Like Representation Algorithm* (or QLRA for short). This algorithm works for probability distribution associated to two discrete random variables having n different outcomes [49, 52, 47]. However, only in the case $n = 2, 3$ simple conditions on the probability distribution of these random variables has been found [53, 47, 48]. In the general case a simple condition has not been found yet, despite the difficulties seems to be more in computational side rather than in the mathematical one. Observations regarding a possible extension of this algorithm to the case of three random variables with $n = 2$ outcomes, and to the case of two continuous random variables can be found in [Cur7]. On the other hand, the method proposed here based on entropic uncertainty relations, is somehow simpler despite not as powerful as QLRA. It does not tell us how to find $|\psi\rangle$ explicitly but, once that conditions $i) - iii)$ of section 3.2.2 are fulfilled, we know that the random phenomena must be described using a non-commutative probability space. Although this is not a tremendous improvement with respect to QRLA, this method allows to study interesting situations, as we will do the next chapters.

The assumption $\alpha(x,y) = |\langle x|y\rangle|^2$ done in (3.3) to compute $\delta(x|Y,C_i)$ need some discussion. Since $|\langle x|y\rangle|^2$ is symmetric in x and y, this assumption implies

$P(X = x|Y = y) = P(Y = y|X = x)$. By the Bayes theorem on the original probability space we have to conclude that X and Y have to be uniformly distributed. So it does not seem possible that X and Y depend on common event C which generate the context. Hence it seem that we are forced to use $P(X = x|Y = y) \neq P(Y = y|X = x)$. In this case, the Hilbert space representation presented in the previous section may still be used. The price to pay is that one cannot define a single scalar products, $\langle\cdot|\cdot\rangle_1$ and $\langle\cdot|\cdot\rangle_2$, namely one have to use two different Hilbert spaces. In one Hilbert space one use the first scalar product in order to have $|\langle y|x\rangle_1|^2 = P(X = x|Y = y)$ while in the other $|\langle x|y\rangle_2|^2 = P(Y = y|X = x)$. Both Hilbert spaces would lead to the same statistical predictions about the two random variables (except for the transition probabilities of course), which are represented as non-commuting operators on both Hilbert spaces. This is for example what happens in QLRA [49, 52, 47] when $P(X = x|Y = y) \neq P(Y = y|X = x)$.

However, a different and more interesting approach is possible. In appendix it is proved how it is possible to construct a representation on a single Hilbert space when $P(X = x|Y = y) \neq P(Y = y|X = x)$ using POVM. The transition probabilities $|\langle x|y\rangle|^2$ can be defined from $P(X = x|Y = y)$ and $P(Y = y|X = x)$ in an unique way. According to the standard interpretation of POVM in quantum mechanics, they describe imprecise measurement. Hence, according to that, choosing $P(X = x|Y = y) \neq P(Y = y|X = x)$ corresponds in assume that the measurement may not reflect the real outcome of the random variable, due to the measurement apparatus inefficiencies. This is reasonable from the physical point of view. To keep the discussion simple, in the rest of the thesis we assume that $|P(X = x|Y = y) - P(Y = y|X = x)| << 1$, which correspond to neglect the apparatus inefficiencies during the measurements. In this way, we can say that $P(X = x|Y = y) \approx P(Y = y|X = x)$ for all practical proposes, despite from the discussion in appendix, one should understand that this condition is not really a limitation.

Before to conclude, we want to observe the following interesting fact. Given $(\Omega, \mathcal{E}(\Omega)_{C_i}, P_{C_i})_{C_i \in \mathcal{C}}$, we cannot reconstruct the original probability space $(\Omega, \mathcal{E}, P)$. Additional information is required: we need $P(C_i)$. In this sense, if $\mathcal{C}$ is the set of all elementary events for a random variable Z, i.e. all events like $C_i := \{Z = z_i\}$, such random variable Z cannot be described with the contextual probability space obtained after conditioning. In this sense, Z is no longer present in the (probabilisitic) model. Because of this fact, we will also say that the random variable Z was *removed* from the model. Such collection of probability spaces thus represents a tool to describe a random phenomenon, after a random variable (representing some feature of such a phenomenon) is eliminated from the description. It is not clear if such elimination procedure implies always an entropic uncertainty relation.

Chapter 4

The kinematics of a "non-commutative" particle

In the previous section we described a method to obtain a non-commutative behavior of probability starting from an ordinary probability space. In this chapter we apply this method in order to derive the basic commutation relation of quantum mechanics: the one between position and momentum,

$$[\hat{Q}, \hat{P}] = i\hbar\hat{\mathbb{I}}. \tag{4.1}$$

We will present two models where the ideas explained in chapters 2 and 3 are applied. In the first model, once that all the ingredients are defined, we use conditioning to create a collection of probability spaces where position, X, and velocity, V, of a particle can be described. Under certain assumptions we prove an entropic uncertainty relation between X and V. This implies the non-commutativity between X and V, once they are represented as operators on a common Hilbert space. The second model, generalize the first to the continuous time case. A comparison with ordinary quantum mechanics is discussed at the end.

4.1 Model A: Discrete-time 1-D kinematics on a random space

Here we will describe a discrete (and finite) random space and a particle moving on it jumping at random from one point of space to another. Space, position, and velocity of the particle at a given time will be treated in the same way: using random variables. The whole model is 1-dimensional. We will show that, once the space process is removed from the model, the position and velocity of the particle can be jointly described in a non-commutative probability space.

4.1.1 The space process

The process describing space in this model (model A) will be called *space process*. The space process is assumed to be a discrete and finite set of points distributed at random. More precisely, at each instant of time, space is a random distribution of $M \in \mathbb{N}$ points over the real line. Such points evolve in time as discrete-time random walks and, in this sense, space is a *stochastic process*. This time evolution has a twofold interpretation. A first possibility is to think it with respect to the real line: a point of the space process is a random walk and it changes its position along the line as time changes. A second possible way to see this time evolution is to look at its effects on the "ordering among points": the points change their distances with respect to a chosen point (the origin) when this distance is "measured on the points" (example of distance of this kind are discussed in [54]). In some sense this second point of view can be considered as an internal description: it describes space as if the observer has no possibility to see the continuous real line. On the other hand, the first possibility should be considered as an external description[a]. For simplicity, we chose to describe the whole model from the first point of view. Nothing forbids to adopt the second point of view for the description despite, at a first look, it seems more complicated.

Let us recall some basic facts about the random walk [55]. Consider a lattice of points having spacing $l \in \mathbb{R}$, say $\mathbb{Z}_l := \{x \in \mathbb{R} \mid x = ln, n \in \mathbb{Z}\}$. Then take a collection of independent, identically distributed Bernulli random variables $\{Y_i\}_{i=1}^{\infty}$, characterised by the probabilities $P[Y_i = -l] = p$ and $P[Y_i = +l] = q = 1 - p$ for all i. Using this collection, we can define the random walk as the process

$$S_N := \sum_{i=0}^{N} Y_i \tag{4.2}$$

where Y_0 is an arbitrary random variable with distribution $\pi(y_0)$ taking value on $\mathbb{Z}_l$, representing the initial position of the random walk. S_N represents the position of the random walk at time N. Let us now derive the probability distribution of the random walk position at time N, i.e. S_N. Consider the random walk at time N and assume, for the moment, that $Y_0 = 0$. Since at each time-step the random walk can move by $+l$ or by $-l$ its position, if for $n < N$ times the random walk moves by $-l$, its final position d will be

$$d = (N - n)l - nl = (N - 2n)l$$

Using this equation we can see that, if at time N the random walk is found in

[a] An interesting analogy can be made between the two possibilities explained here for the description of the space and the description of a manifold. A manifold can be studied using a coordinate system on it (internal description), or imagine that is embedded in a larger space (external description), similarly to what happens here.

d, the number of times the random walk moves by $-l$ is

$$n = \frac{1}{2}\left(\frac{d}{l} + N\right).$$

Clearly, the number of times it moves by $+l$ will be $N - n$. Note that the chronological order of the movements does not make any difference on the final position. Since for a given $d = (N - 2n)l$ the random walk is just the sum of Bernulli random variables, i.e. a binomial process, we can write

$$\begin{aligned} P[S_N = d] &= \binom{N}{n} p^n (1-p)^{N-n} \\ &= \binom{N}{\frac{1}{2}\left(\frac{d}{l} + N\right)} p^{\frac{1}{2}\left(\frac{d}{l}+N\right)} (1-p)^{N-\frac{1}{2}\left(\frac{d}{l}+N\right)} \\ &= \binom{N}{\frac{d+Nl}{2l}} p^{\frac{d+Nl}{2l}} (1-p)^{\frac{Nl-d}{2l}}. \end{aligned}$$

Nevertheless this formula holds only for $d \in [-lN, lN]$. If $d > lN$ or $d < -Nl$, this probability must be zero because these regions of space cannot be reached by the random walk in N time-steps. Restoring Y_0 (hence we simply translate the final position d by $Y_0 = y_0$), we can write that

$$P[S_N - y_0 = d] = \binom{N}{\frac{d+N}{2l}} p^{\frac{d+N}{2l}} (1-p)^{\frac{N-d}{2l}}, \tag{4.3}$$

Note that (4.3) can be used as a probability only when the value of the random variable Y_0 is given: hence it is a conditional probability with respect to the value of Y_0, i.e. $P[S_N - y_0 = d] = P[S_N^{(Y_0)} = d | Y_0 = y_0]$, where $S_N^{(Y_0)}$ denotes the random walk starting at Y_0. To complete the description of the random walk (4.2), using the Bayes theorem we obtain

$$P[S_N = d] = \sum_{y_0 \in \mathbb{Z}_l} P[S_N^{(Y_0)} = d | Y_0 = y_0] \pi(y_0), \tag{4.4}$$

which is the probability to find the random walk at time N in the position $d \in \mathbb{Z}_l$, given that at the initial time it started from the position Y_0, random variable with distribution $\pi(y_0)$. Without loosing generality, we set $l = 1$ for simplicity. We conclude our review on basic facts about the random walk, formalising the description at measure-theoretic level. As for any stochastic process, also for the random walk, there exists a probability space $(\Omega, \mathcal{E}, P)$. The sample space Ω can be imagined as the set of all possible trajectories of the random walk. It is a countable set (provided that the time of the random walk vary over a finite interval), since the random walk is a discrete process. $\mathcal{E}$ is a σ-algebra on Ω, and can be thought as the power set of Ω, i.e. $\mathcal{E} = \mathcal{P}(\Omega)$[b], while P is the probability measure. The random walk on this probability space is the identity

[b] Given a set A, with the symbol $\mathcal{P}(A)$ we label the power set of A.

random variable evaluated at a given time N, i.e. for $s \in \Omega$ the position of the random walk at time N is $S_N(s) = s(N)$.

Let us now come back to the space process. As stated in the beginning, it consists of a collection of M random walks. At any time step N, the random distribution of points of the random walks is the space process of model A at time N. We may start with this preliminary definition.

Definition 15. *Let $\{S_N^{(i)}\}_{i\in I}$ be a collection of independent random walks, where $|I| = M \in \mathbb{N}$, defined as in* (4.2) *and described with the probability distributions* (4.4)*. We call such a collection the* space process *for the Model A.*

A possible realisation of the space process is given in figure 4.1. We label

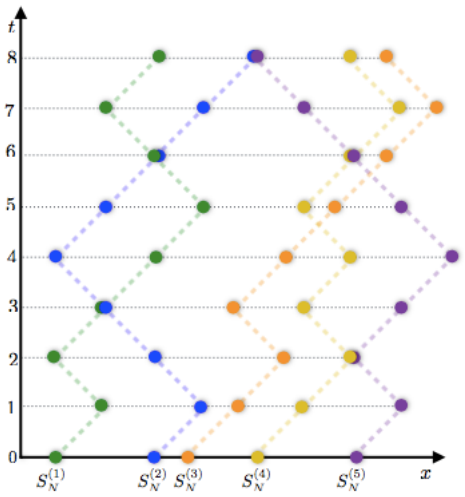

Figure 4.1: A realisation of an $M = 5$ space process $\mathbb{S}^A$ is given. The set of coloured point at any time step represents the random distribution of points of the space process at a fixed time step, $\mathbb{S}^A_N$. The dashed lines liking the points of the same color represent the random walk evolution of each point of the space process. Such evolution changes the random distribution of points of the space process as time changes. Note that at a given time step, points may overlap.

the space process of model A with the symbol $\mathbb{S}^A$, while $\mathbb{S}^A_N$ is the space process at the time-step N (hence a random variable describing the distribution of M points in $\mathbb{R}$). The outcome of the random variable $\mathbb{S}^A_N$, can be thought as a M-tuple, i.e. $\mathbf{S}_N = (s_1(N), \cdots, s_M(N))$ where $s_i(N) \in \mathbb{R}$ is the position of the i-th random walk at time N. Call $\tilde{P}^A$ the probability measure for the space process

$\mathbb{S}^A$. Since the random walks are assumed to be independent, the probability to obtain a specific configuration is given by

$$\tilde{P}^A[\mathbb{S}^A_N = \mathbf{S}_N] = \prod_{i=1}^{M} P[S^{(i)}_N = s_i(N)]. \tag{4.5}$$

For the same reason, it may happen that for some realisation the points overlap. In a similar manner, we can also construct the joint probabilities

$$\begin{aligned}\tilde{P}^A[\mathbb{S}^A_N = \mathbf{S}_N, \mathbb{S}^A_T = \mathbf{S}'_T] = \\ \prod_{i,j=1}^{M} P[S^{(i)}_N = s_i(N), S^{(j)}_T = s'_j(T)].\end{aligned} \tag{4.6}$$

Note that $P[S^{(i)}_N = s_i(N), S^{(j)}_T = s'_j(T)]$ can be constructed using the independence of random walks when $i \neq j$, while for $i = j$ it is just the joint probability distribution of the i-th random walk. Proceeding in this way, we may construct the whole family of finite-dimensional distributions for the space process $\mathbb{S}^A$, which is consistent since the probabilities of the single random walks belongs to consistent families (in the sense of the Kolmogorov extension theorem, see Th. 2.1.5 in [56]). At this point we may replace the preliminary definition of the space process with the following which is more precise.

Definition 16. *Let $\{S^{(i)}_N\}_{i\in I}$ be a collection of $M = |I| \in \mathbb{N}$ independent random walks defined on probability spaces $\{(\Omega_i, \mathcal{E}_i, P_i)\}_{i\in I}$. Let us define*

i) $\Omega_{\mathbb{S}^A} := \Omega_1 \times \cdots \times \Omega_M$;

ii) $\mathcal{E}_{\mathbb{S}^A} = \mathcal{P}(\Omega_{\mathbb{S}^A})$;

iii) $\tilde{P}^A : \mathcal{E}_{\mathbb{S}^A} \to [0, 1]$ *defined from the* $\{P_i\}_{i\in I}$, *as in* (4.5) *or* (4.6) *and generalisation.*

The space process is the stochastic process on $(\Omega_{\mathbb{S}^A}, \mathcal{E}_{\mathbb{S}^A}, \tilde{P}^A)$ defined as the identity function, namely $\mathbb{S}^A(s_1, \cdots, s_M) = (s_1, \cdots, s_M)$.

The set of all the possible configurations of points of the space process at a given time N will be labeled by $\mathcal{S}(N)$.

4.1.2 The particle process

In this model, a particle is considered as a point-like object. At any time-step N, it is completely described by its *position* and its *velocity*, which are assumed to be random variables.

The *position random variable*, labeled by X_N, is interpreted as the actual position of the particle at time N, with respect to a chosen origin (of a reference frame in $\mathbb{S}^A_N$). Let $(\Omega_{\mathbb{S}^A}, \mathcal{E}_{\mathbb{S}^A}, \tilde{P}^A)$ be the probability space for the space process.

On a probability space $(\Omega_I, \mathcal{E}_I, P_I)$ define an integer value discrete-time stochastic process $I_N : \Omega_I \to \{1, \cdots, M\}$, which we call *selection process*. Assume that we place the origin of a reference frame in $\mathbb{S}^A_N$ in the point $S^{(i_O)}_N(\mathbf{S})$. Then we define

$$X_N(\omega) := \pi_{I_N(\omega_I)}(\mathbb{S}^A_N(\mathbf{S})) - S^{(i_O)}_N(\mathbf{S}), \tag{4.7}$$

where π_i is the projector of the i-th component of an M-tuple, and $\omega = (\omega_I, \mathbf{S})$ with $\omega_I \in \Omega_I$ and $\mathbf{S} \in \Omega_{\mathbb{S}^A}$. Thus we have the following definition:

Definition 17. *Consider the probability space $(\Omega_I \times \Omega_{\mathbb{S}^A}, \mathcal{E}_I \otimes \mathcal{E}_{\mathbb{S}^A}, P^A)$ and a measurable space $(\mathbb{Z}, \mathcal{P}(\mathbb{Z}))$. The random variable X_N is the $\mathcal{P}(\mathbb{Z})$-measurable function*

$$X_N : \Omega_I \times \Omega_{\mathbb{S}^A} \to \mathbb{Z}$$

defined as in (4.7)*. X_N represents the position of the particle at time N.*

Note that on the probability space $(\Omega_I \times \Omega_{\mathbb{S}^A}, \mathcal{E}_I \otimes \mathcal{E}_{\mathbb{S}^A}, P^A)$ we can describe also the space process $\mathbb{S}^A$ by simply demanding that $P^A \circ [\mathbb{S}^A]^{-1} = \tilde{P}^A$. From now on in the whole discussion of model A, instead of writing P^A we simply write P if no confusion arises. By construction, X_N is a function of the space process $\mathbb{S}^A$. This implies that X_N and $\mathbb{S}^A_N$ are not independent random variables.

Let us now describe the *velocity random variable*. In order to introduce this process, we need to specify how the particle moves on a physical space described with the space process introduced before. We assume that particle moves by jumps: it jumps from one of the points of the space process at time N to another point of the space process at time $N + 1$. These jumps are described by the transition probabilities

$$P[X_{N+1} = b | X_N = a] = \alpha(b, a), \tag{4.8}$$

where $a, b \in \mathbb{Z}$. Once these transition probabilities are given, we can define the velocity random variable V_N. We set

$$V_N := \frac{X_{N+1} - X_N}{N + 1 - N} = X_{N+1} - X_N. \tag{4.9}$$

This is clearly the discrete-time version of the usual definition of velocity. Note that this physical definition makes sense because, thanks to the transition probabilities (4.8), we can describe V_N from the probabilistic point of view using only information available at time N. More formally, the transition probabilities (4.8) allows to describe V_N on the same probability space of X_N, i.e. $(\Omega_I \times \Omega_{\mathbb{S}^A}, \mathcal{E}_I \otimes \mathcal{E}_{\mathbb{S}^A}, P^A)$.

Definition 18. *Consider the probability space $(\Omega_I \times \Omega_{\mathbb{S}^A}, \mathcal{E}_I \otimes \mathcal{E}_{\mathbb{S}^A}, P^A)$ and the measurable space $(\mathbb{Z}, \mathcal{P}(\mathbb{Z}))$. The velocity random variable V_N is the $\mathcal{P}(\mathbb{Z})$-measurable function*

$$V_N : \Omega_I \times \Omega_{\mathbb{S}^A} \to \mathbb{Z}$$

defined in (4.9)*. V_N represents the velocity of the particle at time N.*

Also the velocity random variable is a function of the space process: thus they are not independent random variables. Let us now derive the relation between the probabilities $P[V_N = c]$ and $P[X_N = a]$, in a way that is consistent with the transition probabilities (4.8). It can be done following this intuitive idea. Suppose that at time N we know that the particle is in the position $X_N = a$. Then the event $A := \{X_N = a\}$ is true, i.e. $P(A) = P[X_N = a] = 1$, which means that $P[X_N = a'|A] = \delta_{a,a'}$. Under the same conditions, one should also write that $V_N = X_{N+1} - a$, and this suggests that the probability to observe $V_N = c$ is equal to the probability to observe $X_{N+1} = a + c$, when A happens. Thus, using (4.8) we can write that

$$P[V_N = c|A] = P[X_{N+1} = a + c|A] = \alpha(a + c, a).$$

More formally, we have the following.

Proposition 1. *Let X_N and V_N be the position and the velocity random variables. If $P[X_{N+1} = b|X_N = a] = \alpha(b, a)$, then $P[V_N = c|A] = \alpha(a + c, a)$ where $A = \{X_N = a\}$.*

Proof. See appendix. □

At this point, we may obtain $P[V_N = c]$ simply using the Law of total probability, namely

$$P[V_N = c] = \sum_a \alpha(a + c, a)P[X_N = a] \tag{4.10}$$

which is consistent with the transition probabilities given in the beginning. Having defined both the position and velocity random variables, we may give a precise definition of what we call particle in model A .

Definition 19. *A particle is a point like-object whose features at time N are completely specified by the position and velocity random variables. More formally, we can say that a particle corresponds to the random vector $\mathscr{P}_N := (X_N, V_N)$. We will refer to $\mathscr{P}_N$ with the name* particle process*, when considered as a function of time.*

An example of particle process is drawn in figure 4.2.

4.1.3 Remove the space from the model

In this section, we will explain what we mean with the expression "remove the space from the model". In section 3.2.3 we observed that, given three random variables X, Y and Z on the same probability space $(\Omega, \mathcal{E}, P)$, one can eliminate one of them, say Z, simply by conditioning with respect to the outcomes of this random variable, i.e. conditioning on the events $\{Z = z\}$. In the collection of probability spaces obtained after conditioning, the description of the random variable Z is not anymore possible unless one adds the probabilities $P[Z = z]$. Such information cannot be obtained from the collection of probability spaces

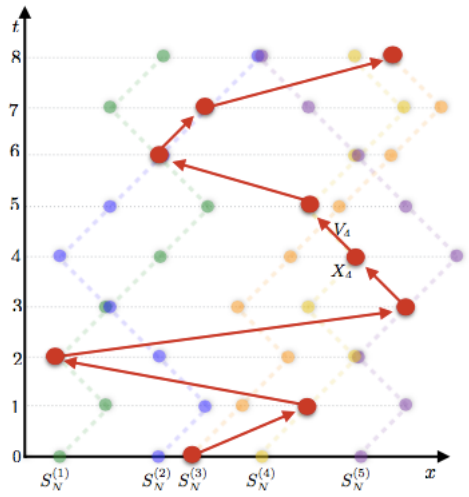

Figure 4.2: A possible realization of the particle process is drawn in red, over the same realization of the space process considered in figure 4.1. The position of the particle at a given time-step is given by the red point while its velocity at the same time-step is represented by the outgoing arrow.

one has after conditioning. The statistical description of the remaining random variables can be done without influence the random variable Z: in this sense Z is not present anymore in the probabilistic model used to describe X and Y. However, when we deal with a stochastic process the elimination of a random variable representing it at given time, do not guarantee that we can manipulate all the remaining random variables without influence the stochastic process (which means that we change the probability distributions of the remaining random variables without modify the probability distribution of the removed process). In this case we need to add additional conditions in order to be sure that the stochastic process is not present anymore in the remaining probabilistic model. Once that we apply a procedure that is capable to do so, we say that the stochastic process is *removed from the model*.

Model A exhibits features that are interesting from the point of view of quantum mechanics when we remove the space process from the model. Before describing how to implement it mathematically, let us first explain the physical principles that motivate this removal. Model A describes a particle that jumps at random over a random distribution of points. Such a random distribution of points is assumed to be the physical space in which the particle moves: the physical space is not anymore a passive background against which physical processes take place. Preparing the particle in a given state means to perform

an experimental procedure after which the statistical properties of the particle's observables are known. In other words, the state preparation is an experimental procedure such that right after it terminates, all random variables associated to the particle's observables have a given probability distribution. Hence, saying that a particle at time N is in a given state, means that at time N all probability distributions of the observables of the particle are fixed. However, assuming that the physical space is random has big consequences. Any experimental procedure happens in such a physical space. If the experimental procedure for the state preparation ends at time N, the probability distributions of the observables are always conditioned to the configuration of space at that time. This is because one prepares the state of the particle at time N, in the configuration the space process assumes at that time. Hence the probability distributions that describe the particle must be always conditioned to some space configuration. If it is not so, to prepare the particle in a given state we need to have control not only on it but also on the whole space. This means that the probability distribution describing the space process at a given time does not depend on the probabilities describing the particle after a preparation procedure. In other words, the change of the probability distributions describing the particle (i.e. the change of the state) does not have to modify the probability distributions describing the space process. When this happens we say that the space process is removed from model A. In order to implement that, we have to require the following:

i) The particle at time N can be described only by using probabilities that are conditioned with respect to some space configuration at that time. This means that to describe the position and velocity random variables we have to use only

$$P_{\mathbf{S}_N}[X_N = a] := P[X_N = a|\mathbb{S}_N^A = \mathbf{S}_N],$$
$$P_{\mathbf{S}_N}[V_N = c] := P[V_N = c|\mathbb{S}_N^A = \mathbf{S}_N],$$

where $\mathbf{S}_N$ is the configuration of the space process at time N.

ii) The transition probabilities of any point of space (i.e. $p_i = P[S_{N+1}^{(i)} = a + 1|S_N^{(i)} = a]$ for all $i \in I$) cannot be changed by the preparation procedure of the particle. This means that changing the conditional probabilities of the particles, the transition probabilities of the single point of space remains fixed.

These two conditions implement the idea that the space process cannot be influenced by the preparation procedure of the particle. Note that the requirement i) is needed in order to avoid that the probability of the space process at time N is changed by the preparation procedure of the particle, while the requirement ii) avoids that such preparation procedure alters the space process probabilities at times $N' \neq N$ (i.e. in the past or in the future). Since we are dealing with non-relativistic systems this last requirement is reasonable from the physical point of view.

Let us now describe the effects of the removal of the space process in model A from the mathematical point of view. We will focus first on the consequences of the requirement i). In order to do so we need to study better the effect of conditioning on a probability space. According to the removal procedure explained above, we can describe the particle using only probabilities that are conditioned to the event $\{\mathbf{S}_N\} := \{\mathbf{S} \in \Omega_{\mathbb{S}^A} | \mathbb{S}^A_N(\mathbf{S}) = \mathbf{S}_N\}$. At the level of the events, this means that for the random variable X_N and V_N we consider only events of this kind: $\{X_N \in A\} \cap \{\mathbb{S}^A_N = \mathbf{S}_N\}$ and $\{V_N \in B\} \cap \{\mathbb{S}^A_N = \mathbf{S}_N\}$. For the position random variable, this means that the conditioning procedure effectively changes the sample space and the σ-algebra of its starting probability space as

$$(\Omega_I \times \Omega_{\mathbb{S}^A}, \mathcal{E}_I \otimes \mathcal{E}_{\mathbb{S}^A}) \to (\Omega_I \times \{\mathbf{S}_N\}, \mathcal{P}(\Omega_I \times \{\mathbf{S}_N\})).$$

Let us call $\Omega_{X_N} := \Omega_I \times \{\mathbf{S}_N\}$ and $\mathcal{E}_{X_N} := \mathcal{P}(\Omega_I \times \{\mathbf{S}_N\})$. It is a known fact from probability theory that the measurable space $(\Omega_{X_N}, \mathcal{E}_{X_N})$ equipped with conditional probability $P_{\mathbf{S}_N}[X_N = \cdot]$ defines a probability space. On this probability space $(\Omega_{X_N}, \mathcal{E}_{X_N}, P_{\mathbf{S}_N})$ the random variable X_N can be described after conditioning on the event $\{\mathbf{S}_N\}$. Everythig we said till now, clearly also holds for the velocity random variable: after conditioning it can be described in a probability space $(\Omega_{V_N}, \mathcal{E}_{V_N}, P_{\mathbf{S}_N})$ defined in a similar manner.

Relevant for our goal is the study of the joint probabilities for X_N and V_N, and its link with the transition probabilities (4.8) after conditioning. By definition X_N and V_N are two random variables defined on the same probability space $(\Omega_I \times \Omega_{\mathbb{S}^A}, \mathcal{E}_I \otimes \mathcal{E}_{\mathbb{S}^A}, P)$. This means that we can always find a joint probability distribution $P[X_N = a, V_N = c]$, which can be used to derive the transition probabilities (4.8) using the usual Bayes formula. Since the space process can be described on the same probability space of X_N and V_N, also the joint probability distribution $P[X_N = a, V_N = c, \mathbb{S}^A_N = \mathbf{S}_N]$ exists. Applying the Bayes formula, we can derive the *conditional* joint probability for X_N and V_N, namely

$$P_{\mathbf{S}_N}[X_N = a, V_N = c] := \frac{P[X_N = a, V_N = c, \mathbb{S}^A_N = \mathbf{S}_N]}{P[\mathbb{S}^A_N = \mathbf{S}_N]},$$

from which one can derive *conditional* transition probabilities

$$\alpha_{\mathbf{S}_N}(c, a) := P_{\mathbf{S}_N}[V_N = c | X_N = a] = \frac{P_{\mathbf{S}_N}[X_N = a, V_N = c]}{P_{\mathbf{S}_N}[X_N = a]}.$$

Note that $\alpha_{\mathbf{S}_N}(c, a) \neq \alpha(c, a)$. From the point of view of the probability spaces, after conditioning we can always describe the two random variables using a single probability space. Such probability space is simply $(\Omega_{X_N} \times \Omega_{V_N}, \mathcal{E}_{X_N} \otimes \mathcal{E}_{V_N}, P_{\mathbf{S}_N})$. On it, we can define a joint probability distribution $P_{\mathbf{S}_N}[X_N = \cdot, V_N = \cdot]$ such that $P_{\mathbf{S}_N}[X_N = \cdot]$ and $P_{\mathbf{S}_N}[V_N = \cdot]$ are the two marginals and $\alpha_{\mathbf{S}_N}(c, a)$ are the transition probabilities between V_N and X_N. Since the joint probability distribution are symmetric under the exchange of the arguments, clearly

$$\alpha_{\mathbf{S}_N}(c, a) P_{\mathbf{S}_N}[X_N = a] = \alpha_{\mathbf{S}_N}(a, c) P_{\mathbf{S}_N}[V_N = c],$$

where $\alpha_{\mathbf{S}_N}(a,c) = P_{\mathbf{S}_N}[X_N = a|V_N = c]$. According to [47], this is a signature that we are working on a single measure-theoretic probability space. A more interesting case happens when we use the the *unconditional* transition probabilities $\alpha(a+c,a)$ and $P[X_N = a|V_N = c]$. In this case we have that

$$\alpha(a+c,a)P_{\mathbf{S}_N}[X_N = a] \neq P[X_N = a|V_N = c]P_{\mathbf{S}_N}[V_N = c] \tag{4.11}$$

in general, which means that we cannot describe X_N and V_N using a single measure-theoretic probability space, *if we choose to use the unconditional transition probabilities after conditioning with respect to the space process at time* N. However, this does not mean that we cannot describe X_N and V_N after conditioning using the transition probabilities $\alpha(a+c,a)$ (we will come back on the physical reason for the use of $\alpha(a+c,a)$ instead of $\alpha_{\mathbf{S}_N}(c,a)$ later). We can do it using two different probability spaces: one for X_N and one for V_N. We have already seen that, after conditioning, we obtain a probability spaces for each random variables, i.e. $(\Omega_{X_N}, \mathcal{E}_{X_N}, P_{\mathbf{S}_N})$ for X_N and $(\Omega_{V_N}, \mathcal{E}_{V_N}, P_{\mathbf{S}_N})$ for V_N. However we cannot construct a joint probability space where $P_{\mathbf{S}_N}[X_N = \cdot]$ and $P_{\mathbf{S}_N}[V_N = \cdot]$ are the two marginals of some joint probability distribution and $\alpha(a+c,a)$ are the transition probabilities that we obtain from the same joint probability distribution. This is exactly the content of (4.11): the joint probability we are looking for would not be symmetric in the exchange of the arguments. This is something that it is not possible in an ordinary measure space since the intersection of events in a sigma algebra is a symmetric operation (i.e. commutative). As a consequence we may conclude that the Bayes theorem cannot be used to relate the two marginals. However a relation between $P_{\mathbf{S}_N}[X_N = \cdot]$ and $P_{\mathbf{S}_N}[V_N = \cdot]$ can still be found [49].

Theorem 11. *Let* $\{P_{\mathbf{S}_N}[X_N = a]\}_{a\in\Omega_{X_N}}$ *be the probabilities describing the position of the particle at time* N *under the condition that the space process at time* N *is* $\mathbf{S}_N$*. If* $P[X_{N+1} = b|X_N = a] = \alpha(b,a)$*, then*

$$P_{\mathbf{S}_N}[V_N = c] = \sum_a \alpha(a+c,a)P_{\mathbf{S}_N}[X_N = a] + \delta(c|X_N, \mathbf{S}_N) \tag{4.12}$$

where

$$\begin{aligned}\delta(c|X_N,\mathbf{S}_N) = \frac{1}{P[\mathbb{S}_N^A = \mathbf{S}_N]} \sum_{\substack{\mathbf{S}'_N \\ \mathbf{S}'_N \neq \mathbf{S}_N}} \Bigg[\sum_a \alpha(a+c,a)\cdot \\ \cdot P[X_N = a, \mathbb{S}_N^A = \mathbf{S}'_N] - P[V_N = c, \mathbb{S}_N^A = \mathbf{S}'_N] \Bigg]\end{aligned} \tag{4.13}$$

which is in general different from zero.

Proof. See appendix □

We can see that, after the conditioning on the space process, the Bayes formula cannot be used anymore to compute $P_{\mathbf{S}_N}[V_N = c]$ from the probabilities

of the position random variable if we want to use the transition probabilities $\alpha(a+c,a)$. We need to add a correction term which contains statistical information about the space process. Note that this correction term has the property

$$\sum_c \delta(c|X_N,\mathbf{S}_N)=0, \tag{4.14}$$

which is necessary in order to preserve the normalisation of probabilities, i.e. $\sum_c P_{\mathbf{S}_N}[V_N=c]=1$. We also note that in general $\delta(c|X_N,\mathbf{S}_N)\in[-1,1]$ and in particular it can be negative. Summarising, given the transition probabilities $\alpha(a+c,a)$ we cannot describe X_N and V_N on a single probability space after conditioning on the space configuration at time N. However, the description X_N and V_N in a single probability space after conditioning can be always done: the price to pay is that we have to change the transition probabilities from $\alpha(a+c,a)$ to $\alpha_{\mathbf{S}_N}(c,a)$.

At this point a legitimate question arises: can we motivate physically the choice to use $\alpha(a+c,a)$ instead $\alpha_{\mathbf{S}_N}(c,a)$? Yes, if we take into account the fact that we want to remove space from the model. Indeed, in order to measure with an experimental procedure $\alpha_{\mathbf{S}_N}(c,a)$, one would have control over space since one has to be able to prepare the space process always in the configuration $\mathbf{S}_N$, in order to measure $\alpha_{\mathbf{S}_N}(c,a)$. Since the removal of space is done exactly to avoid such things, the use of $\alpha(a+c,a)$ is more reasonable from the physical point of view. We want to conclude our analysis on the consequence of the requirement i) with a comment on the particle process. Since it is a random vector parametrized by time, one may be tempted to consider $\mathscr{P}_N$ as a stochastic process. This is certainly possible considering also the space process, namely before conditioning on $\mathbf{S}_N$. Nevertheless, after conditioning and using the transition probabilities $\alpha(a+c,a)$, we just have a collection of probability spaces and it is not trivial to assume that each of these spaces can be seen as, part of a bigger probability space describing the particle only (i.e. with no space process involved in the construction of such probability space) as the Kolmogorov extension theorem [56] would imply. For this reason, considering the particle process as a stochastic process, in this context, should be done with care.

Till now we explored the consequences of the requirement i) for the removal of space. Conditioning with respect to the space configuration $\mathbf{S}_N$, we effectively eliminate the possibility to change the $P[\mathbb{S}^A_N=\mathbf{S}_N]$ by varying the (conditional) probability distribution describing the particle. The requirement ii) is added in order to avoid that by varying the probabilities of the particle we can modify the probabilities $P[\mathbb{S}^A_{N'}=\mathbf{S}_{N'}]$ when $N'\neq N$. The consequences of ii) which are relevant for our analysis will be analyzed in the next section.

4.1.4 The entropic uncertainty relation for X_N and V_N

In this section we will analyze the basic consequence of the requirement ii) for the removal of the space process in model A. From now on, we exclude that the probabilities describing the space process have delta-like distributions. This

implies that the space process of model A is not a deterministic process. We note that the whole removal procedure, which makes model A interesting to study, is meaningless in this case.

Theorem 12. *Let X_N and V_N be the position and velocity random variables of model A. Fixing the transition probabilities $p_i = P[S^{(i)}_{N+1} = a+1 | S^{(i)}_N = a]$ of the points of the space process for all $i \in I$, then*

$$H_{\mathbf{S}_N}(X_N) + H_{\mathbf{S}_N}(V_N) \geqslant D, \tag{4.15}$$

where D is a positive constant which does not depend on $\{P_{\mathbf{S}_N}[X_N = a]\}_{a \in X_N(\Omega_X)}$ and $\{P_{\mathbf{S}_N}[V_N = c]\}_{c \in V_N(\Omega_V)}$.

Proof. See appendix □

We can better grasp the physical meaning of the inequality between entropies proved above, considering a particular case of space process. Assume that all the random walks of the space process are *identically distributed*. This means that if $p_i = P[S^{(i)}_{N+1} = a+1 | S^{(i)}_N = a]$ are the transition probabilities and $\pi^{(i)}$ are the probability distributions of the initial position of all random walks, we have

$$p_1 = p_2 = \cdots = p_M$$
$$\pi^{(1)} = \pi^{(2)} = \cdots = \pi^{(M)}.$$

This implies that $P[S^{(i)}_N = a] = P[S^{(j)}_N = a]$ for any $a \in \mathbb{Z}$, for any $N \geqslant 0$, and any $i, j \in \{1, \cdots, M\}$. From the proof of theorem 12 given in appendix, one has that the value of the constant D given is

$$D = \min\{D_1, D_2\}$$

where

$$D_1 = \min_a \left[\sum_c \min_{ij} \left(- P_{\mathbf{S}_N}[S^{(i)}_{N+1} = a + c | S^{(j)}_N = a] \log P_{\mathbf{S}_N}[S^{(i)}_{N+1} = a + c | S^{(j)}_N = a] \right) \right]$$

$$D_2 = \min_c \left[\sum_a \min_{i,j,d} \left(- P_{\mathbf{S}_N}[S^{(i)}_{N+1} = a + c | S^{(j)}_N = d - c] \log P_{\mathbf{S}_N}[S^{(i)}_{N+1} = a + c | S^{(j)}_N = d - c] \right) \right].$$

Consider the constant D_1. The min's can be eliminated since all the probabilities are equals. The same hold for D_2 Hence

$$D_1 = D_2 = -p \log p - (1-p) \log(1-p)$$

Thus we can conclude that $D = -p \log p - (1-p) \log(1-p)$. This is the so called binary entropy, which vanishes only if $p = 0, 1$ namely if that space is a deterministic process, a case which is excluded from the beginning. The physical meaning of the inequality $H_{\mathbf{S}_N}(X_N) + H_{\mathbf{S}_N}(V_N) \geqslant D$, in this particular case, is now clear: the uncertainty that we have on X_N or V_N must be at least equal to the uncertainty we have on a single point in the future configurations of the space process (given that at time N the configuration is $\mathbf{S}_N$).

4.1.5 Construction of the Hilbert space structure for model A

Theorem 11 implies that after the removal of the space process, X_N and V_N are described using two distinct probability spaces if we want to use the unconditional transition probabilities $\alpha(a+c,a)$. Theorem 12 tell us that under the same assumptions, the position and the velocity of the particle in model A, fulfil an entropic uncertainty relation. At this point, we may proceed algebraically and define the smallest C^*-algebra which is capable to describe both X_N and V_N after conditioning, and the entropic uncertainty relation (12) tells us that this algebra is non-commutative [57]. Then, we can represent these elements of the algebra as two non-commuting operators over a Hilbert space via the GNS theorem. Despite this is a legitimate way to proceed, in this section, using the results collected in chapters 2 and 3, we will use a more constructive approach. In particular, we show how to construct the operators associated to these random variables and how to define a suitable Hilbert space on which they are defined.

Consider the position random variable X_N. After conditioning on a particular configuration of the space process $\mathbf{S}_N$, X_N can be seen as the as the following map between probability spaces

$$(\Omega_I \times \Omega_{\mathbb{S}^A}, \mathcal{E}_I \otimes \mathcal{E}_{\mathbb{S}^A}, P)|_{\mathbf{S}_N} \xrightarrow{X_N} (\Omega_{X_N}, \mathcal{E}_{X_N}, \mu_{X_N})$$

where $\Omega_{X_N} = X_N(\Omega_I \times \{\mathbf{S}_N\})$, $\mathcal{E}_{X_N} = \mathcal{P}(\Omega_{X_N})$ and $\mu_{X_N} := P_{\mathbf{S}_N} \circ X_N^{-1}$. As we have seen in chapter 2, random variables over a probability space form a commutative von-Neumann algebra which is isomorphic to an algebra of multiplicative operators over an Hilbert space. In this particular case, the random variables over $(\Omega_{X_N}, \mathcal{E}_{X_N}, \mu_{X_N})$ (on which X_N is represented by the identity map) form the abelian von-Neumann algebra $\mathcal{V}_c(L_2(\Omega_{X_N}, \mu_{X_N}))$. Seen as element of this algebra, the random variables over $(\Omega_{X_N}, \mathcal{E}_{X_N}, \mu_{X_N})$ are multiplicative operators over $L_2(\Omega_{X_N}, \mu_{X_N})$.

Similar considerations hold for the velocity of the particle. The main difference is the definition of Ω_{V_N}, i.e. the set of all the elementary outcomes. It is not difficult to understand that, if we fix the space process only, Ω_{V_N} seems to contain more outcomes of those one should expect. The number of outcomes of the position process is M^2, i.e. card $\Omega_{X_N} = M^2$. This because the origin and the point of $\mathbb{S}_N^A$ selected by the selection process I_N, can take M different values. For the velocity process similar considerations lead to card$\Omega_{V_N} = M^4$. However, we have to take into account that we cannot detect the movement of the origin: $S_{N+1}^{(i_O)} - S_N^{(i_O)}$ must be set equal to 0, and all the situations where this does not hold must be identified with it [c]. After that the velocity can takes only M^2 different values (the M's of $S_N^{(i_{N+1})}$ times the M's of $S_N^{(i_N)}$). Thus doing that we have card $\Omega_{V_N} = M^2$. After this observation, we may see the velocity

[c] More precisely, we can define an equivalence relation between X_{N+1} and X_N: $X_{N+1} \sim X_N$ if $X_{N+1} - X_N = S_{N+1}^{(i_o)} - S_N^{(i_o)}$. In this way we restrict our attention to the intrinsic motion of the particle.

random variable, after conditioning to $\mathbf{S}_N$, as the map

$$(\Omega_I \times \Omega_{\mathbb{S}^A}, \mathcal{E}_I \otimes \mathcal{E}_{\mathbb{S}^A}, P)|_{\mathbf{S}_N} \xrightarrow{V_N} (\Omega_{V_N}, \mathcal{E}_{V_N}, \mu_{V_N})$$

where $\Omega_{V_N} = V_N(\Omega_I \times \{\mathbf{S}_N\})$, $\mathcal{E}_{V_N} = \mathcal{P}(\Omega_{V_N})$ and $\mu_{V_N} := P_{\mathbf{S}_N} \circ V_N^{-1}$. Also in this case, the random variables over $(\Omega_{V_N}, \mathcal{E}_{V_N}, \mu_{V_N})$, are elements of a commutative von-Neumann algebra $\mathcal{V}_c(L_2(\Omega_{V_N}, \mu_{V_N}))$ (i.e. multiplicative operators on $L_2(\Omega_{V_N}, \mu_{V_N})$).

Thus both X_N and V_N can be represented by multiplicative operators on suitable Hilbert spaces. Note that the two Hilbert spaces are different and depend on the probability measure. In order to construct a common Hilbert space on which both operators are defined, we should invoke the spectral representation theorem, as we saw in chapter 2. We recall that the spectral decomposition theorem tells that, given an operator $\hat{T}$, there exist a surjective isometry $\hat{U}_i : \mathcal{H}_i \to L_2(\sigma(\hat{T}), \mu_i)$ such that $\hat{U}_i^* \hat{T}|_{\mathcal{H}_i} \hat{U}_i$ is a multiplicative operator on $L_2(\sigma(\hat{T}), \mu_i)$, i.e. an element of $\mathcal{V}_c(L_2(\sigma(\hat{T}), \mu_i))$. Consider the position random variable X_N. We know that it is a multiplicative operator on $L_2(\Omega_{X_N}, \mu_{X_N})$, and let us now choose to parametrise the probability measure of the position random variable with the outcome of X_N. This can be achieved in the following way. Take $a \in \Omega_{X_N}$ and consider the probability measure $P^{(a)}_{\mathbf{S}_N}$, which is defined such that $\mu_{X_N}(c) = P^{(a)}_{\mathbf{S}_N} \circ X_N^{-1}(c) = \delta_{a,c}$. We can parametrise the probability measure of X_N with its outcomes defining $\mu_{X_N|a} := P^{(a)}_{\mathbf{S}_N} \circ X_N^{-1}$. Doing that we obtain a collection of Hilbert spaces $\{L_2(\Omega_{X_N}, \mu_{X_N|a})\}_{a \in \Omega_{X_N}}$. Now, the random variable X_N can be represented with an operator $\hat{X}_N$, having spectrum $\sigma(\hat{X}_N) = \Omega_{X_N}$. The spectral decomposition theorem tells that there exists a collection of Hilbert spaces $\{\mathcal{H}_a\}_{a \in \Omega_{X_N}}$ and surjective isometries $\hat{U}_a : \mathcal{H}_a \to L_2(\Omega_{X_N}, \mu_{X_N|a})$, which allows to define the Hilbert space

$$\mathcal{H}(X_N) := \bigoplus_{a \in \Omega_{X_N}} \mathcal{H}_a$$

on which $\hat{X}_N$ can be seen as a multiplicative operator. The spectral representation theorem tells that if $\{|x_N\rangle\}$ is a basis of $\mathcal{H}(X_N)$ such that $|x_N\rangle \in \mathcal{H}_{x_N}$ for any $x_N \in \Omega_{X_N}$, then X_N can be represented by the operator

$$\hat{X}_N = \sum_{x_N \in \Omega_{X_N}} x_N |x_N\rangle\langle x_N|.$$

With similar considerations, for V_N we obtain

$$\mathcal{H}(V_N) := \bigoplus_{c \in \Omega_{V_N}} \mathcal{H}_c$$

on which the operator $\hat{V}_N$ representing the velocity random variable, is diagonal

$$\hat{V}_N = \sum_{v_N \in \Omega_{V_N}} v_N |v_N\rangle\langle v_N|. \tag{4.16}$$

At this point, we impose the condition

$$\mathcal{H}(X_N) = \mathcal{H}(V_N)$$

i.e. that the two Hilbert spaces are *unitary equivalent*. This is possible since the dimension of both Hilbert spaces is M^2: both Hilbert spaces are constructed from the spectrum of $\hat{X}_N$ or $\hat{V}_N$, and both have the same number of elements. Since Hilbert spaces of equal dimension are always isomorphic, there exists a unitary map between them (see [58], Th. 5.3), i.e. there exists

$$\hat{U} : \mathcal{H}(V_N) \to \mathcal{H}(X_N),$$

such that $\hat{U}\hat{U}^* = \hat{\mathbb{I}}_{\mathcal{H}(X_N)}$ and $\hat{U}^*\hat{U} = \hat{\mathbb{I}}_{\mathcal{H}(V_N)}$. This unitary mapping allows to have, on the same Hilbert space, the operators representing the position and the velocity random variables. More precisely, take the velocity operator $\hat{V}_N$ on $\mathcal{H}(V_N)$ defined in (4.16), then the unitary map mentioned above allows us to write

$$\begin{aligned}\hat{V}_N|_{\mathcal{H}(X_N)} &= \hat{U}\Big(\sum_{v\in\Omega_{V_N}} v_N|v_N\rangle\langle v_N|\Big)\hat{U}^* \\ &= \sum_{v_N\in\Omega_{V_N}} v_N\hat{U}|v_N\rangle\langle v_N|\hat{U}^*,\end{aligned}$$

which represents the velocity random variable on $\mathcal{H}(X_N)$, the Hilbert space constructed from the spectrum of the position operator (on which $\hat{X}_N$ is diagonal). The entropic uncertainty relation, ensures that X_N and V_N as operators on the same Hilbert space, do not commute. In fact, it implies [41]

$$\max_{x_N,v_N} |\langle x_N|v_N\rangle| \leqslant e^{-\frac{D}{2}} < 1, \tag{4.17}$$

as already observed in chapter 3. Thus the two operators cannot be diagonalised on the same basis, i.e. they do not commute.

We can also represent on $\mathcal{H}(X_N)$ the velocity random variable directly. Indeed on this Hilbert space, we may always consider a generic basis $\{|w_N\rangle\}_{w_N\in\Omega_{V_N}}$ and impose that $\hat{V}_N$ is diagonal on this basis, i.e.

$$\hat{V}_N = \sum_{w_N\in\Omega_{V_N}} w_N|w_N\rangle\langle w_N|.$$

We can always parametrise the probability measure of μ_{V_N} using the outcome of X_N simply defining $\mu_{V_N|a} := P^{(a)}_{\mathbf{S}_N} \circ V_N^{-1}$. Then we obtain the collection of Hilbert spaces $\{L_2(\Omega_{V_N}, \mu_{V_N|a})\}_{a\in\Omega_{X_N}}$. For a given $a \in \Omega_{X_N}$, the entropic uncertainty relation (4.15) forbids to have delta-like probability measure for both operators. Indeed, considering $\hat{X}_N$, we have

$$\mu_{X_N|a} := \langle\psi|\hat{P}_{\mathcal{H}_{x_N}}\psi\rangle = \langle\psi|x_N\rangle\langle x_N|\psi\rangle = \delta_{x_N,a}$$

which is possible only if $|\psi\rangle = |a\rangle$. On the other hand for $\hat{V}_N$, if $\hat{P}_{\mathcal{H}_{w_N}}$ is the projector on the subspace of $\mathcal{H}(X_N)$ associated to the eigenvalue w_N (i.e. the outcome w_N of the random variable V_N), we have

$$\mu_{V_N|a} := \langle\psi|\hat{P}_{\mathcal{H}_{w_N}}\psi\rangle = \langle a|w_N\rangle\langle w_N|a\rangle = |\langle a|w_N\rangle|^2.$$

Since the entropic uncertainty relation hold, (4.17) forbids that $|w_N\rangle$ and $|x\rangle$ to be vectors belonging to the same orthonormal basis. Again, we conclude that X_N and V_N can be represented on a common Hilbert space, $\mathcal{H}(X_N)$, using two operators $\hat{X}_N$ and $\hat{V}_N$ which cannot be diagonalised on the same basis. Note that this $\hat{V}_N$ coincides exactly with $\hat{V}_N|_{\mathcal{H}(X_N)}$ thanks to the existence of the unitary map $\hat{U} : \mathcal{H}(V_N) \to \mathcal{H}(X_N)$. Clearly, also X_N can be represented on $\mathcal{H}(V_N)$ directly, following a similar procedure. In this sense the whole description is consistent: starting the construction of the Hilbert space from X_N or V_N does not change anything, as it should be.

Finally, we conclude by observing that the probabilistic content is now encoded in the vectors $|\psi\rangle$ of the constructed Hilbert spaces. In fact, given $|\psi\rangle \in \mathcal{H}(X_N)$ (or $\mathcal{H}(V_N)$), we can write that

$$\mathbb{E}_{\mathbf{S}_N}[V_N] = \mathrm{Tr}\left[|\psi\rangle\langle\psi|\hat{V}_N\right] = \sum_v v\,\mathrm{Tr}\left[|\psi\rangle\langle\psi|v\rangle\langle v|\right]$$

where $\mathrm{Tr}\left[|\psi\rangle\langle\psi|v\rangle\langle v|\right] = P_{\mathbf{S}_N}[V_N = v]$ is the probability distribution for V_N after conditioning. The probability distribution for V_N can be related with the distribution of X_N as follows

$$\begin{aligned}
P_{\mathbf{S}_N}[V_N = v] &= \sum_x \langle x|\psi\rangle\langle\psi|v\rangle\langle v|x\rangle \\
&= \sum_x \sum_{x'} \langle x|\psi\rangle\langle\psi|x'\rangle\langle x'|v\rangle\langle v|x\rangle \\
&= \sum_x \alpha(x+v,x) P_{\mathbf{S}_N}[X_N = x] \\
&\quad + \sum_{x\neq x'} \langle x|\psi\rangle\langle\psi|x'\rangle\langle x'|v\rangle\langle v|x\rangle
\end{aligned}$$

where we used $P_{\mathbf{S}_N}[X_N = x] = |\langle x|\psi\rangle|^2$ and $\alpha(x+v,x) = |\langle x|v\rangle|^2$. Note that the second term in the last sum (the interference term) corresponds to the correction term $\delta(V_N|X_N, \mathbf{S}_N)$ in theorem 11. However, the method used here does not provide a way to determine uniquely the objects on $\mathcal{H}(X_N)$ (or $\mathcal{H}(V_N)$) associated to a given set of probability distributions, as already noted. In fact, the method proposed does not provide an explicit way to compute the phase of $\langle x_N|v_N\rangle$ starting from the interference term. However QRLA may indicate a possible way to do that [49, 47, 49, 52, 48].

4.2 Model B: Continuous-time 1-D kinematics on a random space

We have shown in section 4.1 that model A exhibits very interesting features from the point of view of quantum mechanics. Nevertheless, it also has some limitations: time is a discrete parameter and the spectrum of the position operator $\hat{X}_N$ is discrete. They do not allow for a direct comparison with ordinary quantum mechanics. To attempt to derive the commutation of quantum mechanics from a similar model, one may try to generalize the previous construction to continuous-time random variables. Here we will show how to do it.

4.2.1 The space process

In order to generalize model A to the continuous time case, we may start by generalizing the space process. Instead of considering the space process as a collection of random walks, we may consider their "continuous limits", i.e. Wiener processes. Let us recap the basic features of the Wiener process [59], as done for the random walk. A *Wiener process W_t starting at y* is a Gaussian process with mean $\mathbb{E}[W_t] = y$, and covariance $\mathbb{E}[W_t W_s] = \min(t, s)$. This is one of the possible equivalent definitions of a Wiener process, and it implies that (in the 1D case)

$$P[W_t \in A] = \int_A \frac{1}{\sqrt{2\pi t}} e^{-\frac{(x-y)^2}{2t}} dx.$$

As consequence of its definition, the Wiener process W_t is a continuous function of the parameter t for all $t \in \mathbb{R}^+$, in the sense that there exists always a *continuous version* of the Wiener process (with "version of a process X_t " we mean that there exists another process Y_t such that $P[X_t = Y_t] = 1$ for any $t \in \mathbb{R}^+$). For a Wiener process, the trajectories (which can be thought as the function $\omega(t) := W_t(\omega)$) have the following properties [59]:

i) they are nowhere differentiable;

ii) they are never monotone;

iii) they have infinite variation in any interval;

iv) they have quadratic variation equal to t in the interval $[0, t]$.

More generally, let $C(\mathbb{R}^+, \mathbb{R})$ be the space of all functions $t \mapsto f_t$ taking value on $\mathbb{R}$ and continuous for any $t \in \mathbb{R}^+$. $C(\mathbb{R}^+, \mathbb{R})$ can be equipped with a norm, which allows to define open sets (i.e. a topology). As usual these open sets can be used to construct a Borel σ-algebra on $C(\mathbb{R}^+, \mathbb{R})$, say $\mathscr{B}(C(\mathbb{R}^+, \mathbb{R}))$. The Wiener process can be seen as the identity function on $(C(\mathbb{R}^+, \mathbb{R}), \mathscr{B}(C(\mathbb{R}^+, \mathbb{R}), \gamma)$ where γ is the so called *Wiener measure*. The set of all continuous functions $f_t \in C(\mathbb{R}^+, \mathbb{R})$ which does not fulfill *i*) – *iv*) have zero measure under γ. Such a probability space is called *Wiener space*. Finally we conclude by observing that

if also the starting position y is a random variable with distribution $\pi(dy)$ over $\mathbb{R}$, then

$$P[W_t \in A] = \int_A \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi t}} e^{-\frac{(x-y)^2}{2t}} \pi(dy) dx.$$

Let us now consider the space process for this model. As assumed for model A, space is discrete and evolves with time. In particular, we have the following preliminary definition which generalizes the one given for model A.

Definition 20. *Let $\{W_t^{(i)}\}_{i\in I}$ be a collection of independent Wiener processes, where $|I| = M \in \mathbb{N}$. Such collection will be called* space process *for model B.*

We will label this process by $\mathbb{S}^B$. At any given time $t \in \mathbb{R}^+$, the space process is a collection of M points on $\mathbb{R}$, which are the positions of the M Wiener processes: in this sense the space is discrete and evolves, in a continuous way, in time. Because of independence, equation (4.5) holds true if we simply substitute $P[S_N^{(i)} = s_i]$ with $P[W_t^{(i)} \in A_i]$, where $A_i \subset \mathbb{R}$ for any $i \in \{1, \cdots, M\}$, and similarly for (4.6) and generalisation. Because $P[W_t^{(i)} \in A_i]$ can be written as the integral over A_i with respect to a probability density $\rho_{W_t^{(i)}}(x_i)$, equation (4.5) is replaced by the following

$$\rho_{\mathbb{S}_t^B}(\mathbf{S}_t) = \prod_{i=1}^{M} \rho_{W_t^{(i)}}(x_i), \tag{4.18}$$

where $\rho_{\mathbb{S}_t^B}(\mathbf{S}_t)$ is the probability density of the probability measure $\tilde{P}^B[\mathbb{S}_t^B \in A]$. In a similar way one can generalise (4.6) and any other density for the space process. At this point, as done for model A, we may give the following definition for the space process.

Definition 21. *Let $\{W_t^{(i)}\}_{i\in I}$ be a collection of $M = |I| \in \mathbb{N}$ Wiener processes defined on the Wiener spaces $\{(\Omega_i, \mathcal{E}_i, P_i)\}_{i\in I}$. Let us define*

i) $\Omega_{\mathbb{S}^B} := \Omega_1 \times \cdots \times \Omega_M$*;*

ii) $\mathcal{E}_{\mathbb{S}^B}$ *is the Borel σ-algebra generated by the open sets of* $\Omega_{\mathbb{S}^B}$[d]*;*

iii) $\tilde{P}^B : \mathcal{E}_{\mathbb{S}^B} \to [0,1]$ *defined from the $\{P_i\}_{i\in I}$, via the densities as in* (4.18) *and generalisations.*

The space process is the stochastic process on $(\Omega_{\mathbb{S}^B}, \mathcal{E}_{\mathbb{S}^B}, \tilde{P}^B)$ defined as the identity function, namely $\mathbb{S}^B(\omega_1, \cdots, \omega_M) = (\omega_1, \cdots, \omega_M)$.

The set of all possible configurations of the space process at time t will be labeled by the symbol $\mathcal{S}(t)$. This completes our description for the space process in model B.

[d]To define an open set on $\Omega_{\mathbb{S}^B}$ we may use the topology induced by the norm $\|\mathbf{S}\| := \sup_{t\in[0,T]} |\mathbf{S}_t|_M$, where $|\cdot|_M$ is the M-dimensional euclidean norm. This is what is typically done on classical Wiener spaces.

4.2.2 The particle process

Again, a particle is considered as a point-like object. It jumps from one point of space to another and it is completely characterized by the *position* and *velocity* random variables.

The *position random variable*, labeled by X_t, is interpreted as the actual position of the particle at time t with respect to a chosen origin. Hence, if $(\Omega_{\mathbb{S}^B}, \mathcal{E}_{\mathbb{S}^B}, \tilde{P}^B)$ is the probability space of the space process, $(\Omega_I, \mathcal{E}_I, P_I)$ is a probability space on which an integer value stochastic process $I_t : \Omega_I \to \{1, \cdots, M\}$ is defined (called *selection process*), and $W_t^{(i_0)}$ is a chosen origin of a reference frame on $\mathbb{S}_t^B$, then

$$X_t(\omega_X) := \pi_{I_t(\omega_I)}(\mathbb{S}_t^B(\mathbf{S})) - W_t^{(i_0)}, \tag{4.19}$$

where π_i is the projector of the i-th component of an M-tuple, and $\omega_X = (\omega_I, \mathbf{S})$ with $\omega_I \in \Omega_I$ and $\mathbf{S} \in \Omega_{\mathbb{S}^B}$. Thus we have the following definition.

Definition 22. *Consider the probability space $(\Omega_I \times \Omega_{\mathbb{S}^B}, \mathcal{E}_I \otimes \mathcal{E}_{\mathbb{S}^B}, P^B)$ and a measurable space $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. The random variable X_t is the $\mathscr{B}(\mathbb{R})$-measurable function*

$$X_t : \Omega_I \times \Omega_{\mathbb{S}^B} \to \mathbb{R}$$

defined as in (4.19)*. X_t represents the position of the particle at time t.*

Clearly, as any random variable X_t induces a probability distribution $\mu_{X_t} = P^B \circ X_t^{-1}$ and, on the probability space $(\mathbb{R}, \mathscr{B}(\mathbb{R}), \mu_{X_t})$ it can be considered as the identity function. Also in this case the space process can be described on $(\Omega_I \times \Omega_{\mathbb{S}^B}, \mathcal{E}_I \otimes \mathcal{E}_{\mathbb{S}^B}, P^B)$, by simply demanding that $P^B \circ [\mathbb{S}^B]^{-1} = \tilde{P}^B$. Again if no confusion arises, we omit the suffix ^{B} in the probability measure P^B.

In model B, the particle moves by jumps from one point to another. This time the frequency of the jumps is assumed to be infinite, which means that the particle jumps from one point to another at each instant of time. In this way, we can say that it is the continuous time generalization of the kinematics described in model A. We do not generalize the definition of the velocity process given before directly. This time we use the following definition:

$$V_t(t') := \frac{X_{t'} - X_t}{t' - t} \tag{4.20}$$

where we always assume $t' > t$. More formally we adopt the following definition for $V_t(t')$.

Definition 23. *Consider the probability space $(\Omega_I \times \Omega_{\mathbb{S}^B}, \mathcal{E}_I \otimes \mathcal{E}_{\mathbb{S}^B}, P^B)$ and a measurable space $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. Let $t, t' \in \mathbb{R}$ such that $t' > t$, the random variable $V_t(t')$ is the $\mathscr{B}(\mathbb{R})$-measurable function*

$$V_t(t') : \Omega_I \times \Omega_{\mathbb{S}^B} \to \mathbb{R}$$

defined in (4.20)*. $V_t(t')$ represents the mean velocity of the particle in the interval $[t, t']$.*

Also in this case $V_t(t')$ can be seen as the identity random variable on the probability space $(\mathbb{R}, \mathscr{B}(\mathbb{R}), \mu_{V_t(t')})$, where $\mu_{V_t(t')} = P^B \circ V_t(t')^{-1}$. As in model A, for the description of the particle we need to introduce the transition probabilities. These allow to write that

$$\mu_{V_t(t')}(v) = \int_{\mathbb{R}} \alpha(v, x; t') \mu_{X_t}(x) dx \tag{4.21}$$

where $\alpha(v, x; t')$ are the probability densities of $V_t(t')$ *given the event* $\{X_t = x\}$. Note that they depend also on the value of t' used to define $V_t(t')$. In what follows we will omit t' in $\alpha(v, x; t')$ if no confusion arises.

We conclude this section defining the particle process for this model.

Definition 24. *Let X_t and $V_t(t')$ be the position and velocity process. The couple $\mathscr{P}_t(t') = (X_t, V_t(t'))$ is called* particle process *of model B.*

An example of particle process over a space process is drawn in figure 4.3.

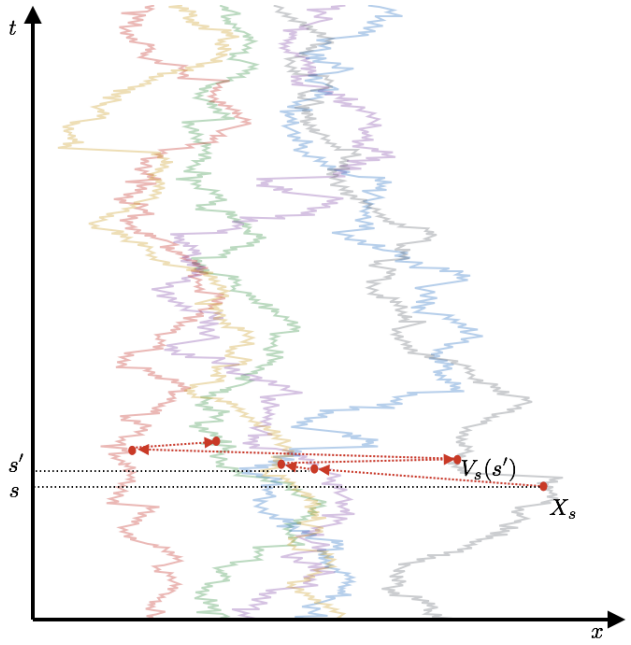

Figure 4.3: The particle process in model B is drawn in red. The position of the particle at a given time is given by the red point while its velocity at the same time is represented by the outgoing arrow. On the back, a possible realization of an $M = 6$ space process.

4.2.3 The removal of the space process

The removal of the space process in model B is done exactly as before:

i) We consider only the conditional probability densities $\mu_{X_t|\mathbf{S}_t}(x)$ and $\mu_{V_t(t')|\mathbf{S}_t}(v)$ for the random variables X_t and $V_t(t')$;

ii) We fix the transition probabilities of the single point of space, i.e. we fix the transition probabilities $p^{(i)}(x,t';y,t) := \rho_{W^{(i)}_{t'}|W^{(i)}_t=y}(x)$ of all the M Wiener processes.

As in model A, requirement $i)$ implies that we will always work with the densities $\mu_{X_t|\mathbf{S}_t}(x)$ and $\mu_{V_t(t')|\mathbf{S}_t}(v)$, namely the probability distributions for X_t and $V_t(t')$ *given the event* $\{\mathbb{S}^B_t = \mathbf{S}_t\}$. Clearly, we can define a joint probability space for X_t and $V_t(t')$ after conditioning on $\{\mathbb{S}^B_t = \mathbf{S}_t\}$ and, on this joint probability space, some *conditional* transition probabilities $\alpha_{\mathbf{S}_t}(v,x)$ can be defined. However, if we insist in using the *unconditional probability density* $\alpha(v,x)$, no joint probability space can be defined. Assuming for simplicity that all the probability distributions admits density with respect to the Lebesgue measure, the analogue of theorem 11 can be easily derived. More precisely, we have

$$\mu_{V_t(t')|\mathbf{S}_t}(v) = \int_{\Omega_{X_t}} \alpha(v,x)\mu_{X_t|\mathbb{S}^B_t}(x,\mathbf{S}_t)dx + \delta(v|X_t,\mathbf{S}_t),$$

where

$$\delta(v|X_t,\mathbf{S}_t) = \frac{1}{\mu_{\mathbb{S}^B_t}(\mathbf{S}_t)}\left[\int_{\Omega_{X_t}} \alpha(v,x)\mu_{X_t,\mathbb{S}^B_t}(x,\mathbf{S}_t)dx - \mu_{V_t(t'),\mathbb{S}^B_t}(v,\mathbf{S}_t)\right].$$

Thus again X_t and $V_t(t')$ cannot be described on the same probability space if we use the unconditional transition probabilities. Let us now analyze the consequences of $ii)$. Consider the velocity random variable of model B. Setting $\delta t := t' - t$ we can write

$$V_t(t') = V_t(t+\delta t) = \frac{X_{t+\delta t} - X_t}{\delta t}.$$

Since t is a parameter, we can always rescale it in order to have $\delta t = 1$. In this case, $V_t(t+1)$ resemble the velocity random variable V_N of model A. This can be done for any value of $t' > t$. To make this correspondence more concrete, we may also discretize the space process. More precisely, since the points of the space process are Wiener processes taking values on the real line, we can partition $\mathbb{R}$ in intervals $\{\Delta_k\}_{k\in K}$ (i.e. $\cup_{k\in K}\Delta_k = \mathbb{R}$) where $K \subset \mathbb{N}$. At this point one can consider the discretized random variable for the space process and the position random variable. If $W^{(i)}_t$ is a point of the space process $\mathbb{S}^B_t$, one can define a new random variable $S^{(i)}_t : \Omega_{\mathbb{S}^B} \to K$ as

$$S^{(i)}_t(\omega) := k \qquad \text{if } W^{(i)}_t(\omega) \in \Delta_k,$$

which simply reveals in which Δ_k the Wiener process is. Clearly, $P[S^{(i)}_t = k] = \int_{\Delta_k} \rho_{W^{(i)}_t}(x)dx$ and given the transition probability densities for the Wiener

process, say $p^{(i)}(x, t+1; y, t)$, the transition probabilities for $S_t^{(i)}$ are given, i.e.

$$P[S_{t+1}^{(i)} = k | S_t^{(i)} = j] = \frac{\int_{\Delta_k} dx \int_{\Delta_j} dy p^{(i)}(x, t+1; y, t) \rho_{W_t^{(i)}}(y)}{\int_{\Delta_j} \rho_{W_t^{(i)}}(y) dy}.$$

At the end of this procedure one ends up with a discretized version of the space process of model B, which is equivalent to the one used in model A. The same discretization procedure can be done for the position and velocity random variables X_t and $V_t(t+1)$. It is not difficult to realize that theorem 12 can be applied and its application does not depend on the size of the sets $\{\Delta_k\}_{k\in K}$. Thus, as in the previous model, the requirement *ii*) implies the entropic uncertainty relation between the position and the velocity random variables. Then as in model A, this relation can be used to prove that X_t and $V_t(t')$, after conditioning on $\mathbf{S}_t$, are representable as two non-commuting operators on the same Hilbert space. This will be discussed in the next section.

Let us now describe a bit further how to obtain the entropic uncertainty relation from the discretization of model B. First of all, if we want to apply theorem 10, we need to be sure that the two random variables are bounded, i.e. the set of all values they can assume is a bounded set. In fact, only in this case, they can be associated to two bounded self-adjoint operators, which are elements of a C^*-algebra, and the relation between non-commutativity and the entropic uncertainty relation holds true. In order to do that, we consider the restriction of the two random variables to a given subset. More precisely, given $\Lambda \subset \mathbb{R} = \Omega_{X_t}$, the *bounded version* of X_t will be the random variable

$$X_t|_\Lambda(\omega) := X_t(\omega)\chi_\Lambda(X_t(\omega))$$

where $\chi_\Lambda(x)$ is the indicator function of the set Λ. Similarly, we can define the bounded version of $V_t(t')|_\Gamma$. At this point we consider the discrete version of these random variables, similarly to what we did for the space process. Given $X_t|_\Lambda$, we can discretise it simply by dividing the set Λ in N parts of equals size, obtaining a partition $\{\Delta_{N,k}^X\}_{k\in K}$, $K \subset \mathbb{N}$, such that $|\Delta_{N,k}^X| = |\Delta_{N,k'}^X|$ for any possible k. We can see that the number of subsets of the partition (i,e. $N = |K|$) determines the width of the sets $\Delta_{N,k}^X$. The bounded and *discrete version* of X_t is then defined as

$$X_t^\Delta|_\Lambda(\omega) := k \qquad \text{if } X_t(\omega)|_\Lambda \in \Delta_{N,k}^X.$$

The same construction can be done for the bounded version of $V_t(t')$, using in general a different partition $\{\Theta_{N',j}^V\}_{j\in J}$, obtaining $V_t(t')^\Theta|_\Gamma$. It is useful to choose the partitions for X_t and $V_t(t')$ compatible with the partition used for the space process. To do that it is enough to set the partition for X_t and $\mathbb{S}_t^B$ equal and choose the partition for $V_t(t')$ consequently. Finally we also chose to set $|\Gamma| = |\Lambda|$, i.e. the size of the two set used to bound the position and velocity random variable coincides. At this point, by discretising time as explained above $X_t^\Delta|_\Lambda$ and $V_t(t+1)^\Theta|_\Gamma$ (for simplicity we simply write $V_t^\Theta|_\Gamma$) become discrete

random variables similar to those used for model A. Then, applying theorem 12, we know that

$$H_{\mathbf{S}_t}(X_t^{\Delta}|_{\Lambda}) + H_{\mathbf{S}_t}(V_t^{\Theta}|_{\Gamma}) \geqslant D, \tag{4.22}$$

where D is a positive constant that in general can depend on the partition chosen but not on the probability distribution of $X_t^{\Delta}|_{\Lambda}$ and $V_t^{\Theta}|_{\Gamma}$, hence on $\mu_{X_t|\mathbf{S}_t}$ and $\mu_{V_t|\mathbf{S}_t}$. The whole construction does not depend on the partitions chosen, once they are chosen in the consistent way explained above. In particular, the above inequality holds for arbitrary partitions having small but finite size.

4.2.4 Construction of the Hilbert space structure for model B

The construction of the Hilbert space structure for model B goes more or less as in model A. However, in this case, we have some additional technicalities due to the use of the partitions for the description of the two random variables involved. The entropic uncertainty relation (4.22), ensures that $X_t^{\Delta}|_{\Lambda}$ and $V_t^{\Theta}|_{\Gamma}$, after conditioning on $\mathbf{S}_t$, can be jointly described only on a non-commutative probability space, i.e. with non-commuting operators. Let us fix for the moment the partitions used. As in model A, the bounded and discrete version of the position random variable can be represented on the Hilbert space

$$\mathcal{H}(X_t|N,\Lambda) = \bigoplus_{k=1}^{N} \mathcal{H}_k,$$

as the diagonal operator

$$\hat{X}_t(N,\Lambda) = \sum_{k=1}^{N} k|k\rangle\langle k|.$$

Here $|k\rangle \in \mathcal{H}_k$ and $\hat{P}_k^{(\hat{X}_t(N,\Lambda))} := |k\rangle\langle k|$ is the PVM such that

$$P[X_t^{\Delta}|_{\Lambda} = k] = P[X_t|_{\Lambda} \in \Delta_{N,k}^{X}] = \langle\psi|\hat{P}_k^{(\hat{X}_t(N,\Lambda))}|\psi\rangle \tag{4.23}$$

for some $\psi \in \mathcal{H}(X_t|N,\Lambda)$. Similarly, the bounded and discrete version of the velocity random variable can be represented on the Hilbert space

$$\mathcal{H}(V_t|N,\Gamma) = \bigoplus_{j=1}^{N} \mathcal{H}_j$$

(note that particular partitions considered implies that $N = N'$) as the diagonal operator

$$\hat{V}_t(N,\Gamma) = \sum_{j=1}^{N} j|j\rangle\langle j|.$$

The two Hilbert spaces $\mathcal{H}(X_t|N,\Lambda)$ and $\mathcal{H}(V_t|N,\Gamma)$ have the same dimension and so they are unitary equivalent, i.e. there exists a unitary map $\hat{U}:\mathcal{H}(V_t|\Gamma)\to\mathcal{H}(X_t|\Lambda)$. Hence we can represent $\hat{V}_t(N,\Gamma)$ on $\mathcal{H}(X_t|N,\Lambda)$ and viceversa. The entropic uncertainty relation (4.22) ensures that

$$[\hat{X}_t(N,\Lambda),\hat{V}_t(N,\Gamma)]\neq 0 \tag{4.24}$$

Let us now analyse what happens when we change the size of the partition. First, we consider the limit $N\to\infty$ which means that the size of the partitions goes to zero. Because the sets $\Delta^X_{N,k}$ shrink to a point, say $\{x\}$, we have

$$\lim_{N\to\infty}\langle\psi|\hat{P}_k^{(\hat{X}_t(N,\Lambda))}|\psi\rangle=\lim_{N\to\infty}P[X_t|_\Lambda\in\Delta^X_{N,k}]=0 \tag{4.25}$$

for any ψ, i.e. any P. This means, by prop 9.14 of [3], $x\in\sigma_c(\hat{X}_t(\Lambda))$ (here $\hat{X}_t(\Lambda):=\hat{X}_t(\infty,\Lambda)$). By the arbitrariness of x we conclude, as expected, that $\hat{X}_t(\Lambda)$ is a bounded operator with purely continuous spectrum. Note that the Hilbert space on which we can define $\hat{X}_t(\Lambda)$ is

$$\mathcal{H}(X_t|\Lambda):=\int_\Lambda^\oplus\mathcal{H}_x dx$$

which is not separable in general. Here, $\hat{X}_t(\Lambda)$ can be written as

$$\hat{X}_t(\Lambda)=\int_\Lambda xP^{(\hat{X}_t(\Lambda))}(dx).$$

Similar conclusions hold for the operator representing the bounded and discrete velocity random variable: $\hat{V}_t(\Gamma):=\hat{V}_t(\infty,\Gamma)$ is a bounded operator with continuous spectrum. Since for any value of N, $\hat{X}_t(N,\Lambda)$ is the operator representing the random variable obtained by discretizing the *same* random variable $X_t|_\Lambda$, also the operators $\hat{X}_t(N,\Lambda)$ can be obtained by discretising the same operator $\hat{X}_t(\Lambda)$. The same holds for $\hat{V}_t(N,\Gamma)$. At this point because (4.24) is valid for any possible partition chosen in the consistent way explained in the previous section (i.e. for any N), we can conclude that

$$[\hat{X}_t(\Lambda),\hat{V}_t(\Gamma)]\neq 0.$$

Since Γ and Λ are arbitrary, with similar considerations we may conclude that

$$[\hat{X}_t,\hat{V}_t]\neq 0 \tag{4.26}$$

where $\hat{X}_t$ is the unbounded operator on a Hilbert space $\mathcal{H}(X_t):=\mathcal{H}(X_t|\mathbb{R})$ such that $\hat{X}_t(\Lambda)=\hat{P}_\Lambda\hat{X}_t\hat{P}_\Lambda$ (here $\hat{P}_\Lambda$ is the projector from $\mathcal{H}(X_t)$ to the Hilbert space $\mathcal{H}(X_t|\Lambda)$) and $\hat{V}_t$ is defined in a similar manner.

We conclude by observing that $\mathcal{H}(X_t|\Lambda)$ and $\mathcal{H}(V_t|\Gamma)$ may be not separable (and so also $\mathcal{H}(X_t)$ and $\mathcal{H}(V_t)$). In general, non-separable infinite-dimensional Hilbert spaces are not mutually isomorphic (see [58], corr. 5.5). Thus in this

case we cannot define a unitary map $\hat{U} : \mathcal{H}(V_t) \to \mathcal{H}(X_t)$ which maps the operator representations of X_t and V_t on $\mathcal{H}(V_t)$ into the corresponding operators in $\mathcal{H}(X_t)$. This is an effect of the possible lack of separability of the Hilbert spaces $\mathcal{H}(X_t)$ and $\mathcal{H}(V_t)$. However this does not mean that we cannot represent the velocity random variable on $\mathcal{H}(X_t|\Lambda)$ and vice-versa: one simply represents the velocity random variable on $\mathcal{H}(X_t|N,\Lambda)$ and then takes the limit. However, to have a consistent description the velocity operator obtained in this limit must be isomorphic to the operator $\hat{V}_t$ diagonal on $\mathcal{H}(V_t|\Gamma)$. We will refer to this problem with the name "separability problem" and we will comment on it in the next section. We conclude by observing that the result obtained here, as explained in the previous section, holds for any value of $t' > t$.

4.3 Comparison with quantum mechanics

In model A it turns out that the Hilbert space is finite dimensional and the dimension is determined by the number of points. Hence, even if this model is capable to reproduce the non-commutativity of quantum mechanics, surely the commutator between $[\hat{X}_N, \hat{V}_N]$ cannot be equal to (4.1). Indeed, it is well known that (4.1) does not hold on a finite dimensional Hilbert space. Taking the limit $M \to \infty$, a comparison with quantum mechanics is not possible since time is a discrete parameter in model A.

On the other hand in model B, the construction presented leads to an infinite dimensional Hilbert space. It may be not separable while in ordinary quantum mechanics the Hilbert space always is. Comparing this model with model A, we can understand that this time the number of points in the space process, M, does not determine the dimension of the Hilbert space. After a bit of thought, one can realize that this is a consequence of the fact that we are using probability measures which admit a density with respect to the Lebesgue measure. Another consequence of this fact is the continuous spectrum of the operators representing the particle process. However, one can always imagine that, if we let the support of the probability measure shrink to a single point (hence obtaining a Dirac measure, which is not absolutely continuous with respect to the Lebesgue measure), the operators have a pure point spectrum. This suggests that the "real" continuity of the spectrum is obtained only in the limit $M = \infty$ and the absolute continuity is possible only in this case. One may observe the following. When $M \to \infty$ we can have two cases:

a) the points increase in a *non dense* way: their number is infinite but in any subset of $\mathbb{R}$ these is just a finite number of them (they behaves as numbers in $\mathbb{N}$ or $\mathbb{Z}$);

b) the points increase in a *dense* way: their number is infinite and in any subset of $\mathbb{R}$ there is an infinite number of them (like numbers in $\mathbb{Q}$). We will refer to this case with the name *dense-point limit*.

Note that in both cases they are assumed to be countable. In the first case, $\hat{X}_t$ can be seen as the limit of a sequence of compact operators: the spectrum is

purely point-like. However, this possibility does not seem to be comparable with the usual position operator in quantum mechanics, which is bounded (and not compact) when we restrict it to a subset of $\mathbb{R}$. On the other hand, the second case is more interesting. Indeed, it may give rise to bounded operators which are not compact. This suggests that to completely recover quantum mechanics, the dense-point limit must be taken.

Despite the observations done above, we still want to try a comparison with non-relativistic quantum mechanics. This time we are really closer to deriving the canonical commutation relation between position and momentum from the quantities of the model, as we will see. Assume the following:

i) The Hilbert space on which we can represent $\hat{X}_t$ is separable and infinite-dimensional, i.e. $L_2(\mathbb{R})$;

ii) There exists a self-adjoint operator

$$\hat{H} = \frac{1}{2m}\nabla_x^2 + V(x)$$

which, together with $\hat{X}_t$, fulfils all the mathematical requirements needed to apply the Ehrenfest theorem (see [60]).

Clearly $\hat{H}$ is nothing but the ordinary hamiltonian operator in quantum mechanics. At this point, by the Ehrenfest theorem, we have the equation

$$\frac{d}{dt}\langle \hat{X}_t \rangle_\psi = \frac{1}{m}\langle \hat{P}_t \rangle_\psi,$$

where m is the mass of the quantum particle and $\psi \in L_2(\mathbb{R})$. Consider now the velocity random variable of the model B

$$V_t(t') = \frac{X_{t'} - X_t}{t' - t}.$$

Note that, after the removal of the space process, the three random variables lies in three different probability spaces and there does not exist a joint probability space where we can describe all of them (we recall that when we remove the space we use the *unconditional* transition probabilities). Thus this expression is purely formal and, in particular, it is not expected to hold at the level of the outcomes of these random variables. However, the following expression makes sense

$$\mathbb{E}[V_t(t')|\mathbf{S}_t] = \frac{\mathbb{E}[X_{t'}|\mathbf{S}_t] - \mathbb{E}[X_t|\mathbf{S}_t]}{t' - t}$$

since the probability measures of each expectation are defined on different probability spaces. Using the procedure explained in the previous section and under the assumption i), we can jointly describe these three random variables using a non-commutative probability space. In particular, we compute the expectation using the Hilbert space structure, writing

$$\mathbb{E}[V_t(t')|\mathbf{S}_t] = \langle \hat{V}_t(t') \rangle_\psi = \frac{\langle \hat{X}_{t'} \rangle_\psi - \langle \hat{X}_t \rangle_\psi}{t' - t},$$

where $\psi \in \mathcal{H}$ with $\mathcal{H}$ Hilbert space constructed as in section 4.2.4. This time an explicit procedure to construct ψ is not known despite some results in this case can be found in [Cur7]. From this equation we can write that

$$\lim_{t' \to t} \langle \hat{V}_t(t') \rangle_\psi = \frac{d}{dt} \langle \hat{X}_t \rangle_\psi = \frac{1}{m} \langle \hat{P}_t \rangle_\psi,$$

which means that the *weak*-limit $t' \to t$ of velocity operator in model B, under assumptions i) and ii), coincides with the momentum operator of non-relativistic quantum mechanics. Note that the assumption i) on the separability of the Hilbert space is crucial for this consideration. Finally, we also note that separability also solves the problem of the non-unitary equivalence of $\mathcal{H}(X_t)$ and $\mathcal{H}(V_t)$ mentioned at the end of section 4.2.4.

Summarising, despite Model B is capable to reproduce the commutation relation between the position and velocity operators of the particle, which resembles the quantum mechanical commutation relation, it did not succeed in the derivation of (4.1). However, if in some other model (similar to model B) we can justify i) and ii) in some way, we can have a correspondence of the model with non-relativistic quantum mechanics.

Chapter 5

Probabilistic intermezzo: point processes

In this chapter we will introduce the notion of point process. A particular class of point processes, the determinantal one, will be analyzed and some interesting features discussed. These notions will we used in chapter 6, where an attempt to justify the assumption *i*) and *ii*) of section 4.3 is presented.

5.1 Point processes

A simple introduction to the general theory of point processes can be found in [61], while a more systematic approach is presented in [62]. These two are the main references for the concepts introduced here.

5.1.1 General structure

Before starting with the formal mathematical description, let us first explain what point processes are and how they are described. Let X be a d-dimensional space and let ξ label a point process on it. Loosely speaking, ξ can be thought as a collection points of X randomly chosen according to some probability distribution. To describe it, on a d-dimensional space, the most natural way is by counting the number of points that fall in a given subset of X. Let $B \subset \mathsf{X}$ and let $N_\xi(B)$ be the number of points of ξ inside B. It is not difficult to imagine that we can completely describe the point process ξ by knowing $N_\xi(B)$ for any subset B of X, i.e. by knowing the collection $\{N_\xi(B)\}_{B\subset\mathsf{X}}$. More precisely, we can say that $\{N_\xi(B)\}_{B\subset\mathsf{X}}$ contains enough information to recover the position of all points of ξ. Note that this is true only if the points of ξ are not too dense: if $N_\xi(B) = \infty$ for any $B \subset \mathsf{X}$ we are not able to reconstruct the single points locations by a counting technique. Let us now formalise this idea.

Definition 25. *Let X be a locally-compact second countable Hausdorff space and let $\mathscr{B}(\mathsf{X})$ be the Borel σ-algebra on it. Let $N : \mathscr{B}(\mathsf{X}) \to \mathbb{N}$ be a non-negative*

integer valued counting measure on X. *When* $N(B) < \infty$ *for any* $B \in \mathscr{B}(\mathsf{X})$ *bounded, we say that the counting measure is* boundely finite. *The set*

$$\mathcal{N}_{\mathsf{X}} := \{N(\cdot) | N(B) < \infty \text{ for all bounded } B \in \mathscr{B}(\mathsf{X})\},$$

is called space of all the boundely finite counting measures on X.

The space $\mathcal{N}_{\mathsf{X}}$ can be equipped with the vague topology, which allows to define open sets that can be used to construct a Borel σ-algebra $\mathscr{B}(\mathcal{N}_{\mathsf{X}})$.

Definition 26. *Let* $(\Omega, \mathcal{E}, P)$ *be a probability space,* X *be a locally-compact second countable Hausdorff space and* $\mathcal{N}_{\mathsf{X}}$ *be the space of all the boundely finite counting measure on it. The* P*-measurable map* $\xi : (\Omega, \mathcal{E}, P) \to (\mathcal{N}_{\mathsf{X}}, \mathscr{B}(\mathcal{N}_{\mathsf{X}}))$ *defined as*

$$\xi : \omega \mapsto \{N_\xi(B; \omega)\}_{B \in \mathscr{B}(\mathsf{X})},$$

is called point process *over* X.

Let us explain better how this definition fits with the idea explained in the beginning. Given $\omega \in \Omega$ the realisation $\xi(\omega)$ of the point process is fixed. Then the number of points of $\xi(\omega)$ that fall in B is $N_{\xi(\omega)}(B) = N_\xi(B; \omega)$. Since we have this information for any Borel set $B \subset \mathsf{X}$ we can reconstruct the whole collection of points of $\xi(\omega)$. If no confusion arises, we will omit the dependence of ξ on ω. Note that in general, such a collection of points cannot be regarded as a discrete random subset of X. Indeed, there can be point processes whose points may overlap, namely $N_\xi(\{x\}) = m > 1$ for some $x \in \mathsf{X}$, and others where this does not happens.

Definition 27. *A point process* ξ *on* X *is said* simple *if* $N_\xi(\{x\}) \leqslant 1$ *for all* $x \in \mathsf{X}$.

In case of simple point processes, ξ can be represented as the (random) subset $\xi := \{x_1, x_2, \cdots\} \subset \mathsf{X}$ without loosing any information. Such collection of points is also called *configuration* of ξ. In general we may always represent the point process ξ as a collection of points, i.e. $\xi := \{x_1, x_2, \cdots\}$, but when the process is not simple they do not form a subset of X because there can be $x_i = x_j$ for some i, j: this information is lost if ξ is thought as a set.

Definition 28. *Let* ξ *be a point process on* X. *If* $N_\xi(\mathsf{X}) = M \leqslant \infty$ *then* ξ *is said* finite, *while if* $N_\xi(\mathsf{X}) = \infty$ *the point process is said* locally-finite.

Note that any point process which is locally finite on X, induces a finite point process on any compact subset $\Lambda \subset \mathsf{X}$ by construction.

Let us now explain how a point process is described from the statistical point of view. Consider a *finite* point process ξ on X, and the probability distibution $\mu_\xi := P \circ \xi^{-1}$ induced on $(\mathcal{N}_{\mathsf{X}}, \mathscr{B}(\mathcal{N}_{\mathsf{X}}))$. When the distribution μ_ξ is used to compute the probabilities of events like

$$\{N(A_1) = n_1\} \cap \cdots \cap \{N(A_r) = n_r\}, \tag{5.1}$$

with $A_i \in \mathscr{B}(\mathsf{X})$ and $n_i \in \mathbb{N}$ for all $i = 1, \cdots, r$, μ_ξ takes a special name.

Definition 29. *Let $\xi : (\Omega, \mathcal{E}, P) \to (\mathcal{N}_{\mathsf{X}}, \mathscr{B}(\mathcal{N}_{\mathsf{X}}))$ be a point process, the probability distributions*

$$\mu_\xi(A_1, n_1; \cdots ; A_r, n_r) := P[N_\xi(A_1) = n_1, \cdots , N_\xi(A_r) = n_r],$$

where $A_i \in \mathscr{B}(\mathsf{X})$ and $n_i \in \mathbb{N}$ for all $i = 1, \cdots , r$, are called finite-dimensional distributions *(or* fidi*s) of the point processes ξ.*

The importance of the fidis for the description of a point process is encoded in the following theorem [61].

Theorem 13. *Let $\xi : (\Omega, \mathcal{E}, P) \to (\mathcal{N}_{\mathsf{X}}, \mathscr{B}(\mathcal{N}_{\mathsf{X}}))$ and $\eta : (\Omega, \mathcal{E}, P) \to (\mathcal{N}_{\mathsf{X}}, \mathscr{B}(\mathcal{N}_{\mathsf{X}}))$ be two point processes on X. If all the fidis of ξ and η coincide, then ξ and η have the same distribution.*

This means that, if the fidis of two point processes are the same then they are equal in distribution, namely $\xi \stackrel{d}{=} \eta$. Typically point processes are not described in terms of fidis directly, but by using two quantities, the Janossy measure and the moment measure, from which the fidis can be derived. Let us introduce the Janossy measure. Consider the event (5.1), used to compute the fidi $\mu_\xi(A_1, n_1; \cdots ; A_r, n_r)$ and assume that $\{A_i\}_{i=1}^r$ is a *finite partition of* X (i.e. $\cup_{i=1}^r A_i = \mathsf{X}$ and $A_i \cap A_j = \{\varnothing\}$ for any $i \neq j$). When this event happens, the point process contains exactly $n = n_1 + \cdots + n_r$ points. This observation allows us to write the following:

$$\mu_\xi(A) = \sum_{n=0}^{\infty} p_n \Pi_n(A)$$

where $A \in \mathscr{B}(\mathcal{N}_{\mathsf{X}})$ is a generic event and

i) $\{p_n\}_{n=0}^{\infty}$ are the probabilities that the point process has exactly n points, thus they fulfil the normalisation condition

$$\sum_{i=0}^{\infty} p_n = 1; \tag{5.2}$$

ii) $\Pi_n(A)$ is a probability distribution on $\mathsf{X}^{(n)}$, i.e. the n-fold product space $\mathsf{X} \times \cdots \times \mathsf{X}$, which can be interpreted as the probability distribution of the position of the points of ξ, given that their number is exactly n.

To implement indistinguishability of the points of ξ, the joint probability distribution Π_n should assign equal weight to any permutation of the coordinates $(x_1, \cdots , x_n)$. If Π_n is not so, we can always implement indistinguishability by introducing the symmetrised form

$$\Pi_n^{sym}(A_1 \times \cdots \times A_n) = \frac{1}{n!} \sum_{\sigma \in P_n} \Pi_n(A_{\sigma(1)} \times \cdots \times A_{\sigma(n)})$$

where P_n is the set of all the permutations of n elements and $A_1, \cdots , A_r \in \mathscr{B}(\mathsf{X})$ form a finite partition of X.

Definition 30. *Let $\xi : (\Omega, \mathcal{E}, P) \to (\mathcal{N}_{\mathsf{X}}, \mathscr{B}(\mathcal{N}_{\mathsf{X}}))$ be a point process on X. Taken a finite partition $A_1, \cdots, A_n \in \mathscr{B}(\mathsf{X})$ the (n-th)* Janossy measure *of ξ is defined as*

$$\begin{aligned} J_n(A_1 \times \cdots \times A_n) :&= p_n \sum_{\sigma \in P_n} \Pi_n(A_{\sigma(1)} \times \cdots \times A_{\sigma(n)}) \\ &= n! p_n \Pi_n^{sym}(A_1 \times \cdots \times A_n). \end{aligned}$$

From the normalisation condition (5.2) we can see that J_n is not a probability measure. In fact, observing that $\Pi_n(\mathsf{X}^{(n)}) = 1$, we can write

$$\sum_{n=0}^{\infty} \frac{J_n(\mathsf{X}^{(n)})}{n!} = 1, \tag{5.3}$$

where we interpret $J_0(\mathsf{X}^{(0)}) = p_0$. For any $n \geqslant 1$ we have

$$J_n(\mathsf{X}^{(n)}) = p_n n! \tag{5.4}$$

It is clear that any family of symmetric measures fulfilling the normalisation condition (5.3), can be used to construct the probability distribution $\{p_n\}_{n=0}^{\infty}$, using (5.4), and so also the sets of symmetric probability distributions $\{\Pi_n^{sym}\}_{n=0}^{\infty}$. Now we want to show explicitly how to construct the fidis of the point process from the Janossy measures. In order to do that, we recall that the multinomial coefficient

$$\binom{n}{n_1, \cdots, n_r} = \frac{n!}{n_1! \cdots n_r!}$$

counts the number of ways we may arrange $n = n_1 + \cdots + n_r$ objects in r different boxes putting n_1 objects in the 1-th box, ..., n_r objects in the r-th box. This implies that

$$\begin{aligned} \mu_\xi(A_1, n_1; \cdots; A_r, n_r) &= p_n \binom{n}{n_1, \cdots, n_r} \Pi_n^{sym}(A_1^{(n_1)} \times \cdots \times A_r^{(n_r)}) \\ &= \frac{J_n(A_1^{(n_1)} \times \cdots \times A_r^{(n_r)})}{n_1! \cdots n_r!}, \end{aligned} \tag{5.5}$$

where $A_1, \cdots A_n$ form a finite partition on X. In general, the sets on which the fidis can be evaluated do not form a partition of X, however we can still find them by the Janossy measure. Suppose that $A_1, \cdots, A_r$ are disjoint sets but they do not from a partition. Hence there exists a set $C = (A_1 \cup \cdots \cup A_r)^c$ such that $\mathsf{X} = A_1 \cup \cdots \cup A_r \cup C$ and clearly $A_i \cap C = \{\varnothing\}$. Thus $A_1, \cdots, A_r, C$ is a partition of X and the previous formula applies. Let $n = n_1 + \cdots + n_r$ and $s \in \mathbb{N}$, we can write that

$$\mu_\xi(A_1, n_1; \cdots; A_r, n_r; C, s) = \frac{J_{n+s}(A_1^{(n_1)} \times \cdots \times A_r^{(n_r)} \times C^{(s)})}{n_1! \cdots n_r! s!}.$$

Using the law of total probability, we have

$$\mu_\xi(A_1, n_1; \cdots; A_r, n_r) = \sum_{s=0}^{\infty} \mu_\xi(A_1, n_1; \cdots; A_r, n_r; C, s)$$

and so

$$\mu_\xi(A_1, n_1; \cdots; A_r, n_r) = \frac{1}{n_1! \cdots n_r!} \sum_{s=0}^{\infty} \frac{J_{n+s}(A_1^{(n_1)} \times \cdots \times A_r^{(n_r)} \times C^{(s)})}{s!}. \tag{5.6}$$

At this point it should be clear that Janossy measures are important tools to describe a point process. However nothing was said about their meaning. We already observed that they are not probability measures, thus a statistical interpretation is not available in general. By the way, in some particular case such interpretation is available.

Definition 31. *Let $\xi : (\Omega, \mathcal{E}, P) \to (\mathcal{N}_{\mathsf{X}}, \mathscr{B}(\mathcal{N}_{\mathsf{X}}))$ be a point process on X and let $J_n(A_1 \times \cdots \times A_n)$ be a n-th Janossy measure of the process. Let μ be a Borel measure on X. The function $j_n(x_1, \cdots, x_n)$ such that*

$$J_n(A_1 \times \cdots \times A_n) = \int_{A_1} \cdots \int_{A_n} j_n(x_1, \cdots, x_n) \mu(dx_1) \cdots \mu(dx_n)$$

is called (n-th) Janossy density *of the point process ξ with respect to μ.*

The Janossy densities have a particularly simple interpretation, in fact $j_n(x_1, \cdots, x_n)\mu(dx_1) \cdots \mu(dx_n)$ represents the *probability that there are exactly n points in the process ξ, one in each of the n distinct infinitesimal regions $(x_i, x_i + dx_i)$*. The notion of Janossy measure (or density, if it exist) can be extended to the case of locally finite point processes by restricting the point process ξ on some compact region $\Lambda \subset \mathsf{X}$. Everything remains the same except that the normalisation condition (5.3) and (5.4) are not computed with $\mathsf{X}^{(n)}$ but by using $\Lambda^{(n)}$. Similarly in the equation (5.6), $A_1, \cdots, A_r$ are a partition of Λ not of X. For this reason, it is typical to make this Λ-dependence explicit by adding a label in the symbol of the Janossy measure (or density), i.e. $J_n(A_1 \times \cdots \times A_n|\Lambda)$ and $j_n(x_1. \cdots, x_n|\Lambda)$ which are called *local Janossy measure* and *density*, respectively. The existence of a the Janossy density when $\mathsf{X} = \mathbb{R}^d$ and the measure μ is the ordinary Lebesgue measure, can be used to define an important class of point processes.

Definition 32. *Let ξ be a point process on $\mathsf{X} \subseteq \mathbb{R}^d$. If, for all $n \geqslant 1$ and some bounded $A \in \mathscr{B}(\mathbb{R}^d)$, the local Janossy measures $J_n(dx_1 \times \cdots \times dx_n|A)$ exist and are absolutely continuous with respect to the Lebesgue measure, the point process is said* regular on A. *If this happens for any bounded $A \in \mathscr{B}(\mathbb{R}^d)$, then ξ is said* regular point process.

Regularity is important because it implies *simpleness*: more precisely, if a point process ξ on $\mathbb{R}^d$ has Janossy measure admitting Janossy densities with respect to the Lebesgue measure, then ξ is simple (see Prop. 5.4 V in [62]).

Let us now describe the second quantity which is typically used to describe a point process: the moment measure. By definition, a point process is described by using *random measures*, i.e. measure-valued random variables. Moment measures are just the moments of these random variables, which are measure-valued. This description of a point process ξ on X, is clearly related to the description via the Janossy density. Let us start with the simplest case: the *intensity measure* or *1st-moment measure*. The intensity measure of a locally finite point process ξ in X is defined as the measure

$$M_1(A) := \mathbb{E}\Big[\sum_{x\in\xi}\chi_A(x)\Big], \tag{5.7}$$

where χ_A is the indicator function of the set $A \in \mathscr{B}(\mathsf{X})$. The intensity measure is by definition the expectation value of the counting measure on A, i.e. $M_1(A) = \mathbb{E}[N_\xi(A)]$, and so it can be interpreted as the expected number of points of ξ in A. Similarly one can define the *2nd-moment measure* as

$$M_2(A_1 \times A_2) := \mathbb{E}\Big[\sum_{x,y\in\xi}\chi_{A_1\times A_2}(x,y)\Big]. \tag{5.8}$$

Also in this case, we can conclude that $M_2(A_1 \times A_2) = \mathbb{E}[N_\xi(A_1)N_\xi(A_2)]$. Recognising that $N_\xi(A_1)N_\xi(A_2)$ is the number of elements of the set $\{(x,y) \in A_1 \times A_2, x, y \in \xi\}$, we can interpret the 2-th moment measure as the intensity measure of a point process on $\mathsf{X}\times\mathsf{X}$. In this definition of M_2, two contributions can be distinguished:

$$\begin{aligned} M_2(A_1 \times A_2) &= \mathbb{E}\Big[\sum_{x,y\in\xi}\chi_{A_1\times A_2}(x,y)\Big] = \mathbb{E}\Big[\sum_{\substack{x,y\in\xi\\ x\neq y}}\chi_{A_1\times A_2}(x,y)\Big] + \mathbb{E}\Big[\sum_{x\in\xi}\chi_{A_1\times A_2}(x,x)\Big] \\ &= \mathbb{E}\Big[\sum_{\substack{x,y\in\xi\\ x\neq y}}\chi_{A_1\times A_2}(x,y)\Big] + M_1(A_1 \cap A_2), \end{aligned}$$

where we used the fact that $\chi_{A_1\times A_2}(x,x)$ is non zero only for $x \in A_1 \cap A_2$, i.e. $\chi_{A_1\times A_2}(x,x) = \chi_{A_1\cap A_2}(x)$. The quantity

$$M_{[2]}(A_1 \times A_2) := \mathbb{E}\Big[\sum_{\substack{x,y\in\xi\\ x\neq y}}\chi_{A_1\times A_2}(x,y)\Big] \tag{5.9}$$

is called *2nd-factorial moment*. Like for the 2nd moment measure, the 2nd factorial moment measure can be seen as the intensity measure of a point process

on $\mathsf{X} \times \mathsf{X}$ consisting of all the 2-tuples of distinct points of the original process ξ. Note that

$$\begin{aligned} M_{[2]}(A \times A) &= M_2(A \times A) - M_1(A) \\ &= \mathbb{E}[N_\xi(A)^2] - \mathbb{E}[N_\xi(A)] \\ &= \mathbb{E}[N_\xi(A)(N_\xi(A) - 1)]. \end{aligned}$$

The moment measures defined by (5.7), (5.8) and (5.9) can be generalised as follows.

Definition 33. *Let ξ be a point process on X, $n \in \mathbb{N}$ and $A_1, \cdots, A_n \in \mathscr{B}(\mathsf{X})$. The measure*

$$M_n(A_1 \times \cdots \times A_n) := \mathbb{E}\Big[\sum_{x_1, \cdots, x_n \in \xi} \chi_{A_1 \times \cdots \times A_n}(x_1, \cdots, x_n) \Big]$$

is said n-th moment measure *of the process ξ, while the measure*

$$M_{[n]}(A_1 \times \cdots \times A_n) := \mathbb{E}\Big[\sum_{\substack{x_1, \cdots, x_n \in \xi \\ x_1 \neq \cdots \neq x_n}} \chi_{A_1 \times \cdots \times A_n}(x_1, \cdots, x_n) \Big]$$

is said n-th factorial moment measure *of the process ξ.*

In general, these measures may not exist for any n (they can be infinte sometimes). The n-th moment measure can be written as the expectation of a product of counting measures. More precisely, let $A_1, \cdots, A_r \in \mathscr{B}(\mathsf{X})$ with $r \leqslant n$, then

$$M_n(A_1^{(n_1)} \times \cdots \times A_r^{(n_r)}) = \mathbb{E}[(N_\xi(A_1))^{n_1} \cdots (N_\xi(A_r))^{n_r}]$$

where $n_i \in \mathbb{N}$ for all $i = 1, \cdots, r$ and $n_1 + \cdots + n_r = n$. A similar formula for the n-th factorial moment is not available in general, but a similar result holds for disjoint sets. Indeed, introducing the *r-th factorial power of x*

$$x^{[r]} := \begin{cases} x(x-1)\cdots(x-r+1) & \text{for } r \leqslant x \\ 0 & \text{otherwhise.} \end{cases}$$

where $x \in \mathbb{R}$ and $r \in \mathbb{N}$, if $A_1, \cdots, A_r \in \mathscr{B}(\mathsf{X})$ with $r \leqslant n$ are disjoint sets, one can write that

$$M_{[n]}(A_1^{(n_1)} \times \cdots \times A_r^{(n_r)}) = \mathbb{E}[(N_\xi(A_1))^{[n_1]} \cdots (N_\xi(A_r))^{[n_r]}] \tag{5.10}$$

where again $n_i \in \mathbb{N}$ for all $i = 1, \cdots, r$ and $n_1 + \cdots + n_r = n$.

Moment measures and the Janossy measures are related. In particular, we can pass from the Janossy measures to the moment measures using

$$M_{[n]}(A_1^{(n_1)} \times \cdots \times A_r^{(n_r)}) = \sum_{s=0}^{\infty} \frac{J_{n+s}(A_1^{(n_1)} \times \cdots \times A_r^{(n_r)} \times \mathsf{X}^{(s)})}{s!}. \tag{5.11}$$

This relation can be inverted provided that all the moments exist (i.e. $M_{[n]}(\mathsf{X}^{(n)}) < \infty$ for any $n \in \mathbb{N}$), getting

$$J_n(A_1^{(n_1)} \times \cdots \times A_r^{(n_r)}) = \sum_{s=0}^{\infty} (-1)^s \frac{M_{[n+s]}(A_1^{(n_1)} \times \cdots \times A_r^{(n_r)} \times \mathsf{X}^{(s)})}{s!}. \quad (5.12)$$

Also in this case it is not necessary that the sets $A_1, \cdots, A_r$ form a partition of X. If the point process is not finite, these last two equations may still be used by replacing X with some subset $\Lambda \subset \mathsf{X}$ and by using the local Janossy measures on Λ. Also in this case these measures, M_n or $M_{[n]}$, may admit densities with respect to some measure μ on X.

Definition 34. *Consider a point process ξ and a measure μ, both defined on X. If, given $A_1, \cdots, A_r \in \mathscr{B}(\mathsf{X})$ disjoint subsets of X and $n_i \in \mathbb{N}$ for all $i = 1, \cdots, r$ such that $n_1 + \cdots + n_r = n$, we can write*

$$M_{[n]}(A_1^{(n_1)} \times \cdots \times A_r^{(n_r)}) = \int_{A_1^{(n_1)} \times \cdots \times A_r^{(n_r)}} \rho_n(x_1, \ldots, x_n) \mu(dx_1) \cdots \mu(dx_n)$$

the function $\rho(x_1, \ldots, x_n)$ is said n-th correlation function *of the process ξ.*

It is not difficult to see that the relations (5.11) and (5.12) can be used to relate these densities with the Janossy densities and viceversa. Also in this case we have an interpretation of the quantity $\rho(x_1, \cdots, x_n)\mu(dx_1) \cdots \mu(dx_n)$: it represents *the probability to find at least n points of the process ξ, one in each of the n distinct infinitesimal regions* $(x_i, x_i + dx_i)$. In contrast with the Janossy densities, in this case the number of points in the intervals is not fixed (for $j_n(x_1, \cdots, x_n)$ the number is *exactly* n). Note that (5.12) and (5.11) can be used to pass from Janossy to moment densities and viceversa.

5.1.2 Marked point processes

A very useful extension of the notion of point process is the one of *marked point process*, whose main features will be briefly presented here.

There can be situations where the point process is not the principal object that one wants to analyze, like for example, when the point process is just a component of a more complex model. In such situations, it is sometimes useful to associate to each point x_i of the point process, an additional variable m_i belonging to some set M. Such variables are called *marks* and can be anything: for example they can be a label (e.g. the time at which an event happened), a random variable or a set. The set M containing all marks is said *mark space*. The resulting point process, whose generic points are represented by the couple (x_i, m_i), is said *marked point process*.

Definition 35. *A* marked point process *on X is a point process on $\mathsf{X} \times \mathsf{M}$ having points $\tilde{\xi} = \{(x_1, m_1), (x_2, m_2), \cdots\}$, such that $N_g(A) := N_{\tilde{\xi}}(A \times \mathsf{M}) < \infty$ for*

any $A \in \mathscr{B}(\mathsf{X})$. *The point process* ξ_g *defined with the measures* $\{N_g(A)\}_{A\in\mathscr{B}(\mathsf{X})}$ *is said* ground process.

Not all point processes on the product space $\mathsf{X} \times \mathsf{M}$ are marked point processes, but only those for which the ground process is still a point process. A rather simple case when this is always possible is when the mark space is a finite set, i.e. $\mathsf{M} := \{1, \cdots, k\}$ for some $k \in \mathbb{N}$. The marked point process in this case is said *multivariate*, and the finiteness condition of the marked point process is always satisfied. In fact, we have

$$N_g(A) = N_{\tilde{\xi}}(A \times \{1, \cdots, k\}) = \sum_{i=1}^{m} N_{\tilde{\xi}}(A \times \{i\}), \tag{5.13}$$

which follows from the additivity of counting measures between disjoint sets (note that $(A \times \{i\}) \cap (A \times \{j\}) = \{\varnothing\}$ for $i \neq j$). Since $N_{\tilde{\xi}}(A \times \{i\}) < \infty$ for all $A \in \mathscr{B}(\mathsf{X})$ then also $N_g(A) < \infty$ for all $A \in \mathscr{B}(\mathsf{X})$, showing that if the space of marks is a finite set, then any point process on $\mathsf{X} \times \mathsf{M}$ is marked. The counting measures $N_i(A) := N_{\tilde{\xi}}(A \times \{i\})$ define point processes on $\mathsf{X} \times \{i\}$, for any $i \in \mathsf{M}$, which are sometimes called *component processes* of $\tilde{\xi}$.

5.1.3 Determinantal point processes

Here we want to describe an interesting class of point processes: the determinantal point processes. Our interest in this particular class of processes is mainly due to the fact that the whole statistical properties are determined by the kernel of a (locally) trace-class operator over a separable infinite-dimensional Hilbert space. The main references are [63, 64, 65, 66, 67, 68].

From now on, we assume $\mathsf{X} = \mathbb{R}^d$ for simplicity. Let $\hat{K}$ be an operator acting on $L_2(\mathbb{R}^d)$ such that

As1) $\hat{K} \in \mathcal{B}_1^{loc}(L_2(\mathbb{R}^d)) \cap \mathcal{B}_2(L_2(\mathbb{R}^d)$, namely $\hat{K}$ admits kernel $K(x, y)$ and

$$Tr(\hat{P}_\Lambda \hat{K} \hat{P}_\Lambda) < \infty$$

where $\Lambda \subset \mathbb{R}^d$ is compact and $\hat{P}_\Lambda : L_2(\mathbb{R}^d) \to L_2(\Lambda)$ is a projector;

As2) $\hat{\mathbb{O}} \leqslant \hat{K} < \hat{\mathbb{I}}$, namely the spectrum of $\hat{K}$ is in $[0, 1)$;

As3) $\hat{K}$ is a self-adjoint operator, which implies that the associated kernel $K(x, y) : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{C}$ is hermitian, namely such that

$$K(x, y) = [K(y, x)]^*$$

for any $x, y \in \mathbb{R}^d$.

Given such operator $\hat{K}$, consider its *local version* on $\Lambda \subset \mathbb{R}^d$, i.e. $\hat{K}_\Lambda := \hat{P}_\Lambda \hat{K} \hat{P}_\Lambda$. By the Mercer theorem (see Th. 4.27 in [3]) the kernel associated to $\hat{K}_\Lambda$ can be written as

$$K_\Lambda(x, y) = \sum_{\mu_\Lambda} \mu_\Lambda \varphi_{\mu_\Lambda}(x) \varphi^*_{\mu_\Lambda}(y) \tag{5.14}$$

where $\mu_\Lambda \in \mathbb{R}$ are eigenvalues and $\varphi_{\mu_\Lambda} \in L_2(\Lambda)$ are the corresponding eigenvectors of $\hat{K}_\Lambda$. Note that, in any case, $K_\Lambda(x,y) = \chi_\Lambda(x)K(x,y)\chi_\Lambda(y)$. Now we are ready to define the point process we are interested in.

Definition 36. *Let ξ be a locally-finite simple point process on $\mathbb{R}^d$ having, for any $n \in \mathbb{N}$, n-th factorial moment density given by*

$$\rho_n(x_1, \cdots, x_n) = \det([K(x_i, x_j)]_{i,j=1\cdots n}),$$

where $x_1, \cdots, x_n \in \mathbb{R}^d$ and $K(x,y)$ is the kernel of an operator $\hat{K}$ on $L_2(\mathbb{R}^d)$ fulfilling As1) − As3). The point process ξ is said determinantal point process (DPP) *on $\mathbb{R}^d$.*

In the above definition, the writing $[K(x_i, x_j)]_{i,j=1\cdots n}$ is a short hand notation for the $n \times n$ matrix whose (i,j)-th element is $K(x_i, x_j)$. By construction, in a DPP all the moment densities are well defined. This implies that all the (local) Janossy densities exist, are well defined and can be found using (5.11). Following [66], we derive the Janossy density in a different way. Given $\hat{K}$ in $L_2(\mathbb{R}^d)$, operator whose kernel define a DPP on $\mathbb{R}^d$, let us define

$$\hat{J}_\Lambda := (\hat{\mathbb{I}} - \hat{K}_\Lambda)^{-1}\hat{K}_\Lambda.$$

This operator is usually called *local interaction operator*. It can be proved that $\hat{J}_\Lambda \in \mathcal{B}_1(L_2(\mathbb{R}^d))$, hence it admits kernel, denoted by the symbol $\mathcal{J}_\Lambda(x,y)$, and in particular from (5.14), we can write that

$$\mathcal{J}_\Lambda(x,y) = \sum_{\mu_\Lambda} \frac{\mu_\Lambda}{1-\mu_\Lambda} \varphi_{\mu_\Lambda}(x)\varphi^*_{\mu_\Lambda}(y).$$

Given a configuation of the DPP, say $\{x_1, \cdots, x_n\}$, define the function

$$\eta_\Lambda(x_1, \cdots, x_n) := \det([\mathcal{J}_\Lambda(x_i, x_j)]_{i,j=1,\cdots,n}). \tag{5.15}$$

Then the Janossy densities can be computed using the following result [66, 68].

Proposition 2. *Let $\hat{K}$ be an operator on $L_2(\mathbb{R}^d)$ fulfilling As1) − As3) and $K(x,y)$ its associated kernel. Let ξ be a DPP on $\mathbb{R}^d$ with kernel $K(x,y)$. Then for all compact subset $\Lambda \subset \mathbb{R}^d$ and $n \in \mathbb{N}/\{0\}$, the local Janossy density of the process ξ is*

$$j_n(x_1, \cdots, x_n|\Lambda) = \det(\hat{\mathbb{I}} - \hat{K}_\Lambda)\eta_\Lambda(x_1, \cdots, x_n)$$

while $j_0(\Lambda) = \det(\hat{\mathbb{I}} - \hat{K}_\Lambda)$.

In the above proposition, $\det(\hat{\mathbb{I}} - \hat{K}_\Lambda)$ must be interpreted as Fredholm determinant (see [69] for mathematical details). Recalling the discussion done in the section 5.1.1, the kernel $K(x,y)$ of a DPP ξ contains all the information needed to completely characterise the process. Indeed, it allows to compute all

the fidis of ξ, but we may also deduce other properties. Moreover, from the 1st moment measure, for any $A \subset \mathbb{R}^d$ we have that

$$\begin{aligned}\mathbb{E}[N_\xi(A)] &= \int_A K(x,x)dx \\ &= \int_{\mathbb{R}^d} \chi_A(x)K(x,x)dx \\ &= Tr(\hat{P}_A\hat{K}) = Tr(\hat{K}_A)\end{aligned}$$

where $\hat{P}_A : L_2(\mathbb{R}^d) \to L_2(A)$ is an orthogonal projector. Setting $A = \mathbb{R}^d$, we can see that the expected number of points is nothing but that trace of $\hat{K}$. This suggests that directly from $\hat{K}$, we can obtain information on the finiteness of the DPP. This theorem formalise exactly this idea [66, 67].

Theorem 14. *Let $\hat{K}$ be an operator on $L_2(\mathbb{R}^d)$ fulfilling $As1) - As3)$ and let ξ the DPP generated by the associated kernel. We have*

i) If $Tr(\hat{K}) < \infty$, i.e. $\hat{K} \in \mathcal{B}_1(L_2(\mathbb{R}^d))$ then $P[N_\xi(\mathbb{R}^d) < \infty] = 1$ which means that the point process is finite with probability 1, while when $Tr(\hat{K}) = \infty$ the point process is only locally-finite, namely $P[N_\xi(\mathbb{R}^d) < \infty] = 0$;

ii) $P[N_\xi(\mathbb{R}^d) \leqslant m] = 1$ for some $m \in \mathbb{N}/\{0\}$ if and only if $\hat{K}$ has finite rank and $Rank(\hat{K}) \leqslant m$;

iii) The number of points of the process is exactly $m \in \mathbb{N}/\{0\}$ if and only if $\hat{K}$ is an orthogonal projector having $Rank(\hat{K}) = m$.

A DPP whose kernel is an orthogonal projector is said *orthogonal*. Finally, we conclude by explaining what happens to a DPP when we remove a point. For a point process, this operation is called *thinning*. DPPs are closed under thinning in the sense that the new point process obtained after this operation is still a DPP. More formally, given a DPP ξ on $\mathbb{R}^d$ we may form a new point process as follows. Given a point $z \in \mathbb{R}^d$, we condition the DPP on the event $z \in \xi$ and then remove this point. With this procedure a new point process is obtained, say $\xi/\{z\}$, which is the DPP ξ without the point z. For the DPP obtained with this procedure the following theorem holds [68, 70].

Theorem 15. *Let ξ be a DPP with kernel $K(x,y)$ and $z \in \mathbb{R}^n$ such that $0 < K(z,z) < \infty$. Then the point process obtained by removing z from ξ as described above is again a DPP with kernel $\tilde{K}(x,y)$ given by*

$$\tilde{K}(x,y) = K(x,y) - \frac{K(x,z)K(z,y)}{K(z,z)}.$$

This theorem can be extended also to points $z \in \mathbb{R}^d$ for which $K(z,z) = 0$ or $K(z,z) = \infty$, if we can approximate z with points s all having $0 < K(s,s) < \infty$. Also in this case the process $\xi/\{z\}$ is still a DPP with kernel given by

$$\tilde{K}(x,y) = K(x,y) - \lim_{s \to z} \frac{K(x,s)K(s,y)}{K(s,s)}.$$

5.1.4 DPP diffusion

In the description of the DPP done in section 5.1.3, we did not mention time at any level. In this sense the DPP (as any point process) can be considered as a "spatial process", i.e. a collection of random variables parametrised by space (see definition 26). In ordinary stochastic processes, the parameter is typically assumed to be the time and time-evolution is stochastic. Here we want to describe, at a very qualitative level, how it is possible to implement a stochastic time-evolution for a DPP. In particular, we will describe how to implement a *diffusive* time-evolution: this will be done by using suitable *Dirichlet forms*.

Let us start with an analogy. The diffusion process we are going to describe can be imagined as follow. A DPP can be thought as a classical gas of fermions[a] and the diffusion process we want to describe is the diffusion of such a gas. Note that, because of the fermionic character, the gas cannot be thought as non-interacting: even by neglecting all the possible interactions between particles two fermions cannot be found in the same place, i.e. there is an exchange-correlation interaction to be taken into account. Such interaction prevents the formation of clusters in the process: in this sense, it is a repulsive interaction between particles. A possible approach to describe this diffusion process can be the following. Given a DPP configuration ξ one marks each point with a label $t \in \mathbb{R}^+$. The t-components of marked point process thus obtained, $\xi_t = \xi \times \{t\}$, is made of couples (x_i, t), with $x_i \in \xi$ and $t \in \mathbb{R}^+$, representing the spatial and temporal location of the point. Then we can imagine that each point in ξ_t evolves in time according to some stochastic differential equation (SDE), one for each point. Since we have an infinite number of points (in general), we have an infinite number of SDEs. In addition, to model the repulsiveness due to the fermionic character of the gas, the SDEs describing the trajectories of each particle should have a drift term which depends on the locations of all the other particles. Thus, if we want to find the configuration of this gas (the position of each particle) at a given time $t \in \mathbb{R}^+$ (knowing its configuration at some time $t_0 < t$), we should solve an infinite system of coupled SDEs. It is not easy to do so, however this problem can be tackled from a different perspective using the so called Dirichlet forms, as explained in [71]. The interested reader, in appendix find a very brief review on Dirichlet forms and its relation with Markov processes.

The Dirichlet forms approach for the diffusion of a DPP is used in [72] for a particular class of DPP, those having translational invariant kernels. The application of the same method to more general cases, can be found in [66] and [73]: here we report the main steps explained in these works, for the construction of such a diffusion. By definition of point process, a DPP is a random variable taking values on the space $(\mathcal{N}_{\mathbb{R}^d}, \mathscr{B}(\mathcal{N}_{\mathbb{R}^d}))$. Let μ_{dpp} be the probability distribution for such a process. If ξ is a configuration of the DPP, take its restriction

[a] The word "classical" in this context should be interpreted as follows: a gas of *classical* fermion is a gas of particles having n-point correlation functions equal to the one of a quantum gas of fermions. However, no quantum description of these particles is needed for the definition of such a gas [63, 67].

to Λ, compact subset of $\mathbb{R}^d$, i.e. $\xi_\Lambda := \xi \cap \Lambda$. Then define $\mathcal{N}_\Lambda^f$ as the subset of $\mathcal{N}_{\mathbb{R}^d}$

$$\mathcal{N}_\Lambda^f := \{N' \in \mathcal{N}_{\mathbb{R}^d} | N'(\Lambda) < \infty\},$$

i.e. the set of all the *finite* simple counting measures on Λ, which is the set on which ξ_Λ takes values. To the restrincted DPP ξ_Λ, one can associate an Hilbert space $L_2(\mathcal{N}_\Lambda^f, \mu_{dpp})$. Consider the set of real valued functions $F : \mathcal{N}_\Lambda^f \to \mathbb{R}$, which can be written as

$$F(\xi_\Lambda) = f_0 \chi_{\{N_\xi(\Lambda)=0\}} + \sum_{i=1}^{n} \chi_{\{N_\xi(D)=i\}} f_i(x_1, \cdots, x_i)$$

where $x_1, \cdots, x_n \in \xi_\Lambda$, $n \geq 1$ is an integer, f_0 a constant and f_i are smooth symmetric functions. Call this set $\mathcal{S}_\Lambda$. It can be proved that $\mathcal{S}_\Lambda$ is dense in $L_2(\mathcal{N}_\Lambda^f, \mu_{dpp})$, hence $F(\xi_\Lambda)$ can seen as the typical element of $L_2(\mathcal{N}_\Lambda^f, \mu_{dpp})$. On $\mathcal{S}_\Lambda$, one can define a gradient operator

$$\nabla_x^{\mathcal{N}_\Lambda^f} F(\xi) := \sum_{k=1}^{n} \chi_{\{N_\xi(\Lambda)=k\}} \sum_{y \in \xi} \chi_{\{y=x\}} \nabla_x f_k(x_1, \cdots, x_k),$$

where ∇_x is the ordinary gradient in $\mathbb{R}^d$. This last definition allows to introduce the following symmetric real bilinear form on $\mathcal{S}(\Lambda)$

$$\varepsilon_\Lambda(F, G) = \mathbb{E}\Big[\sum_{y \in \xi_\Lambda} \nabla_y^{\mathcal{N}_\Lambda^f} F(\xi_\Lambda) \cdot \nabla_y^{\mathcal{N}_\Lambda^f} G(\xi_\Lambda) \Big],$$

where $\cdot$ denotes the ordinary scalar product in $\mathbb{R}^d$. This form can be extended to the whole $L_2(\mathcal{N}_\Lambda^f, \mu_{dpp})$ and, under suitable conditions, can be rewritten as (see lemma 4.2, [73])

$$\mathbb{E}\Big[\sum_{y \in \xi_\Lambda} \nabla_y^{\mathcal{N}_\Lambda^f} F(\xi_\Lambda) \cdot \nabla_y^{\mathcal{N}_\Lambda^f} G(\xi_\Lambda) \Big] = \mathbb{E}[F(\xi_\Lambda) \hat{H}_\Lambda G(\xi_\Lambda)],$$

where $\hat{H}_\Lambda$ is a symmetric non-negative definite operator acting on $L_2(\mathcal{N}_\Lambda^f, \mu_{dpp})$. Considering the closure of such operator, one obtains the from

$$\bar{\varepsilon}_\Lambda(F, G) = \mathbb{E}[F(\xi_\Lambda) \overline{\hat{H}_\Lambda} G(\xi_\Lambda)]. \tag{5.16}$$

Then $(\bar{\varepsilon}_\Lambda, \mathrm{Dom}(\overline{\hat{H}_\Lambda^{1/2}}))$ is a Dirichlet form on $L_2(\mathcal{N}_\Lambda^f, \mu_{dpp})$ (see Th. 4.1, [73]). This Dirichlet form, which is defined on the space $\mathcal{N}_\Lambda^f$ on which the DPP ξ_Λ takes values, can be used to construct a stochastic time-evolution for ξ_Λ which is a diffusion. In particular the following theorem holds (Th. 5.1, [73]).

Theorem 16. *Let ξ be a DPP on $\mathbb{R}^d$ with kernel $K(x, y)$ and probability distribution μ_{dpp}. Let $\hat{K}$ the locally trace-class integrable operator associated to this kernel from which one can construct the function $\eta_\Lambda(x_1, \cdots, x_n)$ as in* (5.15)*, where Λ is a compact subset of $\mathbb{R}^d$. Assuming that*

a) $(x_1, \cdots, x_n) \mapsto \eta_\Lambda(x_1, \cdots, x_n)$ *is continuously differentiable on* Λ^n*;*

b) for any $n \in \mathbb{N}$, $1 \leqslant i, j \leqslant n$ *and* $1 \leqslant h, k \leqslant d$,

$$\int_{\Lambda^n} \left| \frac{\partial_{x_i^{(h)}} \eta_\Lambda(x_1, \cdots, x_n) \partial_{x_j^{(k)}} \eta_\Lambda(x_1, \cdots, x_n)}{\eta_\Lambda(x_1, \cdots, x_n)} \right| \cdot \chi_{\{\eta_\Lambda(x_1, \cdots, x_n) > 0\}}(x_1, \cdots, x_n) dx_1 \cdots dx_n < \infty,$$

where $\partial_{x_i^{(h)}}$ *is the derivative with respect to the h-th component of* $x_i \in \xi$*;*

Then there exists a Markov process $\{\xi_t\}_{t \in \mathbb{R}^+}$ *on* $\mathcal{N}_\Lambda^f$ *such that*

i) $\{\xi_t\}_{t \in \mathbb{R}^+}$ *is a diffusion (i.e. the trajectories are continuous with probability one);*

ii) The transition probabilities of $\{\xi_t\}_{t \in \mathbb{R}^+}$ *can be calculated using the semi-group* $\hat{T}_t$ *associated to the Dirichlet from* $(\bar{\varepsilon}_\lambda, Dom(\overline{\hat{H}_\Lambda^{1/2}}))$ *defined in* (5.16)*;*

iii) $\{\xi_t\}_{t \in \mathbb{R}^+}$ *is unique up to* μ_{dpp}*-null sets;*

iv) $\{\xi_t\}_{t \in \mathbb{R}^+}$ *has* μ_{dpp} *invariant measure.*

Let us explain a bit further the content of this theorem. Given a DPP ξ and taking its restriction on some compact subset Λ, one can construct a diffusion process from the Dirichlet form (5.16). If $\hat{H}_\Lambda$ is the generator of the Dirichlet form and $A \in \mathscr{B}(\mathcal{N}_{\mathbb{R}^d})$ is an event at time t_0 with probability $\mu_{dpp}(A, t_0)$, then the probability of the event A at time $t + t_0$, $p(A, t + t_0)$ can be computed as

$$p(A, t + t_0) = e^{-t\hat{H}_\Lambda} \mu_{dpp}(A, t_0).$$

The theorem ensures that such diffusion process is unique and also that the stochastic process at time $t + t_0$ is still determinantal, namely

$$p(A, t + t_0) = \mu_{dpp}(A, t_0).$$

Hence, such time-evolution preserves the determinantality of the process. Since the determinantal point process is simple, the trajectories of the single points of ξ do not collide during this evolution. In fact, we have the following theorem (see Th. 5.3, [73])

Theorem 17. *Under the assumptions of theorem 16, for* $d \geqslant 2$ *the diffusion is non-colliding.*

Chapter 6

A possible geometry for quantum mechanics

In this chapter we propose a model of space which is capable to justify the assumption *i*) done at the end of chapter 4. Imposing the correct symmetry group, one can justify also the assumption *ii*). To do that, we reconsider model B explained in chapter 4 under a different point of view which makes easier to replace the space process with a generic point process. Using a determinantal point process as space process, in section 6.2 assumption *i*) will be justified.

6.1 3-D generalization of model B and point process structure

Here, we describe model B using the general theory of point processes explained in section 5.1. This enables us to generalize model B from 1-D to 3-D in a quite straightforward way.

6.1.1 Model B: underlying point process structure and generalization to 3D case

In model B, two are the main ingredients: the space process $\mathbb{S}^B_t$ and the particle process $\mathscr{P}_t = (X_t, V_t(t'))$. Let us start from the space process $\mathbb{S}^B_t$. Since $\mathbb{S}^B_t$ is a random distribution of points over the real line, we should be able to describe it with the theory of point processes presented in the previous section. Suppose we have a collection of M independent Wiener processes taking values on $\mathbb{R}$. At a given time t, the subset of $\mathbb{R}$ formed by all positions of each Wiener process defines a point process. Call ξ_t such point process, without assuming *distinguishability* we can easily conclude that the fidis density of such process

are given by

$$\rho_{\xi_t}(x_1, \cdots, x_M) = \frac{1}{M!} \sum_{\sigma \in P_M} \prod_{i=1}^{M} \rho_{W_t^{(\sigma(i))}}(x_i) \tag{6.1}$$

where σ is a permutation of $\{1, \cdots, M\}$, P_M is the set of all the permutations over this set, and $\rho_{W_t^{(i)}}(x)$ is the probability density of the i-th 1-D Wiener process at time t. One may note that, if we assume that the Wiener processes are identically distributed, we recover the probability distribution used in model B for $\mathbb{S}_t^B$ in this particular case. However, without this assumption, this formula does not coincide with the one used in model B: the difference lies in the fact that here we are not assuming the distinguishability of the Wiener process. We will come back on this fact soon, but for the moment we observe that, at this level, we are describing *only* the random pattern formed by these points. Given the expression above, one can immediately realise that the M-th Janossy density is

$$j_M(x_1, \cdots, x_M; t) = \sum_{\sigma \in P_M} \prod_{i=1}^{M} \rho_{W_t^{(\sigma(i))}}(x_i).$$

This holds for any time $t \in \mathbb{R}^+$ hence, the time evolution of the Wiener processes, can be used to obtain the fidis and Janossy densities at any time. Now consider the particle process of model B, and in particular the position random variable X_t (see (22))

$$X_t(\omega) := \pi_{I_t(\omega)}(\mathbb{S}_t^B(\omega)) - W_t^{i_0}(\omega). \tag{6.2}$$

The position random variable selects one of the points of $\mathbb{S}_t^B(\omega)$, and takes the value of the difference between the point selected and the origin of the reference frame. To express this selection procedure with this writing, the point of the space process must be distinguishable: only in this way one can choose the i-th point. Thus, in order to use the definition above starting for the point process ξ_t with fidis given by (6.1), we need to distinguish their points. This can be done introducing by an additional label, say i, attached to each point and considering the marked point process obtained in this way. More precisely, consider the point process ξ_t, call $x_t^{(i)}$ its i-th point according to some arbitrary ordering, and consider the set of points

$$\tilde{\xi}_t = \{(x_t^{(1)}, 1), \cdots, (x_t^{(M)}, M)\}.$$

It is a multivariate marked point process on $\mathbb{R}$ (by construction, the ground process is exactly ξ_t). Using the marked point process $\tilde{\xi}_t$ we can really select a point of ξ_t, thus the definition of X_t used makes sense. More precisely we can say that the point process of model B can be described by using the marked point process $\tilde{\xi}_t$: the fidi densities of $\tilde{\xi}_t$ corresponds to the probability densities of $\mathbb{S}_t^B$. In this sense $\mathbb{S}_t^B$ is a particular case of the point process described here. Using (5.13) we can write

$$N_{\xi_t}(A) = \sum_{i=1}^{M} N_{\tilde{\xi}_t}(A \times \{i\})$$

where $A \in \mathscr{B}(\mathbb{R})$. Each component $N_{\tilde{\xi}_t}(A \times \{i\})$ is a point process consisting of a single point. The selection procedure, due to the random variable I_t, can be represented considering the I_t-th component of $\tilde{\xi}_t$, i.e. $N_{\tilde{\xi}_t}(A \times \{I_t\})$. Defining $N_{X_t}(A) := N_{\tilde{\xi}_t}(A \times \{I_t\})$, the point process $\xi_{X_t} := \{N_{X_t}(A)\}_{A \in \mathscr{B}(\mathbb{R}^3)}$ represents the position of the particle at time t. We can use the counting measure $N_{X_t}(A)$ to determine where the particle is: $N_{X_t}(A) \neq 0$ only for those A containing the point $X_t = \pi_{I_t}(\tilde{\xi}_t)$, and this can be used to determine the position of the particle. Compared with (6.2), we can see that the origin is assumed to be in 0 at time t. This can be done without loss of generality, since to have a non-zero origin one can always perform a suitable translation. From now on we consider the origin always in 0. Also the distribution of X_t can be derived from such a measure. In order to see it, let us define the intensity measure

$$\mu_{X_t}(A) := M_1(A) = \mathbb{E}_{\xi_{X_t}}[N_{X_t}(A)]. \tag{6.3}$$

Since ξ_{X_t} consists of a single point, clearly $N_{X_t}(\mathbb{R}^3) = N_{\tilde{\xi}_t}(\mathbb{R}^3 \times \{I_t\}) = 1$, which implies that $\mu_{X_t}(\mathbb{R}^3) = 1$, i.e. it is a probability measure. This is the probability distribution of the random variable X_t, in fact $\mu_{X_t}(A)$ is the expected number of times we find the particle in the set $A \in \mathscr{B}(\mathbb{R})$, i.e. its probability to be found in A. Note that in (6.3) the expectation is taken with respect to the point process ξ_{X_t}, which depends both on the ground process ξ_t (thus on the marked point process $\tilde{\xi}_t$) and the selection random variable I_t. Hence we can write

$$\begin{aligned}
\mu_{X_t}(A) &= \int_A \mu_{X_t}(dx) \\
&= \int_A \int_{\mathcal{N}_{\mathbb{R}^d}} \nu_{X_t|\xi_t=\zeta}(dx)\nu_{\xi_t}(d\zeta) \\
&= \int_A \int_{\tilde{\mathcal{N}}_{\mathbb{R}^d}} \mu_{X_t|\tilde{\xi}_t=\mathbf{S}_t}(dx)\mu_{\tilde{\xi}_t}(d\mathbf{S}_t)
\end{aligned}$$

Thus the position random variable of model B can be described by using suitable components of the multivariate marked point process constructed from a suitable point process.

Before going on, let us discuss the relation between the indistinguishability of the points, the need of a labeling, and the arbitrariness of this labeling. We have seen that the point process ξ_t describes the random pattern of the space process but, in order to define the position random variable, we need to introduce an arbitrary labeling among the points (i.e. we need to consider the marked point process $\tilde{\xi}_t$). The arbitrariness of this labeling, needed to define $\tilde{\xi}_t$, may sound strange. However, this situation is a quite common situation. In fact, suppose we have a point $P \in \mathbb{R}$. In order to describe it we need a reference frame, namely, we need to specify an origin O and a way to measure the distance between O and P. The choice of these two objects is completely arbitrary. The relation between ξ_t, $\tilde{\xi}_t$ and the labeling is similar to the one just described. In particular, to define this ordering one may choose an arbitrary

point in ξ_t, the origin, and use an arbitrary distance function to label all the other points: the label i of a point $x \in \xi_t$ contains the information needed to describe the point in the correct way with respect to the chosen origin (which is automatically in 0). Thus changing the distance function (which is arbitrary) changes the labeling. In this sense ξ_t can be thought as $\mathbb{R}$ without specifying a distance function, while $\tilde{\xi}_t$ can be thought as $\mathbb{R}$ *equipped with a particular distance function*. In this sense, the need and the arbitrariness of the labeling introduced above should not be considered as a strange thing.

The generalization to the 3-D case is straightforward: one can replace the densities $\rho_{W_t^{(i)}}(x_i)$ associated to the 1-D Wiener process at time t with the corresponding densities associated to the 3-D Wiener process. The M-th Janossy densities (and so the fidi densities) obtained in this way are associated to a point process in $\mathbb{R}^3$, and this holds at any time t. In the 3-D case, the particle position is a 3-vector, namely

$$X_t := (X_t^1, X_t^2, X_t^3). \tag{6.4}$$

As for the 1-D case, equation (6.3) gives us the distribution $\mu_{X_t}(A)$ for some $A \in \mathscr{B}(\mathbb{R}^3)$. The velocity random variable can be obtained by using (6.4) in the definition given for model B. Thus, the velocity random variable is the 3-vector

$$V_t(t') := \frac{X_{t'} - X_t}{t' - t}. \tag{6.5}$$

Then one can proceed as in chapter 4 to prove that, after the removal of the space process, when we represent these random variables on a common Hilbert space, the operators associated to each component do not commute, i.e.

$$[\hat{X}_t^i, \hat{V}_t^i(t')] \neq 0, \tag{6.6}$$

where $i = 1, 2, 3$. Since we want to compare it with ordinary quantum mechanics, we also need to check if the components of $\hat{X}_t$ commute among them. If it is not so, there is no hope to re-obtain ordinary non-relativistic quantum mechanics from any modification of this model. We recall that the non-commutativity is deduced by means of an entropic uncertainty relation, obtained after conditioning on the point process. The strategy used to derive the bound between the sum of $H_{\mathbf{S}_t}(X_t)$ and $H_{\mathbf{S}_t}(V_t(t'))$, is based on the fact that the transition probabilities depend on the point process at different times (say t and t'). Indeed, we condition the position and velocity random variables to a given realization $\mathbf{S}_t$ but $V_t(t')$ depends also on $\mathbf{S}_{t'}$ (since it depends on X_t and $X_{t'}$). Thus the conditioning on $\mathbf{S}_t$ is not sufficient to have a delta-like probability distribution for $V_t(t')$ when this happens for X_t: the minimal uncertainty on $V_t(t')$ is bounded by the uncertainty of $\mathbb{S}_{t'}^B$ given $\mathbf{S}_t$, and this is the origin of the bound in the sum of the two entropies. If we apply this idea to the 3-D generalization, we cannot derive any bound between the components of X_t. In fact given X_t^i and X_t^j, with $i \neq j$, their probability distributions can be derived as marginal of μ_{X_t} and so we can always define a conditional probability like $\mu_{X_t^i | X_t^j}$. However

$\mu_{X_t^i|X_t^j}$ may depend only to a single $\mathbf{S}_t$, hence after conditioning on such point process configuration, we can always shrink the transition probabilities $\mu_{X_t^i|X_t^j}$, to delta-like measures, which implies that $H_{\mathbf{S}_t}(X_t^i) + H_{\mathbf{S}_t}(X_t^j) \geqslant 0$ (after a suitable discretization). Thus

$$[\hat{X}_t^i, \hat{X}_t^j] = 0, \tag{6.7}$$

for all $i, j = 1, 2, 3$. However, using this commutation relation together with (6.4) and (6.5), we have to conclude that

$$[\hat{X}_t^i, \hat{V}_t^j(t')] = \frac{1}{t'-t}[\hat{X}_t^i, \hat{X}_{t'}^j], \tag{6.8}$$

(note the time indices) and

$$[\hat{V}_t^i(t'), \hat{V}_t^j(t')] = -\frac{1}{(t'-t)^2}\Big\{[\hat{X}_t^i, \hat{X}_{t'}^j] + [\hat{X}_{t'}^i, \hat{X}_t^j]\Big\}, \tag{6.9}$$

for all $i \neq j$. Notice that in general $[\hat{X}_t^i, \hat{X}_{t'}^j] \neq 0$, and $[\hat{X}_{t'}^i, \hat{X}_t^j] \neq 0$. This concludes the 3-D generalisation of model B presented in chapter 4.

6.1.2 The role of separability in the 3-D generalization of model B

The commutators (6.6) and (6.7) suggest that the operator representing the position random variable in model B, $\hat{X}_t$, resemble the ordinary position operator of non-relativistic quantum mechanics (such operator will be labeled by $\hat{Q}_t$, to avoid ambiguities). In section 4.1.5 and section 4.2.4, the construction of the Hilbert space representation of the position operator was explained in detail. Since in the 3-D generalization we replace the position random variable with a vector-valued random variable, the Hilbert space on which the operator representing the position random variable acts, is

$$\mathcal{H}(X_t) = \mathcal{H}(X_t^{(1)}) \otimes \mathcal{H}(X_t^{(2)}) \otimes \mathcal{H}(X_t^{(3)}),$$

where $\mathcal{H}(X_t^{(i)})$ is the Hilbert space on which the operator representing the i-th component of X_t is defined. Clearly $\mathcal{H}(X_t)$ is infinite-dimensional and non-separable in general, because each $\mathcal{H}(X_t^{(i)})$ is not. We recall that on this Hilbert space we can describe also the operator $\hat{V}_t(t')$, but it is not diagonalizable on the same basis of $\hat{X}_t$ (which is diagonal by construction in $\mathcal{H}(X_t)$, see section 4.2.4). The importance of the separability was already recognized in section 4.3 where we showed that, assuming the separability of the infinite dimensional Hilbert space on which the operators are represented, we can relate the velocity random variable of the model to the ordinary momentum operator of non-relativistic quantum mechanics $\hat{P}_t$. The same argument can be used also in the 3-D generalization: assuming the separability of $\mathcal{H}(X_t)$, and the existence of a suitable operator $\hat{H}$, by the Ehrenfest theorem we can prove that

$$\lim_{t' \to t} \hat{V}_t^{(i)}(t') = \frac{1}{m}\hat{P}_t^i$$

in weak sense for any $i = 1, 2, 3$, where m is a constant with the dimension of a mass, and $\hat{P}_t^i$ is the i-th component of $\hat{P}_t$. Summarising, if we can justify separability then a direct correspondence between the particle in model B and a quantum particle seems to be possible.

6.2 Model C

In this section, we modify model B in order derive the right commutation relations between position and momentum and establish a correspondence between this new model and non-relativistic quantum mechanics. The model we will obtain will be called model C. Before starting let us point out that the proposed model does not take into account the spin of a quantum particle: this model was not conceived for such a scope. This does not mean that we cannot describe a particle with spin, but simply that the description is done exactly as in ordinary non-relativistic quantum mechanics (hence this model does not add anything to the ordinary description).

The idea behind model C is the following. In this model, we use as space process a DPP, since it is described by using a separable infinite-dimensional Hilbert space. The analysis done in section 6.1, where we recognize the underlying point process structure of model B, allows to replace the point process used for model B with a DPP in a quite straightforward way. Closure under thinning of the DPP allows to derive the right separable Hilbert space structure for the particle process. Since separability is not enough to establish a correspondence between the velocity operator $\hat{V}_t(t')$ and the quantum momentum operator $\hat{P}_t$, one needs also to introduce a suitable operator $\hat{H}$. This can be done by imposing the correct symmetry group for a non-relativistic system, i.e. the Galilean group.

6.2.1 The space process

We start the description of this model from the space process. At a given time, say $t = 0$, the space process is represented by a DPP on $\mathbb{R}^3$ with locally trace-class operator $\hat{K}$ on $L_2(\mathbb{R}^3)$. The number of points of the DPP is assumed to be infinite which means, by the theorem 14, that $\mathrm{Rank}(\hat{K}) = \infty$. The space process evolves via diffusion, as briefly explained in section 5.1.4. Such a diffusion preserves the determinantality of the process: this means that if at time $t = 0$ the space process is determinantal it remains so for any $t > 0$, as theorem 16 shows.

Definition 37. *At a given time t the* space process of model C*, call it $\mathbb{S}_t^C$, is a DPP on $\mathbb{R}^3$ with locally trace-class operator $\hat{K}$ on $L_2(\mathbb{R}^3)$ having Rank$(\hat{K}) = \infty$. The time evolution of $\mathbb{S}_t^C$, i.e. $t \mapsto \mathbb{S}_t^C$, is a DPP diffusion. The stochastic process $\mathbb{S}^C = \{\mathbb{S}_t^C\}_{t\in\mathbb{R}^+}$ is the space process of model C.*

A possible realisation of the space process $\mathbb{S}^C$ is given in figure 6.1.

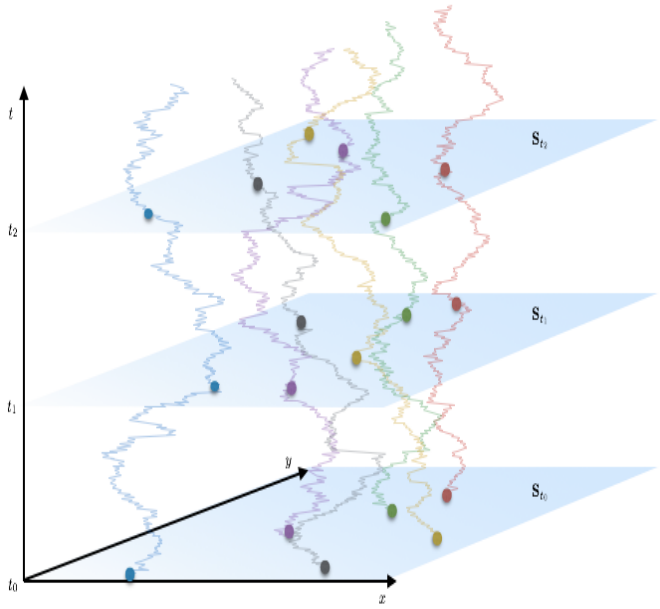

Figure 6.1: A pictorial representation of a realization of the space process of model C. Note that in model C, the point process at different times is a collection of points in a 3-D space, which here is represented by the 2-D planes in light-blue. The colored points belonging to the same plane represent a configuration of the DPP at a given time. The trajectories of the DPP diffusion are also drawn in the back.

6.2.2 The particle process

Let us turn our attention to the particle process at a given time t. We have already seen in section 6.1.1 how we can describe the particle process using a marked point process, whose ground process is the point process of the model. Thus we consider the marked point process

$$\tilde{\mathbb{S}}_t^C := \{(x_t^{(1)}, 1), (x_t^{(2)}, 2), \cdots\},$$

defined exactly as in section 6.1.1. According with the discussion done in that section, this can be considered as the physical space equipped with a particular distance function. By (5.13),

$$N_{\mathbb{S}_t^C}(A) = \sum_{i=0}^{\infty} N_{\tilde{\mathbb{S}}_t^C}(A \times \{i\})$$

which is finite since the ground process is a DPP. This means that, by construction, the sum on the RHS is convergent and so each member of the sum

is a point process too. The point process $\{N_{\tilde{\mathbb{S}}_t^C}(A \times \{i\})\}_{A \in \mathscr{B}(\mathbb{R}^3)}$ has only a single point and is obtained from the initial DPP, $\mathbb{S}_t^C$, by deleting all the points having mark $j \neq i$. Hence, applying repeatedly theorem 15, we can conclude that $\{N_{\tilde{\mathbb{S}}_t^C}(A \times \{i\})\}_{A \in \mathscr{B}(\mathbb{R}^3)}$ is a DPP. Call $\hat{\rho}$ the locally trace-class operator on $L_2(\mathbb{R}^3)$ of this DPP. Since it has a single point, we have to conclude that $\text{Rank}(\hat{\rho}) = 1$ (see theorem 14), which means that the kernel of $\hat{\rho}$ can be written as

$$\rho(x, y) = \psi(x)\psi^*(y),$$

where $\psi(x) \in L_2(\mathbb{R}^3)$. We introduce a selection random variable, i.e. the positive integer valued random variable $I_t : \Omega_I \to \mathbb{N}$, which allows to define the position of the particle at time t as the point described by the DPP $\xi_{X_t} := \{N_{\tilde{\mathbb{S}}_t^C}(A \times \{I_t\})\}_{A \in \mathscr{B}(\mathbb{R}^3)}$.

Definition 38. *Let $\tilde{\mathbb{S}}_t^C$ be the marked point process defined above with ground process $\mathbb{S}_t^C$ and I_t be an $\mathbb{N}$-valued random variable. The* position of the particle at time *t in model C is described by the point process on $\mathbb{R}^3$ defined as*

$$\xi_{X_t} := \{N_{X_t}(A)\}_{A \in \mathscr{B}(\mathbb{R}^3)},$$

where $N_{X_t}(A) := N_{\tilde{\mathbb{S}}_t^C}(A \times \{I_t\})$ for any $A \in \mathscr{B}(\mathbb{R}^3)$.

Let us label by X_t the (single) point of ξ_{X_t}: this is the *position random variable* at time t. Again the origin is assumed to be in 0, without loss of generality.

Definition 39. *Let X_t and $X_{t'}$ be the position random variables at time t and t' obtained from ξ_{X_t} and $\xi_{X_{t'}}$, respectively. The* velocity random variable $V_t(t')$ *is defined as*

$$V_t(t') := \frac{X_{t'} - X_t}{t' - t}.$$

As for model B, $V_t(t')$ should be interpreted as the average velocity of the particle in the time interval $[t, t']$.

As already explained, theorem 15 implies that ξ_{X_t} is a DPP with a rank-1 kernel. This means that the intensity measure of this process, is

$$M_1(A) = \int_A \rho(x, x)dx = \int_A |\psi(x)|^2 dx.$$

Since $\mu_{X_t}(A) := M_1(A)$ is the expected number of times one finds the point of ξ_t in A, we can conclude that X_t has probability density $|\psi(x)|^2$ for some $\psi(x) \in L_2(\mathbb{R}^3)$. However, as in model A and model B, we are not interested in the probability $P[X_t \in A] = \mu_{X_t}(A)$, but rather in the conditional probability with respect to some configuration of the space process. Let $\mathbf{S}_t$ label a possible configuration of the marked point process $\tilde{\mathbb{S}}_t^C$ with fidis density $\mu(\mathbf{S}_t)$, namely we can write

$$\mu(A_1, n_1; \cdots; A_r, n_r) = \int_{A_1^{(n_1)} \times \cdots \times A_r^{(n_r)}} \mu(\mathbf{S}_t) d\mathbf{S}_t.$$

As explained in section 5.1.1, the fidis can be expressed in terms of the Janossy measure, thus the fidis density expressed above can be obtained from the Janossy densities (which always exist for a DPP). We note that once $\mathbf{S}_t$ is fixed, the position random variable can vary only over a discrete set of points, i.e. over the set $\Lambda_{\mathbf{S}_t} := \{x | x \in \mathbf{S}_t\}$ which is a subset of $\mathbb{R}^3$. Thus we can write

$$\mu_{X_t}(A) = \int_A |\psi(x)|^2 dx = \int_{\mathcal{N}_{\mathbb{R}^3}} \int_{A_{\mathbf{S}_t}} |\psi(y, \mathbf{S}_t)|^2 d\mathbf{S}_t dy,$$

where $A_{\mathbf{S}_t}$ is the subset of $\Lambda_{\mathbf{S}_t}$ compatible with the event $\{X_t \in A\}$. As a matter of fact the integral with respect to y is actually a sum of discrete values. As already observed in section 4.3, in order to deal with probability measures absolutely continuous with respect to Lebesgue we need to take the *dense-point limit*. In this limit, the point of space gets denser and denser till the continuum ($\mathbb{R}^3$) is reached. Let us describe more formally how this limit may be taken. The (local) density of the space process $\mathbb{S}_t^C$, can be measured by counting the number of points in an arbitrary set $A \subset \mathbb{R}^3$, namely with $n := N_{\mathbb{S}_t^C}(A)$. The dense point limit is obtained for $n \to \infty$ for any $A \subset \mathbb{R}^3$ (when $n = \infty$ for any $A \subset \mathbb{R}^3$ the point process structure of the space is replaced by the continuum). The density n parametrises also the marked point process $\tilde{\mathbb{S}}_t^C$ (we write $\tilde{\mathbb{S}}_t^C[n]$). The function $y \mapsto \psi(y, \mathbf{S}_t)$ is in $L_2(\Lambda_{\mathbf{S}_t[n]})$ for any n finite. Since $n = \infty$ corresponds to the continuum $\mathbb{R}^3$ and $L_2(A) \subset L_2(B)$ if $A \subset B$ [a], in the dense point limit we have $y \mapsto \psi(y, \mathbf{S}_t) \in L_2(\mathbb{R}^3)$. From now on, we assume to work in the dense-point limit of the space process. In this case we write

$$\begin{aligned}\mu_{X_t}(A) &= \int_{\mathcal{N}_{\mathbb{R}^3}} \int_A |\psi(y, \mathbf{S}_t)|^2 d\mathbf{S}_t dy \\ &= \int_{\mathcal{N}_{\mathbb{R}^3}} \int_A |\psi_{\mathbf{S}_t}(y)|^2 \mu(\mathbf{S}_t) d\mathbf{S}_t dy\end{aligned}$$

where

$$\psi_{\mathbf{S}_t}(y) := \frac{\psi(y, \mathbf{S}_t)}{\sqrt{\mu(\mathbf{S}_t)}}. \tag{6.10}$$

The conditional probability density $\mu_{X_t|\mathbf{S}_t}(x) = |\psi_{\mathbf{S}_t}(x)|^2$ is the probability measure that we can use, repeating the arguments used for model B (see section IV E and section IV F in [54]), to represent the random variable X_t with an operator on an Hilbert space $\mathcal{H}(X_t)$. After conditioning on $\mathbf{S}_t$, since $\psi(x) \in L_2(\mathbb{R}^3)$, then also $\psi_{\mathbf{S}_t}(x)$ is a square integrable function on $\mathbb{R}^3$, i.e. $\psi_{\mathbf{S}_t}(x) \in$

[a] We are assuming that a function $f(x)$ in $L_2(A)$ is extended to $L_2(B)$ setting $f(x) = 0$ when $x \in B/A$.

$L_2(\mathbb{R}^3)|_{\mathbf{S}_t}$ [b]. We have that

$$\begin{aligned}\mathbb{E}[X_t|\mathbf{S}_t] &= \int_{\mathbb{R}^3} x|\psi_{\mathbf{S}_t}(x)|^2 dx \\ &= \int_{\mathbb{R}^3} \psi^*_{\mathbf{S}_t}(x) x \psi_{\mathbf{S}_t}(x) dx \\ &= \langle \psi_{\mathbf{S}_t} | \hat{X}_t \psi_{\mathbf{S}_t} \rangle,\end{aligned} \tag{6.11}$$

where $\hat{X}_t$ is a self-adjoint multiplicative operator on $L_2(\mathbb{R}^3)|_{\mathbf{S}_t}$ whose action is

$$\hat{X}_t \psi_{\mathbf{S}_t}(x) = x \psi_{\mathbf{S}_t}(x), \tag{6.12}$$

and with domain given by $D(\hat{X}_t) := \{\phi \in L_2(\mathbb{R}^3)|_{\mathbf{S}_t} \mid \hat{X}_t \phi \in L_2(\mathbb{R}^3)|_{\mathbf{S}_t}\}$. The domain is defined such that $\hat{X}_t$ is a well defined operator on $L_2(\mathbb{R}^3)|_{\mathbf{S}_t}$. This is exactly the operator representing X_t on an Hilbert space, when we follow the procedure explained in chapter 4. Thus we can conclude that the DPP process induces the separability of the Hilbert space $\mathcal{H}(X_t)$ of model C. It is not difficult to note that $\hat{X}_t$ coincides with the ordinary position operator $\hat{Q}_t$ in non-relativistic quantum mechanics. Finally, also the velocity random variable can be represented on the same Hilbert space, i.e. $\hat{V}_t(t')$ is an operator acting on $L_2(\mathbb{R}^3)|_{\mathbf{S}_t}$, as in the case of model B. Separability ensures that starting from the velocity random variable and constructing the Hilbert space $\mathcal{H}(V_t)$, the operators representing the particle process on this Hilbert space coincide with the position and velocity operators defined on $\mathcal{H}(X_t)$. Summarising, on $L_2(\mathbb{R}^3)|_{\mathbf{S}_t}$ the two operators associated to the particle process are defined. In particular, the position of the particle process coincides with the quantum mechanical operator.

6.2.3 Recovering the momentum operator

Nothing, for the moment, can be said about the relation between $\hat{V}_t(t')$ and the momentum operator of non-relativistic quantum mechanics. In section 4.3, we argue that separability, together with the existence of a suitable operator $\hat{H}$, implies that for $t' \to t$, the operator representing the velocity random variable, reduces to the momentum operator of ordinary non-relativistic quantum mechanics. Here we repeat the argument in a more systematic way, and we introduce the required $\hat{H}$ operator via symmetry group considerations.

We want to show that the operator representing the velocity random variable $\hat{V}_t(t')$, reduces to the ordinary momentum operator of non-relativistic quantum mechanics $\hat{P}_t$. In order to do that, we can use the Ehrenfest theorem. We will not state it in its whole generality, but we consider only the part of this theorem regarding the position operator in quantum mechanics (see Cor 1.2 and Th 1.3 [60]).

[b] It is worth to note that $\psi_{\mathbf{S}_t}(x)$ is a square integrable function on $\mathbb{R}^3$, thus is an element of $L_2(\mathbb{R}^3)$. However, to emphasize the dependence on the configuration of the DPP and distinguish such Hilbert space from the $L_2(\mathbb{R}^3)$ space of the DPP, we choose to use the symbol $L_2(\mathbb{R}^3)|_{\mathbf{S}_t}$.

Theorem 18. *Let $\hat{H}$ be an operator on an Hilbert space $L_2(\mathbb{R}^d)$ of the from*

$$\hat{H} = -\sum_{i=1}^{d} \frac{1}{2m_i} \frac{\partial^2}{\partial q_i^2} + V(q)$$

with domain $D(\hat{H}) = H_2(\mathbb{R}^d)$ [c] where $m_1, \cdots, m_d > 0$, $V : \mathbb{R}^d \to \mathbb{R}$ is a real valued, locally integrable function fulfilling the Kato-Rellich condition[d]. Let ψ_t be a solution of the Schrödinger equation having Hamiltonian $\hat{H}$ written in the $\hat{Q}^i$-representation, where $\hat{Q}^i$ the i-th component of the position operator in non-relativistic quantum mechanics. Then $\langle \hat{Q}^i \rangle_{\psi_t}$ is continuously differentiable with respect to t for any $\psi_t \in D(\hat{H}) \cap D(\hat{Q}^i_t)$ and satisfies the equation

$$\frac{d}{dt}\langle \hat{Q}^i \rangle_{\psi_t} = i \left(\langle \hat{H}\psi_t | \hat{Q}^i \psi_t \rangle - \langle \hat{Q}^i \psi_t | \hat{H} \psi_t \rangle \right).$$

Moreover, if $V : \mathbb{R}^d \to \mathbb{R}$ has also locally weak derivative which is integrable, and also $\sqrt{|\nabla V|}$ satisfies the Kato-Rellich condition, the equation above reduces to

$$\frac{d}{dt}\langle \hat{Q}^i \rangle_{\psi_t} = \frac{1}{m_i} \langle \hat{P}^i \rangle_{\psi_t}. \tag{6.13}$$

First we observe that in model C, $\mu_{X_t|\mathbf{S}_t}(x) = |\psi_{\mathbf{S}_t}(x)|^2$ depends on time (it represents the probability distribution of the position of the particle *at time* t), thus one should write $\psi_{\mathbf{S}_t}(x,t)$ instead of $\psi_{\mathbf{S}_t}(x)$: such time dependence was omitted in section 6.2.2. However, the operator $\hat{X}_t$ associated to the random variable X_t always fulfils (6.12), for any time t. In this sense, it does not evolve in time. Now, computing the expectation value of the velocity random variable, we can write

$$\begin{aligned} \mathbb{E}[V_t(t')|\mathbf{S}_t] &= \frac{\mathbb{E}[X_{t'}|\mathbf{S}_t] - \mathbb{E}[X_t|\mathbf{S}_t]}{t'-t} \\ &= \frac{\langle \hat{X}_{t'} \rangle_{\psi_{t'}} - \langle \hat{X}_t \rangle_{\psi_t}}{t'-t}, \end{aligned}$$

where we used (6.11) and we omitted the $\mathbf{S}_t$ label in the last line. As already observed at the end of [54], the expectation values used in the first line of

[c] $H_2(\mathbb{R}^d)$ is the Sobolev space of square integrable functions on $\mathbb{R}^d$.

[d] Let $\hat{G} : D(\hat{G}) \to \mathcal{H}$ and $\hat{V} : D(\hat{V}) \to \mathcal{H}$ be densely defined operators on the Hilbert space $\mathcal{H}$. If there exist $a, b \in [0, \infty)]$ such that

$$\|\hat{V}\varphi\| \leqslant a\|\hat{G}\varphi\| + b\|\varphi\|,$$

for any $\varphi \in D(\hat{G})$, then $\hat{V}$ is said *$\hat{G}$-bounded*. The greatest lower bound of all the numbers a satisfying the condition above for some b, is said *relative bound* of $\hat{V}$ with respect to $\hat{G}$. We say that $\hat{V}$ satisfies the *Kato-Rellic condition* if, given

$$\hat{G} = -\sum_{i=1}^{d} \frac{1}{2m_i} \frac{\partial^2}{\partial q_i^2},$$

the $\hat{V}$ is $\hat{G}$-bounded with relative bound $a < 1$.

the above formula are defined on different probability spaces: the three random variables cannot be defined on the same probability space after the space process is removed. Taking the limit $t' \to t$, we get

$$\lim_{t' \to t} \mathbb{E}[V_t(t')|\mathbf{S}_t] = \lim_{t' \to t} \frac{\langle \hat{X} \rangle_{\psi_{t'}} - \langle \hat{X} \rangle_{\psi_t}}{t' - t} = \frac{d}{dt} \langle \hat{X} \rangle_{\psi_t}.$$

Since $\hat{Q}_t = \hat{X}_t$, equation (6.13) suggests that

$$\lim_{t' \to t} \mathbb{E}[V_t(t')|\mathbf{S}_t] = \lim_{t' \to t} \langle \hat{V}_t(t') \rangle_{\psi_t} = \frac{1}{m} \langle \hat{P} \rangle_{\psi_t}.$$

At the level of operators, if the relation above holds, we can say that $\lim_{t' \to t} \hat{V}_t(t') = \hat{P}_t/m$ *weakly*. In order to state that, we need to introduce an operator $\hat{H}$ on $L_2(\mathbb{R}^3)|_{\mathbf{S}_t}$ fulfilling the conditions in theorem 18.

To introduce $\hat{H}$ in a physically meaningful manner, we can proceed using the symmetry group. We have already observed in the introduction, that the symmetry group of quantum theory is one of the three underling assumptions of the set of postulates used. The symmetry group of non-relativistic quantum mechanics is the central extension of the Galilean group (see [74], or Ch. 12 of [3]) and we may impose that the probabilistic description of our model is invariant under this symmetry group. A subgroup of the Galilean group is the abelian additive group $(\mathbb{R}, +)$ of time translations. The operator $\hat{H}$ can be introduced, via Stone's theorem, as the generator of the time translation of the Galilean group. In particular, for a single particle, it is

$$\hat{H} = \frac{1}{2m} \hat{P}^2 \tag{6.14}$$

where $\hat{P}$ is the momentum operator in non-relativistic quantum mechanics and such operator fulfils the requirements of theorem 18. This is nothing but the kinetic term for a non-relativistic free quantum particle. This means that $\psi_{\mathbf{S}_t}(x,t) \in L_2(\mathbb{R}^3)|_{\mathbf{S}_t}$ evolves in time as

$$\psi_{\mathbf{S}_{t'}}(x,t') = e^{-i\hat{H}t} \psi_{\mathbf{S}_t}(x,t),$$

or equivalently

$$i \frac{d}{dt} \psi_{\mathbf{S}_t}(x,t) = \hat{H} \psi_{\mathbf{S}_t}(x,t),$$

which is the usual Schrödinger equation for a free non-relativistic quantum particle. Clearly this time evolution for the probability of X_t is not a consequence of model C only, but is a general feature of the Galilean group. In addition by, representing the Galilean group on $L_2(\mathbb{R}^3)|_{\mathbf{S}_t}$,we also obtain the operator representation of other physical observables like, for example, the angular momentum. Summarising, by using considerations on the symmetry group of non-relativistic systems one can introduce the operator $\hat{H}$ with the correct properties in a physically meaningful way. This justifies the conclusion that for a non-relativistic particle,

$$\lim_{t' \to t} \hat{V}_t(t') = \frac{\hat{P}_t}{m}$$

weakly, for some constant $m > 0$ having the dimension of a mass. In this way, we may recover the quantum mechanical momentum operator in model C. Note that this is not a consequence of the Ehenefest theorem and Galilean group only: *model C, assuming a stochastic geometry for the space in which the particle moves, is capable to reproduce the correct Hilbert space structure without postulating it in advance.* Because of this we can conclude that, model C *together with the Galilean group* is capable to reproduce the non-relativistic quantum mechanics of a free particle.

6.2.4 n-particle generalization

In the previous section, we arrived at the conclusion that we can use model C to describe a free single quantum particle. Here we want to extend the discussion to the n-particle case. In particular, we want to understand if we can derive from model C the *tensor product postulate* of non-relativistic quantum mechanics, namely the fact that the Hilbert space associated to two distinguishable particles is the tensor product of the single-particle Hilbert spaces. Here we will consider the 2-particle case only, since considerations to the case $n > 2$ can be straightforwardly generalized. As we will see, despite model C is capable to describe two or more particles, we cannot fully derive the tensor product postulate without adding further conditions (or at least the author was not able to do so).

Following the idea explained in section 6.2.2, where we discussed the position random variable for a single particle, we generalize the description to the case of two particles in the following natural way. Given the space process $\mathbb{S}_t^C$ on $\mathbb{R}^3$ at a given time t, let us consider two marked DPPs $\tilde{\mathbb{S}}_t^C(a)$ and $\tilde{\mathbb{S}}_t^C(b)$, one for each particle, having both as ground process $\mathbb{S}_t^C$. Then we consider the point process obtained from the Cartesian product of $\tilde{\mathbb{S}}_t^C(a)$ and $\tilde{\mathbb{S}}_t^C(b)$, namely

$$\overline{\mathbb{S}}_t^C := \{(x_t^{(1)}, 1) \times (y_t^{(1)}, 1), (x_t^{(1)}, 1) \times (y_t^{(2)}, 2), \cdots\}$$

where $x_t^{(i)}$ and $y_t^{(j)}$ are points of $\mathbb{S}_t^C$. The ground process of $\overline{\mathbb{S}}_t^C$ has counting measure

$$N_g(A \times B) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} N_{\overline{\mathbb{S}}_t^C}(A \times \{i\} \times B \times \{j\}),$$

which is finite, because $N_g(A \times B)$ is just the number of points of $\mathbb{S}_t^C \times \mathbb{S}_t^C$ in $A \times B$, and $\mathbb{S}_t^C$ is a DPP. Thus each component must be $N_{\overline{\mathbb{S}}_t^C}(A \times \{i\} \times B \times \{j\}) < \infty$, and so it defines a point process. Such process can be obtained from the original point process $\mathbb{S}_t^C$ by removing all the points having labeling different from i and j. By theorem 15, the thinned point process is again a DPP and has *at least two points*. However, we can always say that the point process has just a *single point* if we consider, as ground process of $\overline{\mathbb{S}}_t^C$, the Cartesian product of two copies of $\mathbb{S}_t^C$. This is the same observation done in section 5.1.1 which allows to interpret the moment measure M_2 for a point process ξ as the intensity measure M_1 of the point process $\xi \times \xi$. This means that for i and j

fixed, the counting measure $N_{\overline{\mathbb{S}}_t^C}(A \times \{i\} \times B \times \{j\})$ can define a point process on $\mathbb{R}^3 \times \mathbb{R}^3 = \mathbb{R}^6$ having a single point. Now, we introduce two selection random variables, $I_t : \Omega \to \mathbb{N}$ and $J_t : \Omega \to \mathbb{N}$, which may depend on each other, and we define the point process on $\mathbb{R}^6$

$$\xi_{(X_t^a, X_t^b)} := \{N_{\overline{\mathbb{S}}_t^C}(A \times \{I_t\} \times B \times \{J_t\})\}_{A \times B \in \mathscr{B}(\mathbb{R}^6)},$$

as done in section 6.2.2. The joint probability distribution for the positions of two particles at a given time t will be

$$\mu_{(X_t^a, X_t^b)}(A \times B) = \int_{A \times B} \rho_1(x, y) dx dy,$$

where $\rho_1(x, y)$ is the intensity measure density of the point process $\xi_{(X_t^a, X_t^b)}$. At this point a difficulty arises: we are not able to deduce the that $\rho_1(x, y) = |\psi(x, y)|^2$ with $\psi(x, y) \in L_2(\mathbb{R}^6) = L_2(\mathbb{R}^3) \otimes L_2(\mathbb{R}^3)$. The reason for this difficulty lies in the fact that, if a point process ξ is a DPP, then the point process $\xi \times \xi$ is not in general a DPP. As consequence, we cannot define the Hilbert space $L_2(\mathbb{R}^6)|_{\mathbf{S}_t}$, and so in general the Hilbert space on which the operators associated to the two position random variables are defined, may be non-separable. More precisely, we can have two cases:

a) The point process $\overline{\mathbb{S}}_t^C$ is determinantal. If so, as in the single particle case, we know that there exist $\psi(x, y) \in L_2(\mathbb{R}^6) = L_2(\mathbb{R}^3) \otimes L_2(\mathbb{R}^3)$ such that

$$\mu_{(X_t^a, X_t^b)}(A \times B) = \int_A \int_B |\psi(x, y)|^2 dx dy.$$

In this case, using (6.10), we obtain that the (conditional on $\mathbf{S}_t$) probability density of the position of the two particles, $\mu_{(X_t^a, X_t^b)|\mathbf{S}_t}$, is given by the square modulus of a function in $L_2(\mathbb{R}^3) \otimes L_2(\mathbb{R}^3)$. At this point depending on whether $\mu_{(X_t^a, X_t^b)|\mathbf{S}_t}$ factorize, i.e. $\mu_{(X_t^a, X_t^b)|\mathbf{S}_t} = \mu_{X_t^a|\mathbf{S}_t} \mu_{X_t^b|\mathbf{S}_t}$, or not, one can describe particles which are uncorrelated or entangled.

b) The point process $\overline{\mathbb{S}}_t^C$ is not determinantal. In this case we are in a situation which is similar to the one of model B. Once we choose to represent the position random variables with operators, the Hilbert space on which both these operators acts, $\mathcal{H}(X_t^a, X_t^b)$, is in general non-separable, which implies that $\mathcal{H}(X_t^a, X_t^b) \supset L_2(\mathbb{R}^3) \otimes L_2(\mathbb{R}^3)$. However, also in this case we can describe uncorrelated or entangled-like particles depending on whether the probability distribution $\mu_{(X_t^a, X_t^b)}(A \times B)$ after conditioning on $\mathbf{S}_t$ (which, in any case, defines a state on the algebra describing two particles in model C) factorise or not.

We have to admit that the case *a*) seems to be very rare. Thus, we conclude that using the natural generalization of the procedure showed in section 6.2.2, *we cannot say that the Hilbert space on which we can describe two particles*

is always the tensor product of the single-particle Hilbert spaces. To overcome this difficulty we suggest the following approach. Instead of deriving the tensor product structure from the point process structure of space, one may solve this difficulty "probabilistically", as explained below. For a single particle, the ordinary quantum mechanical description of probability applies. This means that the whole machinery of quantum logic and its propositional calculus can be used to describe a single particle in model C. This is almost sufficient to derive the tensor product postulate. In fact, following the work of D. Aerts and I. Daubechies [19], *within the quantum logic framework*, given two quantum systems and requiring that 1) the structure of the two systems is preserved; 2) a measurement on one system does not disturb the other (in the sense that operators associated to the two systems commute); 3) maximal information obtained on both systems separately gives the maximal information on the joint system; one can conclude that the Hilbert space of the joint quantum system is the tensor product of the single particle Hilbert space. Summarizing, by adding these assumptions, the tensor product postulate of quantum mechanics can be used within the framework of model C.

6.2.5 Does the Bell and PBR no-go theorems apply to model C?

The work done till now shows that model C can be put in correspondence with ordinary non-relativistic quantum mechanics. However, the models proposed in chapter 4 and model C use essentially classical probability theory (more precisely, the Kolmogorov's axiom of probability theory), a possibility which seems to be excluded by various no-go theorems. Here we will argue why Bell's theorem [75] and PBR theorem[76] do not rule out the models proposed and in particular model C.

In chapter 3 we described the general probabilistic framework in which the three models were elaborated. In particular we introduce the notion *contextual probabilistic model* elaborated in [47] and [48]. In chapter 4 and also here, each random variable is considered on its own probability space. The position X_t and the velocity $V_t(t')$ are defined initially on a common probability space however, we describe them only when the space is removed. This changes the probability space on which these random variables are considered. X_t is defined on a probability space obtained by conditioning on $\mathbf{S}_t$. Also the velocity $V_t(t')$ can be defined on a probability space obtained conditioning on $\mathbf{S}_t$, but is not the same of X_t because the Bayes theorem does not hold (we use the unconditional transition probabilities). Thus each random variable is considered on its own probability space: this is a consequence of the removal of the space process. The entropic uncertainty relations that we found in the models enforce exactly this conclusion: the position and the velocity random variable *must be in two different probability spaces.* Both Bell's and PBR theorem assume that all the observables can be represented as random variables acting on the *same* probability space (where the Bayes theorem holds): this is the reason why the two theorems cannot be applied directly. For the analysis of the probabilistic argu-

ments used in the Bell's theorem we refer to [47, 48, 77]. There it is explained how fundamental is the assumption of a single probability space for the validity of the Bell's argument. Moreover, as explained in [78, 79, 32] the non commutativity of the position operator at two different times (something that happens in model C) is sufficient to have a maximal violation of the Bell's inequality. A direct application of the PBR argument to model C is not possible at the moment. However since in model C each observable is in its own probability space, the PBR theorem does not seem to be applicable in this case.

Chapter 7

Conclusion

In this last chapter we list and criticize some of the assumptions used in the models presented in [54] and in model C. Let us start with an informal summary of the work done. In chapter 4 we start to develop a series of models to derive, using probabilistic considerations, the commutation relations $[\hat{Q}_i, \hat{P}_i] = i\hat{\mathbb{I}}$ and the Hilbert space structure of quantum mechanics. The physical space was always assumed to be the random distribution of points and not the underlying $\mathbb{R}^3$ space. On the physical space, a particle was defined through two random variables: one for its position and one for its velocity at a given time (which is always an external parameter). This particle moves by jumps from one point of the physical space to the other. Once the physical space is not taken into account in the probabilistic description of the particle (operation that was implemented by conditioning on *a* space configuration, and not averaging over *all* the space configurations), the mathematical model that we can use to describe the particle is a non-commutative probability theory (we can no longer use the ordinary measure-theoretic description consisting in a single probability space). Imposing conditions on the space process (requiring to be a DPP) and on the group of symmetry under which the particle is invariant (the symmetry group is the central extension of the Galilean group), we showed that the description of the particle in this framework coincides with that of a non-relativistic quantum particle. Summarizing, we showed that assuming that the geometry of the physical space is stochastic and discrete, a simple jump-type kinematics for a particle is capable to reproduce the ordinary quantum description in the dense-point limit.

7.1 Critics

Below we list and discuss some possible critics on the proposed model.

i) *Space process.* We chose to model space with a random distribution of points over a $\mathbb{R}^3$. Nothing forbid to start from some kind of "*stochastic manifold*", in the sense of a manifold on which the distance between

points changes at random. For an interesting work in this direction, one may consult [80, 81]. Other interesting structures can be *graphs*, namely a collection of points (called *vertex*) and relations between two points which can be visualized with a straight line linking the two related points (called *edge*). To implement the intrinsic randomness of space, one can imagine to have a probability distribution over the set of all edges, describing the probability that an edge, linking point a to point b, exists. This last approach is particularly interesting since it leaves open the possibility to recover the underlying mathematical space $\mathbb{R}^3$ with some limiting procedure [82]. Connections between DPP and graphs are known since long time [64, 65]. We also note the following curious fact: the dimensionality of the problem is dictated by the dimensionality of the underlying mathematical space and not of the physical space. The use of graphs for the description of the space process may also help to remove this unpleasant feature of the model, for example, using techniques similar to those used for causal set [83, 84, 85].

ii) *Hilbert space representation.* As already discussed in chapter 3, the Hilbert space representation of the random variables of the model is obtained via algebraic methods. Only in some cases, we can represent *explicitly* the probability distribution with vectors of the Hilbert space, i.e. in all the cases where QRLA applies. However, only for dichotomic and trichotomic random variables, simple conditions for the application of QRLA are known. In the more general case, no simple condition is known up to now. In [86] an interesting possibility of how to represent a probability distribution with vectors in $L_2(\mathbb{R}^3)$ is presented: there a vector in $L_2(\mathbb{R}^3)$ from a probability distribution is constructed but nothing is said about a change of basis. This means that, in principle, given a probability distribution of the position, we can construct $\psi(x)$, but we do not know how to change basis and derive the probability distribution of the velocity. In the limit where the velocity of the model coincides with the quantum mechanical momentum, we know how a change of basis can be performed, but in the general case, the algorithm is not available. In [Cur7] another possible approach is presented despite it is only able to construct $\psi(x)$ up to a local phase factor and so not very satisfying.

iii) *Determinantal point process.* In order to induce the correct Hilbert space structure, i.e. an infinite dimensional separable Hilbert space, we used a specific point process: the DPP. Thanks to this structure we were able to derive the correct description of a free quantum particle. However, since space is "removed" from the models, it seems strange that space still plays such a fundamental role. For example, there are other point processes, like the permanental or more generally α-determinantal point process [87], which use kernels of locally trace-class operators in $L_2(\mathbb{R}^3)$. Thus one may expect to obtain the quantum description of a single particle also using these processes. However, results which hold for a DPP do not hold anymore for these processes and so technical difficulties may arise.

Moreover, also the necessity of special classes of point processes to induce the L_2 structure seems to be a rather artificial condition: probably solving the Hilbert space representation problem explained in *ii*) or some algebraic consideration, may remove this constraint leaving us free to choose any point process we like or, more generally, random fields.

iv) *Tensor product postulate.* As observed in section 6.2.4, model C has difficulties in justifying the tensor product postulate. This is not surprising since we use DPP, to induce the correct Hilbert space structure: when we cannot use anymore the DPP, the problem of the correct Hilbert space structure appears. The way out proposed uses results of quantum logic to justify such a postulate [19]. Because quantum logic can be considered as the background logic of probability theory describing a quantum system, this solution has a "probabilistic flavour". Note that the necessity to appeal to a more detailed analysis of the probabilistic structure of the models, is in agreement with the observations done in *ii*) and *iii*) on the L_2 structure.

v) *Dependence of the Hilbert space on the space configuration.* When we constructed the Hilbert space associated to a single particle in section 6.2.2, we obtained a $\mathbf{S}_t$-dependent object. If such a dependence is relevant for the evolution of the state vector, more general time-evolutions of $\psi_{\mathbf{S}_t}$ are possible, for example a stochastic unitary time-evolution. However in this case it is not expected to give rise to any observable effect, as suggested in [Cur6] where a somehow similar situation is considered.

7.2 Future outlook

The critical points recognized are of two kinds: geometric and probabilistic. From the probabilistic side, find simple conditions for the applicability of the QRLA to any couple of random variables may solve the criticality *ii*), *iii*) and *v*). On the other hand, the geometric problem mentioned in *i*) strays in the active research field of quantum gravity. The models proposed are interesting because despite space is removed to get the quantum description, it is an active player of the whole model. Space possesses its own dynamics, something which resembles the situation we have in general relativity (despite the space dynamics used here was not chosen to mimic any general relativistic effect). Moreover, starting from a manifold equipped with a stochastic metric, the stochastic mechanics [88, 89, 90] can be obtained (see [81] for an overview). It is interesting to observe that a similar picture, of a particle jumping at random on the space, emerge in stochastic mechanics from a careful analysis of the stochastic trajectories of a quantum particle[89, 91].

To conclude, we think that an interesting part of the new description presented here is the model of space used. Maybe, by applying this new ingredient in contests which are outside the realm of quantum mechanics, can give rise

to new interesting results. For example, this idea of space may be used in the contest of the early universe, where the effects of such a structure of space may be relevant.

Chapter 8

Appendix

Proof of the theorem 2

Proof. (see [25]) We can easily see that $\mathcal{A} = \{\hat{M}_f | f \in L_\infty(\Omega, \mathcal{E}, P)\}$ is an algebra of operators (with involution). The state ϕ is clearly faithful, hence what we need to prove is that $\mathcal{A}$ is strongly closed. Then if this is true, by the von Neumann's double commutant theorem [3], ϕ is always a normal state. In order to prove the strong closure, let us consider a sequence of functions $\{f_n\}_{n\in\mathbb{N}} \in L_\infty(\Omega, \mathcal{E}, P)$ such that

$$s - \lim_{n\to\infty} \hat{M}_{f_n} = \hat{X}$$

where $\hat{X}$ is some operator. The above expression is equivalent to

$$L_2 - \lim_{n\to\infty} \hat{M}_f \psi = \hat{X}\psi \qquad \forall \psi \in L_2(\Omega, \mathcal{E}, P)$$

We may always assume, without loss of generality that $\|\hat{X}\| = 1$. Now, we need to prove that $\hat{X} = \hat{M}_f$. Let us set $f(\omega) := \hat{X}1(\omega)$, since the identity function $1(\omega) \in L_2(\Omega, \mathcal{E}, P)$. Now, consider the set

$$E_\epsilon := \{\omega \in \Omega | |f(\omega)|^2 \geqslant 1 + \epsilon\}$$

for any $\epsilon > 0$. Clearly, $E_\epsilon \in \mathcal{E}$ and so, using the Cauchy – Schwarz inequality and recalling that $\|\hat{X}\| = 1$, we can write

$$\begin{aligned} P(E_\epsilon) &= \int_\Omega \chi_{E_\epsilon}(\omega) P(d\omega) = \|\chi_{E_\epsilon}(\omega)\|^2_{L_2} \\ &\geqslant \|\hat{X}\chi_{E_\epsilon}(\omega)\|^2_{L_2} = \|f\chi_{E_\epsilon}(\omega)\|^2_{L_2} \\ &= \int_{E_\epsilon} |f(\omega)|^2 P(d\omega) \geqslant (1+\epsilon) P(E_\epsilon) \end{aligned}$$

which implies that $P(E_\epsilon) = 0$. Since this holds for any $\epsilon > 0$, then $|f| \leqslant 1$ almost everywhere with respect to P, from which we conclude that $f \in L_\infty(\Omega, \mathcal{E}, P)$.

For $g \in L_\infty(\Omega, \mathcal{E}, P)$, since $L_\infty(\Omega, \mathcal{E}, P) \subset L_2(\Omega, \mathcal{E}, P)$, we can write that:

$$\begin{aligned}\hat{X}g &= L_2 - \lim_{n\to\infty} f_n g \\ &= \hat{M}_g L_2 - \lim_{n\to\infty} f_n \\ &= \hat{M}_g f = gf.\end{aligned}$$

Since $L_\infty(\Omega, \mathcal{E}, P)$ is dense in $L_2(\Omega, \mathcal{E}, P)$ we can conclude that $\hat{X} = \hat{M}_f$. □

Proof of the theorem 10

Before starting the proof, let us first explain its structure and some technical facts. The proof can be divided in two parts. In the first part (STEP 1 - STEP 3) is just a proof that dispersion free states do not exist in a non abelian algebra. Clearly it is not the first time that this fact is proved, however here we prove this fact in a probabilistic manner: we construct explicitly the random variables and the joint probability spaces. Only in the last part (STEP 4), the entropic uncertainty relations come into play. To explicitly construct the joint probability space and the random variables, a technical point is needed.

Theorem 19 (Th. 9.15 [3]; Cp. IV, Th. 2.3 [14])**.** *Let $\mathcal{H}$ be a separable Hilbert space and let $\hat{A}_1, \cdots, \hat{A}_n$ be a set of self-adjoint mutually commuting bounded operators. Let $\hat{P}^{(\hat{A}_1)}, \cdots, \hat{P}^{(\hat{A}_n)}$ be the associated PVMs, then there exists a unique PVM $\hat{P}^{(\hat{\mathbf{A}})}$ such that*

$$\hat{P}^{(\hat{\mathbf{A}})}(B_1 \times \cdots \times B_n) := \hat{P}^{(\hat{A}_1)}(B_1) \cdots \hat{P}^{(\hat{A}_n)}(B_n)$$

where $B_i \in \mathscr{B}(\mathbb{R})$ for any i. $\hat{P}^{(\hat{\mathbf{A}})} : \mathscr{B}(\mathbb{R}^n) \to \mathcal{B}(\mathcal{H})$ is called joint PVM *of $\hat{A}_1, \cdots, \hat{A}_n$. If $F : \mathbb{R} \to \mathbb{C}$ is bounded measurable function, then*

$$\int_{\mathbb{R}^n} F(x_k(\mathbf{x}))\hat{P}^{(\hat{\mathbf{A}})}(d\mathbf{x}) = \int_{\mathbb{R}} F(x_k)\hat{P}^{(\hat{A}_k)}(dx_k) = F(\hat{A}_k)$$

where $\mathbf{x} = (x_1, \cdots, x_n) \in \mathbb{R}^n$ and $x_k(\mathbf{x})$ is the k-th component of $\mathbf{x}$.

We can say that our proof is a corollary of this theorem. In particular we are interested in the consequences of it on the spectral measure. First we recall that, as a consequence of the spectral decomposition theorem, given a bounded self-adjoint operator $\hat{T}$ then its spectrum is the support of the associated PVM, i.e. $\sigma(\hat{T}) = \text{supp}(\hat{P}^{(\hat{T})})$. Now, the above theorem ensures that $\sigma(\hat{\mathbf{A}}) = \sigma(\hat{A}_1) \times \cdots \times \sigma(\hat{A}_n)$. Using the joint PVM and given a normalised $\psi \in \mathcal{H}$, we can define the *joint spectral measure* simply as:

$$\mu_\psi^{(\hat{\mathbf{A}})}(d\mathbf{x}) = \langle\psi|\hat{P}^{(\hat{A}_1)}(dx_1) \cdots \hat{P}^{(\hat{A}_n)}(dx_n)\psi\rangle$$

which is a probability measure on the probability space $(\sigma(\hat{A}_1)\times\cdots\times\sigma(\hat{A}_n), \mathcal{E}, \mu_\psi^{(\hat{\mathbf{A}})})$, where $\mathcal{E}$ is a borel σ-algebra. This is the tensor product of the probability spaces

associated to the spectral measures one can construct from the single PVMs, i.e. $(\sigma(\hat{A}_k), \mathscr{B}(\sigma(\hat{A}_k)), \mu_\psi^{(\hat{A}_k)}(dx))$. The independence properties of the probability measure $\mu_\psi^{(\hat{\mathbf{A}})}$ depend on ψ.

Hence the existence of a joint PVM of the multiplicative form as described in the theorem, ensures the existence of a common probability space for all commuting operators when thought as random variables. If the operators $\hat{A}_1, \cdots, \hat{A}_n$ do not commute this is not anymore possible: we can always multiply them obtaining again a self-adjoint operator, the spectral theorem ensures the existence of a PVM for such product and so a probability space (i.e. a spectral measure) can be defined, but this probability space cannot be related (at least in a trivial manner, i.e. the one seen above) to the probability spaces associated to each operator.

The second technical fact is the following. In general, not all the GNS representations of a C^*-algebra $\mathcal{A}$ are faithful. Faithfulness is an important property because it allows to think the whole algebra as operators over the *same* Hilbert space. Luckily, there exists the (general) *Gel'fand-Naimark theorem* which tells us how to construct a representation which is always faithful (the so called universal representation).

Theorem 20 (Th. 14.23 [3])**.** *For any C^*-algebra with unit $\mathcal{A}$ there exists an Hilbert space $\mathcal{H}$ and an isometric *-isomorphism $\Pi : \mathcal{A} \to \mathcal{B}_{GN}$, where $\mathcal{B}_{GN} \subset \mathcal{B}(\mathcal{H})$ is a C^*-sub-algebra of $\mathcal{B}(\mathcal{H})$.*

More precisely, the universal representation of $\mathcal{A}$ on $\mathcal{H}$ is defined as the direct sum of all the representations with respect to ω, i.e. $\Pi := \oplus_\omega \pi_\omega$. The Hilbert space is defined in a similar way, i.e. $\mathcal{H} := \oplus_\omega \mathcal{H}_\omega$. This allows to always think about $\mathcal{A}$ as a C^*-sub-algebra if $\mathcal{B}(\mathcal{H})$, but for technical reasons, we will need to consider the von Neumann algebra that we can construct on $\mathcal{H}$ closing $\mathcal{B}_{GN}$. This forces us to the following definition.

Definition 40. *Given a C^*-algebra with unit $\mathcal{A}$, then $\mathcal{A}^{vn} := \overline{\Pi(\mathcal{A})}^s$ is the closure in the strong topology of $\mathcal{A}$ when it is though as an algebra of bounded operators on $\mathcal{H}$, the Hilbert space of the universal representation. $\mathcal{A}^{vn}$ will be called* strong closure of $\mathcal{A}$.

Now we can start with the proof of theorem 10.

Proof. STEP 1: $[a, b] = 0 \Rightarrow a = f_1(c), b = f_2(c)$ FOR $c \in \mathcal{A}^{vn}$.
We will prove that given $a, b \in \mathcal{A}$, if $[a, b] = 0$ then we can always find two maps f_1 and f_2 and an element of the strong closure of the algebra $c \in \mathcal{A}^{vn}$ such that $a = f_1(c)$ and $b = f_2(c)$. Let $\mathcal{A}_c[a, b]$ be the commutative sub-algebra of $\mathcal{A}$ generated by a and b.

Consider a state $\omega : \mathcal{A} \to \mathbb{C}$, then by the GNS theorem we may represent $\mathcal{A}$ using bounded operators over $\mathcal{H}_\omega$. Assume that $\hat{\pi}_\omega$ is faithful (i.e. one-to-one, we will deal with the general case at the end) and let $\hat{A}_\omega = \hat{\pi}_\omega(a)$ and $\hat{B}_\omega = \hat{\pi}_\omega(b)$ be the representation of a and b on it. Since $[\hat{A}_\omega, \hat{B}_\omega] = 0$ then theorem 19 ensures that there exists a joint spectral measure, say $\hat{P}^{(\hat{C})}$,

which is associated to some self-adjoint bounded operator $\hat{C}$ whose existence is guaranteed by the spectral decomposition theorem. Now, take $F : \mathbb{R} \to \mathbb{C}$ as the identity function, then theorem 19 allows us to write

$$\int_{\sigma(\hat{A}_\omega)\times\sigma(\hat{B}_\omega)} x_1(\alpha,\beta)\hat{P}^{(\hat{C})}(d\alpha d\beta) = \\ = \int_{\sigma(\hat{A}_\omega)} \alpha \hat{P}^{(\hat{A}_\omega)}(d\alpha) = \hat{A}_\omega$$

where x_1 is the projector on the 1-th component of the vector (α, β). Thus we can see that, setting $f_1(\cdot) = x_1(\cdot)$, we have $\hat{A}_\omega = f_1(\hat{C})$. Clearly the same holds for $\hat{B}_\omega$. Because by construction $\hat{C}$ commutes with either $\hat{A}_\omega$ and $\hat{B}_\omega$, then $\hat{C} \in \mathcal{A}_c[\hat{A}_\omega, \hat{B}_\omega]$. The chosen representation $\hat{\pi}_\omega : \mathcal{A} \to \mathcal{B}(\mathcal{H}_\omega)$ is faithful, hence we can conclude that $c \in \mathcal{A}_c[a, b] \subset \mathcal{A} \subset \mathcal{A}^{vn}$, $a = f_1(c)$ and $b = f_2(c)$. Assume now that the representation is not faithful. In this case we may invoke the Gel'fand-Naimark theorem and then the same arguments apply. This time c may not belong to the original algebra $\mathcal{A}$ (it belongs to the strong closure of $\mathcal{A}$, $c \in \mathcal{A}_c[\hat{A}_\omega, \hat{B}_\omega] \subset \mathcal{A}^{vn}$ in general) because c this time is a generic element of $\mathcal{B}(\mathcal{H})$, not of $\mathcal{B}_{GN}$.

STEP 2: GIVEN $\Pi : \mathcal{A} \to \mathcal{A}^{vn}$ THEN $\sigma_{\mathcal{A}^{vn}}(\Pi(a)) \subset \sigma_{\mathcal{A}}(a)$.
This is a technical step. We recall that given $a \in \mathcal{A}$, C^*-algebra with unit I, *the spectrum of* a (in $\mathcal{A}$) from the algebraic point of view is the set $\sigma_{\mathcal{A}}(a) := \{\lambda \in \mathbb{C} | \nexists (a - \lambda\mathsf{I})^{-1} \in \mathcal{A}\}$.

Let $\tau : \mathcal{A} \to \mathcal{A}^{vn}$ be a unit-preserving *-homomorphism between $\mathcal{A}$ and its strong closure (it cannot be an isomorphism). An example is $\tau(\cdot) = \Pi(\cdot)$, i.e. the universal representation itself. Let I and I_{vn} be the identities of $\mathcal{A}$ and $\mathcal{A}^{vn}$, respectively. Then by definition of τ, $\mathsf{I}_{vn} = \tau(\mathsf{I})$ and so given $\lambda \in \mathbb{C}$ we have

$$\mathsf{I}_{vn} = \tau((a - \lambda\mathsf{I})^{-1}(a - \lambda\mathsf{I})) = \tau((a - \lambda\mathsf{I})^{-1})(\tau(a) - \lambda\mathsf{I}_{vn})$$

and so we have that $(\tau(a) - \lambda\mathsf{I}_{vn})^{-1} = \tau((a - \lambda\mathsf{I})^{-1})$. Now, if $\lambda \in \sigma_{\mathcal{A}^{vn}}(a)$, then $\nexists(\tau(a) - \lambda\mathsf{I}_{vn})^{-1} \in \mathcal{A}^{vn}$, which is possible if and only if $\nexists(a - \lambda\mathsf{I})^{-1} \in \mathcal{A}$ which means that $\lambda \in \sigma_{\mathcal{A}}(a)$. The converse is not true in general. Thus $\sigma_{\mathcal{A}^{vn}}(\tau(a)) \subset \sigma_{\mathcal{A}}(a)$.

STEP 3: $[a, b] = 0 \Rightarrow \exists\omega$ SUCH THAT $\mu^{(a)}_\omega(\cdot) = \delta_\alpha(\cdot)$ AND $\mu^{(b)}_\omega(\cdot) = \delta_\beta(\cdot)$.
Because $[a, b] = 0$, we know that there exists a probability space where a and b are ordinary random variable. Let $\mu^{(a)}_\omega(dx)$ and $\mu^{(b)}_\omega(dx)$ be the spectral measure of a and b for a given state ω on $\mathcal{A}$. We want to prove that there exists a state ω such that $\mu^{(a)}_\omega(dx) = \delta_\alpha(dx)$ and $\mu^{(b)}_\omega(dx) = \delta_\beta(dx)$ for some suitable $\alpha \in \sigma_{\mathcal{A}}(a)$ and $\beta \in \sigma_{\mathcal{A}}(b)$. For any state $\omega^{vn} : \mathcal{A}^{vn} \to \mathbb{C}$, the universal representation $\Pi : \mathcal{A} \to \mathcal{A}^{vn}$ allows us to define the corresponding state ω on $\mathcal{A}$, setting $\omega := \omega^{vn} \circ \Pi$. Then if $\hat{C} \in \mathcal{A}^{vn}$ and $f : \mathcal{A}^{vn} \to \mathcal{A}$ is a function we can write that:

$$\omega(f(\hat{C})) = \omega^{vn}(\Pi(f(\hat{C}))) \\ = \int_{\sigma_{\mathcal{A}^{vn}}(\hat{C})} [\Pi \circ f](\mathbf{x})\mu^{(\hat{C})}_{\omega^{vn}}(d\mathbf{x}).$$

Suppose that $\hat{C}$ is the operator of STEP 1. By theorem 19 we can also write that:

$$\omega(f(\hat{C})) = \int_{\sigma_{\mathcal{A}^{vn}}(\hat{A}_\omega)\times\sigma_{\mathcal{A}^{vn}}(\hat{B}_\omega)} [\Pi \circ f](\mathbf{x})\mu_{\omega^{vn}}^{(\hat{C})}(d\mathbf{x})$$

Form what we found in STEP 2, we know that $\sigma_{\mathcal{A}^{vn}}(\hat{A}_\omega)\times\sigma_{\mathcal{A}^{vn}}(\hat{B}_\omega) \subset \sigma_{\mathcal{A}}(a)\times\sigma_{\mathcal{A}}(b)$. Let us now choose the state ω'^{vn} such that $\mu_{\omega'^{vn}}^{(\hat{C})}(d\mathbf{x}) = \delta_\gamma(d\mathbf{x})$ for some $\gamma = (\alpha_1, \beta_1)$, which is always possible for a single random variable. Now if we set $f = f_1$ as in STEP 1, we can write that:

$$\begin{aligned}
\omega'(a) &= \int_{\sigma_{\mathcal{A}^{vn}}(\hat{A}_\omega)\times\sigma_{\mathcal{A}^{vn}}(\hat{B}_\omega)} \Pi(f_1(\mathbf{x}))\mu_{\omega'^{vn}}^{(\hat{C})}(d\mathbf{x}) \\
&= \int_{\sigma_{\mathcal{A}^{vn}}(\hat{A}_\omega)\times\sigma_{\mathcal{A}^{vn}}(\hat{B}_\omega)} \Pi(f_1(\mathbf{x}))\delta_\gamma(d\mathbf{x}) \\
&= \int_{\sigma_{\mathcal{A}^{vn}}(\hat{A}_\omega)} \Pi(x)\delta_{\alpha_1}(dx) \\
&= \int_{\sigma_{\mathcal{A}}(a)} \alpha\mu_{\omega'}^{(a)}(d\alpha)
\end{aligned}$$

and so $\mu_{\omega'}^{(a)}(d\alpha) = \delta_{\alpha_1}(d\alpha)$ with $\alpha_1 = f_1(\gamma) \in \sigma_{\mathcal{A}^{vn}}(\hat{A}_\omega) \subset \sigma_{\mathcal{A}}(a)$. The same holds for b setting $f = f_2$. Thus for the same state, we have two delta-like probability measures for commuting observables: this is simply a proof of the existence of dispersion free states for abelian C^*-algebra.

STEP 4: $\inf_\omega[H_\omega(a,\varepsilon) + H_\omega(b,\delta)] = 0 \Leftrightarrow [a,b] = 0$.

Let ω' be the state found in STEP 3 and $\varepsilon = \{E_i\}_{i\in I}$, $\delta = \{F_j\}_{j\in J}$ two partitions. Using ω', clearly $C(\varepsilon,\delta) := \inf_\omega[H_\omega(a,\varepsilon) + H_\omega(b,\delta)] = 0$ for any (ε,δ)-partition. Suppose that $C(\varepsilon,\delta) = 0$ even if $[a,b] \neq 0$ for any (ε,δ)-partition. If this is possible, since the Shannon entropy is always non-negative, the only possibility to have $C(\varepsilon,\delta) = 0$ is to have a delta-like spectral measure for both a and b. In the state where this happens we have a common probability space for a and b (i.e. there exists a joint PVM). But this contradicts theorem 19. Note that we have no contradiction, if $C(\varepsilon,\delta) = 0$ just for some (ε,δ)-partition. Indeed, we can always have $C(\varepsilon,\delta) = 0$ for any partition where the supports of the induced probability measures is completely contained in exactly one set of the partition, if this happens for both a and b, namely, if $\text{supp}\mu_\omega^{(a)} \subset E_i \in \varepsilon$ and $\text{supp}\mu_\omega^{(b)} \subset F_j \in \delta$, then $C(\varepsilon,\delta) = 0$. Thus we can say that if for some partition (ε,δ)

$$H_\omega(a;\varepsilon) + H_\omega(b;\delta) \geqslant C(\varepsilon,\delta) > 0$$

then $[a,b] \neq 0$, and this concludes the proof. □

Proof of the corollary 1

Proof. Let ω be a mixed state, hence it can be written as $\lambda\omega_1 + (1-\lambda)\omega_2$, for $\lambda \in (0,1)$. If $a \in \mathcal{A}$ is a self-adjoint element of the algebra and ε is a given

partition, from the definition of $p_i^{\omega,\varepsilon}$, we can see that:

$$p_i^{\omega,\varepsilon} = \lambda p_i^{\omega_1,\varepsilon} + (1-\lambda) p_i^{\omega_2,\varepsilon}.$$

Because the entropy is a concave function, we can write that $H_\omega(a;\varepsilon) \geqslant \lambda H_{\omega_1}(a;\varepsilon) + (1-\lambda) H_{\omega_2}(a;\varepsilon)$. Hence if a and b fulfil an entropic uncertainty relation for pure states, then it holds also for mixed states, with the same constant $C(\varepsilon,\delta)$. Indeed:

$$\begin{aligned} H_\omega(a;\varepsilon) + H_\omega(b;\delta) &\geqslant \lambda[H_{\omega_1}(a;\varepsilon) + H_{\omega_1}(b;\delta)] \\ &\quad + (1-\lambda)[H_{\omega_2}(a;\varepsilon) + H_{\omega_2}(b;\delta)] \\ &\geqslant \lambda C(\varepsilon,\delta) + (1-\lambda) C(\varepsilon,\delta) = C(\varepsilon,\delta) \end{aligned}$$

which concludes the proof. □

POVM and non-simmetrically conditioned observables

Let A and B be two correlated random variables on a common probability space $(\Omega, \mathcal{E}, P)$ and assume we use some conditional probabilities P_c to describe them: in this way we have a contextual probability space. We also require that they fulfill conditions *ii*) and *iii*) of section 3.2.3 so that we expect that they are described after conditioning using a non-abelian probability space. We want to study which kind of representation for these two random variables is possible on a common Hilbert space when they are non-symmetrically conditioned, namely when

$$p(a|b) \neq p(b|a).$$

When this condition hold we cannot use the transition probabilities to define the square modulus of the scalar product on a common Hilbert space $\mathcal{H}$. Basically, the reason for that is the need to define two scalar products on the same vector space: one such that $|\langle a|b\rangle_1|^2 = p(a|b)$ and one such that $|\langle a|b\rangle_2|^2 = p(b|a)$. In this way one defines two Hilbert spaces on which both the random variables may be represented as operators. Note that a condition needed in order to have a good definition of scalar product is the *double stochasticity* of the transition probabilities, namely

$$\sum_b p(a|b) = \sum_a p(b|a) = 1. \tag{8.1}$$

In the case of binary random variable this implies symmetrically conditioning [47], while for random variables having three or more possible outcomes this is not always true. Here we will show that when $p(a|b) \neq p(b|a)$, instead of using two Hilbert spaces, it is possible to have a representation of the random variables on a single Hilbert space using POVM.

Consider two POVMs on some finite dimensional Hilbert space $\mathcal{H}$. In particular we consider the following POVMs:

$$\begin{aligned} \hat{M}_a &= \sum_{a'} \lambda(a|a')\hat{P}_{a'} \\ \hat{N}_b &= \sum_{b'} \mu(b|b')\hat{Q}_{b'} \end{aligned} \tag{8.2}$$

where $\hat{P}_{a'} = |a'\rangle\langle a'|$, $\hat{Q}_{b'} = |b'\rangle\langle b'|$. $\lambda(a|a')$ and $\mu(b|b')$ are two real and positive numbers such that

$$\sum_a \lambda(a|a') = 1, \qquad \sum_b \mu(b|b') = 1. \tag{8.3}$$

It is know that POVM are tools used to describe realistic (non-ideal) measurement since they are able to describe also the uncertainty due to a lack of efficiency of the experimental apparatus [92, 36]. According to the formalism used in quantum mechanics to describes non-ideal measurements, given a the quantum state $|\psi\rangle$, the probability to measure $B = b$ is given by

$$P(b||\psi\rangle) = \langle\psi|\hat{N}_b|\psi\rangle, \tag{8.4}$$

and a similar equation holds for $P(a||\psi\rangle)$ using $\hat{M}_a$. From this follows that if $|\psi\rangle = |a\rangle$, then

$$\begin{aligned} P(b||a\rangle) &= \sum_{b'} \mu(b|b')\langle a|b'\rangle\langle b'|a\rangle \\ &= \sum_{b'} \mu(b|b')\tilde{p}(a|b') \end{aligned} \tag{8.5}$$

where we set $\tilde{p}(a|b) = |\langle a|b\rangle|^2$. Note that these transition probabilities are different from $p(a|b)$ and $p(b|a)$. In particular they are symmetrically conditioned and double stochastic by construction. We will discuss them in detail later. Similarly for $|\psi\rangle = |b\rangle$ we get

$$P(a||b\rangle) = \sum_{a'} \lambda(a|a')\tilde{p}(a'|b). \tag{8.6}$$

It is not difficult to see that $P(a||b\rangle)$ can be different from $P(b||a\rangle)$, provided that the λs and the μs are different or they do not have both delta-like shape. Hence we can set them equal to $p(a|b)$ and $p(b|a)$, showing that there is the possibility to represent two random variables on a single Hilbert space even if they are not symmetrically conditioned.

Let us now discuss how to choose the transition probabilities $\tilde{p}(a|b)$. In general they are arbitrary. However, if we require that $\tilde{p}(a|b)$ have to be calculated from $p(a|b)$ and $p(b|a)$, they are uniquely determinate. To keep the discussion

simple we consider the case of A and B binary random variables. Let us assume for the moment that $p(a|b)$ and $p(b|a)$ are doubly stochastic. Then in this case we can chose

$$\tilde{p}(a|b) = \frac{p(a|b) + p(b|a)}{2}, \tag{8.7}$$

which are symmetric by construction and normalized because double stochasticity. Note that this is the unique symmetric function of $p(a|b)$ and $p(b|a)$ having the properties of a transition probability, i.e. normalization when we sum over a or b. Let us now see what happens if $p(a|b)$ and $p(b|a)$ are not double stochastic. In this case, a naive application of (8.7) would not give normalized probability. To avoid this problem one can proceed as follow. First one constructs the matrices

$$\mathbf{P}(a|b) = \begin{bmatrix} p(a_1|b_1) & p(a_2|b_1) \\ p(a_1|b_2) & p(a_2|b_2) \end{bmatrix} \tag{8.8}$$

and

$$\mathbf{P}(b|a) = \begin{bmatrix} p(b_1|a_1) & p(b_2|a_1) \\ p(b_1|a_2) & p(b_2|a_2) \end{bmatrix} \tag{8.9}$$

These are positive matrices which are only stochastic (i.e. only the rows sum up to 1). At this point using the Sinkhorn's theorem [93, 94], one can find two diagonal matrix $\mathbf{D}_1$ and $\mathbf{D}_2$ (up to a multiplicative constant) such that the matrix $\mathbf{Q}(a|b) = \mathbf{D}_1\mathbf{P}(a|b)\mathbf{D}_2$ is double stochastic. $\mathbf{Q}(a|b)$ is unique (up to a multiplicative constant). Similarly from $\mathbf{P}(\mathbf{b}|\mathbf{a})$ one can find a double stochastic matrix $\mathbf{Q}(b|a)$. Using the entries of these matrices in (8.7), one can define $\tilde{p}(a|b)$ and define the (square modulus) of the scalar product on the vector space. One obtain in this way a single Hilbert space on which both the POVM are defined. All generalize straightforwardly to the case of n different outcomes, and using functional analysis, also to the continuous case [95].

The only thing that remains to check is if it is really possible to construct the POVMs $\hat{M}_a$ and $\hat{N}_b$. Let us focus on $\hat{M}_a$ (for $\hat{N}_b$ what follows can be applied mutatis mutandis). We have to show that there exist the λs such that (8.6) holds for any choices of $p(a|b)$ and $\tilde{p}(a|b)$. Again we consider first the case of binary random variable. By construction $\hat{M}_{a_2} = \hat{\mathbb{I}} - \hat{M}_{a_1}$, hence we may focus only on $\hat{M}_{a_1}$. To define this POVM, we need to show that the following problem as at least a solution: find $x, y \in [0, 1]$ such that

$$p(a_1|b) = y\tilde{p}(a_1|b) + x\tilde{p}(a_2|b), \tag{8.10}$$

for some b. First consider the case where at least one of the two $\tilde{p}$ is bigger than $p(a_1|b)$, say $\tilde{p}(a_1|b)$. In this case we may set

$$y = \frac{p(a_1|b)}{\tilde{p}(a_1|b)} < 1, \quad \text{and} \quad x = 0,$$

which are a solution of (8.10) fulfilling $x, y \in [0, 1]$. Now consider the case where both the $\tilde{p}$ are less than $p(a_1|b)$. In particular we assume that

$$p(a_1|b) \geqslant \tilde{p}(a_1|b) \geqslant \tilde{p}(a_2|b). \tag{8.11}$$

Note that this condition is not restrictive. Indeed, if $\tilde{p}(a_2|b) \geqslant \tilde{p}(a_1|b)$ one simply exchange x and y. From (8.10) we have that

$$x = \frac{p(a_1|b)}{\tilde{p}(a_2|b)} - y\frac{\tilde{p}(a_1|b)}{\tilde{p}(a_2|b)}.$$

We have $x \geqslant 0$ when $p(a|b) \geqslant y\tilde{p}(a_1|b)$, which is always true for any $y \in [0,1]$ because of (8.11). On the other hand we have $x \leqslant 1$ for

$$y \geqslant \frac{p(a_1|b) - \tilde{p}(a_2|b)}{\tilde{p}(a_1|b)}.$$

The RHS is less or equal to 1 for any $p(a_1|b) \leqslant 1$, which is always the case, while it is bigger than zero only when $p(a_1|b) \geqslant \tilde{p}(a_2|b)$, which is true because of assumption (8.11). This show that there exist an $y \in [0,1]$ solving (8.10) for any $x \in [0,1]$. It is not difficult to see that x, y are not uniquely determined. Hence there can be different POVMs leading to the same transition probability $p(a|b)$. Let us now consider the case of random variables with $n > 2$ outcomes. The previous proof does not generalize in a straightforward manner. However one can observe that (8.6), defines a Markov chain where the matrix with entries $\lambda(a_i|a_j)$ is the associated transition matrix. This matrix can be estimated [96, 97] from a numerical simulation of the initial and final state of the Markov chain. Regarding the continuous case, since each measurable function can be approximated using simple functions [3], there are indications that the result may hold also in this case.

This shows that it is possible to represent on the same Hilbert space two random variables that are neither symmetrically conditioned nor have double stochastic transition probabilities using POVM.

Proof of the proposition 1

Proof. Since $V_N = X_{N+1} - X_N$, clearly X_{N+1} and X_N are conditionally independent under the event $A = \{X_N = a\}$. Let $\varphi_{V_N}(\lambda)|_A$, $\varphi_{X_{N+1}}(\lambda)|_A$ and $\varphi_{X_N}(\lambda)|_A$ be the characteristic functions of the three random variables considered here, computed with the conditional probabilities. By conditional independence we can write that

$$\varphi_{V_N}(\lambda)|_A = \varphi_{X_{N+1}}(\lambda)|_A \cdot \varphi_{-X_N}(\lambda)|_A.$$

Since

$$\varphi_{-X_N}(\lambda)|_A = \sum_{a'} e^{-i\lambda a'}\delta_{a,a'} = e^{-i\lambda a}$$

$$\varphi_{X_{N+1}}(\lambda)|_A = \sum_{b} \alpha(b,a)e^{i\lambda b}$$

we have that

$$\varphi_{V_N}(\lambda)|_A = \sum_{b} \alpha(b,a)e^{i\lambda(b-a)}.$$

Because $b - a \in \mathbb{Z}$, clearly $\varphi_{V_N}(\lambda)|_A = \varphi_{V_N}(\lambda + 2\pi)|_A$ which means that the random variable V_N is a discrete random variable (as expected). The inversion formula of the characteristic function, in this case is

$$P[V_N = c|A] = \lim_{T\to+\infty} \frac{1}{2T} \int_{-T}^{+T} e^{-i\lambda c} \varphi_{V_N}(\lambda)|_A d\lambda.$$

Thus

$$\begin{aligned} P[V_N = c|A] &= \lim_{T\to+\infty} \frac{1}{2T} \int_{-T}^{+T} e^{-i\lambda c} \sum_b \alpha(b,a) e^{i\lambda(b-a)} d\lambda \\ &= \sum_b \alpha(b,a) \lim_{T\to+\infty} \frac{1}{2T} \int_{-T}^{+T} e^{i\lambda(b-a-c)} d\lambda \\ &= \sum_b \alpha(b,a) \lim_{T\to+\infty} \frac{e^{iT(b-a-c)} - e^{iT(b-a-c)}}{2Ti(b-a-c)} \\ &= \sum_b \alpha(b,a) \lim_{T\to+\infty} \operatorname{sinc}(T(b-a-c)) \end{aligned}$$

where $\operatorname{sinc}(x) = \sin x/x$. Since $\lim_{a\to\infty} \operatorname{sinc}(ax) = \delta_{x,0}$ when $x \in \mathbb{Z}$, we conclude that

$$P[V_N = c|A] = \sum_b \alpha(b,a)\delta_{b-a-c,0} = \alpha(a+c,a).$$

This concludes the proof. □

Proof of theorem 11

Proof. Given $P[V_N = c]$, we can always write

$$P[V_N = c] = \sum_{\mathbf{S}'_N} P[V_N = c, \mathbb{S}_N^A = \mathbf{S}'_N]$$

and similarly

$$P[X_N = a] = \sum_{\mathbf{S}'_N} P[X_N = a, \mathbb{S}_N^A = \mathbf{S}'_N].$$

Note that the sum over all possible configurations of the space process at time N is well defined, since the number of configurations is clearly countable (it is a cartesian product of a discrete process taking value on the integers). Substi-

tuting these expressions in (4.10) and dividing by $P[\mathbb{S}_N^A = \mathbf{S}_N]$, we get

$$\sum_{\mathbf{S}'_N} \frac{P[V_N = c, \mathbb{S}_N^A = \mathbf{S}'_N]}{P[\mathbb{S}_N^A = \mathbf{S}_N]} = \sum_a \alpha(a+c,c) \sum_{\mathbf{S}'_N} \frac{P[X_N = a, \mathbb{S}_N^A = \mathbf{S}'_N]}{P[\mathbb{S}_N^A = \mathbf{S}_N]}$$

$$P_{\mathbf{S}_N}[V_N = c] + \sum_{\substack{\mathbf{S}'_N \\ \mathbf{S}'_N \neq \mathbf{S}_N}} \frac{P[V_N = c, \mathbb{S}_N^A = \mathbf{S}'_N]}{P[\mathbb{S}_N^A = \mathbf{S}_N]} = \sum_a \alpha(a+c,c) P_{\mathbf{S}_N}[X_N = a] +$$

$$+ \sum_a \alpha(a+c,c) \sum_{\substack{\mathbf{S}'_N \\ \mathbf{S}'_N \neq \mathbf{S}_N}} \frac{P[X_N = a, \mathbb{S}_N^A = \mathbf{S}'_N]}{P[\mathbb{S}_N^A = \mathbf{S}_N]}$$

Moving the second term of the LHS to the RHS, we obtain (4.12) and (4.13). Note that in general (4.13) is non zero since

$$\sum_a \alpha(a+c,a) P[X_N = a, \mathbb{S}_N^A = \mathbf{S}'_N] \neq P[V_N = c, \mathbb{S}_N^A = \mathbf{S}'_N].$$

This concludes the proof. □

Proof of theorem 12

Proof. The entropy is a non-negative quantity by definition, hence varying with respect to all $P_{\mathbf{S}_N}[X_N = a]$ clearly $H_{\mathbf{S}_N}(X_N) \geqslant 0$. Now consider the entropy for the random variable V_N and let us study what happens when we vary with respect to all $P_{\mathbf{S}_N}[X_N = a]$. Given $P_{\mathbf{S}_N}[X_N = a]$, the probability $P_{V_N}[V_N = c]$ can be computed by means of the formula in theorem 11. On the other hand we are always free to use the conditional transition probabilities $\alpha_{\mathbf{S}_N}(c,a)$, i.e. to work on the joint probability space of X_N and V_N, to study how $H_{\mathbf{S}_N}(V_N)$ change varying with respect to $P_{\mathbf{S}_N}[X_N = a]$. This allow us to write

$$H_{\mathbf{S}_N}(V_N) \geqslant \sum_a P_{\mathbf{S}_N}[X_N = a] H_{\mathbf{S}_N}(V_N | X_N = a),$$

with

$$H_{\mathbf{S}_N}(V_N | X_N = a) = -\sum_c \alpha_{\mathbf{S}_N}(c,a) \log \alpha_{\mathbf{S}_N}(c,a).$$

The conditional transition probabilities $\alpha_{\mathbf{S}_N}(c,a)$ can be rewritten as follows. Consider the joint probability $P_{\mathbf{S}_N}[X_{N+1} = b, X_N = a]$. In what follows, with-

out loss of generality we set $S_N^{io} = 0$ at any time N. We can write the following

$$\begin{aligned}
P_{\mathbf{S}_N}[X_{N+1} = b, X_N = a] &= \sum_{i=1}^{M} P_{\mathbf{S}_N}[X_{N+1} = S_{N+1}^{(i)} | S_{N+1}^{(i)} = b, X_N = a] P_{\mathbf{S}_N}[S_{N+1}^{(i)} = b, X_N = a] \\
&= \sum_{i=1}^{M} P_{\mathbf{S}_N}[X_{N+1} = S_{N+1}^{(i)} | S_{N+1}^{(i)} = b, X_N = a] \cdot \\
&\quad \cdot \left(\sum_{j=1}^{M} P_{\mathbf{S}_N}[S_{N+1}^{(i)} = b | X_N = a, I_N = j] P_{\mathbf{S}_N}[X_N = a, I_N = j] \right) \\
&= \sum_{i=1}^{M} P_{\mathbf{S}_N}[X_{N+1} = S_{N+1}^{(i)} | S_{N+1}^{(i)} = b, X_N = a] \cdot \\
&\quad \cdot \left(\sum_{j=1}^{M} P_{\mathbf{S}_N}[S_{N+1}^{(i)} = b | X_N = a, I_N = j] P_{\mathbf{S}_N}[I_N = j | X_N = a] P_{\mathbf{S}_N}[X_N = a] \right)
\end{aligned}$$

Since the event $\{X_N = a\} \cap \{I_N = j\} = \{S_N^{(j)} = a\}$ by definition and because $P_{\mathbf{S}_N}[X_{N+1} = a + c, X_N = a] = P_{\mathbf{S}_N}[V_N = c, X_N = a]$, from this decomposition we can conclude that

$$\begin{aligned}
\alpha_{\mathbf{S}_N}(c, a) = \sum_{i=1}^{M} & P_{\mathbf{S}_N}[X_{N+1} = S_{N+1}^{(i)} | S_{N+1}^{(i)} = b, X_N = a] \cdot \\
& \cdot \left(\sum_{j=1}^{M} P_{\mathbf{S}_N}[S_{N+1}^{(i)} = b | S_N^{(j)} = a] P_{\mathbf{S}_N}[I_N = j | X_N = a] \right).
\end{aligned}$$

We also note that

$$\begin{aligned}
\sum_{i=1}^{M} P_{\mathbf{S}_N}[X_{N+1} = S_{N+1}^{(i)} | S_{N+1}^{(i)} = b, X_N = a] = 1, \\
\sum_{j=1}^{M} P_{\mathbf{S}_N}[I_N = j | X_N = a] = 1.
\end{aligned} \tag{8.12}$$

In what follows, we set $\gamma(i) := P_{\mathbf{S}_N}[X_{N+1} = S_{N+1}^{(i)} | S_{N+1}^{(i)} = b, X_N = a]$ and $\eta(i, j) := P_{\mathbf{S}_N}[S_{N+1}^{(i)} = b | S_N^{(j)} = a]$ in order to keep the notation compact. From the above decomposition of $\alpha_{\mathbf{S}_N}(c, a)$ we can write that

$$\begin{aligned}
H_{\mathbf{S}_N}(V_N | X_N = a) = -\sum_{c} & \left(\sum_{i=1}^{M} \gamma(i) \left(\sum_{j=1}^{M} \eta(i, j) P_{\mathbf{S}_N}[I_N = j | X_N = a] \right) \right) \cdot \\
& \cdot \log \left(\sum_{i=1}^{M} \gamma(i) \left(\sum_{j=1}^{M} \eta(i, j) P_{\mathbf{S}_N}[I_N = j | X_N = a] \right) \right).
\end{aligned}$$

Note that since only positive probabilities contribute to the entropy, all the $\alpha_{\mathbf{S}_N}(c,a)$ are different from zero. This implies that all the $\gamma(i)$, $\eta(i,j)$ and $P_{\mathbf{S}_N}[I_N = j|X_N = a]$ used to compute the entropy are strictly positive. Since $f(x) = -x\log x$ is a concave function, by the Jensen inequality and using (8.12), we have

$$\begin{aligned}
H_{\mathbf{S}_N}(V_N|X_N = a) &\geqslant \sum_c \sum_{i=1}^M \gamma(i)\bigg(- \bigg(\sum_{j=1}^M \eta(i,j) P_{\mathbf{S}_N}[I_N = j|X_N = a]\bigg)\cdot \\
&\qquad \cdot \log\bigg(\sum_{j=1}^M \eta(i,j) P_{\mathbf{S}_N}[I_N = j|X_N = a]\bigg)\bigg) \\
&\geqslant \sum_c \bigg(\sum_{i=1}^M \gamma(i)\bigg) \min_i \bigg(- \bigg(\sum_{j=1}^M \eta(i,j) P_{\mathbf{S}_N}[I_N = j|X_N = a]\bigg)\cdot \\
&\qquad \cdot \log\bigg(\sum_{j=1}^M \eta(i,j) P_{\mathbf{S}_N}[I_N = j|X_N = a]\bigg)\bigg) \\
&\geqslant \sum_c \min_i \Bigg(\bigg(\sum_{j=1}^M P_{\mathbf{S}_N}[I_N = j|X_N = a]\bigg) \min_j \left(-\eta(i,j)\log\eta(i,j)\right)\Bigg) \\
&= \sum_c \min_{i,j} \left(-\eta(i,j)\log\eta(i,j)\right)
\end{aligned}$$

where $\min_{i,j}$ means the minimum over $i,j \in \{1,\cdots,M\}$ keeping c constant. Summarising, we have that

$$H(V_N|X_N = a) \geqslant \sum_c \min_{i,j} \left(-P_{\mathbf{S}_N}[S^{(i)}_{N+1} = a + c|S^{(j)}_N = a]\log P_{\mathbf{S}_N}[S^{(i)}_{N+1} = a + c|S^{(j)}_N = a]\right).$$

Note that in the RHS there is still a dependence on a, which can be removed by taking the minimum with respect to it. Thus we can write that

$$H_{\mathbf{S}_N}(V_N) \geqslant \sum_a P_{\mathbf{S}_N}[X_N = a] H(V_N|X_N = a) \geqslant D_1,$$

where we set

$$D_1 := \min_a \left[\sum_c \min_{ij} \left(- P_{\mathbf{S}_N}[S^{(i)}_{N+1} = a + c|S^{(j)}_N = a]\log P_{\mathbf{S}_N}[S^{(i)}_{N+1} = a + c|S^{(j)}_N = a]\right)\right]. \tag{8.13}$$

D_1 is a positive number, since $P_{\mathbf{S}_N}[S^{(i)}_{N+1} = a + c|S^{(j)}_N = a] \in (0,1)$ (we exclude the case of *deterministic* space process) and only positive probabilities contribute to the entropy, as said above. To explicitly show that the D_1 does not depend on $P_{\mathbf{S}_N}[X_N = a]$ and $P_{\mathbf{S}_N}[V_N = c]$, let us study in detail the terms $P_{\mathbf{S}_N}[S^{(i)}_{N+1} = a + c|S^{(j)}_N = a]$. Recalling that the random walks are independent

and that $P_{\mathbf{S}_N}[S^{(i)}_{N+1} = a + c | S^{(j)}_N = a] \neq 0$ only for the $S^{(j)}_N \in \mathbf{S}_N$, we can write that

$$P_{\mathbf{S}_N}[S^{(i)}_{N+1} = a + c | S^{(j)}_N = a] = \begin{cases} 0 \text{ if } i = j \text{ and } c \neq \pm 1; \\ p_i \text{ if } i = j \text{ and } c = 1; \\ 1 - p_i \text{ if } i = j \text{ and } c = -1; \\ P_{\mathbf{S}_N}[S^{(i)}_{N+1} = a + c] \text{ if } i \neq j \end{cases}$$

where p_i and $1 - p_i$ are the transition probabilities of the i-th random walk, which are fixed by hypothesis. Again, the first case is excluded since only positive probabilities contribute to the entropy. What we need to check is the last case, namely $P_{\mathbf{S}_N}[S^{(i)}_{N+1} = a + c]$. Since in the configuration $\mathbf{S}_N$ there is also the i-th random walk, this term reduces to

$$P_{\mathbf{S}_N}[S^{(i)}_{N+1} = a + c] = P[S^{(i)}_{N+1} = a + c | S^{(i)}_N = e]$$

for some $e \in \mathbb{Z}$. The only terms of this kind that contribute to the entropy are those having $e = a + c \pm 1$, i.e. the transition probabilities of the i-th random walks, which are fixed by hypothesis. Thus fixing p_i for all $i \in I$ implies that D_1 is a positive constant. Summarising we showed that

$$H_{\mathbf{S}_N}(X_N) + H_{\mathbf{S}_N}(V_N) \geqslant D_1,$$

when we vary over any possible value of $P_{\mathbf{S}_N}[X_N = a]$ and when the transition probabilities of the M random walks are fixed.

To conclude the proof we need to study what happens when we vary over all possible values of $P_{\mathbf{S}_N}[V_N = c]$. Similarly to the previous case, $H_{\mathbf{S}_N}(V_N) \geqslant 0$ while $H_{\mathbf{S}_N}(X_N)$ changes according with the inequality

$$H_{\mathbf{S}_N}(V_N) \geqslant \sum_c P_{\mathbf{S}_N}[V_N = c] H_{\mathbf{S}_N}(X_N | V_N = c),$$

where $H_{\mathbf{S}_N}(X_N | V_N = c)$ is the entropy computed using $\alpha_{\mathbf{S}_N}(a, c) = P_{\mathbf{S}_N}[X_N = a | V_N = c]$. From the definition of X_N and V_N, one can conclude that

$$\{X_N = a\} \cap \{V_N = c\} = \{X_N = a\} \cap \{V_N = c\} \cap \{X_{N+1} = a+c\} = \{V_N = c\} \cap \{X_{N+1} = a+c\}.$$

This implies that $P_{\mathbf{S}_N}[X_N = a, V_N = c] = P_{\mathbf{S}_N}[X_{N+1} = a + c, V_N = c]$, i.e.

$$\begin{aligned} \alpha_{\mathbf{S}_N}(a, c) = P_{\mathbf{S}_N}[X_N = a | V_N = c] &= \frac{P_{\mathbf{S}_N}[X_N = a, V_N = c]}{P_{\mathbf{S}_N}[V_N = c]} \\ &= \frac{P_{\mathbf{S}_N}[X_{N+1} = a + c, V_N = c]}{P_{\mathbf{S}_N}[V_N = c]}. \end{aligned}$$

As before, the whole analysis reduces to the study of this term. Given $P_{\mathbf{S}_N}[X_{N+1} =$

$a+c, V_N = c]$ we can write that

$$
\begin{aligned}
&P_{\mathbf{S}_N}[X_{N+1} = a+c, V_N = c] = \\
&\quad = \sum_{i=1}^{M} P_{\mathbf{S}_N}[X_{N+1} = S_{N+1}^{(i)} | S_N^{(i)} = a+c, V_N = c] P[S_{N+1}^{(i)} = a+c, V_N = c] \\
&\quad = \sum_{i=1}^{M} P_{\mathbf{S}_N}[X_{N+1} = S_{N+1}^{(i)} | S_{N+1}^{(i)} = a+c, V_N = c] \cdot \\
&\qquad \cdot \Big(\sum_{j,d} P_{\mathbf{S}_N}[S_{N+1}^{(i)} = a+c | V_N = c, I_N = j, X_{N+1} = d] P_{\mathbf{S}_N}[V_N = c, I_N = j, X_{N+1} = d] \Big) \\
&\quad = \sum_{i=1}^{M} P_{\mathbf{S}_N}[X_{N+1} = S_{N+1}^{(i)} | S_{N+1}^{(i)} = a+c, V_N = c] \cdot \\
&\qquad \cdot \Big(\sum_{j,d} P_{\mathbf{S}_N}[S_{N+1}^{(i)} = a+c | V_N = c, I_N = j, X_{N+1} = d] \cdot \\
&\qquad\quad \cdot P_{\mathbf{S}_N}[I_N = j, X_{N+1} = d | V_N = c] P_{\mathbf{S}_N}[V_N = c] \Big).
\end{aligned}
$$

Observing that the event $\{V_N = c\} \cap \{I_N = j\} \cap \{X_{N+1} = d\} = \{S_N^{(j)} = d - c\}$, we conclude that

$$
\begin{aligned}
\alpha_{\mathbf{S}_N}(a,c) = \sum_{i=1}^{M} & P_{\mathbf{S}_N}[X_{N+1} = S_{N+1}^{(i)} | S_{N+1}^{(i)} = a+c, V_N = c] \cdot \\
& \cdot \Big(\sum_{j,d} P_{\mathbf{S}_N}[S_{N+1}^{(i)} = a+c | S_N^{(j)} = d-c] P_{\mathbf{S}_N}[I_N = j, X_{N+1} = d | V_N = c] \Big).
\end{aligned}
$$

Note that

$$
\begin{aligned}
\sum_{i=1}^{M} P_{\mathbf{S}_N}[X_{N+1} = S_{N+1}^{(i)} | S_{N+1}^{(i)} = a+c, V_N = c] = 1 \\
\sum_{j,d} P_{\mathbf{S}_N}[I_N = j, X_{N+1} = d | V_N = c] = 1
\end{aligned}
\tag{8.14}
$$

Defining $\tilde{\gamma}(i) := P_{\mathbf{S}_N}[X_{N+1} = S_{N+1}^{(i)} | S_{N+1}^{(i)} = a+c, V_N = c]$ and $\tilde{\eta}(i,j) := P_{\mathbf{S}_N}[S_{N+1}^{(i)} = a+c | S_N^{(j)} = d-c]$, the whole analysis done in the previous case can be repeated. One has simply to replace $\gamma(i)$ with $\tilde{\gamma}(i)$, $\eta(i,j)$ with $\tilde{\eta}(i,j)$ and use (8.14) instead of (8.12), obtaining

$$
\begin{aligned}
H_{\mathbf{S}_N}(X_N | V_N = c) \geqslant \sum_a \min_{i,j,d} \Big(& - P_{\mathbf{S}_N}[S_{N+1}^{(i)} = a+c | S_N^{(j)} = d-c] \cdot \\
& \cdot \log P_{\mathbf{S}_N}[S_{N+1}^{(i)} = a+c | S_N^{(j)} = d-c] \Big).
\end{aligned}
$$

Setting

$$D_2 := \min_c \left[\sum_a \min_{i,j,d} \left(-P_{\mathbf{S}_N}[S_{N+1}^{(i)} = a+c | S_N^{(j)} = d-c] \log P_{\mathbf{S}_N}[S_{N+1}^{(i)} = a+c | S_N^{(j)} = d-c] \right) \right]$$

which is a positive constant, we conclude that

$$H_{\mathbf{S}_N}(X_N) + H_{\mathbf{S}_N}(V_N) \geqslant D_2,$$

when we vary over any possible value of $P_{\mathbf{S}_N}[V_N = c]$ and when the transition probabilities of the M random walks are fixed. Setting $D := \min\{D_1, D_2\}$ the statement of the theorem follows. This concludes the proof. □

Essentials of Dirichlet forms and its relation with Markov processes

Let $\mathcal{H}$ be an Hilbert space and consider a dense subspace $\mathcal{D} \subset \mathcal{H}$. A map $\varepsilon : \mathcal{D} \times \mathcal{D} \to \mathbb{R}$ is said *real bilinear form* if $\varepsilon(\alpha u + \beta v, z) = \alpha\varepsilon(u, z) + \beta\varepsilon(v, z)$, for any $u, v, z \in \mathcal{D}$, and $\alpha, \beta \in \mathbb{R}$. A real bilinear form is said *positive* if $\varepsilon(u, u) \geqslant 0$, for any $u \in \mathcal{D}$, and *symmetric* if $\varepsilon(u, v) = \varepsilon(v, u)$ for any $u, v \in \mathcal{D}$. Let $\langle \cdot | \cdot \rangle$ denote the scalar product on $\mathcal{H}$, and consider the bilinear form $(u, v) := \varepsilon(u, v) + \langle u | v \rangle$ which is defined for all $u, v \in \mathcal{D}$. When ε is symmetric and positive, the bilinear form $(\cdot, \cdot)$ is an inner product on $\mathcal{D}$. We say that the real symmetric positive bilinear form ε is *closed* if $\mathcal{D}$ is an Hilbert space, when equipped with the inner product $(\cdot, \cdot)$ defined above. Now we are ready to define a Dirichlet form.

Definition 41. *Let $(\mathsf{X}, \mathcal{X}, \mu)$ be a σ-finite measure space and set $\mathcal{H} = L_2(\mathsf{X}, \mu)$. Consider a real symmetric positive bilinear form $\varepsilon : \mathcal{D} \times \mathcal{D} \to \mathbb{R}$, where $\mathcal{D} \subset \mathcal{H}$ is dense, which is also closed. If for any $\epsilon > 0$, there exists a real function $\phi_\epsilon(x)$, $x \in \mathbb{R}$ with the following features*

i) $\phi_\epsilon(x) = x$ for $x \in [0, 1]$,

ii) $\phi_\epsilon(x) \in [-\epsilon, 1 + \epsilon]$ for any $x \in \mathbb{R}$,

iii) $\phi_\epsilon(x) - \phi_\epsilon(x') \in [0, x - x']$ whenever $x' < x$,

for which when $u \in \mathcal{D}$ then $\phi_\epsilon(x) \in \mathcal{D}$ and

$$\varepsilon(\phi_\varepsilon(u), \phi_\varepsilon(u)) \leqslant \varepsilon(u, u),$$

the couple $(\varepsilon, \mathcal{D})$ is said Dirichlet form *on $L_2(\mathsf{X}, \mu)$.*

For the rest of this appendix, we assume that X is a locally compact second countable metric space and μ is a Borel measure having support on the whole X. This abstract object is important because of the following theorem (Th 1.3.1, [87])

Theorem 21. *There is a one to one correspondence between a Dirichlet form $(\varepsilon, \mathcal{D})$ on $L_2(\mathsf{X}, \mu)$ and the family of non-positive definite self-adjoint operators on $L_2(\mathsf{X}, \mu)$. The correspondence is the following*

$$\varepsilon(u, v) = \langle [-\hat{H}]^{1/2} u | [-\hat{H}]^{1/2} v \rangle$$

and $D([-\hat{H}]^{1/2}) = \mathcal{D}$, where $D(\hat{H})$ is the domain of the operator $\hat{H}$.

The operator $\hat{H}$ is called *generator* of the Dirichlet form. Note that, because it is self-adjoint we can write

$$\varepsilon(u, v) = -\langle u | \hat{H} v \rangle$$

but note that $\mathrm{Dom}(\hat{H}) \subset \mathcal{D}$. Using such a generator, one can define the operator

$$\hat{T}_t := \exp(-t\hat{H}) \tag{8.15}$$

acting on $L_2(\mathsf{X}, \mu)$. Since $-\hat{H}$ is non-negative, $\hat{T}_t$ is always bounded. It also has the *semigroup property*, i.e. $\hat{T}_{t+s} = \hat{T}_t \hat{T}_s$, and it can be proved that it is also *strongly continuous* on $L_2(\mathsf{X}, \mu)$. The link between these objects and stochastic processes is encoded in the following theorem (Th. 1.4.1, [87]).

Theorem 22. *Let ε be a Dirichelt form on $L_2(\mathsf{X}, \mu)$ with generator $\hat{H}$. Then the operator* (8.15) *is a strongly continuous semigroup such that*

$$0 \leqslant \hat{T}_t u \leqslant 1 \qquad \mu\text{-a.s}$$

whenever $0 \leqslant u \leqslant 1$, μ-a.s., with $u \in L_2(\mathsf{X}, \mu)$.

At this point the connection with stochastic processes starts to appear. Given a Markov process, $\{X_t\}_{t \in \mathbb{R}^+}$ taking values on X with distribution μ_X, consider its transition probability density $p(x, t|s, y)$. Note that we are implicitly assuming that the transition probability admits a density, however the whole argument remains valid even in the general case. Given $p(x, t|s, y)$, we can define the following integral operator

$$\hat{S}_t f(x) := \int_{\mathsf{X}} f(y) p(x, t|s, y) dy,$$

where $f(x)$ is a bounded measurable function. More generally this operator is well defined any for $f \in L_2(\mathsf{X}, \mu_X)$. From the Markov property, the semigroup property follows, i.e. $\hat{S}_{t+s} = \hat{S}_t \hat{S}_s$. We note that $0 < \hat{S}_r f(x) < 1$ for all $x \in \mathsf{X}$ whenever $0 < f(x) < 1$ for all $x \in \mathsf{X}$, i.e. when $f(x)$ is a probability density. In addition $\hat{S}_t 1 = 1$, which is a consequence of the fact that $p(x, t|s, y)$ are transition probability densities. Finally, one can prove that $\hat{S}_t$ is strongly continuous in t when thought as a linear operator on $L_2(\mathsf{X}, \mu_X)$. Taking the generator of $\hat{S}_t$, i.e. the operator

$$-\hat{H}' := \lim_{t \to 0} \frac{\hat{S}_t f(x) - f(x)}{t}$$

where $f(x) \in L_2(\mathsf{X}, \mu_X)$, one can define a Dirichlet form $\varepsilon'(f,g) = -\langle f|\hat{H}'g\rangle$ on $L_2(\mathsf{X}, \mu_X)$. Thus one can study the properties of the Markov process $\{X_t\}_{t\in\mathbb{R}^+}$ using Dirichlet forms. Note that the opposite is not always true: for example in theorem 22, nothing is said on the condition $\hat{T}_t 1 = 1$ which clearly holds for a Markov process. Among the properties of $\{X_t\}_{t\in\mathbb{R}^+}$ that can studied using Dirichlet forms, there are also the path properties. In particular one can verify if the process is a diffusion or not. We conclude by saying that to be sure to obtain a Markov process, one needs to add other conditions on the Dirichlet form $(\varepsilon, \mathcal{D})$: for example, to obtain $\hat{T}_t 1 = 1$, one has to require that $1 \in \mathcal{D}$ and $\varepsilon(1,1) = 0$. For a detailed and complete discussion on the general relation between Dirichlet forms and Markov processes we refer to [87].

List of symbols used

$\mathcal{H}$, generic Hilbert space
$\mathcal{P}(\mathcal{H})$, set of all the orthogonal projectors on $\mathcal{H}$
$\mathcal{B}(\mathcal{H})$, set of all the bounded operators on $\mathcal{H}$
$\mathcal{B}_\infty(\mathcal{H})$, set of all the compact operators on $\mathcal{H}$
$\mathcal{B}_2(\mathcal{H})$, set of all the Hilbert-Schmidt operators on $\mathcal{H}$
$\mathcal{B}_1(\mathcal{H})$, set of all the trace-class operators on $\mathcal{H}$
$\dim A$, dimension on the vector space A

Bibliography

[1] Markus Arndt, Olaf Nairz, Julian Vos-Andreae, Claudia Keller, Gerbrand Van der Zouw, and Anton Zeilinger. Wave–particle duality of c60 molecules. *nature*, 401(6754):680–682, 1999.

[2] Frank Laloe Claude Cohen-Tannoudji, Bernard Dui. *Quantum mechanics*. Wiley-Vch, 1991.

[3] Valter Moretti. *Spectral theory and quantum mechanics: with an introduction to the algebraic formulation*. Springer Science & Business Media, 2013.

[4] Brian C Hall. *Quantum theory for mathematicians*, volume 267. Springer, 2013.

[5] Detlef Dürr and Stefan Teufel. Bohmian mechanics. In *Bohmian Mechanics*. Springer, 2009.

[6] Leslie E Ballentine. *Quantum mechanics: a modern development*. World Scientific, 1998.

[7] Michael A Nielsen and Isaac L Chuang. Quantum computation and quantum information (cambridge series on information and the natural sciences). 2004.

[8] Enrico G Beltrametti and Gianni Cassinelli. *The logic of quantum mechanics*, volume 15. Cambridge University Press, 2010.

[9] Joseph M Jauch and Richard A Morrow. Foundations of quantum mechanics. *American Journal of Physics*, 36(8):771–771, 1968.

[10] Garrett Birkhoff and John Von Neumann. The logic of quantum mechanics. *Annals of mathematics*, pages 823–843, 1936.

[11] Andrew M Gleason. Measures on the closed subspaces of a hilbert space. *Journal of mathematics and mechanics*, 6(6):885–893, 1957.

[12] Franco Strocchi. *An introduction to the mathematical structure of quantum mechanics: a short course for mathematicians*. World Scientific, 2005.

[13] Valter Moretti. Mathematical foundations of quantum mechanics: An advanced short course. *International Journal of Geometric Methods in Modern Physics*, 13(Supp. 1):1630011, 2016.

[14] Eduard Prugovecki. *Quantum mechanics in Hilbert space*, volume 92. Academic Press, 1982.

[15] Marshall H Stone. Linear transformations in hilbert space iii. operational methods and group theory. *Proceedings of the National Academy of Sciences*, 16(2):172–175, 1930.

[16] J v Neumann. Die eindeutigkeit der schrödingerschen operatoren. *Mathematische Annalen*, 104(1):570–578, 1931.

[17] J v Neumann. Uber einen satz von herrn mh stone. *Annals of Mathematics*, pages 567–573, 1932.

[18] George W Mackey et al. A theorem of stone and von neumann. *Duke Mathematical Journal*, 16(2):313–326, 1949.

[19] Diederik Aerts and Ingrid Daubechies. Physical justification for using the tensor product to describe two quantum systems as one joint system. *Helv. Phys. Acta*, 51(5-6), 1978.

[20] Chetan Nayak, Steven H Simon, Ady Stern, Michael Freedman, and Sankar Das Sarma. Non-abelian anyons and topological quantum computation. *Reviews of Modern Physics*, 80(3):1083, 2008.

[21] Luc Bouten, Ramon Van Handel, and Matthew R James. An introduction to quantum filtering. *SIAM Journal on Control and Optimization*, 46(6):2199–2241, 2007.

[22] Robin L Hudson and Kalyanapuram R Parthasarathy. Quantum ito's formula and stochastic evolutions. *Communications in Mathematical Physics*, 93(3):301–323, 1984.

[23] Luigi Accardi, Alberto Frigerio, and John T Lewis. Quantum stochastic processes. *Publications of the Research Institute for Mathematical Sciences*, 18(1):97–133, 1982.

[24] EB Davies. Quantum stochastic processes. *Communications in Mathematical Physics*, 15(4):277–304, 1969.

[25] Hans Maassen. Quantum probability theory (lecture notes). Available at `http://www.math.ru.nl/~maassen/lectures/qp.pdf` (20/04/2017), 1998.

[26] Karl E Petersen. *Ergodic theory*, volume 2. Cambridge University Press, 1989.

[27] Richard V Kadison and John R Ringrose. *Fundamentals of the Theory of Operator Algebras*. American Mathematical Soc., 2015.

[28] Arno Bohm and Manuel Gadella. Dirac kets, gamow vectors and gel'fand triplets. 1989.

[29] Miklós Rédei and Stephen Jeffrey Summers. Quantum probability theory. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 38(2):390–417, 2007.

[30] Luigi Accardi. Quantum probability. Available in italian at `https://art.torvergata.it/retrieve/handle/2108/42206/74620/Einaudi_Prob_quant%20100503.pdf` (20/04/2017), 2010.

[31] Dan V Voiculescu, Ken J Dykema, and Alexandru Nica. *Free random variables*. Number 1. American Mathematical Soc., 1992.

[32] Boris S Cirel'son. Quantum generalizations of bell's inequality. *Letters in Mathematical Physics*, 4(2):93–100, 1980.

[33] John F Clauser, Michael A Horne, Abner Shimony, and Richard A Holt. Proposed experiment to test local hidden-variable theories. *Physical review letters*, 23(15):880, 1969.

[34] Časlav Brukner and Anton Zeilinger. Conceptual inadequacy of the shannon information in quantum measurements. *Physical Review A*, 63(2):022113, 2001.

[35] Christopher G Timpson. On a supposed conceptual inadequacy of the shannon information in quantum mechanics. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 34(3):441–468, 2003.

[36] Michael A Nielsen and Isaac L Chuang. *Quantum computation and Quantum information*. Cambridge University Press India, 2000.

[37] Iwo Białynicki-Birula and Jerzy Mycielski. Uncertainty relations for information entropy in wave mechanics. *Communications in Mathematical Physics*, 44(2):129–132, 1975.

[38] Iwo Bialynicki-Birula. Formulation of the uncertainty relations in terms of the rényi entropies. *Physical Review A*, 74(5):052101, 2006.

[39] David Deutsch. Uncertainty in quantum measurements. *Physical Review Letters*, 50(9):631, 1983.

[40] M Hossein Partovi. Entropic formulation of uncertainty for quantum measurements. *Physical Review Letters*, 50(24):1883, 1983.

[41] Hans Maassen and Jos BM Uffink. Generalized entropic uncertainty relations. *Physical Review Letters*, 60(12):1103, 1988.

[42] M Krishna and KR Parthasarathy. An entropic uncertainty principle for quantum measurements. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 842–851, 2002.

[43] Gerald B Folland. *Real analysis: modern techniques and their applications*. John Wiley & Sons, 2013.

[44] Shunsuke Ihara. *Information theory for continuous systems*, volume 2. World Scientific, 1993.

[45] Curcuraci Luca. Entropy and non commutativity. Master thesis, university of Trento, 2015.

[46] Andrei Y Khrennikov. *Ubiquitous quantum structure*. Springer, 2014.

[47] Andrei Y Khrennikov. *Contextual approach to quantum formalism*, volume 160. Springer Science & Business Media, 2009.

[48] Andrei Khrennikov. *Probability and Randomness: Quantum versus Classical*. World Scientific, 2016.

[49] Andrei Khrennikov. Interference of probabilities in the classical probabilistic framework. *Fuzzy Sets and Systems*, 155(1):4–17, 2005.

[50] Alexander S Holevo. *Probabilistic and statistical aspects of quantum theory*, volume 1. Springer Science & Business Media, 2011.

[51] Diederik Aerts. Quantum structures: An attempt to explain the origin of their appearance in nature. *International Journal of Theoretical Physics*, 34(8):1165–1186, 1995.

[52] Andrei Yu Khrennikov. Interference in the classical probabilistic model and its representation in complex hilbert space. *Physica E: Low-dimensional Systems and Nanostructures*, 29(1-2):226–236, 2005.

[53] Peter Nyman and Irina Basieva. Quantum-like representation algorithm for trichotomous observables. *International journal of theoretical physics*, 50(12):3864–3881, 2011.

[54] Luca Curcuraci. On non-commutativity in quantum theory (ii): toy models for non-commutative kinematics. Available under request, 2017.

[55] Joseph Rudnick and George Gaspari. *Elements of the random walk: an introduction for advanced students and researchers*. Cambridge University Press, 2004.

[56] Bernt Oksendal. *Stochastic differential equations: an introduction with applications*. Springer Science & Business Media, 2013.

[57] Luca Curcuraci. On non-commutativity in quantum theory (i): from classical to quantum probability. Available under request, 2017.

[58] John B Conway. *A course in functional analysis*, volume 96. Springer Science & Business Media, 2013.

[59] Fima C Klebaner. *Introduction to stochastic calculus with applications*. World Scientific Publishing Co Inc, 2005.

[60] Gero Friesecke and Mario Koppen. On the ehrenfest theorem of quantum mechanics. *Journal of Mathematical Physics*, 50(8):082102, 2009.

[61] Adrian Baddeley, Imre Bárány, and Rolf Schneider. Spatial point processes and their applications. *Stochastic Geometry: Lectures given at the CIME Summer School held in Martina Franca, Italy, September 13–18, 2004*, pages 1–75, 2007.

[62] Daryl J Daley and David Vere-Jones. *An introduction to the theory of point processes*. Springer Science & Business Media, 2007.

[63] Odile Macchi. The coincidence approach to stochastic point processes. *Advances in Applied Probability*, 7(01):83–122, 1975.

[64] John Ben Hough, Manjunath Krishnapur, Yuval Peres, and Bálint Virág. *Zeros of Gaussian analytic functions and determinantal point processes*, volume 51. American Mathematical Society Providence, RI, 2009.

[65] J Ben Hough, Manjunath Krishnapur, Yuval Peres, Bálint Virág, et al. Determinantal processes and independence. *Probability surveys*, 3(206-229):9, 2006.

[66] Laurent Decreusefond, Ian Flint, Nicolas Privault, and Giovanni Luca Torrisi. Determinantal point processes. In *Stochastic Analysis for Poisson Point Processes*, pages 311–342. Springer, 2016.

[67] Alexander Soshnikov. Determinantal random point fields. *Russian Mathematical Surveys*, 55(5):923–975, 2000.

[68] Tomoyuki Shirai and Yoichiro Takahashi. Random point fields associated with certain fredholm determinants i: fermion, poisson and boson point processes. *Journal of Functional Analysis*, 205(2):414–463, 2003.

[69] Issa Karambal, Veerle Ledoux, Simon JA Malham, and Jitse Niesen. Introductory fredholm theory and computation. 2010.

[70] Arno BJ Kuijlaars. Lecture notes in: Riemann-hilbert problems and multiple orthogonal polynomials. `https://math.vanderbilt.edu/dept/conf/constructive2014/KuijlaarsNotes.pdf`, 2014.

[71] Hirofumi Osada. Dirichlet form approach to infinite-dimensional wiener processes with singular interactions. *Communications in mathematical physics*, 176(1):117–131, 1996.

[72] Hyun Jae Yoo. Dirichlet forms and diffusion processes for fermion random point fields. *Journal of Functional Analysis*, 219(1):143–160, 2005.

[73] Laurent Decreusefond, Ian Flint, Nicolas Privault, and Giovanni Luca Torrisi. Stochastic dynamics of determinantal processes by integration by parts. *arXiv preprint arXiv:1210.6109*, 2012.

[74] Valentine Bargmann. On unitary ray representations of continuous groups. *Annals of Mathematics*, pages 1–46, 1954.

[75] John S Bell. On the einstein podolsky rosen paradox, 1964.

[76] Matthew F Pusey, Jonathan Barrett, and Terry Rudolph. On the reality of the quantum state. *Nature Physics*, 8(6):475–478, 2012.

[77] Andrei Khrennikov. Bell as the copernicus of probability. *arXiv preprint arXiv:1412.6987*, 2014.

[78] J Kiukas and RF Werner. Maximal violation of bell inequalities by position measurements. *Journal of Mathematical Physics*, 51(7):072105, 2010.

[79] Michele Correggi and Giovanni Morchio. Quantum mechanics and stochastic mechanics for compatible observables at different times. *Annals of Physics*, 296(2):371–389, 2002.

[80] Carlton Frederick. Stochastic space-time and quantum theory. *Physical Review D*, 13(12):3183, 1976.

[81] Sisir Roy. *Statistical geometry and applications to microphysics and cosmology*, volume 92. Springer Science & Business Media, 2012.

[82] Luca Bombelli, Alejandro Corichi, and Oliver Winkler. Semiclassical quantum gravity: obtaining manifolds from graphs. *Classical and Quantum Gravity*, 26(24):245012, 2009.

[83] Luca Bombelli, Joohan Lee, David Meyer, and Rafael D Sorkin. Space-time as a causal set. *Physical review letters*, 59(5):521, 1987.

[84] Graham Brightwell and Ruth Gregory. Structure of random discrete spacetime. *Physical review letters*, 66(3):260, 1991.

[85] David D Reid. Manifold dimension of a causal set: Tests in conformally flat spacetimes. *Physical Review D*, 67(2):024034, 2003.

[86] S. Aerts. Probability Conservation and the State Determination Problem. In A. Khrennikov, editor, *Foundations of Probability and Physics*, pages 39–49, December 2001.

[87] Masatoshi Fukushima, Yoichi Oshima, and Masayoshi Takeda. *Dirichlet forms and symmetric Markov processes*, volume 19. Walter de Gruyter, 2010.

[88] Masao Nagasawa. *Schrödinger equations and diffusion theory*, volume 86. Birkhäuser, 2012.

[89] Masao Nagasawa. *Stochastic processes in quantum physics*, volume 94. Birkhäuser, 2012.

[90] Edward Nelson. Derivation of the schrödinger equation from newtonian mechanics. *Physical review*, 150(4):1079, 1966.

[91] Masao Nagasawa. Transformations of diffusion and schrödinger processes. *Probability Theory and Related Fields*, 82(1):109–136, 1989.

[92] Willem M De Muynck and Hans Martens. Nonideal quantum measurements, simultaneous measurement of incompatible observables, and the bell inequalities. In *Proceedings of the 3rd International Symposium Foundations of Quantum Mechanics in the Light of New Technology: Central Research Laboratory, Hitachi Ltd, Kokubunji, Tokyo, Japan*, pages 28–31, 1989.

[93] Richard Sinkhorn. A relationship between arbitrary positive matrices and doubly stochastic matrices. *The annals of mathematical statistics*, 35(2):876–879, 1964.

[94] Richard Sinkhorn and Paul Knopp. Concerning nonnegative matrices and doubly stochastic matrices. *Pacific Journal of Mathematics*, 21(2):343–348, 1967.

[95] Leonid Gurvits. Classical complexity and quantum entanglement. *Journal of Computer and System Sciences*, 69(3):448–484, 2004.

[96] Thomas R Fleming. Nonparametric estimation for nonhomogeneous markov processes in the problem of competing risks. *The Annals of Statistics*, pages 1057–1070, 1978.

[97] Thomas R Fleming and David P Harrington. Estimation for discrete time nonhomogeneous markov chains. *Stochastic processes and their applications*, 7(2):131–139, 1978.