UNIVERSITÉ DE RENNES 1

Master STS mention mathématiques

Quantum mechanics
Mathematical foundations and applications

Lecture notes

Dimitri Petritis

UFR Mathématiques

Rennes
Preliminary draft of 25 November 2011

Contents

Chapter 1

Physics, mathematics, and mathematical physics

> *La mathématique est une science expérimentale. Contrairement en effet à un contresens qui se répand de nos jours (...), les objets mathématiques préexistent à leurs définitions ; celles-ci ont été élaborées et précisées par des siècles d'activité scientifique et, si elles se sont imposées, c'est en raison de leur adéquation aux objets mathématiques qu'elles modélisent.*
>
> Michel DEMAZURE : *Calcul différentiel*, Presses de l'École Polytechnique, Palaiseau (1979).

Physics relies ultimately on *experiment*. Observation of many different experiments of similar type establishes a *phenomenology* revealing relations between the experimentally measured physical observables. The next step is inductive: *physical models* are proposed satisfying the phenomenological relations. Then, new phenomenology is predicted, experiments designed to verify it, and modelling is proposed. When sufficient data are available, a *physical theory* is proposed verifying all the models that have been developed so far and all the phenomenological relations that have been established. The theory can *deductively* predict the outcome for yet unrealised experiments. If it is technically possible, the experiment is performed. Either the subsequent phenomenology contradicts the theoretical predictions — and the theory must be rejected — or it is in accordance with them — and this precise experiment serves as an additional validity check of the theory.[1] Therefore, physical theories have not a definite status: they are accepted as long as no experiment contradicts them!

[1] See an example of this procedure in *le Monde* of 20 September 2002.

It is a philosophical debate *how mathematical theories* emerge. Some scientists — among them the author of these lines — share the opinion expressed by Michel DEMAZURE (see quotation), claiming that Mathematics is as a matter of fact an experimental science. Accepting, for the time being, this view, hypotheses for particular mathematical branches are the pendants of models. What differentiates strongly mathematics from physics is that once the axioms are stated, the resulting theorems (phenomenology) need not be experimentally corroborated, they exist *per se*. The experimental nature of mathematics is hidden in the mathematician's intuition that served to propose a given set of axioms instead of another.

Mathematical physics is physics, i.e. its truth relies ultimately on experiment but it is also mathematics, in the sense that physical theories are stated as a set of axioms and the resulting physical phenomenology must derive both as theorems and as experimental truth.

A general physical theory must describe all physical phenomena in the *universe*, extending from elementary particles to cosmological phenomena. Numerical values of the fundamental physical quantities, i.e. mass, length, and time span vast ranges: $10^{-31}\text{kg} \leq M \leq 10^{51}\text{kg}$, $10^{-15}\text{m} \leq L \leq 10^{27}\text{m}$, $10^{-23}\text{s} \leq T \leq 10^{17}\text{s}$. Units used in measuring fundamental quantities, i.e. kilogramme (kg), metre (m), and second (s) respectively, were introduced after the French Revolution so that everyday life quantities are expressed with reasonable numerical values (roughly in the range $10^{-3} - 10^{3}$.) The general theory believed to describe the universe[2] is called *Quantum Field Theory*; it contains two fundamental quantities, the speed of light in the vacuum, $c = 2.99792458 \times 10^{8}\text{m/s}$, and the Plank's constant $\hbar = 1.05457 \times 10^{-34}\text{J}\cdot\text{s}$. These constants have extraordinarily atypical numerical values. Everyday velocities are negligible compared to c, everyday actions are overwhelmingly greater than $\hbar$. Therefore, everyday phenomena can be thought as the $c \to \infty$ and $\hbar \to 0$ limits of quantum field theory; the corresponding theory is called *classical mechanics*.

It turns out that considering solely the $c \to \infty$ limit of quantum field theory gives rise to another physical theory called *quantum mechanics*; it describes phenomena for which the action is comparable with $\hbar$. These phenomena are important when dealing with atoms and molecules.

The other partial limit, $\hbar \to 0$, is physically important as well; it describes phenomena involving velocities comparable with c. These phenomena lead to another physical theory called *special relativity*.

Although quantum field theory is still mathematically incomplete, the theories obtained by the limiting processes described above, namely quantum mechanics,

[2] Strictly speaking, there remain unsolved theoretical difficulties in order to succesfully include gravitational phenomena.

special relativity, and classical mechanics are *mathematically closed*, i.e. they can be formulated in a purely axiomatic fashion. All experimental observations made so far (within the range of validity of these theories) are compatible with the derived theorems.

The purpose of this course is twofold. Firstly, the *mathematical foundations of quantum mechanics* are presented. Algebra, analysis, probability, and statistics are necessary to describe and interpret this theory. Its predictions are often totally counter-intuitive. Hence it is interesting to study this theory that provides a useful application of the mathematical tools, a source of inspiration for new developments for the underlying branches of mathematics, and a description of unusual physical phenomena. All these phenomena are verified experimentally nowadays and are even used in breaking through technologies: e.g. tunnel effect has been used for the construction of tunnel effect microscope, a device crucial for the development of *nanotechnology*.

There is however another major technological breakthrough that is foreseen with a tremendous socio-economical impact: if the integration of electronic components continues at the present pace (see figure 1.1), within 10–15 years, only some tenths of silicium atoms will be required to to store a single bit of information. Classical (Boolean) logics does not apply any longer to describe atomic logical gates, quantum (orthocomplemented lattice) logics is needed instead.

Theoretical exploration of this new type of informatics has started and it is proven [11] that some algorithmically complex problems, like the integer prime factoring problem — for which the best known algorithm requires a time is super-polynomial in the number of digits[3] — can be achieved in polynomial time using quantum logic. The present time technology does not yet allow the prime factoring of large integers but it demonstrates that there is no fundamental physical obstruction to its achievement for the rapidly improving computer technology. Should such a breakthrough occur, all our electronic transmissions, protected by classical cryptologic methods could become vulnerable. On the other hand, present day technology allows to securely and unbreakably cipher messages using *quantum cryptologic protocols*. Thus the second purpose of this course is to present the applications of quantum mechanics into the rapidly developing field of *quantum information, computing, communication, and cryptology.*

[3] As a matter of fact, the best known algoritme (Lenstra and Lenstra [4]), requires time $\mathcal{O}(\exp(n^{2/3}\log^2 n))$ to factor a n-digit number.

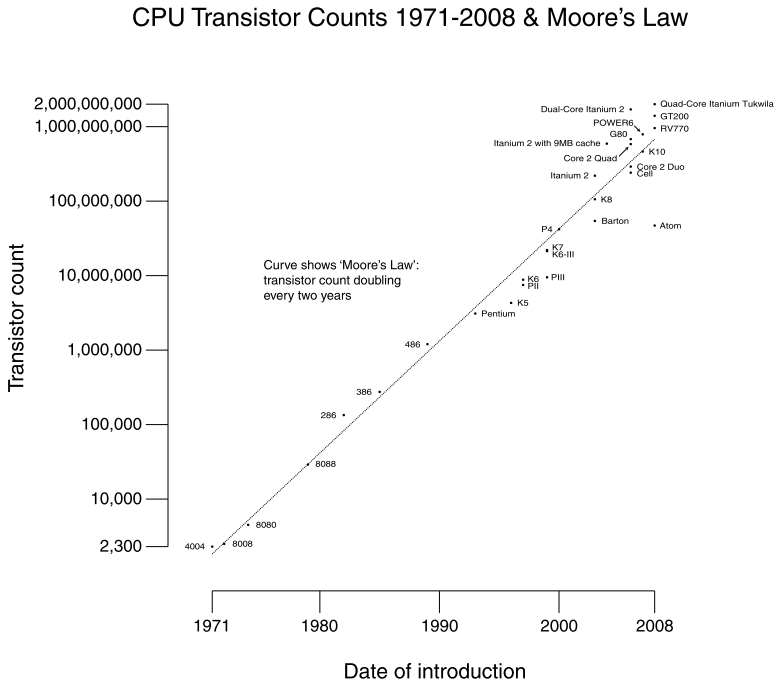

Figure 1.1: The evolution of the number of transistors on integrated circuits in the period 1971–2008, courtesy From Wikipedia: transistor count.

Chapter 2

Phase space, observables, measurements, and yes-no experiments

2.1 Introduction

In experimental sciences, all information on a physical system is obtained through *observation* (also called *measurement*) of the values — within a prescribed set — that can take the physical *observables*. The biggest the set of observables whose values are known, the finest is the knowledge about the physical system. Since crude physical observables (e.g. number of particles, energy, velocity, etc.) can take values in various sets ($\mathbb{N}, \mathbb{R}_+, \mathbb{R}^3$, etc.), to have a unified treatment for general systems, we reduce any physical experiment into a series of measurements of a special class of observables, called *yes-no experiments*. This is very reminiscent of the approximation of any integrable random variable by a sequence of step functions. Therefore, ultimately, we can focus on observables taking values in the set $\{0,1\}$.

To become quantifiable and theoretically exploitable, experimental observations must be performed under very precise conditions, known as *the experimental protocol*. Firstly, the system must be carefully prepared in an initial condition known as the *state* of the system. Mathematically, the state incorporates all the a priori information we have on the system, it belongs to some abstract space of states. Secondly, the system enters in contact with a measuring apparatus, specifically designed to measure the values of a given observable, returning the experi-

mental data with values in some space $(\mathbb{X}, \mathscr{X})$; this is precisely the *measurement process.*

The whole physics relies on the *postulate of statistical reproducibility of experiments*: if the same measurement is performed a very large number of times on a system prepared in the some given state, the experimentally observed data for a given observable are scattered around some mean value in $\mathbb{X}$ with some fluctuations around the mean value. However, when the number of repetitions tends to infinity, the empirical distribution of the observed data tends to some probability distribution on $(\mathbb{X}, \mathscr{X})$. Thus, abstractly, a measurement is a black box transforming states into probability measures on some space of observations.

Dealing with *random variables*, the natural question that arises is: what is the appropriate (abstract) probability space, if any, on which random variables entering a given problem can be defined? For sequences of classical random variables, the answer is well known: such an abstract probability space exists, provided that the sequence verifies the Kolmogorov's compatibility conditions; in that case, there exists a canonical (minimal) realisation of the abstract probability space on which the whole sequence is defined. Elements of this probability space are called *trajectories* of the sequence. The physical analogue of the minimal realisation of the abstract probability space is called *phase space*. Elements of the phase space are called *(pure) phases*. The physical analogue of a random variable is called *observable*.

It turns out that physical observables for classical systems are just random variables so that the phase space for such systems is a genuine probability space, while for quantum systems, observables are (generally non-commuting) *Hermitean operators* acting on an *abstract Hilbert space* that plays the rôle of quantum phase space.

2.2 Classical systems

2.2.1 Some reminders from probability theory

Let us start with the mathematical notion of a random variable.

Definition 2.2.1 (Random variable) Let $(\Omega, \mathscr{F}, \mathbb{P})$ be an abstract probability space and $(\mathbb{X}, \mathscr{X})$ a measurable space[1]. A function $X : \Omega \to \mathbb{X}$ that is $(\mathscr{F}, \mathscr{X})$-measurable is called (a $\mathbb{X}$-valued) *random variable*. The induced probability mea-

[1] The space $\mathbb{X}$ can be any Polish space (i.e. a metric, complete, and separable space.) We shall only consider the case $\mathbb{X} = \mathbb{R}^d$, for some d, in this course.

sure $\mathbb{P}_X$ on $(\mathbb{X}, \mathscr{X})$ (i.e. $\mathbb{P}_X(A) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\}, A \in \mathscr{X})$ is called the *law* (or *distribution*) of X.

Example 2.2.2 Let $\mathbb{X} = \{0,1\}$, $\mathscr{X}$ be the algebra of subsets of $\mathbb{X}$, and $\mathbb{P}_X(\{0\}) = \mathbb{P}_X(\{1\}) = 1/2$ the law of a random variable X (the honest coin tossing). A possible realisation of $(\Omega, \mathscr{F}, \mathbb{P})$ is $([0,1], \mathscr{B}([0,1]), \lambda)$, where λ denotes the Lebesgue measure, and a possible realisation of the random variable X is

$$X(\omega) = \begin{cases} 0 & \text{if } \omega \in [0, 1/2[\\ 1 & \text{if } \omega \in [1/2, 1]. \end{cases}$$

Notice however that the above realisation of the probability space involves the Borel σ-algebra over an uncountable set, quite complicated an object indeed. A much more economical realisation should be given by $\Omega = \{0,1\}$, $\mathscr{F} = \mathscr{X}$, and $\mathbb{P}(\{0\}) = \mathbb{P}(\{1\}) = 1/2$. In the latter case the random variable X should read $X(\omega) = \omega$: on this smaller probability space, the random variable is the identity function. Such a realisation is *minimal*.

Exercise 2.2.3 Generalise the above minimal construction to the case we consider two random variables $X_i : \Omega \to \mathbb{X}$, for $i = 1, 2$. Are there some plausible requirements on the joint distributions for such a construction to be possible?

The canonical construction of the minimal probability space carrying an infinite family of random variables is also possible.

Definition 2.2.4 (Consistency) Let T be an infinite set (countable or uncountable) and for each $t \in T$ denote by $\mathbb{R}_t$ a copy of the real line, indexed by t. Denote by $\mathbb{R}^T = \times_{t \in T} \mathbb{R}_t$ and for $n \geq 1$ by $\tau = (t_1, \ldots, t_n)$ a finite ordered set of distinct indices $t_i \in T, i = 1, \ldots, n$. Denote $\mathbb{P}^\tau$ a probability measure on $(\mathbb{R}^\tau, \mathscr{B}(\mathbb{R}^\tau))$ where $\mathbb{R}^\tau = \mathbb{R}_{t_1} \times \cdots \times \mathbb{R}_{t_n}$. We say that the family $(\mathbb{P}^\tau)$, where τ runs through all finite ordered subsets of T, is *consistent*, if

1. $\mathbb{P}^{(t_1, \ldots t_n)}(A_1 \times \cdots \times A_n) = \mathbb{P}^{(t_{\sigma(1)}, \ldots t_{\sigma(n)}}(A_{\sigma(1)} \times \cdots \times A_{\sigma(n)})$, where σ is an arbitrary permutation of $(1, \ldots, n)$ and $A_i \in \mathscr{B}(\mathbb{R}_{t_i})$, and
2. $\mathbb{P}^{(t_1, \ldots t_n)}(A_1 \times \cdots \times A_{n-1} \times \mathbb{R}) = \mathbb{P}^{(t_1, \ldots t_{n-1})}(A_1 \times \cdots \times A_{n-1})$.

Definition 2.2.5 Let T be a subset of $\mathbb{R}$. A family of random variables $X \equiv (X_t)_{t \in T}$ is called a *stochastic process* with time domain T.

If $T = \mathbb{N}$ or $\mathbb{Z}$, the process is called a discrete time process or random sequence, if $T = [0,1]$ or $\mathbb{R}$ or $\mathbb{R}_+$, the process is a continuous time process.

The natural question that arises is whether there exists a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ carrying the whole process. In other words, if $\mathbb{P}_X$ denotes the distribution of the process X, what are the conditions it must fulfil so that there exists a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ such that $\mathbb{P}(B) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in B\}$ for all $B \in \mathscr{B}(\mathbb{R}^T)$? The answer is given by the following

Theorem 2.2.6 (Kolmogorov's existence) *Suppose that for $n \geq 1$, the family $\mathbb{P}_{(X_{t_1},\ldots,X_{t_n})}$, with $t_1 < \ldots t_n$ and $t_i \in T \subseteq \mathbb{R}$, for $i = 1,\ldots,n$, is a consistent family of probability measures. Then, there are*

1. *a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and*
2. *a stochastic process $X = (X_t)_{t\in T}$ such that $\mathbb{P}_{(X_{t_1},\ldots,X_{t_n})}(]-\infty, x_1] \times \cdots \times]-\infty, x-n] = \mathbb{P}(\{\omega \in \Omega : X_{t_1}(\omega) \leq x_1, \ldots, X_{t_n}(\omega) \leq x_n\}$.*

Proof: See, for instance, in [10], theorem II.2.1, p. 247. □

Remark 2.2.7 The canonical construction of the minimal probability space is $\Omega = \mathbb{R}^T$, $\mathscr{F} = \mathscr{B}(\mathbb{R}^T)$ and for every $t \in T$, $X_t(\omega) = \omega_t$. This minimal space is also called *space of trajectories* of the random process and the realisation of X *coordinate method*.

2.2.2 Classical phase space, observables, and states

A rough definition of the notion of classical phase space is: the (minimal) space on which all legitimate physical observables can be defined, or, equivalently, all legitimate questions can receive a definite answer. From this conceptual view, the classical phase space shares the same indeterminacy as the probability space. The only objects having physical relevance are the physical observables (as is the case for random variables in probability theory.) Therefore, the same system can be minimally described by two different phase spaces, depending on the set of questions to be answered.

Give example of coin tossing by Diaconis, Holmes, Montgomery: Dynamical bias in the coin tossing

Example 2.2.8 (Die rolling) Let the physical system be a die and the complete set of questions to be answered the set $\{Q_1,\ldots,Q_6\}$, where $Q_i, i = 1,\ldots,6$ stands

for the question: "When the die lies at equilibrium on the table, does the top face read i?" An obvious choice for the phase space is $\Omega = \{1, \ldots, 6\}$. The random variable X corresponding to the physical observable "value of the top face" is realised by $X(\omega) = \omega, \omega \in \Omega$ and the questions by $Q_i = \mathbb{1}_{\{X=i\}}$, for $i = 1, \ldots, 6$.

Exercise 2.2.9 Determine the phase space for a point mass in dimension 1 subject to the force exerted by a spring of elastic constant k.

Solution: Recall that a point mass m in dimension 1 obeys Newton's equation:

$$m\frac{d^2x}{dt^2}(t) = F(x(t)),$$

subject to the initial conditions $x(0) = x_0$ and $\dot{x}(0) = v_0$, where $x(t)$ denotes the position of the mass at instant t and $F(y)$ denotes the force exerted on the particle at position y. The kinetic energy, K, of the particle is a quadratic form in the velocity

$$K(\dot{x}) = \frac{m}{2}\dot{x}^2$$

and the potential energy, U, is given by

$$U(x) = -\int_{x_0}^{x} F(y)dy.$$

In order to conclude, we need the following

Theorem 2.2.10 *The total energy $H(x,\dot{x}) = K(\dot{x}) + U(x)$ is a constant of motion, i.e. does not depend on t.*

Proof:

$$\begin{aligned}\frac{d}{dt}(K(\dot{x}) + U(x)) &= m\dot{x}\ddot{x} + \frac{\partial U}{\partial x}(x)\dot{x} \\ &= \dot{x}(m\ddot{x} - F(x)) \\ &= 0.\end{aligned}$$

□

Hence the Newton's equation is equivalent to the system of first order differential equations, known as *Hamilton's equations*:

$$\begin{aligned}\frac{dp}{dt} &= -\frac{\partial H}{\partial q} \\ \frac{dq}{dt} &= \frac{\partial H}{\partial p},\end{aligned}$$

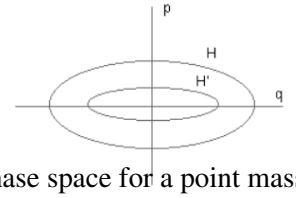

Figure 2.1: The phase space for a point mass in dimension one.

subject to the initial condition

$$\begin{pmatrix} q(0) \\ p(0) \end{pmatrix} = \begin{pmatrix} q_0 \\ p_0 \end{pmatrix},$$

where $p = m\dot{x}$, $q = x$, and $H = \frac{p^2}{2m} + U(q)$. Therefore, the phase space for the point mass in dimension one is $\mathbb{R}^2$ (one dimension for the position, q, and one for the momentum p.) Moreover, this space is stratified according to constant energy surfaces that are ellipses for the case of elastic spring, because potential energy is quadratic in q (see figure 2.1.)

If $\omega(t) = \begin{pmatrix} q(t) \\ p(t) \end{pmatrix} \in \mathbb{R}^2$ represents the coordinate and momentum of the system at time t, the time evolution induced by the system of Hamilton's equations can be thought as the flow on $\mathbb{R}^2$, described by $\omega(t) = T_t\omega(0)$, with initial condition $\omega(0) = \begin{pmatrix} q_0 \\ p_0 \end{pmatrix}$.

Example 2.2.11 Consider the same physical system as in example 2.2.8 and the same set of questions but think of the die as a solid body that can evolve in the space. To completely describe its state, we need 3 coordinates for its barycentre, 3 coordinates for the velocity of the barycentre, 3 coordinates for the angular velocity, and 3 Euler angles for the orientation of the exterior normal at the centre of face "6". Thus, $\Omega = \mathbb{R}^9 \times [0, 2\pi]^3$. Now the realisation $X : \Omega \to \{1, \ldots, 6\}$ is much more involved (but still possible in principle) and the questions are again represented by $Q_i = \mathbb{1}_{\{X=i\}}$, for $i = 1, \ldots, 6$. Yet, obviously, the representation given in example 2.2.8 is much simpler than the present one.

Axiom 2.2.12 *The phase space of a classical system is an abstract measurable space* $(\Omega, \mathscr{F})$. *The states of a classical system are the probability measures on* $(\Omega, \mathscr{F})$. *Pure states correspond to Dirac masses.*

Axiom 2.2.13 *Any time evolution of an isolated classical system is implemented by an invertible measurable transformation $T : \Omega \to \Omega$ leaving the states invariant.*

Axiom 2.2.14 *To any physical observable of a classical system corresponds a random variable $X : \Omega \to \mathbb{X}$, where $(\mathbb{X}, \mathscr{X})$ is a measurable space. Yes-no questions are special observables of the form $Q : \Omega \to \{0,1\}$. Measurement of a classical observable X when the system is in state μ, is the law of the random variable under μ.*

Remark 2.2.15 Questions are special kinds of random variables. They always can be written as $Q = \mathbb{1}_A \circ X$, where $X : \Omega \to \mathbb{X} \subseteq \mathbb{R}$ is a random variable, and $A \in \mathscr{X}$. Any random variable X is termed *physical observable*. Questions are special types of physical observables. Since any question is an indicator, it verifies $Q^2 = Q$ i.e. it is a projector. When dealing with a single random variable (physical observable), the complete set of possible questions is in bijection with the σ-algebra $\mathscr{X}$ of measurable subsets of $\mathbb{X}$. If $Q_A = \mathbb{1}_A \circ X$ and $Q_B = \mathbb{1}_B \circ X$ are two different questions and moreover $A \cap B = \emptyset$ then $Q_A Q_B = 0$, i.e. questions testing disjoint sets in the range of a random variable are orthogonal projectors.

Exercise 2.2.16 Let μ be a state on $(\Omega, \mathscr{F})$, X a $\mathbb{X}$-valued random variable ($\mathbb{X} \subseteq \mathbb{R}$), and Q_A the question $\mathbb{1}_{\{X \in A\}}$ for some fixed $A \in \mathscr{X}$. Compute $\mu(Q_A)$. What happens if μ is a pure state? What happens if $(\Omega, \mathscr{F})$ is minimal for the random variable X?

Solution:

$$\begin{aligned} \mu(Q_A) &= \int_\Omega \mathbb{1}_{\{X \in A\}}(\omega)\mu(d\omega) \\ &= \int_\Omega \mathbb{1}_A(X(\omega))\mu(d\omega) \end{aligned}$$

If $\mu = \delta_{\omega_0}$ for some ω_0.

$$\begin{aligned} \delta_{\omega_0}(Q_A) &= \int_\Omega \mathbb{1}_{\{X \in A\}}(\omega)\delta_{\omega_0}(d\omega) \\ &= \mathbb{1}_A(X(\omega_0)). \end{aligned}$$

If the space is minimal for X, then $X(\omega) = \omega$ and we get respectively: $\mu(Q_A) = \int_\Omega \mathbb{1}_A(X(\omega))\mu(d\omega) = \mu(A)$ and $\delta_{\omega_0}(Q_A) = \delta_{\omega_0}(A)$. □

Exercise 2.2.17 What is the minimal phase space for a mechanical system composed by N point particles in dimension 3?

2.3 Quantum systems

2.3.1 A first axiomatic setting for quantum mechanics

Various formulations of quantum mechanics are possible. We start from the most straightforward one, historically introduced by John VON NEUMANN [16]. Later on, a more general formulation [13], based on C^*-algebras, will be given; this latter formulation has the advantage of allowing a unified treatment for both classical and quantum systems.

For the time being, we proclaim that a quantum system verifies the following axioms.

Axiom 2.3.1 *The phase space of a quantum mechanical system is a complex Hilbert space $\mathbb{H}$. Unit vectors of $\mathbb{H}$ correspond to pure quantum states.*

Axiom 2.3.2 *Any time evolution of an isolated quantum system is described by a unitary operator acting on $\mathbb{H}$. Conversely, any unitary operator acting on $\mathbb{H}$ corresponds to to a possible time evolution of the system.*

Axiom 2.3.3 *With every physical observable, O_X, of a quantum system is associated a self-adjoint operator X acting on the phase space $\mathbb{H}$ of the system. Yes-no questions are special self-adjoint operators that are projections. Measurement of an observable represented by the self-adjoint operator X for a quantum system being in the pure state described by the unit vector ψ corresponds to the spectral measure on the real line induced by $\langle \psi | X\psi \rangle$.*

These axioms will be revisited later. For the time being, it is instructing to illustrate the implications of these axioms on a very simple non-trivial quantum system and try to interpret their significance.

2.3.2 Interpretation of the basic axioms

In this subsection we study a quantum system whose phase space $\mathbb{H} = \mathbb{C}^2$. This is the simplest non-trivial situation that might occur and could describe, for instance, the internal degrees of freedom of an atom having two states. Notice however that in general, even for very simple finite systems, the phase space is not necessarily finite-dimensional.

Interpretation of axiom 2.3.1

Every $f \in \mathbb{H}$ can be decomposed into $f = f_1\varepsilon_1 + f_2\varepsilon_2$ with $f_1, f_2 \in \mathbb{C}$ and $\varepsilon_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\varepsilon_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. If $\|f\| \neq 0$, denote by $\phi = \frac{f}{\|f\|}$ the corresponding normalised vector[2].

Now $\phi = \phi_1\varepsilon_1 + \phi_2\varepsilon_2$ with $|\phi_1|^2 + |\phi_2|^2 = 1$ is a pure state. The numbers $|\phi_1|^2$ and $|\phi_2|^2$ are non-negative reals summing up to 1; therefore, they are interpreted as a probability on the finite set of coordinates $\{1,2\}$. Consequently, the complex numbers $\phi_1 = \langle \varepsilon_1 \,|\, \phi \rangle$ and $\phi_2 = \langle \varepsilon_2 \,|\, \phi \rangle$ are complex probability amplitudes, their squared modulus represents the probability that a system in a pure state ϕ is in the pure state ε_1 or ε_2.

Interpretation of axiom 2.3.2

A unitary operator on $\mathbb{H}$ is a 2×2 matrix U, verifying $UU^* = U^*U = I$. If ϕ is a pure state, then $\psi = U\phi$ verifies $\|\psi\|^2 = \langle U\phi \,|\, U\phi \rangle = \langle \phi \,|\, U^*U\phi \rangle = \|\phi\|^2$. Therefore quantum evolution preserves pure states. Moreover, due to the unitarity of U, we have $\phi = U^*\psi$, and since U^* is again unitary, it corresponds to a possible time evolution (as a matter of fact to the time reversed evolution of the one corresponding to U.) This shows that time evolution of isolated quantum systems is reversible.

Interpretation of axiom 2.3.3

This axiom has the most counter-intuitive consequences. Recall that any linear operator X admits a spectral decomposition $X = \int_{\mathsf{spec}(X)} \lambda P(d\lambda)$: If X is self-adjoint, then $\mathsf{spec}(X) \subseteq \mathbb{R}$. Let us illustrate with a very simple example: chose for X the matrix $X = \begin{pmatrix} 1 & 2i \\ -2i & 2 \end{pmatrix}$. We compute easily

Eigenvalues λ	Eigenvectors $u(\lambda)$	Projectors $P(\{\lambda\})$
-3	$\frac{1}{\sqrt{5}}\begin{pmatrix} -i \\ 2 \end{pmatrix}$	$\frac{1}{5}\begin{pmatrix} 1 & -2i \\ 2i & 4 \end{pmatrix}$
2	$\frac{1}{\sqrt{5}}\begin{pmatrix} 2i \\ 1 \end{pmatrix}$	$\frac{1}{5}\begin{pmatrix} 4 & 2i \\ -2i & 1 \end{pmatrix}$

[2]General (unnormalised) vectors of $\mathbb{H}$ are denoted by small Latin letters f, g, h, etc.; normalised vectors by small Greek letters ϕ, χ, ψ, etc.

Hence

$$\begin{aligned} X &= \sum_{\lambda\in\{-3,2\}} \lambda P(\{\lambda\}) \\ &= (-3)\frac{1}{5}\begin{pmatrix} 1 & -2i \\ 2i & 4 \end{pmatrix} + 2\frac{1}{5}\begin{pmatrix} 4 & 2i \\ -2i & 1 \end{pmatrix}. \end{aligned}$$

The operators $P(\{-3\})$ and $P(\{2\})$ are self-adjoint (hence they correspond to observables) and are projectors to mutually orthogonal subspaces. They play the role of yes-no questions for a quantum system (recall remark 2.2.15.)

Now, let $\psi \in \mathbb{H}$ be a pure phase; since $u(-3)$ and $u(2)$ are two orthonormal vectors of $\mathbb{H}$ (hence also pure phases), they serve as basis to decompose $\psi = \alpha_{-3}u(-3) + \alpha_2 u(2)$, with $\|\psi\|^2 = |\alpha_{-3}|^2 + |\alpha_2|^2 = 1$. Thus any pure state ψ, with probability $|\langle\,\psi\,|\,u(-3)\,\rangle|^2$ is in the pure state $u(-3)$ and with probability $|\langle\,\psi\,|\,u(2)\,\rangle|^2$ is in the pure state $u(2)$.

Compute further

$$\begin{aligned} \langle\,\psi\,|\,X\psi\,\rangle &= \sum_{\lambda,\lambda',\lambda''} \alpha^*_\lambda \alpha_{\lambda''}\lambda'\langle\,u(\lambda)\,|\,P(\lambda')u(\lambda'')\,\rangle \\ &= \sum_{\lambda\in\mathsf{spec}(X)} \lambda|\alpha_\lambda|^2. \end{aligned}$$

Yet $(|\alpha_\lambda|^2)_{\lambda\in\mathsf{spec}(X)}$ can be interpreted as a probability on the set of the spectral values. Hence, the scalar product $\langle\,\psi\,|\,X\psi\,\rangle$ is the expectation of the spectral values with respect to the decomposition of ψ on the basis of eigenvectors. It is worth noticing that expectation of a classical random variable X taking values in a finite set $\{x_1,\ldots,x_n\}$ with probabilities $p_1,\ldots,p_n$ respectively, is

$$\begin{aligned} \mathbb{E}X &= \sum_{i=1}^n x_i p_i \\ &= \sum_{i=1}^n \sqrt{p_i}x_i\sqrt{p_i} \\ &= \sum_{i=1}^n \sqrt{p_i}\exp(-i\theta_i)x_i\sqrt{p_i}\exp(i\theta_i), \end{aligned}$$

with arbitrary $\theta_i \in \mathbb{R}, i = 1,\ldots,n$. Hence, classically, $\mathbb{E}X = \langle\,\psi\,|\,X\psi\,\rangle$ with $\psi = \begin{pmatrix} \sqrt{p_1}\exp(i\theta_1) \\ \vdots \\ \sqrt{p_n}\exp(i\theta_n) \end{pmatrix}$, verifying $\|\psi\| = 1$ and with $X = \begin{pmatrix} x_1 & & 0 \\ & \ddots & \\ 0 & & x_n \end{pmatrix}$. We have moreover seen that classical probability is equivalent to classical physics; thanks to the previous lines, it turns out that that it is also equivalent to quantum physics involving solely diagonal self-adjoint operators as observables. The full flavour of

quantum physics is obtained only when the observables are represented by non-diagonal self-adjoint operators.

Consider now,

$$\begin{aligned} f_\lambda &= P(\{\lambda\})\psi \\ &= \begin{cases} \langle u(\lambda) \,|\, \psi \rangle u(\lambda) & \text{if } \lambda \in \mathsf{spec}(X) \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

The vector f_λ is in general unnormalised; the corresponding normalised state $\phi_\lambda = \frac{P(\{\lambda\})\psi}{\|P(\{\lambda\})\psi\|}$, well defined when $\lambda \in \mathsf{spec}(X)$, has a very particular interpretation. Suppose we ask the question: "does the physical observable O_X takes the value -3?" The answer, as in the classical case, is a probabilistic one: $\mathbb{P}(\{O_X = -3\}) = |\alpha_{-3}|^2 = \langle f_{-3} \,|\, f_{-3} \rangle = \|P(\{-3\})\psi\|^2$. What is new, is that once we have asked this question, the state ψ is projected on the eigenspace $P(\{-3\})\mathbb{H}$ and is represented by the state ϕ_{-3}. This means that asking a question on the system changes its state! This is a totally new phenomenon without classical counterpart. Asking questions about a quantum system corresponds to a *quantum measurement*. Hence, the measurement irreversibly changes (projects) the state of the system.

Summarising the interpretation of the three axioms, we have learnt that

- quantum mechanics has a probabilistic interpretation, generalising the classical probability theory to a quantum (non-Abelian) one,
- quantum evolution is reversible,
- quantum measurement is irreversible.

Were only to consider this generalisation of probability theory to a non-commutative setting and to explore its implications for explaining quantum physical phenomena, should the enterprise be already a fascinating one. But there is even much more fascination about it: there has been demonstrated lately that quantum phenomena can serve to cipher messages in an unbreakable way and these theoretical predictions have already been exemplified by currently working pre-industrial prototypes[3].

In a more speculative perspective, it is even thought that in the near future there will be manufactured computers capable of performing large scale computations using quantum algorithms[4]. Should such a construction be realised, a vast family of

[3] See the article [**?**], articles in *Le Monde* (they can be found on the website of this course), the website `www.idquantique.com` of the company commercialising quantum cryptologic and teleporting devices, etc.

[4] Contrary to the quantum transmission and cryptologic technologies that are already available, the prototypes of quantum computers that have been manufactured so far have still extremely limited scale capabilities.

problems in the (classical) complexity class of "exponential time" could be solved in polynomial time on a quantum computer.

2.4 Quantum explanation of some experiments without classical interpretation

2.4.1 Light polarisers are not classical filters

The experimental setting of this experiment is depicted in the following figure 2.2.

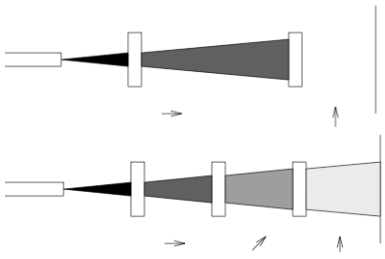

Figure 2.2: The experimental setting with two polarisers perpendicularly crossed and three polarisers with relative orientations differing by 45 degrees.

When natural light passes through a horizontally oriented polariser, half of the initial intensity is transmitted. When a vertical polariser is then placed in the beam, the light is totally absorbed (first setting in figure 2.2.) On the contrary when three polarisers with respective orientations turned by 45 degrees each time, the eighth of the intensity is transmitted.

Assume for the moment that polarisation is a classical $[0,2\pi]$-valued observable X. If the source emits an unpolarised beam, the different photons composing it have random orientations, i.e. the emitted photon polarisation is a random variable $X \in [0,2\pi]$ with distribution $\mathbb{P}_X = \lambda/2\pi$, where λ is the Lebesque measure on $[0,2\pi]$. The minimal probability space is $([0,1],\mathscr{B}([0,1]),\lambda/2\pi)$.

Assume now that the first polariser acts as a classical slit allowing photons with exactly the slit orientation to cross. Since experimental imperfections are always present, we can even assume that the slit is not exactly one-dimensional but some angular aperture is possible. Hence, we assume that the presence of the horizontal polariser is equivalent in asking the question:

$$Q_{\rightarrow} = \mathbb{1}_{X\in[-\delta,\delta]\cup[\pi-\delta,\pi+\delta]},$$

for some arbitrarily small $\delta > 0$. One easily computes $\mathbb{E}Q_{\rightarrow} = \frac{2\delta}{\pi}$ and $\lim_{\delta\to 0}\mathbb{E}X = 0$. The vertical polariser is equivalent to the question:

$$Q_{\uparrow} = \mathbb{1}_{X\in[\pi/2-\delta,\pi/2+\delta]\cup[3\pi/2-\delta,3\pi/2+\delta]}.$$

Obviously $\mathbb{E}(Q_{\rightarrow}Q_{\uparrow}) = 0$ for small δ. Finally,

$$Q_{\nearrow} = \mathbb{1}_{X\in[\pi/4-\delta,\pi/4+\delta]\cup[5\pi/4-\delta,5\pi/4+\delta]}$$

and similarly $\mathbb{E}(Q_{\rightarrow}Q_{\uparrow}Q_{\nearrow}) = 0$ for small δ. If the initial beam contains N photons (for real beams, N is very large), with arbitrary polarisations $(X_i)_{i=1,\ldots,N}$, that are independent and identically distributed random variables of law $\mathbb{P}_X$, the intensity before crossing the polariser is $I_0 = \kappa N$, with κ some constant while after crossing the first polariser is $I_{\rightarrow} = \kappa\sum_{i=1}^{N} Q_{\rightarrow}^{(i)}$. Now, by the law of large numbers, $\lim_{N\to\infty}\frac{I_{\rightarrow}}{I_0} = 2\delta/\pi$, while $\lim_{N\to\infty}\frac{I_{\rightarrow}I_{\uparrow}}{I_0} = 0$ and $\lim_{N\to\infty}\frac{I_{\rightarrow}I_{\uparrow}I_{\nearrow}}{I_0} = 0$. All these results contradict the experimental observations.

Assume now that the system is quantum and any single photon of the original beam is in some state of the form $\psi = \psi_1\varepsilon_1 + \psi_2\varepsilon_2 \in \mathbb{H}$ with $|\psi_1|^2 + |\psi_2|^2 = 1$, and since the photons are unpolarised, it is reasonable to assume that $|\psi_1| = |\psi_2| = 1/\sqrt{2}$. Now consider three one-dimensional subspaces of $\mathbb{H}$, denoted $\mathbb{H}_{\rightarrow} = \{\alpha\varepsilon_1, \alpha\in\mathbb{C}\}$, $\mathbb{H}_{\uparrow} = \{\alpha\varepsilon_2, \alpha\in\mathbb{C}\}$, and $\mathbb{H}_{\nearrow} = \{\alpha(\varepsilon_1+\varepsilon_2), \alpha\in\mathbb{C}\}$. Questions asked by the three polarisers correspond to projections $P_{\rightarrow}$, $P_{\uparrow}$, and $P_{\nearrow}$ to the corresponding subspaces. Now, $\mathbb{E}_{\psi}Q_{\rightarrow} = \langle\,\psi\,|P_{\rightarrow}\psi\,\rangle = \langle\,\psi_1\varepsilon_1+\psi_2\varepsilon_2\,|\,\psi_1\varepsilon_1\,\rangle = |\psi_1|^2 = 1/2$. After the photon has crossed the first polariser, it is in the new state $\psi' = \varepsilon_1$. If the next polariser to cross is a polariser at 45 degrees, we have $\mathbb{E}_{\psi'}Q_{\nearrow} = \langle\,\varepsilon_1\,|P_{\nearrow}\varepsilon_1\,\rangle = \langle\,\varepsilon_1\,|\,\langle\,\frac{\varepsilon_1+\varepsilon_2}{\sqrt{2}}\,|\,\varepsilon_1\,\rangle\frac{\varepsilon_1+\varepsilon_2}{\sqrt{2}}\,\rangle = 1/2$ and the state after the question has been asked is $\psi'' = \frac{\varepsilon_1+\varepsilon_2}{\sqrt{2}}$. When the photon crosses the third polariser, we get $\mathbb{E}_{\psi''}Q_{\uparrow} = \langle\,\frac{\varepsilon_1+\varepsilon_2}{\sqrt{2}}\,|P_{\uparrow}\frac{\varepsilon_1+\varepsilon_2}{\sqrt{2}}\,\rangle = \langle\,\frac{\varepsilon_1+\varepsilon_2}{\sqrt{2}}\,|\,\langle\varepsilon_2|\,\frac{\varepsilon_1+\varepsilon_2}{\sqrt{2}}\,\rangle\varepsilon_2\,\rangle = 1/8$. But if the intermediate polariser is omitted, we must compute $\mathbb{E}_{\psi'}Q_{\uparrow} = \langle\,\varepsilon_1\,|\,\langle\,\varepsilon_1\,|\,\varepsilon_2\,\rangle\varepsilon_1\,\rangle = 0$. Hence the quantum explanation is in complete agreement with the experimental observation.

2.4.2 Heisenberg's uncertainty principle

Spectral decomposition allows computation of the expectation of an operator X, in a pure state, ψ, by

$$\mathbb{E}_{\psi}X = \langle\,\psi\,|X\psi\,\rangle = \sum_{\lambda\in\mathsf{spec}(X)}\lambda\,|\psi_{\lambda}|^2$$

and when the operator X is self-adjoint, the spectrum is real and the expectation is then a real number. What makes quantum probability different from classical one, is (among other things) the impossibility of simultaneous diagonalisation of

two non-commuting operators. Following the probabilistic interpretation, denote by $\mathsf{Var}_\psi(X) = \mathbb{E}_\psi(X^2) - (\mathbb{E}_\psi(X))^2$.

Theorem 2.4.1 (Heisenberg's uncertainty) *Let X, Y be two bounded self-adjoint operators on a Hilbert space $\mathbb{H}$ and suppose a fixed pure state ψ is given. Then*

$$\mathsf{Var}_\psi(X)\mathsf{Var}_\psi(Y) \geq \frac{|\langle \psi | [X,Y]\psi \rangle|}{2}.$$

Proof: First notice that $(i[X,Y])^* = i[X,Y]$ thus the commutator is skew-adjoint. Without loss of generality, we can assume that $\mathbb{E}_\psi X = \mathbb{E}_\psi Y = 0$ (otherwise consider $X - \mathbb{E}_\psi X$ and similarly for Y.) Now, $\langle \psi | XY\psi \rangle = \alpha + i\beta$, with $\alpha, \beta \in \mathbb{R}$. Hence, $\langle \psi | [X,Y]\psi \rangle = 2i\beta$ and obviously

$$\begin{aligned} 0 &\leq 4\beta^2 = |\langle \psi | [X,Y]\psi \rangle|^2 \\ &\leq 4|\langle \psi | XY\psi \rangle|^2 \\ &\leq 4\langle \psi | X^2\psi \rangle \langle \psi | Y^2\psi \rangle, \end{aligned}$$

the last inequality being Cauchy-Schwarz. □

This is a typically quantum phenomenon without classical counterpart. In fact, given two arbitrary classical random variables X, Y on a measurable space $(\Omega, \mathscr{F})$, there exists always states (i.e. probability measures) on $(\Omega, \mathscr{F})$ such that $\mathsf{Var}(X)\mathsf{Var}(Y) = 0$ (for instance chose $\mathbb{P}(d\omega) = \delta_{\omega_0}(d\omega)$.

2.5 Dirac's notation

Usual notation	Dirac's notation
Orthonormal basis $(e_1, \ldots, e_n)$	n symbols, eg. $\{1, \ldots, n\}$ $\|1\rangle, \ldots \|n\rangle$
$\psi = \sum_i \psi_i e_i$ $\langle \phi \| \psi \rangle = \sum \overline{\phi}_i \psi_i$	$\|\psi\rangle = \sum_i \psi_i \|i\rangle$ $\langle \phi \| \psi \rangle = \sum \overline{\phi}_i \psi_i$
$\mathbb{H}^* = \{f : \mathbb{H} \to \mathbb{C}, \text{linear}\}$ $\dagger : \mathbb{H} \to \mathbb{H}^*$ $\dagger : \phi \mapsto f(\phi(\cdot) = \langle \phi \| \cdot \rangle$ $\langle \phi \| \psi \rangle = f_\phi(\psi)$	$\mathbb{H}^* = \{f : \mathbb{H} \to \mathbb{C}, \text{linear}\}$ $\dagger : \mathbb{H} \to \mathbb{H}^*$ $\dagger : \|\phi\rangle \mapsto \langle \phi \|$ $\langle \phi \| \psi \rangle = \langle \phi \| \| \psi \rangle$
$X = X^*$ $\langle \phi \| X\psi \rangle = \langle X^*\phi \| \psi \rangle = \langle X\phi \| \psi \rangle$	$X = X^*$ $\langle \phi \| X \| \psi \rangle$
$Xu(\lambda_i) = \lambda_i u(\lambda_i)$ $P(\{\lambda_i\})$ projector $X = \sum_i \lambda_i P(\{\lambda_i\})$	$X\|\lambda_i\rangle = \lambda_i \|\lambda_i\rangle$ $\|\lambda_i\rangle\langle\lambda_i\|$ $X = \sum_i \lambda_i \|\lambda_i\rangle\langle\lambda_i\|$

Chapter 3

Algebras of operators

3.1 Introduction and motivation

Let $\mathbb{V} = \mathbb{C}^n$, with $n \in \mathbb{N}$. Elementary linear algebra establishes that the set of linear mappings $\mathfrak{L}(\mathbb{V}) = \{T : \mathbb{V} \to \mathbb{V} : T \text{ linear }\}$ is a $\mathbb{C}$-vector space of (complex) dimension n^2, isomorphic to $\mathbf{M}_n(\mathbb{C})$, the space of $n \times n$ matrices with complex coefficients. Moreover, if $S, T \in \mathfrak{L}(\mathbb{V})$, the maps S and Tćan be composed, their composition $T \circ S$ being represented by the corresponding matrix product. Thus, on the vector space $\mathfrak{L}(\mathbb{V})$, is defined an internal multiplication

$$\mathfrak{L}(\mathbb{V}) \times \mathfrak{L}(\mathbb{V}) \ni (T, S) \mapsto T \circ S \in \mathfrak{L}(\mathbb{V})$$

turning this vector space into *an algebra*.

When the underlying vector space $\mathbb{V}$ is of infinite dimension, caution must be paid on defining linear maps. In general, linear mappings $T : \mathbb{V} \to \mathbb{V}$, called *(linear) operators*, are defined only on some proper subset of $\mathbb{V}$ denoted $\mathsf{Dom}(T)$ and called the *domain*[1] of T. When $\mathbb{V}$ is a normed space, there is a natural way to define a norm on $\mathfrak{L}(\mathbb{V})$. We denote by $\mathfrak{B}(\mathbb{V})$ the vector space of *bounded linear operators* on $\mathbb{V}$, i.e. linear maps $T : \mathbb{V} \to \mathbb{V}$ such that $\|T\| < \infty$ (equivalently, verifying $\mathsf{Dom}(T) = \mathbb{V}$.) When $\mathbb{H}$ is a Hilbert space, bounded linear operators on $\mathbb{H}$, whose set is denoted by $\mathfrak{B}(\mathbb{H})$, with operator norm $\|T\| = \sup\{\|Tx\|, x \in \mathbb{H}, \|x\| \leq 1\}$, share the properties of linear operators defined on more algebraic setting. Sometimes it is more efficient to work with explicit representations of operators in $\mathfrak{B}(\mathbb{H})$ (that play the rôle of matrices in the infinite dimensional setting) and some others with abstract algebraic setting.

[1] The set $\mathsf{Dom}(T)$ is generally a *linear manifold*, i.e. algebraically a vector subspace of $\mathbb{V}$ which is not necessarily topologically closed.

Since all operators encountered in quantum mechanics are linear, we drop henceforth the adjective linear.

3.2 Algebra of operators

Definition 3.2.1 An *algebra* is a set $\mathfrak{A}$ endowed with three operations:

1. a *scalar multiplication* $\mathbb{C}\times\mathfrak{A}\ni(\lambda,a)\mapsto\lambda a\in\mathfrak{A}$,
2. a *vector addition* $\mathfrak{A}\times\mathfrak{A}\ni(a,b)\mapsto a+b\in\mathfrak{A}$, and
3. a *vector multiplication* $\mathfrak{A}\times\mathfrak{A}\ni(a,b)\mapsto ab\in\mathfrak{A}$,

such that $\mathfrak{A}$ is a vector space with respect to scalar multiplication and vector addition and a ring (not necessarily commutative) with respect to vector addition and vector multiplication. Moreover, $\lambda(ab)=(\lambda a)b=a(\lambda b)$ for all $\lambda\in\mathbb{C}$ and all $a,b\in\mathfrak{A}$. The algebra is called *commutative* if $ab=ba$, for all $a,b\in\mathfrak{A}$; it is called *unital* if there exists (a necessarily unique) element $e\in\mathfrak{A}$ (often also written $\mathbb{1}$ or $\mathbb{1}_{\mathfrak{A}}$) such that $ae=ea=a$ for all $a\in\mathfrak{A}$;

A linear map from an algebra $\mathfrak{A}_1$ to an algebra $\mathfrak{A}_2$ is a *homomorphism* if it is a ring homomorphism for the underlying rings, it is an *isomorphism* if it is a bijective homomorphism.

Definition 3.2.2 An *involution* on an algebra $\mathfrak{A}$ is a map $\mathfrak{A}\ni a\mapsto a^*\in\mathfrak{A}$ that verifies

1. $(\lambda a+\mu b)^*=\overline{\lambda}a^*+\overline{\mu}b^*$,
2. $(ab)^*=b^*a^*$, and
3. $(a^*)^*=a$.

Involution is also called *adjoint operation* and a^* the *adjoint* of a. An involutive algebra is termed a $*$-*algebra*.

An element $x\in\mathfrak{A}$ is said *normal* if $aa^*=a^*a$, an *isometry* if $a^*a=\mathbb{1}$, *unitary* if both a and a^* are isometries, *self-adjoint or Hermitean* if $a=a^*$. On denoting $h:\mathfrak{A}_1\to\mathfrak{A}_2$ a homomorphism between two $*$-algebras, we call it a $*$-*homomorphism* if it preserves adjoints, i.e. $h(a^*)=h(a)^*$.

A *normed* (respectively *Banach*) algebra $\mathfrak{A}$ is an algebra equipped with a norm map $\|\cdot\|:\mathfrak{A}\to\mathbb{R}_+$ that is a normed (respectively Banach) vector space for the norm and verifies $\|ab\|\leq\|a\|\|b\|$ for all $a,b\in\mathfrak{A}$. $\mathfrak{A}$ is *normed* (respectively *Banach*) $*$-*algebra* if it has an involution verifying $\|a^*\|=\|a\|$ for all $a\in\mathfrak{A}$.

Theorem 3.2.3 *Let $T:\mathbb{H}_1\to\mathbb{H}_2$ be a linear map between two Hilbert spaces $\mathbb{H}_1$ and $\mathbb{H}_2$. Then the following are equivalent:*

1. $\|T\|=\sup\{\|Tf\|_{\mathbb{H}_2}, f\in\mathbb{H}_1, \|f\|_{\mathbb{H}_1}\leq 1\}<\infty$,
2. *T is continuous,*
3. *T is continuous at one point of $\mathbb{H}_1$.*

Proof: Analogous to the proof of the theorem **??** for linear functional. (Please complete the proof!) □

Notation 3.2.4 We denote by $\mathfrak{B}(\mathbb{H}_1,\mathbb{H}_2)$ the algebra of bounded operators with respect to the aforementioned norm:

$$\mathfrak{B}(\mathbb{H}_1,\mathbb{H}_2)=\{T\mathfrak{L}(\mathbb{H}_1,\mathbb{H}_2):\|T\|<\infty\}.$$

When $\mathbb{H}_1=\mathbb{H}_2=\mathbb{H}$, we write simply $\mathfrak{B}(\mathbb{H})$.

Proposition 3.2.5 *Let $\mathbb{H}_1$ and $\mathbb{H}_2$ be two Hilbert spaces and $T\in\mathfrak{B}(\mathbb{H}_1,\mathbb{H}_2)$. Then, there exists a unique bounded operator $T^*:\mathbb{H}_2\to\mathbb{H}_1$ such that*

$$\langle T^*g\,|\,f\rangle=\langle g\,|\,Tf\rangle \text{ for all } f\in\mathbb{H}_1, g\in\mathbb{H}_2.$$

Proof: For each $g\in\mathbb{H}_2$, the map $\mathbb{H}_1\ni f\mapsto\langle g\,|\,Tf\rangle_{\mathbb{H}_2}\in\mathbb{C}$ is a continuous (why?) linear form. By Riesz-Fréchet theorem **??**, there exists a unique $h\in\mathbb{H}_1$ such that $\langle h\,|\,f\rangle_{\mathbb{H}_1}=\langle g\,|\,Tf\rangle_{\mathbb{H}_2}$, for all $f\in\mathbb{H}_1$. Let $T^*:\mathbb{H}_2\to\mathbb{H}_1$ be defined by the assignment $T^*g=h$; it is obviously linear and easily checked to be bounded (exercise!) □

Proposition 3.2.6 *For all $T\in\mathfrak{B}(\mathbb{H}_1,\mathbb{H}_2)$,*

1. $\|T^*\|=\|T\|$,
2. $\|T^*T\|=\|T\|^2$.

Proof:

1. By Cauchy-Schwarz inequality, for all $f \in \mathbb{H}_2$, $g \in \mathbb{H}_1$,

$$\begin{aligned} |\langle f \,|\, Tg \rangle_{\mathbb{H}_2}| &\leq \|f\|_{\mathbb{H}_2} \|Tg\|_{\mathbb{H}_2} \\ &\leq \|T\|_{\mathfrak{B}(\mathbb{H}_1,\mathbb{H}_2)} \|g\|_{\mathbb{H}_1} \|f\|_{\mathbb{H}_2} \end{aligned}$$

   so that

$$\|T\| \geq \sup\{|\langle f \,|\, Tg \rangle| : \|g\| \leq 1, \|f\| \leq 1\}.$$

   Conversely, we may assume that $\|T\| \neq 0$, and therefore choose some $\varepsilon \in]0, \|T\|/2[$. Choose now $g \in \mathbb{H}_1$ with $\|g\| \leq 1$, such that $\|Tg\| \geq \|T\| - \varepsilon$ and $f = \frac{Tg}{\|Tg\|} \in \mathbb{H}_2$, $\|f\| = 1$. For this particular choice of f and g:

$$|\langle f \,|\, Tg \rangle_{\mathbb{H}_2}| \geq \|Tg\| \geq \|T\| - \varepsilon.$$

   Hence,

$$\sup\{|\langle f \,|\, Tg \rangle| : \|g\| \leq 1, \|f\| \leq 1\} \geq \|T\| - \varepsilon.$$

   Since ε is arbitrary, we get $\|T\| = \sup\{|\langle f \,|\, Tg \rangle_{\mathbb{H}_2}| : g \in \mathbb{H}_1, f \in \mathbb{H}_2, \|g\| \leq 1, \|f\| \leq 1\}$. As $\langle f \,|\, Tg \rangle = \langle T^* f \,|\, g \rangle$ for all f and g, we get $\|T^*\| = \|T\|$

2. $\mathfrak{B}(\mathbb{H}_1, \mathbb{H}_2)$ being a normed algebra, $\|T^*T\| \leq \|T^*\|\|T\| = \|T\|^2$. Conversely,

$$\begin{aligned} \|T\|^2 &\leq \sup\{|Tf\| : f \in \mathbb{H}_1, \|f\| \leq 1\} \\ &= \sup\{|\langle Tf \,|\, Tf \rangle| : f \in \mathbb{H}_1, \|f\| \leq 1\} \\ &= \sup\{|\langle f \,|\, T^*Tf \rangle| : f \in \mathbb{H}_1, \|f\| \leq 1\} \\ &\leq \|T^*T\|. \end{aligned}$$

□

Definition 3.2.7 A *C^*-algebra* $\mathfrak{A}$ is an involutive Banach algebra verifying additionally

$$\|a^*a\| = \|a\|^2, \text{ for all } a \in \mathfrak{A}.$$

Example 3.2.8 Let $\mathbb{X}$ be a compact Hausdorff[2] space and $\mathfrak{A} = \{f : \mathbb{X} \to \mathbb{C} \mid f \text{ continuous}\} \equiv C(\mathbb{X})$ Define

1. $\mathbb{C} \times \mathfrak{A} \ni (\lambda, f) \mapsto \lambda f \in \mathfrak{A}$ by $(\lambda f)(x) = \lambda f(x), \forall x \in \mathbb{X}$,
2. $\mathfrak{A} \times \mathfrak{A} \ni (f, g) \mapsto f + g \in \mathfrak{A}$ by $(f + g)(x) = f(x) + g(x), \forall x \in \mathbb{X}$,

[2]Recall that a topological space is called Hausdorff when every two distinct of its points posses disjoint neighbourhoods.

3. $\mathfrak{A} \times \mathfrak{A} \ni (f,g) \mapsto fg \in \mathfrak{A}$ by $(fg)(x) = f(x)g(x), \forall x \in \mathbb{X}$,

4. $\mathfrak{A} \ni f \mapsto f^* \in \mathfrak{A}$ by $f^*(x) = \overline{f(x)}, \forall x \in \mathbb{X}$,

Then $\mathfrak{A}$ is a unital (specify the unit!) C^*-algebra for the norm $\|f\| = \sup_{x\in\mathbb{X}} |f(x)|$. (Prove it!) The algebra $\mathfrak{A}$ is moreover commutative.

Example 3.2.9 Let $\mathbb{H}_1$ and $\mathbb{H}_2$ be two Hilbert spaces. Then $\mathfrak{B}(\mathbb{H}_1, \mathbb{H}_2)$ is a unital C^*-algebra. In general, this algebra is not commutative.

This example has also a converse, given in theorem 3.2.11, below.

Definition 3.2.10 Let $\mathfrak{A}$ be an involutive Banach algebra. A *representation* on a Hilbert space $\mathbb{H}$ of $\mathfrak{A}$ is a $*$-homomorphism of $\mathfrak{A}$ into $\mathfrak{B}(\mathbb{H})$, i.e. a linear map $\pi : \mathfrak{A} \to \mathfrak{B}(\mathbb{H})$ such that

1. $\pi(ab) = \pi(a)\pi(b), \forall a,b \in \mathfrak{A}$,

2. $\pi(a^*) = \pi(a)^*, \forall a \in \mathfrak{A}$,

The space $\mathbb{H}$ is called the *representation space*. We write $(\pi, \mathbb{H})$, or $\mathbb{H}_\pi$ if necessary. Two representations $(\pi_1, \mathbb{H}_1)$ and $(\pi_2, \mathbb{H}_2)$ are said to be *unitarily equivalent* if there exists an isometry $U : \mathbb{H}_1 \to \mathbb{H}_2$ such that for all $a \in \mathfrak{A}$, it holds $U\pi_1(a)U^* = \pi_2(a)$. If moreover for every non zero element of $\mathfrak{A}$, $\pi(a) \neq 0$, then the representation is called *faithful*.

Theorem 3.2.11 (Gel'fand-Naïmark) *If $\mathfrak{A}$ is an arbitrary C^*-algebra, there exists a Hilbert space $\mathbb{H}$ and a linear mapping $\pi : \mathfrak{A} \to \mathfrak{B}(\mathbb{H})$ that is a faithful representation of $\mathfrak{A}$.*

Proof: It can be found in [3], theorem 4.5.6, page 281. □

3.3 Classes of operators

Since any C^*-algebra can be faithfully represented on some Hilbert space $\mathbb{H}$, the different classes of abstract elements of the algebra, introduced in the previous section, have a counterpart in the context of this representation.

3.3.1 Self-adjoint and positive operators

Definition 3.3.1 An operator $T \in \mathfrak{B}(\mathbb{H})$ is called *self-adjoint or Hermitean*[3] if $T = T^*$. The set of Hermitean operators on $\mathbb{H}$ is denoted by $\mathfrak{B}_h(\mathbb{H})$.

Exercise 3.3.2 *The operator $T \in \mathfrak{B}(\mathbb{H})$ is self-adjoint if and only if $\langle f \,|\, Tf \rangle \in \mathbb{R}$ for all $f \in \mathbb{H}$. (Hint: use the polarisation equality* **??**.*)*

Exercise 3.3.3 *If $T \in \mathfrak{B}(\mathbb{H})$ is self-adjoint then $\|T\| = \sup\{\langle f \,|\, Tf \rangle, f \in \mathbb{H}, \|f\| \leq 1\}$.*

Definition 3.3.4 An operator $T \in \mathfrak{B}(\mathbb{H})$ is called *positive* if $\langle f \,|\, Tf \rangle \geq 0$ for all $f \in \mathbb{H}$. Such an operator is necessarily self-adjoint. We denote by $\mathfrak{B}_+(\mathbb{H})$ the set of positive operators.

Exercise 3.3.5 Show that $T \in \mathfrak{B}_+(\mathbb{H})$ if and only if there exists $S \in \mathfrak{B}(\mathbb{H})$ such that $T = S^*S$.

3.3.2 Projections

Definition 3.3.6 An operator $P \in \mathfrak{B}(\mathbb{H})$ is called a *projection* if

1. $P^2 = P$ and
2. $P^* = P$.

Projections are necessary positive (why?). The set of projections is denoted by $\mathfrak{P}(\mathbb{H})$.

Exercise 3.3.7 (A very important one!)

1. Show that there is a bijection between $\mathfrak{P}(\mathbb{H})$ and the set of closed subspaces of $\mathbb{H}$, given by $\mathfrak{P}(\mathbb{H}) \ni P \mapsto P(\mathbb{H}) \subset \mathbb{H}$, $P(\mathbb{H})$ closed.
2. Consequently, show that $\mathfrak{P}(\mathbb{H})$ is partially ordered, i.e. $P_1 \leq P_2$ if $P_1(\mathbb{H})$ subspace of $P_2(\mathbb{H})$ (equivalently $P_1P_2 = P_1$.)

Two projections P_1, P_2 are *orthogonal* if $P_1(\mathbb{H}) \perp P_2(\mathbb{H})$ (equivalently $P_1P_2 = 0$.)

[3] Strictly speaking, the term Hermitean is more general; it applies also to unbounded operators and it means self-adjoint on a dense domain. The two terms coincide for bounded operators.

3.3.3 Isometries

Definition 3.3.8 An operator $T \in \mathfrak{B}(\mathbb{H}_1, \mathbb{H}_2)$ is an *isometry* if $T^*T = \mathbb{1}$ (or equivalently $\|Tf\| = \|f\|$, for all $f \in \mathbb{H}_1$.)

Exercise 3.3.9 Let $\mathbb{H} = \ell^2(\mathbb{N})$ and for $x = (x_1, x_2, x_3, \ldots) \in \mathbb{H}$, define the left and right shifts by

$$Lx = (x_2, x_3, \ldots) \in \mathbb{H},$$

and

$$Rx = (0, x_1, x_2, x_3, \ldots) \in \mathbb{H}.$$

1. Show that $R^* = L$.
2. Show that R is an isometry.
3. Determine $\mathsf{Ran}R$.

This exercise demonstrates that, in infinite dimensional spaces, isometries are not necessarily surjective.

Theorem 3.3.10 *For $T \in \mathfrak{B}(\mathbb{H}_1, \mathbb{H}_2)$, the five following conditions are equivalent:*

1. $(T^*T)^2 = T^*T$,
2. $(TT^*)^2 = TT^*$,
3. $TT^*T = T$,
4. $T^*TT^* = T^*$,
5. *there exist closed subspaces $E_1 \subseteq \mathbb{H}_1$ and $E_2 \subseteq \mathbb{H}_2$ such that $T = I \circ S \circ P$ where $P : \mathbb{H}_1 \to E_1$ is a projection, $S : E_1 \to E_2$ an isometry, and $I : E_2 \to \mathbb{H}_2$ the inclusion map.*

If one (hence all) condition holds then T^*T is the projection $\mathbb{H}_1 \to E_1$ and TT^* is the projection $\mathbb{H}_2 \to E_2$. In this situation T is called a partial isometry with initial space E_1, initial projection T^*T, final space E_2, and final projection TT^*.

Proof: Exercise! (See [1] or [8].) □

3.3.4 Unitary operators

Definition 3.3.11 An operator $U \in \mathfrak{B}(\mathbb{H})$ is *unitary* if $U^*U = UU^* = \mathbb{1}$. The set of unitary operators is denoted by $\mathfrak{U}(\mathbb{H}) = \{U \in \mathfrak{B}(\mathbb{H}) : U^*U = UU^* = \mathbb{1}\}$ (it is in fact a group; for $\mathbb{H} = \mathbb{C}^n$ it is the Lie group denoted by $U(n)$.)

Exercise 3.3.12 Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and $T : \Omega \to \Omega$ a measure preserving transformation i.e. $\mathbb{P}(T^{-1}B) = \mathbb{P}(B)$ for all $B \in \mathscr{F}$. On the Hilbert space $\mathbb{H} = L^2(\Omega, \mathscr{F}, \mathbb{P})$ define $U : \mathbb{H} \to \mathbb{H}$ by $Uf(\omega) = f(T^{-1}\omega)$.

1. Show that U is a partial isometry.
2. Under which condition is U surjective (hence unitary)?

3.3.5 Normal operators

Definition 3.3.13 An operator $T \in \mathfrak{B}(\mathbb{H})$ is *normal* if $T^*T = TT^*$ (or equivalently if $\|T^*f\| = \|Tf\|$ for all $f \in \mathbb{H}$.)

Exercise 3.3.14 A vector $f \in \mathbb{H} \setminus \{0\}$ is called an *eigenvector* corresponding to an *eigenvalue* λ of an operator $T \in \mathfrak{B}(\mathbb{H})$ if $Tf = \lambda f$ for some $\lambda \in \mathbb{C}$. Show that if T is normal and f_1, f_2 are eigenvectors corresponding to different eigenvalues then $f_1 \perp f_2$. (The proof goes as for the finite dimensional case.)

Exercise 3.3.15 Let M be the multiplication operator on $L^2[0,1]$ defined by $Mf(t) = tf(t), t \in [0,1]$. Show that

1. M is self-adjoint (hence normal),
2. M has no eigenvectors.

Exercise 3.3.16 Choose some $z \in \mathbb{C}$ with $|z| < 1$ and consider $z \in \ell^2(\mathbb{N})$ given by $z = (1, z, z^2, z^3, \ldots)$. Let L and R be the left and right shifts defined in exercise 3.3.9.

1. Show that R is not normal,
2. compute R^*z,
3. conclude that R^* has uncountably many eigenvalues.

Chapter 4

Spectral theory in Banach algebras

4.1 Motivation

In linear algebra one often encounters systems of linear equations of the type

$$Tf = g \tag{4.1}$$

with $f, g \in \mathbb{C}^n$ and $T = (t_{i,j})_{i,j=1,\dots,n}$ a $n \times n$ matrix with complex coefficients. Elementary linear algebra establishes that this system of equations has *solutions* provided that the map $Tf \mapsto f$ is surjective and the solution is unique provided that this map is injective. Thus the system has a unique solution for each $g \in \mathbb{C}^n$ provided that the map is bijective, or equivalently the matrix T is invertible. This happens precisely when $\det T \neq 0$. However, this criterion of invertibility is of limited practical use even for the elementary (finite-dimensional) case because det is too complicated an object to be efficiently computed for large n. For infinite dimensional cases, this criterion becomes totally useless since there is no infinite dimensional analogue of det that discriminates between invertible and non-invertible operators T (see exercise 4.1.1 below!)

Another general issue connected with the system (4.1) is that of *eigenvalues*. For every $\lambda \in \mathbb{C}$, denote by $V_\lambda = \{f \in \mathbb{C}^n : Tf = \lambda f\}$. For most choices of λ, the subspace V_λ is the trivial subspace $\{0\}$; this subspace is not trivial only when $T - \lambda \mathbb{1}$ is not injective (i.e. $\ker(T - \lambda \mathbb{1}) \neq \{0\}$.) On defining the *spectrum* of T by

$$\mathsf{spec}(T) = \{\lambda \in \mathbb{C} : T - \lambda \mathbb{1} \text{ is not invertible } \},$$

one easily shows that $\mathsf{spec}(T) \neq \emptyset$ and $\mathsf{card}\,\mathsf{spec}(T) \leq n$ (why?) Not always the family $(V_\lambda)_{\lambda \in \mathsf{spec}(T)}$ spans the whole space $\mathbb{C}^n$. When it does, on decomposing

$g = g^{(1)} + \ldots + g^{(k)}$ where $g^{(j)} \in V_{\lambda_j}$ and $\mathsf{spec}(T) = \{\lambda_1, \ldots, \lambda_k\}$, the solution of (4.1) is given by

$$f = \frac{g^{(1)}}{\lambda_1} + \ldots + \frac{g^{(k)}}{\lambda_k}.$$

(Notice that $\lambda_i \neq 0$, for all $i = 1, \ldots, k$; why?) When the family $(V_\lambda)_{\lambda \in \mathsf{spec}(T)}$ does not span $\mathbb{C}^n$, the problem is more involved but the rôle of the spectrum remains fundamental.

A final issue involving the spectrum of T is the *functional calculus* associated with T. If $p \in \mathbb{R}[t]$, this polynomial can be naturally extended on $\mathfrak{B}(\mathbb{H})$. In fact, if $p(t) = a_n t^n + \ldots a_0$ is the expression of the polynomial p; the expression $p(T) = a_n T^n + \ldots a_0 \mathbb{1}$ is well defined for all $T \in \mathfrak{B}(\mathbb{H})$. Moreover, if $T \in \mathfrak{B}_h(\mathbb{H})$ then $p(T) \in \mathfrak{B}_h(\mathbb{H})$. Suppose now that $T \in \mathfrak{B}_h(\mathbb{H})$, $m = \inf_{\|f\|=1} \langle f \,|\, Tf \rangle$, $M = \sup_{\|f\|=1} \langle f \,|\, Tf \rangle$, and $p(t) \geq 0$ for all $t \in [m, M]$; then $p(T) \in \mathfrak{B}_+(\mathbb{H})$. Now every $f \in C[m, M]$ can be uniformly approximated by polynomials, i.e. there is a sequence $(p_l)_{l \in \mathbb{N}}$, with $p_l \in \mathbb{R}[t]$ such that for all $\varepsilon > 0$, there exists $n_0 \in \mathbb{N}$ such that for $l \geq n_0$, $\max_{t \in [m,M]} |f(t) - p_l(t)| < \varepsilon$. It is natural then to define $f(T) = \lim_l p_l(T)$. However, the computations involved in the right hand side of this equation can be very complicated. Suppose henceforth that $\mathbb{H} = \mathbb{C}^n$ and T is a Hermitean $n \times n$ matrix that is diagonalisable, i.e. $T = UDU^*$ with $D = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$ and U unitary. Then $p_l(T) = U p_l(D) U^*$ and letting $l \to \infty$ we get $f(T) = U f(D) U^*$. Thus, if T is diagonalisable, the computation of $f(T)$ is equivalent to the knowledge of $f(t)$ for $t \in \mathsf{spec}(T)$. For the infinite dimensional case, the problem is more involved but again the spectrum remains fundamental.

The rest of this chapter, based on [1], is devoted to the appropriate generalisation of the spectrum for infinite dimensional operators.

Exercise 4.1.1 (Infinite-dimensional determinant) Let $\mathbb{H} = \ell^2(\mathbb{N})$ and $(t_n)_{n \in \mathbb{N}}$ be a fixed numerical sequence. Suppose that there exist constants $K_1, K_2 > 0$ such that $0 < K_1 \leq t_n \leq K_2 < \infty$ for all $n \in \mathbb{N}$. For every $x \in \ell^2(\mathbb{N})$ define $(Tx)_n = t_n x_n, n \in \mathbb{N}$.

1. Show that $T \in \mathfrak{B}(\mathbb{H})$.

2. Exhibit a bounded operator S on $\mathbb{H}$ such that $ST = TS = \mathbb{1}$.

3. Assume henceforth that $(t_n)_{n \in \mathbb{N}}$ is a monotone sequence. Let $\Delta_n(T) = t_1 \cdots t_n$. Show that $\Delta_n(T)$ converges to a non-zero limit $\Delta(T)$ if and only if $\sum_n (1 - t_n) < \infty$.

4. Any plausible generalisation, δ, of det in the infinite dimensional setting should verify $\delta(\mathbb{1}) = 1$, $\delta(AB) = \delta(A)\delta(B)$, and if T is diagonal $\delta(T) = \Delta(T)$. Choosing $t_n = \frac{n}{n+1}$, for $n \in \mathbb{N}$, conclude that although T is diagonal and invertible, it has $\delta(T) = 0$.

4.2 The spectrum of an operator acting on a Banach space

Let $\mathbb{V}$ be a $\mathbb{C}$-Banach space. Denote by $\mathfrak{B}(\mathbb{V})$ the set of bounded operators $T : \mathbb{V} \to \mathbb{V}$. This space is itself a unital Banach algebra for the induced operator norm.

Exercise 4.2.1 If $\mathbb{X}$ and $\mathbb{Y}$ are metric spaces and $d_{\mathbb{X}}$ and $d_{\mathbb{Y}}$ denote their respective metrics

1. verify that

$$d_p((x_1,y_1),(x_2,y_2)) = (d_{\mathbb{X}}(x_1,x_2)^p + d_{\mathbb{Y}}(y_1,y_2)^p)^{1/p},$$

with $p \in [1,\infty[$ and

$$d_\infty((x_1,y_1),(x_2,y_2)) = \max(d_{\mathbb{X}}(x_1,x_2), d_{\mathbb{Y}}(y_1,y_2))$$

are metrics on $\mathbb{X} \times \mathbb{Y}$; (the corresponding metric space $(\mathbb{X} \times \mathbb{Y}, d_p)$, $p \in [1,\infty]$ is denoted[1] $\mathbb{X} \oplus \mathbb{Y}$)

2. show that the sequence $(x_n, y_n)_n$ in $\mathbb{X} \times \mathbb{Y}$ converges to a point $(\xi, \psi) \in \mathbb{X} \times \mathbb{Y}$ with respect to any of the metrics d_p if and only if $d_{\mathbb{X}}(x_n, \xi) \to 0$ and $d_{\mathbb{Y}}(y_n, \psi) \to 0$.

Exercise 4.2.2 Let $\mathbb{X}$ and $\mathbb{Y}$ be metric spaces and $f : \mathbb{X} \to \mathbb{Y}$ be a continuous map. We denote by

$$\Gamma(f) = \{(x, f(x)) : x \in \mathbb{X}\}$$

the graph of f. Show that $\Gamma(f)$ is closed (i.e. if $(x_n)_n$ is a sequence in $\mathbb{X}$ and if there exists $(x,y) \in \mathbb{X} \times \mathbb{Y}$ such that $x_n \to x$ and $f(x_n) \to y$, then necessarily $y = f(x)$.)

Exercise 4.2.3 (The closed graph theorem) Suppose $\mathbb{X}$ and $\mathbb{Y}$ are Banach spaces and $T : \mathbb{X} \to \mathbb{Y}$ a linear map having closed graph. Show that T is continuous.

Theorem 4.2.4 *For every $T \in \mathfrak{B}(\mathbb{V})$, the following are equivalent:*

[1] more precisely $\mathbb{X} \oplus_{\ell^p} \mathbb{Y}$.

1. *for every $y \in \mathbb{V}$ there is a unique $x \in \mathbb{V}$ such that $Tx = y$,*

2. *there is an operator $S \in \mathfrak{B}(\mathbb{V})$ such that $ST = TS = \mathbb{1}$.*

Proof: Only the part $1 \Rightarrow 2$ is not trivial to show. Condition 1 implies that T is invertible; call S its inverse. The only thing to show is the boundedness of S. As a subset of $\mathbb{V} \oplus \mathbb{V}$, the graph of S is related to the graph of T. In fact

$$\Gamma(S) = \{(y, Sy) : y \in \mathbb{V}\} = \{(Tx, x), x \in \mathbb{V}\}.$$

Now T is bounded, hence continuous, so that that the set $\{(Tx,x), x \in \mathbb{V}\}$ is closed (see exercise 4.2.2.) Thus the graph of S is closed, and by the closed graph theorem (see exercise 4.2.3), S is continuous hence bounded. □

Definition 4.2.5 Let $T \in \mathfrak{B}(\mathbb{V})$ where $\mathbb{V}$ is a Banach space.

1. T is called *invertible* if there exists an operator $S \in \mathfrak{B}(\mathbb{V})$ such that $ST = TS = \mathbb{1}$.

2. The *spectrum* of T, denoted by $\mathsf{spec}(T)$, is defined by

$$\mathsf{spec}(T) = \{\lambda \in \mathbb{C} : T - \lambda \mathbb{1} \text{ is not invertible}\}.$$

3. The *resolvent set* of T, denoted by $\mathsf{Res}(T)$, is defined by

$$\mathsf{Res}(T) = \mathbb{C} \setminus \mathsf{spec}(T).$$

Notice that in finite dimension, invertibility of an operator R reduces essentially to injectivity of R since surjectivity of R can be trivially verified if we reduce the space V into $\mathsf{Ran}(R)$. In infinite dimension, several things can go wrong: of course injectivity may fail as in finite dimension; but a new phenomenon can appear when $\mathsf{Ran}(R)$ is not closed: in this latter case, $\mathsf{Ran}(R)$ can further be dense in V or fail to be dense in V. All these situations may occur and correspond to different types of sub-spectra.

Definition 4.2.6 Let $T \in \mathfrak{B}(\mathbb{V})$ where $\mathbb{V}$ is a Banach space.

1. The *point spectrum* of T is defined by $\mathsf{spec}_p(T) = \{\lambda \in \mathbb{C} : T - \lambda \mathbb{1} \text{ is not injective}\}$. Every $\lambda \in \mathsf{spec}_p(T)$ is called an *eigenvalue* of T.

2. The *continuous spectrum*, $\mathsf{spec}_c(T)$, of T is defined as the complex values λ such that $T - \lambda \mathbb{1}$ is injective but not surjective and $\mathsf{Ran}(T - \lambda \mathbb{1})$ is dense in $\mathbb{V}$.

3. The *residual spectrum*, $\mathsf{spec}_(T)$, of T is defined as the complex values λ such that $T-\lambda\mathbb{1}$ is injective but not surjective and $\mathsf{Ran}(T-\lambda\mathbb{1})$ is not dense in $\mathbb{V}$.

Example 4.2.7 Let $\mathbb{V}$ be a finite dimensional Banach space and $T:\mathbb{V}\to\mathbb{V}$ a linear transformation (hence bounded.) Since $\dim\ker(T-\lambda\mathbb{1})+\dim\mathsf{Ran}(T-\lambda\mathbb{1})=\dim\mathbb{V}$, it follows that $T-\lambda\mathbb{1}$ is injective if and only if $\mathsf{Ran}(T-\lambda\mathbb{1})=\mathbb{V}$. Therefore $\mathsf{spec}_r(T)=\emptyset$. Further, if $T-\lambda\mathbb{1}$ is injective, then it has an inverse on $\mathbb{V}$. Since any linear transformation of a finite dimensional space is continuous, it follows that $(T-\lambda\mathbb{1})^{-1}$ is continuous, hence $\mathsf{spec}_c(T)=\emptyset$. Therefore, in finite dimension we always have $\mathsf{spec}(T)=\mathsf{spec}_p(T)$.

Exercise 4.2.8 Let $\mathbb{V}=\ell^2(\mathbb{N})$ and consider the right shift, R, on $\mathbb{V}$.

1. Show that $R-\lambda\mathbb{1}$ is injective for all $\lambda\in\mathbb{C}$. Conclude that $\mathsf{spec}_p(R)=\emptyset$.
2. Show that for $|\lambda|>1$, $\mathsf{Ran}(R-\lambda\mathbb{1})=\mathbb{V}$. Conclude that all $\lambda\in\mathbb{C}$ with $|\lambda|>1$ belong to $\mathsf{Res}(R)$.
3. For $|\lambda|<1$, show that $\mathsf{Ran}(R-\lambda\mathbb{1})$ is orthogonal to the vector $\Lambda=(1,\lambda,\lambda^2,\ldots)$. Show that for $|\lambda|<1$, $\mathsf{Ran}(R-\lambda\mathbb{1})=\{y\in\mathbb{V}:y\perp\Lambda\}$. Conclude that all $\lambda\in\mathbb{C}$ with $|\lambda|<1$ belong to $\mathsf{spec}_r(R)$.
4. The case $|\lambda|=1$ is the most difficult. Try to show that $\mathsf{Ran}(R-\lambda\mathbb{1})$ is dense in $\mathbb{V}$ so that the unit circle coincides with $\mathsf{spec}_c(R)$.

4.3 The spectrum of an element of a Banach algebra

In the previous section we studied spectra of bounded operators acting on Banach spaces. They form a Banach algebra with unit. Spectral theory can be established also abstractly on Banach algebras. Before stating spectral properties, it is instructive to give some more examples.

Example 4.3.1 Let $C_K(\mathbb{R})$ be the set of continuous functions on $\mathbb{R}$ which vanish outside a bounded interval; it is a normed vector space (with respect to the L^1 norm for instance; its completion is the Banach space $L^1(\mathbb{R},\lambda)$, where λ stands for the Lebesgue measure.) A product can be defined by the convolution

$$f\star g(x)=\int_{\mathbb{R}}f(y)g(x-y)dy$$

turning this space into a commutative Banach algebra. This algebra is not unital (this can be seen by solving the equation $f \star f = f$ in L^1), but it has an approximate unit (i.e. a sequence $(f_n)_n$ of integrable functions with $\|f_n\| = 1$ for all n and such that for all $g \in L^1(\mathbb{R})$, $\|g \star f_n - g\| \to 0$. (Give an explicit example of such an approximate unit!)

Example 4.3.2 The algebra $\mathbf{M}_n(\mathbb{C})$ is a unital non-commutative algebra. There are many norms that turn it into a finite-dimensional Banach algebra, for instance:

1. $\|A\| = \sum_{i,j=1}^n |a_{i,j}|$
2. $\|A\| = \sup_{\|x\| \leq 1} \frac{\|Ax\|}{\|x\|}$.

Definition 4.3.3 Let $\mathfrak{A}$ be a unital Banach algebra. (We can always assume that $\|\mathbb{1}\| = 1$, may be after re-norming the elements of $\mathfrak{A}$.) An element $a \in \mathfrak{A}$ is called *invertible* if there is an element $b \in \mathfrak{A}$ such that $ab = ba = \mathbb{1}$. The set of all invertible elements of $\mathfrak{A}$ is denoted by $\mathsf{GL}(\mathfrak{A})$ and called the *general linear group of invertible elements* of $\mathfrak{A}$.

Theorem 4.3.4 *Let $\mathfrak{A}$ be a unital Banach algebra. If $a \in \mathfrak{A}$ and $\|a\| < 1$ then $\mathbb{1} - a$ is invertible and*

$$(\mathbb{1} - a)^{-1} = \sum_{n=0}^{\infty} a^n.$$

Moreover,

$$\|(\mathbb{1} - a)^{-1}\| \leq \frac{1}{1 - \|a\|}$$

and

$$\|\mathbb{1} - (\mathbb{1} - a)^{-1}\| \leq \frac{\|a\|}{1 - \|a\|}.$$

Proof: Since $\|a^n\| \leq \|a\|^n$ for all n, we can define $b \in \mathfrak{A}$ as the sum of the absolutely convergent series $b = \sum_{n=0}^{\infty} a^n$. Moreover, $b(\mathbb{1} - a) = (\mathbb{1} - a)b = \lim_{N \to \infty} \sum_{n=0}^{N} b^n = \lim_{N \to \infty}(\mathbb{1} - b^{N+1}) = \mathbb{1}$. Hence $\mathbb{1} - a$ is invertible and $(\mathbb{1} - a)^{-1} = b$. The first majorisation holds because $\|b\| \leq \sum_{n=0}^{\infty} \|a\|^n = \frac{1}{1-\|a\|}$. The second one follows from remarking that $\mathbb{1} - b = -\sum_{n=1}^{\infty} a^n = -ab$, hence $\|\mathbb{1} - b\| \leq \|a\|\|b\|$. □

Exercise 4.3.5

1. Prove that $\mathsf{GL}(\mathfrak{A})$ is an open set in $\mathfrak{A}$ and that the mapping $a \mapsto a^{-1}$ is continuous on $\mathsf{GL}(\mathfrak{A})$.
2. Justify the term "general linear group" of invertible elements, i.e. show that $\mathsf{GL}(\mathfrak{A})$ is a topological group in the relative norm topology.

Definition 4.3.6 Let $\mathfrak{A}$ be a unital Banach algebra. For every $a \in \mathfrak{A}$, the *spectrum* of a is the set

$$\mathsf{spec}(a) = \{\lambda \in \mathbb{C} : a - \lambda \mathbb{1} \notin \mathsf{GL}(\mathfrak{A})\}.$$

In the rest of this section, $\mathfrak{A}$ will be a unital algebra and we shall write $a - \lambda$ instead of $a - \lambda \mathbb{1}$.

Proposition 4.3.7 *For every $a \in \mathfrak{A}$, the set $\mathsf{spec}(a)$ is a closed subset of the disk $\{\lambda \in \mathbb{C} : |\lambda| \leq \|a\|\}$.*

Proof: Consider the resolvent set

$$\mathsf{Res}(a) = \{\lambda \in \mathbb{C} : a - \lambda \in \mathsf{GL}(A)\} = \mathbb{C} \setminus \mathsf{spec}(a).$$

Since the set $\mathsf{GL}(\mathfrak{A})$ is open (see exercise 4.3.5) and the map $\mathbb{C} \ni \lambda \mapsto a - \lambda \in \mathfrak{A}$ continuous, the set $\mathsf{Res}(a)$ is open hence the set $\mathsf{spec}(a)$ is closed. Moreover, if $|\lambda| > \|a\|$, on writing $a - \lambda = (-\lambda)[1 - a/\lambda]$ and remarking that $\|a/\lambda\| < 1$, we conclude that $a - \lambda \in \mathsf{GL}(\mathfrak{A})$. □

Theorem 4.3.8 *For every $a \in \mathfrak{A}$, the set $\mathsf{spec}(a)$ is non-empty.*

Proof: For $\lambda_0 \notin \mathsf{spec}(a)$, the $\mathfrak{A}$-valued function $\lambda \mapsto (a - \lambda)^{-1}$ is well defined for all λ sufficiently close to λ_0 because the set $\mathsf{Res}(a)$ is open. Moreover, for $\lambda, \lambda_0 \in \mathsf{Res}(a)$,

$$\begin{aligned}(a-\lambda)^{-1} - (a-\lambda_0)^{-1} &= (a-\lambda)^{-1}[(a-\lambda_0) - (a-\lambda)](a-\lambda_0)^{-1} \\ &= (\lambda - \lambda_0)(a-\lambda)^{-1}(a-\lambda_0)^{-1}.\end{aligned}$$

Thus

$$\lim_{\lambda \to \lambda_0} \frac{1}{\lambda - \lambda_0}[(a-\lambda) - (a-\lambda_0)] = (a-\lambda_0)^{-2}.$$

Assume now that $\mathsf{spec}(a) = \emptyset$ and choose an arbitrary bounded linear functional $\phi : \mathfrak{A} \to \mathbb{C}$. Then, the scalar function $f : \mathbb{C} \to \mathbb{C}$ defined by $\lambda \mapsto f(\lambda) = \phi((a-\lambda)^{-1})$ is defined on the whole $\mathbb{C}$. By linearity, the function f has everywhere a complex derivative, satisfying $f'(\lambda) = \phi((a-\lambda)^{-2})$. Thus f is an entire function. Notice moreover that f is bounded and for $|\lambda| > \|a\|$, by theorem 4.3.4,

$$\begin{aligned}\|(a-\lambda)^{-1}\| &= \frac{\|(1-a/\lambda)^{-1}\|}{|\lambda|} \\ &\leq \frac{1}{|\lambda|(1 - \|a\|/|\lambda|)} \\ &= \frac{1}{|\lambda| - \|a\|}.\end{aligned}$$

Thus $\lim_{\lambda\to\infty} f(\lambda) = 0$ and since this function is bounded and entire, by Liouville's theorem (see [**?**] for instance), it is constant, hence $f(\lambda) = 0$ for all $\lambda \in \mathbb{C}$ and every linear functional ϕ. The Hahn-Banach theorem implies then that $(a-\lambda)^{-1} = 0$ for all $\lambda \in \mathbb{C}$. But this is absurd because $(a-\lambda)$ is invertible and $\mathbb{1} \neq 0$ in $\mathfrak{A}$. □

Definition 4.3.9 For every $a \in \mathfrak{A}$, the *spectral radius* of a is defined by $r(a) = \sup\{|\lambda| : \lambda \in \mathsf{spec}(a)\}$.

Exercise 4.3.10
1. Let $p \in \mathbb{R}[t]$ and $a \in \mathfrak{A}$. Show that $p(\mathsf{spec}(a)) \subseteq \mathsf{spec}(p(a))$. (Hint: if $\lambda \in \mathsf{spec}(a)$, the map $\lambda' \mapsto p(\lambda') - p(\lambda)$ is a polynomial vanishing at $\lambda' = \lambda$. Conclude that $p(a) - p(\lambda)$ cannot be invertible.)
2. For every $a \in \mathfrak{A}$ show that $r(a) = \lim_{n\to\infty} \|a^n\|^{1/n}$.

4.4 Relation between diagonalisability and the spectrum

Motivated again by elementary linear algebra, we recall that a self-adjoint $n \times n$ matrix T can be diagonalised, i.e. it is possible to find a diagonal matrix $D = \begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{pmatrix}$ and a unitary matrix U such that $T = UDU^*$; we have then $\mathsf{spec}(T) = \{d_1, \ldots, d_n\}$. We shall generalise this result to infinite dimensional spaces.

An orthonormal basis for $\mathbb{H}$ is a sequence $\mathscr{E} = (e_1, e_2, \ldots)$ of mutually orthogonal unit vectors of $\mathbb{H}$ such that[2] $\overline{\mathsf{span}}\mathscr{E} = \mathbb{H}$. On fixing such a basis, we define a unitary operator $U : \ell^2(\mathbb{N}) \to \mathbb{H}$ by

$$Uf = \sum_{i\in\mathbb{N}} f_i e_i$$

for $f = (f_1, f_2, \ldots)$. Specifying a particular orthonormal basis in $\mathbb{H}$ is equivalent to specifying a particular unitary operator U. Suppose now that $T \in \mathfrak{B}(\mathbb{H})$ is a normal operator and admits the basis vectors of $\mathscr{E}$ as eigenvectors, i.e. $Te_k = t_k r_k$, $t_k \in \mathbb{C}$, $k \in \mathbb{N}$. Then $t = (t_k)_k \in \ell^\infty(\mathbb{N})$ and $U^*TU = M$ where M is the multiplication operator defined by $(Mf)_k = (U^*TUf)_k = (U^{-1}TUf)_k = (U^{-1}T\sum_i f_i e_i)_k = f_k t_k$. Thus an operator T on $\mathbb{H}$ is diagonalisable in a given basis $\mathscr{E}$ if the unitary operator associated with $\mathscr{E}$ implements an equivalence between T and a multiplication operator M acting on $\ell^2(\mathbb{N})$. This notion is still inadequate since it involves only normal operators with pure point spectrum; it can nevertheless be appropriately generalised.

[2] Recall that $\mathbb{H}$ is always considered separable.

Definition 4.4.1 An operator T acting on a Hilbert space $\mathbb{H}$ is said *diagonalisable* if there exist a (necessarily separable) σ-finite measure space $(\Omega, \mathscr{F}, \mu)$, a function $m \in L^\infty(\Omega, \mathscr{F}, \mu)$, and a unitary operator $U : L^2(\Omega, \mathscr{F}, \mu) \to \mathbb{H}$ such that

$$UM_m = TU$$

where M_m denotes the multiplication operator by m, defined by $M_m f(\omega) = m(\omega) f(\omega)$, for all $\omega \in \Omega$ and all $f \in L^2(\Omega, \mathscr{F}, \mu)$

Example 4.4.2 Let $\mathbb{H} = L^2([0,1])$ and $T : \mathbb{H} \to \mathbb{H}$ defined by $Tf(t) = tf(t)$, for $t \in [0,1]$ and $f \in \mathbb{H}$. This operator is diagonalisable since it is already a multiplication operator.

Notice that a diagonalisable operator is always normal because the multiplication operator is normal. The following theorem asserts the converse.

Theorem 4.4.3 *Every normal operator acting on a Hilbert space is diagonalisable.*

Proof: Long but without any particular difficulty; it can be found in [1], pp. 52–55. □

4.5 Spectral measures and functional calculus

Start again from some heuristic ideas. Let $(\Omega, \mathscr{F}, \mu)$ be a probability space and $f : \Omega \to \mathbb{R}$ a bounded measurable function. Standard integration theory states that f can be approximated by simple functions. More precisely, for every $\varepsilon > 0$, there exists a finite family $(E_i)_i$ of disjoint measurable sets $E_i \in \mathscr{F}$ and a finite family of real numbers $(\alpha_i)_i$ such that $|f(\omega) - \sum_i \alpha_i \mathbb{1}_{E_i}(\omega)| < \varepsilon$ for all $\omega \in \Omega$. It is instructive to recall the main idea of the proof of this elementary result.

Let $m = \inf f(\omega)$, $M = \sup f(\omega)$, and subdivide the interval $[m, M]$ into a finite family of disjoint intervals $(I_j)_j$, with $|I_j| < \varepsilon$ (see figure 4.1.) For each j, select an arbitrary $\alpha_j \in I_j$; in the subset $f^{-1}(I_j) \in \mathscr{F}$, the values of f lie within ε from α_j. Therefore, we get the desired result by setting $E_j = f^{-1}(I_j)$. If for every Borel set $B \in \mathscr{B}(\mathbb{R})$, we define $P(B) = \mathbb{1}_{f^{-1}(B)}$ (this is a function-valued set function!), the approximation result can be rewritten as

$$|f(\omega) - \sum_j \alpha_j P(I_j)(\omega)| < \varepsilon, \ \forall \omega \in \Omega.$$

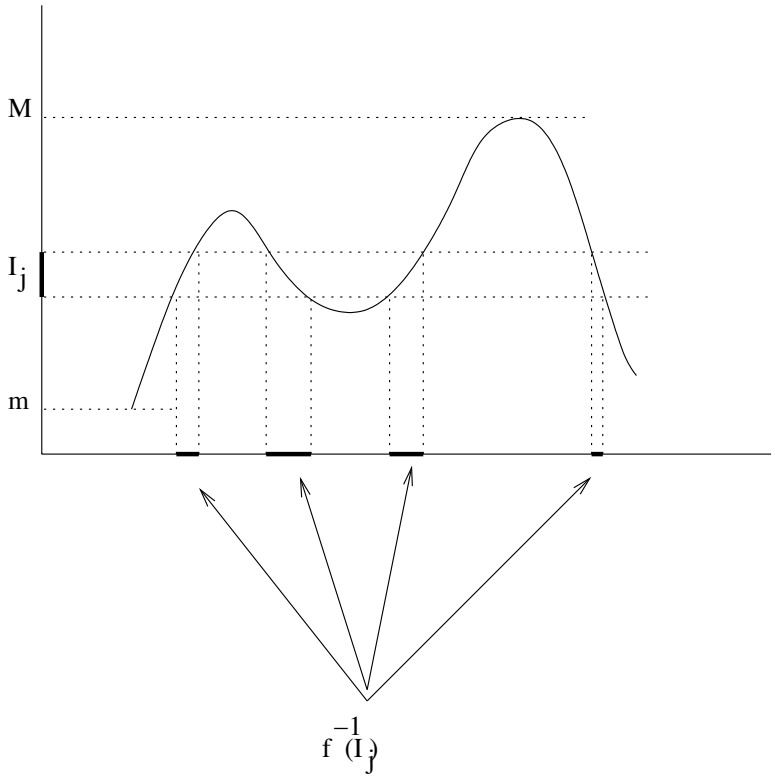

Figure 4.1: The approximation of a bounded measurable function by simple functions

Now, P is set function (a measure actually) and the sum $\sum_j \alpha_j P(I_j)$ tends (in some sense[3]) to $\int \alpha P(d\alpha)$.

Summarising the heuristics developed so far: the approximability of a real-valued, bounded, measurable function f by simple functions can be expressed by writing $f = \int \alpha P(d\alpha)$, where P is the *function-valued measure*

$$\mathscr{B}(\mathbb{R}) \ni B \mapsto P(B) = \mathbb{1}_{f^{-1}(B)}$$

with the following properties:

1. P is idempotent: i.e. $P(B)^2 = P(B)$ for all $B \in \mathscr{B}(\mathbb{R})$,
2. P is multiplicative: i.e. $P(B \cap C) = P(B)P(C)$ for all $B, C \in \mathscr{B}(\mathbb{R})$,

[3] As a matter of fact, it is possible to construct a descent theory of integration in which $\int \alpha P(d\alpha)$ acquires a precise meaning.

3. P is supported by $\mathsf{Ran}(f)$: i.e. $P(B) \equiv 0$ for all $B \in \mathscr{B}(\mathbb{R})$ such that $B \cap \mathsf{Ran}(f) = \emptyset$.

The measure P reflects the properties of f; it is called the *spectral measure of* f.

In the non-commutative setting, the analogue of a bounded, real-valued, measurable function is a bounded Hermitean operator on $\mathbb{H}$. Idempotence, characterising indicators in the commutative case, is verified by projections belonging to $\mathfrak{P}(\mathbb{H})$. Hence, we are seeking approximations of bounded Hermitean operators by complex finite combinations of projections. Now we can turn into precise definitions.

Definition 4.5.1 Let $(\mathbb{X}, \mathscr{F})$ be a measurable space and $\mathbb{H}$ a Hilbert space. A function $P : \mathscr{F} \to \mathfrak{P}(\mathbb{H})$ is called a *spectral measure* on $(\mathbb{X}, \mathscr{F})$ if

1. $P(\mathbb{X}) = \mathbb{1}$,
2. if $(F_n)_{n\in\mathbb{N}}$ is a sequence of disjoint elements in $\mathscr{F}$, then $P(\cup_{n\in\mathbb{N}} F_n) = \sum_{n\in\mathbb{N}} P(F_n)$.

Example 4.5.2 Let $(\mathbb{X}, \mathscr{F}, \mu)$ be a probability space and $\mathbb{H} = L^2(\mathbb{X}, \mathscr{F}, \mu)$. Then the mapping $\mathscr{F} \ni F \mapsto P(F) \in \mathfrak{P}(\mathbb{H})$, defined by $P(F)f = \mathbb{1}_F f$ for all $f \in \mathbb{H}$, is a spectral measure.

Exercise 4.5.3 If P is a spectral measure on $(\mathbb{X}, \mathscr{F})$, then $P(\emptyset) = 0$ and P is finitely disjointly additive.

Theorem 4.5.4 *Let $(\mathbb{X}, \mathscr{F})$ be a measurable space and $\mathbb{H}$ a Hilbert space. If P is a finitely disjointly additive function $\mathscr{F} \to \mathfrak{P}(\mathbb{H})$ such that $P(\mathbb{X} = \mathbb{1}$ then (for $F, G \in \mathscr{F}$)*

1. *P is monotone: $F \subseteq G \Rightarrow P(F) \leq P(G)$,*
2. *P is subtractive: $F \subseteq G \Rightarrow P(G \setminus F) = P(G) - P(F)$,*
3. *P is modular: $P(F \cup G) + P(F \cap G) = P(F) + P(G)$,*
4. *P is multiplicative: $P(F \cap G) = P(F)P(G)$.*

Proof: The statements 1 and 2 are immediate by noticing that $F_1 \subseteq F_2 \Rightarrow F_2 = F_1 \sqcup (F_2 \setminus F_1)$.

3) Since $F \cup G = (F \setminus G) \sqcup (F \cap G) \sqcup (G \setminus F)$ we have: $P(F \cup G) + P(F \cap G) = [P(F \setminus G) + P(F \cap G)] + [P(G \setminus F) + P(G \cap F)] = P(F) + P(G)$.

4) By 1)

$$P(F \cap G) \leq P(F) \leq P(F \cup G). \qquad (*)$$

Multiplying the first inequality of (*) by $P(F \cap G$, we get $P(F \cap G) \leq P(F)P(F \cap G)$ and since $P(F) \leq \mathbb{1}$, the right hand side of the latter inequality is bounded further by $P(F \cap G)$. Hence $P(F)P(F \cap G) = P(F \cap G)$. Similarly, multiplying the second inequality of (*) by $P(F)$ and since again $P(F \cup G) \leq \mathbb{1}$, we get $P(F)P(F \cup G) = P(F)$. Adding the thus obtained equalities, we get:

$$P(F)[P(F \cup G) + P(F \cap G)] = P(F \cap G) + P(F)$$

and we conclude by modularity. □

Exercise 4.5.5 Show that for all $F, G \in \mathscr{F}$, we have $[P(F), P(G)] = 0$.

Theorem 4.5.6 *Let $(\mathbb{X}, \mathscr{F})$ be a measurable space and $\mathbb{H}$ a Hilbert space. A map $P : \mathscr{F} \to \mathfrak{P}(\mathbb{H})$ is a spectral measure if and only if*

1. *$P(\mathbb{X}) = \mathbb{1}$, and*
2. *for all $f, g \in \mathbb{H}$, the set function $\mu_{f,g} : \mathscr{F} \to \mathbb{C}$, defined by*

   $$\mu_{f,g}(F) = \langle f \,|\, P(F)g \rangle, F \in \mathscr{F},$$

   is countably additive.

Proof:

($\Rightarrow$): If P is a spectral measure, then statements 1 and 2 hold trivially.

($\Leftarrow$): Suppose, conversely, that 1 and 2 hold. If $F \cap G = \emptyset$ then $\langle f \,|\, P(F \cup G)g \rangle = \langle f \,|\, P(F)g \rangle + \langle f \,|\, P(G)g \rangle = \langle f \,|\, [P(F) + P(G)]g \rangle$, hence P is finitely additive (hence multiplicative). Let now $(F_n)_n$ be a sequence of disjoint sets in $\mathscr{F}$. Multiplicativity of P implies $(P(F_n))_n$ is a sequence of orthogonal projections and hence $(P(F_n)g)_n$ a sequence of orthogonal vectors for any $g \in \mathbb{H}$. Let $F = \cup_n F_n$. Hence, for all $f, g \in \mathbb{H}$, we have: $\langle f \,|\, P(F)g \rangle = \langle f \,|\, \sum_n P(F_n)g \rangle$, due to the countable additivity property of $\mu - f, g$. We are tempted to conclude that $P(F) = \sum_n P(F_n)$. Yet, it may happen that $\sum_n P(F_n)$ does not make any sense because weak convergence does not imply convergence in the operator norm. However, $\sum_n \|P(F_n)g\|^2 = \sum_n \langle g \,|\, P(F_n)g \rangle = \langle g \,|\, P(F)g \rangle = \|P(F)g\|^2$. It follows that the sequence $(P(f_n)g)_n$ is summable. If we write $\sum_n P(f_n)g = Tg$, it defines a bounded operator T coinciding with $P(F)$.

□

Notation 4.5.7 Let $(\mathbb{X}, \mathscr{F})$ be a measurable space and $F : \mathbb{X} \to \mathbb{C}$. We denote by $\|F\| \equiv \sup\{|F(x)| : x \in \mathbb{X}\}$, and $\mathfrak{B}(\mathbb{X}) = \{F : \mathbb{X} \to \mathbb{C} \mid \text{measurable}, \ \|F\| < \infty\}$.

Henceforth, the Hilbert space $\mathbb{H}$ will be fixed and $\mathfrak{B}(\mathbb{H})$ (respectively $\mathfrak{P}(\mathbb{H})$) will denote as usual the set of bounded operators (respectively projections) on $\mathbb{H}$.

Theorem 4.5.8 *Let $(\mathbb{X}, \mathscr{F})$ be a measurable space and $\mathbb{H}$ a Hilbert space. If P is a spectral measure on $(\mathbb{X}, \mathscr{F})$ and $F \in \mathfrak{B}(\mathbb{X})$, then there exists a unique operator $T_F \in \mathfrak{B}(\mathbb{H})$ such that*

$$\langle f \,|\, T_F g \rangle = \int_{\mathbb{X}} F(x) \langle f \,|\, P(dx) g \rangle,$$

for all $f, g \in \mathbb{H}$. We write $T_F = \int_{\mathbb{X}} F(x) P(dx)$.

Proof: The boundedness of F implies that the right hand side of the integral gives rise to a well-defined sesquilinear functional $\phi(f,g) = \int_{\mathbb{X}} F(x) \langle f \,|\, P(dx) g \rangle$, for $f, g \in \mathbb{H}$. Moreover, $|\phi(f,f)| \leq \int_{\mathbb{X}} |F(x)| \|P(dx) f\|^2 \leq \|F\| \|f\|^2$, hence the functional ϕ is bounded. Existence and uniqueness of T_F follows from the Riesz-Fréchet theorem. □

Theorem 4.5.9 (Spectral decomposition theorem) *If $T \in \mathfrak{B}_h(\mathbb{H})$ then there exists a spectral measure on $(\mathbb{C}, \mathscr{B}(\mathbb{C}))$, supported by $\mathsf{spec}(T) \subseteq \mathbb{R}$, such that*

$$T = \int_{\mathsf{spec}(T)} \lambda P(\lambda).$$

Proof: Let $p \in \mathbb{R}[t]$ and $f, g \in \mathbb{H}$ be two arbitrary vectors. Denote by $L_{f,g}(p) = \langle f \,|\, p(T) g \rangle$. Then $|L_{f,g}(p)| \leq \|p(T)\| \|f\| \|g\|$ and since $p(T) \in \mathfrak{B}(\mathbb{H})$ we have also $\|p(T)\| = \sup\{|p(\lambda)| : \lambda \in \mathsf{spec}(T)\}$ (exercise!). Since $\mathsf{spec}(T)$ is a bounded set, $\|p(T)\| < \infty$ for all $p \in \mathbb{R}[t]$. Hence the linear functional $L_{f,g}$ is a bounded linear functional on $\mathbb{R}[t]$. By Riesz-Fréchet theorem, there exists consequently a unique complex measure $\mu_{f,g}$, supported by $\mathsf{spec}(T)$, such that

$$L_{f,g}(p) \equiv \langle f \,|\, p(T) g \rangle = \int_{\mathsf{spec}(T)} p(\lambda) \mu_{f,g}(d\lambda),$$

for all $p \in \mathbb{R}[t]$, verifying $|\mu_{f,g}(B)| \leq \|f\| \|g\|$, for all $B \in \mathscr{B}(\mathbb{C})$. Using the uniqueness of $\mu_{f,g}$, it is immediate to show that for every $B \in \mathscr{B}(\mathbb{C})$, $S_B(f,g) = \mu_{f,g}(B)$ is a sesquilinear form. Now, $|S_B(f,g)| = |\mu_{f,g}(B)| \leq \|f\| \|g\|$, for all B. Hence the sesquilinear form is bounded; therefore, there exists an operator $P(B) \in \mathfrak{B}_h(\mathbb{H})$

such that $S_B(f,g) = \langle f \,|\, P(B)g \rangle$ for all $f,g \in \mathbb{H}$. Recall that neither $\mu_{f,g}$, nor S_B, nor P depend on the initially chosen polynomial p. Choosing $p_0(\lambda) = 1$, we get $\int_{\mathsf{spec}(T)} \langle f \,|\, P(d\lambda)g \rangle = \langle f \,|\, P(\mathsf{spec}(T))g \rangle = \langle f \,|\, g \rangle$ and choosing $p_1(\lambda) = \lambda$, we get $\int_{\mathsf{spec}(T)} \langle f \,|\, \lambda P(d\lambda)g \rangle = \langle f \,|\, Tg \rangle$, for all $f,g \in \mathbb{H}$. To complete the proof, it remains to show that P is a projection-valued measure. It is enough to show the multiplicativity property. For any fixed pair $f,g \in \mathbb{H}$ and any fixed real polynomial q, introduce the auxiliary complex measure $\nu(B) = \int_B q(\lambda)\langle f \,|\, P(d\lambda)g \rangle$, with $B \in \mathscr{B}(\mathbb{C})$. For every real polynomial p, we have

$$\begin{aligned} \int p(\lambda)\nu(d\lambda) &= \int p(\lambda)q(\lambda)\langle f \,|\, P(d\lambda)g \rangle \\ &= \langle f \,|\, P(p(T)q(T)g \rangle \\ &= \langle q(T)f \,|\, p(T)g \rangle \\ &= \int p(\lambda)\langle q(T)f \,|\, P(d\lambda)g \rangle. \end{aligned}$$

Therefore,

$$\begin{aligned} \nu(B) &= \int q(\lambda)\mathbb{1}_B(\lambda)\langle f \,|\, P(d\lambda)g \rangle \\ &= \langle q(T)f \,|\, P(B)g \rangle \\ &= \langle f \,|\, q(T)P(B)g \rangle \\ &= \int q(\lambda)\langle f \,|\, P(d\lambda)P(B)g \rangle. \end{aligned}$$

Since q is arbitrary,

$$\begin{aligned} \langle f \,|\, P(B\cap C)g \rangle &= \int_C \langle f \,|\, P(d\lambda)P(B)g \rangle \\ &= \langle f \,|\, P(B)P(C)g \rangle, \end{aligned}$$

and since $f,g \in \mathbb{H}$ are arbitrary, we get $P(B\cap C) = P(B)P(C)$. □

Theorem 4.5.10 *If T is a normal operator in $\mathfrak{B}(\mathbb{H})$, then there exists a necessarily unique complex spectral measure on $(\mathbb{C}, \mathscr{B}(\mathbb{C}))$, supported by $\mathsf{spec}(T)$, such that*

$$T = \int_{\mathsf{spec}(T)} \lambda P(d\lambda).$$

Proof: Exercise! (Hint: $T = T_1 + iT_2$ with $T_1, T_2 \in \mathfrak{B}_h(\mathbb{H})$.) □

4.6 Some basic notions on unbounded operators

The operators arising in quantum mechanics are very often unbounded.

Definition 4.6.1 Let $\mathbb{H}$ be a Hilbert space. An *operator* on $\mathbb{H}$, possibly unbounded, is a pair $(\mathsf{Dom}(T), T)$ where $\mathsf{Dom}(T) \subseteq \mathbb{H}$ is a linear manifold and $T : \mathsf{Dom}(T) \to \mathbb{H}$ is a linear map. The set of operators on $\mathbb{H}$ is denoted $\mathfrak{L}(\mathbb{H})$.

The *graph* of an operator $T \in \mathfrak{L}(\mathbb{H})$ is the linear sub-manifold of $\mathbb{H} \oplus \mathbb{H}$ of the form

$$\Gamma(T) = \{((f, Tf) \in \mathbb{H} \times \mathbb{H} : f \in \mathsf{Dom}(T)\}.$$

The operator T is *closed* if $\Gamma(T)$ is closed. The operator T is *closable* if there exists $\hat{T} \in \mathfrak{L}(\mathbb{H})$ such that $\Gamma(\hat{T}) = \overline{\Gamma(T)}$ in $\mathbb{H} \oplus \mathbb{H}$. Such an operator is unique and is called the *closure* of T. An operator T is said *densely defined* if $\overline{\mathsf{Dom}(T)} = \mathbb{H}$.

If $T_1, T_2 \in \mathfrak{L}(\mathbb{H})$ with $\mathsf{Dom}(T_1) \subseteq \mathsf{Dom}(T_2)$ and $T_1 f = T_2 f$ for all $f \in \mathsf{Dom}(T_1)$, then T_2 is called an *extension* of T_1 and T_1 the *restriction* of T_2 on $\mathsf{Dom}(T_1)$; we write $T_1 \subseteq T_2$. If T is bounded on its domain and $\overline{\mathsf{Dom}(T)} = \mathbb{H}$, then T can be extended by continuity on the whole space.

The definitions of null space and range are also modified for unbounded operators:

$$\begin{aligned} \ker(T) &= \{f \in \mathsf{Dom}(T) : Tf = 0\} \\ \mathsf{Ran}(T) &= \{Tf \in \mathbb{H} : f \in \mathsf{Dom}(T)\}. \end{aligned}$$

The operator T is *invertible* if $\ker(T) = \{0\}$ and its inverse, T^{-1} is the operator defined on $\mathsf{Dom}(T^{-1}) = \mathsf{Ran}(T)$ by $T^{-1}(Tf) = f$ for all $f \in \mathsf{Dom}(T)$.

If $T_1, T_2 \in \mathfrak{L}(\mathbb{H})$, then $T_1 + T_2$ is defined on $\mathsf{Dom}(T_1 + T_2) = \mathsf{Dom}(T_1) \cap \mathsf{Dom}(T_2)$ by $(T_1 + T_2)f = T_1 f + T_2 f$. Similarly, the product $T_1 T_2$ is defined on $\mathsf{Dom}(T_1 T_2) = \{f \in \mathsf{Dom}(T_2) : T_2 \in \mathsf{Dom}(T_1)\}$ by $(T_1 T_2)f = T_2(T_1 f)$.

Definition 4.6.2 Suppose that T is densely defined. Then T is the *adjoint operator* with $\mathsf{Dom}(T^*) = \{g \in \mathbb{H} : \sup |\langle g \,|\, Tf \rangle| < \infty, f \in \mathsf{Dom}(T), \|f\| = 1\}$; since $\overline{\mathsf{Dom}(T)} = \mathbb{H}$, by Riesz theorem, there exists a unique $g^* \in \mathbb{H}$ such that $\langle g^* \,|\, f \rangle = \langle g \,|\, Tf \rangle$ for all $f \in \mathsf{Dom}(T)$. We define then $T^* g = g^*$.

Example 4.6.3 (The position operator) Let $(\Omega, \mathscr{F}, \mu)$ be any separable, σ-finite measure space, $\mathbb{H} = L^2(\Omega, \mathscr{F}, \mu; \mathbb{C})$, and $f \in \mathbb{H}$ measurable. Let $T \in \mathfrak{L}(\mathbb{H})$ be the operator defined by $\mathsf{Dom}(T) = \{g \in \mathbb{H} : \int (1 + |f|^2)|g|^2 d\mu < \infty\}$ and $Tg(\omega) = f(\omega)g(\omega)$ for $g \in \mathsf{Dom}(T)$ and $\omega \in \Omega$. Then T is closed, densely defined, with $\mathsf{Dom}(T^*) = \mathsf{Dom}(T)$ and $T^* g(\omega) = \overline{f}(\omega)g(\omega)$. When $\Omega = \mathbb{R}$, $\mathscr{F} = \mathscr{B}(\mathbb{R})$, and μ is the Lebesque measure, we say that T is the *position operator*; it is obviously self-adjoint.

Example 4.6.4 (The momentum operator) Let $\mathbb{H} = L^2(\mathbb{R})$. A function $u : \mathbb{R} \to \mathbb{R}$ is called *absolutely continuous*, (a.c.) if there exists a function $v : \mathbb{R} \to \mathbb{R}$ such that

$$u(b) - u(a) = \int_a^b v(x)dx, \text{ for all } a < b.$$

In such a case, we write $u' = v$, u' is called the derivative of u. The function v is determined almost everywhere. Define now $T \in \mathfrak{L}(\mathbb{H})$ on

$$\mathsf{Dom}(T) = \{f \in \mathbb{H} : f \text{ a.c.}, \int (|f|^2 + |f'|^2)dx < \infty\}$$

by $Tf = f'$. Then T is a closed, densely defined operator with $T^* = -iT$. The operator $-iT$ is called the *momentum operator*.

Exercise 4.6.5 Let q be the position operator, p the momentum operator. Show that $[q, p] \subseteq i\mathbb{1}$.

Exercise 4.6.6 (Heisenberg's uncertainty principle) Denote by $\mathbf{S}(\mathbb{R})$ the Schwartz space of indefinitely differentiable functions of rapid decrease. If $f \in \mathbf{S}(\mathbb{R})$, denote by $\hat{f}$ its Fourier transform $\hat{f}(\xi) = \int_{\mathbb{R}} f(x)\exp(-i\xi x)dx$. Let $p : \mathbf{S}(\mathbb{R}) \to \mathbf{S}(\mathbb{R})$ be defined by $pf = -if'$ and $q : \mathbf{S}(\mathbb{R}) \to \mathbf{S}(\mathbb{R})$ by $qf(x) = xf(x)$, for all $x \in \mathbb{R}$. Show that $[q, p] = i\mathbb{1}$. If $\langle \cdot \,|\, \cdot \rangle$ denotes the L^2 scalar product on $\mathbf{S}(\mathbb{R})$, show that

$$|\langle f \,|\, f \rangle| \leq 2\|pf\|_2\|qf\|_2.$$

Conclude that for any $f \in \mathbf{S}(\mathbb{R})$,

$$\|f\|_2 \leq 4\pi\|xf\|_{L^2(\mathbb{R})}\|\xi\hat{f}\|_{L^2(\hat{\mathbb{R}})}.$$

Below are depicted the graphs of pairs $|f(x)|^2$ and $|\hat{f}(\xi)|^2$, chosen among a class of Gaussian functions, for different values of some parameter. How do you interpret these results?

Chapter 5

Propositional calculus

5.1 Introduction

Phenomenology is an essential step in constructing physical theories. Phenomenological results are of the following type: if a physical system is subject to conditions $A, B, C, \ldots$, then the effects $X, Y, Z, \ldots$ are observed. We further introduced yes-no experiments consisting in measuring questions in given states. However, there may exist questions that depend on other questions and hold independently of the state in which they are measured. More precisely, suppose for instance that Q_A denotes the question: "does the physical particle lie in A, for some $A \in \mathscr{B}(\mathbb{R}^3)$?" Let now $B \supseteq A$ be another Borel set in $\mathbb{R}^3$. Whenever Q_A is true (i.e. for every state for which Q_A is true) Q_B is necessarily true. This remark defines a natural order relation in the set of questions. Considering questions on given physical system more abstractly, as a logical propositions, it is interesting to study first the abstract properties of a partially ordered set of propositions. This abstract setting allows the statement of the basic axioms for classical or quantum systems on an equal footing.

5.2 Lattice of propositions

Let Λ be a set of propositions and for any two propositions a and b, denote by $a \leq b$ the implication "whenever a is true, it follows that b is true"

Definition 5.2.1 The pair $(\Lambda, \leq)$ is a partially ordered set (poset) if the relation $\leq$ is a partial order (i.e. a reflexive, transitive, and antisymmetric binary operation).

For $a, b \in \Lambda$, we say that u is a *least upper bound* if

1. $a \leq u$ and $b \leq u$,
2. if $a \leq v$ and $b \leq v$ for some $v \in \Lambda$, then $u \leq v$.

If a least upper bound of two elements a and b exists, then it is unique and denoted by $\sup(a, b) \in \Lambda$,

Definition 5.2.2 A *lattice* is a set Λ with two binary operations, denoted respectively by $\vee$ ('join') and $\wedge$ ('meet'), and two constants $0 \in \Lambda$ and $1 \in \Lambda$, satisfying, for all $a, b, c \in \Lambda$ the following properties:

1. idempotence: $a \wedge a = a = a \vee a$,
2. commutativity: $a \wedge b = b \wedge a$ and $a \vee b = v \vee a$,
3. associativity: $a \wedge (b \wedge c) = (a \wedge b) \wedge c$ and $a \vee (b \vee c) = (a \vee b) \vee c$,
4. identity: $a \wedge 1 = a$ and $a \vee 0 = a$,
5. absorption: $a \wedge (a \vee b) = a = a \vee (a \wedge b)$.

Theorem 5.2.3 *Let $(\Lambda, \leq)$ be a poset. Suppose that*

1. *Λ has a least element 0 and a greatest element 1, i.e. for all $a \in \Lambda$, we have $0 \leq a \leq 1$,*
2. *any two elements $a, b \in \Lambda$ have a least upper bound in Λ, denoted by $a \vee b$, and a greatest lower bound in Λ, denoted by $a \wedge b$. Then $(\Lambda, \wedge, \vee, 0, 1)$ is a lattice.*

Conversely, if $(\Lambda, \wedge, \vee, 0, 1)$ is a lattice, then, on defining $a \leq b$ whenever $a \wedge b = b$, the pair $(\Lambda, \leq)$ is a poset verifying properties 1 and 2 of definition 5.2.1

Proof: : Exercise! □

Definition 5.2.4 A lattice $(\Lambda, \wedge, \vee, 0, 1)$ is called *distributive* if it verifies, for all $a, b, c \in \Lambda$,

$$a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c),$$

and

$$a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c).$$

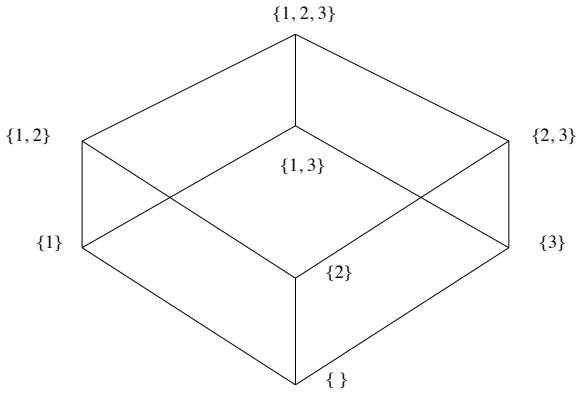

Figure 5.1: The Hasse diagram of the lattice of subsets of the set $\{1,2,3\}$.

Remark 5.2.5 A finite lattice (or finite poset) can be represented by its *Hasse diagram* in the plane. The points of the lattice are represented by points in the plane arranged so that if $a \leq b$ then the representative of b lies higher in the plane than the representative of a. We join the representatives of a and b by a segment when b covers a, i.e. when $a \leq b$ but there is no $c \in \Lambda$ such that $a < c < b$.

Example 5.2.6 Let S be a finite set and $\mathscr{P}(S)$ the collection of its subsets. Then $(\mathscr{P}(S), \subseteq)$ is a poset, equivalent to the lattice $(\mathscr{P}(S), \cap, \cup, \emptyset, S)$, called the *lattice of subsets* of S. This lattice is distributive. For the particular choice $S = \{1,2,3\}$ its Hasse diagram is depicted in figure 5.1.

Exercise 5.2.7 Let $\mathbb{V} = \mathbb{R}^2$ (viewed as a $\mathbb{R}$-vector space) and E_1, E_2, E_3 be three distinct one-dimensional subspaces of $\mathbb{V}$. Denote by $\leq$ the order relation "be a vector subspace of". Show that there is a finite set S of vector subspaces of $\mathbb{V}$ containing E_1, E_2, and E_3 such that $(S, \leq)$ is a lattice. Is this lattice distributive?

In any lattice Λ, a *complement* of $a \in \Lambda$ is an element $a' \in \Lambda$ such that $a \wedge a' = 0$ and $a \vee a' = 1$. Complements may fail to exist and they may be not unique. However, in a distributive lattice, any element has at most one complement.

Definition 5.2.8 A *Boolean algebra* is a complemented distributive lattice (i.e. a distributive lattice in which any element has a — necessarily unique — complement.)

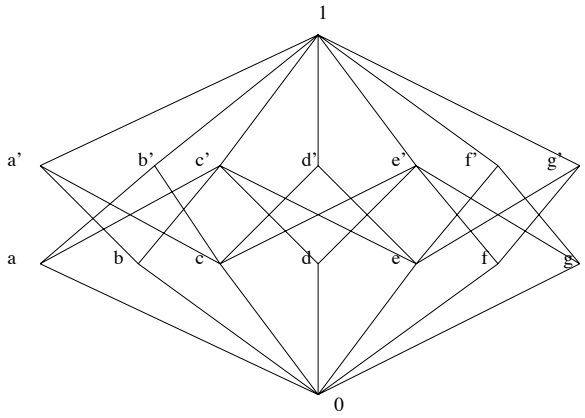

Figure 5.2: The Hasse diagram of the Dilworth lattice.

When the lattice Λ is infinite, one can consider infinite subsets $F \subseteq \Lambda$. When both $\wedge_{a\in F} a$ and $\vee_{a\in F} a$ exist (in Λ) for any countable subset F, the lattice is called σ-*complete*. A Boolean σ-algebra is a Boolean algebra that is σ-complete.

Definition 5.2.9 A lattice Λ is called *modular* of it satisfies the modularity condition:

$$a \leq c \Rightarrow \forall b \in \Lambda, a \vee (b \wedge c) = (a \vee b) \wedge c.$$

A complemented lattice is called *orthomodular* if satisfies the orthomodularity condition: for every complement a' of a,

$$a \leq b \Rightarrow b = a \vee (a' \wedge b).$$

Example 5.2.10 The *Dilworth lattice*, whose Hasse diagram is depicted in figure 5.2, is orthomodular but not distributive.

Exercise 5.2.11 Show that a Boolean algebra is always modular.

Definition 5.2.12 An *atom* in a lattice is a minimal non-zero element, i.e. $a \in \Lambda$ is an atom if $a \neq 0$ and if $x < a$ for some $x \in \Lambda$ then $x = 0$. A lattice is *atomic* if every point is the join of a finite number of atoms.

Definition 5.2.13 A *homomorphism* from a complemented lattice Λ_1 into a complemented lattice Λ_2 is a map $h : \Lambda_1 \to \Lambda_2$ such that

1. $h(0_1) = 0_2$ and $h(1_1) = 1_2$,
2. $h(a') = h(a)'$ for all $a \in \Lambda_1$,
3. $h(a \vee b) = h(a) \vee h(b)$ and $h(a \wedge b) = h(a) \wedge h(b)$, for all $a, b \in \Lambda_1$

An *isomorphism* is a lattice homomorphism that is bijective. If the condition 3 above holds also for countable joins and meets, h is called a σ-homomorphism. If $\Lambda_1 = \Lambda_2$ a lattice isomorphism is called *lattice automorphism*.

Theorem 5.2.14 *Let Λ be a Boolean σ-algebra. Then there exist an abstract set $\mathbb{X}$, a σ-algebra, $\mathscr{X}$, of subsets of $\mathbb{X}$ and a σ-homomorphism $h : \mathscr{X} \to \Lambda$.*

Proof: It is first given in [6] and later reproduced in [14]. □

This theorem serves to extend the notion of measurability, defined for maps between measurable spaces, to maps defined on abstract Boolean σ-algebras. Recall that if $\mathbb{X}$ is an arbitrary set of points equipped with a Boolean σ-algebra of subsets $\mathscr{X}$, and $\mathbb{Y}$ a complete separable metric space equipped with its Borel σ-algebra $\mathscr{B}(\mathbb{Y})$, a map $f : \mathbb{X} \to \mathbb{Y}$ is called *measurable* if for all $B \in \mathscr{B}(\mathbb{Y})$, $f^{-1}(B) \in \mathscr{X}$.

Definition 5.2.15 Let Λ be an abstract Boolean σ-algebra and $(\mathbb{Y}, \mathscr{B}(\mathbb{Y}))$ a complete separable metric space equipped with its Borel σ-algebra. A $\mathbb{Y}$-*valued classical observable* associated with Λ is a σ-homomorphism $h : \mathscr{B}(\mathbb{Y}) \to \Lambda$. If $\mathbb{Y} = \mathbb{R}$, the observable is called real-valued.

The careful reader will have certainly remarked that the previous definition is compatible with axiom 2.2.14. As a matter of fact, with every real random variable X on an abstract measurable space $(\Omega, \mathscr{F})$ is associated a family of propositions $Q_B^X = \mathbb{1}_{\{X \in B\}}$, for $B \in \mathscr{B}(\mathbb{R})$. The aforementioned σ-homomorphism $h : \mathscr{B}(\mathbb{R}) \to \mathscr{F}$ is given by

$$h(B) = \{\omega \in \Omega : Q_B^X(\omega) = 1\} = X^{-1}(B) \in \mathscr{F}.$$

Notice that this does not hold for quantum systems where some more general notion is needed.

5.3 Classical and quantum logics, observables, and states

5.3.1 Logics

Definition 5.3.1 Let $(\Lambda, \leq)$ be a poset (hence a lattice). By an *orthocomplementation* on Λ is meant a mapping $\perp\colon \Lambda \ni a \mapsto a^{\perp} \in \Lambda$, satisfying for $a, b \in \Lambda$:

1. $\perp$ is injective,
2. $a \leq b \Rightarrow b^{\perp} \leq a^{\perp}$,
3. $(a^{\perp})^{\perp} = a$,
4. $a \wedge a^{\perp} = 0$.

A lattice with an orthocomplementation operation is called *orthocomplemented*.

We remark that from condition 2 it follows that $0^{\perp} = 1$ and $1^{\perp} = 0$. From condition 3 it follows that $\perp$ is also surjective. Finally, conditions 1, 2, and 3 imply that $a \vee a^{\perp} = 1$.

Definition 5.3.2 An orthocomplemented lattice, Λ, is said to be a *logic* if

1. for any countable sequence $(a_n)_{n\in\mathbb{N}}$ of elements of Λ, both $\vee_{n\in\mathbb{N}} a_n$ and $\wedge_{n\in\mathbb{N}}$ exist in Λ,
2. if $a_1, a_2 \in \Lambda$ and $a_1 \leq a_2$, then there exists $b \in \Lambda$, such that $b \leq a_1^{\perp}$ and $b \vee a_1 = a_2$.

Without loss of generality, we can always assume that an orthocomplemented lattice verifies orthomodularity for $a^{\perp} = a'$. Remark also that the element whose existence is postulated in item 2 of the previous definition is unique and equal in fact to $a_1^{\perp} \wedge a_2$. In fact, if $b \leq a_1^{\perp}$ is such that $b \vee a_1 = a_2$, then necessarily, $a_1 \leq b^{\perp}$ and $b^{\perp} \wedge a_1^{\perp} = a_2^{\perp}$. Using orthomodularity, $a_1 \vee (b^{\perp} \wedge a_1^{\perp}) = b^{\perp}$ and substituting the left hand side parenthesis by $a_2^{\perp}$, we get the dual of the required equality. Dualising, we conclude.

The element $a^{\perp}$ is called the *orthogonal complement* of a in Λ. If $a \leq b^{\perp}$ and $b \leq a^{\perp}$, then a and b are said *orthogonal* and we write $a \perp b$.

Exercise 5.3.3 Assume that $(\Lambda, \leq)$ is a poset (hence a lattice) that is orthocomplemented. Let $a, b \in \Lambda$ be such that $a < b$. Denote by

$$\Lambda[a,b] = \{c \in \Lambda : a \leq c \leq b\}.$$

Show that

1. $\Lambda[0,b]$ becomes a lattice in which countable joins and meets exist and whose zero element is 0 and unit element is b,
2. if we define, for $x \in \Lambda[0,b]$, $x' = x^{\perp} \wedge b$, then the operation $' : \Lambda[0,b] \to \Lambda[0,b]$ is an orthocomplementation,
3. conclude that $\Lambda[0,b]$ is a logic.

Example 5.3.4 Any Boolean σ-algebra is a logic provided we define, for any element a, its orthocomplement to be its complement a'. Boolean σ-algebras are called *classical logics*.

Example 5.3.5 Let $\mathbb{H}$ be a $\mathbb{C}$-Hilbert space. Let Λ be the collection of all Hilbert subspaces of $\mathbb{H}$. If $\leq$ is meant to denote "be a Hilbert subspace of" and $\perp$ the orthogonal complementation in the Hilbert space sense, then Λ is a logic, called *standard quantum logic*.

Axiom 5.3.6 *In any physical system (classical or quantum), the set of all experimentally verifiable propositions is a logic (classical or standard quantum).*

5.3.2 Observables associated with a logic

Suppose that Λ is the logic of verifiable propositions of a physical system and let X be any real physical quantity relative to this system. Denoting $x(B)$ the proposition "the numerical results of the observation of X lie in B", it is natural and harmless to consider that $B \in \mathscr{B}(\mathbb{R})$; obviously then, x is a mapping $x : \mathscr{B}(\mathbb{R}) \to \Lambda$. We regard to physical quantities X and X' as identical whenever the corresponding maps $x, x' : \mathscr{B}(\mathbb{R}) \to \Lambda$ are the same. If $f : \mathbb{R} \to \mathbb{R}$ is a Borel function, we mean by $X' = f \circ X$ a physical quantity taking value $f(r)$ whenever X takes value r. The corresponding map is given by $\mathscr{B}(\mathbb{R}) \ni B : x' \mapsto x'(B) = x(f^{-1}(B)) \in \Lambda$. Hence we are led naturally to the following

Definition 5.3.7 Let Λ be a logic. A real *observable* associated with Λ is a mapping $x : \mathscr{B}(\mathbb{R}) \to \Lambda$ verifying:

1. $x(\emptyset) = 0$ and $x(\mathbb{R}) = 1$,
2. if $B_1, B_2 \in \mathscr{B}(\mathbb{R})$ with $B_1 \cap B_2 = \emptyset$ then $x(B_1) \perp x(B_2)$,
3. if $(B_n)_{n\in\mathbb{N}}$ is a sequence of mutually disjoint Borel sets, then $x(\cup_{n\in\mathbb{N}} B_n) = \vee_{n\in\mathbb{N}} x(B_n)$.

We write $\mathscr{O}(\Lambda)$ for the set of all real observables associated with Λ.

Exercise 5.3.8 Let Λ be a logic and $x \in \mathscr{O}(\Lambda)$. Show that for any sequence of Borel sets $(B_n)_{n\in\mathbb{N}}$ we have

$$x(\cup_{n\in\mathbb{N}} B_n) = \vee_{n\in\mathbb{N}} x(B_n)$$

and

$$x(\cap_{n\in\mathbb{N}} B_n) = \wedge_{n\in\mathbb{N}} x(B_n).$$

Definition 5.3.9 Let Λ be a logic and $\mathscr{O}(\Lambda)$ the set of its associated observables. A real number λ is called a *strict value* of an observable $x \in \mathscr{O}(\Lambda)$, if $x(\{\lambda\}) \neq 0$. The observable $x \in \mathscr{O}(\Lambda)$ is called *discrete* if there exists a countable set $C = \{c_1, c_2, \ldots\}$ such that $x(C) = 1$; it is called *constant* if there exists $c \in \mathbb{R}$ such that $x(\{c\}) = 1$. It is called *bounded* if there exists a compact Borel set K such that $x(K) = 1$.

Definition 5.3.10 We call *spectrum* of $x \in \mathscr{O}(\Lambda)$ the closed set defined by

$$\mathsf{spec}(x) = \cap_{C \text{ closed } : x(C)=1} C.$$

The numbers $\lambda \in \mathsf{spec}(x)$ are called *spectral values* of x.

Any strict value is a spectral value; the converse is not necessarily true.

Exercise 5.3.11 Show that $\lambda \in \mathsf{spec}(x)$ if and only if any open set U containing λ verifies $x(U) \neq 0$.

If $(a_n)_{n\in\mathbb{N}}$ is a partition of unity, i.e. a family of mutually orthogonal propositions in Λ such that $\vee_{n\in\mathbb{N}} a_n = 1$, there exists a unique discrete observable admitting as spectral values a given discrete subset $\{c_1, c_2, \ldots\}$ of the reals. In fact, it is enough to define for all $n \in \mathbb{N}$, $x(\{c_n\}) = a_n$ and for any $B \in \mathscr{B}(\mathbb{R})$, $x(B) = \vee_{n : c_n \in B} a_n$. Notice however that discrete observables do not exhaust all the physics of quantum mechanics; important physical phenomena involve continuous observables.

5.3.3 States on a logic

We have seen that to every classical system is attached a measurable space $(\Omega, \mathscr{F})$ (its phase space); observables are random variables and states are probability measures that may degenerate to Dirac masses on particular points of the phase space. This description is incompatible with the experimental observation for quantum systems. For the latter, the Heisenberg's uncertainty principle stipulates that no matter how carefully the system is prepared, there always exist observables whose values are distributed according to some non-trivial probability distribution.

Definition 5.3.12 Let Λ be a logic and $\mathscr{O}(\Lambda)$ its set of associated observables. A *state function* is a mapping $\rho : \mathscr{O}(\Lambda) \ni x \mapsto \rho_x \in \mathscr{M}_1^+(\mathbb{R}, \mathscr{B}(\mathbb{R}))$.

For every Borel function $f : \mathbb{R} \to \mathbb{R}$, for every observable x, and every Borel set B on the line, we have:

$$\rho_{f\circ x}(B) = \rho_x(f^{-1}(B)).$$

Denoting by o the zero observable and 0 the zero of $\mathbb{R}$, we have that $\rho_o = \delta_0$. In fact, suppose that $f : \mathbb{R} \to \mathbb{R}$ is the identically zero map. Then $f \circ o = o$ and

$$f^{-1}(B) = \begin{cases} \mathbb{R} & \text{if } 0 \in B \\ \emptyset & \text{otherwise.} \end{cases}$$

Hence, if $0 \in B$, then $\rho_o(B) = \rho_{f\circ o}(B) = \rho_o(f^{-1}(B)) = 1$, because ρ_o is a probability on $\mathbb{R}$; if $0 \notin B$ then similarly $\rho_o(B) = 0$. Therefore, in all circumstances, $\rho_o(B) = \delta_0(B)$.

If $x \in \mathscr{O}(\Lambda)$ is any observable and $B \in \mathscr{B}(\mathbb{R})$ is such that $x(B) = 0 \in \Lambda$, then $\rho_x(B) = 0$. In fact, for this B, we have $\mathbb{1}_B \circ x = o$ and $\rho_x(B) = \rho_o(\{1\}) = \delta_0(\{1\}) = 0$. This implies that if x is discrete, the measure ρ_x is supported by the set of the strict values of x.

Definition 5.3.13 An observable $q \in \mathscr{O}(\Lambda)$ is a *question* if $q(\{0,1\}) = 1$. A question is the necessarily discrete. If $q(\{1\}) = a \in \Lambda$, then q is the only question such that $q(\{1\}) = a$; we call it *question associated with the proposition a* and denote by q_a if necessary.

Definition 5.3.14 Let Λ be a logic. A function $p : \Lambda \to [0,1]$ satisfying

1. $p(0) = 0$ and $p(1) = 1$,

2. if $(a_n)_{n\in\mathbb{N}}$ is a sequence of mutually orthogonal propositions of Λ, and $a = \vee_{n\in\mathbb{N}} a_n$, then $p(a) = \sum_{n\in\mathbb{N}} p(a_n)$

is called *state (or probability measure) on the logic* Λ. The set of states on Λ is denoted by $\mathscr{S}(\Lambda)$.

The concept of probability measure on a logic coincides with a classical probability measure when the logic is a Boolean σ-algebra. For non distributive logics however, the associated probability measures are genuine generalisations of the classical probabilities. For standard quantum logics, the associated states are called quantum probabilities.

Theorem 5.3.15 *Let $p \in \mathscr{S}(\Lambda)$, where Λ is a logic.*

1. *On defining a map $\rho^p : \mathscr{O}(\Lambda) \to \mathscr{M}_1^+(\mathbb{R}, \mathscr{B}(\mathbb{R}))$, by the formula: for every $x \in \mathscr{O}(\Lambda)$ and for every $B \in \mathscr{B}(\mathbb{R})$, $\rho_x^p(B) = p(x(B))$, then ρ^p is a state function.*

2. *Conversely, if ρ is an arbitrary state function, then for every $x \in \mathscr{O}(\Lambda)$, then there exists a unique probability measure $p \in \mathscr{S}(\Lambda)$ such that for every $x \in \mathscr{O}(\Lambda)$ and for every $B \in \mathscr{B}(\mathbb{R})$, $\rho_x(B) = p(x(B))$.*

Proof:

1. The map $\rho_x^p : \mathscr{B}(\mathbb{R}) \to [0,1]$ is certainly a σ-additive, non-negative map. Moreover, $\rho_x^p(\mathbb{R}) = p(1) = 1$, hence it is a probability. If $f : \mathbb{R} \to \mathbb{R}$ is a Borel function,

$$\rho_{f\circ x}^p(B) = p(f \circ x(B)) = p(x(f^{-1}(B))) = \rho_x^p((f^{-1}(B)).$$

   Hence ρ^p is a state function.

2. Let ρ be a state function. If $a \in \Lambda$ and $q_a \in \mathscr{O}(\Lambda)$ the question associated with proposition a, then ρ_{q_a} is a probability measure on $\mathscr{B}(\mathbb{R})$. Since q_a is a question, $\rho_{q_a}(\{0,1\}) = 1$. Define $p(a) = \rho_{q_a}(\{1\})$. Obviously, for all $a \in \Lambda$, $p(a)$ is well defined and is taking values in $[0,1]$. It remains to show that p is a probability measure on Λ, that is to say verify σ-additivity and normalisation. For $0 \in \Lambda$, $q_0(\{1\}) = 0$. Hence $\rho_{q_0}(\{1\}) = 0 = p(0)$. Similarly, we show that $= p(1) = 1$. This shows normalisation.

   Let $(a_n)_{n\in\mathbb{N}}$ be a sequence of mutually orthogonal elements of Λ, and denote by $a = \vee_{n\in\mathbb{N}} a_n$. Let $x^i n \mathscr{O}(\Lambda)$ be the discrete observable defined by $x(\{0\}) =$

$a^\perp$ and $x(\{n\}) = a_n$, for $n = 1,2,\ldots$. Then, $\mathbb{1}_{\{n\}} \circ x(\{1\}) = x(\{n\}) = a_n$. Hence $q_{a_n} = \mathbb{1}_{\{n\}} \circ x$ and $p(a_n) = \rho_x(\{n\})$. Since ρ_x is a probability measure, $\sum_n p(a_n) = \rho_x(\{1,2,3,\ldots\}) = \rho_x(\mathbb{N})$. Similarly, $\mathbb{1}_{\mathbb{N}} \circ x = q_a$ because $\mathbb{1}_{\mathbb{N}} \circ x(\{1\}) = x(\mathbb{N}) = \vee_{n\in\mathbb{N}} x(\{n\}) = \vee_{n\in\mathbb{N}} a_n = a$. Hence, finally, $p(a) = \sum_n p(a_n)$ establishing thus σ-additivity of p. Finally, for $x \in \mathscr{O}(\Lambda)$ and $B \in \mathscr{B}(\mathbb{R})$,

$$\rho_x(B) = \rho_{\mathbb{1}_B \circ x}(\{1\}) = \rho_{q_{x(B)}}(\{1\}) = p(x(B)).$$

□

If $p \in \mathscr{S}(\Lambda)$ and $x \in \mathscr{O}(\Lambda)$, the map $\mathscr{B}(\mathbb{R}) \ni B \mapsto p(x(B)) \in [0,1]$ defines a probability measure on $\mathscr{B}(\mathbb{R})$. It is called the *probability distribution* induced on the space of its values by the observable x when the system is in state p and is denoted ρ_x^p. The *expected value* of x in state p is

$$\mathbb{E}_p(x) = \int_{\mathbb{R}} t\rho_x^p(dt)$$

and for a Borel function $f : \mathbb{R} \to \mathbb{R}$, we have

$$\mathbb{E}_p(f \circ x) = \int_{\mathbb{R}} f(t)\rho_x^p(dt)$$

(provided the above integrals exist.) If $\mathbb{E}_p(x^2) < \infty$, the *variance* of x in p is $\mathsf{Var}_p(x) = \mathbb{E}_p(x^2) - (\mathbb{E}_p(x))^2$.

Axiom 5.3.16 *Observables of a physical system described by the logic Λ are $\mathscr{O}(\Lambda)$.*

Axiom 5.3.17 *States of a physical system described by the logic Λ are $\mathscr{S}(\Lambda)$.*

Axiom 5.3.18 *Measuring whether the values of a physical observable $x \in \mathscr{O}(\Lambda)$ lie in $B \in \mathscr{B}(\mathbb{R})$ when the system is prepared in state $p \in \mathscr{S}(\Lambda)$ means determining $\rho_x^p(B)$.*

5.4 Pure states, superposition principle, convex decomposition

Proposition 5.4.1 *Let $\mathscr{S}(\Lambda)$ be the set of states on the logic Λ. Let $(p_n)_{n\in\mathbb{N}}$ be a sequence in $\mathscr{S}(\Lambda)$ and $(c_n)_{n\in\mathbb{N}}$ a sequence in $\mathbb{R}_+$ such that $\sum_{n\in\mathbb{N}} c_n = 1$. Then $p = \sum_{n\in\mathbb{N}} c_n p_n$, defined by $p(a) = \sum_{n\in\mathbb{N}} c_n p_n(a)$ for all $a \in \Lambda$, is a state.*

Proof: Exercise! □

Corollary 5.4.2 *For any logic Λ, the set $\mathscr{S}(\Lambda)$ is convex.*

Remark 5.4.3 Notice that if $p = \sum_{n\in\mathbb{N}} c_n p_n$ as above, for every $x \in \mathscr{O}(\Lambda)$, we have that $\rho_x^p = \sum_{n\in\mathbb{N}} c_n \rho_x^{p_n}$. In fact, for all $B \in \mathscr{B}(\mathbb{R})$,

$$\rho_x^p(B) = p(x(B)) = \sum_{n\in\mathbb{N}} c_n p_n(x(B)) = \sum_{n\in\mathbb{N}} c_n \rho_x^{p_n}(B).$$

This decomposition has the following interpretation: the sequence $(c_n)_{n\in\mathbb{N}}$ defines a classical probability on $\mathbb{N}$ meaning that in the sum defining p, each p_n is chosen with probability c_n. Therefore, for each integrable observable $x \in \mathscr{O}(\Lambda)$, the expectation $\mathbb{E}_p(x) = \sum_{n\in\mathbb{N}} c_n \mathbb{E}_{p_n}(x)$ consists in two averages: a classical average on the choice of p_n and a (may be) quantum average $\mathbb{E}_{p_n}(x)$.

Exercise 5.4.4 Give a plausible definition of the notion of *integrable observable* used in the previous remark and then prove the claimed equality: $\mathbb{E}_p(x) = \sum_{n\in\mathbb{N}} c_n \mathbb{E}_{p_n}(x)$

Definition 5.4.5 A state $p \in \mathscr{S}(\Lambda)$ is said to be *pure* if the equation $p = cp_1 + (1-c)p_2$, for $p_1, p_2 \in \mathscr{S}(\Lambda)$ and $c \in [0,1]$ implies $p = p_1 = p_2$. We write $\mathscr{S}_p(\Lambda)$ for the set of pure states of Λ. Obviously $\mathscr{S}_p(\Lambda) = \mathsf{Extr}\,\mathscr{S}(\Lambda)$.

Definition 5.4.6 Let $\mathscr{D} \subseteq \mathscr{S}(\Lambda)$ and $p_0 \in \mathscr{S}(\Lambda)$. We say that p_0 is a *superposition of states in $\mathscr{D}$* if for $a \in \Lambda$,

$$\forall p \in \mathscr{D}, p(a) = 0 \Rightarrow p_0(a) = 0.$$

It is an exercise to show that the state $p = \sum_{n\in\mathbb{N}} c_n p_n$ defined in the proposition **??** is a superposition of states in $\mathscr{D} = \{p_1, p_2, \ldots\}$. In the case Λ is a Boolean σ-algebra, the next theorem 5.4.7 shows that this is in fact the only kind of possible superposition. This implies, in particular, the *unicity* of the decomposition of a classical state into extremal (pure) states. If Λ is a standard quantum logic, unicity of the decomposition does not hold any longer!

Theorem 5.4.7 *Let Λ be a Boolean σ-algebra of subsets of a space $\mathbb{X}$. Suppose that*

1. *Λ is separable*[1],

[1] i.e. there is a countable collection of subsets $A_n \subseteq \mathbb{X}$, $n \in \mathbb{N}$, generating Λ by complementation, intersections, and unions.

2. *for all* $a \in \mathbb{X}$, $\{a\} \in \Lambda$.

For any $a \in \mathbb{X}$ *and any* $A \subseteq \mathbb{X}$, *let* δ_a *be the state defined by*

$$\delta_a(A) = \begin{cases} 1 & \text{if } a \in A \\ 0 & \text{otherwise.} \end{cases}$$

Then, $(\delta_a)_{a\in\mathbb{X}}$ *is precisely the set of all pure states in* Λ. *If* $\mathscr{D} \subseteq \mathscr{S}_p(\Lambda)$ *and* $p_0 \in \mathscr{S}_p(\Lambda)$, *then* p_0 *is a superposition of states in* $\mathscr{D}$ *if and only if* $p_0 \in \mathscr{D}$.

Proof: Denote $\{A_1, A_2, \ldots\}$ a denumerable collection of subsets of $\mathbb{X}$ generating Λ. Purity of δ_a is trivially verified. Suppose that p is a pure state. If for some $A_0 \in \Lambda$ we have $0 < p_0(A) < 1$, then, on putting for $A \in \Lambda$

$$p_1(A) = \frac{1}{p(A_0)} p(A \cup A_0) \qquad (*)$$

and

$$p_2(A) = \frac{1}{1 - p(A_0)} p(A \cap A_0^c), \qquad (**)$$

we get $p(A) = p(A_0)p_1(A) + (1 - p(A_0))p_2(A)$. Yet, applying (*) and (**) to A_0, we get $p_1(A_0) = 1$ and $p_2(A_0) = 0$, hence $p_1 \neq p_2$. This is in contradiction with the assumed purity of p. Therefore, we conclude that for all $A \in \Lambda$, we have $p(A) \in \{0, 1\}$. Replacing A_n by A_n^c if necessary, we can assume without loss of generality that $p(A_n) = 1$ for all the sets of the collection generating Λ. Let $B = \cap_n A_n$. Then $p(B) = 1$ and consequently B cannot be empty. Now B cannot contain more than one point either. In fact, the collection of all sets $C \in \Lambda$ such that either $B \subseteq C$ or $B \cap C = \emptyset$ is a σ-algebra containing all the sets A_n, $n \in \mathbb{N}$. Hence, it coincides with Λ. As singletons are members of Λ, the set B must be a singleton, i.e. $B = \{a\}$ for some $a \in \mathbb{X}$. Put then $p = \delta_a$. Finally, let p_0 be a superposition of states in $\mathscr{D}$ (all its elements are pure states). If $p_0 = \delta_{a_0}$ but $p_0 \notin \mathscr{D}$, then $p(\{a_0\}) = 0$ for all $p \in \mathscr{D}$ but $p_0(\{a_0\}) \neq 0$, a contradiction. □

5.5 Simultaneous observability

In quantum systems, the Heisenberg's uncertainty principle, already shown in chapter 2, there are observables that cannot be simultaneously observed with arbitrary precision.

Definition 5.5.1 Let $a, b \in \Lambda$. Propositions a and b are said to be *simultaneously verifiable*, denoted by $a \leftrightarrow b$, if there exists elements $a_1, b_1, c \in \Lambda$ such that

1. a_1, b_1, c are mutually orthogonal and,

2. $a = a_1 \vee c$ and $b = b_1 \vee c$ hold.

Observables $x, y \in \mathscr{O}(\Lambda)$ are *simultaneously observable* if for all $B \in \mathscr{B}(\mathbb{R})$, $x(B) \leftrightarrow y(B)$. For $A, B \subseteq \Lambda$, we write $A \leftrightarrow B$ if for all $a \in A$ and all $b \in B$ we have $a \leftrightarrow b$.

Lemma 5.5.2 *Let $a, b \in \Lambda$. The following are equivalent:*

1. $a \leftrightarrow b$,

2. $a \wedge (a \wedge b)^{\perp} \perp b$,

3. $b \wedge (a \wedge b)^{\perp} \perp a$,

4. *there exist $x \in \mathscr{O}(\Lambda)$ and $A, B \in \mathscr{B}(\mathbb{R})$ such that $x(A) = a$ and $x(B) = b$,*

5. *there exists a Boolean sub-algebra of Λ containing a and b.*

Proof:

$1 \Rightarrow 2$:

$$\begin{aligned} a \leftrightarrow b &\Leftrightarrow a = a_1 \vee c \text{ and } b = b_1 \vee c \\ &\Rightarrow c \leq a \text{ and } c \leq b \\ &\Rightarrow c \leq a \wedge b. \end{aligned}$$

From the definition 5.3.2 (logic), it follows that there exists $d \in \Lambda$ such that $c \perp d$ and $c \vee d = a \wedge b$.

Now $d \leq c \vee d = a \wedge b \leq a$ and $d \leq c^{\perp}$ (since $d \perp d$.) Hence, $d \leq a \wedge c^{\perp} = a_1$ (see remark immediately following the definition 5.3.2.) Similarly, $d \leq b_1 \Rightarrow d \leq b_1 \wedge q_1 = 0$. Therefore $d = 0$ and consequently $c = a \wedge b$. It follows $a_1 = a \wedge (a \wedge b)^{\perp}$. Yet, $a_1 \perp c$ and $a_1 \perp b_1$ so that $a_1 \perp (b_1 \perp c) = b$. Summarising, $a \wedge (a \wedge b)^{\perp} \perp b$.

$1 \Rightarrow 3$: By symmetry.

$2 \Rightarrow 1$: Since $a \wedge (a \wedge b)^{\perp} \perp b$, on writing $a_1 = a \wedge (a \wedge b)^{\perp}$, $b_1 = b \wedge (a \wedge b)^{\perp}$, and $c = a \wedge b$, we find $a = a_1 \vee c$ and $b = b_1 \vee c$. Since $a_1 \perp b$, it follows that $a_1 \perp b_1$ and $a_1 \perp c$, while, by definition, $c \perp b_1$ which proves the implication.

Henceforth, the equivalence $1 \Leftrightarrow 2 \Leftrightarrow 3$ is established.

$1 \Rightarrow 4$: If $a = a_1 \vee c$, $b = b_1 \vee c$ and a_1, b_1, c mutually orthogonal, write $d = a_1 \vee b_1 \vee c$ and define x to be the discrete observable such that $x(\{0\}) = a_1$, $x(\{1\}) = b_1$, $x(\{2\}) = c$, and $x(\{3\}) = d$. Then $x(\{0,2\}) = a$ and $x(\{1,2\}) = b$.

$4 \Rightarrow 5$: $x(A \cap (A \cap B)^c) = a \wedge (a \wedge b)^\perp$ and $x(B \cap (A \cap B)^c) = b \wedge (a \wedge b)^\perp$. On writing $a_1 = a \wedge (a \wedge b)^\perp$, $a_2 = a \wedge b$, $a_3 = b \wedge (a \wedge b)^\perp$, and $a_4 = (a \vee b)^\perp$, we see that $(a_i)_{i=1,\ldots,4}$ are mutually orthogonal and $a_1 \vee a_2 \vee a_3 \vee a_4 = 1$. If

$$\mathscr{A} = \{a_{i_1} \vee \ldots \vee a_{i_k} : k \leq 4; 1 \leq i_1 \leq \ldots \leq i_k \leq 4\},$$

it is easily verified that $\mathscr{A}$ is Boolean sub-algebra of Λ. Since $a, b \in \mathscr{A}$, this proves the implication.

$5 \Rightarrow 2$: Let $\mathscr{A}$ be a Boolean sub-algebra of Λ containing a and b. Now, $[a \wedge (a \wedge b)^\perp] \wedge b = 0$. As $a, b, a \wedge (a \wedge b)^\perp, b^\perp \in \mathscr{A}$, it follows that

$$\begin{aligned} a \wedge (a \wedge b)^\perp &= [(a \wedge (a \wedge b)^\perp) \wedge b] \\ &\quad \vee [(a \wedge (a \wedge b)^\perp) \wedge b^\perp] \\ &= [(a \wedge (a \wedge b)^\perp) \wedge b^\perp] \\ &\leq b^\perp. \end{aligned}$$

Therefore $a \wedge (a \wedge b)^\perp \perp b$.

□

The significance of this lemma is that if two propositions are simultaneously verifiable, we can operate on them as if they were classical.

Theorem 5.5.3 *Let Λ be any logic and $(x_\lambda)_{\lambda \in D}$ a family of observables. Suppose that $x_\lambda \leftrightarrow x_{\lambda'}$ for all $\lambda, \lambda' \in D$. Then there exist a space $\mathbb{X}$, a σ-algebra $\mathscr{X}$ of subsets of $\mathbb{X}$, a family of measurable functions $g_\lambda : \mathbb{X} \to \mathbb{R}$, $\lambda \in D$, and a σ-homomorphism $\tau : \mathscr{X} \to \Lambda$ such that $\tau(g_l^{-1}(B)) = x_\lambda(B)$ for all $\lambda \in D$ and all $b \in \mathscr{B}(\mathbb{R})$. Suppose further that either Λ is separable or D is countable. Then, for all $\lambda \in D$, there exist a $x \in \mathscr{O}(\Lambda)$ and a measurable function $f_\lambda : \mathbb{R} \to \mathbb{R}$ such that $x_\lambda = f_\lambda \circ x$.*

The proof of this theorem is omitted. Notice that it allows to construct functions of several observables that are simultaneously observable. This latter result is also stated without proof.

Theorem 5.5.4 *Let Λ be any logic and $(x_1, \ldots, x_n)$ a family of observables that are simultaneously observable. Then there exists a σ-homomorphism $\tau : \mathscr{B}(\mathbb{R}^n) \to \Lambda$ such that for all $B \in \mathscr{B}(\mathbb{R})$ and all $i = 1, \ldots, n$,*

$$x_i(B) = \tau(\pi_i^{-1}(B)), \qquad (*)$$

where $\pi_i : \mathbb{R}^n \to \mathbb{R}$ *is the projection* $\pi(t_1,\ldots,t_n) = t_i$, $i = 1,\ldots,n$. *If* g *is a Borel function on* $\mathbb{R}^n$, *then* $g \circ (x_1,\ldots,x_n)(B) = \tau(g^{-1}(B))$ *is an observable. If* $g_1,\ldots,g_k$ *are real valued Borel functions on* $\mathbb{R}^n$ *and* $y_i = g_i \circ (x_1,\ldots,x_n)$, *then* $y_1,\ldots,y_k$ *are simultaneously observable and for any real valued Borel function* h *on* $\mathbb{R}^k$, *we have* $h \circ (y_1,\ldots,y_k) = h(g_1,\ldots,g_k) \circ (x_1,\ldots,x_n)$ *where, for* $\mathbf{t} = (t_1,\ldots,t_n)$, $h(g_1,\ldots,g_k)(\mathbf{t}) = h(g_1(\mathbf{t}),\ldots,g_k(\mathbf{t}))$.

An immediate consequence of this theorem is that if p is a probability measure on Λ, then $\rho^p_{x_1,\ldots,x_n}(B) = p(\tau(B))$, for $B \in \mathscr{B}(\mathbb{R}^n)$, is the *joint probability distribution* of $(x_1,\ldots,x_n)$ in state p.

5.6 Automorphisms and symmetries

Let Λ be a logic. The set $\mathsf{Aut}(\Lambda)$, of automorphisms of Λ, acquires as usual a group structure; they induce naturally automorphisms on $\mathscr{S}(\Lambda)$, called *convex automorphisms*.

Let, in fact, $\alpha \in \mathsf{Aut}(\Lambda)$ and $p \in \mathscr{S}(\Lambda)$. If we define $\tilde{\alpha}$ to be the induced action of α on p, by $\tilde{\alpha}(p)(a) = p(\alpha^{-1}(a))$, for all $a \in \Lambda$, then $\tilde{\alpha}$ is a convex automorphism of $\mathscr{S}(\Lambda)$.

Definition 5.6.1 A map $\beta : \mathscr{S}(\Lambda) \to \mathscr{S}(\Lambda)$ is a *convex automorphisms* if

1. β is bijective and
2. if $(c_n)_{n\in\mathbb{N}}$ is a sequence of non-negative reals such that $\sum_{n\in\mathbb{N}} c_n = 1$ and $(p_n)_{n\in\mathbb{N}}$ is a sequence of states in $\mathscr{S}(\Lambda)$, then

$$\beta(\sum_{n\in\mathbb{N}} c_n p_n) = \sum_{n\in\mathbb{N}} c_n \beta(p_n).$$

The set of convex automorphisms of $\mathscr{S}(\Lambda)$ is denoted $\mathsf{Aut}(\mathscr{S}(\Lambda))$.

Lemma 5.6.2 *Let* $\alpha \in \mathsf{Aut}(\Lambda)$. *Then the induced automorphism* $\tilde{\alpha}$ *on* $\mathscr{S}(\Lambda)$ *is convex.*

Proof: Bijectivity of $\tilde{\alpha}$ follows immediately from the bijectivity of α. If $p = \sum_{n\in\mathbb{N}} c_n p_n \in \mathscr{S}(\Lambda)$ (with the notation of definition 5.6.1), then $\tilde{\alpha}(p)(a) = p(\alpha^{-1}(a)) = \sum_{n\in\mathbb{N}} c_n p_n(\alpha^{-1}(a)) = \sum_{n\in\mathbb{N}} c_n \tilde{\alpha}(p_n)(a)$ for all $a \in \Lambda$. □

Remark 5.6.3 It is obvious that convex automorphisms map pure states of $\mathscr{S}_p(\Lambda)$ into pure states.

Dynamics, i.e. time evolution of a system described by a logic Λ can be defined in the following manner. For each $t \in \mathbb{R}$, there exists a unique map $D(t) : \mathscr{S}(\Lambda) \to \mathscr{S}(\Lambda)$ having the following interpretation: if $p \in \mathscr{S}(\Lambda)$ is the state of the system at time t_0, then $D(t)(p)$ will represent the state of the system at time $t + t_0$.

Definition 5.6.4 Let G be a locally compact topological group. By a *representation* of G into $\mathsf{Aut}(\mathscr{S}(\Lambda))$, we mean a map $\pi : G \to \mathsf{Aut}(\mathscr{S}(\Lambda))$ such that

1. $\pi(g_1 g_2) = \pi(g_1)\pi(g_2)$ for all $g_1, g_2 \in G$,
2. for each $a \in \Lambda$ and each $p \in \mathscr{S}(\Lambda)$, the mapping $g \mapsto \pi(g)(p)(a)$ is $\mathscr{B}(G)$-measurable.

Axiom 5.6.5 *Time evolution of an isolated physical system described by a logic Λ, is implemented by a map $\mathbb{R} \ni t \mapsto D(t) \in \mathsf{Aut}(\mathscr{S}(\Lambda))$. This map provides a representation of the Abelian group $(\mathbb{R}, +)$ into $\mathsf{Aut}(\mathscr{S}(\Lambda))$. More generally, any physical symmetry, implemented by the action of a locally compact topological group G, induces a representation into $\mathsf{Aut}(\mathscr{S}(\Lambda))$.*

Here is an interpretation and/or justification of this axiom. If $p = \sum_{n\in\mathbb{N}} c_n p_n$ represents the initial state of the system, we can realise this state as follows. First chose an integer $n \in \mathbb{N}$ with probability c_n and prepare the system at state p_n. Let the system evolve under the dynamics. Then at time t it will be at state $p'_n = D(t)(p_n)$ with probability c_n. Assuming now that $D(t)$ is a convex automorphism means that $D(t)(p) = \sum_{n\in\mathbb{N}} c_n D(t)(p_n)$, i.e. at time t, the system is in state $p'_n = D(t)(p_n)$ with probability c_n, exactly the result we obtained with the first procedure.

To further exploit the notions of logic, states, observables, and convex automorphisms, we must specialise the physical system.

Chapter 6

Standard quantum logics

We recall that a standard quantum logic Λ was defined in chapter 4 to be the set of Hilbert subspaces of $\mathbb{C}$-Hilbert space $\mathbb{H}$. For every Hilbert subspace $M \in \Lambda$, we denote by P_M the orthogonal projection to M. If $x \in \mathscr{O}(\Lambda)$, then $B \mapsto P_{x(B)}$, for $B \in \mathscr{B}(\mathbb{R})$, is a projection-valued measure on $\mathscr{B}(\mathbb{R})$. Conversely, for every projection-valued measure P on $\mathscr{B}(\mathbb{R})$, there exists an observable $x \in \mathscr{O}(\Lambda)$ such that $P(B) = P_{x(B)}$, for all $B \in \mathscr{B}(\mathbb{R})$. We identify henceforth Hilbert subspaces with the orthogonal projectors mapping the whole space on them (recall exercise 3.3.7.)

6.1 Observables

Lemma 6.1.1 *Let $M_1, M_2 \in \Lambda$. Then propositions associated with M_1 and M_2 are simultaneously verifiable if and only if $[P_{M_1}, P_{M_2}] = 0$.*

Proof:

- ($\Rightarrow$): Propositions M_1 and M_2 are simultaneously verifiable if there exist mutually orthogonal elements $N_1, N_2, N \in \Lambda$ such that $M_i = N_i \vee N$, for $i = 1,2$. Then $P_{M_i} = P_{N_i} + P_N$ and the commutativity of the projectors follows immediately.
- ($\Leftarrow$): If $[P_{M_1}, P_{M_2}] = 0$, let $P = P_{M_1} P_{M_2}$. Then P is a projection. Define $Q_i = P_{M_i} - P$, for $i = 1,2$; it is easily verified that Q_i are projections and $PQ_i = Q_i P = 0$. Therefore $Q_1 Q_2 = Q_2 Q_1 = 0$. If we define $N_i = Q_i(\mathbb{H})$, for $i = 1,2$ and $N = P(\mathbb{H})$, then N_1, N_2, N are mutually orthogonal and $M_i = N_i \vee N$ which proves that $M_1 \leftrightarrow M_2$.

□

Theorem 6.1.2 *Let Λ be a standard logic with associated Hilbert space $\mathbb{H}$. For any $x \in \mathscr{O}(\Lambda)$, denote X the self-adjoint (not necessarily bounded) operator on $\mathbb{H}$ with spectral measure given by the mapping $\mathscr{B}(\mathbb{R}) \ni B \mapsto P_{x(B)} \in \Lambda$. Then*

1. *the map $x \mapsto X$ is a bijection between $\mathscr{O}(\Lambda)$ and self-adjoint operators on $\mathbb{H}$,*
2. *the observable x is bounded if and only if $X \in \mathfrak{B}_h(\mathbb{H})$,*
3. *two bounded observables x_1 and x_2 are simultaneously observable if and only if the corresponding bounded operators X_1 and X_2 commute,*
4. *if x is a bounded observable and $Q \in \mathbb{R}[t]$, then the operator associated with $Q \circ x$ is $Q(X)$,*
5. *more generally, if $x_1, \ldots, x_r$ are bounded observables any two of them being simultaneously observable, and $Q \in \mathbb{R}[t_1, \ldots, t_2]$, then the observable $Q \circ (x_1, \ldots, x_r)$ has associated operator $Q(X_1, \ldots, X_r)$.*

Proof: Assertions 1–4 are simple exercises based on the spectral theorem for self-adjoint operators. Assertion 5 is a direct consequence of theorem 5.5.4. □

6.2 States

In chapter 2, we defined (pure) quantum states to be unit vectors of $\mathbb{H}$. In chapter 5, states have been defined as probability measures on a logic. We first show that in fact rays correspond to states viewed as probability measures on Λ.

Unit vectors of $\mathbb{H}$ are called *rays*. Let $\xi \in \mathbb{H}$, with $\|\xi\| = 1$ be a ray and denote by $p_\xi : \Lambda \to [0,1]$ the map defined by

$$\Lambda \ni M \mapsto p_\xi(M) = \langle \xi \,|\, P_M \xi \rangle = \|P_M \xi\|^2.$$

We have: $p_\xi(1) \equiv p_\xi(\mathbb{H}) = 1$, $p_\xi(0) \equiv p_\xi(\{0\}) = 0$, and if $(M_n)_{n\in\mathbb{N}}$ is a sequence of mutually orthogonal Hilbert subspaces of $\mathbb{H}$ and $M = \vee_{n\in\mathbb{N}} M_n$, then

$$p_\xi(M) = \|P_M \xi\|^2 = \sum_{n\in\mathbb{N}} \langle \xi \,|\, P_{M_n} \xi \rangle = \sum_{n\in\mathbb{N}} p_\xi(M_n).$$

Hence $p_\xi \in \mathscr{S}(\Lambda)$. If $c \in \mathbb{C}$, with $|c| = 1$, then $p_{c\xi} = p_\xi$.

Theorem 6.2.1 *Let $\mathbb{H}$ be a Hilbert space, $(\varepsilon_n)_{n\in\mathbb{N}}$ an orthonormal basis in it an $T \in \mathfrak{B}_+(\mathbb{H})$. We define the trace of T by*

$$\mathsf{tr}(T) = \sum_{n\in\mathbb{N}} \langle \varepsilon_n \,|\, T\varepsilon_n \rangle \in [0,+\infty].$$

Then for all $T, T_1, T_2 \in \mathfrak{B}_+(\mathbb{H})$ the trace has the following properties

1. *is independent of the chosen basis,*
2. $\mathsf{tr}(T_1 + T_2) = \mathsf{tr}(T_1) + \mathsf{tr}(T_2)$,
3. $\mathsf{tr}(\lambda T) = \lambda\, \mathsf{tr}(T)$ *for all* $\lambda \geq 0$,
4. $\mathsf{tr}(UTU^*) = \mathsf{tr}(T)$, *for all* $U \in \mathfrak{U}(\mathbb{H})$.

Proof: (To be filled in a later version.) □

Definition 6.2.2 Let $T \in \mathfrak{B}(\mathbb{H})$. The operator T is called *trace-class operator* if $\mathsf{tr}(|T|) < \infty$. The family of trace-class operators is denoted by $\mathfrak{T}^1(\mathbb{H})$.

Lemma 6.2.3 *The space $\mathfrak{T}^1(\mathbb{H})$ is a two-sided ideal of $\mathfrak{B}(\mathbb{H})$ and $\mathsf{tr}(TB) = \mathsf{tr}(BT)$ for all $B \in \mathfrak{B}(\mathbb{H})$.*

Proof: (To be filled in a later version.) □

Definition 6.2.4 If D is a bounded, self-adjoint, non-negative, trace-class operator on $\mathbb{H}$, then D is called a *von Neumann* operator. If further $\mathsf{tr}(D) = 1$, then D is said to be a *density matrix (operator)*. The set of density matrices on $\mathbb{H}$ is denoted by $\mathfrak{D}(\mathbb{H})$.

The states p_ξ, for ξ a ray of $\mathbb{H}$, can also be described in another way. Let D_ξ be the projection operator on the one-dimensional subspace[1] $\mathbb{C}\xi$. Then D_ξ is trace-class and for every $X \in \mathfrak{B}(\mathbb{H})$, it follows that $D_\xi X$ is also trace-class. Let $(\varepsilon_n)_{n\in\mathbb{N}}$ be an arbitrary orthonormal basis of $\mathbb{H}$; without loss of generality, we can then assume that $\varepsilon_1 = \xi$. We have

$$\begin{aligned} \mathsf{tr}(D_\xi X) &= \mathsf{tr}(XD_\xi) \\ &= \sum_{n\in\mathbb{N}} \langle \varepsilon_n \,|\, XD_\xi \varepsilon_n \rangle \\ &= \langle \xi \,|\, X\xi \rangle \\ &= \mathbb{E}_\xi(X). \end{aligned}$$

[1] We recall that the term subspace always means closed subspace.

In particular, if $X = P_M$ for $M \in \Lambda$,

$$p_\xi(M) = \langle\, \xi \,|\, P_M \xi \,\rangle = \mathsf{tr}(D_\xi P_M).$$

Lemma 6.2.5 *Let $(\xi_n)_{n\in\mathbb{N}}$ be an arbitrary sequence of rays in $\mathbb{H}$ and $(c_n)_{n\in\mathbb{N}}$ an arbitrary sequence of non-negative reals such that $\sum_{n\in\mathbb{N}} c_n = 1$. Denote by D_n the projection operator on the one-dimensional subspace $\mathbb{C}\xi_n$, for $n \in \mathbb{N}$. Then*

$$D = \sum_{n\in\mathbb{N}} c_n D_n$$

is a well defined density matrix.

Proof: Exercise. □

Exercise 6.2.6 Show that $\mathfrak{D}(\mathbb{H})$ is convex.

Lemma 6.2.7 *Let D be a density matrix defined as in lemma 6.2.5 and $p : \Lambda \to \mathbb{R}$ the mapping defined by $\Lambda \ni M \mapsto p(M) = \mathsf{tr}(P_M D)$. Then $p \in \mathscr{S}(\Lambda)$ and moreover it can be decomposed into $p = \sum_{n\in\mathbb{N}} c_n p_{\xi_n}$.*

Proof: First the superposition property follows from the linearity of the trace: for all $M \in \Lambda$, we have $p(M) = \mathsf{tr}(P_M D) = \sum_{n\in\mathbb{N}} c_n \mathsf{tr}(P_M D_n) = \sum_{n\in\mathbb{N}} c_n p_{\xi_n}(M)$. It is now obvious that p is a state: in fact, $p(0) = p(\{0\}) = 0$ and $p(1) = p(\mathbb{H}) = 1$. □

Conversely, if D is any density matrix, then the map $\Lambda \ni M \mapsto p(M) = \mathsf{tr}(DP_M)$ is a state in $\mathscr{S}(\Lambda)$. States of this type are called *tracial states*. The natural question is whether every state in $\mathscr{S}(\Lambda)$ arises as a tracial state. The answer to this question is one of the most profound results in the mathematical foundations of quantum mechanics, the celebrated Gleason's theorem:

Theorem 6.2.8 (Gleason) *Let $\mathbb{H}$ be a complex separable Hilbert space with $3 \leq \dim\mathbb{H} \leq \aleph_0$, $\mathfrak{D}(\mathbb{H})$ the convex set of density matrices on $\mathbb{H}$, and Λ the logic of subspaces of $\mathbb{H}$. Then*

1. *the map $\mathfrak{D}(\mathbb{H}) \ni D \mapsto \rho_D \in \mathscr{S}(\Lambda)$, defined by $\rho_D(M) = \mathsf{tr}(DP_M)$ for all $M \in \Lambda$, is a convex isomorphism of $\mathfrak{D}(\mathbb{H})$ on $\mathscr{S}(\Lambda)$,*
2. *a state $p \in \mathscr{S}(\Lambda)$ is pure if and only if $p = p_\xi$ for some ray ξ in $\mathbb{H}$,*

3. *two pure states p_ξ and p_ζ are equal if and only if there exists a complex number c with $|c| = 1$ such that the rays ξ and ζ verify $\xi = c\zeta$.*

The proof, lengthy and tricky, is omitted. It can be found, extending over 13 pages (!), in [12], pages 147–160.

6.3 Symmetries

Definition 6.3.1 A linear map $S : \mathbb{H} \to \mathbb{H}$ is a *symmetry* if

1. S is bijective, and
2. for all $f, g \in \mathbb{H}$, the scalar product is preserved: $\langle Sf \,|\, Sg \rangle = \langle f \,|\, g \rangle$.

Exercise 6.3.2 Let $\alpha \in \mathsf{Aut}(\Lambda)$ where Λ is the standard quantum logic associated with a given Hilbert space $\mathbb{H}$. Show that

1. there exists a symmetry $S \in \mathfrak{B}(\mathbb{H})$ such that for all $M \in \Lambda$, $\alpha(M) = SM$,
2. if S' is another symmetry corresponding to the same automorphism α, then there exists a complex number c, with $|c| = 1$ such that $S' = cS$,
3. if S is any symmetry of $\mathbb{H}$, the map $\Lambda \ni M \mapsto SM \in \Lambda$ is an automorphism of Λ.

Notice that unitaries are obviously symmetries. It turns out that they are the only symmetries encountered in elementary quantum systems[2].

[2] In general, anti-unitaries may also occur as symmetries. They are not considered in this course.

Chapter 7

Two illustrating examples

7.1 The harmonic oscillator

In chapter 5, a general formalism, covering both classical and quantum logics, has been introduced. Here we present a simple physical example, the harmonic oscillator, in its classical and quantum descriptions. Beyond providing a concrete illustration of the formalism developed so far, this example has the advantage of being completely solvable and illustrating the main similarities and differences between classical and quantum physics.

7.1.1 The classical harmonic oscillator

The system is described by a mass m attached to a spring of elastic constant k. The motion is assumed frictionless on the horizontal direction and the mass originally equilibrates at point 0. The spring is originally elongated to position q_0 and the system evolves then freely under the equations of motion. The setting is described in figure 7.1. The system was already studied in chapter 2. The equation of motion, giving the elongation $q(t)$ as a function of time t, is

$$\begin{aligned} m\ddot{q}(t) &= f(q(t)) = -kq(t) \\ q(0) &= q_0 \\ \dot{q}(0) &= v_0 = 0. \end{aligned}$$

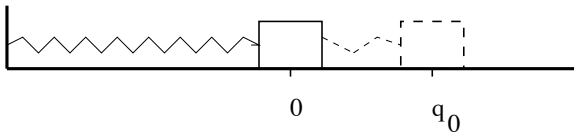

Figure 7.1: The experimental setting of the one-dimensional harmonic oscillator.

Introducing the new variable $p = m\dot{q}$ and transforming the second order differential equation into a system of first order equations, we get the vector equation

$$\frac{d\omega}{dt}(t) = A\omega(t), \qquad (*)$$

where

$$\omega(t) = \begin{pmatrix} q(t) \\ p(t) \end{pmatrix}, \text{ with initial condition } \omega(0) = \begin{pmatrix} q_0 \\ p_0 \end{pmatrix}$$

and

$$A = \begin{pmatrix} 0 & \frac{1}{m} \\ -k & 0 \end{pmatrix}.$$

The solution to equation $(*)$ is given by a flow on the phase space $\Omega = \mathbb{R}^2$ given by

$$\omega(t) = T^t\omega(0),$$

where

$$T^t = \exp(tA) = \begin{pmatrix} \cos(\mu t) & \frac{\sin(\mu t)}{m\mu} \\ -\frac{k}{\mu}\sin(\mu t) & \cos(\mu t) \end{pmatrix},$$

and $\mu = \sqrt{k/m}$. Since $\det T^t = 1$, it follows that the evolution is invertible and $(T^t)^{-1} = T^{-t}$. The orbit of the initial condition $\omega(0) = \begin{pmatrix} q_0 \\ 0 \end{pmatrix}$ under the flow reads $(T^t\omega)_{t\in\mathbb{R}}$, where $\omega(t) = T^t\omega = \begin{pmatrix} q_0\cos(\mu t) \\ -q_0\frac{k}{\mu}\sin(\mu t) \end{pmatrix}$.

The system is classical, hence its logic Λ is a Boolean σ-algebra; the natural choice is $\Lambda = \mathscr{B}(\mathbb{R}^2)$. Now observables in $\mathscr{O}(\Lambda)$ are mappings $x : \mathscr{B}(\mathbb{R}) \to \Lambda \equiv \mathscr{B}(\mathbb{R}^2)$. Identify henceforth indicator functions with Borel sets in $\mathscr{B}(\mathbb{R}^2)$ (i.e. for any Borel set $B \in \mathscr{B}(\mathbb{R})$, instead of considering $x(B) = F \in \mathscr{B}(\mathbb{R}^2)$ we shall identify $x(B) = \mathbb{1}_F$.)

Let now $X:\Omega\to\mathbb{R}$ be any measurable bounded mapping and chose as $x(B)=\mathbb{1}_{X^{-1}(B)}$ for all $B\in\mathscr{B}(\mathbb{R})$. Then, on defining $X=\int\lambda x(d\lambda)$, a bijection is established between x and X. Now since $(T^t\omega)_{t\in\mathbb{R}}=(\exp(tA)\omega)_{t\in\mathbb{R}}$ is the orbit of the initial condition ω_0 in Ω, the value $X(T^t\omega)$ is well defined for all $t\in\mathbb{R}$; we denote by $X_t(\omega)\equiv X(T^t\omega)$. Then

$$\begin{aligned}\frac{dX_t}{dt}(\omega) &= \partial_1X(T^t\omega)\frac{d(T^t\omega)_1}{dt}+\partial_2X(T^t\omega)\frac{d(T^t\omega)_2}{dt}\\ &= \partial_1X(T^t\omega)\frac{dq}{dt}(t)+\partial_2X(T^t\omega)\frac{dp}{dt}(t),\end{aligned}$$

provides the evolution of X under the flow $(T^t)_t$.

The Hamiltonian is a very particular measurable bounded map on the phase space (hence an observable) $H:\Omega\to\mathbb{R}$, having the formula $H(\omega)=k\omega_1^2/2+\omega_2^2/2m$. It evolves also under the flow $(T^t)_t$: Then

$$\begin{aligned}\frac{dH_t}{dt}(\omega) &= kq(t)\dot{q}(t)+\frac{p(t)}{m}\dot{p}(t)\\ &= kq(t)\dot{q}(t)+\dot{q}(t)(-k\dot{q}(t))\\ &= 0.\end{aligned}$$

Thus, the Hamiltonian is a constant of motion. Physically it represents the energy of the system. Initially, $H(q_0,p_0)=\frac{kq_0^2}{2}=E$ and during the flow, the energy always remains E, so that *the energy takes arbitrary (but constant with respect to the flow) values* $E\in\mathbb{R}_+$. Moreover, $\partial_1H(T^t\omega)=kq(t)=-\dot{p}(t)$ and $\partial_2H(T^t\omega)=\frac{p(t)}{m}=\dot{q}(t)$. Hence we recover the Hamilton equations

$$\begin{aligned}\frac{dq}{dt}(t) &= \frac{\partial H}{\partial p}=\partial_2H\\ \frac{dp}{dt}(t) &= -\frac{\partial H}{\partial q}=-\partial_1H.\end{aligned}$$

Therefore, $\frac{dX_t}{dt}=\partial_1X\partial_2H+\partial_2X(-\partial_1H)=L_HX$ with $L_H=-(\partial_1H\partial_2-\partial_2H\partial_1)$. Hence, denoting for every two function $f,g\in C^1(\Omega)$ by $\{f,g\}=\partial_1f\partial_2g-\partial_2f\partial_1g$ the Poisson's bracket, we have for the flow of an observable, assuming integrability of the evolution equation, $X_t=\exp(tL_H)X$. This means that the flow $(T^t\omega)_t)$ on Ω induces a flow $(\exp(tL_H)X)_t$ on observables. Notice also that $X_t=\exp(tL_H)X$ is a shorthand notation for

$$X_t=\sum_{n=0}^{\infty}\frac{(-t)^n}{n!}\{H,\{H,\dots\{H,X\}\dots\}\}.$$

Theorem 7.1.1 (Liouville's theorem) *Let μ be the Lebesgue measure on Ω, i.e. $\mu(d\omega_1d\omega_2)=d\omega_1d\omega_2$. Then*

1. *the measure μ is invariant under T^t, i.e. $\mu(T^tB) = \mu(B)$ for all $B \in \mathscr{B}(\mathbb{R}^2)$ and all $t \in \mathbb{R}$,*

2. *the operator L_H is formally skew-adjoint on $L^2(\Omega, \mathscr{F}, \mu)$.*

Proof:

1. $\mu(T^tB) = \int_{T^tB} d\omega_1 d\omega_2$. Now, if $\omega \in T^tB \Rightarrow T^{-t}\omega \in B$. Hence, denoting $(x_1, x_2) = T^{-t}(\omega_1, \omega_2)$, we have

$$\begin{aligned} \int_{T^tB} d\omega_1 d\omega_2 &= \int_B \frac{\partial(\omega_1, \omega_2)}{\partial(x_1, x_2)} dx_1 dx_2 \\ &= \int_B dx_1 dx_2 = \mu(B), \end{aligned}$$

   because the Jacobian verifies

$$\frac{\partial(\omega_1, \omega_2)}{\partial(x_1, x_2)} = \det \exp(tA) = 1.$$

2. L_H is not bounded on $L^2(\Omega, \mathscr{F}, \mu)$. It can be defined on dense subset of $L^2(\Omega, \mathscr{F}, \mu)$, for instance the Schwartz space $\mathbf{S}(\mathbb{R}^2)$. For $f, g \in \mathbf{S}(\mathbb{R}^2)$, we have

$$\begin{aligned} \langle f | L_H g \rangle &= \int \overline{f}(\omega) L_H g(\omega) \mu(d\omega) \\ &= -\int \overline{L_H f}(\omega) g(\omega) \mu(d\omega) + \text{ bdry terms.} \end{aligned}$$

   Now the boundary terms vanish because f and g vanish at infinity. Hence, on $\mathbf{S}(\mathbb{R}^2)$, the operator is skew-adjoint $L_H^* = -L_H$ and hence formally skew-adjoint on $L^2(\Omega, \mathscr{F}, \mu)$.

□

Notice that, as a consequence of the previous theorem, $\exp(tL_H)$ is formally unitary on $L^2(\Omega, \mathscr{F}, \mu)$.

Any probability measure p on Λ is a state. We have for all $B \in \mathscr{B}(\mathbb{R})$, $\rho_x(B) = p(x(B)) = p(X^{-1}(B))$ while $\rho_{x_t} = p(x_t(B)) = p(X_t^{-1}(T^{-t}B)) = p(x(T^{-t}B))$. Hence the flow T^t on Ω induces a convex automorphism $\tilde{\alpha}(p)(x(B)) = p(x((T^{-t}B))$ on states.

7.1.2 Quantum harmonic oscillator

Standard quantum logic Λ coincides with the family of subspaces of an infinite-dimensional Hilbert space $\mathbb{H}$. Since all separable Hilbert spaces are isomorphic, we can chose any of them. The Schrödinger's choice for the one-dimensional harmonic oscillator is $\mathbb{H} = L^2(\mathbb{R})$. States are probability measures $p : \Lambda \to [0,1]$ and thanks to Gleason's theorem, we can limit ourselves to tracial states, i.e.

$$\Lambda \ni M \mapsto p(M) = \mathsf{tr}(P_M D) = p_D(M),$$

for some $D \in \mathscr{D}(\mathbb{H})$. Symmetries are implemented by unitary operators on $\mathbb{H}$ (automorphisms on Λ.) Let $U \in \mathfrak{U}(\mathbb{H})$. Then $\alpha : M \mapsto \alpha(M) = UM$ induces a projection $P_{UM} = U^* P_M U$. Subsequently, the automorphism α induces a convex automorphism on $\mathscr{S}(\Lambda)$, given by

$$\begin{aligned} \tilde{\alpha}(p)(M) &= p_D(\alpha(M)) \\ &= \mathsf{tr}(P_{UM} D) \\ &= \mathsf{tr}(U^* P_M U D) \\ &= \mathsf{tr}(P_M D^{(U)}), \end{aligned}$$

with $D^{(U)} = UDU^*$. Physics remains invariant under time translations. Hence time translation (evolution) must be a symmetry implemented by a unitary operator $U(t)$ acting on $\mathbb{H}$. Define $U(t) = \exp(-itH/\hbar)$ (this a definition of H.) Then H is formally self-adjoint, hence an observable (a very particular one!) generating the Lie group of time translations. It will be shown below that H is time invariant. Now $U(t)$ acts on rays of $\mathbb{H}$ to give a flow. Denoting $\psi(t)U(t)\psi$, we have the *Schrödinger's evolution equation* in the *Schrödinger's picture*:

$$i\hbar \frac{d\psi}{dt}(t) = H\psi(t).$$

Thanks to the spectral theorem (and, identifying for $x \in \mathscr{O}(\Lambda)$ and $B \in \mathscr{B}(\mathbb{R})$, $x(B)$ with the projection-valued measure corresponding to the subspace $x(B)$), there is a bijection between $x \in \mathscr{O}(\Lambda)$ and self-adjoint operators on $\mathbb{H}$ through $X = \int \lambda x(d\lambda)$. For every tracial states p_D, we have $\mathbb{E}_{p_D}(X) = \int \lambda \, \mathsf{tr}(x(d\lambda)D)$ and

$$\begin{aligned} \mathbb{E}_{\tilde{\alpha}(p_D)}(X) &= \int \lambda \, \mathsf{tr}(x(d\lambda) D^{U(t)}) \\ &= \int \lambda \, \mathsf{tr}(U^*(t) x(d\lambda) U(t) D) \\ &= \mathbb{E}_{p_D}(X_t), \end{aligned}$$

where we defined $X_t = U^*(t) X U(t)$. Hence the flow $U(t)\psi$ on $\mathbb{H}$ induces a flow on observables satisfying

$$\frac{dX_t}{dt} = \frac{i}{\hbar}[H, X] = L_H X$$

with $L_H(\cdot) = \frac{i}{\hbar}[H,\cdot]$. Notice incidentally that $dH_t/dt = 0$ proving the claim that H is a constant of motion. Moreover, H has dimensions $M.L^2/T^2$ (energy), therefore H is interpreted as the quantum Hamiltonian. If the flow is integrable, we have

$$\begin{aligned} X_t &= \exp(tL_H)X \\ &= \sum_{n=0}^{\infty} \frac{(it)^n}{n!}[H,[H,\ldots,[H,X]\ldots]]. \end{aligned}$$

Physics remains invariant also by space translations. Hence they must correspond to a symmetry implemented by a unitary transformation.

Lemma 7.1.2 *The operator ∇_x is formally skew-adjoint on $L^2(\mathbb{R})$.*

Proof: For all $f,g \in \mathbf{S}(\mathbb{R})$ (dense in $L^2(\mathbb{R})$), we have, $\langle f \,|\, \nabla_x g \rangle = \int \overline{f}(x)\frac{d}{dx}g(x)dx = -\int \frac{d}{dx}\overline{f}(x)g(x)dx + fg|_{-\infty}^{\infty}$. □

Consequently, the operator $\exp(x\cdot\nabla_x)$ is formally unitary and since $\exp(x\cdot\nabla_x)\psi(y) = \psi(y+x)$, ∇_x is the generator of space translations. If we write $p = \frac{\hbar}{i}\nabla_x$ then p is formally self-adjoint, has dimensions $L\cdot M\cdot(L/T^2)\cdot(1/L) = M\cdot L/T$ (momentum), and $\exp(ix\cdot\nabla_x/\hbar)$ is unitary and implements space translations.

Define $H_{\text{osc}} = p^2/2m + kq^2$ as the formally self-adjoint operator on $L^2(\mathbb{R})$, with $p = \frac{\hbar}{i}\nabla_x$ and $q\psi(x) = x\psi(x)$, the multiplication operator. Introduce $\mu = \sqrt{k/m}$, $Q = \sqrt{m\mu/\hbar}q$, $P = (1/\sqrt{m\mu\hbar})p$, and $H = (1/\hbar\mu)H_{\text{osc}}$. Then $H = (1/2)(P^2+Q^2)$ where $P = -i\nabla$ and Q is the multiplication operator; these two latter operators are formally self-adjoint and verify the commutation relation $[P,Q] = -i\mathbb{1}$.

Definition 7.1.3 (Creation and annihilation operators) Define the *creation operator* $A^* = \frac{1}{\sqrt{2}}(P+iQ)$ and the *annihilation operator* $A = \frac{1}{\sqrt{2}}(P-iQ)$.

Exercise 7.1.4 For the creation and annihilation operators, show

1. $[A,A^*] = \mathbb{1}$,
2. $H = A^*A + \mathbb{1}/2$,
3. $[H,A] = -A$,
4. $[H,A^*] = A^*$,
5. for $n \in \mathbb{N}$, $[H,(A^*)^n] = n(A^*)^n$.

Lemma 7.1.5 *If $\psi_0 \in \mathbf{S}(\mathbb{R})$ is a ray (in the L^2 sense) satisfying $A\psi_0 = 0$ then*

1. $\psi_0(x) = \pi^{-1/4}\exp(-x^2/2)$,
2. $H\psi_0 = \psi_0/2$, *and*
3. $H(A^*)^n\psi_0 = (1/2+n)A^*n\psi_0$, *for all* $n \in \mathbb{N}$.

Proof:

$$\begin{aligned} A\psi_0 = 0 \quad &\Rightarrow \quad \frac{1}{\sqrt{2}}(P - iQ)\psi_0 \\ &\Rightarrow \quad -i\frac{d}{dx}\psi_0(x) - ix\psi_0(x) = 0 \\ &\Rightarrow \quad \psi_0(x) = x\exp(-x^2/2), \end{aligned}$$

and by normalisation, $c = \pi^{-1/4}$. □

Lemma 7.1.6 *Denote, for $n \in \mathbb{N}$, $\psi_n = \frac{1}{\sqrt{n!}}A^{*n}\psi_0$. Then*

1. $(\psi_n)_{n\in\mathbb{N}}$ *is an orthonormal sequence,*
2. $A^*\psi_n = \sqrt{n+1}\psi_{n+1}$, *for* $n \geq 0$,
3. $A\psi_n = \sqrt{n}\psi_{n-1}$, *for* $n \geq 1$, *and*
4. $A^*A\psi_n = n\psi_n$, *for* $n \geq 0$.

Proof: All the assertions can be shown by similar arguments. It is enough to show the arguments leading to orthonormality:

$$\begin{aligned} \langle\, \psi_0 \,|\, A^n A^{*n} \psi_0 \,\rangle &= \langle\, \psi_0 \,|\, A^{n-1} A A^* A^{*n-1} \psi_0 \,\rangle \\ &= \langle\, \psi_0 \,|\, A^{n-1}(\mathbb{1} + A^*A) A^{*n-1} \psi_0 \,\rangle \\ &\vdots \\ &= n\langle\, \psi_0 \,|\, A^{n-1} A^{*n-1} \psi_0 \,\rangle \\ &\vdots \\ &= n!\langle\, \psi_0 \,|\, \psi_0 \,\rangle. \end{aligned}$$

□

Theorem 7.1.7 *The sequence $(\psi_n)_{n\in\mathbb{N}}$ is a complete orthonormal sequence in $\mathbb{H}$.*

The proof is based on an analogous result for Hermite polynomials that can be shown using the two following lemmata.

Lemma 7.1.8 *Let* $c_{n,j} = \frac{n!}{(n-2j)!2^j j!}$, *for* $n \in \mathbb{N}$, *and* $j \in \mathbb{N}$ *such that* $0 \leq j \leq n/2$. *Then*

$$c_{n,j} = (1 - \frac{2j}{n+1})c_{n+1,j} = \frac{2(j+1)}{(n+1)(n-2j)})c_{n+1,j+1}$$

and if

$$\eta_n(x) = \sum_{j=0}^{[n/2]} (-1)^j c_{n,j} x^{n-2j},$$

then

$$(x - \frac{d}{dx})\eta_n(x) = \eta_{n+1}(x)$$

while $x^n = \sum_{j=0}^{[n/2]} c_{n,j}\eta_{n-2j}(x)$.

Proof: Substitute and make induction. □

Lemma 7.1.9 $(A^{*n}\psi_0)(x) = \eta_n(\sqrt{2}x)\psi_0(x)$.

Proof: True for $n = 0$. Conclude by induction. □

Corollary 7.1.10 $\mathsf{spec}(H) = 1/2 + \mathbb{N}$.

Therefore the energy is quantised in quantum mechanics i.e. it can take only discrete values. It is this surprising phenomenon that gave its adjective *quantum* to the term quantum mechanics.

Exercise 7.1.11 Using Dirac's notation $|n\rangle \equiv \psi_n$, for $n \in \mathbb{N}$,

1. $H|n\rangle = (1/2+n)|n\rangle$,
2. $A^*|n\rangle = \sqrt{n+1}|n+1\rangle$,
3. $A|n\rangle = \sqrt{n}|n-1\rangle$, and
4. $A^*A|n\rangle = n|n\rangle$.

7.2 Tunnel effect

Chapter 8

Turing machines, algorithms, computing, and complexity classes

All computers, from Babbage's never constructed project of analytical machine (1833) to the latest model of supercomputer, are based on the same principles. A *universal computer* uses some *input* (a sequence of bits) and a *programme* (a sequence of instructions) to produce an *output* (another sequence of bits.) Universal computers are modelled by *Turing machines.* Never forget however that an abstract Turing machine never computed anything. We had to wait until the first ENIAC was physically constructed to obtain the first output of numbers.

8.1 Deterministic Turing machines

There are several variants of deterministic Turing machines; all of them are equivalent in the sense that a problem solvable by one variant is also solvable by any other variant within essentially the same amount of time (see below, definition **??** and section 8.3.) A Turing machine is a model of computation; it is to be thought as a finite state machine disposing of an infinite scratch space (an external tape[1].) The tape consists of a semi-infinite or infinite sequence of squares, each of which can hold a single symbol. A tape-head can read a symbol from the tape, write a symbol on the tape, and move one square in either direction (for semi-infinite tape, the head cannot cross the origin.) More precisely, a Turing machine is defined as follows.

[1] Mind that during Turing's times no computer was physically available. The external tape was invented by Alan Turing — who was fascinated by typewritters — as an external storage device.

Definition 8.1.1 A *deterministic Turing machine* is a quadruple (A, S, u, s_0) where

1. A is a finite set, the *alphabet*, containing a particular symbol called the *blank symbol* and denoted by $\sharp$; the alphabet deprived from its blank symbol, denoted $A_b = A \setminus \{\sharp\}$, is assumed non-empty,
2. S is a finite non-empty set, the *states* of the machine, partitioned into the set S_i of intermediate states and the set S_f of final states,
3. $D = \{L, R\} \equiv \{-1, 1\}$ is the *displacement set*,
4. $u : A \times S \to A \times S \times D$ is the *transition function*, and
5. $s_0 \in S_i$ the *initial state* of the machine.

The set of deterministic Turing machines is denoted by DTM.

The machine is presented an input, i.e. a finite sequence of contiguous non-blank symbols, and either it stops by producing an output, i.e. another finite sequence of symbols, else the programme does never halt.

Example 8.1.2 (A very simple Turing machine) Let $M \in \mathsf{DTM}$ with $A = \{0, 1, \sharp\}$, $S = S_i \cup S_f$ where $S_i = \{\mathsf{go}\}$, $S_f = \{\mathsf{halt}\}$, and transition function $u(a, s) = (a', s', d)$ defined by the following table:

a	s	a'	s'	d
0	go	0	go	L
1	go	1	go	L
$\sharp$	go	$\sharp$	halt	R

If the programme, described by this Turing machine, starts with the head over any non-blank symbol of the input string, it ends with the head over the leftmost non-blank symbol while the string of symbols remains unchanged.

Other equivalent variants of the deterministic Turing machine may have displacement sets with a 0 (do not move) displacement, have their alphabet A partitioned into external and internal alphabet, etc. The distinction into internal and external alphabet is particularly useful in the case of semi-infinite tape, an internal character $*$, identified as "first symbol", can be used to prevent the head from going outside the tape. It is enough to define $U(*, \mathsf{go}) = (*, \mathsf{go}, R)$.

Notation 8.1.3 If W is a finite set, we denote by $W^* = \cup_{n\in\mathbb{Z}_+} W^n$ and $W^\infty = \partial W = W^{\mathbb{Z}_+}$. Notice that $\mathbb{Z}_+ = \{0,1,2,\ldots\} \neq \mathbb{N} = \{1,2,\ldots\}$ and that $W^0 = \{\emptyset\}$. Elements of W^* are called *words* of finite length over the alphabet W. For every $w \in W^*$, there exists $n \in \mathbb{Z}_+$ such that $w \in W^n$; we denote then by $|W| = n$ the *length* of the word w.

For every $\alpha \in A_b^*$, we denote by $\overline{\alpha} \in A^\infty$ the completion of the word α by blanks, namely $\overline{\alpha} = (\alpha_1, \ldots, \alpha_{|\alpha|}, \sharp, \sharp, \sharp, \ldots)$.

Considering the example 8.1.2, we can, without loss of generality, always assume that the machine starts at the first symbol of the input string $\alpha = \alpha \in A_b^*$. Starting from $(\alpha, s_0, h_0 = 1)$, successive applications of the transition function U induce a dynamical system on $\mathbb{X} = A^* \times S \times \mathbb{Z}$. A configuration is an instantaneous description of the word written on the tape, the internal state of the machine, and the position of the head, i.e. an element of $\mathbb{X}$.

Let $\tau_\alpha = \inf\{n \geq 1 : s_n \in S_f\}$. The programme starting from initial configuration $(\alpha, s_0, h_0 = 1)$ stops running if $\tau_\alpha < \infty$, it never halts when $\tau_\alpha = \infty$. While $1 \leq n < \tau_\alpha$, the sequence $(\alpha^{(n)}, s_n, h_n)_{n\leq\tau_\alpha}$ is defined by updates of single characters; if, for $0 \leq n < \tau_\alpha$, we have $u(\alpha_{h_n}^{(n)}, s_n) = (a', s', d)$, then $(\alpha^{(n+1)}, s_{n+1}, h_{n+1})$, is defined by

$$\begin{aligned} s_{n+1} &= s' \\ h_{n+1} &= h_n + d \\ \alpha^{(n+1)} &= (\alpha_1^{(n)}, \ldots, \alpha_{h_n-1}^{(n)}, a', \alpha_{h_n+1}^{(n)}, \ldots, \alpha_{|\alpha^{(n)}|}^{(n)}). \end{aligned}$$

If the machine halts at some finite instant, the output is obtained by reading the tape from left to right until the first blank character. The sequence of words $(\alpha^{(n)})_n$ is called a *computational path* or *computational history* starting from α.

8.2 Computable functions and decidable predicates

Every $M \in \mathsf{DTM}$ computes a particular partial function $\phi_M : A_b^* \to A_b^*$. Since the value of $\phi_M(\alpha)$ remains undetermined when the programme M does not halt, the function ϕ_M is termed partial because in general $\mathsf{Dom}(\phi_M) \subset A_b^*$.

Definition 8.2.1 A partial function $f : A_b^* \to A_b^*$ is called *computable* if there exists a $M \in \mathsf{DTM}$ such that $\phi_M = f$. In that case, f is said to be computed by the programme M.

Exercise 8.2.2 Show that there exist non-computable functions.

Definition 8.2.3 A *predicate*, P, is a function taking Boolean values 0 or 1. A *language*, L, over an alphabet A is a subset of A_b^*.

Thus, for predicates P with $\mathsf{Dom}(P) = A_b^*$, the set $\{\alpha \in A_b^* : P(\alpha)\}$ is a language. Hence predicates are in bijection with languages.

Definition 8.2.4 A predicate $P : A_b^* \to \{0,1\}$ is *decidable*, if the function P is computable.

Let P be a predicate and L the corresponding language. The predicate is decidable if there exists a $M \in \mathsf{DTM}$ such that for every word α, the programme halts after a finite number of steps and

- if $\alpha \in L$, then the machine halts returning 1, and
- if $\alpha \notin L$, then the machine halts returning 0.

Definition 8.2.5 Let $M \in \mathsf{DTM}$ and $s_M, t_M : \mathbb{Z}_+ \to \mathbb{R}_+$ be given functions. If for every $\alpha \in A_b^*$, the machine stops after having visited at most $s_M(|\alpha|)$ cells, we say that it works in *computational space* s_M. We say that it works in *computational time* t_M if $\tau_\alpha \leq t_M(|\alpha|)$.

8.3 Complexity classes

Computability of a function does not mean effective computability since the computing algorithm can require too much time or space. We say that $r : \mathbb{N} \to \mathbb{R}_+$ is of *polynomial growth* if there exist constants $c, C > 0$ such that $r(n) \leq Cn^c$, for large n. We write symbolically $r(n) = \mathsf{poly}(n)$.

Henceforth, we shall assume $A_b \equiv \mathbb{A} = \{0,1\}$.

Definition 8.3.1 The *complexity class* P consists of all languages L whose predicates P are *decidable in polynomial time*, i.e. for every L in the class, there exists a machine $M \in \mathsf{DTM}$ such that $\phi_M = P$ and $t_M(|\alpha|) = \mathsf{poly}(|\alpha|)$ for all $\alpha \in \mathbb{A}^*$.

Similarly, we can define the class PSPACE of languages whose predicates are *decidable in polynomial space*. functions computable in polynomial space.

Other complexity classes will be determined in the subsequent sections. Obviously P $\subseteq$ PSPACE.

Conjecture 8.3.2 P $\neq$ PSPACE.

8.4 Non-deterministic Turing machines and the NP class

Definition 8.4.1 A *non-deterministic Turing machine* is a quadruplet $(A, S, \mathbf{u}, s_0)$ where A, S and s_0 are as in definition 8.1.1; $\mathbf{u}$ is now a multivalued function, i.e. there are r different branches $u_i, i = 1, \ldots, r$ and $u_i : A \times S \to A \times S \times D$. For every pair $(a, s) \in A \times S$ there are different possible outputs $(a'_i, \sigma'_i, d_i)_{i=1,\ldots,r}$, the choice of a particular branch can be done in a non-deterministic way at each moment. All such choices are legal actions. The set of non-deterministic Turing machines is denoted by NTM.

A computational path for a $M \in$ NTM is determined by a choice of one legal transition at every step. Different steps are possible for the same input. Notice that NTM do not serve as models of practical devices but rather as logical tools for the formulation of problems rather than their solution.

Definition 8.4.2 A language L (or its predicate P) belongs to the NP class if there exists a $M \in$ NTM such that

- if $\alpha \in L$ (i.e. $P(\alpha) = 1$) for some $\alpha \in \mathbb{A}^*$, then there exists a computational path with $\tau_\alpha \leq \text{poly}(|\alpha|)$ returning 1,
- if $\alpha \notin L$ (i.e. $P(\alpha) = 0$) for some $\alpha \in \mathbb{A}^*$, then there exists no computational path with this property.

It is elementary to show that P $\subseteq$ NP. Clay Institute offers you[2] USD 1 000 000 if you solve the following

Exercise 8.4.3 Is it true that P = NP?

8.5 Probabilistic Turing machine and the BPP class

Definition 8.5.1 Let $\tilde{\mathbb{R}}$ be the set of real numbers computable by a deterministic Turing machine within accuracy 2^{-n} in poly(n) time. A *probabilistic Turing machine* is a quintuple $(A, S, \mathbf{u}, \mathbf{p}, s_0)$ where A, S, $\mathbf{u}$, and s_0 are as in definition 8.4.1

[2] http://www.claymath.org/millennium/

while $\mathbf{p} = (p_1, \ldots, p_r) \in \tilde{\mathbb{R}}_+$, with $\sum_{i=1}^r p_i = 1$ is a probability vector on the set of branches of $\mathbf{u}$. All branches correspond to legal actions; at each step, the branch i is chosen with probability p_i, independently of previous choices. The set of probabilistic Turing machine is denoted by PTM.

Each $\alpha \in \mathbb{A}^*$ generates a family of computational paths. The local probability structure on the transition functions induces a natural probability structure on the computational path space. The evolution of the machine is a Markov process with the state space $A_b^* \times S \times \mathbb{Z}$ and stochastic evolution kernel determined by the local probability vector $\mathbf{p}$. Hence, any input gives a set of possible outputs each of them being assigned a probability of occurrence. A machine in PTM is also called a *Monte Carlo algorithm*.

Definition 8.5.2 Let $\varepsilon \in]0, 1/2[$. A predicate P (hence a language L) belongs to the BPP class if there exists a $M \in$ BPP such that for any $\alpha \in \mathbb{A}^*$, $\tau_\alpha \leq \mathsf{poly}(|\alpha|)$ and

- if $\alpha \in L$, then $\mathbb{P}(P(\alpha) = 1) \geq 1 - \varepsilon$, and
- if $\alpha \notin L$, then $\mathbb{P}(P(\alpha) = 1) \leq \varepsilon$.

Exercise 8.5.3 Show that the definition of the class does not depend on the choice of ε provided it lies in $]0, 1/2[$.

Church-Turing thesis ...

8.6 Boolean circuits

Notation 8.6.1 For $b \in \mathbb{N}$ and $\mathbb{Z}_b = \{0, \ldots, b-1\}$, we denote by $x = \langle x_{n_1} \cdots x_0 \rangle_b$ the mapping defined by

$$\mathbb{Z}_b^n \ni (x_0, \ldots, x_n) \mapsto x = \langle x_n \cdots x_0 \rangle_b = \sum_{k=0}^{n-1} x_k b^k \in \mathbb{Z}_{b^n}.$$

Since conversely for every $x \in \mathbb{Z}_{b^n}$ the sequence $(x_0, \ldots, x_n) \in \mathbb{Z}_b^n$ is uniquely determined, we identify x with the sequence of its digits. For $b = 2$ we omit the basis subscript and we write simply $\langle \cdot \rangle$.

Definition 8.6.2 Let $f : \mathbb{A}^n \to \mathbb{A}^m$ be a Boolean function of n entries and m outputs. Let **B** be a fixed set of Boolean functions of different arities. We call *Boolean circuit* of f in terms of the basis **B** a representation of f in terms of functions from **B**.

Example 8.6.3 (Addition with carry of 2 binary 2-digit numbers) Let $x = \langle x_1 x_0 \rangle$ and $y = \langle y_1 y_0 \rangle$. We wish to express $z = x + y = \langle z_2 z_1 z_0 \rangle$ in terms of Boolean functions in $\mathbf{B} = \{\text{XOR}, \text{AND}\} = \{\oplus, \wedge\}$. The truth table is given in table 8.1. We

x_1	x_0	y_1	y_0	z_2	z_1	z_0
0	0	0	0	0	0	0
0	1	0	0	0	0	1
1	0	0	0	0	1	0
1	1	0	0	0	1	1
0	0	0	1	0	0	1
0	1	0	1	0	1	0
1	0	0	1	0	1	1
1	1	0	1	1	0	0
0	0	1	0	0	1	0
0	1	1	0	0	1	1
1	0	1	0	1	0	0
1	1	1	0	1	0	1
0	0	1	1	0	1	1
0	1	1	1	1	0	0
1	0	1	1	1	0	1
1	1	1	1	1	1	0

Table 8.1: The truth table of the Boolean function $\mathbb{A}^4 \to \mathbb{A}^3$ implementing the addition with carry of two binary 2-digit numbers.

verify immediately that:

$$\begin{aligned} z_0 &= x_0 \oplus y_0 \\ z_1 &= (x_0 \wedge y_0) \oplus (x_1 \oplus y_1) \\ z_2 &= (x_1 \wedge y_1) \oplus [(x_1 \oplus y_1) \wedge (x_0 \wedge y_0)] \end{aligned}$$

Consequently, the Boolean circuit is depicted in figure **??**.

A basis **B** is *complete* if any Boolean function f can be constructed as a circuit with gates from **B**.

Example 8.6.4 $\{\text{NOT}, \text{OR}, \text{AND}\}$ is a complete but redundant basis; $\{\text{NOT}, \text{OR}\}$, $\{\text{NOT}, \text{AND}\}$, and $\{\text{AND}, \text{XOR}\}$ are complete minimal bases.

Definition 8.6.5 The minimal number of gates from **B** needed to compute f, denoted by $c_B(f)$, is *circuit complexity* of f in **B**.

The function implementing the addition with carry of table 8.1 over the basis $\mathbf{B} = \{\text{AND}, \text{XOR}\}$, has circuit complexity 7.

Any DTM can be implemented by circuits.

8.7 Classical information, entropy, and irreversibility

The information content of a message is a probabilistic notion. The less probable a message is, the more information it carries. Let X be a random variable defined on $(\Omega, \mathscr{F}, \mathbb{P})$ taking values in the finite set $\mathbb{X} = \{x_1, \ldots, x_n\}$ Let $\mathsf{PV}_n = \{\mathbf{p} \in \mathbb{R}_+^n : \sum_{i=1}^n p_i = 1\}$. To each element $\mathbf{p} \in \mathsf{PV}_n$ corresponds a probability measure $\mathbb{P}_X^{\mathbf{p}}$ defined by $\mathbb{P}_X^{\mathbf{p}}(x_i) = p_i$, for $i = 1, \ldots, n$. Ask about the information content carried by the random variable X is the same thing as trying to quantify the predictive power of the law $\mathbb{P}_X$. The main idea is that the information content of X is equal to the average information missing in order to decide the outcome value of X when the only thing we know is its law $\mathbb{P}_X$. Some reasonable requirements on the information content of X are given below:

- Suppose that all $p_i, i = 1, \ldots, n$ but one are 0 and $p_j = 1$, for some j. Then $\mathbb{P}_X^{\mathbf{p}}(x_j) = 1$ and no information is missing, there is no uncertainty about the possible outcome of X;
- Suppose on the contrary that $p_i = 1/n, i = 1, \ldots, n$. Our perplexity is maximal and this perplexity increases with n.
- If S is to be interpreted as a missing information associated with a probability vector $\mathbf{p} \in \mathsf{PV}_n$, on denoting $\mathsf{PV} = \cup_{n \in \mathbb{N}} \mathsf{PV}_n$, the function $S : \mathsf{PV} \to \mathbb{R}_+$ and the first statement implies that $S(1, 0, 0, \ldots, 0) = 0$ while $S(1/n, \ldots, 1/n)$ is an increasing function of n.
- The function S must be invariant under permutations of its arguments i.e. $S(p_{\sigma(1)}, \ldots, p_{\sigma(n)}) = S(p_1, \ldots, p_n)$ for all the permutations $\sigma \in S_n$.
- If we split the possible outcome values into two sets, the function S must verify the *grouping property*, i.e.

$$\begin{aligned} S(p_1, \ldots, p_n; p_{n+1}, \ldots, p_N) &= S(q_A, q_B) \\ &\quad + q_A S\left(\frac{p_1}{q_A}, \ldots, \frac{p_n}{q_A}\right) \\ &\quad + q_B S\left(\frac{p_{n+1}}{q_B}, \ldots, \frac{p_N}{q_B}\right), \end{aligned}$$

where $q_A = p_1 + \ldots + p_n$ and $q_B = p_{n+1} + \ldots + p_N$.

- Finally, we require $S(p_1,\ldots,p_n;0,\ldots,0) = S(p_1,\ldots,p_n)$.

Theorem 8.7.1 *The only function $S : PV \to \mathbb{R}_+$ satisfying the above requirements is the function defined by*

$$PV \ni (p_1,\ldots,p_n) \mapsto S(p_1,\ldots,p_n) = -k\sum_{i=1}^{n} p_i \log p_i,$$

where k is an arbitrary non-negative constant and the convention $0\log 0 = 0$ is used. The function S is called the (classical) entropy of the probability vector.

Proof: (To be filled in a later version.) □

Entropy is closely related to irreversibility since the second principle of thermodynamics states: *Entropy of an isolated system is a non decreasing function of time. It can remain constant only for reversible evolutions.* For a system A undergoing an irreversible transformation the entropy increases; however the system can be considered as part of a larger isolated composite system (A and environment), undergoing globally a reversible transformation. In that case the total entropy (of the system A and of the environment) remains constant but since the entropy of A must increase, the entropy of the environment must decrease[3] hence the missing information decreases. In other words, when the system A undergoes an irreversible transformation, the environment gains information.

This leads to the Landauer's principle: *When a computer erases a single bit of information, the environment gains at least $k\ln 2$ units of information, where $k > 0$ is a constant.*

Classical computers are based on gates $\{\text{XOR}, \text{AND}\}$ for example. It is easily shown that these gates are irreversible. Therefore it is intuitively clear why classical computers can produce information. What is much less intuitively clear is how quantum processes can produce information since they are reversible (unitary).

In 1973, BENNETT predicted that it is possible to construct reversible universal gates. In 1982, FREDKIN exemplifies such a reversible gate. *Fredkin's gate* is a 3 inputs - 3 outputs gate, whose truth tableau is given in table **??**. This gate produces both AND (since inputs $0,x,y$ return outputs $x\wedge y, \bar{x}\wedge y, x$) and NOT gates (since inputs $1,0,x$ return outputs $\bar{x},x,x$.) The gates AND and NOT forming a complete basis for Boolean circuits, the universality of Freidkin's gate is established.

[3] Notice that this assertion is not in contradiction with the second principle of thermodynamics because the environment is not isolated.

Input			Output		
a	b	c	a'	b'	c'
0	0	0	0	0	0
0	0	1	0	0	1
0	1	0	0	1	0
0	1	1	1	0	1
1	0	0	1	0	0
1	0	1	0	1	1
1	1	0	1	1	0
1	1	1	1	1	1

Table 8.2: The truth table of Fredkin's gate. We remark that $c' = c$ and if $c = 0$ then ($a' = a$ and $b' = b$) else ($a' = b$ and $b' = a$.)

In 1980, BENIOFF describes how to use quantum mechanics to implement a Turing machine, in 1982, FEYNMAN proves that there does not exist a Turing machine (either deterministic or probabilistic) on which quantum phenomena can be efficiently simulated; only a quantum Turing machine could do so. Finally, in 1985, DEUTSCH constructs (on paper) a universal quantum Turing machine.

8.8 Composite quantum systems, tensor products, and entanglement

Definition 8.8.1 Let V, W be two vector spaces. Their *tensor product* is a vector space, denoted $V \otimes W$, and a bilinear map $\otimes : V \times W \to V \otimes W$, satisfying the following universality property: for any vector space Z and any bilinear map $\beta : V \times W \to Z$, there is a unique linear map $f : V \otimes W \to Z$, such that $\beta(v,w) = f(v \otimes w)$ for all $v \in C$ and $w \in W$.

Remark 8.8.2 The meaning of this definition for finite dimensional spaces is the following: if $(e_1, \ldots, e_{\dim V})$ and $(f_1, \ldots, f_{\dim W})$ are bases of V and W, then $(e_i \otimes f_j)_{1 \leq i \leq \dim V; 1 \leq j \leq \dim W}$ is a basis of $\mathsf{Var} \otimes W$ and if $v = \sum_i v_i e_i$ and $w = \sum_j w_j f_j$, then $\otimes(v,w) = v \otimes w = \sum_{i,j} v_i w_j e_i \otimes f_j$.

Corollary 8.8.3 $\dim(V \otimes W) = \dim V \dim W$.

Let $(|0\rangle,|1\rangle) \equiv (|x\rangle)_{x\in\mathbb{A}}$, where $\mathbb{A}=\{0,1\}$, be a basis in $\mathbb{H}=\mathbb{C}^2$. Then a basis of $\mathbb{H}^{\otimes n}$ is given by

$$\begin{aligned}(|0\rangle\otimes\cdots\otimes|0\rangle, \quad &\ldots \quad ,|1\rangle\otimes\cdots\otimes|1\rangle)\\ &= \quad (|x_n\cdots x_1\rangle)_{x_i\in\mathbb{A}}\\ &= \quad (|x\rangle)_{x\in\{0,\ldots,2^n-1\}}).\end{aligned}$$

Hence any basis vector represent the integer $x=\sum_{i=1}^{n} x_i 2^{i-1}\in\mathbb{Z}_{2^n}$. Therefore the vector $\psi=\sum_{x=0}^{2^n-1}\psi_x|x\rangle$, with $\psi_x=1/2^{n/2}$ represents *simultaneously* 2^n integers.

An operator X acting on $\mathbb{H}$ can be represented in an arbitrary orthonormal basis $(|i\rangle)_{i=1,\ldots,\dim\mathbb{H}}$ by the matrix (possibly infinite-dimensional) of matrix elements x_{jk} by

$$X=\sum_{j,k} x_{j,k}|j\rangle|k\rangle,$$

where $x_{jk}=\langle j|Xk\rangle$.

The scalar product in $\mathbb{H}^{\otimes n}$ reads

$$\begin{aligned}&\langle(|\psi_1\rangle\otimes\cdots\otimes|\psi_n\rangle)\,|\,|\xi_1\rangle\otimes\cdots\otimes|\xi_n\rangle)\rangle\\ =\;&\langle\psi_1\cdots\psi_n|\xi_1\cdots\xi_n\rangle\\ =\;&\langle\psi_1|\xi_1\rangle\cdots\langle\psi_n|\xi_n\rangle.\end{aligned}$$

The tensor product of operators is defined similarly

$$(A\otimes B)|\psi\rangle|\xi\rangle=(A|\psi\rangle)\otimes(B|\xi\rangle).$$

If $A=\sum a_{j,k}|j\rangle k$ and $B=\sum b_{j,k}|j\rangle k$, then $C=A\otimes B$ has matrix elements

$$c_{(jk)(lm)}=a_{jl}b_{km}.$$

Definition 8.8.4 Let $\psi\in\mathbb{H}_1\otimes\cdots\otimes\mathbb{H}_n$, for $n\geq 2$. The pure state ψ is called *entagled* if it cannot be written as a tensor product $\psi=\psi_1\otimes\cdots\otimes\psi_n$, with $\psi_i\in\mathbb{H}_i$, for all $i=1,\ldots,n$.

Example 8.8.5 Let $n=2$ and $\mathbb{H}_1=\mathbb{H}_2=\mathbb{C}^2$. A basis of $\mathbb{H}^{\otimes 2}$ is given by $(|00\rangle,|01\rangle,|10\rangle,|11\rangle)$. An arbitrary vector $\psi\in\mathbb{H}^{\otimes 2}$ is decomposed as

$$\psi=\psi_0|00\rangle+\psi_1|01\rangle+\psi_2|10\rangle+\psi_3|11\rangle.$$

If $\psi_2=\psi_3=0$, then $\psi=\psi_0|00\rangle+\psi_1|01\rangle=|0\rangle\otimes(\psi_0|0\rangle+\psi_1|1\rangle)$ and the state is not entangled. If $\psi_1=\psi_2=0$ while $\psi_0\psi_3\neq 0$ then the state is entangled since it cannot be written as a tensor product.

8.9 Quantum Turing machines

Definition 8.9.1 Let $\tilde{\mathbb{C}}$ be the set of complex numbers whose real and imaginary part can be computed by a deterministic algorithm with precision 2^{-n} within $\mathrm{poly}(n)$ time. A *pre-quantum Turing machine* is a quadruple (A,S,c,s_0), where A,S,s_0 are as for a deterministic machine and $c : (A\times S)^2 \times D \to \tilde{C}$, where D is the displacement set.

Any configuration x of the machine is represented by a triple $x = (\alpha,s,h) \in A^* \times S \times \mathbb{Z} = \mathbb{X}$. The quantum configuration space $\mathbb{H}$ is decomposed into $\mathbb{H}_T \otimes \mathbb{H}_S \otimes \mathbb{H}_H$, where the indices T,S,H stand respectively for tape, internal states, and head. The space $\mathbb{H}$ is spanned by the orthonormal system $(|\,\psi\rangle)_{\psi\in\mathbb{X}} = (|\,\alpha sh\rangle)_{\alpha\in A^*,s\in S,h\in\mathbb{Z}}$.

Define now onservables having $(|\,\alpha\rangle)_{\alpha\in A^*}$, $(|\,s\rangle)_{s\in S}$, and $(|\,h\rangle)_{h\in\mathbb{Z}}$ as respective eigenvectors. To do so, identify the sets A with $\{0,\ldots,|A|-1\}$ and S with $\{0,\ldots,|S|-1\}$. Denoty by $\hat{T}$, $\hat{S}$, and $\hat{H}$ the self-adjoint operators describing these observables, i.e.

$$\begin{aligned}
\hat{S} &= \sum_{s=0}^{|S|-1} s|\,s\rangle\langle\,s| \\
\hat{H} &= \sum_{h\in\mathbb{Z}} h|\,h\rangle\langle\,h| \\
\hat{T} &= \otimes_{i\in\mathbb{Z}}\hat{T}_i \text{ where } \hat{T}_i = \sum_{a=0}^{|A|-1} a|\,a\rangle\langle\,a|.
\end{aligned}$$

Due to the linearity of quantum flows, it is enough to describe the flow on the basis vectors $\psi = |\,\alpha,s,h\rangle; a \in A^{\mathbb{Z}}, s \in S, h \in \mathbb{Z}$. The machine is prepared at some initial pure state $\psi = |\,\alpha,s,h\rangle$, with α a string of contiguous non blank symbols and we assume that the time is discretised:

$$|\,\psi_n\rangle = U^n|\,\psi\rangle.$$

Suppose that the displacement set D reads $\{-1,0,1\}$. Then for $\psi = |\,\alpha,s,h\rangle$ and $\psi' = |\,\alpha',s',h'\rangle$

$$\begin{aligned}
U_{\psi,\psi'} &= \langle\,\alpha',s',h'\,|\,U\,\alpha,s,h\rangle \\
&= [\delta_{h',h+1}c(\alpha_h,s,\alpha'_h,s',1) \\
&\quad +\delta_{h',h}c(\alpha_h,s,\alpha'_h,s',0) \\
&\quad +\delta_{h',h-1}c(\alpha_h,s,\alpha'_h,s',-1)]\prod_{j\in\mathbb{Z}\setminus\{h\}}\delta_{\alpha_j,\alpha'_j}.
\end{aligned}$$

Definition 8.9.2 A pre-quantum Turing machine is called a *quantum Turing machine* if the function c is such that the operator U is unitary.

Exercise 8.9.3 Find the necessary and sufficient conditions on the function c so that U is unitary.

Wavelets, Cuntz-Krieger algebras, Bratteli-Jørgensen ...

To halt the machine, we can not perform intermediate measurements of the composite state because quantum mechanical measurement perturbs the system. To proceed, suppose that $S_f = \{\mathsf{halt}\} \equiv \{0\}$ and introduce a *halting flag* operator $\hat{F} = |0\rangle\langle 0|$. Once the state s is set to 0, the function c is such that U does not any longer change either the state s or the result of the computation.

A *predicate* is a projection operator $P_\alpha = |\alpha\rangle\langle\alpha|$. Let the machine evolve for some time n: it is at the state $|\psi_n\rangle = U^n|\psi\rangle$. Perform the measurement $\langle\psi_n|P_\alpha \otimes \hat{F} \otimes I\psi_n\rangle = p \in [0,1]$.

Definition 8.9.4 A language L belongs to the BQP complexity class if there is a machine $M \in \mathsf{QTM}$ such that

- if $\alpha \in L$, then the machine accepts with probability $p > 2/3$,
- if $\alpha \notin L$, then the machine rejects with probability $p > 2/3$,

within a running time $\mathsf{poly}(|\alpha|)$.

Theorem 8.9.5 P $\subseteq$ BPP $\subseteq$ BQP $\subseteq$ PSPACE.

Chapter 9

Cryptology

Cryptology, grouping cryptography and cryptanalysis, is an old preoccupation of mankind because information is, as a matter of fact, a valuable resource. Nowadays classical technology allows secure ciphering of information that cannot be deciphered in real time. However, the cryptologic protocols used nowadays are all based on the unproven conjecture that factoring large integers is a hard computational task. Should this conjecture be proved false, and an efficient polynomial factorisation algorithm be discovered, the security of our communication networks could become vulnerable. But even without any technological breakthrough, the ciphered messages we exchange over public channels (internet, commutated telephone network, fax, SMS, etc.) can be deciphered by spending 8–10 months of computing time; hence our information exchange is already vulnerable for transporting information that remains important 10 months after its transmission.

Quantum information acquired an unprecedented impetus when Peter SHOR [11] proved that on a quantum computer, factoring is a polynomially hard problem. On the other hand, quantum communication can use the existing technology to securely cipher information. It is therefore economically and strategically important to master the issues of advanced cryptography and to invent new cryptologic methods.

9.1 An old idea: the Vernam's code

In 1917, Gilbert VERNAM proposed [15] the following ciphering scheme[1] Let A be a finite alphabet, identified with the set $\{0,\ldots,|A|-1\}$ and m a message of length N over the alphabet A, i.e. a word $m \in A^N$. The Vernam's ciphering algorithm uses a ciphering key of same length as m, i.e. a word $k \in A^N$ and performs character-wise addition as explained in the following

---

Algorithm 9.1.1 VernamsCiphering

Require: Original message $m \in A^N$ and UNIFRANDOMGENERATOR(A^N)
Ensure: Ciphered message $c \in A^N$
  Choose randomly ciphering key $k \in A^N$
  $i \leftarrow 1$
  repeat
    Add character-wise $c_i = m_i + k_i \mod |A|$
    $i \leftarrow i+1$
  until $i > N$

The recipient of the ciphered message c, knowing the ciphering key k performs the following

---

Algorithm 9.1.2 VernamsDeciphering

Require: Ciphered message $c \in A^N$ and ciphering key $k \in A^N$
Ensure: Original message $m \in A^N$
  $i \leftarrow 1$
  repeat
    Subtract character-wise $m_i = c_i - k_i \mod |A|$
    $i \leftarrow i+1$
  until $i > N$

As far as the ciphering key is used only once and the key word has the same length as the message, the Vernam's algorithm is proved [9] to be perfectly secure. The main problem of the algorithm is how to securely communicate the key k?

9.2 The classical cryptologic scheme RSA

Theorem 9.2.1 *(Fermat's little theorem) Let p be a prime. Then*

[1] Appeared as a first patent US Patent 1310719 issued on 22 July 1919, and further improved in a series of patents: US Patent 1416765, US Patent 1584749, and US Patent 1613686.

1. *any integer a satisfies* $a^p = a \mod p$,

2. *any integer a, not divisible by p, satisfies* $a^{p-1} = 1 \mod p$.

Definition 9.2.2 The *Euler's function* $\phi : \mathbb{N} \to \mathbb{N}$ is defined by

$$\phi(n) = \mathsf{card}\{0 < a < n : \gcd(a,n) = 1\}, n \in \mathbb{N}.$$

In particular, if p is prime, then $\phi(p) = p - 1$.

Theorem 9.2.3 *(Euler's) If* $\gcd(a,m) = 1$, *then* $a^{\phi(m)} = 1 \mod m$.

Proposition 9.2.4 *Let m be an integer, strictly bigger than 1, without square factors, and r a multiple of* $\phi(m)$. *Then*

- $a^r = 1 \mod m$, *for all integers a relatively prime with respect to m, and*
- $a^{r+1} = a \mod m$ *for all integers.*

The proofs of all the previous results are straightforward but outside the scope of the present course; they can be found in pages 50–60 of [**?**].

The RSA protocols, named after its inventors RIVEST, SHAMIR, AND ADLEMAN [7], involves two legal parties: Alice and Bob, and an eavesdropper, Eve. Bob produces by the classical key distribution algorithm a private key d and a public key π. Alice uses the public key of Bob to cipher the message and Bob uses his private key to decipher it. Eve, even if she intercepts the ciphered message, cannot decipher it in real time.

Algorithm 9.2.5 ClassicalKeyDistribution

Require: Two primes p and q
Ensure: Public, π, and private, d, keys of Bob
$n \leftarrow pq$ (hence $\phi(n) = (p-1)(q-1)$)
Choose any $e < n$, such that $\gcd(e, \phi(n)) = 1$
$d \leftarrow e^{-1} \mod \phi(n)$
$\pi \leftarrow (e,n)$

Bob publishes his public key π on his internet page. Alice uses π to cipher the message m using the following

Algorithm 9.2.6 Ciphering

Require: Public key $\pi = (e,n)$ and message $m \in \mathbb{N}$, with $m < n$
Ensure: Ciphered text $c \in \mathbb{N}$
$c \leftarrow m^e \mod n$

Alice transmits the ciphered text c through a vulnerable public channel to Bob. He uses his private key to decipher by using the following

Algorithm 9.2.7 Deciphering

Require: Private key d and ciphered message $c \in \mathbb{N}$
Ensure: Deciphered text $\mu \in \mathbb{N}$
$\mu \leftarrow c^d \mod n$

Theorem 9.2.8 $\mu = m$

Proof:

$$\begin{aligned} c^d &= m^{ed} \mod n \\ ed &= 1 + k\phi(n), \text{ for some } k \in \mathbb{N} \\ m^{ed} &= m^{1+k\phi(n)}, \end{aligned}$$

and since $n = pq$ has no square factors, by using proposition 9.2.4, we get $m^{1+k\phi(n)} \mod n = m \mod n$. □

If Eve intercepts the message, to compute d she must know $\phi(n)$, hence the factoring of n into primes. Security of the protocol is based on the conjecture that it is algorithmically hard to factor n. If we denote by $N = \log n$, then it is worth noticing that when the RSA protocol has been introduced, the best known algorithm of factor n run in $\exp(N)$ time. The best [2] known algorithm nowadays [4] runs in $\exp(N^{1/3}(\log N)^{2/3})$ time. This algorithmic improvement, combined with the increasing in the computational capabilities of computers, allows the factoring of a 1000 digits number in ca. 8 months instead of a time exceeding the age of the universe at the moment the algorithm has been proposed. Until May 2007, the RSA company ran an international contest offering several hundreds thousand dollars to whoever could factor multi-digit numbers they provided on line. When the contest stopped the company gave the official reasons explained in RSA factoring challenge.

[2] See also [5] for an updated state of the art.

9.3 Quantum key distribution

Theorem 9.3.1 *(No cloning theorem) Let $|\phi\rangle$ and $|\psi\rangle$ be two rays in $\mathbb{H}$ such that $\langle\phi|\psi\rangle \neq 0$ and $|\phi\rangle \neq \exp(i\theta)|psi\rangle$. Then there does not exist any quantum device allowing duplication of ϕ and ψ.*

Proof: Suppose that such a device exists. Then, for some $n \geq 1$, there exists a unitary $U : \mathbb{H}^{\otimes(n+1)} \to \mathbb{H}^{\otimes(n+1)}$ and some ancillary ray $|\alpha_1 \cdots \alpha_n\rangle \in \mathbb{H}^{\otimes n}$ such that we get

$$\begin{aligned} |\phi\phi\beta_1 \cdots \beta_{n-1}\rangle &= U|\phi\alpha_1 \cdots \alpha_n\rangle \\ |\psi\psi\gamma_1 \cdots \gamma_{n-1}\rangle &= U|\psi\alpha_1 \cdots \alpha_n\rangle. \end{aligned}$$

Then

$$\begin{aligned} \langle\psi|\phi\rangle &= \langle\psi\alpha_1 \cdots \alpha_n|U^*U|\phi\alpha_1 \cdots \alpha_n\rangle \\ &= \langle\psi|\phi\rangle^2 \prod_{i=1}^{n-1} \langle\gamma_i|\beta_i\rangle. \end{aligned}$$

Since $\langle\phi|\psi\rangle \neq 0$ we get $\langle\psi|\phi\rangle \prod_{i=1}^{n-1} \langle\gamma_i|\beta_i\rangle = 1$ and since $|\phi\rangle \neq \exp(i\theta)|psi\rangle$, it follows that $0 < |\langle\psi|\phi\rangle| < 1$. Subsequently, $\prod_{i=1}^{n-1} |\langle\gamma_i|\beta_i\rangle| > 1$ but this is impossible since for every i, $|\langle\gamma_i|\beta_i\rangle| \leq 1$. □

This theorem is at the basis of the BB84 quantum key distribution protocol [2]. Alice and Bob communicate through a quantum and a classical public channels; they agree publicly to use two different orthonormal bases of $\mathbb{H} = \mathbb{C}^2$ (describing the photon polarisation):

$$\begin{aligned} B_+ &= \{\varepsilon_0^+ = |0\rangle, \varepsilon_1^+ = |1\rangle\} \\ B_\times &= \{\varepsilon_0^\times = \frac{|0\rangle - |1\rangle}{\sqrt{2}}, \varepsilon_1^\times = \frac{|0\rangle + |1\rangle}{\sqrt{2}}\}. \end{aligned}$$

The first element of each basis is associated with the bit 0, the second with the bit 1. Moreover Alice and Bob agree on some integer $n = (4+\delta)N$ with some $\delta > 0$, where N is the length of the message they wish to exchange securely; it will be also the length of their key. Alice needs also to know the function $T : \{0,1\}^2 \to \mathbb{H}$ defined by

$$T(x,y) = \begin{cases} \varepsilon_0^+ & \text{if } (x,y) = (0,0) \\ \varepsilon_1^+ & \text{if } (x,y) = (0,1) \\ \varepsilon_0^\times & \text{if } (x,y) = (1,0) \\ \varepsilon_1^\times & \text{if } (x,y) = (1,1). \end{cases}$$

Algorithm 9.3.2 AlicesKeyGeneration

Require: UNIFRANDOMGENERATOR($\{0,1\}$), T, n
Ensure: Two strings of n random bits $\mathbf{a},\mathbf{b} \in \{0,1\}^n$ and a sequence of n qubits $(|\psi_i\rangle)_{i=1,\ldots,n}$
  Generate randomly $a_1,\ldots,a_n$
  $\mathbf{a} \leftarrow (a_1,\ldots,a_n) \in \{0,1\}^n$
  Generate randomly $b_1,\ldots,b_n$
  $\mathbf{b} \leftarrow (b_1,\ldots,b_n) \in \{0,1\}^n$
  $i \leftarrow 1$
  repeat
    $|\psi_i\rangle \leftarrow T(a_i,b_i)$
    Transmit $|\psi_i\rangle$ to Bob via public quantum channel
    $i \leftarrow i+1$
  until $i > n$

On reception of the ith qubit, Bob performs a measurement of the projection operator $P^\sharp = |\varepsilon_1^\sharp\rangle\langle\varepsilon_1^\sharp|$, where $\sharp \in \{+,\times\}$.

Algorithm 9.3.3 BobsKeyGeneration

Require: UNIFRANDOMGENERATOR($\{0,1\}$), n, sequence $|\psi_i\rangle$ for $i = 1,\ldots,n$, $P^\sharp$ for $\sharp \in \{+,\times\}$
Ensure: Two strings of n bits $',\mathbf{b}' \in \{0,1\}^n$
  Generate randomly $b'_1,\ldots,b'_n$
  $\mathbf{b}' \leftarrow (b'_1,\ldots,b'_n) \in \{0,1\}^n$
  $i \leftarrow 1$
  repeat
    if $b'_i = 0$ **then**
      ask whether P^+ takes value 1
    else
      ask whether $P^\times$ takes value 1
    end if
    if Counter triggered **then**
      $a'_i \leftarrow 1$
    else
      $a'_i \leftarrow 0$
    end if
    $i \leftarrow i+1$
  until $i > n$
  $\mathbf{a}' \leftarrow (a'_1,\ldots,a'_n) \in \{0,1\}^n$
  Transmit string $\mathbf{b}' \in \{0,1\}^n$ to Alice via public classical channel

When Alice receives the string $\mathbf{b}$, she performs the conciliation algorithm de-

scribed below.

Algorithm 9.3.4 Conciliation

Require: Strings $\mathbf{b}, \mathbf{b}' \in \{0,1\}$
Ensure: Sequence $(k_1, \ldots, k_L)$ with some $L \leq n$ of positions of coinciding bits
  $\mathbf{c} \leftarrow \mathbf{b} \oplus \mathbf{b}'$
  $i \leftarrow 1$
  $k \leftarrow 1$
  repeat
    $k \leftarrow \min\{j : k \leq j \leq n \text{ such that } c_j = 0\}$
    if $k \leq n$ **then**
      $k_i \leftarrow k$
      $i \leftarrow i+1$
    end if
  until $k > n$
  $L \leftarrow i-1$
  transmit $(k_1, \ldots, k_L)$ to Bob via public classical channel

Theorem 9.3.5 *If there is no eavesdropping on the quantum channel then*

$$\mathbb{P}((a'_{k_1}, \ldots, a'_{k_L}) = (a_{k_1}, \ldots, a_{k_L}) | \boldsymbol{a}, \mathbf{b}) = 1.$$

Proof: Compute $\langle \psi_i | P^+ \psi_i \rangle$ and $\langle \psi_i | P^\times \psi_i \rangle$ for all different possible choices of $\psi_i \in B^+ \cup B^\times$. We observe that for those i's such that $b'_i = b_i$ we have $\mathbb{P}(a'_i = a_i) = 1$. Hence on deciding to consider only the substrings of $\mathbf{a}$ and $\mathbf{a}'$ defined on the locations where $\mathbf{b}$ and $\mathbf{b}'$ coincide, we have the certainty of sharing the same substrings, although $\mathbf{a}$ and $\mathbf{a}'$ have never been exchanged. □

Lemma 9.3.6 *If there is no eavesdropping, for N large enough, L is of the order* $2N$.

Proof: Elementary use of the law of large numbers. □

If Eve is eavesdropping, since she cannot copy quantum states (no-cloning theorem), she can measure with the same procedure as Bob and in order for the leakage not to be apparent, she re-emits a sequence of qubits $|\tilde{\psi}_i\rangle$ to Bob. Now again L is of the order $2N$ but since Eve's choice of the $\mathbf{b}$'s is independent of the choices of Alice and Bob, the string $\mathbf{a}'$ computed by Bob will coincide with Alice's string $\mathbf{a}$ at only $L/2 \simeq N$ positions.

Hence to securely communicate, Alice and Bob have to go through the eavesdropping detection procedure and reconciliation.

Bob randomly chooses half of the bits of the substring $(a'_{k_1},\ldots,a'_{k_L})$, i.e. $(a'_{r_1},\ldots,a'_{r_{L/2}})$ with $r_i \in \{k_1,\ldots,k_L\}$ and $r_i \neq r_j$ for $i \neq j$, and sends the randomly chosen positions $(r_1,\ldots,r_{L/2})$ and the corresponding bit values $(a'_{r_1},\ldots,a'_{r_{L/2}})$ to Alice. If $(a'_{k_1},\ldots,a'_{k_L}) = (a_{k_1},\ldots,a_{k_L})$ (reconciliation) then Alice announces this fact to Bob and they use the complementary substring of $(a'_{k_1},\ldots,a'_{k_L})$ (of length $L/2 \simeq N$) as their key to cipher with Vernam's algorithm. Else, they restart BB84 protocol.

Notice that Alice and Bob never exchanged the ultimate substring of N bits they use as key.

Chapter 10

Elements of quantum computing

In this chapter, B denotes the set $\{0,1\}$ and elements $b \in B$ are called *bits*; $\mathbb{H}$ will denote $\mathbb{C}^2$ and rays $|\psi\rangle \in \mathbb{H}$ are called *qubits*. Similarly, arrays of n bits are denoted by $\mathbf{b} = (b_1,\ldots,b_n) \in B^n$; arrays of n qubits by $|\psi\rangle = |\psi_1 \cdots \psi_n\rangle \in \mathbb{H}^{\otimes n}$.

10.1 Classical and quantum gates and circuits

A *classical circuit* implements a Boolean mapping $f : B^n \to B^n$ by using elementary gates of small arities[1], chosen from a family G; A *quantum circuit* implements a unitary mapping $U : \mathbb{H}^{\otimes n} \to \mathbb{H}^{\otimes n}$ by using unitary elementary gates of small arities[2], chosen from a family G.

Definition 10.1.1 Let $U : \mathbb{H}^{\otimes n} \to \mathbb{H}^{\otimes n}$ for some n and G be a fixed family of unitary operators of different arities. A *quantum circuit* over G is a product of operators from G acting on appropriate qubit entries.

It is usually assumed that G is closed under inversion.

Definition 10.1.2 Let $V : \mathbb{H}^{\otimes n} \to \mathbb{H}^{\otimes n}$ be a unitary operator. This operator is said to be *realised* by a unitary operator $W : \mathbb{H}^{\otimes N} \to \mathbb{H}^{\otimes N}$, with $N \geq n$ entries, acting on n qubits and $N-n$ ancillary qubits, if for all $|\xi\rangle \in \mathbb{H}^{\otimes n}$,

$$W(|\xi\rangle \otimes |0^{N-n}\rangle) = (V|\xi\rangle) \otimes |0^{N-n}\rangle.$$

[1] usually acting on $\mathcal{O}(1)$ bits.
[2] usually acting on $\mathcal{O}(1)$ qubits.

Ancillary qubits correspond to some memory in a fixed initial state we borrow for intermediate computations that is returned into the same state. Returning ancillary qubits into the same state can be relaxed. What cannot be relaxed is that ancilla must not be entangled with the n qubits (it must remain in tensor form); otherwise the anicllary subsystem could not be forgotten.

Quantum circuits are supposed to be more general than classical circuits. However, arbitrary Boolean circuits cannot be considered as classical counterparts of quantum ones because the classical analogue of a unitary operator on $\mathbb{H}^{\otimes n}$ is an invertible map on B^n, i.e. a permutation $\pi \in S_{2^n}$. Since to any n-bit array $\xi = (\xi_1 \cdots \xi_n) \in B^n$ corresponds a basis vector $|\,\xi\,\rangle = |\,\xi_1 \cdots \xi_n\,\rangle \in \mathbb{H}^{\otimes n}$, to every permutation $\pi \in S_{2^n}$ naturally corresponds a unitary operator $\hat{\pi}$, defined by

$$\hat{\pi}|\,\xi\,\rangle = |\,\pi(\xi)\,\rangle,$$

with $\hat{\pi}^* = \hat{\pi}^{-1} = \widehat{\pi^{-1}}$. Hence we can define:

Definition 10.1.3 Let $G \subseteq S_{2^n}$. A *reversible circuit* over G is a sequence of permutations from G.

An arbitrary Boolean function $F : B^m \to B^n$ can be extended to a function $F_\oplus : B^{m+n} \to B^{m+n}$, defined by

$$F_\oplus(x,y) = (x, y \oplus F(x)),$$

where the symbol $\oplus$ in the right hand side stands for the bit-wise addition modulo 2. It is easily checked that $F_\oplus$ is a permutation. Moreover $F_\oplus(x,0) = (x,F(x))$.

Notice that 2-bit permutation gates do not suffice the realise all functions of the form $F_\oplus$. On the contrary $G = \{\text{NOT}, \Lambda\}$ with $\Lambda : B^3 \to B^3$ the *Toffoli gate*, defined by $\Lambda(x,y,z) = (x,y,z \oplus (x \wedge y))$, is a basis.

10.2 Approximate realisation

There are uncountably many unitary operators $U : \mathbb{H}^{\otimes n} \to \mathbb{H}^{\otimes n}$. Hence if a quantum computer is to be constructed, the notion of exact realisation of a unitary operator must be weakened to an approximate realisation. The same *rationale* prevails also in classical computing, instead of all real functions (uncountably many), only Boolean functions are implemented.

Lemma 10.2.1 *An arbitrary unitary operator $U : \mathbb{C}^m \to \mathbb{C}^m$ can be represented*

as a product $V = \prod_{i=1}^{m(m-1)/2} V^{(i)}$ *of matrices of the form*

$$\begin{pmatrix} 1 & & & & & & & \\ & \ddots & & & & & & \\ & & 1 & & & & & \\ & & & \begin{pmatrix} a & b \\ c & d \end{pmatrix} & & & & \\ & & & & 1 & & & \\ & & & & & \ddots & & \\ & & & & & & 1 \end{pmatrix}, \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbf{U}(2).$$

Moreover, the sequence of matrices appearing in the product can be explicitly constructed in a running time $\mathcal{O}(m^3)\mathit{poly}(\log(1/\delta))$ *where* $\delta = \|U - V\|$.

Proof: An exercise if one recalls that for all $c_1, c_2 \in \mathbb{C}$, there exists a unitary operator $W \in \mathbf{U}(2)$ such that

$$W \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} \sqrt{|c_1|^2 + |c_2|^2} \\ 0 \end{pmatrix}.$$

□

Basic properties of the operator norm are recalled below:

$$\begin{aligned} \|XY\| &\leq \|X\|\|Y\| \\ \|X\| &= \|X\| \\ \|U\| &= 1 \\ \|X \otimes Y\| &= \|X\|\|Y\|, \end{aligned}$$

where X and Y are arbitrary operators and U is a unitary.

Definition 10.2.2 A unitary operator U' *approximates* a unitary operator U within δ if $\|U - U'\| \leq \delta$.

Lemma 10.2.3 *If a unitary* U' *approximates a unitary* U *within* δ*, then* U'^{-1} *approximates* U^{-1} *within* δ.

Proof: Since $U'^{-1}(U' - U)U^{-1} = U^{-1} - U'^{-1}$, it follows that $\|U^{-1} - U'^{-1}\| \leq \|U' - U\| \leq \delta$. □

Lemma 10.2.4 *If unitary operators* $(U'_k)_{k=1,\ldots,L}$ *approximate unitary operators* $(U_k)_{k=1,\ldots,L}$ *within* δ_k*, then* $U' = U'_L \cdots U'_1$ *approximates* $U = U_L \cdots U_1$ *within* $\sum_{k=1}^{L} \delta_k$.

Proof: $\|U_2'U_1' - U_2U_1\| \leq \|U_2'(U_1' - U_1) + (U_2' - U_2)U_1\| \leq \delta_1 + \delta_2$. □

Definition 10.2.5 A unitary operator $U : \mathbb{H}^{\otimes n} \to \mathbb{H}^{\otimes n}$ *is approximated* by a unitary operator $U : \mathbb{H}^{\otimes N} \to \mathbb{H}^{\otimes N}$, with $N \geq n$, within δ if for all $|\,\xi\,\rangle \in \mathbb{H}^{\otimes n}$

$$\|U'(|\,\xi\,\rangle \otimes |\,0^{N-n}\,\rangle) - U|\,\xi\,\rangle \otimes |\,0^{N-n}\,\rangle\| \leq \delta\|\xi\|.$$

Definition 10.2.6 For every unitary operator $U : \mathbb{H}^{\otimes n} \to \mathbb{H}^{\otimes n}$ there exists a unitary operator $C(U) : \mathbb{H} \otimes \mathbb{H}^{\otimes n} \to \mathbb{H} \otimes \mathbb{H}^{\otimes n}$, called the *controlled-U operator*, defined for all $|\,\xi\,\rangle \in \mathbb{H}^{\otimes n}$ by

$$C(U)|\,\varepsilon\,\rangle \otimes |\,\xi\,\rangle = \begin{cases} |\,\varepsilon\,\rangle \otimes |\,\xi\,\rangle & \text{if } \varepsilon = 0 \\ |\,\varepsilon\,\rangle \otimes U|\,\xi\,\rangle & \text{if } \varepsilon = 1 \end{cases}$$

Similarly, multiply controlled-U $C^k(U) : \mathbb{H}^{\otimes k} \otimes \mathbb{H}^{\otimes n} \to \mathbb{H}^{\otimes n} \otimes \mathbb{H}^{\otimes n}$, is defined by

$$C^k(U)|\,\varepsilon_1 \cdots \varepsilon_k\,\rangle \otimes |\,\xi\,\rangle = \begin{cases} |\,\varepsilon_1 \cdots \varepsilon_k\,\rangle \otimes |\,\xi\,\rangle & \text{if } \varepsilon_1 \cdots \varepsilon_k = 0 \\ |\,\varepsilon_1 \cdots \varepsilon_k\,\rangle \otimes U|\,\xi\,\rangle & \text{if } \varepsilon_1 \cdots \varepsilon_k = 1 \end{cases}$$

Example 10.2.7 Let $\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ be the unitary operator corresponding to the classical NOT gate. Then $C^2(\sigma_1) = \hat{\Lambda}$, where Λ is the Toffoli gate.

Definition 10.2.8 The set

$$G = \{H, K, K^{-1}, C(\sigma_1), C^2(\sigma_1)\},$$

with $H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ (Hadamard gate) and $K = \begin{pmatrix} 1 & 0 \\ 0 & -i \end{pmatrix}$ (phase gate), is called *the standard computational basis*.

Theorem 10.2.9 *Any unitary operator $U : \mathbb{H}^{\otimes n} \to \mathbb{H}^{\otimes n}$ can be approximated within δ by a $\mathsf{poly}(\log(1/\delta))$-size circuit over the standard basis using ancillary qubits. There is a $\mathsf{poly}(n)$-time algorithm describing the construction of the approximating circuit.*

Proof: An exercise, once you have solved the exercise 10.2.10 below. □

Exercise 10.2.10 Let $\sigma_{0,\ldots,3}$ be the 3 Pauli matrices augmented by the identity matrix, H the Hadamard gate, and $\Phi(\phi) = \begin{pmatrix} 1 & 0 \\ 0 & \exp(2i\phi) \end{pmatrix}$.

1. Show that if $A \in \mathscr{M}_2(\mathbb{C})$ with $A^2 = \mathbb{1}$ and $\phi \in \mathbb{R}$, then

$$\exp(i\phi A) = \cos\phi\,\sigma_0 + i\sin\phi A.$$

2. Let $R_j(\theta) = \exp(-i\frac{\theta}{2}\sigma_j)$, for $j = 1,2,3$ and $R_{\hat{n}}(\theta) = \exp(-i\frac{\theta}{2}\hat{n}\cdot\vec{\sigma})$, where $\hat{n} = (N_1, n_2, n_3)$ with $n_1^2 + n_2^2 + n_3^2 = 1$ and $\vec{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$. Express $R_j(\theta)$ and $R_{\hat{n}}(\theta)$ on the basis $\sigma_0, \ldots, \sigma_3$.

3. Show that $H = \exp(i\phi)R_1(\alpha)R_3(\beta)$, for some ϕ, α, β to be determined.

4. If $|\,\xi\,\rangle \in \mathbb{C}^2$ is a ray represented by a vector of the Bloch sphere $\mathbb{S}^2 = \{x \in \mathbb{R}^3 : \|x\|_2 = 1\}$, show that

$$R_{\hat{n}}(\theta)|\,\xi\,\rangle = |\,T_{\hat{n}}(\theta)x\,\rangle$$

   where $T_{\hat{n}}(\theta)x$ is the rotation of x around $\hat{n}$ by an angle θ.

5. Show that every $U \in \mathbf{U}(2)$ can be written as

$$U = \exp(i\alpha)R_{\hat{n}}(\theta)$$

   for some $\alpha, \theta \in \mathbb{R}$.

6. Show that every $U \in \mathbf{U}(2)$ can be written as

$$U = \exp(i\alpha)R_3(\beta)R_2(\gamma)R_3(\delta)$$

   for some $\alpha, \beta, \gamma, \delta \in \mathbb{R}$.

7. Suppose that $\hat{m}$ and $\hat{n}$ are two not parallel vectors of $\mathbb{S}^2$. Show that every $U \in \mathbf{U}(2)$ can be written as

$$U = \exp(i\alpha)R_{\hat{n}}(\beta_1)R_{\hat{m}}(\gamma_1)R_{\hat{n}}(\beta_2)R_{\hat{m}}(\gamma_2)\cdots.$$

8. Establish identities

$$\begin{aligned} H\sigma_1 H &= \sigma_3 \\ H\sigma_2 H &= -\sigma_2 \\ H\sigma_3 H &= \sigma_1 \\ H\Phi(\frac{\pi}{8})H &= \exp(i\alpha)R_1(\frac{\pi}{4}) \end{aligned}$$

   for some α.

10.3 Examples of quantum gates

10.3.1 The Hadamard gate

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

$$H|\varepsilon\rangle = \frac{1}{\sqrt{2}}((-1)^{\varepsilon}|\varepsilon\rangle + |1-\varepsilon\rangle), \varepsilon \in B.$$

$$H^{\otimes 3}|000\rangle = \frac{1}{\sqrt{8}} \sum_{x=0}^{7} |x\rangle.$$

10.3.2 The phase gate

$$\Phi(\phi) = \begin{pmatrix} 1 & 0 \\ 0 & \exp(2i\phi) \end{pmatrix}.$$

$$\Phi(\phi)|\varepsilon\rangle = \exp(2i\varepsilon\phi)|\varepsilon\rangle$$

$$\Phi(\frac{\pi}{4} + \frac{\phi}{2})H\Phi(\theta)H|0\rangle = \cos\theta|0\rangle + \exp(i\phi)\sin\theta|1\rangle.$$

10.3.3 Controlled-NOT gate

$$C(\sigma_1) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

For any $x \in B$, $C(\sigma_1)|x0\rangle = |xx\rangle$, but for arbitrary $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$,

$$C(\sigma_1)|\psi 0\rangle = \alpha|00\rangle + \beta|11\rangle \neq |\psi\psi\rangle.$$

10.3.4 Controlled-phase gate

$$C(\Phi(\phi)) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \exp(2i\phi) \end{pmatrix}.$$

For $x, y \in B$,

$$C(\Phi(\phi))|xy\rangle = \exp(2i\phi xy)|xy\rangle.$$

10.3.5 The quantum Toffoli gate

For all $x, y, z \in B$,

$$C^2(\sigma_3)|xyz\rangle = |x, y, (x \wedge y) \oplus z\rangle.$$

Suppose that $f : B^m \to B^n$ is a Boolean function, implemented by the unitary operator $U_f : \mathbb{H}^{\otimes(n+m)} \to \mathbb{H}^{\otimes(n+m)}$. If $|\psi\rangle = \frac{1}{2^{m/2}} \sum_{\varepsilon_1,\ldots,\varepsilon_m \in B} |\varepsilon_1, \ldots, \varepsilon_m\rangle$ then

$$U_f|\psi\rangle \otimes |0^n\rangle = \frac{1}{2^{m/2}} \sum_{x=0}^{2^m-1} |x, f(x)\rangle.$$

Hence computing *simultaneously* all values of f over its domain of definition requires the same computational effort as computing the value over a singleton of the domain.

Chapter 11

The Shor's factoring algorithm

Bibliography

[1] William Arveson. *A short course on spectral theory*, volume 209 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2002. 25, 28, 35

[2] C. H. Bennett and G. Brassard. Quantum public key distribution system. *IBM Technical disclosure bulletin*, 28:3153–3163, 1985. 93

[3] Richard V. Kadison and John R. Ringrose. *Fundamentals of the theory of operator algebras. Vol. I*, volume 15 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 1997. Elementary theory, Reprint of the 1983 original. 23

[4] A. K. Lenstra and H. W. Lenstra, Jr. Algorithms in number theory. In *Handbook of theoretical computer science, Vol. A*, pages 673–715. Elsevier, Amsterdam, 1990. 3, 92

[5] Arjen K. Lenstra. Integer factoring. *Des. Codes Cryptogr.*, 19(2-3):101–128, 2000. Towards a quarter-century of public key cryptography. 92

[6] L. H. Loomis. The lattice theoretic background of the dimension theory of operator algebras. *Mem. Amer. Math. Soc.*, 1955(18):36, 1955. 47

[7] R. L. Rivest, A. Shamir, and L. Adleman. A method for obtaining digital signatures and public-key cryptosystems. *Comm. ACM*, 21(2):120–126, 1978. 91

[8] Walter Rudin. *Functional analysis*. International Series in Pure and Applied Mathematics. McGraw-Hill Inc., New York, second edition, 1991. 25

[9] Claude Shannon. Communication theory of secrecy systems. *Bell System Technical Journal*, 28:656–715, 1949. 90

[10] A. N. Shiryayev. *Probability*, volume 95 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1984. Translated from the Russian by R. P. Boas. 8

[11] Peter W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM J. Comput.*, 26(5):1484–1509, 1997. 3, 89

[12] V. S. Varadarajan. *Geometry of quantum theory. Vol. I*. D. Van Nostrand Co., Inc., Princeton, N.J.-Toronto, Ont.-London, 1968. The University Series in Higher Mathematics. 65

[13] V. S. Varadarajan. *Geometry of quantum theory*. Springer-Verlag, New York, second edition, 1985. 12

[14] V. S. Varadarajan. *Geometry of quantum theory*. Springer-Verlag, New York, second edition, 1985. 47

[15] Gilbert S. Vernam. *Cipher printing telegraph systems for secret wire and radio telegraphic communications*, volume 55. 1926. 90

[16] John von Neumann. *Mathematical foundations of quantum mechanics*. Princeton Landmarks in Mathematics. Princeton University Press, Princeton, NJ, 1996. Translated from the German and with a preface by Robert T. Beyer, Twelfth printing, Princeton Paperbacks. 12