Quantum Information Theory

Lecture Notes for Math 992 Spring 2022

Jeffrey Schenker[1]

Michigan State University
Spring 2022
schenke6@msu.edu

[1]©2022 Jeffrey Schenker

This work is licensed under a
Creative Commons Attribution-NonCommercial 4.0 International License

Contents

Comments and course information **vii**

I An introduction to quantum theory **1**

1 Classical Physics **3**
1.1 Classical mechanics 4
1.2 The algebra of observables in classical physics 5
1.3 States of classical physics 6

2 From classical to quantum **9**
2.1 An alternative picture of classical motion 9
2.2 Developing the ideas of quantum theory 10
2.3 States and observables in quantum mechanics 12
2.4 A finite-dimensional matrix model of quantum mechanics 14

3 Finite dimensional models of quantum mechanics **17**
3.1 Poisson bracket and dynamics for the matrix model 17
3.2 The spectral theorem for self-adjoint matrices 18
3.3 Dynamics on the matrix model 20
3.4 Heisenberg uncertainty relations 21

4 Example systems: a Qubit and an infinite dimensional model of quantum mechanics **23**
4.1 A Qubit — an example 23
4.2 An Infinite Dimensional Model of Quantum Mechanics 26

5 Trace Class Operators **29**
5.1 Compact Operators, Positive Operators, and the Polar Decomposition 29
5.2 Trace Class Operators 33

6 States on the infinite dimensional model of quantum mechanics **35**
6.1 Defining the states 35

7 Dynamics and undbounded operators **39**
7.1 Unbounded Operators . . . 39
7.2 Dynamics and Stone's Theorem . . . 40
7.3 The Weyl form of the CCR and the Stone-von Neumann Theorm . . . 42

8 Examples **45**
8.1 Spectral representation of the momentum and position operators . . . 45
8.2 The Free Hamiltonian . . . 46
8.3 Particle in a potential . . . 47
8.4 The Harmonic Oscillator . . . 48
8.5 Heisenberg evolution of a harmonic oscillator . . . 50
8.6 Coherent states . . . 51

II Composite and Open Quantum Systems **55**

9 Modeling composite systems **57**
9.1 Composite systems in classical mechanics . . . 57
9.2 Tensor Products . . . 58

10 Composite systems in quantum mechanics **63**
10.1 Tensor product of observables . . . 63
10.2 Multiple tensor products . . . 64
10.3 Composite systems in quantum mechanics . . . 65

11 Measurements, entanglement, the EPR argument and Bell's inequality **67**
11.1 Measurement . . . 67
11.2 Measurements of a qubit . . . 68
11.3 Measurements in composite systems . . . 69
11.4 Bell's inequality . . . 70

12 Postulates of quantum mechanics and properties of density matrices **73**
12.1 Postulates of Quantum Mechanics . . . 73
12.2 Interpreting mixed states . . . 74
12.3 Schmidt decomposition of pure states . . . 74
12.4 Purification of mixed states . . . 74

13 Open Quantum Systems **75**
13.1 Markov processes . . . 77
13.2 Operator-Sum Representation of an Open Quantum System . . . 78

14 Open Quantum Systems II **81**
14.1 Interpreting the operator sum representation . . . 81
14.2 Completely Positive Maps . . . 83

15 Non uniqueness of the operator sum representation and Examples of Quantum Operations **87**
15.1 Non uniqueness of the operator sum representation 87

16 Quantum Markov Semigroups **89**
16.1 The true quantum evolution of an open system 89
16.2 An example 90
16.3 Heuristic derivation of the Markov semigroup 92

17 Quantum Markov Semigroup for a Qubit **99**

18 Lindblad's Theorem **103**

19 Quantum Trajectories: unraveling the Lindblad dynamics **111**
19.1 Background: a bit about classical Markov jump processes 111

20 Quantum Trajectories: unraveling the Lindblad dynamics (II) **115**
20.1 An example 119

III Quantum Information Theory **121**

21 A bit of classical information theory **123**
21.1 Shannon Entropy 123
21.2 Shannon's Noiseless Coding Theorem 124
21.3 Relative and Conditional Entropies and Mutual Information 127

22 The von Neumann Entropy **131**
22.1 Relative Entropy 131
22.2 Projective measurements increase entropy 132
22.3 Entropies in Composite Systems 133
22.4 Concavity of the entropy and the entropy of mixtures 134

23 Strong Subadditivity of the Entropy **137**

24 Quantum Source Coding **139**

A Hilbert Spaces **143**
1.1 The Riesz Representation Theorem 144

Bibliography **147**

Comments and course information

These are lecture notes for Quantum Information Theory (special topics in analysis, MTH 992) in Spring 2022. The material will be taken from several texts:

- L. D. Faddeev and O. A. Yakubovskiĭ. *Lectures on Quantum Mechanics for Mathematics Students*. English ed. Student Mathematical Library v. 47. Providence, R.I: American Mathematical Society, 2009. 234 pp.
- M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. 10th anniversary ed. Cambridge ; New York: Cambridge University Press, 2010. 676 pp.
- J. Watrous. *The Theory of Quantum Information*. Cambridge, United Kingdom: Cambridge University Press, 2018. 590 pp.

  Other useful references include:

- References on quantum information/quantum computing:
  - A. Kitaev, A. Shen, and M. Vyalyi. *Classical and Quantum Computation*. Vol. 47. Graduate Studies in Mathematics. Providence, Rhode Island: American Mathematical Society, May 31, 2002
  - J. Preskill. *Physics 219 Course Information*. URL: `http://theory.caltech.edu/~preskill/ph229/` (visited on 01/08/2022)
- For background on quantum theory, I highly recommend reading the classic text of Dirac:
  - P. A. M. Dirac. *The Principles of Quantum Mechanics*. 4. ed. (rev.), repr. International Series of Monographs on Physics 27. Oxford: Clarendon Press, Oxford University Press, 2010. 314 pp.
- For background on quantum theory aimed at a mathematical audience, the following are also useful references:
  - B. C. Hall. *Quantum Theory for Mathematicians*. Graduate Texts in Mathematics 267. New York: Springer, 2013. 554 pp.
  - G. Teschl. *Mathematical Methods in Quantum Mechanics: With Applications to Schrödinger Operators*. Vol. 99. American Mathematical Society, 2009

Part I

An introduction to quantum theory

Lecture 1

Classical Physics

Classical mechanics — developed starting with Newton's Principia and through the work of Euler, Laplace, Lagrange, Hamilton and others in the 18th and 19th centuries — provided a remarkable framework for the understanding of the motion of objects in space. It is truly remarkable that a theory developed by Newton in the 17th century is accurate enough that, with the advent of 20th century computational tools, we can predict the motion of the planets well enough to send spacecraft to Mars. Of course, the computers used to make those predictions rely on solid state technology which required 20th century science (quantum mechanics) to be understood and developed. And it is worth noting that NASA still employed *human computers*, armed with slide rules, pencil and paper, during the Apollo missions.

The 19th century brought an understanding of electricity and magentism, developed by Faraday, Gauss, Ampere and others, and thermodynamics, developed by Carnot, Claussius, Thomson and others. Ultimately, Maxwell showed that electricity and magnetism were two aspects of one *field* that propagated through space according to rules still governed by classical mechanical theory; and the work of Maxwell, Boltzmann and Gibbs showed that thermodynamics could be explained by applying statistical arguments to systems of mechanical particles.

Still there were things that classical mechanics could not correctly describe, namely the discrete nature of atomic "spectra," and more generally the stability of atoms and molecules. Why don't negatively charged electrons crash into the positively charged nucleus as predicted by classical electricity and magnetism? The Maxwell-Boltzmann statistical theory makes incorrect predictions about the spectrum of radiation from a "black body," i.e., a perfectly absorbent body emitting electromagnetic radiation in thermal equilibrium.

Explaining these discrepancies ultimately led to a complete revolution in physics and how we understand the nature of the universe. This was the quantum theory developed in the 1920s by Bohr, Heisenberg, Pauli, Schrödinger and others, building on earlier ideas of Planck and Einstein. To explain the quantum theory we should first revisit quickly what classical mechanics is and how classical physics described the world.

1.1 Classical mechanics

Consider a "point like" particle in space. "Point like" means that the particle is so small that we can neglect its structure and imagine it concentrated at a single point. (Depending on the scales being considered, a planet might be "point like".) In classical physics the "state" of such a particle is specified by six numbers: it's *position* $\boldsymbol{x} = (x_1, x_2, x_3)$ and *velocity* $\boldsymbol{v} = (v_1, v_2, v_3)$. More precisely this is a point in a tangent bundle TM, where $\boldsymbol{x} \in M$, the manifold of points in space, and $\boldsymbol{v}$ is a tangent vector at $\boldsymbol{x}$. Typically $M = \mathbb{R}^3$, but one can consider classical mechanics on other manifolds — and this is useful, e.g., to describe the rotating bodies. The manifold M is typically called the *configuration* space of the system.

Newton's second law asserts that the particle moves according to the equation

$$\boldsymbol{v} \;=\; \frac{d\boldsymbol{x}}{dt} \quad \text{and} \quad \boldsymbol{F} \;=\; m\frac{d\boldsymbol{v}}{dt} \tag{1.1}$$

where t is time, m is mass of the particle, and $\boldsymbol{F}$ is the force on the particle at time t. In more detail, we have a system of coupled differential equations

$$\frac{d\boldsymbol{x}(t)}{dt} \;=\; \boldsymbol{v}(t) \quad \text{and} \quad m\frac{d\boldsymbol{v}(t)}{dt} \;=\; \boldsymbol{F}(\boldsymbol{x}(t), \boldsymbol{v}(t), t), \tag{1.2}$$

where we have imagined that the force could depend instantaneously on the position and velocity of the particle and might explicitly change in time.

The typical situation is to consider a force that is a function only of the particle position (although velocity dependent forces are used to provide a phenomenological description of friction), and is also *conservative* meaning that

$$\boldsymbol{F}(\boldsymbol{x}) \;=\; -\frac{\partial V}{\partial \boldsymbol{x}}(\boldsymbol{x}) \tag{1.3}$$

where V is a function on configuration space, called the potential (or potential energy), and $\frac{\partial V}{\partial \boldsymbol{x}} = (\frac{\partial V}{\partial x_1}, \frac{\partial V}{\partial x_2}, \frac{\partial V}{\partial x_3})$. When this holds the equations of motion read

$$\frac{d\boldsymbol{x}}{dt} \;=\; \boldsymbol{v} \quad \text{and} \quad m\frac{d\boldsymbol{v}}{dt} \;=\; -\frac{\partial V}{\partial \boldsymbol{x}}\,.$$

One introduces a new variable $\boldsymbol{p} = m\boldsymbol{v}$, the *momentum*, and writes the equations as follows

$$\frac{d\boldsymbol{x}}{dt} \;=\; \frac{\boldsymbol{p}}{m} \quad \text{and} \quad \frac{d\boldsymbol{p}}{dt} \;=\; -\frac{\partial V}{\partial \boldsymbol{x}}\,,$$

which can be expressed as follows

$$\frac{d\boldsymbol{x}}{dt} \;=\; \frac{\partial H}{\partial \boldsymbol{p}} \quad \text{and} \quad \frac{d\boldsymbol{p}}{dt} \;=\; -\frac{\partial H}{\partial \boldsymbol{x}}\,,$$

where $H(\boldsymbol{x}, \boldsymbol{p}) = \frac{1}{2m}\boldsymbol{p}^2 + V(\boldsymbol{x})$. The minus sign in (1.3) is there to produce a positive sign here, so that H is positive if V is positive.

All of the above can be generalized to the following situation. Consider a Riemannian manifold M (typically $M = \mathbb{R}^{3n}$) and a function H defined on the *cotangent bundle*, called the Hamiltonian. In coordinates $H = H(q_1, \dots, q_m; p_1, \dots, p_m)$ where $(q_1, \dots, q_m)$ are coordinates for a point in space M and $(p_1, \dots, p_m)$ is a co-vector at $(q_1, \dots, q_m)$. The cotangent bundle T^*M (typically just $\mathbb{R}^{3n} \times \mathbb{R}^{3n}$) is called *phase space*. The equations of motion

$$\frac{dq_i}{dt} = \frac{\partial H}{\partial p_i} \quad \text{and} \quad \frac{p_i}{dt} = -\frac{\partial H}{\partial q_i}$$

are called *Hamilton's equations* and describe a large class of physical systems. One verifies easily that along the trajectory of such a system one has

$$\frac{dH}{dt} = \sum_{i=1}^{m} \frac{\partial H}{\partial q_i}\frac{dq_i}{dt} + \frac{\partial H}{\partial p_i}\frac{dp_i}{dt} = 0 .$$

This is the principle of conservation of energy, where we identify the Hamiltonian function at a point in phase space as the total energy of the system with that configuration.

In a typical example we have n points in $\mathbb{R}^3$ with Hamiltonian

$$H = \sum_{i=1}^{n} \frac{\boldsymbol{p}_i^2}{2m_i} + \sum_{i<j}^{n} V_{ij}(\boldsymbol{x}_i - \boldsymbol{x}_j) + \sum_{i=1}^{n} V_i(\boldsymbol{x}_i)$$

where

- V_{ij} is the potential energy of interaction between the i^{th} and j^{th} particles, and
- V_i is an *external potential* acting on the i^{th} particle.

This system could be used, for example, to describe the motion of the planets and their moons in the solar system. If we make the approximation that the sun is a fixed point, then it creates an "external potential" affecting all of the planets and moons.

1.2 The algebra of observables in classical physics

An *observable* of a classical system is just a function on phase space. Let $\mathfrak{A}$ denote the algebra of real valued infinitely differentiable functions on phase space. Given $f \in \mathfrak{A}$, let $f_t = f(\boldsymbol{q}_t, \boldsymbol{p}_t)$ along a trajectory of Hamilton's equations. Then

$$\frac{df_t}{dt} = \sum_{i=1}^{m} \frac{\partial H}{\partial p_i}\frac{\partial f}{\partial q_i} - \frac{\partial H}{\partial q_i}\frac{\partial f}{\partial p_i} .$$

This motivates the definition of the *Poisson Bracket*:

$$\{f, g\} = \sum_{i=1}^{m} \frac{\partial f}{\partial p_i}\frac{\partial g}{\partial q_i} - \frac{\partial f}{\partial q_i}\frac{\partial g}{\partial p_i} .$$

Then we have

$$\frac{df_t}{dt} = \{H, f\} . \tag{1.4}$$

The Poisson Bracket has several basic properties

1. $\{f, g + \lambda h\} = \{f, g\} + \lambda\{f, h\}$ (linearity),

2. $\{f, g\} = -\{g, f\}$ (skew symmetry);

3. $\{f, \{g, h\}\} + \{g, \{h, f\}\} + \{h, \{f, g\}\} = 0$ (Jacobi identity);

4. $\{f, gh\} = g\{f, h\} + \{f, g\}h$ (derivation).

Put succinctly, the algebra $\mathfrak{A}$ of observables is a real Lie algebra with respect to the bracket $\{\cdot, \cdot\}$ and also an algebra with respect to the product of functions, so that the bracket is a derivation.

The Hamiltonian H is an element of $\mathfrak{A}$ which generates dynamics. We can "forget" about phase space and consider the dynamics just on the algebra of observables, using the equation (1.4). We can define a mapping $U_t : \mathfrak{A} \to \mathfrak{A}$ be solving these equations. Namely:

$$U_t(f) = f_t \quad \text{where} \quad \frac{df_t}{dt} = \{H, f_t\} \quad \text{and} \quad f_0 = f \ .$$

The operation U_t is linear and preserves the algebraic structure of $\mathfrak{A}$:

- $U_t(f + \lambda g) = U_t(f) + \lambda U_t(g)$;
- $U_t(fg) = U_t(f)U_t(g)$; and
- $U_t(\{f, g\}) = \{U_t(f), U_t(g)\}$.

1.3 States of classical physics

In an experiment, we can measure a property of a physical system. This amounts to determining, with some precision, the value of one or more physical observables. The *state* of the system describes the possible outcomes of such experiments. Experience dictates that every experiment is subject to some uncertainty; when repeated under what appear to us to be identical conditions the results are slightly different. Good experiments have lower uncertainty, but there are always small fluctuations. One can imagine two possibilities for describing the uncertainty:

1. In principle, if we were just clever enough and had enough time and resources, we could make the experiment as precise as we would like.

2. The physics of the system we are studying preclude making completely precise measurements.

Classical physics tacitly held, without any clear justification, that (1) holds. Heisenberg's great observation is that quantum physics asserts that (2) is correct. (Even in a system for which (2) holds, (1) plays a role. Errors and lack of knowledge on our part can increase the uncertainty.)

Before moving to quantum physics we should consider the states of a classical system. The above description suggests that the *state* of a system should take an observable $f \in \mathfrak{A}$

and return a probability distribution for the possible values of f in an experiment. Thus a state ω is a map $\omega : \mathfrak{A} \to \mathcal{P}(\mathbb{R})$, the set of Borel probability measures on $\mathbb{R}$. We denote the value of ω at f by ω_f. Given a Borel set $E \subset \mathbb{R}$, $\omega_f(E)$ is the probability that a measurement of f yields a value in E. We expect the map ω specifying a state to have certain properties:

1. $\omega_{\lambda f}(E) \;=\; \omega_f(\frac{E}{\lambda})$ (change of scale or units)

2. More generally if $\phi : \mathbb{R} \to \mathbb{R}$ is any smooth map then, $\omega_{\phi \circ f}(E) = \omega_f(\phi^{-1}(E))$.

3. If $f \equiv C$ is the constant function, equal to C, then $\omega_f = \delta_C$, the delta measure at C.

4. If ω_f is supported on $\mathrm{ran} f = f(M)$. In particular, if $f \geq 0$ then ω_f is supported on $[0, \infty)$.

Given any two states, specified by maps ω^0 and ω^1, we can form the convex combination ω^λ for any $0 \leq \lambda \leq 1$:

$$\omega_f^\lambda(E) \;=\; (1-\lambda)\omega_f^0(E) + \lambda\omega_f^1(E)\,. \tag{1.5}$$

It is clear that the function on the right hand side produces a probability measure on $\mathbb{R}$. This seemingly *ad-hoc* construction has a natural interpretation in terms of possibility (1) above. Suppose that the systems we measure are produced by some random process that produces a system with state ω^0 with probability $(1-\lambda)$ and a system with state ω^1 with probability λ. Then, to us as experimenters, the systems are seen to have state ω^λ, which is termed a *mixed state*.

Thus the set of all states has a *convex structure*. Among all the states are those that cannot be decomposed as in (1.5) with $\omega^0 \neq \omega^1$. Such states are called *pure* states. All other states are *mixed states*.

Note that we made a subtle shift from considering the state of *a* system to considering the state of *an ensemble* of systems above. This is natural when we consider probabilistic statements. It is hard to pin down what we mean when we discuss probabilities for the outcome of a *single* experiment. To give an operational definition to probabilities we need to be able to talk about the frequency of outcomes in repeated experiments that are independent but "identical."

Associated to any observable, we can compute the *mean value* in state ω:

$$\langle f|\omega\rangle \;=\; \int_{\mathbb{R}} \lambda d\omega_f(\lambda)\,.$$

(Let us consider bounded observables, so the integral is guaranteed to converge for any ω.) This *dual pairing* between $\mathfrak{A}$ and the set of states satisfies

$$\langle f|\omega\rangle \;\geq\; 0 \quad \text{if } f \geq 0$$

and

$$|\langle f|\omega\rangle| \;\leq\; \sup_x |f(x)|\,.$$

If we also require the pairing to be linear (as seems natural):

$$\langle f + \lambda g|\omega\rangle \;=\; \langle f|\omega\rangle + \lambda\,\langle g|\omega\rangle \;,$$

then the Riesz-Markov theorem tells us that there is, associated to ω, a unique Borel probability measure μ_ω on phase space $\mathcal{M}$ such that

$$\langle f|\omega\rangle \;=\; \int_{\mathcal{M}} f(q,p)d\mu_\omega(q,p)\;.$$

Thus the states of a classical system are probability distributions on phase space. We will use distributional notation henceforth, writing $d\mu_\omega = \rho(q,p)dqdp$, and identify the state ω with the density (generalized) function. It is possible to show that the *pure states* are precisely the Dirac measures at a specific point of phase space:

$$\rho^{q_0,p_0}(p,q) \;=\; \delta(q-q_0)\delta(p-p_0),$$

with δ the Dirac delta "function." In this state, every observable has a precise value: $f \in \mathfrak{A}$ takes the values $f(q_0,p_0)$ with probability one. Thus classical physics falls solidly into category (1) as a theory. It posits that if we could somehow reduce all uncertainty as much as possible then we could pin the value of every observable down precisely.

In practice, mixed states play a very important role, in particular in the statistical theory of thermodynamics. It is observed that the relevant properties of a mechanical system with many many degrees of freedom are much more efficiently described by a mixed state, such as the *Gibb's ensemble*

$$\rho(q,p) \;=\; \frac{1}{Z}e^{-\beta H(q,p)}$$

where $\beta = \frac{1}{k_B T}$, with T the temperature and k_B = Boltzmann's constant. Even if the system is in principle in a specific pure state, specified by pinning down the positions and momenta of 10^{23} different particles, the properties we care about (total energy, density, pressure) are given by those of the Gibb's state, for the overwhelming majority of pure states.

Lecture 2

From classical to quantum

2.1 An alternative picture of classical motion

Consider the evolution $df_t = \{H, f_t\}dt$ on $\mathfrak{A}$, paired with a pure state ρ^{q_0,p_0}. One easily verifies that

$$\langle f_t | \rho^{q_0,p_0} \rangle = f(q_t, p_t)$$

where (q_t, p_t) solve Hamilton's equations with q_0 and p_0 the given point defining the pure state. Let $\Phi_t : \mathcal{M} \to \mathcal{M}$ denote the map of phase space sending (q_0, p_0) to the point on the corresponding trajectory at time t. This map is a diffeomorphism for each t. Then, for a mixed state with density ρ, we have $f_t(q,p) = f(\Phi_t(q,p))$ and

$$\int_{\mathcal{M}} f(\Phi_t(q,p))\rho(q,p)dqdp = \int_{\mathcal{M}} f(q,p)\rho(\Phi_t^{-1}(q,p))|\mathrm{Jac}(\Phi_t^{-1})(q,p)|dqdp\,,$$

where

$$\mathrm{Jac}(\Phi_t^{-1})(p,q) = \det\begin{bmatrix} \frac{\partial \Phi_t^{-1}}{\partial q} \\ \frac{\partial \Phi_t^{-1}}{\partial p} \end{bmatrix}$$

is the Jacobian determinant giving the infinitesimal change of volume around a point in phase space under Φ_t. It is a theorem, due to Liouville, that this determinant is identically one. That is that the flow Φ_t is volume preserving. Thus we have

$$\int_{\mathcal{M}} f(\Phi_t(q,p))\rho(q,p)dqdp = \int_{\mathcal{M}} f(q,p)\rho(\Phi_t^{-1}(q,p))dqdp\,.$$

The function $\rho_t = \rho \circ \Phi_t^{-1}$ satisfies

$$\frac{d\rho_t}{dt} = -\{H, \rho_t\}\,, \tag{2.1}$$

which is the same as (1.4) but with a minus sign. Eq. (2.1) is called *Liouville's equation*. It describes the evolution of a mixed state on phase space.

The proof of Liouville's theorem is very straightforward. Let $\mathrm{Jac}_t = \mathrm{Jac}(\Phi_t(p,q))$. Since $\Phi_0(q,p) = (q,p)$, we have $\mathrm{Jac}_0 = 1$ and it suffices to show that $\frac{d\mathrm{Jac}_t}{dt} = 0$. Since $t \mapsto \Phi_t$ is a

one-parameter group, i.e., $\Phi_{t+h} = \Phi_h \circ \Phi_t$, we have

$$\mathrm{Jac}_{t+h} \;=\; (\mathrm{Jac}_h \circ \Phi_t)\mathrm{Jac}_t$$

by the chain rule. Thus

$$\frac{d\mathrm{Jac}_t}{dt} \;=\; \left(\left.\frac{d\mathrm{Jac}_s}{ds}\right|_{s=0} \circ \Phi_t \right) \mathrm{Jac}_t \; .$$

Thus it suffices to prove $\frac{d\mathrm{Jac}_t}{dt} = 0$ at $t = 0$. Now for $t \to 0$ we have

$$\Phi_t(q,p) \;=\; (q,p) + t\left(\frac{\partial H}{\partial p}, -\frac{\partial H}{\partial q} \right) + O(t^2) \; .$$

Thus

$$\mathrm{Jac}_t \;=\; \det\left(I + t \begin{bmatrix} \frac{\partial^2 H}{\partial q \partial p} & -\frac{\partial^2 H}{\partial q^2} \\ \frac{\partial^2 H}{\partial p^2} & -\frac{\partial^2 H}{\partial p \partial q} \end{bmatrix} + O(t^2) \right) \; .$$

Thus

$$\left.\frac{d\mathrm{Jac}_t}{dt}\right|_{t=0} \;=\; \mathrm{tr} \begin{bmatrix} \frac{\partial^2 H}{\partial q \partial p} & -\frac{\partial^2 H}{\partial q^2} \\ \frac{\partial^2 H}{\partial p^2} & -\frac{\partial^2 H}{\partial q \partial q} \end{bmatrix} \;=\; \sum_{i=1}^{m} \frac{\partial^2 H}{\partial q_i \partial p_i} - \frac{\partial^2 H}{\partial p_i \partial q_i} \;=\; 0 \; ,$$

by equality of mixed second order partials.

So we have two pictures of the evolution in classical physics:

1. We can consider the *state* of the system ρ to be fixed and evolve observables according $\frac{df}{dt} = \{H, f\}$;
2. We can consider the observables to be fixed and evolve states according to $\frac{d\rho_t}{dt} = -\{H, \rho_t\}$.

We have the equivalence between the two pictures, because the dual pairing satisfies

$$\langle f_t | \rho \rangle \;=\; \langle f | \rho_t \rangle \; .$$

This structure will be preserved in quantum mechanics, leading to the Heisenberg and Schrödinger pictures respectively.

2.2 Developing the ideas of quantum theory

The first "quantum hypothesis" came in the work of Planck 1901 on the electromagnetic radiation emitted by a *black body*, a hypothetical perfectly absorbing object emitting radiation in thermal equilibrium at temperature $T > 0$. Such an object will emit radiation at all possible frequencies ν in varying density. Classical statistical mechanics together with the Maxwell equations for the electro-magnetic field predicts the *Rayleigh-Jeans law*

$$B(\nu, T) \;=\; \frac{8\pi\nu^2 k_B T}{c^2}$$

for the power emitted per unit frequency per area of the surface of the black body (here c is the speed of light and $k_B \approx 1.38 \times 10^{-23} \mathrm{J\,K}^{-1}$ is Boltzmann's constant). This expression leads to the nonsensical prediction that a black body emits an infinite amount of energy (since $\int_{\nu_0}^{\infty} \nu^2 d\nu = \infty$). This was called the ultraviolet catastrophe and was a major problem in physics during the late 19th century. Planck solved this problem by proposing that electromagnetic radiation came in discrete units with an energy

$$E \;=\; h\nu,$$

where h is a certain fundamental constant of nature, now called *Planck's constant*. Because the field could not have arbitrarily small increments of energy at high frequencies, Planck was able to derive the following formula for the powerper frequency per area

$$B(\nu, T) \;=\; \frac{8\pi h\nu^3}{c^2} \frac{1}{e^{\frac{h\nu}{k_B T}} - 1}\,.$$

For small ν, this expression reproduces the Rayleigh-Jeans law, but for large ν the spectrum is exponentially damped, leading to a sensible prediction for the total emission. Furthermore Planck's expression was found to be in agreement with the actual spectrum of hot objects. Planck's constant is experimentally found to be $h \approx 6.626 \times 10^{-34} \mathrm{J\,s}$. The constant $\hbar = \frac{h}{2\pi}$ appears in many formulas and is called "h-bar." (Warning: Faddeev and Yakubovskiĭ 2009 use h for $\hbar$!)

Einstein 1905 took Planck's idea and applied it to the *photo-electric effect*, a phenomenon by which electrons are ejected from a metal that is irradiated by a beam of light. Einstein suggested that there is a certain amount of work W required eject an electron from a metal and that the total kinetic energy T of electrons emitted should obey the relation

$$T + W \;=\; h\nu$$

where ν is the frequency of the radiating light. In particular, if $h\nu \leq W$, then no electrons should be emitted. Furthermore, the kinetic energy of the emitted electrons should *not* depend on the intensity of the radiation. These predictions turn out to be correct, and Einstein's theory was validated by work of Millikan 1914.

The quanta of light proposed by Einstein and Planck became known as *photons*. If photons carry a discrete unit of energy, they should carry a a discrete amount of momentum too. Einstein proposed that the momentum of a photon should be $\boldsymbol{p} = \hbar \boldsymbol{k}$, where $\boldsymbol{k}$ is the *wave vector*. That is $\boldsymbol{k}$ points in the direction of propagation and has magnitude $|\boldsymbol{k}| = \frac{2\pi}{\lambda}$ where λ is the wave length. The wave vector and frequency are related by $2\pi\nu = c|\boldsymbol{k}|$.

Through the early twentieth century, various experiments uncovered discrete phenomena in the behavior of matter as well. Some of these are discussed briefly in Faddeev Yakubovskii. The culmination of this was the realization that particles of matter, such as electrons or protons or neutrons (only discovered in 1931), have a wave like character. De Broglie suggested in 1924 that a free particle with momentum $\boldsymbol{p}$ and energy E should (in some sense that wasn't quite clear) be associated to a wave with frequency ν and wave vector $\boldsymbol{k}$ satisfying

$$E = h\nu \quad \text{and} \quad \boldsymbol{p} = \hbar \boldsymbol{k}\,.$$

Ultimately, Davisson and Germer 1928 demonstrated in 1928 diffraction of electrons, proving experimentally that there was a wave like character to the motion of matter.

At around the same time, Heisenberg 1927 proposed that their was a fundamental uncertainty in the position and/or momentum of a particle in any experiment given by

$$\Delta \boldsymbol{p} \Delta \boldsymbol{x} \ \sim \ h$$

where $\Delta \boldsymbol{p}$, and $\Delta \boldsymbol{x}$ denote the uncertainties in position and momentum. The basic idea behind this is that in order to precisely measure the position, we must inevitably affect the wave like behavior of the particle in such a way that the momentum becomes more uncertain. If we use light, then a photon with wave length λ carries a momentum h/λ, some of which is transferred to the particle through the scattering used to determine its location. To locate the particle with precision λ we need a photon with wave length of this magnitude and therefore alter the momentum. A similar argument can be made, considering wave packets and the Fourier transform, for passing the particle through a slit to localize it.

2.3 States and observables in quantum mechanics

Having considered the physical basis for quantum effects, let us know develop some ideas about what the mathematical structure of quantum theory must have. We will adopt, as in classical mechanics, the notion that we have *states* ω for our system and *observables* $a, b, \ldots$ that can be measured. We denote by $\mathfrak{A}$ the set of observables; as yet it is not an algebra. As before, we suppose that a state and an observable together yield a probability distribution ω_a on the real line which gives the statistics of possible outcomes, and we have the pairing

$$\langle \omega | a \rangle \ = \ \int_{\mathbb{R}} t d\omega_a(t) \ .$$

(Note that, as before, we are implicitly considering a situation in which we have an *ensemble* of functionally identical systems so that we can make useful statistical predictions.)

Given an observable a and a function $f : \mathbb{R} \to \mathbb{R}$ that is, say, continuous, we can define the observable $f(a)$. This is the observable that takes value $f(t)$ if a takes value t. It should have distribution

$$\omega_{f(a)} \ = \ \omega_a \circ f^{-1}$$

in state ω. In particular, we have scalar multiplication on the set of observables, $a \mapsto \lambda a$.

We also have some notion of which sets observables are *simultaneously measurable*, meaning that we can (in principle) measure all observables in a set together with arbitrary precision. Clearly a and $f(a)$ are simultaneously measurable. More generally the whole family $C(a) \ = \ \{ f(a) \ : f \in C(\mathbb{R}) \}$ is measurable. This is the first inkling that their is some sort of algebraic/analytic structure on the space of observables, because $C(a)$ itself is an algebra (we can multiply, and add, and scalar multiply).

Can we make sense of addition of two observables $a + b$? If they are simultaneously measurable, then the answer is yes and we can go further assigning to each state a probability distribution $\omega_{a,b}$ on $\mathbb{R}^2$ giving their *joint distribution* in a measurement (essentially

this is what we mean by simultaneously measurable). Their sum is the observable with distribution

$$\omega_{a+b}(E) \;=\; \omega_{a,b}(\{(s,t) \;:\; s+t \in E\}) \;.$$

But if they are *not* simultaneously measurable, then we have no such joint distribution.

To proceed we need a little bit more structure. It is natural to ask that the dual pairing $(a,\omega) \mapsto \langle a|\omega\rangle$ *separates points* on observables and on states. That is, if

$$\langle a|\omega\rangle = \langle b|\omega\rangle$$

for all states ω then a and b are, in fact, the same observables. After all, "a difference, to be a difference, has to make a difference."[1] If there is no way to tell the observables apart, what sense is there in distinguishing them. Similarly, if

$$\langle a|\omega\rangle = \langle a|\omega'\rangle$$

for all observables a, then ω and ω' are, in fact, the same state. If there is no operational way to distinguish two states for a system, what would be the sense in calling them different states?

With these assumptions, we define the sum $a+b$ to be the observable with dual pairing

$$\langle a+b|\omega\rangle \;=\; \langle a|\omega\rangle + \langle b|\omega\rangle \;.$$

This does not suggest a way to measure $a+b$. Furthermore, as we will see $a+b$ can have quite a different character than either a or b in quantum systems. But with this notion, we now have the structure of a vector space on our set of observables $\mathfrak{A}$. There is no natural product on $\mathfrak{A}$, although we can, using functions of observables and addition, make the definition

$$a \circ b = \frac{(a+b)^2 - (a-b)^2}{4} \;.$$

This makes $\mathfrak{A}$ an algebra. If a and b are simultaneously observable, then $a \circ b$ is easily seen to reduce to the product $ab = ba$ defined by

$$\omega_{ab}(E) \;=\; \omega_{(a,b)}(\{(s,t) \;:\; st \in E\}) \;.$$

The other important structure we would like for our set of observables is the Poisson bracket. We will assume that the algebra of observables has a Lie bracket $\{\cdot,\cdot\}$ satisfying asymmetry, linearity, the derivation property (with respect to $\circ$), and the Jacobi identity.

In formulating models of quantum mechanics, we will work with algebras $\mathfrak{A}$ and sets of states having the above structures, mirroring those of classical physics. The key difference will be in the existence of observables that fail to be simultaneously observable, and in the implementation of the Poisson bracket. This must be done in a way that allows us to introduce uncertainty relations such as Heisenberg proposed.

[1] This phrase is attributed to Gertrude Stein.

2.4 A finite-dimensional model $\mathfrak{A}_n$ of quantum mechanics

Let us take for $\mathfrak{A}$ the set

$$\mathfrak{A}_n \;=\; \left\{a \in \mathbb{C}^{n\times n} \,\middle|\, a = a^\dagger\right\}$$

of self-adjoint $n \times n$ matrices. Here $M^\dagger$ denotes the conjugate transpose matrix:

$$a^\dagger_{i,j} \;=\; a^*_{j,i}$$

where z^*= complex conjugate of z. The set $\mathfrak{A}_n$ is clearly a real vector space. It is *not* an algebra under matrix multiplication, but is an algebra under the symmetric product:

$$a \circ b \;=\; \frac{1}{2}(ab + ba).$$

In particular, if $a \in \mathfrak{A}_n$ then $a^2 \in \mathfrak{A}_n$ and, in fact, any real polynomial $p(a) \in \mathfrak{A}_n$.

Because $\mathfrak{A}_n$ is finite dimensional, we can identify the space of linear functionals on $\mathfrak{A}_n$ with itself. One way to do this is to define the following inner product on $\mathfrak{A}_n$:

$$\langle a|b\rangle \;:=\; \operatorname{tr} ab\,,$$

where tr denotes the trace. Explicitly,

$$\langle a|b\rangle \;:=\; \sum_{i,j} a_{i,j}b_{j,i}\,.$$

One easily verifies that the map $(a, b) \mapsto \langle a|b\rangle$ has all the properties of an inner product. The Riesz representation theorem then states that for any linear functional ω on $\mathfrak{A}_n$ there is $b \in \mathfrak{A}_n$ such that $a \overset{\omega}{\mapsto} \operatorname{tr} ab$.

Let ω be a state of our system, as specified by the dual pairing

$$a \mapsto \langle a|\omega\rangle\;.$$

This is a linear functional on $\mathfrak{A}_n$. By the Riesz representation theorem, there is $\rho_\omega \in \mathfrak{A}_n$ such that

$$\langle a|\omega\rangle \;=\; \operatorname{tr} a\rho_\omega\,.$$

However, ω is not an arbitrary linear functional, it has two key properties:

1. $\langle I|\omega\rangle = 1$ (a constant observable is mapped to its value), and
2. $\left\langle a^2|\omega\right\rangle \geq 0$ for any $a \in \mathfrak{A}_n$ (positive observables, have positive expectation).

From the first we conclude that

$$\operatorname{tr}\rho_\omega \;=\; 1,$$

while from the second, applied to the observable $a = \psi\psi^\dagger$, with $\psi \in \mathbb{C}^n$, we find that

$$\psi^\dagger\rho_\omega\psi \;\geq\; 0$$

for any $\psi \in \mathbb{C}^n$. That is ρ_ω is positive semi-definite with trace one.

We have obtained the following simple result

Proposition 2.1. *The set of states associated the algebra $\mathfrak{A}_n$ one-to-one correspondence with the set positive semi-definite matrices of trace one, where a matrix ρ corresponds to the state $a \mapsto \operatorname{tr} a\rho$.*

Positive semi-definite matrices of trace one are called *density matrices*. They provide the most general formulation of the states of a finite dimensional quantum system with algebra $\mathfrak{A}$. Let $\mathfrak{A}_n^{+,1}$ denote the set of density matrices. It is easily seen to be a convex set (as it should be, being the set of states). It is natural to ask for the extreme points.

Proposition 2.2. *The set $\mathfrak{A}_n^{+,1}$ is convex and compact (in the usual topology on $\mathfrak{A}_n$)[2]. Furthermore, the cextreme points of $\mathfrak{A}_n^{+,1}$ are precisely the rank one projections $\psi\psi^\dagger$ with $|\psi| = 1$.*

Proof. It is left as an exercise to show that $\mathfrak{A}_n^{+,1}$ is compact, using the Heine Borel theorem. Convexity is clear.

Suppose that $\rho \in \mathfrak{A}_n^{+,1}$. By the spectral theorem for self-adjoint matrices we have a decomposition

$$\rho \;=\; \sum_{j=1}^{m} \lambda_j \psi_j \psi_j^\dagger$$

where $\lambda_1, \ldots, \lambda_m$ are the non-zero eigenvalues of ρ, with corresponding normalized eigenvectors $\psi_1, \ldots, \psi_m$. Since ρ is positive semi-definite, we have $\lambda_j > 0$, and since $\operatorname{tr} \rho = 1$, we have $\sum_{j=1}^m \lambda_j = 1$. Thus if $m \geq 2$, then ρ is not an extreme point.

It remains to show that any rank one projection is an extreme point. Suppose that $\rho = \psi\psi^\dagger$ and that $\rho = t\rho_1 + (1-t)\rho)$. Let $\phi \perp \psi$. Since ρ_1 and ρ_2 are both positive semi-definite and $\rho\phi = 0$, we must have $\rho_1\phi = \rho_2\phi = 0$. Thus, both ρ_1 and ρ_2 are rank one with range in span ψ. Since they are both trace one, we have $\rho_1 = \rho_2 = \rho$, i.e., ρ is extreme.

□

So the pure states of our quantum system with algerbra $\mathfrak{A}_n$ are in one-to-one correspondence with rank one projections of $\mathbb{C}^n$. This can also be seen as the projective space $\mathbf{CP}^{n-1}$, i.e. $\mathbb{C}^n \setminus \{0\}$ modulo multiplication by a constant. Note that the spectral theorem for a general state ρ gives an expression for ρ as a convex combination of pure states!

[2]Since $\mathfrak{A}_n$ is finite dimensional all of the possible norms it has as a vector space, or as a space of linear operators, generate the same topology, which is the topology generated by the coordinate maps $a \mapsto a_{i,j}$.

Lecture 3

Finite dimensional models of quantum mechanics

3.1 Poisson bracket and dynamics for $\mathfrak{A}_n$

Recall that we introduced the algebra

$$\mathfrak{A}_n \;=\; \left\{a \in \mathbb{C}^{n\times n} \,\middle|\, a = a^\dagger\right\}$$

of observables, with the space of states

$$\mathfrak{A}_n^{+,1} \;=\; \{\rho \in \mathfrak{A}_n \mid \operatorname{tr}\rho = 1 \quad \text{and } \rho \geq 0\}\ ,$$

and dual pairing

$$\langle a|\rho\rangle \;=\; \operatorname{tr} a\rho\ .$$

We have yet to write down the Poisson bracket for this system.

A natural object to look at is the commutator $[a,b] = ab - ba$, which is well known to be a Lie bracket on the space of $n \times n$ matrices. It is easily seen to satisfy

1. $[a,b] = -[b,a]$ (anti-symmetry),
2. $[a + \lambda b, c] = [a,b] + \lambda[a,c]$ (linearity)
3. $[a, b \circ c] = [a,b] \circ c + b \circ [a,c]$ (derivation property)
4. $[a,[b,c]] + [b,[c,a]] + [c,[a,b]] = 0$ (Jacobi identity).

In fact the derivation property even holds with respect to the usual matrix product:

$$[a,bc] = abc - bca = (ab - ba)c + b(ac - ca) = [a,b]c + b[a,c]$$

— note that order of matrices very much matters here. However, $[a,b]$ is not self-adjoint if a and b are. Instead we have

$$([a,b])^\dagger \;=\; (ab - ba)^\dagger \;=\; b^\dagger a^\dagger - a^\dagger b^\dagger \;=\; ba - ab = -[a,b]$$

if $a, b \in \mathfrak{A}_n$. This is easily remedied by using $i[a,b]$ in place of the commutator (since complex conjugation introduces a minus sign).

But there is still something missing — it doesn't quite make sense to take $\{a,b\} = i[a,b]$. Recall that we want to have the equation of motion

$$\frac{d}{dt}a_t \;=\; \{H, a_t\}$$

with H the Hamiltonian observable. The observable H will have units of energy (any explicit expression for it will depend on the choice of measurement units). The equation $\frac{d}{dt}a_t = i[H,a_t]$ has a problem: the left hand side has units $\frac{\text{units of } a}{\text{time}}$ while the right hand side has units units of $a \times$ energy. We conclude that the correct equation must be of the form $\frac{d}{dt}a_t = \frac{i}{c}[H,a_t]$ where c is a physical constant with units energy $\times$ time (also called *action*). The correct choice for c, as it turns out, is $c = \hbar$. We conclude that the Poisson bracket should be

$$\{a,b\} \;=\; \frac{i}{\hbar}[a,b]\;.$$

We have obtained the following finite dimensional model for a quantum system:

1. Observables: $\mathfrak{A}_n = n \times n$ self-adjoint matrices,
2. States: $\mathfrak{A}_n^{+,1} = n \times n$ positive semi-definite matrices of trace one, and
3. Poisson bracket: $\{a,b\}_\hbar \;=\; \frac{i}{\hbar}[a,b]$.

3.2 The spectral theorem for self-adjoint matrices

Before discussing the dynamics on $\mathfrak{A}_n$ we should look at the structure of self-adjoint matrices. The Euclidean inner product between two vectors $\boldsymbol{z}, \boldsymbol{w} \in \mathbb{C}^n$ is defined by

$$\langle \boldsymbol{z} | \boldsymbol{w} \rangle \;=\; \boldsymbol{z}^\dagger \boldsymbol{w} \;=\; \sum_{j=1}^{n} z_j^* w_j\;. \tag{3.1}$$

A self adjoint matrix $A \in \mathfrak{A}_n$ satisfies

$$\langle \boldsymbol{z} | A\boldsymbol{w} \rangle \;=\; \langle A\boldsymbol{z} | \boldsymbol{w} \rangle\;.$$

Definition 3.1. A collection $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_m$ of vectors in $\mathbb{C}^n$ is *orthonormal* if

$$\left\langle \boldsymbol{z}_i \middle| \boldsymbol{z}_j \right\rangle \;=\; \delta_{i,j}$$

for $j = 1, \ldots, m$

An orthonormal set is linearly independent, as one can readily check. Thus if $m = n$, then the set is a basis for $\mathbb{C}^n$, and will be called an *orthonormal basis*.

Definition 3.2. An $n \times n$ matrix U is *unitary* if $U^\dagger U = UU^\dagger = I$.

The columns of a unitary matrix form an orthonormal basis. Conversely, given an orthonormal basis $z_1, \dots, z_n$, the matrix

$$U \;=\; [z_1, \dots, z_n] \tag{3.2}$$

is a basis.

A very important result for quantum mechanics is the following

Theorem 3.3 (Spectral Theorem for Self-Adjoint Matrices)**.** *Let $A \in \mathfrak{A}_n$. Then there is an orthonormal basis $z_1, \dots, z_n$ of eigenvectors for A, with corresponding eigenvalues $\lambda_1, \dots, \lambda_n \in \mathbb{R}$. If $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n)$ and U is the corresponding unitary, as in* (3.2)*, then*

$$A \;=\; U \operatorname{diag}(\boldsymbol{\lambda}) U^\dagger \,. \tag{3.3}$$

Remark. Here $\operatorname{diag}(\boldsymbol{\lambda})$ denotes the diagonal matrix with $\boldsymbol{\lambda}$ along the diagonal (and zeros off the diagonal).

Proof. First note that (3.3) is equivalent to $AU = U \operatorname{diag}(\boldsymbol{\lambda})$, which is just a reexpression of the eignvector equations $Az_j = \lambda_j z_j$, $j = 1, \dots, n$. Thus it suffices to prove there is an orthonormal basis of eigenvectors and the eigenvalues are real.

Second note that

$$\lambda_j \;=\; \big\langle z_j \big| A z_j \big\rangle \;=\; \big\langle A z_j \big| z_j \big\rangle \;=\; \big\langle \lambda z_j \big| z_j \big\rangle \;=\; \lambda_j^* \,.$$

It follows that all eigenvalues are real. So it suffices to prove that A has an orthonormal basis of eigenvectors.

To prove this, we start by noting that A has at least one eigenvector. Indeed, any matrix has an eigenvector, since the equation $\det(A - \lambda)$ is a polynomial of degree n in λ, so has at least one root $\lambda_1 \in \mathbb{C}$. For A self-adjoint, λ_1 must be real by the above argument. Let z_1 be the corresponding eigenvector. Let

$$\mathcal{H}_1 \;=\; \{z_1\}^\perp \;=\; \{z \in \mathbb{C}^n \;:\; \langle z_1 | z \rangle \;=\; 0\} \,.$$

If $z \in \mathcal{H}_1$, then

$$\langle z_1 | Az \rangle \;=\; \langle A z_1 | z \rangle \;=\; \lambda_1 \langle z_1 | z \rangle \;=\; 0 \,.$$

So $\mathcal{H}_1$ is an invariant subspace for A. Since $\mathcal{H}_1$ is a subspace of dimension $n-1$, the restriction of A to $\mathcal{H}_1$ can be expressed as an $(n-1) \times (n-1)$ matrix once we pick an orthonormal basis for $\mathcal{H}_1$. This matrix is easily seen to be self-adjoint. Repeating the argument we find in this way a sequence of eigenvectors $z_1, \dots, z_n$ with $z_j \in \{z_1, \dots, z_{j-1}\}^\perp$ for $j = 2, \dots, n$. After n steps, we have $\{z_1, \dots, z_{j-1}\}^\perp = \{0\}$ and the argument terminates.

□

The set $\{\lambda_1, \dots, \lambda_n\}$ is called the *spectrum* of A, and is denoted $\sigma(A)$. The spectral theorem allows us to define $f(A)$ for *an arbitrary function* $f : \sigma(A) \to \mathbb{C}$. Indeed, we simply take

$$f(A) \;:=\; U \operatorname{diag}(f(\boldsymbol{\lambda})) U^\dagger$$

with $f(\boldsymbol{\lambda}) = (f(\lambda_1), \dots, f(\lambda_n))$. This induces an algebra homomorphism from the commutative algebra $C(\sigma(A))$ of alla functions on $\sigma(A)$ into the set of $n \times n$ matrices. That is, we have

1. $(\alpha f)(A) = \alpha(f(A))$ for $\alpha \in \mathbb{C}$ and $f \in C(\sigma(A))$,
2. $(f+g)(A) = f(A) + g(A)$ for $f, g \in C(\sigma(A))$, and
3. $(fg)(a) = f(A)g(A)$ for $f, g \in C(\sigma(A))$.

To prove the last fact it is useful to note that

$$\operatorname{diag}(fg(\boldsymbol{\lambda})) = \operatorname{diag}(f(\boldsymbol{\lambda}))\operatorname{diag}(g(\boldsymbol{\lambda})) = \operatorname{diag}(f(\boldsymbol{\lambda}))UU^\dagger \operatorname{diag}(g(\boldsymbol{\lambda}))\,,$$

since U is unitary. The image of this homomorphism

$$C(A) := \{f(A) : f \in C(\sigma(A))\}$$

is the set of all matrices that have the same orthonormal basis of eigenvectors as A. It is a commutative sub-algebra of $\mathbb{C}^{n\times n}$. Every element of $C(A)$ is a *normal* matrix, meaning it commutes with its adjoint. Indeed, since $f(A)^\dagger = f^*(A)$ this follows directly from the above.

We should check that this definition of $f(A)$ agrees with other definitions one might take. Note that

- If $f(x) \equiv c$, then $f(A) = cI$.
- If $f(x) = x$, then $f(A) = A$.

From these two observations, and the fact that $A \mapsto f(A)$ is an algebra homomorphism, it follows that if f is a polynomial, $f(x)c_0 + c_1 x + \cdots + c_k x^k$, we have

$$f(A) = c_0 + c_1 A + \cdots + c_k A^k\,.$$

By taking limits, it follows that if f is analytic in an open disk containing a point $\{\lambda_0\} \cup \sigma(A)$, then

$$f(A) = \sum_{k=0}^{\infty} c_k (A - \lambda_0 I)^k$$

where $f(z) = \sum_{k=0}^{\infty} c_k (z-\lambda_0)^k$ is the power series of f at λ_0.

3.3 Dynamics on $\mathfrak{A}_n$

For a given system, there should be a distinguished observable H, called the *Hamiltonian* that generates the dynamics. As in the classical case, we have two possible pictures of the evolution of the system

1. The Heisenberg picture: the states of the system remain fixed in time, and observables evolve according to the equation

$$\frac{d}{dt}a_t = \frac{i}{\hbar}[H, a_t]\,. \tag{3.4}$$

   The expected value of an observable at time t in the state with density matrix ρ is given by $\operatorname{tr} a_t \rho$.

2. The Schrödinger picture: the observables remain fixed in time, while states (density matrices) evolve according to

$$\frac{d}{dt}\rho_t \;=\; -\frac{i}{\hbar}[H,\rho_t]\,. \tag{3.5}$$

The expected value of an observable a at time t in the state with density matrix ρ_t is $\operatorname{tr} a\rho_t$.

Eq. (3.4) is called the *Heisenberg equation.* The solution to this equation, with $a_0 = a$, is easily seen to be given by

$$a_t \;=\; U_t^\dagger a U_t\,, \quad U_t \;=\; \mathrm{e}^{-\frac{it}{\hbar}H}\,,$$

Here the exponential of a matrix can be defined by the spectral theorem or by the power series

$$\mathrm{e}^{-\frac{it}{\hbar}H} \;=\; I + \sum_{n=1}^{\infty}\frac{1}{n!}\left(\frac{-itH}{\hbar}\right)^n\,,$$

and one sees that U_t is *unitary*, $U_t^\dagger U_t = I$. Similarly, the solution to eq. (3.5) with $\rho_0 = \rho$ is seen to be

$$\rho_t \;=\; U_t \rho U_t^\dagger\,.$$

Since

$$\operatorname{tr} a_t\rho \;=\; \operatorname{tr} U_t^\dagger a U_t \rho \;=\; \operatorname{tr} a U_t \rho U_t^\dagger \;=\; \operatorname{tr} a\rho_t$$

by *cylcity of the trace*. So the two pictures of quantum mechanics are equivalent.

For a pure state $\rho = \psi\psi^\dagger$ we have

$$\rho_t \;=\; \psi_t\psi_t^\dagger, \quad \psi_t = U_t\psi\,.$$

Thus a pure state remains pure for all time, and the corresponding vector solves the *Schrödinger equation*:

$$\frac{d}{dt}\psi_t \;=\; -\frac{i}{\hbar}H\psi_t\,. \tag{3.6}$$

3.4 Heisenberg uncertainty relations

Let us consider the uncertainty of an observable in a state ω. One way to measure the uncertainty is to look at root mean square deviation the variance of the observable

$$\Delta_\omega(a) \;=\; \sqrt{\langle a^2|\omega\rangle - \langle a|\omega\rangle^2}\,. \tag{3.7}$$

Note that

$$\Delta_\omega(a) \;=\; \sqrt{\big\langle (a - \langle a|\omega\rangle\, I)^2\big|\omega\big\rangle}\,. \tag{3.8}$$

For a pure state of classical mechanics one has $\Delta(a)$ for any observable, since the corresponding measure in phase space is concentrated at a single point. This is not so in quantum mechanics! Indeed for the pure state $\omega = \psi\psi^\dagger$ we have

$$\Delta_\omega^2(a) \;=\; \psi^\dagger a^2\psi - (\psi^\dagger a\psi\psi^\dagger a\psi) \;=\; \psi^\dagger a(I - \psi\psi^\dagger)a\psi\,.$$

This can be expressed as follows

$$\Delta_\omega(a) \;=\; \sqrt{\psi^\dagger a P_\psi^\perp a \psi}$$

where $P_\psi^\perp$ is projection on the subspace orthogonal to ψ. We conclude that, *unless ψ is an eigenvector of a the uncertainty of a in the state $\psi\psi^\dagger$ is strictly positive.*

Heisenberg's uncertainty relation relates the uncertainties of two observables a and b to their commutator $[a, b]$.

Proposition 3.4 (General Heisenberg Uncertainty Relation)**.** *Let $a, b \in \mathfrak{A}_n$ and let $\rho \in \mathfrak{A}_n^{+,1}$ be a state. Then*

$$\Delta_\rho(a)\Delta_\rho(b) \;\geq\; \frac{1}{2}|\left\langle [a,b]|\rho\right\rangle|\,. \tag{3.9}$$

We will use the following in the proof:

Lemma 3.5 (Cauchy-Schwarz inequality for matrix inner product)**.** *Let $A, B \in \mathbb{C}^{n\times n}$. Then*

$$|\operatorname{tr} A^\dagger B|^2 \;\leq\; \operatorname{tr} A^\dagger A \operatorname{tr} B^\dagger B\,.$$

This Lemma follows from the usual Cauchy-Schwarz inequality, once one observes that $\langle A,\ B\rangle = \operatorname{tr} A^\dagger B$ obeys all the requirements of an inner product on $\mathbb{C}^{n\times n}$.

Proof of the Heisenberg Uncertainty relations. Let $a' = a - \langle a|\rho\rangle\, I$ and $b' = b - \langle b|\rho\rangle\, I$. Then $\Delta_\rho^2(a) \;=\; \left\langle a'^2|\rho\right\rangle$, $\Delta_\rho^2(b) \;=\; \left\langle b'^2|\rho\right\rangle$ and $[a, b] \;=\; [a', b']$. Thus eq. (3.9) follows from the following equation

$$|\left\langle [a,b]|\rho\right\rangle| \;\leq\; 2\sqrt{\left\langle a^2|\rho\right\rangle \left\langle b^2|\rho\right\rangle}\,. \tag{3.10}$$

(applied with $a \mapsto a'$ and $b \mapsto b'$)

We have

$$\left\langle [a,b]|\rho\right\rangle \;=\; \operatorname{tr} ab\rho - \operatorname{tr} ba\rho.$$

By cyclicity of the trace and the Cauchy-Schwartz inequality, we have

$$|\operatorname{tr} ab\rho| \;=\; |\operatorname{tr}\sqrt{\rho}ab\sqrt{\rho}| \;\leq\; \sqrt{\operatorname{tr}\sqrt{\rho}a^2\sqrt{\rho}\operatorname{tr}\sqrt{\rho}b^2\sqrt{\rho}} \;=\; \sqrt{\left\langle a^2|\rho\right\rangle \left\langle b^2|\rho\right\rangle}$$

and similarly,

$$|\operatorname{tr} ba\rho| \;\leq\; \sqrt{\left\langle a^2|\rho\right\rangle \left\langle b^2|\rho\right\rangle}\,.$$

Eq. (3.10) follows.

□

Lecture 4

Example systems: a Qubit and an infinite dimensional model of quantum mechanics

4.1 A Qubit — an example

Although we obtained the model $\mathfrak{A}$ in a speculative way, it turns out to be useful. The case $n = 2$ is exactly the correct system to model the internal spin degrees of freedom of a particle of "spin $\frac{1}{2}$", such as an electron, a proton, or a neutron. More generally, it can model any "two level" system, a *qubit*.

Consider the algebra $\mathfrak{A}_2$, a four dimensional vector space over $\mathbb{R}$. This space has a basis given by the identity

$$\sigma_0 = I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

together with the *Pauli* matrices:

$$\sigma_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_2 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \text{and} \quad \sigma_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{4.1}$$

Consider a system with $\mathfrak{A}_2$ for its algebra of observables. Let H be the Hamiltonian of the system. By the spectral theorem, H is unitarily equivalent to a diagonal matrix, i.e.,

$$H \;=\; U\left(\alpha I + \beta\sigma_3\right)U^\dagger \;=\; \alpha I + \beta U\sigma_3 U^\dagger\,.$$

If $\beta = 0$, then the Hamiltonian $H = \alpha I$ commutes with every operator — this is a trivial situation with no dynamics. If $\beta \neq 0$, then by making suitable changes of variables, we can take $\alpha = 0$, $U = I$ and $\beta = 1$:

1. Adding a constant multiple of the identity to the Hamiltonian does not change the Heisenberg equation of motion or the Liouville/Schrödinger equation for the density matrix (solutions to the Schrödinger equation for a vector get multiplied by and irrelevant overall time dependent phase). So we may "shift the origin of the energy" to take $\alpha = 0$.

2. The explicit expression for U depends on the orthonormal basis used to describe the space $\mathbb{C}^2$. By working in the basis of eigenvectors for H, we may take $U = I$.

3. We can take $\beta = 1$ by changing our units for the energy.

Thus the only interesting Hamiltonian on $\mathfrak{A}_2$ is $H = \sigma_3$. How does the system evolve? With one further change of coordinates (changing units for time), we may take $\hbar = 1$. Then

$$U_t \;=\; \mathrm{e}^{-it\sigma_3} \;=\; \begin{bmatrix} \mathrm{e}^{-it} & 0 \\ 0 & \mathrm{e}^{it} \end{bmatrix} .$$

Heisenberg picture

The identity I and σ_3 both commute with $H = \sigma_3$, so do not evolve in time. For σ_1 and σ_2 we have

$$\sigma_1(t) \;=\; \begin{bmatrix} 0 & \mathrm{e}^{2it} \\ \mathrm{e}^{-2it} & 0 \end{bmatrix} \quad \text{and} \quad \sigma_2(t) \;=\; \begin{bmatrix} 0 & -i\mathrm{e}^{2it} \\ i\mathrm{e}^{-2it} & 0 \end{bmatrix} . \tag{4.2}$$

A general observable $a = \left[\begin{smallmatrix} \alpha & \gamma^* \\ \gamma & \beta \end{smallmatrix}\right]$ evolves according to

$$a_t \;=\; \begin{bmatrix} \alpha & \gamma^*\mathrm{e}^{2it} \\ \gamma\mathrm{e}^{-2it} & \beta \end{bmatrix} .$$

Eq. (4.2) can be determined explicitly from $\sigma_j = U_t^\dagger \sigma_j U_t$, but could also be solved for using the Heisenberg equation. Indeed, note that

$$[\sigma_3, \sigma_1] \;=\; \begin{bmatrix} 0 & 2 \\ -2 & 0 \end{bmatrix} \;=\; -2i\sigma_2 \quad \text{and} \quad [\sigma_3, \sigma_2] \;=\; \begin{bmatrix} 0 & -2i \\ -2i & 0 \end{bmatrix} \;=\; -2i\sigma_1 \, .$$

It follows that

$$\dot\sigma_1(t) \;=\; 2\sigma_2(t) \quad \text{and} \quad \dot\sigma_2(t) \;=\; -2\sigma_1(t) \, .$$

The form (4.2) of the solutions follows.

Schrödinger Picture

A density matrix of the spin-1/2 system is a 2×2 positive semi-definite matrix. Such a matrix is of the following form:

$$\rho \;=\; \begin{bmatrix} \alpha & \gamma^* \\ \gamma & \beta \end{bmatrix} \tag{4.3}$$

with $\alpha, \beta \geq 0$, $\alpha + \beta = 1$, and $\det \rho = \alpha\beta - |\gamma|^2 \geq 0$. Furthermore, ρ is represents a pure state if and only if $\det \rho = \alpha\beta - |\gamma|^2 = 0$.

Proposition 4.1 (Bloch sphere representation)**.** *Let ρ be the density matrix of a state of $\mathfrak{A}_2$. Then ρ has the form*

$$\rho \;=\; \tfrac{1}{2}\left[I + \boldsymbol{v} \cdot \boldsymbol{\sigma}\right] \;=\; \tfrac{1}{2}\left[I + v_1\sigma_1 + v_2\sigma_2 + v_3\sigma_3\right] , \tag{4.4}$$

with $\boldsymbol{v} = (v_1, v_2, v_3)$ and $|\boldsymbol{v}|^2 \leq 1$. Furthermore, ρ is a pure state if and only if $|\boldsymbol{v}| = 1$.

Proof. Let ρ have the form (4.3). Taking $v_1 = 2\Re(\gamma)$, $v_2 = 2\Im(\gamma)$ and $v_3 = \alpha - \beta$, we find that

$$\rho \;=\; \tfrac{1}{2}\left[I + v_1\sigma_1 + v_2\sigma_2 + v_3\sigma_3\right] \,,$$

with

$$v_1^2 + v_2^2 + v_3^2 \;=\; 4|\gamma|^2 + \alpha^2 - 2\alpha\beta + \beta^2 \;\leq\; (\alpha+\beta)^2 \;=\; 1\,.$$

It also follows that ρ is pure, with $\alpha\beta - |\gamma|^2 = 0$, if and only if $|\boldsymbol{v}|^2 = 1$.

□

The Schrödinger evolution of a density matrix ρ of the form (4.3) with Hamiltonian $H = \sigma_3$ is just

$$\rho(t) \;=\; U_t\rho U_t^\dagger \;=\; \begin{bmatrix} \alpha & \gamma^*\mathrm{e}^{-2it} \\ \gamma\mathrm{e}^{2it} & \beta \end{bmatrix}.$$

In terms of the vector $\boldsymbol{v}(t) = (v_1(t), v_2(t), v_3(t))$ in the Bloch sphere, we have

$$\boldsymbol{v}(t) \;=\; \boldsymbol{v}(0)\begin{bmatrix} \cos(2t) & \sin(2t) & 0 \\ -\sin(2t) & \cos(2t) & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

That is $\boldsymbol{v}(t)$ rotates counter clockwise around the x_3 axis.

Note that the states $\frac{1}{2}[I + \alpha\sigma_3]$ are all fixed by the evolution. These are the states that are diagonal in the basis of eigenvectors for $H = \sigma_3$. All other states rotate with frequency π.

Uncertainty relations

Consider the state $\rho_v = \frac{1}{2}\left[I + \boldsymbol{v}\cdot\boldsymbol{\sigma}\right]$. Let us compute the uncertainty of the observable σ_1 in ρ. Since

$$\sigma_1^2 = I, \quad \sigma_1\sigma_2 \;=\; i\sigma_3, \quad \text{and} \quad \sigma_1\sigma_3 \;=\; -i\sigma_2\,,$$

we have

$$\Delta_v^2(\sigma_1) \;=\; \operatorname{tr}\sigma_1^2\rho_v - \left(\operatorname{tr}\sigma_1\rho_v\right)^2 \;=\; 1 - v_1^2.$$

By a similar computation for $j = 2, 3$, we find that

$$\Delta_v(\sigma_j) \;=\; \sqrt{1 - v_j^2} \tag{4.5}$$

for $j = 1, 2, 3$.

Note that at most one of the uncertainties $\Delta_v(\sigma_j)$, $j = 1, 2, 3$, can vanish. Indeed $\Delta_v(\sigma_1) = 0$ if and only if $v_1 = \pm 1$, which forces $\Delta_v(\sigma_2) = \Delta_v(\sigma_3) = 1$. So there is no state in which any two of the spin observables, σ_j, $j = 1, 2, 3$, take a sharp value. These observables are not simultaneously measurable in a strong way.

It is a short computation to see that the Pauli matrices have the following commutation relations

$$[\sigma_i, \sigma_j] \;=\; 2\mathrm{i}\varepsilon_{ijk}\sigma_k \tag{4.6}$$

where we take the summation convention on the right hand side and ε_{ijk} is the totally antisymmetric symbol, satisfying $\varepsilon_{ijk} = 0$ if any two entries indices are equal and

$$\varepsilon_{123} = \varepsilon_{231} = \varepsilon_{312} = -\varepsilon_{213} = -\varepsilon_{132} = -\varepsilon_{321} = 1$$

i.e., $\varepsilon_{\tau(123)} = \text{sign}(\tau)$ for any permutation $\tau \in S_3$. The Heisenberg uncertainty relations for the Pauli matrices thus state that

$$\Delta_v(\sigma_1)\Delta_v(\sigma_2) \;\geq\; |\text{tr}\, \sigma_3 \rho_v| \,.$$

Plugging in the uncertainties, we see that

$$\sqrt{(1-v_1^2)(1-v_2^2)} \;\geq\; |v_3| \,.$$

Since

$$|v_3| \leq \sqrt{1-v_1^2-v_2^2} \leq \sqrt{1-v_1^2-v_2^2+v_1^2v_2^2} \,,$$

we see that there is strict inequality unless $|v_3| = 1$. However, for $v_3 = \pm 1$, the Heisenberg uncertainty relations are sharp.

4.2 An Infinite Dimensional Model of Quantum Mechanics

The models $\mathfrak{A}_n$ are readily generalized to an infinite dimensional model, which is rich enough to describe many real quantum systems. In practice, however, the finite dimensional models are sufficient to understand most quantum phenomena, and we will focus on them for much of this course. However, it is useful to have the idea of the infinite dimensional system in mind.

Let $\mathcal{H}$ be a separable, infinite dimensional Hilbert space, and let $\mathfrak{A}_\infty$ denote the space of bounded, self-adjoint maps on $\mathcal{H}$. That is $A \in \mathfrak{A}_\infty$ satisfies

1. $A : \mathcal{H} \to \mathcal{H}$ is a linear map;
2. there is $c < \infty$ such that $\|A\psi\| \leq c\|\psi\|$ for all $\psi \in \mathcal{H}$ — the smallest such c is called the (operator) norm of A, denoted $\|A\|$; and
3. $\langle \phi | A\psi \rangle = \langle A\phi | \psi \rangle$ for all ϕ, ψ.

The space $\mathfrak{A}_\infty$ will be the algebra of observables for our model; it is a vector space over $\mathbb{R}$ with the commutative product

$$A \circ B \;=\; \tfrac{1}{2}(AB + BA)$$

and Poisson bracket

$$\{A, B\} \;=\; \frac{i}{\hbar}[A, B] \,.$$

Recall that an *orthogonal projection* P on $\mathcal{H}$ is an element of $\mathfrak{A}_\infty$ such that $P^2 = P$. Orthogonal projections are in one-to-one correspondence with closed subspaces of $\mathcal{H}$ via the

map $P \mapsto \operatorname{ran} P$. Let $\mathcal{P}(\mathcal{H})$ denote the set of (orthogonal) projections on $\mathcal{H}$. There is a natural partial order on $\mathcal{P}(\mathcal{H})$ given by the relation

$$P_1 \leq P_2 \quad \text{if} \quad \operatorname{ran} P_1 \subset \operatorname{ran} P_1 \ ;$$

equivalently $P_1 P_2 = P_2 P_1 = P_1$. One has the following

Theorem 4.2 (Spectral Theorem for Bounded Self Adjoint Operators)**.** *Let $A \in \mathfrak{A}_\infty$. Then there is a map $P_A : \mathbb{R} \to \mathcal{P}(\mathcal{H})$ such that*

1. *$\lambda \mapsto P_A(\lambda)$ is increasing, i.e., if $\lambda \leq \mu$, then $P_A(\lambda) \leq P_A(\mu)$;*
2. *$P_A(\lambda) = 0$ for $\lambda < -\|A\|$ and $P_A(\lambda) = 1$ for $\lambda \geq \|A\|$;*
3. *$P_A(\lambda)$ is* strongly right continuous*;*
4. *$P_A(\lambda)$ commutes with any bounded operator B such that $[B, A] = 0$; and*
5. *$A = \int_{\mathbb{R}} \lambda dP_A(\lambda)$ in the sense that*

$$\langle \phi | A | \psi \rangle \ = \ \int_{\mathbb{R}} \lambda d \langle \phi | P_A(\lambda) | \psi \rangle \tag{4.7}$$

for any $\phi, \psi \in \mathcal{H}$.

Remarks.

1. A map $\lambda \overset{F}{\mapsto} \mathcal{B}(\mathcal{H})$, the space of bounded operators on $\mathcal{H}$, is *strongly right continuous* if

$$\lim_{\mu \downarrow \lambda} F(\mu)\phi \ = \ F(\lambda)\phi$$

for every $\phi \in \mathcal{H}$ and $\lambda \in \mathbb{R}$. It does not follow, and need not be, that $\|F(\lambda) - F(\mu)\|$ converges to zero. *Strong left continuity* and *strong continuity* are defined similarly.

2. When $\phi = \psi$ in (4.7), the function $\langle \phi | P_A(\lambda) | \phi \rangle$ is an increasing function on the real line, while for $\phi \neq \psi$, the *polarization identity*

$$\begin{aligned} \langle \phi | P_A(\lambda) | \psi \rangle \ = \ \tfrac{1}{4} \Big[& \langle \phi + \psi | P_A(\lambda) | \phi + \psi \rangle - \langle \phi - \psi | P_A(\lambda) | \phi - \psi \rangle \\ & - i \langle \phi + i\psi | P_A(\lambda) | \phi + i\psi \rangle + i \langle \phi + \psi | P_A(\lambda) | \phi + \psi \rangle \Big] \end{aligned} \tag{4.8}$$

shows that $\langle \phi | P_A(\lambda) | \psi \rangle$ is a linear combination of increasing functions, and thus of *bounded variation*. The measure $d \langle \phi | P_A(\lambda) | \psi \rangle$ is the associated Lebesgue-Stieltjes measure.

3. We have introduced here the Dirac notation $\langle \phi | A | \psi \rangle$ for the operator A in the inner product $\langle \phi | A\psi \rangle \ = \ \langle A^\dagger \phi | \psi \rangle$.

An example of the spectral theorem – finite dimensions

Let $A \in \mathfrak{A}_n$, with orthonormal eigenvectors $\psi_1, \ldots, \psi_n$ and eigenvalues $\lambda_1 \leq \cdots \leq \lambda_n$, which we may choose to write in increasing order. Since

$$|\langle\psi| A |\psi\rangle| \;\leq\; \|\psi\|\|A\psi\| \;\leq\; \|A\|\|\psi\|^2$$

for any $\psi \in \mathbb{C}^n$, we see that $|\lambda_j| \leq \|A\|$ for all j (because $\lambda_j = \langle\psi_j| A |\psi_j\rangle$). Define the projection valued measure

$$P(\lambda) \;=\; \sum_{j:\lambda_j \leq \lambda} |\psi_j\rangle\langle\psi_j| \;,$$

where $|\psi_j\rangle\langle\psi_j|$ denotes the rank-one projection onto ψ_j. Note that $P(\lambda)$ is increasing, $P(\lambda) = 0$ for $\lambda < \lambda_1$, and $P(\lambda) = I$ for $\lambda \geq \lambda_n$. Furthermore, P is discontinuous at each λ_j, and we have

$$P(\lambda) = P(\lambda_j) \;, \quad \text{for } \lambda_j \leq \lambda < \lambda_{j+1}.$$

In particular, $P(\lambda)$ is continuous from the right everywhere. One has,

$$dP(\lambda) \;=\; \sum_{j=1}^{n} |\psi_j\rangle\langle\psi_j| \, \delta(\lambda - \lambda_j) d\lambda \;,$$

with $\delta(\lambda - \lambda_j)d\lambda$ the Dirac δ measure at λ_j. Thus

$$\int_{\mathbb{R}} \lambda dP(\lambda) \;=\; \sum_{j=1}^{n} \lambda_j |\psi_j\rangle\langle\psi_j| \;=\; A \;.$$

This shows how the spectral theorem state above generalizes the result for finite dimensional operators.

Lecture 5

Trace Class Operators

What are the states on the algebra of observables $\mathfrak{A}_\infty$, consisting of the bounded self-adjoint operators on a separable Hilbert space $\mathcal{H}$? Following the finite-dimensional model, we expect to have pure states of the form $\langle A|\omega\rangle = \langle\psi|A|\psi\rangle$ for $\psi \in \mathcal{H}$. Formally, for the moment, we can think of this state as the map $A \mapsto \operatorname{tr} A|\psi\rangle\langle\psi|$. To make sense of such an expression we should define a trace on suitable operators on $\mathcal{H}$.

First a word of caution: the trace on operators over $\mathcal{H}$ is a slightly subtle object. One cannot simply define it to be the sum of diagonal entries and be done. There is no reason for the sum of diagonal entries to be convergent. Nor, when the sum is convergent, does one necessarily obtain the same expression for the trace in different bases. These facts only hold for a special class of operators, *the trace class operators*, which we will now define.

5.1 Compact Operators, Positive Operators, and the Polar Decomposition

Before defining the trace class operators, we need a bit of terminology and some general facts about operators on $\mathcal{H}$. Let $\mathcal{B}(\mathcal{H})$ denote the set of *bounded linear operators on* $\mathcal{H}$ (see appendix A), with the *operator norm* of an operator B defined by

$$\|B\| = \sup_{\psi\neq 0} \frac{\|B\psi\|}{\|\psi\|} . \tag{5.1}$$

Recall that the unit ball $B_1 = \{\psi \in \mathcal{H} : \|\psi\| \leq 1\}$ in $\mathcal{H}$ is *not compact*. To see this note that if $(\psi_j)_{j=1}^\infty$ is an orthonormal basis, then this sequence has no convergent subsequence, because $\|\psi_j - \psi_k\| = \sqrt{2}$ whenever $j \neq k$. A bounded operator $A : \mathcal{H} \to \mathcal{H}$ is *compact* if $A(B_1)$ is a compact set. Thus A is compact if for any $(\psi_j)_{j=1}^\infty$ is a bounded sequence in $\mathcal{H}$, the sequence $(A\psi_j)_{j=1}^\infty$ has a convergent subsequence.

Theorem 5.1 (See, e.g., Reed and Simon 1980, Theorem VI.13)**.** *An operator $A \in \mathcal{B}(\mathcal{H})$ is trace class if and only if there is a sequence $(F_n)_{n=1}^\infty$ of finite rank operators on $\mathcal{H}$ such that $\lim_{n\to\infty} \|A - F_n\| = 0$.*

Also, the compact operators are an *ideal*

Lemma 5.2. *Let $A, M \in \mathcal{B}(\mathcal{H})$. If A is compact, then AM and MA are compact.*

Sketch of proof. For MA we use the fact that if $K \subset \mathcal{H}$ is a compact set then $M(K)$ is compact for any bounded operator. This can be shown by verifying sequential compactness for $M(K)$ (since sequential compactness is equivalent to compactness on metric spaces).

For AB we note that

1. The image $A(B_r(0))$ of any closed ball $B_r(0) = \{\psi \,:\, \|\psi\| \leq r\}$ is compact.
2. $AM(B_1(0))$ is a closed subset of $B_{\|M\|}(0)$.

□

An operator $A \in \mathcal{B}(\mathcal{H})$ is *positive semi-definite* if $\langle\psi|A\psi\rangle \geq 0$ for all $\psi \in \mathcal{H}$. We denote this by $A \geq 0$, and often abbreviate the terminology to *positive*. Given two operators we say $A \geq B$ if $A - B \geq 0$. Every bounded positive operator on $\mathcal{H}$ is self adjoint. More generally one has

Proposition 5.3. *Let $A \in \mathcal{B}(\mathcal{H})$. If $\langle\psi|A\psi\rangle \in \mathbb{R}$ for all $\psi \in \mathcal{H}$, then A is self-adjoint.*

Proof. We will use the *polarization identity*:

$$\begin{aligned}\langle\psi|\phi\rangle \;=\; \tfrac{1}{4}[&\langle\psi+\phi|\psi+\phi\rangle - \langle\psi-\phi|\psi-\phi\rangle \\ &+i\langle\psi-i\phi|\psi-i\phi\rangle - i\langle\psi+i\phi|\psi+i\phi\rangle]\,,\end{aligned} \tag{5.2}$$

which recovers the inner product from its diagonal terms. For $A^\dagger$ one has $\langle\psi|A^\dagger\phi\rangle = \langle A\psi|\phi\rangle = (\langle\phi|A\psi\rangle)^*$. Applying the polarization identity to $\langle\phi|A\psi\rangle$ we find that

$$\begin{aligned}\langle\psi|A^\dagger\phi\rangle \;=\; (\langle\phi|A\psi\rangle)^* \;=\; \tfrac{1}{4}[&\langle\phi+\psi|A(\phi+\psi)\rangle - \langle\phi-\psi|A(\phi-\psi)\rangle \\ &- i\langle\phi-i\psi|A(\phi-i\psi)\rangle + i\langle\phi+i\psi|A(\phi+i\psi)\rangle]\,,\end{aligned}$$

where we have used the fact that all the inner products on the right are real. Factoring -1 out of the factors in the second term, $-i$ out of the factors in the third term and i out of the factors in the final term, we find that

$$\begin{aligned}\langle\psi|A^\dagger\phi\rangle \;=\; \tfrac{1}{4}[&\langle\psi+\phi|A(\psi+\phi)\rangle - \langle\psi-\phi|A(\psi-\phi)\rangle \\ &- i\langle\psi+i\phi|A(\psi+i\phi)\rangle + i\langle\psi-i\phi|A(\psi-i\phi)\rangle]\,,\end{aligned}$$

which is equal to $\langle\psi|A\phi\rangle$ by (5.2).

□

We can take the square root of positive operators:

Lemma 5.4 (*The Square Root Lemma* — see, e.g., Reed and Simon 1980, Theorem VI.9)**.** *If $A \in \mathcal{B}(\mathcal{H})$ with $A \geq 0$, then there is a unique $B \in \mathcal{B}(\mathcal{H})$ with $B \geq 0$ and $B^2 = A$. Furthermore, B commutes with every bounded operator that commutes with A.*

One way to construct B is to use the spectral theorem, defining

$$B \;=\; \int_{\mathbb{R}} \sqrt{\lambda} dP_A(\lambda)\;, \tag{5.3}$$

where the final equality means

$$\langle \psi | B\phi \rangle \;=\; \int_{\mathbb{R}} \sqrt{\lambda} d\,\langle \psi | P_B(\lambda)\psi \rangle\;.$$

To do this, one should first show that the projection valued measure of A is supported on $[0,\infty)$ (so $\sqrt{\lambda}$) makes sense. For an alternate, more elementary proof using the power series of the square root around 1, see Reed and Simon 1980.

Given any operator $A \in \mathcal{B}(\mathcal{H})$, the operator $A^\dagger A \geq 0$, since $\left\langle \psi | A^\dagger A \psi \right\rangle = \|A\psi\|^2$. We define its *absolute value* by

$$|A| \;=\; \sqrt{A^\dagger A}\,, \tag{5.4}$$

where $\sqrt{A^\dagger A}$ is the operator whose existence is given by the square root lemma. This allows us to define the *polar decomposition*

Theorem 5.5 (Polar decomposition — see, e.g., Reed and Simon 1980, Theorem VI.10)**.** *Let $A \in \mathcal{B}(\mathcal{H})$. Then there is a unique* partial isometry *U such that $A = U|A|$.*

Remark. A *partial isometry* U is a map such that the restriction of U to $(\ker U)^\perp$ is an isometry (unitary map) from this space to $\operatorname{ran} U$. U is a partial isometry if and only if U^*U and UU^* are both orthogonal projections, onto $(\ker U)^\perp$ and $\operatorname{ran} U$, respectively. A unitary map is a partial isometry with $\ker U = \{0\}$ and $\operatorname{ran} U = \mathcal{H}$.

Sketch of proof. One simply defines $U : \operatorname{ran}|A| \to \operatorname{ran} A$ as follows. Given $\psi \in \operatorname{ran}|A|$ we have $\psi = |A|\phi$ for some ϕ. Let $U\psi = A\phi$. One has to check that this is well defined. Once this is done, we have

$$\|U\psi\|^2 \;=\; \|A\phi\|^2 \;=\; \left\langle \phi | A^\dagger A \phi \right\rangle \;=\; \langle |A|\phi | |A|\phi \rangle \;=\; \|\psi\|^2\,,$$

so U is an isometry on $\operatorname{ran}|A|$. Extend U by defining it to be zero on $\ker|A| = (\operatorname{ran}|A|)^\perp$.

□

Using the polar decomposition, we may show

Proposition 5.6. *A is compact if and only $|A|$ is compact.*

Proof. One has $A = U|A|$, from the polar decomposition, as well as $|A| = U^\dagger|A|$, since the constructed U has $U^\dagger U =$ projection onto the range of $|A|$. Thus the proposition follows from the ideal lemma 5.2

□

Finally, we have

Theorem 5.7 (Spectral theorem for compact self-adjoint operators)**.** *If $A \in \mathcal{B}(\mathcal{H}))$ is compact and self-adjoint, then A has an orthonormal basis of eigenvectors $(\psi_j)_{j=1}^{\infty}$ with eigenvalues $(\lambda_j)_{j=1}^{\infty}$ that converge to zero as $j \to \infty$. Conversely, if A is self-adjoint and has an orthonormal basis of eigenvectors, with eigenvalues that converge to zero, then A is compact.*

Sketch of proof. The backwards direction follows from Thm. 5.1, since if $A = \sum_j \lambda_j \left|\psi_j\right\rangle \left\langle\psi_j\right|$ with $\lambda_j \to 0$, then

$$\left\| A - \sum_{j=1}^{N} \lambda_j \left|\psi_j\right\rangle \left\langle\psi_j\right| \right\| \;\leq\; \sup_{j>N} |\lambda_j| \,,$$

so A is a norm limit of finite rank operators, hence compact.

The forward direction can be proved almost as in the finite dimensional case. However, we cannot easily use the determinant to construct eigenvectors. Instead we will look for a vectors $\psi \in B_1(0)$ that maximize $|\langle\psi|A\psi\rangle|$. We will need the following lemma

Lemma 5.8. *Let A be a compact operator on $\mathcal{H}$. Then there is a unit vector ψ_0 such that*

$$|\langle\psi_0|A\psi_0\rangle| \;=\; \sup_{\|\psi\|=1} |\langle\psi|A\psi\rangle| \,.$$

To prove the lemma, let $(\psi_j)_{j=1}^{\infty}$ be a sequence of unit vectors such that $\left|\left\langle\psi_j\right|A\psi_j\right\rangle\right|$ converges to the supremum. By the Banach Alaoglu Theorem (see, e.g., Reed and Simon 1980, Thm. IV.21), by passing to a subsequence we may assume that $(\psi_j)_{j=1}^{\infty}$ *converges weakly* to some ψ_0, i.e., $\lim_j \left\langle\phi\right|\left.\psi_j\right\rangle = \langle\phi|\psi_0\rangle$. Since A is compact, it maps the weakly convergent sequence $(\psi_j)_{j=1}^{\infty}$ to a norm-convergent sequence (see Reed and Simon 1980, Theorem VI.11). Thus $\lim_j \left\|A\psi_j - A\psi_0\right\| = 0$. Thus

$$\left\langle\psi_j\right|A\psi_j\left.\right\rangle \;=\; \left\langle\psi_j\right|A(\psi_j - \psi_0)\left.\right\rangle + \left\langle\psi_j\right|A\psi_0\left.\right\rangle \;,$$

where the first term converges to zero since $\left\|\psi_j\right\| \leq 1$ and the second term converges to $\langle\psi_0|A\psi_0\rangle$ because $(\psi_j)_{j=1}^{\infty}$ converges weakly.

Returning to the spectral theorem, we define the vectors using a *min-max* principle. By the lemma there is a unit vector ψ_1 such that

$$\sigma_1 \;:=\; \sup_{\|\psi\|=1} |\langle\psi|A\psi\rangle| \;=\; |\langle\psi_1|A\psi_1\rangle| \,.$$

Given $\psi_1, \ldots, \psi_{j-1}$ and $\sigma_1, \ldots, \sigma_{j-1}$ define ψ_j to be a unit vector perpendicular to $\{\psi, \ldots, \psi_{j-1}\}$ such that

$$\sigma_j \;:=\; \sup_{\psi \perp \{\psi, \ldots, \psi_{j-1}\},\ \|\psi\|=1} |\langle\psi|A\psi\rangle| \;=\; \left|\left\langle\psi_j\right|A\psi_j\left.\right\rangle\right| \,.$$

Note that $\sigma_1 \geq \sigma_2 \geq \sigma_3 \cdots$ and the existence of ψ_j is guaranteed by the lemma. By construction, the sequence $(\phi_j)_{j=1}^{\infty}$ is ortho-normal and, hence, converges weakly to zero. Thus $(A\psi_j)_{j=1}^{\infty}$ converges to zero in norm and $\lim_j \sigma_j = \lim_j \left\langle\psi_j\right|A\psi_j\left.\right\rangle = 0$.

For each j, let $\lambda_j = \langle \psi_j | A\psi_j \rangle$, so $\lambda_j = \pm\sigma_j$. If $\lambda_j = \sigma_j$ then

$$\langle \psi_j | A\psi_j \rangle \;=\; \max_{\psi \perp \{\psi,\ldots,\psi_{j-1}\},\ \|\psi\|=1} \langle \psi | A\psi \rangle \ .$$

Using the method of Lagrange multipliers, and taking small variations of ψ_j within the subspace $\{\psi, \ldots, \psi_{j-1}\}^\perp$, one can verify that $A\psi_j = \lambda_j \psi_j$. Similarly, if $\lambda_j = -\sigma_j$ then

$$\langle \psi_j | A\psi_j \rangle \;=\; \min_{\psi \perp \{\psi,\ldots,\psi_{j-1}\},\ \|\psi\|=1} \langle \psi | A\psi \rangle$$

and $A\psi_j = \lambda_j \psi_j$.

Let $\mathcal{H}^\perp = \{\psi_1, \psi_2, \ldots\}^\perp$. Since $\sigma_j \to 0$, it follows that $A\psi = 0$ for all $\psi \in \mathcal{H}^\perp$. It follows that

$$A \;=\; \sum_{j=1}^{\infty} \lambda_j \, |\psi_j\rangle \langle \psi_j| \ ,$$

as claimed.

□

5.2 Trace Class Operators

Let $(\phi_j)_{j=1}^\infty$ be an orthonormal basis for $\mathcal{H}$. For a *positive* operator A, define $\operatorname{tr} A = \sum_{j=1}^\infty \langle \phi_j | A\phi_j \rangle$. Then one has (see Reed and Simon 1980, Theorem VI.18):

1. $\operatorname{tr}(A + B) = \operatorname{tr} A + \operatorname{tr} B$, for $A, B \geq 0$;
2. $\operatorname{tr}(\lambda) = \lambda \operatorname{tr} A$, for $\lambda \geq 0$ and $A \geq 0$);
3. The value $\operatorname{tr} A$ for $A \geq 0$ does not depend on the choice of orthonormal basis and $\operatorname{tr}(UAU^{-1}) = \operatorname{tr} A$ for any unitary U and $A \geq 0$;
4. If $0 \leq A \leq B$, then $\operatorname{tr} A \leq \operatorname{tr} B$.

Definition 5.9. An operator $A \in \mathcal{B}(\mathcal{H})$ is *trace class* if $\operatorname{tr}|A| < \infty$. The family of trace class operators is denoted by $\mathcal{I}_1$.

Theorem 5.10 (see Reed and Simon 1980, Theorem VI.19)**.** *$\mathcal{I}_1$ is a $*$-ideal in $\mathcal{B}(\mathcal{H})$, that is*

1. *$\mathcal{I}_1$ is a vector space;*
2. *If $A \in \mathcal{I}_1$ and $B \in \mathcal{B}(\mathcal{H})$, then AB and BA are in $\mathcal{I}_1$; and*
3. *If $A \in \mathcal{I}_1$, then $A^* \in \mathcal{I}_1$.*

We define the *trace norm* by $\|A\|_1 = \operatorname{tr}|a|$.

Theorem 5.11 (see Reed and Simon 1980, Theorem VI.20)**.** *$\mathcal{I}_1$ is a Banach space under the norm $\|\cdot\|_1$ and $\|A\| \leq \|A\|_1$ for all $A \in \mathcal{I}_1$.*

Theorem 5.12 (see Reed and Simon 1980, Theorem VI.21)**.** *Every $A \in \mathcal{I}_1$ is compact. A compact operator A is trace class if and only if the non-zero eigenvalues $\sigma_1 \leq \sigma_2 \leq \cdots$ of $|A|$ (called the* singular values of A*) satisfy $\sum_{j=1}^{\infty} \sigma_j < \infty$.*

Finally, we can introduce the trace for non-positive operators

Theorem 5.13. *If $A \in \mathcal{I}_1$ and $(\phi_n)_{n=1}^{\infty}$ is an orthonormal basis, then $\sum_{j=1}^{\infty} \langle \phi_j | A\phi_j \rangle$ converges absolutely and the limit is independent of the choice of basis.*

Sketch of proof. By the polar decomposition, square root lemma, and Cauchy-Schwarz inequalty

$$\left|\langle \phi_j | A\phi_j \rangle\right| = \left|\left\langle |A|^{\frac{1}{2}} U^* \phi_j \middle| |A|^{\frac{1}{2}} \phi_j \right\rangle\right| \leq \left\| |A|^{\frac{1}{2}} U^* \phi_j \right\| \left\| |A|^{\frac{1}{2}} \phi_j \right\| = \left(\langle \phi_j | U|A|U^* \phi_j \rangle \langle \phi_j | |A| \phi_j \rangle \right)^{\frac{1}{2}} .$$

By another application of Cauchy-Schwarz to the sum of j, we find that

$$\sum_{j=1}^{\infty} \left|\langle \phi_j | A\phi_j \rangle\right| \leq \left(\operatorname{tr} U|A|U^* \operatorname{tr} |A|\right)^{\frac{1}{2}} \leq \operatorname{tr} |A| .$$

The proof that the result is independent of the basis follows the finite dimensional case once absolute convergence has been shown.

□

Definition 5.14. We define the map $\operatorname{tr} : \mathcal{I}_1 \to \mathbb{C}$ by $\operatorname{tr} A = \sum_j \langle \phi_j | A\phi_j \rangle$ for any orthonormal basis

Warning: It can happen that $\sum_{j=1}^{\infty} \left|\langle \phi_j | A\phi_j \rangle\right| < \infty$ for some orthonormal basis although the operator is not trace class. For example if

$$A = \begin{bmatrix} 0 & 1 & 0 & & & \cdots \\ 1 & 0 & 1 & 0 & & \\ 0 & 1 & 0 & 1 & 0 & \\ & 0 & 1 & 0 & 1 & \\ \vdots & & & \ddots & \ddots & \ddots \end{bmatrix}$$

then A defines a self-adjoint bounded operator on ℓ^2 that is not even compact, much less trace class.

Lecture 6

States on the infinite dimensional model of quantum mechanics

6.1 Defining the states

We wish to define a state ω of our quantum system with algebra of observables $\mathfrak{A}_\infty$ to be a map $\omega : \mathfrak{A}_\infty \to \mathbb{R}$ satisfying the following:

1. $A \mapsto \omega(A)$ is a linear map;
2. $\omega(A^2) \geq 0$ for any observable A; and
3. $\omega(I) = 1$, for the identity observable I.

Since any observable A satisfies $-\|A\| \leq A \leq \|A\| I$, it follows from the above that

$$-\|A\| \ \leq \omega(A) \ \leq \ \|A\| \, . \tag{6.1}$$

We can and will extend a state ω to act on the algebra $\mathcal{B}(\mathcal{H})$ of all bounded operators by taking

$$\omega(A) \ := \ \frac{1}{2}\omega((A + A^\dagger)) + \frac{i}{2}\omega(\frac{1}{i}(A - A^\dagger)) \, , \tag{6.2}$$

since $\frac{1}{2}(A + A^\dagger)$ and $\frac{1}{2i}(A - A^\dagger)$ are both self-adjoint.

Exercise 6.1. Show that if $\omega : \mathfrak{A}_\infty \to \mathbb{R}$ satisfies (1-3) above, then the extended functional, defined in (6.2), is a complex linear map from $\mathcal{B}(\mathcal{H}) \to \mathbb{C}$ that satisfies $\omega(A^\dagger A) \geq 0$ for all $A \in \mathcal{B}(\mathcal{H})$ and $\omega(I) = 1$.

We claim that the extension of ω to $\mathcal{B}(\mathcal{H})$ is a *bounded linear functional of norm one*, i.e.,

$$\|\omega\| \ := \ \sup_{A \neq 0} \frac{|\omega(A)|}{\|A\|} \ = \ 1 \, . \tag{6.3}$$

To see this, suppose that $A \in \mathcal{B}(\mathcal{H})$ and let $\theta \in \mathbb{R}$ such that $\mathrm{e}^{-i\theta}\omega(A) \ = \ |\omega(A)|$. Since the extension is complex linear,

$$|\omega(A)| \ = \ \Re\left[\mathrm{e}^{-i\theta}\omega(A)\right] \ = \ \Re\omega(\mathrm{e}^{-i\theta}A) \ = \ \frac{1}{2}\omega\left(\mathrm{e}^{-i\theta}A + \mathrm{e}^{i\theta}A^\dagger\right)$$
$$\leq \ \left\|\mathrm{e}^{-i\theta}A + \mathrm{e}^{i\theta}A^\dagger\right\| \ \leq \ \|A\|,$$

where in the second to last step we have applied (6.1) to the observable $e^{-i\theta}A + e^{i\theta}A^\dagger$. By (6.3), a state is necessarily continuous *with respect to the operator norm on* $\mathcal{B}(\mathcal{H})$.

However, the continuity given by (6.3) is not enough for our purposes — we need a stronger property. To motivate this recall that we want a state to define a probability measure for each observable A and that every observable A gives rise to an increasing projection valued function $P_A(\lambda)$ from the spectral theorem. In a standard abuse of notation, we will use P_A to also denote the projection valued measure

$$P_A(E) \;=\; \int_E dP_A(\lambda)\;, \tag{6.4}$$

for Borel sets E, where as usual the equation should be understood as an identity for matrix elements. Before proceeding, let us consider a typical example:

Exercise 6.2. Consider the Hilbert space $L^2(0,1)$ and let A be the observable $Af(x) = xf(x)$. Show that the projection valued measure associated to A is the following

$$P_A(E)f(x) \;=\; \chi_E(x)f(x)\;,$$

where $\chi_E(x)$ is the characteristic function of E. Thus if $F \subset E$ then $P_A(E)f - P_A(F)f = \chi_{E\setminus F}f$. Show that this implies

$$\|P_A(E) - P_A(F)\| \;=\; 1$$

whenever the Lebesgue measure of $E \setminus F$ is positive. In particular, if E is a disjoint union $\cup_j E_j$ then the series $P_A(E) = \sum_j P_A(E_j)$ *does not converge in norm*. However, if $f \in L^2$ is any vector then the series $\sum_j P_A(E_j)f$ *does converge in* L^2 to $P_A(E)f$.

What the above exercise shows is that we cannot expect the projection valued measure P_A to be countable additive in norm. However, the Spectral Theorem 4.2 guarantees that it is *strongly* countably additive, meaning that if $(E_j)_{j=1}^\infty$ are disjoint then

$$P_A(\cup_j E_j)\psi \;=\; \sum_j P_A(E_j)\psi \tag{6.5}$$

for any vector $\psi \in \mathcal{H}$. Eq.4.2 is (6.5) is countable additivity in the *strong operator topology*, which is the weakest topology on $\mathcal{B}(\mathcal{H})$ such that the maps $A \mapsto \|A\psi\|$ are continuous for each $\psi \in \mathcal{H}$.

Exercise 6.3. The Spectral Theorem 4.2 as stated above gives countable additivity in the weak operator topology, which is

$$\langle\phi|\, P_A(\cup_j E_j)\,|\psi\rangle \;=\; \sum_j \langle\phi|\, P_A(E_j)\,|\psi\rangle\;.$$

Show that this implies eq. (6.5). Hint: use the fact that $\|P_A(E)\psi\|^2 = \langle\psi|\, P_A(E)\,|\psi\rangle$, since $P_A(E)$ is a projection.

Given an observable A, the map $\mu_{\omega,A}(E) = \omega(P_A(E))$ should define the probability that a measure of A in state ω yields an outcome in the Borel set $E \subset \mathbb{R}$. With the assumptions (1-3) outlined above, we can prove that this map is finitely additive, but we cannot show that it is countably additive. To obtain countable additivity, we need another condition, which is usually phrased in terms of *monotone sequences of positive operators*:

Definition 6.1. A *monotone sequence of positive operators* is a sequence $(A_j)_{j=1}^{\infty}$ in $\mathcal{B}(\mathcal{H})$ that satisfies $0 \leq A_1 \leq A_2 \leq \cdots$. The sequence is *bounded* if $\sup_j \|A_j\| < \infty$.

Lemma 6.2. *If $(A_j)_{j=1}^{\infty}$ be a bounded monotone sequence of positive operators, then there is a positive operator $A \in \mathcal{B}(\mathcal{H})$ such that $A\psi = \lim_j A_j\psi$ for any $\psi \in \mathcal{H}$.*

Sketch of Proof. Note that $0 \leq A_j \leq \sup_j \|A_j\| I$ for each j. It follows that $0 \leq \langle\psi| A_1 |\psi\rangle \leq \langle\psi| A_2 |\psi\rangle \leq \cdots$ is a bounded sequence of real numbers for each $\psi \in \mathcal{H}$. Let $q(\psi)$ denote its limit; this is a bounded quadratic form on $\mathcal{H}$ and hence is given by $q(\psi) = \langle\psi| A |\psi\rangle$ with A a bounded self-adjoint operator (see Cor. A.2).

□

ypically phrased as follows:

4. If $0 \leq A_1 \leq A_2 \leq \cdots$ is a monotone sequence of positive operators with $\sup_j \|A\|_j < \infty$, then $\lim_j \omega(A_j) = \omega(A)$ where $A = \sup_j A_j$ is the operator defined by

$$\langle\psi| A |\psi\rangle \;=\; \lim_{j\to\infty} \langle\psi| A_j |\psi\rangle \;.$$

Note that if $0 \leq A_1 \leq A_2 \leq \cdots$ is monotone sequence of positive operators, then
A linear functional on $\mathcal{B}(\mathcal{H})$ is called:

1. *positive* if $\omega(A^\dagger A) \geq 0$ for any A, and

2. *normalized* if $\omega(I) = 1$.

A positive linear functional is called *normal* if it satisfies (4) above. A *state* on $\mathcal{B}(\mathcal{H})$ is a positive, normalized, normal linear functional.[1]

Now we are ready for the big theorem

Theorem 6.3. *Let ω be a state on $\mathcal{B}(\mathcal{H})$. Then there is a positive trace class operator ρ such that $\omega(A) = \operatorname{tr} \rho A$.*

Proof. Let $\phi \in \mathcal{H}$ and consider the operator $|\phi\rangle\langle\phi|$. Define $q(\phi) = \omega(|\phi\rangle\langle\phi|)$. The map $\phi \mapsto q(\phi)$ is a quadratic form on $\mathcal{H}$, meaning that it satisfies the parallelogram law and is homogeneous of order 2 under scaling (see Appendix A). Furthermore q is bounded, since $0 \leq q(\psi) \leq 1$. If follows from the Riesz representation theorem (see Cor. A.2) that there is a self-adjoint operator ρ such that $q(\psi) = \langle\psi| \rho |\psi\rangle$. Since $q(\psi) \geq 0$, ρ is seen to be positive. Finally, if $(\psi_j)_{j=1}^{\infty}$ is an orthonormal basis, then $Q_n = \sum_{j=1}^{n} |\psi_j\rangle\langle\psi_j|$ is a monotone sequence of positive operators that converge to the identity. Thus

$$\sum_{j=1}^{\infty} \langle\psi_j| \rho |\psi_j\rangle \;=\; \lim_{n\to\infty} \sum_{j=1}^{n} \langle\psi_j| \rho |\psi_j\rangle \;=\; \lim_{n\to\infty} \omega(Q_n) \;=\; \omega(I) = 1 \,,$$

[1] In other contexts, one introduces "states" that are positive, normalized linear functionals and refers the objects defined here as "normal states." However, for a specific model of quantum mechanics we really need a normal state to obtain a the probability distributions for observables.

so ρ is trace class and has $\operatorname{tr}\rho = 1$.

It remains to show that $\omega(A) = \operatorname{tr}\rho A$ for any A. First note that by the polarization identity we have

$$\omega(|\psi\rangle\langle\phi|) = \operatorname{tr}\rho|\psi\rangle\langle\phi| = \langle\phi|\rho|\psi\rangle$$

for any rank-one operator $|\psi\rangle\langle\phi|$. For general A, it suffices to prove the identity assuming A to be self-adjoint A, and use (6.2) to extend it to all A. For self adjoint A we have $A + \|A\|I \geq 0$. Thus it suffices to prove the result for non-negative A. If $A \geq 0$, then the operators $A_n = Q_n A Q_n$, with $Q_n = \sum_{j=1}^{n} |\psi_j\rangle\langle\psi_j|$ for some orthonormal basis, are an increasing sequence with $\sup_n A_n = A$. Furthermore A_n is finite rank, so

$$\omega(A_n) = \sum_{i,j=1}^{n} \langle\psi_i|A|\phi_j\rangle\,\omega(|\psi_i\rangle\langle\phi_j|) = \sum_{i,j=1}^{n} \langle\psi_i|A|\phi_j\rangle\langle\phi_j|\rho|\psi_i\rangle = \operatorname{tr}\rho A_n\,.$$

It follows that

$$\omega(A) = \lim_n \omega(A_n) = \lim_n \operatorname{tr}\rho A_n = \operatorname{tr}\rho A\,.$$

□

Lecture 7

Dynamics and undbounded operators

Unfortunately the algebra $\mathfrak{A}_\infty$ is not quite enough to describe everything we need. Heisenberg 1927 famously postulated that the quantum description of a particle should include *position* Q_j and *momentum* P_j operators, $j = 1, 2, 3$, satisfying the *canonical commutation relations*:

$$[Q_j, P_i] \;=\; i\hbar I \,. \tag{7.1}$$

There are, however, no bounded operators even on an infinite dimensional Hilbert space satisfying (7.1).

Exercise 7.1. Show that we cannot have $[Q, P] = icI$ for a pair of bounded opearators Q and P. Hint: show that $[Q, P^n] = icnP^{n-1}$, from which it follows that $2\|Q\|\|P\|^n \geq cn\|P\|^{n-1}$. This leads to a contradiction.

However, one easily notes that for functions on $\mathbb{R}^3$ the operators

$$Q_i f(x) = x_i \quad \text{and} \quad -i\hbar\frac{\partial}{\partial x_i} f(x)$$

satisfy (7.1). However, these expressions are not defined for every $f \in L^2(\mathbb{R}^3)$: for $Q_i f$ to be defined f must decay at infinity fast enough for $\int |x_i|^2 |f(x)|^2 dx < \infty$, while for $P_i f$ to be defined f must have a derivative with respect to x_i in a suitable sense.

7.1 Unbounded Operators

The solution to this difficult to relax, slightly, our notion of an observable.

Definition 7.1. A *partially defined* operator A on $\mathcal{H}$ is a map $A : \mathcal{D}_A \to \mathcal{H}$ defined on a (not necessarily closed) subspace $\mathcal{D}_A \subset \mathcal{H}$. The operator is *densely defined* if $\mathcal{D} \subset \mathcal{H}$ is a dense subspace of $\mathcal{H}$. The *adjoint* of a densely defined operator A is the map $A^\dagger$ defined to be $\langle A^\dagger\psi|\phi\rangle = \langle\psi|A\phi\rangle$ using the Riesz Representation Theorem A.1 on the domain

$$\mathcal{D}_{A^\dagger} \;=\; \{\psi \in \mathcal{H} \,|\, |\langle\psi|A\phi\rangle| \leq c\|\phi\| \quad \text{for some } c < \infty\} \,.$$

A *self-adjoint operator* is a densely defined operator such that $A = A^\dagger$. That is

1. $\langle A\psi|\phi\rangle = \langle\psi|A\phi\rangle$ for all $\phi, \psi \in \mathcal{D}_A$, *and*

2. $\mathcal{D}_{A^\dagger} = \mathcal{D}_A$.

One has

Theorem 7.2 (Spectral Theorem for Self Adjoint Operators)**.** *Let $A : \mathcal{D}_A \to \mathcal{H}$ be a self-adjoint operator. Then there is a map $P_A : \mathbb{R} \to \mathcal{P}(\mathcal{H})$ such that*

1. *$\lambda \mapsto P_A(\lambda)$ is increasing, i.e., if $\lambda \leq \mu$, then $P_A(\lambda) \leq P_A(\mu)$;*

2. *$\lim_{\lambda\to-\infty} P_A(\lambda)\psi = 0$ and $\lim_{\lambda\to\infty} P_A(\lambda)\psi = \psi$ for any $\psi \in \mathcal{H}$,*

3. *$P_A(\lambda)$ is* strongly right continuous*;*

4. *$P_A(\lambda)$ commutes with any bounded operator B such that $[B, A] = 0$; and*

5. *$A = \int_{\mathbb{R}} \lambda dP_A(\lambda)$ in the sense that if $\psi \in \mathcal{D}_A$ then*

$$\langle\phi| A |\psi\rangle \;=\; \int_{\mathbb{R}} \lambda d \langle\phi| P_A(\lambda) |\psi\rangle \tag{7.2}$$

for any $\phi, \psi \in \mathcal{H}$.

7.2 Dynamics and Stone's Theorem

Based on our experience with the finite dimensional model, it is natural to think that the dynamics of an isolated system in the infinite dimensional model should be given by the formal Liouville-Schrödinger equation

$$\partial_t \rho_t \;=\; -i[H, \rho_t]\,, \tag{7.3}$$

in units with $\hbar = 1$. Here H should be the Hamiltonian. A prototype for the sort of Hamiltonian that we would like to be able to handle is the harmonic oscillator $H = \frac{1}{2}\left(P^2 + Q^2\right)$ with $[Q, P] \;=\; i$. So it is clear we do not expect the Hamiltonian to be bounded, and (7.3) requires interpretation.

If H were bounded, then we could solve (7.3) as follows. Let $\rho_t \;=\; U_t \rho (U_t)^\dagger$, where $U_t \;=\; \mathrm{e}^{-itH}$, defined via the power series or by the spectral theorem. One checks (for H bounded) that

$$\frac{d}{dt}\mathrm{e}^{-itH} \;=\; -iH\mathrm{e}^{-itH} \;=\; -i\mathrm{e}^{-itH}H.$$

Here the derivative is taken with respect to operator norm convergence, that is we have

$$\lim_{\delta\to 0} \frac{\left\|\mathrm{e}^{-i(t+\delta)H} - \mathrm{e}^{-itH} + i\delta H\mathrm{e}^{-itH}\right\|}{|\delta|} \;=\; 0\,.$$

Furthermore $U_t = \mathrm{e}^{-itH}$ is a *norm continuous, one parameter group*, meaning

1. $U_{t+s} = U_t U_s$ for all $s, t \in \mathbb{R}$ (one parameter group),

2. $\lim_{t\to s} \|U_t - U_s\| = 0$ (norm continuous).

If H is self-adjoint, then U_t is also unitary, since then $U_t^\dagger = \mathrm{e}^{itH} = U_t^{-1}$, by the group property.

Theorem 7.3 (Norm continuous one parameter groups)**.** *Let $S_t : \mathbb{R} \to \mathcal{B}(\mathbf{H})$ be a norm continuous, one parameter group. Then there is an operator $A \in \mathcal{B}(\mathbf{H})$ such that $S_t = \mathrm{e}^{tA}$. The group is unitary if and only if $A = -iH$ with H a self-adjoint operator.*

Sketch of Proof. To construct A, we will use the following

Lemma 7.4. *Let $\mathcal{B}_{\ln 2} = \{A \in \mathcal{B}(\mathcal{H}) \ : \ \|A\| < \ln 2\}$. Then the exponential $\exp(A) = \mathrm{e}^A$, defined by the power series, is one-to-one on $\mathcal{B}_1$ and $\exp(\mathcal{B}_1)$ is an open neighborhood of the identity.*

To prove the lemma, define the logarithm as a power series on $\mathcal{U} := \{B \ : \ \|B - I\| < 1\}$ and show that $\exp \circ \log = \log \circ \exp$.

Returning to the proof of the theorem, note that by continuity of S_t there is $\epsilon > 0$ such that $S_t \in \exp(\mathcal{B}_{\ln 2})$ for $|t| < \epsilon$. Fix $0 < t_0 < \epsilon$ and let

$$A = \frac{1}{t_0} \log\left(S_{t_0}\right) .$$

Using the group property for S one shows that the identity $S_t = \mathrm{e}^{tA}$ holds for dyadic rational multiples of t_0. Then by continuity the identity extends to all t. The fact that S_t is unitary if and only if $A = -iH$ with H self adjoint follows from the discussion above the statement of the theorem.

□

Based on the above we see that we do not expect the dynamics U_t of a quantum system to be norm-continuous. However there is a generalization of the above result to *strongly continuous* one parameter groups, which satisfy

(2′) $\lim_{t\to s} U_t\phi = U_s\phi$ for all $s \in \mathbb{R}$ and $\phi \in \mathcal{H}$.

There is an imporant generalization of the above theorem due to Marshall Stone:

Theorem 7.5 (Stone 1932)**.** *Let U_t be a strongly continuous one-parameter unitary group. Then there is a self adjoint operator $A : \mathcal{D}_A \to \mathcal{H}$ such that*

$$U_t \ = \ \int_{\mathbb{R}} \mathrm{e}^{-it\lambda} dP_A(\lambda) \ =: \ \mathrm{e}^{-itA} \, , \tag{7.4}$$

where P_A is the projection valued measure for A. The domain $\mathcal{D}_A$ is given by

$$\mathcal{D}_A \ := \ \left\{ \psi \in \mathcal{H} \ \middle| \ \lim_{\epsilon\to 0} \frac{1}{\epsilon} \left(U_\epsilon \psi - \psi\right) \ \textit{exists} \right\} \, .$$

Conversely, any self-adjoint operator A defines a strongly continuous one-parameter group via (7.4)

Sketch of Proof. We define

$$A\psi \;=\; i\lim_{\epsilon\to 0}\frac{1}{\varepsilon}\,(U_\varepsilon\psi-\psi)$$

on $\mathcal{D}_A$.

To show that $\mathcal{D}_A$ is a dense subspace one notes that for any vector of the form

$$\phi_f \;=\; \int_{\mathbb{R}} f(t)U_t\phi dt$$

with f a Schwarz function[1] is in $\mathcal{D}_A$. Indeed, one has

$$U_\epsilon\phi_f \;=\; \int_{\mathbb{R}} f(t)U_{t+\epsilon}\phi dt \;=\; \int_{\mathbb{R}} f(t-\epsilon)U_t\phi dt\,,$$

from which it follows that $\phi_f\in\mathcal{D}_A$ and that

$$A\phi_f \;=\; -i\int_{\mathbb{R}} f'(t)U_t\phi \;=\; -i\phi_{f'}\,.$$

The set of vectors of the form ϕ_f is easily seen to be dense, using a sequence of Schwarz functions converging to the Dirac measure at 0 and the strong continuity of U_t.

The symmetry $\langle\psi|A\phi\rangle=\langle A\psi|\phi\rangle$ is a consequence of the unitarity of U_t. To prove self-adjointess one must show more: namely that the domain of $A^\dagger$ is also equal to $\mathcal{D}_A$. For details, see the §7.3 of Simon 2015a.

□

The upshot is that if H is a self-adjoint operator on $\mathcal{H}$ we can interpret (7.3) as having solutions $\rho_t = \mathrm{e}^{-itH}\rho_0\mathrm{e}^{itH}$ where e^{-itH} is a strongly continuous one parameter unitary group on $\mathcal{H}$. As we will see later there is a more general form of the evolution that can be introduced on density matrices, called the *Lindblad equation* that can incorporate certain effects arising from the interaction of a quantum system with its environment.

7.3 The Weyl form of the CCR and the Stone-von Neumann Theorm

Let us return to the canonical commutation relations (7.1), which for a single conjugate pair Q and P read

$$[Q,P] \;=\; i\hbar I \tag{7.5}$$

As indicated above, this equation has no bounded solutions. Furthermore, it is slightly tricky to formulate in the unbounded setting because of domain issues. However, the lesson from Stone's theorem is that every self-adjoint operator is associated to a unitary group, so it might be better to rephrase the equation in terms of the unitary groups $U(t)=\mathrm{e}^{itQ}$ and $V(t)\mathrm{e}^{itP}$. This obervation was originally made by Weyl 1927, who by a formal calculation suggested the following formulation

$$U(t)V(s) \;=\; \mathrm{e}^{-i\hbar st}V(s)U(t)\,, \tag{7.6}$$

[1] So f is smooth and it and all its derivatives vanish faster than any power.

for all $s, t \in \mathbb{R}$. This identity is known as the *Weyl form of the CCR*, or the *Weyl relations*.

If we represent Q as multiplication by the independent variable on $L^2(\mathbb{R})$, $Q\psi(x) = x\psi(x)$, then $U(t) = \mathrm{e}^{itx}\psi(x)$. In this case the Weyl relations are satisfied by taking $V(t)$ to be the *translation group* $V(t)\psi(x) = \psi(x + \hbar t)$. The generator of the translation group is the derivative, so we have $P\psi(x) = -i\hbar\partial_x\psi(x)$. The *Stone von Neumann* theorem states that this is essentially the only representation.

Theorem 7.6 (von Neumann 1931 and Stone 1930)**.** *Let $\mathcal{H}$ be a Hilbert space and let $U(t)$ and $V(s)$ be strongly continuous, one parameter unitary groups such that the Weyl relations* (7.6) *hold. Then there is a finite or countable direct sum decomposition $\mathcal{H} = \oplus_{j=1}^{N}\mathcal{H}_j$ such that for each $\mathcal{H}_j$ there is a unitary map $W_j : L^2(\mathbb{R}) \to \mathcal{H}_j$ for which*

$$U(t)W_j = W_j\mathrm{e}^{itx} \quad \textit{and} \quad V(t)W_j = W_j\mathrm{e}^{\hbar t\partial_x}\,.$$

Lecture 8

Examples

8.1 Spectral representation of the momentum and position operators

Consider $\mathcal{H} = L^2(\mathbb{R})$ with $P = -i\frac{d}{dx}$ and $Q =$ multiplication by x. These are each self adjoint operators with domains

$$\mathcal{D}_P = H^1(\mathbb{R}) = \{f \in \mathcal{H} : f' \in \mathcal{H}\} \quad \text{and} \quad \mathcal{D}_Q = \{f \in \mathcal{H} : xf(x) \in \mathcal{H}\}\,.$$

In the definition of the domain of $\mathcal{D}_P$, the derivative f' denotes the derivative in the sense of distributions, i.e., f' is defined to be the linear functional on the Schwartz class $\mathcal{S}$ given by $\langle f'|\phi\rangle = -\int_{\mathbb{R}} f\phi dx$ for $\phi \in \mathcal{S}$.

The cannonical commutation relation $[Q,P] = -iI$ holds in a weak sense on the intersection of the domains $\mathcal{D}_{\text{CCR}} = \mathcal{D}_P \cap \mathcal{D}_Q$:

$$\langle\phi|\,[Q,P]\,|\psi\rangle = -i\,\langle\phi|\,\psi|\quad\rangle \text{ for } \phi,\psi \in \mathcal{D}_{\text{CCR}}\,. \tag{8.1}$$

Exercise 8.1. Use the weak form of the CCR (8.1) to prove the following form of the Heisenberg uncertainty relations. Show that $\mathcal{D}_{\text{CCR}} = \mathcal{D}_P \cap \mathcal{D}_Q$ is dense in $\mathcal{H}$ and that for $\phi \in \mathcal{D}_{P^2} \cap \mathcal{D}Q^2$ with $\|\phi\| = 1$ one has

$$\Delta_\phi(P)\Delta_\phi(Q) \geq \frac{1}{2}\,,$$

where

$$\Delta_\phi(A) = \|A\phi\|^2 - (\langle\phi|\,A\,|\phi\rangle)^2 \quad A = P, Q\,.$$

Hint: follow the proof of the Heisenberg uncertainty relations Prop. 3.4 and use (8.1).

Remark. Note that formally $\|A\phi\|^2 = \langle\phi|\,A^2\,|\phi\rangle$. However, the operator domain $\mathcal{D}_{A^2}\{\phi \in \mathcal{D}_A : A\phi \in \mathcal{D}_A\}$ is typically smaller than the domain of $\mathcal{D}_A$, so $\|A\phi\|$ is defined for a larger class of vectors.

Exercise 8.2. Show that the projection valued measure for Q is given by

$$P_Q(\lambda)f = \chi_{(-\infty,\lambda)}f$$

where χ_E denotes the characteristic function of a set E and that

$$[F(Q)f](x) \;=\; F(x)f(x)$$

for any bounded, Borel measurable function F.

The spectral representation of P can be found by taking the Fourier transform

$$\hat{f}(k) \;=\; \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{-ikx} f(x) dx \;.$$

Note that

$$\widehat{Pf}(k) \;=\; -i \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{-ikx} f'(x) dx \;=\; k\hat{f}(k)$$

by integration by parts. Thus

$$\widehat{P_P(\lambda)f}(k) \;=\; \chi_{(-\infty,\lambda)}(k)\hat{f}(k) \;.$$

For a Borel function F of P, one has $[F(P)\hat{f}](k) \;=\; F(k)\hat{f}(k)$. One can use the inverse Fourier transform to write the projection valued measure in real space:

$$P_P(\lambda)f(x) \;=\; \frac{1}{2}f(x) + i\frac{1}{2\pi}\mathrm{P.V.} \int_{\mathbb{R}} \frac{1}{x'-x} \mathrm{e}^{-i(x'-x)\lambda} f(x') dx' \;,$$

where "P.V. $\int_{\mathbb{R}}$" denotes the principle value integral $\lim_{\epsilon \downarrow 0} \int_{|x-x'|>\epsilon}$.

8.2 The Free Hamiltonian

Consider $H \;=\; \frac{1}{2}P^2$. This is the Hamiltonian describing the motion of a free (one dimensional) particle of mass 1 (with mass m we would have $\frac{1}{2m}P^2$).

The Heisenberg relations for H are easy to solve and correspond exactly to the classical equations of motion. Note that (formally)

$$[H,P] = 0 \quad \text{and} \quad [H,Q] \;=\; \frac{1}{2}(P[P,Q] + [P,Q]P) \;=\; -iP \;.$$

Thus the Heisenberg equations are just

$$\dot{P} \;=\; 0 \quad \text{and} \quad \dot{Q} \;=\; P \;,$$

which are easily solved to give

$$P(t) \;=\; P \quad \text{and} \quad Q(t) \;=\; Q + tP \;.$$

As a result, we see that the mean values $p(t) \;=\; \mathrm{tr}\, P(t)\rho$ and $q(t) \;=\; \mathrm{tr}\, Q(t)\rho$ satisfy the classical equations

$$p(t) = p(0) \quad \text{and} \quad q(t) \;=\; q(0) + tp(0)$$

for any initial state. Furthermore, the uncertainty in P remains unchanged for all time $\Delta_\rho(P(t)) = \Delta_\rho(P)$. However, the uncertainty in position satisfies (operating formally, assuming that all traces are finite)

$$\begin{aligned}\Delta_\rho^2(Q(t)) &= \operatorname{tr}\rho(Q+tP)^2 - (\operatorname{tr}\rho(Q+tP))^2 \\ &= \operatorname{tr}Q^2 + t\operatorname{tr}\rho(QP+PQ) + t^2\operatorname{tr}P^2 - (\operatorname{tr}\rho P)^2 - 2t\operatorname{tr}\rho P\operatorname{tr}\rho Q - t^2\operatorname{tr}\rho P^2 \\ &= \Delta_\rho^2(Q) + t^2\Delta_\rho^2(P) + t\left[\operatorname{tr}\rho(QP+PQ) - 2\operatorname{tr}\rho P\operatorname{tr}\rho Q\right] .\end{aligned}$$

We see that the position uncertainty grows linearly with time as $t \to \infty$.

One can solve the Schrödinger evolution using the Fourier transform. One has

$$\widehat{\mathrm{e}^{-itH}\psi}(k) = \mathrm{e}^{-it\frac{k^2}{2}}\widehat{\psi}(k) ,$$

which gives via the inverse Fourier transform

$$\mathrm{e}^{-itH}\psi(x) = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} \mathrm{e}^{ixk}\mathrm{e}^{-it\frac{k^2}{2}}\widehat{\psi}(k) .$$

Based on this equation, one sees that formally $\mathrm{e}^{-itH}\psi(x) = \int_{\mathbb{R}} W_t(x,y)\psi(y)$, where the "propagator" is given by the oscillatory integral

$$W_t(x,y) = \frac{1}{2\pi}\int_{\mathbb{R}} \mathrm{e}^{i(x-y)k - it\frac{k^2}{2}} .$$

Because this is a Gaussian integral (with a quadratic exponent), one can compute the integral exactly and obtain the identity

$$W_t(x,y) = \frac{\mathrm{e}^{-i\frac{\pi}{4}}}{\sqrt{2\pi t}}\mathrm{e}^{i\frac{(x-y)^2}{2t}} .$$

Although the above derivation was formal, for a function $\psi \in L^1 \cap L^2$ the identity

$$\mathrm{e}^{-itH}\psi(x) = \frac{\mathrm{e}^{-i\frac{\pi}{4}}}{\sqrt{2\pi t}}\int_{\mathbb{R}} \mathrm{e}^{i\frac{(x-y)^2}{2t}}\psi(y)dy$$

does in fact hold.

8.3 Particle in a potential

A general (one dimensional) particle in a potential has Hamiltonian $H = P^2 + V(Q)$. To develop an idea of what can happen it is useful to work the following exercise:

Exercise 8.3. Let $V(x) = -u\chi_{(-1,1)}(x)$ with $u > 0$ a real number. Show that

1. There is a finite number of normalized L^2 solutions to the eigenvalue equation $H\psi = E\psi$ with energies $E < 0$. Because V vanishes outside $(-1,1)$ these solutions decay exponentially at $\pm\infty$.

2. For $E \geq 0$ any solution to $H\psi = E\psi$ does not decay at ∞. For each $E \geq 0$ show that there is two dimensional vector space of solutions.

The Hamiltonian H has a projection valued measure with two parts: a discrete part with jumps at the negative eigenvalues and a positive part which is a sort of generalization of the Fourier transformed expression for the projection valued measure of the free particle.

The "particle in a box" mentioned above can be studied exactly. For other potentials that decay at infinity (or are compactly supported) the above qualitative behavior is correct, although exact expressions are not usually avaiable. In higher dimensions, the nature of the negative energy bound states and the positive energy "scattering states" can be a bit more complicated. In particular, in three dimensions it is possible for a small negative potential to fail to have any negative energy bound states.

8.4 The Harmonic Oscillator

A potential that grows as $x \to \pm\infty$ can lead to all states being bound states. The prototypical model for this situation is the harmonic oscillator $H = \frac{1}{2}(P^2 + Q^2)$ on $L^2(\mathbb{R})$. Note that

$$H\mathrm{e}^{-\frac{1}{2}x^2} \;=\; \tfrac{1}{2}\mathrm{e}^{-\frac{1}{2}x^2} \;.$$

Thus $\frac{1}{2}$ is an eigenvalue of H and the corresponding eigenfunction is $\psi_0(x) = \frac{1}{\pi^{1/4}}\mathrm{e}^{-\frac{1}{2}x^2}$. This is the *ground state* of the Harmonic oscillator.

The standard trick for computing the spectrum of the harmonic oscillator is to define the *raising and lowering operators*

$$a \;=\; \frac{1}{\sqrt{2}}(iP + Q) \quad \text{and} \quad a^\dagger = \frac{1}{\sqrt{2}}(-iP + Q) \;.$$

These are two densely defined, but not self-adjoint operators. To avoid domain questions, let us take them to be defined on $\mathcal{S} = \{\text{Schwarz functions}\}$. Then for $f \in \mathcal{S}$, we have $af, a^\dagger f \in \mathcal{S}$,

$$[a, a^\dagger]f \;=\; aa^\dagger f - a^\dagger a f \;=\; f, \tag{8.2}$$

and

$$Hf \;=\; (a^\dagger a + \tfrac{1}{2})f \;.$$

Note that the ground state ψ_0 satisfies $a\psi_0 = 0$, and is the unique function (up to a scalar multiple) that does so. By (8.2), and the fact that $\mathcal{S}$ is preserved by a and $a^\dagger$, we have

$$\begin{aligned} a(a^\dagger)^n\psi_0 \;&=\; [a, (a^\dagger)^n]\psi_0 \\ &=\; \left\{[a, a^\dagger](a^\dagger)^{n-1} + a^\dagger[a, a^\dagger](a^\dagger)^{n-2} + \cdots + (a^\dagger)^{n-1}[a, a^\dagger]\right\}\psi_0 \;=\; n(a^\dagger)^{n-1}\psi_0 \;. \end{aligned}$$

Thus

$$H(a^\dagger)^n\psi_0 \;=\; (n + \tfrac{1}{2})(a^\dagger)^n\psi_0 \;,$$

i.e., $(a^\dagger)^n\psi_0$ is an eigenvector with eigenvalue $n+\frac{1}{2}$. We can use the commutation relations to compute the norm of this vector, since

$$\left\langle (a^\dagger)^n\psi_0 | (a^\dagger)^n\psi_0 \right\rangle \;=\; \langle \psi_0 |\, (a)^n (a^\dagger)^n \,|\psi_0\rangle \;=\; n!\,\langle \psi_0|\psi_0\rangle \;=\; n!\,,$$

by repeated use of (8.2). Thus $\psi_n = \frac{1}{\sqrt{n!}}(a^\dagger)^n\psi_0$ is a normalized eigenfunction.

So we have shown that H has eigenvectors $(\psi_n)_{n=0}^\infty$ with eigenvalues $n+\frac{1}{2}$. In fact, this is a complete description of the spectrum of H:

Lemma 8.1. *The eigenvectors $(\psi_n)_{n=0}^\infty$ are an orthonormal basis for $L^2(\mathbb{R})$.*

Proof. We chose the vectors to be normalized. That they are orthogonal follows by direct computation using (8.2), or because they are eigenvectors of a self-adjoint operator with distinct eigenvalues.

It remains to show that the eigenvectors span the Hilbert space. Note that $\psi_n(x) = p_n(x)\psi_0(x)$ where $p_n(x)$ is a polynomial of degree n. The first few polynomials are

$$p_0(x) = 1\,, \quad p_1(x) = x\,, \quad p_2(x) = \frac{1}{\sqrt{2}}(x^2-1)\,, \quad \dots\,.$$

In fact, the sequence of polynomials generated this way are the *Hermite polynomials*, one of the classical sequences of orthogonal polynomials. We do not need that fact, however. Instead we need only note that

$$p_n(x) = c_n x^n + \text{lower order terms} \quad \text{with } c_n \neq 0\,.$$

It follows that $\{p_0, \dots, p_n\}$ is a basis for the polynomials of order n and thus that any function of the form $p(x)\psi_0(x)$ with p a polynomial can be written as a finite linear combination of $(\psi_n)_{n=0}^\infty$.

To complete the proof, we will show that

$$\mathcal{P} \;=\; \{p(x)\psi_0(x) \;:\; p \text{ is a polynomial}\}$$

is dense in $L^2(\mathbb{R})$. This is trickier than one might think. The following proof is a modification of one that can be found in §6.4 of Simon 2015b.

First we claim that the functions $f_z(x) = \mathrm{e}^{zx}\psi_0(x)$ for $z \in \mathbb{C}$ are in $\overline{\mathcal{P}}$, the closure of $\mathcal{P}$. To see this note that

$$p_{N,z}(x) \;=\; \sum_{n=0}^{N} \frac{z^n x^n}{n!}\mathrm{e}^{zx} \;\to\; \mathrm{e}^{zx}$$

for every x, and

$$\left|p_{N,z}(x)\right| \;\leq\; \mathrm{e}^{|zx|} \;\leq\; 2\cosh(|z||x|)\,.$$

It follows, by the dominated convergence theorem, that

$$\lim_{N\to\infty} \left\|p_{N,z}\psi_0 - f_z\right\|_{L^2} \;=\; 0\,.$$

Next we note that linear combinations, and hence integrals, of the f_z also lie in $\overline{\mathcal{P}}$. In particular, given a $f \in L^1(\mathbb{R})$ we have

$$\int_{\mathbb{R}} \mathrm{e}^{itx} f(t)\psi_0(x)dt \;=\; \hat{f}(x)\psi_0(x) \;\in \overline{\mathcal{P}}\,,$$

where $\hat{f}$ is the Fourier transform of f. Since the Schwarz class $\mathcal{S}$ is mapped onto itslef by the Fourier transform,[1] we find that $f\psi_0 \in \overline{\mathcal{P}}$ for any $f \in \mathcal{S}$.

Finally, given $f \in L^2$ we may find $h_n \in C_c^\infty(\mathbb{R})$, the compactly supported continuous functions, such that $\lim_n \|f - h_n\|_{L^2} \to 0$ — this may be accomplished by truncating f to a sequence of large intervals and convolving with a δ-sequence of smooth bump functions. Since h_n is compactly supported, we have $\frac{h_n}{\psi_0} \in \mathcal{S}$ and thus $h_n \in \mathcal{S}\psi_0 \subset \overline{\mathcal{P}}$. Thus $f \in \overline{\mathcal{P}}$ and the proof is complete.

□

Since the harmonic oscillator has purely discrete spectrum, its representation by the spectral theorom is particularly simple. Given any function f we have

$$f(H) \;=\; \sum_{n=0}^{\infty} f(\tfrac{1}{2}+n)\,|\psi_n\rangle\langle\psi_n|\;.$$

The projection valued measure $P_H(\lambda)$ is given by

$$P_H(\lambda) \;=\; \begin{cases} 0 & \lambda < \frac{1}{2}\,, \\ \sum_{j=0}^{n} |\psi_j\rangle\langle\psi_j| & \frac{1}{2}+n \le \lambda < \frac{1}{2}+n+1\,, \end{cases} \tag{8.3}$$

and one does indeed have $f(H) = \int_{\mathbb{R}} f(\lambda)dP_H(\lambda)$. It is a striking fact of quantum mechanics that the spectrum of the oscillator P^2+Q^2 is purely discrete when the component operators P^2 and Q^2 both have continuous spectrum on the entire interval $[0,\infty)$.

8.5 Heisenberg evolution of a harmonic oscillator

The Heisenberg equations for the harmonic oscillator are remarkably simple. In fact, they look *exactly* like the classical equations for the harmonic oscillator. Since $P = \frac{i}{\sqrt{2}}(a^\dagger - a)$ and $Q = \frac{1}{\sqrt{2}}(a^\dagger + a)$, we have

$$[H,P] \;=\; \frac{i}{\sqrt{2}}(a^\dagger + a) \;=\; iQ \quad \text{and} \quad [H,Q] \;=\; \frac{1}{\sqrt{2}}(a^\dagger - a) \;=\; -iP\,.$$

Thus the Heisenberg equations are (with $\hbar = 1$)

$$\dot{P} \;=\; -Q \quad \text{and} \quad \dot{Q} \;=\; P\,.$$

[1] This may be proved in an elementary way, since for $f \in \mathcal{S}$ we have

$$\int_{\mathbb{R}} \mathrm{e}^{itx}(1+x^2)^n f(x)dx \;=\; (1-\partial_t^2)^n \hat{f}(t) \quad \text{and} \quad (1+t^2)^n \hat{f}(t) \;=\; \int_{\mathbb{R}} \mathrm{e}^{itx}(1-\partial_x^2)^n f(x)dx\,.$$

The solution to these equations is easily seen to be

$$P(t) \;=\; \cos(t)P - \sin(t)Q \quad \text{and} \quad Q(t) \;=\; \cos(t)Q + \sin(t)P\,.$$

This calculation was somewhat formal, but it is absolutely correct that

$$\mathrm{e}^{itH}P\mathrm{e}^{-itH} \;=\; \cos(t)P - \sin(t)Q \quad \text{and} \quad \mathrm{e}^{itH}Q\mathrm{e}^{-itH} \;=\; \cos(t)Q + \sin(t)P\,.$$

Note that as a result, the mean values of momentum and position in any quantum state are given by solutions to the *classical* harmonic oscillator equation. This is a remarkable parallel between the classical and quantum harmonic oscillators, which does not hold for general quantum systems.

8.6 Coherent states

There is a beautiful picture of the harmonic oscillator if one considers particular set of vector states, given formally by

$$|C_z\rangle \;=\; \mathrm{e}^{-\frac{1}{4}|z|^2}\mathrm{e}^{\frac{1}{\sqrt{2}}za^\dagger}\,|\psi_0\rangle\;,$$

for $z \in \mathbb{C}$. The operator $a^\dagger$ is unbounded, so it is not obvious that this expression makes sense. However, ψ_0 is in the domain of $(a^\dagger)^n$ for every n, and we may simply use the power series expansion to write

$$|C_z\rangle \;:=\; \mathrm{e}^{-\frac{1}{4}|z|^2}\sum_{n=0}^{\infty}\frac{1}{2^{n/2}n!}z^n(a^\dagger)^n\,|\psi_0\rangle \;=\; \mathrm{e}^{-\frac{1}{4}|z|^2}\sum_{n=0}^{\infty}\frac{1}{\sqrt{2^n n!}}z^n\,|\psi_n\rangle\;. \tag{8.4}$$

One computes that $\||C_z\rangle\|^2 = \mathrm{e}^{-\frac{1}{2}|z|^2}\sum_{n=0}^{\infty}\frac{1}{2^n n!}|z|^{2n} = 1$.

The states defined in (8.4) are called *coherent states*. They were introduced in the early 1960s by Klauder 1960 and Glauber 1963. Using $[a,(a^\dagger)^n] = n(a^\dagger)^{n-1}$, we find that

$$a\,|C_z\rangle \;=\; \mathrm{e}^{-\frac{1}{4}|z|^2}\sum_{n=1}^{\infty}\frac{1}{2^{n/2}(n-1)!}z^n(a^\dagger)^{n-1}\,|\psi_0\rangle \;=\; \frac{z}{\sqrt{2}}\,|z\rangle\;.$$

That is $|C_z\rangle$ is an eigenstate of the lowering operator a, with eigenvalue $\frac{z}{\sqrt{2}}$. For the raising operator $a^\dagger$, we have

$$a^\dagger\,|C_z\rangle \;=\; \mathrm{e}^{-\frac{1}{4}|z|^2}\sum_{n=0}^{\infty}\frac{1}{2^{n/2}n!}z^n(a^\dagger)^{n+1}\,|\psi_0\rangle \;=\; \sqrt{2}\mathrm{e}^{-\frac{1}{4}|z|^2}\frac{d}{dz}\mathrm{e}^{\frac{1}{4}|z|^2}\,|C_z\rangle \;=\; \sqrt{2}\left(\frac{d}{dz}+\frac{1}{4}\overline{z}\right)|C_z\rangle\;.$$

Thus the harmonic oscillator hamiltonian H satisfies

$$H\,|C_z\rangle \;=\; (a^\dagger a+\frac{1}{2})\,|C_z\rangle \;=\; \left(\frac{d}{dz}z+\frac{1}{4}|z|^2+\frac{1}{2}\right)|C_z\rangle \;=\; \left(z\frac{d}{dz}+\frac{1}{4}|z|^2+\frac{3}{2}\right)|C_z\rangle\;.$$

Amazingly, one may solve the Schrödinger equation with the *ansatz* that the time dependent solution is of the form $\mathrm{e}^{-i\alpha t}\left|C_{z_0\mathrm{e}^{-it}}\right\rangle$. Note that

$$\frac{d}{dt}\left|C_{z_0\mathrm{e}^{-it}}\right\rangle \;=\; -i\mathrm{e}^{-\frac{1}{4}|z_0|^2}\sum_{n=0}^{\infty}\frac{1}{2^{n/2}n!}nz_0^n\mathrm{e}^{-int}(a^\dagger)^n\left|\psi_0\right\rangle \;=\; -i\left(z\frac{d}{dz}+\frac{1}{4}|z|^2\right)\left|C_z\right\rangle\Big|_{z=z_0\mathrm{e}^{-it}}\,.$$

We see that $\mathrm{e}^{-i\alpha t}\left|C_{z_0\mathrm{e}^{-it}}\right\rangle$ solves the Schrödinger equation $\frac{d}{dt}\left|\psi_t\right\rangle = -iH\left|\psi_t\right\rangle$ provided $\alpha = \frac{3}{2}$. It follows that

$$\mathrm{e}^{-itH}\left|C_z\right\rangle \;=\; \mathrm{e}^{-i\frac{3}{2}t}\left|C_{\mathrm{e}^{-it}z}\right\rangle\,. \tag{8.5}$$

Thus, with the family of states $|C_z\rangle$, the motion of the oscillator is just given by circulation in the complex plane (up to an overall phase).

One can relate the above motion to the expect position and momentum in the state by looking more closely at the form of $|C_z\rangle$. Using the eigenvalue eqution $a\left|C_z\right\rangle$ one can solve for the function $C_z \in L^2$, since we have $(\partial_x + x)C_z(x) = zC_z(x)$. Thus $C_z(x) = f_z(x)\psi_0(x)$ with $\partial_x f_z(x) = zf_z(x)$. We conclude that $C_z(x) = \mathrm{e}^{-(\Re z)^2}\mathrm{e}^{zx}\psi_0(x)$, where the normalization $\mathrm{e}^{-(\Re z)^2}$ guarantees that

$$\int_{\mathbb{R}}|C_z(x)|^2dx \;=\; \frac{1}{\sqrt{\pi}}\mathrm{e}^{-(\Re z)^2}\int_{\mathbb{R}}\mathrm{e}^{-|x|^2+2\Re zx}dx \;=\; \frac{1}{\sqrt{\pi}}\mathrm{e}^{-|x-\Re z|^2}dx \;=\; 1\,.$$

Let us rewrite the expression for C_z as follows, taking $z = q + ip$ with $q, p \in \mathbb{R}$,

$$C_{q+ip}(x) \;=\; \frac{1}{\pi^{1/4}}\mathrm{e}^{ipx}\mathrm{e}^{-\frac{1}{2}(x-q)^2}\,. \tag{8.6}$$

This wave function has

1. average position $\left\langle C_{q+ip}\right| Q\left|C_{q+ip}\right\rangle \;=\; q$, and
2. average momentum $\left\langle C_{q+ip}\right| P\left|C_{q+ip}\right\rangle \;=\; p$.

Thus the value of $z = p + iq$ is simple the average position and momentum in phase space, realized as the complex plane. The oscillatory motion $z\mathrm{e}^{-it}$ is just the oscillation of mean values observed above via solutions to the Heisenberg equations.

The uncertainties of position and momentum in the state C_{q+ip} can also be computed:

$$\Delta^2_{q+ip}(Q) \;=\; \frac{1}{\sqrt{\pi}}\int_{\mathbb{R}}(x-q)^2\mathrm{e}^{-(x-q)^2}dx \;=\; \frac{1}{2}\,,$$

and

$$\Delta^2_{q+ip}(P) \;=\; \frac{1}{\sqrt{\pi}}\int_{\mathbb{R}}\mathrm{e}^{-\frac{1}{2}(x-q)^2}\mathrm{e}^{-ipx}\left(-i\frac{d}{dx}-p\right)^2\mathrm{e}^{ipx}\mathrm{e}^{-\frac{1}{2}(x-q)^2}$$
$$\;=\; \frac{1}{\sqrt{\pi}}\int_{\mathbb{R}}\left[\frac{d}{dx}\mathrm{e}^{-\frac{1}{2}(x-q)^2}\right]^2 \;=\; \frac{1}{2}\,.$$

It follows that the coherent states satisfy

$$\Delta_{q+ip}(Q)\Delta_{q+ip}(Q) \;=\; \frac{1}{2}\,,$$

which is the minimum uncertainty allowed by the Heisenberg commutation relations, since $[Q,P]=i$. Coherent states are states in $L^2(\mathbb{R})$ with minimal uncertainty.

In fact one may use the above expression to find the solution to the Schrödinger equation for a general state. The key is the observation that

$$\frac{1}{2\pi}\int_{\mathbb{C}} |C_z\rangle\langle C_z|\, d^2z \;=\; \frac{1}{2\pi}\sum_{n,m=0}^{\infty}\int_{\mathbb{C}} \mathrm{e}^{-\frac{1}{2}|z|^2}\frac{1}{\sqrt{2^{n+m}n!m!}}\bar{z}^n z^m d^2z\,|\psi_n\rangle\langle\psi_n| \;=\; \sum_{n=0}^{\infty}|\psi_n\rangle\langle\psi_n|\,, \tag{8.7}$$

where we have noted that

$$\frac{1}{2\pi}\int_{\mathbb{C}} \mathrm{e}^{-\frac{1}{2}|z|^2}\bar{z}^m z^n d^2z \;=\; \delta_{nm}\int_0^\infty r^{2n+1}\mathrm{e}^{-\frac{1}{2}r^2}dr \;=\; \delta_{nm}2^n\int_0^\infty s^n\mathrm{e}^{-s}ds \;=\; \delta_{nm}2^n n!\,.$$

Eq. (8.7) expresses the fact that the coherent states are an *over complete basis*. It can be used to express *any* state as a superposition of coherent states:

$$|\psi\rangle \;=\; \frac{1}{2\pi}\int_{\mathbb{C}} |C_z\rangle\langle C_z|\psi\rangle\, d^2z\,.$$

Thus

$$\mathrm{e}^{-itH}|\psi\rangle \;=\; \frac{1}{2\pi}\int_{\mathbb{C}} \mathrm{e}^{-i\frac{3}{2}t}\,|C_{\mathrm{e}^{-it}z}\rangle\langle C_z|\psi\rangle\, d^2z\,.$$

One can use this formula to derive an exact expression for the propagator of e^{-itH}, analgous to that for the free particle. However we will have no need for it.

Part II

Composite and Open Quantum Systems

Lecture 9

Modeling composite systems

9.1 Composite systems in classical mechanics

Suppose we have two classical systems with configuration spaces M_1 and M_2. How do we describe the phase space of two systems M_1 and M_2 considered together?

The phase space of a single classical system is the tangent bundle TM over the configuration space M, where tangent vector describe the velocity of the system. For example, imagine a classical system consisting of a single particle in a box in space. The configuration space is B, the open interior of the box. Imagine that there is very steep potential energy that diverges to infinity at the boundary of the box, so the particle never leaves B. The phase space of this classical system is the tangent space $TB \cong B \times \mathbf{R}^3$.

Suppose we have two particles, each in separate boxes B_1 and B_2. To specify the state of the composite system we should specify the positions $q_j \in B_j$ of each particle. And to specify the specify the velocity of the configuration (q_1, q_2), we just need to specify the velocity of each particle separately: (v_1, v_2). So the phase space of the composite system is just $TB_1 \times TB_2$.

In general, the configuration space of the composite of two classical systems is the product $M_1 \times M_2$ of their configuration spaces, and the phase space is the tangent space $T(M_1 \times M_2)$. Because, the tangent space at a point of $M_1 \times M_2$ satisfies

$$T_(q_1, q_2)(M_1 \times M_2) \;=\; (T_{q_1} M_1) \oplus (T_{q_2} M_2)\,,$$

we see that $T(M_1 \times M_2) \cong (TM_1) \times (TM_2)$ is also just the product of the individual phase spaces. (Note: the cartesian product of vector spaces corresponds to the direct sum operation.)

How do we phrase this in terms of observables? The full algebra of observables for the composite system is the set of smooth functions on the phase space $\mathcal{A} = C^\infty(T(M_1 \times M_2))$. There is a natural imbedding of $\mathcal{A}_j = C^\infty(TM_j)$ into $\mathcal{A}$: just consider $f \in \mathcal{A}_j$ to be a function on $T(M_1 \times M_2)$ that does not depend on the position or velocity of the other system. More formally, we let $\Pi_j : T(M_1 \times M_2) \to TM_j$ denote the coordinate maps

$$(q_1, q_2, v_1, v_2) \overset{\Pi_j}{\mapsto} (q_j, v_j)\,.$$

Then we embed $\mathcal{A}_j \hookrightarrow \mathcal{A}$ by mapping $f \mapsto f \circ \Pi_j$.

Note that $\mathcal{A} \neq \mathcal{A}_1 \times \mathcal{A}_2$. Instead, $\mathcal{A}$ is the *tensor product* $\mathcal{A}_1 \otimes \mathcal{A}_2$ — or, more correctly, it is the completion of the algebraic tensor product in a certain metric. Put in a more pedestrian fashion, any observable $f : T(M_1 \times M_2) \to \mathbb{R}$ can be expressed as a limit of finite linear combinations of the form

$$\sum_{j=1}^{n} g_j(q_1, v_1) h_j(q_2, v_2) \ ,$$

where $(g_j)_{j=1}^n$ are observables on M_1 and $(h_j)_{j=1}^n$ are observables on M_2. The fact that any function on $T(M_1 \times M_2)$ can be written as such a limit can be derived from the Stone-Weirstrass theorem. We will see that this tensor product structure carries over to quantum mechanics.

What about states? A state on the composite system is a probability measure on $T(M_1 \times M_2)$. Again, such a measure need not be a product measure can be written as a convex combination of product measures. In fact, the miracle of classical mechanics is that any *pure state* on $T(M_1 \times M_2)$ (which, after all, is just a point in $T(M_1 \times M_2)$) is indeed a product state.

9.2 Tensor Products

Before devloping the description of composite systems in quantum mechanics, it is helpful to do a bit of algebra and functional analysis. We start with the algebra. Given a vector space V, the (algebraic) dual of V is the set V^* is the set of all linear functionals $\ell : V \to \mathbb{C}$. There is a natural embedding of $V \hookrightarrow (V^*)^*$, the dual of the dual, given by $x(\ell) = \ell(x)$. If V is finite dimensional, this embedding is surjective but in infinite dimensions that is not the case.

Let V_1 and V_2 be vector spaces. A *bilinear* map on $V_1 \times V_2$ is a function $F : V_1 \times V_2 \to \mathbb{C}$ such that

$$x_1 \mapsto F(x_1, x_2) \quad \text{and} \quad x_2 \mapsto F(x_1, x_2)$$

are linear. Given $x_1 \in V_1$ and $x_2 \in V_2$, let $x_1 \otimes x_2$ be the bilinear functional on $V_1^* \times V_2^*$ defined by

$$x_1 \otimes x_2(\ell_1, \ell_2) \ = \ \ell_1(x_1)\ell_2(x_2) \ .$$

The *algebraic tensor product* $V_1 \widehat{\otimes} V_2$ is the linear span of $\{x_1 \otimes x_2 \ : \ x_1 \in V_1, x_2 \in V_2\}$. That is

$$V_1 \widehat{\otimes} V_2 \ = \ \left\{ \sum_{j=1}^{n} x_j \otimes y_j \ : \ n \in \mathbb{N} \text{ and } x_j \in V_1 \ , \ y_j \in V_2, \ j = 1, \dots, n \right\} \ . \tag{9.1}$$

Remarks:

- There is no need for scalars in the linear combinations on the right hand side of (9.1), since $\alpha(x_j \otimes y_j) = (\alpha x_j) \otimes y_j = x_j \otimes (\alpha y_j)$.
- If V_1 and V_2 are finite dimensional, then $V_1 \widehat{\otimes} V_2$ is the set of all bilinear maps on $V_1^* \times V_2^*$.

- One could define the tensor product $x \otimes y$ purely formally as an equaivalence class of $V_1 \times V_2$ modulo the relations induced by requiring it to be linear in each entry.

Let $\mathcal{H}$ be a separable[1] Hilbert space. By the Riesz representation theorem, there is a surjective conjugate linear map $R : \mathcal{H}^* \to \mathcal{H}$, given by $\langle R\ell | x \rangle = \ell(x)$. We make the dual $\mathcal{H}^*$ a Hilbert space with the inner product

$$\langle \ell | \ell' \rangle_{\mathcal{H}^*} \;:=\; \langle R\ell' | R\ell \rangle_{\mathcal{H}} \;.$$

(Note that the order of the vectors is reversed, so that the inner product on $\mathcal{H}^*$ is conjugate linear in the left hand entry as required.) Let $B(\mathcal{H}_1^*, \mathcal{H}_2^*)$ denote the set of all *bounded bilinear maps* from $\mathcal{H}_1^* \times \mathcal{H}_2^* \to \mathbb{C}$. That is $F \in B(\mathcal{H}_1^*, \mathcal{H}_2^*)$ is a bilinear map $F : \mathcal{H}_1^* \times \mathcal{H}_2^* \to \mathbb{C}$ such that

$$|F(\ell_1, \ell_2)| \;\leq c \|\ell_1\| \|\ell_2\|$$

for some $c < \infty$. Given $F \in B(\mathcal{H}_1^*, \mathcal{H}_2^*)$, let $\tilde{F} : \mathcal{H}_1 \times \mathcal{H}_2 \to \mathbb{C}$ denote the map

$$\tilde{F}(\phi_1, \phi_2) \;=\; F(R_1^{-1}\phi_1, R_2^{-1}\phi_2) \;, \tag{9.2}$$

with R_j the corresponding Riesz maps. The map $\tilde{F}$ is a conjugate bilinear map, i.e., $\tilde{F}(\phi_1 + \alpha\phi_1', \phi_2) = \tilde{F}(\phi_1, \phi_2) + \alpha^* \tilde{F}(\phi_1', \phi_2)$ and similarly for the second factor.

Given $\phi_1, \phi_2 \in \mathcal{H}$, the tensor product $\phi_1 \otimes \phi_2$ is a bounded bilinear map

$$|\phi_1 \otimes \phi_2(\ell_1, \ell_2)| \;=\; |\ell_1(\phi_1)||\ell_2(\phi_2)| \;\leq\; \|\phi_1\|\|\phi_2\|\|\ell_1\|\|\ell_2\| \,.$$

Thus the algebraic tensor product $\mathcal{H}_1 \widehat{\otimes} \mathcal{H}_2 \subset B(\mathcal{H}_1, \mathcal{H}_2)$. If $F = \phi_1 \otimes \phi_2$, let $\phi_1 \widetilde{\otimes} \phi_2 = \tilde{F}$, with $\tilde{F}$ as in (9.2). Then

$$\phi_1 \widetilde{\otimes} \phi_2(\psi_1, \psi_2) \;=\; \langle \psi_1 | \phi_1 \rangle_1 \, \langle \psi_2 | \phi_2 \rangle_2 \;.$$

Proposition 9.1. *There is a unique inner product on $\mathcal{H}_1 \widehat{\otimes} \mathcal{H}_2$ such that*

$$\langle \psi_1 \otimes \psi_2 | \phi_1 \otimes \phi_2 \rangle_{1 \otimes 2} \;=\; \langle \psi_1 | \phi_1 \rangle_1 \, \langle \psi_2 | \phi_2 \rangle_2 \;. \tag{9.3}$$

Proof. Let

$$\alpha \;=\; \sum_{j=1}^{J} \psi_1^{(j)} \otimes \psi_2^{(j)} \quad \text{and} \quad \beta \;=\; \sum_{k=1}^{K} \phi_1^{(k)} \otimes \phi_2^{(k)}$$

be arbitrary elements of $\mathcal{H}_1 \widehat{\otimes} \mathcal{H}_2$. We would like to take

$$\langle \alpha | \beta \rangle \;=\; \sum_{j=1}^{J} \sum_{k=1}^{K} \left\langle \psi_1^{(j)} \middle| \phi_1^{(k)} \right\rangle \left\langle \psi_2^{(j)} \middle| \phi_2^{(k)} \right\rangle \;. \tag{9.4}$$

Unfortunately, it is unclear that this is well defined, since the representation of the elements $\alpha, \beta \in \mathcal{H}_1 \widehat{\otimes} \mathcal{H}_2$ as a linear combination is not unique. However, note that we can

[1] From now on, all Hilbert spaces will be separable unless otherwise mentioned

rewrite the right hand side of (9.4) as $\sum_{j=1}^{J} \tilde{\beta}(\psi_1^{(j)}, \psi_2^{(j)})$ or as $\left(\sum_{k=1}^{K} \tilde{\alpha}(\phi_1^{(k)}, \phi_2^{(k)})\right)^*$. These identites can be used to show: 1) for $(\psi_1^{(j)}, \psi_2^{(j)})_{j=1}^{J}$ fixed the right hand side of (9.4) does not depend on the representation of β and is linear in β, 2) $(\phi_1^{(k)}, \phi_2^{(k)})_{k=1}^{K}$ fixed the right hand side of (9.4) does not depend on the representation of α and is conjugate linear in α. It follows, by telescoping differences, that the right hand side of (9.4) does not depend on the representation of either α or β and is conjugate linear in α and linear in β.

To complete the proof we must show that $\langle \alpha | \alpha \rangle \geq 0$ and that it equals zero if and only if $\alpha = 0$. For this purpose we will use the following

Proposition 9.2. *Let $\alpha \in \mathcal{H}_1 \widehat{\otimes} \mathcal{H}_2$. Then there are orthonormal families $(\phi_1^{(j)})_{j=1}^{J}$ in $\mathcal{H}_1$ and $(\phi_2^{(k)})_{k=1}^{K}$ in $\mathcal{H}_2$ such that*

$$\alpha \;=\; \sum_{j=1}^{J} \sum_{k=1}^{K} c_{j,k} \phi_1^{(j)} \otimes \phi_2^{(k)}$$

and $\langle \alpha | \alpha \rangle = \sum_{j,k} |c_{j,k}|^2$.

Exercise 9.1. Prove prop. 9.2. (Hint: use Gram-Schmidt.)

From the above, we see directly that $\langle \alpha | \alpha \rangle \geq 0$. Furthermore if equality is attained then $c_{j,k} = 0$ for all j, k, whence $\alpha = 0$.

□

Definition 9.3. Let $\mathcal{H}_1$, $\mathcal{H}_2$ be Hilbert spaces. The (Hilbert space) *tensor product* $\mathcal{H}_1 \otimes \mathcal{H}_2$ is the completion of $\mathcal{H}_1 \widehat{\otimes} \mathcal{H}_2$ in the inner product defined in Prop. 9.1.

Lemma 9.4. *The space $B(\mathcal{H}_1^*, \mathcal{H}_2^*)$ of bounded bilinear maps from $\mathcal{H}_1^* \times \mathcal{H}_2^* \to \mathbb{C}$ is a Banach space under the norm*

$$\|F\|_B \;=\; \sup_{\|\ell_j\|_{\mathcal{H}_j}^* = 1,\, j=1,2} |F(\ell_1, \ell_2)| \,.$$

For $\alpha \in \mathcal{H}_1 \widehat{\otimes} \mathcal{H}_2$ we have

$$\|\alpha\|_B \;\leq\; (\langle \alpha | \alpha \rangle)^{\frac{1}{2}} \,.$$

Exercise 9.2. Prove that $\|\cdot\|_B$ is a norm and that any Cauchy sequence in this norm converges to a bounded bilinear map.

Proof. It remains to show the inequality for the norms. By Prop. 9.2 we have

$$|\alpha(\ell_1, \ell_2)| \;=\; \left| \sum_{j,k=1}^{J,K} c_{j,k} \ell_1(\phi_1^{(j)}) \ell_2(\phi_2^{(k)}) \right| ,$$

with $|\alpha|\alpha = \sum_{j,k} = |c_{j,k}|^2$ and $(\phi_1^{(j)})_j$ $(\phi_2^{(k)})_k$ orthonormal. Thus, by Cauchy-Schwarz,

$$|\alpha(\ell_1, \ell_2)| \;\leq\; \left(\sum_{j,k=1}^{J,K} |c_{j,k}|^2 \right)^{\frac{1}{2}} \left(\sum_{j=1}^{J} \left| \ell_1(\phi_1^{(j)}) \right|^2 \right)^{\frac{1}{2}} \left(\sum_{k=1}^{K} \left| \ell_2(\phi_2^{(k)}) \right|^2 \right)^{\frac{1}{2}} \;\leq\; (\langle \alpha | \alpha \rangle)^{\frac{1}{2}} \, \|\ell_1\| \|\ell_2\| \,,$$

by Bessel's inequality.

□

Theorem 9.5. *There is a continuous, one-to-one embedding of $\mathcal{H}_1 \otimes \mathcal{H}_2$ into $B(\mathcal{H}_1^*, \mathcal{H}_2^*)$. Furthermore, if $(\phi_1^{(j)})_{j=1}^J$ and $(\phi_2^{(k)})_{k=1}^K$ are orthonormal bases for $\mathcal{H}_1$ and $\mathcal{H}_2$, with $J, K \in \mathbb{N} \cup \{\infty\}$, then $(\phi_1^{(j)} \otimes \phi_2^{(k)})_{j,k=1}^{J,K}$ is an orthonormal basis for $\mathcal{H}_1 \otimes \mathcal{H}_2$.*

Proof. Let $(\alpha_j)_{j=1}^\infty$ be a Cauchy sequence in $\mathcal{H}_1 \widehat{\otimes} \mathcal{H}_2$. By Lemma 9.4, we have that $(\alpha_j)_{j=1}^\infty$ is also Cauchy in $B(\mathcal{H}_1^*, \mathcal{H}_2^*)$. Since the latter space is a Banach space, the sequence has a limit in the space and we may identify this point with the corresponding point in $\mathcal{H}_1 \otimes \mathcal{H}_2$. The bound of Lemma 9.4 shows that this identification is continuous.

Clearly $(\phi_1^{(j)} \otimes \phi_2^{(k)})_{j,k}$ is an orthonormal sequence. So it suffices to show closure of its span is $\mathcal{H}_1 \otimes \mathcal{H}_2$. For this purpose, it suffices to show that any simple tensor product $\phi \otimes \psi$ is a limit of finite linear combinations of basis vectors. But we have

$$\phi \otimes \psi \;=\; \lim_{N \to \infty} \sum_{j,k=1}^{N} \alpha_j \beta_k \phi_1^{(j)} \otimes \phi_2^{(k)} \;,$$

with $\alpha_j = \left\langle \phi_1^{(j)} \middle| \phi \right\rangle$ and $\beta_k = \left\langle \phi_2^{(k)} \middle| \psi \right\rangle$, as one may easily show.

□

Lecture 10

Composite systems in quantum mechanics

10.1 Tensor product of observables

Let $\mathcal{H}$ be a Hilbert space and let $A \in \mathcal{B}(\mathcal{H})$ be a bounded operator. The *transpose of* A is the map on the dual $\mathcal{H}^*$ defined by $(A^T\ell)(\phi) = \ell(A\phi)$.[1] Note that

$$\left\langle RA^T\ell|\phi\right\rangle \;=\; \langle R\ell|A\phi\rangle \;=\; \left\langle A^\dagger R\ell|\phi\right\rangle \;.$$

Thus $A^T = R^{-1}A^\dagger R$, where $R : \mathcal{H}^* \to \mathcal{H}$ is the Riesz map. Note that

$$\left\|A^T\right\| \;=\; \sup_{\|\ell\|=1}\left\|A^T\ell\right\| \;=\; \sup_{\|\ell\|=\|\phi\|=1}\left|(A^T\ell)(\phi)\right| \;=\; \sup_{\|\ell\|=\|\phi\|=1}|\ell(A\phi)| \;=\; \sup_{\|\phi\|=1}\|A\phi\| \;=\|A\|.$$

Consider two Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$. Let A_1 and A_2 be bounded operators on $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively. If $F : \mathcal{H}_1^* \times \mathcal{H}_2^* \to \mathbb{C}$ is a bounded bilinear form, we define

$$((A_1 \otimes A_2)\cdot F)(\ell_1,\ell_2) \;=\; F(A_1^T\ell_1, A_2^T\ell_2)\;.$$

Note that

$$\|(A_1\otimes A_2)\cdot F\| \;=\; \sup_{\|\ell_1\|=1,\ \|\ell_2\|=1}\left|F(A_1^T\ell_1,A_2^T\ell_2)\right| \;\le\; \|F\|\|A_1\|\|A_2\|\,,$$

so $(A_1 \otimes A_2)\cdot F$ is again a bounded bilinear form. If $\phi_j \in \mathcal{H}_j$, $j = 1,2$, then one has

$$(A_1\otimes A_2)\cdot(\phi_1\otimes\phi_2) \;=\; (A_1\phi_1)\otimes(A_2\phi_2)\;m \tag{10.1}$$

It follows that $A_1 \otimes A_2 : \mathcal{H}_1 \widehat{\otimes} \mathcal{H}_2 \to \mathcal{H}_1 \widehat{\otimes} \mathcal{H}_2$, and therefore since it is bounded, that $A_1 \otimes A_2 \in \mathcal{B}(\mathcal{H}_1 \otimes \mathcal{H}_2)$. Note that the bounded operator on $\mathcal{H}_1 \otimes \mathcal{H}_2$ satisfying (10.1) is unique, since (10.1) already specifies the values of the operator on an orthonormal basis of product vectors. We have obtained

[1] The terminology "transpose" is not standard. A more proper term would be "Banach space adjoint."

Proposition 10.1. *Let $A_j \in \mathcal{B}(\mathcal{H}_j)$, $j = 1, 2$. Then there is a unique operator $A_1 \otimes A_2 \in \mathcal{B}(\mathcal{H}_1 \otimes \mathcal{H}_2)$ such that* (10.1) *holds for all $\phi_j \in \mathcal{H}_j$, $j = 1, 2$.*

Definition 10.2. Let $\mathfrak{A}_j \subset \mathcal{B}(\mathcal{H}_j)$ be an algebra on the Hilbert space $\mathcal{H}_j$, $j = 1, 2$. The *tensor product* $\mathfrak{A}_1 \otimes \mathfrak{A}_2$ is the algebra in $\mathcal{B}(\mathcal{H}_1 \otimes \mathcal{H}_2)$ obtained by completing $\{A_1 \otimes A_2 \ : \ A_j \in \mathfrak{A}_j \ , \ j = 1, 2\}$ in the *strong operator topology*.

Theorem 10.3. *Let $\mathcal{H}_j$, $j = 1, 2$ be Hilbert spaces. We have*

1. $\mathcal{B}(\mathcal{H}_1) \otimes \mathcal{B}(\mathcal{H}_2) = \mathcal{B}(\mathcal{H}_1 \otimes \mathcal{H}_2)$.

2. *If $\mathfrak{A}_j = \{A \in \mathcal{B}(\mathcal{H}_j) : A = A^\dagger\}$, $j = 1, 2$, then $\mathfrak{A}_1 \otimes \mathfrak{A}_2$ is the set of self-adjoint elements of $\mathcal{B}(\mathcal{H}_1 \otimes \mathcal{H}_2)$*

Remark 10.4. There are multiple definitions of tensor products in infinite dimensions. The one we take here, with the strong operator topology, is the *von Neumann* tensor product and is the one most relevant to quantum mechanics. If instead we complete $\mathcal{B}(\mathcal{H}_1) \otimes \mathcal{B}(\mathcal{H}_2)$ in the operator norm, we obtain the *minimal tensor product* $\mathcal{B}(\mathcal{H}_1) \otimes_{\min} \mathcal{B}(\mathcal{H}_2)$, which is strictly smaller than $\mathcal{B}(\mathcal{H}_1 \otimes \mathcal{H}_2)$ if both $\mathcal{H}_1$ and $\mathcal{H}_2$ are infinite dimensional. As the name suggests, there is also a *maximal tensor product*, which is somewhat trickier to define.

10.2 Multiple tensor products

Given $\mathcal{H}_j$, $j = 1, \dots, n$, one might consider the tensor product of the spaces by grouping the Hilbert spaces in pairs in various ways, for example $(\mathcal{H}_1 \otimes \mathcal{H}_2) \otimes \mathcal{H}_3$ or $\mathcal{H}_1 \otimes (\mathcal{H}_2 \otimes \mathcal{H}_3)$. However, an alternative is to follow the above argument and define the tensor product $\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n$ as a space of multilinear maps on $\mathcal{H}_1^* \times \cdots \times \mathcal{H}_n^*$. Namely, given $\phi_j \in \mathcal{H}_j$ we define $\phi_1 \otimes \cdots \otimes \phi_n$ to be the multilinear map

$$\phi_1 \otimes \cdots \otimes \phi_n(\ell_1, \dots, \ell_n) \ = \ \prod_{j=1}^{n} \ell_j(\phi_j) \ .$$

As above, we can define an inner product and complete the space of finite linear combinations of such maps to obtain $\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n$.

Exercise 10.1. Check that there are natural isomorphisms

1. $(\mathcal{H}_1 \otimes \mathcal{H}_2) \otimes \mathcal{H}_3 \cong \mathcal{H}_1 \otimes (\mathcal{H}_2 \otimes \mathcal{H}_3) \cong \mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \mathcal{H}_3$, and

2. $(\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_{n-1}) \otimes \mathcal{H}_n \cong \mathcal{H}_1 \otimes \cdots \mathcal{H}_n$.

Conclude that the tensor product does not depend on the order it is built up.

10.3 Composite systems in quantum mechanics

Now consider a composite quantum system with two parts $j = 1, 2$, with corresponding algebras of observables $\mathfrak{A}_j \subset \mathcal{B}(\mathcal{H}_j)$ consisting of all self-adjoint operators on $\mathcal{H}_j$. We will assume henceforth that $\dim \mathcal{H}_j < \infty$. It is a postulate of quantum mechanics that the algebra of observables for the composite system is $\mathfrak{A}_1 \otimes \mathfrak{A}_2$, which in this case consists of all self-adjoint operators on $\mathcal{H}_1 \otimes \mathcal{H}_2$.

States on the compostie system are density matrices on $\mathcal{H}_1 \otimes \mathcal{H}_2$. The pure states are rank-1 matrices $|\Psi\rangle\langle\Psi|$ with $\Psi \in \mathcal{H}_1 \otimes \mathcal{H}_2$ and $\|\Psi\| = 1$. It is important to note, that in contrast to classical mechanics, the pure states need not be product states. Let $(\phi_j)_{j=1}^{J}$ and $(\psi_k)_{k=1}^{K}$ be orthonormal bases for $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively. Then $(\phi_j \otimes \psi_k)_{j,k=1}^{J,K}$ is an orthonormal basis for $\mathcal{H}_1 \otimes \mathcal{H}_2$ and we can expand

$$|\Psi\rangle = \sum_{j,k=1}^{J,K} \alpha_{j,k} \left|\psi_j \otimes \psi_k\right\rangle ,$$

where $\alpha_{j,k} = \left\langle \psi_j \otimes \psi_k \middle| \Psi \right\rangle$. Each $\alpha_{j,k}$ is a complex number and there are JK such numbers need to specify a vector. However, the projection $\frac{|\Psi\rangle\langle\Psi|}{\|\Psi\|^2}$ is invariant under multiplication of Ψ by a scalar. Thus the set of states is a $JK - 1$ dimensional complex manifold. On the other hand, a product state corresponds to a vector of the form

$$|\Psi\rangle = \left(\sum_{j=1}^{J} a_j \left|\phi_j\right\rangle\right) \otimes \left(\sum_{k=1}^{K} b_j \left|\psi_j\right\rangle\right) .$$

Since the corresponding projection is invariant under scalar multiplication of either vector, the corresponding manifold of product states has complex dimension $J + K - 2$. Since $J, K \geq 2$, we always have $JK - 1 > J + K - 2$. proone complex number is eliminated by the normalization of Ψ and the invariance of the state $|\Psi\rangle\langle\Psi|$ under s Ψ by a complex Entanglement. Entropy.

Lecture 11

Measurements, entanglement, the EPR argument and Bell's inequality

Reading: §2.2 of Nielsen and Chuang

11.1 Measurement

In the Schrödinger picture, an isolated quantum system evolves via the unitary evolution $\rho_t = \mathrm{e}^{-itH}\rho_0\mathrm{e}^{itH}$, where H is the Hamiltonian. By itself, however, this evolution does not describe the outcome of experiments. For this one needs an additional idea.

A fundamental observation is that the results of measurement of quantum systems are random. The outcome of a particular experiment cannot be accurately predicted, but the distribution of outcomes from many identical experiments can be. There are difficulty theoretical/physical issues at setting the boundary between the "quantum world" of a quantum system and the "classical world" of experimental apparati. However, there is a simple formulation of the idea of quantum measurement that is robust enough to be descriptive without needing to fully resolve these issues.

Definition 11.1. A *quantum measurement ensemble* (QME) is collection of operators $(M_a)_{a\in S}$ in $\mathcal{B}(\mathcal{H})$ indexed by $a \in S$, the set of *measurement outcomes*, such that $\sum_{a\in S} M_a^\dagger M_a = I$.

We say a *measurment of a quantum system* is described by a QME $(M_a)_{a\in S}$ if, when the system is in state ρ

1. outcome a is observed with probability $p_a = \operatorname{tr} M_a\rho M_a^\dagger$, and
2. if outcome is observed the state of the system after measurement is given by $\frac{1}{p_a}M_a\rho M_a^\dagger$.

Note that the *completeness equation* $\sum_{a\in S} M_a^\dagger M_a = I$ guarantees that the sum of probabilities $\sum_a p_a = 1$.

The most important class of measurements (and the original class to be considered) are the *projective measurements*:

Definition 11.2. A *projective measurement ensemble* (PME) is a QME $(M_a)_{a\in S}$ for which M_a is an orthogonal projection for every a.

Assuming that $\mathcal{H}$ is finite dimensional, then every observable $A \in \mathfrak{A}$ gives rise to a PME $(P_a)_{a\in\sigma(A)}$ via the spectral representation

$$A \;=\; \sum_{a\in\sigma(A)} aP_a \,,$$

where $\sigma(A)$ = the set of eigenvalues of A and P_a = the projection onto the subspace of eigenvectors of eigenvalue a. A *measurement of A* is a measurement described by the PME $(P_a)_{a\in\sigma(A)}$.

On readily checks that the probabilistic expectation of a in the measurement of an observable A agrees with the quantum expectation that we have been using:

$$\sum_a ap_a \;=\; \sum_a a \operatorname{tr} P_a\rho P_a \;=\; \operatorname{tr}\left[\sum_a aP_a\rho\right] \;=\; \operatorname{tr} A\rho \,.$$

Similarly, the variance of a is given by the square of the qunatum uncertainty:

$$\sum_a a^2p_a - \left(\sum_a aP_a\right)^2 \;=\; \operatorname{tr} A^2\rho - (\operatorname{tr} A\rho)^2 \;.$$

For a system in the pure state $\rho = |\psi\rangle\langle\psi|$ we can think of the state as being indexed by the unit vector $\psi \in \mathcal{H}$. Since $M_a|\psi\rangle\langle\psi|M_a$ is also rank one for any $M_a \in \mathcal{B}(\mathcal{H})$, we see that a system that is in a pure statement before a measurement remains in a pure state after the measurement. In terms of the unit vector ψ, we have

1. the probability of outcome a is $\langle\psi|M_a^\dagger M_a|\psi\rangle$, and
2. if the outcome is a, then the state after measurement is the pure state associated to the vector $\frac{1}{\sqrt{p_a}}M_a\psi$.

The change of ψ after a measurement is sometimes called, somewhat melodramatically, "wave function collapse."

However, recall that the overall phase of ψ does not matter — ψ and $e^{i\theta}\psi$ generate the same state. Similarly, there is a gauge invariance to QME's: $(M_a)_{a\in S}$ and $(e^{i\theta_a}M_a)_{a\in S}$ describe the same measurement process.

11.2 Measurements of a qubit

Suppose we have one qubit, with Hilbert space $\mathcal{H} = \mathbb{C}$. Let the vectors $|0\rangle$ and $|1\rangle$ correspond to the matrices $\left(\begin{smallmatrix}1\\0\end{smallmatrix}\right)$ and $\left(\begin{smallmatrix}0\\1\end{smallmatrix}\right)$, respectively. Recall the Pauli matrices (4.1); we introduce a new notation here to accomodate indexing by quibit number below:

$$X = \sigma_1 = \left(\begin{smallmatrix}0&1\\1&0\end{smallmatrix}\right), \quad Y = \sigma_2 = \left(\begin{smallmatrix}0&-i\\i&0\end{smallmatrix}\right), \quad \text{and} \quad Z = \sigma_3 = \left(\begin{smallmatrix}1&0\\0&-1\end{smallmatrix}\right).$$

Suppose the system is in the pure state corresponding to the vector $|0\rangle$. Consider the measurement of each Pauli matrix:

1. If we measure Z, then we obtain the result $a = 1$ with probability 1 and the state remains unchanged after measurement.

2. If we measure X, then we obtain the results ± 1, each with probability $1/2$, and the state after measurement is $\frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle)$.

3. If we measure Y, then we obtain the results ± 1, each with probability $1/2$, and the state after measurement is $\frac{1}{\sqrt{2}}(|0\rangle \pm i\,|1\rangle)$.

Note that measurements of X and Y have the same distribution of outcomes, but the resulting states after measurement are different.

The ovservable $\boldsymbol{v}\cdot\boldsymbol{\sigma}$, with $\boldsymbol{\sigma} = (X, Y, Z)$ and $|\boldsymbol{v}| = 1$ corresponds to the spin in direction $\boldsymbol{v}$. It has eigenvalues ± 1, which can be seen by computing $(\boldsymbol{v}\cdot\boldsymbol{\sigma})^2 = I$. Recall that a general state of a qubit can be written as $\rho_{\boldsymbol{w}} = \frac{1}{2}[I + \boldsymbol{w}\cdot\boldsymbol{\sigma}]$ with $|\boldsymbol{w}| \leq 1$. What happens if we measure the spin in direction $\boldsymbol{v}\cdot\boldsymbol{\sigma}$ in the state $\rho_{\boldsymbol{w}}$? We can easily solve this problem by noting that the probabilities are rotation invariant. Thus we may choose a rotation O with $O\boldsymbol{v} = \boldsymbol{e}_3$ and compute the probabilities for measuring Z in the state $\rho_{\boldsymbol{w}'}$ with $\boldsymbol{w}' = O\boldsymbol{w}$. Thus we obtain $\mathrm{Prob}[\boldsymbol{v}\cdot\boldsymbol{\sigma} = \pm 1] = \frac{1}{2}(1 \pm w_3') = \frac{1}{2}(1 \pm \boldsymbol{v}\cdot\boldsymbol{w})$.

Proposition 11.3. *In a measurement of $\boldsymbol{v}\cdot\boldsymbol{\sigma}$ in the state $\rho_{\boldsymbol{w}} = \frac{1}{2}[I + \boldsymbol{w}\cdot\boldsymbol{\sigma}]$ we obtain the result ± 1 with probability*

$$\mathrm{Prob}[\boldsymbol{v}\cdot\boldsymbol{\sigma} = \pm 1] = \tfrac{1}{2}(1 \pm \boldsymbol{v}\cdot\boldsymbol{w}) ,$$

and after the measurement the quibit is in the state $\rho_{\pm\boldsymbol{v}} = \frac{1}{2}[I \pm \boldsymbol{v}\cdot\boldsymbol{\sigma}]$.

11.3 Measurements in composite systems

Let us now consider measurement in a composite system. Suppose that we have two qubits, with Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$ (both equal to $\mathbb{C}^2$). We will label operators on the composite system by the index $j = 1, 2$ to indicate that they act on the first or second qubit. For instance

$$Z_1 = Z \otimes I , \quad Z_2 = I \otimes Z ,$$

and similarly for other Pauli matrices. We will also use the abreviated notation

$$|\sigma\sigma'\rangle = |\sigma\rangle \otimes |\sigma'\rangle \qquad \sigma, \sigma' \in \{0, 1\} .$$

Suppose that by some means our composite system is placed in the pure state corresponding to

$$|\Psi\rangle = \frac{1}{\sqrt{2}}\left(|01\rangle - |10\rangle\right) , \tag{11.1}$$

and that we measure the operator Z_1. One readily computes that the result is ± 1 with probability $1/2$. If we measure 1, the state after measurement is $|00\rangle$, while if we measure -1, the state after measurement is $|11\rangle$. We obtain the same results for measurement of Z_2. Suppose now that we make two measurements in succession: first Z_1 with result $s_1 = \pm$

and then Z_2 with result $s_2 = \pm 1$. This is where things start to get interesting, because the results are correlated: *with absolute certainty, we have* $s_2 = -s_1$.

Let us now consider measuring the spin in a different direction. Consider the operators $v \cdot \sigma_j = v_1 X_j + v_2 Y_j + v_3 Z_j$ with $|v| = 1$ and $j = 1, 2$. The ± 1 spectral projections for this operator are $P_{v,j,\pm 1} = \frac{1}{2}\left[I \pm v \cdot \sigma_j\right]$. We can compute the reduced density matrix for each qubit; they are

$$\rho_1 \;=\; \mathrm{tr}_2 \left|\Psi\right\rangle \left\langle\Psi\right| \;=\; \tfrac{1}{2}\,\mathrm{tr}_2 \left(\left|01\right\rangle \left\langle 01\right| + \left|10\right\rangle \left\langle 10\right| - \left|01\right\rangle \left\langle 10\right| - \left|10\right\rangle \left\langle 01\right|\right) \;=\; \tfrac{1}{2} I$$

and similarly $\rho_2 \;=\; \mathrm{tr}_1 \left|\Psi\right\rangle \left\langle\Psi\right| = \frac{1}{2} I$, by a similar calculation. We see that if we measure *any* operator of the form $v \cdot \sigma_j$ with $|v| = 1$, we obtain ± 1 as outcomes with equal probability. Again, if we measure $v \cdot \sigma_1$ with result s_1 and then $v \cdot \sigma_2$ with result s_2, then *with absolute certainty* $s_2 = -s_1$

So we have the following strange situation. Suppose that Charlie, at his laboratory in Chicago, entangles two quibits, A and B, placing them in the state Ψ in (11.1). Charlie then sends one quibt to Alice in Albuquerque and one to Bob in Bogota. Alice measures the spin of qubit A along the axis v. According to the quantum measurement theory we have proposed, the state of quibit B, in Bogota, is immediately changed so that it is perfectly anti-aligned with the resulting spin from Alice's experiment.

Einstein, Podolsky, and Rosen 1935 argued that the notion that a measurement in one place may instantaneously change the state of a particle far away cannot be a correct description of reality. Based on this they argue that the wave function "does not provide a complete description of physical reality." Einstein remained uncomfortable with this phenomenon, which he termed "spooky action at a distance" in a 1947 letter to Max Born, throuhgout his life.

With regard to the specific measurements described above, there is no need for "spooky action at a distance." Suppose that the actual outcome of the experiments of measuring the spins $v \cdot \sigma_j$, $j = 1, 2$, were determined initially when Charlie prepared the state $|\Psi\rangle$. These describing the outcome are described by some auxiliary "hidden" variables that are not part of the wave function. ALl that is required is that we have a statistical ensemble in which $s_2 = -s_1$ with probability one. There are many such statistical ensembles.

11.4 Bell's inequality

Bell 1964 realized that one could understand this thought experiment more deeply by allowing Alice and Bob to measure the spin along different axes. Suppose we first measure $v \cdot \sigma_2$ for quibit one, obtaining the results $s_1 = \pm 1$ each with probability $\frac{1}{2}$. After measurement, the system is in the pure state with vector

$$\begin{aligned} \left|\Psi\right\rangle_1 \;&=\; \tfrac{1}{2}\left(\left|01\right\rangle - \left|10\right\rangle + s_1 v_1 (\left|11\right\rangle - \left|00\right\rangle) + s_1 i v_2 (\left|11\right\rangle + \left|00\right\rangle) + s_1 v_3 (\left|01\right\rangle + \left|10\right\rangle)\right) \\ &=\; \tfrac{1}{2}\left(\left(1 + s_1 v_3\right)\left|01\right\rangle - \left(1 - s_1 v_3\right)\left|10\right\rangle + s_1 (v_1 + i v_2)\left|11\right\rangle + s_1 (v_1 - i v_2)\left|00\right\rangle\right) \;. \end{aligned}$$

Let us then measure Z_2 for the second quibit, obtaining the results $s_2 = \pm 1$ with probalities:

$$\text{Prob}(s_2 = 1) \;=\; \tfrac{1}{4}\left((1 - s_1 v_3)^2 + v_1^2 + v_2^2\right) \;=\; \tfrac{1}{2}\,(1 - s_1 v_3)$$
$$\text{and} \quad \text{Prob}(s_2 = -1) \;=\; \tfrac{1}{2}\,(1 + s_1 v_3)\ .$$

Using the rotation invariance, we find the following general formula

$$\text{Prob}\left(v_2 \cdot \sigma_2 = s_2 \,|\, v_1 \cdot \sigma_1 = s_1\right) \;=\; \tfrac{1}{2}\,(1 - s_1 s_2 v_1 \cdot v_2)\ . \tag{11.2}$$

Now consider an experiment. Suppose that Charlie sends Alice and Bob many different copies of the entangled qubits. When Alice receives the j^{th} qubit, she decides, at random, to measure $Q = Z_1$ or $R = X_1$, records the result, and notes which measurement she took. When Bob receives the j^{th} quibit, he decides, at random, to measure $S = \frac{-Z_2 - X_2}{\sqrt{2}}$ or $T = \frac{Z_2 - X_2}{\sqrt{2}}$, records the result, and notes the measurement taken as $M_2^{(j)} = S$ or T. When they are done, they get on Zoom and compare their results.

Let us suppose that, as Einstein, Podolsky, and Rosen proposed, there is at the time Charlie produces the entangled pair a joint probability distribution describing the outcome of experiments $p(q, r, s, t)$, for $q, r, s, t \in \{-1, 1\}$, such that if for the j^{th} qubit Alice measures Q and Bob measures S, then

$$\text{Prob}[Q = q \text{ and } S = s] \;=\; \sum_{r,t=\pm 1} p(q, r, s, t)\ ,$$

and similarly for other combinations of measurements. Bell 1964 derived an inequality for the outcome of such experiments, assuming the existence of such a probability function. Bell's inequality was later reformulated by Clauser et al. 1969 in a form we now present. Note that $QS + RS + RT - QT = (Q + R)S + (R - Q)T$. Since $Q, R, S, T = \pm 1$, we have either $Q + R = 0$ or $Q - R = 0$, and thus $(Q + R)S + (R - Q)T = \pm 2$. If we average the possible outcomes we obtain

$$\mathbb{E}(QS) + \mathbb{E}(RS) + \mathbb{E}(RT) - \mathbb{E}(QT) \;=\; \mathbb{E}(QS + RS + RT - QT) \;\leq\; 2\,. \tag{11.3}$$

This is the CHSH inequality, derived in Clauser et al. 1969, sometimes also called *Bell's inequality* (although, Bell 1964 derived a somewhat different, more complicated, result).

Let us return to our thought experiment. Suppose Alice and Bob compare notes after measuring many correlated quibits and compute $\mathbb{E}(QS) + \mathbb{E}(RS) + \mathbb{E}(RT) - \mathbb{E}(QT)$. This is easy to do — they just have to divide their data into four groups corresponding to each type of measurement and average the product of their outcomes for each. If they measure many N quibits, then they should have roughly $\frac{N}{4}$ measurements of each type by the law of large numbers. And if quantum mechanics is correct, then each empirical average should be close to the theoretical average (again, by the law of large numbers).

So what does quantum mechanics say about the theoretical average? Let us start with $\mathbb{E}(QS)$. Since $Q = Z_1$ and $S = -\frac{Z_2 + X_2}{\sqrt{2}}$, we can use eq. (11.2) to compute

$$\mathbb{E}(QS) \;=\; \tfrac{1}{4}\left[\left(1 + \tfrac{1}{\sqrt{2}}\right) - \left(1 - \tfrac{1}{\sqrt{2}}\right) - \left(1 - \tfrac{1}{\sqrt{2}}\right) + \left(1 + \tfrac{1}{\sqrt{2}}\right)\right] \;=\; \frac{1}{\sqrt{2}}\,.$$

By a similar calculuation, we obtain

$$\mathbb{E}(QT) \;=\; -\frac{1}{\sqrt{2}}\,, \quad \mathbb{E}(RS) \;=\; \frac{1}{\sqrt{2}}\,, \quad \text{and} \quad \mathbb{E}(RT) \;=\; \frac{1}{\sqrt{2}}\,.$$

Thus

$$\mathbb{E}(QS) + \mathbb{E}(RS) + \mathbb{E}(RT) - \mathbb{E}(QT) \;=\; 2\sqrt{2} \;>\; 2\,, \tag{11.4}$$

in violation of the CHSH inequality (11.3).

What has happened!? As in any proof by contradiction, we have shown that something we assumed to be true is, in fact, false. In this case, the questionable assumption was the existence of a probability distribution $p(q, r, s, t)$ describing the joint probabilities of possible outcomes. Provided that (11.4) is observed to be true in nature, we conclude that *such a description of the system does not exist at the time when Alice and Bob preform their measurements.* So the underlying variables that Einstein, Podolsky, and Rosen hoped would preclude the "spooky action at a distance" do not exist!

To date numerous experiments have shown that nature *does indeed* violate the CHSH inequality and other Bell type inequalities. Most of these measurements have used entanglement of the polarization in pairs of photons, which is a bit different from but conceptually similar to the thought experiment considered here. The earliest demonstration was by Freedman and Clauser 1972. Entanglement and violation of Bell's inequality have even been demonstrated in pairs of photons emitted 7.8 billion years ago by a distant quasar by Rauch et al. 2018.

Lecture 12

Postulates of quantum mechanics and properties of density matrices

Reading: §I.2 in Nielsen and Chuang

12.1 Postulates of Quantum Mechanics

Dirac, in his 1930 book *The Principles of Quantum Mechanics*, formulated various postulates for the description of quantum mechanics

1. Associated to any isolated quantum system is a Hilbert space $\mathcal{H}$. The states of an isolated quantum system are described by unit vectors $|\psi\rangle$ in the Hilbert space $\mathcal{H}$.
2. Every observable quantity of the system is described by a self-adjoint operator on the Hilbert space $\mathcal{H}$.
3. The result of measuring an observable A must be one of the eigenvalues of A.
4. If the system is in state $|\psi\rangle$, the probability of observing the eigenvalue a in a measurement of the observable A is $\langle\psi| P_a |\psi\rangle$, where P_a is the corresponding orthogonal projection onto the subspace of eigenvectors of eigenvalue a.
5. The time evolution of the system is described by unitary maps $U(t,s)$, such that if the system is in state $|\psi(s)\rangle$ and time s then it is in state $|\psi(t)\rangle = U(t,s)|\psi(s)\rangle$ at time t.

Von Neumann developed the spectral theory of self-adjoint operators in his 1931 paper, and 1932 book *Mathematische Grundlagen der Quantenmechanik*, for the purpose of extending these postulates to operators without discrete spectrum.

These postulates, as stated, do not address composite systems, mixed states, or generalized measurements. The following are updated postulates of Quantum Mechanics, a la Nielsen and Chuang 2010:

> **Postulate 1**: Associated to an isolated quantum system is a complex Hilbert space $\mathcal{H}$. A state of the system is fully described by a *density matrix*, which is a positive operator ρ on $\mathcal{H}$ of trace one.

Postulate 2: The evolution of an isolated quantum system is described by unitary maps $U(t,s)$ such that if the system is in state ρ at time s, then it is in state $U(t,s)\rho U(s,t)^\dagger$ at time t.

Postulate 3: A measurement of the system is described by a *quantum measurement ensemble*, consisting of a finite list of bounded operators $(M_a)_{a\in S}$ on $\mathcal{H}$ satisfying the completeness relation $\sum_{a\in S} M_a^\dagger M_a$. If the system is in state ρ immediately before the measurement, then the outcome $a \in S$ is observed with probability $p_a = \operatorname{tr} M_a^\dagger M_a \rho$ and the state of the system after the measurement is $\frac{1}{p_a} M_a \rho M_a^\dagger$.

Postulate 4: The Hilbert space for a composite system consisting of n distinct parts, each with with Hilbert space $\mathcal{H}_j$, $j = 1, \dots, n$, is the tensor product $\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_2$. If each subsystem is isolated and in state ρ_j, $j = 1, \dots, n$, then the state of the composite system is $\rho_1 \otimes \cdots \otimes \rho_n$.

Nielsen and Chuang formulate these postulates in a finite dimensional context. In infinite dimensions, one should add the requirement that $t \mapsto U(t,s)$ is strongly continuous for each s to postulate 2. The unitary maps $U(t,s)$ necessarily satisfy the composition rule: $U(t,s)U(s,r) = U(t,r)$. If $U(t,s) = V(t-s)$ for some one-parameter family of unitaries $(V(t))_{t\in\mathbb{R}}$, then we say the system is *time translation invariant*. By the composition rule $V(t)V(s) = V(t+s)$, so V is a one-parameter group. It follows from Stone's Theorem 7.5 that $V(t) = \mathrm{e}^{-itH}$ for some self-adjoint operator, called the *Hamiltonian*.

12.2 Interpreting mixed states

There are two ways in which

1. State for part of a composite system.
2. Ensembles of pure states.

Definition 12.1. Ensemble of pure sates $(p_i, |\psi_i\rangle)_{i=1}^n$.

Average result of measurements.
Non-uniqueness in terms of ensemble.

Theorem 12.2. *Two ensembles*

12.3 Schmidt decomposition of pure states

12.4 Purification of mixed states

Lecture 13

Open Quantum Systems

Reading: Ch. 8 of Nielsen and Chuang

Everything we have discussed so far pertains to isolated, or closed, quantum systems, with the exception of measurements. Real systems are *never* perfectly isolated from their environment, so this notion is at best an idealization approximately valid over some time scale.

To develop the ideas here, think of a classical simple pendulum. Let us begin with an idealized, theoretical pendulum of mass m hanging from a rigid massless rod of length L. This system has equation of motion

$$L\frac{d^2}{dt^2}\theta + g\sin(\theta) \;=\; 0$$

where θ is the angle from verticle and g is the gravitational acceleration. This system has a Hamiltonian $H(p,\theta) = \frac{p^2}{2L} + g(1-\cos(\theta))$ where p is the conjugate momentum to θ. If we imagine starting the pendulum at rest $p = 0$ at angle $\theta_0 > 0$. The pendulum will oscillate, first falling to $\theta = 0$, then rising to $\theta = -\theta_0$, and then repeating the motion in reverse. The period is given by the elliptic integral

$$T = 4\sqrt{\frac{L}{2g}}\int_0^{\theta_0}\frac{1}{\sqrt{\cos\theta - \cos\theta_0}}d\theta_0\ ,$$

which for small amplitude $\theta_0 << 1$ gives the classical expression

$$T \;\approx\; 2\pi\sqrt{\frac{L}{g}}\ .$$

Because energy is conserved a theroetical pendulum witll oscillate, periodically over the same trajectory for ever.

Now consider a real pendulum. We would be very surprised if we set it oscillating and came back a year later to find it still moving, with the same amplitude. The real pendulum is subject to friction which damps it oscillations, coming from its interaction with the environment, e.g., friction at the pivot point of the string, air resistance on the ball. Similarly,

if we place our real pendulum at rest, we might find it oscillating slightly at a later time, perhaps due to a seismological tremor or a gust of air.

The pendulum is a key example to keep in mind when we talk about "open systems." An open system is any system subject to *friction* and *noise* due to its interactions with its surroundings. For the pendulum, one might model friction and noise by supposing it to be subject to a *Langevin equation*

$$\frac{d}{dt}p(t) \;=\; -g\sin(\theta) - \nu p + \eta(t) \quad \text{and} \quad L\frac{d}{dt}\theta(t) \;=\; p(t)\,, \tag{13.1}$$

with $\nu > 0$ a friction coefficient and $\eta(t)$ a noise term representing random forces. If one considers small amplitude fluctuations, with $\sin(\theta) \approx \theta$, this equation reduces to the following Langevin equation for a damped, noisy harmonic oscillator:

$$\frac{d}{dt}p(t) \;=\; -g\theta(t) - \nu p(t) + \eta(t) \quad \text{and} \quad L\frac{d}{dt}\theta(t) \;=\; p(t)\,. \tag{13.2}$$

Let us think about the consequences of (13.2) for a mixed state of the oscillator. Let $\rho_t(p,\theta)$ be a non-negative function on phase space representing the probability density of finding the oscillator with that phase at time t. Because of the noise term in the Langevin equation there are two things we might mean by this:

1. We might ask for the probability $\rho_t(p,\theta|\{\eta(s)\}_{s\in\mathbb{R}})$ *conditioned* on the noise. This is the *sample specific* or *quenched* density. It describes the motion of a single oscillator subject to a particular realization of environmental noise.

2. We might ask for the probability $\rho_t(p,\theta)$ to find the oscillator at phase (p,θ) at time t, *averaged over realizations of the noise*. This is the *averaged* or *annealed* density.

A typical noise model is to assume that $\eta(t)$ is *white noise*, meaning that $\eta(t)$ is a Gaussian process satisfying

$$\mathbb{E}\left(\eta(t)\eta(s)\right) \;=\; D\delta(t-s)\,.$$

Such a Gaussian process exists as a distribution valued process $\eta(t) \;=\; \sqrt{D}b'(t)$, where $b(t)$ is the Wiener process. The Wiener process is actually no-where differentiable, and technically the resulting distribution η is not a function with a defined value at points. However, one can interpret (13.2) as a stochastic differential equation

$$dp(t) \;=\; -(g\theta(t) + \nu p(t))dt + \sqrt{D}db(t) \quad \text{and} \quad Ld\theta(t) = p(t)dt\,. \tag{13.3}$$

The resulting equation for the *averaged density* is the *Focker-Planck* equation

$$\frac{\partial\rho_t}{\partial t}(p,\theta) \;=\; D\frac{\partial^2\rho_t}{\partial p^2}(p,\theta) - \frac{p}{L}\frac{\partial\rho_t}{\partial_\theta}(p,\theta) + \frac{\partial}{\partial p}\left((g\theta+\nu p)\rho_t(p,\theta)\right)\,. \tag{13.4}$$

13.1 Markov processes

To obtain a somewhat simplified view point, let us discreteize both space and time. Imagine a classical system with only finitely many possible states $X = \{x_1, \dots, x_n\}$ and consider the system at times $t = 0, 1, \dots$. We may suppose that the closed system is static — if the system is in state x_j at time 0 then the system stays in that state forever. More generally, the motion might be periodic: there might be a permutation $\sigma : X \to X$ such that if the system is in state x_j at time t then at time $t + 1$ is is found in state $\sigma(x_j)$.

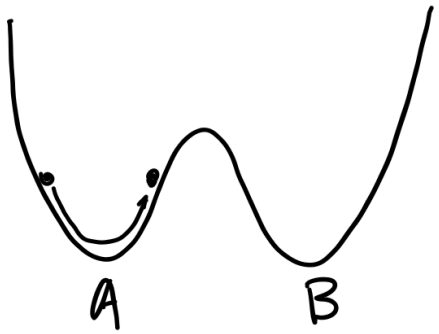


Figure 13.1: A double-well potential with coarse-grained states A and B.

Such a system may come about by *coarse graining* a classical system. For example consider a particle in the double well in Figure 13.1. If we only care about whether the particle is in the left (A) or right (B) well, then we can replace the state space $\{A, B\}$ If the energy is below the internal maximum, then the particle will stay in its given well for all time. Suppose however, that our double well is in a seismic zone and occasionally the potential well shakes such that the particle transitions from one well to the other. Then a trajectory of the system will be a sequence $w_j \in \{A, B\}$ of random points describing which well the particle is in at times $0, 1, 2, \dots$.

A natural assumption to make is that the random process is a *Markov process*, which is to say that the probability distribution of w_j given the trajectory up to time $j - 1$ depends only on w_{j-1}. In symbols,

$$\text{Prob}(w_j = A | (w_1, \dots, w_{j-1})) \;=\; \text{Prob}(w_j = A | w_{j-1}) \,.$$

In this case there is a transition matrix

$$T_j \;=\; \begin{pmatrix} \text{Prob}(w_j = A | w_{j-1} = A) & \text{Prob}(w_j = A | w_{j-1} = B \\ \text{Prob}(w_j = B | w_{j-1} = A) & \text{Prob}(w_j = B | w_{j-1} = B) \end{pmatrix} ,$$

so that

$$\begin{pmatrix} \text{Prob}(w_j = A) \\ \text{Prob}(w_j = B) \end{pmatrix} \;=\; T_j \begin{pmatrix} \text{Prob}(w_{j-1} = A) \\ \text{Prob}(w_{j-1} = B) \end{pmatrix} . \tag{13.5}$$

Eq. (13.5) is the analogue of the Fokker-Planck eq. (13.4) for such discrete Markov processes.

More generally, given a finite state space $X = \{x_1, \ldots, x_n\}$, an initial point $w_0 \in X$, and a sequence $T_1, T_2, \ldots$ of $n \times n$ transition matrices, satisfying

1. $\langle j| T_m |k\rangle \geq 0$, and
2. $\sum_j \langle j| T_m |k\rangle = 1$,

there is a Markov process given by a random sequence of points $w_1, w_2, \ldots$ such that

$$\boldsymbol{p}_m \;=\; T_m \boldsymbol{p}_{m-1} \tag{13.6}$$

where $\boldsymbol{p}_m = (\text{Prob}(w_m = x_1), \ldots, \text{Prob}(w_m = x_n))^T$.

13.2 Operator-Sum Representation of an Open Quantum System

Consider now a quantum system S with finite dimensional Hilbert space $\mathcal{H}_S$. Suppose the system is interacting with an environment R, with Hilbert space $\mathcal{H}_R$.[1] We will suppose that the Hilbert space $\mathcal{H}_R$ is also finite dimensional.

We will also suppose that the initial state of the compositie system is a product state $\rho \otimes \rho_R$. This assumption is natural, because we imagine that S is a system that we can isolate and manipulate. If so, when we prepare the system in some state it is initially not entangled with the environment. At a later time, the compound system is in state

$$U(\rho \otimes \rho_R)U^\dagger$$

where $U = \mathrm{e}^{-itH}$ with H the Hamiltonian for the compound system (or perhaps the propagator for a time dependent dynamics). In general the Hamiltonian contains terms coupling the two systems and the state at time t is not a product state. The state of the system at time t is obtained by taking the partial trace

$$\rho' \;=\; \mathrm{tr}_R \left[U(\rho \otimes \rho_R)U^\dagger \right] \;. \tag{13.7}$$

Note that (13.7) can be expressed as $\rho' = \mathcal{E}(\rho)$, where $\mathcal{E} : \mathfrak{A}_1^+ \to \mathfrak{A}_1^+$ is the map on the left hand side. The map $\mathcal{E}$ is the analogue in the context of quantum mechanics of the Markov transition matrices discussed above.

Eq. (13.7) is completely general. It applys to any open system that starts in a product state with its environment. However, it is not very useful. We would prefer to have an intrinsic representation of the transition operator. This can be accomplished as follows. Let $|1; R\rangle, \ldots |N; R\rangle$ be an orthonormal basis for the environment space $\mathcal{H}_R$ consisting of eigenvectors of ρ_R. Thus

$$\rho_R \;=\; \sum_{k=1}^{N} \lambda_k \, |k; R\rangle \langle k; R| \;.$$

[1]The symbol R for the environment stands for "Reservoir." This comes from thermodynamics, where often the environment plays the role of a heat reservoir bringing the system into thermal equilibrium.

Inserting this into eq. (13.7), we have

$$\mathcal{E}(\rho) \;=\; \sum_k \lambda_k \,\mathrm{tr}_R \left[U\rho \otimes |k;R\rangle \langle k;R| \, U^\dagger \right] \;.$$

For $n,k \in \{1,\ldots,N\}$ define the operator $E(n,k) = \sqrt{\lambda_k}\,\langle n;R|\,U\,|k;R\rangle$ on $\mathcal{H}_S$ by

$$\langle \psi|\, E(n,k)\,|\phi\rangle \;=\; \sqrt{\lambda_k}\,\langle \psi| \otimes \langle n;R|\,U\,|\phi\rangle \otimes |k;R\rangle \;.$$

Then we have the following *operator sum representation* for $\mathcal{E}$:

$$\mathcal{E}(\rho) \;=\; \sum_{n,k=1}^{N} E(n,k)\rho E(n,k)^\dagger \;. \tag{13.8}$$

The advantage of (13.8) is that the Hilbert space of the environment is completely absent — the operators $E(n,k)$ act only on $\mathcal{H}_S$.

Lecture 14

Open Quantum Systems II

Reading: Nielsen and Chuang 2010, Ch. 8.

14.1 Interpreting the operator sum representation

We have seen that the dynamics of an open system, over an discrete interval of time, can be given the *operator sum representation*

$$\mathcal{E}(\rho) \;=\; \sum_{k=1}^{n} E(k)\rho E(k)^\dagger \,, \tag{14.1}$$

where $E(k)$ are operators on the system Hilbert space $\mathcal{H}_S$ satisfying the completeness relation

$$\sum_{k=1}^{n} E(k)^\dagger E(k) \;=\; I \,. \tag{14.2}$$

The completeness relation guarantees that the map $\mathcal{E}$ is trace preserving.

A useful generalization of the above is to consider $\mathcal{E}$ of the form (14.1) but with

$$\sum_{k=1}^{n} E(k)^\dagger E(k) \;\leq\; I \,. \tag{14.3}$$

We will call such a map a *quantum operation*. The trace of the density matrix goes down under a quantum operation:

$$\operatorname{tr}\mathcal{E}(\rho) \;=\; \sum_{k=1}^{n} \operatorname{tr} E(k)\rho E(k)^\dagger \;=\; \sum_{k=1}^{n} \operatorname{tr} E(k)^\dagger E(k)\rho \;\leq\; \operatorname{tr}\rho \,.$$

This allows us to interpret $\operatorname{tr}\mathcal{E}(\rho)$ as the probability that the operation occurs.

For example, consider a scattering experiment in which our system is a single quantum particle interacting with a target. Perhaps we are using the particle to observe properties of the target. Figure 14.1 shows the incoming particle approaching from the left. We have a detector that can capture the particle if it is scattered to the right into the detector. If it is

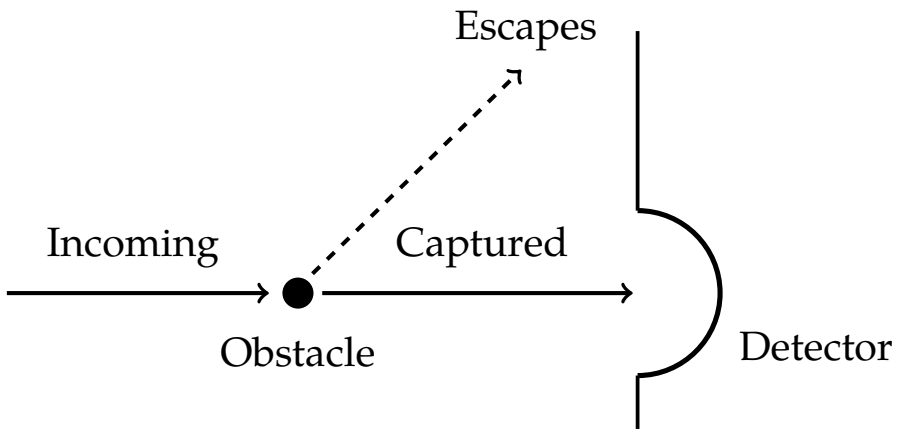


Figure 14.1: Scattering outcomes: capture by the detector or escape.

scattered upward, it misses the detector and is lost. It is natural to associate to this system a quantum operation

$$\mathcal{E}(\rho) \;=\; \mathrm{tr}_R\, P_D U \rho \otimes \rho_R U^\dagger P_D$$

where U is the dynamics and P_D is a projector indicating that the particle is in the projector after the experiment. The trace $\mathrm{tr}\,\mathcal{E}(\rho)$ gives the parobability that the particle is captured by the detector.

There is an elegant interpretation of what happens in a quantum operation. Suppose that the environment is in the pure state $|e_1\rangle$ and let $|e_1\rangle, \ldots, |e_n\rangle$ be an orthonormal basis for the environment Hilbert space. There is actually no loss in assuming the environment to be in a pure state, since we can always imagine it to be part of a larger Hilbert space on which the state is pure, as shown above. An operator sum representation for the dynamics can be given by

$$\mathcal{E}(\rho) \;=\; \mathrm{tr}_R\, U\rho \otimes |e_1\rangle\langle e_1| U^\dagger \;=\; \sum_{k=1}^{n} \langle e_k| U |e_1\rangle \rho \langle e_1| U^\dagger |e_k\rangle \;=\; \sum_{k=1}^{n} E_k \rho E_k^\dagger$$

with $E_k = \langle e_k| U |e_1\rangle$. We can think of this sum as follows

$$\mathcal{E}(\rho) \;=\; \sum_{k=1}^{n} p(k)\rho_k \tag{14.4}$$

where

$$p(k) \;=\; \mathrm{tr}\, E_k \rho E_k^\dagger \;=\; \mathrm{tr}\, \langle e_k| U |e_1\rangle \rho \langle e_1| U^\dagger |e_k\rangle$$

is the probability of finding the environment in state $|e_k\rangle$ and

$$\rho_k \;=\; \frac{E_k \rho E_k^\dagger}{p(k)}\,.$$

Thus the action of the (trace preserving) quantum operator $\mathcal{E}$ is equivalent to picking $k \in \{1, \ldots, n\}$ at random with distribution $p(k)$ and replacing the state ρ by ${}^{E_k \rho E_k^\dagger}/_{p(k)}$. This

interpretation is reminiscent of the evolution of Markov chains, such as have been used in classical information theory to model the transmission of information on noisy channels. For this reason trace preserving quantum operations are revered to as *quantum channels*. In fact, given an operator sum representation of a quantum channel we can define a Markov chain on density matrices with transition map

$$\text{Prob}(\rho \mapsto {}^{E_k \rho E_k^\dagger}/_{\text{tr}\, E_k \rho E_k^\dagger}) \;=\; \text{tr}\, E_k \rho E_k^\dagger \,.$$

We will return to these chains letter when we discuss trajectories for repeated measurements

Finally, note that (14.4) is the same as the process we obtain if we measure the environment in the projective measurement ensemble $(\left|e_j\right\rangle \left\langle e_j\right|)_{j=1}^n$. In some sense, the system is observing the environment. Similarly, this is the process we obtain by a generalized measurement of the system in the ensemble $(E_k)_{k=1}^n$.

14.2 Completely Positive Maps

Let us now take a different approach to the dynamics of an open quantum system. Suppose that we have a given system with Hilbert space $\mathcal{H}_S$ and a transition map $\mathcal{E}$ such that if the system has state ρ at time 0 then at time t it is found with state $\mathcal{E}(\rho)$.

We may wish to allow our map $\mathcal{E}$ to take operators on one Hilbert space to operators on another Hilbert space. For instance, in the scattering experiment sketched above perhaps our particle is absorbed by the detector and some property of the particle (perhaps its spin) becomes encoded in the quantum state of some part of the detector. Then we wish to consider an operation taking density matrices on $\mathcal{H}_S$ (the particle Hilbert space) to density matrices on $\mathcal{H}_D$ (the Hilbert space of the spin in our detector). These Hilbert spaces may even have different dimension.

Let $\mathcal{H}_I$ and $\mathcal{H}_F$ denote a pair of Hilbert spaces and let $\mathcal{D}_I$ denote the density matrices on $\mathcal{H}_I$. We want to consider a map $\mathcal{E} : \mathcal{D}_I \to \mathcal{B}(\mathcal{H}_F)$ defining a quantum operation. What sort of properties should we assume about $\mathcal{E}$?

1. We should have $\text{tr}\, \mathcal{E}(\rho) \leq 1$ for any $\rho \in \mathcal{D}_I$.
2. We expect that $\mathcal{E}$ is *convex linear*: $\mathcal{E}(t\rho_1 + (1-t)\rho_0) = t\mathcal{E}(\rho_1) + (1-t)\mathcal{E}(\rho_0)$.

The first property is very natural since, as we have seen, it allows us consider processes that include loss (like the scattering experiment sketched above). The requirement that $\text{tr}\, \mathcal{E}(\rho) \leq 1$ stems from the interpretation of this trace as the probability that the operation $\mathcal{E}$ occured.

The second property is physical. If we select a state ρ_i at random from an ensemble $\{p_i, \rho_i\}_{i=1}^n$ of states, then the average behavior should be described by the action of $\mathcal{E}$ on the density matrix $\rho = \sum_i p_i \rho_i$. This is convex linearity.

We extend the action of $\mathcal{E}$ to any positive operator Q by defining $\mathcal{E}(Q) = \text{tr}\, Q \mathcal{E}({}^{Q}/_{\text{tr}\, Q})$. This is consistent with (1) and (2). We can further extend $\mathcal{E}$ to act on any observable A by $\mathcal{E}(A) = \mathcal{E}(A_+) - \mathcal{E}(A_-)$, with $A_\pm$ the positive parts of A. Because of (2), this extended map is *linear*.

There is one additional property needed. Namely, $\mathcal{E}(\rho)/\operatorname{tr}\mathcal{E}(\rho)$ should be a density matrix for any density matrix ρ. For this, we must have $\mathcal{E}(\rho)$ a positive operator. However, quantum mechanics is more subtle than this. For our system, perhaps unknown to us, might be entangled with some other system, with Hilbert space $\mathcal{H}_O$ (O for "other"). The true state of the composite system (I, O) might be a density matrix Q that cannot be written as a product. If the operation $\mathcal{E}$ applies to the system, this results in the map $\mathcal{E} \otimes \mathcal{I}$ applied to Q. Here the map $\mathcal{E} \otimes \mathcal{I}$ is defined by

$$\mathcal{E} \otimes \mathcal{I}(A \otimes B) \;=\; \mathcal{E}(A) \otimes B$$

for tensor products of operators. Since any operator on $\mathcal{H}_I \otimes \mathcal{H}_O$ is a linear combination of product operators, we can extend $\mathcal{E} \otimes \mathcal{I}$ to all operators (in particular density matrices) on $\mathcal{H}_I \otimes \mathcal{H}_O$ by linearity. Note that we need linearity and non-positive operators here. It is not possible to write an arbitrary density matrix on $\mathcal{H}_I \otimes \mathcal{H}_O$ as a *convex* combination of product states. The resulting extension had better map positive operators to positive operators, else we run into a problem interpreting $\mathcal{E} \otimes \mathcal{I}(Q)/\operatorname{tr}\mathcal{E} \otimes \mathcal{I}(Q)$ as a density matrix.

3. We expect $\mathcal{E}$ to be a *completely positive map*. That is if $\mathcal{H}_O$ is any other system (of any dimension) then the map $\mathcal{E} \otimes \mathcal{I}$ takes positive operators to positive operators.

Theorem 14.1. *Assume that* $\dim \mathcal{H}_I$ *and* $\dim \mathcal{H}_F < \infty$. *A map* $\mathcal{E} : \mathcal{B}(\mathcal{H}_I) \to \mathcal{B}(\mathcal{H}_F)$ *satisfies* (1), (2), *and* (3) *above if and only if*

$$\mathcal{E}(\rho) \;=\; \sum_{i=1}^{n} E_i \rho E_i^\dagger \tag{14.5}$$

for some set of operators $E_i \in \mathcal{B}(\mathcal{H}_I, \mathcal{H}_F)$ *with* $\sum_i E_i^\dagger E_i \le I$.

Remark. This is sometimes called *Kraus's Theorem*. Kraus 1971 proved a more general result, which includes infinite dimensional Hilbert spaces.

Proof. Suppose that $\mathcal{E}$ has the form (14.5). The $\mathcal{E}$ is clearly linear, so convex-linear on $\mathcal{D}_I$. Furthermore

$$\operatorname{tr}\mathcal{E}(\rho) \;=\; \sum_{i=1}^{n} \operatorname{tr} E_i \rho E_i^\dagger \;=\; \sum_{i=1}^{n} \operatorname{tr} E_i^\dagger E_i \rho \;\le\; \operatorname{tr}\rho \;=\; 1$$

for $\rho \in \mathcal{D}_I$. We need to show that $\mathcal{E}$ is completely positive. Let Q be a positive operator on the extended system $\mathcal{H}_I \otimes \mathcal{H}_O$. Let $|\Psi\rangle \in \mathcal{H}_F \otimes \mathcal{H}_O$ and let $|\phi_i\rangle = E_i^\dagger \otimes I_O |\Psi\rangle \in \mathcal{H}_I \otimes \mathcal{H}_O$. Then

$$\langle\Psi| (E_i \otimes I_O) Q (E_i^\dagger \otimes I_O) |\Psi\rangle \;=\; \langle\phi_i| Q |\phi_i\rangle \;\ge\; 0\,,$$

so

$$\langle\Psi| \mathcal{E}(Q) |\Psi\rangle \;=\; \sum_i \langle\Psi| (E_i \otimes I_O) Q (E_i^\dagger \otimes I_O) |\Psi\rangle \;\ge\; .$$

That is $\mathcal{E}(Q)$ is a positive operator, so $\mathcal{E}$ is completely positive.

Suppose now that $\mathcal{E}$ satisfies (1), (2) and (3). We need to construct the operator sum representation for $\mathcal{E}$. Let R be an auxiliary system with $\dim \mathcal{H}_R = \dim \mathcal{H}_I$. Let $(|j_R\rangle)_{j=1}^n$ and $(|j_I\rangle)_{j=1}^n$ be orthonormal bases for $\mathcal{H}_R$ and $\mathcal{H}_I$. Define the vector

$$|\alpha\rangle = \sum_{j=1}^{n} |j_I\rangle \otimes |j_R\rangle \ . \tag{14.6}$$

Now let

$$\sigma = \mathcal{E} \otimes \mathcal{I}_R(|\alpha\rangle \langle\alpha|) \ .$$

σ is a bounded operator on $\mathcal{H}_F \otimes \mathcal{H}_O$.

Given $|\psi\rangle = \sum_j \psi_j |j_I\rangle$ in $\mathcal{H}_I$, let $|\widetilde{\psi}\rangle = \sum_j \psi_j |j_R\rangle$ in $\mathcal{H}_R$ and consider the operator $\langle\widetilde{\psi}| \sigma |\widetilde{\psi}\rangle$ on $\mathcal{H}_F$. We have

etc...

□

Lecture 15

Non uniqueness of the operator sum representation and Examples of Quantum Operations

15.1 Non uniqueness of the operator sum representation

Lecture 16

Quantum Markov Semigroups

16.1 The true quantum evolution of an open system

Consider a quantum system S, with Hilbert space $\mathcal{H}_S$, interacting with an environment E, with Hilbert space $\mathcal{H}_E$. The total Hamiltonian H_T (T for "total") of the system+environment is assumed to be of the form

$$H_T \;=\; H_S \otimes \mathbb{1}_E + \mathbb{1}_S \otimes H_E + gH_I$$

where

1. H_S is the *system Hamiltonian*,
2. H_E is the *environment Hamiltonian*, and
3. gH_I is the *interaction Hamiltonian*, with g a coupling constant assumed to be small.

The density matrix ρ_T for the system+environment satisfies the Schrödinger equation

$$\dot{\rho}_T(t) \;=\; -i\left[H_T, \rho_T(t)\right] \,.$$

Because of the coupling gH_I, this equation is typically difficult to solve even if we know exact solutions to the Schrödinger equations for the system and environment separately.

We now introduce the *ineraction picture* for the dynamics. Let $H_0 = H_S \otimes \mathbb{1}_E + \mathbb{1}_S \otimes H_E$ and set

$$\hat{\rho}_T(t) \;=\; \mathrm{e}^{itH_0}\rho_T(t)\mathrm{e}^{-itH_0} \,.$$

So $\hat{\rho}_T(t)$ is the initial density matrix we would need to take at time 0 for the non-interacting dynamics to reach the final state $\rho_T(t)$ at time t. A short calculation shows that

$$\frac{d\hat{\rho}(t)}{dt} \;=\; -ig\left[\widehat{H}_I(t), \hat{\rho}_T(t)\right] \,,$$

where

$$\widehat{H}_I(t) \;=\; \mathrm{e}^{itH_0}H_I\mathrm{e}^{-itH_0} \,. \tag{16.1}$$

Eq. (16.1) is the Schrödinger equation in the *interaction picture*. Integrating, and noting that $\hat{\rho}_T(0) = \rho_T(0)$, we see that it is equivalent to the integral equation

$$\hat{\rho}_T(t) \;=\; \rho_T(0) - ig\int_0^t ds\,\big[\widehat{H}_I(s), \hat{\rho}_T(s)\big]\;. \tag{16.2}$$

This equation is the starting point for the development of formal perturbation theory for interacting systems. It can be iterated to produce a formal series solution

$$\hat{\rho}_T(t) \;=\; \rho_T(t) + \sum_{n=1}^{\infty}(-ig)^n \int_{\Sigma_n(t)} \big[\widehat{H}_I(s_n), \big[\widehat{H}_I(s_{n-1}), \cdots \big[\widehat{H}_I(s_1), \rho_T(0)\big]\cdots\big]\big]\, ds_1 \ldots ds_n\,,$$

where $\Sigma(t) \;=\; \{(s_1, \cdots, s_n) : 0 \le s_1 \le \cdots \le s_n \le t\}$. If the interaction Hamiltonian is bounded, then this series is absolutely convergent and allows one to interpret the total evolution as a sequence of interactions between the system and the environment. As beautiful as this formula and its interpretation are, they are not so useful to directly understand the evolution of the system itself, since they give the dynamics of the system+environment.

Instead of expanding in a series, we may use (16.2) in (16.1) to obtain

$$\frac{d\hat{\rho}_T(t)}{dt} \;=\; -ig\big[\widehat{H}_I(t), \rho(0)\big] - g^2\int_0^t ds\,\big[\widehat{H}_I(t), \big[\widehat{H}_I(s), \hat{\rho}_T(s)\big]\big]\;. \tag{16.3}$$

Eq. (16.3) is *exact*, but still intractible. We are interested in the *system* density matrix, given (in the ineraction picture) by $\hat{\rho}_S(t) = \mathrm{tr}_E\,\hat{\rho}_T(t)$. If we now take the partial trace we are left with

$$\frac{d\hat{\rho}_S(t)}{dt} \;=\; -ig\,\mathrm{tr}_E\left(\big[\widehat{H}_I(t), \rho(0)\big]\right) - g^2\int_0^t ds\,\mathrm{tr}_E\left(\big[\widehat{H}_I(t), \big[\widehat{H}_I(s), \hat{\rho}_T(s)\big]\big]\right)\;, \tag{16.4}$$

which is still exact, and still impractical as the right hand side depends on the total density matrix $\hat{\rho}_T$.

We would really like an autonomous equation for the system density matrix $\hat{\rho}_S(t)$, hopefully something that is local in time

$$\frac{d\hat{\rho}_S(t)}{dt} \;=\; \mathcal{L}(\hat{\rho}_S(t))\,, \tag{16.5}$$

where $\mathcal{L}$ is a "super-operator" mapping density matrices to density matrices. To get there we will need to make some approximations. But first let us look at an example.

16.2 An example

The simplest example we might take is for both S and E to be a qubit, with $\mathcal{H}_S = \mathcal{H}_E = \mathbb{C}^2$. Let us take $H_S = -Z \otimes \mathbb{1}$ and $H_E = -\mathbb{1} \otimes Z$ with Z = Pauli z matrix. For an interaction Hamiltonian, let we take $H_I = X \otimes X$. With respect to the basis

$$|e_1\rangle = |0,0\rangle\,, \quad |e_2\rangle = |1,0\rangle\,, \quad |e_3\rangle = |0,1\rangle\,, \quad |e_4\rangle = |1,1\rangle\;,$$

the matrices of these operators are

$$H_S + H_E \;=\; \begin{pmatrix} -2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix} \quad \text{and} \quad H_I \;=\; \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

Thus the total Hamiltonian is

$$H_T \;=\; \begin{pmatrix} -2 & 0 & 0 & g \\ 0 & 0 & g & 0 \\ 0 & g & 0 & 0 \\ g & 0 & 0 & 2 \end{pmatrix} \cong \begin{pmatrix} -2 & g \\ g & 2 \end{pmatrix} \oplus \begin{pmatrix} 0 & g \\ g & 0 \end{pmatrix}.$$

Thus the eigenvalues and corresponding eigenvectors of H_T are

$$\lambda_0 \;=\; -\sqrt{4+g^2}\,, \quad |v_0\rangle \;=\; \frac{1}{\sqrt{2\sqrt{4+g^2}}}\left(-\sqrt{\sqrt{4+g^2}+2}\,|0,0\rangle + \frac{1}{\sqrt{\sqrt{4+g^2}+2}}\,|1,1\rangle\right),$$

$$\lambda_1 \;=\; -g\,, \quad |v_1\rangle \;=\; \frac{1}{\sqrt{2}}\left(|1,0\rangle - |0,1\rangle\right), \quad \lambda_2 \;=\; g\,, \quad |v_2\rangle \;=\; \frac{1}{\sqrt{2}}\left(|1,0\rangle + |0,1\rangle\right), \text{and}$$

$$\lambda_3 \;=\; \sqrt{4+g^2}\,, \quad |v_3\rangle \;=\; \frac{1}{\sqrt{2\sqrt{4+g^2}}}\left(-\sqrt{\sqrt{4+g^2}-2}\,|0,0\rangle + \frac{1}{\sqrt{\sqrt{4+g^2}-2}}\,|1,1\rangle\right).$$

Thus the unitary propagator is

$$\begin{aligned} U_t \;=\mathrm{e}^{-itH_T} \;&=\; \mathrm{e}^{it\sqrt{4+g^2}}\,|v_0\rangle\langle v_0| \;+\; \mathrm{e}^{itg}\,|v_1\rangle\langle v_1| \\ &\quad+\; \mathrm{e}^{-itg}\,|v_2\rangle\langle v_2| \;+\; \mathrm{e}^{-it\sqrt{4+g^2}}\,|v_3\rangle\langle v_3| \\ &= \begin{pmatrix} \cos\left(\sqrt{4+g^2}t\right) + \frac{2i}{\sqrt{4+g^2}}\sin\left(\sqrt{4+g^2}t\right) & 0 & 0 & -\frac{ig}{\sqrt{4+g^2}}\sin\left(\sqrt{4+g^2}t\right) \\ 0 & \cos(gt) & -i\sin(gt) & 0 \\ 0 & -i\sin(gt) & \cos(gt) & 0 \\ -\frac{ig}{\sqrt{4+g^2}}\sin\left(\sqrt{4+g^2}t\right) & 0 & 0 & \cos\left(\sqrt{4+g^2}t\right) - \frac{2i}{\sqrt{4+g^2}}\sin\left(\sqrt{4+g^2}t\right) \end{pmatrix}. \end{aligned}$$

Suppose that the initial density matrix is of the form

$$\rho(0) = \begin{pmatrix} a & 0 \\ 0 & 1-a \end{pmatrix} \otimes \langle 0|0\rangle \;=\; \begin{pmatrix} a & 0 & 0 & 0 \\ 0 & 1-a & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

and

$$\rho(t) = U_t\rho(0)U_t^\dagger = a\begin{pmatrix} |F(t)|^2 & 0 & 0 & \frac{ig}{\sqrt{4+g^2}}F(t)\sin\left(\sqrt{4+g^2}t\right) \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -\frac{ig}{\sqrt{4+g^2}}F(t)^*\sin\left(\sqrt{4+g^2}t\right) & 0 & 0 & \frac{g^2}{4+g^2}\sin^2(\sqrt{4+g^2}t) \end{pmatrix}$$
$$+ (1-a)\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \cos^2(gt) & i\cos(gt)\sin(gt) & 0 \\ 9 & -i\cos(gt)\sin(gt) & \sin^2(gt) & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

with $F(t) = \cos\left(\sqrt{4+g^2}t\right) + \frac{2i}{\sqrt{4+g^2}}\sin\left(\sqrt{4+g^2}t\right)$. The reduced density matrix for the system is thus given by

$$\begin{aligned} \rho_S(t) &= \mathrm{tr}_E\,\rho(t) \\ &= a\begin{pmatrix} 1-\frac{g^2}{4+g^2}\sin^2(\sqrt{4+g^2}t) & 0 \\ 0 & \frac{g^2}{4+g^2}\sin^2(\sqrt{4+g^2}t) \end{pmatrix} + (1-a)\begin{pmatrix} \sin^2(gt) & 0 \\ 0 & \cos^2(gt) \end{pmatrix} \\ &= \begin{pmatrix} a & 0 \\ 0 & 1-a \end{pmatrix} + \left((1-a)\sin^2(gt) - a\frac{g^2\sin^2(\sqrt{4+g^2}t)}{4+g^2}\right)Z\,. \end{aligned}$$

Exercise 16.1. Show that the evolution on the right hand side is not the result of an autonomous equation $\cdot\rho_S(t) = \mathcal{L}(\rho_S(t))$ with $\mathcal{L}$ a linear operator on the space of 2×2 matrices.

16.3 Heuristic derivation of the Markov semigroup

In this section we will derive an approximation to the true evolution, known as a "Lindblad master equation." We follow the approach in Manzano 2020. To proceed we will make several approximations. First, note that

$$\hat\rho_T(s) = \hat\rho_T(t) + ig\int_s^t \left[\widehat{H}_I(r), \hat\rho_I(t)\right]dr = \hat\rho_T(t) + O(g)\,. \tag{16.6}$$

Thus

$$\frac{d\hat\rho_T(t)}{dt} = -ig\left[\widehat{H}_I(t), \rho(0)\right] - g^2\int_0^t ds\left[\widehat{H}_I(t), \left[\widehat{H}_I(s), \hat\rho_T(t)\right]\right] + O(g^3)\,. \tag{16.7}$$

In this approximation, we have neglected the dependence of the $O(g^3)$ term on t. If we are more careful we would see that it is $O(g^3t)$ and that this approximation is rigrously valid only if $g << t^{-1/3}$.

To go any further, we need to specify the form of $\rho(0)$ and H_I. Regarding the initial state we shall assume

Assumption 16.1. The initial state $\rho(0)$ is a product state $\rho(0) = \rho_S(0) \otimes \rho_E$, with $\rho_S(0)$ an arbitrary initial state on $\mathcal{H}_S$ and ρ_E a state on $\mathcal{H}_E$ satisfying $[H_E, \rho_E] = 0$.

This is quite natural if the system is prepared in state S and then "brought into contact" with the environment at equilibrium — if the environment is in equilibrium (without contact with the system) then we should have $[H_E, \rho_E] = 0$, so that ρ_E is a time independent solution of the Schrödinger equation.

As for the interaction Hamiltonian, it can always be expressed in the form

$$H_I \;=\; \sum_i A_i \otimes B_i \tag{16.8}$$

with $A_i \in \mathcal{B}(\mathcal{H}_S)$ and $B_i \in \mathcal{B}(\mathcal{H}_E)$. The individual operators A_i and B_i do not need to be self-adjoint, but H_I must be. Thus we must have

$$H_I \;=\; \sum_i A_i^\dagger \otimes B_i^\dagger \,. \tag{16.9}$$

We shall assume that

$$\operatorname{tr} B_i \rho_E(0) \;=\; 0 \,. \tag{16.10}$$

There is no loss of generality in assuming (16.10): if it does not hold then

$$H_I \;=\; \sum_i A_i \otimes \widetilde{B}_i \;+\; \left(\sum_i b_i A_i \right) \otimes \mathbb{1}_E$$

where $b_i = \operatorname{tr} B_i \rho_E(0)$ and $\widetilde{B}_i = B_i - b_i \mathbb{E}$. The second term on the right can be incorporated into the system Hamiltonian H_S.

We have freedom in the choice of the terms in the expansion (16.8). We will make a special choice of the A_i's. To proceed, let us work in the eigenbasis for H_S on $\mathcal{H}_S$, $|\alpha\rangle$, $\alpha = 1, \ldots, d$, with eigenvalues E_α and let $\mathcal{F} = \{E_\alpha - E_{\alpha'} : \alpha, \alpha' = 1, \ldots, d\}$ denote the spectrum of energy differences for the system. We can expand each operator A_i as a sum

$$A_i \;=\; \sum_{\omega \in \mathcal{F}} A_i(\omega)$$

where $\langle \alpha | A_i(\omega) | \beta \rangle \;=\; I[E_\alpha - E_\beta = \omega] \langle \alpha | A_i | \beta \rangle$ so that $A_i(\omega)$ is an eigenvector for the commutator with H_S: $[H_S, A_i(\omega)] = \omega A_i(\omega)$. Thus we have

$$H_I \;=\; \sum_i \sum_\omega A_i(\omega) \otimes B_i \;=\; \sum_i A_i^\dagger(\omega) \otimes B_i^\dagger \,, \tag{16.11}$$

so that

$$\widehat{H}_I(t) \;=\; \sum_{i,\omega} \mathrm{e}^{it\omega} A_i(\omega) \otimes \hat{B}_i(t) \;=\; \sum_{i,\omega} \mathrm{e}^{-it\omega} A_i^\dagger(\omega) \otimes \hat{B}_i(t) \tag{16.12}$$

where $\hat{B}_i(t) = \mathrm{e}^{itH_E} B_i \mathrm{e}^{-itH_E}$.

With these assumptions, we return to (16.4). The first term on the right hand side of (16.4) is now seen to be zero. Indeed it is a linear combination of terms of the form

$$\begin{aligned}\mathrm{tr}_E\left(\left[A_i(\omega)\otimes\hat{B}_i(t),\rho_S(0)\otimes\rho_E\right]\right) &= \mathrm{tr}_E\left(\left[A_i(\omega)\otimes\hat{B}_i(t),\rho_S(0)\otimes\rho_E\right]\right)\\ &= A_i(\omega)\rho_S(0)\,\mathrm{tr}\left(\hat{B}_i(t)\rho_E\right)-\mathrm{tr}\left(\rho_E\hat{B}_i(t)\right)\rho_S(0)A_i(\omega)\\ &= \mathrm{tr}\left(\mathrm{e}^{-itH_E}\rho_E\mathrm{e}^{itH_E}B_i\right)\left[A_i(\omega),\rho_S(0)\right]\\ &= \mathrm{tr}\left(\rho_E B_i\right)\left[A_i(\omega),\rho_S(0)\right] \;=\; 0\,,\end{aligned}$$

where we have used cyclicity of the trace over E and the assumption that $[H_E,\rho_E]=0$. Thus, (16.4) reduces to

$$\frac{d\hat{\rho}_S(t)}{dt} \;=\; -g^2\int_0^t ds\,\mathrm{tr}_E\left(\left[\widehat{H}_I(t),\left[\widehat{H}_I(s),\hat{\rho}_T(t)\right]\right]\right)+O(g^3)\,. \tag{16.13}$$

We now introduce the following

Assumption 16.2. The density matrix $\hat{\rho}_T(t)$ is approximately the product $\hat{\rho}_S(t)\otimes\rho_E$.

Clearly, we cannot literally have $\hat{\rho}_T(t)=\hat{\rho}_S(t)\otimes\rho_E$. This physically motivated assumption is that the difference is neglible. For this to be reasonable the environment should be large, so that the overall effect of the system is small, and the environment should be *dispersive* so that local perturbations of the environment "radiate away to infinity." For the time begin we will simply take this as a formal approximation and see where it leads.

With the product state approximation, we have

$$\frac{d\hat{\rho}_S(t)}{dt} \;\approx\; -g^2\int_0^t ds\,\mathrm{tr}_E\left(\left[\widehat{H}_I(t),\left[\widehat{H}_I(s),\hat{\rho}_S(t)\otimes\rho_E\right]\right]\right)\,. \tag{16.14}$$

We now change variables $s\mapsto t-s$ in the integral and replace the domain of integration by $[0,\infty)$ — this last approximation is reasonable if we are concerned with the long time behavior of solutions. The resulting equation is

$$\frac{d\rho_S(t)}{dt} \;\approx\; -g^2\,\mathrm{tr}_E\left(\left[\widehat{H}_I(t),\left[\int_0^\infty ds\widehat{H}_I(t-s),\rho_S(t)\otimes\rho_E\right]\right]\right)\,. \tag{16.15}$$

We now expand $\widehat{H}_I$ using (16.12) and consider the four terms of the double commutator on the right hand side of (16.15). The first term of the commutator is

$$\begin{aligned}&\int_0^\infty \mathrm{tr}_E\left(\widehat{H}_I(t)\widehat{H}_I(t-s)\hat{\rho}_S(t)\otimes\rho_E\right)ds\\ &\qquad= \sum_{i,j,\omega,\omega'}\left[\mathrm{e}^{-it\omega'}\int_0^\infty \mathrm{e}^{i(t-s)\omega}\,\mathrm{tr}\left(\hat{B}_j^\dagger(t)\hat{B}_i(t-s)\rho_E\right)ds\right]A_j^\dagger(\omega')A_i(\omega)\hat{\rho}_S(t)\\ &\qquad= \sum_{i,j,\omega,\omega'}\left[\mathrm{e}^{it(\omega-\omega')}\int_0^\infty \mathrm{e}^{-is\omega}\,\mathrm{tr}\left(\hat{B}_i(-s)\rho_E B_j^\dagger\right)ds\right]A_j^\dagger(\omega')A_i(\omega)\hat{\rho}_S(t)\\ &\qquad= \sum_{i,j,\omega,\omega'}\left[\mathrm{e}^{it(\omega-\omega')}\int_0^\infty \mathrm{e}^{-is\omega}Q_{i,j}(-s)ds\right]A_j^\dagger(\omega')A_i(\omega)\hat{\rho}_S(t)\,,\end{aligned}$$

where we have introduced the *environment correlation function*

$$Q_{i,j}(s) \;=\; \mathrm{tr}\left(\hat{B}_i(s)\rho_E B_j^\dagger\right) \;=\; \mathrm{tr}\left(B_i\rho_E\hat{B}_j^\dagger(-s)\right)\,. \tag{16.16}$$

In a similar way, the second term is

$$\begin{aligned}\int_0^\infty \mathrm{tr}_E\left(\widehat{H}_I(t)\hat{\rho}_S(t)\otimes\rho_E\widehat{H}_I(t-s)\right)ds&\\ = \sum_{i,j,\omega,\omega'}\left[\mathrm{e}^{it(\omega'-\omega)}\int_0^\infty \mathrm{e}^{is\omega}\,\mathrm{tr}\left(B_i\rho_E\hat{B}_j^\dagger(-s)\right)ds\right]A_i(\omega')\hat{\rho}_S(t)A_j^\dagger(\omega)&\\ = \sum_{i,j,\omega,\omega'}\left[\mathrm{e}^{it(\omega'-\omega)}\int_0^\infty \mathrm{e}^{is\omega}Q_{i,j}(s)ds\right]A_i(\omega')\hat{\rho}_S(t)A_j^\dagger(\omega)\,.&\end{aligned}$$

Similarly, the third and fourth terms are

$$\begin{aligned}\int_0^\infty \mathrm{tr}_E\left(\widehat{H}_I(t-s)\hat{\rho}_S(t)\otimes\rho_E\widehat{H}_I(t)\right)ds&\\ = \sum_{i,j,\omega,\omega'}\sum_{\alpha',\beta'}\left[\mathrm{e}^{it(\omega-\omega')}\int_0^\infty \mathrm{e}^{-is\omega}Q_{i,j}(-s)ds\right]A_i(\omega)\hat{\rho}_S(t)A_j^\dagger(\omega')\,,&\end{aligned}$$

and

$$\begin{aligned}\int_0^\infty \mathrm{tr}_E\left(\hat{\rho}_S(t)\otimes\rho_E\widehat{H}_I(t-s)\widehat{H}_I(t)\right)ds&\\ = \sum_{i,j,\omega,\omega'}\left[\mathrm{e}^{it(\omega'-\omega)}\int_0^\infty \mathrm{e}^{is\omega}Q_{i,j}(s)ds\right]\hat{\rho}_S(t)A_j^\dagger(\omega)A_i(\omega')\,.&\end{aligned}$$

Let $\widehat{Q}_{i,j}(\omega)=\int_0^\infty \mathrm{e}^{is\omega}Q_{i,j}(s)ds$ denote the Fourier tansform of the environment correlation function and note that

$$\begin{aligned}\widehat{Q}_{i,j}(\omega)^* \;&=\; \int_0^\infty \mathrm{e}^{-is\omega}Q_{i,j}(s)^*ds\\ =\; \int_0^\infty \mathrm{e}^{-is\omega}\,\mathrm{tr}\,B_j\rho_E\hat{B}_i^\dagger(s)ds \;&=\; \int_0^\infty \mathrm{e}^{-is\omega}Q_{j,i}(-s)ds\,.\end{aligned} \tag{16.17}$$

Thus, putting together the above computations, we have

$$\begin{aligned}\frac{d}{dt}\hat{\rho}_S(t) \;&=\; -g^2\int_0^\infty \mathrm{tr}_E\left(\left[\widehat{H}_I(t),\left[\widehat{H}_I(t-s),\rho_S(t)\otimes\rho_E\right]\right]\right)ds\\ &=\; g^2\sum_{i,j,\omega,\omega'}\mathrm{e}^{it(\omega'-\omega)}\widehat{Q}_{i,j}(\omega)\left[A_i(\omega'),\hat{\rho}_S(t)A_j^\dagger(\omega)\right]\\ &+g^2\sum_{i,j,\omega,\omega'}\mathrm{e}^{it(\omega-\omega')}\widehat{Q}_{j,i}(\omega)^*\left[A_i(\omega)\hat{\rho}_S(t),A_j^\dagger(\omega')\right]\,.\end{aligned} \tag{16.18}$$

We now introduce one final approximation: we will drop all terms from the right hand side of (16.18) with $\omega\neq\omega'$. These terms all contain the rapidly oscillating factor $\mathrm{e}^{\pm it(\omega-\omega')}$. When integrated out in the differential equation for $\hat{\rho}_S(t)$ these terms will give a smaller

contribution due to phase cancellations. Dropping this term is refered to as the *rotating wave* approximation. With this approximation we have

$$\frac{d}{dt}\hat{\rho}_S(t) \;=\; g^2 \sum_{i,j,\omega} \left(\widehat{Q}_{i,j}(\omega) \left[A_i(\omega), \hat{\rho}_S(t) A_j^\dagger(\omega) \right] \;+\; \widehat{Q}_{j,i}(\omega)^* \left[A_i(\omega) \hat{\rho}_S(t), A_j^\dagger(\omega) \right] \right) \;.$$

Let us transform back to the Schrödinger picture by conjugating the equation with e^{-itH_S}. Since

$$\mathrm{e}^{-itH_S} A_i(\omega) \mathrm{e}^{itH_S} \;=\; \mathrm{e}^{-it\omega} A_i(\omega) \quad \text{and} \quad \mathrm{e}^{-itH_S} A_j(\omega)^\dagger \mathrm{e}^{itH_S} \;=\; \mathrm{e}^{it\omega} A_j^\dagger(\omega) \;,$$

we obtain

$$\begin{aligned} \frac{d}{dt}\rho_S(t) \;=\; -i \left[H_S, \rho_S(t) \right] \;+\; g^2 \sum_{i,j,\omega} \Big(& \widehat{Q}_{i,j}(\omega) \left[A_i(\omega), \rho_S(t) A_j^\dagger(\omega) \right] \\ & + \widehat{Q}_{j,i}(\omega)^* \left[A_i(\omega) \rho_S(t), A_j^\dagger(\omega) \right] \Big). \end{aligned} \tag{16.19}$$

Let us rexpress this in terms of the Hermitian and anti-Hermitian parts of $\widehat{Q}_{i,j}(\omega)$,

$$\widehat{Q}_{i,j}(\omega) \;=\; \gamma_{i,j}(\omega) + i\sigma_{i,j}(\omega)$$

with $\gamma_{i,j}(\omega) = \frac{1}{2}\left[\widehat{Q}_{i,j}(\omega) + \widehat{Q}_{j,i}^*(\omega) \right]$ and $\sigma_{i,j}(\omega) = \frac{1}{2i}\left[\widehat{Q}_{i,j}(\omega) - \widehat{Q}_{j,i}^*(\omega) \right]$. Then

$$\gamma_{i,j}(\omega) \;=\; \frac{1}{2} \int_{-\infty}^{\infty} \mathrm{e}^{is\omega} Q_{i,j}(s) ds$$

and

$$\sigma_{i,j}(\omega) \;=\; \frac{1}{2i} \int_{-\infty}^{\infty} \mathrm{e}^{is\omega} \operatorname{sgn}(s) Q_{i,j}(s) ds \;.$$

Because $Q_{i,j}(\omega)$ has an analytic extension to the upper half plane, the two functions $\gamma_{i,j}(\omega)$ and $\sigma_{i,j}(\omega)$ satisfy the *Kramer's Kronig* relation

$$\sigma_{i,j}(\omega) \;=\; \frac{1}{\pi} \mathrm{p.v.} \int_{-\infty}^{\infty} \frac{1}{\omega - \omega'} \gamma_{i,j}(\omega') d\omega' \;,$$

p.v. $\int$ denotes a principal value integral. With these definitions we have

$$\frac{d}{dt}\rho_S(t) \;=\; -i \left[H_S + g^2 H_L, \rho_S(t) \right] \;+\; g^2 \mathcal{L}(\rho_S(t)) \tag{16.20}$$

where

$$H_L \;=\; -\sum_{i,j,\omega} \sigma_{i,j}(\omega) A_j^\dagger(\omega) A_i(\omega) \;, \tag{16.21}$$

is the *Lamb shift Hamiltonian* and

$$\mathcal{L}(\rho) \;=\; \sum_{i,j,\omega} 2\gamma_{i,j}(\omega) \left[A_i(\omega) \rho A_j^\dagger(\omega) - \frac{1}{2} \left\{ A_j^\dagger(\omega) A_i(\omega) \;,\; \rho \right\} \right] \tag{16.22}$$

is the *Lindbladian*, where $\{S,T\} = ST + TS$.

We note that the Lindbladian can be written in the form

$$\mathcal{L}(\rho) \;=\; \Phi(\rho) - \frac{1}{2}\left\{\Phi^T(\mathbb{1})\,,\,\rho\right\}\,,$$

where

$$\Phi(\rho) \;=\; \sum_{i,j,\omega} 2\gamma_{i,j}(\omega) A_i(\omega)\rho A_j^\dagger(\omega)\,, \tag{16.23}$$

and Φ^T is defined by $\mathrm{tr}(\Phi^T(S)T) = \mathrm{tr}(S\Phi(T))$, so that

$$\Phi^T(\mathbb{1}) \;=\; \sum_{i,j,\omega} 2\gamma_{i,j}(\omega) A_j^\dagger(\omega) A_i(\omega)\,.$$

Note that we explicitly have

$$\mathrm{tr}\,\mathcal{L}(\rho) \;=\; \mathrm{tr}\,\Phi(\rho) - \frac{1}{2}\,\mathrm{tr}\{\Phi^T(\mathbb{1}),\rho\} \;=\; \mathrm{tr}\,\Phi(\rho) - \mathrm{tr}\,\Phi(\rho) \;=\; 0\,,$$

as is required by conservation of probability since $\mathrm{tr}\,\dot\rho(t) = 0$.

Proposition 16.3. *For each ω, the matrix $\gamma_{i,j}(\omega)$ is non-negative definite.*

Proof. We will use

Theorem (Bochner)**.** *A function $f : \mathbb{R} \to \mathbb{C}$ has a non-negative Fourier transform if and only if f is positive definite in the sense that for any finite set of points $t_1, \dots, t_n$ the matrix*

$$M_{i,j} = f(t_i - t_j)$$

is non-negative definite.

Remark. For a proof, and a more general statement involving distributions, see Simon 2015b, Thm. 6.6.6.

Let $v_i \in \mathbb{C}$, $i = 1, \dots, d$. We must show that $f(t) = \sum_{i,j} Q_{i,j}(t) v_i * v_j$ is a positive-definite function. Let $t_1, \dots, t_n \in \mathbb{R}$ and $w_i \in \mathbb{C}$, $i = 1, \dots, n$. We must show that

$$\sum_{k,\ell} f(t_k - t_\ell) w_k^* w_\ell \;=\; \sum_{k,\ell,\ i,j} Q_{i,j}(t_k - t_\ell) w_k^* v_i^* w_\ell v_j \;\geq 0\,.$$

This follows because

$$\begin{aligned}\sum_{k,\ell,\ i,j} Q_{i,j}(t_k - t_\ell) w_k^* v_i^* w_\ell v_j \;&=\; \sum_{k,\ell,\ i,j} \mathrm{tr}_E\left(\hat B_i(t_k - t_\ell)\rho_E B_j^\dagger\right) w_k^* v_i^* w_\ell v_j \\ &=\; \mathrm{tr}_E\left(\sum_{i,k} w_k^* v_i^* \hat B_i(t_k)\rho_E \sum_{\ell,j} w_\ell v_j \hat B_j(t_\ell)\right) \;=\; \mathrm{tr}_E\left(S^\dagger S \rho_E\right)\,,\end{aligned}$$

where $S = \sum_{i,k} w_k^* v_i^* \hat B_i(t_k)$.

□

By Prop. 16.3 and the Kraus representation, we see that Φ is a completely positive map. Thus the equation of motion for ρ (with the Markov, product state, long time and rotating wave approximations) is of the form

$$\frac{d}{dt}\rho_t \;=\; -i\,[H,\rho_t] + \Phi(\rho) - \frac{1}{2}\{\Phi^T(\mathbb{1})\;,\;\rho\} \tag{16.24}$$

with H a self-adjoint operator and Φ a completely positive map. As we will see below the solution to such an equation is given by $\rho_t = \mathcal{S}_t(\rho_0)$ where $\mathcal{S}_t$ is a trace preserving, completely positive semigroup. That is

1. $\operatorname{tr}\mathcal{S}_t(\rho_0) = \operatorname{tr}\rho_0$ for each t,
2. $\mathcal{S}_t$ is a completely positive map for each t, and
3. $\mathcal{S}_{t'} \circ \mathcal{S}_t \;=\; \mathcal{S}_{t+t'}$ for all t, t'.

Lecture 17

Quantum Markov Semigroup for a Qubit

Let us examine the formalism developed in the previous lecture in the context of a system consisting of a single qubit, with system Hamiltonian $H_S = -\alpha Z$ and interaction Hamiltonian $H_I = X \otimes \xi$. Here ξ is a self-adjoint operator on the environment, and we take $g = 1$, absorbing the coupling into ξ if necessary. The Pauli X operator satisfies

$$X \;=\; \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \;=\; \underbrace{\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}}_{a} \;+\; \underbrace{\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}}_{a^\dagger},$$

where $[H_S, a] = -2\alpha a$ and $\left[H_S, a^\dagger\right] = 2\alpha a^\dagger$. Hence $X(2\alpha) = a^\dagger$ and $\quad X(-2\alpha) = a$, with $X(\omega)$ as in the discussion preceeding (16.11). The Lamb Shift hamiltonian (16.21) is thus

$$H_L \;=\; -\sigma(2\alpha) a a^\dagger - \sigma(-2\alpha) a^\dagger a \;=\; \begin{pmatrix} -\sigma(2\alpha) & 0 \\ 0 & -\sigma(-2\alpha) \end{pmatrix} \;=\; -\overline{\sigma} I - \lambda Z\,,$$

where $\overline{\sigma} = \frac{\sigma(2\alpha)+\sigma(-2\alpha)}{2}$ and $\lambda = \frac{\sigma(2\alpha)-\sigma(-2\alpha)}{2}$.

The constant $\overline{\sigma}$ is an irrelevant overall shift of the energy, but λ describes a change $2\alpha \mapsto 2(\alpha + \lambda)$ of the splitting between the two system eigenvalues. The name "Lamb Shift Hamiltonian" stems from this observation. Lamb and Retherford 1947 observed a splitting of the levels of the hydrogen atom that could not be explained by the Dirac equation. The shift became known as the "Lamb shift." It was quicky realized that it was likely due to interaction with the electromagnetic field, but attempts to perturbatively calculate from existing theories led to an infinite answer. Bethe 1947 realized that the correct result should be obtained by ignoring the infinity (which was already incorporated into the electron mass) and looking at the next leading term. This eventually led to the idea of *renormalization* which allowed for the development of quantum electrodynamics and other quantum field theories.

The Lindbladian (16.22) for our qubit is

$$\mathcal{L}(\rho) \;=\; \Phi(\rho) - \frac{1}{2}\left\{\Phi^T(\mathbb{1})\,,\,\rho\right\}\,, \quad \text{with} \quad \Phi(\rho) \;=\; 2\gamma(2\alpha)\, a^\dagger \rho a + 2\gamma(-2\alpha)\, a \rho a^\dagger\,.$$

We have

$$\Phi^T(\mathbb{1}) \;=\; 2\gamma(2\alpha) a a^\dagger + 2\gamma(-2\alpha) a^\dagger a \;=\; \Gamma\mathbb{1} + \Delta Z\,,$$

with $\Gamma = \gamma(2\alpha) + \gamma(-2\alpha)$ and $\Delta = \gamma(2\alpha) - \gamma(-2\alpha)$. We can rewrite Φ as follows

$$\Phi(\rho) \;=\; (\Gamma+\Delta)a^\dagger\rho a \;+\; (\Gamma-\Delta)a\rho a^\dagger \;=\; \frac{\Gamma}{2}(X\rho X + Y\rho Y) + i\frac{\Delta}{2}(X\rho Y - Y\rho X)\,,$$

where in the last step we have used that $a = \frac{1}{2}(X+iY)$ and $a^\dagger = \frac{1}{2}(X-iY)$.

Let us now look at the evolution in the Bloch sphere representation. We have

$$\mathcal{L}(\mathbb{1}) \;=\; -2\Delta Z\,, \quad \mathcal{L}(X) \;=\; -\Gamma X\,, \quad \mathcal{L}(Y) \;=\; -\Gamma Y\,, \quad \mathcal{L}(Z) \;=\; -2\Gamma Z\,.$$

Thus, if $\rho(t) = \frac{1}{2}\left[\mathbb{1} + \boldsymbol{v}(t)\cdot\boldsymbol{\sigma}\right]$, then

$$\begin{aligned}\dot{\boldsymbol{v}}\cdot\boldsymbol{\sigma} \;&=\; -iv_1(\alpha+\lambda)\,[Z,X] - iv_2(\alpha+\lambda)\,[Z,Y] - 2\Delta Z - \Gamma v_1 X - \Gamma v_2 Y - 2\Gamma v_3 Z \\ &=\; -(\Gamma v_1 + 2(\alpha+\lambda)v_2)X - (\Gamma v_2 - (\alpha+\lambda)v_1)Y - (2\Delta + 2\Gamma v_3)Z\,.\end{aligned}$$

Hence

$$\dot{\boldsymbol{v}} \;=\; -2\tilde{\alpha}\boldsymbol{e}_3\times\boldsymbol{v} - \Gamma\begin{pmatrix}1&0&0\\0&1&0\\0&0&2\end{pmatrix}\boldsymbol{v} - 2\Delta\boldsymbol{e}_3\,, \tag{17.1}$$

with $\tilde{\alpha} = \alpha + \lambda$. The solution to (17.1) is given by

$$\boldsymbol{v}(t) \;=\; \begin{pmatrix}\mathrm{e}^{-\Gamma t}\cos(2\tilde{\alpha}t) & -\mathrm{e}^{-\Gamma t}\sin(2\tilde{\alpha}t) & 0\\ \mathrm{e}^{-\Gamma t}\sin(2\tilde{\alpha}t) & \mathrm{e}^{-\Gamma t}\cos(2\tilde{\alpha}t) & 0\\ 0 & 0 & \mathrm{e}^{-2\Gamma t}\end{pmatrix}\begin{pmatrix}v_1(0)\\v_2(0)\\v_3(0)\end{pmatrix} - \begin{pmatrix}0\\0\\(1-\mathrm{e}^{-2\Gamma t})\frac{\Delta}{\Gamma}\end{pmatrix}. \tag{17.2}$$

Under the evolution (17.1), $\boldsymbol{v}(t) \to \begin{pmatrix}0 & 0 & \frac{\Delta}{\Gamma}\end{pmatrix}^T$ as $t\to\infty$, which is to say

$$\rho(t) \;\to\; \frac{1}{2}\left[\mathbb{1} + \frac{\Delta}{\Gamma}Z\right] \;=\; \frac{1}{2}\begin{pmatrix}1+\frac{\Delta}{\Gamma} & 0\\ 0 & 1-\frac{\Delta}{\Gamma}\,.\end{pmatrix}$$

The limiting state is a mixture of $|0\rangle\langle 0|$ and $|1\rangle\langle 1|$ with differing probabilities for each state depending on the ratio $\frac{\Delta}{\Gamma}$.

The dynamics in (17.1) can be expressed as $\rho(t) = \Psi_t(\mathrm{e}^{i\tilde{\alpha}tZ}\rho(0)\mathrm{e}^{-it\tilde{\alpha}Z})$ where Ψ_t is the generalized amplitude damping channel

$$\begin{aligned}\Psi_t(\rho) \;=\; &\tfrac{1}{2}\left(1+\tfrac{\Delta}{\Gamma}\right)\left[\begin{pmatrix}1&0\\0&\mathrm{e}^{-\Gamma t}\end{pmatrix}\rho\begin{pmatrix}1&0\\0&\mathrm{e}^{-\Gamma t}\end{pmatrix} + \begin{pmatrix}0&\sqrt{1-\mathrm{e}^{-2\Gamma t}}\\0&0\end{pmatrix}\rho\begin{pmatrix}0&0\\\sqrt{1-\mathrm{e}^{-2\Gamma t}}&0\end{pmatrix}\right] \\ &+\; \tfrac{1}{2}\left(1-\tfrac{\Delta}{\Gamma}\right)\left[\begin{pmatrix}\mathrm{e}^{-\Gamma t}&0\\0&1\end{pmatrix}\rho\begin{pmatrix}\mathrm{e}^{-\Gamma t}&0\\0&1\end{pmatrix} + \begin{pmatrix}0&0\\\sqrt{1-\mathrm{e}^{-2\Gamma t}}&0\end{pmatrix}\rho\begin{pmatrix}0&\sqrt{1-\mathrm{e}^{-2\Gamma t}}\\0&0\end{pmatrix}\right].\end{aligned}$$

Thus the dynamics is a mixture of the unitary evolution due to the Hamiltonian $\widetilde{H} = -\tilde{\alpha}Z$ and the semigroup Ψ_t of completely positive amplitude damping channels.

The coefficients Δ and Γ are related to the correlation function $\gamma(\omega) = \mathrm{Re}\widehat{Q}(\omega)$ of the environment, where $\widehat{Q}(\omega) = \int_0^\infty \mathrm{e}^{i\omega t}\,\mathrm{tr}(\hat{B}(t)\rho_E B)dt$. Note that $\widehat{Q}(\omega)$ is an analytic function of ω in the upper half plane. Usually the integral defining $\widehat{Q}$ is not absolutely convergent for real ω, but must be understood as the boundary value

$$\widehat{Q}(\omega+i0) \;=\; \lim_{\epsilon\downarrow 0}\widehat{Q}(\omega+i\epsilon) \;=\; \lim_{\epsilon\downarrow 0}\int_0^\infty \mathrm{e}^{i(\omega+i\epsilon)t}\,\mathrm{tr}(\hat{B}(t)\rho_E B)dt\,.$$

Let us consider two typical situations.

Environment in a ground state. Suppose $\rho_E = |\psi_0\rangle\langle\psi_0|$ with ψ_0 a ground state of the environment Hamiltonian H_E. Then

$$\mathrm{tr}_E\, \hat{\xi}(t)\rho_E\xi = \langle\psi_0|\, \xi \mathrm{e}^{it(H-E_0)}\xi\, |\psi_0\rangle$$

where E_0 is the ground state energy. As a result

$$\widehat{Q}(\omega) = \frac{1}{i}\lim_{\varepsilon\to 0}\langle\psi_0|\, \xi(E_0 - H - \omega - i\varepsilon)^{-1}\xi\, |\psi_0\rangle\ .$$

Thus

$$\gamma(\omega) = \mathrm{Im}\,\langle\psi_0|\, \xi(E_0 - H - \omega - i0)^{-1}\xi\, |\psi_0\rangle = \pi\,\langle\psi_0|\, \xi\delta(H - E_0 + \omega)\xi\, |\psi_0\rangle\ , \quad (17.3)$$

where we have used the identity

$$\lim_{\varepsilon\to 0}\mathrm{Im}\frac{1}{x - i\varepsilon} = \lim_{\varepsilon\to 0}\frac{\varepsilon}{x^2+\varepsilon^2} = \pi\delta(x)\ .$$

If the environment is idealized as an infinite system with absolutely continuous spectrum, (17.3) will give a finite result for γ. Because $H - E_0 \geq 0$, since E_0 is the ground state energy, $\gamma(\omega)$ vanishes for ω positive. Thus we have $\Gamma = \Delta = \gamma(-2\alpha)$ and the semigroup Ψ_t is a family of amplitude damping channels

$$\Phi_t(\rho) = \left(\begin{smallmatrix}1 & 0\\ 0 & \mathrm{e}^{-\Gamma t}\end{smallmatrix}\right)\rho\left(\begin{smallmatrix}1 & 0\\ 0 & \mathrm{e}^{-\Gamma t}\end{smallmatrix}\right) + \left(\begin{smallmatrix}0 & \sqrt{1-\mathrm{e}^{-2\Gamma t}}\\ 0 & 0\end{smallmatrix}\right)\rho\left(\begin{smallmatrix}0 & 0\\ \sqrt{1-\mathrm{e}^{-2\Gamma t}} & 0\end{smallmatrix}\right)\ .$$

If the system is put into any state other than the ground state $|0\rangle\langle 0|$, it emits energy to the environment and decays to the ground state as $t \to \infty$.

Environment in a thermal state Another typical situation involves an environment at thermal equilibrium in the *Gibbs state*

$$\rho_E = \frac{1}{Z(\beta)}\mathrm{e}^{-\beta H_E}\ ,$$

with $\beta = 1/k_B T$ the inverse temperature and $Z(\beta) = \mathrm{tr}\,\mathrm{e}^{-\beta H_E}$. This state satisfies the *Kubo-Martin-Schwinger*(KMS) condition:

$$\mathrm{tr}\,\alpha_t(A)B\rho_E = \mathrm{tr}\,B\alpha_{t+i\beta}(A)\rho_E \quad (17.4)$$

where $\alpha_t(A) = \mathrm{e}^{itH_E}A\mathrm{e}^{-itH_E}$ is the Heisenberg dynamics on the environment.

In fact, to properly define the thermal state in an infinite system (necessary to get the idealization of a correlation function $Q(t)$ that vanishes as $t \to \infty$), one needs to take the *thermodynamic*, or infinite volume, limit. In the thermodynamic limit one has a C^*-algebra of observables $\mathfrak{A}$ on which the dynamics α_t is defined. However, the dynamics is not, in general, given by "inner automorphisms" of the algebra, i.e., conjugation by unitary

maps. Likewise the state is not given by trace against a trace class operator (there is no natural trace on the infinite volume algebra). Instead the state is a positive linear functional $A \mapsto \langle A \rangle$ which satisfies the KMS condition

$$\langle \alpha_t(A)B \rangle \;=\; \langle B\alpha_{t+i\beta}(A) \rangle \,. \tag{17.5}$$

There is a method, known as the Gelfand Naimark Segal (GNS) construction, to create a representation of $\mathfrak{A}$ on a new Hilbert space such that

$$\langle A \rangle \;=\; \langle \psi_0 | \, A \, | \psi_0 \rangle$$

for a given vector $|\psi_0\rangle$. On this space there is a self-adjoint operator, the *GNS-Liouvillian* L_{GNS}, that implements the dynamics in the sense that

$$\langle \alpha_t(A)B \rangle \;=\; \langle \psi_0 | \, A \mathrm{e}^{-itL_{GNS}} B \, | \psi_0 \rangle$$

and

$$\langle A\alpha_t(B) \rangle \;=\; \langle \psi_0 | \, A \mathrm{e}^{itL_{GNS}} B \, | \psi_0 \rangle \,.$$

In general L_{GNS} is unbounded from above and below — it should be thought of as the thermodynamic limit of the Liouvillian $A \mapsto [H_E, A]$. By the KMS condition, we have

$$\langle \psi_0 | \, A \mathrm{e}^{-itL_{GNS}} B \, | \psi_0 \rangle \;=\; \langle \psi_0 | \, B \mathrm{e}^{i(t+i\beta)L_{GNS}} A \, | \psi_0 \rangle \,.$$

Thus the correlation function Q satisfies

$$Q(t) \;=\; \langle \psi_0 | \, \xi \mathrm{e}^{-itL_{GNS}} \xi \, | \psi_0 \rangle \;=\; \langle \psi_0 | \, \xi \mathrm{e}^{i(t+i\beta)L_{GNS}} \xi \, | \psi_0 \rangle \;=\; Q(-t - i\beta) \,.$$

It follows that

$$\gamma(\omega) \;=\; \pi \, \langle \psi_0 | \, \xi \delta(L_{GNS} - \omega) \xi \, | \psi_0 \rangle \;=\; \pi \, \langle \psi_0 | \, \xi \delta(L_{GNS} + \omega) \mathrm{e}^{-\beta L_{GNS}} \xi \, | \psi_0 \rangle \;=\; \mathrm{e}^{\beta\omega} \gamma(-\omega) \,.$$

In other words, γ satisfies the thermal equilbrium condition

$$\frac{\gamma(\omega)}{\gamma(-\omega)} \;=\; \mathrm{e}^{\beta\omega} \,.$$

As a result we have

$$\frac{\Delta}{\Gamma} \;=\; \frac{1 - \mathrm{e}^{-2\beta\alpha}}{1 + \mathrm{e}^{-2\beta\alpha}}$$

so that

$$\frac{1}{2}\left(1 + \frac{\Delta}{\Gamma}\right) \;=\; \frac{1}{1 + \mathrm{e}^{2\beta\alpha}} \quad \text{and} \quad \frac{1}{2}\left(1 - \frac{\Delta}{\Gamma}\right) \;=\; \frac{\mathrm{e}^{-2\beta\alpha}}{1 + \mathrm{e}^{-2\beta\alpha}} \,.$$

Thus the density matrix converges as $t \to \infty$ to the thermal state

$$\rho(t \to \infty) \;=\; \frac{1}{\mathrm{tr}\, \mathrm{e}^{-\beta H_S}} \mathrm{e}^{-\beta H_S} \,.$$

That is, our Lindbladian evolution describes the process of thermal equilibration on a system in contact with a heat bath.

Lecture 18

Lindblad's Theorem

Consider a finite dimensional quantum system with Hilbert space $\mathcal{H}_S \cong \mathbb{C}^n$.

Definition 18.1. A *quantum dynamical semigroup* is a family of linear maps $\Phi_t : \mathcal{B}(\mathcal{H}_S) \to \mathcal{B}(\mathcal{H}_S)$ for $t \geq 0$ such that

1. Φ_0 is the identity, map $\Phi_0(A) = A$ for all $A \in \mathcal{B}(\mathcal{H}_S)$,
2. $\Phi_t \circ \Phi_s = \Phi_{t+s}$ for all $t, s \geq 0$,
3. Φ_t is completely positive for every $t \geq 0$,
4. $t \mapsto \Phi_t(A)$ is continuous for each t and $A \in \mathcal{B}(\mathcal{H})$

A quantum dynamical semi-group Φ_t is called *trace preserving* if $\operatorname{tr} \Phi_t(A) = \operatorname{tr} A$ for all A and is *unital* if $\Phi_t(\mathbb{1}) = \mathbb{1}$.

Remarks.

1) It is sufficient to assume that the semigroup is continuous at $t = 0$. Continuity for $t > 0$ follows from this and property (2).

2) If we take $\Phi_t(A) = PAP$ for $t > 0$ and $\Phi_0(A) = A$ with P a proper projection, then Φ_t satisfies assumptions (1-3) but is not continuous. So the continuity assumption matters here even though we are in finite dimensions.

3) For infinite dimensional algebras, one can distinguish between norm continuous and weakly continuous semi-groups. In our finite dimensional context that is not necessary.

Any continuous semigroup Φ_t on a finite dimensional vector space, such as $\mathcal{B}(\mathcal{H}_S)$, is in fact differentiable and has a *generator*

$$L(A) \;=\; \lim_{t \downarrow 0} \frac{1}{t} \left(\Phi_t(A) - A \right) \;.$$

The semigroup, in turn, is given by

$$\Phi_t(A) \;=\; \mathrm{e}^{tL}(A) \;,$$

where the exponential can be defined via the power series

$$e^{tL}(A) \;=\; \sum_{n=0}^{\infty} \frac{1}{n!} t^n L^n(A) \;.$$

Lindblad 1976 derived the following result which characterizes the generators of quantum dynamical semigroups:

Theorem 18.2 (Lindblad 1976)**.** *If Φ_t is a unital quantum dynamical semigroup on $\mathcal{B}(\mathcal{H})$, then there is a completely positive map Ψ and a self-adjoint operator H such that the generator $\mathcal{L}$ of Φ_t is*

$$\mathcal{L}(A) \;=\; \Psi(A) - \frac{1}{2}\left\{\Psi(\mathbb{1})\;,\; A\right\} + i\left[H, A\right] \;. \tag{18.1}$$

Lindblad's Theorem 18.2 describes the Heisenberg picture of a quantum dynamical semigroup, acting on the observables of the system. In the Schrödinger picture, we have a trace preserving semigroup Φ_t acting on density matrices as above. In this case, the transposed semigroup Φ_t^T defined on observables via

$$\operatorname{tr} \Phi_t^T(A)\rho \;=\; \operatorname{tr} A\Phi_t(A)$$

is a unital semigroup. As a result we obtain the following

Corollary 18.3. *If Φ_t is a trace preserving quantum dynamical semigroup on $\mathcal{B}(\mathcal{H})$, then there is a completely positive map Ψ and a self-adjoint operator H such that the generator $\mathcal{L}$ of Φ_t is*

$$\mathcal{L}(\rho) \;=\; \Psi(\rho) - \frac{1}{2}\left\{\Psi^T(\mathbb{1})\;,\; \rho\right\} - i\left[H, \rho\right] \;. \tag{18.2}$$

Lindblad's 1976 paper is quite readable. In it he actually proved the more general result that the representation (18.2) holds for the generator of a norm-continuous completely positive semigroup on $\mathcal{B}(\mathcal{H})$, with $\mathcal{H}$ any separable Hilbert space.

We will now sketch the main details of Lindblad's argument. The first step is the following inequality for completely positive maps

Lemma 18.4 (Lindblad 1976, Lemma 5)**.** *Let $\Phi : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$ be a completely positive map. Then for all $X \in \mathcal{B}(\mathcal{H})$,*

$$\Phi(X^\dagger)\Phi(\mathbb{1})^{-1}\Phi(X) \;\leq\; \Phi(X^\dagger X) \;. \tag{18.3}$$

If $\Phi(\mathbb{1}) = \mathbb{1}$, then

$$\Phi(X^\dagger)\Phi(X) \;\leq\; \Phi(X^\dagger X) \;. \tag{18.4}$$

Proof. First note that (18.3) follows from (18.4) applied to the map

$$\Psi(X) \;=\; \Phi(\mathbb{1})^{-\frac{1}{2}}\Phi(X)\Phi(\mathbb{1})^{-\frac{1}{2}} \;.$$

Eq. (18.3) can be found in Størmer 1974, Theorem 3.1. Here is a simple proof using the Kraus representation.

Since Φ is CP with $\Phi(\mathbb{1}) = \mathbb{1}$ we have

$$\Phi(X) \;=\; \sum_{j=1}^{n} K_j^\dagger X K_j \quad \text{with} \quad \sum_j K_j^\dagger K_j = \mathbb{1}\,.$$

Let $V : \mathcal{H} \to \mathcal{H}^n$ be the map $V\boldsymbol{u} = (K_1\boldsymbol{u}, \dots, K_n\boldsymbol{u})$. Note that V is an isometry, since

$$\|V\boldsymbol{u}\|^2 \;=\; \sum_{j=1}^{n} \left\|K_j\boldsymbol{u}\right\|^2 \;=\; \langle \boldsymbol{u}| \sum_{j=1}^{n} K_j^\dagger K_j\, |\boldsymbol{u}\rangle \;=\; \|\boldsymbol{u}\|^2\,.$$

It follows that $VV^\dagger$ is the projection of $\mathcal{H}^n$ onto the range of V and that $\left\|V^\dagger\boldsymbol{\xi}\right\| \le \|\boldsymbol{\xi}\|$ for all $\boldsymbol{\xi} \in \mathcal{H}^n$, where $V^\dagger(\boldsymbol{u}_1, \dots, \boldsymbol{u}_n) = \sum_{j=1}^n K_j^\dagger \boldsymbol{u}_j$. Given $X \in \mathcal{B}(\mathcal{H})$ let $\pi(X)$ denote the coordinatewise action of X on $\mathcal{H}_n$: $\pi(X)(\boldsymbol{u}_1, \dots, \boldsymbol{u}_n) = (X\boldsymbol{u}_1, \dots, X\boldsymbol{u}_n)$. Then

$$\begin{aligned}\langle \boldsymbol{u}|\, \Phi(X^\dagger X)\, |\boldsymbol{u}\rangle \;&=\; \sum_{j=1}^{N} \langle \boldsymbol{u}|\, K_j^\dagger X^\dagger X K_j\, |\boldsymbol{u}\rangle \;=\; \sum_{j=1}^{n} \left\|XK_j\boldsymbol{u}\right\|^2 \\ &=\; \|\pi(X)V\boldsymbol{u}\|^2 \;\ge\; \left\|V^\dagger\pi(X)V\boldsymbol{u}\right\|^2 \;=\; \langle \boldsymbol{u}|\, \Phi(X^\dagger)\Phi(X)\, |\boldsymbol{u}\rangle\,.\end{aligned}$$

□

Let $\Phi_t = e^{tL}$ be a norm continuous semigroup of completely positive maps with $\Phi_t(\mathbb{1}) = \mathbb{1}$. Consider the semigroup $\Phi_t^{(n)} = \Phi_t \otimes I_n$ on

$$\mathcal{B}_n(\mathcal{H}) \;=\; \mathcal{B}(\mathcal{H}) \otimes \mathcal{B}(\mathbb{C}^n) \;=\; \{n \times n \text{ matrices with entries in } \mathcal{B}(\mathcal{H})\}\,. \tag{18.5}$$

It has generator $L_n = L \otimes I_n$. Differentiating (18.4) for $\Phi_t^{(n)}$ at $t = 0$ we obtain

$$L_n(X^\dagger X) - L_n(X^\dagger)X - X^\dagger L_n(X) \;\ge\; 0\,. \tag{18.6}$$

Given a linear map $L : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$, we introduce the *dissipation function* $D(L)$

$$D(L; X, Y) \;=\; L(X^\dagger Y) - L(X^\dagger)Y - X^\dagger L(Y)$$

for $X, Y \in \mathcal{B}(\mathcal{H})$.

Definition 18.5. A linear map $L : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$ is a *star map* if $L(X^\dagger) = L(X)^\dagger$ for all $X \in \mathcal{B}(\mathcal{H})$ and is *completely dissipative* if it is a star map with $L(I) = 0$ and $D(L_n; X, X) \ge 0$ for all $X \in \mathcal{B}_n(\mathcal{H})$.

Proposition 18.6 (Lindblad 1976, Proposition 4)**.** *Let $L : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$ be a star map and let $\Phi_t = \mathrm{e}^{tL}$. The following are equivalent*

1. *$\Phi_t(X^\dagger)\Phi_t(X) \le \Phi_t(X^\dagger X)$ for all X and $\Phi_t(\mathbb{1}) = \mathbb{1}$.*
2. *$D(L; X, X) \ge 0$ for all X and $L(\mathbb{1}) = 0$.*

In the proof of Prop. 18.6 we will make use of three lemmas. The first is a simple observation that elements of $\mathcal{B}(\mathcal{H})$ are linear combinations of unitaries. Let $\mathcal{U}(\mathcal{H})$ denote the set of unitary elements of $\mathcal{B}(\mathcal{H})$.

Lemma 18.7. *Every $X \in \mathcal{B}(\mathcal{H})$ can be expressed as $\frac{\|X\|}{2}(U_1 + U_2)$ with $U_1, U_2 \in \mathcal{U}(\mathcal{H})$.*

Remark. Russo and Dye 1966 proved a vast generalization, namely that the convex hull of the unitary elements is dense in the unit ball of any von Neumann algebra.

Proof. If $X = 0$ the result is simple. For non-zero X it suffices to consider $\|X\| = 1$, by scaling. Let $X = V|X|$ be the polar decomposition of X, where we take V to be unitary by defining it suitably on $\ker X$. Let λ_j, ψ_j, $j = 1, \dots, n$, be the eigenvalues and eigenvectors of $|X|$. We have $0 \le \lambda_j \le 1$ since $\|X\| = \||X|\| = 1$. Then

$$|X| \;=\; \tfrac{1}{2}(U_+ + U_-) \quad \text{with} \quad U_\pm = \sum_{j=1}^{n} \mathrm{e}^{\pm i \arccos(\lambda_j)} \left|\psi_j\right\rangle \left\langle\psi_j\right| \,.$$

Thus $X = \frac{1}{2}(VU_+ + VU_-)$.

□

The second lemma is a standard fact from the theory of C^* algebras:

Lemma 18.8. *Let $\sigma : \mathcal{B}(\mathcal{H}) \to \mathbb{C}$ be a linear map with $\sigma(\mathbb{1}) \le 1$ and $\|\sigma\| = 1$. Then σ is a positive, i.e., $\sigma(A) \ge 0$ if $A \ge 0$.*

Remarks. 1) Since $\|\sigma\| \ge |\sigma(1)|$ it follows that $\|\sigma\| = 1$. 2) For the corresponding result and proof in the context of C^* algebras, see Kadison and Ringrose 1997, Theorem 4.3.2.

Proof. We have $\sigma(A) = \operatorname{tr} SA$ for some $S \in \mathcal{B}(\mathcal{H})$. Clearly $\operatorname{tr} S = \sigma(\mathbb{1}) = 1$. Let $S = V|S|$ be the polar decomposition of S. Then

$$\phi(V^\dagger) = \operatorname{tr} V^\dagger V|S| = \operatorname{tr}|S| \,.$$

Because $1 \le |\operatorname{tr} S| \le \operatorname{tr}|S| \le \|\phi\| = 1$ we have $\operatorname{tr}|S| = 1$. Let λ_j, ψ_j, $j = 1, \dots, n$, be the eigenvectors and eigenvalues of $|S|$. Then

$$1 \;=\; \operatorname{tr} S \;=\; \sum_{j=1}^{n} \left\langle\psi_j\right| V|S| \left|\psi_j\right\rangle \;=\; \sum_{j=1}^{n} \lambda_j \left\langle\psi_j\right| V \left|\psi_j\right\rangle \,.$$

Since $\sum_j \lambda_j = \operatorname{tr}|S| = 1$ and $\left|\left\langle\psi_j\right| V \left|\psi_j\right\rangle\right| \le 1$ this can only hold if $\left\langle\psi_j\right| V \left|\psi_j\right\rangle = 1$ for each j. Because V is an isometry, this implies $\left|\psi_j\right\rangle = V \left|\psi_j\right\rangle$, so $V = \mathbb{1}$ and σ is positive.

□

The third lemma adapts Russo and Dye 1966, Corollary 1 to the simpler context of the finite dimensional algebra $\mathcal{B}(\mathcal{H})$:

Lemma 18.9 (Russo and Dye)**.** *The norm of a linear mapping $\phi : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$ is equal to $\sup_{U \in \mathcal{U}(\mathcal{H})} \|\phi(U)\|$. Moreover, if $\phi(\mathbb{1}) = \mathbb{1}$, then ϕ is positive if and only if $\|\phi\| = 1$.*

Proof. Let $\|\phi\|_U = \sup_{U\in\mathcal{U}(\mathcal{H})} \|\phi(U)\|$. Clearly $\|\phi\|_U \le \|\phi\|$. Conversely, if $X \in \mathcal{B}(\mathcal{H})$, then by Lem. 18.7

$$\|\phi(X)\| \;\le\; \tfrac{\|X\|}{2}\left(\|\phi(U_1)\| + \|\phi(U_2)\|\right) \;\le\; \|\phi\|_U \|X\|\,,$$

from which the reverse inequality follows.

Before proving the second part of the lemma, let us note that ϕ is positive if and only if ϕ^T is positive. Indeed if ϕ is positive and $A \ge 0$ then

$$\langle \boldsymbol{u}|\, \phi^T(A)\, |\boldsymbol{u}\rangle \;=\; \operatorname{tr} \phi^T(A)\, |\boldsymbol{u}\rangle\langle \boldsymbol{u}| \;=\; \operatorname{tr} A\phi(|\boldsymbol{u}\rangle\langle \boldsymbol{u}|) \;=\; \operatorname{tr} A^{\frac{1}{2}} |\boldsymbol{u}\rangle\langle \boldsymbol{u}| A^{\frac{1}{2}} \;\ge\; 0\,.$$

The reverse implication is the same with the roles of ϕ and ϕ^T interchanged.

Now suppose $\phi(\mathbb{1}) = \mathbb{1}$ and that ϕ is positive. To prove that $\|\phi\| = 1$ it suffices to show that $|\operatorname{tr}\rho\phi(A)| \le \|A\|$ for any density matrix ρ and any $A \in \mathcal{B}(\mathcal{H})$. Since $\operatorname{tr}\phi^T(\rho) = \operatorname{tr}\rho\phi(\mathbb{1}) = 1$ we see that $\phi^T(\rho)$ is a density matrix. Thus

$$|\operatorname{tr}\rho\phi(A)| \;=\; \left|\operatorname{tr}\phi^T(\rho)A\right| \;\le\; \operatorname{tr}\phi^T(\rho)\|A\| \;=\; \|A\|\,.$$

Conversely, suppose that $\phi(\mathbb{1}) = \mathbb{1}$ and that $\|\phi\| = 1$. Let $A \ge 0$ be a positive element of $\mathcal{B}(\mathcal{H})$ and let $\boldsymbol{u}$ be a unit vector. Then $\sigma(A) = \langle \boldsymbol{u}|\, \phi(A)\, |\boldsymbol{u}\rangle$ is a linear functional with $\sigma(\mathbb{1}) = 1$ and $\|\sigma\| \le 1$. Thus σ is positive. We have shown that $\langle \boldsymbol{u}|\, \phi(A)\, |\boldsymbol{u}\rangle \ge 0$ for any $\boldsymbol{u}$. That is, $\Phi(A) \ge 0$.

□

The third lemma combines a standard fact about linear functionals on C^* algebras with the second part of Russo and Dye 1966, Corollary 1:

Sketch of Proof of Prop. 18.6. We proved (1) $\implies$ (2) above. For the proof that (2) $\implies$ (1), note that the semigroup can be obtained as the limit

$$\Phi_t \;=\; \lim_{n\to\infty} (I + \tfrac{1}{n}L)^n\,. \tag{18.7}$$

We want to show that Φ_t is a contraction on $\mathcal{B}(\mathcal{H})$. By the Russo-Dye Lemma 18.7,

$$\begin{aligned}\|I + tL\|^2 \;&=\; \sup_{U\in\mathcal{B}(\mathcal{U})} \|U + tL(U)\|^2 \;=\; \sup_{U\in\mathcal{B}(\mathcal{U})} \left\|I + tU^\dagger L(U) + tL(U^\dagger) + t^2 L(U)^\dagger L(U)\right\| \\ &\le\; \sup_{U\in\mathcal{B}(\mathcal{U})} \left\|I + tL(U^\dagger U) + t^2 L(U)^\dagger L(U)\right\| \;\le\; 1 + t^2 \sup_{U\in\mathcal{B}(\mathcal{U})} \|L(U)\|^2 \;=\; 1 + t^2\|L\|^2\,,\end{aligned}$$

where we have noted that $L(U^\dagger U) = L(\mathbb{1}) = 0$. Setting $t = \frac{1}{n}$ and taking the square root we find that $\left\|I + \frac{1}{n}L\right\| \le 1 + \frac{1}{2n^2}\|L\|^2$. It follows from (18.7) that $\|\Phi_t\| \le 1$. Since $\Phi_t(\mathbb{1}) = \mathbb{1}$, we have $\|\Phi_t\| = 1$ and Φ_t is a positive map by the Russo-Dye Lemma 18.9.

Now, let

$$S_\lambda = \lambda \int_0^\infty \mathrm{e}^{-\lambda t}\Phi_t dt = \lambda(\lambda - K)^{-1}, \quad \text{for } \lambda > 0\,.$$

Since Φ_t is positive, so is S_λ. Furthermore, we have $\|S_\lambda\| \le 1$ and $S_\lambda(\mathbb{1}) = \mathbb{1}$ so $\|S_\lambda\| = 1$. Let $X \in \mathcal{B}(\mathcal{H})$ and let $Y = (\lambda - L)^{-1}(X)$. Then $X = \lambda Y - L(Y)$ so that

$$\begin{aligned}S_\lambda(X^\dagger X) \;&=\; S_\lambda[\lambda^2 Y^\dagger Y - \lambda(L(Y^\dagger)Y + Y^\dagger L(Y)) + L(Y^\dagger)L(Y)] \\ &\ge\; S_\lambda[\lambda^2 Y^\dagger Y - \lambda L(Y^\dagger Y)] \;=\; \lambda S_\lambda[(\lambda - L)(Y^\dagger Y)] \;=\; \lambda^2 Y^\dagger Y = S_\lambda(X^\dagger)S_\lambda(X)\,.\end{aligned}$$

Because S_λ is positive, it follows by induction that $S_\lambda^n(X^\dagger X) \geq S_\lambda^n(X^\dagger) S_\lambda^n(X)$. The identity now follows for Φ_t because we have $\Phi_t = \lim_n S_{\frac{t}{n}}^n$.

□

We may now prove

Theorem 18.10. *Let $L : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$ be a $*$-map and let $\Phi_t = \mathrm{e}^{tL}$. Then Φ_t is unital quantum dynamical semigroup if and only if L is completely dissipative.*

Proof. Apply Prop. 18.6 to L_n and $\Phi_t^{(n)}$ for each n.

□

To complete the proof of Lindblad's Theorem 18.2, we show

Proposition 18.11. *A map L on $\mathcal{B}(\mathcal{H})$ is completely dissipative map if and only if there is a completely positive map Ψ and a self-adjoint $H \in \mathcal{B}(\mathcal{H})$ such that for all $x \in \mathcal{B}(\mathcal{H})$.*

$$L(X) \;=\; \Psi(X) = \frac{1}{2}\{\Psi(X)\,,\, X\} + i\,[H, X]\;.$$

Proof. We start with the forward implication. Let $\mathcal{U}(\mathcal{H})$ denote the set of unitary elements of $\mathcal{B}(\mathcal{H})$ and let h denote Haar measure on this group. Define

$$K \;=\; \int_{\mathcal{U}(\mathcal{H})} L(U^\dagger) U dh(U)\;.$$

Then $\|K\| \leq \|L\|$. By the invariance of Haar measure we have

$$KV \;=\; \int_{\mathcal{U}(\mathcal{H})} L(U^\dagger) UV dh(U) \;=\; \int_{\mathcal{U}(\mathcal{H})} L(VU^\dagger) U dh(U)\;.$$

By Lem. 18.7 we have

$$KX \;=\; \int_{\mathcal{U}(\mathcal{H})} L(XU^\dagger) U dh(U)$$

for every X.

For $X, Y \in \mathcal{B}(\mathcal{H})$ we have

$$\int_{\mathcal{U}(\mathcal{H})} D(L; UX, UY) dh(U) \;=\; L(X^\dagger Y) - KX^\dagger Y - X^\dagger Y K^\dagger\;.$$

Define $\Psi : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$ by

$$\Psi(X) \;=\; L(X) - KX - XK^\dagger\;.$$

Because L is completely disipative, we have

$$\Psi(X^\dagger X) \;=\; \int_{\mathcal{U}(\mathcal{H})} D(L; UX, UX) dh(U) \;\geq\; 0\,,$$

so Ψ is positive. Applying the same argument to $\mathcal{B}_n(\mathcal{H})$, we find that

$$K \otimes I_n \;=\; \int_{\mathcal{U}_n(\mathcal{H})} L_n(U^\dagger) U dh_n(U)\;,$$

where h_n now denotes Haar measure on $\mathcal{U}_n(\mathcal{H})$. Similarly, we find that $\Psi\otimes_n$ is positive. So Ψ is completely positive.

Thus we have $L(X) = \Psi(X) + KX + XK^\dagger$ with Ψ a completely positive map. To complete the proof, note that

$$\Psi(\mathbb{1}) \;=\; L(\mathbb{1}) - K - K^\dagger \;=\; -(K+K)^\dagger\;.$$

If we take $H = \frac{1}{2i}(K - K^\dagger)$ it follows that

$$KX + XK^\dagger \;=\; -\frac{1}{2}\{\Psi(\mathbb{1})\;,\; X\} + i\,[H, X]\;.$$

Conversely, suppose that Ψ is completely positive and H is self adjoint and let $L(X) = \Psi(x) - \frac{1}{2}\{\Psi(\mathbb{1}), X\} + i\,[H, X]$. Clearly $L(\mathbb{1}) = 0$ and L is a star map. As in the proof of Lemma 18.4, we can write $\Psi(X) = W^\dagger \pi(A) W$ with π a representation of $\mathcal{B}(\mathcal{H})$ on another Hilbert space $\mathcal{K}$ and W a bounded linear map from $\mathcal{H}$ to $\mathcal{K}$, with $W^\dagger W = \psi(\mathbb{1})$. Then

$$\begin{aligned} L(X^\dagger X) - X^\dagger L(X) - L(X^\dagger)X \;&=\; \Psi(X^\dagger X) - X^\dagger \Psi(X) - \Psi(X^\dagger)X + X^\dagger \Psi(\mathbb{1}) X \\ &=\; W^\dagger \pi(X^\dagger)\pi(X) W - X^\dagger W^\dagger \pi(X) W - V^\dagger \pi(X) W X + X^\dagger W^\dagger W X \\ &=\; (\pi(X)W - WX)^\dagger (\pi(X)W - WX) \;\geq\; 0\,, \end{aligned}$$

so that $D(L; X, X) \geq 0$. Repeating the argument with Ψ_n we see that L is completely dissipative.

□

Lecture 19

Quantum Trajectories: unraveling the Lindblad dynamics

"Quantum Trajectory theory" was developed by H. Carmichael 1993 to study phenomena in quantum optics, in particular counting of scattered photons from an optical cavity. See the review paper H. J. Carmichael 1999. It leads to predictions about the nature of the "quantum jumps" associated with measurements, that were ultimately observed experimentally Minev et al. 2019. In the next two lectures we will develop some of the associated mathematics.

19.1 Background: a bit about classical Markov jump processes

Let $R(x,y)$ be a non-negative, Borel measurable function on $\mathbb{R}^d \times \mathbb{R}^d$. We would like to consider a random process $\xi(t)$ on $\mathbb{R}^d$ that has the following dexcription: *if $\xi(t) = x$ at time t then the probability that $\xi(t+dt) \in K$ for some set $K \subset \mathbb{R}^d$, given the trajectory of ξ up to time t, is asymptotically $[\int_K R(x,y)dy]dt$ as $dt \to 0$.* Here $\int_K R(x,y)dy$ gives the *rate* at which a particle at x jumps into K. In order for this to make sense, we need the total rate for jumping away from x, which is $D(x) = \int_{\mathbb{R}^d} R(x,y)dy$, to be finite for each x. We will go further here and assume that $\sup_x D(x) < \infty$.

Can we construct such a process? First a word about what we mean by a *process*. Let Ω, Prob be a probability space. A *stochastic process* on Ω is a measurable map $\xi : [0,\infty) \times \Omega \to \mathbb{R}^d$. That is, for each realization $\omega \in \Omega$, we have a trajectory $t \mapsto \xi(t,\omega)$ in $\mathbb{R}^d$ (with the Borel σ algebra on $\mathbb{R}$ and $\mathbb{R}^d$). Typically we supress the variable ω and just write $\xi(t)$, understanding that it is a random variable depending on the realization ω.

Suppose the process we have in mind exists. It starts at some point $p_0\xi(0)$ and makes jumps at times $0 < \tau_1 < \tau_2 < \cdots$. The process is discontinuous at these times. We will take the convention that $\xi(t)$ is continuous from the right. Let τ_1 be the time of its first jump.

For $t > 0$ let $h(t) = \text{Prob}(\tau_1 \geq t)$. Then according to the above requirement

$$\frac{h(t) - h(t+dt)}{h(t, \mathbb{R}^d)} \;=\; \text{Prob}(\tau_1 < t + dt | \tau_1 \geq t) \;\approx\; D(p_0)dt\,.$$

Thus $h'(t) = D(p_0)h(t)$. Since $h(0) = 1$, we find that

$$\text{Prob}(\tau_1 > t) \;=\; h(t) \;=\; \mathrm{e}^{-D(p_0)t}\,.$$

We have come to the conclusion that the jump time τ_1 is an *exponential random variable* with rate $D(p_0)$. Let $p_1 = \xi(\tau_1)$ — this is the point to which the process jumps after leaving p_0. We have

$$\text{Prob}(p_1 \in K \text{ and } \tau_1 < t + dt | \tau_1 \geq t) \;\approx\; \left[\int_K R(p_0, y)dy\right]dt\,.$$

Thus

$$\text{Prob}(p_1 \in K \text{ and } t \leq \tau_1 < t + dt) \;=\; \left[\int_K R(p_0, y)dy\right]\mathrm{e}^{-D(p_0)t}dt\,.$$

Integrating from t to ∞, we find that

$$\begin{aligned}\text{Prob}(p_1 \in K \text{ and } \tau_1 \geq t) \;&=\; \left[\frac{1}{D(p_0)}\int_K R(p_0, y)dy\right]\mathrm{e}^{-D(p_0)t} \\ &=\; \left[\frac{1}{D(p_0)}\int_K R(p_0, y)dy\right]\text{Prob}(\tau_1 \geq t)\,.\end{aligned}$$

We conclude that p_1 is *independent* of τ_1 and

$$\text{Prob}(p_1 \in K) \;=\; \frac{1}{D(p_0)}\int_K R(p_0, y)dy\,.$$

Similar considerations apply to the subsequent jump times τ_j, $j = 2, \ldots$, and the points $\xi(\tau_j) = p_j$. Let $\mathcal{F}_j$ denote the sigma algebra generated by $(\tau_i, p_i)_{i=0}^j$ — that is by $\xi(\cdot)$ up to the time of the j-th jump. Conditioned on $\mathcal{F}_j$, τ_{j+1} and p_{j+1} are independent and satisfy

$$\text{Prob}(\tau_{j+1} > \tau_j + t | \mathcal{F}_j) \;=\; \mathrm{e}^{-D(p_{j-1})t} \tag{19.1}$$

and

$$\text{Prob}(p_{j+1} \in K | \mathcal{F}_j) \;=\; \frac{1}{D(p_j)}\int_K R(p_j, y)dy\,. \tag{19.2}$$

Theorem 19.1. *Let $R(x, y) \geq 0$ be a non-negative, Borel measurable function of $\mathbb{R}^d$ and let $D(x) = \int_{\mathbb{R}^d} R(x, y)dy$. If $\sup_x D(x) < \infty$, then for there is a probability space Ω, Prob such that for any $p_0 \in \mathbb{R}^d$ there is an $\mathbb{R}^d$ valued stochastic process on Ω with right continuous paths that satisfies $\xi(0) = p_0$ and*

$$\text{Prob}(\xi(t + h) \in K | (\xi(s))_{0 \leq s \leq t}) \;=\; h\int_K R(\xi(t), y)dy \;+\; o(h)\,. \tag{19.3}$$

Furthermore, any such process satisfies the following

1. *$\xi(t)$ is a piecewise constant function of t, with jumps at a random sequence of times $0 < \tau_1 < \tau_2 < \tau_3 < \cdots$.*

2. *Let $p_j = \xi(\tau_j)$, $j = 1, 2, \ldots$. Conditioned on the (τ_i, p_i) for $0 \le i \le j$, p_{j+1} and τ_{j+1} are independent random variables satisfying* (19.1) *and* (19.2).

3. *If $\rho(x,t)dy$ denotes the probability distribution of $\xi(t)$ at time t then*

$$\frac{\partial \rho}{\partial t}(x,t) \;=\; \int_{\mathbb{R}^d} R(y,x)\rho(y,t)dy \;-\; D(x)\rho(x,t)\;. \tag{19.4}$$

Sketch of proof. We have already seen that (1) and (2) have to hold for any such process. The main point is to construct the process. Let $(w_j)_{j=1}^{\infty}$ be a sequence of independent, identically distributed exponential variables with rate 1, i.e., $\mathrm{Prob}(w_j > t) = \mathrm{e}^{-t}$. Let $(p_j)_{j=1}^{\infty}$ be the Markov chain on $\mathbb{R}^d$ with

$$\mathrm{Prob}(p_j \in K | (p_i)_{i=0}^{j-1}) \;=\; \mathrm{Prob}(p_j \in K | p_{j-1}) \;=\; \frac{1}{D(p_j)} \int_K R(p_{j-1}, y)dy\;.$$

Take a probability space Ω on which $(w_j)_{j=1}^{\infty}$ and $(p_j)_{j=1}^{\infty}$ are realized independently. One could take this space to be $\Omega = \Omega_w \times \Omega_p$ with $\Omega_w = \mathbb{R}^{\mathbb{N}}$ and $\Omega_p = (\mathbb{R}^d)^{\mathbb{N}}$ with the correct probability measures on each. Define $\tau_j = \sum_{i=1}^{j} \frac{w_i}{D(p_{i-1})}$ and $\xi(t) \;=\; p_j \quad \text{if} \quad \tau_j \le t < \tau_{j+1}$. It is left as an exercise to check that (19.3) holds.

Now suppose that $f : \mathbb{R}^d \to \mathbb{R}$ is a smooth function that is, say, compactly supported. We have $\mathbb{E}\,(f(\xi(t))) = \int f(x)\rho(x,t)$. According to (19.3) we have

$$\mathbb{E}\,(f(\xi(t+dt))|\xi(t)) \;\approx\; \Big[\int R(\xi(t), y) f(y) dy\Big] dt + (1 - D(\xi(t))dt) f(\xi(t))\;,$$

where the second term accounts for the value of $f(\xi(t+dt))$ if no jump occurs in $[t, t+dt]$. Averaging over the value of $\xi(t)$ we obtain

$$\mathbb{E}\,(f(\xi(t+dt)) - f(\xi(t))) \;\approx\; \Big[\int \mathbb{E}\,(R(\xi(t), y)) f(y) dy - \mathbb{E}\,(D(\xi(t)) f(\xi(t)))\Big] dt\;.$$

Thus

$$\begin{aligned}\int f(x) \frac{\partial \rho}{\partial t}(x,t)dx \;&=\; \int \Big[\int R(x,y) f(y) dy - D(x) f(x)\Big] \rho(x) dx \\ &=\; \int \Big[\int R(y,x)\rho(y)dy - D(x)\rho(x)\Big] f(x)dx\;,\end{aligned}$$

from which (19.4) follows.

□

Lecture 20

Quantum Trajectories: unraveling the Lindblad dynamics (II)

Now are ready to consider the Lindblad equation:

$$\frac{d\rho_t}{dt} \;=\; -i\,[H,\rho_t] \;+\; \sum_{m=1}^{n} K_m \rho_t K_m^\dagger - \tfrac{1}{2}\{K_m^\dagger K_m\,,\;\rho_t\}\,. \tag{20.1}$$

As we will see there is a natural jump process associated to this equation that gives ρ_t after averaging. We will assume that

$$\left[H, K_m^\dagger K_m\right] \;=\; 0\,, \quad \text{for } m = 1,\dots,n\,. \tag{20.2}$$

As we saw above, this occurs naturally for the Lindblad equations derived by the Markov and rotating wave approximations. In this context it simplifies things a great deal. Below we will comment on how to adapt the discussion in case this is violated. A key player here is the *dissipation operator*

$$D \;:=\; \sum_{j=1}^{m} K_m^\dagger K_m. \tag{20.3}$$

It is a non-negative self-adjoint operator that commutes with H.

Our main goal is to find a jump process $|\psi(t)\rangle$ on pure states so that a solution to (20.1) with $\rho_0 = |\psi(0)\rangle\langle\psi(0)|$ is given by

$$\rho_t \;=\; \mathbb{E}\left[|\psi(t)\rangle\langle\psi(t)|\right]\,, \tag{20.4}$$

where $\mathbb{E}$ denotes disorder averaging. The motivating idea is that we want $|\psi(t)\rangle$ to evolve according to a "normal" quantum evolution in between jumps that occur at rates given by

$$\mathrm{Prob}(|\psi(t+dt)\rangle \propto K_m\,|\psi(t)\rangle\,|(|\psi(s)\rangle)_{0\le s\le t}) \;=\; \langle\psi(t)|\,K_m^\dagger K_m\,|\psi(t)\rangle\,dt + o(dt)\,. \tag{20.5}$$

Let us suppose that (20.5) holds and that in between jumps, $|\psi(t)\rangle$ is propagated by e^{itG} with G some "Hamiltonian." You might guess that we should take $G = H$, but that turns out not to be right: the correct G turns out to be non-self adjoint! Assuming (20.5), the *total* jump rate at time t is given by

$$D(t) \;:=\; \sum_{m=1}^{n} \langle\psi(t)|\,K_m^\dagger K_m\,|\psi(t)\rangle \;=\; \langle\psi(t)|\,D\,|\psi(t)\rangle\,, \tag{20.6}$$

and

$$\begin{aligned}
&\mathbb{E}\left[|\psi(t+dt)\rangle\langle\psi(t+dt)| \,\middle|\, (|\psi(s)\rangle)_{0\le s\le t}\right] \\
&\quad = (1-D(t)dt)\frac{\mathrm{e}^{-idtG}|\psi(t)\rangle\langle\psi(t)|\mathrm{e}^{idtG^\dagger}}{\langle\psi(t)|\mathrm{e}^{idt(G^\dagger-G)}|\psi(t)\rangle} \\
&\qquad + \left[\sum_{m=1}^n \langle\psi(t)|K_m^\dagger K_m|\psi(t)\rangle \frac{K_m|\psi(t)\rangle\langle\psi(t)|K_m^\dagger}{\langle\psi(t)|K_m^\dagger K_m|\psi(t)\rangle}\right]dt \;+\; o(dt) \\
&\quad = |\psi(t)\rangle\langle\psi(t)| \;+\; \Big[\big(\langle\psi(t)|i(G-G^\dagger)|\psi(t)\rangle - D(t)\big)|\psi(t)\rangle\langle\psi(t)| \\
&\qquad - iG|\psi(t)\rangle\langle\psi(t)| \;+\; i|\psi(t)\rangle\langle\psi(t)|G^\dagger \;+\; \sum_{m=1}^n K_m|\psi(t)\rangle\langle\psi(t)|K_m^\dagger\Big] + o(dt).
\end{aligned}$$

We see that by taking $G = H - i\frac{1}{2}D$ that

$$\begin{aligned}
&\mathbb{E}\left[|\psi(t+dt)\rangle\langle\psi(t+dt)| \,\middle|\, (|\psi(s)\rangle)_{0\le s\le t}\right] \\
&= |\psi(t)\rangle\langle\psi(t)| \;+\; \Big[-i\left[H,|\psi(t)\rangle\langle\psi(t)|\right] \;-\; \tfrac{1}{2}\left\{D\,,|\psi(t)\rangle\langle\psi(t)|\right\} \\
&\qquad + \sum_{m=1}^n K_m|\psi(t)\rangle\langle\psi(t)|K_m^\dagger\Big]dt \;+\; o(dt)\,.
\end{aligned} \tag{20.7}$$

Thus, upon averaging, we see that $\mathbb{E}\left[|\psi(t)\rangle\langle\psi(t)|\right]$ satisfies (20.1).

Theorem 20.1. *Consider the Lindblad equation* (20.1) *and suppose that* (20.2) *holds. If* $\rho_0 = |v_0\rangle\langle v_0|$*, with* $|v_0\rangle$ *a normalized vector, then there is a pure state valued stochastic process* $|\psi(t)\rangle$ *such that* $|\psi(0)\rangle = |v_0\rangle$ *and* (20.4) *holds. Furthermore,*

1. *The process is continuous from the right with limits from the left,*[1] *and is discontinuous at finitely or countably many times* $0 < \tau_1 < \tau_2 < \ldots$*, called the* jump times.
2. *We have* $\lim_j \tau_j = \infty$ (*if there are finitely many jumps, we take* $\tau_j = \infty$ *after the last jump*).
3. *If* D *is strictly positive there are countably many jumps.*
4. *If* $\left|v_j\right\rangle = \left|\psi(\tau_j)\right\rangle$, $j = 1, \ldots$*, then for* $\tau_j \le t < \tau_{j+1}$*, we have*

$$|\psi(t)\rangle \;\propto\; \mathrm{e}^{-(t-\tau_j)(iH+\frac{1}{2}D)}\left|v_j\right\rangle\,. \tag{20.8}$$

5. *Conditioned on* $(\tau_i, v_i)_{i=1}^j$*, we have*

$$\mathrm{Prob}(\tau_{j+1} > t+\tau_j | (\tau_i,v_i)_{i=1}^j) \;=\; \left\langle v_j\right|\mathrm{e}^{-tD}\left|v_j\right\rangle\,, \tag{20.9}$$

[1] That is $|\psi(t)\rangle = \lim_{s\downarrow t}|\psi(s)\rangle$ and $\lim_{s\uparrow t}|\psi(t)\rangle$ exists for all t. Such processes are called *càdlàg*, for *contiue à droite, limite à gauche*, in the probability literature.

and, with $\left|u_j\right\rangle = \lim_{t\uparrow\tau_{j+1}} |\psi(t)\rangle$*, we have for* $1 \le m \le n$*,*

$$\text{Prob}\left(\left|v_{j+1}\right\rangle \propto K_m \left|u_j\right\rangle \middle| (\tau_i, v_i)_{i=1}^{j}, \tau_{j+1}\right) \;=\; \frac{\left\langle u_j\right| K_m^\dagger K_m \left|u_j\right\rangle}{\left\langle u_j\right| D \left|u_j\right\rangle} \,. \tag{20.10}$$

Remarks. The generator $iH + \frac{1}{2}D$ is not self-adjoint, so the evolution (20.8) between jump times τ_j and τ_{j+1} is not unitary. The vector $|\phi(t)\rangle = \mathrm{e}^{-(t-\tau_j)(iH+\frac{1}{2}D)}\left|v_j\right\rangle$ is not normalized. However, because $[H, D] = 0$ we have

$$\langle\phi(t)|\phi(t)\rangle \;=\; \left\langle v_j\right| \mathrm{e}^{-(t-\tau_j)D} \left|v_j\right\rangle \,.$$

The state vector $|\psi(t)\rangle$ is

$$|\psi(t)\rangle \;=\; \frac{1}{\sqrt{\langle\phi(t)|\phi(t)\rangle}} |\phi(t)\rangle \;=\; \frac{1}{\sqrt{\langle v_j|\mathrm{e}^{-(t-\tau_j)D}|v_j\rangle}} \mathrm{e}^{-(t-\tau_j)(iH+\frac{1}{2}D)} \left|v_j\right\rangle \,.$$

The total jump rate $r(t)$ at time t is the derivative

$$D(t) \;=\; -\frac{d}{dt} \log \left\langle v_j\right| \mathrm{e}^{-(t-\tau_j)D} \left|v_j\right\rangle \;=\; \frac{\langle\phi(t)| D |\phi(t)\rangle}{\langle\phi(t)|\phi(t)\rangle} \;=\; \langle\psi(t)| D |\psi(t)\rangle \,,$$

as anticipated in (20.6). Using (20.10) we see that the rate of jumping from $|\psi(t)\rangle$ to a state proportional to $K_m |\psi(t)\rangle$ is given by

$$D(t) \frac{\langle\psi(t)| K_m^\dagger K_m |\psi(t)\rangle}{\langle\psi(t)| D |\psi(t)\rangle} \;=\; \langle\psi(t)| K_m^\dagger K_m |\psi(t)\rangle \,,$$

which is what we wanted. Thus (20.9) and (20.10) imply (20.5).

Proof. Let Ω be a probability space with $(w_j)_{j=1}^\infty$ independent, identically distributed exponential random variables with rate 1 and $(y_j)_{j=1}^\infty$ independent, identically distributed variables uniformly distributed in $[0, 1]$. Choose Ω and these variables so that the y's and w's are independent (e.g., you could take a product space).

We will define $(\tau_j, \left|v_j\right\rangle)_{j=1}^N$ recursively. Given $(\tau_i, |v_i\rangle)_{i=1}^j$, let $\tau_{j+1} = \tau_j + \delta_{j+1}$ with δ_{j+1} the unique value such that

$$\left\langle v_j\right| \mathrm{e}^{-\delta_{j+1}D} \left|v_j\right\rangle \;=\; \mathrm{e}^{-w_{j+1}} \,,$$

if such a value exists. If not, we take $\tau_{j+1} = \infty$ and stop the process. Note that δ_{j+1} exists provided

$$w_{j+1} < -\lim_{\delta\to\infty} \log \left\langle v_j\right| \mathrm{e}^{-\delta D} \left|v_j\right\rangle \;=\; -\log \left\langle v_j\right| P_0 \left|v_j\right\rangle \,,$$

with P_0 = projection onto the kernel of D. Thus, δ_{j+1} always exists if D is strictly positive. Clearly we have

$$\text{Prob}(\tau_{j+1} > t + \tau_j | (\tau_i, |v_i\rangle)_{i=0}^j) \;=\; \text{Prob}(w_{j+1} > -\log \left\langle v_j\right| \mathrm{e}^{-tD} \left|v_j\right\rangle) \;=\; \left\langle v_j\right| \mathrm{e}^{-tD} \left|v_j\right\rangle \,.$$

If $\tau_{j+1} < \infty$, let $\left|u_j\right\rangle = \mathrm{e}^{-(\tau_{j+1}-\tau_j)(iH+\frac{1}{2}D)}\left|v_j\right\rangle$ and take

$$p_j(m) \;=\; \frac{\left\langle u_j\right| K_m^\dagger K_m \left|u_j\right\rangle}{\left\langle u_j\right| D \left|u_j\right\rangle}\;, \quad m = 1, \ldots, n\;.$$

Note that $\sum_{m=1}^n p_j(m) = 1$. Choose m_{j+1} to be the smallest value of m with $\sum_{m'=1}^m p_j(m') > y_j$ and take

$$\left|v_{j+1}\right\rangle \;=\; \frac{1}{\sqrt{\langle u_j|K_{m_{j+1}}^\dagger K_{m_{j+1}}|u_j\rangle}} K_{m_{j+1}} \left|u_j\right\rangle\;,$$

so that we have

$$\mathrm{Prob}\left[\left|v_{j+1}\right\rangle \;=\; \frac{1}{\sqrt{\langle u_j|K_m^\dagger K_m|u_j\rangle}} K_m \left|u_j\right\rangle \,\middle|\, (\tau_i, |v_i\rangle)_{i=0}^j\,, \tau_{j+1}\right] \;=\; p_j(m) \;=\; \frac{\left\langle u_j\right| K_m^\dagger K_m \left|u_j\right\rangle}{\left\langle u_j\right| D \left|u_j\right\rangle}\;,$$

for $m = 1, \ldots, n$.

Let $d_m > 0$ be the largest eigenvalue of D. Then we have $\left\langle v_j\right| \mathrm{e}^{-tD} \left|v_j\right\rangle \geq \mathrm{e}^{-td_m}$ and thus $\delta_{j+1} \geq \frac{w_{j+1}}{d_m}$. It follows that

$$\tau_{j+1} \;\geq\; \frac{1}{d_m} \sum_{i=1}^{j+1} w_i \;\to\; \infty$$

as $j \to \infty$ with probability one. (Note that by the law of large numbers, $\frac{1}{j+1}\sum_{i=1}^{j+1} w_i \to 1$ with probability one.)

Now, for $t > 0$, let

$$|\phi(t)\rangle \;=\; \mathrm{e}^{-(t-\tau_j)(iH+\frac{1}{2}D)}\left|v_j\right\rangle \quad \text{if } \tau_j \leq t < \tau_{j+1}\;,$$

and set

$$|\psi(t)\rangle \;=\; \frac{1}{\sqrt{\langle\phi(t)|\phi(t)\rangle}}\,|\phi(t)\rangle\;.$$

Using (20.9) and (20.10) and the discussion before the theorem, we see that (20.7) and thus (20.1) hold.

□

If D is not strictly positive, then the $\ker D$ is an invariant subspace for the jump process. It can happen that $|\psi(t)\rangle$ lands in this subspace after finitely many steps, after which the evolution is simply given by e^{-itH}. It can also happen that $|\psi(t)\rangle$ fails to jump after some time, because the exponential variable w_j is too large. Then the evolution is given by $\mathrm{e}^{-(t-\tau_j)(iH+D)}\,|\psi(t)\rangle$ after the last jump, which converges as $t \to \infty$ to an element of $\ker D$. It can also happen that the process jumps infinitely often and does not converge to the kernel of D.

The requirement that $\left[H, K_m^\dagger K_m\right] = 0$ was convenient for defining the process. However it is possible to dispense with it. After the j-th jump, with $\left|\psi(\tau_j)\right\rangle = \left|v_j\right\rangle$, the rate of jumping shou

$$\begin{aligned} D(t) &= \langle\psi(t)|\, D\,|\psi(t)\rangle = \frac{\left\langle v_j\right| \mathrm{e}^{-(t-\tau_j)(-iH+\frac{1}{2}D)} D \mathrm{e}^{-(t-\tau_j)(iH+\frac{1}{2}D)} \left|v_j\right\rangle}{\left\langle v_j\right| \mathrm{e}^{-(t-\tau_j)(-iH+\frac{1}{2}D)} \mathrm{e}^{-(t-\tau_j)(iH+\frac{1}{2}D)} \left|v_j\right\rangle} \\ &= -\frac{d}{dt} \log \left\langle v_j\right| \mathrm{e}^{-(t-\tau_j)(-iH+\frac{1}{2}D)} \mathrm{e}^{-(t-\tau_j)(iH+\frac{1}{2}D)} \left|v_j\right\rangle = -\frac{d}{dt} \left\| \mathrm{e}^{-(t-\tau_j)(iH+\frac{1}{2}D)} \left|v_j\right\rangle \right\|^2 . \end{aligned}$$

We then define $\tau_{j+1} = \tau_j + \delta_{j+1}$ with δ_{j+1} such that

$$\left\| \mathrm{e}^{-\delta_{j+1}(iH+\frac{1}{2}D)} \left|v_j\right\rangle \right\|^2 = \mathrm{e}^{-w_{j+1}} ,$$

if such δ_{j+1} exsits. Given the jump time τ_{j+1} we choose $\left|v_{j+1}\right\rangle$ similarly, with

$$\left|v_{j+1}\right\rangle \propto K_m \left|u_j\right\rangle \quad \text{with probability } \frac{\left\langle u_j\right| K_m^\dagger K_m \left|u_j\right\rangle}{\left\langle u_j\right| D \left|u_j\right\rangle}$$

with $\left|u_j\right\rangle = \lim_{t\uparrow\tau_{j+1}} |\psi(t)\rangle$. Essentially the analysis is the same, but the distribution of the waiting times is more complicated.

20.1 An example

Consider the Lindblad equation for a qubit:

$$\frac{d\rho}{dt} = i\alpha\left[Z,\rho\right] + (\Gamma-\Delta) a^\dagger \rho a + (\Gamma+\Delta) a\rho a^\dagger - \tfrac{1}{2}\left\{\Gamma I - \Delta Z\,,\ \rho\right\} , \tag{20.11}$$

with $|\Delta| \le \Gamma$. The dissipation operator $D = \Gamma I + \Delta Z \ge 0$ and is strictly positive if $|\Delta| < \Gamma$. Since $a = |0\rangle\langle 1|$ and $a^\dagger |1\rangle\langle 0|$, after the first jump the quibit is in one of the two states $|0\rangle$ or $|1\rangle$. These states are eigenstates of the generator $i\alpha Z - \frac{1}{2}\Delta Z + \frac{1}{2}\Gamma$.

Let $|v_0\rangle\langle v_0|$ be the initial state with $|v_0\rangle = a|0\rangle + b|1\rangle$. Suppose that $|\Delta| < \Gamma$. Then the first jump occurs at a time τ_1 with

$$\mathrm{Prob}(\tau_1 > t) = |a|^2 \mathrm{e}^{-(\Gamma-\Delta)t} + |b|^2 \mathrm{e}^{-(\Gamma+\Delta)t} .$$

The state at time τ_1 is $|v_1\rangle$ with $v_1 = 0$ or 1. We have

$$|v_1\rangle \propto \begin{cases} |1\rangle & \text{with probability } \frac{|a|^2(\Gamma-\Delta)\mathrm{e}^{\Delta\tau_1}}{|a|^2(\Gamma-\Delta)\mathrm{e}^{\Delta\tau_1}+|b|^2(\Gamma+\Delta)\mathrm{e}^{-\Delta\tau_1}} , \\ |0\rangle & \text{with probability } \frac{|b|^2(\Gamma+\Delta)\mathrm{e}^{-\Delta\tau_1}}{|a|^2(\Gamma-\Delta)\mathrm{e}^{\Delta\tau_1}+|b|^2(\Gamma+\Delta)\mathrm{e}^{-\Delta\tau_1}} . \end{cases}$$

Subsequent jumps occur just between the states $|0\rangle$ and $|1\rangle$:

1. If $\left|v_j\right\rangle \propto |0\rangle$, then $\left|v_{j+1}\right\rangle \propto |1\rangle$,

$$\text{Prob}(\tau_{j+1} > t + \tau_j | (\tau_i, v_i)_{i=1}^{j}) \;=\; \mathrm{e}^{-(\Gamma-\Delta)t}\,,$$

and $|\psi(t)\rangle \;=\; \mathrm{e}^{i\alpha(t-\tau_j)}\left|v_j\right\rangle$ for $t \in [\tau_j, \tau_{j+1}]$.

2. If $\left|v_j\right\rangle \propto |1\rangle$, then $\left|v_{j+1}\right\rangle \propto |0\rangle$ and

$$\text{Prob}(\tau_{j+1} > t + \tau_j | (\tau_i, v_i)_{i=1}^{j} \;=\; \mathrm{e}^{-(\Gamma+\Delta)t}\,,$$

and $|\psi(t)\rangle \;=\; \mathrm{e}^{-i\alpha(t-\tau_j)}\left|v_j\right\rangle$ for $t \in [\tau_j, \tau_{j+1}]$.

After the first jump, the $|\psi(t)\rangle\langle\psi(t)|$ is piecewise constant in time t and jumps back and forth between $|0\rangle\langle 0|$ and $|1\rangle\langle 1|$ at the jump times.

Now suppose $\Delta = \Gamma$. With probability $|a|^2$, no jump at all occurs and the evolved state is simply

$$|\psi(t)\rangle \;=\; \frac{a\mathrm{e}^{i\alpha t+\frac{1}{2}\Delta t}\,|0\rangle + b\mathrm{e}^{-i\alpha t-\frac{1}{2}\Delta t}\,|1\rangle}{\sqrt{|a|^2\mathrm{e}^{\Delta t} + |b|^2\mathrm{e}^{-\Delta t}}}\,.$$

With probability $|b|^2 = 1 - |a|^2$ a single jump occurs at a time τ_1 with exponential distribution

$$\text{Prob}(\tau_1 > t) \;=\; \mathrm{e}^{-2\Delta t}\,.$$

After that time the state $|\psi(t)\rangle \propto |0\rangle$ and no further jumps occur.

Part III

Quantum Information Theory

Lecture 21

A bit of classical information theory

21.1 Shannon Entropy

Reading: Nielsen and Chuang 2010, §11.1

Let S be a finite set and let $\mathcal{P}(S)$ denote the set of Probability distributions on S, i.e, the set of maps $P : S \to [0,1]$ with $\sum_{s\in S} P(s) = 1$. The *Shannon entropy* of $P \in \mathcal{P}(S)$ is the quantity

$$H(P) \;=\; -\sum_{s\in S} P(s)\log P(s)\;. \tag{21.1}$$

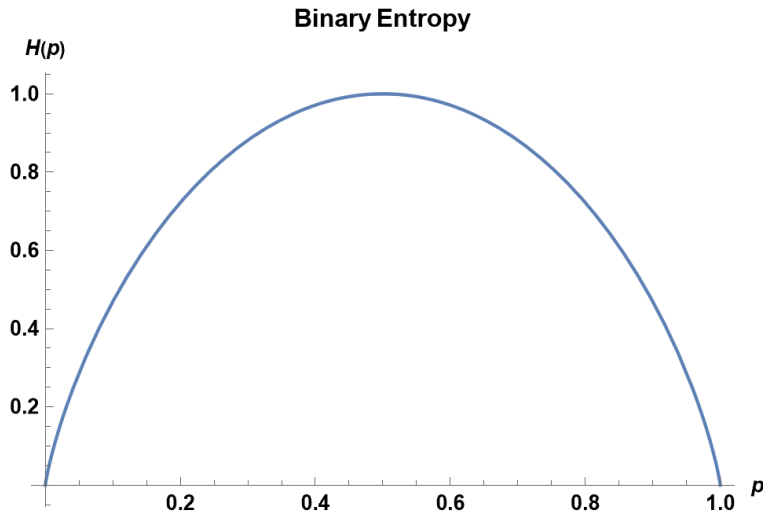


Figure 21.1: Binary entropy as a function of the probability of one outcome.

Consistent with Nielsen and Chuang 2010, we will take the log to denote *log based two*, which is standard in computer science and quantum information theory. Note that the choice of base for the logarithm affects the value of H only through a multiplicative constant. We take $0\log 0 \equiv 0$, consistent with the limiting behavior of $t\log t$. Thus $H(P) = 0$ if (and only if) $P(S) \subset \{0,1\}$.

The *binary entropy* of a two-outcome probability distribution $(p, 1-p)$ is

$$H_{\text{bin}}(p) \;:=\; -p\log p - (1-p)\log(1-p) \;=\; H(p, 1-p)\,. \tag{21.2}$$

The curve is shown in Figure 21.1. When clear from context, we will write $H(p)$ for $H_{\text{bin}}(p)$. Note that $H(p) = H(1-p)$ and $\max_p H(p) = H(\frac{1}{2}) = 1$. The simple value 1 for the maximum justifies the choice of the log based 2 — if we had taken the natural log the maximum would be $\ln 2$. Because

$$H''(p) \;=\; -\frac{1}{p} - \frac{1}{1-p} \;=\; -\frac{1}{p(1-p)} \;<\; 0\,,$$

the binary entropy is *strictly concave*

$$H(tp + (1-t)q) \;>\; tH(p) + (1-t)H(q) \qquad \text{if } p \neq q \text{ and } 0 < t < 1\,. \tag{21.3}$$

Exercise 21.1. Let $S = \{s_1, \dots, s_n\}$.

1. Show that
$$\max_{P \in \mathcal{P}(S)} H(P) \;=\; H(\tfrac{1}{n}, \cdots, \tfrac{1}{n}) \;=\; \log n\,.$$

2. Show that the Shannon entropy H is strictly concave on $\mathcal{P}(S)$. That is, if $P, Q \in \mathcal{P}(S)$ with $P \neq Q$ and $0 < t < 1$, then $H(tP + (1-t)Q) > tH(P) + (1-t)H(Q)$.

We can also speak of the entropy of random variables. Let (Ω, Prob) be a probability space. A *simple random variable* is a measurable map X defined on Ω taking only finitely many values. Let $\operatorname{ran} X$ denote the possible values of X and let p_X denote the *distribution of* X given by $P_X(x) = \text{Prob}[X = x]$ for $x \in \operatorname{ran} X$. The *entropy of* X is the defined to be $H(X) = H(P_X)$.

21.2 Shannon's Noiseless Coding Theorem

Reading: Nielsen and Chuang 2010, §12.2 and Watrous 2018, §5.3.1

The Shannon entropy was introduced by Shannon 1948 in his seminal work on communication theory. In this section we describe his *noiseless coding theorem*.

Source coding deals with data transmission and compression. Suppose we have an alphabet Σ, a finite set. A *coding scheme* of rate R for Σ is a sequence of pairs of mappings

$$\Sigma^n \;\xrightarrow{C_n}\; \{0,1\}^{\lfloor nR \rfloor} \;\xrightarrow{D_n}\; \Sigma^n\,.$$

The idea is that we wish to take source strings in Σ^n, code them using C_n to binary strings in $\{0,1\}^{\lfloor nR \rfloor}$, transmit them and then decode them using D_n. Thus we would like $D_n \circ C_n$ to be the identity on as many relevant strings as possible. One possibility would be to associate to each element of $s \in \Sigma$ a binary string of length $\lceil \log |\Sigma| \rceil$. This naturally defines a perfect coding scheme

$$\Sigma^n \;\xrightarrow{C_n}\; \{0,1\}^{n\lceil \log |\Sigma| \rceil} \;\xrightarrow{D_n}\; \Sigma_n\,,$$

with $D_n \circ C_n = \mathrm{Id}$. But we would also like to be efficient, and take m as small as possible. Can we do better than $R = \lceil \log |\Sigma| \rceil$?

As with the Latin alphabet in English, different elements of Σ are used with different frequencies. So we suppose give an *a priori* probability distribution $P \in \mathcal{P}(\Sigma)$ giving the frequency with which given letters in Σ are used. We will make the simplifying assumption that the source is an independent identically distributed source — that is, strings in Σ^n occur with the probability distribution

$$P_n(\boldsymbol{a}) \;=\; P(a_1) \cdots P(a_n) \;.$$

A *reliable coding scheme* is a coding scheme such that

$$\lim_{n\to\infty} \mathrm{Prob}\left[D_n(C_n(\boldsymbol{a})) \;=\; \boldsymbol{a}\right] \;=\; 1 \;, \tag{21.4}$$

where we suppose that the entries of $\boldsymbol{a} = (a_1, \ldots, a_n)$ are chosen independently from the distribution P. That is $\sum_{\boldsymbol{a}\in G_n} P_n(\boldsymbol{a}) = 1$, where

$$G_n \;=\; \{\boldsymbol{a} = (a_1, \ldots, a_n) \in \Sigma^n \;: D_n(C_n(\boldsymbol{a})) \;=\; \boldsymbol{a}\}$$

is the "good set" of strings in Σ^n for which the coding scheme is succesful.

Theorem 21.1 (Shannon's noiseless source coding theorem)**.** *Let Σ be an alphabet and let $P \in \mathcal{P}(\Sigma)$. If $R > H(P)$, then there exists a reliable coding scheme for P of rate R. If $R < H(P)$ then there is no reliable coding scheme.*

To sketch the proof of this theorem, let us introduce the idea of a *typical sequence*. Suppose $\boldsymbol{a} \in \Sigma^n$. Typically a given letter $x \in \Sigma$ should appear in the sequence roughly $nP(x)$ times. So we expect that for a typical sequence $\boldsymbol{a}$ we should have

$$P_n(\boldsymbol{a}) \;\approx\; \prod_{x\in\Sigma} P(x)^{nP(x)} \;\approx\; 2^{-nH(P)} \;.$$

With this heuristic in mind, we say that $\boldsymbol{a}$ is *ϵ-typical* if

$$2^{-n(H(P)+\epsilon)} \;\le\; P_n(\boldsymbol{a}) \;\le\; 2^{-n(H(X)-\epsilon)} \;,$$

which is to say that

$$\left|\frac{1}{n}\log\frac{1}{P_n(\boldsymbol{a})} - H(P)\right| \;\le\; \epsilon \;. \tag{21.5}$$

Theorem 21.2. *Let $\epsilon > 0$ be given. Then*

$$\lim_{n\to\infty} \mathrm{Prob}[\boldsymbol{a} \in \Sigma^n \textit{ is } \epsilon\textit{- typical}] \;=\; 1 \;.$$

That is

$$\lim_{n\to\infty} \sum_{\boldsymbol{a}\in T(n,\epsilon)} P_n(\boldsymbol{a}) \;=\; 1,$$

where $T(n,\epsilon)$ denotes the set of ϵ-typical sequences.

Proof. This follows from the Law of Large Numbers. The random variables $-\log P(a_j)$, $j = 1, 2, \ldots$, are independent an identically distributed. Thus for any $\epsilon, \delta > 0$, we have for sufficiently large n

$$\text{Prob}\left(\left|-\frac{1}{n}\sum_{j=1}^{n}\log P(a_j) + \mathbb{E}(\log P(a))\right| \leq \epsilon\right) \;\geq\; 1-\delta\,,$$

where

$$\mathbb{E}(\log P(a)) \;=\; \sum_{x\in\Sigma} P(x)\log P(x) \;=\; -H(P)\,,$$

and $\sum_{j=1}^{n}\log P(a_j) = \log P_n(\boldsymbol{a})$. The theorem follows using (21.5)

□

Corollary 21.3. *For each n, let $S(n)$ be a collection of at most 2^{nR} elements of Σ^n, with $R < H(P)$ fixed. Then*

$$\lim_{n\to\infty}\sum_{\boldsymbol{a}\in S(n)} P_n(\boldsymbol{a}) \;=\; 0\,.$$

Proof. Choose $\epsilon > 0$ such that $R + \epsilon < H(X)$. Let $T(n) = S(n) \cap T(n,\epsilon)$ and let $A(n) = S(n) \setminus T(n,\epsilon)$. We have

$$\lim_{n\to\infty}\sum_{\boldsymbol{a}\in A(n)} P_n(\boldsymbol{a}) \;\leq\; \lim_{n\to\infty}\sum_{\boldsymbol{a}\notin T(n,\epsilon)} P_n(\boldsymbol{a})\;0\,,$$

by Thm. 21.2. On the other hand

$$\sum_{\boldsymbol{a}\in T(n)} P_n(\boldsymbol{a}) \;\leq\; |S(n)|2^{-n(H(P)-\epsilon)} \;\leq\; 2^{-n(H(P)-R-\epsilon)} \;\to\; 0\,.$$

□

This corollary immediately implies the second part of Shannon's Thm. 21.1, since the size of the good set G_n is necessarily smaller than $|\{0,1\}^{\lfloor Rn\rfloor}| \leq 2^{nR}$. Thus any scheme with $R < H(P)$ is necessarily not reliable.

For the first part of the theorem, note that the number of ϵ-typical strings can be bounded as follows

$$|T(n,\epsilon)| \;\leq\; 2^{n(H(P)+\epsilon)}\sum_{\boldsymbol{a}\in T(n,\epsilon)} P_n(\boldsymbol{a}) \;\leq\; 2^{n(H(P)+\epsilon)}\,.$$

Given $R > H(P)$, pick $\epsilon > 0$ with $H(P)+\epsilon < R$. Pick some one-to-one map from $T(n,\epsilon)$ into $\{0,1\}^{\lfloor nR\rfloor}$, extend it to all of Σ^n by mapping all atypical sequences to some fixed sequence that we haven't used. The decoding map is any map that inverts the typical sequences. This defines a coding scheme that works perfectly for typical sequences. So $G_n \supset T(n,\epsilon)$ which proves the first part of the theorem.

21.3 Relative and Conditional Entropies and Mutual Information

Let S be a finite set and let $P, Q \in \mathcal{P}(S)$. The *relative entropy* of P with respect to Q is

$$H(P\|Q) \;:=\; \sum_{s\in S} P(s)\log\tfrac{P(s)}{Q(s)} \;=\; H(P) - \sum_s P(s)\log Q(s)\,, \tag{21.6}$$

with $H(P\|Q) = \infty$ if we have $P(s) \neq 0$ and $Q(s) = 0$ for some s.

Theorem 21.4. *The relative entropy is non-negative $H(P\|Q) \geq 0$ with equality if and only if $P \equiv Q$.*

Proof. Because the logarithm is strictly concave we have

$$\log x \;<\; c(x-1)$$

for all $x \neq 0$, where $c = \frac{1}{\ln 2} = \frac{d}{dx}\log x|_{x=1}$. Thus

$$\begin{aligned} H(P\|Q) \;&=\; -\sum_s P(S)\log\tfrac{Q(s)}{P(s)} \\ &\geq\; c\sum_s P(s)\left(1-\tfrac{Q(s)}{P(s)}\right) \;=\; c\sum_s (P(s)-Q(s)) \;=\; c(1-1) \;=\; 0\,. \end{aligned}$$

Equality can hold here if and only if $\frac{Q(s)}{P(s)} = 1$ for all s.

□

Given two random variables X and Y, each taking finitely many values, the *joint entropy* of X and Y is

$$H(X,Y) \;=\; -\sum_{x\in\operatorname{ran} X,\; y\in\operatorname{ran} Y} P(x,y)\log P(x,y)\,, \tag{21.7}$$

with $p(x,y) = \text{Prob}[X = x, Y = y]$ the joint distribution of X and Y. Thus the joint entropy is just the entropy of the random variable (X,Y). The *entropy of X conditioned on Y* is defined by

$$H(X|Y) \;:=\; H(X,Y) - H(X) \;=\; \sum_{x,y} P(x,y)\left[\log P(x,y) - \log P_X(x)\right]\,, \tag{21.8}$$

where $P_X(x) = \sum_y P(x,y)$.

The conditional entropy $H(X|Y)$ measures the uncertainty in X given knowledge of the outcome of Y. It is assymetric in X and Y. It is useful to have a symmetric measure of how much X and Y determine each other. The *mutual information of X and Y* is

$$H(X:Y) \;:=\; H(X) + H(Y) - H(X,Y)\,. \tag{21.9}$$

Theorem 21.5. *Let X and Y be random variables. Then*

1. *$H(X,Y) = H(Y,X)$ and $H(X:Y) = H(Y:X)$.*

2. $H(X:Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$.

3. $H(Y|X) \geq 0$ *and thus* $H(X) \leq H(X,Y)$ *and* $H(X:Y) \leq H(Y)$*, with equality if and only if* Y *is a function of* X*, i.e.,* $Y = f(X)$ *for some deterministic function* $f : \operatorname{ran} X \to \operatorname{ran} Y$.

Proof. (1) and (2) follow directly from the definitions. To prove (3), let $P(x,y)$ denote the joint probability distribution of X and Y, let $P(x) = \sum_y P(x,y)$ denote the distribution of X, and let $P(y|x) = \frac{P(x,y)}{P_X(x)}$ denote the distribution of Y given that $X = x$. Then we have $P(x,y) = P(x)P(y|x)$, so

$$\begin{aligned} H(X,Y) &= -\sum_{x,y} P(x,y) \log p(x)P(y|x) \\ &= -\sum_x P(x) \log P(x) - \sum_x P(x) \sum_y P(y|x) \log P(y|x) \\ &= H(X) + \sum_x P(x) H(P(\cdot|x)) \,. \end{aligned}$$

We conclude that $H(Y|X) = \sum_x P(x) H(P(\cdot|x)) \geq 0$, with equality if and only $H(P(\cdot|x)) = 0$ for each x, from which we conclude that $Y = f(X)$.

□

Theorem 21.6 (Subbaditivity and Strong Subadditivity)**.** *Let* X*,* Y*, and* Z *be random variables. Then*

1. ***Subadditivity****:* $H(X,Y) \leq H(X) + H(Y)$ *with equality if and only if* X *and* Y *are independent.*

2. ***Strong Subadditivity****:* $H(X,Y,Z) + H(Y) \leq H(X,Y) + H(Y,Z)$*, with equality if and only if* $Z \to Y \to X$ *is a Markov chain.*

Remarks.

1) Recall that X and Y are *indenpendent* if and only if their joint distribution $P(X,Y) = P_X(x)P_Y(y)$.

2) $Z \to Y \to X$ is a *Markov chain* if the conditional distribution of X given Y and Z depends only on Y. That is

$$\text{Prob}(X = x | Y = y,\ Z = z) \ = \ \text{Prob}(X = x | Y = y)$$

for all tripples (x,y,z).

Proof. To prove subadditivity, note that

$$\begin{aligned} H(X) + H(Y) - H(X,Y) &= \sum_{x,y} P(x,y) \log \frac{P_X(x)P_Y(y)}{P(x,y)} \\ &\leq \frac{1}{\ln 2} \sum_{x,y} P(x,y) \left(\frac{P_X(x)P_Y(y)}{P(x,y)} - 1 \right) \\ &\leq \frac{1}{\ln 2} \sum_{x,y} (P_X(x)P_Y(y) - P(x,y)) \ = \ 1 - 1 \ = \ 0 \,, \end{aligned}$$

with equality if and only if $P_X(x)P_Y(y) = P(x,y)$ for every x, y.

To prove strong subadditivity, note that

$$\begin{aligned} H(X,Y) + H(Y,Z) - H(X,Y,Z) - H(Y) &= \sum_{x,y,z} P(x,y,z) \log \frac{P_{X,Y}(x,y)P_{Y,Z}(y,z)}{P(x,y,z)P_Y(y)} \\ &= \sum_{x,y,z} P(x,y,z) \log \frac{P_{X|Y}(x|y)P_{Y,Z}(y,z)}{P(x,y,z)} \end{aligned}$$

where $P_{X|Y}(x|y) = \frac{P_{X,Y}(x,y)}{P_Y(y)}$ is the conditional probability that $X = x$ given that $Y = y$. Thus

$$\begin{aligned} H(X,Y) + H(Y,Z) - H(X,Y,Z) - H(Y) &\leq \frac{1}{\ln 2} \sum_{x,y,z} (P_{X|Y}(x|y)P_{Y,Z}(y,z) - P(x,y,z)) \\ &= \frac{1}{\ln 2} \sum_{x,y} (P_{X|Y}(x|y)P_Y(y) - P_{X,Y}(x,y)) \;=\; 0\,, \end{aligned}$$

with equality if and only if

$$P_{X|Y}(x|y) \;=\; \frac{P(x,y,z)}{P_{Y,Z}(y,z)} \;=\; \mathrm{Prob}(X = x | Y = y, Z = z)\,,$$

which is equivalent to $Z \to Y \to X$ being a Markov chain.

□

Lecture 22

The von Neumann Entropy

Reading: Nielsen and Chuang 2010, Chapter 11

The *von Neumann entropy* of a quantum state with density matrix ρ is defined to be

$$S(\rho) \;:=\; -\operatorname{tr}[\rho \log \rho]\;. \tag{22.1}$$

As above we will take log based two, and interpret $0 \log 0$ as 0. If λ_j, $j = 1, \ldots, n$ are the eigenvalues of ρ then

$$S(\rho) \;=\; -\sum_j \lambda_j \log \lambda_j\;,$$

which is just the Shannon entropy of the probability distribution $(\lambda_1, \ldots, \lambda_n)$. Thus we can tink of the von Neumann entropy as measuring the uncertainty as to which pure state is in, with respect to a measurement in the eigenbasis of ρ.

Theorem 22.1 (Basic Properties of the von Neumann Entropy)**.** *Let ρ be a state of quantum system with Hilbert space $\mathcal{H}$ of dimension d.*

1. *$S(\rho) \geq 0$ with equality if and only if ρ is a pure state.*
2. *$S(\rho) \leq \log d$, with equality if and only if $\rho = \frac{1}{d}\mathbb{1}$.*
3. *If $\rho = \sum_{i=1}^m p_i \rho_i$ where $\boldsymbol{p} = (p_1, \ldots, p_m)$ is a probability distribution and the states ρ_i are supported on orthogonal subspaces, then*

$$S(\rho) \;=\; H(\boldsymbol{p}) \;+\; \sum_i p_i S(\rho_i)\;. \tag{22.2}$$

The proof is left as an exercise — a proof of (3) can be found in Nielsen and Chuang 2010.

22.1 Relative Entropy

Given two density matrices ρ and σ we define the *relative entropy* of ρ to σ by

$$S(\rho \| \sigma) \;:=\; \operatorname{tr}[\rho \log \rho] - \operatorname{tr}[\rho \log \sigma]\;. \tag{22.3}$$

The second term should be interpreted as

$$-\operatorname{tr}[\rho \log \sigma] \;=\; \sum_{j=1}^{n} \left\langle \phi_j \middle| \rho \middle| \phi_j \right\rangle \log \mu_j \,,$$

with $(\mu_j, \phi_j)_{j=1}^n$ the eigenvalues and eigenvectors for σ. If $\mu_j = 0$ and $\left\langle \phi_j \middle| \rho \middle| \phi_j \right\rangle \neq 0$ for some j, then this term and the relative entropy are equal to ∞.

Theorem 22.2 (Klein 1931)**.** *The quantum relative entropy $S(\rho\|\sigma)$ is non-negative and equal to zero if and only if $\rho = \sigma$.*

Proof. Let $\rho = \sum_i \lambda_i |\psi_i\rangle \langle\psi_i|$ and $\sigma = \sum_j \mu_j \left|\phi_j\right\rangle \left\langle\phi_j\right|$ be the spectral decompositions for ρ and σ. Then

$$\begin{aligned} S(\rho\|\sigma) \;&=\; \sum_i \lambda_i \log \lambda_i (1 - \langle\psi_i| \log \sigma \, |\psi_i\rangle) \\ &=\; \sum_i \lambda_i \log \lambda_i \left(1 - \sum_j \log\big(\mu_j\big) P_{i,j}\right) \,, \end{aligned}$$

with $P_{i,j} = \left|\left\langle \psi_i \middle| \phi_j \right\rangle\right|^2$. Note that $P_{i,j} \geq 0$, $\sum_i P_{i,j} = 1$, and $\sum_j P_{i,j} = 1$. By the strict concavity of log, we have

$$\sum_j P_{i,j} \log\big(\mu_j\big) \;\leq\; \log\left(\underbrace{\sum_j P_{i,j}\mu_j}_{r_i}\right) \,,$$

with equality if and only if $P_{i,j} = 1$ for some j. Thus

$$S(\rho\|\sigma) \;\geq\; \sum_i \lambda_i \log \frac{\lambda_i}{r_i}$$

with equality if and only if for each i there is a j such that $P_{i,j} = 1$, i.e., if and only if $(P_{i,j})$ is a permutation matrix. Note that $\sum_i r_i = \sum_{i,j} \mu_j P_{i,j} = \operatorname{tr} \sigma = 1$. Thus the right hand side is the classical relative entropy of the probability distribution $(\lambda_1, \ldots, \lambda_n)$ with respect to $(r_1, \ldots, r_n)$. By Thm. 21.4, this is bounded below by 0, with equality if and only if $r_i = \lambda_i$. Thus we have $S(\rho\|\sigma) \geq 0$ with equality if and only if $P_{i,j}$ is a permutation matrix and $r_i = \lambda_i$ for all i, which is to say if and only if $\rho = \sigma$.

□

22.2 Projective measurements increase entropy

Theorem 22.3. *Suppose that $(P_i)_{i=1}^n$ is a projective measurement ensemble (a complete set of orthogonal projections) and ρ is a density matrix. Let $\rho_i = \sum_i P_i \rho P_i$. Then $S(\rho') \geq S(\rho)$, with equality if and only if $\rho = \rho'$*

Remark. The state ρ' is the mixed state obtained after the projective measurement associated to $(P_i)_{i=1}^n$.

Proof. We apply Klein's inequality (Thm. 22.2) to ρ and ρ',

$$0 \;\leq\; S(\rho'\|\rho) \;=\; S(\rho) - \operatorname{tr}(\rho \log \rho') \;.$$

We will show that $S(\rho') = -\operatorname{tr}(\rho \log \rho')$ this will prove the result. We have

$$-\operatorname{tr}(\rho \log \rho') \;=\; -\operatorname{tr} \sum_i P_i \rho \log \rho' \;=\; -\sum_i \operatorname{tr}(P_i \rho \log \rho' P_i) \;,$$

using $\sum_i P_i = \mathbb{1}$, $P_i^2 = P_i$ and cyclicity of the trace. Using that $P_i P_j = 0$ for $i \neq j$ and $P_i^2 = P_i$ again we see that $\rho' P_i \;=\; P_i \rho P_i = P_i \rho'$, i.e., ρ' commutes with P_i. Thus

$$-\operatorname{tr}(\rho \log \rho') \;=\; -\sum_i \operatorname{tr}(P_i \rho P_i \log \rho') \;=\; -\operatorname{tr} \rho' \log \rho' \;=\; S(\rho') \;.$$

□

22.3 Entropies in Composite Systems

Suppose we have a composite system (A, B) with Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$. The *joint entropy* $S(A, B)$ for the composite system is just the von Neumann entropy $S(A, B) \;=\; -\operatorname{tr} \rho_{AB} \log \rho_{AB}$, with $\rho_{A,B}$ the state of the system. The entropy of subsystem A is $S(A) := S(\rho_A)$ where $\rho_A = \operatorname{tr}_B \rho_{A,B}$ is the reduced density matrix for A. The entropy of B is defined similarly. The *conditional entropy* of A relative to B and *mutual information* of (A, B) are given by

$$S(A|B) \;:=\; S(A, B) - S(B) \;, \tag{22.4}$$

and

$$S(A : B) \;:=\; S(A) + S(B) - S(A, B) \;=\; S(A) - S(A|B) \;=\; S(B) - S(B|A) \;. \tag{22.5}$$

The key new feature of quantum systems is *entanglement*. For instance, if $\rho = |\Psi\rangle \langle\Psi|$ is a pure state, the reduced density matrices need not be pure. Thus even if $S(A, B) = 0$, we may have $S(B)$ and $S(A)$ not equal to zero. In particular, the conditional entropy can be negative — learning information about B may increase our uncertainty!

Proposition 22.4. *If the composite system (A, B) is in the pure state $|\Psi\rangle \langle\Psi|$ then $S(A) = S(B)$.*

Proof. Recall that $|\Psi\rangle$ has a *Schmidt Decomposition*

$$|\Psi\rangle \;=\; \sum_{i=1}^{m} \alpha_i \, |\phi_i\rangle \otimes |\psi_i\rangle$$

with $\alpha_i > 0$ and $(\phi_i)_i$, $(\psi_i)_i$ orthonormal sequences in $\mathcal{H}_A$, $\mathcal{H}_B$, repsectively. It follows that

$$\rho_A \;=\; \sum_{i=1}^{m} \alpha_i \, |\phi_i\rangle \langle\phi_i| \quad \text{and} \quad \rho_B \;=\; \sum_{i=1}^{m} \alpha_i \, |\psi_i\rangle \langle\psi_i| \;.$$

Thus ρ_A and ρ_B have the same eigenvalues. Since the von Neumann entropy only depends on the eigenvalues of the density matrix, the result follows.

□

Definition 22.5. The *bipartite entanglement entropy* of a pure state $\Psi \in \mathcal{H}_A \otimes \mathcal{H}_B$, with respect to the decomposition (A, B), is the entropy $S(A) = S(B)$.

Theorem 22.6. *Let the composite quantum system (A, B) have joint state ρ_{AB}. Then*

1. **Subadditivity***: $S(A, B) \leq S(A) + S(B)$, with equality if and only if (A, B) is in a product state, i.e., $\rho_{AB} = \rho_A \otimes \rho_B$.*
2. **Triangle Inequality***: $S(A, B) \geq |S(A) - S(B)|$*

Remark. The triangle inequality is due to Araki and Lieb 1970.

Exercise 22.1. Suppose the composite system is in a product state. Prove that $S(A, B) = S(A) + S(B)$.

Proof. Let $\sigma = \rho_A \otimes \rho_B$ and $\rho = \rho_{A,B}$. By Klein's inequality (Thm. 22.2),

$$\begin{aligned} S(\rho) \leq -\operatorname{tr}[\rho \log \sigma] &= -\operatorname{tr} \rho(\log \rho_A + \log \rho_B) \\ &= -\operatorname{tr} \rho_A \log \rho_A - \operatorname{tr} \rho_B \log \rho_B = S(A) + S(B) . \end{aligned}$$

Subbaditivity follows. Note that, by Thm. 22.2, equality holds if and only if $\rho = \sigma$.

To prove the triangle inequality, let R be an auxiliary system with Hilbert space $\mathcal{H}_R$ such that $\rho = \rho_{A,B}$ has a purification $|\Psi\rangle\langle\Psi|$ on $\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_R$. That is $\rho_{A,B} = \operatorname{tr}_R |\Psi\rangle\langle\Psi|$. Then we have $S(R) + S(A) \geq S(A, R)$ by subbaditivity. Furthermore, since (A, B, R) is in a pure state, we have $S(A, R) = S(B)$ and $S(R) = S(A, B)$ by Prop. 22.4. Thus

$$S(A, B) = S(R) \geq S(A, R) - S(A) = S(B) - S(A) .$$

Similarly, one shows that $S(A, B) \geq S(A) - S(B)$. The triangle inequality follows.

□

22.4 Concavity of the entropy and the entropy of mixtures

Proposition 22.7 (Concavity of the entropy)**.** *Let ρ be the state of a quantum system A. If $\rho = \sum_{i=1}^n p_i \rho_i$ with $(p_i)_{i=1}^n$ probabilities and $(\rho_i)_{i=1}^n$ states, then*

$$\sum_{i=1}^n p_i S(\rho_i) \leq S(\rho) ,$$

with equality if and only if $\rho_i = \rho$ for all i with $p_i \neq 0$.

Proof. Let B be an auxiliary system with dimension n and basis $(|i\rangle)_{i=1}^n$. Introduce a state $\rho_{AB} = \sum_i p_i \rho_i \otimes |i\rangle\langle i|$. The $\rho_A = \rho$ while $\rho_B = \sum_i p_i |i\rangle\langle i|$. Thus

$$S(A) = S(\rho) \quad \text{and} \quad S(B) = S\left(\sum_i p_i |i\rangle\langle i|\right) = H(\boldsymbol{p}) ,$$

with $\boldsymbol{p} = (p_1, \ldots, p_n)$.

Let λ_i^j, $j = 1, \ldots, d$, be the eigenvalues of ρ_i for $i = 1, \ldots, n$, with eigenvectors $\left|\phi_i^j\right\rangle$. Then the eigenvalues of $\rho_{A,B}$ are seen to be $p_i\lambda_i^j$ for $i = 1, \ldots, n$ and $j = 1, \ldots, d$, with corresponding eigenvectors $\left|\phi_i^j\right\rangle \otimes |i\rangle$. Thus

$$S(\rho_{A,B}) \;=\; -\sum_{i=1}^{n}\sum_{j=1}^{d} p_i\lambda_i^j(\log p_i + \log\lambda_i^j) \;=\; H(\boldsymbol{p}) + \sum_{i=1}^{n} p_i S(\rho_i)\;.$$

Thus by subadditivity

$$\sum_{i=1}^{n} p_i S(\rho_i) + H(\boldsymbol{p}) \;\leq\; S(\rho) + H(\boldsymbol{p})$$

with equality if and only if ρ_i is independent of i (so that $\rho_{A,B}$ is a product state).

□

Theorem 22.8 (Entropy of Mixtures)**.** *Suppose that $\rho = \sum_i p_i\rho_i$ with $\boldsymbol{p} = (p_1, \ldots, p_n)$ probabilities and ρ_i density matrices. Then*

$$S(\rho) \;\leq\; \sum_i p_i S(\rho_i) + H(\boldsymbol{p})$$

with equality if and only if the states ρ_i are supported on orthogonal subspace.

Proof. Suppose to begin that $\rho_i = |\psi_i\rangle\langle\psi_i|$ is a pure state for each $i = 1, \ldots, n$. Call our system A and introduce an auxiliary system B with orthonormal basis $|i\rangle$, $i = 1, \ldots, n$. Define

$$|\Psi\rangle \;:=\; \sum_{i=1}^{N} \sqrt{p_i}\,|\psi_i\rangle\,|i\rangle\;.$$

Because $|\Psi\rangle$ is a pure state we have

$$S(B) \;=\; S(A) \;=\; S(\rho_A) \;=\; S\left(\sum_i p_i\,|\psi_i\rangle\langle\psi_i|\right) \;=\; S(\rho)\;.$$

Now consider a projective measurement of B in the basis $|i\rangle$. The density matrix ρ' after measurement is given by

$$\rho' \;=\; \sum_{i=1}^{n} p_i\,|i\rangle\langle i|\Psi\rangle\langle\Psi|i\rangle\langle i| \;=\; \sum_{i=1}^{n} p_i\,|\psi_i\rangle\langle\psi_i| \otimes |i\rangle\langle i|\;,$$

so that the state of system B is

$$\rho'_B \;=\; \sum_i p_i\,|i\rangle\langle i|\;.$$

Thus $S(\rho'_B) = H(\boldsymbol{p})$ and

$$S(\rho) \;=\; S(A) \;=\; S(B) \;\leq\; S(\rho'_B) \;\leq\; H(\boldsymbol{p})\;.$$

□

Lecture 23

Strong Subadditivity of the Entropy

Lecture 24

Quantum Source Coding

Appendix

Appendix A

Hilbert Spaces

Recall that an *inner product* $\langle\cdot|\cdot\rangle$ on a vector space $\mathcal{H}$ over $\mathbb{C}$ is a map from $\mathcal{H}\times\mathcal{H}\to\mathbb{C}$ such that

1. for fixed $\phi\in\mathcal{H}$, the map $\psi\mapsto\langle\phi|\psi\rangle$ is linear;
2. the map is skew-symmetric, $\langle\psi|\phi\rangle = \langle\phi|\psi\rangle^*$; and
3. for each $\phi\in\mathcal{H}$, $\langle\phi|\phi\rangle\geq 0$, with equality if and only if $\phi=0$.

Note that skew-symmetry implies, by (2), that for fixed $\psi\in\mathcal{H}$, the map $\phi\mapsto\langle\phi|\psi\rangle$ is conjugate linear, meaning that

$$\langle\phi_1+\alpha\phi_2|\psi\rangle = \langle\phi_1|\psi\rangle+\alpha^*\langle\phi_2|\psi\rangle$$

for $\alpha\in\mathbb{C}$. Associated to the inner product is the norm

$$\|\phi\| = \sqrt{\langle\phi|\phi\rangle}\,,$$

and the associated metric $d(\phi,\psi) = \|\phi-\psi\|$. Any inner product satisies the *Cauchy-Schwartz* inequality

$$|\langle\phi|\psi\rangle| \leq \|\phi\|\|\psi\|\,, \tag{A.1}$$

as can be seen by minimizing the positive function $\|\phi+\alpha\psi\|^2$, where α ranges over $\mathbb{C}$. Recall that a Hilbert space $\mathcal{H}$ is a vector space of $\mathbb{C}$ with an *inner product* $\langle\cdot|\cdot\rangle$ such that $\mathcal{H}$ is a complete metric space with respect to the distance $d(\phi,\psi)=\|\phi-\psi\|$.

A linear map $B:\mathcal{H}_1\to\mathcal{H}_2$ is called a *bounded* if there is a constant $c<\infty$ such that

$$\|B\psi\|_{\mathcal{H}_2} \leq c\|\psi\|_{\mathcal{H}_1}\,.$$

The smallest such constant is called the *norm of* B, and is denoted $\|B\|$. The set of all bounded linear maps from $\mathcal{H}_1\to\mathcal{H}_2$ is denoted $\mathcal{B}(\mathcal{H}_1,\mathcal{H}_2)$, or by $\mathcal{B}(\mathcal{H}_1)$ in case $\mathcal{H}_1=\mathcal{H}_2$. A bounded linear map from $\mathcal{H}\to\mathbb{C}$ is called a *bounded linear functional*. The *Riesz-Frechet representation theorem* states that every bounded linear functional on a Hilbert space $\mathcal{H}$ is

of the form $\phi \mapsto \langle \psi | \phi \rangle$ for some $\psi \in \mathcal{H}$. Given $B \in \mathcal{B}(\mathcal{H}_1, \mathcal{H}_2$, the Riesz-Frechet theorem allows one to define the *adjoint map* $B^\dagger \in \mathcal{B}(\mathcal{H}_2, \mathcal{H}_1)$ via

$$\left\langle B^\dagger \psi | \phi \right\rangle_{\mathcal{H}_1} \;=\; \langle \psi | B\phi \rangle_{\mathcal{H}_2}$$

for $\phi \in \mathcal{H}_1$ and $\psi \in \mathcal{H}_2$. A map $B \in \mathcal{B}(\mathcal{H})$ is *self-adjoint* if $B = B^\dagger$ and $B \in \mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$ is *unitary* if $B^\dagger B = I_{\mathcal{H}_1}$ and $BB^\dagger = I_{\mathcal{H}_2}$. Equivalently, B is unitary if it is surjective and $\|B\phi\|_{\mathcal{H}_2}$ for all $\phi \in \mathcal{H}_1$.

A set $S \subset \mathcal{H}$ is called orthonormal if

1. $\|\phi\| = 1$ for every $\phi \in S$, and
2. $\langle \phi | \psi \rangle = 0$ for $\phi, \psi \in S$ with $\phi \neq \psi$.

For any orthonormal set S one has *Bessel's inequality*

$$\sum_{\phi \in S} |\langle \phi | \psi \rangle|^2 \;\leq\; \|\psi\|^2 \tag{A.2}$$

for every $\psi \in \mathcal{H}$. An orthonormal set is an *orthonormal basis* if equality holds in Bessel's inequality for every ψ, or equivalently that S is a maximal orthonormal set. Using Zorn's lemma one can show that every Hilbert space has an orthonormal basis.

A Hilbert space $\mathcal{H}$ is called *separable* if it has a countable dense subset. It is a fact that $\mathcal{H}$ is separable if and only if either

1. $\dim \mathcal{H} < \infty$, or
2. $\mathcal{H}$ has a countable orthonormal basis.

Most Hilbert spaces that arise in practice are separable. For example $L^2(\mathbb{R}^n)$ is separable for every n and $\ell^2(\mathbb{N})$ is separable, etc. Any two separable, infinite dimensional Hilbert spaces are isomorphic as Hilbert spaces, i.e., there is a unitary map between them.[1] To see this, let $\mathcal{H}_j$, $j = 1, 2$ be separable and infinite with orthonormal bases $\phi_1^{(j)}, \phi_2^{(j)}, \ldots$ for $j = 1, 2$. Given $\psi \in \mathcal{H}_1$ let

$$U\psi \;=\; \sum_{k=1}^{\infty} \left\langle \phi_k^{(1)} \middle| \psi \right\rangle \phi_k^{(2)} \,.$$

One can show that the series on the right hand side is absolutely convergent and the resulting map is unitary.

1.1 The Riesz Representation Theorem

The following theorem is an important result in Hilbert space theory

[1] More generally, one can define the Hilbert space dimension of a Hilbert space to be the cardinality of a maximal orthonormal set. Any two Hilbert spaces of the same cardinality are isomorphic.

Theorem A.1 (Riesz Representation Theorem)**.** *Let $\ell : \mathcal{H} \to \mathbb{C}$ be a bounded linear functional on a Hilbert space $\mathcal{H}$. Then there is a unique $\psi \in \mathcal{H}$ such that $\ell(\phi) = \langle \psi | \phi \rangle$*

Remark. Recall that ℓ is bounded if there is a constant $c < \infty$ such that $|\ell(\phi)| \leq c\|\phi\|$ for all ϕ. Bounded linear functionals are precisely the continuous linear maps from $\mathcal{H}$ to $\mathbb{C}$.

Proof. Let ℓ be a bounded linear functional, and let $Y = \ker \ell$ be its kernel. Then $Y = \ell^{-1}(0)$ is a closed subspace of $\mathcal{H}$. If $\ell \equiv 0$, then $Y = \mathcal{H}$ and we must take $\psi = 0$. If $\ell \not\equiv 0$, then by elementary linear algebra, Y is of co-dimension 1. It follows that $Y^\perp = \{v : \langle v | \phi \rangle = 0 \text{ for all } \phi \in Y\}$ is one-dimensional. Let $\widehat{\psi}$ be a unit vector in $Y^\perp$. If $\phi \in \mathcal{H}$ is a general vector then we have

$$\phi \;=\; \left\langle \widehat{\psi} | \phi \right\rangle \widehat{\psi} + \phi'$$

with $\phi' = \phi - \left\langle \widehat{\psi} | \phi \right\rangle \widehat{\psi}$. Thus

$$\ell(\phi) \;=\; \left\langle \widehat{\psi} | \phi \right\rangle \ell(\widehat{\psi}) \;=\; \left\langle (\ell(\widehat{\psi}))^* \widehat{\psi} | \phi \right\rangle .$$

Thus we may take $\psi = (\ell(\widehat{\psi}))^* \widehat{\psi}$.

Uniqueness follows because if $\langle \psi | \phi \rangle = \langle \psi' | \phi \rangle$ for all ϕ, then taking $\phi = \psi - \psi'$ yields

$$\|\psi - \psi'\|^2 \;=\; 0 .$$

□

One key use for the Riesz representation theorem is to prove the existence of operators associated to *quadratic forms* on $\mathcal{H}$. A *quadratic form* on $\mathcal{H}$ is a map $q : \mathcal{H} \to \mathbb{R}$ that is

1. *homogeneous of order* 2, $q(\alpha\psi) = |\alpha|^2 q(\psi)$ for all $\alpha \in \mathbb{C}$ and $\psi \in \mathcal{H}$, and
2. satisfies the parallelogram identity, $q(\psi + \phi) + q(\psi - \phi) = 2q(\phi) + 2q(\psi)$.

A quadratic form q is *bounded* if $q(\psi) \leq c\|\psi\|^2$ for some $c < \infty$. Associated to any bounded quadratic form q is a *bounded sesqui-linear* form $\tilde{q} : \mathcal{H} \times \mathcal{H} \to \mathbb{C}$ such that $q(\phi) = \tilde{q}(\phi, \phi)$.

Exercise A.1 (See exercise 4, §3.1 of Simon 2015b.). Suppose that q is a bounded quadratic form. Define $\tilde{q}$ by the polarization identity

$$\tilde{q}(\phi, \psi) \;=\; \frac{1}{4}\left[q(\phi + \psi) - q(\phi - \psi) + iq(\phi + i\psi) - iq(\phi - i\psi)\right] . \tag{A.3}$$

Prove that

1. $\tilde{q}$ is *sesqui-linear*, i.e.
   (a) $\tilde{q}$ is skew-symmetric, $\tilde{q}(\phi, \psi) = \tilde{q}(\psi, \phi)^*$ and
   (b) linear in the second argument $\tilde{q}(\phi, \psi + \alpha\psi') = \tilde{q}(\phi, \psi) + \alpha\tilde{q}(\phi, \psi')$.
2. q is *bounded*: $|\tilde{q}(\phi, \psi)| \leq c\|\phi\|\|\psi\|$ with $c < \infty$.

Skew-symmetry of $\tilde{q}$ follows directly from (A.3). However, linearity of q in the second factor requires some effort. One can proceed by using the parallelogram law and the definition (A.3) to prove succesively that: 1) $\tilde{q}(\phi, 2\psi) = 2\tilde{q}(\phi, \psi)$, 2) $\tilde{q}(\phi, \psi + \psi') = \frac{1}{2}\tilde{q}(2\phi, \psi + \psi') = \tilde{q}(\phi, \psi) + \tilde{q}(\phi, \psi')$,3) $\tilde{q}(\phi, \pm i\psi) = \pm i\tilde{q}(\phi, \psi)$, and 4) $\tilde{q}(\phi, \alpha\psi) = \alpha\tilde{q}(\phi, \psi)$ for all dyadic rationals in $\mathbb{C}$. To complete the proof we need to verify boundedness and use this to approximate scaling for $\alpha \in bbC$ by scaling for dyadic rationals. From the definition (A.3) and boundedness of q, one obtains easily $|\tilde{q}(\phi, \psi)| \leq c(\|\phi\|^2 + \|\psi\|^2)$. Let $\alpha > 0$ be a dyadic rational, then we have

$$|\tilde{q}(\phi, \psi)| = \frac{1}{\alpha}|\tilde{q}(\phi, \alpha\psi)| \leq \frac{1}{\alpha}\|\phi\|^2 + \alpha\|\psi\|^2 .$$

Choosing $\frac{\|\phi\|}{\|\psi\|} \leq \alpha \leq 2\frac{\|\phi\|}{\|\psi\|}$, we obtain boundedness of $\tilde{q}$:

$$|\tilde{q}(\phi, \psi)| \leq \frac{3}{2}c\|\phi\|\|\psi\| .$$

Finally we may use the additivity and the boundedness of $\tilde{q}$ to obtain scaling $\tilde{q}(\phi, \alpha\psi) = \alpha\tilde{q}(\phi, \psi)$ for arbitrary $\alpha \in \mathbb{C}$ by approximating $\tilde{q}(\phi, \alpha\psi)$ with a sequence $\tilde{q}(\phi, \alpha_n\psi)$ with α_n dyadic rationals.

Conversely, if $\tilde{q}$ is a sesquilinear form then $q(\psi) = \tilde{q}(\psi, \psi)$ is easily seen to be a quadratic form.

1. $q(\phi, \psi) = q(\psi, \phi)^*$ for all $\phi, \psi \in \mathcal{H}$ (skew-symmetry), and
2. $q(\phi, \psi + \alpha\psi') = q(\phi, \psi) + \alpha q(\phi, \psi')$ for all $\phi, \psi, \psi' \in \mathcal{H}$ and $\alpha \in \mathbb{C}$ (linearity).

Note that it follows that $\phi \mapsto q(\phi, \psi)$ is skew-linear. The quadratic form is *bounded* if $|q(\phi, \psi)| \leq C\|\phi\|\|\psi\|$ for some $C < \infty$.

Corollary A.2 (Bounded Quadratic Forms are Operators)**.** *Let $q : \mathcal{H} \to \mathbb{C}$ be a bounded quadratic form. Then there is a unique self-adjoint $A \in \mathcal{B}(\mathcal{H})$ such that $q(\phi) = \langle\phi| A |\phi\rangle$.*

Proof. Let $\tilde{q}$ be the associated sesqui-linear form. For each $\phi \in \mathcal{H}$, define $A\phi$ by the Riesz representation theorem A.1 to be the unique element of $\mathcal{H}$ such that

$$\langle A\phi|\psi\rangle = \tilde{q}(\phi, \psi) .$$

It follows from the skew-symmetry of the quadratic form and the inner product that

$$\langle\psi|A\phi\rangle = \tilde{q}(\psi, \phi) .$$

From this identity, one sees that A is a linear map. Finally, A is bounded since

$$\|A\phi\| = \sup_{\|\psi\|\leq 1} |\langle\psi|A\phi\rangle| = \sup_{\|\psi\|\leq 1} \tilde{q}(\psi, \phi) \leq c\|\phi\| ,$$

since $\tilde{q}$ is bounded.

□

Bibliography

Texts

Dirac, P. A. M. *The Principles of Quantum Mechanics*. 4. ed. (rev.), repr. International Series of Monographs on Physics 27. Oxford: Clarendon Press, Oxford University Press, 2010. 314 pp.

Faddeev, L. D. and O. A. Yakubovskiĭ. *Lectures on Quantum Mechanics for Mathematics Students*. English ed. Student Mathematical Library v. 47. Providence, R.I: American Mathematical Society, 2009. 234 pp.

Hall, B. C. *Quantum Theory for Mathematicians*. Graduate Texts in Mathematics 267. New York: Springer, 2013. 554 pp.

Kitaev, A., A. Shen, and M. Vyalyi. *Classical and Quantum Computation*. Vol. 47. Graduate Studies in Mathematics. Providence, Rhode Island: American Mathematical Society, 31, 2002.

Nielsen, M. A. and I. L. Chuang. *Quantum Computation and Quantum Information*. 10th anniversary ed. Cambridge ; New York: Cambridge University Press, 2010. 676 pp.

Preskill, J. *Physics 219 Course Information*. URL: `http://theory.caltech.edu/~preskill/ph229/` (visited on 01/08/2022).

Reed, M. and B. Simon. *Methods of Modern Mathematical Physics I: Functional Analysis*. San Diego, CA: Academic Press, 1980. 400 pp.

Simon, B. *Operator Theory*. Providence, Rhode Island: American Mathematical Society, 19, 2015.

— *Real Analysis*. Providence, Rhode Island: American Mathematical Society, 4, 2015.

Teschl, G. *Mathematical Methods in Quantum Mechanics: With Applications to Schrödinger Operators*. Vol. 99. American Mathematical Society, 2009.

Watrous, J. *The Theory of Quantum Information*. Cambridge, United Kingdom: Cambridge University Press, 2018. 590 pp.

Other References

Araki, H. and E. H. Lieb. "Entropy Inequalities". *Commun.Math. Phys.* 18.2 (1, 1970), pp. 160–170.

Bell, J. S. "On the Einstein Podolsky Rosen Paradox". *Physics Physique Fizika* 1.3 (1, 1964), pp. 195–200.

Bethe, H. A. "The Electromagnetic Shift of Energy Levels". *Phys. Rev.* 72.4 (15, 1947), pp. 339–341.
Carmichael, H. J. "Quantum Jumps Revisited: An Overview of Quantum Trajectory Theory". *Quantum Future Volta Como Present Beyond*. Ed. by P. Blanchard and A. Jadczyk. Berlin, Heidelberg: Springer, 1999, pp. 15–36.
Carmichael, H. *An Open Systems Approach to Quantum Optics: Lectures Presented at the Université Libre de Bruxelles October 28 to November 4, 1991*. Red. by H. Araki et al. Vol. 18. Lecture Notes in Physics Monographs. Berlin, Heidelberg: Springer Berlin Heidelberg, 1993.
Clauser, J. F. et al. "Proposed Experiment to Test Local Hidden-Variable Theories". *Phys. Rev. Lett.* 23.15 (13, 1969), pp. 880–884.
Davisson, C. J. and L. H. Germer. "Reflection of Electrons by a Crystal of Nickel". *Proceedings of the National Academy of Sciences* 14.4 (1, 1928), pp. 317–322.
Einstein, A. "Über Einen Die Erzeugung Und Verwandlung Des Lichtes Betreffenden Heuristischen Gesichtspunkt". *Ann. Phys.* 322.6 (1905), pp. 132–148.
Einstein, A., B. Podolsky, and N. Rosen. "Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?" *Phys. Rev.* 47.10 (15, 1935), pp. 777–780.
Freedman, S. J. and J. F. Clauser. "Experimental Test of Local Hidden-Variable Theories". *Phys. Rev. Lett.* 28.14 (3, 1972), pp. 938–941.
Glauber, R. J. "Coherent and Incoherent States of the Radiation Field". *Phys. Rev.* 131.6 (15, 1963), pp. 2766–2788.
Heisenberg, W. "Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik". *Z. Physik* 43.3-4 (1927), pp. 172–198.
Kadison, R. and J. Ringrose. *Fundamentals of the Theory of Operator Algebras. Volume I*. Vol. 15. Graduate Studies in Mathematics. Providence, Rhode Island: American Mathematical Society, 22, 1997.
Klauder, J. R. "The Action Option and a Feynman Quantization of Spinor Fields in Terms of Ordinary C-Numbers". *Annals of Physics* 11.2 (1, 1960), pp. 123–168.
Klein, O. "Zur quantenmechanischen Begründung des zweiten Hauptsatzes der Wärmelehre". *Z. Physik* 72.11 (1, 1931), pp. 767–775.
Kraus, K. "General State Changes in Quantum Theory". *Ann. Phys.* 64.2 (1, 1971), pp. 311–335.
Lamb, W. E. and R. C. Retherford. "Fine Structure of the Hydrogen Atom by a Microwave Method". *Phys. Rev.* 72.3 (1, 1947), pp. 241–243.
Lindblad, G. "On the Generators of Quantum Dynamical Semigroups". *Commun.Math. Phys.* 48.2 (1, 1976), pp. 119–130.
Manzano, D. "A Short Introduction to the Lindblad Master Equation". *AIP Adv.* 10.2 (2020), p. 025106.
Millikan, R. A. "A Direct Determination of "h."". *Phys. Rev.* 4.1 (1, 1914), pp. 73–75.
Minev, Z. K. et al. "To Catch and Reverse a Quantum Jump Mid-Flight". *Nature* 570.7760 (7760 2019), pp. 200–204.
Planck, M. "Ueber das Gesetz der Energieverteilung im Normalspectrum". *Ann. Phys.* 309.3 (1901), pp. 553–563.
Rauch, D. et al. "Cosmic Bell Test Using Random Measurement Settings from High-Redshift Quasars". *Phys. Rev. Lett.* 121.8 (20, 2018), p. 080403.

Russo, B. and H. A. Dye. “A Note on Unitary Operators in C*-Algebras”. *Duke Math. J.* 33.2 (1966), pp. 413–416.
Shannon, C. E. “A Mathematical Theory of Communication”. *Bell Syst. Tech. J.* 27.3 (1948), pp. 379–423.
Stone, M. H. “Linear Transformations in Hilbert Space: III. Operational Methods and Group Theory”. *PNAS* 16.2 (1, 1930), pp. 172–175. pmid: 16587545.
— “On One-Parameter Unitary Groups in Hilbert Space”. *The Annals of Mathematics* 33.3 (1932), p. 643. JSTOR: 1968538.
Størmer, E. “Positive Linear Maps of Cu*-Algebras”. *Foundations of Quantum Mechanics and Ordered Linear Spaces: Advanced Study Institute Marburg 1973*. Ed. by A. Hartkämper and H. Neumann. Lecture Notes in Physics. Berlin, Heidelberg: Springer, 1974, pp. 85–106.
Von Neumann, J. “Die Eindeutigkeit der Schrödingerschen Operatoren”. *Math. Ann.* 104.1 (1, 1931), pp. 570–578.
— *Mathematische Grundlagen der Quantenmechanik*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1996.
Weyl, H. “Quantenmechanik und Gruppentheorie”. *Z. Physik* 46.1 (1, 1927), pp. 1–46.