

Some Physics for Mathematicians: Mathematics 7120, Spring 2011

Len Gross

with the assistance of

Mihai Bailesteianu, Cristina Benea, Joe Chen, Nate Eldridge,
Chi-Kwong (Alex) Fok, Igors Gorbovickis, Amanda Hood,
John Hubbard, Yasemin Kara, Tom Kern, Janna Lierl,
Yao Liu, Shisen Luo, Peter Luthy, Justin Moore,
Maicol Ochoa Daza, Milena Pabiniak, Eyvindar Palsson,
Ben Steinhurst, Baris Ugurcan, Hung Tran, Jim West,
Daniel Wong, Tianyi Zheng,...

June 23, 2011

These notes are based on a previous incarnation of this seminar:

Physics for Mathematicians:

Mathematics 712, Spring 2003

by Len Gross

with the assistance of

Treven Wall, Roland Roeder, Pavel Gyra,
Dmitriy Leykekhman, Will Gryc, Fernando Schwartz,
Todd Kemp, Nelia Charalambous, ...
and the participation of
Kai-Uwe Bux, Martin Dindos, Peter Kahn, Bent Orsted,
Melanie Pivarski, Josè Ramirez, Reyer Sjamaar, Sergey Slavnov,
Brian Smith, Aaron Solo, Mike Stillman, ...

Contents

1	Introduction	5
2	Newtonian Mechanics	7
2.1	Work, energy and momentum for one particle	7
2.2	Work, energy and momentum for N particles	10
2.3	Angular Momentum	12
2.4	Rigid bodies and $\text{SO}(3)$	14
2.4.1	Angular velocity	15
2.4.2	Moment of Inertia and Angular Momentum	16
2.5	Configuration spaces and the Newtonian flow	18
2.6	Lagrangian Mechanics	21
2.6.1	Linear systems	21
2.6.2	General configuration spaces. (Nonlinear systems.) . .	24
2.7	Hamiltonian mechanics	26
2.7.1	The Legendre transform	28
2.7.2	From Lagrange to Hamilton via the Legendre transform	30
2.8	SUMMARY	31
2.9	References	31
2.10	Solutions to problems	32
3	Electricity and Magnetism	34
3.1	Lodestones and amber from antiquity till 1600	34
3.2	Production, transfer and storage of electrostatic charge. . . .	36
3.2.1	Production of electrostatic charge, 1704-1713. Hauksbee	36
3.2.2	Transfer of electrostatic charge, 1731. Grey and Dufay	37
3.2.3	Storage of electrostatic charge, 1746. Musschenbroek of Leyden.	37
3.3	Quantitative measurement begins: COULOMB	42
3.3.1	How Coulomb did it.	42
3.3.2	Mathematical consequences	43
3.4	The production of steady currents, 1780-1800	45
3.5	The connection between electricity and magnetism	47
3.5.1	Oersted, Ampère, Biot and Savart	47
3.5.2	FARADAY	51
3.6	MAXWELL puts it all together, 1861	55
3.7	Maxwell's equations ala differential forms	57

3.8	Electromagnetic forces ala Newton, Lagrange and Hamilton	59
3.9	References	63
4	Quantum Mechanics.	64
4.1	Spectroscope	64
4.2	100 years of spectroscopy: radiation measurements that defied Newton ∪ Maxwell	66
4.3	The rules of quantum mechanics	69
4.4	The Heisenberg commutation relations and uncertainty principle	75
4.5	Dynamics and stationary states	79
4.6	Hydrogen	81
4.7	Combined systems, Bosons, Fermions	85
4.8	Observables from group representations	87
4.8.1	Example: Angular momentum	88
4.9	Spin	90
4.10	Pictures: Heisenberg vs Schrödinger	92
4.11	Conceptual status of quantum mechanics	95
4.12	References	96
5	Quantum field theory	97
5.1	The harmonic oscillator	98
5.2	A quantized field; heuristics.	99
5.3	The ground state transformation	102
5.4	Back to the quantized field.	107
5.5	The time dependent field	113
5.6	Many, many particles: Fock space	117
5.6.1	Creation and annihilation operators	118
5.6.2	The canonical commutation relations	122
5.6.3	Occupation number bases	125
5.6.4	Time evolution on \mathcal{F}_b and \mathcal{F}_f	128
5.7	The Particle-Field isomorphism.	130
5.8	Pre-Feynman diagrams	134
6	The electron-positron system	141
6.1	The Dirac equation	141
6.2	Dirac hole theory	145
6.3	Pair production	151

7	The Road to Yang-Mills Fields	152
7.1	Quantization of a particle in an electromagnetic field	153
7.2	Two classic debuts of connections in quantum mechanics	155
7.2.1	The Aharonov-Bohm experiment	156
7.2.2	The Dirac magnetic monopole	157
8	The Standard Model of Elementary Particles	160
8.1	The neutron and non-commutative structure groups. Timeline.	163
9	Appendices	164
9.1	The Legendre transform for convex functions	164
9.1.1	The Legendre transform for second degree polynomials	167
9.2	Poisson's equation	168
9.3	Some matrix groups and their Lie algebras	171
9.3.1	Connection between $SU(2)$ and $SO(3)$	174
9.3.2	The Pauli spin matrices	175
9.4	grad, curl, div and d	176
9.5	Hermite polynomials	178
9.6	Special relativity	180
9.7	Other interesting topics, not yet experimentally confirmed . .	180
9.7.1	The Higgs particle for the standard model	180
9.7.2	Supersymmetry	182
9.7.3	Conformal field theory	182
9.7.4	String theory	182
9.8	How to make your own theory	182
10	Yet More History	182
10.1	Timeline for electricity vs magnetism.	182
10.2	Timeline for radiation measurements	184
10.3	Planck, Einstein and Bohr on radiation. Timeline.	185
10.4	List of elementary particles	188

1 Introduction

There have been a number of good books aimed at mathematicians, describing the mathematical structures that arise in quantum mechanics and quantum field theory. Here are just a few. Most have the words “Quantum” and “Mathematicians” in the title. [21, 28, 3, 46, 57, 58, 59, 73, 74]. Some aim to explain how these mathematical structures build on those that represent an earlier physical theory . Some aim to give a mathematically precise exposition of various topics for mathematicians who want to understand the meaning of terms and ideas developed by physicists. It is no secret that the writing styles of mathematicians and physicists are not conducive to cross-communication.

This seminar is aimed at describing the experiments and observational data that led physicists to make up the successful theories that explained the observed phenomena. Of course we want to describe what these theories are and how they solved some of the experimentally produced problems. To this end we will make use of some of the common mathematical background of the participants (supplemented when necessary), and thereby save time which would otherwise have to be used in a physics course. The listed prerequisites for this seminar were a) highschool physics and b) two years of graduate mathematics courses. This tells you something.

There is really no substitute for going into a laboratory and doing experiments with real objects. But our room is not well equipped for this. Instead I’m going to try to substitute some history of electricity, magnetism and radiation, where possible, and describe the experimental facts that demanded explanation. There is, I think, some physical insight to be gained from understanding who tried to do what, who succeeded and who failed and how and why. As to whether or not a semi-historical exposition is really in the least bit helpful for conveying the notions of physics, even to a mathematically sophisticated audience, remains to be seen. It would be easy and quick just to write down Maxwell’s equations and give the definitions of the electric and magnetic fields. We will, however, repeat in class Faraday’s great experiment of 1831, which put the final touch on the experimental facts that Maxwell used 30 years later to bring the theory of electricity and magnetism into its final form. Fortunately it is possible to do this experiment with rather primitive equipment, which is all that we have.

One failed goal of this seminar was to reach a description, however prim-

itive, of the current status of elementary particle physics. A list of omitted important topics on the straight road to this goal would very likely be longer than the table of contents. Maybe next time.

Here is the progression of topics necessary to understand in order to get at least a little background for current elementary particle theory.

$$[F = ma] \rightarrow \text{Lagrangian Mechanics} \rightarrow \text{Hamiltonian Mechanics} \rightarrow \\ \text{Quantum Mechanics} \rightarrow \text{Quantum Field Theory}$$

Parallel to this one needs to understand two other “classical” theories and their relation.

$$\text{Maxwell's equations} \rightarrow \text{Yang-Mills equations}$$

2 Newtonian Mechanics

2.1 Work, energy and momentum for one particle

The objective of this section is to review the notions of linear momentum, angular momentum, energy and the conservation laws that they satisfy.

Recall Newton's equation:

$$F = ma.$$

We are going to elaborate on the various forms that this equation takes in gradually more sophisticated systems. To begin, consider a single particle moving in space, which we, perhaps understandably, will take to be \mathbb{R}^3 .

Denote the position of the particle at time t by $x(t)$. Let m be a strictly positive constant. We will refer to m as the mass of the particle. The acceleration of the particle is by definition d^2x/dt^2 . Newton's equation then reads

$$F = m d^2x(t)/dt^2$$

where F is the force acting on the particle at time t .

In the preceding paragraph several mathematically undefined words have been used, namely “particle”, “time”, “mass” and “force”. These words acquire meaning only from experience with the physical world. I'm going to assume that the reader has some familiarity with their physical meaning and I'll use them without further ado.

The force F is a vector in \mathbb{R}^3 and may depend on time, on the position, $x(t)$, of the particle, on the velocity $dx(t)/dt$ and could, in principle, depend on higher derivatives of the position. But the existence theory for solutions of 1) is greatly affected if F depends on higher derivatives than the first. If the particle is charged and there are electromagnetic forces then the force on the particle can depend on t , on the position of the particle and also on its velocity. But the fundamental notions of mechanics are best understood in the simplest contexts first.

Definition 2.1 The **work** done on a particle by a constant force is

$$\text{Work} = \text{force times distance.}$$

Here “force” refers to the component of the force in the direction of motion.

Example 2.2 If you lift a 10 pound weight up 3 feet then you have done 30 footpounds of work. If you lower the ten pound weight 3 feet then you have done -30 footpounds of work. (The weight has done 30 footpounds of work on you.) If you move the weight 3 feet horizontally you have done no work if you didn't have to overcome some friction.

The extension of this notion of work in case the force is not constant and the path is not straight is given by

Definition 2.3 (Work done by a variable force) The work done by a force in moving a particle along a trajectory $x(\cdot)$ from $t = a$ to $t = b$ is

$$W = \int_a^b F \cdot dx(t).$$

This clearly reduces to “force times distance” if the force in the direction of motion is constant and the trajectory is a straight line.

Example 2.4 If one moves a ten pound weight around a vertical rectangle then the total work done on the weight is zero.

Although the force in Definition 2.3 can depend on the position, velocity of the particle and the time t , we will consider, for a while, only forces that depend on the position of the particle and not on its velocity or explicitly on time. In this case the work done in moving a particle along a curve C is clearly given by a parametrization independent line integral

$$W = \int_C F(x) \cdot dx.$$

Definition 2.5 A force field $F(\cdot)$ defined in an open set $U \subset \mathbb{R}^3$ is called **conservative** if

$$\int_C F(x) \cdot dx = 0$$

for every closed curve in U .

Theorem 2.6 *In a connected region U of \mathbb{R}^3 a force field F is conservative if and only if there is a function $V : U \rightarrow \mathbb{R}$ such that*

$$F = -\text{grad } V \quad \text{in } U. \tag{2.1}$$

Proof. Pick $x_0 \in U$ and define $V(x) = - \int_{x_0}^x F(x) \cdot dx$ along any curve in U joining x_0 to x . V is well defined because the integral is path independent. As you have proved many times before, the identity (2.1) holds for this function.

■

Terminology: V is called the **potential** of the force field F . V is clearly determined by F only up to an additive constant in the connected region U .

Example 2.7 (Gravitational forcefield of the earth in the interior of this room.) Choose coordinates x, y, z aligned with the left front corner of the room. Since the height of the room is rather small compared with the radius of the earth the gravitational force is “pretty much” constant and is downward. Take $V(x, y, z) = mgz$ where g is a constant to be determined by experiment and m is the mass of the particle acted upon. Then $-\text{grad}V = -(0, 0, mg)$, which agrees with the assertion that the gravitational force is “locally” constant and downward.

Example 2.8 (Gravitational force field of the earth at a big distance, r , from the center of the earth.) Measurements suggest that the force on a particle of mass m at distance r from the center is proportional to $1/r^2$ and is pointed toward the center of the earth. The function $V(x, y, z) = mG/r$ gives such a force and is believed to be accurate in the absence of very huge gravitating masses.

Example 2.9 (Harmonic oscillator.) A particle constrained to move on a line is called a harmonic oscillator if it is subject to a force which, choosing the origin properly, always acts toward the origin and is proportional to the distance. I.e., $F(x) = -kx$ for some constant $k > 0$. In this case the function $V(x) = kx^2/2$ is a potential.

Definition 2.10 The **energy** of a particle, at a point x and having velocity v , in a conservative force field is

$$E(x, v) = (1/2)m|v|^2 + V(x) \quad (2.2)$$

Theorem 2.11 (*Conservation of energy.*) *If a particle of mass m moves under the influence of a conservative force $F(x) = -\text{grad } V(x)$ in accordance with Newton’s equations then its energy is constant along orbits.*

Proof.

$$\begin{aligned}
 (d/dt)E(x(t), v(t)) &= m(v(t), dv/dt) + (gradV, dx/dt) \\
 &= (v, ma - F) \\
 &= 0
 \end{aligned}$$

■

Definition 2.12 The **momentum** of a particle of mass m and velocity v is

$$p = mv.$$

Note: In terms of momentum Newton's equations read

$$dp/dt = F.$$

Neat, huh? We'll see better reasons for introducing the notion of momentum later.

2.2 Work, energy and momentum for N particles

Consider N particles of masses m_1, \dots, m_N respectively. Let $x_j(t)$ denote the position of the j th particle at time t . Write $v_j = dx_j/dt$ for the velocity of the j th particle.

Definition 2.13 The **momentum** of the j th particle is

$$p_j = m_j v_j$$

The **total momentum** of the system is

$$P = \sum_{j=1}^N p_j$$

Two body forces. A frequently arising kind of force in such a system is that in which the total force on the j th particle is a sum

$$F_j = \sum_{k \neq j} F_{jk}$$

where F_{jk} is the force exerted on the j th particle by the k th particle, e.g., gravitational forces. These so called “two body forces” are said to obey “action and reaction” if the j th particle “pulls back” on the k th particle by the same force. That is

$$F_{jk}(x_j, x_k) = -F_{kj}(x_k, x_j).$$

Theorem 2.14 (*Conservation of momentum.*) *In a system with two body forces obeying action and reaction the total momentum is conserved during the Newtonian flow.*

Proof.

$$\begin{aligned} dP/dt &= \sum_{j=1}^N dp_j/dt \\ &= \sum_j F_j \quad (\text{by Newton's equations}) \\ &= \sum_{k \neq j} F_{jk} \\ &= 0. \end{aligned}$$

■ {The following theorem is done in greater generality in the section on Lagrangians. So change the following later.}

Theorem 2.15 (*Conservation of energy.*) *Assume that each of the forces F_{jk} is conservative and depends only on the relative position of the two particles. That is,*

$$F_{jk} = -\text{grad}_j V(x_j - x_k)$$

[For simplicity we take V the same for each pair.] Then the total energy of the system,

$$E = \sum_{j=1}^N (1/2)m_j|v_j|^2 + (1/2) \sum_{k \neq j} V(x_j - x_k)$$

is conserved under the Newtonian flow.

Proof.

$$\begin{aligned} dE(t)/dt &= \sum_{j=1}^N m_j \dot{v}_j \cdot v_j + (1/2) \sum_{k \neq j} ((\text{grad } V)(x_j - x_k)) \cdot (v_j - v_k) \\ &= \sum_{j=1}^N F_j \cdot v_j - (1/2) \sum_{k \neq j} F_{jk} \cdot (v_j - v_k) \\ &= 0 \end{aligned}$$

■

END OF DAY 1. 1/25/2011

SUMMARY. We defined momentum and energy and showed that for certain isolated systems momentum and energy are both invariant (conserved) under the Newtonian flow.

Exercise 2.16 Suppose that a cubical piece of wood of mass 5 grams is sitting on a table. From the center of the right side of the cube a strong but weightless rope is attached, which extends horizontally to a frictionless pulley at the right side of the table. The rope descends from the pulley down to a wooden ball of mass 3 grams, to which it is attached at the top. When the horizontal and vertical portions of the rope are taut and the system is stationary, the ball and cube are released. Find the acceleration of the ball as a function of time.

Hint: Although there are no point particles apparent in this problem you may apply the ideas from the preceding section by using your physical intuition and good judgement. (You're welcome.)

2.3 Angular Momentum

We are going to have to review some elementary facts about the rotation group $SO(3)$ because it plays a central role in understanding angular momentum and spin as well as the classification of elementary particles.

Definition 2.17 The **angular momentum about the origin**, of a particle at x moving with momentum p is

$$L = x \times p$$

The **torque about the origin** of a force F acting on the particle is

$$N = x \times F$$

Lemma 2.18 (*Newton's Equations for angular momentum*)

$$dL/dt = N \quad [\text{Note the similarity to } dp/dt = F.]$$

Proof.

$$\begin{aligned} dL/dt &= dx/dt \times mv + x \times dp/dt \\ &= 0 + x \times F \\ &= N \end{aligned}$$

■

Theorem 2.19 *The total angular momentum of a system of N particles is conserved under the Newtonian flow if the forces are 2-body forces that obey “action and reaction” and in addition act along the line joining the pair of particles.*

Proof. Case n =2:

$$\begin{aligned} dL/dt &= (d/dt)(x_1 \times p_1 + x_2 \times p_2) \\ &= x_1 \times dp_1/dt + x_2 \times dp_2/dt \quad (\text{because } v_j \times p_j = 0) \\ &= x_1 \times F_1 + x_2 \times F_2. \end{aligned}$$

But by assumption $F_2 = -F_1$ and also these forces act along the line joining x_1 and x_2 . Hence

$$dL/dt = (x_1 - x_2) \times F_1 = 0$$

Case n =n.

$$\begin{aligned} dL/dt &= \sum_{j=1}^n (d/dt)(x_j \times p_j) \\ &= \sum_{j=1}^n (v_j \times p_j + x_j \times dp_j/dt) \\ &= \sum_{j=1}^n (x_j \times F_j) \\ &= \sum_{j=1}^n (x_j \times \sum_{k \neq j} F_{jk}) \\ &= \sum_{\substack{\text{unordered pairs}}} (n=2 \text{ case}) \\ &= 0. \end{aligned}$$

■

2.4 Rigid bodies and $\text{SO}(3)$

Definition 2.20 A rigid body is a discrete or continuous system of particles such that the distance between any two particles can not vary with time. [This definition is taken from Goldstein first edition, page 10.]

The discrete or continuous distribution of particles can be represented by a finite measure μ on \mathbb{R}^3 , which we will take to have compact support, and which should be interpreted as the mass distribution of the body. The only motions that a rigid body can make are

- a. a translation. This translates the measure μ .
- b. a rotation. This rotates the measure μ .
- c. a composition of these two (i.e. a proper Euclidean motion.)
- d. a continuously changing family of proper Euclidean motions.

Facts about the rotation group $SO(3)$.

1. $SO(3)$ is homeomorphic to $B_\pi := \{v \in \mathbb{R}^3 : |v| \leq \pi\}$ with opposite points on the surface identified. “Proof:” Any rotation is a rotation around some axis v . Use the right hand rule to determine the direction of the rotation and choose $|v|$ to be the amount of right-handed rotation. (If right thumb points along v then the rotation is in the direction of one’s fingers.) Then make the obvious identifications.
2. The Lie algebra of $SO(3)$ is $so(3) := \{3 \times 3 \text{ skew symmetric real matrices}\}$.
3. $so(3)$ is isomorphic as a Lie algebra to \mathbb{R}^3 in the cross product.
4. $\pi_1(SO(3)) = \mathbb{Z}_2$.
5. $SU(2)$ is the covering group of $SO(3)$ by the covering map (spell this out).

To be precise, let E^3 denote the proper Euclidean group of \mathbb{R}^3 . Thus a map $A : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ is in E^3 if its of the form $Ax = Lx + a$ where $L \in SO(3)$ and $a \in \mathbb{R}^3$. (So $E^3 = SO(3)$ semidirect product with \mathbb{R}^3 .)

Let $g(t)$ be a smooth curve in the group E^3 . Then the one parameter family of measures $g(t)_*\mu$ is the typical allowed motion of the rigid body. In the special case that one point in the body is fixed in space, say at the origin, then translation of this point cannot occur. In this case $g(t)$ lies in the subgroup $SO(3)$ for all t .

What form do Newton's equations take for this system? For a rigid body the analog of linear velocity is angular velocity. We need to digress for a moment to discuss angular velocity.

END OF DAY 2 1/27/2011

2.4.1 Angular velocity

Define $A_\omega x = \omega \times x$ for ω and x in R^3 .

Recall the famous identities $a \times (b \times c) = b(a \cdot c) - c(a \cdot b)$ and $a \cdot (b \times c) = (a \times b) \cdot c$.

Theorem 2.21

$$A_\omega^* = -A_\omega. \quad (2.3)$$

and

$$A_{u \times v} = [A_u, A_v] \quad (2.4)$$

Proof. $(A_\omega x, y) = (\omega \times x) \cdot y = -(x \times \omega) \cdot y = -x \cdot (\omega \times y) = -(x, A_\omega y)$. To prove the second equality observe that

$$\begin{aligned} u \times (v \times x) &= v(u \cdot x) - x(u \cdot v) \\ v \times (u \times x) &= u(v \cdot x) - x(u \cdot v) \\ (u \times v) \times x &= v(x \cdot u) - u(x \cdot v) \\ &= u \times (v \times x) - v \times (u \times x) \end{aligned}$$

■

Now choose coordinate axes such that $\omega = (0, 0, a)$ with $a > 0$. If $x = (x_1, x_2, x_3)$ then $\omega \times x = (-ax_2, ax_1, 0)$. Hence

$$A_\omega = \begin{pmatrix} 0 & -a & 0 \\ a & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (2.5)$$

and therefore

$$\exp(tA_\omega) = \begin{pmatrix} \cos ta & -\sin ta & 0 \\ \sin ta & \cos ta & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (2.6)$$

So $\exp(tA_\omega)$ is a rotation in the x, y plane by an amount $\theta = \theta(t) = ta$. Hence $d\theta/dt = a$. Thus the one parameter group of rotations $\exp(tA_\omega)$ rotates R^3 around the axis ω with speed equal to $|\omega|$. The direction of rotation is determined from ω by the right hand rule. (If right thumb points along ω then the rotation is in the direction of one's fingers.) A_ω is the infinitesimal generator of this one parameter group of rotations and ω is called the ANGULAR VELOCITY of this motion if t is time. The previous theorem shows that the map $\omega \rightarrow A_\omega$ is a Lie algebra homomorphism of R^3 (in the cross product) to $so(3)$. It is clearly an isomorphism because dimension $so(3) = 3$. The map $\omega \rightarrow \exp(A_\omega)$ maps $B_\pi \equiv \{\omega : |\omega| \leq \pi\}$ onto $SO(3)$ because every rotation of R^3 is a rotation around some unique axis by at most π radians. But if $|\omega| = \pi$ then $\exp(A_{\pm\omega})$ are the same rotation. So this map is a diffeomorphism from B_π with antipodal points identified onto $SO(3)$.

The rotational motion of a point $x \in R^3$ under this 1-parameter rotation group is given by $x(t) = \exp(tA_\omega)x$. The velocity of this motion is then

$$v(t) = (d/dt) \exp(tA_\omega)x = A_\omega x(t) = \omega \times x(t).$$

The expression

$$v = \omega \times x$$

is the usual form of this velocity used in the physics literature. It makes the angular velocity ω apparent. Otherwise put, the vector field

$$x \rightarrow \omega \times x$$

is the vector field that generates the one parameter diffeomorphism group $\exp(tA_\omega)$ of R^3 .

2.4.2 Moment of Inertia and Angular Momentum

Definition 2.22 Let μ be the measure representing the mass density of a rigid body. The **moment of inertia** about a point, which we take to be the origin, is the linear transformation $M : R^3 \rightarrow R^3$ defined by

$$M\omega = \int_{R^3} x \times (\omega \times x) d\mu(x)$$

Exercise 2.23

- a. Show that M is a nonnegative symmetric operator.
- b. Show that M is invertible if and only if the body does not lie in a plane through the origin. (I.e. μ is not supported in such a plane.)

Theorem 2.24 *The kinetic energy of a body with one point fixed in space and with angular velocity ω is*

$$T = (1/2)(M\omega, \omega)_{\mathbb{R}^3}$$

Proof. (Take note of the similarity to $E = (1/2)m|v|^2$.) The velocity of a point x in the rigid body is $v = \omega \times x$. The kinetic energy of the body is therefore

$$T = \int_{\mathbb{R}^3} (1/2)|v|^2 d\mu(x) \tag{2.7}$$

$$= (1/2) \int |\omega \times x|^2 d\mu(x) \tag{2.8}$$

$$= (1/2) \int (\omega \times x, \omega \times x) d\mu(x) \tag{2.9}$$

$$= (1/2) \int (x \times (\omega \times x), \omega) d\mu(x) \tag{2.10}$$

$$= (1/2)(M\omega, \omega) \tag{2.11}$$

In the next to the last line I used the fact that the cross product is skew-symmetric in the \mathbb{R}^3 inner product. This is a special case of the fact that $ad A_\omega$ is skew symmetric on any Ad invariant inner product on $so(3)$. ■

Definition 2.25 The **angular momentum** of a rigid body rotating with angular velocity ω about a point in space (which we take to be the origin) is

$$L = M\omega,$$

where M is the moment of inertia about the same point.

Definition 2.26 The **torque** on a rigid body about the origin is the vector sum of the torques acting at each point of the body. (If the body has handles this can be a complicated force to describe.) Since the body is rigid and may

even have very light handles or rods attached to the main mass the torque need not act at the mass points.

$$N = \int_{\mathbb{R}^3} x \times F(dx) \quad (2.12)$$

where F is a vector valued measure on \mathbb{R}^3 , possibly unrelated to the mass density measure μ . But if the force is entirely gravitational then they are clearly related, as in the next Exercise.

Lemma 2.27 *Newton's equations for a rigid body with one point held fixed at the origin are*

$$dL/dt = N$$

Proof. You do this one. ■

Exercise 2.28 Seesaw : A narrow seesaw of length 20 feet is supported at its center. The seesaw is made out of wood of density 3 lbs. per linear foot. Person A sits on the far right end of the seesaw (10 feet from the fulcrum.) Person A weighs 40 lbs. Person B sits 8 feet to the left of the fulcrum. Person B weighs 60 lbs. The right side of the seesaw tilts upward by an angle θ . Assume that $\theta = 0$ at $t = 0$. Find $d^2\theta/dt^2$ at $t = 0$.

Exercise 2.29 Offset wheel on an axle. Compute torque. Write short essay on static versus dynamical balancing. You may consult with classmates, but not with local tire shops.

Exercise 2.30 Use the product rule for differentiation of $SO(3)$ valued functions to explain Coriolis force. Ref. Goldstein, Chapter 5.

2.5 Configuration spaces and the Newtonian flow

So far we have discussed the form that Newton's equations take for some simple mechanical systems; n particles free to move in R^3 under various kinds of forces, and a rigid body. We saw that there are some “constants of the motion”, total energy, total momentum, total angular momentum, whose invariance under the Newtonian flow (under some circumstances on the forces) seems “reasonable” given our developed sense of intuition - 300 years after Newton.

We want to consider now the general structure of more complicated systems in order to force a better understanding of the general structure of Newtonian flows. Our goal is to provide a jumping off point to quantum mechanics.

Definition 2.31 The *configuration space* of a mechanical system is a C^∞ manifold whose points are in one-to-one correspondence with the possible positions (alias configurations) of the system.

This is the kind of definition linking a precise mathematical notion to an intuitive physical notion that can only be understood by examples. Here are some examples.

Example 2.32 (Examples of configuration spaces.)

1. Physical system: a particle free to move in \mathbb{R}^3 under some forces. The configuration space C is \mathbb{R}^3 .
2. Physical system: N particles, each free to move in \mathbb{R}^3 under some forces. $C = \mathbb{R}^{3N}$. (You could remove the coincident points from this product to avoid having two or more particles at the same point if you wish.)
3. Physical system: a particle constrained to move on a sphere of radius five feet (also known as a spherical pendulum). C is a sphere in \mathbb{R}^3 of radius five feet.
4. Physical system: a pendulum on an arm of length 5 feet allowed to move in a plane. C is a circle of radius five feet.
5. Physical system: a double pendulum (the arm of the second pendulum is attached to the mass of the first pendulum.) $C = S^1 \times S^1$. (Each S^1 can have a different radius.)
6. Physical system: a rigid body. $C = E^3$.
7. Physical system: a rigid body with one point fixed in space. $C = SO(3)$.

END of DAY 3 2/1/2011

A curve $t \rightarrow q(t)$ in C is the mathematical object representing the trajectory of the system. Denote by $T(C)$ the tangent bundle. Then the curve $t \rightarrow \dot{q}(t) \in T_{q(t)}(C)$ is a map from \mathbb{R} to $T(C)$. Newton's equations, being second order in $q(t)$, are first order in the pair $q(t), \dot{q}(t)$. For example the

equation of a particle moving in \mathbb{R}^3 in accordance with $m\ddot{x}(t) = F(x)$ can be equivalently described as a solution to the first order system

$$(d/dt)(x(t), v(t)) = (v(t), F(x(t))/m) \quad (2.13)$$

because the first component of this equation forces the identification $v(t) = \dot{x}(t)$. Define a vector field X on $T(\mathbb{R}^3)$ (which can be identified with \mathbb{R}^6 , at some small risk) by

$$X_{(x,v)} = (v, F(x)/m), \quad x \in \mathbb{R}^3, \quad v \in \mathbb{R}^3 = T_x(\mathbb{R}^3). \quad (2.14)$$

Then we may rewrite (2.13) as

$$(d/dt)(x(t), v(t)) = X_{x(t), v(t)} \quad (2.15)$$

As in this example, in general systems the forces in the system, as well as the masses (or e.g. moments of inertia, if some rigid bodies are involved) can be encoded into a vector field X over $T(C)$, after determining what manifold C represents the configuration space of the physical system. Newton's equations then take the form

$$db(t)/dt = X_{b(t)}, \quad (2.16)$$

where $b(t) = (q(t), \dot{q}(t))$ comprises the instantaneous position and velocity of the system. By the “state of the system” one means the pair $q \in C$ and $v \in T_q(C)$. Since the initial position and velocity of the system determines the solution to (2.16), the initial state of the system determines the state of the system for all time. Of course we are assuming here that Equation (2.16) has solutions which exist for all time and are unique. In this case the solution to (2.16) defines a 1-parameter group of diffeomorphisms ϕ_t of $T(C)$ given by

$$(d/dt)\phi_t(a) = X_{\phi_t(a)} \quad (2.17)$$

The group ϕ_t is the Newtonian flow on $T(C)$ determined by the vector field X .

Identifying the vector field X on $T(C)$ tends to be a messy business if the system (i.e. C and forces) is a little bit complicated. Our aim in the next section is to show that the vector field X can be deduced from a knowledge of the kinetic and potential energies, even for a quite general system. Both of these are functions on $T(C)$ and encode the masses and forces.

Example 2.33 (Sphere in \mathbb{R}^3 .) Take C to be the centered sphere in \mathbb{R}^3 of radius R . Let $m > 0$. Define the kinetic energy to be $T_q(v) = (1/2)m|v|^2$ for $v \in T_q(C)$. The norm on v is the Euclidean norm on \mathbb{R}^3 restricted to the tangent space to C at q . Define $V(q, v) = mg q_3$, where g is the gravitational constant in your favorite units and $q = (q_1, q_2, q_3)$. Both T and V are functions on $T(C)$, although V doesn't depend on v . The force specified by the potential V is then $F(q) = -\nabla V = -mg(\partial/\partial q_3)$, which is the constant, downward pointing, gravitational force.

Exercise 2.34 (Harmonic oscillator) Take $C = \mathbb{R}$, $T(v) = (1/2)mv^2$, and $V(x) = 3x^2$. Compute the force and find the solution to Newton's equation, given that the state of the system at time zero is (x_0, v_0) .

2.6 Lagrangian Mechanics

Our goal here is to see how the kinetic and potential energy of a mechanical system determine the Newtonian flow. We really need to do this for a general configuration space so that the transition to Hamiltonian mechanics will be more understandable. But first we are going to practice with linear systems, where we can make computations easily, including an actual integration by parts.

2.6.1 Linear systems

Notation 2.35 Let k be a strictly positive integer, and for each integer $j \in \{1, \dots, k\}$ let m_j be a strictly positive constant. Define a quadratic form $T : \mathbb{R}^k \rightarrow [0, \infty)$ by

$$T(v) = (1/2) \sum_{j=1}^k m_j v_j^2, \quad v = (v_1, \dots, v_k) \tag{2.18}$$

Let $V \in C^2(\mathbb{R}^k)$. We have met special cases of this two function structure in Section 2.2. There we were concerned with the case in which $k = 3N$, N being the number of particles moving in \mathbb{R}^3 , the present constants m_1, m_2, m_3 being all equal to the mass of the first particle, whose velocity is given by the first three components of the k vector v above. And so on. We want to free the computations in this section from the unnecessary notational complexity of that earlier case. To this end we will also write simply $q = (q_1, \dots, q_k)$ for

the Cartesian coordinates in \mathbb{R}^k . If the coordinates of the system of particles is given by the point q then $V(q)$ is to be interpreted as the potential of the system. The j th component of the usual force $F = -\text{grad } V$ (for one particle) may now be written $F_j(q) = -\partial V(q)/\partial q_j$ for the system. To say that the system point $q(t)$ moves in accordance with Newton's equations means

$$m_j \ddot{q}_j(t) = F_j(q(t)) = -(\partial V/\partial q_j)(q(t)) \quad (\text{Newton's equations}) \quad (2.19)$$

The total energy at time t along the orbit is by definition

$$E(t) = T(\dot{q}(t)) + V(q(t)) \quad (2.20)$$

Here we have written $\dot{q}(t)$ for the velocity $v(t)$ of the system in the kinetic energy term.

Conservation of energy in the Newtonian flow holds in this seemingly more general context than that of Section 2.2, but has an equally easy proof:

Theorem 2.36

$$dE(t)/dt = 0 \quad (2.21)$$

if (2.19) holds.

Proof. In view of (2.18) we see that

$$\begin{aligned} \frac{dE(t)}{dt} &= \sum_{j=1}^k m_j \dot{q}_j(t) \ddot{q}_j(t) + (d/dt)V(q(t)) \\ &= \sum_{j=1}^k \dot{q}_j \{ m_j \ddot{q}_j + (\partial V/\partial q_j)(t) \} \\ &= 0 \end{aligned}$$

■ Be aware, however, that conservation of momentum does not hold unless the forces have some kind of special structure, such as two body forces obeying action and reaction, discussed in Section 2.2.

Definition 2.37 The *Lagrangian* of the system specified in (2.18) is

$$L(q, v) = T(v) - V(q). \quad (2.22)$$

It would be well to keep in mind that L is a function of two variables, q and v , each of which varies over \mathbb{R}^k . In keeping with the usual conventions (namely $f_x(x, y)$ means the partial derivative of f with respect to x , evaluated at (x, y)) we will write $(L_q)(q, v)$ and $(L_v)(q, v)$ for the two partial directional derivatives. If the q and v variable are linked then the usual chain rule applies. For example if $q(t)$ is a curve in \mathbb{R}^k and $v(t) = \dot{q}(t)$ is its velocity then a variation of the curve $q(\cdot)$ by a small variation, giving the nearby curve $q(t) + sh(t)$, will also vary the velocity: $v(t) \mapsto (d/dt)(q(t) + sh(t)) = v(t) + sh'(t)$. Consequently, by the chain rule, at $s = 0$,

$$\begin{aligned} & (d/ds)L((q + sh)(t), (d/dt)(q + sh(t))) \\ &= (L_q)(q(t), v(t))\langle h(t) \rangle + (L_v)(q(t), v(t))\langle \dot{h}(t) \rangle \end{aligned} \quad (2.23)$$

$$= - \sum_{j=1}^k V_j(q(t))h_j(t) + \sum_{j=1}^k m_j \dot{q}_j(t) \dot{h}_j(t) \quad (2.24)$$

for each t , where $h(t) = (h_1(t), \dots, h_k(t))$.

Theorem 2.38 *Let $t_0 < t_1$ and let a_0 and a_1 be two points in \mathbb{R}^k . Denote by \mathcal{P} the set of C^2 paths in \mathbb{R}^k over $[t_0, t_1]$ joining a_0 to a_1 . Let $Y = \{h \in C^2([t_0, t_1], \mathbb{R}^k) : h(t_0) = h(t_1) = 0\}$. Then Y is a linear space (in fact a Banach space in the C^2 norm) and if $q(\cdot) \in \mathcal{P}$ and $h \in Y$ then $t \mapsto q(t) + sh(t)$ is in \mathcal{P} for all real s . Suppose that $q(\cdot) \in \mathcal{P}$. Then the following are equivalent.*

1) *Newton's equations:*

$$m_j \ddot{q}_j(t) = -(\partial V / \partial q_j)(q(t)) \quad (2.25)$$

2) *Lagrange's equations:*

$$(d/dt)[(L_v)(q(t), \dot{q}(t))] - L_q(q(t)) = 0 \quad \forall t \in [t_0, t_1] \quad (2.26)$$

3) *Least action principle:*

The "action" functional $A : \mathcal{P} \rightarrow \mathbb{R}$ given by

$$A(y) = \int_{t_0}^{t_1} L(y(t), \dot{y}(t)) dt \quad (2.27)$$

has a critical point at $y(\cdot) = q(\cdot)$

(N.B. Both terms in (2.26) are linear functionals on R^{2k} .)

Proof. Inserting $y(t) = q(t) + sh(t)$ into $f(y)$ and using (2.23) we find

$$\begin{aligned}
& (d/ds)|_{s=0} A(q(t) + sh(t)) \\
&= \int_{t_0}^{t_1} \left\{ (L_q)(q(t), v(t)) \langle h(t) \rangle + (L_v)(q(t), v(t)) \langle \dot{h}(t) \rangle \right\} dt \\
&= \int_{t_0}^{t_1} \left\{ (L_q)(q(t), v(t)) \langle h(t) \rangle - \left((d/dt) \left[(L_v)(q(t), v(t)) \right] \right) \langle h(t) \rangle \right\} dt \\
&= \int_{t_0}^{t_1} \left\{ (L_q)(q(t), v(t)) - (d/dt) \left[(L_v)(q(t), v(t)) \right] \right\} \langle h(t) \rangle dt
\end{aligned} \tag{2.28}$$

To derive the third line we used an integration by parts along with $h(t_0) = h(t_1) = 0$ to get rid of the boundary terms. (If you feel queasy about dealing with integration by parts for such linear functionals rewrite it in coordinates, as in (2.24).)

Thus $q(\cdot)$ is a critical point of $A(\cdot)$ if and only if the integral in (2.28) is zero for all of the allowed functions $h(\cdot)$. Since these constitute quite a hefty set of functions, the integral is zero for all such h if and only if the factor in braces in the integrand is zero, that is, if and only if Lagrange's equation (2.26) holds. Moreover if we put $L = T - V$ in (2.27) and write this equation in coordinates we find

$$\begin{aligned}
0 &= \left\{ (d/dt) \left[(L_v)(q(t), \dot{q}(t)) \right] + V_q(q(t)) \right\} \langle h(t) \rangle \\
&= \sum_{j=1}^k \left\{ m_j \ddot{q}_j(t) + V_j(q(t)) \right\} h_j(t)
\end{aligned}$$

for all allowed functions $h_j(t)$. This is equivalent to Newton's equations (2.25). ■

2.6.2 General configuration spaces. (Nonlinear systems.)

Notation 2.39 Let $L : T(C) \rightarrow \mathbb{R}$ be any smooth function. In addition to the one form dL on $T(C)$ there is another natural one form determined by L . Let $\pi : T(C) \rightarrow C$ be the natural projection. Then $\pi_* : T(T(C)) \rightarrow T(C)$. For each $q \in C$ let $L^q : T_q \rightarrow \mathbb{R}$ be the restriction of L to $T_q(C)$, which is of course contained in $T(C)$. Since $T_q(C)$ is linear we may regard $d(L^q)$ as a

linear functional on $T_q(C)$ for each point v in $T_q(C)$. If $w \in T_{q,v}T(C)$ then $\pi_*w \in T_q(C)$ and so

$$\theta_L(w) \doteq d(L^q)(\pi_*w)$$

is a well defined real number. θ_L is therefore a well defined one form on $T(C)$.
{ Give its expression in local coordinates.}

Theorem 2.40 *Let $L : T(C) \rightarrow \mathbb{R}$ be a smooth function. Let $\pi : T(C) \rightarrow C$ be the natural projection and define the associated 1-form θ_L as in Notation 2.39. Assume that for each point $q \in C$ the function L^q on T_q has no critical points. (I.e. its second derivative matrix in the linear coordinates is invertible at each point $v \in T_q$.) (This is in the spirit of saying that all masses are strictly positive.)*

Then there exists a unique vector field X on $T(C)$ such that

$$\pi_*X_{q,v} = v \quad (2.29)$$

and

$$\mathcal{L}_X\theta_L - dL = 0 \quad (2.30)$$

where \mathcal{L}_X denotes the Lie derivative.

Moreover in any local coordinate system (q_1, \dots, q_k) and induced coordinates (v_1, \dots, v_k) on $T_q(C)$ over the coordinate patch, the flow in $T(C)$ $t \mapsto q(t), \dot{q}(t) \in T_{q(t)}$, induced by the vector field X , satisfies Lagrange's equations

$$(d/dt)[(\partial L/\partial v_i)(q(t), \dot{q}(t))] - (\partial L/\partial q_i)(q(t), \dot{q}(t)) = 0, \quad i = 1, \dots, k \quad (2.31)$$

Proof. This statement is based on a lecture by Roland Roeder (in Math. 712, (2003)) Very likely there is a proof in either [39] or in [43]. ■

Note: A converse of this theorem also holds in the sense that, in the presence of (2.29), the Lagrange equations, (2.31), imply (2.30), when properly formulated.

Significance: We saw in Theorem 2.38 that if $C = \mathbb{R}^k$ and if $L = T - V$ on $T(C)$ then Lagrange's equations, (2.31), are equivalent to Newton's equations. Therefore the vector field X in Theorem 2.40 gives the Newtonian flow on $T(C)$ in this (linear) case.

Example 2.41 (Spherical pendulum again.) We saw in Example 2.33 how easy it was to write down the Lagrangian for the spherical pendulum:

$$L(q, v) = (1/2)m|v|^2 - mg q_3 \quad (2.32)$$

But just try writing down Newton's equations, say in spherical coordinates!

See John Hubbard's exposition of the spherical pendulum via both the Lagrangian and Hamiltonian approaches. It is available on the course website.

2.7 Hamiltonian mechanics

Let $\pi : T^*(C) \rightarrow C$ be the natural projection. So $\pi_* : T(T^*(C)) \rightarrow T(C)$. Let b be a point of $T^*(C)$, We may write $b = (q, p)$ where $q \in C$ and $p \in T_q^*(C)$. Although any point $p \in T^*(C)$ is a point in $T_q^*(C)$ for some unique point q we will nevertheless be redundant and specify q also, as is customary in the physics literature. For any vector $u \in T_b(T^*(C))$ we have $\pi_* u \in T_q(C)$. Hence $\langle p, \pi_* u \rangle$ is well defined and is linear in u . Thus the equation

$$\alpha(u) = \langle p, \pi_* u \rangle \quad (2.33)$$

defines a one-form on $T^*(C)$. Let

$$\omega = d\alpha.$$

Lemma 2.42 *If dimension $C = n$ then $\omega^n \neq 0$.*

Proof. Aside from proving the assertion of the lemma this proof is intended to give some insight into the two-form ω for future use. Choose a local coordinate system q_1, \dots, q_n in an open set U in C . A point p in $T_q^*(C)$ may then be written uniquely in the form $p = \sum_{j=1}^n p_j dq_j$ and consequently $q_1, \dots, q_n, p_1, \dots, p_n$ form a coordinate system in $\pi^{-1}(U)$. The vectors $\partial/\partial q_j, \partial/\partial p_j|_{j=1}^n$ form a basis of $T_b(T^*(C))$. Since $\pi(q_1, \dots, q_n, p_1, \dots, p_n) = q_1, \dots, q_n$ we find

$$\pi_* \partial/\partial p_j = 0$$

and

$$\pi_* \partial/\partial q_j = \partial/\partial q_j.$$

Be aware that $\partial/\partial q_j$ has a different meaning on the two sides of the last equation. It follows that

$$\alpha \left\langle \sum_{j=1}^n (a_j \partial/\partial p_j + v_j \partial/\partial q_j) \right\rangle = \sum_{j=1}^n p_j v_j. \quad (2.34)$$

In other words, in terms of the local coordinates on $\pi^{-1}U$ we have

$$\alpha = \sum_{j=1}^n p_j dq_j. \quad (2.35)$$

Hence

$$\omega = \sum_{j=1}^n dp_j \wedge dq_j \quad (2.36)$$

It now follows that

$$\omega^n = \Pi_1^n(dp_j \wedge dq_j)$$

which is clearly nowhere zero. ■

The fundamental two-form ω sets up an isomorphism between $T_b(T^*C)$ and T_b^* by means of the correspondence

$$T_b^* \ni \beta \rightarrow z \in T_b \text{ if } \beta \langle u \rangle = \omega \langle u, z \rangle \text{ for all } u \in T_b. \quad (2.37)$$

The correspondence is an isomorphism because ω is nondegenerate.

Suppose then that $H : T^*(C) \rightarrow \mathbb{R}$ is a smooth function. Then the correspondence (2.37) determines a vector field $Y \equiv Y_H$ on $T^*(C)$ from the one-form dH . Thus Y is the unique vector field on $T^*(C)$ given by

$$\langle dH, u \rangle = \omega \langle u, Y \rangle \quad \forall u \in T(T^*(C)) \quad (2.38)$$

If ω were an inner product then it would be customary to write $Y = \nabla H$. But in fact ω is a skew symmetric bilinear form. It is customary to write

$$Y = \nabla_\omega H \quad (2.39)$$

The ω gradient of H is therefore defined by the identity

$$\langle dH, u \rangle = \omega \langle u, \nabla_\omega H \rangle \quad \forall u \in T(T^*(C)). \quad (2.40)$$

We intend to show that if one chooses

$$H = T + V \quad (2.41)$$

then the flow in $T^*(C)$ determined by the ω gradient $\nabla_\omega H$ on $T^*(C)$ “is” the Newtonian flow. However, since the kinetic energy T has not yet been defined on $T^*(C)$, and the vector field X_L of the previous section, denoted

simply X there, is also not defined on $T^*(C)$, it behooves us to explain the meaning of the word “is”. To this end we will show that (under suitable conditions) the Lagrangian L itself sets up an isomorphism between $T_q(C)$ and $T_q^*(C)$, which extends to a diffeomorphism of $T(C)$ with $T^*(C)$. This diffeomorphism, in turn, interchanges X_L with $\nabla_\omega H$. In this sense the flow of $\nabla_\omega H$ on $T^*(C)$ is equivalent to the flow of X_L on $T(C)$, which in turn, is equivalent to the Newtonian flow when this statement makes sense, namely $C = \mathbb{R}^k$ (see Theorem 2.38) or $C =$ configuration space for some constrained system in \mathbb{R}^k see spherical pendulum in Examples 2.33 and 2.41.

It seems well worth pointing out that, quite aside from the connections with Lagrangian and Newtonian mechanics, the theory of flows over such a cotangent space $T^*(C)$, as well as geometrical questions and applications to other parts of mathematics is a highly developed subject in itself. In fact any even dimensional manifold carrying a symplectic form, analogous to ω , already has a rich structure, that has been explored by some of your classmates.

2.7.1 The Legendre transform

Let Y be a finite dimensional real vector space. Denote by Y^* its dual space. We want to consider a class of functions on Y which will capture the typical velocity dependence of a Lagrangian $L(q, v)$ at a fixed point q , in the presence of velocity dependent forces. Y should be interpreted as the tangent space to configuration space at the point q .

For a smooth function $f : Y \rightarrow \mathbb{R}$ its derivative $f'(v) \in Y^*$ is defined as usual, for each $v \in Y$, by the prescription

$$f'(v)\langle w \rangle = \frac{d}{dt} f(v + tw)|_{t=0} \quad \text{for } w \in Y. \quad (2.42)$$

We will use brackets $\langle w \rangle$ to emphasize that the argument is linear in w .

Definition 2.43 (Legendre transform) Suppose that $f : Y \rightarrow \mathbb{R}$ is a smooth function and that the map

$$Y \ni v \mapsto f'(v) \in Y^* \quad (2.43)$$

is one-to-one and onto. Define a function

$$f^* : Y^* \rightarrow \mathbb{R} \quad (2.44)$$

by the prescription

$$f^*(p) = \langle p, v \rangle - f(v) \quad p \in Y^*, \quad (2.45)$$

where v is the unique solution to the equation

$$p = f'(v). \quad (2.46)$$

f^* is called the *Legendre transform* of f .

Example 2.44 Take $Y = R^n$ with its standard inner product. Let

$$f(v) = \frac{1}{2}m|v|^2. \quad (2.47)$$

Then $f'(v)\langle w \rangle = (mv, w)$. Upon identifying Y with Y^* via the inner product we may therefore write $f'(v) = mv$. Hence the map $v \mapsto f'(v)$ is one-to-one and onto. Solving the equation $p = f'(v)$ for v in terms of p gives $v = p/m$. Therefore

$$f^*(p) = (p, \frac{p}{m}) - \frac{m}{2}|\frac{p}{m}|^2 = \frac{1}{2m}|p|^2 \quad (2.48)$$

Example 2.45 Take $Y = R^n$ with its standard inner product again. Let A be a vector in Y^* . Define

$$f(v) = \frac{1}{2}m|v|^2 + \langle A, v \rangle \quad (2.49)$$

Then

$$f^*(p) = \frac{1}{2m}|p - A|^2 \quad (2.50)$$

Proof. $f'(v)\langle w \rangle = (mv, w) + \langle A, w \rangle$. Hence

$$f'(v) = mv + A \quad (2.51)$$

where mv again denotes the element of Y^* gotten by identifying Y with Y^* as in the preceding example. Thus, solving the equation $p = f'(v) = mv + A$ we find $v = (p - A)/m$. Therefore

$$\begin{aligned} f^*(p) &= \langle p, v \rangle - f(v) \\ &= \langle p, \frac{p - A}{m} \rangle - \left(\frac{1}{2}m|\frac{p - A}{m}|^2 + \langle A, \frac{p - A}{m} \rangle \right) \\ &= \frac{1}{2m}|p - A|^2 \end{aligned}$$

■

It is this example that is responsible for forcing connections on vector bundles into quantum mechanics in the presence of velocity dependent forces.

Example 2.46 Take $Y = \mathbb{R}$ and define $f(v) = v^2 + v^4$. Then $f'(v)\langle w \rangle = (2v + 4v^3)w$. That is, $f'(v) = 2v + 4v^3$. The range of this function is clearly all of \mathbb{R} . Moreover $f''(v) = 2 + 12v^2 \geq 2$. Hence f' is one-to one and onto. Its Legendre transformation is therefore well defined.

Remark 2.47 As a cultural matter you should be aware that the definition 2.43, which assumes that the map $Y \ni v \mapsto f'(v) \in Y^*$ is one-to-one and onto, can be replaced by a very clean and general definition of the Legendre transform, for convex functions f which need not be differentiable, nor even defined on all of Y , but only on some convex subset of Y . The Legendre transform is then another such function and one has the nice theorem that $f^{**} = f$. Such an extension of our present discussion is very useful in thermodynamics and other parts of mathematics, but is not needed by us for mechanics. A more careful survey of this extension is outlined in Appendix 9.1. Notice by the way that the previous three examples are all convex.

2.7.2 From Lagrange to Hamilton via the Legendre transform

Suppose that $L : T(C) \rightarrow \mathbb{R}$ is a smooth function satisfying the non-degeneracy condition of Theorem 2.40; $v \mapsto L(q, v)$ has no critical points on $T_q(C)$. Although this is all that's needed for the basic theory we will assume more. Namely we assume that for each $q \in C$ the function $v \mapsto L(q, v)$ is quadratic plus linear and furthermore, for each element $q \in C$ the function $T_q \ni v \mapsto L(q, v)$ has a nonsingular quadratic part, as in Section 9.1.1. In all of the examples that we've looked at so far the operator M that appears in Section 9.1.1 is positive definite, since it just comes from a kinetic energy. When we introduce electromagnetic forces later we will have to add on a linear term in v also. We are now ready to define the Hamiltonian function on $T^*(C)$. It is, for each $q \in C$, the Legendre transform of the Lagrangian In the velocity variable. That is, if $L^q = L|T_q(C)$ then

$$H(q, p) = (L^q)^*(p). \quad (2.52)$$

H is a function on $T^*(C)$. This is the function to which we want to apply the theory of Section 2.7.

The relation between the Newtonian, Lagrangian and Hamiltonian approaches to classical mechanics is developed further in John Hubbard's notes. These are now available on the course website. See also Goldstein, "Classical Mechanics" [23] for the standard approach used by physicists

2.8 SUMMARY

<i>Newton</i>	<i>Lagrange</i>	<i>Hamilton</i>
Forces and masses	$L : T(C) \rightarrow \mathbb{R}$	$H : T^*(C) \rightarrow \mathbb{R}$
Second order equ.	First order system	First order system
$\mathbf{F} = m\mathbf{a}$	$\mathcal{L}_X \mathbf{d}(L^q) - dL = \mathbf{0}$	$\omega(\mathbf{Y}, \nabla_{\omega} \mathbf{H}) = (\mathbf{d}\mathbf{H})(\mathbf{Y})$
Local versions:	$\frac{d}{dt} \left[\frac{\partial L}{\partial \dot{q}} \right] - \frac{\partial L}{\partial q} = 0$	$\dot{q}_i = \frac{\partial H}{\partial p_i}, \dot{p}_i = -\frac{\partial H}{\partial q_i}$
Newton's equations	Lagrange's equations	Hamilton's equations

Table 1: Newton to Lagrange to Hamilton.

We saw that all the Lagrangians $L = T - V$ on $T(C)$ in our examples have the property that the restriction, L_q , of L to T_q is a positive quadratic function (the kinetic energy) plus a constant ($-V(q)$) and is therefore smooth and strictly convex on the finite dimensional linear space $T_q(C)$. Later we will add on a linear term to incorporate electromagnetic forces. The Lagrangian therefore sets up a diffeomorphism (actually an affine map) ϕ_q between T_q and T_q^* for each $q \in C$, in accordance with Section 9.1.1. Putting all these fiber maps together yields a diffeomorphism between $T(C)$ and $T^*(C)$. For each point $q \in C$ the Hamiltonian function $H(q, \cdot)$ is the conjugate function to L^q , as defined in (2.52).

See also John Hubbard's notes on these topics. These are now available on the course website.

We owe a debt of gratitude to Roland Roeder, who gave the initial lectures on the Lagrangian and Hamiltonian approaches in the 2003 manifestation of this seminar.

END of DAY 6, 2/10/2011

2.9 References

References:

V. I. Arnold, *Mathematical methods of classical mechanics* [4]

Cushman and Bates, *Global aspects of classical integrable systems*.[14]

Gerald Folland, *Quantum field theory, A tourist guide for mathematicians*.[21]

This book contains a chapter on classical mechanics, aimed at preparing the reader for quantum mechanics.

Goldstein, *Classical mechanics*. [23]

Brian Hall, *An introduction to quantum theory for mathematicians*. [28].

This book contains a chapter on classical mechanics, aimed at preparing the reader for quantum mechanics. An updated version of this book can be downloaded from the 7120 website.

José and Saletan, *Classical dynamics, A contemporary approach*. [39]

Marsden and Ratiu, *Introduction to mechanics and symmetry*, A basic exposition of classical mechanical systems. [43]

2.10 Solutions to problems

Exercise 2.16 . Solution was presented by Peter Luthy.

Exercise 2.23 Solution presented by Alex Fok

1. Show that M is a nonnegative symmetric operator.
2. Show that M is invertible iff the body does not lie in a line through the origin. (i.e. μ is not supported in such a line.)

Solution:

- (a) It suffices to show that $(M\omega, \nu) = (M\nu, \omega)$ and $(M\omega, \omega) \geq 0$ for any $\omega \in \mathbb{R}^3$. Note that

$$\begin{aligned}(M\omega, \nu) &= \int_{\mathbb{R}^3} (x \times (\omega \times x), \nu) d\mu(x) \\&= \int_{\mathbb{R}^3} (\nu \times x, \omega \times x) d\mu(x) \\&= \int_{\mathbb{R}^3} (x \times (\nu \times x), \omega) d\mu(x) \\&= (M\nu, \omega) \\(M\omega, \omega) &= \int_{\mathbb{R}^3} \|\omega \times x\|^2 d\mu(x) \\&\geq 0\end{aligned}$$

Thus M is a nonnegative symmetric operator.

- (b) Since M is a symmetric operator, M is not invertible iff there exists $\omega_0 \neq 0$ such that $(M\omega_0, \omega_0) = 0$. This is equivalent to saying that

$$\int_{\mathbb{R}^3} \|\omega_0 \times x\|^2 d\mu(x) = 0,$$

which is equivalent to saying that $\omega_0 \times x = 0$ for all $x \in \text{supp}(\mu)$ because of continuity in x of the integrand. But

$$\begin{aligned} & \omega_0 \times x = 0 \text{ for all } x \in \text{supp}(\mu) \\ \iff & x \in \mathbb{R}\omega_0 \text{ for all } x \in \text{supp}(\mu) \\ \iff & \text{supp}(\mu) \subset \mathbb{R}\omega_0 \end{aligned}$$

3 Electricity and Magnetism

3.1 Lodestones and amber from antiquity till 1600

History: Ref. Duane Roller “The de Magnete of William Gilbert” (1959)

Lodestones were known at least as far back as 900 BC (Homer mentioned them) Aristotle, Plato, Pliny and the usual gang made reference to them in passing, in a way that made it clear that the reader was expected to be familiar with them. There were various theories about how they worked.

One of the early theories regarded the material attracted by the amber as “food for amber”.

But the two most long lasting theories were the following.

Theory Number 1. A “sympathy” exists between the lodestone and the iron that it attracts.

Theory Number 2. The lodestone emits some stuff that removes air from around it. A piece of iron nearby will then fall into the empty space.

The first theory is anthropomorphic while the second theory in effect rejects action at a distance.

Some highlights:

1. **Lucretius** [c. 60 BC] describes Theory Number 2 for lodestones thus. It is from his poem “On the nature of things”. [35, p464]

First, from this stone there must stream a shoal of seeds in a current
Driving away with its blows all the air 'twixt itself and the iron.
Soon as this space is made void and a vacuum fashioned between them,
Instantly into that void the atoms of iron tumble headlong
All in a lump; thus the ring itself moves bodily forward.

2. **Plutarch** [c. 50 AD] Here is Plutarch’s version of Theory Number 2 for amber. [35, p465]

In amber there is a flammuous and spirituous nature,
and this by rubbing on the surface is emitted by hidden passages,
and does the same that lodestone does.

3. Not much progress in understanding or using lodestones till c. 1200 when compasses were invented in China and the technology gradually spread to Europe. However there have been many medical and social applications

found for lodestone. One needed only to tie a piece of lodestone onto a diseased part of the body. This was particularly advised for curing gout and epilepsy. Here are some other known recommendations.

4. **Marbode, (11th century).** To determine whether a wife is chaste or unchaste apply a lodestone to her head. If she is unchaste she will fall out of bed. [50, p27]

5. **St. Hildegard of Bingen, (12th century).** To cure insanity just tie on a lodestone and sing a suitable incantation. [50, p28]

6. **Unidentified source, Ithaca, NY, (21st century).**

- a. To cure carpal tunnel syndrome sleep with a magnet bound to your wrist.
- b. To cure back pains sleep on a pad filled with magnets.

1600 AD. George Gilbert published the first organized hard data about lodestone-like and amber-like objects in 1600. He emphasized that magnetic phenomena and electric phenomena were different. He presented hard experimental data for both effects. The data was *not quantitative*. But he listed many more amber-like materials than were previously known (which he named “electrics”). He had invented a very sensitive detection device - a versorium (a needle carefully balanced at its center and allowed to rotate under the influence of weak forces) He had his own theory to explain these phenomena the “effluvia theory”. (Effluvia means emanations.) This was a modification of Theory Number 2 above. Here is Gilbert’s own theory presented in his own words.[35, p468]

The effluvia spread in all directions: they are specific and peculiar, and, sui generis, different from the common air; generated from humor; called forth by calorific motion and rubbing, and attenuation; they are as it were material rods - hold and take up straws, chaff, twiggs, till their force is spent or vanishes; and then these small bodies, being set free again, are attracted by the earth itself and fall to the ground.

Some other events before electricity and magnetism really got off the ground:

1632-1642: Galileo is kept under house arrest.

1687: Newton publishes “Principia”

END of DAY 7 = 2/15/2011

3.2 Production, transfer and storage of electrostatic charge.

In order to do experiments with electricity, whatever that is, its real handy to have a goodly supply of the stuff under your control. That means that you have to be able to produce it, store it, and transfer it to a useful place. You know what transfer means - conduction of electricity. The devices invented between 1700 and 1750 for these purposes were not suitable for quantitative measurements. But they paved the way. The inventors of these devices were concerned with the stuff (electrostatic charge) that pulls or pushes little pieces of this or that.

3.2.1 Production of electrostatic charge, 1704-1713. Hauksbee

I'm afraid that the hair combing method that we used in class to produce electrostatic charge, and which we successfully detected with our own electroscope, is not as useful for further experiments as one might hope. Francis Haucksbee (1666-1713) constructed a very efficient rubbing device for making charge. It was a glass globe on a horizontal axis with a crank at one end that you could turn, while holding something in contact with the glass globe, such as some silk, or leather, or your own hand. His most advanced model (version 1710.3) was constructed around 1710. (When you have free time you might try to identify all the objects in Figure 1. He used this to make many observations of how charged, and sometimes uncharged, materials attract or repel. For example uncharged strings, held at one end, would all align themselves radially so as to point toward a charged sphere at the center. He also showed how two such spheres, one charged, produce a mysterious green light, first observed in mercury barometers. It was actually this mysterious light that first turned him on to these investigations. This "electric machine", as he called it, was used by generations of experimenters. See [49] Chapter 3 if you would like to know more about what he said he did and why he did it. He himself attempted to explain his observations in his 45 published papers. But, sadly, his extension of Gilbert's "effluvia" theory did not stand the test of time.



Figure 1: Hauksbee’s electrostatic generator

3.2.2 Transfer of electrostatic charge, 1731. Grey and Dufay

Some materials conduct the “electrical fire” quite readily and some don’t, conductors and nonconductors, as we call them nowadays. It was Grey and Dufay who discovered this in **1731**. Moreover they accumulated enough real data to enable Dufay to propose a reasonable theory of what charge is. It was called the two fluid theory, which captures some of our present theory of positive and negative charge. Of course no such proposal about things you can’t see is likely to be accepted right away by any scientific community. Our own Ben Franklin later proposed a “one fluid” theory. This was before he got involved with politics and revolution. Physicists took sides on the “n-fluid” issue, of course. Coulomb, for example, after surviving the French revolution, argued for the two fluid theory.

3.2.3 Storage of electrostatic charge, 1746. Musschenbroek of Leyden.

The accidental discovery by Musschenbroek of Leyden, in **1746**, of a method of storing large quantities of the “electrical fire” changed the landscape in electrical science.

We tend to think of the development of science in the seventeenth and eighteenth centuries as proceeding at a slow “gentlemanly” pace. The reason for this view, is that the really significant developments, the ones that survived the test of time, were few and far between, with the lesser supporting discoveries not even brought to our attention, in spite of the fact that their discoverers were scrambling to discover and their lesser discoveries were influential on those who finally put the pieces together.

The events leading up to and immediately following the accidental discovery of the Leyden jar are particularly fun to read about because the general public was getting into the swing of frontline research on electric-

ity. University lectures on electricity were attended by the general public, even to the extent that the registered students couldn't find seats. (Paris, I suppose.) Theaters offered kisses from pretty (electrically charged) women swinging from wires on stage, a shocking experience for willing members of the audience.

In the 1750s a home without a charge generator prominently displayed on a coffee table could not claim to be cultured. More on this may be found in Chapter 6 of Roller's book [49] and in [35, Section 26.7].

Within months of Musschenbroek's discovery the news traveled across the Atlantic Ocean to Ben Franklin. He immediately began his own experiments with stored charge. He constructed a version of the Leyden jar which produced very big sparks. By November, **1747**, just in time for a feast, he was able to kill a turkey.

Before you read what Musschenbroek actually did you might like to read the following extract from the popular writings of Henry Smith Williams (1863-1943), describing other effects of the discovery of the Leiden jar on public entertainment. This was written about 1904 in Harper's Magazine. For more by this author see

<http://www.worldwideschool.org/library/books/sci/history/AHistoryofScienceVolumeII/chap49.html>

"The advent of the Leyden jar, which made it possible to produce strong electrical discharges from a small and comparatively simple device, was followed by more spectacular demonstrations of various kinds all over Europe. These exhibitions aroused the interest of the kings and noblemen, so that electricity no longer remained a "plaything of the philosophers" alone, but of kings as well. A favorite demonstration was that of sending the electrical discharge through long lines of soldiers linked together by pieces of wire, the discharge causing them to "spring into the air simultaneously" in a most astonishing manner. A certain monk in Paris prepared a most elaborate series of demonstrations for the amusement of the king, among other things linking together an entire regiment of nine hundred men, causing them to perform simultaneous springs and contortions in a manner most amusing to the royal guests¹. But not all the experiments being made were of a purely spectacular character, although most of them accomplished little except in a negative way. The famous Abbe Nollet, for example, combined useful experiments with spectacular demonstrations, thus keeping up popular interest while aiding the cause of scientific electricity."

¹Perhaps this gives the reader some insight into the origin of the French Revolution.

1746: Musschenbroek of Leyden discovers the Leyden jar.

The accidental discovery of a method for storing (what we now call) charge was a shocking experience for its discoverer, Peter van Musschenbroek of Leyden. For the benefit of readers with a slightly sadistic streak I'm going to excerpt below the relevant portion of Roller's book on the development of the concept of electric charge. [49][p52].

In order to understand what Musschenbroek did it would be good to understand, at an intuitive level, how a condenser (\equiv capacitor) works.

HOW A CAPACITOR WORKS: Envision a 3 volt battery with a wire sticking out of the negative side. (That's the casing.) The battery tries to push electrons out along this wire. But they have no place to go. A little more precisely, a few electrons do get pushed out along the wire, but since the electrons repel each other they will push back until no more electrons can get pushed onto the wire by the battery. (If it were a nine volt battery a few more electrons would be pushed onto the wire.) Similarly a wire attached to the positive end of the battery (that's the little dimple at the other end of the battery) will try to pull electrons into the battery. But once a few electrons are pulled out of the wire the remaining positive ions in the wire pull back on the electrons, stopping the very little current flow. Suppose now that one attaches these two wires to two big identical flat metal plates (say 5 inch by 5 inch squares) and puts the two plates parallel to each other with a thin piece of glass in between. If the glass were not there the two plates would be in contact all along their 25 square inch surface and current would flow. I.e. electrons would move from the negative side of the battery, through the plates, and into the positive side of the battery. But with the glass separating the two plates what actually happens is this: a few electrons get pushed onto plate A from the negative side of the battery and a few electrons get pulled off plate B into the positive end of the battery. But the extra electrons on plate A are very close to the positive ions on plate B and consequently each (partly) neutralizes the push or pull of the other thereby allowing more electrons to get pushed onto plate A and more electrons to be pulled off plate B. The result is that some significant current flows out of the battery, but only for a very short time, until the charge built up on plate A, even after being partly neutralized by the positive ions on plate B, push back on new incoming electrons with a "force" of 3 volts. If you now cut the two wires the negative charges (the electrons) are trapped on plate A while the positive charges (the ions) are trapped on plate B. You have now

STORED CHARGE. Such an arrangement of plates is called a *condenser*, or equivalently, a *capacitor*. This is what Musschenbroek discovered. If now you separate the two plates the positive and negative charges no longer neutralize each other. The electrons on plate A now want to get off plate A very badly (yes this is anthropomorphism). A voltmeter attached between plate A and plate B would have shown a push of 3 volts when the plates were near. But after being separated there is a much higher push. The voltmeter will show a *much higher* voltage between plate A and plate B than 3 volts. (Keep this mind.) The arrangement of plates can be varied. For example if the piece of glass is actually a jar one could replace one of the plates by water in the jar (best to add a little salt to make it a good conductor) and one could replace the other plate by ones own hand (which is indeed a good conductor), in which one is holding the jar.

Here is Pieter van Musschenbroek's letter (of January 1746) to his friend J.A. Nollet, who included part of it in a paper of his own, published in the Memoires of the French Academy early in 1746. (Pretty short publication time, isn't it?) The following is from [49], page 52.

"I am going to tell you about a new but terrible experiment which I advise you not to try for yourself... I was making some investigations on the force of electricity. For this purpose I had suspended by two threads of blue silk, a gun barrel, which received by communication the electricity of a glass globe that was turned rapidly on its axis while it was rubbed by the hands placed against it. From the other end of the gun barrel there hung freely a brass wire, the end of which passed into a glass flask, partly filled with water. This flask I held in my right hand, while with my left hand I attempted to draw sparks from the gun barrel. Suddenly my right hand was struck so violently that all my body was affected as if it had been struck by lightning ... The arm and all the body are affected in a terrible way that I cannot describe; in a word, I thought that it was all up with me ... "

Peter van Musschenbroek nevertheless didn't give up these experiments. He went on to describe to Nollet variations of this experiment, in which, for example he placed the jar on a metal plate sitting on a wooden table, while the experimenter stood on the wooden floor when drawing the spark. There was very little effect in this case unless the experimenter touches the metal plate at the same time. Further, if one person holds the jar while another draws the spark (by touching the gun barrel with hand or metal rod) then the shock is very small. Thus he writes

"The person who tries the experiment may simply stand on the floor, but it is important that the same man hold the flask in one hand and try to draw the spark with the other: the effect is very slight if these actions were performed by two different persons. If the flask is placed on a metal support on a wooden table, then the one who touches this metal even with the end of his finger and draws the spark with his other hand receives a great shock."

For videos of a Leyden jar in action go to

<http://www.magnet.fsu.edu/education/tutorials/java/electrostaticgenerator/index.html>
<http://www.magnet.fsu.edu/education/tutorials/java/leydenjar/index.html>

The first shows how to charge up a Leyden jar. The second shows how to discharge it. Both are interactive. Be careful.

SUMMARY

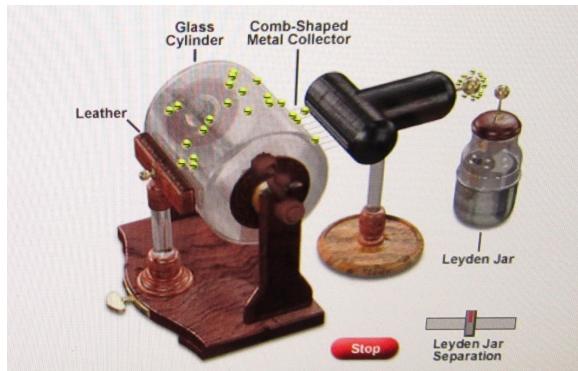


Figure 2: Production - Transfer - Storage of charge

These discoveries of the first half of the 18th century were qualitative: if you do this you will get more of that. (E.g. if you rotate this faster you will get more attraction or repulsion and bigger sparks. If you connect a wire the attraction or repulsion will show up at the other end. If you make a capacitor (Leyden jar) you can store lots and lots of the electrical fire (charge).

But none of this is quantitative. What can one measure anyway? What about spark length? Daniel Gralath made many such measurements of spark

length (about 1770) but was unable to embed his measurements into any consistent theory. Ideas on what one might usefully measure evolved. See [35, Chapter 26] on the evolution of ideas from 1760 onward that finally culminated in the measurement of FORCE between charged particles, by Coulomb, which, for the first time, put electrical science on a firm quantitative footing.

A small timeline of this period.

1776: The Colonies revolt against King George.

1785: Coulomb measures the force between charges.

1789: The French revolution. [Coulomb survived the French Revolution.]

3.3 Quantitative measurement begins: COULOMB

CHARLES AUGUSTIN COULOMB (1736-1806) invented the torsion balance, a very sensitive instrument for measuring small forces. He had already used it for measuring frictional forces, when, in 1785, he adapted it to the measurement of forces between charged particles.

3.3.1 How Coulomb did it.

There is available now on the web pictures of his instruments and explanations of how they worked. The Univrsity of Pavia website

<http://ppp.unipv.it/Coulomb/>

contains pictures of his actual instruments, modern versions, a biography of Coulomb, copies of his original articles (in French), translations into English, and finally an interactive version in which you can start his measurement yourself in dry air, normal air or wet air and see what happens. The URL for the latter is

<http://ppp.unipv.it/Coulomb/Pages/e5StrFrm.htm>

Upon arriving at this website click on the image of Coulomb's torsion balance on the left side of the page.

Here is the first quantitative result in the theory of electricity and magnetism, Coulomb's discovery. The following force between two charges at distance r was actually measured by Coulomb in case $q_1 = q_2$. The formula

itself can then be used to define units of charge in terms of already established units of force.

$$|F| = \frac{q_1 q_2}{r^2} \quad \text{Coulomb's law.} \quad (3.1)$$

Coulomb also verified accurately that an earlier measurement by Michell was correct: magnetic poles also attract or repel by an inverse square law.

3.3.2 Mathematical consequences

In the context of Newtonian mechanics Lagrange had shown (1777) that the force exerted on a massive body by several other massive bodies could be expressed in the form $F = -\text{grad } V$ for some function V . [We called V the potential of this force field in the earlier chapter. The name “potential” was given to it by George Green in 1828.] Laplace then showed (1782) that the function V satisfies the equation

$$\Delta V = 0 \quad \text{Laplace's equation in empty space, [1782].} \quad (3.2)$$

But after the discovery by Coulomb that the force on charges has a similar form to the force on masses [$1/r^2$ law] Simeon Denis Poisson (1781-1840) showed that the corresponding potential V satisfies, even in a charged region,

$$\Delta V = -4\pi\rho \quad \text{Poisson's Equation, [1812]} \quad (3.3)$$

To be precise:

Theorem 3.1 *Let ρ be a distribution on \mathbb{R}^3 with compact support. Define*

$$V(x) = \int_{\mathbb{R}^3} \frac{\rho(y)}{|x - y|} dy \quad (3.4)$$

Then (3.3) holds in the sense of distributions.

Proof. See Appendix 9.2 for a complete proof. ■

Let us jump ahead in time and interpret Coulomb's law from the point of view that was later pushed by Faraday and Maxwell and is now generally accepted. It is **field theory**. One thinks of the charge q_2 at the point y as generating an **electric field** $E : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ given by

$$E(x) = q_2 \frac{x - y}{|x - y|^3} \quad (3.5)$$

Then (3.1) says that a charged particle at x of size q_1 experiences a force $F = q_1 E(x)$. This equation clearly gives the direction of the force correctly as well as its magnitude. If you wish to give the vector field E some further physical meaning by thinking of it as measuring a stressed state of some kind of "aether" you will be in the company of many great physicists of the nineteenth century. But the implications of such a viewpoint disagree with later experiments (e.g the Michelson -Morley experiment.)

Now it is an assumption of linearity (to be checked by experiment) that the force on a charged particle at x produced by a number of charges is the sum of their separate forces. That is,

$$E(x) = \sum_{j=1}^N q_j \frac{(x - y_j)}{|x - y_j|^3}. \quad (3.6)$$

For a continuously distributed charge of density ρ the analog (and indeed Riemann limit) is

$$E(x) = \int_{\mathbb{R}^3} \frac{(x - y)}{|x - y|^3} \rho(y) dy \quad (3.7)$$

The force exerted by this distribution of charge on a small "test" charge at x and of charge q is then

$$F = qE(x) \quad (3.8)$$

Interchanging the gradient operator with the integral in (3.4) we see that $E(x) = -\text{grad } V(x)$. Since $\Delta = \text{div grad}$, Poisson's equation, (3.3) immediately gives the following formula.

Theorem 3.2 (*Gauss, 1777-1855*)

$$\text{div } E = 4\pi\rho \quad \text{Gauss' law} \quad (1837) \quad (3.9)$$

Equations (3.9) and (3.7) are *equivalent* in the sense that the only solution of (3.9) (that vanishes at infinity) is given by (3.7). Thus Gauss' law is exactly the differential version of Coulomb's law (3.1).

We are going to formulate all of the measured force laws as differential equations in this way in order to reach Maxwell's theory.

END of DAY 8 = 2/17/2011

3.4 The production of steady currents, 1780-1800

So far, all the experimental results and theories that we have discussed deal with charges sitting still - so called electrostatics. Never mind that there was movement of charge in sparks. These discharges couldn't be controlled anyway. The next stage in understanding of electrical phenomena began by yet another accident, culminating in the discovery by Volta of a battery, which could be used to produce a steady movement of charge - a current. As you may have heard somewhere, the ability to produce controlled currents was the vital ingredient in making the connection between electricity and magnetism and between electricity and chemistry.

1780: Luigi Galvani was an anatomist with an interest in electrical phenomena. For many years before the event that made him famous he had been studying the susceptibility of the nerves to irritation; and having been formerly a student of Beccaria, he was also interested in electrical experiments. In the latter part of the year 1780 he had, as he tells us, 'dissected and prepared a frog, and laid it on a table, on which, at some distance from the frog, was an electric machine. It happened by chance that one of my assistants touched the inner crural nerve of the frog, with the point of a scalpel; whereupon at once the muscles of the limbs were violently convulsed.'

‘Another of those who used to help me in electrical experiments thought he had noticed that at this instant a spark was drawn from the conductor of the machine. I myself was at the time occupied with a totally different matter; but when he drew my attention to this, I greatly desired to try it for myself, and discover its hidden principle. So I, too, touched one or other of the crural nerves with the point of the scalpel, at the same time that one of those present drew a spark; and the same phenomenon was repeated exactly as before.’

The preceding is taken from Whittaker [67, pp 67-68], which contains the following footnote. {¹ According to a story which has often been repeated, but which rests on no sufficient evidence, the frog was one of a number which had been procured for the Signora Galvani who, being in poor health, had been recommended to take a soup made of these animals as a restorative.}

As to whether the electrical effects discovered by Galvani were of animal origin or not sharply divided physicists. Alexander von Humboldt supported the animal origin viewpoint and did experiments on himself to prove it. Unfortunately they didn’t work. Here is an excerpt from Humboldt’s description of his experiment on himself:

“ I raised two blisters on my back, each the size of a crown piece and covering the trapezius and deltoid muscles respectively. Meanwhile I lay flat on my stomach. When the blisters were cut and contact was made with the zinc and silver electrodes, I experienced a sharp pain, which was so severe that the trapezius muscle swelled considerably, and the quivering was communicated upwards to the base of the skull and the spinous processes of the vertebra. Contact with silver produced thee or four single throbings which I could clearly separate. Frogs place upon my back were observed to hop.”

A reader interested in gore can read the rest of his description in [8, pages 32-33]. The next few pages in this book contain interesting gossip about Goethe and Schiller.

1793-1800: Volta was one of the physicists who believed that the effect discovered by Galvani had nothing to do with living matter, but depended somehow only on the contact of two different metals. By 1800 he discovered how to increase this effect. Whereas two discs, one copper and one zinc,

placed in contact, was known to produce a small but detectable effect on an electroscope, a “pile” of these could produce a shock. (Pile: a repeating sequence of discs: copper, zinc, moistened pasteboard, copper, zinc, moistened pasteboard, etc.)

As to whether the shocks from Volta’s battery were of the same electrical origin as had been explored so intensely in the preceding 100 years was settled by Volta thus:

- 1) Water could be decomposed by currents produced the old way (rubbing amber, store it in a Leyden jar etc.) or by Volta’s battery.
- 2) A Leyden jar could be charged up by Volta’s battery with similar shocking effect.
- 3) Electricities produced by opposite ends of a pile attract while electricities produced by corresponding ends of different piles repel.

Conclusion (1801): The electricity produced by Volta’s piles is the same stuff that’s produced by rubbing two materials together.

3.5 The connection between electricity and magnetism

What happened next really illustrates well how discoveries are dependent on previous discoveries. In the years after Volta invented the pile, the availability of a steady source of current made possible many discoveries concerning the relation between electrical charge and chemical reactions. But of interest for us was the accidental discovery in 1820 by Oersted that a current produces a magnetic field. As the story goes, he was showing his class that a current does not affect a compass. He had aligned the wire carrying the current perpendicular to the compass. No effect. A student suggested aligning the wire parallel to the compass. Big effect! The compass needle moved quickly to a perpendicular position. (Interesting story if true.) Here is a timeline of the following cross-channel communications.

3.5.1 Oersted, Ampère, Biot and Savart

1820: Oersted, Biot, Savart, Ampère. Immediately after the discovery by Oersted (July 1820), that a wire carrying current can influence

a compass, Biot and Savart and independently Ampère made careful measurements to determine the law of force of a current on a magnetic pole and (Ampère) on another current carrying loop. The dates and fast reaction of Biot and Savart and Ampère to Oersted's discovery are interesting because they show how intense the competition was. A physicist by the name of Arago came back from a meeting in London and described at a meeting in Paris, on September 11, 1820, the results that Oersted had discovered. A week later, on September 18, Ampère presented his ideas on how to quantify Oersted's discovery. On October 30 Biot and Savart gave their formula for the magnetic field produced by a wire carrying current. By October 30, 1820 all the main ideas were out on the table.

Biot and Savart gave a formula for the force that an infinitely long straight wire carrying a current would exert on one pole of a long bar magnet and therefore also on a compass needle. Ampère gave what could be regarded as a generalization of this. He gave the law of force by which one closed loop of current acts on another closed loop of current. For our purposes its best to break up their "action at a distance" laws of force into two parts, which reflects the later thinking of Faraday, Maxwell, and us. In the latter point of view the first loop generates a "magnetic field". The magnetic field then acts on the second loop. Here is a precise definition.

Definition 3.3 The **magnetic field** B at a point $x \in \mathbb{R}^3$ is the force exerted on a unit magnetic pole at x . Equivalently, it is the torque exerted on a small magnet at x of unit magnetic moment.

Its time for a little more precision. In measuring the strength of an electric field at a point x in space its important to place a very small test charge at the point x because a large test charge could affect the distribution of charges whose field we want to measure. Moreover the test charge should be located "at x ", not just distributed in some small neighborhood of x . Otherwise one will be measuring an average force over the neighborhood. In other words one should use a test particle of "vanishingly small" charge and spatial extent. Now in reality neither of these requirements can be met. We know now that the smallest magnitude of a charge is the charge on an electron. Moreover even a single atom has nonzero spatial extent. Thus the concept of a "vanishingly small" point charge is an idealization. Its a useful idealization when studying the classical theory of fields, as we are doing. Later, when we study the quantized theory of fields it will be both

conceptually and technically necessary to talk only about the “average” force of the field on a smoothly distributed test charge around x .

For the present therefore we should think of a measurement of the electric field at x as being carried out by a sequence of increasingly precise average measurements using smaller and smaller pith balls around x of smaller and smaller total charge. In practice this is just what one does. Measurements are never 100% accurate anyway. It's convenient to separate the two limiting operations conceptually: first let the pith ball shrink to a point, thereby obtaining what is usually called a POINT CHARGE. Then imagine repeating measurements with point charges of smaller and smaller charge.

The in-principle method of measuring a magnetic field is similar but with an extra complication. Take a very small compass needle and suspend it from its center by a thread so that it can rotate in any direction. Since the magnetic field pushes one end of the needle one way and the other end the other way the magnetic field will exert a torque on the compass needle so as to align the needle with the direction of the field. This determines the direction of B at the point x , where the center of the needle is located. The needle should be very short so that the field at one pole is “substantially” the same as at the other pole. Otherwise there will be a net force on the needle in addition to the torque and the needle will translate as well as rotate. The idealization of such a small compass needle (the analog of point charge) is called a MAGNETIC DIPOLE. A magnetic dipole is prescribed by a vector μ in \mathbb{R}^3 and should be conceptualized as a limit of small compass needles of length ϵ pointing in the direction μ and of pole strength $p > 0$ (north pole at the tip of μ) and pole strength $-p$ at the other end chosen so that $p\epsilon = |\mu|$. Thus $p \rightarrow \infty$ as $\epsilon \rightarrow 0$. This relation between ϵ and p is necessary in order for the limiting torque of the magnetic field on the tiny compass needle to exist and not vanish identically. The vector μ is called the *magnetic moment* of the magnetic dipole.

Exercise 3.4 Show that in a constant magnetic field B the torque on the small compass needles μ_ϵ about any point is independent of ϵ and of the point and is given by

$$N = \mu \times B$$

Exercise 3.5 Suppose that $B(x)$ is a continuously differentiable vector field on \mathbb{R}^3 . Show that the torque N_ϵ on the compass needle μ_ϵ described above satisfies

$$\lim_{\epsilon \rightarrow 0} N_\epsilon = N$$

But there is also a translational force. This underlies the Stern-Gerlach experiment. Find a formula showing that the translational force on a magnetic dipole located (exactly) at x in the presence of a non constant magnetic field is proportional to the magnetic moment and to the first derivative of the magnetic field at x .

Exercise 3.6 The mathematical meaning of these limiting configurations is most easily defined in terms of generalized functions. Let $\phi \in C_c^\infty(\mathbb{R}^3)$. A magnetic dipole μ at the origin is the linear functional

$$\phi \rightarrow \mu\phi \equiv -\mu \cdot \operatorname{grad} \phi(0).$$

Prove that

$$\mu\phi = \lim_{\epsilon \rightarrow 0} p[\phi(u\epsilon/2) - \phi(-u\epsilon/2)]$$

where u is a unit vector in the direction of μ .

The formulas of Biot and Savart and Ampère

Let C and C' be two oriented curves carrying currents i and i' respectively in the direction of the orientation. Then the current in curve C produces a magnetic field B given by

$$B(x) = i \int_C \frac{1}{|x-y|^3} d\mathbf{s} \times (x-y) \quad (3.10)$$

and the force on an element $d\mathbf{s}$ (at x) of the curve C' is

$$\mathbf{F} = (i' \times d\mathbf{s}') \times B(x). \quad (3.11)$$

Just as Poisson's equation (3.3) and Gauss' law (3.9) follow from Coulomb's law, so also Ampère's formula gives both of the following equations.

Corollary 3.7

$$\operatorname{div} B = 0 \quad (3.12)$$

$$\operatorname{curl} B = \frac{4\pi}{c} \mathbf{j}(x) \quad \text{Ampère's law} \quad (3.13)$$

where $\mathbf{j}(x)$ is the current density defined by

$$\begin{aligned} \mathbf{j}(x) \cdot \mathbf{n} dA &= \text{total charge passing through} \\ &\text{the element of surface } dA \text{ per second.} \end{aligned} \quad (3.14)$$

and c is some constant to be determined by measurements on condensers and coils of wire.

3.5.2 FARADAY

As Oersted showed, a current in a loop of wire generates a magnetic field. It stands to reason then that a magnetic field should generate a current in a loop of wire. Does it not? Faraday carried out the experiment to test this, in 1824, but found no effect. In our class we carried out this experiment also. In Figure 4 you can see a coil of wire in series with a meter ($50 \mu\text{A}$ full scale). The instructor (me) brought the magnet over to the coil of wire and held it there for several seconds. No movement of the meter was observed during all the time that I held it there. Thus we verified Faraday's 1824 observation that a magnetic field need not generate a current in a loop of wire.

It happened that several alert students, sitting in the first few rows of our class, noticed that there actually was a movement of the meter as I moved the magnet into the vicinity of the coil and as I moved it away. By a remarkable coincidence, Faraday noticed this effect also. He repeated his experiment in 1831. But this time he tested the hypothesis that a *changing* magnetic field should produce a current in a loop of wire. His report on its successful outcome was received by the Royal Society on November 24, 1831. If you look back at all the experiments done up to that point in electricity and magnetism you will see that all dealt with stationary circumstances - stationary charge, stationary current, stationary magnetic field. It's true that there were lots of sparks, etc. generated in the previous 130 years, and these represented sudden motion of charges. But these were not controlled enough to do quantitative experiments. Thus Faraday's experiment was the first one to deal with time dependent phenomena in electricity and magnetism.

We repeated Faraday's 1831 experiment in class and tested out two hypotheses. Along with Faraday we assumed that the current induced in the coil of wire depends on the first derivative, \dot{B} , of the magnetic field at the location of the wire. There are two obvious ways to get a number from combining \dot{B} with the geometry of the coil and which are invariant under Euclidean motions of the apparatus. Namely,

$$\text{Current} = \text{const.} \int_C \dot{B}_{\text{tangential}} dx \quad (3.15)$$

and

$$\text{Current} = \text{const.} \int_S \dot{B}_{\text{normal}} dA \quad (3.16)$$

where C denotes the closed curve which is the coil of wire and S is a surface

with boundary C . The integral over the surface is independent of the choice of S because $\operatorname{div} \vec{B} = 0$ by (3.12).

By moving the magnet perpendicular to the axis of the coil one can expect to get a large current if (3.15) is correct and a small current if (3.16) is correct. By moving the magnet parallel to the axis of the coil one should expect to get a large current if (3.16) is correct and a small current if (3.15) is correct. (Picture these two cases thoughtfully.) Three independent teams of students and faculty made observations of meter needle motion under the two types of magnet movement as above. All three teams concluded, independently, and well within the limits of experimental accuracy, that (3.16) can be the only correct one among these two formulas. By a remarkable coincidence, Faraday reached the same conclusion.

{This discussion is repeated a little differently after Figure 3. Leave it there for now.}

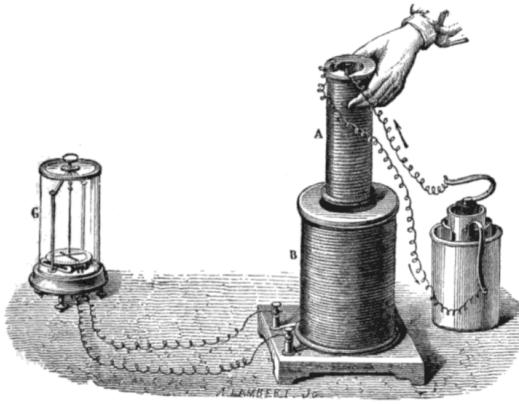


Figure 3: Faraday’s experiment showing induction between coils of wire: The liquid battery (right) provides a current which flows through the small coil (A), creating a magnetic field. When the coils are stationary, no current is induced. But when the small coil is moved in or out of the large coil (B), the magnetic flux through the large coil changes, inducing a current which is detected by the galvanometer (G).

Here is *our* version of Faraday’s experiment. We tested *two* hypotheses: given the assumption that a changing magnetic field $B(x, t)$, produces an electric field, which in turn produces a current in a loop of wire, does the current resulting from this electric field depend on the integral of the tangential component of the time derivative, \dot{B} , around the loop C ? or does it depend on the integral of the normal component of \dot{B} over some surface S with boundary C ? In the latter case, Equation (3.12), together with Gauss’s theorem, shows that the integral is independent of which surface one chooses. These two kinds of formulas were the only simple, translation and rotation invariant formulas we could think of that combine a closed curve and a vector field to produce a number. Our three expert lab assistants tested hypothesis #1 by moving the magnet quickly, perpendicular to the coil near its center (to cancel non tangential contributions.) They also tested hypothesis #2 by moving one pole of the magnet parallel to the axis very quickly. They independently arrived at the same conclusion as each other: Hypothesis # 2 is correct.

See Figure 4 for a picture of our equipment.



Figure 4: Differences from Faraday's version, compare with Fig. 3:

1. We used a magnet instead of an electromagnet (A).
2. We used a (nearly) modern ammeter instead of a galvanometer (G).
3. We had more wires hanging out than Faraday.
4. We tested *two* hypotheses; Hyp. 1, Current = constant $\int_C \dot{B}(x) \cdot dx$;
Hyp. 2, Current = constant $\int_S \dot{B}(x) \cdot \mathbf{n} d(Area)$ where S is a surface with boundary C .

By choosing the current loop C small, it now follows from our experimental verification of hypothesis #2 and from Stokes' theorem that

$$\text{curl } E = -\frac{1}{c} \partial B / \partial t \quad \text{Faraday's law} \quad (3.17)$$

for some constant c .

To be perfectly honest, the appearance of the electric field E in (3.17), rather than some version of current, as in (3.16), needs explaining. But the connection between these two was just being explored by Ohm at the time that Faraday was doing his experiment. Ohm did some experiments suggesting that there is an “electromotive force” that pushes current through wires. Just how much current will depend on both the nature of the wire (resistance as we would say nowadays), and the strength of the electromotive force. The electromotive force in the wire is in turn proportional to the electric field produced by the changing magnetic field, as in (3.17).

STATUS

From Coulomb, Oersted, Biot, Savart, Ampère and Faraday we now have relations between charge, electric field, current and magnetic field. The ex-

pression of these relations by means of differential equations is

$$\operatorname{div} E = 4\pi\rho \quad \text{Gauss' law} \quad (3.18)$$

$$\operatorname{div} B = 0 \quad \text{no magnetic monopoles} \quad (3.19)$$

$$\operatorname{curl} B = \frac{4\pi}{c} \mathbf{j}(x) \quad \text{Ampère's law} \quad (3.20)$$

$$\operatorname{curl} E = -\frac{1}{c} \partial B / \partial t \quad \text{Faraday's law} \quad (3.21)$$

3.6 MAXWELL puts it all together, 1861

1861: MAXWELL (1831-1879)

In truth, Ampère derived his force law from experiments involving steady currents. Neither the currents nor the charges were changing with time. There was no condenser on which charge might gradually accumulate. Consider a movement of the electric fluid (“charge”, as we would say today) whose flow is given by the vector field $\mathbf{j}(x)$ as defined in (3.14). But now allow the vector field to depend on time also. Denote by $\rho(x, t)$ the charge density. In a bounded open set V (with smooth boundary) the total charge inside is $\int_V \rho(x, t) dx$ while the rate of charge flow out of V is $\int_{\partial V} \mathbf{j}(x, t) \cdot \mathbf{n} dA$. Under the assumption that no electric fluid (i.e. charge) can be created or destroyed it follows that $\partial/\partial t \int_V \rho(x, t) dx + \int_{\partial V} \mathbf{j}(x, t) \cdot \mathbf{n} dA = 0$. By Gauss' theorem we find that $\int_V \partial/\partial t \rho(x, t) dx + \int_V \operatorname{div} \mathbf{j}(x, t) dx = 0$. Since V is arbitrary we find

$$\operatorname{div} \mathbf{j}(x, t) + \partial \rho(x, t) / \partial t = 0 \quad \text{conservation of charge.} \quad (3.22)$$

This equation has a different character from the equations of Gauss, Ampère and Faraday because it just reflects the assumption that there is some “electric fluid” which can move but not increase or decrease in total.

Now Ampère's law, (3.13), is not really consistent with the equation (3.22) of conservation of charge when the charge density is changing. Indeed, taking the divergence of (3.13) and using the identity $\operatorname{div} \operatorname{curl} = 0$ we find

$$0 = \operatorname{div} \operatorname{curl} B = \frac{4\pi}{c} \operatorname{div} \mathbf{j} = -\frac{4\pi}{c} \partial \rho / \partial t,$$

which is a contradiction unless $\partial \rho / \partial t = 0$.

Maxwell resolved the inconsistency by simply adding another term to Ampère's law, yielding the following internally consistent equations.

MAXWELL'S EQUATIONS

$$\operatorname{div} E = 4\pi\rho \quad \text{Gauss' law} \quad (3.23)$$

$$\operatorname{div} B = 0 \quad \text{no magnetic monopoles} \quad (3.24)$$

$$\operatorname{curl} B = \frac{4\pi}{c} \mathbf{j}(x) + \frac{1}{c} \partial E / \partial t \quad \text{Ampère-Maxwell} \quad (3.25)$$

$$\operatorname{curl} E = -\frac{1}{c} \partial B / \partial t \quad \text{Faraday's law} \quad (3.26)$$

This glib description of what Maxwell did and why must be taken cum grano salis. Maxwell in fact had some considerable physical reasons for adding on the extra term in (3.25). Imagine a closed loop of wire broken at two points. At one point insert a battery. At the other point insert a condenser. A current will flow for a while until the condenser is charged up. But is it "right" to say that a current is flowing across the gap in the condenser when in fact no charge is moving across this gap? (It's best to consider the gap evacuated for this discussion so that one doesn't digress onto polarization of some intermediate insulator between the condenser plates.) Maxwell said, yes, the term $c^{-1}\partial/\partial t E$ "represents" a current, the so called "displacement current" and with it the circuit can now be considered closed. Many of Maxwell's contemporaries did not look kindly on this viewpoint and continued to try to develop their own theory of forces on moving charges.

Implications of Maxwell's equations.

1. Electromagnetic influences propagate with a finite speed, c .

Proof: For simplicity consider Maxwell's equations in a region with no charges or currents. I.e. $\rho = 0$ and

$j = 0$. Then Maxwell's equations yield

$$\begin{aligned} \frac{1}{c^2} \partial^2 E / \partial t^2 &= \partial / \partial t \frac{1}{c} \operatorname{curl} B \\ &= -\operatorname{curl}^2 E \\ &= \Delta E - \operatorname{grad} \operatorname{div} E \\ &= \Delta E \end{aligned} \quad (3.27)$$

Thus E satisfies the wave equation $\frac{1}{c^2} \partial^2 E / \partial t^2 = \Delta E$. A similar computation shows that B also satisfies the same wave equation. As is well known, (just take the Fourier transform in the x variables) this equation implies that E is a superposition (Fourier transform) of waves that move at the speed c .

Now Faraday and some others before him suspected that light itself was a manifestation of magnetic waves. Maxwell confidently asserted this on the basis of his equations. Two simple (in-principle) laboratory experiments can be used to determine the constant c . Experiment No. 1 is purely electrical: Make a condenser of two parallel plates of measured area and measured distance. Connect a battery of voltage V and measure how much charge, Q , can be pushed onto one plate by the battery. Define $\epsilon = \alpha Q/V$ where α is a constant that depends only on the geometric measurements made before. Experiment No. 2 is purely magnetic: Pass a current through a long straight wire and measure the magnetic field produced one inch away from the wire. Insert the data into the formula of Biot and Savart to find the constant μ needed in their formula (but omitted in (3.13)) to make the magnetic field strength come out in agreement with this measurement. Had we written Maxwell's equations in terms of standard units, centimeters, Coulombs and seconds, these measured constants would have appeared in the equations. The constant c in (3.27) would have been replaced by $c^{-2} = \epsilon\mu$. In this way *one can measure c by experiments involving only stationary charge and stationary current.* **Experimental results.** Over the next 30 years (that is into the 1890's) more and more accurate measurements of ϵ , μ and the actual speed of light showed that the constant c is indeed (to within experimental error) the speed of light!

3.7 Maxwell's equations ala differential forms

It is remarkable that after 170 years of electric machines, conductors, shocking jars, kites, dead turkeys, balances, accidental currents and cleverly moved coils of wire, the entire edifice can be summarized in a couple of very simple looking equations (thanks to Cartan's calculus of differential forms), that one could almost guess at (in hindsight).

Choose coordinates x_1, x_2, x_3 and $x_4 = ct$ for \mathbb{R}^4 . The *Minkowski metric* on \mathbb{R}^4 is defined by its values on 1-forms by the definition

$$\langle dx_j, dx_k \rangle = g_{jk} \quad (3.28)$$

where $g_{jk} = \text{diag } 1, 1, 1, -1$. We are going to rewrite Maxwell's equations,

(3.23) - (3.26) in terms of differential forms. To this end we will define a 1-form $E^{(1)}$ which is a form version of the vector field E and define a 2-form $B^{(2)}$ which is the correct form version of the vector field B as follows.

$$E(x, y, z, t)^{(1)} = \sum_{i=1}^3 E_i(x, y, z, t) dx_i \quad (3.29)$$

$$B(x, y, z, t)^{(2)} = \sum_{(i,j,k)} B_i(x, y, z, t) dx_j \wedge dx_k \quad (3.30)$$

The sum over (i, j, k) refers to a sum over the three cyclic permutations of $(1, 2, 3)$. Then $E^{(1)}$ is a time dependent 1-form on \mathbb{R}^3 , which can just as well be interpreted as a 1-form on \mathbb{R}^4 with no dx_4 component. $B^{(2)}$ is a time dependent 2-form on \mathbb{R}^3 which we may and will interpret as a 2-form on \mathbb{R}^4 . The mapping from vectors to 1-forms or 2-forms is discussed further in Appendix 9.4 where the relations between the usual vector calculus operators div, curl, gradient and the exterior derivative on forms is reviewed. Define

$$F = E^{(1)} \wedge dx_4 + B^{(2)} \quad (3.31)$$

and

$$J = \sum_{i=1}^3 j_i dx_i - \rho dx_4. \quad (3.32)$$

Theorem 3.8 Denote by D the exterior derivative operator on forms over \mathbb{R}^4 and by D^* its adjoint with respect to the Minkowski metric (3.28). Then Maxwell's equations

$$\text{div } B = 0 \quad \text{no magnetic monopoles} \quad (3.33)$$

$$\text{curl } E + \frac{1}{c} \frac{\partial B}{\partial t} = 0 \quad \text{Faraday's law} \quad (3.34)$$

$$\text{div } E = 4\pi\rho \quad \text{Gauss' law} \quad (3.35)$$

$$\text{curl } B - \frac{1}{c} \frac{\partial E}{\partial t} = \frac{4\pi}{c} \mathbf{j}(x) \quad \text{Ampère-Maxwell} \quad (3.36)$$

hold if and only if

$$DF = 0, \quad \text{and} \quad (3.37)$$

$$D^*F = (4\pi/c)\mathbf{J} \quad (3.38)$$

Proof. Denote by d the exterior derivative on forms over \mathbb{R}^3 . Then $D\omega = d\omega + dx_4 \wedge (\partial\omega/\partial x_4)$ for any form ω over \mathbb{R}^4 . Hence

$$DF = (dE^{(1)}) \wedge dx_4 + dB^{(2)} + dx_4 \wedge (\partial B^{(2)}/\partial x_4) \quad (3.39)$$

Since $dx_4 \wedge$ commutes with the 2-form $(\partial B^{(2)}/\partial x_4)$ it follows that $DF = 0$ if and only if $dB^{(2)} = 0$ and $(dE^{(1)}) + (\partial B^{(2)}/\partial x_4) = 0$. These two equations are the form versions of (3.33) and (3.34), respectively, as shown in Appendix 9.4. Hence the equation $DF = 0$ is equivalent to the pair of equations (3.33) and (3.34).

Now let $\phi = \sigma^{(1)} + \tau dx_4$ be a C_c^∞ 1-form on \mathbb{R}^4 where $\sigma^{(1)} = \sum_{i=1}^3 \sigma_i dx_i$. Then, mindful of (3.28), we have

$$\begin{aligned} \int_{\mathbb{R}^4} \langle D^*F, \phi \rangle &= \int_{\mathbb{R}^4} \langle F, D\phi \rangle \\ &= \int_{\mathbb{R}^4} \langle F, d\sigma^{(1)} + (d\tau - (\partial\sigma^{(1)}/\partial x_4) \wedge dx_4) \rangle \\ &= \int_{\mathbb{R}^4} \langle B^{(2)}, d\sigma^{(1)} \rangle + \int_{\mathbb{R}^4} \langle E^{(1)} \wedge dx_4, (d\tau - (\partial\sigma^{(1)}/\partial x_4) \wedge dx_4) \rangle \\ &= \int_{\mathbb{R}^4} \langle d^*B^{(2)}, \sigma^{(1)} \rangle - \int_{\mathbb{R}^4} \langle E^{(1)}, (d\tau - \partial\sigma^{(1)}/\partial x_4) \rangle \\ &= \int_{\mathbb{R}^4} \langle d^*B^{(2)}, \sigma^{(1)} \rangle - \int_{\mathbb{R}^4} \left((d^*E^{(1)})\tau + \langle (\partial E^{(1)}/\partial x_4), \sigma^{(1)} \rangle \right) \\ &= \int_{\mathbb{R}^4} \langle d^*B^{(2)} - (\partial E^{(1)}/\partial x_4), \sigma^{(1)} \rangle - \int_{\mathbb{R}^4} (d^*E^{(1)})\tau \end{aligned}$$

On the other hand

$$\int_{\mathbb{R}^4} \langle J, \phi \rangle = \int_{\mathbb{R}^4} \langle j^{(1)}, \sigma^{(1)} \rangle + \int_{\mathbb{R}^4} \rho\tau \quad (3.40)$$

Therefore $D^*F = (4\pi/c)J$ if and only if $-d^*E^{(1)} = 4\pi\rho$ and $d^*B^{(2)} - (\partial E^{(1)}/\partial x_4) = (4\pi/c)j^{(1)}$. (A c is missing from somewhere.) These two equations are the form versions of the two remaining (inhomogeneous) equations (3.35) and (3.36), respectively, as shown in Appendix 9.4. ■

3.8 Electromagnetic forces ala Newton, Lagrange and Hamilton

In order to incorporate electromagnetic forces into quantum mechanics it will be necessary to reformulate the Newtonian mechanics for these forces

in Hamiltonian form. The force on a charged particle of charge e in the presence of electric and magnetic fields is given by $F = eE$ if the particle is standing still. This is the very definition of the electric field E . But if the particle is moving we have not yet written down the force. In Definition 3.3 we wrote down the force of a magnetic field on a single magnetic pole (with the other end of the magnet “far away”) or, equivalently, the torque on a very small magnet (magnetic dipole), which we took as the definition of the magnetic field. Four years after Hertz’ definitive experiments, showing that an electromagnetic wave from a spark really does propagate, all the way across a room, H.A. Lorentz proposed (1892) that the force on a moving charge is

$$\mathbf{F} = eE + (e/c)v \times B, \quad (3.41)$$

where B is the magnetic field at the position of the particle and v is the velocity of the particle. Take note of the fact that there is no magnet lurking in this definition even though the magnetic field is producing a force. The many consequences of Lorentz’ proposed force law are consistent with experiment and with Maxwell’s equations (provided quantum mechanics doesn’t have to be taken into account). Lorentz’ force law is now accepted.

In order to incorporate the Lorentz force (3.41) into the quantum mechanics of a charged particle it will be necessary to carry out the following procedure:

- 1) Express forces in terms of a potential.
- 2) Express Newton’s equations in terms of Lagrange’s equations.
- 3) Express Lagrange’s equations in terms of Hamilton’s equations.
- 4) Convert from Hamilton’s equations in classical mechanics to Schrödinger’s equation in quantum mechanics.

Step 4) will be described in the next chapter for velocity independent forces, in the section on the rules of quantum mechanics. But for a velocity dependent force, such as the electromagnetic force (3.41), the transition to quantum mechanics will be described in detail in Chapter 7. In the remainder of this chapter we will carry out Steps 1), 2) and 3).

Since the Lorentz force is velocity dependent it does not fall among the types of examples we considered in Chapter 2. It is not necessarily true that every type of force law in Newtonian mechanics can be transcribed into an equivalent Lagrangian formulation along the lines we described in Chapter 2. For example friction forces, which typically also depend on the velocity in a linear way, require a substantial change in the Lagrangian formalism, a

change which does not lend itself to quantization

But the very special form of the Lorentz force, in combination with Maxwell's equations, yields a successful Hamiltonian formulation, which, as we will see later, leads to a resulting Schrödinger equation that begs for a geometric interpretation in terms of connections on vector bundles. In fact this is the classical jumping off point to the appearance of Yang-Mills fields and to the currently most widely accepted theory of elementary particles.

Here are the details for Steps 1), 2), and 3), for a particle in the presence of an electromagnetic field. Step 4) will be carried out in Chapter 7.

Step 1. The electromagnetic potentials.

When the force on a particle depends on the velocity of the particle the simple equation Force = $-\nabla V$ can't hold for some function V of space only. Our discussion of the Lagrange formalism has to be modified. Since the 2-form F defined in the preceding section satisfies $DF = 0$ and \mathbb{R}^4 is cohomologically trivial there is a 1-form A such that

$$F = DA \quad (3.42)$$

The 1-form A is called an *electromagnetic potential* for the electromagnetic field E, B . A is highly non-unique: if f is any smooth function on \mathbb{R}^4 one may replace A by $A + Df$ without changing E and B because $D^2f = 0$. We may write $A = \sum_{j=1}^3 A_j dx_j + \phi dt$ where the components A_j and ϕ are real valued functions of x and t . The 2-form F has six independent components. The equation (3.42) may be written, as we know from Appendix 9.4,

$$B = \text{curl } \mathbf{A} \quad (3.43)$$

$$E = -\text{grad}\phi - (1/c)\partial\mathbf{A}/\partial t \quad (3.44)$$

where $\mathbf{A} = (A_1, A_2, A_3)$ are the three spatial components of the 1-form A . This is the customary way to write the equation (3.42) in the physics literature. *The 1-form A is going to replace our old potential $V(x)$. Whereas the potential V was unique up to an additive constant, the electromagnetic potential is unique only up to an exact 1-form, df .*

Step 2. Newton to Lagrange.

Here is a Lagrangian that allows the Lorentz force (3.41) to be incorporated into Lagrange's form of Newtonian mechanics. Define

$$L(x, v, t) = (1/2)m|v|^2 - e\phi + e(v/c) \cdot \mathbf{A} \quad (3.45)$$

Theorem 3.9 *With the force given by (3.41) Newton's equations are equivalent to the Lagrange equations*

$$\frac{d}{dt} \frac{\partial L}{\partial v_i} = \frac{\partial L}{\partial x_i}, \quad i = 1, 2, 3 \quad (3.46)$$

Proof. Referring to (3.45) we see that $\partial L / \partial v_i = m\dot{v}_i + (e/c)A_i(x)$. Hence

$$\frac{d}{dt} \frac{\partial L}{\partial v_i} = m\dot{v}_i + \frac{e}{c} \left(\frac{\partial A_i}{\partial t} + \sum_{j=1}^3 \frac{\partial A_i}{\partial x_j} \dot{x}_j \right) \quad (3.47)$$

$$= m\dot{v}_i + \frac{e}{c} \left(\frac{\partial A_i}{\partial t} + (v \cdot \nabla) A_i \right) \quad (3.48)$$

Also

$$\frac{\partial L}{\partial x_i} = -e \frac{\partial \phi}{\partial x_i} + e(v/c) \cdot \frac{\partial \mathbf{A}}{\partial x_i} \quad (3.49)$$

$$= -e \frac{\partial \phi}{\partial x_i} + \frac{e}{c} \frac{\partial (v \cdot \mathbf{A})}{\partial x_i} \quad (3.50)$$

since x_i and v are independent variables.

Therefore, taking into account (3.43) and (3.44), we find

$$\begin{aligned} \frac{d}{dt} \frac{\partial L}{\partial v_i} - \frac{\partial L}{\partial x_i} &= m\dot{v}_i + \frac{e}{c} \left(\frac{\partial A_i}{\partial t} + (v \cdot \nabla) A_i \right) + e \frac{\partial \phi}{\partial x_i} - (e/c) \frac{\partial (v \cdot \mathbf{A})}{\partial x_i} \\ &= m\dot{v}_i + e \left(\frac{\partial \phi}{\partial x_i} + (1/c) \frac{\partial A_i}{\partial t} \right) + (e/c) \left((v \cdot \nabla) A_i - \frac{\partial (v \cdot \mathbf{A})}{\partial x_i} \right) \\ &= m\dot{v}_i - eE_i + (e/c) \left((v \cdot \nabla) \mathbf{A} - \nabla(v \cdot \mathbf{A}) \right)_i \\ &= m\dot{v}_i - eE_i - (e/c)(v \times (\text{curl } \mathbf{A}))_i \\ &= m\dot{v}_i - eE_i - (e/c)(v \times B)_i \\ &= m\dot{v}_i - F_i \end{aligned}$$

where F is the Lorentz force, given by (3.41). We have used the identity $v \times (\nabla \times \mathbf{A}) = \nabla(v \cdot \mathbf{A}) - (v \cdot \nabla)\mathbf{A}$. ((Some readers might prefer to write this identity as $-i_v(dA) = d(i_v A) - \mathcal{L}_v A$, where \mathcal{L}_v is the Lie derivative, as in [40, Proposition 3.10].))) This proves the theorem. ■

Ref. Goldstein pages 19-21.

Step 3. Lagrange to Hamilton.

Next we must pass from Lagrangian mechanics to Hamiltonian mechanics. To this end we must compute the Legendre transform of the Lagrange function (3.45). Recall that one holds x and t fixed and carries out the Legendre transform on the fiber over x . In Example 2.45 we carried out the Legendre transform for the crucial part, $(1/2)m|v|^2 + e(v/c) \cdot \mathbf{A}$, and since the potential ϕ is independent of v it just goes along for the ride. Hence the Hamiltonian is

$$H(x, p, t) = \frac{1}{2m} \sum_{j=1}^3 (p_j - \frac{e}{c} \mathbf{A}_j(x, t))^2 + e\phi(x, t), \quad (3.51)$$

where I have written now H instead of L^* in accordance with the customary respect for Hamilton.

SUMMARY: Starting with Lorentz's force law, (3.41) for a moving charged particle in an electromagnetic field, we carried out the transitions from Newtonian mechanics to Lagrangian mechanics to Hamiltonian mechanics for Lorentz's force law. After we learn what the rules of quantum mechanics are, we will be ready to insert our Hamiltonian formulation of the electromagnetic force into quantum mechanics. This will be done in Section 7.

3.9 References

- Modern electrodynamics: J.D. Jackson [37]
- Maxwell's great treatise (1873) [44]
- History books
 - E.T. Whitaker (History of Aether and Electricity) [67, 68]
 - Roller and Roller (History just from the ancient Greeks to Coulomb) [49]
 - Holton and Roller [35].

4 Quantum Mechanics.

4.1 Spectroscope

The devices that were invented in the 18th century for producing, storing and measuring electric charge were simple enough for us to understand without much background on our part. But by the time of Galileo the development of optical devices was well along (telescopes by Galileo, then microscopes later). We need to understand, albeit at a superficial level, how one particular optical device works, whose purpose was to analyze the “quality” of light. As you know (I assume), a beam of light hitting a surface of glass at an angle will change direction as it enters the glass. In fact blue light will bend more than red light, with the result that if the hunk of glass is shaped like a prism the light emerging from the prism will be split up according to its color. As a result a beam of white light, hitting a side of a prism made of glass will split up, by the time it comes out, into its color components. See Figure 5

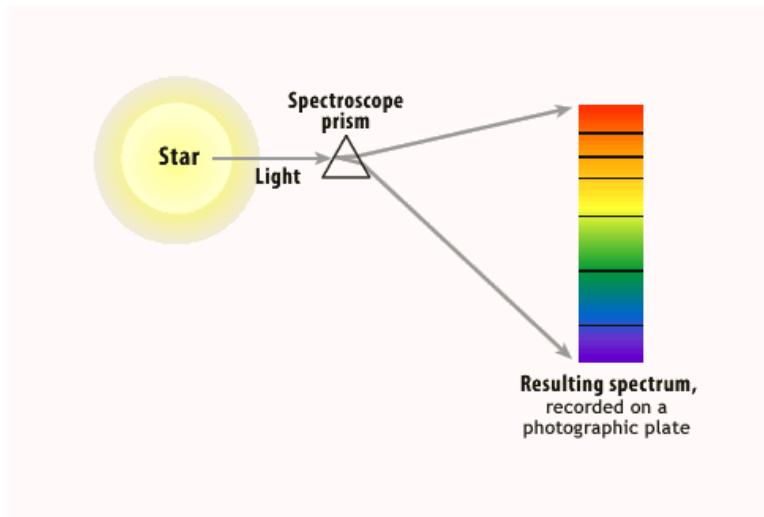


Figure 5: Spectral analysis by prism

Not shown in this simple diagram are three technical features, one of which is really important for us to understand.

First, the beam of light from the star has to be put through some lenses so that it will head toward the prism in a well formed beam of parallel light rays. That's just optics.

Second, the star can be replaced by any other source of light you might like to use. For example you might wish to put some hydrogen in a jar and pass a spark through it so that it will emit light. Of course you have to pass this light through the lenses so that the light will head in a well formed beam toward the prism. Or instead of hydrogen, or a star, you might like to vaporize some pure sodium in a jar, pass a spark through it, and direct the resulting light through the lenses toward the prism.

Third, and most important for us to understand, is that between the lenses and the prism you must place a vertical plate with a narrow horizontal slit in it. The incoming beam of light from the lenses, which you can think of as having a circular cross section, will be mostly stopped by this plate. What little comes through the horizontal slit will have a cross section which is just a horizontal line segment. Now this beam with a linear cross section will pass through the prism and break up into a bunch of horizontal line segments of different colors. The six horizontal dark lines in Figure 5 represent *missing* colors in the incoming beam from the star. In Figure 6 you can see what would result if the source is just hydrogen, instead of a star.

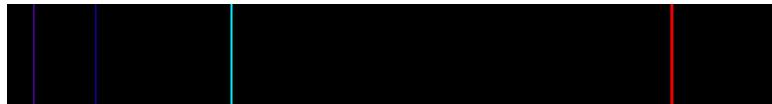


Figure 6: The four visible lines in the hydrogen spectrum

There are four lines visible on this dark background. Each one is an image of the horizontal slit. By measuring the positions of these lines you can figure out their wavelengths and therefore their frequencies, since (frequency) times (wavelength) = speed of light. Because of the appearance of such line segments in the measuring apparatus of frequencies in a light beam it is customary to refer to them as *spectral lines*. Of course each such line is just an image of the horizontal slit. In Figure 5 all colors are present in the incoming starlight *except* the six colors given by the dark lines. The missing colors (equivalently frequencies) (equivalently wavelengths) (equivalently lines) are called *absorption* lines. The four lines in the hydrogen spectrum are called *emission* lines. *Spectroscopy* is the study of spectra of light. The simple spectroscope described above was invented by Newton in 1666 and already used by him to study the radiation from stars (i.e. starlight). But as we saw above, one can also use it to study the radiation from heated chemical elements. There is yet another kind of radiation that can be analyzed in this

way (black body radiation, see next section). In the next section we will describe the spectroscopic effects discovered in the 100 years before quantum mechanics finally explained them.

4.2 100 years of spectroscopy: radiation measurements that defied Newton \cup Maxwell

While Faraday, Maxwell and friends were discovering the laws of interaction between electric charges, electric currents, magnetic fields and light, there was an independent line of discoveries being made concerning the interaction of matter with light. The emission and absorption of light by matter is what we mean by radiation. We saw in the preceding section how one can measure the ‘quality’ (spectral composition) of light emitted from various kinds of sources (stars, heated chemical elements). There is another kind of radiation source that we need to understand also. Namely, black body radiation, which is described in item 3 below. Here is a summary of spectroscopic knowledge accumulated over a hundred years.

1. A vaporized chemical element emits a spectrum peculiar to that element. The spectra of two different chemical elements are distinct.
2. If white light passes through a (cool) gas made of a chemical element then, after passing through the gas, the white light will have *dark* lines exactly where the (heated) chemical element had produced light lines as in item 1. The cool gas has absorbed radiation of the same wavelengths as the heated chemical element can emit.
3. A **blackbody**, e.g a piece of soot covered iron, emits a red glow when heated a little bit (think electric stove range). If heated hotter it will turn orange and if heated hot enough it will turn bright yellow (think lamp filament). A spectroscope shows that, actually, at any temperature, all frequencies are present in the emitted light. But at lower temperatures there is a preponderance of red light (i.e., low frequency light) while at higher temperatures the preponderance of light is more blueish (i.e., higher frequencies, equivalently, longer wavelengths). So the distribution moves toward the blue end (i.e., higher frequencies) as the body is heated. In Figure 7 you can see four light intensity curves at four different temperatures. Notice that the maximum point on a curve is more bluish at higher temperatures. Thus at 6000 K the peak is in the yellow, which is more toward the blue end than the peak

at 4000 K. The intensity distribution, as a function of frequency (or wavelength), depends only on the temperature and not on which “blackbody” is used. (Yes, its one word.)

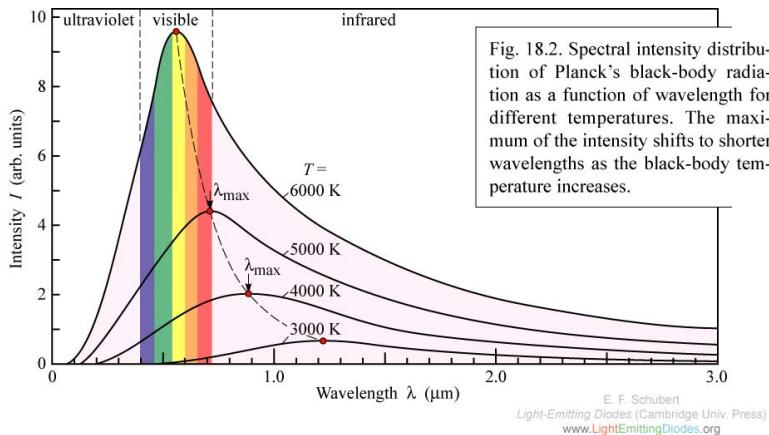


Figure 7: Blackbody radiation at different temperatures

Concerning the spectrum of individual chemical elements, it happens that some manipulations with the chemical element can change its spectrum. Thus:

4. (Zeeman effect, 1893.) The spectrum of chemical elements can be shifted by placing the test element in a magnetic field while examining its spectrum. For example take an element that's easy to deal with such as sodium. It has a very bright line (called the D line) in the yellow portion of its spectrum. This is why sodium vapor street lights cast a yellow light. This line has a wavelength (in vacuum) of approximately 5891.583264 Angstroms (1 Angstrom = 10^{-8} cm.)

Now place a weak magnetic field near the vaporized sodium. The D line splits into 3 distinct lines. Why? (The splitting increases as the magnetic field increases).

5. (Stark effect, 1913). Same effect as in item 4, except replace the magnetic field by an electric field.

6. Hydrogen spectrum. Most of the hydrogen spectrum is outside the visible spectrum. One needs a more sophisticated version of our spectroscope to measure the wavelengths of those lines that are not visible. But it can be

done. Here are five of the wavelengths that follow a remarkable pattern: the wavelengths are in nanometers (nm). ($1 \text{ nm} = 10 \text{ \AA}$.)

$$\begin{aligned}\lambda_3 &= 656 \text{ nm} \\ \lambda_4 &= 486 \text{ nm} \\ \lambda_5 &= 434 \text{ nm} \\ \lambda_6 &= 410 \text{ nm} \\ \lambda_7 &= 397 \text{ nm}\end{aligned}$$

The first four are in the visible spectrum. These four lines are shown in Figure 6. The red line is very bright. This is why passing sparks through a tube with pure hydrogen in it will produce a red light. *It just so happens* that all five of these wavelengths can be described by a simple formula, namely

$$\lambda_n^{-1} = R\left(\frac{1}{4} - \frac{1}{n^2}\right), \quad n = 3, 4, \dots, 7 \quad (4.1)$$

Is this simple formula just a fluke? After all we could add on to the right side any function that vanishes on the set $\{3, 4, 5, 6, 7\}$ and still have a valid formula. Well there are a lot of other spectral lines of hydrogen that fit this pattern also. The general formula

$$\lambda_{n',n}^{-1} = R\left(\frac{1}{(n')^2} - \frac{1}{n^2}\right), \quad n > n' \geq 1 \quad \text{Balmer-Rydberg formula} \quad (4.2)$$

fits a large number of observed spectral lines of hydrogen. In Figure 8 you can see many more of the spectral lines on hydrogen. The four lines in the visible range are shown in black. Don't be fooled by the colors in Figure 8. Ultraviolet is on the far left. Infrared is on the far right. {By the way it takes a lot of experimental effort to measure and analyze the hydrogen spectrum to find even one of these series (i.e. fixed n' .) We are indebted for $n' = 2$ to Balmer (1885), for $n' = 1$ to Lyman (1906-1914), for $n' = 3$ to Paschen (1908), for $n' = 4$ to Brackett (1922) for $n' = 5$ to Pfund (1924), and for $n' = 6$ to Humphreys.}

SUMMARY: In order to understand radiation one must explain an awful lot of very different kinds of observations. We just scratched the surface in our description of the previous six kinds of experiments. At least hundreds

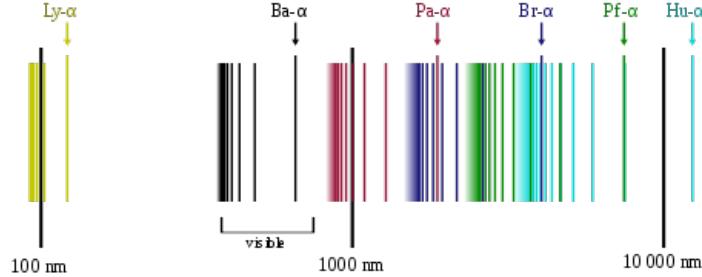


Figure 8: The four visible H lines are in black, starting from Ba - α . UV is to the left. IR is to the right.

(if not thousands) of unsuccessful attempts to explain these experiments on the basis of Newtonian mechanics and Maxwell’s theory of electricity and magnetism were made between (say) 1850 and 1925. It would be well to dwell on this for a few minutes before going on to the next section.

4.3 The rules of quantum mechanics

One’s physical intuition is based not only on experience but is also molded considerably by the mathematical models we make to explain our experience. For the motion of bodies that we are familiar with, under the influence of forces that we are familiar with, the mathematical model is classical mechanics. Table 2 describes some physical notions for which we have some intuitive ideas, based on our experience. In the second and third columns these intuitive notions are given mathematical representations in the two forms we’ve already studied, namely Lagrange’s and Hamilton’s formulations of Newtonian mechanics. We need to get the conceptual organization of these notions straight before passing to the corresponding mathematical representation of these notions in quantum mechanics.

Concerning line (i), the “instantaneous state” of the system is, by definition, the information necessary to determine the instantaneous state at all future times. Since Newton’s equations are second order in time one needs to know, typically, the initial positions and velocities of all “particles” (e.g. including rigid bodies) in the system. Both Lagrange and Hamilton’s formulation include such data, since the instantaneous state is a point $(q, v) \in T(C)$ or a point $(q, p) \in T^*(C)$, respectively.

Physical Notion	Interpreted by Lagrange	Interpreted by Hamilton
(i) Instantaneous state	A point $(q, v) \in T(C)$	A point (q, p) in $T^*(C)$.
(ii) Observable	A function $g : T(C) \rightarrow \mathbb{R}$	A function $f : T^*(C) \rightarrow \mathbb{R}$
(iii) Value in a state	$g(q, v)$	$f(q, p)$
(iv) Dynamics	Newton ala Lagrange	Newton ala Hamilton

Table 2: Classical mechanics ala Lagrange and Hamilton.

Concerning line (ii), the mathematical objects corresponding to things we observe are (usually) given by functions on the instantaneous state space. See Table 3 for lots of examples in case configuration space is just \mathbb{R}^3 . The “axiom” in line (iii) is pathetically self explanatory.

Concerning line (iv), the forces, masses, moments of inertia, etc. are needed to determine the motion of the system, given the instantaneous state. The motion is then determined by Newton’s equations, or, equivalently, by Lagrange’s equations or Hamilton’s equations.

The first three lines in Table 2 merely describe what system is under our purview. This is called *kinematics*. Line (iv) is the only line dealing with the motion of the system. The information governing the motion of the system is called *dynamics*.

Example 4.1 Configuration space = \mathbb{R}^3 : For a particle moving in \mathbb{R}^3 under the influence of some forces the configuration space is of course just \mathbb{R}^3 . The tangent bundle and cotangent bundle are both $\mathbb{R}^3 \times \mathbb{R}^3$, upon identifying \mathbb{R}^3 with its dual space. In Table 3 we have listed a bunch of observables and the functions on $T^*(\mathbb{R}^3)$ that represent them in classical mechanics. Later we will make the same table over again, but giving the mathematical objects that represent these same observables in quantum mechanics. The following table contains examples for the third column in Table 2, line (ii).

Quantum mechanical analog of Table 2.

This table is very non-definitive because the complex Hilbert space \mathcal{H} has not been specified. Nor have the physical meanings of those self-adjoint operators A in line (ii). Of course one must specify what the mechanical system under consideration is in order to determine \mathcal{H} . Just as with our definition of “configuration space” one can only learn how the Hilbert space \mathcal{H} is to be chosen by looking at examples.

The observable:	The function $f : T^*(C) \rightarrow \mathbb{R} :$
1.“ the j th coordinate of position”	$f((x, p)) = x_j$
2.“the j th coordinate of momentum”	$f((x, p)) = p_j$
3.“the potential energy ”	$f((x, p)) = V(x)$
4.“the kinetic energy”	$f((x, p)) = T = \frac{1}{2m} \sum_{j=1}^3 p_j^2$
5.“the z component of angular momentum”	$f((x, p)) = L_z = x_1 p_2 - x_2 p_1$
6.“total angular momentum”	$f((x, p)) = L^2 = L_x^2 + L_y^2 + L_z^2$
7.“Question: Is the particle in the Borel set $B \subset \mathbb{R}^3$?	$f((x, p)) = \chi_B(x)$
8.“the total energy”	$f((x, p)) = E = \frac{1}{2m} \sum_{j=1}^3 p_j^2 + V(x)$

Table 3: Some classical observables for a single free particle in \mathbb{R}^3 .

Physical Notion	Quantum Mechanics
(i) Instantaneous state	Unit vector ψ in Hilbert space \mathcal{H} .
(ii) Observable	Self-adjoint operator A on \mathcal{H}
(iii) Value in a state	Expected value: $(A\psi, \psi)$
(iv) Dynamics	Schrödinger equation

Table 4: Quantum Mechanics.

Example 4.2 Configuration space = \mathbb{R}^3 : How to choose the entries in the second column of Table 4 for a particle free to move in \mathbb{R}^3 under a force given by $F = -\text{grad } V$.

First: Take $\mathcal{H} = L^2(\mathbb{R}^3)$.

Second: Use the following table to determine which operators on \mathcal{H} have what physical interpretation. Compare these operators with the functions given in Table 3. In particular note that if a function $f((x, p))$ depends only on x , say $f(x, p) = g(x)$, then the quantum mechanical operator corresponding to that classical mechanical observable is just the operator of multiplication by g on $L^2(\mathbb{R}^3)$. Notation: $(M_g\psi)(x) = g(x)\psi(x)$. The observables in lines 1., 3. and 7. have this form. But all the others are very far from being multiplication operators. They are in fact differential operators. This is the time to dwell, for a few minutes, on these operators, especially the momentum operator P_j in line 2.

Third: In Table 4, line 3 you see “Expected value” instead of “Value”.

The observable:	The operator on \mathcal{H}
1.“ the j th coordinate of position”	$Q_j = M_{x_j}$
2.“the j th coordinate of momentum”	$P_j = -i\hbar\partial/\partial x_j$
3.“the potential energy ”	M_V
4.“the kinetic energy”	$T = \frac{1}{2m} \sum_{j=1}^3 P_j^2 = \frac{\hbar^2}{2m}(-\Delta)$
5.“the z component of angular momentum”	$L_z = -i\hbar(x\partial/\partial y - y\partial/\partial x)$
6.“total angular momentum”	$L^2 = L_x^2 + L_y^2 + L_z^2$
7.“Question: Is it in the Borel set $B \subset \mathbb{R}^3$?	M_{χ_B}
8.“the total energy”	$H = -\frac{\hbar^2}{2m}\Delta + M_V$

Table 5: Some observables for a single free particle in \mathbb{R}^3 and the corresponding operators on $\mathcal{H} \equiv L^2(\mathbb{R}^3)$.

This means that the theory does not actually predict what you will get in a measurement of the observable whose operator is A when the system is in the state ψ . Rather, it predicts that if you make the measurement many, many times and take the average you will get $(A\psi, \psi)$.

Fourth: The dynamics for the quantum mechanical system is given by the equation

$$i\hbar \frac{d\psi(t)}{dt} = H\psi(t) \quad \text{Schrödinger equation} \quad (4.3)$$

Here $t \mapsto \psi(t)$ is a function on \mathbb{R} into \mathcal{H} . Since H is a partial differential operator this ODE is really a PDE in disguise. We may write it explicitly as

$$i\hbar \frac{\partial\psi(t, x)}{\partial t} = -\frac{\hbar^2}{2m}\Delta\psi(t, x) + V(x)\psi(t, x) \quad \text{Schrödinger equation} \quad (4.4)$$

Advisement on Table 4 rows (i) and (iii). Projective space. If c is a complex number of absolute value one and ψ is a unit vector in \mathcal{H} then $c\psi$ is also a unit vector. Since $(A(c\psi), (c\psi)) = (A\psi, \psi)$, the expected value of “ A ” in the state $c\psi$ is the same as in the state ψ . But according to Table 4, the only information about the state ψ that the theory predicts is given by such inner products $(A\psi, \psi)$. (You can vary A .) Therefore ψ and $c\psi$ contain exactly the same physical information. So the set of instantaneous states is not really $\{\text{unit vectors in } \mathcal{H}\}$ but rather this set modulo the action

of the circle group on \mathcal{H} . This, of course, is exactly the infinite dimensional projective space, $P(\mathcal{H})$, based on \mathcal{H} , i.e., the set of one dimensional subspaces of \mathcal{H} .

On the one hand, it would be a technical nuisance to deal with the projective space $P(\mathcal{H})$ too extensively, instead of the nice Hilbert space \mathcal{H} , while on the other hand it is essential for the physical interpretation to realize that ψ and $c\psi$ represent the same physical state.

Direct physical interpretation of ψ . Line (iii) in Table 4 provides the (only) basis for interpretation of the wave function ψ when we have for our use the choices for A given in Table 5, and their powers. For ease in reading lets drop down to one dimension from three. Thus we will consider a particle free to move on the line. In this case the state space is $\mathcal{H} = L^2(\mathbb{R}^3)$. We can drop some subscripts in Table 5 and just write $Q = M_x$ and $P = -i\hbar d/dx$. Let n be a non-negative integer. According to the rule (iii) in Table 4, a measurement of the “observable” x^n should give on average the value $(Q^n\psi, \psi)$. Writing this out explicitly we find

$$(Q^n\psi, \psi) = \int_{\mathbb{R}} x^n |\psi(x)|^2 dx \quad n \geq 0 \quad (4.5)$$

Since $\int_{\mathbb{R}} |\psi(x)|^2 dx = 1$ the measure $|\psi(x)|^2 dx$ is a probability measure on the line and $(Q^n\psi, \psi)$ is just its n th moment.

Interpretation: $|\psi(x)|^2$ is the probability density for finding the particle at the point x under repeated measurements when the particle is in the state ψ . Neat, huh?

Generalization of this interpretation. We can easily abstract the preceding example, replacing our special Hilbert space $L^2(\mathbb{R})$ by an arbitrary Hilbert space \mathcal{H} and replacing Q by an arbitrary self-adjoint operator A on \mathcal{H} . In this case we may write, by the spectral theorem

$$A = \int_{-\infty}^{\infty} \lambda dE(\lambda) \quad (4.6)$$

where $E(\cdot)$ is a projection valued measure on the Borel sets of \mathbb{R} . Let ψ be a unit vector in \mathcal{H} and define

$$P(B) = (E(B)\psi, \psi) \quad B = \text{Borel set in } \mathbb{R}. \quad (4.7)$$

Then $P(\mathbb{R}) = (I\psi, \psi) = 1$, so P is a probability measure on \mathbb{R} . Moreover

$$(A^n\psi, \psi) = \int_{-\infty}^{\infty} \lambda^n P(d\lambda). \quad (4.8)$$

Just as in the preceding example, we may interpret (4.8) to mean that a measurement of the observable “ O ” represented by the operator A will be found to lie in a Borel set $B \subset \mathbb{R}$ with probability $(E(B)\psi, \psi)$.

For technical reasons (uniqueness of the moment problem) it is stronger to make an axiom based on the probability measure $B \mapsto (E(B)\psi, \psi)$ then line (iii) in Table 4. So if you really want to make some axiom system for quantum mechanics, you might as well replace line (iii) of Table 4 by the following.

Axiom of measurement. Given a self-adjoint operator A on the Hilbert state space, \mathcal{H} , and a unit vector ψ in \mathcal{H} , let

$$A = \int_{-\infty}^{\infty} \lambda dE(\lambda) \quad (4.9)$$

be the spectral resolution of A . Then any measurement of the observable “ O ” whose corresponding operator is A will lie in the Borel set $B \subset \mathbb{R}$ with probability

$$P(B) = (E(B)\psi, \psi) \quad (4.10)$$

when the system is in the state ψ . (As noted above, $P(B) \geq 0$ and $P(\mathbb{R}) = (I\psi, \psi) = 1$. So P is a probability measure on the line.)

Probability density for momentum. Lets return to the simple case of one particle moving on the line. We already have an interpretation of the wave function: $|\psi(x)|^2 dx$ is the probability measure for finding the particle at a point x . We got this by applying the Axiom of measurement to the position operator Q . Now lets apply this axiom to the momentum operator $A = P$. According to the Axiom of measurement the probability of of finding the momentum of the particle in some Borel set B in \mathbb{R} can be found from the spectral resolution of P . But the spectral resolution is known because the Fourier transform “diagonalizes” P .

Let

$$\hat{\psi}(p) := (2\pi\hbar)^{-1/2} \int_{\mathbb{R}} \psi(x) e^{ipx/\hbar} dx$$

be the Fourier transform of ψ . (The factor \hbar in the exponent makes the exponent dimensionless.) Differentiating under the integral sign we see that

$$\widehat{(P\psi)}(p) = p\hat{\psi}(p) \quad (4.11)$$

In other words, under the Fourier transform, which is of course a unitary map of $L^2(\mathbb{R}^3)$ onto itself, the differential operator P goes over to the multiplication operator M_p . Explicitly, $\mathcal{F}P\mathcal{F}^{-1} = M_p$. We may therefore proceed exactly as in the case of Q , which was multiplication by x , and conclude that $|\hat{\psi}(p)|^2$ is the probability density of finding a measurement of momentum to be p when the system is in the state ψ .

SUMMARY:

1. The rule (iii) in Table 4 is (informally) equivalent to the Axiom of measurement, (4.10).
2. For a particle moving on the line, and whose wave function is ψ
 - a) the probability that a measurement of position lies in an interval $[a, b]$ is

$$\int_a^b |\psi(x)|^2 dx \quad (4.12)$$

and

- b) the probability that a measurement of momentum lies in $[a, b]$ is

$$\int_a^b |\hat{\psi}(p)|^2 dp. \quad (4.13)$$

4.4 The Heisenberg commutation relations and uncertainty principle

There to be learned from our simple case of one particle moving on the line under some, as yet unspecified, force, even before we get to dynamics.. Suppose that the wave function is supported in a small interval $[c, d]$ of length at most ϵ . In that case measurements of position will find the particle somewhere in this interval on average because

$$c \leq \int_{\mathbb{R}} x |\psi(x)|^2 dx \leq d \quad (4.14)$$

How widely dispersed can these (many) measurements be? The usual quantitative measure of dispersion of measurements is the variance of the probability distribution: Let $x_0 = \int_{\mathbb{R}} x |\psi(x)|^2 dx$. Of course x_0 lies in the interval $[c, d]$ by (4.14). The *variance* of the distribution $|\psi(x)|^2 dx$ is, by definition,

$$Var(Q, \psi) = \int_{\mathbb{R}} (x - x_0)^2 |\psi(x)|^2 dx. \quad (4.15)$$

Since $(x - x_0)^2 \leq (\epsilon)^2$ on the support of $|\psi|^2$ it follows that

$$Var(Q, \psi) \leq \epsilon^2. \quad (4.16)$$

Undoubtedly you have been wondering whether quantum mechanics allows the existence of states for which repeated measurements of position will come out close to the same value every time, just as in classical mechanics. You see that for the wave function ψ supported in the small interval $[c, d]$ the variance of these measurements will be small. In fact the higher moments, $\int_{\mathbb{R}} (x - x_0)^{2n} dx \leq \epsilon^{2n}$ which are also small. So here is a quantum state of the particle in which not only is the average value of many measurements equal to x_0 , but the deviation of these measurements from x_0 is also small in the usual statistical sense. In fact, the probability that a measurement of position will lie outside the interval $[c, d]$ is zero by the axiom of measurement, isn't it.

Now the dispersion of measurements of the momentum in this same state ψ can be computed the same way, using (4.13) instead of (4.12). The variance of momentum measurements in the state ψ is therefore

$$Var(P, \psi) = \int_{\mathbb{R}} (p - p_0)^2 |\hat{\psi}(p)|^2 dp \quad (4.17)$$

where

$$p_0 = \int_{\mathbb{R}} p |\psi(p)|^2 dp \quad (4.18)$$

is the expected value of these measurements. We were able to choose ψ so that the variance $Var(Q, \psi)$ was small. But having chosen ψ , its Fourier transform is no longer under your control. The next theorem shows that you can't measure both position and momentum with arbitrarily good accuracy.

Theorem 4.3 Heisenberg uncertainty principle *For any unit vector $\psi \in L^2(\mathbb{R})$,*

$$Var(Q, \psi) Var(P, \psi) \geq \hbar^2/4. \quad (4.19)$$

This theorem looks like a theorem in Fourier analysis, which it is. But it is also a direct consequence of the following fundamental identity of Heisenberg, which itself follows immediately from the product rule for derivatives.

The Heisenberg commutation relations

$$PQ - QP = -i\hbar \text{ (Identity operator on } \mathcal{H}). \quad (4.20)$$

Indeed, since $P = -i\hbar(d/dx)$ and $(Q\psi)(x) = x\psi(x)$ this identity merely says that $(-i\hbar)(x\psi(x))' - x\psi'(x) = (-i\hbar)\psi(x)$.

All formulations of quantum mechanics build in the Heisenberg commutation relations (4.20) in one way or another. It is precisely the non-commutativity encompassed by (4.20) that distinguishes classical mechanics from quantum mechanics, in the view of many.

We will show now that (4.20) implies (4.19). First lets get straight the general definition of variance.

Definition 4.4 Let A be a self-adjoint operator on a Hilbert space \mathcal{H} and let ψ be a unit vector in the domain of A . The *mean* and *variance* of A in the state ψ are defined by

$$a = (A\psi, \psi) \quad (\text{Mean of } A \text{ in the state } \psi). \quad (4.21)$$

$$\text{Var}(A, \psi) = \|(A - a)\psi\|^2 \quad (\text{Variance of } A \text{ in the state } \psi). \quad (4.22)$$

This is a good time to verify that this definition agrees with our previous definitions of $\text{Var}(Q, \psi)$ and $\text{Var}(P, \psi)$.

Theorem 4.5 *Let A and B be self-adjoint operators on a Hilbert space \mathcal{H} . Assume that for some complex number c and for all vectors $\psi \in \text{Dom}(AB) \cap \text{Dom}(BA)$ there holds*

$$(AB - BA)\psi = c\psi \quad (4.23)$$

Then, for all unit vectors $\psi \in \text{Dom}(AB) \cap \text{Dom}(BA)$

$$\|A\psi\|\|B\psi\| \geq |c|/2 \quad (4.24)$$

and

$$\text{Var}(A, \psi)\text{Var}(B, \psi) \geq |c|^2/4 \quad (4.25)$$

Proof. {Reference: von Neumann [60, page 234].} If $\psi \in Dom(AB) \cap Dom(BA)$ then the following computation is justified:

$$\begin{aligned} 2\|A\psi\|\|B\psi\| &\geq 2 \operatorname{Im}(A\psi, B\psi) \\ &= (-i)\left[(A\psi, B\psi) - (B\psi, A\psi)\right] \\ &= (i)((AB - BA)\psi, \psi) \\ &= (i)(c\psi, \psi) \\ &= ic\|\psi\|^2 \end{aligned}$$

Take absolute values to deduce (4.24) in case $\|\psi\| = 1$. By the way, notice that the last four equalities imply that c must be purely imaginary if (4.23) holds.

Now let $a = (A\psi, \psi)$ and $b = (B\psi, \psi)$. Then $A - aI$ and $B - bI$ also satisfy the commutation relations (4.23). Therefore (4.24) shows that

$$\|(A - aI)\psi\|\|(B - b)\psi\| \geq c/2 \quad (4.26)$$

This proves (4.39). ■

Taking now $A = P$ and $B = Q$ and $c = i\hbar$ we see that (4.19) is just a special case of (4.39). But while we're at it lets formulate the Heisenberg commutation relations for a system whose configuration space is \mathbb{R}^n . (E.g. $n = 3, 6, \dots$ are interesting.) You probably don't need me to tell you that for a system of particles whose configuration space is \mathbb{R}^n , the Hilbert state space should be taken as $\mathcal{H} = L^2(\mathbb{R}^n)$. Moreover the position operator corresponding to the the j th coordinate, x_j on \mathbb{R}^n should be taken to be $Q_j = M_{x_j}$ while the observable "jth component of momentum" is represented by the operator $P_j = -i\hbar\partial/\partial x_j$. Now the product rule for derivatives gives, just as before, (compare (4.20))

$$P_j Q_k - Q_k P_j = -i\hbar\delta_{jk} I_{\mathcal{H}}. \quad (4.27)$$

For any unit vector ψ we have therefore the joint uncertainty relations

$$Var(Q_k, \psi)Var(P_j, \psi) \geq (\hbar^2/4)\delta_{jk} \quad (4.28)$$

Of course if $j \neq k$ then (4.27) says that P_j and Q_k commute while (4.28) says nothing.

END of DAY 14 = 3/10/2011

4.5 Dynamics and stationary states

We have so far discussed only lines (i), (ii) and (iii) in Table 4. These deal only with the description of the quantum analog of instantaneous states, i.e., kinematics. Now we want to discuss the rules by which quantum mechanics asserts that states change with time. Line (iv) in Table 4 specifies that the time evolution is guided by the Schrödinger equation.

$$i\hbar \frac{d\psi(t)}{dt} = H\psi(t) \quad \psi(0) = \psi_0 \quad \text{Schrödinger equation} \quad (4.29)$$

where H is the total energy operator on the state space \mathcal{H} . In our example of a particle moving in \mathbb{R}^3 under the force $F = -\text{grad } V$ the total energy operator is

$$H = -\frac{\hbar^2}{2m}\Delta + M_V \quad (4.30)$$

as we see from Table 5.

It is appropriate to write d/dt in (4.29) because one should think of $\psi(\cdot)$, in the abstract, as a function of one variable, t , into H . But when H is itself a function space, as in Table 2, it is appropriate to write (4.29) as a partial differential equation:

$$i\hbar \partial\psi/\partial t = -\frac{\hbar^2}{2m}\Delta\psi(x) + V(x)\psi(x), \quad (4.31)$$

which is the concrete form that the Schrödinger equation takes for our Examples 4.1 and 4.2.

The solution to (4.29) is simply given, by the spectral theorem,

$$\psi(t) = e^{-it\hbar^{-1}H}\psi_0 \quad (4.32)$$

The operators $e^{-i(t/\hbar)H}$ are unitary operators on \mathcal{H} for each real t , again by the spectral theorem. Of course it is one thing to cite a powerful theorem like the spectral theorem to ensure the existence of a solution to the Schrödinger equation and quite another to actually find the solution explicitly. But the abstract solution (4.32) allows us to understand some of the physical meaning of the dynamics.

Definition 4.6 An eigenfunction for the energy operator H is called a *stationary state* of the system.

So why should an eigenfunction be called a stationary state? Well, if $H\psi_0 = \lambda\psi_0$ then the solution (4.32) reads

$$\psi(t) = e^{-i(t/\hbar)\lambda}\psi_0 \quad (4.33)$$

because $e^{itH}\psi_0 = e^{it\lambda}\psi_0$ for such an eigenfunction. BUT, for each real t , $e^{it\lambda}\psi_0$ is just a multiple of ψ_0 by a constant of absolute value one. Consequently $\psi(t)$ and ψ_0 represent the same physical state, as we have already observed above. In other words although the vector $\psi(t)$ changes with time, the state $\psi(t)$ does not change with time. So we call it a stationary state. Isn't this a reasonable terminology? Yes.

Here is some physical interpretation of stationary states that relates to our classical view of one particle orbiting another (earth around sun, or electron around nucleus.) A classical particle orbiting another in an elliptical orbit is of course not standing still. Its position is changing with time. *But the orbit is not changing.* It remains the same ellipse forever. What Heisenberg realized (or conjectured, or believed, or thought, or theorized) while he was on the island Helgoland during hay fever season in May, 1925 was that you shouldn't really talk about the orbit of the electron in too much detail because you can't measure the electron's position anyway, without destroying the orbit. E.g. if you "look at" the electron, i.e. bounce even just one photon off of it, you will change its velocity so much that it will no longer be in the same orbit. By contrast, if you "look at" a planet, i.e reflect a few trillion trillion photons into your telescope, it will not change the orbit of the planet perceptibly. So you can see where the planet is at any instant. Heisenberg once complained that his school chemistry book explained the bonding of chemical elements by showing the atoms with hooks attached so that one atom would grab on to the hook of another atom. In his view this was, shall we say politely, misleading. But he viewed the classical picture of elliptic orbits for electrons as misleading also because you could not, even in principle, measure the position of the electron in such an orbit. Instead he took the whole orbit itself as the physical object to study rather than the position of the electron in the orbit. In his view the only thing that could actually be measured was the energy of these orbits and the rate of transition from one orbit to another (under the influence of bumping by photons).

4.6 Hydrogen

As we know, the potential that gives the force between two charged particles is proportional to their inverse distance. (See (??).) We want to consider the example which models the hydrogen atom. We are going to take the proton fixed at the origin and allow the electron to wander around \mathbb{R}^3 subject to the attractive Coulomb force exerted by the proton. The potential is given by

$$V(x) = -\frac{\gamma}{|x|}, \quad x \in \mathbb{R}^3 \quad (4.34)$$

for some positive constant γ , which will be discussed some more later. You might like to check that this sign assures that the force, $-\nabla V$, always points toward the origin. This artifice of fixing the proton at the origin reduces our mechanical system to one whose Hilbert space is $L^2(\mathbb{R}^3)$ and allows us to apply the machinery of Table 4. Fixing the proton at the origin, instead of allowing the two particles to move freely, which would entail taking configuration space to be \mathbb{R}^6 , forces a minor correction in the mass of the electron, which will be explained later. The Hamiltonian operator for the hydrogen atom is

$$H = -\frac{\hbar^2}{2m}\Delta - \frac{\gamma}{|x|}, \quad (4.35)$$

where m is the mass of the electron. (This will be corrected later.) H is a self-adjoint operator on $L^2(\mathbb{R}^3)$, when its domain is chosen right. The eigenvalue equation for this operator is the partial differential equation

$$-\frac{\hbar^2}{2m}\Delta\psi - \frac{\gamma}{|x|}\psi = E\psi \quad (4.36)$$

where E is a real number and ψ is subject to the condition $\psi \in Dom(H)$. The solutions to this equation may be found by using spherical coordinates (r, θ, ϕ) . One seeks a solution, as usual, in the form $\psi(r, \theta, \phi) = u(r)$ times functions of θ and ϕ . This leads to an ODE for each of the three functions. You can consult a standard quantum mechanics book, such as Schiff [51, pages 80-87] for details on how to carry this out, if you really want to know. Here is the result. Let

$$E_n = -\frac{\kappa}{n^2}, \quad n = 1, 2, \dots, \quad (4.37)$$

where $\kappa = m\gamma^2/(2\hbar^2)$. Then there are functions of the form

$$\psi_{n,l,m}(r, \theta, \phi) = u_{n,l}(r)Y_{l,m}(\theta, \phi), \quad l = 0, 1, \dots, n-1, \quad -l \leq m \leq l \quad (4.38)$$

such that

$$-\frac{\hbar^2}{2m}\Delta\psi_{n,\ell,m} - \frac{\gamma}{|x|}\psi_{n,\ell,m} = E_{n+\ell}\psi_{n,\ell,m} \quad (4.39)$$

(As a matter of culture you might like to know that the functions $Y_{\ell,m}$ are the spherical harmonics on S^2 , while the functions $u_n(r)$ are essentially the Laguerre functions on $(0, \infty)$.) The set $\{E_1, E_2, \dots\}$ is the entire point spectrum of H while the functions $\psi_{n,\ell,m}$ form an orthonormal basis of the subspace of $L^2(\mathbb{R}^3)$ spanned by all eigenfunctions. The remainder of the spectrum of H is the interval $[0, \infty)$.

What does this have to do with the line spectrum of hydrogen? The numbers, E_n , that we arrived at are energies, whereas the positions of the spectral lines are determined by frequencies (equivalently, wavelengths), as in the Rydberg formula, (4.2). Here is the connection.

Planck's hypothesis

- a) Light comes in little packets called photons.
- b) The energy of a photon of frequency ν is

$$E = h\nu, \quad (4.40)$$

for some constant h independent of ν and to be determined by experiment.

Now we know that a hydrogen atom, left to itself, and starting in a stationary state, such as $\psi_{n,\ell,m}$ will remain in that state permanently, because that's what the rules in Table 4 tell us. But just **suppose** that, for some reason, the orbiting electron suddenly drops into a lower energy orbit, $\psi_{n',\ell',m'}$. That is, $E_{n'+\ell'} < E_{n+\ell}$. What happens to the energy that it loses? Well, just **suppose** that the energy that it loses is somehow radiated away by the emission of a photon. By Planck's hypothesis the frequency of the emitted photon is given by (4.40). Therefore the frequency of the emitted photon is $\nu = h^{-1}(E - E')$. In view of (4.37) and (4.39) the frequency of the emitted photon is

$$\nu = \frac{m\gamma^2}{2h\hbar^2} \left(\frac{1}{(m')^2} - \frac{1}{m^2} \right) \quad (4.41)$$

where $m' = n' + \ell'$, $m = n + \ell$ and γ , which is defined in (4.34), is proportional to the product of the electric charges (electron and proton) in suitable units. Thus we have derived the Balmer-Rydberg formula, (4.2), including the value of Rydberg's constant R ! ! ! (Of course we have to know Planck's constant.)

SUMMARY: We have derived the line spectrum of hydrogen from the rules of quantum mechanics (Tables 4 and 5) in conjunction with Planck's hypothesis, (4.40).

Let's back up a little and admit that some things are just a little fishy with this "derivation" of the hydrogen spectrum.

First, where did Planck's hypothesis come from? Answer: After previous failed attempts (by Wien (1893) and Lord Raleigh (1900)) to derive a formula for the experimentally observed graph, Figure 7, for blackbody radiation, Planck derived a formula that fit the data, using the hypothesis (4.40) as an input. The value of h that makes this formula fit the data can be deduced from the experimental data. Planck announced this in December of 1900, which some identify as the beginning of quantum mechanics. Moreover, in 1905, Einstein used Planck's hypothesis to show that a beam of light, shining on a piece of metal in an evacuated tube, will knock electrons out of the metal if and only if the frequency of the light is high enough. This is the photoelectric effect. Einstein's thinking was that the energy of little corpuscles of light, i.e. photons as we call them now, must be high enough to overcome the forces at the surface of the metal, which are trying to prevent electrons from escaping. But according to Planck (4.40), this means that the frequency of the impinging light must be high enough. Einstein's deductions were soon confirmed experimentally (by others) and Einstein won a Nobel Prize. Thus for this reason and other nice ways in which Planck's hypothesis was used successfully, the formula (4.40) was widely accepted by 1925, when our Tables 4 and 5 originated.

Second, why should the electron suddenly drop down from a higher energy orbit to a lower one? And if all the electrons in all the atoms keep dropping down to lower energy orbits there won't be any higher energy electrons left, soon, to drop down. There must be some mechanism operating that makes them jump up to a higher energy orbit. What is it?

Third, the mechanism for explaining why and how the atom emits a photon when jumping is entirely missing. The electromagnetic field that represents the photon mathematically must somehow enter the equations. But the electromagnetic field is nowhere in sight in Table 5 or in the computations at the beginning of this section. Furthermore, looking back at Chapter 3 on electricity and magnetism, where do you see any mathematical structure that

suggests such a discrete notion as “one photon, two photons . . .”? Answer: You don’t.

Resolution: Not only must our notion of quantization be applied to the electron orbiting the proton but also to the electromagnetic field that is produced by the electron and in turn influences the electron. Whereas a particle moving in \mathbb{R}^3 has only three (or six) degrees of freedom, any classical field theory has infinitely many degrees of freedom (e.g. the space of initial conditions). We can already expect, therefore that the quantum Hilbert state space for a field theory will be something like $L^2(\mathbb{R}^\infty)$ instead of $L^2(\mathbb{R}^n)$ for some finite n . This is what we will start on in the next chapter.

Nevertheless the simple theory of the hydrogen atom developed above is a vital ingredient in much of chemistry and should not be looked down on just because we haven’t incorporated the interaction with the electromagnetic field. In fact the sets $\{x \in \mathbb{R}^3 : |\psi_{n,\ell,m}(x)|^2 \geq a\}$, for various $a > 0$, are very important to understand for chemistry and are graphed extensively in chemistry books and on the web. Your classmate Yao Liu has suggested the following website, which contains lots of nice pictures of these sets, called orbitals.

<http://winter.group.shef.ac.uk/orbitron/>

Here is the orbital for $\psi_{2,1,1}$. The wave function is positive in the green zone and negative in the blue zone.

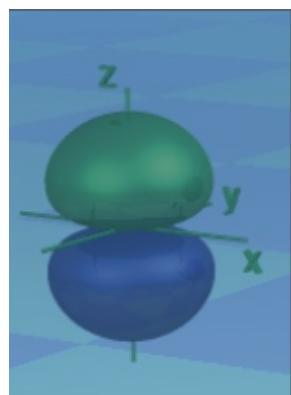


Figure 9: Orbital for $\psi_{2,1,1}$

END of DAY 15 = 3/15/11

4.7 Combined systems, Bosons, Fermions

There is a loose set of rules explaining how to construct the Hilbert space representing a system composed of two independent systems in terms of the Hilbert spaces associated to each subsystem.

In classical mechanics the rule is simple: if C_1 and C_2 are the configuration spaces for two systems then the Cartesian product $C_1 \times C_2$ is the configuration space for the combined system. Moreover the total energy is, typically, the sum of the energies of the two subsystems plus another term that represents the energy of interaction. For example there may be forces acting between the two subsystems in addition to the original forces acting within each subsystem.

In quantum mechanics the rule is more or less similar: if \mathcal{H}_1 and \mathcal{H}_2 are the Hilbert state spaces for two systems then the Hilbert space for the combined system is the tensor product

$$\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2. \quad (4.42)$$

This is consistent with the combining rule for classical mechanics because of the natural isomorphism

$$L^2(C_1 \times C_2) = L^2(C_1) \otimes L^2(C_2), \quad (4.43)$$

which holds for whatever measures you use on the factors, if you use the product measure on the product. For example the Hilbert state space for a system of two particles, each free to move in \mathbb{R}^3 is $L^2(\mathbb{R}^6)$ because the configuration space of the system composed of two particles, each moving in \mathbb{R}^3 , is $\mathbb{R}^3 \times \mathbb{R}^3 = \mathbb{R}^6$. But by the natural isomorphism (4.43) this Hilbert space can be just as well described as $L^2(\mathbb{R}^3) \otimes L^2(\mathbb{R}^3)$.

Here are the two exceptions to the rule (4.42). Suppose that we have two identical particles. In classical mechanics they are still distinguishable, even though identical: you can paint a 1 or a 2 on two identical billiard balls. (Use a thin layer of paint so as not to alter their mechanical properties.) But you can't paint anything on an electron or a photon. *In quantum mechanics two identical particles are not only identical but also indistinguishable.* If \mathcal{H} is the quantum Hilbert space for each of the two identical particles then the Hilbert space for the pair is not $\mathcal{H} \otimes \mathcal{H}$. Rather, denoting, the interchange operator $P : \mathcal{H} \otimes \mathcal{H} \rightarrow \mathcal{H} \otimes \mathcal{H}$, which is the unique unitary operator extending the transposition map $P(u \otimes v) = v \otimes u$, one insists that the state space of the two

particles is just the set of those vectors in $\mathcal{H} \otimes \mathcal{H}$ which are indistinguishable under this transposition. There are two known types of particles which are subject to this rule:

A *Boson* is a particle which, when combined with a particle identical to itself, has for its combined state space

$$\{\psi \in \mathcal{H} \otimes \mathcal{H} : P\psi = \psi\}. \quad \text{Bosons} \quad (4.44)$$

A *Fermion* is a particle which, when combined with a particle identical to itself, has for its combined state space

$$\{\psi \in \mathcal{H} \otimes \mathcal{H} : P\psi = -\psi\}. \quad \text{Fermions} \quad (4.45)$$

The fact that the particle interchange operator for Fermions changes ψ to $-\psi$ for Fermions should not cause you any worry as to whether interchanging particles gives you a different state, because , after all ψ and $-\psi$ define the same state: no measurement can distinguish them. You might want to review Table 4 and its “Advisement” following equation (4.4).

A very different example of this product rule arises when taking into account the idea that an electron, although regarded as a point, has some kind of “internal structure”. The Hilbert state space for an electron with spin is

$$\mathcal{H} = L^2(\mathbb{R}^3) \otimes \mathbb{C}^2. \quad (4.46)$$

The first factor reflects the fact that the classical configuration space for the electron is \mathbb{R}^3 . The second factor, which is just two dimensional, does not have any classical analog. What is it doing there? ² With it, you can explain the periodic table, the Zeeman effect, heat capacity of metals, etc. etc. Without it

The physical significance of the extra factor \mathbb{C}^2 in (4.46) can be understood in a way that is predictive of things to come, concerning the classification of elementary particles, that we will address later. We will continue this discussion later in the section on spin.

²1924-1926, Pauli, Goudsmit, Uhlenbeck (former father-in-law of our own Karen Uhlenbeck)

4.8 Observables from group representations

Symmetry is everywhere in physics. Symmetry usually means that there is a group acting on some set and that two points a and b are physically equivalent if there is a group element g such that $g(a) = b$. For example if the set consists of the positions and velocities five particles moving in \mathbb{R}^3 then the induced action of $SO(3)$ on this 30 dimensional space converts one such instantaneous state into another one with isomorphic physics because the laws of physics are rotation invariant. (If you want to include a gravitational force from the earth you must include the earth in the system.) The same invariance of classical mechanics holds also for the Euclidean group, which is six dimensional. You can also change the zero setting of your clock, thereby allowing the Galilean group, which is seven dimensional. If you want to do correct physics at very high speeds, however, you will have to replace the Galilean group by the Poincare group, which is ten dimensional. In this case you are into special relativity. See Section 9.6.

In quantum mechanics the isomorphisms are the unitary operators. Thus if \mathcal{H} is the Hilbert state space for a system and A is the operator representing an observable of interest, and if $U : \mathcal{H} \rightarrow \mathcal{K}$ is a unitary operator, then the Hilbert space \mathcal{K} can just as well be taken as the Hilbert state space of the system, provided you now represent the previous observable by the induced operator UAU^{-1} . The expected value $(A\psi, \psi) = (UAU^{-1}(U\psi), (U\psi))$ doesn't change in the corresponding state. Therefore quantum mechanical symmetry with respect to some group G usually means that there is a unitary representation of the group lurking nearby. Recall that a unitary representation of a group G is a homomorphism from G into the unitary group on some Hilbert space. If G is a Lie group then one has to assume that the homomorphism is continuous in some suitable sense.

The link between group representations and observables descends from the following theorem.

Theorem 4.7 (Stone-von Neumann theorem) *Let $t \mapsto U(t)$ be a function from \mathbb{R} into the group of unitary operators on a complex Hilbert space \mathcal{H} . Assume*

- a) $U(t+s) = U(t)U(s)$, for all $t, s \in \mathbb{R}$
- b) $t \mapsto U(t)\psi$ is a continuous function on \mathbb{R} to \mathcal{H} for each vector $\psi \in \mathcal{H}$.

Then there exists a unique self-adjoint operator A on \mathcal{H} such that

$$U(t) = e^{itA} \quad \text{for all } t \in \mathbb{R}. \tag{4.47}$$

Moreover A is given by

$$iA\psi = \frac{d}{dt} \Big|_{t=0} U(t)\psi \quad (4.48)$$

for all ψ in the domain of A .

Proof. Find it in any book that has “quantum” and “mathematicians” in the title. ■

Many of the fundamental observables that arise in quantum mechanics and quantum field theory are produced naturally by this theorem. We are going to show first how the angular momentum operator arises in exactly this way. In the next section we will do the same for spin.

4.8.1 Example: Angular momentum

The rotation group $SO(3)$ acts naturally on $L^2(\mathbb{R}^3)$ by the map

$$U_R : \psi \mapsto \psi \circ R^{-1} \quad \psi \in L^2(\mathbb{R}^3), \quad R \in SO(3) \quad (4.49)$$

Since a rotation preserves Lebesgue measure, and $U_R^{-1} = U_{R^{-1}}$, the map U_R is unitary. Moreover it is a simple exercise (if you pay a little attention to the definition) to show that $U_{R_1 R_2} = U_{R_1} U_{R_2}$. Therefore the map $R \mapsto U_R$ is a unitary representation of $SO(3)$ on $L^2(\mathbb{R}^3)$.

Consider the one-parameter group of rotations given by

$$R(\theta) = \begin{pmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (4.50)$$

By (4.49) these rotations act naturally on $L^2(\mathbb{R}^3)$ giving the following function of θ into L^2 ,

$$\theta \mapsto \psi \circ (R(\theta)^{-1}), \quad \theta \in \mathbb{R} \quad (4.51)$$

for each function $\psi \in L^2$.

Lets see what observable we get from this one-parameter group of unitaries by the recipe (4.48). Of course we replace t by θ in (4.48). Here is the computation.

$$\begin{aligned} \frac{\partial}{\partial \theta} \psi(R(\theta)^{-1}(x, y, z)) &= \frac{\partial}{\partial \theta} \psi(x \cos \theta + y \sin \theta, -x \sin \theta + y \cos \theta, z) \\ &= (-x \sin \theta + y \cos \theta) \psi_x + (-x \cos \theta - y \sin \theta) \psi_y \end{aligned}$$

So at $\theta = 0$ we find

$$\frac{\partial}{\partial \theta} \Big|_{\theta=0} \psi(R(\theta)^{-1}(x, y, z)) = (y \frac{\partial}{\partial x} - x \frac{\partial}{\partial y})\psi \quad (4.52)$$

Now look all the way back at Table 5 line 5., which describes the z component of the angular momentum operator. Back then we derived that operator from the classical angular momentum $x \times p$ by just making the usual replacement $p \rightarrow -i\hbar\nabla$. But now we see that we have arrived at the same operator from the action of the rotation group on $L^2(\mathbb{R}^3)$. Specifically, comparing with (4.52) we see that

$$(L_z\psi)((x, y, z)) = (-i)\hbar \frac{\partial}{\partial \theta} \Big|_{\theta=0} \psi(R(\theta)^{-1}(x, y, z)). \quad (4.53)$$

Lets put these pieces together. The exponential of the derivative gives back the one parameter group of unitaries in accordance with (4.47). We therefore have

$$e^{i\theta(L_z/\hbar)}\psi = \psi \circ R(\theta)^{-1} \quad \forall \theta \in \mathbb{R} \quad (4.54)$$

Furthermore, by (2.6), the one parameter group $R(\theta)$ itself is

$$R(\theta) = e^{\theta A_\omega}, \quad \omega = (0, 0, 1) \quad (4.55)$$

where A_ω is the element of the Lie algebra $so(3)$ given by (2.5) with $a = 1$.

SUMMARY:

a) If we write A_ω as the element (2.5) of the Lie algebra $so(3)$ which “generates” rotation around the z axis, i.e., (4.55) holds, and if L_z is the operator on $L^2(\mathbb{R}^3)$ corresponding to the observable “z-component of angular momentum” then

$$e^{i\theta(L_z/\hbar)}\psi = \psi \circ e^{-\theta A_\omega} \quad \forall \theta \in \mathbb{R} \quad (4.56)$$

and

$$(i/\hbar)L_z\psi = \frac{\partial}{\partial \theta} \Big|_{\theta=0} \psi \circ e^{-\theta A_\omega}. \quad (4.57)$$

b) Starting now with an arbitrary vector $\omega \in \mathbb{R}^3$ the previous computations show that the procedure of differentiating the one parameter group of unitaries $\theta \mapsto (\psi \rightarrow \psi \circ e^{-\theta A_\omega})$ at $\theta = 0$ gives a skew adjoint operator $(i/\hbar)L_\omega$ where L_ω is the angular momentum operator in the direction ω , (times $|\omega|$)

of course). We thereby have a linear map from the Lie algebra of $SO(3)$ to the (self-adjoint) angular momentum operators:

$$so(3) \ni A_\omega \mapsto L_\omega = \text{angular momentum operator in direction } \omega, \quad (4.58)$$

where $A_\omega x = \omega \times x$.

c) For any unitary representation of any Lie group G , Theorem 4.7 induces a map from \mathfrak{g} , the Lie algebra of G to operators on the representation Hilbert space. We will see over and over that this map assigns to the elements of certain Lie algebras, operators whose physical interpretation is central to quantum mechanics and quantum field theory. The next section will do this for electron spin.

4.9 Spin

We asserted in Section 4.7 that the Hilbert state space for an electron with spin is $\mathcal{H} \equiv L^2(\mathbb{R}^3) \otimes \mathbb{C}^2$. This is not a statement with content until one knows what operators on \mathcal{H} have what physical interpretation. Here is the physical significance of the factor \mathbb{C}^2 . Let $a \in SU(2)$. The homomorphism $\rho : SU(2) \rightarrow SO(3)$ described in Appendix 9.3 immediately gives us a representation of $SU(2)$ on \mathcal{H} by the formula

$$W(a) = U_{\rho(a)} \otimes a : L^2(\mathbb{R}^3) \otimes \mathbb{C}^2 \rightarrow L^2(\mathbb{R}^3) \otimes \mathbb{C}^2 \quad (4.59)$$

where U_R is the unitary operator on $L^2(\mathbb{R}^3)$ defined in (4.49). You should verify at your leisure (like now for example) that $W(ab) = W(a)W(b)$ and that $W(a)$ is actually unitary. So W is indeed a unitary representation of $SU(2)$. We are going to apply Theorem 4.7 and see what we get.

Let $s_z = (i/2) \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ as in Appendix 9.3. Let us compute the observable corresponding to the one parameter group of unitaries $\theta \mapsto W(e^{\theta s_z})$. We've already done the hard part in Section 4.8.1. The product rule for differentiation gives, in view of (4.57),

$$\begin{aligned} \frac{\partial}{\partial \theta} W(e^{\theta s_z}) &= \frac{\partial}{\partial \theta} \left(U_{\rho(e^{\theta s_z})} \otimes e^{\theta s_z} \right) \\ &= (i/\hbar) L_z \otimes I_{\mathbb{C}^2} + I_{L^2} \otimes s_z \end{aligned} \quad (4.60)$$

at $\theta = 0$. Therefore

$$-i\hbar \frac{\partial}{\partial \theta} \Big|_{\theta=0} W(e^{\theta s_z}) = L_z + (-i\hbar s_z). \quad (4.61)$$

CAUTION: I've omitted the identity operator factors in (4.61), as is common in the physics literature. But they should be there, just as in (4.60).

Define

$$J_z = L_z + (-i\hbar s_z), \quad \text{which operates on } L^2(\mathbb{R}^3) \otimes \mathbb{C}^2. \quad (4.62)$$

Terminology: J_z is called the (z component of the) *total angular momentum*. It is a sum of L_z , which is called the (z component of the) *orbital angular momentum*, and $-i\hbar s_z$, which is called the (z component of the) *spin angular momentum*. These names are themselves suggestive and therefore useful to know. By accepting J_z as the "(z component of) the quantum mechanical angular momentum operator", we acknowledge that the straight-forward quantization procedure, $p \mapsto -i\hbar\nabla, x \mapsto M_x$, which led us to define the angular momentum operator in Table 5 from the classical angular momentum $x \times p$, may not capture the physically correct operator corresponding to a given classical observable. In our present example there is simply no classical analog of a point particle of finite mass, for which spinning around its center makes sense.

Remarkably, the mechanism we illustrated in Section 4.8.1 for producing an angular momentum operator without spin, and in the present section for producing an angular momentum operator with spin, seems to apply to all of the fundamental observables in quantum mechanics and quantum field theory. One starts with some Lie group G and some unitary representation of G and uses the mechanism of Theorem 4.7 to map the Lie algebra of G to operators corresponding to some possibly non-classical observables. But which group G and which unitary representation? Ah! There's the rub. This brings us to the front lines of the classification of elementary particles. But we will develop this procedure in the context of relativity theory first, before going on to elementary particles.

SUMMARY:

Taking the viewpoint that the angular momentum operator should always be *defined* as the image of a Lie algebra element in some representation of the rotation group, or more generally its covering group $SU(2)$, we saw that the natural representation of $SU(2)$ on $L^2(\mathbb{R}^3) \otimes \mathbb{C}^2$ produces an angular momentum operator which is a sum of the classical angular momentum operator, coming from the motion of the point electron around the origin,

plus a term which seems to come from the point itself spinning around its center! Go figure.

But it works: if you place an atom in a magnetic field then the interaction of the magnetic field with the current (produced of course by the moving electron) changes the Hamiltonian and therefore the spectrum. This shows up experimentally in that some of the spectral lines change when a magnetic field is turned on around the atom. In fact where there was one spectral line before, there are now three, or five, or more. The change of Hamiltonian produced by the interaction of the magnetic field with the orbital angular momentum is adequate to explain the split into three lines (Zeeman effect) but not into five or more lines (anomalous Zeeman effect). However the spinning of the electron does the trick: the electron becomes a little magnet independently of its orbital motion. This magnet interacts with the magnetic field, changing the Hamiltonian even more. Result: spectral lines split, according to the modified Hamiltonian, (almost) precisely in agreement with experiment. Why (almost)? Because we have not yet taken special relativity into account.

A final word on terminology in the physics literature: ignore the orbital motion of the electron for the moment and consider just its internal Hilbert state space, which is \mathbb{C}^2 . Each of the three Hermitian spin angular momentum operators $-i\hbar s_x, -i\hbar s_y, -i\hbar s_z$ have two one dimensional eigenspaces in \mathbb{C}^2 . These eigenspaces are not in any sense parallel because these three operators don't commute. But in many computations it is convenient and customary to single out the z axis (e.g. one might take the magnetic field along the z axis). And then the z component of the spin angular momentum $-i\hbar s_z$ plays a special role. Since this operator is proportional to the matrix $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ its eigenvectors are $(1, 0)$ and $(0, 1)$, the first corresponding to the spin axis pointing up along the z axis and the second pointing down. Its customary to refer to the first as a "spin up" state and to the second as a "spin down" state. But in the absence of a computational or expositional reason to single out the z axis all directions in \mathbb{C}^2 are of equal physical significance.

4.10 Pictures: Heisenberg vs Schrödinger

If $\phi_t : T^*(C) \rightarrow T^*(C)$ is the one-parameter diffeomorphism group giving the Newtonian flow and b is a point in $T^*(C)$ then a classical observable

(which you may recall is specified by a function f on $T^*(C)$) takes the value $f(\phi_t(b))$ at time t if the state at time $t = 0$ was b . Mathematically there is no difference in saying that the observable itself is changing, that is; $f \rightarrow f \circ \phi_t$, while the state b is fixed, because after all $(f \circ \phi_t)(b) = f(\phi_t(b))$. There is a precise analog of these dual views in quantum mechanics. Suppose that H is the Hamiltonian operator on our Hilbert state space, \mathcal{H} . Lets put $\hbar = 1$ for readability. The time evolution of a state ψ_0 under the Schrödinger evolution is $\psi(t) = e^{-itH}\psi_0$. Suppose that A is a self-adjoint operator on \mathcal{H} corresponding to some observable. Then the expected value of A after time t is $(A\psi(t), \psi(t)) = (Ae^{-itH}\psi_0, e^{-itH}\psi_0) = (e^{itH}Ae^{-itH}\psi_0, \psi_0)$. So define

$$A_t = e^{itH}Ae^{-itH}. \quad (4.63)$$

Then $(A\psi(t), \psi(t)) = (A_t\psi_0, \psi_0)$. Since these inner products determine all the information one can get about the physics one could interpret this equation by saying that one might just as well regard the states as fixed while the observables change with time. A convenient mathematical nicety of this viewpoint is that the map $A \rightarrow A_t$ is a $*$ preserving automorphism of the algebra of all bounded operators on \mathcal{H} . The flow of diffeomorphisms in classical mechanics is thereby replaced by a flow of automorphisms of a non-commutative algebra. This is a technically and conceptually useful point of view, both for mathematical purposes and for quantum statistical mechanics, as it happens. The view presented in Table 4, in which the states ψ change in accordance with Schrödinger's equation, is called the "Schrödinger picture". The view presented in this section, in which the states ψ remain fixed and the observables change, $A \rightarrow A_t$, is called the "Heisenberg picture". These dual views of the time evolution are mathematically equivalent. But each viewpoint has advantages.

Lets consider a single particle free to move in one dimension, for simplicity. Suppose that it has mass m and is subject to a force $F = -V'(x)$, as usual. The Hilbert state space is $\mathcal{H} = L^2(\mathbb{R}, dx)$ and the Hamiltonian is of course

$$H = \frac{1}{2m}P^2 + M_V.$$

Lets put $\hbar = 1$ for readability again. Then

$$H = -\frac{1}{2m}d^2/dx^2 + M_V.$$

Define

$$Q(t) = e^{itH}Qe^{-itH} \quad (4.64)$$

and

$$P(t) = e^{itH} P e^{-itH}. \quad (4.65)$$

Theorem 4.8 (*Newton's equations in the Heisenberg picture.*) When applied to a wave function $\psi \in C_c^\infty(\mathbb{R})$ the previous operators satisfy the following equations.

$$m \frac{dQ(t)/dt}{dt} = P(t) \quad (4.66)$$

$$\frac{dP(t)/dt}{dt} = -V' \quad (4.67)$$

$$[P(t), Q(t)] = -iI_{\mathcal{H}}, \quad (\text{equal time CCR}) \quad (4.68)$$

Proof.

$$\frac{dQ(t)/dt}{dt} = e^{itH} i(HQ - QH) e^{-itH} \quad (4.69)$$

But, since multiplication operators commute,

$$(HQ - QH)\psi = -(1/2m)(d^2/dx^2 x\psi - xd^2/dx^2\psi) \quad (4.70)$$

$$= -(1/m)d\psi/dx. \quad (4.71)$$

This proves (4.66). Similarly, $\frac{dP(t)/dt}{dt} = e^{itH} i(HP - PH) e^{-itH}$. But, since P commutes with the kinetic energy term, we find

$$i(HP - PH)\psi = i(V(x)(-id/dx) - (-id/dx)V(x))\psi(x) \quad (4.72)$$

$$= -V'(x)\psi(x). \quad (4.73)$$

This proves (4.67). Finally,

$$(P(t)Q(t) - Q(t)P(t))\psi = e^{itH}(-id/dx x - x(-id/dx))e^{-itH}\psi \quad (4.74)$$

$$= e^{itH}(-iI_{\mathcal{H}})e^{-itH} \quad (4.75)$$

$$= -iI_{\mathcal{H}} \quad (4.76)$$

■

SUMMARY: As we know, with $p = mv$, Newton's equations take the form $dp/dt = -V'(x)$ when the force is given as $F(x) = -V'(x)$. The equations (4.66) and (4.67) are therefore precisely Newton's equations, but for *operator valued* functions of time. The (equal time) canonical commutation relations (4.68) show, however, that these are not identifiable with classical solutions. Some would say that the commutation relations, (4.68), are the real difference between classical mechanics and quantum mechanics.

4.11 Conceptual status of quantum mechanics

Although quantum mechanics gives fantastically good agreement with experiment in all areas of physics, its conceptual foundations are regarded by many (mostly mathematicians) as incomplete and imprecise. In his classical book [60], von Neumann proposed a mathematical interpretation of the measurement process. If, for example, a wave function is a linear combination of two (normalized) eigenstates ψ_1 and ψ_2 of the Hamiltonian with distinct eigenvalues a and b , say $\psi = 2^{-1/2}\psi_1 + 2^{-1/2}\psi_2$, then a measurement of the system will find the system to be either in state ψ_1 or in state ψ_2 , each with probability 1/2. (You must make the measurement many times, starting with the state ψ each time.) But after the measurement the system will remain in state ψ_1 , if that's what you found. Thus in the measurement process the system has undergone a sudden change from ψ to ψ_1 . This is called collapse of the wave function. To avoid possible inconsistencies associated with collapse of the wave function some leading physicists endorse the notion of “decoherence”. A fast exposition of this notion and lots of references can be found on the Wikipedia website

<http://en.wikipedia.org/wiki/Decoherence>

There is a vast literature aiming to put quantum mechanics on a well motivated and clearly organized axiomatic foundation. The algebra of bounded observables is no longer commutative, as it is in classical mechanics. Moreover the lattice of “Question observables” (for which the answer is yes or no) is very different in classical mechanics from that in quantum mechanics. In classical mechanics the typical question is “Does the system lie in the Borel set $B \subset T^*(C)$?”. These questions form a lattice (a partially ordered set in which any two elements have at least one upper bound. E.g. $B_1 \leq B_2$ means $B_1 \subset B_2$. Clearly any two elements have a supremum, namely $B_1 \cup B_2$.) In quantum mechanics the question observables are specified by projections onto subspaces of the state Hilbert space \mathcal{H} . ($P_1 \leq P_2$ means the usual.) The properties of these two lattices are very different. One can impose one's favorite conditions on an abstract lattice and ask what conditions imply that the lattice is just like the quantum mechanical one. Axiomatic approaches based on such lattices are ubiquitous. One approach, containing some of this idea, but augmented by more concrete axioms can be found in the paper of George Mackey [41].

The earliest attempt to formulate mechanics in terms of the algebra of

observables, from which one could hope to show that the only examples are classical mechanics or quantum mechanics, is the C^* algebra approach of I. E. Segal [54]. Subsequent work showed that examples other than these two could satisfy the axioms. But Segal's approach was later adapted to form a useful conceptual and technical tool in quantum field theory.

Another approach, that goes under the name of “Geometric quantization” asks how one might map the space of functions on phase space, $T^*(C)$, into corresponding operators on the state Hilbert space \mathcal{H} . The classic book on this approach is by Woodhouse [70], from which one can trace back to the fundamental work of B. Kostant. For a recent survey of this approach see B. Hall, [29]

Perhaps quantum mechanics is just an “averaging” over some deeper theory in which there are quantities, “hidden variables”, that we just don't know how to measure as yet. This is a very extensively investigated question. See for example the book [47] by E. Nelson on stochastic quantum mechanics.

There is a recent work by John Conway and Simon Kochen, which relates to hidden variables theory and connects these issues to the classical Free Will question! See for example

<http://www.ams.org/notices/200902/rtx090200226p.pdf>

Research on the conceptual and axiomatic foundations of quantum mechanics is an ongoing activity. For a lead-in to some of this research see, for example, J.M. Jauch [38] or [5, 6, 7]. And/or type in [Title] ⟨The interpretation of quantum mechanics⟩ on mathscinet. This will bring up (at least) 44 references.

4.12 References

Treatments of the basic concepts of quantum mechanics, aimed at mathematicians can be found in

Gerald Folland, *Quantum field theory, A tourist guide for mathematicians.*[21]

Brian Hall, *An introduction to quantum theory for mathematicians.* [28].

An updated version of Brian Hall's forthcoming book can be downloaded from the 7120 website. These two books give a more serious and detailed treatment of many of the topics we've covered.

Victor Guillemin (the father of our Victor Guillemin) “The story of quantum mechanics” [27]. This is a fun book to read.

B.L. van der Waerden has a source book in which some of the original papers of Heisenberg et al are reprinted. van der Waerden also makes insightful commentaries of his own, concerning what happened in those days, the 1920's. Source book: [61]

Hermann Weyl, in 1928, was already into quantum mechanics. His classic book “Gruppentheorie und Quantenmechanik” has been revised several times since the 1928 edition. It was translated into english in 1931 and is now available in a Dover edition. [65].

Jagdish Mehra and Helmut Rechenberg, [45], have written 2000 pages of recollections, gossip, and a record of the changing viewpoints of the founders of quantum mechanics. This six volume set also contains detailed expositions of many of the fundamental papers. They explain the relation between the contributions of the founders and the influence of one on the other. This work is based on extensive interviews with the founders as well as on the papers themselves.

5 Quantum field theory

We are going to construct the simplest quantum field. Our aim is to illustrate not only what the typical structure of a quantum field is, but to show also how a “free” quantum field can be regarded as a sequence of harmonic oscillators. The heuristics behind this “equivalence” is itself extremely illuminating and still dominates a large part of the textbook literature on quantum field theory. The heuristics includes the first, and most easily understood, subtraction of an infinite quantity. Such subtractions (so-called renormalizations) are ubiquitous in quantum field theory. Many physicists regard them (at the present time, 2011) as vital ingredients of the theory, rather than an artifact of the present day formulation. In the preface to his reprint book of the classical papers, “Quantum electrodynamics” [53, page xv], Julian Schwinger points out that “it took the labors of more than a century to develop the methods that express fully the mechanical principles laid down by Newton.” Now quantum field theory is barely 80 years old, and is a much more complicated mathematical system than Newtonian mechanics. Even ignoring the possibility of major fundamental modifications of the present version of quantum field theory, such as e.g. string theory, which may be required for understanding higher energy phenomena, 80 years is a pretty short time to get the mathematics straight. Moreover it may be impossible to

get the present versions of quantum field theory into a mathematically consistent form without including future modifications. Will another hundred years do the trick? Or is that too optimistic.

5.1 The harmonic oscillator

One harmonic oscillator. The restoring force for a harmonic oscillator on the line is given by Hooke's law $F(x) = -kx$, as already mentioned in Example 2.9. Newton's equation, $F = ma$, reduces, in this case to

$$-kx = m\ddot{x} \quad (5.1)$$

where x is the oscillator position, $\dot{x} = dx/dt$, and k is the spring constant. The force F is $-\text{grad } V$, with $V = \frac{k}{2}x^2$. The energy of the oscillator is therefore

$$E = \frac{1}{2}m\dot{x}^2 + \frac{k}{2}x^2 \quad (5.2)$$

or, equivalently,

$$E = \frac{1}{2m}p^2 + \frac{k}{2}x^2, \quad (5.3)$$

where $p = mv = m\dot{x}$ is the momentum. Define $\omega = \sqrt{k/m}$. Then Newton's equation, (5.1), reads $\ddot{x} + \omega^2x = 0$, for which the general solution is

$$x(t) = A \cos \omega t + B \sin \omega t. \quad (5.4)$$

So $\nu \equiv \omega/(2\pi)$ is the frequency of this periodic motion. The angular frequency ω is of more interest to us than the spring constant. So we will always write harmonic oscillator equations in terms of ω . Thus, since $k = m\omega^2$, we will write the energy (5.3) as

$$E = \frac{1}{2m}p^2 + \frac{m\omega^2}{2}x^2, \quad (5.5)$$

Quantization of this system yields the Hamiltonian

$$H = -\frac{1}{2m} \frac{d^2}{dx^2} + \frac{m\omega^2}{2}x^2, \quad (5.6)$$

which is to be interpreted as a self-adjoint operator on $L^2((-\infty, \infty); dx)$ (and which is obtained from (5.3) by the usual substitution $p \rightarrow \frac{1}{i}\frac{\partial}{\partial x}$, $x \rightarrow \text{mult. by } x$). We are going to set $\hbar = 1$ in this section.

n harmonic oscillators. Similarly the energy and Hamiltonian for a system consisting of n independent harmonic oscillators of masses m_j and spring constants k_j may be derived from the corresponding Newton's equations,

$$m_j \frac{d^2x_j}{dt^2} = -k_j x_j \quad j = 1, \dots, n. \quad (5.7)$$

The total energy is

$$E = \sum_{j=1}^n \left(-\frac{1}{2m_j} p_j^2 + \frac{m_j \omega_j^2}{2} x_j^2 \right), \quad (5.8)$$

wherein we have again expressed the energy in terms of the angular frequency $\omega_j = \sqrt{k_j/m_j}$. The Hamiltonian is therefore given by

$$H = \sum_{j=1}^n \left(-\frac{1}{2m_j} \frac{\partial^2}{\partial x_j^2} + \frac{m_j \omega_j^2}{2} x_j^2 \right). \quad (5.9)$$

H is to be interpreted as a self-adjoint operator on a suitable domain in $L^2(R^n; d^n x)$ (e.g., it can be defined as the closure of its restriction to $C_c^\infty(R^n)$).

5.2 A quantized field; heuristics.

We will show now how the quantization of a classical (noninteracting) field can be regarded as just an assembly of infinitely many harmonic oscillators.

To avoid technical problems that will only obscure the big ideas, we are going to consider only a field that has just one component (not six like the electromagnetic field), that lives on a one dimensional space (not on \mathbb{R}^3), and that lives in fact on just a finite interval (not all of \mathbb{R}). In the absence of any charges or currents, which is the case of interest to us, any component of the electromagnetic field satisfies the wave equation $\frac{\partial^2 u}{\partial t^2} = \Delta u$. It is this feature, a hyperbolic wave equation, that must be preserved in any reasonable example. Whereas quantum mechanics arises from quantizing an ordinary differential equation, namely Newton's equation, quantum field theory arises from quantizing a partial differential equation (the field equation).

We are going to quantize the “field equation”

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2}, \quad (5.10)$$

which is the equation for a vibrating string. Moreover we will take space to be the interval $(0, \pi)$ and assume that the string is fixed at the endpoints. I.e.,

$$u(0, t) = 0 \quad u(\pi, t) = 0 \quad \text{for all } t. \quad (5.11)$$

In accordance with the method of separation of variables a general solution of (5.10) and (5.11) can be written in the form

$$u(x, t) = \sum_{j=1}^{\infty} q_j(t) u_j(x) \quad (5.12)$$

where $u_j(x)$ and $q_j(t)$ each satisfy ordinary differential equations and u_j is zero at the endpoints. The ordinary differential equation for u_j is the eigenvalue equation $d^2u/dx^2 = \lambda u(x)$ with $u(0) = u(\pi) = 0$. The eigenfunctions, normalized in $L^2((0, \pi))$, are

$$u_j(x) = \sqrt{\frac{2}{\pi}} \sin jx \quad j = 1, 2, \dots \quad (5.13)$$

The eigenvalues are $\lambda = -j^2$, $j = 1, 2, \dots$. That is, $u_j'' = -j^2 u_j$. The functions u_j form an orthonormal basis of $L^2((0, \pi))$ and any initial condition can therefore be written as

$$u_0(x) = \sum_{j=1}^{\infty} q_j u_j(x) \quad (5.14)$$

for some suitable choice of the real numbers q_j , which we leave unspecified for now. The time evolution of the initial data u_0 is found, as usual, by letting the coordinates q_j depend on time as in equation (5.12). Then the field equation (5.10) reads

$$\begin{aligned} \sum_{j=1}^{\infty} \ddot{q}_j(t) u_j(x) &= \partial^2 u(x, t) / \partial t^2 \\ &= \partial^2 u(x, t) / \partial x^2 = \sum_{j=1}^{\infty} q_j(t) u_j''(x) \\ &= - \sum_{j=1}^{\infty} j^2 q_j(t) u_j(x). \end{aligned}$$

Since the functions u_j are orthonormal it follows that

$$\frac{d^2q_j}{dt^2} = -j^2 q_j \quad j = 1, 2, \dots . \quad (5.15)$$

Of course $u(x, t)$ is determined by the values $u(x, 0)$ and $\dot{u}(x, 0)$ which, in view of (5.12), are determined by the two sequences $\{q_j(0)\}$ and $\{\dot{q}_j(0)\}$.

Comparison of equations (5.15) with (5.7) shows that the boundary value problem (5.10), (5.11) is equivalent to a mechanical system consisting of infinitely many harmonic oscillators with masses m_j all equal to one and spring constants $k_j = j^2$. The canonical coordinates are $\{q_j\}_{j=1}^\infty$ and $\{p_j = m_j \dot{q}_j = \dot{q}_j\}_{j=1}^\infty$.

According to the principles of quantum mechanics one should therefore quantize this system by taking the Hilbert space to be

$$\mathcal{K} = L^2(\mathbb{R}^\infty, dq_1 dq_2 \cdots) \quad (5.16)$$

and the Hamiltonian to be (in analogy to (5.9) with all $m_j = 1$)

$$H = \sum_{j=1}^{\infty} \frac{1}{2} \left(-\frac{\partial^2}{\partial q_j^2} + \omega_j^2 q_j^2 \right) \quad (5.17)$$

where $\omega_j = j$. Moreover, if Q_j denotes the quantization of q_j , i.e. $Q_j =$ multiplication by q_j on \mathcal{K} , the principles of quantum mechanics assert, in view of (5.12), that the quantized field φ at time zero is given by

$$\varphi(x, 0) = \sum_{j=1}^{\infty} Q_j u_j(x). \quad (5.18)$$

The field is therefore an operator valued function on $(0, \pi)$. That is to say, the field is an operator on \mathcal{K} for each $x \in (0, \pi)$.

There are a few little problems with the preceding discussion.

First, the definition of \mathcal{K} in (5.16) is meaningless because of the appearance of infinite dimensional Lebesgue measure. *Don't worry.* We will change the Lebesgue measure to a well defined probability measure in a natural way.

Second, the series defining the operator H in (5.17) diverges to $+\infty$ in every reasonable sense. *Don't worry.* We will subtract off an infinite constant that will make H a meaningful operator on a meaningful Hilbert space

Third, the series (5.18) defining the quantum field $\varphi(x)$ diverges for almost every $x \in (0, \pi)$. *Don't worry.*

SUMMARY: The astute reader can very likely appreciate now the word “heuristics” in the title of this subsection. The heuristic discussion of the present section still remains today a good guide for the construction, and even definition, of what constitutes a quantum field. Precise and meaningful definitions of “quantum field” have been given and explored extensively. See e.g. Streater and Wightman [56]. Studies dealing with mathematically meaningful definitions of quantum fields go by the name “axiomatic quantum field theory”. Unfortunately no axioms have yet been proposed which seem likely (to many) to be satisfied by the actual fields (gauge fields) that have been most successfully corroborated by experiment. For the time being, therefore, it's important to understand the heuristic basis for our current example as well as a meaningful version of it, to which we now turn. In order to rescue the preceding heuristic discussion of our quantized field we are going to return temporarily to harmonic oscillators. In one form or another the following rescue operation still appears in modern textbooks on quantum field theory for physicists.

5.3 The ground state transformation

One harmonic oscillator again. Consider a single harmonic oscillator with $m = 1$ and $k = \omega^2$. Its Hamiltonian, according to (5.6), is

$$\hat{H} = -\frac{1}{2m} \frac{d^2}{dq^2} + \frac{1}{2}m\omega^2q^2 \quad \text{on } L^2((-\infty, \infty); dq). \quad (5.19)$$

For any constant $a > 0$ the easily verified identity

$$\hat{H}e^{-ax^2/2} = \frac{a}{2m}e^{-ax^2/2} + \left(\frac{m\omega^2}{2} - \frac{a^2}{2m}\right)x^2e^{-ax^2/2} \quad (5.20)$$

shows that $e^{-ax^2/2}$ is an eigenfunction of \hat{H} if and only if $a = m\omega$. Let

$$\psi_0(q) = (m\omega/\pi)^{1/4}e^{-m\omega q^2/2}. \quad (5.21)$$

Then (5.20) shows that

a) $\hat{H}\psi_0 = (\omega/2)\psi_0$

Moreover an easy Gaussian integral shows that

$$b) \int_{-\infty}^{\infty} \psi_0(q)^2 dq = 1.$$

So ψ_0 is a normalized eigenfunction for \hat{H} and, since ψ_0 is positive, $\omega/2$ is the lowest eigenvalue of \hat{H} and ψ_0 is the ground state of \hat{H} . A direct computational proof that $\omega/2$ is the lowest eigenvalue will be given in the next theorem.

Let

$$d\mu(q) = \psi_0(q)^2 dq. \quad (5.22)$$

Then μ is a probability measure on R . Define

$$U : L^2(R^n, dq) \rightarrow L^2(R^n, \mu)$$

by

$$(U\psi)(q) = \psi(q)/\psi_0(q). \quad (5.23)$$

U is clearly unitary. We assert that

$$U\hat{H}U^{-1} = \left(-\frac{1}{2m} \frac{d^2}{dq^2} + \omega q \frac{d}{dq} \right) + \frac{\omega}{2} I. \quad (5.24)$$

Indeed, let $f(x) = (U\psi)(x) = \psi(x)/\psi_0(x)$. Then $(U^{-1}f)(x) = \psi(x) = f(x)\psi_0(x)$ and therefore

$$(\hat{H}\psi)(x) = f(x)\hat{H}\psi_0 - \frac{1}{2m} \left(f''(x)\psi_0(x) + 2f'(x)\psi'_0(x) \right) \quad (5.25)$$

$$= f(x) \frac{\omega}{2} \psi_0(x) - \frac{1}{2m} \left(f''(x) - 2f'(x)(m\omega x) \right) \psi_0(x) \quad (5.26)$$

Hence $(U\hat{H}U^{-1}f)(x) = (\hat{H}\psi)(x)/\psi_0(x) = (\omega/2)f(x) - \frac{1}{2m}f''(x) + \omega x f'(x)$, which is (5.24).

Define

$$H = -\frac{1}{2m} \frac{d^2}{dq^2} + \omega q \frac{d}{dq} \quad \text{on } L^2(R, \mu) \quad (5.27)$$

Then (5.24) asserts that

$$H = U(\hat{H} - \frac{\omega}{2})U^{-1} \quad (5.28)$$

Theorem 5.1

$$(Hf, g)_{L^2(\mu)} = \frac{1}{2m} \int_R f'(q) \bar{g}'(q) d\mu(q) \quad \text{and} \quad (5.29)$$

$$(Hf, f)_{L^2(\mu)} \geq 0 \quad \text{for all sufficiently smooth } f \in L^2(\mu). \quad (5.30)$$

Moreover $\inf(\text{spectrum } H) = 0$ and $\inf(\text{spectrum } \hat{H}) = \omega/2$.

Proof. In view of the definition (5.21) an integration by parts gives

$$\begin{aligned} \int_R f'(q) \bar{g}'(q) d\mu(q) &= \int_R f'(q) \bar{g}'(q) \psi_0(q)^2 dq \\ &= \int_R (-f''(q) + 2m\omega q f'(q)) \bar{g}(q) \psi_0(q)^2 dq \\ &= 2m(Hf, g)_{L^2(\mu)}. \end{aligned}$$

This proves (5.29). Put $g = f$ in (5.29) to deduce (5.30). Since the constant functions are in $L^2(\mu)$ and $H1 = 0$ the bottom of the spectrum is indeed zero and is an eigenvalue. Finally (5.28) shows that $\hat{H} = U^{-1}HU + \omega/2$. Since $U^{-1}HU$ has the same spectrum as H this completes the proof. ■

Remark 5.2 The operator U commutes with multiplication, as is clear from the simple formula (5.23). For example

$$UM_qU^{-1} = M_q. \quad (5.31)$$

(The domain of M_q has changed however. But never mind that.) But U does not commute with differentiation. Thus if $f(q) = \psi(q)/\psi_0(q)$ then $(U^{-1}f)(q) = \psi(q) = f(q)\psi_0(q)$ and therefore

$$(U \frac{d}{dq} U^{-1} f)(q) = U \left(\frac{d}{dq} (f \cdot \psi_0) \right) \quad (5.32)$$

$$= \left(f' \psi_0 + f \psi'_0 \right) / \psi_0 \quad (5.33)$$

$$= f' - \omega q \quad (5.34)$$

Therefore our momentum operator, which was given by $P = -i\hbar(d/dq)$ when we used Lebesgue measure to define the Hilbert state space is now given by

$$P = -i\hbar \left(\frac{d}{dq} - \omega q \right) \quad (5.35)$$

as an operator in $L^2(\mathbb{R}; \gamma_\omega)$. That's what happens when you change the representation of the Hilbert state space. we saw a more drastic change in the formula for the momentum operator when we transformed the Hilbert state space by the Fourier transform. P just became a multiplication operator. There is a lesson here: In spite of the changed appearance of the formulas for P and Q , the all important commutation relations $[P, Q] = -iI$ are unchanged.

There are now five naturally occurring operators in this “ground state” representation of the Hilbert state space for a harmonic oscillator. The relations between them dominate a large part of quantum field theory. Define

$$\alpha = \frac{d}{dq} \quad \text{acting on } L^2(\mathbb{R}; \gamma_\omega) \quad (5.36)$$

Then an integration by parts shows, just as in the proof of (5.29), that

$$\alpha^* = -\frac{d}{dq} + 2\omega q \quad (5.37)$$

Consequently, in view of (5.35) and (5.27), we have the relations

$$\alpha + \alpha^* = 2\omega Q, \quad (5.38)$$

$$\alpha - \alpha^* = 2(i/\hbar)P \quad \text{and} \quad (5.39)$$

$$(1/2)\alpha^*\alpha = H. \quad (5.40)$$

In short, all of the operators we have discussed so far can be expressed neatly in terms of the differentiation operator α and its adjoint in the Hilbert space $L^2(\mathbb{R}; \gamma_\omega)$.

SUMMARY: We see that $\omega/2$ is the bottom of the spectrum of the original Hamiltonian \hat{H} that we've been using in preceding sections. If we lower the entire spectrum by subtracting off the so called “zero point energy”, $\omega/2$ then the resulting operator is unitarily equivalent to the operator H . The quadratic form of H has the really simple formula given in (5.29). An operator whose quadratic form has this nice first derivative look is called a Dirichlet form operator. The description (5.29) is equivalent to writing H in the form (5.40). Such operators are especially easy to deal with. Now notice that subtracting the constant $\omega/2$ from the Hamiltonian \hat{H} does not change the physics because this just amounts to changing the potential $m\omega^2q^2/2$ by

an additive constant. Moreover a unitary transformation is an isomorphism for quantum mechanics and therefore the operator H represents the harmonic oscillator Hamiltonian just as well as our original choice \hat{H} does.

n harmonic oscillators again. Let us return now to a system of n independent harmonic oscillators all of mass one and spring constants ω_j^2 . By (5.9) we find the Hamiltonian in this case to be

$$\hat{H}_n = (1/2) \sum_{j=1}^n (-\partial^2/\partial q_j^2 + \omega_j^2 q_j^2) \quad \text{acting in } L^2(\mathbb{R}^n, dq) \quad (5.41)$$

Let

$$\psi_j(q_j) = (\omega_j/\pi)^{1/4} e^{-\omega_j q_j^2/2}$$

and

$$\psi_0(q) = \prod_{j=1}^n \psi_j(q_j), \quad q = (q_1, \dots, q_n)$$

Then

$$\hat{H}_n \psi_0 = \left[(1/2) \sum_{j=1}^n \omega_j \right] \psi_0 \quad (5.42)$$

because \hat{H}_n is a sum of independent differential operators. So ψ_0 is the ground state of \hat{H}_n and the “zero point energy” (i.e. inf spectrum \hat{H}_n) is

$$E_n \equiv (1/2) \sum_{j=1}^n \omega_j. \quad (5.43)$$

Define

$$d\mu_n(q) = \prod_{j=1}^n [\psi_j(q)^2 dq_j] \quad \text{on } \mathbb{R}^n \quad (5.44)$$

Then μ_n is a probability measure on \mathbb{R}^n and the map

$$\psi \rightarrow U_n \psi = \psi / \psi_0 \quad (5.45)$$

is again unitary from $L^2(\mathbb{R}^n, dq)$ onto $L^2(\mathbb{R}^n, \mu_n)$. Just as for a single oscillator one sees that

$$U_n (\hat{H}_n - E_n) U_n^{-1} = \sum_{j=1}^n \left(-(1/2) \partial^2 / \partial q_j^2 + \omega_j q_j \partial / \partial q_j \right). \quad (5.46)$$

So define

$$H_n = \sum_{j=1}^n \left(-\frac{1}{2} \frac{\partial^2}{\partial q_j^2} + \omega_j q_j \frac{\partial}{\partial q_j} \right) \quad \text{acting in } L^2(\mathbb{R}^n, \mu_n). \quad (5.47)$$

Then

$$U_n \hat{H}_n U_n^{-1} = H_n + E_n. \quad (5.48)$$

Let us simply record here the obvious extension of the formulas (5.35) - (5.40) to n harmonic oscillators. Define

$$\alpha_j = \frac{\partial}{\partial q_j} \quad \text{acting on } L^2(\mathbb{R}^n; \mu_n) \quad (5.49)$$

Then

$$\alpha_j^* = -\frac{\partial}{\partial q_j} + 2\omega_j q_j \quad (5.50)$$

$$\alpha_j + \alpha_j^* = 2\omega_j Q_j \quad (5.51)$$

$$\alpha_j - \alpha_j^* = 2(i/\hbar) P_j \quad (5.52)$$

$$(1/2) \sum_{j=1}^n \alpha_j^* \alpha_j = H \quad (5.53)$$

SUMMARY: H_n differs from our first choice of Hamiltonian, (5.41) in two ways. First, it incorporates the subtraction of the zero point energy E_n , which does not affect the physics because it just changes the potential by an additive constant. Second, it incorporates the unitary transform U_n , which also does not affect the physics. Thus the assembly of quantum harmonic oscillators are just as well served by H_n acting in $L^2(\mathbb{R}^n; \mu_n)$ as by \hat{H}_n acting in $L^2(\mathbb{R}^n; d^n q)$.

We are ready to return to an infinite assembly of harmonic oscillators.

5.4 Back to the quantized field.

We can see now, from (5.41) and (5.43) that the operator H “defined” in (5.17) is bounded below by $(1/2) \sum_{j=1}^{\infty} \omega_j$, which diverges because $\omega_j = j$ in that field theory. So we can say, informally, that $H \geq +\infty$, which shows that H is at best meaningless. The customary resolution of this problem goes like this. Any potential is defined only up to an additive constant because only

$\text{grad } V$ has direct physical meaning. Thus there is no change in the physics if we subtract the infinite constant $(1/2) \sum_{j=1}^{\infty} \omega_j$ from the operator H . We may therefore attempt to give the infinite sum in (5.17) a mathematical interpretation by passing to the limit $n \rightarrow \infty$ in the system (5.47), which after all is physically equivalent to the system (5.41).

To this end note first that the measure μ_n in (5.44) is just a product of probability measures, an infinite product of which is a perfectly respectable measure. Define

$$d\mu(q) = \prod_{j=1}^{\infty} [\psi_j(q_j)^2 dq_j] \quad \text{on } R^{\infty} \quad (5.54)$$

Let

$$H = \sum_{j=1}^{\infty} \left(-\frac{1}{2} \frac{\partial^2}{\partial q_j^2} + \omega_j q_j \frac{\partial}{\partial q_j} \right) \quad \text{acting in } L^2(R^{\infty}, \mu). \quad (5.55)$$

Since there are no zeroth order terms in H (unlike (5.41)) the infinite sum makes perfectly good sense when applied to many functions on the infinite product space. For example if $f : \mathbb{R}^{\infty} \rightarrow \mathbb{C}$ depends on only finitely many coordinates, say $f(q) = g(q_1, \dots, q_n)$, with $g \in C_c^{\infty}(\mathbb{R}^n)$, then there are only finitely many nonzero terms in the sum for Hf . Such functions are dense in $L^2(\mathbb{R}^{\infty}, \mu)$. So H is densely defined by the formula (5.55). Thus by subtracting the (infinite) zero point energy from (5.17) and unitarily transforming to the ground state measure μ we have given a meaning to (5.17) as a densely defined operator in $L^2(\mathbb{R}^{\infty}, \mu)$, namely as the operator H defined in (5.55). Moreover we have also “given” a meaning to the Hilbert space \mathcal{K} “defined” in (5.16) since \mathcal{K} can just be replaced (informally) by the well defined Hilbert space $L^2(\mathbb{R}^{\infty}, \mu)$. After all, the latter is just the image under the “unitary operator $\lim_{n \rightarrow \infty} U_n$ ” (which however makes no sense because its domain is meaningless). (If you are getting the feeling that the boundary between heuristics and theorems is being stretched you’re right. But you have to get used to it. The infinite subtraction and concomitant change of Hilbert space description is among the mildest in this subject.) We have now solved the first two of the three little problems raised at the end of Section 5.2. Before addressing the third lets understand the operator H a little better and the measure μ .

The operator H is particularly nicely related to the measure μ . Just as in the one harmonic oscillator case, (5.29), the operator H_n in (5.47)

clearly satisfies

$$(H_n f, g)_{L^2(\mu_n)} = (1/2) \int_{\mathbb{R}^n} (\nabla f \cdot \nabla \bar{g}) d\mu \quad (5.56)$$

With just a little thought you can see that this persists in the limit as $n \rightarrow \infty$ for functions f and g which depend on only finitely many coordinates. Thus for such a dense set of functions we have

$$(Hf, g)_{L^2(\mathbb{R}^\infty, \mu)} = (1/2) \int_{\mathbb{R}^\infty} \sum_{j=1}^{\infty} (\partial f / \partial q_j)(\partial \bar{g} / \partial q_j) d\mu(x) \quad (5.57)$$

If one wished to pursue the functional analytic questions left untouched in this discussion, such as whether H actually has a self-adjoint version in $L^2(\mu)$, the description of H as the operator associated to a Dirichlet form, as in (5.57), offers a quick and easy mechanism to prove such nice properties because Dirichlet forms in finite and infinite dimensions are well understood. But we won't carry this out here.

To connect with parts of the physics literature it is illuminating to deal in the following informal way with the measure μ defined in (5.54). Write $Dq = \prod_{j=1}^{\infty} dq_j$ for Lebesgue “measure” on \mathbb{R}^∞ . Then (5.54) gives

$$d\mu(x) = [\prod_{j=1}^{\infty} (\omega_j / \pi)^{1/2}] [e^{-\sum_{j=1}^{\infty} \omega_j q_j^2}] Dq. \quad (5.58)$$

In the case of primary interest to us we have $\omega_j = j$. So the first of the three factors in (5.58) is infinite. The second factor happens to be zero a.e. with respect to μ . And the third factor is meaningless. It is therefore particularly edifying that the product of all three factors makes perfectly good sense as a measure on \mathbb{R}^∞ . The re-association of factors from (5.54) to (5.58) converts a meaningful probability measure μ into a meaningless expression (5.58). But get used to it. In the physics literature this measure is usually written as (5.58).

Returning now to the informally quantized field $\varphi(x)$ defined in (5.18) here is how we will give it a well defined meaning. Recall that in defining the classical electric field $E(x)$ one must, in principle, place a small charged piece of matter around the point $x \in \mathbb{R}^3$ and measure the force on the object. This gives the average force on the object. One must then, in principle, repeat this measurement with a smaller object, a ball centered at x say, and take a sequence of measurements as the radius of the ball goes to zero while the

charge also goes to zero. This way you get the field value E exactly at x and uninfluenced by the change of the sources of the field that may be produced by the testing charge itself. This is an OK viewpoint in classical mechanics. But in the actual world you can't take the charge to zero because the minimum charge is the charge on an electron. And of course you can't take the radius of the ball smaller than the radius of an atom (or proton, or quark if you want to stretch it.) One can only hope that the averages themselves are good enough to make a well defined theory. We are going to carry this out in our vibrating string example.

Let $f \in C_c^\infty((0, \pi))$ be real valued. Think of f as the density of charge on the little test ball (test interval in our case). We want to give meaning to $\int_0^\pi \varphi(x)f(x)dx$ when φ is given by (5.18). Interchanging the integral with the sum we *define*

$$\varphi_f = \sum_{j=1}^{\infty} Q_j \int_0^\pi f(x)u_j(x)dx \quad (5.59)$$

We will show that this series of operators on $L^2(\mathbb{R}^\infty; \mu)$ converges in a reasonable sense. And then we may write, informally,

$$\int_0^\pi f(x)\phi(x)dx = \varphi_f. \quad (5.60)$$

Since the series defining $\varphi(x)$ does not actually converge, the integrand on the left is meaningless. But the operator on the right is well defined. This circumstance is to be interpreted by saying that the quantized field $\varphi(x)$ is not an operator valued function after all but rather an operator valued distribution. This is typical for quantized fields.

Theorem 5.3 *Let f_j be a sequence of real numbers. For $\psi \in L^\infty(\mathbb{R}^\infty; \mu)$ define*

$$(Q_j\psi)(q_1, q_2, \dots) = q_j\psi(q_1, q_2, \dots). \quad (5.61)$$

Then the series

$$\sum_1^{\infty} Q_j f_j \psi \quad (5.62)$$

converges in $L^2(\mathbb{R}^\infty; \mu)$ for all $\psi \in L^\infty$ if and only if

$$\sum_{j=1}^{\infty} f_j^2 / \omega_j < \infty \quad (5.63)$$

In particular, if $f \in C_c^\infty((0, \pi))$ then the series in (5.59) converges on the dense set $L^\infty(\mathbb{R}^\infty, \mu)$.

Proof. Each coordinate function q_j is a mean zero Gaussian random variable with

$$\|q_j\|_{L^2(\mu)}^2 = \frac{1}{2\omega_j}. \quad (5.64)$$

Since they are also independent we have

$$\left\| \sum_{j=1}^n q_j f_j \right\|_{L^2(\mu)}^2 = \sum_{j=1}^n f_j^2 / (2\omega_j). \quad (5.65)$$

Since any series of orthogonal functions converges in L^2 if and only if the series of their square norms converges, it follows that the series $\sum_{j=1}^\infty q_j f_j$ converges in $L^2(\mu)$ if and only if (5.63) holds. But

$$\sum_1^\infty Q_j f_j \psi = (\sum_{j=1}^\infty q_j f_j) \psi. \quad (5.66)$$

So for bounded ψ the series on the left converges in L^2 for each ψ if and only if the series $(\sum_{j=1}^\infty q_j f_j)$ converges in L^2 .

Now if $f \in C_c^\infty((0, \pi))$ then it is also in $L^2((0, \pi))$ and therefore $\sum_{j=1}^\infty f_j^2 < \infty$. Hence (5.63) holds because $\omega_j \rightarrow \infty$. This proves the last assertion of the theorem. ■

Side remark: The functions $s_j = \sqrt{2\omega_j} q_j$ are orthonormal coordinates in $L^2(\mathbb{R}^\infty, \mu)$, as is shown in (5.64).

SUMMARY

We have now followed a heuristic path, starting with analogies to infinite dimensional quantum mechanics and its attendant infinities and meaningless spaces, and ending with a well defined Hilbert space, a well defined Hamiltonian on it, and well defined field operators. In this way we have united the spirit of the vibrating string equation (5.10) with the spirit of quantum mechanics.

Putting all the heuristics behind us now, here is what we actually constructed. The following summarizes the notation of this section and the statement of theorems. It is self contained except for the proofs. There are no divergences and no infinities to subtract.

Notation 5.4 Define

$$u_j(x) = \sqrt{2/\pi} \sin(jx), \quad 0 \leq x \leq \pi, \quad j = 1, 2, \dots \quad (5.67)$$

$$\gamma_\omega(dq) = \sqrt{\omega/\pi} e^{-\omega q^2} dq, \quad q \in \mathbb{R}, \quad \omega > 0 \quad (5.68)$$

$$\omega_j = j \quad (5.69)$$

$$\mathbf{q} = (q_1, q_2, \dots) \quad (5.70)$$

$$\gamma(d\mathbf{q}) = \prod_{j=1}^{\infty} \gamma_{\omega_j}(dq_j) \quad (5.71)$$

$$\mathcal{H} = L^2(\mathbb{R}^\infty, \gamma) \quad (5.72)$$

$$(Q_j \psi)(\mathbf{q}) = q_j \psi(\mathbf{q}), \quad \text{for } \psi : \mathbb{R}^\infty \rightarrow \mathbb{C} \quad (5.73)$$

Theorem 5.5 Let $f \in C_c^\infty((0, \pi))$. Define

$$\phi_f = \sum_{j=1}^{\infty} (f, u_j)_{L^2((0, \pi))} Q_j \quad (5.74)$$

This series of (unbounded) operators converges in $L^2(\mathbb{R}^\infty, \gamma)$ when applied to a bounded function ψ . Moreover the map

$$C_c^\infty((0, \pi)) \ni f \mapsto \phi_f \psi \quad (5.75)$$

is a linear map from $C_c^\infty((0, \pi))$ into $L^2(\mathbb{R}^\infty; \gamma)$ for each function $\psi \in L^\infty(\mathbb{R}^\infty; \gamma)$. Furthermore there is a densely defined, non-negative, essentially self-adjoint operator H on $L^2(\mathbb{R}^\infty; \gamma)$ given by the series

$$H\psi = \sum_{j=1}^{\infty} \left(-\frac{1}{2} \frac{\partial^2}{\partial q_j^2} + \omega_j q_j \frac{\partial}{\partial q_j} \right) \psi \quad (5.76)$$

which converges in $L^2(\mathbb{R}^\infty; \gamma)$ for a dense set of ψ . (E.g. ψ that depend on only finitely many q_j and smoothly also.)

Notice that this is a clear cut theorem with no divergences and no heuristics. The linear map $f \mapsto \phi_f$ is a map from test functions to operators on \mathcal{H} . So it is an operator valued distribution over $(0, \pi)$. The Hilbert space \mathcal{H} is a well defined Hilbert space and the Hamiltonian H is as nice as can be. The operator valued distribution $f \mapsto \phi_f$ is the simplest example of a typical quantum field. Based on this example you are almost ready to make your own set of axioms for quantum field theory.

However we have not yet produced time dependent field operators $\phi(x, t)$ which can claim to be quantum analogs of solutions to the field equation (5.10). You can't make axioms for relativistic quantum field theory if you don't have time playing a role similar to that of space.

5.5 The time dependent field

As we know from Section 4.10, we have the option of letting the initial wave function ψ_0 evolve in time by the Schrödinger equation, giving $\psi(t) = e^{-itH}\psi_0$, or keeping ψ_0 fixed and letting the observables change with time (Heisenberg picture). In quantum field theory it is more insightful to use the Heisenberg picture. To this end let us put the time dependence into the quantized field φ_f . In accordance with the prescription (??), define

$$\varphi_f(t) = e^{itH} \varphi_f e^{-itH}.f \in C_c^\infty((0, \pi)) \quad (5.77)$$

This is the Heisenberg field operator. The use of the honest operator ϕ_f instead of its informal expression

$$\phi_f = \int_0^\pi \phi(x)f(x)dx \quad (5.78)$$

is unfortunately cumbersome and actually makes some simple equations hard to read. We have to learn how to write these equations in terms of such informal symbols as $\phi(x)$ and how then to give these equations precise meanings in the distribution sense. The time to do this is now.

To begin with, let us informally manipulate with the definition (5.77) thus:

$$\phi_f(t) = e^{itH} \int_0^\pi \phi(x)f(x)dx e^{-itH} \quad (5.79)$$

$$= \int_0^\pi \left(e^{itH} \phi(x) e^{-itH} \right) dx \quad (5.80)$$

$$= \int_0^\pi \phi(x, t)f(x)ds \quad (5.81)$$

where

$$\phi(x, t) = e^{itH} \phi(x) e^{-itH}, \quad 0 \leq x \leq \pi, \quad t \in \mathbb{R}. \quad (5.82)$$

Since $x \mapsto \phi(x)$ is not really a function, what does (5.82) mean?

Answer #1. $\phi(x, t)$ is a distribution in x for each t . In other words $\int_0^\pi \phi(x, t)f(x)dx$ is a well defined operator for each t , namely it is $\phi_f(t)$.

Answer #2. Since $\phi(x, t)$ depends on both x and t it would be reasonable to interpret it as an operator valued distribution over $(0, \pi) \times \mathbb{R}$. In this interpretation one would want to give meaning to

$$\phi(g) = \int_{(0, \pi) \times \mathbb{R}} \phi(x, t)g(x, t)dxdt. \quad (5.83)$$

as an honest operator for each test function $g \in C_c^\infty((0, \pi) \times \mathbb{R})$. This is easy enough to do in our case because we need only do the x integral first in (5.83) to find

$$\phi(g) = \int_{\mathbb{R}} \phi_{g_t}(t) dt, \quad (5.84)$$

where $g_t(x) = g(x, t)$. The integrand is a well defined operator valued function of t . So the integral *might* make sense. Why is there any question about this? Well, these are all unbounded operators. Their domains may change with t . Thus if one hopes to interpret (5.84) as meaning

$$\phi(g)\psi = \int_{\mathbb{R}} \phi_{g_t}(t)\psi dt \quad (5.85)$$

for some set of $\psi \in \mathcal{H}$ then one has to be sure to use only those ψ which are in the domains of *all* the operators $\phi_{g_t}(t)$, $t \in \mathbb{R}$. Fortunately, there does indeed exist a dense subspace $\mathcal{D} \subset \mathcal{H}$ which not only lies in all of these domains but for which the function $t \mapsto \phi_{g_t}(t)\psi$ is a continuous as a function into \mathcal{H} . The integral in (5.85) makes clear sense for such ψ and the equation (5.84) therefore has a meaningful interpretation. It would take us too far afield to go any further into these domain issues. But a reader who would like to pursue these technical problems herself might consider the functions $\psi \in \mathcal{H} \equiv L^2(\mathbb{R}^\infty; \gamma)$ such that $\psi(\mathbf{q})$ depends only on finitely many coordinates, q_1, q_2, \dots, q_N and as a function of these coordinates is in $C^\infty(\mathbb{R}^N)$ with bounded derivatives. This dense subspace “works” for all choices of the test function g .

Having now alerted the reader as to how to interpret the following definitions and equations we can now state the most significant properties of our quantum field.

Theorem 5.6 Define $\phi(x)$ as in the preceding section. Let

$$\phi(x, t) = e^{itH}\phi(x)e^{-itH}, \quad 0 < x < \pi, \quad t \in \mathbb{R} \quad (5.86)$$

Define

$$\pi(x, t) = (\partial/\partial t)\phi(x, t) \quad (5.87)$$

Then

$$\ddot{\phi}(x, t) = \phi''(x, t) \quad \text{and} \quad (5.88)$$

$$[\pi(x, t), \phi(y, t)] = -i\delta(x - y)I_{\mathcal{H}} \quad (5.89)$$

Remark 5.7 (More pep talk.) Before proving this theorem lets review the meaning of these equations yet again. Let $g \in C_c^\infty((0, \pi) \times \mathbb{R})$. The equation (5.88) means this: multiply by $g(x, t)$, integrate with respect to x and t , do two integration by parts with respect to x and two with respect to t . Ignore boundary terms because g is zero near the boundary of $(0, \pi) \times \mathbb{R}$. Then we find

$$\int_{(0, \pi) \times \mathbb{R}} \phi(x, t) \ddot{g}(x, t) dx dt = \int_{(0, \pi) \times \mathbb{R}} \phi(x, t) g''(x, t) dx dt. \quad (5.90)$$

We already know what this equation means because there are no derivatives of ϕ involved. By definition this equation is what (5.88) means. To interpret (5.89) let f and h be two real functions in $C_c^\infty((0, \pi))$. Multiply (5.89) by $f(x)h(y)$ and integrate with respect to x and y . We find

$$[\pi_f(t), \phi_h(t)] = -i(f, h)_{L^2((0, \pi))} I_{\mathcal{H}} \quad (5.91)$$

By definition, this is the meaning of (5.89).

Remark 5.8 (Yet more pep talk.) There are lots of real valued solutions to the vibrating string equation (5.10), which is the same equation as (5.88). But (5.89) forces ϕ to be operator valued, in order get a non zero commutation. Remarkably, the pair of equations (5.88) and (5.89) together have only one solution, namely the one we constructed. Here is just a little more precision about this.

Hypotheses: a) Suppose that $\hat{\phi}(x, t)$ and $\hat{\pi}(x, t) \equiv (\partial/\partial t)\hat{\phi}(x, t)$ are operator valued distributions over $(0, \pi) \times \mathbb{R}$, acting on some Hilbert space \mathcal{K} and satisfying (5.88) and (5.89). b) \mathcal{K} contains no closed subspace which is invariant under all the operators $\hat{\phi}(x, t)$. c) technical conditions.

Conclusion: There is a unitary operator $V : \mathcal{K} \rightarrow \mathcal{H}$ such that $V\hat{\phi}(x, t)V^{-1} = \phi(x, t)$ for all x and t .

MORAL: We have constructed the only solution to the pair of equations (5.88) and (5.89) up to unitary equivalence.

Proof of Theorem 5.6. Differentiating (5.86) with respect to t we find $\pi(x, t) = ie^{itH}(H\phi(x) - \phi(x)H)e^{-itH}$ because $(d/dt)e^{itH}$ is equal both to $e^{itH}iH$ and to iHe^{itH} . As far as the meaning of this equation is concerned, notice that if one multiplies this equation by $f(x)$ and integrates over $(0, \pi)$ one gets the well defined operator ϕ_f in two places. So this equation makes

good sense. We are going to leave out this kind of observation henceforth, leaving such niceties to the reader.

If we differentiate (5.86) again we arrive at a second commutator. Thus we have

$$\pi(x, t) = \dot{\phi}(x, t) = ie^{itH}[H, \phi(x)]e^{-itH} \quad (5.92)$$

$$\ddot{\phi}(x, t) = -e^{itH}[H, [H, \phi(x)]]e^{-itH}. \quad (5.93)$$

We will evaluate these commutators. Now $\phi(x) = \sum_{j=1}^{\infty} u_j(x)Q_j$ and $H = \sum_{k=1}^{\infty} \left(-(1/2)\partial_k^2 + \omega_k q_k \partial_k \right)$. All of the terms in H commute with Q_j except the term $k = j$. It is straightforward to compute the one non-zero commutator because it is just a one dimensional computation. So also for the second commutator. Here is the result.

$$[H, Q_j] = -\partial_j + \omega_j Q_j \quad (5.94)$$

$$[H, [H, Q_j]] = \omega_j^2 Q_j. \quad (5.95)$$

Therefore

$$\begin{aligned} [H, \phi(x)] &= \sum_{j=1}^{\infty} u_j(x)[H, Q_j] \\ &= \sum_{j=1}^{\infty} u_j(x)(-\partial_j + \omega_j Q_j) \end{aligned} \quad (5.96)$$

and

$$\begin{aligned} [H, [H, \phi(x)]] &= \sum_{j=1}^{\infty} u_j(x)\omega_j^2 Q_j \\ &= -\sum_{j=1}^{\infty} u_j''(x)Q_j \\ &= -\phi''(x), \end{aligned} \quad (5.97)$$

since $u_j'' + \omega_j^2 u_j = 0$ by (5.13). This proves (5.88).

Since (5.89) contains a delta function we will have to adhere to the more precise meaning of (5.89), which is (5.91). Let $f_j = (f, u_j)_{L^2((0, \pi))}$ and $h_j =$

$(h, u_j)_{L^2((0,\pi))}$. As before, the conjugation by e^{itH} can be taken outside the commutator. Since e^{itH} commutes with the identity operator it suffices to prove (5.91) at $t = 0$. Thus we need to prove that

$$i[[H\phi_f], \phi_h] = -i(f, h)_{L^2((0,\pi))} I_{\mathcal{H}}$$

By (5.96) we have

$$i[[H\phi_f], \phi_h] = i\left[\sum_{j=1}^{\infty} f_j(-\partial_j + \omega_j Q_j), \sum_{k=1}^{\infty} h_k Q_k\right] \quad (5.98)$$

$$= i \sum_{j=1}^{\infty} [f_j(-\partial_j + \omega_j Q_j), h_j Q_j] \quad (5.99)$$

$$= -i \sum_{j=1}^{\infty} f_j h_j I_{\mathcal{H}} \quad (5.100)$$

$$= -1(f, h)_{L^2((0,\pi))} I_{\mathcal{H}} \quad (5.101)$$

We have of course used $[(-\partial_j + \omega_j Q_j), Q_k] = 0$ if $k \neq j$. This proves (5.91) and therefore (5.89). ■

5.6 Many, many particles: Fock space

There are processes in which a particle suddenly appears that wasn't there before. For example when the electron in a hydrogen atom drops into a lower energy state (orbit) light is emitted in the form of a photon that wasn't around before. It can also happen that a photon wandering by is absorbed by the atom, putting the electron into a higher energy state. In this case the photon disappears. The interactions between particles can be described in this way in general: some particles disappear while other particles appear. The Hilbert state space appears to be changing suddenly, corresponding to the change of particles being described. This won't do. The dynamics of such transitions from one multi-particle state to another, with different particles and different numbers of them coming in and going out, has been incorporated into the usual format; operators on a single Hilbert space. This will be a "big" Hilbert space, one that has subspaces corresponding to all the particles involved and all the possible numbers of them. There will be operators whose interpretations are to create or destroy particles. The mathematical structure that does the trick was implicit in Dirac's paper [?] and made explicit in V.

A. Fock's paper [20]. A fully rigorous account of this mathematical structure was first given by J.M. Cook [13].

The mathematical substance of this structure is largely algebraic and a little bit functional analytic. Moreover much of the structure is highly specific to quantum field theory. Our exposition of this topic will therefore be self contained.

5.6.1 Creation and annihilation operators

Notation 5.9 Let K be a complex, separable Hilbert space. \mathcal{F}_0 will denote the space of algebraic tensors over K . Thus an element $\beta \in \mathcal{F}_0$ is of the form

$$\beta = \sum_{j=0}^{\infty} \beta_j \quad (\text{finite sum}), \text{ where } \beta_j \in K^{\otimes_{alg} j}. \quad (5.102)$$

Here $K^{\otimes_{alg} j}$ denotes the algebraic j tensors over K for $j \geq 1$ and denotes \mathbb{C} for $j = 0$. We will also write

$$\mathcal{F} = \sum_{j=0}^{\infty} K^{\otimes j} \quad (5.103)$$

\mathcal{F} is a Hilbert space and \mathcal{F}_0 is dense in \mathcal{F} .

Notation 5.10 (Left interior product) Let $u \in K$ and let $\alpha = x_1 \otimes \cdots \otimes x_n \in K^{\otimes n}$. Define

$$i_u \alpha = (x_1, u) x_2 \otimes \cdots \otimes x_n \quad (5.104)$$

and extend i_u linearly to $K^{\otimes_{alg} n}$. Then

$$(i_u \alpha, \beta) = (\alpha, u \otimes \beta) \quad (5.105)$$

for decomposable α and therefore for all algebraic n -tensors α and all $\beta \in K^{\otimes(n-1)}$. Hence

$$|(i_u \alpha, \beta)| = |(\alpha, u \otimes \beta)| \leq \|\alpha\| \|u\| \|\beta\|$$

Therefore i_u extends uniquely to a bounded operator $i_u : K^{\otimes n} \rightarrow K^{\otimes(n-1)}$ with norm at most $\|u\|$.

Notation 5.11 (Permutations) Denote by S_n the group of permutations of $\{1, 2, \dots, n\}$. For $\sigma \in S_n$ define

$$P_\sigma(u_1 \otimes \cdots \otimes u_n) = u_{\sigma^{-1}1} \otimes \cdots \otimes u_{\sigma^{-1}n}, \quad u_j \in K, \quad j = 1, \dots, n \quad (5.106)$$

P_σ is easily seen to extend linearly to a unitary operator on $K^{\otimes n}$. Moreover $P_{\sigma\tau} = P_\sigma P_\tau$. So $\sigma \mapsto P_\sigma$ is a unitary representation of S_n .

Let

$$P_b = \frac{1}{n!} \sum_{\sigma \in S_n} P_\sigma \quad \text{and} \quad (5.107)$$

$$P_f = \frac{1}{n!} \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) P_\sigma, \quad (5.108)$$

where $\operatorname{sgn}(\sigma)$ is one if σ is even and is minus one if σ is odd.

Lemma 5.12 P_b and P_f are orthogonal projections on $K^{\otimes n}$.

Proof. If $\phi : S_n \rightarrow \{-1, 1\}$ is a homomorphism then

$$\begin{aligned} \left(\frac{1}{n!} \sum_{\sigma \in S_n} \phi(\sigma) P_\sigma \right)^2 &= \frac{1}{n!^2} \sum_{\sigma, \tau} \phi(\sigma) \phi(\tau) P_\sigma P_\tau \\ &= \frac{1}{n!^2} \sum_{\eta \in S_n} \sum_{\sigma \tau = \eta} \phi(\sigma \tau) P_{\sigma \tau} \\ &= \frac{1}{n!} \sum_{\eta \in S_n} \phi(\eta) P_\eta. \end{aligned}$$

In case $\phi(\sigma) \equiv 1$ then this asserts that $P_b^2 = P_b$. In case $\phi(\sigma) = \operatorname{sgn}(\sigma)$ this asserts that $P_f^2 = P_f$. Since the adjoint of any summand is another summand ($P_\sigma^* = P_{\sigma^{-1}}$) both of these projections are Hermitian. ■

Lemma 5.13 Let $1 \leq k \leq n+1$ and denote by $[1, k]$ the cyclic permutation $1 \rightarrow k \rightarrow k-1 \rightarrow \cdots \rightarrow 2 \rightarrow 1$. Let $u \in K$. Then, writing $K^{\otimes(n+1)} \equiv K \otimes K^{\otimes n}$, we have

$$P_b^{(n+1)} = \frac{1}{n+1} \sum_{k=1}^{n+1} P_{[1,k]}^{(n+1)} \left(I \otimes P_b^{(n)} \right), \quad \text{and} \quad (5.109)$$

$$P_f^{(n+1)} = \frac{1}{n+1} \sum_{k=1}^{n+1} (-1)^{k-1} P_{[1,k]}^{(n+1)} \left(I \otimes P_f^{(n)} \right). \quad (5.110)$$

Moreover

$$P_\tau i_u = i_u \left(I \otimes P_\tau \right) \quad \text{for any } \tau \in S_n, \quad \text{and} \quad (5.111)$$

$$i_v P_{[1,j+1]}^{(n+1)}(u \otimes \alpha) = P_{[1,j]}^{(n)} \left(u \otimes (i_v \alpha) \right), \quad \text{for all } \alpha \in K^{\otimes n}, \quad 1 \leq j \leq n. \quad (5.112)$$

Proof. Every permutation σ in S_{n+1} can be written uniquely in the form $\sigma = [1, k]\tau$ for some permutation τ that leaves 1 invariant. Just take $\tau = [1, k]^{-1}\sigma$ if $\sigma(1) = k$. Now $\text{sgn}([1, k]) = (-1)^{k-1}$. So $\text{sgn}(\sigma) = (-1)^{k-1} \text{sgn}(\tau)$. The identity (5.107), with $n+1$ in place of n , gives

$$P_b^{(n+1)} = \frac{1}{(n+1)!} \sum_{k=1}^{n+1} P_{[1,k]}^{(n+1)} \sum_{\tau} P_{\tau} \quad (5.113)$$

where the sum over τ runs over the permutation group of $\{2, 3, \dots, n+1\}$. This is (5.109). Similarly, the identity (5.108) gives (5.110).

It suffices to prove (5.111) and (5.112) in case $\alpha = x_1 \otimes \dots \otimes x_{n+1}$. But then the left side of (5.111) is $P_\tau(x_1, u)(x_2 \otimes \dots \otimes x_{n+1})$ while the right side is $(x_1, u)P_\tau(x_2 \otimes \dots \otimes x_{n+1})$. To prove (5.112) observe that for $1 \leq k \leq n+1$ we have

$$P_{[1,k]}^{(n+1)}(u \otimes \alpha) = x_1 \otimes \dots \otimes x_{k-1} \otimes u \otimes \dots \otimes x_n \quad (\text{u in } k^{\text{th}} \text{ position})$$

Hence, writing $k = j+1$, we find

$$\begin{aligned} i_v P_{[1,k]}^{(n+1)}(u \otimes \alpha) &= (x_1, v)x_2 \otimes \dots \otimes x_{k-1} \otimes u \otimes \dots \otimes x_n \quad (\text{u in } j^{\text{th}} \text{ position}) \\ &= (x_1, v)P_{[1,j]}^{(n)}(u \otimes x_2 \otimes \dots \otimes x_n) \\ &= P_{[1,j]}^{(n)} \left(u \otimes (i_v \alpha) \right). \end{aligned}$$

■

Lemma 5.14 Let u and v be in K and let $\alpha \in K^{\otimes n}$. Denote by $P_{(1,2)}$ the permutation operator on $K^{\otimes n}$ which permutes the first two factors. Then

$$i_u i_v P_{(1,2)} \alpha = i_v i_u \alpha \quad (5.114)$$

Proof. Since all three operators in (5.114) are bounded it suffices to prove (5.114) in case $\alpha = x_1 \otimes \dots \otimes x_n$. But in that case the left side of (5.114) is $(x_1, u)(x_2, v)x_3 \otimes \dots \otimes x_n$ while the right side is $(x_2, v)(x_1, u)x_3 \otimes \dots \otimes x_n$ ■

Definition 5.15 A tensor $\alpha \in K^{\otimes n}$ is *symmetric* if $P_\sigma \alpha = \alpha$ for all $\sigma \in S_n$. α is *anti-symmetric* if $P_\sigma \alpha = \text{sgn}(\sigma) \alpha$ for all $\sigma \in S_n$. Any element of $K^{\otimes 0} \equiv \mathbb{C}$ or of K is both symmetric and anti-symmetric.

Corollary 5.16 Suppose that $\alpha \in K^{\otimes n}$. Then

$$(i_u i_v - i_v i_u)\alpha = 0 \quad \text{if } \alpha \text{ is symmetric} \quad (5.115)$$

$$(i_u i_v + i_v i_u)\alpha = 0 \quad \text{if } \alpha \text{ is anti-symmetric} \quad (5.116)$$

Proof. Since the transposition $(1, 2)$ is odd, the identity (5.114) proves both (5.115) and (5.116). ■

Definition 5.17 The *Boson Fock space* is the subspace \mathcal{F}_b of \mathcal{F} consisting of symmetric tensors.

The *Fermion Fock space* is the subspace \mathcal{F}_f of \mathcal{F} consisting of anti-symmetric tensors. Thus if

$$\alpha = \sum_{n=0}^{\infty} \alpha_n, \quad \alpha_n \in K^{\otimes n} \quad (5.117)$$

then

$\alpha \in \mathcal{F}_b$ if each α_n is symmetric.

$\alpha \in \mathcal{F}_f$ if each α_n is anti-symmetric.

Definition 5.18 Let $u \in K$. The *annihilation operator* associated to u is given by

$$a(u)\alpha = \sum_{n=1}^{\infty} \sqrt{n} i_u \alpha_n \quad (5.118)$$

when α is given by (5.117). The domain of $a(u)$ consists of those α for which the series (5.118) converges. The interior product i_u is defined in Notation 5.10. The annihilation operators will be of interest for us only when operating on \mathcal{F}_b or \mathcal{F}_f . The *creation operator* associated to u is

$$c(u)\alpha = \sum_{n=0}^{\infty} \sqrt{n+1} P_b(u \otimes \alpha_n) \quad \text{for } \alpha \in \mathcal{F}_b \quad (5.119)$$

$$c(u)\alpha = \sum_{n=0}^{\infty} \sqrt{n+1} P_f(u \otimes \alpha_n) \quad \text{for } \alpha \in \mathcal{F}_f \quad (5.120)$$

As before, the domain of $c(u)$ consists of all α in the respective space \mathcal{F}_b or \mathcal{F}_f for which the series converges. These two creation operators are called the *Boson creation operator* and *Fermion creation operator*, respectively.

Lemma 5.19 \mathcal{F}_b and \mathcal{F}_f are each invariant under $a(u)$ for all $u \in K$. Moreover

$$c(u)^* = a(u) \quad \text{on } \mathcal{F}_b \text{ or } \mathcal{F}_f, \text{ respectively.} \quad (5.121)$$

Or maybe write

$$\left(a(u) | \mathcal{F}_b \right)^* = c(u) \quad \text{on } \mathcal{F}_b \quad (5.122)$$

$$\left(a(u) | \mathcal{F}_f \right)^* = c(u) \quad \text{on } \mathcal{F}_f \quad (5.123)$$

Proof. If α is in $K^{\otimes n}$ and $\tau \in S_{n-1}$ then $P_\tau i_u \alpha = i_u (I \otimes P_\tau) \alpha$ by (5.111). Thus if α is symmetric then so is $i_u \alpha$ and if α is anti-symmetric then so is $i_u \alpha$. So \mathcal{F}_b and \mathcal{F}_f are each invariant under $a(u)$.

If $\beta \in K^{\otimes(n-1)}$ then $(i_u \alpha, \beta) = (\alpha, u \otimes \beta)$ by (5.105). Multiply by \sqrt{n} to find $(a(u)\alpha, \beta) = \sqrt{n}(\alpha, u \otimes \beta) = (\alpha, \sqrt{n}(P_b(u \otimes \beta))$ if α is symmetric. So if α and β are both symmetric then $(a(u)\alpha, \beta) = (\alpha, c(u)\beta)$. This proves (5.122). This also proves (5.123) if we simply replace symmetric by anti-symmetric and P_b by P_f .

Since $a(u)$ lowers rank by one while $c(u)$ raises rank by one its clear that the identity $(a(u)\alpha, \beta) = (\alpha, c(u)\beta)$ holds for all finite rank symmetric tensors α and β and for all finite rank anti-symmetric tensors α and β . This proves (5.122) and (5.123) on the space of finite rank tensors. In particular $a(u)$ and $c(u)$ have densely defined adjoints and therefore their restrictions to finite rank tensors have closed extensions. Its clear from the definitions that the closure of these operators have the domains specified Definition 5.18. ■

END of DAY 23 = 4/19/11

5.6.2 The canonical commutation relations

Theorem 5.20 (*Boson commutation relations*) For any vectors u and v in K we have, on \mathcal{F}_b ,

$$a(u)a(v) - a(v)a(u) = 0 \quad (5.124)$$

$$c(u)c(v) - c(v)c(u) = 0 \quad (5.125)$$

$$a(u)c(v) - c(v)a(u) = (v, u)I_{\mathcal{F}_b} \quad (5.126)$$

Proof. We are going to prove these commutations relations when applied just to finite rank vectors. To this end we need only consider a tensor $\alpha \in K^{\otimes n}$ which is symmetric. In this case

$$(a(u)a(v) - a(v)a(u))\alpha = \sqrt{(n(n-1))}(i_u i_v - i_v i_u)\alpha = 0$$

by (5.115). This proves (5.124). (5.125) now follows from Lemma 5.19. (Say something about domains.)

Using (5.109) and the symmetry of α , we find

$$\begin{aligned} a(v)c(u)\alpha &= (\sqrt{n+1}i_v)\sqrt{n+1}P_b^{(n+1)}(u \otimes \alpha) \\ &= i_v \sum_{k=1}^{n+1} P_{[1,k]}^{(n+1)}(u \otimes \alpha) \\ &= i_v(u \otimes \alpha) + \sum_{k=2}^{n+1} i_v P_{[1,k]}^{(n+1)}(u \otimes \alpha) \\ &= (u, v)\alpha + \sum_{j=1}^n P_{[1,j]}^{(n)}\left(u \otimes (i_v \alpha)\right), \end{aligned}$$

wherein we have used (5.111) in the last step. On the other hand, by (5.109) with n instead of $n+1$, we have

$$\begin{aligned} c(u)a(v)\alpha &= nP_b^{(n)}\left(u \otimes (i_v \alpha)\right) \\ &= \sum_{j=1}^n P_{[1,j]}^{(n)}\left(u \otimes (i_v \alpha)\right). \end{aligned}$$

This proves (5.126) on symmetric tensors in $K^{\otimes n}$ and therefore on all symmetric tensors of finite rank. ■

Theorem 5.21 (*Fermion commutation relations*) *For any vectors u and v in K we have, on \mathcal{F}_f ,*

$$a(u)a(v) + a(v)a(u) = 0 \tag{5.127}$$

$$c(u)c(v) + c(v)c(u) = 0 \tag{5.128}$$

$$a(u)c(v) + c(v)a(u) = (v, u)I_{\mathcal{F}_b} \tag{5.129}$$

Proof. The proofs of (5.127) and (5.128) are the same as those of (5.124) and (5.125) if one uses (5.116) instead of (5.115). Next, using (5.110) and the anti-symmetry of α we find

$$\begin{aligned} a(v)c(u)\alpha &= (\sqrt{n+1}i_v)\sqrt{n+1}P_f^{(n+1)}(u \otimes \alpha) \\ &= i_v \sum_{k=1}^{n+1} (-1)^{k-1} P_{[1,k]}^{(n+1)}(u \otimes \alpha) \\ &= i_v(u \otimes \alpha) + \sum_{k=2}^{n+1} (-1)^{k-1} i_v P_{[1,k]}^{(n+1)}(u \otimes \alpha) \\ &= (u, v)\alpha + \sum_{j=1}^n (-1)^j P_{[1,j]}^{(n)}\left(u \otimes (i_v\alpha)\right), \end{aligned}$$

wherein we have used (5.111) in the last step. On the other hand, by (5.110) with n instead of $n+1$, we have

$$\begin{aligned} c(u)a(v)\alpha &= nP_b^{(n)}\left(u \otimes (i_v\alpha)\right) \\ &= - \sum_{j=1}^n (-1)^j P_{[1,j]}^{(n)}\left(u \otimes (i_v\alpha)\right). \end{aligned}$$

This proves (5.126) on anti-symmetric tensors in $K^{\otimes n}$ and therefore on all anti-symmetric tensors of finite rank. ■

Mildly hidden in the commutation relations (5.124) - (5.126) are the Heisenberg canonical commutation relations for infinitely many P s and Q s. Indeed these identities immediately give

$$[c(u) + a(u), c(v) + a(v)] = \left((v, u) - (u, v)\right)I = 2i\Im(v, u)I \quad (5.130)$$

So let $\{e_1, e_2, \dots\}$ be an orthonormal basis of K and define

$$\hat{Q}_j = \frac{c(e_j) + a(e_j)}{\sqrt{2}} \quad \text{and} \quad (5.131)$$

$$\hat{P}_k = \frac{c(ie_k) + a(ie_k)}{\sqrt{2}} = i \frac{c(e_k) - a(e_k)}{\sqrt{2}} \quad (5.132)$$

Then all of these operators are clearly symmetric. (actually self-adjoint, but never mind that.) Moreover (5.130) shows that

$$[\hat{P}_k, \hat{Q}_j] = i\delta_{j,k}I, \quad j, k = 1, 2, \dots \quad (5.133)$$

Thus the space \mathcal{F}_b supports a countable family of operators satisfying the Heisenberg commutation relations. We are going to see in the next section that this is just what one needs to pass from a system with finitely many degrees of freedom, i.e., mechanics, to a system with infinitely many degrees of freedom, i.e., a field.

5.6.3 Occupation number bases

The following bases of \mathcal{F}_b and \mathcal{F}_f are conceptually illuminating and a big step in understanding some of the notation common in the physics literature.

Theorem 5.22 (*Occupation number basis of \mathcal{F}_b*) *Let e_1, e_2, \dots be an orthonormal basis of K . For any vector $u \in K$ let*

$$u^n = u \otimes u \otimes \cdots \otimes u, \quad (n \text{ factors}) \quad (5.134)$$

Let n_1, n_2, \dots be a finitely nonzero sequence of non-negative integers and let $n = \sum_{j=1}^{\infty} n_j$. (This is a finite sum.) Define

$$|n_1, n_2, \dots\rangle = \left(\frac{n!}{n_1! n_2! \dots} \right)^{1/2} P_b(e_1^{n_1} \otimes e_2^{n_2} \otimes \cdots) \quad (5.135)$$

Then the set of vectors $\{|n_1, n_2, \dots\rangle : n_j \geq 0, \sum_{j=1}^{\infty} n_j < \infty\}$ is an orthonormal basis of \mathcal{F}_b .

Proof. Since P_b is a projection,

$$\|P_b(e_1^{n_1} \otimes e_2^{n_2} \otimes \cdots)\|^2 = (P_b(e_1^{n_1} \otimes e_2^{n_2} \otimes \cdots), e_1^{n_1} \otimes e_2^{n_2} \otimes \cdots). \quad (5.136)$$

Referring to (5.107), observe that if a permutation σ carries the subset $\{1, \dots, n_1\}$ into it self and carries the set $\{n_1 + 1, \dots, n_2\}$ into itself and so on then $(P_\sigma(e_1^{n_1} \otimes e_2^{n_2} \otimes \cdots), (e_1^{n_1} \otimes e_2^{n_2} \otimes \cdots)) = ((e_1^{n_1} \otimes e_2^{n_2} \otimes \cdots), (e_1^{n_1} \otimes e_2^{n_2} \otimes \cdots))$. The number of permutations of this form is $n_1! n_2! \dots$. For any permutation not of this form the last inner product is clearly zero. Hence $\|P_b(e_1^{n_1} \otimes e_2^{n_2} \otimes \cdots)\|^2 = (n_1! n_2! \dots)/n!$. This shows that the vectors $|n_1, n_2, \dots\rangle$ are normalized. Moreover $((e_1^{n_1} \otimes e_2^{n_2} \otimes \cdots), (e_1^{m_1} \otimes e_2^{m_2} \otimes \cdots)) = 0$ if any $m_j \neq n_j$ no matter what order these factors are placed in. Hence the vectors $|n_1, n_2, \dots\rangle$ are also orthogonal. Finally, if w is a symmetric n tensor which is orthogonal to all of the $|n_1, n_2, \dots\rangle$ then $(w, e_{j_1} \otimes e_{j_2} \otimes \cdots e_{j_n}) = (P_b w, e_{j_1} \otimes e_{j_2} \otimes \cdots e_{j_n}) =$

$(w, P_b(e_{j_1} \otimes e_{j_2} \otimes \cdots e_{j_n})) = 0$ because $P_b(e_{j_1} \otimes e_{j_2} \otimes \cdots e_{j_n})$ is a multiple of one of the basis elements in (5.135). ■

Terminology One says that the vector $|n_1, n_2, \dots\rangle$ is a state in which there are present n_1 particles in state e_1 , n_2 particles in state e_2 , and so on. Otherwise put, the state $|n_1, n_2, \dots\rangle$ is occupied by n_1 particles in state e_1 , n_2 particles in state e_2 , and so on.

Theorem 5.23 (*Action of annihilation and creation operators.*) *For any orthonormal basis e_1, e_2, \dots of K we have*

$$c(e_j)|n_1, n_2, \dots\rangle = \sqrt{n_j + 1} |n_1, n_2, \dots, n_j + 1, n_{j+1}, \dots\rangle \quad (5.137)$$

$$a(e_j)|n_1, n_2, \dots\rangle = \sqrt{n_j} |n_1, n_2, \dots, n_j - 1, n_{j+1}, \dots\rangle \quad (5.138)$$

wherein we have made the convention that $|n_1, n_2, \dots\rangle = 0$ if any $n_i < 0$.

For the Fermion occupation number basis our readers may find it more familiar to describe the Fermion Fock space as the completion of the exterior algebra over K in the direct sum norm that we have been using. In order to make algebraic manipulations with ease we will denote by $\Lambda(K)$ the algebraic anti-symmetric tensors over the Hilbert space K .

Recall the standard notation, [40, pages 7 and 28],

$$u_1 \wedge \cdots \wedge u_n = P_f(u_1 \otimes \cdots \otimes u_n), \quad u_j \in K \quad (5.139)$$

$$\alpha \wedge \beta = P_f(\alpha \otimes \beta), \quad \alpha, \beta \in \Lambda(K) \quad (5.140)$$

$\Lambda(K)$ is an associative algebra in this product, as follows from the identity $(u_1 \wedge \cdots \wedge u_k) \wedge (v_1 \wedge \cdots \wedge v_j) = (u_1 \wedge \cdots \wedge u_k \wedge v_1 \wedge \cdots \wedge v_j)$, which in turn follows from the easily verified identity $P_f^{(k+j)}(P_f^{(k)} \otimes P_f^{(j)})$ (exercise).

The creation operator can be described in these simple algebraic terms thus: If α is an n -tensor in $\Lambda(K)$ then

$$c(u)\alpha = \sqrt{n+1} u \wedge \alpha, \quad \alpha \in \Lambda(K) \cap K^{\otimes n}, \quad (5.141)$$

which follows from the definitions (5.120) and (5.140). In particular, induction now shows that

$$c(u_1) \cdots c(u_n) 1 = \sqrt{n!} u_1 \wedge \cdots \wedge u_n \quad (5.142)$$

The norm of this element is easy to compute if $\{u_1, \dots, u_n\}$ is orthonormal because $(P_\sigma(u_1 \otimes \dots \otimes u_n), P_\tau(u_1 \otimes \dots \otimes u_n)) = 0$ unless $\sigma = \tau$, and in this case the value is one. So

$$\begin{aligned} n! \| u_1 \wedge \dots \wedge u_n \|^2 &= n! \| P_f(u_1 \otimes \dots \otimes u_n) \|^2 \\ &= \frac{1}{n!} \sum_{\sigma} \sum_{\tau} (\text{sgn } \sigma)(\text{sgn } \tau) (P_\sigma(u_1 \otimes \dots \otimes u_n), P_\tau(u_1 \otimes \dots \otimes u_n)) \\ &= 1 \end{aligned}$$

Hence

$$\|c(u_1) \cdots c(u_n)1\| = 1 \quad \text{if } \{u_1, \dots, u_n\} \text{ is orthonormal} \quad (5.143)$$

An occupation number basis for Fermions can be readily defined. Let e_1, e_2, \dots be an orthonormal basis of K . Let (n_1, n_2, \dots) be a finitely non-zero sequence of zeros and ones. Define

$$|n_1, n_2, \dots\rangle = \sqrt{n!} e_1 \wedge e_2 \wedge \dots \quad (5.144)$$

where $e_j^{n_j}$ should be simply omitted from the wedge product if $n_j = 0$. Then (5.142) and (5.143) show that these are unit vectors. Any two of them are clearly orthogonal and their span is dense in \mathcal{F}_f . So they form an orthonormal basis of \mathcal{F}_f . Moreover (5.142) shows that

$$c(e_j)|n_1, n_2, \dots\rangle = \begin{cases} (-1)^{\sum_{i=1}^{j-1} n_i} |n_1, n_2, \dots, n_j + 1, n_{j+1}, \dots\rangle & \text{if } n_j = 0 \\ 0 & \text{if } n_j = 1 \end{cases} \quad (5.145)$$

It is a simple exercise to deduce from this and from the commutation relations that

$$a(e_j)|n_1, n_2, \dots\rangle = \begin{cases} (-1)^{\sum_{i=1}^{j-1} n_i} |n_1, n_2, \dots, n_j - 1, n_{j+1}, \dots\rangle & \text{if } n_j = 1 \\ 0 & \text{if } n_j = 0 \end{cases} \quad (5.146)$$

Of course no more than one particle can occupy a state e_j because $e_j \wedge e_j = 0$. This is reflected in the choice of indexing sequences, namely zeros and ones, in contrast with the Boson case.

The Pauli exclusion principle, which asserts that no more than one electron, (or other Fermion) can occupy a given state, is also reflected in this choice of indexing sequences.

5.6.4 Time evolution on \mathcal{F}_b and \mathcal{F}_f .

A Schrödinger evolution on K naturally induces a Schrödinger evolution on \mathcal{F}_b and \mathcal{F}_f as follows.

Let U be a unitary operator on \mathcal{K} . Then $U^{\otimes n} \equiv U \otimes \cdots \otimes U$ (n factors) is a unitary operator on $K^{\otimes n}$. The direct sum

$$\Gamma(U) = I \oplus U \oplus U^{\otimes 2} \oplus \dots \quad (5.147)$$

is a unitary operator on \mathcal{F} . It is clear that $\Gamma(U)$ commutes with permutations, and therefore leaves \mathcal{F}_b and \mathcal{F}_f invariant. Moreover Γ is a group homomorphism in that $\Gamma(UV) = \Gamma(U)\Gamma(V)$ for any two unitaries U and V on \mathcal{K} . Furthermore, if A is a self-adjoint operator on K and e^{itA} is the one parameter group that it generates, then $\Gamma(e^{itA})$ is also a one parameter group on \mathcal{F} . For any vectors $x_1, \dots, x_n \in \mathcal{K}$ each factor in $(e^{itA}x_1) \otimes \cdots \otimes (e^{itA}x_n)$ is a continuous function of t and therefore so is the product. Therefore $\Gamma(e^{itA})w$ is a continuous function of t for a dense set of vectors $w \in \mathcal{F}$ and therefore for all $w \in \mathcal{F}$ because $\Gamma(e^{itA})$ is uniformly bounded. Thus the map $t \mapsto \Gamma(e^{itA})$ is a strongly continuous one-parameter unitary group, and, by the Stone-von Neumann theorem, Theorem 4.7, has a self-adjoint generator $\gamma(A)$. That is,

$$\Gamma(e^{itA}) = e^{it\gamma(A)}. \quad (5.148)$$

If we simply differentiate $(e^{itA}x_1) \otimes \cdots \otimes (e^{itA}x_n)$ with respect to t at $t = 0$ we find a formula for $\gamma(A)$ on n -tensors. Namely,

$$\gamma(A) = A \otimes I \otimes I \otimes \cdots \otimes I + I \otimes A \otimes I \otimes \cdots \otimes I + \dots \quad (5.149)$$

on n -tensors which are in the span of products of vectors in the domain of A . Actually this domain is a core for the self-adjoint operator $\gamma(A)$. But we will omit any discussion of this technical issue here. You could look up J.M. Cook's paper [13] on this point if you wish.

If A is the Hamiltonian for some system whose state space is K then $A \otimes I + I \otimes A$ is the Hamiltonian for the system consisting of two particles "of type K ". Of course if this particle is a Boson you must restrict this operator to the symmetric tensors in $K \otimes K$. And if the particle of type K is a Fermion you must restrict this operator to the anti-symmetric tensors in $K \otimes K$. Clearly the operator $A \otimes I + I \otimes A$ leaves both of these subspaces invariant. Similarly, the operator $\gamma(A)$ is the Hamiltonian for the system

consisting of an arbitrary number of identical particles, whether they be all Bosons or all Fermions.

There is an important physical point to be understood in this construction of a Hamiltonian for a family of identical particles from the Hamiltonian for one of them. The Hamiltonian $\gamma(A)$ corresponds to adding together the energy of several particles, say n , without including any energy of interaction between the particles. The n particles move “freely” without interfering with each other. This is not really “reality”. In a realistic system particles interact, either directly or through some other intermediate particle. Typically, such interactions between otherwise independently moving particles is incorporated by adding an additional term to the Hamiltonian $\gamma(A)$. We are going to explore simple cases of this in Section 5.8.

REFERENCES:

Cook [13],
Fock [20],

Reed and Simon, [48, Section X.7]. This reference contains much more about Fock spaces than we have discussed.

5.7 The Particle-Field isomorphism.

In Section 5.5 we constructed a quantized field $\varphi(x, t)$. On the one hand it was a solution to our field equation

$$\ddot{\varphi}(x, t) = \varphi''(x, t)$$

On the other hand it was “quantized” in the sense that not only do $\varphi(x, t)$ and $\varphi(y, t')$ not commute with each other but the “position-like” operators $\varphi(x, t)$ and the “momentum-like” operators $\dot{\varphi}(y, t)$ satisfy the continuum version of the usual Heisenberg commutation relations $[P_j, Q_k] = i\delta_{jk}I$. Namely,

$$[\dot{\varphi}(x, t), \varphi(y, t)] = -i\delta(x - y)I$$

cf. (5.89). Of course the appearance of the delta function just reflects the fact that $x \mapsto \varphi(x, t)$ is an operator valued distribution on $(0, \pi)$, not an operator valued function.

The structure of the space $L^2(\mathbb{R}^\infty, \mu)$ on which these field operators act, contrasts sharply with that in the previous section, which was all about particles, many particles. In this section we are going to show that the field picture of Section 5.5 and the particle picture of Section 5.6 are isomorphic in a way that preserves all the quantum mechanical structures. A little more precisely, we will construct a unitary operator from the field Hilbert space $L^2(\mathbb{R}^\infty, \mu)$ onto the appropriate Boson Fock space \mathcal{F}_b , which interchanges the field operators $\varphi(x, t)$ with simple expressions in terms of the annihilation and creation operators. To this end we will apply the machinery of Section 5.6 to the appropriate Hilbert space K , which underlies the definition of the quantized field φ constructed in Section 5.5.

Definition 5.24 (The field space.) Denote by $\mathbf{h}_{1/2}$ the space of sequences $\mathbf{q} \equiv (q_1, q_2, \dots)$ of real numbers such that

$$\|\mathbf{q}\|_{\mathbf{h}_{1/2}}^2 \equiv 2 \sum_{j=1}^{\infty} \omega_j q_j^2 < \infty. \quad (5.150)$$

Denote by $H_{1/2}$ the set of real valued functions u on $(0, \pi)$ of the form

$$u(x) = \sum_{j=1}^{\infty} q_j u_j(x), \quad x \in (0, \pi), \quad (q_1, q_2, \dots) = \mathbf{q} \in \mathbf{h}_{1/2} \quad (5.151)$$

and define

$$\|u\|_{H_{1/2}} = \|\mathbf{q}\|_{\mathbf{h}_{1/2}} \quad (5.152)$$

Since $\omega_j \rightarrow \infty$, the definition (5.150) implies that $\mathbf{h}_{1/2} \subset \ell^2$. The series (5.151) therefore converges in $L^2((0, \pi))$. Its worth noting that (5.152) gives

$$\|u_j\|_{H_{1/2}} = \sqrt{2\omega_j}. \quad (5.153)$$

A reader familiar with Sobolev spaces might notice that $H_{1/2}$ is exactly the Sobolev space of order 1/2 over $(0, \pi)$ because the functions u_j satisfy $u_j'' = -\omega_j^2 u_j$ with $\omega_j = j$.

Preview: The following theorem has quite a few different proofs. Here are sketches of the main ones.

a) (not for us) The von Neumann algebra generated by the operators \hat{Q}_j defined in Section 5.6 is a maximal abelian algebra. The vacuum state, $1 \in K^{\otimes 0} \equiv \mathbb{C}$, is a cyclic vector for this algebra and induces the measure γ on the spectrum of the algebra when the spectrum is identified with \mathbb{R}^∞ .

b) (not for us) The infinite dimensional heat semigroup $e^{t\Delta_\infty}$ takes $L^2(\mathbb{R}^\infty; \gamma)$ to functions with holomorphic extensions to \mathbb{C}^∞ . The Taylor coefficients of such functions, computed at zero, are exactly the symmetric tensors \mathcal{F}_b . Wick ordering lurks in this approach

c) (for us) We will map a natural orthonormal basis of $L^2(\mathbb{R}^\infty, \gamma)$ onto the occupation number basis of \mathcal{F}_b defined in Section 5.6 and extend the map unitarily.

Theorem 5.25 (Particle-Field isomorphism) Take $K = H_{1/2}$ in Section 5.6 and denote by \mathcal{F}_b the corresponding Boson Fock space over $H_{1/2}$. There is a unique unitary operator

$$U : L^2(\mathbb{R}^\infty, \gamma) \rightarrow \mathcal{F}_b \quad (5.154)$$

such that

$$a) \quad U1 = 1 \quad (\text{the zero rank tensor in } \mathcal{F}_b) \quad \text{and} \quad (5.155)$$

$$b) \quad U(\partial/\partial q_j)U^{-1} = a(u_j), \quad j = 1, 2, \dots \quad (5.156)$$

Further, let

$$a_j = (2\omega_j)^{-1/2} a(u_j), \quad j = 1, 2, \dots \quad (5.157)$$

denote the normalized annihilation operators. Then U induces the following maps on operators

$$UQ_j U^{-1} = \omega_j^{-1/2} \hat{Q}_j, \quad (5.158)$$

$$UP_j U^{-1} = \omega_j^{1/2} \hat{P}_j \quad \text{and} \quad (5.159)$$

$$UHU^{-1} = \sum_{j=1}^{\infty} \omega_j a_j^* a_j, \quad (5.160)$$

where \hat{Q}_j and \hat{P}_j are defined in (5.131) and (5.132).

Proof. Referring to the notation in Section 5.5, let $e_j = (1/\sqrt{2\omega_j})u_j$. Then e_j is an orthonormal basis of $H_{1/2}$, as we see from (??). This basis defines in turn the orthonormal basis $|n_1, n_2, \dots\rangle$ defined in Theorem ??.

Referring now to Appendix 9.5 on Hermite polynomials, define functions $\psi_{n,j} : \mathbb{R}^\infty \rightarrow \mathbb{R}$ by

$$\psi_{n,j}(q_1, q_2, \dots) = \frac{1}{\sqrt{n!}} H_n(\sqrt{2\omega_j} q_j) \quad (5.161)$$

where H_n is the n^{th} order Hermite polynomial. It was shown in Appendix 9.5 that for each fixed j the functions $q_j \mapsto (1/\sqrt{n!})H_n(\sqrt{2\omega_j} q_j)$, $n = 0, 1, \dots$ form an orthonormal basis of $L^2(\mathbb{R}, \gamma_{\omega_j})$. Since the functions $q \mapsto q_j$ are independent with respect to γ , the finite products $\Pi_{j=1}^N \psi_{n_j,j}(q)$ are orthonormal with respect to γ . In fact they form an orthonormal basis of $L^2(\mathbb{R}^\infty, \gamma)$. (Exercise for the reader). Define

$$U\left(\prod_{j=1}^N \psi_{n_j,j}\right) = |n_1, n_2, \dots\rangle \quad (5.162)$$

Then U extends to a unitary operator from $L^2(\mathbb{R}^\infty; \gamma)$ onto $\mathcal{F}_b(H_{1/2})$. It remains to show that U has the asserted properties.

To this end note first that if $n = 0$ then $\psi_{n,j} = 1$ for all j . Therefore $U(1) = |0, 0, \dots\rangle = 1$. This proves (5.155).

Next, observe that (9.69) shows that

$$(\partial/\partial q_j)\psi_{n,j} = \sqrt{2\omega_j} \sqrt{n} \psi_{n-1,j}. \quad (5.163)$$

For any occupation number basis vector $|n_1, n_2, \dots\rangle$ the application of U^{-1} yields a product of functions $\psi_{n_k,k}(\mathbf{q})$ and since $\partial/\partial q_j$ differentiates just

one factor, (5.163) shows that the effect of differentiation is to lower n_j by one and to multiply by $\sqrt{2\omega_j}\sqrt{n_j}$. In accordance with (5.138) this yields $\sqrt{2\omega_j}a_j|n_1, n_2, \dots\rangle$, which, by (5.157), equals $a(u_j)|n_1, n_2, \dots\rangle$. Thus (5.156) holds when applied to any occupation number basis vector and therefore holds on a dense set in $\mathcal{F}_b(H_{1/2})$. We omit the technicalities addressing the exact domain on which (5.156) holds.

The remaining three properties (5.158) - (5.160) are really just rewrites of the identities (5.49) - (5.53) proved in Section 5.3 along with the definitions (5.131) and (5.132) once one knows the defining relation (5.156). Thus (5.156) implies that $U(\partial/\partial q_j)^*U^{-1} = a(u_j)^*$. Therefore

$$U\left(\partial/\partial q_j + (\partial/\partial q_j)^*\right)U^{-1} = a(u_j) + c(u_j) \text{ and} \quad (5.164)$$

$$U\left(\partial/\partial q_j - (\partial/\partial q_j)^*\right)U^{-1} = a(u_j) - c(u_j) \quad (5.165)$$

Keeping in mind the definition (5.157), divide both of these equations by $2\sqrt{\omega_j}$ to find

$$U\left(\frac{1}{2\sqrt{\omega_j}}(\partial/\partial q_j + (\partial/\partial q_j)^*)\right)U^{-1} = \frac{(a_j + c_j)}{\sqrt{2}} \text{ and} \quad (5.166)$$

$$U\left(\frac{1}{2\sqrt{\omega_j}}(\partial/\partial q_j - (\partial/\partial q_j)^*)\right)U^{-1} = \frac{(a_j - c_j)}{\sqrt{2}} \quad (5.167)$$

Therefore, in view of (5.38), (5.39), (5.131) and (5.132) we have

$$U\left(\omega_j^{1/2}Q_j\right)U^{-1} = \hat{Q}_j \text{ and} \quad (5.168)$$

$$U\left(\omega_j^{-1/2}P_j\right)U^{-1} = \hat{P}_j. \quad (5.169)$$

which proves (5.158) and (5.159). Finally, the equation (5.53) yields (5.160) if one replaces α_j , which, by (5.156), is the unitary transform of $a(u_j)$, by a_j , in accordance with (5.157). ■

SUMMARY: In the past four sections we constructed an operator valued distribution $\phi(x, t)$ over $(0, \pi) \times \mathbb{R}$ acting on a Hilbert space $L^2(\mathbb{R}^\infty, \gamma)$, which satisfies our wave equation (5.88) and also the Heisenberg canonical commutation relations (5.89). It is the latter identity that distinguishes the quantum field ϕ from a classical (i.e. real valued) solution to the wave equation. Where are the “photons” associated to the quantum field ϕ ? In Section

5.6 we constructed a Hilbert space \mathcal{F}_b whose elements correspond to a variable number of particles, each of which has Hilbert state space \mathcal{K} , a given Hilbert space. The idea that \mathcal{F}_b contains vectors representing two particles (of type \mathcal{K} , lets call them) or three particles is very visible in the structure of \mathcal{F}_b because \mathcal{F}_b contains a subspace $\mathcal{K} \otimes \mathcal{K}$ and a subspace $\mathcal{K} \otimes \mathcal{K} \otimes \mathcal{K}$ (actually just the symmetric tensors in these spaces). Such a particle structure is not visible in the Hilbert space $L^2(\mathbb{R}^\infty, \gamma)$, on which the quantum field $\phi(x, t)$ acts. However, as we saw in Theorem 5.25, there is a natural unitary map U from $L^2(\mathbb{R}^\infty, \gamma)$ onto \mathcal{F}_b if we take $\mathcal{K} = H_{1/2}$. Here the word “natural” means that U induces a map on operators which interchanges $\phi(x, t)$ with certain easy to describe operators on \mathcal{F}_b . In fact $U\phi(x, t)U^{-1}$ can be written as linear combinations of creation and annihilation operators on \mathcal{F}_b . (Take this as an easy exercise based on Theorem 5.25.) Pulling back the particle structure from \mathcal{F}_b to $L^2(\mathbb{R}^\infty, \gamma)$, we should try to understand which functions $\psi : \mathbb{R}^\infty \rightarrow \mathbb{C}$ correspond to one photon states, two photon states, etc.

First, dismiss from your mind the idea that you may have, that the harmonic oscillators that we used to construct the space $L^2(\mathbb{R}^\infty, \gamma)$, back in Section 5.3, “are” themselves the particles (photons) in question. Lets look at some examples of what the Particle-Field isomorphism does to some simple wave functions in $L^2(\mathbb{R}^\infty; \gamma)$. Recall that the function $\mathbf{q} \mapsto \psi_{n,j}(\mathbf{q})$ is a function of the j^{th} coordinate q_j of \mathbf{q} and as a function of q_j it is a Hermite polynomial of order n , with its argument adjusted appropriately for the frequency ω_j , as in (5.161). The orthonormal basis of $H_{1/2}$ that we used for construction of the isomorphism is $e_j = \sqrt{2\omega_j} u_j$. The isomorphism takes e.g. $\psi_{1,j}$ to the 1-photon state e_j . So no matter which harmonic oscillator (these are indexed by j) we start with we “get” only a one photon state. For different j these photons differ only in their color (i.e. frequency), not in their number. Next consider the function $\psi_{2,5}$. This maps to $e_5 \otimes e_5$ under the isomorphism (5.162). This is a two photon state. Both photons have frequency ω_5 (which is just 5 in our vibrating string example.) Next, consider the function $\psi_{1,5}(\mathbf{q})\psi_{1,7}(\mathbf{q})$. This maps to $const.(e_5 \otimes e_7 + e_7 \otimes e_5)$, which is also a two photon state. One of the photons has frequency ω_5 and the other has frequency ω_7 .

5.8 Pre-Feynman diagrams

Suppose that A and V are two operators on a Hilbert space. We are going to make some computations that are easily justified if A and V are bounded.

But this is not the case of interest to us. Nevertheless we will proceed as if all steps can be justified, ignoring all technical issues. First we need to establish a simple identity. Let $B = A + V$. Then

$$e^{tB} = e^{tA} + \int_0^t e^{(t-s)A} V e^{sB} ds. \quad (5.170)$$

Proof: The product rule for derivatives gives

$$\frac{d}{ds}(e^{-sA} e^{sB}) = e^{-sA}(B - A)e^{sB}$$

Integrate this identity from $s = 0$ to $s = t$ to find $e^{-tA} e^{tB} - I = \int_0^t e^{-sA} V e^{sB} ds$. Multiply both sides by e^{tA} on the left to deduce (5.170).

We can iterate this identity now by replacing the factor e^{sB} in the integrand by (5.170) itself, with t replaced by s and the s in the integrand replaced by σ , say. We find

$$e^{tB} = e^{tA} + \int_0^t e^{(t-s)A} V \left(e^{sA} + \int_0^s e^{(s-\sigma)A} V e^{\sigma B} d\sigma \right) ds \quad (5.171)$$

$$= e^{tA} + \int_0^t e^{(t-s)A} V e^{sA} ds + \int_{0 \leq \sigma \leq s \leq t} e^{(t-s)A} V e^{(s-\sigma)A} V e^{\sigma B} d\sigma ds \quad (5.172)$$

The exponential of B does not occur in the first two terms, and we can obviously continue in this manner, replacing next $e^{\sigma B}$ in the last term by an expression based on use of (5.170) again. If, in fact, A and B are bounded then it is quite easy to see that the $(n+1)^{st}$ term of this series is bounded by $e^{t\|A\|} \|V\|^n / n!$ and therefore the series converges in operator norm. But enough of this kind of technicality! Let us write out the three terms that we get by applying this procedure just once more. We find

$$e^{tB} = e^{tA} + \int_0^t e^{(t-s)A} V e^{sA} ds + \int_{0 \leq \sigma \leq s \leq t} e^{(t-s)A} V e^{(s-\sigma)A} V e^{\sigma A} d\sigma ds \quad (5.173)$$

+ higher degree terms in V

The case of interest for us is that in which a quantum system moves in accordance with a Hamiltonian consisting of an easy-to-deal-with part, H_0 , and a not-so-easy-to-deal-with part V . That is

$$H = H_0 + V. \quad (5.174)$$

Put $B = iH$ and $A = iH_0$ in (5.173). Then we have

$$e^{itH} = e^{itH_0} + \int_0^t e^{i(t-s)H_0}(iV)e^{isH_0}ds + \int_{0 \leq \sigma \leq s \leq t} e^{i(t-s)H_0}(iV)e^{i(s-\sigma)H_0}(iV)e^{i\sigma H_0}d\sigma ds + \dots \quad (5.175)$$

It is convenient to organize the ensuing computations by symbolizing the various terms in this series by a suggestive diagram. For the first term we will use

which symbolizes the operator e^{itH_0} , customarily called the “free” propagator for the interval $[0, t]$. The integrand in the second term consists of “free” propagation for a time s , followed by the interaction operator (iV), followed then by more free propagation for a time $t - s$. You can see that the total amount of free propagation time is t . And in the third term there is similarly, propagation, interaction, propagation, another interaction and a final propagation. The total propagation time is again t . The time (or times) at which interaction occurs is integrated over. The diagram that symbolizes the second term in (5.175) is

The third term is symbolized by

$$||||||| X ||||| X ||||| \quad (5.178)$$

Both symbols include the understanding that the intermediate time (s) (or times, (σ, s)) are integrated over. The equation (5.175) can therefore be written symbolically as

$$e^{itH} = \cancel{\text{XXXXX}} + \cancel{\text{XXXXX}} X \cancel{\text{XXXXX}} + \cancel{\text{XXXXX}} X \cancel{\text{XXXXX}} X \cancel{\text{XXXXX}} + \dots \quad (5.179)$$

We want to see now how these diagrams actually look for a system consisting of a particle (call it an electron) interacting with a quantized field. (Call it the electromagnetic field, but our model will be much simpler.) The quantized field, as we saw in the preceding section, “can be regarded as” (of course we are talking isomorphism here) an assembly of photons whose number may alter in interactions. We will allow the electron to move in “space”

which we take to be the interval $(0, \pi)$. The quantized field must, of course, also be a field that exerts forces on particles moving in this space. We have just such a field at our fingertips, namely the quantized field constructed in Section 5.5. The Hilbert state space for the electron is $\mathcal{H}_e = L^2((0, \pi))$ while the Hilbert state space for the quantized field (representing photons), and which will be denoted by \mathcal{H}_p , is $\mathcal{H}_p \equiv \mathcal{F}_b$. The combined system, electron plus field has state space

$$\mathcal{H} = \mathcal{H}_e \otimes \mathcal{H}_p, \quad (5.180)$$

in accordance with the combining principle given in (4.42). We need to describe the Hamiltonian that guides the time evolution of the system. Denote by H_e the Hamiltonian for the electron by itself. This is an operator on \mathcal{H}_e . For example if there are no forces on the electron, other than those which are keeping it in the interval $(0, \pi)$, then the Hamiltonian is $H_e = -1/(2m)d^2/dx^2$ with Dirichlet boundary conditions. One could add on a potential V if there are additional forces. But this will not matter for us. Lets just take $V = 0$. In any case the electron wave function moves in \mathcal{H}_e , as usual, by the 1-parameter unitary group e^{itH_e} . We have already constructed the Hamiltonian for the quantized field ϕ . Denote it by H_p . You may recall that this Hamiltonian was determined from our assumption that the classical version of the field ϕ evolves by the vibrating string equation, (5.10). The Hamiltonian is given by (5.55) in the field space representation and by (5.160) in the particle space representation. We are going to stick to the particle space representation henceforth because we want to see how the mathematics results in photons being created and destroyed by the interaction with electrons. You may recall that when we studied the hydrogen atom back in Section 4.6, we had no mechanism for explaining how an electron, jumping from one orbit (= state) to another, emitted light, i.e. produced a photon. It is precisely this mechanism that we want to understand now. To this end we need first to transfer the quantized field operators $\phi(x, t)$ from operating on $L^2(\mathbb{R}^\infty; \gamma)$ to operating on \mathcal{H}_p . Equation (5.158) shows that

$$\hat{\phi}(x) \equiv U\phi(x, 0)U^{-1} = \sum_{j=1}^{\infty} u_j(x)\omega_j^{-1/2}\hat{Q}_j \quad (5.181)$$

where $\phi(x, 0) = \phi(x)$ is given by the definition (5.18). We see also from the definition (5.131) that each \hat{Q}_j is a linear combination of $a_j \equiv a(e_j)$ and

$c_j \equiv c(e_j)$. Thus we may write

$$\hat{\phi}(x) = \sum_{j=1}^{\infty} u_j(x) b_j(c_j + a_j) \quad (5.182)$$

where each $b_j > 0$. It is this expression of the field in terms of creation and annihilation operators that we need to keep in mind. We already know that this series doesn't converge pointwise in any simple sense. We are going to proceed informally in order to understand just what the diagrams are good for. A reader who feels queasy about this can just assume that the coefficients b_j decrease to zero quickly enough to get convergence in her favorite sense, or are even zero after $n = 10$. Let

$$H_0 = H_e \otimes I + I \otimes H_p \text{ and} \quad (5.183)$$

$$H = H_0 + H_I \quad (5.184)$$

where H_I denotes the interaction part of the total Hamiltonian H . We will define H_I in a moment. But first observe that the combined system, electron plus field, would propagate, under the Hamiltonian H_0 by itself, without any interference between the electron and photons because

$$e^{itH_0} = e^{itH_e} \otimes e^{itH_p} \quad (5.185)$$

It is the term H_I which will cause the presence of an electron to influence the time evolution of the photon field and vice versa.

To describe the action of H_I on \mathcal{H} we need to use the natural isomorphism between $L^2((0, \pi)) \otimes \mathcal{H}_p$ and $L^2((0, \pi); \mathcal{H}_p)$. The latter space is the space of functions $\psi : (0, \pi) \rightarrow \mathcal{H}_p$ such that $\|\psi\|^2 = \int_0^\pi \|\psi(x)\|_{\mathcal{H}_p}^2 dx < \infty$. It is easy to see that the map $f \otimes w \mapsto \psi$ where the $\psi(x) = f(x)w$ extends linearly to a unitary map of $L^2((0, \pi)) \otimes \mathcal{H}_p$ onto $L^2((0, \pi); \mathcal{H}_p)$. We will identify these two spaces. Define

$$(H_I\psi)(x) = \hat{\phi}(x)\psi(x) \quad (5.186)$$

So H_I is just a multiplication operator by an operator valued “function” $\hat{\phi}$.

Let us see now what the diagrams in (5.179) signify. We will expand the total propagator e^{itH} in accordance with (5.175), where now $V = H_I$. The diagram (5.176) just means “propagate the initial state by e^{itH_0} ”. Since the initial state always contains our one (and only) electron and may contain several photons (or none), and since e^{itH_0} preserves the number of photons (and

of course preserves the fact that there is one electron, even though (5.185) shows that the state of the electron can be changing), we can symbolize the evolution corresponding to (5.176) by a diagram such as in Figure 10. It shows that we started with one electron and two photons and by the end we still have one electron and two photons.

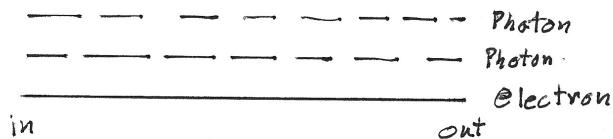


Figure 10: Feynman 1

Next, consider the diagram in (5.177). The initial state propagates freely as in the preceding diagram and then the interaction operates. The definitions (5.181) and (5.186) show that the interaction could increase the number of photons by one, these are the c_j terms, or decrease the number of photons by one, these are the a_j terms. In Figure 11 a you see no photons coming in and one photon coming out. Of course the one electron comes in and goes out. By the time we get to the end of diagram (5.177) the electron has changed its state. But it's still just one electron. However a photon has been created if there was none there before. And if there was a photon present at the beginning then the interaction may destroy it. More politely, one says that the electron has absorbed the photon. Moreover in the case that there is a photon present before the interaction begins, it may be destroyed as in Figure 11 b or the creation operator terms in (5.182) may create a second photon. This would be symbolized by the diagram in Figure 11 c.

The two diagrams in Figures 11 a and b determine all the things that can happen in this interaction. One need only repeat them for the higher order terms in the expansion (5.179).

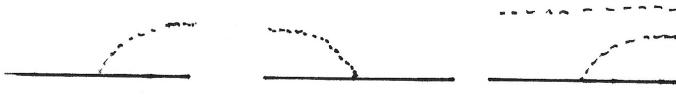


Figure 11: Feynman 2a,b,c

Corresponding to the second order diagram (5.178) there are many combinations of creation and annihilation of photons. Here are just a few of the many possibilities. Lets start with some simple initial states, say zero or one photon. We read the diagrams from left to right in spite of the fact that the operators on the right in (5.175) act first.

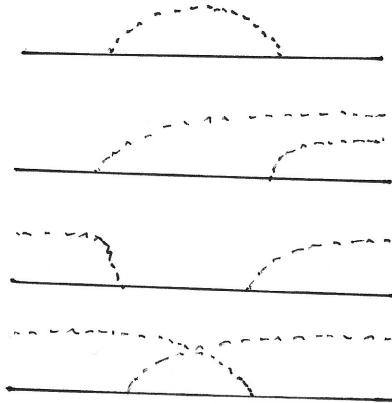


Figure 12: Feynman 3a,b,c,d

MORAL: The complicated terms in (5.175) can be organized by these kind of diagrams. It is believed by many that H_I is “small” in some sense and that the first few terms therefore accurately predict the outcomes of experiments. (So far this seems to be right.) Thus an interaction between the electron and “electromagnetic field” consists of a repeated creation and annihilation of photons while the electron goes its merry way changing its

state dramatically every time it emits or absorbs a photon, and changing gradually in between.

So much for the simple hydrogen atom of Section 4.6. We have a theory now that explains how the electron emits a photon when it falls into a lower energy state. We also understand now why an electron sometimes jumps into a higher energy state - it absorbs a photon.

REFERENCES:

The recent book by Folland, [21], goes much more deeply and honestly into the mathematical structure of quantum field theory than the version in these notes.

Here is a quick introduction to quantum field theory by Paul Federbush, aimed in part at mathematicians. [19]

For a book on quantum field theory by a real physicist with a mathematical orientation see the three volume exposition by Steven Weinberg, [62, 63, 64].

END of DAY 26 = 4/28/11

6 The electron-positron system

Within this section, I only did Section 6.1 in class. I added the remaining two subsections after the end of classes.

6.1 The Dirac equation

The simplest Lorentz invariant wave equation is the Klein-Gordon equation.

$$\partial^2 u / \partial t^2 - \Delta u + m^2 u = 0, \quad m \geq 0 \quad (6.1)$$

(Actually the case $m = 0$ is called the wave equation.) It was proposed to use the Klein-Gordon equation as a Lorentz invariant replacement of the Schrödinger equation (4.4). But the appearance of a second time derivative caused conceptual problems that could not be resolved. In his great paper [16], Dirac argued that one could still use Equ. (6.1) as a starting point for seeking a Lorentz invariant substitute for the Schrödinger equation by “factoring” (6.1) into two first order equations.

You may recall how dependent each step in the development of Maxwell's equations was on the immediately preceding experimental discovery. The final step, by Maxwell, followed thirty years after Faraday's great experiment of 1831 and was a mathematical implementing of Faraday's heuristic view of a magnetic field, based on his own experiment. By contrast, Dirac's discovery of the Dirac equation was based on a search for formal structure, namely Lorentz invariance plus first order in time. Thus:

We seek matrices β and $\alpha_j, j = 1, 2, 3$ such that the first order differential operator

$$H = \sum_{j=1}^3 \alpha_j \partial_j + \beta m \quad (6.2)$$

is a square root of $m^2 - \Delta$. Now

$$H^2 = \left(\sum_{j=1}^3 \alpha_j^2 \partial_j^2 + \beta^2 m^2 \right) + \sum_{j < k} (\alpha_j \alpha_k + \alpha_k \alpha_j) \partial_j \partial_k + \sum_{j=1}^3 m (\alpha_j \beta + \beta \alpha_j) \partial_j \quad (6.3)$$

Therefore $H^2 = m^2 - \Delta$ if and only if

$$\begin{aligned} \beta^2 &= 1, \quad \alpha_j^2 = -1, \quad j = 1, 2, 3 \\ \alpha_j \alpha_k + \alpha_k \alpha_j &= 0, \quad j \neq k, \\ \alpha_j \beta + \beta \alpha_j &= 0, \quad j = 1, 2, 3 \end{aligned} \quad (6.4)$$

Here is an example of 4×4 matrices which satisfy all of these conditions. In Appendix 9.3.2 we defined the Pauli spin matrices as

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (6.5)$$

Define 4×4 matrices by

$$\beta = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix} \quad \text{and} \quad \alpha_j = \begin{pmatrix} 0 & \sigma_j \\ -\sigma_j & 0 \end{pmatrix}, \quad j = 1, 2, 3 \quad (6.6)$$

Since each $\sigma_j^2 = I$, while $\sigma_j \sigma_k + \sigma_k \sigma_j = 0$ for $j \neq k$, the identities (6.4) are satisfied.

Concerning the uniqueness of this particular example we have the following theorem.

Theorem 6.1 Suppose that S is a finite dimensional vector space supporting operators α_j, β , $j = 1, 2, 3$, satisfying (6.4). Then $\dim S$ is a multiple of 4. If $\dim S = 4$ then any solution to the identities (6.4) is similar to (6.6). That is, any other solution has the form $T\alpha_j T^{-1}, T\beta T^{-1}$ for some invertible linear transformation T . If $\dim S > 4$ then S is a direct sum of invariant 4-dimensional subspaces on each of which the restrictions of the α_j and β are similar to the example (6.6).

Proof. See Chevalley [12]. (There must be a better reference for this proof.) ■

The *Dirac equation* is

$$\frac{\partial \psi}{\partial t} = iH\psi \quad \text{Dirac equation} \quad (6.7)$$

where H is given in (6.2) and the coefficients satisfy (6.4). Here $\psi(x, y, z, t)$ lies in the four dimensional complex vector space S on which the matrices α_j, β act. In the case of the choice (6.6) we may of course take $S = \mathbb{C}^4$.

Notice that if ψ is an S valued solution to (6.7) then

$$\begin{aligned} (\partial^2/\partial t^2)\psi &= (\partial/\partial_t)(iH\psi) \\ &= (iH)^2\psi \\ &= -(m^2 - \Delta)\psi \end{aligned} \quad (6.8)$$

So ψ is an S valued solution to the Klein-Gordon equation (6.1).

Lemma 6.2 Choose the \mathbb{C}^4 inner product on S and choose the Dirac matrices as in (6.6). Let $\mathcal{H} = L^2(\mathbb{R}^3; S)$. Then the operator H (cf. (6.2)) is self-adjoint on its natural domain in \mathcal{H} . Moreover

$$\text{spectrum } H = (-\infty, -m] \cup [m, \infty) \quad (6.9)$$

Proof. Note first that $\beta^* = \beta$ and $\alpha_j^* = -\alpha_j$ for all j . Let f and g be in

$C_c^\infty(\mathbb{R}^3; S)$. Then

$$(Hf, g)_{L^2(\mathbb{R}^3; S)} = \int_{\mathbb{R}^3} \langle (Hf)(x), g(x) \rangle_S dx \quad (6.10)$$

$$= \int_{\mathbb{R}^3} \left\langle \left(\sum_{j=1}^3 \alpha_j \partial_j f(x) + \beta m f(x), g(x) \right) \right\rangle_S dx \quad (6.11)$$

$$= \int_{\mathbb{R}^3} \left\langle f(x), - \sum_{j=1}^3 \alpha_j^* \partial_j g(x) + \beta^* m g(x) \right\rangle_S dx \quad (6.12)$$

$$= (f, Hg)_{L^2(\mathbb{R}^3; S)} \quad (6.13)$$

So H is symmetric on this domain. We omit technical domain issues because we are going to Fourier transform H anyway.

Let

$$\hat{\psi}(p) = \frac{1}{\sqrt{(2\pi)^3}} \int_{\mathbb{R}^3} e^{-ip \cdot x} \psi(x) dx \quad (6.14)$$

denote the three dimensional S valued Fourier transform of ψ . Then

$$\widehat{(H\psi)}(p) = \left(\sum_{j=1}^3 \alpha_j(ip_j) + \beta m \right) \hat{\psi}(p). \quad (6.15)$$

Thus H is unitarily equivalent to the operator of multiplication by the matrix valued function

$$H(p) = \sum_{j=1}^3 \alpha_j(ip_j) + \beta m \quad (6.16)$$

The same computation that gave (6.8) now shows that

$$H(p)^2 = \left(m^2 + |p|^2 \right) I_S \quad (6.17)$$

The Hermitian 4×4 matrix $H(p)$ has spectrum lying, therefore, in the set $\{\pm\sqrt{m^2 + |p|^2}\}$. In fact the spectrum consists of both of these numbers. To see this let $\gamma_5 = \alpha_1 \alpha_2 \alpha_3 \beta$. Clearly γ_5 anti-commutes with each α_j and with β . Consequently $\gamma_5 H(p) \gamma_5^{-1} = -H(p)$. So the spectrum of $H(p)$ contains a if it contains $-a$. This proves (6.9). ■

Remark 6.3 Let $\gamma_0 = i\beta$ and $\gamma_j = \beta\alpha_j$, $j = 1, 2, 3$. Multiply (6.7) by $i\beta$ to find

$$\left(\gamma_0\partial_t + \sum_{j=1}^3 \gamma_j\partial_j\right)\psi + m\psi = 0. \quad \text{Dirac Equation} \quad (6.18)$$

This equation looks more symmetric between t and the spatial coordinates than does (6.7). In fact, there exists a representation of the covering group of the Lorentz group $\rho : SL(2, \mathbb{C}) \rightarrow End(S)$ whose induced action on the operators $\gamma_0, \dots, \gamma_3$ agrees with the action of the corresponding Lorentz transformation on a basis of \mathbb{R}^4 . This is the basis of Lorentz invariance of the Dirac equation.

STATUS: On the one hand Dirac produced a Lorentz invariant substitute for the Schrödinger equation, namely equation (6.7), and, since H is self-adjoint, the induced propagation e^{itH} in $L^2(\mathbb{R}^3; S)$ is unitary, as it had better be. On the other hand the Hamiltonian H has the very unpleasant feature of being unbounded below, as we see from (6.9). In fact on the negative energy subspace the energy as a function of momentum is $E(p) = -\sqrt{m^2 + |p|^2}$. Thus as the momentum goes up the energy goes down. No way can this be taken seriously. Dirac pointed this out in his first paper, [16], but didn't propose a fix till two years later, [17]. Dirac fixed this by inventing an infinite dimensional version of the Hodge star operator. This, in turn, suggested the existence of an as yet undiscovered particle, the positron. And this, in turn, was discovered a few years later in cosmic rays. Positrons are now available at your local hospital in large numbers if you should need some for a PET scan. The infinite dimensional Hodge star operator goes under the name of the Dirac hole theory.

6.2 Dirac hole theory

Dirac proposed a fix for the negative energy problem, mentioned above, in a second paper, [17], which he submitted two years after the first, [16]. His resolution of the problem could be described in modern terms as the invention of an infinite dimensional Hodge star operator.

Here is a slight variant of the Hodge star operator in finite dimensions, which we will later extend to infinite dimensions. Let V be a finite dimensional complex vector space of dimension n . Let u_1, \dots, u_n be a basis of V

and let v_1, \dots, v_n be the dual basis of V^* . You may already recognize the map

$$R : \sqrt{k!} u_1 \wedge \cdots \wedge u_k \mapsto \sqrt{(n-k)!} v_{k+1} \wedge \cdots \wedge v_n \quad (6.19)$$

as the Hodge star operator, aside from the numerical factors and the lack of an arbitrary choice of subset of the basis. The numerical factors are needed to make R a unitary map from $\Lambda(V)$ onto $\Lambda(V^*)$. You can see from (5.142) and (5.143) that these factors are required if R is to have a unitary extension. We are going first to characterize the desired map R in a basis independent way, avoiding the use of an inner product on V and will scrupulously avoid identifying V with V^* . This is not just for the sake of mathematical purity, but rather, in the application to the electron-positron system, it will be important to use the dual space to the negative energy subspace \mathcal{H}_- , discussed in Section 6.1, as the support space of the contragredient representation of the orthochronous Lorentz group, in order to explain why electrons and positrons have different “intrinsic parity”.

To this end let us observe first that the annihilation operator $a(u)$ is a conjugate linear function of u because u appears in (5.104) on the right side in the inner product. (See also (5.118).) Had we defined the interior product i_w in (5.104) using $\langle x, w \rangle$, with $w \in K^*$ instead of (x, u) with $u \in K$, then i_w as well as a_w would be linear in w and none of the computations would change. The commutation relations (5.129) would read $a_w c(v) + c(v) a_w = \langle v, w \rangle I_{\mathcal{F}_f}$, which captures nicely the bilinearity in v and w . In the following Proposition the roles of K and K^* are interchanged from the previous discussion.

Proposition 6.4 *Let V be a complex vector space of finite dimension n . Choose a non-zero vector $\omega \in \Lambda^n(V^*)$. There exists a unique linear transformation $R : \Lambda(V) \rightarrow \Lambda(V^*)$ such that*

$$a) \quad R(1) = \omega \quad \text{and} \quad (6.20)$$

$$b) \quad a(u)R = Rc(u) \quad \text{for all } u \in V \quad (6.21)$$

Moreover, if V is an inner product space and $\|\omega\| = 1$ then R is unitary.

Proof. Induction, together with a) and b) show that

$$Rc(u_1) \cdots c(u_r)1 = a(u_1) \cdots a(u_r)\omega \quad (6.22)$$

for any vectors $u_1, \dots, u_r \in V$. In view of (5.142) this asserts that

$$\sqrt{r!} R(u_1 \wedge \cdots \wedge u_r) = a(u_1) \cdots a(u_r)\omega \quad (6.23)$$

Given any basis of V we may allow the set $\{u_1, \dots, u_r\}$ to run over all subsets of the basis. The identity (6.23) shows then that R is uniquely determined on a basis of $\Lambda(V)$ by the conditions a) and b). We may use (6.23) to construct R by defining it on that basis and extending it linearly to all of $\Lambda(V)$. The so constructed R is easily shown to satisfy a) and b) on the basis elements of V and therefore for all $u \in V$. Let us emphasize that we have shown that the so constructed map R is basis independent.

Now if V is an inner product space and $\|\omega\| = 1$ then we can choose any orthonormal basis of V , say $\{v_1, \dots, v_n\}$, and take $\omega = c\sqrt{n!} v_1 \wedge \dots \wedge v_n$ for some constant c of absolute value one. Without loss of generality we may take $c = 1$ because it can be absorbed into v_1 . But in view of (5.144) and (5.146) the assertion (6.23) simply says that

$$R(\sqrt{r!} u_1 \wedge \dots \wedge u_r) = \pm \sqrt{(n-r)!} u'_1 \wedge \dots \wedge u'_{n-r} \quad (6.24)$$

where the primed vectors form the complementary set in $\{v_1, \dots, v_n\}$ of the unprimed vectors. By (5.142) and (5.143), R therefore takes an orthonormal basis to an orthonormal basis. ■

Lemma 6.5 *Suppose that A is a diagonalizable linear operator on V . Denote its transpose on V^* by A' . Then*

$$R\gamma(A)R^{-1} = \text{trace}(A) - \gamma(A') \quad (6.25)$$

Proof. Suppose that x_1, \dots, x_n is a basis of V with $Ax_j = \lambda_j x_j$ for $j = 1, \dots, n$. Then $A'y_j = \lambda_j y_j$ for the dual basis. Let $u = x_{i_1} \wedge \dots \wedge x_{i_k}$ and $v = Ru$. Define $\lambda = \sum_1^k \lambda_{i_j}$. Then $\gamma(A)u = \lambda u$ and $\gamma(A')v = (\text{trace}(A) - \lambda)v$. So $R\gamma(A)R^{-1}v = R\gamma(A)u = R\lambda u = \lambda v = (\text{trace}(A)v - \gamma(A')v)$ ■

How to use the identities (6.21) and (6.25). Suppose that V is the one-particle Hilbert space for some Fermion and that A is its Hamiltonian. We will temporarily take V to be finite dimensional. Since the particle is a Fermion the Hilbert space for an indefinite number of such particles is the exterior algebra $\Lambda(V)$. The total Hamiltonian is $\gamma(A)$ (see (5.149)), which acts on $\Lambda(V)$. As usual, without changing the physics, one can change the Hilbert state space by a unitary operator if one also transfers the observable operators. By Lemma 6.5 we can change the Hilbert space by the unitary operator R to $\Lambda(V^*)$ and use for the total Hamiltonian $\gamma(-A') + \text{trace}(A)I_{\Lambda(V')}$

Since changing a Hamiltonian by an additive constant does not change the physics we can drop the last term and take

$$H = \gamma(-A') \quad \text{acting on} \quad \Lambda(V^*) \quad (6.26)$$

as the Hamiltonian for our system of an indefinite number of Fermions. As will be explained shortly, the case of interest for us is that in which A is a negative operator. In that case $-A'$ is a positive operator. Thus we have made a unitary transform and added a constant and thereby transformed a negative Hamiltonian, $\gamma(A)$, into a positive Hamiltonian $\gamma(-A')$. At the same time we have also interchanged the creation and annihilation operators, as we see from (6.21) and the following exercise. Exercise: For all $y \in V^*$ there holds

$$c(y)R = Ra(y) \quad (6.27)$$

Interpretation. The creation, annihilation and energy operators are automatically transformed by the unitary operator R onto a new Hilbert space $\Lambda(V^*)$. The significance of the states in $\Lambda(V^*)$ need interpretation. The zero rank tensor $1 \in \Lambda^0(V^*)$ is the image under R of the top rank element $w \in \Lambda^n(V)$, as can be seen immediately from (6.24), with $r = n$. w is an n particle state for the Fermion whose state space is V , even though it is now represented, via R , by the zero rank tensor $1 \in \Lambda^0(V^*)$. In the state w all possible energy levels of the one-particle Hamiltonian A are filled.

Terminology. The n^{th} rank tensor $w \in \Lambda^n(V)$ is called a “sea” of V particles. Under the Hodge map R , the sea is mapped to the zero rank tensor $1 \in \Lambda^0(V')$. Thus the zero rank tensor $1 \in \Lambda^0(V')$ now represents the “sea” of particles whose 1-particle state space is V . Hence forth we will refer to this particle as a positron. So one says that the zero rank tensor $1 \in \Lambda^0(V')$ represents a “sea” of positrons, rather than a state with no particles in it. Similarly, a rank one tensor $y \in \Lambda^1(V')$ represents a state of $n - 1$ positrons, i.e., a missing positron, because it is the image under R of a rank $n - 1$ tensor in $\Lambda(V)$. This is a “hole” in the sea. A second rank tensor in $\Lambda^2(V')$ is similarly not a two particle state but represents two holes in the sea of n positrons.

Our case. We saw in our analysis of the Dirac equation that the Dirac Hamiltonian H , defined in (6.2), decomposes $L^2(\mathbb{R}^3; S)$ into two orthogonal subspaces \mathcal{H}_- and \mathcal{H}_+ on which H has spectrum $(-\infty, -m]$ or $[m, \infty)$,

respectively. Denote by H_- the restriction of H to \mathcal{H}_- . This is a strictly negative operator and is unbounded below. It is unacceptable as a quantum mechanical Hamiltonian. But if we apply the procedure developed above to the vector space $V = \mathcal{H}_-$ and $A = H_-$ then we will get an (informally) equivalent description of the quantization of H_- on \mathcal{H}_- by using instead $-H'_-$ on $(\mathcal{H}_-)^*$. We have then a one “particle” Hamiltonian $-H'_-$ which is non-negative. So all is well. But the correct physical interpretation of the zero rank tensor $1 \in \Lambda^0((\mathcal{H}_-)^*)$ is now not that of a no particle state but that of a “sea” of infinitely many positrons. (infinitely many because $\dim \mathcal{H}_- = \infty$.) And a rank one tensor is a hole in that sea. The total Hamiltonian for the Hilbert space of holes, $\Lambda((\mathcal{H}_-)^*)$ “is” $\gamma(-H'_-)$, which acts on $\Lambda((\mathcal{H}_-)^*)$ (actually on its completion, \mathcal{F}_f) and is a non-negative operator. In this discussion of “our case” we have evaded the fact that $\text{trace } H_- = -\infty$. We used the finite dimensional heuristic to motivate our choice of state space as $\Lambda((\mathcal{H}_-)^*)$ and total Hamiltonian as $\gamma(-H'_-)$. In the physics literature the same conclusion is reached, but with the computation made directly in the infinite dimensional case. Thus one has to subtract an infinite “constant” from the total Hamiltonian at some point. Some readers may find that disturbing. But we have already been through a similar heuristic procedure back in Section 5.4, when we subtracted the infinite ground state energy. In both cases we developed a well defined finite dimensional approximation, made a meaningful subtraction of a finite constant from some Hamiltonian, and then took the informal limit as a *definition* of the desired structure. In our present case the change of Hilbert space to $\Lambda((\mathcal{H}_-)^*)$ entails a big conceptual change in the physical meaning of the state vectors. This change turns out to be a serious guide as to what happens physically when the full interaction is put into the theory. (See next section on pair production.)

The infinite dimensional Hodge star operation of Dirac. For any real or complex Hilbert space K denote by $\Lambda(K)$ the Fermion Fock space \mathcal{F}_f over K . As we know, (make sure this is stated in Fock space section) the creation and annihilation operators $c(u)$ and $a(u)$ are bounded operators on $\Lambda(K)$. Denote by $\mathcal{A}(K)$ the C^* algebra of operators on $\Lambda(K)$ generated by all the operators $\{c(u), a(u) : u \in K\}$. Proposition 6.4 asserts that, in case K is finite dimensional, then the unitary operator R constructed there interchanges the algebra $\mathcal{A}(K)$ with $\mathcal{A}(K^*)$ by mapping

$$c(u) \mapsto a(u) \equiv R c(u) R^{-1} \quad (6.28)$$

If $\dim K = \infty$ then there is no unitary operator R because there is no highest rank vector $\omega \in \Lambda^\infty(K^*)$. However the isomorphism of C^* algebras induced by R , as in (6.28), continues to make sense in infinite dimensions. This is the isomorphism proposed by Dirac, [17]. Here is a precise statement.

Theorem 6.6 (*Dirac-Hodge hole theorem*) *Let K be a real or complex Hilbert space. Then there exists a unique algebra isomorphism*

$$\beta : \mathcal{A}(K) \rightarrow \mathcal{A}(K^*) \quad (6.29)$$

such that

$$\beta(B^*) = (\beta(B))^* \quad \text{for all } B \in \mathcal{A}(K) \quad (6.30)$$

and

$$\beta(c(u)) = a(u) \quad \text{for all } u \in K \quad (6.31)$$

In particular $\beta(c(u)^*) = a(u)^*$ for all $u \in K$.

Proof. See the appendix of [25]. Note: $a(u)$ is the annihilation operator on $\Lambda(K^*)$ defined by an element $u \in K$. See discussion preceding Proposition 6.4. ■

Remark 6.7 The action of the isomorphism β on a negative one particle Hamiltonian A , as in (6.25), is best understood in the context of a group representation on K and its contragredient representation on K^* . For example if A is a negative self-adjoint operator on K and $g(t) = e^{itA}$ is the one parameter group that it generates, then the contragredient representation of whatever Lie group, G , might contain this time translation generator would map these operators to $(g(t)^{-1})^{\text{transpose}}$, which equals $e^{-itA'}$. Thus, passing to the contragredient representation changes the sign of generators. The case of interest for the Dirac hole theory is that in which G is the orthochronous Poincaré group. It acts unitarily on Dirac's Hilbert space $\mathcal{H} \equiv L^2(\mathbb{R}^3; \mathbb{C}^4)$ and leaves invariant the positive and negative energy subspaces \mathcal{H}_+ and \mathcal{H}_- . We may restrict this representation to the negative energy subspace \mathcal{H}_- and then take its contragredient representation on $(\mathcal{H}_-)^*$. The resulting representation of G is now a positive energy representation. It is the representation that belongs to positrons (and actually characterizes positrons). It happens that this representation of G is unitarily equivalent to the representation of G gotten by restricting our original representation on \mathcal{H} to the positive energy subspace \mathcal{H}_+ , but only for the connected component of G , not for all of G .

The other component of G is generated by space reflection. Thus one says that a positron differs from an electron not only in charge but also in parity. See the Appendix of [25] for more details.

6.3 Pair production

The interchange of electron creation and annihilation operators described in the previous section on hole theory leads to an interaction Hamiltonian containing terms which represent the mutual destruction of a positron and electron and the replacement of such a pair by a photon. Other terms represent the annihilation of a photon and the creation of an electron-positron pair in its place. This greatly enlarges the family of Feynman diagrams that enter at any order. Compare the diagrams in Section 5.8 for simple versions of such diagrams. In Section 5.8 the interaction does not change the number of electrons.

{In some future version of this seminar this section should be greatly enlarged.}

7 The Road to Yang-Mills Fields

Today's question: Where do connections on vector bundles come from in quantum mechanics?

Today's answer: (Outline)

1) The Lorentz force (3.41) on a charged particle depends on the velocity of the particle. In order to incorporate this force into quantum mechanics one must modify the rules of quantization that we previously discussed for velocity independent forces in Section 4.3, so as to include this force. As a first step one must express the time dependent force with the help of suitable potentials. We did this by

2) first expressing Maxwell's equations in the form $DF = 0$, $D^*F = J$ where F is the 2-form on space-time constructed from E and B as in (3.31). D denotes the exterior derivative on forms over \mathbb{R}^4 (\equiv space-time). See Section 3.8 for this.

3) Since $DF = 0$ there is a 1-form A on space-time such that

$$F = DA. \quad (7.1)$$

The 1-form A replaces the potential V , which we used earlier to express a velocity independent force, $\mathbf{F} = -\text{grad } V$. Whereas V was unique up to an additive constant, the 1-form A is unique only up to an additive exact 1-form. The big mathematical problems which will be induced by this large degree of non-uniqueness was, admittedly, not even mentioned earlier.

4) Having now an electromagnetic analog of the potential V , we made the transitions from Newton to Lagrange to Hamilton for the Lorentz force in Section 3.8, with the resulting Hamiltonian function (cf. (3.51))

$$H(x, p, t) = \frac{1}{2m} \sum_{j=1}^3 (p_j - \frac{e}{c} \mathbf{A}_j(x, t))^2 + e\phi(x, t), \quad (7.2)$$

where

$$A = \sum_{j=1}^3 \mathbf{A}_j dx_j + \phi dt. \quad (7.3)$$

5) We will carry out the quantization procedure for the Hamiltonian function (7.2) in Section 7.1.

6) We will then discuss the geometric meaning of the resulting Schrödinger equation. Specifically, we will see that the Schrodinger equation in the presence of electromagnetic forces begs for an interpretation of the electromagnetic potential A as a connection form for a complex line bundle $\mathbb{C} \times \mathbb{R}^4 \rightarrow \mathbb{R}^4$ with structure group $U(1)$.

7) Yet to come: Whereas the Schrödinger equation that we will derive in Section 7.1 lends itself to interpretation as an equation for sections of a complex line bundle over \mathbb{R}^4 with structure group $U(1)$, the discovery of the neutron in 1932 by Chadwick and the ensuing attempts to understand the strongly attractive forces between n- n, p-p and n-p, culminated in the suggestion by Yang and Mills (1954) to replace the complex line bundle with $U(1)$ structure group by a vector bundle with a more general compact structure group. The correct choice of structure group has to be determined by experiment. At the present time the choice $SU(3) \times SU(2) \times U(1)$ seems to fit best all the experimental data, and is the basis for the standard model of elementary particles. I will elaborate a bit on this item 7) in Section 8. Some more of the history of the period 1932-1954 is in the timeline of Section 8.1.

7.1 Quantization of a particle in an electromagnetic field

We will apply the usual rules of quantization, as in Section 4.3, to a charged particle without spin, subject to electromagnetic forces. The electromagnetic field exerting the force is assumed to satisfy Maxwell's equations (3.33) - (3.36), or equivalently (3.37), (3.38). There is, therefore, an electromagnetic potential A , which is a 1-form on space-time. We will take over the notation from (7.2) and (7.3).

Let

$$\mathcal{H} = L^2(\mathbb{R}^3). \quad (7.4)$$

In accordance with the rules of Section 4.3 one should replace p_j in (7.2) by

$-i\hbar(\partial/\partial x_j)$. Then we obtain the time dependent Hamiltonian operator

$$H = \frac{1}{2m} \sum_{j=1}^3 (-i\hbar(\partial/\partial x_j) - (e/c)\mathbf{A}_j)^2 + e\phi(\cdot, t) \quad (7.5)$$

$$= -\frac{\hbar^2}{2m} \sum_{j=1}^3 \left(\frac{\partial}{\partial x_j} - i\frac{e}{c\hbar}\mathbf{A}_j \right)^2 + e\phi(\cdot, t) \quad (7.6)$$

The Schrödinger equation (4.3) may therefore be written

$$i\hbar \left(\frac{\partial}{\partial t} + i\frac{e}{\hbar}\phi \right) \psi = -\frac{\hbar^2}{2m} \sum_{j=1}^3 \left(\frac{\partial}{\partial x_j} - i\frac{e}{c\hbar}\mathbf{A}_j \right)^2 \psi \quad (7.7)$$

This is the
Schrödinger equation in the presence of an electromagnetic field.

Define

$$D_0 = \left(\frac{\partial}{\partial t} + i\frac{e}{\hbar}\phi \right) \text{ and} \quad (7.8)$$

$$D_j = \left(\frac{\partial}{\partial x_j} - i\frac{e}{c\hbar}\mathbf{A}_j \right) \quad (7.9)$$

{Check signs.} In terms of these operators we may write the Schrödinger equation as

$$i\hbar D_0 \psi = -\frac{\hbar^2}{2m} \sum_{j=1}^3 D_j^2 \quad (7.10)$$

Don't think that (7.10) is just a notationally slick way of writing (7.7). Notice that the differential operators D_k , $k = 0, 1, 2, 3$ differ from the usual first derivatives, $\partial/\partial t, \partial/\partial x_j$ by the addition of purely imaginary functions. So? Well, the circle group, $U(1) \equiv \{e^{i\theta} : \theta \in \mathbb{R}\}$, has Lie algebra $i\mathbb{R}$. Therefore the operators D_k can be interpreted as covariant derivatives for sections of a complex line bundle over \mathbb{R}^4 with structure group $U(1)$. That is, if one regards $\psi(x, t)$ not as a complex valued function but rather as a section of the bundle $\mathbb{C} \times \mathbb{R}^4 \rightarrow \mathbb{R}^4$, then specification of an electromagnetic force field amounts to choosing a particular covariant derivative on sections of this bundle - a covariant derivative different from the trivial one.

Is this interpretation just a stretch of ones imagination? Think back to the non-uniqueness inherent in the representation (7.1), of the fields E and B , by the potential A . For any reasonable (say smooth) function $f : \mathbb{R}^4 \rightarrow \mathbb{R}$ the 1-form $\hat{A} \equiv A + Df$ represents the same electromagnetic field as A because $D^2f = 0$. Now the function $e^{if(x,t)}$ is a $U(1)$ valued function on \mathbb{R}^4 . Therefore the map $\psi \mapsto \hat{\psi} = e^{if}\psi$ is a bundle automorphism. That is, ψ and $\hat{\psi}$ represent the same section, but in different (global) trivializations of the bundle. Now look at this simple identity!

$$(\partial_j - i(\mathbf{A}_j + \partial_j f))e^{if}\psi = e^{if}(\partial_j - i(\mathbf{A}_j))\psi, \quad j = 0, 1, 2, 3 \quad (7.11)$$

This says that the non-uniqueness of the potential A is equivalent to the non-uniqueness of the representative of a section. Otherwise said, it is the covariant derivative, $\partial_j - i(\mathbf{A}_j)$, that has physical meaning rather than the potential A itself. The Lie algebra valued 1-form iA is the connection form that represents the covariant derivative in a particular (global) trivialization. Change the global trivialization and you change the connection form without changing the covariant derivative.

CONCLUSION: The quantum mechanical procedure for incorporating electromagnetic forces into the Schrödinger equation is commensurable with (or is suggestive of, or indeed forces) an interpretation of the electromagnetic potential iA as a connection form on \mathbb{R}^4 for a complex line bundle $\mathbb{C} \times \mathbb{R}^4 \rightarrow \mathbb{R}^4$ with structure group $U(1)$. In short, one could fairly say

Quantum Mech. + E&M = Covariant derivative.

Remark 7.1 Hermann Weyl [66] seems to be the first to have advanced this view of electromagnetism. His objective was to find a way to unite general relativity with electromagnetism.

Reference: A useful source for background on connections on vector bundles is [15].

7.2 Two classic debuts of connections in quantum mechanics

Before leaving S^1 bundles there are two important examples to understand in which the underlying manifold is not topologically trivial. If the configuration

space \mathcal{C} is not \mathbb{R}^3 but, say, an open subset of \mathbb{R}^3 then space-time is $\mathcal{C} \times \mathbb{R}$, a 4-manifold, and all of the preceding discussions about Maxwell's equation, gauge potentials and complex line bundles go through if one does not insist on global trivializations but only local trivializations. In the following two examples \mathcal{C} is, respectively, \mathbb{R}^3 minus a cylinder and \mathbb{R}^3 minus a point. Each case arises naturally in a physical experiment.

Before leaving this transition from electromagnetic fields to geometric objects here is a summary of the terminology used in physics and the corresponding mathematical terminology. This is taken from the "dictionary paper" of Wu and Yang [71].

Dictionary

4-potential = gauge potential = connection form

Field strength= gauge field = curvature

Nonintegrable phase factor = path dependent parallel transport

7.2.1 The Aharonov-Bohm experiment

Denote by S the closed cylinder in \mathbb{R}^3 centered on the z axis and of radius a . That is $S = \{(x, y, z) : x^2 + y^2 \leq a^2\}$. Take configuration space to be $\mathcal{C} = \mathbb{R}^3 - S$ Let b be a real constant and define

$$A = b \begin{cases} (-y, x, 0)/(x^2 + y^2) & \text{outside } S \\ (-y, x, 0)a^{-2} & \text{inside } S \end{cases} \quad (7.12)$$

The magnetic field, $B = \text{curl } A$, is zero in \mathcal{C} and is equal to $(0, 0, 2ba^{-2})$ for all time inside the cylinder S , as one can readily compute. Moreover we will take $\phi = 0$ on \mathbb{R}^4 . Both E and B are time independent. So we will ignore time and focus just on the geometry over \mathcal{C} .

In the Aharonov-Bohm experiment one shoots a beam of electrons parallel to the x axis and observes the pattern on a vertical plate behind the cylinder and perpendicular to the x axis. Make the cylinder surface out of a (near) perfect conductor, such as gold, so that no electrons can get inside the cylinder. Those that hit the cylinder are conducted off to infinity. Thus all electrons stay in a region, \mathcal{C} , where $E = B = 0$. Nevertheless one observes a diffraction pattern on the screen which varies with b . The Schrodinger

equation, (take $e/c\hbar = 1$)

$$-i\partial\psi/\partial t = -(1/2m) \sum_{j=1}^3 (\partial_j + iA_j)^2 \psi(x, t)$$

in $\mathcal{C} \times \mathbb{R}$ (Dirichlet boundary conditions on ∂S) accurately predicts the diffraction pattern! [Refs., Christianne Martin and Nintendo toy company]

Conclusion: Even though the magnetic and electric forces are zero in \mathcal{C} , where the electrons are constrained to lie, they nevertheless sense the connection form itself. Geometrically speaking, the curvature of this connection is zero while the holonomy group of A in \mathcal{C} is not zero. In any ball in \mathcal{C} one can change the local trivialization to reduce A to zero in the ball without changing the physics. Of course this does not change the holonomy group. Thus one could say that the electrons really sense the holonomy group of the connection. This is the observation made by Wu and Yang [71].

Experimental confirmation: The prediction made by Aharonov and Bohm, [1] in 1959, was experimentally confirmed by R. G. Chambers in 1960 [11].

For some other interesting paradoxes see the book by Aharonov and Rohrlich [2].

Exercise 7.2 Show that the holonomy around a circle of radius $r > a$ enclosing the cylinder is $e^{2\pi i b}$.

7.2.2 The Dirac magnetic monopole

The equation defining electric charge as the source of an electric field is Gauss' law (3.9), $\text{div } E = 4\pi\rho$ where ρ is the electric charge density. If magnetic charge exists, its density, M is defined by $\text{div } B = 4\pi M$ because magnetic charge is, by definition, a source of the magnetic field, B . We want to consider a point magnetic charge, i.e. a magnetic monopole, sitting still at the origin. Gauss' law would then be $\text{div } B(x) = 4\pi p\delta(x)$ where p is the pole strength (magnetic charge) of the monopole. The fundamental solution to this equation is $B(x) = px/(|x|^3)$, $x \in \mathbb{R}^3 - \{0\}$. In order to formulate the Schrödinger equation for this force law we need to use, as usual, a corresponding gauge potential A . In order to avoid ambiguities generated by the singularity in B at the origin of \mathbb{R}^3 one is forced to take configuration space to be $\mathcal{C} = \mathbb{R}^3 - \{0\}$. This immediately makes the topology of \mathcal{C} nontrivial. As

we will see this nontrivial topology offers a way out of a serious technical and conceptual issue that will soon insert itself. The structure we want to investigate is time independent because the magnetic field B is time independent, the electric field is identically zero, and the magnetic monopole is sitting still at the origin. We will therefore ignore the time factor in $\mathcal{C} \times \mathbb{R}$ and focus only on the relevant geometry over the configuration space \mathcal{C} . In particular we need to find a connection on the bundle $\mathcal{C} \times \mathbb{C} \rightarrow \mathcal{C}$ whose curvature is B . Now $\text{div } B = 0$ in \mathcal{C} because the monopole is not in \mathcal{C} . Hence locally there exist 1-forms A such that $B = dA$. But in fact there is no globally defined 1-form A on \mathcal{C} for which $B = dA$. Here is the best we can do.

Let $0 < \delta < \pi/2$ and, in spherical coordinates, $r \geq 0, \theta \in [0, \pi], \phi \in [0, 2\pi]$, let

$$N = \{(r, \theta, \phi) : r > 0, 0 \leq \theta < \pi/2 + \delta\} \quad (7.13)$$

$$S = \{(r, \theta, \phi) : r > 0, \pi/2 - \delta < \theta \leq \pi\} \quad (7.14)$$

Then $N \cup S = \mathcal{C}$ gives an open cover of \mathcal{C} by contractible sets. $N \cap S$ intersects any sphere around the origin in a band around the equator. Define

$$A^N = p(1 - \cos \theta)d\phi \text{ on } N \quad (7.15)$$

$$A^S = -p(1 + \cos \theta)d\phi \text{ on } S \quad (7.16)$$

We assert that $B = dA^N$ on N and $B = dA^S$ on S . To see this easily observe that $dA^N = p \sin \theta d\theta \wedge d\phi$ on N and $dA^S = p \sin \theta d\theta \wedge d\phi$ on S . Each of these is therefore equal to p times the element of area on the sphere of radius 1. But B (after applying the Hodge star to convert it into a 2-form) is $B = (p/r^3)\{xdy \wedge dz + ydz \wedge dx + zdx \wedge dy\}$. At the north pole of the sphere of radius r we have $x = y = 0, z = r$. Consequently, near the north pole, the factor in braces is $r \times$ element of area $= r^3$ times element of area of the unit sphere. Thus, by spherical symmetry of B , A^N and A^S both have curvature B .

Let $h(r, \theta, \phi) = A^N - A^S$ on $N \cap S$. Then $h = 2pd\phi$ on $N \cap S$. Assume now that $2p$ is an integer and let $g(r, \theta, \phi) = e^{i2p\phi}$. This is a well defined smooth function on $N \cap S$ because $2p$ is an integer. Moreover $iA^N = iA^S + g^{-1}dg$. Consequently iA^N and iA^S together with g define a connection on a complex line bundle over \mathcal{C} with local trivializations $N \times \mathbb{C}$ and $S \times \mathbb{C}$ and with transition function g . Moreover the curvature of this connection is B . If we reinsert the physical constants $e/(c\hbar)$ in front of the electromagnetic potential

A then the condition on the pole strength needed to make $ie/(\hbar c)A^N$ and $ie/(\hbar c)A^S$ into a well defined connection on a well defined complex line bundle over \mathcal{C} as above is

$$\frac{2pe}{\hbar c} = \text{integer}$$

This is Dirac's quantization condition [18] relating pole strength p and electron charge e . This condition is also necessary, as we shall see in the next theorem. But first, it may be illuminating to see why A^N cannot be extended to all of \mathcal{C} . Convert A^N to cartesian coordinates by using $\phi = (\tan)^{-1}(y/x)$, which implies $d\phi = (xdy - ydx)/(r^2 - z^2)$, and multiply (7.15) by r/r to find

$$A^N = \frac{p(xdy - ydx)}{r(r + z)}.$$

Thus A^N has a singularity along the entire negative z axis. A similar computation shows that, if one chooses $\delta = \pi/2$, then A^S has a singularity along the entire positive z axis. Dirac's view of this, (1931), was that the monopole forces the singularity in the electromagnetic vector potential to extend along a curve from the origin to infinity. But Wu and Yang [71] (1975) emphasized the viewpoint which we have described here: the complex line bundle over \mathcal{C} , associated to a magnetic monopole, does not have a global trivialization.

The following theorem adds to the previous discussion, showing in what sense Dirac's quantization condition, $2p = \text{integer}$, is necessary.

Theorem 7.3 *There exists an S^1 connection over $\mathbb{R}^3 - \{0\}$ whose curvature is (Hodge *) $(px/|x|^3)$ if and only if $2p$ is an integer.*

Proof. The sufficiency has already been proved. Define N and S as above. Assume that there exists a connection over \mathcal{C} whose curvature is $B \equiv *px/|x|^3$. Since N and S are contractible and $dB = 0$ on \mathcal{C} there exist 1-forms A^N on N and A^S on S satisfying $dA^N = B$ and $dA^S = B$ on N and S respectively. Since these forms represent the same connection on the presumed bundle over \mathcal{C} there exists a smooth function $g : N \cap S \rightarrow S^1$ such that $A^N = A^S + g^{-1}dg$

on $N \cap S$. Then

$$\begin{aligned}
4\pi p &= \int_{|x|=1} p \sin \theta d\theta \wedge d\phi \\
&= \int_{|x|=1} B \\
&= \int_{0 \leq \theta \leq \pi/2} dA^N + \int_{\pi/2 \leq \theta \leq \pi} dA^S \\
&= \int_{\theta=\pi/2} A^N - \int_{\theta=\pi/2} A^S \\
&= \int_{\theta=\pi/2} g(1, \pi/2, \phi)^{-1} \partial_\phi g \\
&= 2\pi \text{ winding number of } \phi \rightarrow g(1, \pi/2, \phi) \\
&= 2\pi \text{ integer}
\end{aligned}$$

Hence $2p = \text{integer}$. ■

Remark 7.4 There is a huge literature on the experimental and theoretical aspects of magnetic monopoles. For a survey up to 1990 see [22]. An inquiry to mathscinet: Anywhere magnetic monopoles brings up 970 matches, as of June 7, 2011.

8 The Standard Model of Elementary Particles

Ran out of time, although this topic was the real goal of the seminar.

Fortunately, you have available Chapter 9 of Folland, [21], which gives tremendous perspective, both historical and technical, into the evolution and structure of the standard model.

Another excellent book describing the relation between group theory and, among other things, the classification of elementary particles is the book [55] by Shlomo Sternberg.

Maybe the next time this seminar is taught this topic can be squeezed in. It will require an exposition of connections on vector bundles. See for example [15].

The exposition of connections on vector bundles in [42] is long on the mathematics and its relation to important topics within mathematics, e.g., application to Donaldson polynomials, Jones-Witten invariants, etc. But it also explains the relation of this mathematics to physicist's gauge field theory.

Anyway, the main idea linking what we have done to the currently most successful classification of elementary particles is now simple to describe.

In Section 7.1 we saw that the combination of electromagnetic theory with quantum mechanics leads to an interpretation of the electromagnetic potential as a connection form for the vector bundle $\mathbb{C} \times \mathbb{R}^4 \rightarrow \mathbb{R}^4$ with structure group $U(1)$, the circle group. Consider the following simple generalization of this. Take your favorite compact Lie group K and an irreducible unitary representation of K on a finite dimensional inner product space V . The map $V \times \mathbb{R}^4 \rightarrow \mathbb{R}^4$ is then a vector bundle with structure group K . Denote by \mathfrak{k} the Lie algebra of K . Then a \mathfrak{k} valued 1-form A on \mathbb{R}^4 determines a covariant derivative for this bundle. Of course a change in the global trivialization of this bundle will induce a change in the 1-form A , if we interpret this 1-form as a connection form for this bundle, as we did in the electromagnetic case, and as we will do now. The associated covariant derivative on this bundle now represents some kind of force field as it did in the electromagnetic case. In the electromagnetic case the particles that implemented the electromagnetic force were photons. But in this more general case they are some other kind of particles, called gluons. The interesting case is that in which K is non-commutative. The curvature of a covariant derivative can be expressed in terms of a connection form A for it as

$$R = dA + A \wedge A. \quad (8.1)$$

This is a \mathfrak{k} valued 2-form on \mathbb{R}^4 . If $\mathfrak{k} = i\mathbb{R}$ then the second term in (8.1) is zero and the curvature reduces to the 2-form defining the electric and magnetic fields, as in (3.31). In the non-commutative case we should interpret the six independent \mathfrak{k} valued components of R as "non-commutative" "electric" and "magnetic" fields. These are the force fields associated to the connection form A . The Maxwell equations, (3.37), are now replaced by the Bianchi identity, $DR = 0$, which is automatic when R is the curvature, (8.1). And the Maxwell equations (3.38) are replaced by the equation $D^*R = J$ - the *Yang-Mills equation*. In this geometric way we arrive at an extension of

electromagnetic theory. At the present time these new fields seem to provide the best basis for a comprehensive theory of elementary particles.

What are these force fields acting on? They are acting on whatever particles are being represented by sections of the vector bundle $V \times \mathbb{R}^4 \rightarrow \mathbb{R}^4$. In the electromagnetic case V was one dimensional and the electric and magnetic fields were acting on one charged particle, say an electron. But if V is e.g. eight dimensional then a section represents the state of any one of eight particles, one for each of your favorite basis vectors of V . Of course, given the group K , the vector space V is limited to those that support an irreducible unitary representation of K . Moreover the elements of the Lie algebra of K will transfer to observables, as we learned in Section 4.8. Here are the names of some of these observables: isotopic spin, hypercharge, strangeness, beauty, charm. { The last three are not quite right. See Griffith's book, [24]. } { Concerning spin, the previous discussion is oversimplified. Spin comes not from the group K but from a more geometric source, the action of the Lorentz group on Dirac wave functions, as in Section 6. From the point of view of the preceding discussion, one can just replace V by $\mathbb{C}^4 \otimes V$. But K acts only on V . } In the standard model one takes $K = SU(3) \times SU(2) \times U(1)$.

I'm reluctant to leave the reader with the impression that this simple switch from structure group $U(1)$ to a general compact structure group was invented by someone just "thinking upon it", as Newton claims to have done with some of his own theories. In fact the evolution of this extension from electromagnetism to non-commutative structure group was driven by experiments. Even though Rutherford (1911) used alpha particles to determine the structure of an atom (small nucleus inside big atom) he didn't really know what an alpha particle was. Yes, it had charge +2 and mass = 4 proton masses. It could have been 4 protons stuck together with 2 electrons embedded. Or it could have been 2 protons + 2 neutrons (if they exist). It wasn't till 1932 that the British experimental physicist, Chadwick, showed that neutrons exist. In any case the electric forces between any two protons are pushing the protons apart with tremendous force because the two protons are so close and have the same charge. So what's holding the nucleus together? There must be another force between any of these nucleons (protons or neutrons), an attractive force, even stronger than the electric force, that beats out the repulsive electric force. Heisenberg, immediately upon hearing of Chadwick's discovery, made a (rather ad hoc, I would say) mathematical formalism to allow the possibility that a proton and a neutron are

really the same particle, but in different states. The proton is in a charged state. The mysterious attractive forces between nucleons are independent of their charge, proposed Heisenberg. The charge only affects the repulsive electric force, [31]. See also Heisenberg's related papers [32, 33, 34].

This hypothesis of charge independence of the nuclear force can be tested experimentally. It predicts that the isotopes N_{15} and O_{15} should have a similar radiation spectrum from their nuclei even though N_{15} has 7 protons and 8 neutrons while O_{15} has 8 protons and 7 neutrons. The prediction was experimentally confirmed between 1932 and 1936 (get dates). This was an impetus for Eugene Wigner to amend Heisenberg's simple formalism and promote invariance of nuclear forces under Heisenberg's interchange operator (proton \leftrightarrow neutron) to invariance under $SU(2)$. You might like to read the extract from Wigner's paper, reproduced in the next section, to be sure that I'm not just putting words in his mouth, or thoughts in his mind.

It was another 17 years before Yang and Mills promoted Wigner's global $SU(2)$ invariance to "local" invariance" which for us means vector bundles with non-commutative structure group.

8.1 The neutron and non-commutative structure groups. Timeline.

1911 (May): Rutherford determined the structure of an atom: small nucleus (radius = 10^{-13} cm) compared to the overall radius of the atom (10^{-8} cm.)

1932 (May 10): Chadwick provided convincing evidence that neutrons exist, [10, 9].

1932 : Heisenberg, motivated by Chadwick's discovery, made an ad hoc theory of nuclear forces. It predicts that the isotopes N_{15} and O_{15} should have similar radiation spectrum even though N_{15} has 7 protons and 8 neutrons while O_{15} has 8 protons and 7 neutrons, [31]. See also Heisenberg's related papers [32, 33, 34]. The prediction was experimentally confirmed between 1932 and 1936 (get dates).

1937: Eugene Wigner, [69], promoted Heisenberg's invariance of neutron-proton interactions from invariance under Z_2 (interchange of proton and neutron wave functions) to invariance under $SU(2)$. Thereafter $SU(2)$ is referred to as the *isotopic spin group*, and must be distinguished from the ordinary

spin group $SU(2)$ (which covers $SO(3)$ and thereby acquires an understandable geometric meaning).

From the first paragraph of Wigner's paper.

Recent [experimental] investigations [two references are given] appear to show that the forces between all pairs of constituents of the nucleus are approximately equal. This makes it desirable to treat the protons and neutrons on an equal footing. A scheme for this was devised in his original paper by W. Heisenberg, who considered protons and neutrons as different states of the same particle. Heisenberg introduced a variable τ which we shall call the isotopic spin. The value -1 of this variable can be assigned to the proton state of the particle, the value +1 to the neutron state. The assumption that the forces between all pairs of particles are equal is equivalent, then to the assumption that the forces do not depend on τ , or that the Hamiltonian does not involve the isotopic spin.

In Heisenberg's scheme τ is the 2×2 matrix $\text{diag}(-1,1)$ acting on \mathbb{C}^2 . Wigner gave arguments, based on the experimental evidence, for asserting that a potential operator V on $L^2(\mathbb{R}^3) \otimes \mathbb{C}^2$ should be invariant not just under $I \otimes Z_2$ but also under $I \otimes SU(2)$ if V is to represent nuclear forces consistently with experiment.

1954: Yang and Mills [72] promoted Wigner's $SU(2)$ invariance to a local $SU(2)$ invariance. This means that, whereas Wigner allowed invariance of his and Heisenberg's nuclear force theory to be invariant under the map $\psi(x, t) \mapsto g\psi(x, t)$ for any fixed element $g \in SU(2)$ (recall that ψ takes values now in \mathbb{C}^2), Yang and Mills allowed g to depend on space and time. This local $SU(2)$ "gauge invariance" is linked to the existence of $SU(2)$ "gauge fields" in the same way that the non-uniqueness of the electromagnetic potential A is linked to the electromagnetic field. See the discussion surrounding equation (7.11).

9 Appendices

9.1 The Legendre transform for convex functions

4/27/11 This appendix contains some redundancies. Compare with Section 2.7.1. I replaced these two sections by Section 2.7.1 on 4/27/11. So the ap-

pendix on the Legendre transform (in general and for quadratic polynomials) isn't needed at all for these notes.

As a cultural matter you should know that the formulas derived in Section 9.1.1 for quadratic plus linear functions of velocity fit into a very clean and general scheme of transforms. I will sketch this here, so that you don't think that the formulas of Section 9.1.1 are just ad hoc computations. However we will only need the results of Section 9.1.1 for our applications to Hamiltonian mechanics. The more general case is needed for thermodynamics and other parts of mathematics. You can skip this section without too much guilt.

The Legendre transform is a map from convex functions on a real vector space V to convex functions on the dual space. Here are some definitions. We take V to be finite dimensional. A function f with domain $D(f) \subset V$ is **convex** if $D(f)$ is convex and $f(ax + (1 - a)y) \leq af(x) + (1 - a)f(y)$ for all $x, y \in D(f)$ and for all $a \in [0, 1]$. The **epigraph** of f is the set $\{(v, y) \in D(f) \times \mathbb{R} : y \geq f(v)\}$. A convex function is **closed** if its epigraph is closed in $V \times \mathbb{R}$. Let $p \in V^*$. A hyperplane in $V \times \mathbb{R}$ with slope p is the graph of the function $v \rightarrow \langle p, v \rangle + c$ for some constant $c \in \mathbb{R}$. This hyperplane is a **support plane** for f if $\langle p, v \rangle + c \leq f(v)$ for all $v \in D(f)$ and c is the largest constant for which this inequality holds. In other words, $c = \inf\{f(v) - \langle p, v \rangle : v \in D(f)\} = -\sup\{\langle p, v \rangle - f(v) : v \in D(f)\}$ (which could turn out to be $-\infty$.) The hyperplane clearly intersects the y axis at c .

Definition 9.1 The *Legendre transform* of f is the function f^* given by

$$f^*(p) = \sup\{\langle p, v \rangle - f(v) : v \in D(f)\}, \quad (9.1)$$

with domain $D(f^*)$ consisting of all those $p \in V^*$ for which the sup is less than $+\infty$. (In other words $c > -\infty$.) f^* is called the *conjugate function* to f .

These functions are related by the following easy to state theorem.

Theorem 9.2 *If f is a closed convex function on its domain in V then f^* is a closed convex function on its domain in V^* . Moreover*

$$f^{**} = f \quad (9.2)$$

This is a nice classical theorem with a nice geometric flavor. The proof is easy and we refer the reader to [26][page 146] for a proof. { A proof is in the

Appendices, but commented out.) But it is not exactly the theorem we need because we are really interested only in smooth, strictly convex functions f which are defined on all of V . Of course the theorem applies to this restricted class of functions. But we can say a little more in this case. Draw a picture of a parabola as a model for a nice convex function on \mathbb{R} and you will see that the support plane (line in this case) touches the graph of f at exactly one point $(v, f(v))$ just because f is strictly convex. (If f had flat spots the support line could touch all along the flat spot.) At this unique point v we clearly have

$$f'(v) = p. \quad (9.3)$$

This equation determines v as a function of p , say $v = \phi(p)$, because f' is strictly increasing, since $f'' > 0$. If one only had $f'' \geq 0$ then there could be flat spots and in this case f^* is still well defined on V^* but the diffeomorphism ψ^{-1} constructed in Section 9.1.1 no longer exists. This failure is actually a conceptually important part of thermodynamics because it reflects the existence of phase transformations. But we will need to use the diffeomorphism ψ in our application to mechanics as well as the Legendre transform $f \mapsto f^*$.

Example 9.3 Take $V = \mathbb{R}$ and let

$$f(v) = (1/2)mv^2 + av \quad (9.4)$$

for some constants $m > 0$ and $a \in \mathbb{R}$. The equation (9.3) reduces to

$$mv + a = p. \quad (9.5)$$

So $\phi(p) = m^{-1}(p - a)$. In accordance with (9.1) the conjugate function f^* is given by

$$\begin{aligned} f^*(p) &= [pv - f(v)]_{v=\phi(p)} \\ &= \frac{1}{2m}(p - a)^2 \end{aligned} \quad (9.6)$$

Thus in this example the familiar relation $p = mv$ fails and is replaced by (9.5). Indeed we will see later that (9.5) more accurately captures the relation between momentum and velocity than $p = mv$ does when the particle is acted on by an electromagnetic field. You might like to know that (9.6) can be partly blamed for forcing the appearance of connections on vector bundles in physics.

All of this discussion leading to the construction of the function ϕ applies to higher dimensions also under the strict convexity assumption $f''(v)\langle w, w \rangle > 0$ (for $w \neq 0$.) The equation (9.3) should now be interpreted, of course, to assert that the linear functional on the left equals the linear functional on the right. We are only going to consider functions f that are quadratic, in the sense of Example 9.3. This is done in Section 9.1.1. (The formula based version in Section 9.1.1 doesn't require convexity.) The function ϕ will therefore be affine, just as in the example, and in particular, ϕ will always be a diffeomorphism from V^* onto V . The formula (9.6) is clearly a special case of (9.13). If you are inclined to find a more general class of smooth, strictly convex functions for which ϕ is a global diffeomorphism tell me what you find out.

9.1.1 The Legendre transform for second degree polynomials

This section computes the Legendre transform for a slightly more general class of quadratic functions than those given in the examples of Section 2.7.1. It is not needed by us.

Let Y be a finite dimensional real vector space and denote by Y^* its dual space. We want to consider a class of functions on Y which will be typical for the velocity dependence of a Lagrangian $L(q, v)$ at a fixed point q . Let $M : Y \rightarrow Y^*$ be an invertible linear transformation and let $\alpha \in Y^*$. Consider the function $f : Y \rightarrow \mathbb{R}$ given by

$$f(v) = (1/2)\langle Mv, v \rangle + \langle \alpha, v \rangle \quad v \in Y \quad (9.7)$$

We may identify Y^{**} with Y by the natural isomorphism and then, since M^* maps Y^{**} to Y^* , we may write $M^* : Y \rightarrow Y^*$. In (9.7) we can assume without loss of generality that $M^* = M$ because $\langle Mv, v \rangle = \langle v, M^*v \rangle$ anyway.

The derivative of f in a direction u is a linear functional of u given by

$$\langle p, u \rangle \equiv \partial_u f(v) = \langle Mv, u \rangle + \langle \alpha, u \rangle \quad \forall u \in Y. \quad (9.8)$$

Here, p is an element of Y^* depending on v . It is given by

$$p = \psi(v) = Mv + \alpha. \quad (9.9)$$

Since we have assumed that M is invertible we can solve for v in terms of p , finding

$$v = \psi^{-1}(p) = M^{-1}(p - \alpha). \quad (9.10)$$

The *Legendre transformation* of f is the function $f^* : Y^* \rightarrow \mathbb{R}$ given by

$$f^*(p) = \langle p, v \rangle - f(v), \quad (9.11)$$

wherein it is understood that v should be replaced by the function of p derived in (9.10). Thus

$$\begin{aligned} f^*(p) &= \langle p, M^{-1}(p - \alpha) \rangle \\ &- \{(1/2)\langle MM^{-1}(p - \alpha), M^{-1}(p - \alpha) \rangle + \langle \alpha, M^{-1}(p - \alpha) \rangle\} \\ &= (1/2)\langle M^{-1}(p - \alpha), (p - \alpha) \rangle \end{aligned} \quad (9.12)$$

SUMMARY: If $f(v) = (1/2)\langle Mv, v \rangle + \langle \alpha, v \rangle$, $v \in Y$, then its Legendre transform, $f^* : Y^* \rightarrow \mathbb{R}$ is defined by (9.11) with the insertion of (9.10). It is given by

$$f^*(p) = (1/2)\langle M^{-1}(p - \alpha), (p - \alpha) \rangle \quad \text{for } p \in Y^* \quad (9.13)$$

Moreover the map $\psi : Y \rightarrow Y^*$ is a diffeomorphism.

9.2 Poisson's equation

We will write as usual $r = |x|$ in \mathbb{R}^3 .

Theorem 9.4

$$\Delta \frac{1}{r} = -4\pi\delta. \quad (9.14)$$

in the distribution sense. That is,

$$\Delta L_{1/r} = -4\pi L_\delta$$

We will break the proof up into several small steps.

Lemma 9.5 At $r \neq 0$

$$\Delta(1/r) = 0$$

Proof. $\partial(1/r)/\partial x = -x/r^3$ and $\partial^2(1/r)/\partial x^2 = -1/r^3 + 3\frac{x^2}{r^5}$. So

$$\Delta(1/r) = -3/r^3 + 3\frac{x^2 + y^2 + z^2}{r^5} = -3/r^3 + 3/r^3 = 0.$$

QED.

In view of this lemma you can see that we have only to deal now with the singularity at $r = 0$. Our notion of weak derivative is just right for doing this.

The trick is to avoid the singularity until after one does some clever integration by parts (in the form of the divergence theorem). In case you forgot your vector calculus identities a self contained review is at the end of this section. I want to warn you that this is not the kind of proof that you are likely to invent yourself. But the techniques are so frequently occurring that there is some virtue in following it through at least once in one's life.

Recall the usual notation: $\mathcal{D} = C_c^\infty(\mathbb{R}^3)$.

Lemma 9.6 *Let $\phi \in \mathcal{D}$. Then*

$$\int_{\mathbb{R}^3} (1/r)\Delta\phi(x)dx = \lim_{\epsilon \rightarrow 0} \int_{r \geq \epsilon} (1/r)\Delta\phi dx$$

Proof: The difference between the left and the right sides before taking the limit is at most (use spherical coordinates in the next step)

$$\left| \int_{r \leq \epsilon} (1/r)\Delta\phi d^3x \right| \leq \max_{x \in \mathbb{R}^3} |\Delta\phi(x)| \int_{r \leq \epsilon} (1/r)d^3x = \max_{x \in \mathbb{R}^3} |\Delta\phi(x)| 2\pi\epsilon^2 \rightarrow 0$$

QED.

Before really getting down to business lets apply the definitions.

$$\begin{aligned} \Delta T_{1/r}(\phi) &= \sum_{j=1}^3 (\partial^2/\partial x_j^2) T_{1/r}(\phi) \\ &= - \sum_{j=1}^3 (\partial/\partial x_j) T_{1/r}(\partial\phi/\partial x_j) \\ &= T_{1/r}(\Delta\phi) \\ &= \int_{\mathbb{R}^3} (1/r)\Delta\phi(x)d^3x \\ &= \lim_{\epsilon \rightarrow 0} \int_{r \geq \epsilon} (1/r)\Delta\phi(x)d^3x. \end{aligned}$$

So what we really need to do is show that this limit is $-4\pi\phi(0)$. To this end we are going to apply some standard integration by parts identities in the “OK” region $r \geq \epsilon$.

$$\begin{aligned} C_\epsilon &:= \int_{r \geq \epsilon} (1/r)\Delta\phi(x)d^3x \\ &= \int_{r \geq \epsilon} \nabla \cdot \left(\frac{1}{r} \nabla \phi - \phi \nabla \left(\frac{1}{r} \right) \right) d^3x \quad \text{by identity (9.18)} \\ &= \int_{r=\epsilon} \left(\frac{1}{r} \nabla \phi \cdot \mathbf{n} - \phi \left(\nabla \frac{1}{r} \right) \cdot \mathbf{n} \right) dA \quad \text{by the divergence theorem} \end{aligned}$$

where \mathbf{n} is the unit normal pointing toward the origin. The other boundary term in this integration by parts identity is zero because we can take it over a sphere so large that ϕ is zero on and outside it.

Now

$$\begin{aligned} \left| \int_{r=\epsilon} (1/r)(\nabla \phi \cdot \mathbf{n})dA \right| &= \frac{1}{\epsilon} \left| \int_{r=\epsilon} (\nabla \phi \cdot \mathbf{n})dA \right| \\ &\leq \frac{1}{\epsilon} (\max |\nabla \phi|) 4\pi \epsilon^2 \\ &\rightarrow 0 \end{aligned}$$

as $\epsilon \downarrow 0$. This gets rid of one of the terms in C_ϵ in the limit. For the other one just note that $(\nabla \frac{1}{r}) \cdot \mathbf{n} = -\partial(1/r)/\partial r = 1/r^2$. So

$$\begin{aligned} - \int_{r=\epsilon} \phi \left(\nabla \frac{1}{r} \right) \cdot \mathbf{n} dA &= -\frac{1}{\epsilon^2} \int_{r=\epsilon} \phi(x) dA \\ &= -\frac{1}{\epsilon^2} \int_{r=\epsilon} \phi(0) dA - \frac{1}{\epsilon^2} \int_{r=\epsilon} (\phi(x) - \phi(0)) dA \\ &= -4\pi\phi(0) - \frac{1}{\epsilon^2} \int_{r=\epsilon} (\phi(x) - \phi(0)) dA \end{aligned}$$

Only one more term to get rid of!

$$\frac{1}{\epsilon^2} \left| \int_{r=\epsilon} (\phi(x) - \phi(0)) dA \right| \leq \max_{|x|=\epsilon} |\phi(x) - \phi(0)| \cdot 4\pi \rightarrow 0$$

because ϕ is continuous at $x = 0$. This proves (9.14).

Vector calculus identities.

If f is a real valued function and G is a vector field, both defined on some region in \mathbb{R}^3 then

$$\nabla \cdot (fG) = (\nabla f) \cdot G + f \nabla \cdot G \quad (9.15)$$

Application #1. Take $f = 1/r$ and $G = \nabla\phi$. Then we get

$$\nabla \cdot \left(\frac{1}{r} \nabla\phi \right) = \left(\nabla \frac{1}{r} \right) \cdot \nabla\phi + \frac{1}{r} \Delta\phi \text{ wherever } r \neq 0. \quad (9.16)$$

Application #2. Take $f = \phi$ and $G = \nabla\frac{1}{r}$. Then we get

$$\nabla \cdot \left(\phi \nabla \frac{1}{r} \right) = (\nabla\phi) \cdot \left(\nabla \frac{1}{r} \right) + \phi \Delta \frac{1}{r} \text{ wherever } r \neq 0 \quad (9.17)$$

But $\Delta \frac{1}{r} = 0$ wherever $r \neq 0$. So subtracting (9.17) from (9.16) we find

$$\frac{1}{r} \Delta\phi = \nabla \cdot \left(\frac{1}{r} \nabla\phi - \phi \nabla \frac{1}{r} \right) \text{ wherever } r \neq 0. \quad (9.18)$$

This is the identity we need in the proof of (9.14).

9.3 Some matrix groups and their Lie algebras

This appendix is an extraction of a very small part of the very readable book [30] by Brian Hall on Lie groups. The following is self-contained in the sense that the exercises, which are a guide to the facts, are easily doable by today's students.

Recall (or prove yourself):

If A is an $n \times n$ matrix with real or complex entries then the initial value problem

$$(\text{O.D.E.}) \quad \frac{du(t)}{dt} = Au(t), \quad t \in R, \quad u(t) \in \mathbb{R}^n \quad (9.19)$$

$$(\text{Initial value}) \quad u(0) = v, \quad v \in R^n \quad (9.20)$$

has a unique solution. It is given by

$$u(t) = e^{tA}v \quad (9.21)$$

where

$$e^{tA} \equiv \sum_{n=0}^{\infty} \frac{(tA)^n}{n!}. \quad (9.22)$$

Check (at your leisure) that the series converges in operator norm and that the sum satisfies

$$\frac{de^{tA}}{dt} = Ae^{tA}, \quad (9.23)$$

which ensures that (9.21) gives a solution to (9.19).

Exercise 9.7

a) If A is an $n \times n$ matrix show that

$$\frac{d}{dt} \det(e^{tA}) \Big|_{t=0} = \text{trace } A \quad (9.24)$$

Hint: By (9.22), $e^{tA} = 1+tA$ up to order two in t . Expand $\det(1+tA)$ by its first column and use induction.

b) Prove, using a), that

$$\frac{d}{dt} \det(e^{tA}) = (\det e^{tA})(\text{trace } A) \quad \text{for all } t. \quad (9.25)$$

c) Prove, using b), that

$$\det e^{tA} = 1 \quad \text{for all real } t, \text{ if and only if } \text{trace } A = 0. \quad (9.26)$$

Exercise 9.8

a. Prove: e^{tA} is *orthogonal* for all t if and only if $A^* = -A$.

b. Prove: e^{tA} is *unitary* for all t if and only if $A^* = -A$.

Hint: Differentiate $(e^{tA})^* e^{tA}$.

Notation 9.9

$$O(n) = \{n \times n \text{ real matrices } T : T^*T = 1\} \quad (9.27)$$

$$SO(n) = \{T \in O(n) : \det(T) = 1\} \quad (9.28)$$

$$U(n) = \{n \times n \text{ complex matrices } T : T^*T = 1\} \quad (9.29)$$

$$SU(n) = \{T \in U(n) : \det T = 1\} \quad (9.30)$$

Notation 9.10

$$o(n) = \{n \times n \text{ real matrices } A : A^* = -A\} \quad (9.31)$$

$$so(n) = \{A \in o(n) : \text{trace } A = 0\} \quad (9.32)$$

$$u(n) = \{n \times n \text{ complex matrices } A : A^* = -A\} \quad (9.33)$$

$$su(n) = \{A \in u(n) : \text{trace } A = 0\} \quad (9.34)$$

Exercise 9.11

Prove

- a) $A \in o(n) \iff e^{tA} \in O(n) \forall t \in \mathbb{R}$
- b) $A \in so(n) \iff e^{tA} \in SO(n) \forall t \in \mathbb{R}$
- c) $A \in u(n) \iff e^{tA} \in U(n) \forall t \in \mathbb{R}$
- d) $A \in su(n) \iff e^{tA} \in SU(n) \forall t \in \mathbb{R}$

Notation 9.12 For any two $n \times n$ matrices define

$$[A, B] = AB - BA \quad (9.35)$$

Exercise 9.13 Show that each of the four sets defined in Notation 9.10 is closed under the commutator product, $A, B \mapsto [A, B]$. Thus each of the four linear sets of matrices is an algebra with respect to the product (9.35). Unfortunately (in the view of some) this product is not associative.

Exercise 9.14 Although the commutator product (9.35) is not associative we do have instead the following identity for any three $n \times n$ matrices.

$$[A, [B, C]] + [B, [C, A]] + [C, [A, B]] = 0. \quad \text{Jacobi's identity.} \quad (9.36)$$

Prove this. Notice the easy to remember cyclicity in this identity.

Definition 9.15 A real *Lie algebra* is a real vector space \mathcal{L} with a bilinear pairing $\mathcal{L} \times \mathcal{L} \ni A, B \mapsto [A, B]$ which is skew symmetric (i.e. $[A, B] = -[B, A]$) and which satisfies Jacobi's identity (9.36).

You showed in Exercises 9.13 and 9.14 that each of the four linear spaces in Notation 9.10 is a Lie algebra with respect to the commutator product (9.35)

Terminology. The linear space $o(n)$, with the product $A, B \mapsto [A, B]$ is the *Lie algebra* of $O(n)$. Similarly $so(n)$ is the Lie algebra of $SO(n)$, while $u(n)$ is the Lie algebra of $U(n)$ and $su(n)$ is the Lie algebra of (make a bold guess).

The adjoint representation.

If G is any one of the four groups defined in Notation 9.9 and \mathfrak{g} is its Lie algebra define

$$R(g)A = gAg^{-1} \quad \text{for } g \in G \text{ and } A \in \mathfrak{g} \quad (9.37)$$

Exercise 9.16 a) Show that for each of the four groups in Notation 9.9 the map $R(g)$ carries \mathfrak{g} back into itself for each $g \in G$.

b) Show that the map $g \mapsto R(g)$ is a homomorphism of G into the group of invertible operators on \mathfrak{g} .

Terminology. R is called the *adjoint representation*.

Status. You have now shown that for certain matrix groups G , namely those defined in Notation 9.9, the set \mathcal{L} , of matrices, A , for which $e^{tA} \in G$ \forall real t , forms a real vector space which is closed under the commutator product $[A, B]$.

Fact. This is true for all Lie groups of matrices.

9.3.1 Connection between $SU(2)$ and $SO(3)$.

Exercise 9.17 If x, y, z are real numbers then the matrix $A \equiv \begin{pmatrix} z & x + iy \\ x - iy & -z \end{pmatrix}$ is a Hermitian 2×2 matrix with trace zero. Moreover any Hermitian 2×2 matrix with trace zero is clearly of this form for some unique x, y, z .

- a. Show that $\det(A) = -(x^2 + y^2 + z^2)$.
- b. Show that if U is a unitary 2×2 matrix and $A' = UAU^*$ then A' is Hermitian and has trace zero.
- c. Write $A' = \begin{pmatrix} z' & x' + iy' \\ x' - iy' & -z' \end{pmatrix}$ where A' is given in part b. Clearly (x', y', z') depends linearly on (x, y, z) . Denote this linear transformation by $\rho(U)$. Show that $\rho(\cdot)$ is a homomorphism from $SU(2)$ into $O(3)$.
- d. Show that if $AB = BA$ for all $n \times n$ traceless Hermitian matrices A (where B is an $n \times n$ matrix) then B is a scalar multiple of the identity matrix.
- e. Prove that the kernel of ρ (defined in part c.) consists of $\{I, -I\}$.

9.3.2 The Pauli spin matrices

Define

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (9.38)$$

These are the Pauli spin matrices. Their significance lies in the fact that the following multiples of them form a very convenient basis of the Lie algebra $su(2)$. Define

$$s_x = (i/2)\sigma_x, \quad s_y = (i/2)\sigma_y, \quad s_z = (i/2)\sigma_z \quad (9.39)$$

Observe first that the three Pauli matrices are Hermitian and have trace zero. Therefore the three s matrices are skew-adjoint and have trace zero. In view of Notation 9.10 the three s matrices lie in $su(2)$. When you have a few seconds to spare you should check that $su(2)$ is actually three dimensional (over \mathbb{R}) and that the three s matrices actually span $su(2)$. The reason for the factors of $1/2$ is that the following nifty commutation relations hold.

$$[s_x, s_y] = s_z, \quad [s_y, s_z] = s_x, \quad [s_z, s_x] = s_y. \quad (9.40)$$

[Check these signs]

Theorem 9.18 Identify \mathbb{R}^3 with the set of 2×2 Hermitian matrices of trace zero by the map $(x, y, z) \mapsto A \equiv \begin{pmatrix} z & x + iy \\ x - iy & -z \end{pmatrix}$. For $U \in SU(2)$ define $\rho(U) \in SO(3)$ as in Exercise 9.17. Then the homomorphism

$$SU(2) \mapsto SO(3) \text{ given by } g \mapsto \rho(g) \quad (9.41)$$

is homomorphism of $SU(2)$ onto $SO(3)$ with kernel $\{I, -I\}$.

Proof. Most of this has been already proven in Exercise 9.17. It remains only to prove the surjectivity. To this end it suffices to prove that any rotation around some axis is in the image of ρ . We will do this for rotation around the z axis. This computation will be very useful later in understanding spin.

We already know that for any real number θ the matrix $\exp(\theta s_z)$ lies in $SU(2)$. (by Exercise 9.11.) Moreover, since s_z is diagonal we can compute the exponential easily. We find

$$\exp(\theta s_z) = \begin{pmatrix} e^{(i\theta/2)} & 0 \\ 0 & e^{(-i\theta/2)} \end{pmatrix} \quad (9.42)$$

and therefore

$$\exp(\theta s_z) \begin{pmatrix} z & x+iy \\ x-iy & -z \end{pmatrix} \exp(-\theta s_z) = \begin{pmatrix} z & e^{(i\theta)}(x+iy) \\ e^{(-i\theta)}(x-iy) & -z \end{pmatrix} \quad (9.43)$$

What is the meaning of this matrix? Just write $x+iy = re^{(i\phi)}$ and then you see that the upper right entry is $re^{(i(\theta+\phi))}$. This represents a rotation in the x, y plane by an amount θ . Thus every rotation around the z axis lies in the range of the homomorphism ρ . I'll leave it to your trust in symmetry to conclude that every rotation around any axis is in the range of ρ . Therefore ρ is surjective. ■

It will be real useful to write the result of the computation in the preceding proof in the form

$$\rho(e^{\theta s_z}) = e^{\theta A_\omega}, \quad \omega = (0, 0, 1) \quad (9.44)$$

where A_ω is defined in (2.5). As we know, $e^{\theta A_\omega}$ is the matrix (2.6), which is rotation in the x, y plane through an angle θ .

Peculiarity We see that as θ moves from 0 to 2π , the operator $\rho(e^{\theta s_z})$ runs through rotations starting from the identity rotation and back to the identity rotation. But what is $e^{\theta s_z}$ doing? At $\theta = 2\pi$ we have, by (9.42)

$$\exp(2\pi s_z) = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} = -I \quad (9.45)$$

In fact the curve $\theta \mapsto \exp(\theta s_z)$ clearly returns to the identity operator in $SU(2)$ only when θ reaches 4π . But then its image has returned to the identity operator in $SO(3)$ twice. This is consistent, to say the least, with the assertion in Theorem 9.18 that the kernel of the homomorphism ρ is $\{\pm I\}$.

9.4 grad, curl, div and d

These days its good to be able to switch back and forth between the vector differential operators, gradient, curl and divergence on the one hand and the exterior derivative d over \mathbb{R}^3 on the other. For a nice exposition of these relations see Hubbard and Hubbard [36], Section 6.8.

Here is a short summary. If $v = (v_1, v_2, v_3) \in \mathbb{R}^3$ we may associate with it a 1-form and a 2-form:

$$v^{(1)} = \sum_{j=1}^3 v_j dx_j, \quad \text{which is a 1-form, and} \quad (9.46)$$

$$v^{(2)} = \sum_{(i,j,k)} v_i dx_j \wedge dx_k, \quad \text{which is a 2-form.} \quad (9.47)$$

The sum is over all three triples (i, j, k) which are cyclic permutations of $(1, 2, 3)$. The maps $v \mapsto v^{(1)}$ and $v \mapsto v^{(2)}$ are isomorphisms of \mathbb{R}^3 with $\Lambda^1(\mathbb{R}^3)$ and $\Lambda^2(\mathbb{R}^3)$ respectively.

The Hodge star operator $* : \Lambda(\mathbb{R}^3) \rightarrow \Lambda(\mathbb{R}^3)$ is given by

$$*a = a dx_1 \wedge dx_2 \wedge dx_3, \quad a \in \mathbb{R}, \quad *v^{(1)} = v^{(2)}, \quad ** = \text{Identity.} \quad (9.48)$$

In the physics literature the coordinate vector v is referred to as a “polar” vector or “axial” vector according to how its coordinates transform under the inversion $(x_1, x_2, x_3) \mapsto (-x_1, -x_2, -x_3)$. Under this inversion it’s clear that $v^{(1)} \mapsto -v^{(1)}$ while $v^{(2)} \mapsto v^{(2)}$. To say that v is a polar vector means that under rotations the triple (v_1, v_2, v_3) transforms like the 1-form $v^{(1)}$. To say that v is an axial vector means that the triple (v_1, v_2, v_3) transforms like the 2-form $v^{(2)}$. Some would say that axial vectors should have been identified as 2-forms in the first place. Excuse: “polar” and “axial” were invented long before 2-forms.

Now suppose that each component v_j is a C^1 function on \mathbb{R}^3 . The usual definitions of curl and div are

$$(\text{curl } v(x))_i = \frac{\partial v_k}{\partial x_j} - \frac{\partial v_j}{\partial x_k}, \quad (i, j, k) \text{ cyclic} \quad (9.49)$$

$$\text{div } v(x) = \sum_{i=1}^3 \frac{\partial v_i(x)}{\partial x_i} \quad (9.50)$$

All three operations, grad, curl, div, are special cases of the exterior derivative d if one simply identifies a vector field v with a 1-form or 2-form correctly. Thus:

$$(\nabla \phi)^{(1)} = d\phi \quad (9.51)$$

$$(\text{curl } v)^{(2)} = dv^{(1)} \quad (9.52)$$

$$*(\text{div } v) = dv^{(2)} \quad (9.53)$$

Of course with the help of the Hodge $*$ operator we can write also

$$(\operatorname{curl} v)^{(1)} = *d v^{(1)}, \quad \text{and} \quad (9.54)$$

$$\operatorname{div} v = *d v^{(2)} \quad (9.55)$$

{ Note: If I had defined $a^{(3)} = *a$ then (7.21) would have read $(\operatorname{div} v)^{(3)} = d v^{(2)}$. Cute. }

9.5 Hermite polynomials

Definition 9.19 (Hermite polynomials). The entire function $z \mapsto e^{zx - z^2}$ can be expanded in a power series in z with coefficients that depend on x thus:

$$e^{(zx - \frac{z^2}{2})} = \sum_{n=0}^{\infty} H_n(x) z^n / n! \quad (9.56)$$

Since the coefficient of $z^n/n!$ in the factor e^{zx} is x^n , the coefficient of $z^n/n!$ in (9.56) is a polynomial in x of degree at most n . Moreover the highest order term in $H_n(x)$ is exactly x^n and $H_0(x) = 1$. Since the n th derivative of (9.56) with respect to z at $z = 0$ is exactly $H_n(x)$, Cauchy's integral formula gives

$$H_n(x) = \frac{n!}{2\pi i} \int_{|z|=1} z^{-(n+1)} e^{xz - \frac{z^2}{2}} dz \quad (9.57)$$

The polynomials $H_n, n = 0, 1, 2, \dots$ are the *Hermite polynomials*.

Example 9.20 Comparing the coefficients of z^n in (9.56) for $n = 0, 1, 2$ one sees immediately that

$$H_0(x) = 1, \quad H_1(x) = x, \quad H_2(x) = x^2 - 1.$$

Lemma 9.21 Let $\gamma(dx) = (1/\sqrt{2\pi})e^{-x^2/2}dx$. Then

$$\int_{-\infty}^{\infty} H_n(x) H_k(x) \gamma(dx) = n! \delta_{kn} \quad (9.58)$$

Proof. From the identity

$$(zx - \frac{z^2}{2}) + (\zeta x - \frac{\zeta^2}{2}) - x^2/2 = -\frac{1}{2}(x - \frac{z+\zeta}{2})^2 + z\zeta$$

it follows that

$$\int_{\mathbb{R}} e^{xz - \frac{z^2}{2}} e^{x\zeta - \frac{\zeta^2}{2}} \gamma(dx) = e^{z\zeta}.$$

Multiply by $\frac{n!k!}{(2\pi i)^2} z^{-(n+1)} \zeta^{-(k+1)}$ and integrate over the product of unit circles to find

$$\begin{aligned} \int_{\mathbb{R}} H_n(x) H_k(x) \gamma(dx) &= \frac{n!k!}{(2\pi i)^2} \int_{|z|=1} \int_{|\zeta|=1} z^{-(n+1)} \zeta^{-(k+1)} e^{z\zeta} d\zeta dz \\ &= \frac{n!}{2\pi i} \int_{|z|=1} z^{-(n+1)} z^k dz \\ &= n! \delta_{kn} \end{aligned}$$

■

Lemma 9.22 (*Identities*)

$$H'_n(x) = nH_{n-1}(x) \quad (9.59)$$

$$H_n(x) = xH_{n-1}(x) - (n-1)H_{n-2}(x) \quad (9.60)$$

$$-H''_n(x) + xH'_n(x) = nH_n(x) \quad (9.61)$$

Proof. Differentiate (9.57) to arrive at (9.59). Differentiate (9.56) with respect to z to find

$$(x-z)e^{xz - \frac{z^2}{2}} = \sum_{n=1}^{\infty} H_n(x) \frac{z^{n-1}}{(n-1)!} \quad (9.62)$$

and therefore

$$(x-z) \sum_{n=1}^{\infty} H_n(x) z^n / n! = \sum_{n=1}^{\infty} H_n(x) \frac{z^{n-1}}{(n-1)!}. \quad (9.63)$$

Comparing the coefficients of z^{n-1} on both sides of this identity we find

$$\frac{xH_{n-1}(x)}{(n-1)!} - \frac{H_{n-2}(x)}{(n-2)!} = \frac{H_n(x)}{(n-1)!}.$$

Multiply by $(n-1)!$ to arrive at (9.60). Differentiate (9.59) once more and use (9.60) and (9.59) again to deduce

$$H''_n = n\{(n-1)H_{n-2}\} = n\{xH_{n-1} - H_n\} = xH'_n - nH_n,$$

which is (9.61). ■

Corollary 9.23 (*Conversion to frequency ω*) Let $\omega > 0$. Define

$$\gamma_\omega(dq) = \left(\frac{\omega}{\pi}\right)^{1/2} e^{-\omega q^2} dq \quad \text{and} \quad \psi_n(q) = \frac{1}{\sqrt{n!}} H_n(\sqrt{2\omega}q) \quad (9.64)$$

Then

$$a) \quad \{\psi_n\}_0^\infty \text{ is an orthonormal basis of } L^2(\mathbb{R}, \gamma_\omega) \quad (9.65)$$

$$b) \quad \left(-\frac{1}{2} \frac{d^2}{dq^2} + \omega q \frac{d}{dq}\right) \psi_n = (n\omega) \psi_n, \quad n = 0, 1, \dots \quad (9.66)$$

$$c) \quad \text{Spectrum} \left(-\frac{1}{2} \frac{d^2}{dq^2} + \omega q \frac{d}{dq} \right) = \{n\omega : n = 0, 1, 2, \dots\} \quad (9.67)$$

$$d) \quad \left(\left(-\frac{1}{2} \frac{d^2}{dq^2} + \omega q \frac{d}{dq} \right) f, g \right)_{L^2(\gamma_\omega)} = \frac{1}{2} \int_{\mathbb{R}} f'(q) \bar{g}'(q) \gamma_\omega(dq) \quad (9.68)$$

$$e) \quad \psi'_n(q) = \sqrt{2\omega} \sqrt{n} \psi_{n-1}(q) \quad (9.69)$$

$$f) \quad \sqrt{2\omega} q \psi_{n-1}(q) = \sqrt{n} \psi_n(q) + \sqrt{(n-1)} \psi_{n-2}(q) \quad (9.70)$$

$$g) \quad \int_{\mathbb{R}} q^2 \gamma_\omega(dq) = 1/(2\omega) \quad (9.71)$$

Proof. Put $x = \sqrt{2\omega} q$ in (9.58) to deduce orthonormality of the ψ_n . Do the same to derive b) from (9.61), to derive e) from (9.59), and to derive f) from 9.60. Take completeness of the ψ_n as an exercise. Item c) follows from a) and b). Item d) is a straightforward integration by parts. Item g) is a simple Gaussian integral, but can also be deduced from Item a) by taking $n = 1$, since $\psi_1(q) = \sqrt{2\omega} q$. ■

9.6 Special relativity

Ran out of time. If you're concerned about the clock (\equiv twin) paradox see Alfred Schild, [52] 1959. There is no paradox.

9.7 Other interesting topics, not yet experimentally confirmed

9.7.1 The Higgs particle for the standard model

The conformal group of Minkowski space is a fifteen dimensional Lie group which contains the ten dimensional Poincaré group as a subgroup. Three of

the dimensions of the Poincaré group correspond to changing to a coordinate system which is moving with constant velocity with respect to yours. Three of the dimensions of the conformal group correspond to changing to a coordinate system which is moving with constant acceleration with respect to yours. If a physical theory is invariant under the conformal group then there should be no difference between a system which is stationary and a system which is moving with constant acceleration. So how can Newton's equation $F = ma$ have a meaning if $a = 0$ is physically equivalent to $a \neq 0$? It can't, unless $m = 0$. Now the Yang-Mills equation is invariant under the conformal group , as are the various equations for matter fields associated with it. Consequently a theory based on these equations must include only particles of mass zero. This is problematical because there are particles of strictly positive mass. In a quantum theory, invariance under the conformal group, G lets call it, means that there is a unitary representation of G on the Hilbert state space of the system which commutes with some of the observables, including especially the Hamiltonian. But these unitary operators do not leave every vector invariant. Therefore even though the overall theory is conformally invariant, there may be subspaces which are not invariant under the unitary representation of G but are nevertheless invariant under the Hamiltonian. In that case one could conceivably restrict ones attention to such a subspace and hope that within this subspace the conformal invariance of the overall theory is sufficiently undermined so as not to force all masses to be zero. The standard model includes a mechanism for implementing this idea. One hypothesizes that there exists a particle in the theory, the so called Higgs particle, which has a non-unique ground state. Such a ground state need not be invariant under the conformal group, which will just take the ground state into a different ground state. The subspace of the Hilbert state space "generated" by this ground state will be a non-conformally invariant subspace of the kind we would like. One says that the conformal symmetry is broken in this subspace. The theory shows that the Higgs particle is quite a heavy particle and therefore not easy to produce. It is widely hoped that the Large Hadron Collider (LHC), which recently went into operation in Switzerland, will find it, thereby lending tremendous support to the standard model. As of the recent summer solstice at the latitude of Ithaca, the Higgs particle has not yet been found.

9.7.2 Supersymmetry

9.7.3 Conformal field theory

9.7.4 String theory

9.8 How to make your own theory

Quantum mechanics entails a vast conceptual revision of our understanding of matter, an understanding that we learned from classical mechanics. Many open minded citizens find quantum mechanics hard to accept. If you are among them, there is no reason why you shouldn't make your own theory. To help you with this, I've prepared a short list of observations and experiments whose outcomes must fit into your theory.

1. Hydrogen spectrum.
2. Blackbody radiation formula.
3. Dark lines from stars
4. Zeeman effects. (atom in a magnetic field)
 - a) normal Zeeman effect (one line splits into three lines)
 - b) anomalous Zeeman effect (one line splits into five or more lines)
5. Stark effect. (atom in an electric field)
6. Stern-Gerlach experiment. (magnetic moment of a neutral atom comes in discrete multiples of a unit.)
7. Specific heat of metals (Pauli)

10 Yet More History

10.1 Timeline for electricity vs magnetism.

1731: Lightning struck a box of knives and forks and magnetized them.

1805: Hachette and Desormes attempted to determine whether an insulated voltaic pile , freely suspended, is oriented by terrestrial magnetism. It is not.

1820: Oersted, Biot, Savart, Ampère. The sequence of missteps taken by Oersted and his predecessors, prior to Oersted's discovery of what actually works, is very illuminating. Don't let anyone tell you that if you, the intelligent reader, will just sit and think about this, you will arrive at the theory that works. In fact Oersted tried at first to produce a magnetic field

from a battery without closing the circuit !! It was years before he thought of closing the circuit and producing a magnetic field from a current rather than just from a battery by itself. And even then he oriented the compass incorrectly for seeing an effect. But he had company in this kind of misstep. Hachette and Desormes tried to make a compass needle by just suspending a battery from its middle by a thread, (1805), hoping that it would orient to the earth's axis like a compass needle. It didn't. So much for the production of magnetic fields from electricity - till 1820. You can read more detail about these travails in Whitaker [W1], pages 80-85.

1824: FARADAY: First attempt to produce a current from a magnetic field. Failure. [Do the experiment.]

For Faraday's reasoning for doing this experiment you can read [W1] p170, line 3* to p171 line 6*. He already has the idea of "lines of magnetic force".

1831 Faraday's second attempt. Success. [Do the experiment.]

1861: MAXWELL (1831-1879) puts it all together in 1861

1887: Hertz(1857-1894) settles it. See [W1] p318-325, but skip the mathematics.

Views and gossip.

1. Maxwell intended to find a mechanical model explaining all the previous discoveries. See [W1] P246. and footnote Number 2.
2. Maxwell's mechanical view. See [W1] P250, 251.
3. Maxwell's view of the velocity of light. See [W1] p253 , last paragraph.
4. Maxwell's disdainful view of Ampère's "too mathematical" method. See [44, Vol. 2], page 175 bottom to page 176 top.
5. Maxwell's theory was accepted with reluctance. See [W1] p254, last paragraph.
6. Lord Kelvin never accepted Maxwell's theory. See [W1] p266 "This was not .." to p267, line4*.
7. The Michelson-Morley experiment, (1887). The aether theory was abandoned after this experiment and along with it any mechanical significance of the electric and magnetic fields. (such as "stress in the aether") You have a choice of movies illustrating the Michelson-Morley experiment. Just enter Michelson-Morley into Google and take your pick. Wikipedia has, as usual, a good exposition of this experiment.

{More cute facts about the aftermath of Maxwell's discovery are commented out next.}

10.2 Timeline for radiation measurements

If you really want to know who did what and when, here is a chart. But you can skip it if this kind of history doesn't excite you. The thing to take away from this list is that a large number of physicists evolved a large collection of spectroscopic facts over a long period of time, and that's not even counting Galileo and Newton.

1752: Thomas Melville noticed that the yellow light produced by throwing some table salt into a flame was of a "definite degree of refranglbility". That is, shining the light through a prism turned it a definite amount.

1802: Wollaston noticed that the spectrum of sunlight was crossed by seven dark lines.

1814: Fraunhofer: Also noticed dark lines in the spectrum of sunlight. In 1821 he measured their wavelength with a diffraction grating. He found that the sharpest dark line of this (yellow) sunlight had a wavelength of 5887.7 Angstroms. [One Angstrom = 10^{-8} cm.]. He also found that many flames produced a bright line of this same wavelength. See Figure 5 for a sample of dark lines.

1826: W.H. Fox Talbot placed a slit between a flame and the prism and found that he could distinguish between different elements thrown into the flame by the different line structures of different elements. In this way he could distinguish between strontium and lithium, both of which give red flames.

1827: William Miller showed that light transmitted through a gas produced dark lines exactly where the gas would have produced bright lines if it had been heated. Conclusion: A gas that emits light at wavelength λ also absorbs light of wavelength λ passing through it. This explains the dark lines found in the sun's spectrum: gasses in the sun's outer layers *absorb* light produced deeper in by the hot sun. In this way one could now begin to analyze the chemical composition of stars.

1853: Angstrom showed that a gas radiates and absorbs light of the same wavelength.

1858: Balfour Stewart introduced the notion of equilibrium of radiation of each wavelength. (**Blackbody radiation**).

1885: Balmer found an empirical formula for the wavelengths of light emitted by heated hydrogen. Namely

$$\lambda^{-1} = \text{constant } \left(\frac{1}{4} - \frac{4}{m^2} \right), \quad m = 3, 4, 5, \dots \quad (10.1)$$

1890: Rydberg gave a general formula of a similar form for many other chemical elements.

1893: Wien derived a formula for the energy density of blackbody radiation from “fundamental” physical principles. But it agreed with experiment only for low frequencies. In particular it didn’t agree with the curves in Figure 7.

1900: Lord Raleigh derived another formula for the energy density of black body radiation from other “fundamental” physical principles. But it, too, disagreed with experiment and in particular, with the curves in Figure 7. Moreover it was not even in $L^1(0, \infty)$!!! (Falls off too slowly for large frequencies).

1896: Zeeman discovered that the D line of sodium breaks into three lines when a magnetic field is imposed on the source of the sodium light. Lorentz gave a theoretical derivation of this effect. But Lorentz’ theory couldn’t explain the later discovery that the bright lines of other chemical elements break into 5, 7 or even 11 lines when a magnetic field is imposed on the source. This breakup is called the **Zeeman effect**. (Three lines is called the normal Zeeman effect. Five or more lines is called the anomalous Zeeman effect.)

1913: Johannes Stark discovered the electric analogy of the Zeeman effect: lines split when the element is placed in an electric field. This is now called the **Stark effect**.

Bottom line: To understand radiation requires explaining an awful lot of very different kinds of observations.

10.3 Planck, Einstein and Bohr on radiation. Timeline.

This needs revision and maybe different placement.

October 19, 1900: Planck (1858 -1947), at a meeting of the Berlin Academy of Science, gave a formula for the energy density of black body radiation. Namely,

$$E(T, d\nu) = \frac{8\pi h}{c^3} \frac{\nu^3 d\nu}{e^{h\nu/kT} - 1} \quad \text{Planck's formula} \quad (10.2)$$

where k is Boltzmann’s constant (already well known at that time), and h is some other constant (which we call Planck’s constant nowadays.)

Ref. [W2, p.78 -i]

But the derivation of (10.2) that Planck presented was itself ad hoc: He just assumed that the entropy of the radiation field, as a function of its energy satisfied a particular differential equation (chosen largely for simplicity) from which (10.2) followed. Although (10.2) agreed well with measurements it was still just an empirical formula.

December 14, 1900 Planck gave another derivation of (10.2). This time he based it on the **physical assumption** that the energy of a simple harmonic Hertzian oscillator of frequency ν is proportional to ν .

$$E = h\nu \quad \text{Planck's hypothesis} \quad (10.3)$$

1905: **Einstein** used Planck's hypothesis, (10.2), to explain why light has to be of a high enough frequency before it will knock electrons out of a metal. This is the photoelectric effect.

1911 (May): Rutherford determined the structure of an atom: small nucleus (10^{-13} cm) compared to the overall radius (10^{-8} cm.)

1912: **Millikan** (the oil drop experiment) showed that charge comes in multiples of the charge on an electron.

1913: **Niels Bohr** (1885- 1962) As we have seen, Planck's assumption amounts to the assertion that for a harmonic oscillator of frequency ν the only allowed amplitudes are those for which the energy is $h\nu$. Now Bohr applied this idea to atoms. He made the assumption that not all of the usual classical orbits (remember the elliptical orbits of Newton?) exist in an atom.

Bohr's hypothesis: In an atom only those closed orbits exist for which

$$\int_{\text{orbit}} \alpha = nh, \quad (10.4)$$

where α is the natural one form on $T^*(\mathbb{R}^3)$ defined in (2.33), h is Planck's constant and n is an integer. (10.4) is called the quantization condition for Bohr orbits. (Actually Bohr considered only circular orbits. But Bohr's orbit condition was generalized to elliptical orbits by Sommerfeld (1915) and W. Wilson (1915) independently.)

Result: Bohr's semi-geometric hypothesis (10.4) limits the allowed energies of elliptical (classical) orbits to a computable sequence, $\{E_n\}$. Now apply Planck's hypothesis (10.3) to the energy differences $\{E_n - E_m\}$ to find a double sequence of allowed radiation frequencies,

$$E_n - E_m = h\nu_{n,m} \quad (10.5)$$

These frequencies agree exactly with Balmer's empirical formula.

READ [W2] p. 109 "The culmination .." to p. 110.

1914-1918: World War I.

Other successes of Bohr's theory (after some modifications.)

1924: Pauli (Exclusion principle) No two electrons can be in the same state at the same time. [We will elaborate on this later.]

1925: Pauli (electron spin) shows that an electron must be regarded as a little magnet in order to account for the Zeeman effect properly. This invalidates Sommerfeld's earlier explanation of the Zeeman effect. **So in order to explain the Zeeman effect it is necessary to regard an electron as spinning around some internal axis.** [We will elaborate on this later.]

Bose [W2 p.219]

1926: Fermi [W2 p. 224]

10.4 List of elementary particles

v · d · e		Particles in physics	
Elementary	Fermions	Quarks	u · d · c · s · t · b
	Leptons	e ⁻ · e ⁺ · μ ⁻ · μ ⁺ · τ ⁻ · τ ⁺ · ν _e · ν̄ _e · ν _μ · ν̄ _μ · ν _τ · ν̄ _τ	
	Bosons	Gauge	γ · g · W [±] · Z
Composite		Hadrons	Baryons / Hyperons N (p · n) · Δ · Λ · Σ · Ξ · Ω
Composite			Mesons / Quarkonia π · ρ · η · η' · φ · ω · J/ψ · Υ · θ · K · B · D · T

Figure 13: Elementary particles

References

- [1] Y. Aharonov and D. Bohm, *Significance of electromagnetic potentials in the quantum theory*, Phys. Rev. (2) **115** (1959), 485–491, {This is the famous paper describing the Aharonov-Bohm effect. See [11] for experimental verification.}. MR 0110458 (22 #1336)
- [2] Yakir Aharonov and Daniel Rohrlich, *Quantum paradoxes*, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, 2005, Quantum theory for the perplexed. MR 2327477 (2008i:81002) { This book addresses fundamental conceptual issues in quantum mechanics.}
- [3] Huzihiro Araki, *Mathematical theory of quantum fields*, International Series of Monographs on Physics, vol. 101, Oxford University Press, Oxford, 2009, Translated from the 1993 Japanese original by Ursula Carow-Watamura, Reprint of the 1999 edition [MR1799198]. MR 2542202
- [4] V. I. Arnold, *Mathematical methods of classical mechanics*, Springer-Verlag, New York, 1978, Translated from the Russian by K. Vogtmann and A. Weinstein, Graduate Texts in Mathematics, 60. MR 0690288 (57 #14033b)
- [5] Ph. Blanchard and A. Jadczyk, *Theory of events*, Nonlinear, deformed and irreversible quantum systems (Clausthal, 1994), World Sci. Publ., River Edge, NJ, 1995, pp. 265–272. MR 1457551 (98f:81025)
- [6] Ph. Blanchard and A. Jadczyk (eds.), *Quantum future*, Lecture Notes in Physics, vol. 517, Berlin, Springer-Verlag, 1999, From Volta and Como to the present and beyond. MR 1727766 (2000g:81010)
- [7] Ph. Blanchard, L. Jakóbczyk, and R. Olkiewicz, *Measures of entanglement based on decoherence*, J. Phys. A **34** (2001), no. 41, 8501–8516. MR 1876610 (2002k:81024)
- [8] Douglas Botting, *Humboldt and the cosmos*, Prestel, Munich and New York, 1994.
- [9] J. Chadwick, *The existence of a neutron.*, Proc. Roy. Soc. London. **136** (1932), 692–708, { +Q41 L84, (Refs. are from Schweber’s book “QED and the men who made it.”) }.

- [10] ———, *Possible existence of a neutron.*, Nature **129** (1932), 312–313, { Q1 N282 }.
- [11] R.G. Chambers, *Shift of an electron interference pattern by enclosed magnetic flux*, Phys. Rev. Lett. **5** (1960), 3 – 5, {This is the paper that verified experimentally the Aharonov-Bohm effect. It was cited 150 times.}.
- [12] Claude Chevalley, *The algebraic theory of spinors and Clifford algebras*, Springer-Verlag, Berlin, 1997, Collected works. Vol. 2, Edited and with a foreword by Pierre Cartier and Catherine Chevalley, With a postface by J.-P. Bourguignon. MR 1636473 (99f:01028)
- [13] J. M. Cook, *The mathematics of second quantization*, Trans. Amer. Math. Soc. **74** (1953), 222–245. MR 0053784 (14,825h)
- [14] Richard H. Cushman and Larry M. Bates, *Global aspects of classical integrable systems*, Birkhäuser Verlag, Basel, 1997. MR 1438060 (98a:58083)
- [15] R. W. R. Darling, *Differential forms and connections*, Cambridge University Press, Cambridge, 1994. MR 1312606 (95j:53038)
- [16] P. A. M. Dirac, *The quantum theory of the electron*, Proc. R. Soc. of London **A117** (1928), 610–624, Received January 2, 1928. This is the great paper that introduces the Dirac equation.
- [17] ———, *A theory of electrons and protons*, Proc. R. Soc. of London **A126** (1929), 360–366, Received December 6, 1929. This is the great paper that introduces hole theory. Dirac believed the negative energy particles were protons, in this paper.
- [18] ———, *Quantized singularities in the electromagnetic field*, Proc. of the London Math. Soc. **A133** (1931), 60–72, This is the great paper on magnetic monopoles. [anywhere, magnetic monopoles, brings up 857 references on mathscinet. 4/19/06. This beats the bibliographies of Stevens and of Goldhaber and Trower].
- [19] Paul Federbush, *Quantum field theory in ninety minutes*, Bull. Amer. Math. Soc. (N.S.) **17** (1987), no. 1, 93–103. MR 888881 (88k:81106)

- [20] V.A. Fock, *Konfigurationsraum und zweite quantelung*, Zeitschrift für Physik **75** (1932), 622–647.
- [21] Gerald B. Folland, *Quantum field theory*, Mathematical Surveys and Monographs, vol. 149, American Mathematical Society, Providence, RI, 2008, A tourist guide for mathematicians. MR 2436991 (2010a:81001)
- [22] A.S. Goldhaber and W.P. Trower, *Resource letter mm-1: Magnetic monopoles*, Am. J. Phys. **58** (1990), 429–439.
- [23] Herbert Goldstein, *Classical mechanics*, second ed., Addison-Wesley Publishing Co., Reading, Mass., 1980, Addison-Wesley Series in Physics. MR 575343 (81j:70001)
- [24] David J. Griffiths, *Introduction to elementary particles*, second ed., Weinheim : Wiley-VCH, c2008.
- [25] Leonard Gross, *On the formula of Mathews and Salam*, J. Functional Analysis **25** (1977), no. 2, 162–209. MR 0459403 (56 #17596)
- [26] ———, *Thermodynamics, statistical mechanics and random fields*, Tenth Saint Flour Probability Summer School—1980 (Saint Flour, 1980), Lecture Notes in Math., vol. 929, Springer, Berlin, 1982, pp. 101–204. MR 665596 (84m:82005)
- [27] Victor Guillemin, *The story of quantum mechanics*, first ed., Charles Scribner’s Sons, New York, 1968, {1. This fun book to read was written by the physicist Victor Guillemin, father of our Victor Guillemin.} {2. Contains no mathematics, but quotes Sir James Jeans (1930) (page 232) “... the Great Architect of the Universe ... is a pure mathematician.”}.
- [28] Brian Hall, *An introduction to quantum theory for mathematicians*, In preparation, 2011.
- [29] Brian C. Hall, *Harmonic analysis with respect to heat kernel measure*, Bull. Amer. Math. Soc. (N.S.) **38** (2001), no. 1, 43–78 (electronic). MR 1803077 (2002e:22015)
- [30] ———, *Lie groups, Lie algebras, and representations*, Graduate Texts in Mathematics, vol. 222, Springer-Verlag, New York, 2003, An elementary introduction. MR 1997306 (2004i:22001)

- [31] Werner Heisenberg, *Über den bau der atomkerne*, Zeits. Phys. **77** (1932), 1–11, {This is the paper that introduces isotopic spin.} [Reprinted in Heisenberg, Collected works, (1989)].
- [32] _____, *Über den bau der atomkerne*, Zeits. Phys. **78** (1933), 156–164, [Reprinted in Heisenberg, Collected works, (1989)].
- [33] _____, *Über den bau der atomkerne*, Zeits. Phys. **80** (1933), 587–596, {This is the paper that introduces isotopic spin?} [Reprinted in Heisenberg, Collected works, (1989)] Check these titles.
- [34] _____, *Gesammelte Werke. Abteilung A. Teil II*, Springer-Verlag, Berlin, 1989, Wissenschaftliche Originalarbeiten. [Original scientific papers], Edited and with a preface by W. Blum, H.-P. Dürr and H. Rechenberg. MR 1041794 (91h:01096)
- [35] Gerald Holton and Duane H. D. Roller, *Foundations of modern physical science*, Addison-Wesley Pub. Co., Reading, Mass., 1958.
- [36] John Hamal Hubbard and Barbara Burke Hubbard, *Vector calculus, linear algebra, and differential forms*, Prentice Hall Inc., Upper Saddle River, NJ, 1999, A unified approach. MR 1657732 (99k:00002)
- [37] J. D. Jackson, *Classical electrodynamics*, 1999, Third Edition.
- [38] Josef M. Jauch, *Foundations of quantum mechanics*, Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 1968. MR 0218062 (36 #1151)
- [39] Jorge V. José and Eugene J. Saletan, *Classical dynamics*, Cambridge University Press, Cambridge, 1998, A contemporary approach. MR 1640663 (99g:70001)
- [40] Shoshichi Kobayashi and Katsumi Nomizu, *Foundations of differential geometry. Vol. I*, Wiley Classics Library, John Wiley & Sons Inc., New York, 1996, Reprint of the 1963 original, A Wiley-Interscience Publication. MR 1393940 (97c:53001a)
- [41] George W. Mackey, *The mathematical foundations of quantum mechanics: A lecture-note volume*, W., A. Benjamin, Inc., New York-Amsterdam, 1963. MR 0155567 (27 #5501)

- [42] K. B. Marathe and G. Martucci, *The mathematical foundations of gauge theories*, Studies in Mathematical Physics, vol. 5, North-Holland Publishing Co., Amsterdam, 1992. MR 1173210 (93k:58002)
- [43] Jerrold E. Marsden and Tudor S. Ratiu, *Introduction to mechanics and symmetry*, Texts in Applied Mathematics, vol. 17, Springer-Verlag, New York, 1994, A basic exposition of classical mechanical systems. MR 1304682 (95i:58073)
- [44] James Clerk Maxwell, *A treatise on electricity and magnetism*, vol. 1 and 2, (Dover has its recent version), First edition, 1873, Second edition, 1891.
- [45] Jagdish Mehra and Helmut Rechenberg, *The historical development of quantum theory.*, Springer-Verlag, New York, 1982–2001, About 2000 pages in 6 volumes. MR 676476 (85h:01025a)
- [46] Mikio Nakahara, *Geometry, topology and physics*, second ed., Graduate Student Series in Physics, Institute of Physics, Bristol, 2003. MR 2001829 (2004e:58001)
- [47] Edward Nelson, *Dynamical theories of Brownian motion*, Princeton University Press, Princeton, N.J., 1967. MR 0214150 (35 \#5001)
- [48] Michael Reed and Barry Simon, *Methods of modern mathematical physics. II. Fourier analysis, self-adjointness*, Academic Press [Harcourt Brace Jovanovich Publishers], New York, 1975. MR 0493420 (58 \#12429b)
- [49] Duane Roller and Duane H. D. Roller, *The development of the concept of electric charge; Electricity from the Greeks to Coulomb*,, Cambridge, Harvard University Press., 1954, {QC507 .R74 , Annex and Mann Libs.}.
- [50] Duane H. D. Roller, *The De Magnete of William Gilbert*, Menno Hertzberger, Amsterdam, 1959.
- [51] Leonard I. Schiff, *Quantum mechanics*, McGraw-Hill Book Co. New York, 1949.

- [52] Alfred Schild, *The clock paradox in relativity theory*, Amer. Math. Monthly **66** (1959), 1–18. MR 0136417 (24 #B2455)
- [53] Julian Schwinger, *Selected papers on quantum electrodynamics*, Dover Publications, New York, 1958.
- [54] I. E. Segal, *Postulates for general quantum mechanics*, Ann. of Math. (2) **48** (1947), 930–948. MR 0022652 (9,241b)
- [55] S. Sternberg, *Group theory and physics*, Cambridge University Press, Cambridge, 1994. MR 1287387 (95i:20001)
- [56] R. F. Streater and A. S. Wightman, *PCT, spin and statistics, and all that*, Princeton Landmarks in Physics, Princeton University Press, Princeton, NJ, 2000, Corrected third printing of the 1978 edition. MR 1884336 (2003f:81154)
- [57] F. Strocchi, *An introduction to the mathematical structure of quantum mechanics*, second ed., Advanced Series in Mathematical Physics, vol. 28, World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2008, A short course for mathematicians. MR 2484367 (2010e:81001)
- [58] Leon A. Takhtajan, *Quantum mechanics for mathematicians*, Graduate Studies in Mathematics, vol. 95, American Mathematical Society, Providence, RI, 2008. MR 2433906 (2010c:81003)
- [59] R. Ticciati, *Quantum field theory for mathematicians*, Encyclopedia of Mathematics and its Applications, vol. 72, Cambridge University Press, Cambridge, 1999. MR 1699269 (2000h:81002)
- [60] John von Neumann, *Mathematical foundations of quantum mechanics*, Princeton Landmarks in Mathematics, Princeton University Press, Princeton, NJ, 1996, Translated from the German and with a preface by Robert T. Beyer, Twelfth printing, Princeton Paperbacks. MR 1435976 (98b:81006)
- [61] B. L. van der (Bartel Leendert) Waerden, *Sources of quantum mechanics*, Dover Publications, New York, 1968, c1967, edited with a historical introd. by B. L. van der Waerden.

- [62] Steven Weinberg, *The quantum theory of fields. Vol. I*, Cambridge University Press, Cambridge, 2005, Foundations. MR 2148466
- [63] ———, *The quantum theory of fields. Vol. II*, Cambridge University Press, Cambridge, 2005, Modern applications. MR 2148467
- [64] ———, *The quantum theory of fields. Vol. III*, Cambridge University Press, Cambridge, 2005, Supersymmetry. MR 2148468
- [65] Hermann Weyl, *The theory of groups and quantum mechanics*, Dover Publications, Inc., 1928 1st ed, 1930 2nd ed. translated 1931. MR MR0450450 (56 #8744)
- [66] ———, *What is the title?*, Z. Phys. **56** (1929), 330–??, {This is the paper in which he introduces E and M into general relativity via connections. This ref. is from Dirac 1931 and also from Wu and Yang (1975) [List of publications by Hermann Weyl, Hermann Weyl, 1885-1985 (Eidgenössische Tech. Hochschule, Zrich, 1986), 109-119.] }.
- [67] Edmund Whittaker, *A history of the theories of aether and electricity. Vol. I: The classical theories*, Harper & Brothers, New York, 1960. MR 0154779 (27 #4724)
- [68] ———, *A history of the theories of aether and electricity. Vol. II: The modern theories, 1900–1926*, Harper & Brothers, New York, 1960. MR 0154780 (27 #4725)
- [69] Eugene Wigner, *On the consequences of the symmetry of the nuclear hamiltonian on the spectroscopy of nuclei*, Phys. Rev. **51** (1937), 106–119, {This is the paper that introduces isotopic spin.}.
- [70] N. M. J. Woodhouse, *Geometric quantization*, second ed., Oxford Mathematical Monographs, The Clarendon Press Oxford University Press, New York, 1992, Oxford Science Publications. MR 1183739 (94a:58082)
- [71] T.T. Wu and C.N.Yang, *Concept of nonintegrable phase factors and global formulation of gauge fields*, Phys. Rev. D **12** (1975), 3845–3857, {This is the dictionary paper: Nonintegrable phase factor = path dependent parallel transport. Gauge field = connection. Field strength = curvature, etc.}.

- [72] C.N. Yang and R.L. Mills, *Conservation of isotopic spin and isotopic gauge invariance*, Phys. Rev. **96** (1954), 191–195, {This is THE paper.}.
- [73] Eberhard Zeidler, *Quantum field theory. I. Basics in mathematics and physics*, Springer-Verlag, Berlin, 2006, A bridge between mathematicians and physicists. MR 2257528 (2008b:81002)
- [74] _____, *Quantum field theory. II. Quantum electrodynamics*, Springer-Verlag, Berlin, 2009, A bridge between mathematicians and physicists. MR 2456465 (2010a:81002)