A Second Course on Classical Mechanics and Chaos
Based on F2011PHY411

S. G. Rajeev

March 13, 2012

Contents

1 The Variational Principle **2**
- 1.1 Euler-Lagrange equations 4
- 1.2 The Varational Principle of Mechanics 5

2 Conservation Laws **7**
- 2.1 Generalized Momenta 7
- 2.2 Conservation Laws 7
- 2.3 Conservation of Energy 8
- 2.4 Minimal Surface of Revolution 9

3 The Simple Pendulum **11**
- 3.1 Algebraic Formulation 12
- 3.2 Primer on Jacobi Functions 13
- 3.3 Elliptic Curves 14
- 3.4 Addition Formula 15
- 3.5 Imaginary Time 16
  - 3.5.1 The case of $H = 1$ 17
- 3.6 The Arithmetic-Geometric Mean 17
  - 3.6.1 The Arithmetic-Harmonic Mean is the Geometric Mean . 18
- 3.7 Doubly Periodic functions 20

4 The Kepler Problem **21**
- 4.1 The orbit of a planet lies on a plane which contains the Sun . . 21
- 4.2 The line connecting the planet to the Sun sweeps equal areas in equal times 22
- 4.3 Planets move along elliptical orbits with the Sun at a focus . . . 22
- 4.4 The ratio of the cube of the semi-major axis to the square of the period is the same for all planets 23
- 4.5 The shape of the orbit 24

5 The Rigid Body **26**
- 5.1 The Moment of Inertia 27
- 5.2 Angular Momentum 27
- 5.3 Euler's Equations 28

5.4 Jacobi's Solution 29

6 Hamilton's Principle **31**
6.1 Generalized Momenta 31
6.2 Vector Fields and Integral Curves 33
6.2.1 The Algebraic Formulation of a Vector Field 34
6.3 Phase Space 36
6.3.1 The non-commutative algebra of quantum observables . . 37
6.4 Canonical Transformations 38
6.5 Infinitesimal canonical transformations 39
6.6 Symmetries and Conservation Laws 42

7 Geodesics **44**
7.1 The Metric 44
7.2 The Variational Principle 45
7.2.1 Curves minimizing the action and the length are the same 45
7.2.2 The Geodesic Equation 46
7.3 The Sphere 47
7.3.1 The sphere can also be identified with the complex plane, with the point at infinity added 49
7.4 Hyperbolic Space 50
7.5 Hamiltonian Formulation of Geodesics 51
7.6 Geodesic Formulation of Newtonian Mechanics 51
7.6.1 Keplerian Orbits As Geodesics 53
7.7 Geodesics in General Relativity 53
7.7.1 The Perihelion Shift 55

8 Hamilton-Jacobi Theory **57**
8.1 Conjugate Variables 57
8.2 The Hamilton-Jacobi Equation 58
8.3 The Euler Problem 59
8.4 The Classical Limit of the Schrodinger Equation 61
8.5 Hamilton-Jacobi Equation in Riemannian manifolds 62

9 Integrable Systems **63**
9.1 The Simple Harmonic Oscillator 64
9.2 The General One-Dimensional System 65
9.3 Bohr-Sommerfeld Quantization 65
9.4 The Kepler Problem 66
9.4.1 A Contour Integral 67
9.5 The Relativistic Kepler Problem 68
9.6 Several Degrees of Freedom 68

10 The Three Body Problem **70**
10.1 Preliminaries . . . 71
10.1.1 Scale Invariance . . . 71
10.2 Jacobi Co-Ordinates . . . 72
10.2.1 Orbits as Geodesics . . . 74
10.3 The $\frac{1}{r^2}$ potential . . . 75
10.3.1 Montgomery's Pair of Pants . . . 76

11 The Restricted Three Body Problem **78**
11.1 The Motion of the Primaries . . . 78
11.1.1 A useful identity . . . 80
11.1.2 Equilibrium points . . . 80
11.1.3 Hill's Regions . . . 81
11.1.4 The second derivative of the potential . . . 81
11.1.5 Stability Theory . . . 82

12 Magnetic Fields **84**
12.1 The Equations of Motion . . . 84
12.2 Hamiltonian Formalism . . . 84
12.3 Canonical Momentum . . . 86
12.4 The Lagrangian . . . 86
12.5 The Magnetic Monopole . . . 87
12.6 Quantization of electric charge . . . 89
12.7 The Penning Trap . . . 89

13 Discrete Time **91**
13.1 First Order Symplectic Integrators . . . 92
13.2 Second Order Symplectic Integrator . . . 93
13.3 Chaos With One Degree of Freedom . . . 95

14 Dynamics in One Real Variable **97**
14.1 Maps . . . 97
14.2 Doubling Modulo One . . . 99
14.3 Stability of Fixed Points . . . 100
14.4 Unimodal Maps . . . 101
14.4.1 One Stable Fixed Point: $1 > \mu > 0$ . . . 102
14.4.2 One Stable and One Unstable Fixed Point: $1 < \mu < 3$ . . . 102
14.4.3 Stable Orbit of period two : $3 < \mu < 1 + \sqrt{6}$ . . . 102
14.4.4 Stable Orbit of period four : $3.44949... < \mu < 3.54409..$ . . . 103
14.4.5 $3.54409.. < \mu < \mu_\infty$ . . . 104

15 Dynamics in One Complex Variable **105**
15.1 The Riemann Sphere . . . 105
15.2 Mobius Transformations . . . 106
15.3 Dynamics of a Mobius Transformation . . . 107
15.4 The map $z \mapsto z^2$ . . . 109

15.5 The map $z \mapsto z^2 + i$ 109
15.6 Metric on the Sphere 110
15.7 Convergence 111
15.8 Julia and Fatou Sets 111

16 Stability and Curvature **113**
16.1 Geodesic Deviation 113
16.2 The Laplace Operator 115

17 Fluid Mechanics **117**
17.1 Euler's Equation for an Ideal Fluid 117
17.2 Lie Algebra of Vector Fields 117
17.3 Euler Flow as Geodesics 117
17.4 Arnold's Computation of the Curvture of the Euler Flow 117
17.5 Dissipative Fluid Flow: Navier Stokes Equations 117

Chapter 1

The Variational Principle

Many problems in physics involve finding the minima (more generally extrema) of functions. For example, the equilibrium positions of a static system are the extrema of its potential energy: stable equilibria correspond to local minima. It is a surprise that even dynamical systems, whose positions depend on time, can be understood in terms of extremizing a quantity that depends on the paths: the action. In fact, all the fundamental physical laws of classical physics follow from such variational principles. There is even a generalization to quantum mechanics, based on averaging over paths where the paths of extremal action make the largst contribution.

In essence, the calculus of variations is the differential calculus of functions that depend on an infinite numer of variables. For example, suppose we want to find the shortest curve connecting two different points on the plane. Such a curve can be thought of as a function $(x(t), y(t))$ of some parameter (like time). It must satisfy the boundary conditions

$$x(t_1) = x_1, y(t_1) = y_1$$

$$x(t_2) = x_2, y(t_2) = y_2$$

where the initial and final points are given. The length is

$$S[x, y] = \int_{t_1}^{t_2} \sqrt{\dot{x}^2 + \dot{y}^2} dt$$

This is a function of an infinite number of points because we can make some small changes $\delta x(t), \delta y(t)$ at each time t independently. We can define a differential, the infinitesimal change of the length under such a change:

$$\delta S = \int_{t_1}^{t_2} \frac{\dot{x}\delta\dot{x} + \dot{y}\delta\dot{y}}{\sqrt{\dot{x}^2 + \dot{y}^2}} dt$$

Generalizing the idea from the calculus of several variables, we expect that at the extremum, this quantity will vanish for any $\delta x, \delta y$. This condition leads to a differential equation whose solution turns out to be (no surprise) a straightline.There are two key ideas here. First of all the variation of the time derivative is the time derivative of the variation:

$$\delta\dot{x} = \frac{d}{dt}\delta x$$

This is essenially a postulate on the nature of the variation. (It can be further justified if you want.) The second idea is an integration by parts, remembering that the variation must vanish at the boundary (we are not changing the initial and final point.)

$$\delta x(t_1) = \delta x(t_2) = 0 = \delta y(t_1) = \delta y(t_2)$$

Now,

$$\frac{\dot{x}}{\sqrt{\dot{x}^2+\dot{y}^2}}\frac{d}{dt}\delta x = \frac{d}{dt}\left[\frac{\dot{x}}{\sqrt{\dot{x}^2+\dot{y}^2}}\delta x\right] - \frac{d}{dt}\left[\frac{\dot{x}}{\sqrt{\dot{x}^2+\dot{y}^2}}\right]\delta x$$

and similarly with δy. Then

$$\delta S = \int_{t_1}^{t_2}\frac{d}{dt}\left[\frac{\dot{x}}{\sqrt{\dot{x}^2+\dot{y}^2}}\delta x + \frac{\dot{y}}{\sqrt{\dot{x}^2+\dot{y}^2}}\delta y\right]dt$$

$$-\int_{t_1}^{t_2}\left\{\frac{d}{dt}\left[\frac{\dot{x}}{\sqrt{\dot{x}^2+\dot{y}^2}}\right]\delta x + \frac{d}{dt}\left[\frac{\dot{y}}{\sqrt{\dot{x}^2+\dot{y}^2}}\right]\delta y\right\}dt$$

The first term is a total derivative and becomes

$$\left[\frac{\dot{x}}{\sqrt{\dot{x}^2+\dot{y}^2}}\delta x + \frac{\dot{y}}{\sqrt{\dot{x}^2+\dot{y}^2}}\delta y\right]_{t_1}^{t_2} = 0$$

because δx and δy both vanish at the boundary. Thus

$$\delta S = -\int_{t_1}^{t_2}\left\{\frac{d}{dt}\left[\frac{\dot{x}}{\sqrt{\dot{x}^2+\dot{y}^2}}\right]\delta x + \frac{d}{dt}\left[\frac{\dot{y}}{\sqrt{\dot{x}^2+\dot{y}^2}}\right]\delta y\right\}dt$$

In order for this to vanish for any variation, we must have

$$\frac{d}{dt}\left[\frac{\dot{x}}{\sqrt{\dot{x}^2+\dot{y}^2}}\right] = 0 = \frac{d}{dt}\left[\frac{\dot{y}}{\sqrt{\dot{x}^2+\dot{y}^2}}\right]$$

That is because we can choose a variation that is only non-zero in some tiny (as small you want) neighborhood of a particular value of t. Then the quantity multiplying it must vanish, independently at each value of t. These differential equations simply say that the vector $(\dot{x}, \dot{y})$have constant direction:$(\frac{\dot{x}}{\sqrt{\dot{x}^2+\dot{y}^2}}, \frac{\dot{y}}{\sqrt{\dot{x}^2+\dot{y}^2}})$ is just the unit vector along the tangent. So the solution is a straightline. Why did we do all this work to prove an intuitively obvious fact? For, sometimes intuitively obvious facts are wrong. Also, this method generalizes to situations where the answer is not at all obvious: what is the curve of shortest length between two points that lie entirely on the surface of a sphere?

1.1 Euler-Lagrange equations

In many problems, we will have to find the extremum of a quantity

$$S[q] = \int_{t_1}^{t_2} L[q, \dot{q}, t] dt$$

where $q^i(t)$ are a set of functions of some parameter t. We will call them position and time respectively, although the actual physical meaning may be something else in a particular case. The quantity $S[q]$ whose extremum we want to find is called the action. It depends on an infinite number of independent variables, the values of q at various times t. It is the integral of a function of position and velocity at a given time, integrated on some interval. It can also depend explicitly on time; if it does not, there are some special tricks we can use to simplify the solution of the problem.

As before we note that at an extremum S must be unchanged under small variations of q. Also we assume the identity

$$\delta\dot{q}^i = \frac{d}{dt}\delta q^i$$

We can now see that

$$\delta S = \int_{t_1}^{t_2} \sum_i \left[\delta\dot{q}^i \frac{\partial L}{\partial \dot{q}^i} + \delta q^i \frac{\partial L}{\partial q^i}\right] dt$$

$$= \int_{t_1}^{t_2} \sum_i \left[\frac{d\delta q^i}{dt} \frac{\partial L}{\partial \dot{q}^i} + \delta q^i \frac{\partial L}{\partial q^i}\right] dt$$

We then do an integration by parts,

$$= \int_{t_1}^{t_2} \sum_i \frac{d}{dt}\left[\delta q^i \frac{\partial L}{\partial \dot{q}^i}\right] dt$$

$$+ \int_{t_1}^{t_2} \sum_i \left[-\frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i} + \frac{\partial L}{\partial q^i}\right] \delta q^i \, dt$$

Again in physical appications, the boundary values of q at times t_1 and t_2 are given. So

$$\delta q^i(t_1) = 0 = \delta q^i(t_2)$$

Thus

$$\int_{t_1}^{t_2} \sum_i \frac{d}{dt}\left[\delta q^i \frac{\partial L}{\partial \dot{q}^i}\right] dt = \left[\delta q^i \frac{\partial L}{\partial \dot{q}^i}\right]_{t_1}^{t_2} = 0$$

and at an extremum,

$$\int_{t_1}^{t_2} \sum_i \left[-\frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i} + \frac{\partial L}{\partial q^i}\right] \delta q^i \; dt = 0$$

Since these have to be true for all variations, we get the differential equations

$$-\frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i} + \frac{\partial L}{\partial q^i} = 0.$$

This ancient argument is due to Euler and Lagrange, in the pioneering generation that figured out the consequences of Newton's Laws. The calculation we did earlier is a special case. As an exercise, rederive the equations for minimizing the length of a curve using the Euler-Lagrange equations.

Exercise 1. Find the solution to the Euler Lagrange equations that minimize

$$S[q] = \frac{1}{2}\int_0^a \dot{q}^2 dt$$

subject to the boundary conditions

$$q(0) = q_0, \quad q(a) = q_1$$

1.2 The Varational Principle of Mechanics

Newton's equation of motion of a particle of mass m and position q moving on the line, under a potential $V(q)$ is

$$m\ddot{q} = -\frac{\partial V}{\partial q}$$

There is a quantity $L(q, \dot{q})$such that the Euler-Lagrange equation for minimizing $S = \int L[q, \dot{q}]dt$ are just these equations.

We can write this equation as

$$\frac{d}{dt}[m\dot{q}] + \frac{\partial V}{\partial q} = 0.$$

So if we had

$$m\dot{q} = \frac{\partial L}{\partial \dot{q}}, \quad \frac{\partial L}{\partial q} = -\frac{\partial V}{\partial q}$$

we would have the right equations. A choice is

$$L = \frac{1}{2} m\dot{q}^2 - V(q)$$

This quantity is called the Lagrangian. Note that it is the **difference** of kinetic and potential energies, and not the sum.

More generally, the co-ordinate q may be replaced by a collection of numbers $q^i, i = 1, \cdots n$ which together describe the instantaneous position of a system of particles. The number n of such variables needed is called the number of degrees of freedom. Part of the advantage of the Lagrangian formalism over the older Newtonian one is that it allows even curvilinear co-ordinates: all you have to know arethe kinetic energy and potential energy in these co-ordinates. To be fair, the Newtonian formalism is more general in another direction, as it allows forces that are not conservative (a system can lose energy).

Example 2. The kinetic energy of a particle in spherical polar co-ordinates is

$$\frac{1}{2} m \left[\dot{r}^2 + r^2\dot{\theta}^2 + r^2 \sin^2\theta \dot{\phi}^2 \right].$$

Thus the Lagrangian of the Kepler problem is

$$L = \frac{1}{2} m \left[\dot{r}^2 + r^2\dot{\theta}^2 + r^2 \sin^2\theta \dot{\phi}^2 \right] + \frac{GMm}{r}$$

Chapter 2

Conservation Laws

2.1 Generalized Momenta

Recall that if q is a Cartesian co-ordinate,

$$p = \frac{\partial L}{\partial \dot{q}}$$

is the momentum in that direction. More generally, for any co-ordinate q^i the quantity

$$p_i = \frac{\partial L}{\partial \dot{q}^i}$$

is called the **generalized momentum** conjugate to q^i. For example, in spherical polar co-ordinates the momentum conjugate to ϕ is

$$p_\phi = mr^2\dot{\phi}.$$

You can see that this has the physical meaning of angular momentum around the third axis.

2.2 Conservation Laws

This definition of generalized momentum is motivated in part by a direct consequence of it: if L happens to be independent of a particular co-ordinate q^i (but might depend on $\dot{q}^i$), then the momentum conjugate to it is independent of time: is conserved:

$$\frac{\partial L}{\partial q^i} = 0 \implies \frac{d}{dt}\left[\frac{\partial L}{\partial \dot{q}^i}\right] = 0.$$

For example, p_ϕ is a conserved quantity in the Kepler problem. This kind of information is precious in solving a mechanics problem; so the Lagrangian formalism which identifies such conserved quantities is very convenient to actually solve for the equations of a system.

2.3 Conservation of Energy

L can have a time dependence through its dependence of $q, \dot{q}$ as well as explicitly. The total time derivative is

$$\frac{dL}{dt} = \sum_i \dot{q}^i \frac{\partial L}{\partial q^i} + \sum_i \ddot{q}^i \frac{\partial L}{\partial \dot{q}^i} + \frac{\partial L}{\partial t}$$

The E-L equations imply

$$\frac{d}{dt}\left[\sum_i p_i \dot{q}^i - L\right] = -\frac{\partial L}{\partial t}, \quad p_i = \frac{\partial L}{\partial \dot{q}^i}$$

In particular, if L has no explicit time dependence, the quantity called the **hamiltonian**,

$$H = \sum_i p_i \dot{q}^i - L$$

is conserved.

$$\frac{\partial L}{\partial t} = 0 \implies \frac{dH}{dt} = 0.$$

What is its physical meaning? Consider the example of a particle in a potential

$$L = \frac{1}{2} m \dot{q}^2 - V(q)$$

Since the kinetic energy T is a quadratic function of $\dot{q}$, and $V$is independent of $\dot{q}$,

$$p\dot{q} = \dot{q}\frac{\partial T}{\partial \dot{q}} = 2T$$

Thus

$$H = 2T - (T - V) = T + V.$$

Thus the hamiltonian, in this case, is the total energy.

More generally, if the kinetic energy is quadratic in the generalized velocities $\dot{q}^i$ (which is true very often) and if the potential energy is independent of velocities (also true often), the hamiltonian is the same as energy. There are some cases where the hamiltonian and energy are not the same though: for example, when we view a system in a reference frame that is not inertial. But these are unusual situations.

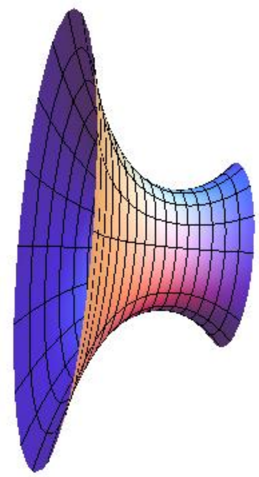

Figure 2.1:

2.4 Minimal Surface of Revolution

Although the main use of the variational calculus is in mechanics, it can also be used to solve some interesting geometric problems. A *minimal surface* is a surface whose area is unchanged under small changes of its shape. You might know that for a given volume, the sphere has minimal area. Another interesting question in geometry is to ask for a surface of minimal area which has a given curve (or disconnected set of curves) as boundary. The first such problem was solved by Euler. What is the surface of revolution of minimal area, with given radii at the two ends? Recall that a surface of revolution is what you get by taking some curve $y(x)$ and rotating it around the $x-$ axis. The cross-section at x is a circle of radius $y(x)$, so we assume that $y(x) > 0$. The boundary values $y(x_1) = y_1$ and $y(x_2) = y_2$ are given. We can, without loss of generality, assume that $x_2 > x_1$ and $y_2 > y_1$.What is the value of the radius $y(x)$in between $x_1$and x_2 that will minimize the area of this surface?

The area of a thin slice between x and $x + dx$ is $2\pi y(x)ds$ where $ds = \sqrt{1 + y'^2}dx$ is the arc-length of the cross-section.Thus the quantity to be minimized is

$$S = \int_{x_1}^{x_2} y(x)\sqrt{1 + y'^2}dx$$

This is the area divided by 2π.

We can derive the Euler Lagrange equation as before: y is analogous to q

and x is analogous to t. But it is smarter to exploit the fact that the integrand is independent of x: there is a conserved quantity

$$H = y' \frac{\partial L}{\partial y'} - L. \quad L = y(x)\sqrt{1+y'^2}$$

That is

$$H = y\frac{y'^2}{\sqrt{1+y'^2}} - y\sqrt{1+y'^2}$$

$$H\sqrt{1+y'^2} = -y$$

$$y' = \sqrt{\frac{y^2}{H^2} - 1}$$

$$\int_{y_1}^{y} \frac{dy}{\sqrt{\frac{y^2}{H^2} - 1}} = x - x_1$$

The substitution

$$y = H\cosh\theta$$

evalutes the integral:

$$H[\theta - \theta_1] = x - x_1$$

$$\theta = \frac{x - x_1}{H} + \theta_1$$

$$y = H\cosh\left[\frac{x - x_1}{H} + \theta_1\right]$$

The constants of integration are fixed by the boundary conditions

$$y_1 = H\cosh\theta_1$$

$$y_2 = H\cosh\left[\frac{x_2 - x_1}{H} + \theta_1\right]$$

The curve $y = H\cosh\left[\frac{x}{H} + \text{constant}\right]$ is called a catenary; the surface you get by revolving it around the x-axis is the catenoid. If we keep the radii fixed and move the boundaries far apart along the $x-$axis, at some critical distance, the surface will cease to be of minimal area. The minimal area is given by the disconnected union of two disks with the circles as boundaries. If we imagine a soap bubble bounded by two circles that are moved apart, at some distance it will break into two flat circles. Can you find the critical distance in terms of the bounding radius, assuming for simplicity that $y_1 = y_2$?

Chapter 3

The Simple Pendulum

Consider a mass g suspended from a fixed point by a rigid rod of length l. Also, it is only allowed to move in a fixed vertical plane.

The angle θ from the lowest point on its orbit serves as a position co-ordinate. The kinetic energy is

$$T = \frac{1}{2} m l^2 \dot{\theta}^2$$

and the potential energy is

$$V(\theta) = mgl(1 - \cos\theta).$$

Thus

$$T - V = ml^2 \left[\frac{1}{2}\dot{\theta}^2 - \frac{g}{l}(1 - \cos\theta) \right]$$

The overall constant will not matter to the equations of motion. So we can choose as Lagrangian

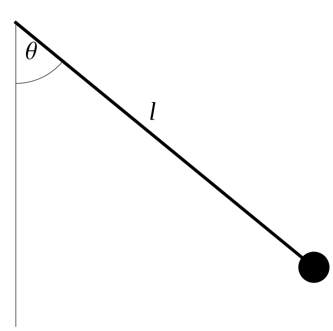

Figure 3.1:

$$L = \frac{1}{2}\dot{\theta}^2 - \frac{g}{l}(1 - \cos\theta)$$

This leads to the equation of motion

$$\ddot{\theta} + \frac{g}{l}\sin\theta = 0,$$

For small angles $\theta << \pi$ this is the equation for a harmonic oscillator with angular frequency

$$\omega = \sqrt{\frac{g}{l}}.$$

But for large amplitudes of oscillation the answer is quite different. To simplify calculations let us choose a unit of time such that ,$g = l$; i.e., such that $\omega = 1$. Then

$$L = \frac{1}{2}\dot{\theta}^2 - (1 - \cos\theta)$$

We can make progress in solving this system using the conservation of energy

$$H = \frac{\dot{\theta}^2}{2} + [1 - \cos\theta].$$

The key is to understand the critical points of the potential. The potential energy has a minimum at $\theta = 0$ and a maximum at $\theta = \pi$. The latter corresponds to an unstable equilibrium point: the pendulum standing on its head. If the energy is less than this maximum value,

$$H < 2$$

the pendulum oscillates back and forth around its equilibrium point. At the maximum angle, $\dot{\theta} = 0$ so that it is given by a transcendental equation

$$1 - \cos\theta_0 = H$$

The motion is periodic, with a period T that depends on energy.That is, we have

$$\sin\theta(t + T) = \sin\theta(t).$$

3.1 Algebraic Formulation

It will be useful to use a variable which takes some simple value at the maximum deflection; also we would like it to be periodic function of the angle. The condition for maximum deflection can be written

$$\sqrt{\frac{2}{H}}\sin\frac{\theta_0}{2} = \pm 1$$

This suggests that we use the variable

$$x = \sqrt{\frac{2}{H}} \sin\frac{\theta}{2}$$

so that at maximum deflection, we simply have $x = \pm 1$. Define also a quantity that parametrizes the energy

$$k = \sqrt{\frac{H}{2}}, \quad x = \frac{1}{k}\sin\frac{\theta}{2}.$$

Changing variables,

$$\dot{x} = \frac{1}{2k}\cos\frac{\theta}{2}\dot{\theta}, \quad \dot{x}^2 = \frac{1}{4k^2}\left(1 - \sin^2\frac{\theta}{2}\right)\dot{\theta}^2 = \frac{1}{4}\left(\frac{1}{k^2} - x^2\right)\dot{\theta}^2$$

Conservation of energy becomes

$$2k^2 = 2\frac{\dot{x}^2}{k^{-2} - x^2} + 2k^2x^2$$

Thus we get the differential equation

$$\dot{x}^2 = (1 - x^2)(1 - k^2x^2)$$

This can be solved in terms of Jacobi functions, which generalize trigonometric functions such as sin and cos.

3.2 Primer on Jacobi Functions

The functions $\mathrm{sn}(u,k), \mathrm{cn}(u,k), \mathrm{dn}(u,k)$ are defined as the solutions of the coupled ODE

$$\mathrm{sn}' = \mathrm{cn}\ \mathrm{dn}, \quad \mathrm{cn}' = -\mathrm{sn}\ \mathrm{dn}, \quad \mathrm{dn}' = -k^2 \mathrm{sn}\ \mathrm{cn}$$

with initial conditions

$$\mathrm{sn} = 0, \quad \mathrm{cn} = 1, \quad \mathrm{dn} = 1, \quad \text{at } u = 0.$$

It follows that

$$\mathrm{sn}^2 + \mathrm{cn}^2 = 1, \quad k^2\mathrm{sn}^2 + \mathrm{dn}^2 = 1$$

Thus

$$\mathrm{sn}'^2 = (1 - \mathrm{sn}^2)(1 - k^2\mathrm{sn}^2)$$

Thus we see that

$$x(t) = \mathrm{sn}(t, k)$$

is the solution to the pendulum. The inverse of this function (which expresses t as a function of x) can be expressed as an integral

$$t = \int_0^x \frac{dy}{\sqrt{(1-y^2)(1-k^2y^2)}}$$

This kind of integral first appeared when people tried to find the perimeter of an ellipse. So it is called an **elliptic integral.**

Detrour: Show that the perimeter of the ellipse $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ is equal to $4a \int_0^1 \sqrt{\frac{1-k^2x^2}{1-x^2}} dx$, where $k = \sqrt{1 - \frac{b^2}{a^2}}$.

The functions $\mathrm{sn}, \mathrm{cn}, \mathrm{dn}$ are called **elliptic functions**. The name is a bit unfortunate, because these functions appear even when there is no ellipse in sight, such as in our case. The parameter k is called the **elliptic modulus.**

Clearly, if $k = 0$, these functions reduce to trigonometric functions:

$$\mathrm{sn}(u, 0) = \sin u, \quad \mathrm{cn}(u, 0) = \cos u, \quad \mathrm{dn}(u, 0) = 1.$$

Thus, for small energies $k \to 0$ and our solution reduces to that of the harmonic oscillator.

From the connection with the pendulum it is clear that the functions are periodic, at least when $0 < k < 1$ (so that $0 < H < 2$ and the pendulum oscillates around the equilibrium point). The period of oscillation is four times the time it takes to go from the bottom to the point of maximum deflection

$$T = 4K(k), \quad K(k) = \int_0^1 \frac{dy}{\sqrt{(1-y^2)(1-k^2y^2)}}$$

This integral is called the **complete elliptic integral.** When $k = 0$, it evaluates to $\frac{\pi}{2}$ so that the period is 2π. That is correct, since we chose the unit of time such that $\omega = \sqrt{\frac{l}{g}} = 1$ and the period of the harmonic oscillator is $\frac{2\pi}{\omega}$. As k grows, the period increases: the pendulum oscillates with larger amplitude. As $k \to 1$ the period tends to infinity: the pendulum has just enough energy to get to the top of the circle, with velocity going to zero as it gets there.

3.3 Elliptic Curves

Given the position x, and velocity $\dot{x}$ at any instant, they are determined for all future times by the equations of motion. Thus it is convenient to think of a space whose co-ordinates are $(x, \dot{x})$ .The conservation of energy determines the shape of the orbit in phase space.

$$\dot{x}^2 = (1-x^2)(1-k^2x^2)$$

In the case of a pendulum, this is an extremely interesting thing called an **elliptic curve**. The first thing to know is that **an elliptic curve is not an**

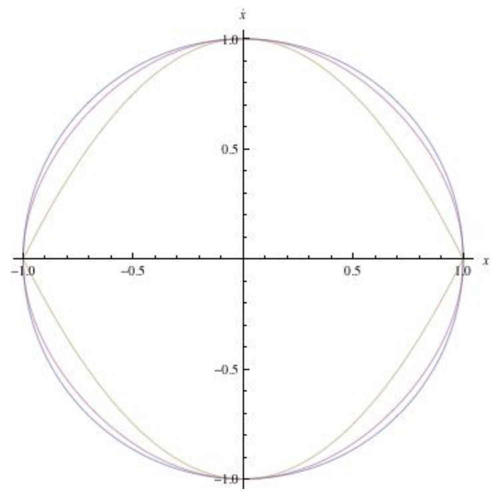

Figure 3.2:

ellipse. It is called that because elliptic functions can be used to parametrically decribe points on this curve:

$$\dot{x} = \mathrm{sn}'(u, k), \quad x = \mathrm{sn}(u, k)$$

For small k an elliptic curve looks more or less like a circle; but as $k > 0$ it is deformed into a more interesting shape. When $k \to 1$ it tends to a parabola.

Only the part of the curve with real x between 1 and -1 has a physical significance in this application.But, as usual to understand any algebraic curve it helps to analytically continue into the complex plane. The surprising thing is that the curve is then a torus; this follows from the double peridocity of sn, which we prove below.

3.4 Addition Formula

Suppose x_1 is the position of the pendulum after a time t_1 has elapsed, assuming that at at time zero, $x = 0$ as well. Similarly let x_2 be the position at some time t_2. If $t_3 = t_1 + t_2$, and x_3 is the position at time t_3, it should not be surprising that x_3 can be found once we know x_1 and x_2 . What is surprising is that x_3 is an *algebraic* function of the positions. This is because of the addition formula for elliptic functions:

$$\int_0^{x_1} \frac{dy}{\sqrt{(1-y^2)(1-k^2y^2)}} + \int_0^{x_2} \frac{dy}{\sqrt{(1-y^2)(1-k^2y^2)}} = \int_0^{x_3} \frac{dy}{\sqrt{(1-y^2)(1-k^2y^2)}}$$

$$x_3 = \frac{x_1\sqrt{(1-x_2^2)(1-k^2x_2^2)} + x_2\sqrt{(1-x_1^2)(1-k^2x_1^2)}}{1-k^2x_1^2x_2^2}$$

If $k = 0$ this is just the addition formula for sines: $\sin[\theta_1+\theta_2] = \sin\theta_1\sqrt{1-\sin^2\theta_2}+\sin\theta_2\sqrt{1-\sin^2\theta_1}$. This operation $x_1, x_2 \mapsto x_3$ satisfies the conditions for an abelian group.The point of stable equilibrium $x = 0, \dot{x} = 1$ is the identity element. The inverse of x_1 just $-x_1$. You can have fun trying to prove algebraically that the above operation is associative. (Or die trying.)

By adding x to itself using the above formula we can get a kind of multiplication of an integer with x. (If n is negative, you add $-x$ to itself $|n|$ times.) If $n > 1$ the algebraic formula for this multiplication is quite complicated: the sequence of points $x,\ 2*x,\ 3*x,\ 4*x\cdots$. If we let a strobe light shine at regular intervals and note the positions of the pendulum at those times, this is the sequence we will get. Suppose you know the initial position x and a much later position $n*x$; can you determine n? This is very hard, because for large enough n the sequence of positions will look almost random, unless the strobe period is a rational multiple of the period of the pendulum. Thus even a simple, exactly solvable system like the pendulum can led to almost random sequences of position.

But for certain special choices of x ,which correspond to a strobe light set to a rational fraction of the period of oscillation, the multiples of x will form a cyclic subgroup.

3.5 Imaginary Time

The replacement $t \to it$ has the effect of reversing the sign of the potential in Newton's equations, $\ddot{q} = -V'(q)$. In our case, $\ddot{\theta} = -\sin\theta$,it amount to revesing the direction of the gravitational field. In term of of co-ordinates, this amounts to $\theta \to \theta + \pi$. Under the transformations $t \to it, \theta \to \theta + \pi$, the conservation of energy

$$k^2 = \frac{\dot{\theta}^2}{4} + \sin^2\frac{\theta}{2}$$

goes over to

$$1 - k^2 = \frac{\dot{\theta}^2}{4} + \sin^2\frac{\theta}{2}$$

The quantity $k' = \sqrt{1-k^2}$ is called the **complementary modulus**. In summary, the simple pendulum has a symmetry

$$t \to it, \theta \to \theta + \pi, k \to k'.$$

This transformation maps an oscillation of small amplitude (small k)to one of large amplitude (k close to 1).

This means that if we analytically continue the solution of the pendulum into the complex $t-$plane, it must be periodic with period $4K(k)$ in the real direction and $4K(k')$ in the imaginary direction.

Exercise 3. Using the change of variable $x \mapsto \frac{1}{\sqrt{1-k'^2x^2}}$, show that $K(k') = \int_1^{\frac{1}{k}} \frac{dx}{\sqrt{[x^2-1][1-k^2x^2]}}$.

3.5.1 The case of $H = 1$

The minimum value of energy is zero and the maximum value for an oscillation is 2. Exactly half way is the oscillation whose energy is 1 ; the maximum angle is $\frac{\pi}{2}$. This orbit is invariant under the above transformation that inverts the potential: either way you look at it, the pendulum bob is horizontal at maximum deflection. In this case the real and imaginary periods are of equal magnitude.

3.6 The Arithmetic-Geometric Mean

Landen, and later, Gauss. found a surprising symmetry for the elliptic integral $K(k)$ that allows a calculation of its value by iterating simple algebraic operations. In our context it means that the period of a pendulum is unchanged if the energy H and angular frequency ω are changed in a certain way that decreases their values. By iterating this we can make the energy tend to zero: but in this limit we know that the period is just 2π over the angular frequency. **In this section we do not set $\omega = 1$ but we continue to factor out ml^2 from the Lagrangian as before.** Then the Lagrangian $L = \frac{1}{2}\dot{\theta}^2 - \omega^2[1-\cos\theta]$ and H have dimensions of the square of frequency.

Let us go back and look at the formula for the period:

$$T = \frac{4}{\omega}K(k), \quad K(k) = \int_0^1 \frac{dx}{\sqrt{(1-x^2)(1-k^2x^2)}}, \quad 2k^2 = \frac{H}{\omega^2}.$$

If we make the substitution

$$x = \sin\phi$$

this becomes

$$T = \frac{4}{\omega}\int_0^{\frac{\pi}{2}} \frac{d\phi}{\sqrt{1-k^2\sin^2\phi}}$$

That is,

$$T(\omega, b) = 4\int_0^{\frac{\pi}{2}} \frac{d\phi}{\sqrt{\omega^2\cos^2\phi + b^2\sin^2\phi}}$$

where

$$b = \sqrt{\omega^2 - \frac{H}{2}}.$$

Note that $\omega > b$ with $\omega \to b$ implying $H \to 0$.The surprising fact is that the integral remains unchanged under the transformations

$$\omega_1 = \frac{\omega + b}{2}, \quad b_1 = \sqrt{\omega b}.$$

$$T(\omega, b) = T(\omega_1, b_1)$$

Exercise 4. Prove this identity. First put $y = b \tan \phi$ to get $T(\omega, b) = 2\int_{-\infty}^{\infty} \frac{dy}{\sqrt{(\omega^2+y^2)(b^2+y^2)}}$. Then make the change of variable $y = z + \sqrt{z^2 + \omega b}$. This proof, due to Newman, was only found in 1985. Gauss' and Landen's proofs were much clumsier. For further explanation, see *Elliptic Curves* by H. McKean and V. Moll .

That is,ω is replaced by the arithmetic mean and b by the geometric mean. Recall that given two numbers a>b>0, the arithmetic mean is defined by

$$a_1 = \frac{a + b}{2}$$

and the geometric mean is defined as

$$b_1 = \sqrt{ab}.$$

As an exercise it is easy to prove that in general $a_1 \geq b_1$. If we iterate this transformation,

$$a_{n+1} = \frac{a_n + b_n}{2}, \quad b_{n+1} = \sqrt{a_n b_n}$$

The two sequences converge to the same number,$a_n \to b_n$ as $n \to \infty$. This limiting value is called the **Arithmetic-Geometric Mean** $\mathrm{AGM}(a, b)$

Thus, the energy of the pendulum tends to zero under this iteration applied to $\omega$and b, since $\omega_n \to b_n$; and the period is the limit of $\frac{2\pi}{\omega_n}$:

$$T(\omega, b) = \frac{2\pi}{\mathrm{AGM}(\omega, b)}.$$

The convergence of the sequence is quite fast, and gives a very accurate, and elementary, way to calculate the period of a pendulum; i.e., without having to calculate any integral.

Exercise 5. Calculate the period of the pendulum with $\omega = 1, H = 1$ by calculating the Arithmetic-Geometric mean. How many iterations do you need to get an accuracy of five decimal places for the AGM?

3.6.1 The Arithmetic-Harmonic Mean is the Geometric Mean

Why would Gauss have thought of the Arithmetic-Geometric Mean? This is perhaps puzzling to a modern reader brought up on calculators. But it is not so strange if you know how to calculate square roots by hand.

Recall that the harmonic mean of a pair of numbers is the reciprocal of the mean of their reciprocals. That is

$$\mathrm{HM}(a,b) = \frac{1}{\frac{1}{2}\left(\frac{1}{a}+\frac{1}{b}\right)} = \frac{2ab}{a+b}$$

Using $(a+b)^2 > 4ab$ it follows that $\frac{a+b}{2} > \mathrm{HM}(a,b)$.Suppose we define an iterative process where by we take the successive arithmetic and harmonic means:

$$a_{n+1} = \frac{a_n+b_n}{2}, \quad b_{n+1} = \mathrm{HM}(a_n,b_n)$$

These two sequences approach each other and the limit can be defined to be the Arithmetic-Harmonic Mean.

$$\mathrm{AHM}(a,b) = \lim_{n\to\infty} a_n$$

In other words, $\mathrm{AHM}(a,b)$ is defined the invariance property

$$\mathrm{AHM}(a,b) = \mathrm{AHM}\left(\frac{a+b}{2}, \mathrm{HM}(a,b)\right)$$

What is this quantity? It is none other than the geometric mean! Simply verify that

$$\sqrt{ab} = \sqrt{\frac{a+b}{2}\frac{2ab}{a+b}}.$$

Thus iterating the arithmetic and harmonic means with 1 is a good way to calculate the square root of any number. (Try it.)

Now you see that it is natural to wonder what we would get if we do the same thing one more time, iterating the Arithmetic and the Geometric Means.

$$\mathrm{AGM}(a,b) = \mathrm{AGM}\left(\frac{a+b}{2}, \sqrt{ab}\right).$$

I don't know if this is how Gauss discovered it: but it is not such a strange idea.

Exercise 6. Relate the Harmonic-Geometric Mean, defined by the invariance below to the AGM.

$$\mathrm{HGM}(a,b) = \mathrm{HGM}\left(\frac{2ab}{a+b}, \sqrt{ab}\right)$$

Exercise 7. (Suggested research probem.) What is the function defined by the following invariance property?

$$\mathrm{AAGM}(a,b) = \mathrm{AAGM}\left(\frac{a+b}{2}, \mathrm{AGM}(a,b)\right)$$

3.7 Doubly Periodic functions

This leads to the modern definition: **an elliptic function is a doubly periodic analytic function of a complex variable**. We allow for poles, but not branch cuts: thus, to be precise, an elliptic function is a doubly periodic meromorphic function of a complex variable.

$$f(z + m_1\tau_1 + m_2\tau_2) = f(z)$$

for integer m_1, m_2 and complex numbers τ_1, τ_2 which are called the periods. The points at which $f$takes the value form a lattice in the complex plane; once we know the values of $f$in a parallelogram whose sides are τ_1 and τ_2, we will know it every where by translation by some integer linear combination of the periods. In the case of the simple pendulum above, one of the periods is real and the other is purely imaginary. More generally, they could both be complex numbers; as long as the area of the fundamental parallelolgram is non-zero, we will get a lattice. By a rotation and a rescaling of the variable, we can always choose one of the periods to be real. The ratio of the two periods

$$\tau = \frac{\tau_2}{\tau_1}$$

is thus the quantity that determines the shape of the lattice. It is possibe to take some rational function and sum over its values at the points $z + m_1\tau_1 + m_2\tau_2$ to get a doubly periodic function, provided that this sum converges. An example is

$$\mathcal{P}'(z) = -2 \sum_{m_1, m_2 = -\infty}^{\infty} \frac{1}{(z + m_1\tau_1 + m_2\tau_2)^3}$$

The power 3 in the denominator is the smallest one for which this sum converges; the factor of -2 in front is there to agree with some conventions.. It has triple poles at the origin and all of its translates $m_1\tau_1 + m_2\tau_2$. It is the derivative of another elliptic function called $\mathcal{P}$, the **Weierstrass elliptic function**. It is possible to express the Jacobi elliptic functions in terms of the Weierstrass function: these two approaches complement each other. See *Elliptic Curves* by H. McKean and V. Moll for more on these matters.

Chapter 4

The Kepler Problem

Much of mechanics was developed in order to understand the motion of planets. Long before Copernicus, many astronomers knew that the apparently erratic motion of the planets can be simply explained as circular motion around the Sun. For example, the *Aryabhateeyam* written in 499 AD gives many calculations based on this model. But various religious taboos and superstitions prevented this simple picture from being universally accepted. It is ironic that the same superstitions (e.g., astrology) were the prime cultural motivation for studying planetary motion. Kepler himself is a transitional figure. He was originally motivated by astrology, yet had the scientific sense to insist on pecise agreement between theory and observation.

Kepler used Tycho Brahe's accurate measurements of planetary positions to find a set of important refinements of the heliocentric model. The three laws of planetary motion he discovered started the scientific revolution which is still continuing. We will rearrange the order of presentation of the laws of Kepler to make the logic clearer. Facts are not always discovered in the correct logical order: reordering them is essential to understanding them.

4.1 The orbit of a planet lies on a plane which contains the Sun

We may call this zeroth law of planetary motion: this is a significant fact in itself. If the **direction** of angular momentum is preserved, the orbit would have to lie in a plane. Since $\mathbf{L} = \mathbf{r} \times \mathbf{p}$, this plane is normal to the direction of $\mathbf{L}$. In polar co-ordinates in this plane, the angular momentum is

$$L = mr^2 \frac{d\phi}{dt}$$

That is, the moment of inertia times the angular velocity. In fact, all the planetary orbits lie on the *same* plane to a good approximation. This plane is normal to the angular momentum of the original gas cloud that formed the solar system.

4.2 The line connecting the planet to the Sun sweeps equal areas in equal times

This is usually called the second Law of Kepler. Since the rate of change of this area is $\frac{r^2}{2}\frac{d\phi}{dt}$, this is the statement that

$$r^2\frac{d\phi}{dt} = \text{constant}$$

This can be understood as due to the conservation of angular momentum. If the force is always directed towards the Sun, this can be explained.

4.3 Planets move along elliptical orbits with the Sun at a focus

This is the famous first Law of Kepler. It is significant the motion is a closed curve and that it is periodic: for most central potentials neither statement is true.

An **ellipse** is a curve on the plane defined by the equation, in polar co-ordinates r, ϕ

$$\frac{\rho}{r} = 1 + \epsilon \cos\phi$$

The parameter ϵ must be between 0 and 1 and is called the **eccentricity**. It measures the deviation of an ellipse from a circle: if $\epsilon = 0$ the curve is a circle of radius ρ . In the opposite limit $\epsilon \to 1$ (keeping ρ fixed) it approaches a parabola. The parameter ρ measures the size of the ellipse.

A more geometrical description of the ellipse is this: Choose a pair of points on the plane F_1, F_2, the **Focii**. If we let a point move on the plane such that the sum of its distances to F_1 and F_2 i s a constant, it will trace out an ellipse.

Exercise 8. Derive the equation for the ellipse above from this geometrical description. (Choose the origin of the polar co-ordinate system to be F_1. What is the position of the other focus ?)

The line connecting the two farthest points on an ellipse is called its **major axis**; this axis passes through the focii. The perpendicular bisector to the major axis is the **minor axis**. If these are equal in length, the ellipse is a circle; in this case the focii coincide. Half of the length of the major axis is called a usually. Similarly, the semi-minor-axis is called b.

Exercise 9. Show that the major axis is $\frac{2\rho}{1-\epsilon^2}$ and that the eccentricity is $\epsilon = \sqrt{1 - \frac{b^2}{a^2}}$.

The eccentricity of planetary orbits is quite small: a few percent. Comets, some asteroids and planetary probes have very eccentric orbits. If the eccentricity is greater then one, the equation describes a curve that is not closed, called a **hyperbola**.

In the *Principia*, Newton proved that an elliptical orbit can be explained by a force directed towards the Sun, which is inversely proportional to the square of distance. Where did he get the idea of a force proportional to the square of distance? The third law of Kepler provides a clue.

4.4 The ratio of the cube of the semi-major axis to the square of the period is the same for all planets

It took seventeen years of hard work for Kepler to go from the second Law to this third law. Along the way, he considered and discarded many ideas on planetary distances that came from astrology and Euclidean geometry (Platonic solids).

If we ignore the eccentricity (which is anyway small) for the moment and consider just a circular orbit of radius r, this is saying that

$$T^2 \propto r^3$$

We already know that the force on the Planet must be pointed toward the Sun, from the conservation of angular momentum. What is the dependence of the force on distance that wil give this dependence of the period? Relating the force to the centripetal acceleration,

$$m\frac{v^2}{r} = F(r)$$

Now, $v = r\dot{\theta}$ and $\dot{\theta} = \frac{2\pi}{T}$ for uniform circular motion.Thus

$$T^2 \propto \frac{r}{F(r)}$$

So we see that $F(r) \propto \frac{1}{r^2}$. Hooke, a much less renowned scientist than Newton, verified using a mechanical model that orbits of particles in this force are ellipses. Newton did not understand this at the time. He discovered an amazing proof of this fact using only geometry (no calculus) while he was writing the *Principia*.

From the fact that the ratio is independent of the planet, we can conclude that the acceleration is independent of the mass of the plant: that the force is proportional to the product of masses. Thus we arrive at Newton's Law of Gravity:

> The gravitational force on a body due to another is pointed along the line connecting the bodies; it has magnitude proportional to the product of masses and inversely to the square of the distance.

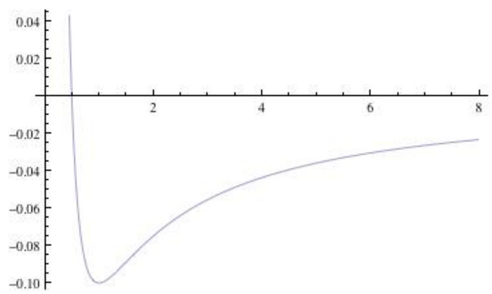

Figure 4.1:

4.5 The shape of the orbit

We now turn to deriving the shape of a planetary orbit from Newton's law of gravity. The Lagrangian is, in plane polar co-ordinates centered at the Sun,

$$L = \frac{1}{2}m\left[\dot{r}^2 + r^2\dot{\phi}^2\right] + \frac{GMm}{r}$$

From this we deduce the momenta

$$p_r = m\dot{r}, \quad p_\phi = mr^2\dot{\phi}$$

and the hamiltonian

$$H = \frac{p_r^2}{2m} + \frac{p_\phi^2}{2mr^2} - \frac{GMm}{r}.$$

Since $\frac{\partial H}{\partial \phi} = 0$ it follows right away that p_ϕ is conserved.

$$H = \frac{p_r^2}{2m} + V(r)$$

where

$$V(r) = \frac{p_\phi^2}{2mr^2} - \frac{GMm}{r}$$

is an **effective potential,** the sum of the gravitational potential and the kinetic energy due to angular motion.

So,

$$\dot{r} = \frac{p_r}{m}$$

$$\dot{p_r} = -V'(r)$$

Right away, we see that there is a circular orbit at the minimum of the potential:

$$V'(r) = 0 \implies r = \frac{p_\phi^2}{GMm^2}.$$

More generally, when $H < 0$, we should expect an oscillation around this minimum., between the **turning points** , which are the roots of $H - V(r)=0$. For $H > 0$ the particle will come in from infinity and, after reflection at a turning point, escape back to infiinity.

The shape of the orbit is given by relating r to ϕ. Using

$$\frac{dr}{dt} = \frac{d\phi}{dt}\frac{dr}{d\phi} = \frac{p_\phi}{mr^2}\frac{dr}{d\phi}$$

This suggests the change of variable

$$u = A + \frac{\rho}{r}, \implies \frac{dr}{dt} = -\frac{p_\phi}{m\rho}\frac{du}{d\phi} = -\frac{p_\phi}{m\rho}u'$$

for some constants A, ρ we will choose for convenience later. We can express the conservation of energy

$$H = \frac{1}{2}m\dot{r}^2 + V(r)$$

as

$$H = \frac{p_\phi^2}{2m\rho^2}u'^2 + \frac{p_\phi^2}{2m\rho^2}(u-A)^2 - \frac{GMm}{\rho}(u-A),$$

$$\frac{2m\rho^2 H}{p_\phi^2} = u'^2 + (u-A)^2 - \frac{2GMm^2\rho}{p_\phi^2}(u-A),$$

We can now choose the constants so that the term linear in u cancels out

$$A = -1, \quad \rho = \frac{p_\phi^2}{GMm^2}$$

and

$$u'^2 + u^2 = \epsilon^2$$

$$\epsilon^2 = 1 + \frac{2p_\phi^2 H}{(GM)^2 m^3}$$

A solution is now clear

$$u = \epsilon\cos\phi$$

or

$$\frac{\rho}{r} = 1 + \epsilon\cos\phi.$$

This is the equation for a conic section of eccentricity ϵ. If $H < 0$ so that the planet cannot escape to infinity, this is less than one, giving an ellipse as the orbit.

Chapter 5

The Rigid Body

If the distance between any two points on a body remains constant as it moves, it is a rigid body. Any configuration of the rigid body can be reached from the initial one by a translation of its center of mass and a rotation around it. Since we are mostly interested in the rotational motion, we will only consider the case of a body on which the total force is zero: the center of mass moves at a constant velocity. In this case we can transform to the reference frame in which the center of mass is at rest: the origin of our co-ordinate system can be placed there. It is not hard to put back in the translational degree of freedom once rotations are understood.

The velocity of one of the particles making up the rigid body can be split as

$$\mathbf{v} = \mathbf{\Omega} \times \mathbf{r}$$

The vector $\mathbf{\Omega}$ is the **angular velocity** :Its direction is the axis of rotation, and its magnitude the rate of change of its angle. The kinetic energy of this particle inside the body is

$$\frac{1}{2}\left[\mathbf{\Omega} \times \mathbf{r}\right]^2 \rho(\mathbf{r}) d^3\mathbf{r}$$

Here $\rho(\mathbf{r})$ is the mass density at the position of the particle; we assume that it occupies some infinitesimally small volume $d^3\mathbf{r}$. Thus the total rotational kinetic energy is

$$T = \frac{1}{2}\int \left[\mathbf{\Omega} \times \mathbf{r}\right]^2 \rho(\mathbf{r}) d^3\mathbf{r}$$

Now, $(\mathbf{\Omega} \times \mathbf{r})^2 = \Omega^2 r^2 - (\mathbf{\Omega} \cdot \mathbf{r})^2 = \Omega_i \Omega_j \left[r^2 \delta_{ij} - r_i r_j\right]$we get

$$T = \frac{1}{2}\Omega_i \Omega_j \int \rho(\mathbf{r}) \left[r^2 \delta_{ij} - r_i r_j\right] d^3\mathbf{r}$$

5.1 The Moment of Inertia

Define the **moment of inertia** to be the symmetric matrix

$$I_{ij} = \int \rho(\mathbf{r}) \left[r^2 \delta_{ij} - r_i r_j\right] d^3\mathbf{r}$$

Thus

$$T = \frac{1}{2}\Omega_i \Omega_j I_{ij}$$

Being a symmetric matrix, there is an orthogonal co-ordinate system in which the moment of inetria is diagonal:

$$T = \frac{1}{2}\left[I_1\Omega_1^2 + I_2\Omega_2^2 + I_3\Omega_3^2\right]$$

The eigenvalues I_1, I_2, I_3 are called the **principal moments of inertia**. They are positive numbers because I_{ij} is a positive matrix; i.e., $u^T I u \geq 0$ for any u.

Exercise 10. Show that the sum of any two principal moments is greater than or equal to the third one. $I_1 + I_2 \geq I_3$ etc.

The shape of the body and how mass is distributed inside it, determines the moment of inertia. The simplest case is when all three are equal. This happens if the body is highly symmetric: a sphere, a regular solid such as a cube. The next simplest case is when two of the moments are equal and the third is different. This is a body that has one axis of symmetry: a cylinder, a prism whose base is a regular polygon etc. The most complicated case is when the three eigenvalues are all unequal. This is the case of the asymmetrical top.

5.2 Angular Momentum

The angular momentum of a small particle inside the rigid body is

$$dM\mathbf{r} \times \mathbf{v} = \left[\rho(\mathbf{r})d^3\mathbf{r}\right] \mathbf{r} \times (\mathbf{\Omega} \times \mathbf{r})$$

Using the identity $\mathbf{r} \times (\mathbf{\Omega} \times \mathbf{r}) = \mathbf{\Omega} r^2 - \mathbf{r}\,(\mathbf{\Omega} \cdot \mathbf{r})$ we get the total angular momentum of the body to be

$$\mathbf{L} = \int \rho(\mathbf{r})[r^2\mathbf{\Omega} - \mathbf{r}\,(\mathbf{\Omega} \cdot \mathbf{r})]d^3\mathbf{r}$$

In terms of components

$$L_i = I_{ij}\Omega_j$$

Thus the moment of inertia relates angular velocity to angular momentum, just as mass relates velocity to momentum. The important difference is that moment of inetria is a matrix so that $\mathbf{L}$ and $\mathbf{\Omega}$ do not have to point in the same direction. Recall that the rate of change of angular momentum is the torque, if they are measured in an inertial reference frame.

Now here is a tricky point. We would like to use a co-ordinate system in which the moment of inetria is a diagonal matrix: that would simplify the relation of angular momentum to angular velocity:

$$L_1 = I_1\Omega_1$$

etc. But this may not an inertial co-ordinate system, as it's axes have to rotate with the body. So we must relate the change of a vector (such as $\mathbf{L}$) in a frame that is fixed to the body to an inertial frame. The difference between the two is a rotation of the body itself, so that

$$\left[\frac{d\mathbf{L}}{dt}\right]_{\text{inertial}} = \frac{d\mathbf{L}}{dt} + \mathbf{\Omega}\times\mathbf{L}$$

This we set equal to the torque acting on the body as a whole.

5.3 Euler's Equations

Even in the special case when the torque is zero the equations of motion of a rigid body are non-linear, since $\mathbf{\Omega}$ and $\mathbf{L}$ are proportional to each other:

$$\frac{d\mathbf{L}}{dt} + \mathbf{\Omega}\times\mathbf{L} = 0$$

In the co-ordinate system with diagonal moment of inertia

$$\Omega_1 = \frac{L_1}{I_1}$$

these become

$$\frac{dL_1}{dt} + a_1L_2L_3 = 0, \quad a_1 = \frac{1}{I_2} - \frac{1}{I_3}$$

$$\frac{dL_2}{dt} + a_2L_3L_1 = 0, \quad a_2 = \frac{1}{I_3} - \frac{1}{I_1}$$

$$\frac{dL_3}{dt} + a_3L_1L_2 = 0, \quad a_3 = \frac{1}{I_1} - \frac{1}{I_2}$$

These equations were originally derived by Euler. Clearly, if all the principal moments of inertia are equal these are trivial to solve: $\mathbf{L}$ is a constant.

The next simplest case

$$I_1 = I_2 \neq I_3$$

is not too hard either. Then $a_3 = 0$ and $a_1 = -a_2$.
It follows that L_3 is a constant. Also, L_1 and L_2 **precess** around this axis:

$$L_1 = A\cos\omega t, \quad L_2 = A\sin\omega t$$

with

$$\omega = a_1 L_3.$$

An example of such a body is the Earth. It is not quite a sphere, because it bulges at the equator compared to the poles. The main motion of the Earth is its rotation around the North-South axis once every 24 hours. But this axis itself precesses once every 26000 years. This means that the axis was not always aligned with the Pole star in distant past. Also, the times of the equinoxes change by a few minutes each year. As early as 280BC Aristarchus described this precsession of the equinoxes. It was Newton who finally explained it physically.

5.4 Jacobi's Solution

The general case of unequal moments can be solved in terms of Jacobi elliptic functions: in fact these functions were invented for this purpose. But before we do that it is useful to find the constants of motion. It is no surprise that the energy

$$H = \frac{1}{2}I_1\Omega_1^2 + \frac{1}{2}I_2\Omega_2^2 + \frac{1}{2}I_3\Omega_3^2 = \frac{L_1^2}{2I_1} + \frac{L_2^2}{2I_2} + \frac{L_3^2}{2I_3}$$

is conserved. You can verify that the magnitude of angular momentum is conserved as well:

$$L^2 = L_1^2 + L_2^2 + L_3^2.$$

Exercise 11. Calculate the time derivatives of $H$and L^2 and verify that they are zero.

Recall that

$$\mathrm{sn}' = \mathrm{cn}\ \mathrm{dn}, \quad \mathrm{cn}' = -\mathrm{sn}\ \mathrm{dn}, \quad \mathrm{dn}' = -k^2\mathrm{sn}\ \mathrm{cn}$$

with initial conditions

$$\mathrm{sn} = 0, \quad \mathrm{cn} = 1, \quad \mathrm{dn} = 1, \quad \text{at } u = 0.$$

Moreover

$$\mathrm{sn}^2 + \mathrm{cn}^2 = 1, \quad k^2\mathrm{sn}^2 + \mathrm{dn}^2 = 1$$

Make the ansatz

$$L_1 = A_1\mathrm{cn}(\omega t, k) \quad L_2 = A_2\mathrm{sn}(\omega t, k), \quad L_3 = A_3\mathrm{dn}(\omega t, k)$$

We get conditions

$$-\omega A_1 + a_1 A_2 A_3 = 0$$

$$\omega A_2 + a_2 A_3 A_1 = 0$$

$$-\omega k^2 A_3 + a_3 A_1 A_2 = 0$$

We want to express the five constants A_1, A_2, A_3, ω, k that appear in the solution in terms of the five physical parameters H, L, I_1, I_2, I_3. Some serious algebra will give[1]

$$\omega = \sqrt{\frac{(I_3 - I_2)(L^2 - 2HI_1)}{I_1 I_2 I_3}}$$

$$k^2 = \frac{(I_2 - I_1)(2HI_3 - L^2)}{(I_3 - I_2)(L^2 - 2HI_1)}$$

and

$$A_2 = \sqrt{\frac{(2HL_3 - L^2)I_2}{I_3 - I_2}}$$

etc.

The quantum mechanics of the rigid body is of much interest in molecular physics. So it is interesting to reformulate this theory in a way that makes the passage to quantum mechanics more natural. The Poisson brackets of angular momentum derived later give such a formulation.

[1] We can label our axes such that $I_3 > I_2 > I_1$.

Chapter 6

Hamilton's Principle

William Rowan Hamilton was the Royal Astronomer for Ireland. In this capacity, he worked on two important problems of mathematical interest: the motion of celestial bodies and the properties of light needed to design telescopes. Amazingly, he found that the laws of mechanics and those of ray optics were, in the proper mathematical framework, remarkably similar. But ray optics is only an approximation, valid when the wavelength of light is small. He probably wondered in the mid nineteenth century: could mechanics be the short wavelength approximation of some wave mechanics?

The discovery of quantum mechanics brought this remote oupost of theoretical physics into the very center of modern physics.

6.1 Generalized Momenta

Recall that to each co-ordinate q^i we can associate a momentum variable,

$$p_i = \frac{\partial L}{\partial \dot{q}^i}$$

p_i is said to be **conjugate** to q^i. It is possible to eliminate the velocities and write the equations of motion in terms of q^i, p_i In this language the equations will be a system of first order ODEs. Recall that from the definition of the hamiltonian

$$L = \sum_i p_i \dot{q}^i - H.$$

So if we view $H(q,p)$ as a function of position and momentum, we get a formula for the action

$$S = \int \left[\sum_i p_i \dot{q}^i - H(q,p,t) \right] dt$$

Suppose we find the condition for the action to be an extremum, treating q^i, p_i as independent variables:

$$\delta S = \int \sum_i \left[\delta p_i \dot{q}^i + p_i \frac{d}{dt}\delta q^i - \delta p_i \frac{\partial H}{\partial p_i} - \delta q^i \frac{\partial H}{\partial q^i} \right] dt = 0$$

We get the system of ODE

$$\frac{dq^i}{dt} = \frac{\partial H}{\partial p_i}, \quad \frac{dp_i}{dt} = -\frac{\partial H}{\partial q^i}$$

These are called **Hamilton's equations.** They provide an alternative formulation of mechanics.

Example 12. A particle moving on the line under a potential has $L = \frac{1}{2}m\dot{q}^2 - V(q)$ and $H = \frac{p^2}{2m} + V(q)$

It follows that

$$\frac{dq}{dt} = \frac{p}{m}, \quad \frac{dp}{dt} = -V'(q)$$

Clearly, these are equivalent to Newton's second law.

Example 13. For the simple pendulum

$$L = \frac{1}{2}\dot{\theta}^2 - [1 - \cos\theta], \quad H = \frac{p_\theta^2}{2} + 1 - \cos\theta$$

In terms of the variable $x = \frac{1}{k}\sin\frac{\theta}{2}$

$$L = 2k^2 \left[\frac{\dot{x}^2}{1 - k^2x^2} - x^2 \right]$$

It follows that

$$p = \frac{\partial L}{\partial \dot{x}} = 4k^2 \frac{\dot{x}}{1 - k^2x^2}, \quad H = 2k^2 \left[\left(1 - k^2x^2\right) p^2 + x^2 \right]$$

If we choose the parameter k such that $H = 2k^2$ the relation between p and x becomes

$$p^2 = \frac{1 - x^2}{1 - k^2x^2}.$$

This is another description of the elliptic curve, related rationally to the more standard one:

$$y = (1 - k^2x^2)\frac{p}{4k^2}, \quad y^2 = (1 - x^2)(1 - k^2x^2).$$

Example 14. A particle moving under a potential in three dimensions has

$$L = \frac{1}{2}m\dot{\mathbf{r}}^2 - V(\mathbf{r})$$

so that

$$H = \frac{\mathbf{p}^2}{2m} + V(\mathbf{r})$$

$$\dot{\mathbf{r}} = \frac{\mathbf{p}}{m}, \quad \dot{\mathbf{p}} = -\nabla V$$

For the Kepler problem

$$V(\mathbf{r}) = -\frac{GMm}{|\mathbf{r}|}$$

6.2 Vector Fields and Integral Curves

Suppose we have a space[1] M on which there is a co-ordinate system x^μ. The number of components of the co-ordinates is the **dimension** of the space. An infinitesimal change of the co-ordinate at each point defines a vector field

$$\delta x^\mu = V^\mu(x)$$

Given a vector field V and an initial point x_0, we can construct a curve on M : start at x_0 and move to the point $x_1^\mu = x_0^\mu + V^\mu(x_0)\Delta t$ at a short time Δt later. Then we move to $x_2^\mu = x_1^\mu + V^\mu(x_1)\Delta t$ at time $2\Delta t$. Repeating this, we will be at $x_n^\mu = x_{n-1}^\mu + V^\mu(x_{n-1})\Delta t$ at time $n\Delta t$. In the limit as $\Delta t \to 0$ the sequence of points x_n merge into a continuous curve which satisfies the ODE

$$\frac{dx^\mu}{dt} = V^\mu(x)$$

with initial condition $x^\mu(0) = x_0^\mu$. This is called the integral curve of the vector field V^μ. The infinitesimal change of a function under a small displacement along V is

$$Vf = V^\mu \partial_\mu f$$

In fact, in modern differential geometry, we regard a vector field as this differential operator

$$V = V^\mu \partial_\mu$$

[1] By "space" we mean "differential manifold".

6.2.1 The Algebraic Formulation of a Vector Field

The idea of a vector field has a more algebraic formulation that has turned out to be useful. The set of functions on the space M form a commutative algebra over the real number field: we can multiply them by constants (real numbers); and we can multiply two functions to get another:

$$fg(x) = f(x)g(x)$$

A **derivation** of a commutative algebra A is a linear map $V : A \to A$ that satisfies the **Leibnitz rule**

$$V(fg) = (Vf)g + fV(g)$$

We can add two derivations to get another. Also we can multiply them by real numbers to get another derivation. The set of derivations of a commutative algebra form a **module** over A; i.e., the left multiplication fV is also a derivation. In our case of functions on a space, each component of the vector field is multiplied by the scalar f.

In this case, a derivation is the same as a first order differential operator or **vector field:**

$$Vf = V^{\mu}\partial_{\mu}f$$

The coefficient of the dervative along each co-ordinate is the component of the vector field in the direction.

The product of two derivations is not in general a derivation: it does not satisfy the Leibnitz rule:

$$V\left(W(fg)\right) = V\left((Wf)g + fW(g)\right)$$

$$= V(W(f))g + 2(Wf)(Vg) + fV(W(g))$$

But the commutator of two derivations, defined as

$$[V, W] = VW - WV$$

is always another derivation:

$$V\left(W(fg)\right) - W\left(V(fg)\right) = V(W(f))g + fV(W(g)) - W(V(f))g - fW(V(g))$$

In terms of components,

$$[V, W]^{\mu} = V^{\nu}\partial_{\nu}W^{\mu} - W^{\nu}\partial_{\nu}V^{\mu}$$

We sum over repeated indices as is the convention in geometry. The commutator of vector fields satisfies the identities

$$[V, W] + [W, V] = 0$$

$$[[U, V], W] + [[V, W], U] + [[W, U], V] = 0$$

Any algebra that satsfies these identities is called a **Lie algebra.** The set of vector fields on a manifold is the basic example. Lie algebras describe infinitesimal transformations. Many of them are of great interst in physics, as they describe symmetries. More on this later.

Example 15. The vector field

$$V = \frac{\partial}{\partial x}$$

generates translations along x. Its integral curve is $x(t) = x_0 + t$.

Example 16. On the other hand,

$$W = x\frac{\partial}{\partial x}$$

generates scaling. That is, its integral curve is

$$x(u) = e^u x_0.$$

We see that these two vector fields have the commutator

$$[V, W] = V$$

Example 17. Infinitesimal rotations around the three Cartesian axes on R^3 are described by the vector fields

$$L_x = -y\frac{\partial}{\partial z} + z\frac{\partial}{\partial y}, \quad L_y = -z\frac{\partial}{\partial x} + x\frac{\partial}{\partial z}, \quad L_z = -x\frac{\partial}{\partial y} + y\frac{\partial}{\partial x}$$

They satisfy the commutation relations

$$[L_x, L_y] = L_z, \quad [L_y, L_z] = L_x, \quad [L_z, L_x] = L_y.$$

Given two vector fields V, W, we can imagine moving from x_0 along the integral curve of V for a time δ_1 and then along that of $W$for some time δ_2. Now suppose we reverse the order by first going along the integral curve of $W$for a time δ_2 and then along that of V for a time δ_2. The difference between the two endpoints is order $\delta_1\delta_2$, but is not in general zero. It is equal to the commutator:

$$[V^\nu \partial_\nu W^\mu - W^\nu \partial_\nu V^\mu]\,\delta_1\delta_2$$

Thus, the commutator of vector fields which we defined above algebraically, has a geometric meaning as the difference between moving along the integral curves in two different ways. Such an interplay of geometry and algebra enriches both fields. Usually geometry helps us imagine things better or relate the mathematics to physical situations. The algebra is more abstract and allows generalizations to new physical situations that were previously unimaginable. For example, the transition from classical to quantum mechanics involves non-commutative algebra. These days we are circling back and constructing a new kind of geometry, **non-commutative geometry**, which applies to quantum systems.

6.3 Phase Space

Being first order ODE, the solution for Hamilton's equations is determined once the value of (q^i, p_i) is known at one instant. The space M whose co-ordinates are (q^i, p_i) is called **phase space**. Each point of phase space determines a solution of Hamilton's equation, which we call the **orbit** through that point. Hamilton's equations tell us how a given point in phase space evolves under an infinitesimal time translation: they define a vector field on the phase space. By compounding such infinitesimal transformations, we can construct time evolution over finite time intervals: the orbit is the **integral curve** of Hamilton's vector field.

$$V_H = \frac{\partial H}{\partial p_i}\frac{\partial}{\partial q^i} - \frac{\partial H}{\partial q_i}\frac{\partial}{\partial p_i}.$$

Since the state of the system is completely specified by a point on the phase space, any physical observable must be a function $f(q,p)$; that is, a function $f: M \to R$ of position and momentum[2] . The hamiltonian itself is an example of an observable;perhaps the most important one.

We can work out an interesting formula for the total time derivative of an observable:

$$\frac{df}{dt} = \sum_i \left[\frac{dq^i}{dt}\frac{\partial f}{\partial q^i} + \frac{dp_i}{dt}\frac{\partial f}{\partial p_i}\right]$$

Using Hamilton's equations this becomes

$$\frac{df}{dt} = \sum_i \left[\frac{\partial H}{\partial p_i}\frac{\partial f}{\partial q^i} - \frac{\partial H}{\partial q^i}\frac{\partial f}{\partial p_i}\right]$$

Given any pair of observables, we define their **Poisson bracket** to be

$$\{g, f\} = \sum_i \left[\frac{\partial g}{\partial p_i}\frac{\partial f}{\partial q^i} - \frac{\partial g}{\partial q^i}\frac{\partial f}{\partial p_i}\right]$$

Thus

[2]Sometimes we would allow also an explicit dependence in time, but we ignore that possibility for the moment.

$$\frac{df}{dt} = \{H, f\}$$

A particular example is when f is one of the co-ordinates themselves:

$$\{H, q^i\} = \frac{\partial H}{\partial p_i}, \quad \{H, p_i\} = -\frac{\partial H}{\partial q^i}.$$

You may verify the **canonical relations**

$$\{p_i, p_j\} = 0 = \left\{q^i, q^j\right\}, \quad \left\{p_i, q^j\right\} = \delta_i^j$$

Here $\delta_i^j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$ is the **Kronecker symbol.**

Exercise 18. Show that the Poisson bracket satisfies the conditions for a Lie algebra:

$$\{f, g\} + \{g, f\} = 0, \quad \{\{f, g\}, h\} + \{\{g, h\}, f\} + \{\{h, f\}, g\} = 0$$

and in addition that

$$\{f, gh\} = \{f, g\} h + g \{f, h\}.$$

Together these relations define a **Poisson algebra**.

6.3.1 The non-commutative algebra of quantum observables

In quantum mechanics, observables are still represented as functions on the phase space. However the rule for multiplying observables is no longer the obvious one: it is a non-commuatative operation. Without explaining how it is derived, we can exhibit the formula for this quantum multiplication law in the case of one degree of freedom:

$$f \circ g = fg - \frac{i\hbar}{2}\left[\frac{\partial f}{\partial p}\frac{\partial g}{\partial q} - \frac{\partial f}{\partial q}\frac{\partial g}{\partial p}\right]$$

$$+\frac{1}{2}\left(\frac{i\hbar}{2}\right)^2\left[\frac{\partial^2 g}{\partial p^2}\frac{\partial^2 f}{\partial q^2} - \frac{\partial^2 g}{\partial q^2}\frac{\partial^2 f}{\partial p^2}\right] + \cdots$$

Or,

$$f \circ g = fg + \sum_{r=1}^{\infty}\frac{1}{r!}\left(-\frac{i\hbar}{2}\right)^r\left[\frac{\partial^{2r} g}{\partial p^r}\frac{\partial^r f}{\partial q^r} - \frac{\partial^r g}{\partial q^r}\frac{\partial^r f}{\partial p^r}\right]$$

This is an associative, but not commutative, multiplication: $(f \circ g) \circ h = f \circ (g \circ h)$. (It can be proved using a Fourier integral representation.) Note that the zeroth order term is the usual multiplication and that the Poisson bracket is the first correction.In particular, in quantum mechanics we have the Heisenberg commutation relations

$$[p, q] = -i\hbar.$$

where the **commutator** is defined as $[p, q] = p \circ q - q \circ p$.

Thus, Poisson algebras are approximations to non-commutative but associative algebras. Is there a non-commutative generalization of geometric ideas such as co-ordinates and vector fields? This is the subject of non-commutative geometry, being actively studied by mathematicians and physicists.This approach to quantization, which connects hamiltonian mechanics to Heisenberg's formulation of quantum mechanics, is called **deformation quantization**. Every formulation of classical mechanics has its counterpart in quantum mechanics; each such bridge between the two theories is convenient approach to certain problems. Deformation quantization allows us to discover not only non-commutative geometry but also new kinds of symmetries of classical and quantum systems where the rules for combining conserved quantities of isolated systems is non-commutative: **quantum groups.** This explained why certain systems that did not have any obvious symmetry could be solved by clever folks such as Bethe, Yang and Baxter. Once the principle is discovered, it allows solution of even more problems . But now we are entering deep waters.

6.4 Canonical Transformations

Suppose we make a change of variables

$$q^i \mapsto Q^i(q,p), \quad p_i \mapsto P_i(q,p)$$

in the phase space. What happens to the Poisson brackets of a pair of observables under this? Using the chain-rule of differentiation

$$\frac{\partial f}{\partial q^i} = \frac{\partial f}{\partial Q^j}\frac{\partial Q^j}{\partial q^i} + \frac{\partial f}{\partial P_j}\frac{\partial P_j}{\partial q^i}$$

$$\frac{\partial f}{\partial p_i} = \frac{\partial f}{\partial Q^j}\frac{\partial Q^j}{\partial p_i} + \frac{\partial f}{\partial P_j}\frac{\partial P_j}{\partial p_i}$$

Using this, and some elbow grease, you can show that

$$\{f,g\} = \left\{Q^i, Q^j\right\}\frac{\partial f}{\partial Q^i}\frac{\partial g}{\partial Q^j} + \{P_i, P_j\}\frac{\partial f}{\partial P_i}\frac{\partial g}{\partial P_j} + \left\{P_i, Q^j\right\}\left\{\frac{\partial f}{\partial P_i}\frac{\partial g}{\partial Q^j} - \frac{\partial f}{\partial Q^j}\frac{\partial g}{\partial P_i}\right\}$$

So if the new variables happen to satisfy the canonical relations as well:

$$\{P_i, P_j\} = 0 = \left\{Q^i, Q^j\right\}, \quad \left\{P_i, Q^j\right\} = \delta_i^j$$

the Poisson brackets are still given by a similar expression:

$$\{f, g\} = \sum_i \left[\frac{\partial f}{\partial P_i} \frac{\partial g}{\partial Q^i} - \frac{\partial f}{\partial Q^i} \frac{\partial g}{\partial P_i} \right]$$

Such transformations are called **canonical transformations;** they are quite useful in mechanics because they preserve the mathematical structure of mechanics. For example, Hamilton's equations remain true after a canonical transformation:

$$\frac{dQ^i}{dt} = \frac{\partial H}{\partial P_i}$$

$$\frac{dP_i}{dt} = -\frac{\partial H}{\partial Q_i}$$

Example 19. The case of one degree of freedom. The interchange of position and momentum variables is an example of a canonical transformation:

$$P = -q, \quad Q = p$$

Notice the sign.

Another example is the scaling

$$Q = \lambda q, \quad P = \frac{1}{\lambda} p$$

Notice the inverse powers. More generally, the condition for a transformation $(q, p) \mapsto (Q, P)$ to be canonical is that the area element $dqdp$ be transformed to $dQdP$. This is because in the case of one degree of freedom, the Poisson bracket happens to be the Jacobian determinant:

$$\{P, Q\} \equiv \frac{\partial P}{\partial p} \frac{\partial Q}{\partial q} - \frac{\partial Q}{\partial p} \frac{\partial P}{\partial q} = \det \begin{bmatrix} \frac{\partial Q}{\partial q} & \frac{\partial Q}{\partial p} \\ \frac{\partial P}{\partial q} & \frac{\partial P}{\partial p} \end{bmatrix}$$

For more degrees of freedom, is still true that the volume element in phase space in invariant, $\prod_i dq^i dp_i = \prod_i dQ^i dP_i$,under canonical transformations, a result known as Liouville's theorem. But the invariance of the phase space volume no longer guarantees that a transformation is canonical: the conditions for that are stronger.

6.5 Infinitesimal canonical transformations

The composition of two canonical transformations is also a canonical transformation. Sometimes we can break up a canonical transformation as the compoistion of infinitesimal transformations. For example, the transformation

$$Q = \lambda q, \quad P = \lambda^{-1} p$$

You can check that for every λ this is a canonical transformation. A scaling through λ_1 followed by another through λ_2 is equal to one by $\lambda_1\lambda_2$. Conversely, we can think of a scaling through $\lambda = e^{\theta}$ as made of up of a large number n of scalings, each through a small value $\frac{\theta}{n}$.For an infinitesimally small $\Delta\theta$,

$$, \quad \Delta Q = q\Delta\theta, \quad \Delta P = -p\Delta\theta$$

These infinitesimal changes of co-ordinates define a vector field

$$V = q\frac{\partial}{\partial q} - p\frac{\partial}{\partial p}$$

That is, the effect of an inifinitesimal rotation on an arbitrary observable is

$$Vf = q\frac{\partial f}{\partial q} - p\frac{\partial f}{\partial p}$$

Now, note that this can be written as

$$Vf = \{pq, f\}$$

This is a particular case of a more general fact: every infinitesimal canonical transformation can be thought of as the Poisson bracket with some function, called its **generating function.**

Let us write an infinitesimal canonical transformation in terms of its components

$$V = V^i\frac{\partial}{\partial q^i} + V_i\frac{\partial}{\partial p_i}$$

The position of the indices is chosen that V^i is the infinitesimal change in q^i and V_i the change in p_i.

$$q^i \mapsto Q^i = q^i + V^i\Delta\theta, \quad p_i \mapsto P_i = p_i + V_i\Delta\theta$$

for some infinitesimal parameter $\Delta\theta$. Let us calculate to first order in $\Delta\theta$:

$$\left\{Q^i, Q^j\right\} = \left(\left\{V^i, q^j\right\} + \left\{q^i, V^j\right\}\right)\Delta\theta$$

$$\{P_i, P_j\} = (\{V_i, p_j\} + \{p_i, V_j\})\,\Delta\theta$$

$$\left\{P_i, Q^j\right\} = \delta_i^j + \left(\left\{V_i, q^j\right\} + \left\{p_i, V^j\right\}\right)\Delta\theta$$

So the conditions for V to be an infinitesimal canonical transformation are

$$\left\{V^i, q^j\right\} + \left\{q^i, V^j\right\} = 0,$$

$$\{V_i, p_j\} + \{p_i, V_j\} = 0,$$

$$\left\{V_i, q^j\right\} + \left\{p_i, V^j\right\} = 0.$$

In terms of partial derivatives

$$\frac{\partial V^i}{\partial p_j} - \frac{\partial V^j}{\partial p_i} = 0$$

$$\frac{\partial V_i}{\partial q^j} - \frac{\partial V_j}{\partial q^j} = 0$$

$$\frac{\partial V_i}{\partial p_j} + \frac{\partial V^j}{\partial q^i} = 0 \tag{6.1}$$

The above conditions are satisfied if [3]

$$V^i = \frac{\partial G}{\partial p_i}, \quad V_i = -\frac{\partial G}{\partial q^i} \tag{6.2}$$

for some function G . The proof is a straightforward computation of second partial derivatives:

$$\frac{\partial V^i}{\partial p_j} - \frac{\partial V^j}{\partial p_i} = \frac{\partial^2 G}{\partial p_i \partial p_j} - \frac{\partial^2 G}{\partial p_i \partial p_j} = 0$$

etc.

Conversely, if (6.5.1) implies (6.5.2), we can produce the required function f as a line integral from the origin to the point (q, p) along some curve:

$$G(q,p) = \int_{(0,0)}^{(q,p)} \left[V_i \frac{dq^i}{ds} - V^i \frac{dp_i}{ds} \right] ds$$

In general such integrals will depend on the path taken, not just the endpoint. But the conditions (6.5.1) are exactly what is needed to ensure independence on the path.[4]

Exercise 20. Prove this by varying the path infinitesimally.

Thus an infinitesimal canonical transformation is the same as the Poisson bracket with some function, called its generator. By composing such infinitesimal transformations, we get a curve in the phase space:

$$\frac{dq^i}{d\theta} = \left\{G, q^i\right\}, \quad \frac{dp_i}{d\theta} = \left\{G, p_i\right\}$$

Now we see that Hamilton's equations are just a special case of this. Time evolution is a canonical transformation too, whose generator is the hamiltonian.

[3]There is an analogy with the condition that the curl of a vector field is zero; such a vector field would be the gradient of a scalar.

[4]We are assuming that any two curves connecting the origin to (q, p) can be deformed continuously into each other. In topology, the result we are using is called the Poincare lemma.

Every observable (i.e., function on phase space) generates its own canonical transformation.

Example 21. A momentum variable generates translations in its conjugate position variable.

Example 22. The generator of rotations is angular momentum along the axis of rotation. For a rotation around the $z$axis

$$x \mapsto \cos\theta x - \sin\theta y, \quad y \mapsto \sin\theta x + \cos\theta y$$

$$p_x \mapsto \cos\theta p_x + \sin\theta p_y, \quad p_y \mapsto -\sin\theta p_x + \cos\theta p_y$$

So we have

$$\frac{dx}{d\theta} = -y, \quad \frac{dy}{d\theta} = x$$

$$\frac{dp_x}{d\theta} = p_y, \quad \frac{dp_y}{d\theta} = -p_x$$

The generator is

$$L_z = xp_y - yp_x$$

6.6 Symmetries and Conservation Laws

Suppose that the Hamiltonian is independent of a certain co-ordinate q^i; then the corresponding momentum is conserved.

$$\frac{\partial H}{\partial q^i} = 0 \implies \frac{dp_i}{dt} = 0.$$

This is the beginning of a much deeper theorem of Noether that asserts that every continuous symmetry implies a conservation law. A symmetry is any canonical transformation of the variables $(q^i, p_i) \mapsto (Q^i, P_i)$ that leaves the hamiltonian unchanged:

$$\{P_i, P_j\} = 0 = \left\{Q^i, Q^j\right\}, \quad \left\{P_i, Q^j\right\} = \delta_i^j$$

$$H(Q(q,p), P(q,p)) = H(q,p)$$

A continuous symmetry is one that can be built up as a composition of infinitesimal transformations. We saw that every such canonical transformation is generated by some observable G. The change of any other observable f under this canonical transformation is given by

$$\{G, f\}$$

In particular the condition that the hamiltonian be unchanged is

$$\{G, H\} = 0.$$

But we saw earlier that the change of G under a time evolution is

$$\frac{dG}{dt} = \{H, G\}$$

So, the invariance of Hunder the canonical transformation generated by G is equivalent to the condition that $G$is conserved under time evolution.

$$\frac{dG}{dt} = 0 \iff \{G, H\} = 0.$$

Example 23. Let us return to the Kepler problem $H = \frac{\mathbf{p}^2}{2m} + V(\mathbf{r})$, where $V(\mathbf{r})$ is a function only of the distance $|\mathbf{r}|$. The components L_x, L_y, L_z of angular momentum

$$\mathbf{L} = \mathbf{r} \times \mathbf{p}$$

generate rotations around the axes x, y, z respectively. Since the hamiltonian is invariant under rotations

$$\{\mathbf{L}, H\} = 0$$

Thus the three components of angular momentum are conserved:

$$\frac{d\mathbf{L}}{dt} = 0.$$

This fact can also be verified directly as we did before.

Exercise 24. Show that the hamiltonian of the Kepler problem in spherical polar co-ordinates is

$$H = \frac{p_r^2}{2m} + \frac{L^2}{2mr^2} - \frac{k}{r}, \quad L^2 = p_\theta^2 + \frac{p_\phi^2}{\sin^2\theta}$$

Show that L is the magnitude of angular momentum and that $p_\phi = L_z$ is one of its components. Thus, show that $\left\{L^2, L_z\right\} = 0 = \left\{H, L^2\right\}$.

Chapter 7

Geodesics

A basic problem in geomety is to find the curve of shortest length that passes through two given points. Such curves are called **geodesics**. On the plane this is a straightline. But if we look at some other surface, such as the sphere the answer is more intricate. Gauss developed a general theory for geodesics on surfaces. Riemann then generalized it to higher dimensions. With the discovery of relativity it became clear that space and time are to be treated on the same footing. Einstein discovered that the Riemannian geometry of space-time provides a relativistic theory of gravitation. The theory of geodesics can be thought of a hamiltonian system, and ideas from mechanics are useful to understand properties of geodesics. Also, geometry is essential to understand the motion of particles in a gravitational field. In another direction, it turns out that the motion of even non-relativistic particles of a given energy in a potential can be understood as geodesics of a certain metric (Maupertuis metric). Thus no study of mechanics is complete without a theory of geodesics.

7.1 The Metric

Let x^i for $i = 1, \cdots n$ be the co-ordinates on some space. In Riemannian geometry, the distance ds between two nearby points x^i and $x^i + dx^i$ is postulated to be a quadratic form[1]

$$ds^2 = g_{ij}(x)dx^i dx^{\nu j}$$

For Cartesian co-ordinates in Euclidean space, g_{ij} are constants,

$$ds^2 = \sum_i \left[dx^i\right]^2, \quad g_{ij} = \delta_{ij}.$$

[1]It isa convention in geometry to place the indices on co-ordinates above, as superscripts. Repeated indices are summed over. Thus $g_{ij}(x)dx^i dx^j$ stands for $\sum_{ij} g_{ij}(x)dx^i dx^{\nu j}$. For this to make sense, you have to make sure that no index occurs more than twice in any factor.

The matrix $g_{ij}(x)$ is called the **metric.** The metric must be a symmetric matrix with an inverse. The inverse is denoted by g^{ij} , with superscripts. Thus

$$g^{ij}g_{jk} = \delta^i_k.$$

Although Riemann only allowed for positive metrics, we now know that the metric of space-time is not positive: along the time-like directions, ds^2 is positive and along space-like directions it is negative.

7.2 The Variational Principle

A curve is given parametrically by a set of functions $x^i(\tau)$ of some real parameter. The length of this curve will be

$$l[x] = \int \sqrt{g_{ij}(x)\frac{dx^i}{d\tau}\frac{dx^{\nu j}}{d\tau}}d\tau$$

This is the quantity to be minimized, if we are to get an equation for geodesics. But it is simpler (and turns out to be equvalent) to minimize instead the related quantity

$$S = \frac{1}{2}\int g_{ij}(x)\frac{dx^i}{d\tau}\frac{dx^j}{d\tau}d\tau$$

If we look at a finite sum instead of an integral, $\frac{1}{2}\sum x_a^2$ and $\sum |x_a|$ are minimized by the same choice of x_a. But x_a^2 is a much nicer quantity than $|x_a|$: for example, it is differentiable. Similarly, S is a more convenient quantity to differentiate.

7.2.1 Curves minimizing the action and the length are the same

This can be proved using a trick using Lagrange multipliers. First of all, we note that the length can be thought of the minimum of

$$S_1 = \frac{1}{2}\left[\int g_{ij}\frac{dx^i}{d\tau}\frac{dx^j}{d\tau}\lambda^{-1}d\tau + \int \lambda d\tau\right]$$

over all non-zero functions λ. Minimizing gives $\lambda^{-2}|\dot{x}|^2 = 1 \implies \lambda = |\dot{x}|$. At this minimum $S_1[x] = l[x]$. Now S_1 is invariant under changes of parameters

$$\tau \to \tau'(\tau), \lambda' = \lambda\frac{d\tau}{d\tau'}$$

Choosing this parameter to be the arc length, S_1 reduces to the action. Thus they describe equivalent variational problems. Moreover, at the minimum S, S_1, l all agree.

7.2.2 The Geodesic Equation

This leads to a differential equation

$$\frac{d}{d\tau}\left[g_{ij}\frac{dx^j}{d\tau}\right]-\frac{1}{2}\partial_i g_{jk}\frac{dx^j}{d\tau}\frac{dx^k}{d\tau}=0.$$

Straightforward application of the Euler-Lagrange equation

$$\frac{d}{d\tau}\frac{\partial L}{\partial \dot{x}^i}-\frac{\partial L}{\partial x^i}=0$$

with Lagrangian

$$L=\frac{1}{2}g_{jk}\dot{x}^j\dot{x}^k$$

$$\frac{\partial L}{\partial \dot{x}^i}=g_{ij}\dot{x}^j$$

An equivalent formulation is

$$\frac{d^2x^i}{d\tau^2}+\Gamma^i_{jk}\frac{dx^j}{d\tau}\frac{dx^k}{d\tau}=0, \quad \Gamma^i_{jk}=\frac{1}{2}g^{il}\left[\partial_j g_{kl}+\partial_k g_{jl}-\partial_l g_{jk}\right]$$

The Γ^i_{jk} are called Christoffel symbols. Calculating them for some given metric is one of the joys of Riemannian geometry; an even greater joy is to get someone else to do the calculation for you.

Proposition 25. *Given an initial point P and a vector V at that point, there is a geodesic that starts at P with V as its tangent*

This just follows from standard theorems about the local existence of solutions of ODEs. The behavior for large τ can be complicated: geodesics are chaotic except for metrics with a high degree of symmetry.

Remark 26. The following are more advanced points that you will understand only during a second reading , or after you have already learned some Riemannian geometry.

Proposition 27. *On a connected manifold, any pair of points are connected by at least one geodesic*

Connected means that there is a continuous curve connecting any pair of points (to define these ideas precisely we wil need first a definition of a manifold, which we will postpone for the moment). Typically there are several geodesics connecting a pair of points, for example on the sphere there are at least two for every (unequal) pair of points: one direct route and that goes around the world.

Proposition 28. *The shortest length of all the geodesics connecting a pair of points is the distance between them*

It is a deep result that such a minimizing geodesic exists. Most geodesics are extrema.

Gauss and Riemann realized that only experiments can determine whether space is Euclidean. They even commissioned an experinent to look for departures from Euclidean geometry; and found none. The correct idea turned out to be to include time as well.

7.3 The Sphere

The geometry of the sphere was studied by the ancients There were two spheres of interest to astronomers: the surface of the Earth and the celestial sphere, upon which we see the stars. Eratosthenes (3rd century BC) is said to have invented the use of the latitude and longitude as co-ordinates on the sphere. The (6th century AD) Sanskrit treatise *Aryabhatiya*, uses this co-ordinate system for the sphere as well (with the city of Ujjaini on the prime meridian) in solving several problems of spherical geometry. Predicting sunrise and sunset times, eclipses, calculating time based on the length of the shadow of a rod, making tables of positions of stars, are all intricate geometric problems.

The metric of a sphere S^2 in polar co-ordinates is

$$ds^2 = d\theta^2 + \sin^2\theta d\phi^2$$

We just have to hold r constant in the expression for distance in R^3 in polar co-ordinates. The sphere was the first example of a curved space.There are no straightlines on a sphere: any straightline of R^3 starting at a point in S^2 will leave it. There are other subspaces of R^3 such as the cylinder or the cone which contain some straightlines. The question arises: what is the shortest line that connects two points on the sphere? Such questions were of much interest to map makers of the nineteenth century, an era when the whole globe was being explored. In the mid nineteenth century Gauss took up the study of the geometry of distances on curved surfaces metrics which was later generalized by Riemann to higher dimensions. Einstein realized a variant of Riemannian geometry, allowing for $ds^2$to be negative or zero as well, is the basis of a relativistic theory of gravity. For technical reasons, we will study a slightly different function than the length of a curve.

The action of a curve on the sphere is defined to be

$$S = \frac{1}{2}\int \left[\dot\theta^2 + \sin^2\theta\dot\phi^2\right] d\tau$$

Note that this is not quite the same thing as the length of the curve:

$$l = \int \left[\dot\theta^2 + \sin^2\theta\dot\phi^2\right]^{\frac{1}{2}} d\tau$$

It turns out that S is a simpler function on the space of curves than l . This is similar to the fact that x^2 is a differentiable function while $|x|$ is not.(Its derivative has a jump discontinuity at the origin.) But the same curves minimize S and l . (Again, both x^2 and $|x|$ are minimized at $x = 0$.)

Remark 29. Some mathematicians, making a confused analogy with mechanics, call S the 'energy' of the curve instead of its action.

This definition of a geodesic does not require it to be a minimum of the action or of distance: in fact many interesting geodesics are saddle points of S . The Euler-Lagrange equations of this variational principle give

$$\delta S = \int \left[\dot\theta\delta\dot\theta + \sin\theta\cos\theta\dot\phi^2\delta\theta + \sin^2\theta\dot\phi\delta\dot\phi\right] d\tau$$

$$-\ddot\theta + \sin\theta\cos\theta\dot\phi^2 = 0$$

$$\frac{d}{d\tau}\left[\sin^2\theta\dot\phi\right] = 0$$

The key to solving any system of ODEs is to identify conserved quantities. The obvious conserved quantity is

$$L = \sin^2\theta\dot\phi$$

The solution is simplest when $L = 0$. For these geodesics, $\phi$is a constant. Then θ is a linear function of τ. These are the lines of meridian of constant longitude. They are also called great circles. Geometrically they are the intersection of a plane passing through the center of the circle with the circle itself.

Proposition 30. *Any pair of points on the circle lie on such a great circle. Thus geodesics are the same as arcs of great circles.*

Using the symmetry of the sphere under rotations, we can always choose a co-ordinate system such that the two points lie along a longitude. So we don't actually have to solve the differential equations to see this fact. But if we have to find the equation of a geodesic with a given choice of axes,

It is possible to solve the equations for any value of L

$$-\ddot\theta + \frac{\cos\theta L^2}{\sin^3\theta} = 0$$

Multiply by $\dot\theta$ and integrate once to get

$$\frac{1}{2}\dot\theta^2 + \frac{L^2}{2\sin^2\theta} = E$$

another constant of motion. Solving

$$\dot\theta = \sqrt{2E - \frac{L^2}{\sin^2\theta}}$$

$$\tau = \int \frac{d\theta}{\sqrt{2E - \frac{L^2}{\sin^2\theta}}}$$

which can be evaluated in terms of trig functions.

Corollary 31. *The equator is a geodesic.*

Corollary 32. *We can form a triangle with geodesics as sides, all of whose interior angles are right angles*

Start at the North Pole; go down to the equator along a meridian; go along the equator for a quarter of the circumference; then move along the meridian back to the North Pole.

In Euclidean geometry, the sum of the interior angles must be π. In spherical geometry, it depends on the area enclosed by the sides.A small geodesic triangle will have angles adding up to π as in Euclidean geometry. For small distances, geodesics appear to be straightlines and the sphere looks flat. This is why people thought the Earth was flat in olden days.

Gauss found the correct measure of the curvature of a surface whose metric is given

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu.$$

7.3.1 The sphere can also be identified with the complex plane, with the point at infinity added

Identify the complex plane with the tangent plane to the sphere at the South plane. Given a point on the sphere, we can draw a straight line in R^3 that connects the North pole to that line: continuing that line, we get a point on the complex plane. This is the co-ordinate of the point. Thus the South pole is at the origin and the North point corresponds to infinity.

The metric of S^2 is

$$ds^2 = 4\frac{d\bar{z}dz}{(1+\bar{z}z)^2}, \quad z = \tan\frac{\theta}{2}e^{i\phi}$$

The isometries of the sphere are fractional linear transformations by $SU(2)$

$$z \mapsto \frac{az+b}{cz+d}, \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} \bar{a} & \bar{c} \\ \bar{b} & \bar{d} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Problem 33. Verify by direct calculations that these leave the metric unchanged.

This is one way of seeing that $SU(2)/\{1,-1\}$ is the group of rotations.

7.4 Hyperbolic Space

The metric of hyperbolic geometry is

$$ds^2 = d\theta^2 + \sinh^2\theta d\phi^2$$

It describes a space of negative curvature. What this means is that two geodesics that start at the same point in slightly different directions will move apart at a rate faster than in Euclidean space. On a sphere, they move apart slower than in Euclidean space so it has positive curvature. Just as the sphere is the set of points at a unite distance from the center,

Proposition 34. *The hyperboloid is the set of unit time-like vectors in Minkowski geometry $R^{1.2}$*

There is the co-ordinate system analogous to the spherical polar co-ordinate system valid in the time-like interior of the light cone:

$$(x^0)^2-(x^1)^2-(x^2)^2=\tau, \quad x^0=\tau\cosh\theta,\ x^1=\tau\sinh\theta\cos\phi,\ x^2=\tau\sinh\theta\sin\phi$$

The Minkowski metric becomes

$$ds^2 = d\tau^2 - \tau^2\left[d\theta^2 + \sinh^2\theta d\phi^2\right]$$

Thus the metric induced on the unit hyperboloid

$$(x^0)^2-(x^1)^2-(x^2)^2=\tau,$$

is the one above.

Proposition 35. *The hyperboloid can also be thought of as the upper half plane with the metric*

$$ds^2 = \frac{dx^2+dy^2}{y^2}, \quad y>0$$

Proposition 36. *The isometries are fractional linear transformations with real parameters a, b, c, d:*

$$z \mapsto \frac{az+b}{cz+d}, \quad ad-bc=1$$

Problem 37. Verify that these are symmetries of the metric.

Proposition 38. *The geodesics are circles orthogonal to the real line.*

If two points have the same value of x, the geodesic is just the line parallel to the imaginary axis that contains them. Using the isometry above we can bring any pair of points to this configuration. It is also possible to solve the geodesic equations to see this fact.

Proposition 39. *The hyperboloid can also be thought of as the interior of the unit disk*

$$ds^2 = \frac{dzd\bar{z}}{(1-\bar{z}z)^2}, \quad \bar{z}z < 1$$

Problem 40. What are the geodesics in this description?

7.5 Hamiltonian Formulation of Geodesics

The analogy with mechanics is clear in the variational principle of geometry. The Lagrangian

$$L = \frac{1}{2}g_{ij}\frac{dx^i}{d\tau}\frac{dx^j}{d\tau}$$

leads to the "momenta"

$$p_i = g_{ij}\frac{dx^j}{d\tau}$$

The hamiltonian is

$$H = p_i\frac{dx^i}{d\tau} - L$$

$$= \frac{1}{2}g^{ij}p_i p_j.$$

Thus H has the physical meaning of half the square of the mass of the particle.

It follows that the geodesic equations can be written as

$$p_i = g_{ij}\frac{dx^j}{d\tau}, \quad \frac{d}{d\tau}p_i = \frac{1}{2}[\partial_i g^{jk}]p_j p_k$$

It is obvious from mechanics that if the metric happens to be independent of a co-ordinate, its conjugate momentum is conserved. This can be used to solve equations for a geodesic on spaces like the sphere which have such symmetries.

A better way point of view is to use the Hamilton-Jacobi equation, a first order PDE. When this equation is separable, the geodesics can be determined explicitlty.

7.6 Geodesic Formulation of Newtonian Mechanics

In the other direction, there is a way to think of the motion of a particle in a potential with a fixed energy as geodesics. Suppose we have a particle (or collection of particles) whose kinetic energy is given by a quadratic function of co-ordinates

$$T = \frac{1}{2} m_{ij}(q) \frac{dq^i}{dt} \frac{dq^j}{dt}$$

For example, in the three body problem of celestial mechanics, i, j takes values 1 through 9: the first three for the position of the first particle and so on. Then the metric is a constant matrix whose diagonal entries are the masses:

$$m_{ij} = \begin{bmatrix} m_1 1_3 & 0 & 0 \\ 0 & m_2 1_3 & 0 \\ 0 & 0 & m_3 1_3 \end{bmatrix}$$

The first 3×3 block gives the kinetic energy of the first particle, the next that of the second particle and so on.

If the potential energy is $V(x)$ we have the condition for the conservation of energy

$$\frac{1}{2} m_{ij}(q) \frac{dq^i}{dt} \frac{dq^j}{dt} + V(q) = E$$

If we only consider paths of a given energy E, Hamilton's principle takes the form of minimizing

$$S = \int_{t_1}^{t_2} p_i \frac{dq^i}{dt} dt$$

since $\int H dt = E[t_2 - t_1]$ is constant. Solving for p_i in terms of $\dot{q}^i$ this becomes

$$S = \int [E - V(q)] m_{ij}(q) \frac{dq^i}{ds} \frac{dq^j}{ds} ds$$

where the parameter ds is defined by

$$\frac{ds}{dt} = [E - V(q)]$$

This can be thought of as the variational principle for geodesics of the metric

$$g_{ij} = 2[E - V(q)] m_{ij}(q) dq^i dq^j$$

Of course, this only makes sense in the region of space with $E > V(q)$: that is the only part that is accessible to a particle of total energy E. This version of the variational principle is older than Hamilton's and is due to Euler who was building on ideas of Fermat and Maupartius in ray optics. Nowadays it is known as the Maupertuis principle.

7.6.1 Keplerian Orbits As Geodesics

Consider the planar Kepler problem with Hamiltonian

$$H = \frac{p_r^2}{2} + \frac{p_\phi^2}{2r^2} - \frac{k}{r}$$

The orbits of this system can be thought of as geodesic of the metric

$$ds^2 = 2\left[E + \frac{k}{r}\right]\left[dr^2 + r^2 d\phi^2\right]$$

There is no singularity in this metric at the collision point $r = 0$: it can be removed ("regularized") by transforming to the co-ordinates ρ, χ:

$$r = \rho^2, \quad \theta = 2\chi, \implies ds^2 = ds^2 = 8\left[E\rho^2 + k\right]\left[d\rho^2 + \rho^2 d\chi^2\right]$$

This is just what we would have found for the harmonic oscillator (for $E < 0$): the Kepler problem can be transformed by a change of variables to the harmonic oscillator.

When $E = 0$ (parabolic orbits) this is just the flat metric on the plane: parabola are mapped to straightlines by the above change of variables. For $E > 0$ (hyperbolic orbits) we get a metric of negative curvature and for $E < 0$ (elliptic orbits) one of positive curvature. These curvatures are not constant, however.

7.7 Geodesics in General Relativity

By far the most important application of Riemannian geometry to physics is General Relativity, Einstein's theory of gravitation. The gravitational field is described by the metric tensor of space-time.

The path of a particle is given by the geodesics of this metric. Of special importance is the metric of a spehrically symmetric mass distribution, called the Schwarschild metric.

$$ds^2 = \left(1 - \frac{r_s}{r}\right)dt^2 - \frac{dr^2}{1 - \frac{r_s}{r}} - r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right)$$

The parameter r_s is proportional to the mass of the source of the gravitational field. For the Sun it is about 1 km. To solve any mechanical problem we must exploit conservation laws. Often symmetries provide clues to these conservation laws.

A time-like geodesic satisfies

$$\left(1 - \frac{r_s}{r}\right)\dot{t}^2 - \frac{\dot{r}^2}{1 - \frac{r_s}{r}} - r^2\left(\dot{\theta}^2 + \sin^2\theta\dot{\phi}^2\right) = H$$

Here the dot denotes derivatives w.r.t. τ. The constant H has to be positive; it can be chosen to be one by a choice of units of τ.

Proposition 41. *Translations in t and rotations are symmetries of the Schwarschild metric*

The angular dependence is the same as for the Minkwoski metric. The invariance under translations in t is obvious

Corollary 42. *Thus the energy and angular momentum of a particle moving in this gravitational field are conserved*

The translation in t gives the conservation of energy per unit mass

$$E = p_t = \left(1 - \frac{r_s}{r}\right)\dot{t}$$

We can choose co-ordinates such that the geodesic lies in the plane $\theta = \frac{\pi}{2}$. By looking at the second component of the geodesic equation

$$\frac{d}{d\tau}\left[r^2 \frac{d\theta}{d\tau}\right] = r^2 \sin\theta\cos\theta\left[\frac{d\phi}{d\tau}\right]^2$$

We can rotate the co-ordinate system so that any plane passing through the center corresponds to $\theta = \frac{\pi}{2}$.The conservation of angular momentum, which is a 3-vector, implies also that the orbit lies in the plane normal to it.We are simply choosing the $z-$ axis to point along the angular momentum.Thus

$$\left(1 - \frac{r_s}{r}\right)\dot{t}^2 - \frac{\dot{r}^2}{1 - \frac{r_s}{r}} - r^2\dot{\phi}^2 = H$$

Rotations in ϕ lead to the conservation of the third component of angular momentum per unit mass

$$L = -p_\phi = r^2\dot{\phi}.$$

This is an analogue of Kepler's law of areas. To determine the shape of the orbit we must detemine r as a function of ϕ

In the Newtonian limit these are conic sections: ellipse, parabola or hyperbola. Let $u = \frac{r_s}{r}$ .Then

$$\dot{r} = r'\dot{\phi} = \frac{r'}{r^2}L = -lu'.$$

Here prime denotes derivative w.r.t. ϕ. Also $l = \frac{L}{r_s}$. So ,

$$\frac{E^2}{1-u} - \frac{l^2u'^2}{1-u} - l^2u^2 = H$$

We get an ODE for the orbit

$$l^2u'^2 = E^2 + H(u-1) - l^2u^2 + l^2u^3$$

This is the Weierstrass equation, solved by the elliptic integral . Since we are interested in the case where the last term (which is the GR correction) is small a different strategy is more convenient. Differentiate the equation to eliminate the constants:

$$u'' + u = \frac{H}{2l^2} + \frac{3}{2}u^2$$

Proposition 43. *In the Newtonian approximation the orbit is periodic.*

The Newtonian approximation is

$$u_0'' + u_0 = \frac{H}{2l^2} \implies$$

$$u_0 = \frac{H}{2l^2} + B\sin\phi$$

for some constant of integration B. Recall the equation for an ellipse in polar co-ordinates

$$\frac{1}{r} = \frac{1}{b} + \frac{\epsilon}{b}\sin\phi$$

Here, ϵ is the eccentricity of the ellipse: if it is zero the equation is that of a circle of radius b. In general b is the semi-latus rectum of the ellipse. If $1 > \epsilon > 0$, the closest and farthest approach to the origin are at $\frac{1}{r_{1,2}} = \frac{1}{b} \pm \frac{\epsilon}{b}$ so that the major axis is $r_2 + r_1 = \frac{2b}{1-\epsilon^2}$. So now we know the meaning of l and B in terms of the Newtonian orbital parameters.

$$b = 2r_s l^2, \quad B = \frac{\epsilon}{b}r_s$$

7.7.1 The Perihelion Shift

Putting

$$u = u_0 + u_1$$

to first order (we choose units with $H = 1$ for convenience)

$$u_1'' + u_1 = \frac{3}{2}u_0^2$$

$$= \frac{3}{8l^4} + \frac{3B}{2l^2}\sin\phi + \frac{3}{2}B^2\sin^2\phi$$

$$u_1'' + u_1 = \frac{3}{8l^4} + \frac{3}{4}B^2 + 3\frac{B}{2l^2}\sin\phi - \frac{3}{4}B^2\cos 2\phi$$

Although the driving terms are periodic, the solution is not periodic,because of the resonant term $\sin\phi$ in the r.h.s.

$$u_1 = \text{periodic} + \text{constant}\phi\sin\phi$$

Proposition 44. *In GR the orbit is not closed.*

Thus GR predicts that as a planet returns to the perihelion its angle has suffered a net shift. After rewriting B, l, r_s,in terms of the parameters a, ϵ, T of the orbit, the perihelion shift is found to be

$$\frac{24\pi^2 a^2}{(1-\epsilon^2)c^2T^2}$$

where a is the semi-major axis and T is the period of the orbit.

Exercise 45. Express the period of $u(\phi)$ in terms of a complete elliptic integral and hence the Arithmetic Geometric Mean. Use this to get the perihelion shift in terms of the AGM.

This perihelion shift agrees with the measured anomaly in the orbit of Mercury

At the time Einstein proposed his theory, such a shift in the perihelion of Mercury was already known-and unexplained- for a hundred years! The prediction of GR, $43''$ of arc per century, exactly agreed with the observation: its first experimental test. For the Earth the shift of the perihelion is even smaller: $3.8''$ of arc per century. Much greater accuracy has been possible in determining the orbit of the Moon through laser ranging. The resuls are a quantitative vindication of GR to high precision.

Chapter 8

Hamilton-Jacobi Theory

We saw that the formulation of classical mechanics in terms of Poisson brackets allows a passage into quantum mechanics: the Poisson bracket measures the infinitesimal departure from commutativity of observables. There is also a formulation of mechanics that is connected to the Schrodinger form of quantum mechanics. Hamiltonon discovered this originally through the analogy with optics. In the limit of small wavelength, the wave equation (which is a second order linear PDE) becomes a first order (but nonlinear) equation, called the eikonal equation. Hamilton and Jacobi found an analogous point of in mechanics. In modern language, it is the short wavelength limit of Schrodinger's wave equation.

Apart fom its conceptual value in connection with quantum mechanics, the Hamilton-Jacobi equation also provides powerful technical tools for solving problems of classical mechanics.

8.1 Conjugate Variables

Recall that we got the Euler-Lagrange equations by minimizing the action

$$S = \int_{t_1}^{t_2} L(q, \dot{q}) dt$$

over paths with fixed endpoints. It is interesting also to hold the initial point fixed and ask how the action varies as a function of the endpoint. Let us change notation slightly and call the end time t , and the variable of integration τ . Also let us call $q(t) = q$, the ending position.

$$S(t, q) = \int_{t_1}^{t} L(q(\tau), \dot{q}(\tau) d\tau$$

From the definition of the integral, we see that

$$\frac{dS}{dt} = L$$

But,

$$\frac{dS}{dt} = \frac{\partial S}{\partial t} + \frac{\partial S}{\partial q^i}\dot{q}^i$$

so that

$$\frac{\partial S}{\partial t} = L - \frac{\partial S}{\partial q^i}\dot{q}^i$$

If we vary the path

$$\delta S = \left[\frac{\partial L}{\partial \dot{q}^i}\delta q^i\right]_{t_1}^{t} - \int_{t_1}^{t}\left[\frac{\partial L}{\partial q^i} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i}\right]dt$$

In deriving the E-L equations we could ignore the first term because th evariation vanished at the endpoints. But looking at the dependence on the ending position, and recalling that $\frac{\partial L}{\partial \dot{q}^i} = p_i$, we get

$$\frac{\partial S}{\partial q^i} = p_i$$

Thus,

$$\frac{\partial S}{\partial t} = L - p_i\dot{q}^i$$

In other words

$$\frac{\partial S}{\partial t} = -H.$$

So we see that the final values of the variables conjugate to t, q^i are given by the derivatives of S.

8.2 The Hamilton-Jacobi Equation

This allows us rewrite content of the action principle as a partial differential equation: we replace p_i by $\frac{\partial S}{\partial q^i}$ in the hamiltonian to get

$$\frac{\partial S}{\partial t} + H\left(q, \frac{\partial S}{\partial q}\right) = 0.$$

Example 46. For the free particle $H = \frac{p^2}{2m}$ and

$$\frac{\partial S}{\partial t} + \frac{1}{2m}\left(\frac{\partial S}{\partial q}\right)^2 = 0$$

A solution to this equation is

$$S(t, q) = -Et + pq$$

for a pair of constants E, p satisfying

$$E = \frac{p^2}{2m}.$$

Thus, the solution to the H-J equation in this case is a sum of terms each depending only one of the variables: it is **separable**. Whenever the H-J equation can be solved by such a separation of variables, we can decompose the motion into one dimensional motions, each of which can be solved separately.

Example 47. The planar Kepler Problem can be solved by separation of variables as well. In polar co-ordinates

$$H = \frac{p_r^2}{2m} + \frac{p_\phi^2}{2mr^2} - \frac{k}{r}$$

so that the H-J equation is

$$\frac{\partial S}{\partial t} + \frac{1}{2m}\left[\frac{\partial S}{\partial r}\right]^2 + \frac{1}{2mr^2}\left[\frac{\partial S}{\partial \phi}\right]^2 - \frac{k}{r} = 0$$

Since t, ϕ do not appear explicitly (i.e., they are **cyclic** variables), their conjugates can be assumed to be constants. So we make the ansatz

$$S(t, r, \theta, \phi) = -Et + R(r) + L\phi$$

yielding

$$\frac{1}{2m}\left[\frac{dR}{dr}\right]^2 + \frac{L^2}{2mr^2} - \frac{k}{r} = E$$

Exercise 48. Show that the H-J equation can be solved by separation of variables

$$S(t, r, \theta, \phi) = T(t) + R(r) + \Theta(\theta) + \Phi(\phi)$$

in spherical polar co-ordinates for any potential of the form $V(r, \theta, \phi) = a(r) + \frac{b(\theta)}{r^2} + \frac{c(\phi)}{r^2 \sin^2\theta}$. The Kepler problem is a special case of this.

8.3 The Euler Problem

Euler solved many problems in mechanics. One of them was the motion of a body under the influence of the gravitational field of two fixed bodies. This does not occur in astronomy, as the two bodies will themselves have to move under each others gravitational field. But centuries later, exactly this problem occured in studying the molecular ion H_2^+: an electron orbiting two protons at fixed positions. Heisenberg dusted off Euler's old method and solved its Schrodinger equation: the only exact solution of a molecule.

The trick is to use a generalization of polar co-ordinates, in which the curves of constant radii are ellipses instead of circles. Place the two fixed masses at points $\pm\sigma$ along the $z-$ axis. If r_1 and r_2 are distances of a point from these points, the potential is

$$V = \frac{\alpha_1}{r_1} + \frac{\alpha_2}{r_2}.$$

with

$$r_{1,2} = \sqrt{(z \mp \sigma)^2 + x^2 + y^2}$$

We can use $\xi = \frac{r_1+r_2}{2\sigma}, \eta = \frac{r_2-r_1}{2\sigma}$ as co-ordinates.

Exercise 49. Note that $|\xi| \geq 1$ while $|\eta| \leq 1$. What are the surfaces where ξ is a constant and where η is a constant?

As the third co-ordinate we can use the angle ϕ of the cylindrical polar system:

$$x = \sigma\sqrt{(\xi^2-1)(1-\eta^2)}\cos\phi, \quad y = \sigma\sqrt{(\xi^2-1)(1-\eta^2)}\sin\phi, \quad z = \sigma\xi\eta$$

This is an orthogonal co-ordinate system; i.e., the metric is diagonal:

$$ds^2 = \sigma^2(\xi^2-\eta^2)\left[\frac{d\xi^2}{\xi^2-1} + \frac{d\eta^2}{1-\eta^2}\right] + \sigma^2(\xi^2-1)(1-\eta^2)d\phi^2$$

Exercise 50. Prove this form of the metric. It is useful to start with the metric in cylindrical polar cordinates $ds^2 = d\rho^2 + \rho^2 d\phi^2 + dz^2$ and make the change of variables $\rho = \sigma\sqrt{(\xi^2-1)(1-\eta^2)}$ and z as above.

Now the Lagrangian is

$$L = \frac{1}{2}m\sigma^2(\xi^2-\eta^2)\left[\frac{\dot{\xi}^2}{\xi^2-1} + \frac{\dot{\eta}^2}{1-\eta^2}\right] + \frac{1}{2}m\sigma^2(\xi^2-1)(1-\eta^2)\dot{\phi}^2 - V(\xi,\eta)$$

leading to the hamiltonian

$$H = \frac{1}{2m\sigma^2(\xi^2-\eta^2)}\left[(\xi^2-1)p_\xi^2 + (1-\eta^2)p_\eta^2 + \left(\frac{1}{\xi^2-1} + \frac{1}{1-\eta^2}\right)p_\phi^2\right] + V(\xi,\eta)$$

and the H-J equation

$$E = \frac{1}{2m\sigma^2(\xi^2-\eta^2)}\left[(\xi^2-1)\left(\frac{\partial S}{\partial\xi}\right)^2 + (1-\eta^2)\left(\frac{\partial S}{\partial\eta}\right)^2 + \left(\frac{1}{\xi^2-1} + \frac{1}{1-\eta^2}\right)\left(\frac{\partial S}{\partial\theta}\right)^2\right] + V(\xi,\eta)$$

The potential can be written as

$$V(\xi,\eta) = -\frac{1}{\sigma}\left\{\frac{\alpha_1}{\xi-\eta} + \frac{\alpha_2}{\xi+\eta}\right\} = \frac{1}{\sigma(\xi^2-\eta^2)}\left\{(\alpha_1+\alpha_2)\xi + (\alpha_1-\alpha_2)\eta\right\}$$

Since ϕ is cyclic we set $\frac{\partial S}{\partial \phi} = L$ a constant: it is the angular momentum around the axis connecting the two fixed bodies. The H-J equation becomes

$$2m\sigma^2(\xi^2-\eta^2)E = (\xi^2-1)\left(\frac{\partial S}{\partial \xi}\right)^2 + 2m\sigma(\alpha_1+\alpha_2)\xi + \frac{L^2}{\xi^2-1} + (1-\eta^2)\left(\frac{\partial S}{\partial \eta}\right)^2 + 2m\sigma(\alpha_1-\alpha_2)\eta + \frac{L^2}{1-\eta^2}$$

or

$$\left\{(\xi^2-1)\left(\frac{\partial S}{\partial \xi}\right)^2 + 2m\sigma(\alpha_1+\alpha_2)\xi + \frac{L^2}{\xi^2-1} + 2mE\sigma^2(\xi^2-1)\right\} +$$
$$\left\{(1-\eta^2)\left(\frac{\partial S}{\partial \eta}\right)^2 + 2m\sigma(\alpha_1-\alpha_2)\eta + \frac{L^2}{1-\eta^2} + 2mE\sigma^2(1-\eta^2)\right\} = 0.$$

This suggests the separation of variables

$$S = A(\xi) + B(\eta)$$

where each satisfies the ODE

$$(\xi^2-1)A'^2 + 2m\sigma(\alpha_1+\alpha_2)\xi + \frac{L^2}{\xi^2-1} + 2mE\sigma^2(\xi^2-1) = K$$

$$(1-\eta^2)B'^2 + 2m\sigma(\alpha_1-\alpha_2)\eta + \frac{L^2}{1-\eta^2} + 2mE\sigma^2(1-\eta^2) = -K$$

The solutions are elliptic integrals.

8.4 The Classical Limit of the Schrodinger Equation

Recall that the Schrodinger equation of a particle in a potential is

$$-\frac{\hbar^2}{2m}\nabla^2\psi + V\psi = i\hbar\frac{\partial \psi}{\partial t}$$

In the limit of small $\hbar$ (i.e., when quantum effects are small) this reduces to the H-J equation. The idea is to make the change of variables

$$\psi = e^{\frac{i}{\hbar}S}$$

so that the the equation becomes

$$-\frac{i\hbar}{2m}\nabla^2 S + \frac{1}{2m}(\nabla S)^2 + V + \frac{\partial S}{\partial t} = 0$$

If we ignore the first term we get the H-J equation.

Co-ordinate systems and potentials in which the H-J is separable also allow the solution of the Schrodinger equation by separation of variables. A complete list is given in Landau-Lifshitz Vols one and three.

8.5 Hamilton-Jacobi Equation in Riemannian manifolds

Given any metric $ds^2 = g_{ij}dq^i dq^j$ in configuration space we have the Lagrangian

$$L = \frac{1}{2}mg_{ij}dq^i dq^j - V(q)$$

The momenta are

$$p_i = mg_{ij}\dot{q}^j$$

and the hamiltonian is

$$H = \frac{1}{2}g^{ij}p_i p_j + V(q)$$

The Hamilton-Jacobi equation becomes, for a given energy

$$\frac{1}{2m}g^{ij}\frac{\partial S}{\partial q^i}\frac{\partial S}{\partial q^j} + V = E$$

If the metric is diagonal, ("orthogonal co-ordinate system") the inverse is easier to calculate.

In the absence of a potential this becomes

$$g^{ij}\frac{\partial S}{\partial q^i}\frac{\partial S}{\partial q^j} = \text{constant}$$

which is the H-J version of the geodesic equation.

Even when there is a potential, we can rewrite this as

$$\tilde{g}^{ij}\frac{\partial S}{\partial q^i}\frac{\partial S}{\partial q^j} = 1, \quad \tilde{g}^{ij} = \frac{1}{2m[E - V(q)]}.$$

Thus the motion of a particle in a potential can be thought of as geodesic motion in an effective metric

$$d\tilde{s}^2 = 2m[E - V(q)]g_{ij}dq^i dq^j$$

This is related to the Maupertuis principle we discussed earlier. Note that only the classically allowed region $E > V(q)$ is accessible to these geodesics.

Chapter 9

Integrable Systems

In quantum mechanics with a finite dimensional number of states, the hamiltonian is a hermitian matrix. It can be diagonalized by a unitary transformation. The analogue in classical physics is a canonical transformation that brings the hamiltonian to **normal form**: so that it depends only on variables P_i that commute with each other. In this form hamilton's equations are trivial to solve:

$$\frac{dQ^i}{dt} = \frac{\partial H}{\partial P_i}, \quad \frac{dP_i}{dt} = 0, \quad i = 1, \cdots n.$$

$$P_i(t) = \mathrm{P_i}(0), \quad \mathrm{Q^i(t)} = \mathrm{Q^i}(0) + \omega_\mathrm{i}\mathrm{t}, \quad \omega_\mathrm{i} = \frac{\partial \mathrm{H}}{\partial \mathrm{P_i}}$$

Thus, the whole problem of solving the equations of motion amounts to finding such a canonical transformation. The position variables Q^i are often (not always) periodic and so are called "angle" variables. By convention, we normalize them so that the periods are all equal to 2π. Their conjugates P_i (on which the hamiltonian depends) have the meaning of the action of a closed path (when the system goes around one period of Q^i) so they were called the "action" variables. In modern parlance, the Q^i variables are co-ordinates on a torus embedded in phase space, when they are periodic. Every initital condition lies on some torus and time evolution keeps it on this torus: thus the tori are invariant under time evolution.

In the early days of mechanics, it was believed that every system can be brought to normal form. For example, you could expand around a stable fixed point in a power series and bring the hamiltonian to normal form order by order. Only after the development of modern analysis, it was realized that there is a catch: the infinite series may not converge. There are only a handful of systems for which a reduction to normal is possible. These are called integrable systems.

Nevertheless, a less ambitious form of the original idea survives in the modern theory. If the hamiltonian is a small perturbation from an integrable system, some of the tori continue to be preserved under time evolution. But inside these there will be regions in the phase space where the dynamics is not confined to any torus: the system will wander around filling all $2n$ dimensions. Such chaotic regions might contain islands of invariant tori , which in turn contain chaotic regions and so on: a kind of fractal structure in phase space.

This explains the co-existence of stability and chaotic behavior in the solar system: the planets are in very predictable orbits while asteroids and the particles in the rings of Saturn are not.

There is a similar catch in the idea that every quantum mechanical hamiltonian can be diagonalized. We still don't know enough functional analysis to state precisely the difference between the integrable and chaotic quantum systems, when the quantum Hilbert space is infinite dimensional. Unravelling this distinction between the solvable and unsolvable in quantum mechanics is one of the frontiers of contemporary physics. At this time, the best we can do is to work out examples at the edge of this frontier and work towards a more general theory. This is the quantum analogue of the work of astronomers in the nineteenth century, such as Hill's theory of the Moon. Instead of astronomy, it is condensed matter systems (quantum dots, artificial atoms, nano-materials) that are giving us examples. Unlike in astronomy, we can manipulate these systems, so there is a good chance that in the next generation, much of this mystery will be unravelled.

9.1 The Simple Harmonic Oscillator

This is the prototype of an integrable system. We choose units such that the mass is equal to unity.

$$H = \frac{p^2}{2} + \frac{1}{2}\omega^2 q^2$$

The orbits are ellipses in phase space.

$$q(t) = \frac{\sqrt{2H}}{\omega}\cos\omega t, \quad p(t) = \sqrt{2H}\sin\omega t$$

This suggests that we choose as the position variable

$$Q = \arctan\frac{p}{\omega q}$$

since it evolves linearly in time.

Its conjugate variable is

$$P = \frac{1}{\omega}\left[\frac{p^2}{2} + \frac{1}{2}\omega^2 q^2\right]$$

Exercise 51. Verify that $\{P, Q\} = 1$. Recall that the Poisson bracket is the Jacobian determinant in two dimensions, so what you need to show is that $dpdq = dPdQ$.

Solution One way to see this quickly is to recall that if we go from cartesion to polar co-ordinates

$$dxdy = d\left[\frac{r^2}{2}\right]d\theta$$

Now put $x = \omega q, y = p, Q = \theta$.
Thus we can write the hamiltonian in normal form

$$H = \omega P.$$

9.2 The General One-Dimensional System

Consider now any hamiltonian system with one degree of freedom

$$H = \frac{1}{2m}p^2 + V(q)$$

Assume for simplicity that the curves of constant energy are closed; i.e., that the motion is periodic. Then we look for a co-ordinate system Q, P in the phase space such that the period of Q is 2π and which is canonical

$$dpdq = dPdQ.$$

In this case the area enclosed by the orbit of a fixed value of P will be just $2\pi P$. On the other hand, this area is just $\oint_H pdq$ over a curve constant energy in the original co-ordinates (Stokes' theorem).Thus we see that

$$P = \frac{1}{2\pi}\oint_H pdq = \frac{1}{\pi}\int_{q_1(H)}^{q_2(H)} \sqrt{2m\left[H - V(q)\right]}dq$$

where $q_{1,2}(H)$ are **turning points**; i.e., the roots of the equation $H - V(q) = 0$.

If we can evaluate this integral, we will get P as a function of H. Inverting this function will give H as a function of P , which is its normal form. By comparing with the H-J equation, we see that P is simply $\frac{1}{2\pi}$ times the change in the action over one period of the q variable (i.e., from q_1 to q_2 and back again to q_1):

$$\frac{1}{2m}\left[\frac{\partial S}{\partial q}\right]^2 + V(q) = H, \quad S(q) = \int_{q_1}^{q} \sqrt{2m[H - V(q)}dq$$

This is why P is called the **action variable.**

9.3 Bohr-Sommerfeld Quantization

Once the hamiltonian is brought to normal form, there is a natural way to quantize the system that explicitly displays its energy eigenvalues. The Schrodinger equation becomes

$$H\left(-i\hbar\frac{\partial}{\partial Q}\right)\psi = E\psi$$

The solutions is a "plane wave"

$$\psi = e^{\frac{i}{\hbar}PQ}$$

If the Q-variables are angles (as in the SHO) with period 2π we must require that $P = \hbar n$ for $n = 0, 1, 2\cdots$ in order that the wave function is single-valued. Thus the spectrum of the quantum hamiltonian is

$$E_n = H(\hbar n).$$

In the case of the SHO we get this way

$$E_n = \hbar\omega n, \quad n = 0, 1, \cdots$$

This is almost the exact answer we would get by solving the Schrodinger equation in terms of q: only an additive constant $\frac{1}{2}\hbar\omega$ is missing.

Using the above formula for P, we see that the quantization rule for energy can be expressed as

$$\oint pdq = 2\pi\hbar n.$$

This is known as Bohr-Sommerfeld quantization and provides a semi-classical approximation to the quantum spectrum. In some fortuitous cases (Such as the SHO or the hydrogen atom) it gives almost the exact answer.

9.4 The Kepler Problem

We already know that p_ϕ and $L^2 = p_\theta^2 + \frac{p_\phi^2}{\sin^2\theta}$ (the component of angular momentum in some direction, say z, and the total angular momentum) are a pair of commuting conserved quantities. So this is a problem with just one degree of freedom.

$$H = \frac{p_r^2}{2m} + \frac{L^2}{2mr^2} - \frac{k}{r}$$

To find the normal form we need to evaluate the integral

$$P = \frac{1}{\pi}\int_{r_1}^{r_2}\sqrt{2m\left[H - \frac{L^2}{2mr^2} + \frac{k}{r}\right]}dr$$

between turning points. This is equal to (see below)

$$P = -L - \frac{\sqrt{2m}k}{2\sqrt{-H}}$$

Thus

$$H = -\frac{mk^2}{2(P+L)^2}.$$

Within the B-S approximation, the action variables P, L are both integers.

So the quantity $P+L$ has to be an integer in the Bohr-Sommerfeld quantization: it is the **principal quantum number** of the hydrogenic atom and the above formula gives its famous spectrum. If we include the effects of special relativity, the spectum depends on P, L separately, not just on the sum: this is the **fine structure** of the hydrogenic atom.

9.4.1 A Contour Integral

It is possible to evaluate the integral by trigonometric substitutions, but it is a mess. Since we only want the integral between turning points, there is a trick involving contour integrals. Consider the integral $\oint f(z)dz$ over a counter clockwise contour of the function

$$f(z) = \frac{\sqrt{Az^2 + Bz + C}}{z}$$

On the Riemann sphere, $f(z)dz$ has a branch cut along the line connecting the zeros of the quadratic under the square roots. It has a simple pole at the origin. This integrand also has a pole at infinity; this is clear if we transform to $w = \frac{1}{z}$

$$\frac{\sqrt{Az^2 + Bz + C}}{z}dz = -\frac{\sqrt{Cw^2 + Bw + A}}{w^2}dw$$

The residue of the pole $w = 0$ is

$$-\frac{B}{2\sqrt{A}}.$$

The integral $\oint f(z)dz$ over a contour that surrounds all of these singularities must be zero: it can be shrunk to some point on the Riemann sphere. So the sum of the residues on the two simple poles plus the integral of the discontinuity across the branchcut must be zero:

$$2\int_{z_1}^{z_2} \frac{\sqrt{Az^2 + Bz + C}}{z}dz = 2\pi i\left[\sqrt{C} - \frac{B}{2\sqrt{A}}\right]$$

With the choice

$$A = 2mH, \quad B = 2mk, \quad C = -L^2$$

we get

$$\int_{r_1}^{r_2} \sqrt{2m\left[H - \frac{L^2}{2mr^2} + \frac{k}{r}\right]}dr = i\pi\left[\sqrt{-L^2} - \frac{2mk}{2\sqrt{(2mH)}}\right]$$

$$= \pi\left[-L - \frac{\sqrt{2m}k}{2\sqrt{-H}}\right]$$

9.5 The Relativistic Kepler Problem

Sommerfeld worked out the effect of special relativity on the Kepler problem, which explained the fine structure of the Hydrogen atom within the Bohr model. A similar calculatin can also be done for the General Relativistic problem, but as yet it does not have a physical realization: gravitational effects are negligible in the atom and quantum effects are so in planetary dynamics. We start with the relation of momentum to energy in special relativity for a free particle:

$$p_t^2 - c^2\mathbf{p}^2 = m^2c^4.$$

In the presence of an electrostatic potential this is modified to

$$[p_t - eV(r)]^2 - c^2\mathbf{p}^2 = m^2c^4$$

In spherical polar co-ordinates

$$[p_t - eV(r)]^2 - c^2p_r^2 - \frac{c^2L^2}{r^2} = m^2c, \quad L^2 = p_\theta^2 + \frac{p_\phi^2}{\sin^2\theta}.$$

Since p_t, L, p_ϕ are still commuting quantities this still reduces to a one-dimensional problem. So we still define

$$P = \frac{1}{2\pi}\oint p_r dr$$

as before. With the Coulomb potential $eV(r) = -\frac{k}{r}$ we again have a quadratic equation for p_r.The integral can be evaluated by contour method again.

Exercise 52. Derive the relativistic formula for the spectrum of the hydrogen atom by applying the Bohr-Sommerfeld quantization rule.

9.6 Several Degrees of Freedom

As long as the H-J equation is separable, there is a generalization of the above procedure to a system with several degrees of freedom.

$$S = \sum_i S_i(q_i)$$

$$H = \sum_i H_i\left(q_i, \frac{\partial S_i}{\partial q_i}\right)$$

$$H_i\left(q_i, \frac{\partial S_i}{\partial q_i}\right) + \frac{\partial S_i}{\partial t} = 0.$$

In essence, separation of variables breaks up the system into decoupled one dimensional systems, each of which can be solved as above. This is essentially what we did when we dealt with the Kepler problem above. The momentum ("action') variables are the integrals

$$P_i = \frac{1}{2\pi} \oint p_i dq_i$$

Exercise 53. Find the spectrum of the hydrogen molecular ion H_2^+ within the Bohr-Sommerfeld approximation. Use elliptic polar co-ordinates to separate the H-J equation; express the action variables in terms of complete elliptic integrals.

Chapter 10

The Three Body Problem

Having solved the two body problem, Newton embarked on a solution of the three body problem: the effect of the Sun on the orbit of the Moon. It defeated him. The work was continued by many generations of mathematical astronomers: Euler, Lagrange, Airy,Hamilton, Jacobi, Hill, Poincare', Kolmogorov, Arnold, Moser..It still continues. The upshot is that it is not possible to solve the system in "closed form": more precisely that the solution is not a real analytic funtion of time. But a solution valid for fairly long times was found by perturbation theory around the two body solution: the series will eventually breakdown as it is only asymptotic and not convergent everywhere. There are regions of phase space where it converges, but these regions interlace those where it is divergent. The problem is that there are resonances whenever the frequencies of the unperturbed solution are rational multiples of each other.

The most remarkable result in this subject is a special exact solution of Lagrange: there is a stable solution in which the three bodies revolve around their center of mass, keeping their positions at the vertices of an equilateral triangle. This solution exists even when the masses are not equal: i.e., the three-fold symmetry of the equilateral triangle holds even if the masses are not equal! Lagrange thought that such special orbits would not appear in nature. But we now know that Jupiter has captured some asteroids (Trojans) into such a resonant orbit. Recently, it was found that the Earth also has such a co-traveller at one of its Lagrange points.

The theory is mainly of mathematical (conceptual) interest these days as it is easy to solve astronomical cases numerically. As the first example of a chaotic system, the three body problem remains fascinating to mathematicians. New facts are still being discovered. For example, Simo, Chenciner,Montgomery found a solution ("choreography") in which three bodies of equal mass follow each other along a common orbit that has the shape of a figure eight.

10.1 Preliminaries

Let $\mathbf{r}_a$, for $a = 1, 2 \cdots n$ be the positions of n bodies interacting through the gravitational force. The Lagrangian is

$$L = \frac{1}{2}\sum_a m_a \dot{\mathbf{r}}_a^2 - U, \quad U = -\sum_{a<b} \frac{G m_a m_b}{|\mathbf{r}_a - \mathbf{r}_b|}$$

Immediately we note the conservations laws of energy (Hamiltonian)

$$H = \sum_a \frac{\mathbf{p}_a^2}{2m_a} + U, \quad \mathbf{p}_a = m\dot{\mathbf{r}}_a$$

total momentum

$$\mathbf{P} = \sum_a \mathbf{p}_a$$

and angular momentum

$$\mathbf{L} = \sum_a \mathbf{r}_a \times \mathbf{p}_a.$$

10.1.1 Scale Invariance

Another symmetry is a scale invariance. If $\mathbf{r}_a(t)$ is a solution, so is

$$\lambda^{-\frac{2}{3}} \mathbf{r}_a(\lambda t).$$

Under this transformation,

$$\mathbf{p}_a \to \lambda^{\frac{1}{3}} \mathbf{p}_a, \quad H \to \lambda^{\frac{2}{3}} H, \quad \mathbf{L} \to \lambda^{-\frac{1}{3}} \mathbf{L}$$

In the two body problem, this leads to Kepler's scaling law $T^2 \propto R^3$relating period to the semi-major axis of the ellipse. There is no conserved quantity corresponding to this symmetry, as it does not leave the Poisson brackets invariant. But it does lead to an interesting relation for the moment of inertia about the center of mass

$$I = \frac{1}{2}\sum_a m_a \mathbf{r}_a^2.$$

Clearly,

$$\frac{dI}{dt} = \sum_a \mathbf{r}_a \cdot \mathbf{p}_a \equiv D$$

It is easily checked that

$$\{D, \mathbf{r}_a\} = \mathbf{r}_a, \quad \{D, \mathbf{p}_a\} = -\mathbf{p}_a$$

So that

$$\{D, T\} = -2T$$

for Kinetic energy and

$$\{D, U\} = -U$$

for potential energy. In other words

$$\{D, H\} = -2T - U = -2H + U$$

That is

$$\frac{dD}{dt} = 2H - U$$

or

$$\frac{d^2 I}{dt^2} = 2H - U.$$

If the potential had been proportional to the inverse square distance (unlike the Newtonian case) this would have said instead

$$\ddot{I} = 2H.$$

We wll return to this case later.

10.2 Jacobi Co-Ordinates

Recall that the Lagrangian of the two body problem

$$L = \frac{1}{2} m_1 \dot{\mathbf{r}}_1^2 + \frac{1}{2} m_2 \dot{\mathbf{r}}_2^2 - U(|\mathbf{r}_1 - \mathbf{r}_2|)$$

can be written as

$$L = \frac{1}{2} M_1 \dot{\mathbf{R}}_1^2 - U(|\mathbf{R}_1|) + \frac{1}{2} M_2 \dot{\mathbf{R}}_2^2$$

where

$$\mathbf{R}_1 = \mathbf{r}_2 - \mathbf{r}_1, \quad \mathbf{R}_2 = \frac{m_1 \mathbf{r}_1 + m_2 \mathbf{r}_2}{m_1 + m_2},$$

and

$$M_1 = \frac{m_1 m_2}{m_1 + m_2}, \quad M_2 = m_1 + m_2.$$

This separates the center of mass co-ordinate $\mathbf{R}_2$ from the relative co-ordinate $\mathbf{R}_1$.

Jacobi found a generalization to three particles:

$$\mathbf{R}_1 = \mathbf{r}_2 - \mathbf{r}_1, \quad \mathbf{R}_2 = \mathbf{r}_3 - \frac{m_1\mathbf{r}_1 + m_2\mathbf{r}_2}{m_1 + m_2}, \quad \mathbf{R}_3 = \frac{m_1\mathbf{r}_1 + m_2\mathbf{r}_2 + m_3\mathbf{r}_3}{m_1 + m_2 + m_3}$$

$\mathbf{R}_2$ is the position of the third particle relative to the cm. of the first pair. The advantage of this choice is that the kinetic energy

$$T = \frac{1}{2}m_1\dot{\mathbf{r}}_1^2 + \frac{1}{2}m_2\dot{\mathbf{r}}_2^2 + \frac{1}{2}m_3\dot{\mathbf{r}}_3^2$$

remains diagonal (i..e., no terms such as $\dot{\mathbf{R}}_1 \cdot \dot{\mathbf{R}}_2$):

$$T = \frac{1}{2}M_1\dot{\mathbf{R}}_1^2 + \frac{1}{2}M_2\dot{\mathbf{R}}_2^2 + \frac{1}{2}M_3\dot{\mathbf{R}}_3^2$$

with

$$M_1 = \frac{m_1 m_2}{m_1 + m_2}, \quad M_2 = \frac{(m_1 + m_2)m_3}{m_1 + m_2 + m_3}, \quad M_3 = m_1 + m_2 + m_3$$

Moreover

$$\mathbf{r}_2 - \mathbf{r}_3 = \mu_1\mathbf{R}_1 - \mathbf{R}_2, \quad \mathbf{r}_1 - \mathbf{r}_3 = -\mu_2\mathbf{R}_1 - \mathbf{R}_2$$

with

$$\mu_1 = \frac{m_1}{m_1 + m_2}, \quad \mu_1 + \mu_2 = 1.$$

This procedure has a genearlization to arbitrary number of bodies.

Exercise 54. The construction of Jacobi co-ordinates is an application of Gram-Schmidt orthogonalization, a standard algorithm of linear algebra. Let the mass matrix be $m = \begin{pmatrix} m_1 & 0 & 0 \\ 0 & m_2 & 0 \\ 0 & 0 & m_3 \end{pmatrix}$. Starting with $\mathbf{R}_1 = \mathbf{r}_2 - \mathbf{r}_1$, find a linear combination $\mathbf{R}_2$ such that $\mathbf{R}_2^T \cdot m\mathbf{R}_1 = 0$. Then find $\mathbf{R}_3$ such that $\mathbf{R}_3^T m\mathbf{R}_1 = 0 = \mathbf{R}_3^T m\mathbf{R}_1$. Apply the linear transformation $\mathbf{R}_a = L_{ab}\mathbf{r}_a$ to get the reduced masses $M = L^T m L$. Because of orthogonality, it will be diagonal $M = \begin{pmatrix} M_1 & 0 & 0 \\ 0 & M_2 & 0 \\ 0 & 0 & M_3 \end{pmatrix}$

Thus, the Lagrangian of the three body problem with pairwise central potentials

$$L = \frac{1}{2}m_1\dot{\mathbf{r}}_1^2 + \frac{1}{2}m_2\dot{\mathbf{r}}_2^2 + \frac{1}{2}m_3\dot{\mathbf{r}}_3^2 - U_{12}(|\mathbf{r}_1 - \mathbf{r}_2|) - U_{13}(|\mathbf{r}_1 - \mathbf{r}_3|) - U_{23}(|\mathbf{r}_2 - \mathbf{r}_3|)$$

becomes

$$L = \frac{1}{2}M_1\dot{\mathbf{R}}_1^2 + \frac{1}{2}M_2\dot{\mathbf{R}}_2^2 - U_{12}(|\mathbf{R}_1|) - U_{13}(|\mathbf{R}_2 + \mu_2\mathbf{R}_1|) - U_{23}(|\mathbf{R}_2 - \mu_1\mathbf{R}_1|) + \frac{1}{2}M_3\dot{\mathbf{R}}_3^2$$

Again the c.m. co-ordinate $\mathbf{R}_3$ satisfies

$$\ddot{\mathbf{R}}_3 = 0$$

So we can pass to a reference frame in which it is at rest; and choose the origin at the c.m.:

$$\mathbf{R}_3 = 0$$

Thus the lagrangian reduces to

$$L = \frac{1}{2}M_1\dot{\mathbf{R}}_1^2 + \frac{1}{2}M_2\dot{\mathbf{R}}_2^2 - U_{12}(|\mathbf{R}_1|) - U_{13}(|\mathbf{R}_2 + \mu_2\mathbf{R}_1|) - U_{23}(|\mathbf{R}_2 - \mu_1\mathbf{R}_1|)$$

The Hamiltonian is

$$H = \frac{\mathbf{P}_1^2}{2M_1} + \frac{\mathbf{P}_2^2}{2M_2} + U_{12}(|\mathbf{R}_1|) + U_{13}(|\mathbf{R}_2 + \mu_2\mathbf{R}_1|) + U_{23}(|\mathbf{R}_2 - \mu_1\mathbf{R}_1|)$$

The total angular momentum

$$\mathbf{L} = \mathbf{R}_1 \times \mathbf{P}_1 + \mathbf{R}_2 \times \mathbf{P}_2$$

is conserved as well.

10.2.1 Orbits as Geodesics

The Hamilton-Jacobi equation becomes

$$\frac{1}{2M_1}\left[\frac{\partial S}{\partial \mathbf{R}_1}\right]^2 + \frac{1}{2M_2}\left[\frac{\partial S}{\partial \mathbf{R}_2}\right]^2 + U_{12}(|\mathbf{R}_1|) + U_{13}(|\mathbf{R}_2 + \mu_2\mathbf{R}_1|) + U_{23}(|\mathbf{R}_2 - \mu_1\mathbf{R}_1|) = E$$

Or,

$$[E - \{U_{12}(|\mathbf{R}_1|) + U_{13}(|\mathbf{R}_2 + \mu_2\mathbf{R}_1|) + U_{23}(|\mathbf{R}_2 - \mu_1\mathbf{R}_1|)\}]^{-1}\left\{\frac{1}{2M_1}\left[\frac{\partial S}{\partial \mathbf{R}_1}\right]^2 + \frac{1}{2M_2}\left[\frac{\partial S}{\partial \mathbf{R}_2}\right]^2\right\} = 1$$

This describes geodesics of the metric

$$ds^2 = [E - \{U_{12}(|\mathbf{R}_1|) + U_{13}(|\mathbf{R}_2 + \mu_2\mathbf{R}_1|) + U_{23}(|\mathbf{R}_2 - \mu_1\mathbf{R}_1|)\}]\left\{M_1 d\mathbf{R}_1^2 + M_2 d\mathbf{R}_2^2\right\}$$

The curvature of this metric ought to give insights into the stability of the three body problem. Much work can still be done in this direction.

In the special case $E = 0, U(r) \propto \frac{1}{r}$ this metric has a scaling symmetry: $\mathbf{R}_a \to \lambda\mathbf{R}_a$, $ds^2 \to \lambda ds^2$. If $E \neq 0$ we can use this symmetry to set $E = \pm 1$ thereby choosing a unit of time and space as well.

10.3 The $\frac{1}{r^2}$ potential

If the potential were $\frac{1}{r^2}$ and not $\frac{1}{r}$ as in Newtonian gravity, dilations are a symmetry. Since we are interested in studying the three body problem as a model of chaos and not for astronomical applications any more, we can study this simpler example instead. Poincare' initiated this study in 1897, as part of his pioneering study of chaos. Montgomery has obtained interesting new results in this direction more than a hundred years later.

$$H(\mathbf{r},\mathbf{p}) = \sum_a \frac{\mathbf{p}_a^2}{2m_a} + U, \quad U(\mathbf{r}) = -\sum_{a<b} \frac{k_{ab}}{|\mathbf{r}_a - \mathbf{r}_b|^2}$$

has the symmetry

$$H(\lambda\mathbf{r}, \lambda^{-1}\mathbf{p}) = \lambda^{-2} H(\mathbf{r},\mathbf{p}).$$

This leads to the "almost conservation" law for the generator of this canonical transformation

$$D = \sum_a \mathbf{r}_a \cdot \mathbf{p}_a$$

$$\frac{dD}{dt} = 2H$$

Since

$$D = \frac{d}{dt} I, \quad I = \frac{1}{2}\sum_a m_a \mathbf{r}_a^2$$

we get

$$\frac{d^2}{dt^2} I = 2H.$$

Consider the special case that the total angular momentum (which is conserved) is zero

$$\mathbf{L} = \sum_a \mathbf{r}_a \times \mathbf{p}_a = 0.$$

This has drastic consequences for the stability of the system. If $H > 0$ the moment of interia is a convex function of time: the system will eventually expand to infinite size. If $H < 0$ we have the opposite behavior and the system will shrink to its center of mass in a finite amount of time. Thus the only stable situation is when $H = 0$.

10.3.1 Montgomery's Pair of Pants

See R. Montgomery arxiv:math/040514v1[Math.DS].

In the case $H = \mathbf{L} = 0$, we can reduce the planar three body orbits with the $\frac{1}{r^2}$ potential and equal masses to the geodesics of a metric on a four dimensional space (i.e., two complex dimensions if we think of $R_{1,2}$ as complex numbers $z_{1,2}$).

$$ds^2 = \left[\frac{1}{|z_1|^2} + \frac{1}{|z_2 + \frac{1}{2}z_1|^2} + \frac{1}{|z_2 - \frac{1}{2}z_1|^2}\right]\left\{|dz_1|^2 + \frac{2}{3}|dz_2|^2\right\}$$

There is an isometry (symmetry) $z_a \to \lambda z_a, \quad 0 \neq \lambda \in C$, which combines rotations and scaling. We can use this to remove two dimensions to get a metric on C

$$ds^2 = U(z)|dz|^2$$

where the effective potential $U(z)$ is a positive function of

$$z = \frac{z_2 + \frac{1}{2}z_1}{z_2 - \frac{1}{2}z_1}$$

singular at the points $z = 0, 1, \infty$, corresponding to pairwise collisions. These singular points are at an infinite distance away.(In mechanics, this distance has the meaning of action) Near each singularity the metric looks asymptotically like a cylinder: rather like a pair of pants for a tall thin person.Thus, we get a metric that is complete on the Riemann sphere with three points removed. Topologically, this is the same as the plane with two points removed: $C - \{0, 1\}$.

Exercise 55. Find the function $U(z)$ explicitly.Compute the curvature (Ricci scalar) in terms of derivatives of U.

From Riemannian geometry (Morse theory) we know that there is a minimizing geodesic in each homotopy class: there is an orbit that **minimizes** the action with a prescribed sequence of turns around each singularity. Let A be the homotopy class of curves that wind around 0 once in the counter clockwise direction and B one that winds around 1 . Then A^{-1} and B^{-1} wind around $0, 1$ in the clockwise direction. Any homotopy class of closed curves corresponds to a finite word made of these two letters

$$A^{m_1}B^{n_1}A^{m_2}B^{n_2}\cdots$$

or

$$B^{n_1}A^{m_1}A^{m_2}B^{n_2}\cdots$$

with non-zero m_a, n_a. These form a group F_2, the free group on two generators: A and B do not commute. Indeed they satisfy no relations among each other at all. There are an exponentially large number of distinct words of a given length: F_2 is a **hyperbolic group**. Given each such word we have a minimizing orbit that winds around 0 a certain number m_1 times then around 1 a certain number $n_1$times then again around 0 some number $m_1$times and so on.

Moreover, the curvature of the metric is negative everywhere except at two points (Lagrange points) where it is zero. Thus the geodesics diverge from each other every where . A small change in the initial conditions can make the orbit careen off in some unpredictable direction, with a completely different sequence of A's and B's. This is chaos. The more realistic $\frac{1}{r}$ potential is harder to analyze, but is believed to have similar qualitative behavior.

Research Problem: Find the action of the minimizing geodesic for each element of F_2. Use this to evaluate Gutzwiller's trace formula for $\zeta(s)$, the sum over closed orbits. Compare with the Selberg-zeta function of Riemann surfaces.

Chapter 11

The Restricted Three Body Problem

A particular case of the three body problem is of special historical importance in astronomy: when one of the bodies is of infinitesimal mass m_3 and the other two bodies (the **primaries** of masses m_1, m_2) are in circular orbit around their center of mass; moreover, the orbit of the small body lies in the same plane as this circle. This is a good approximation for a satellite moving under the influence of the Earth and the Moon; an asteroid with the Sun and Jupiter; a particle in a ring of Saturn influenced also by one of its moons. The basic results are due to Lagrange, but there are refinements (e.g., "halo orbits") being discovered even in our time.

11.1 The Motion of the Primaries

Since the secondary has infinitesimal mass, its effect on the primaries can be ignored. Choose a reference frame where the center of mass of the primaries is at rest at the origin. The relative co-ordinate will describe an ellipse. We assume that the eccentricity of this orbit is zero, a circle centered at the origin. If R is the radius (the distance between the primaries) , and Ω the angular velocity,

$$\frac{m_1 m_2}{m_1+m_2} R\Omega^2 = \frac{G m_1 m_2}{R^2}, \implies \Omega^2 = \frac{G(m_1+m_2)}{R^3}.$$

This is just Kepler's third law. The distance of the first primary from the c.m. is νR with $\nu = \frac{m_2}{m_1+m_2}$. We can assume that $m_1 > m_2$ so that $\nu < \frac{1}{2}$.The other primary will be at a distance $(1-\nu)R$ in the opposite direction. Thus the positions of the primaries are, in polar co-ordinates, $(\nu R, \pi - \Omega t)$ and $(-[1-\nu]R, \Omega t)$.

The secondary will move in the gravitational field created by the two primaries. This field is time dependent, with a period equal to $\frac{2\pi}{\Omega}$. The Lagrangian is,(after dividing out a common factor m_3)

$$L = \frac{1}{2}\dot{r}^2 + \frac{1}{2}r^2\dot{\phi}^2 + G(m_1 + m_2)\left[\frac{1-\nu}{\rho_1(t)} + \frac{\nu}{\rho_2(t)}\right].$$

where $\rho_{1,2}(t)$ is the distances to the primaries:

$$\rho_1(t) = \sqrt{\left[r^2 + \nu^2R^2 + 2\nu rR\cos[\phi - \Omega t]\right]}, \quad \rho_2(t) = \sqrt{\left[r^2 + (1-\nu)^2R^2 - 2(1-\nu)rR\cos[\phi - \Omega t]\right]}$$

Since the Lagrangian is time dependent, energy is not conserved: the secondary can extract energy from the rotation of the primaries. But we can make a transformation to a rotating co-ordinate

$$\chi = \phi - \Omega t$$

to emilinate this time dependence.

$$L = \frac{1}{2}\dot{r}^2 + \frac{1}{2}r^2[\dot{\chi} + \Omega]^2 + G[M + m]\left[\frac{1-\nu}{r_1} + \frac{\nu}{r_2}\right]$$

where

$$r_1 = \sqrt{\left[r^2 + \nu^2R^2 + 2\nu rR\cos\chi\right]}, \quad r_2 = \sqrt{\left[r^2 + (1-\nu)^2R^2 - 2(1-\nu)rR\cos\chi\right]}$$

We pay a small price for this: there are terms in the Lagrangian that depend on $\dot{\chi}$ linearly. These lead to velocity dependent forces (the **Coriolis force**) in addition to the more familiar centrifugal force. Neverthless, we gain a conseved quantity, the hamiltonian.

$$H = \dot{r}\frac{\partial L}{\partial \dot{r}} + \dot{\chi}\frac{\partial L}{\partial \dot{\chi}} - L$$

$$H = \frac{1}{2}\dot{r}^2 + \frac{1}{2}r^2\dot{\chi}^2 - G[m_1 + m_2]\left[\frac{r^2}{2R^3} + \frac{1-\nu}{r_1} + \frac{\nu}{r_2}\right]$$

This \is the sum of kinetic energy and a potential energy; it is often called the **Jacobi integra**l. ("Integral" in an old term for a conserved quantity.) The Coriolis force does no work, being normal to the velocity always; so it does not contribute to the energy. It is important this is the energy measured in a non-intertial reference frame, which is why it includes the term $\propto r^2$, the centrifugal potential.

It is also useful to write the Lagrangian in rotating Caretsian co-ordinates

$$H = \frac{1}{2}\dot{x}^2 + \frac{1}{2}\dot{y}^2 + V(x,y), \quad L = \frac{1}{2}\dot{x}^2 + \frac{1}{2}\dot{y}^2 + \Omega\left[x\dot{y} - y\dot{x}\right] - V(x,y).$$

$$V(x,y) = -G[m_1+m_2]\left[\frac{x^2+y^2}{2R^3}+\frac{1-\nu}{r_1}+\frac{\nu}{r_2}\right]$$
$$r_1 = \sqrt{(x-\nu R)^2+y^2}, \quad r_2\sqrt{(x+[1-\nu]R)^2+y^2}.$$

11.1.1 A useful identity

It is useful to express the potential energy in terms of r_1 and r_2, eliminating r. From the definition of $r_{1,2}$ we can verify that

$$\frac{1}{\nu}r_1^2+\frac{1}{1-\nu}r_2^2 = \frac{1}{\nu(1-\nu)}r^2+R^2.$$

Thus

$$V(r_1,r_2) = -G\left[m_1\left\{\frac{r_1^2}{2R^3}+\frac{1}{r_1}\right\}+m_2\left\{\frac{r_2^2}{2R^3}+\frac{1}{r_2}\right\}\right]$$

up to an irrelevant constant.

11.1.2 Equilibrium points

There are points where the forces are balanced such that the secondary can be at rest. (In the inertial frame, it will then rotate at the same rate as the primaries.) In studying the potential, we can use the distances r_1 and r_2 themselves as co-ordinates in the plane: the potential is separable with this choice. But beware that this system breaks down along the line connecting the primaries, as along there r_1 and r_2 are not independent variables. Also, these variables cover only one half of the plane, the other half being obtained by reflection about the line connecting the primaries.

A short exercise in calculus will show that there is a **maximum** of the potential when

$$r_1 = r_2 = R$$

That is, when the thee bodies are located along the verices of an equilateral triangle. There are actually two such points, on either side of the primary line. They are called **Lagrange points** L_4 and L_5. There are three more equilibrium points L_1, L_2, L_3 that lie along the primary line $y = 0$. They are not visible in terms of r_1 and r_2 because that system breaks down there. But in the Cartesian co-ordinates, it is clear by the symmetry $y \to -y$ that

$$\frac{\partial V}{\partial y} = 0, \quad \text{if } y = 0.$$

Then

$$V(x,0) = -G[m_1+m_2]\left[\frac{x^2}{2R^3} + \frac{1-\nu}{|x-\nu R|} + \frac{\nu}{|x+[1-\nu]R|}\right]$$

This function has three extrema, L_1, L_2, L_3. As functions of $x$these are maxima, but are minima along y: they are **saddle points** of V. There are no other equilibrium points.

11.1.3 Hill's Regions

It is already clear that for a given H, only the regin with $H - V > 0$ is accesible to the secondary particle. For small H the curve $H = V$ is disconnected, with regions near $m_1$and m_2 and near infinity: these are the places where the potential goes to $-\infty$. As Hgrows to the value of the potential at L_1 (the saddle point in between the two primaries) the two regions around the primaries touch; as H grows higher, they merge into a single region. It is only as Hgrows larger than the potential at $L_{4,5}$ that all of space is avaliable for the secondary.

For example if a particle is to from a point near m_1 to one near m_2, the least amount of energy it needs to have is the potential at the Lagrange point in between them. The saddle point is like a mountain pass that has to be climbed to go from deep valley to another. This has interesting implications for space travel; many of which have been explored in fiction. For example, in the imagination of many authors, Lagrange points would have strategic importance (like straits that separate continents) to be guarded. There is also more scientifically interesting study by Belbruno, Marsden and others on transfer orbits of low fuel cost.

11.1.4 The second derivative of the potential

Th equilibrium points L_4, L_5, being at the maximum of a potential, would ordinarily be unstable. But an amazing fact discovered by Lagrange is that the velocity dependence of the Coriolis force can (if ν is not too close to a half) make them **stable** equilibria. Such a reversal of fortune does not happen for L_1, L_2, L_3: small departures from these points will grow. But it has been found recently (numerically) that there are orbits near these points, called Halo orbits which do not cost much in terms of rocket fuel to maintain.

To understand the stability of $L_4$and L_5 we must expand the Lagrangian to second order around them and get an equation for small perturbations. The locations of $L_{4,5}$ are

$$x = \frac{R}{2}, y = \pm R\sqrt{\frac{3}{4} - \nu(\nu-1)}.$$

The second derivative of V at $L_{4,5}$ is

$$V'' \equiv K = -\Omega^2 \begin{bmatrix} \frac{3}{4} & \pm\frac{\sqrt{27}}{4}[1-2\nu] \\ \pm\frac{\sqrt{27}}{4}[1-2\nu] & \frac{9}{4} \end{bmatrix}$$

Note that $K < 0$: both its eigenvalues are negative. On the other hand for $L_{1,2,3}$ we will have a diagonal matrix for K with one positive eigenvalue (along y) and one negative eigenvalue (along x).

11.1.5 Stability Theory

The Lagrangian takes the form, calling the departure from equilibrium (q_1, q_2)

$$L \approx \frac{1}{2}\dot{q}_i\dot{q}_i + \frac{1}{2}B_{ij}q_i\dot{q}_j - \frac{1}{2}K_{ij}q_iq_j$$

where

$$B = 2\Omega \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

comes from the Coriolis force. For small q, the equations of motion become

$$\ddot{q} + B\dot{q} + Kq = 0.$$

We seek solutions of the form

$$q(t) = e^{i\omega t}A$$

for some constant vector and frequency ω. Real values of ω would describe stable perturbations. The eigenvalue equation is somewhat unusual

$$[-\omega^2 + Bi\omega + K]A = 0$$

in that it involves both ω and ω^2.Thus the characteristic equation is

$$\det[-\omega^2 + Bi\omega + K] = 0.$$

Or

$$\det \begin{bmatrix} K_{11} - \omega^2 & K_{12} + 2i\Omega\omega \\ K_{12} - 2i\Omega\omega & K_{22} - \omega^2 \end{bmatrix} = 0$$

which becomes

$$\omega^4 - [4\Omega^2 + \mathrm{tr}K]\omega^2 + \det K = 0$$

There are two roots for ω^2.

$$(\omega^2 - \lambda_1)(\omega^2 - \lambda_2) = 0$$

The condition for stabilityis that both roots must be real and positive. This is equivalent to requiring that the discriminant s positive and also that $\lambda_1\lambda_2 > 0, \lambda_1 + \lambda_2 > 0$. Thus

$$[\mathrm{tr}K + 4\Omega^2]^2 - 4\det K > 0, \quad \mathrm{tr}K + 4\Omega^2 > 0, \quad \det K > 0$$

So,

$$\det K > 0, \quad 4\Omega^2 + \mathrm{tr}K > 2\sqrt{\det K}$$

The first condition cannot be satisfied by a saddle point of V: the eigenvalues of K has opposite signs. So $L_{1,2,3}$ are unstable equilibria.

But surprisingly, it can be satisfied by a maximum as well as the expected minimum of a potential. For $L_{4.5}$

$$\mathrm{tr}K = -3\Omega^2, \quad \det K = \frac{27}{4}\nu(1-\nu)\Omega^4$$

so that the condition becomes

$$27\nu(1-\nu) < 1.$$

In other words,

$$\nu < \frac{1}{2}\left[1 - \sqrt{1 - \frac{4}{27}}\right] \approx 0.03852$$

(Recall that we chose $\nu < \frac{1}{2}$; by calling m_1 the mass of the larger primary). Thus we get stability if the masses of the two primaries are sufficiently different from each other. In this case the frequencies are given by

$$\omega^2 = \frac{1 \pm \sqrt{1 - 27\nu(1-\nu)}}{2}$$

When $\nu << 1$, one of these frequencies will be very small: meaning that the orbit is nearly synchronous with the primaries.

For the Sun-Jupiter system, $\nu = 9.5388 \times 10^{-4}$ so the Lagrange points are stable. The periods of libration (the small oscillations around the equilibrium) follow from the orbital period of Jupiter (11.86 years) : 147.54 years or 11.9 years.

For the Earth-Moon system $\nu = \frac{1}{81}$ is still small enough for stability. The orbital period being 27.32 days, we have libration periods of 90.8days and 28.6 days.

Lagrange discovered something even more astonishing: the equilateral triangle is a stable exact solution for the **full three body problem**, not assuming one of the bodies to be infinitesimally small. He thought that these special solutions were artificial and that they would never be realized in nature. But we now know that there are asteroids (**Trojan asteroids**) that form an equilateral triangle with Sun and Jupiter. This summer (2011 June) even the Earth is found to have such a co-traveller at its Lagrange point with the Sun.

Chapter 12

Magnetic Fields

The force on a charged particle in a magnetic field is normal to its velocity. So it does no work on the particle. The total energy of the particle is not affected by the magnetic field: the hamiltonian as a function of position and velocities does not involve the magnetic field. Can Hamiltonian mechanics still be used to describe such systems? If so, where does the information in the magnetic field go in? It turns out that the magnetic field modifies the Poisson Brackets and not the hamiltonian.

12.1 The Equations of Motion

$$\frac{d\mathbf{r}}{dt} = \mathbf{v}, \quad \frac{d}{d}[m\mathbf{v}] = e\mathbf{v} \times \mathbf{B}$$

Or in terms of components

$$m\frac{dx^i}{dt} = v^i, \quad \frac{d}{dt}[mv^i] = e\epsilon_{ijk}v^j B^k$$

Here is completely anti-symmetric and

$$\epsilon_{123} = 1$$

Let us assume that the magnetic field does not depend on time, only on the psoition.

12.2 Hamiltonian Formalism

The energy (Hamiltonian) is just

$$H = \frac{1}{2}mv^i v^i$$

We want however,

$$\{H, x^i\} = v^i, \quad \{H, v_i\} = \frac{e}{m}\epsilon_{ijk}v^j B^k$$

The first is satisfied if

$$\{p^i, x^j\} = \frac{1}{m}\delta^{ij}$$

which is the usual relation following from canonical relations between position and momentum. So we want

$$\left\{\frac{1}{2}mv^j v^j, v^i\right\} = \frac{e}{m}\epsilon_{ijk}v^j B^k$$

Using the Leibnitz rule this becomes

$$v^j \{v^j, v_i\} = \frac{e}{m^2}\epsilon_{ijk}v^j B^k$$

It is therefore sufficient that

$$\{v_j, v_i\} = \frac{e}{m^2}\epsilon_{ijk}B^k$$

This is **not** what follows from canonical relations: the different components of momentum would commute then. To make this distinction clear, let us denote momentum by

$$\pi_i = mv_i$$

Then :

$$\{x^i, x^j\} = 0, \quad \{\pi^i, x^j\} = \delta^{ij}, \quad \{\pi_i, \pi_i\} = -eF_{ij}$$

where

$$F_{ij} = \epsilon_{ijk}B_k$$

The brackets are anti-symmetric. The Jacobi identity is automatic for all choices except one:

$$\{\{\pi_i, \pi_j\}, \pi_k\} = -e\{F_{ij}, mv_k\}$$

$$= e\partial_k F_{ij}$$

Taking the cyclic sum, we get

$$\partial_i F_{jk} + \partial_j F_{ki} + \partial_k F_{ij} = 0$$

If you work out in components you willsee that this is the condition

$$\nabla \cdot \mathrm{B} = 0$$

which is one of Maxwell's equations.Thus the Jacobi identity is satisfied as long this Maxwell's equation is satisfied.

If we have an electrostatic as well as a magnetic field the Hamiltonian will be

$$H = \frac{\pi_i \pi_i}{2m} + eV$$

again with the commutation relations above.

12.3 Canonical Momentum

It i possible to bring the commutation relations back to the standard form

$$\{x^i, x^j\} = 0, \quad \left\{p^i, x^j\right\} = \delta^{ij}, \quad \{p_i, p_i\} = 0$$

in those cases where the magnetic field is a curl. Recall that locally, every field satisfying

$$\nabla \cdot \mathbf{B} = 0$$

is of the form

$$\mathbf{B} = \nabla \times \mathbf{A}$$

for some vector field $\mathbf{A}$. This is not unique: a change (**gauge transformation)**

$$\mathbf{A} \to \mathbf{A} + \nabla \Lambda$$

leaves $\mathbf{B}$ unchanged. Now if we define

$$\pi_i = p_i - eA_i$$

then the canonical relations imply the relations for π_i.

12.4 The Lagrangian

This suggests that we can find a Lagrangian in terms of A_i. We need

$$p_i = \frac{\partial L}{\partial \dot{x}^i}$$

or

$$\frac{\partial L}{\partial \dot{x}^i} = m\dot{x}_i + eA_i$$

Thus we propose

$$L = \frac{1}{2} m \dot{x}^i \dot{x}^i + eA_i \dot{x}^i - eV$$

as the Lagrangian for a particle in an electromagnetic field.

Exercise 56. Show that the Lorentz force equations follows from this Lagrangian.

An important principle of electromagnetism is that the equations of motion should be invariant under gauge transformations $A_i \to A_i + \partial_i \Lambda$. Under this change the action changes to

$$S = \int_{t_1}^{t_2} L dt \to S + \int_{t_1}^{t_2} e\dot{x}^i \partial_i \Lambda dt$$

The extra terms is a total derivative, hence only depends on end-points:

$$\int_{t_1}^{t_2} e\frac{d\Lambda}{dt} dt = e\left[\Lambda(x(t_2)) - \Lambda(x(t_1))\right].$$

Since we hold the endpoints fixed during a variation, this will not affect the equations of motion.

12.5 The Magnetic Monopole

Recall that the electric field of a point particle satisfies

$$\nabla \cdot \mathbf{E} = 0$$

everywhere but its location. Can there be point particles that can serve as sources of magnetic fields the same way? None have been discovered to date:only magnetic dipoles have been found, a combination of North and South poles. Dirac discovered that the existence of even one such magnetic monopole the somewhere in the universe would explain a remarkable fact about nature: that electric charges appear as multiples of a fundamental unit of charge. To understand this let us study the dynamics of an electrically charged particle in the field of a magnetic monopole, an analysis due to M. N. Saha.

$$\frac{d\mathbf{r}}{dt} = \mathbf{v}, \quad \frac{d}{d}[m\mathbf{v}] = eg\mathbf{v} \times \frac{\mathbf{r}}{r^3}$$

where r is the strength of the magnetic monopole.The problem has spherical symmetry, so we should expect angular momentum to be conserved. But we can check that

$$\frac{d}{dt}[\mathbf{r} \times m\mathbf{v}] = eg\mathbf{r} \times \left[\mathbf{v} \times \frac{\mathbf{r}}{r^3}\right]$$

is not zero. What is going on? Now, recall that identity

$$\frac{d}{dt}\left[\frac{\mathbf{r}}{r}\right] = \frac{\mathbf{v}}{r} - \frac{\mathbf{r}}{r^2}\dot{r}$$

But

$$\dot{r} = \frac{1}{2r}\frac{d}{dt}[\mathbf{r}\cdot\mathbf{r}]$$

$$= \frac{1}{r}\mathbf{r}\cdot\mathbf{v}$$

So

$$\frac{d}{dt}\left[\frac{\mathbf{r}}{r}\right] = \frac{r^2\mathbf{v} - \mathbf{r}(\mathbf{v}\cdot\mathbf{r})}{r^3}$$

or

$$\frac{d}{dt}\left[\frac{\mathbf{r}}{r}\right] = \frac{1}{r^3}\mathbf{r}\times[\mathbf{v}\times\mathbf{r}]$$

Thus we get a new conservation law

$$\frac{d}{dt}\left[\mathbf{r}\times m\mathbf{v} - eg\frac{\mathbf{r}}{r}\right] = 0.$$

The conserved angular momentum is the sum of the orbital angular momentum and a vector pointed along the line connecting the charge and the monopole. This can be understood as the angular momentum contained in the elecromagnetic field. When an electric and a magnetic field exist together, they carry not only energy but also momentum and angular momentum. If you integrate the angular momentum denisty over all of space in the situation abov you will get exactly this extra term.

$$\mathbf{J} = \mathbf{r}\times m\mathbf{v} - eg\frac{\mathbf{r}}{r}$$

is a fixed vector in space. The orbit does not lie in the plane normal to to $\mathbf{J}$. Instead it lies on a cone, whose axis is along $\mathbf{J}$. The angle α of this cone is given by

$$J\cos\alpha = \mathbf{J}\cdot\frac{\mathbf{r}}{r} = -eg.$$

Exercise 57. Verify that $\mathbf{J}$ satisfies the commutation relations of angular momentum

$$\{J_i, J_j\} = \epsilon_{ijk}J_k$$

Does orbital angular momentum $\mathbf{r}\times m\mathbf{v}$ satisfy these relations?

Exercise 58. Determine the orbit by exploiting the conservation of energy and angular momentum.

12.6 Quantization of electric charge

If we quantize this system, we know that an eigenvalue of J_3 is an integer or half-integer multiple of $\hbar$ and that the eigenvalues of J^2 are $j(j+1)$ where j is also such a multiple. On the other hand $\mathbf{L} = \mathbf{r} \times m\mathbf{v}$ also is quantized in the same way. It follows that the vector $eg\frac{\mathbf{r}}{r}$ must have magnitude which is a multiple of $\hbar$

$$eg = n\hbar, \quad n = \frac{1}{2}, 1, \frac{3}{2}, 2 \cdots$$

Thus, if there is even one magnetic monopole somewhere in the universe, electric charge has to be a quantized in multiples of $\hbar$. We do see that electric charge is quantized this way, the basic unit being the magnitude of the charge of the electron. We do not yet know if the reason for this is the existence of a magnetic monopole.

12.7 The Penning Trap

A static electromagnetic field can be used to trap charged particles. You can bottle up antimatter this way; or use hold an electron or ion in place to make precise measurements on it.

It is not possible for a static electric field by itself to provide a stable equilibrium point: the potential must satisfy the Laplace equation $\nabla^2 V = 0$. So at any point the sum of the eigenvalues of the Hessian matrix V''(second dervatives) must vanish. At a stable minimum they would all have to be positive. It is possible to have one negative and two positive eigenvalues. This is true for example for a quadrupole field

$$V(x) = \frac{k}{2}[2x_3^2 - x_1^2 - x_2^2]$$

Such a field can be created by using electrically charged conducting plates shaped like a hyperboloid:

$$x_1^2 + x_2^2 - 2x_3^2 = \text{constant}$$

So the motion in the $x_1 - x_2$ plane is unstable but that along the x_3 axis is stable. Now we can put a constant magnetic field pointed along the $x_3$axis. If it is strong enough, we get a stable equilibrium point at the origin.

Look at the equation of motion of a particle in a constant magnetic field and an electrostatic potential that is a quadratic function of position $V(x) = \frac{1}{2}x^T Kx$.

$$\ddot{q} + F\dot{q} + \frac{e}{m}Kq = 0$$

Here F is an antisymmetric matrix proportional to the magnetic field. For a field of magnitude B alng the x_3 axis

$$F = \frac{e}{m}\begin{pmatrix} 0 & B & 0 \\ -B & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

If we assume the ansatz

$$q = Ae^{i\omega t}$$

the equation becomes

$$\left[-\omega^2 + i\omega F + \frac{e}{m}K\right] A = 0$$

which gives the characterstic equation

$$\det\left[-\omega^2 + i\omega F + \frac{e}{m}K\right] = 0$$

If themagnetic field is along the third axis and if the other matrix has the form

$$K = \begin{pmatrix} k_1 & 0 & 0 \\ 0 & k_2 & 0 \\ 0 & 0 & k_3 \end{pmatrix}$$

this equation factorizes

$$(-\omega^2 + k_3)\det\begin{bmatrix} \frac{e}{m}k_1 - \omega^2 & i\omega\frac{eB}{m} \\ -i\omega\frac{eB}{m} & \frac{e}{m}k_2 - \omega^2 \end{bmatrix} = 0$$

The condition for stability is that all roots for ω^2 are positive. Obe of the roots is k_3 so it must be positive. It is now enough that the discrimant as well as the sum and product of the other two roots are positive. This amounts to

$$k_3 > 0, \quad k_1 k_2 > 0, \quad \frac{eB^2}{m} > 2\sqrt{k_1 k_2} - (k_1 + k_2).$$

The physical meaning is that the electrostatic potential stabilizes the motion in the $x_3$direction. Although the electric field pushes the particle away from the origin, the magnetic force pushes it back in.

A collection of particles moving in such a trap will have frequencies dependent on the ratios $\frac{e}{m}$. These can be measured by Fourier transforming the electric current they induce on a probe. Thus we can measure the ratios $\frac{e}{m}$, a technique called **Fourier transform mass spectrometry.**

Exercise 59. Find the libration frequencies.

Chapter 13

Discrete Time

Is the flow of time continuousor is it discrete, like the sand in an hourglass? As far as we know it is continuous. Yet, it is convenient in many situations to think of it as discrete. For example, in solving a differential equation numerically, it is convenient to calculate the finite change over a small time interval and iterate this process. It is important that we retain the symmetries and conservation laws of the differential equation in making this discrete approximation. In particular, each time step must be a canonical (also called symplectic) transformation. Such symplectic integrators are commonly used to solve problems in celestial mechanics.

Another reason to study discrete time evolution is more conceptual. It goes back to Poincare's pioneering study of chaos. Suppose we have a system with several degrees of freedom. We can try to get a partial understanding by projecting to one degree of freedom (say (p_1, q_1). That is, we look only at initial conditions where the degrees of freedom are fixed at some values (say $(p_i, q_i) = 0$ for $i > 1$). Then we let the system evolve in the full phase space and ask when it will return to this subspace (i.e., what is the next value of (p_1, q_1) for which $(p_i, q_i) = 0$?). This gives a canonical transformation (called the **Poincare' map**) of the plane (p_1, q_1) to istelf. We can then iterate this map to get an orbit, an infinite sequence of points in the plane. In simple cases (like the harmonic oscillator) this orbit will be periodic. If the system is chaotic, the orbit will wander all over the plane, and it is interesting to ask for the density of its distribution: how many points in the orbit are there in some given area? This distribution often has an interesting fractal structure: there are islands that contain points in the orbit surrounded by regions that do not contain any. But if we were to magnify these islands, we will see that they contain other islands surrounded by empty regions and so on to the smallest scales.

13.1 First Order Symplectic Integrators

Symplectic transformation is just another name for a canonical transformation; i.e., a transformation that preserves the Poisson brackets. In many cases of physical interest the hamiltonian of a mechanical system is of the form

$$H = A + B$$

where the dynamics of A and B are separately are easy to solve. For example, if the hamiltonian is

$$H = T(p) + V(q)$$

with a kinetic energy T that depends on momentum variables alone and a potential energy V that depends on position alone, it is easy to solve the equations for each separately:

$$\frac{\partial p_i}{\partial t_1} = \{T, p_i\} = 0, \quad \frac{\partial q^i}{\partial t_1} = \{T, q^i\} = \frac{\partial T}{\partial p_i}, \implies (p_i(t_1), q^i(t_1)) = \left(p_i(0), q^i(0) + \frac{\partial T}{\partial p_i} t_1\right)$$

$$\frac{\partial p_i}{\partial t_2} = \{V, p_i\} = -\frac{\partial V}{\partial q^i}, \quad \frac{dq^i}{dt_2} = \{V, q^i\} = 0, \implies (p_i(t_2), q^i(t_2)) = \left(p_i(0) - \frac{\partial V}{\partial q^i} t_2, q^i(0)\right)$$

The problem is of course that these two canonical transformations do not commute. So we cannot solve the dynamics of H by combining them.

But the commutator of these transformations is of order $t_1 t_2$. Thus for small time intervals it is small, and it might be a good approximation to ignore the lack of commutativity. This gives a first order approximation. If we split the time into small enough intervals, the iteration of this naive approximation might be good enough. We then iterate this time evolution to get an approximation to the orbit. Later we will see how to improve on this by including the effects of the commutator to the next order.

If we perform the above two canonical transformations consecutively[1], choosing equal time steps ϵ, we get the discrete time evolution

$$p_i' = p_i - \epsilon V_i(q), \quad q^{i\prime} = q^i + \epsilon T^i(p')$$

where $V_i = \frac{\partial V}{\partial q^i}, T^i = \frac{\partial T}{\partial p_i}$. In many cases this simple "first order symplectic integrator" already gives a good numerical integration scheme.even here, it is important to keep the

Example 60. The consider the example of the simple pendulum $T = \frac{p^2}{2}$, $V = \omega^2 [1 - \cos q]$. If we perform the above two canonical transformations consecutively, choosing equal time steps ϵ, we get the discrete time evolution

[1] We transform p first then q because usually it leads to simpler formulas. See example.

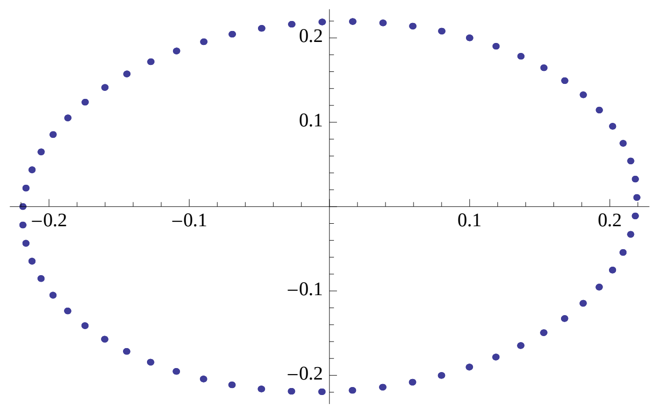

Figure 13.1:

$$p' = p - \epsilon\omega^2 \sin q, \quad q' = q + \epsilon p'$$

Equivalently,

$$p' = p - \epsilon\omega^2 \sin q, \quad q' = q + \epsilon p - \epsilon^2\omega^2 \sin q$$

This is a canonical transformation:

$$\det \begin{bmatrix} \frac{\partial p'}{\partial p} & \frac{\partial p'}{\partial q} \\ \frac{\partial q'}{\partial p} & \frac{\partial q'}{\partial q} \end{bmatrix} = \det \begin{bmatrix} 1 & -\epsilon\omega^2 \cos q \\ \epsilon & 1 - \epsilon^2\omega^2 \cos q \end{bmatrix} = 1.$$

Iterating this map we get a discrete approximation to the time evolution of the pendulum. It gives a nice periodic orbit as we expect for the pendulum.

Notice that if we had made another discrete approximation (which attempts to do the T and V transformations together) we would not have obtained a canonical transformation:

$$p' = p - \epsilon\omega^2 \sin q, \quad q' = q + \epsilon p$$

$$\det \begin{bmatrix} \frac{\partial p'}{\partial p} & \frac{\partial p'}{\partial q} \\ \frac{\partial q'}{\partial p} & \frac{\partial q'}{\partial q} \end{bmatrix} = \det \begin{bmatrix} 1 & -\epsilon\omega^2 \cos q \\ \epsilon & 1 \end{bmatrix} \neq 1$$

The orbit of this map goes wild, not respecting conservation of energy or of area:

This is why we need to use sympletic integrators.

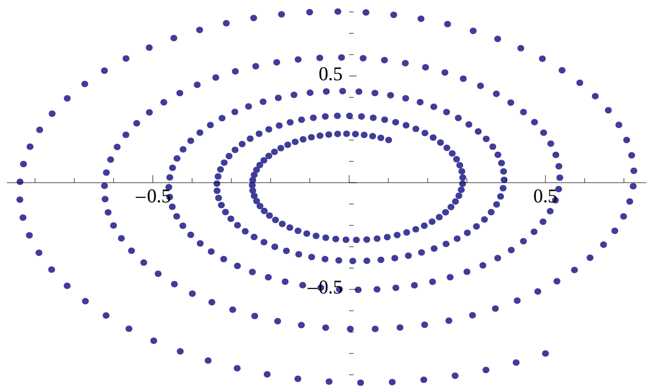

Figure 13.2:

13.2 Second Order Symplectic Integrator

Suppose we are solving a linear system of differential equations

$$\frac{dx}{dt} = Ax$$

for some constant matrix A. The solution is

$$x(t) = e^{tA} x_0$$

where the exponential of a matrix is defined by the series

$$e^{tA} = 1 + tA + \frac{1}{2!}t^2A^2 + \frac{1}{3!}t^3A^3 \cdots$$

Solving nonlinear differential equations

$$\frac{dx^i}{dt} = A^i(x), \quad x^i(0) = x_0^i$$

is the same idea, except that the matrix is replaced by a vector field whose components can depend on x. The solution can be thought of still as an exponential of the vector field, defined by a similar power series. For example, the change of a function under time evolution has the Taylor series expansion

$$f(x(t) = f(x_0) + tAf(x_0) + \frac{1}{2!}t^2A^2f(x_0) + \frac{1}{3!}t^3A^3f(x_0)$$

Here

$$Af = A^i\frac{\partial f}{\partial x^i}, \quad A^2f = A^i\frac{\partial}{\partial x^i}\left(A^j\frac{\partial f}{\partial x^j}\right), \cdots$$

Now, just as for matrices,

$$e^{A+B} \neq e^A e^B$$

in general, because A and Bmay not commute. Up to second order in t we can ignore this effect

$$e^{t(A+B)} = e^{tA}e^{tB} + \mathrm{O}(t^2)$$

The first order symplectic integrator we described earlier is based on this.

There is a way to correct for the comutator, called the Baker-Campbell-Hausdorff formula (or Poincare'-Birkhoff-Witt lemma). To second order, it gives

$$e^{tA}e^{tB} = \exp\left\{tA + tB + \frac{t^2}{2}[A,B] + \mathrm{O}(t^3)\right\}$$

It is possible to determine higher order corrections as well, but we will content ourselves with the first one. It follows that

$$e^{t(A+B)} = e^{\frac{1}{2}tA}e^{tB}e^{\frac{1}{2}tA} + \mathrm{O}(t^3)$$

In our case, A and B will be generators of canonical transformations whose Hamilton's equations are easy to solve (i.e., e^{tA} and e^{tB} are known.) We can then find an approximation for the solution of hamilton's equations for $A + B$. As an example, if

$$H(q,p) = T(p) + V(q)$$

we can deduce a second order symplectic integrator (choosing $A = T, B = V$). Define an intermediate variable

$$z_i = p_i - \frac{1}{2}\epsilon V_i(q)$$

Then a step of the iteration is

$$q^{i\prime} = q^i + \epsilon T^i(z)$$

$$p_i' = z_i - \frac{1}{2}\epsilon V_i(q^{i\prime})$$

See H. Yoshida Phys. Lett. A150, 262(1990) for higher orders.

13.3 Chaos With One Degree of Freedom

A hamiltonian system with a time independent hamiltonian always has at leat one conserved quantity: the hamiltonian itself. Thus we wxpect all such systen with one degree of freedom to be integrable. But if the hamiltonin is time dependent this is not the case any more. For example, if the hamiltonian is periodic funcion of time, in each period the system will evolve by a canonical transformation. Iterating this we will get an orbit of a symplectic map which can often be chaotic. G. D. Birkhoff, Acta. Math. **43**, 1 (1920) is an early study of this situation.

Example 61. The Chirikov Standard Map

$$p_{n+1} = p_n + K \sin q_n, \quad q_{n+1} = q_n + p_{n+1}$$

This can be thought of either as a discrete approximation to the pendulum (see above) or as the time evolution over one period of a **kicked rotor** with the periodic time dependent hamiltonian

$$H = \frac{p^2}{2} + K \cos q \sum_{n \in Z} \delta(t - n)$$

Although the evolution of the pendulum itself is integrable, this discrete evolution can be chaotic. Here are plots of the orbits of the same initial point $(p, q) = (0.1, 1)$ for various values of K. Thus we see that iterations of a canonical transformations can have quite unpredictable behavior even with just one degree of freedom.

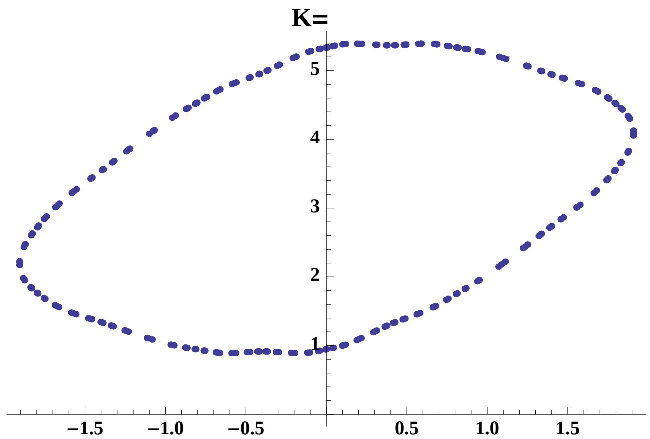

Figure 13.3:

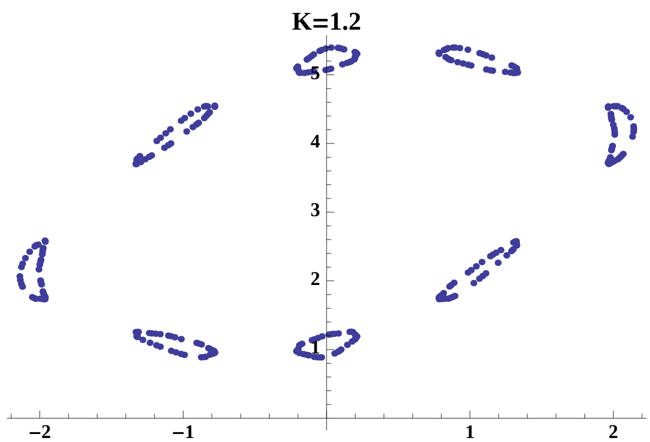

Figure 13.4:

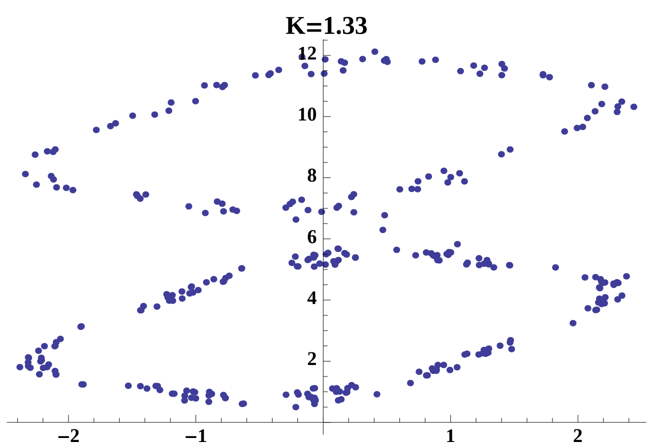

Figure 13.5:

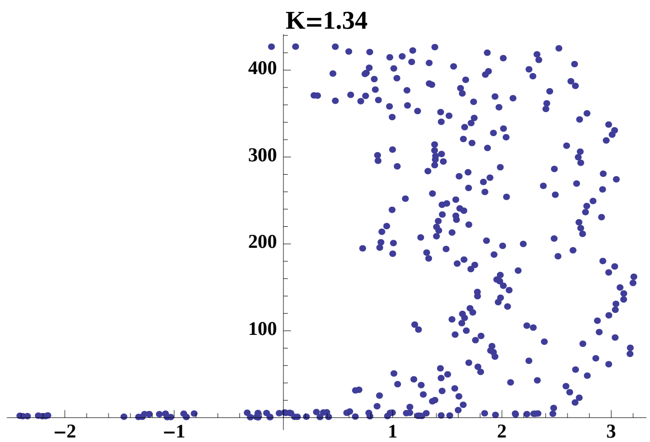

Figure 13.6:

Chapter 14

Dynamics in One Real Variable

It became clear towards the end of the nineteenth century that most systems are not integrable: we will never get a solution in terms of simple functions (trigonometric, elliptic, Painleve' etc.) The focus now shifts to studying statistical properties, such as averages over long time. And to universal properties, whch are independent of the details of the dynamics. It is useful to study the simplest case of dynamics, the iterations of a function of a single real variable, which maps the interval $[0, 1]$ to itself. . Even a simple function like $\mu x(1-x)$ (for a constant μ) leads to chaotic behavior. Only after the advent of digital computers has it become possible to understand this in some detail. But it is not the people who had the biggest of fastest computers that made the important advances: using a hand-held calculator to do numerical experiments, pointed Feigenbaum to patterns which led to a beautiful general theory. The best computer is still the one between your ears.

See S. H. Strogarz *Nonlinear Dynamics and Chaos* Westview Press (1994)

14.1 Maps

A map is just another word for a function $f : X \to X$ that takes some set to itself. Since the domain and range are the same, we can iterate this: given an initial point $x_0 \in X$ we can define an infinite sequence

$$x_0,\ x_1 = f(x_0),\ x_2 = f(x_1), \cdots$$

i.e.,

$$x_{n+1} = f(x_n)$$

This is an **orbit** of f . A **fixed poin**t of f is a point that is mapped to itself:

$$f(x) = x.$$

The orbit of a fixed point is really boring $x, x, x, \cdots$.

A **periodic orbit** satisfies

$$x_{n+k} = x_n$$

for some k. The smallest number for which this is true is called its period. For example, if the orbit of some point is periodic with period two, it will look like

$$x_0, x_1, x_0, x_1, x_0, x_1 \cdots, \quad x_0 \neq x_1$$

Given a function we can define its iterates

$$f_2(x) = f(f(x)), \quad f_3(x) = f(f(f(x))), \quad \cdots$$

A moment's thought will show that a fixed point of one of the iterates $f_k(x)$ is the same thing as a periodic orbit of $f(x)$. For example, if x_0 is not a fixed point of $f(x)$ but is one for f_2, then its orbit is periodic with period two:

$$f(x_0) = x_1 \neq x_0, \quad x_0 = f(f((x_0))$$

gives the orbit

$$x_0, x_1, x_0, x_1 \cdots$$

So far we have not assumed anything abot the space X of the function f. It will often be useful to put some additional conditions such as

- X is a topological space (which allows us to talk of continuous functions) and $f : X \to X$ is a continuous function
- X is a differentiable manifold and $f : X \to X$ is a differentiable function
- $X$is a complex manifold and $f : X \to X$ is a complex analytic (also called **holomorphic**) function
- X carries a Poisson structure $f : X \to X$ is a canonical (also called **symplectic**) transformation

In addition if f is injective (that is, there is only one solution x to the equation $f(x) = y$ for a given y) we can define its inverse. This allows us to extend the definition of an orbit backwards in time:

$$f(x_{-1}) = x_0, \quad x_{-1} = f_{-1}(x_0)$$

etc.

14.2 Doubling Modulo One

Consider the map of tne unit interval $[0,1]$ to itself

$$f(x) = 2x \text{ mod } 1$$

That is, we double the number and keep its fractional part. Clearly, it has a fixed point at $x = 0$.

A simple way to understand this map is to expand every number in base two. We will get a sequence

$$x = 0.a_1 a_2 a_3 \cdots$$

where a_k are either 0 or 1. Doubling $x$is the same as shifting this sequence by one step:

$$x = a_1.a_2 a_3 \cdots$$

Taking modulo one amounts to ignoring the piece to the left of the binary point:

$$f(0.a_1 a_2 a_3 \cdots) = 0.a_2 a_3 a_3 \cdots$$

This map is not injective: two values of x are mapped to the same value $f(x)$ since the information in the first digit is lost. A fixed point occurs when all the entries are equal: either 0 or 1 repeated. But both of these represent 0 modulo one. (Recall that $0.11111.. = 1.0 = 0$ mod1.) So we have just the one fixed point.

An orbit of period two is a sequence

$$0.a_1 a_2 a_1 a_2 a_1 a_2 \cdots$$

We need $a_1 \neq a_2$ so that this is not a fixed point.

Thus we get $0.01010101.. = \frac{1}{3}$ and $0.101010 \cdots = \frac{2}{3}$ which are mapped into each other to get an orbit of period two.Alternatively, they are fixed points of the iterate

$$f_2(x) = f(f(x)) = 2^2 x \text{ mod } 1.$$

More generally we see that

$$f_n(x) = 2^n x \text{ mod } 1$$

which as fixed points at

$$x = \frac{k}{2^n - 1}, \quad k = 1, 2, 3 \cdots.$$

There are a countably infinite number of such points lying on periodic orbits.

Every rational number has a binary expansion that terminates with a repeating sequence. Thus they lie on orbits that are attracted to some periodic orbit.

Since the rational numbers are countable, there are a countably infinite number of such orbits with a predictable behavior. But the irrational numbers in the interval $[0,1]$ which are a much bigger (i.e., uncountably infinite) set, have no pattern at all in their binary expansion: they have a chaotic but deterministic behavior.

14.3 Stability of Fixed Points

If the function is differentiable, we can ask whether a fixed point is stable or not; i.e., whether a small change in initial condition will die out with iterations or not.

Consider again $f:[0,1]\to[0,1]$. Suppose x^* is a fixed point

$$f(x^*)=x^*$$

Under a small change fom the fixed point

$$x=x^*+\epsilon$$

Then

$$f(x^*+\epsilon)=x^*+\epsilon f'(x^*)+\mathrm{O}(\epsilon^2)$$

$$f_2(x^*+\epsilon)=x^*+\epsilon f_2'(x^*)+\mathrm{O}(\epsilon^2)$$

But

$$\frac{d}{dx}f(f(x))=f'(f(x))f'(x)$$

by the chain rule. At a fixed point

$$f_2'(x^*)=f'(x^*)f'(x^*)=\left[f'(x^*)\right]^2$$

More generally

$$f_n'(x^*)=\left[f'(x^*)\right]^n$$

Thus the distance that a point near x^* moves after $n$iterations is

$$|f_n(x^*+\epsilon)-x^*|=\left[f'(x^*)\right]^n\epsilon+\mathrm{O}(\epsilon^2)$$

This will decrease to zero if

$$|f'(x^*)|<1$$

This is the condition for a **stable fixed point.** On the other hand, if

$$|f'(x^*)|>1$$

we have an **unstable fixed point**. The case is

$$|f'(x^*)| = 1$$

is **marginal**: we have to go to higher orders to determine the behavior near x^*.

Example 62. Suppose $f(x) = x^2$ is a map of the closed interval $(0, 1)$. Then 0 is a fixed point. It is stable, as $f'(0) = 0$. But the fixed point at $x = 1$ is unstable as $f'(1) = 2 > 1$. The orbit of 0.99 is driven to zero:

0.99,0.9801,0.960596,0.922745,0.851458,0.72498,0.525596,0.276252,0.076315, 0.00582398,0.0000339187,$1.15048*10^{-9}$,$1.3236*10^{-18}\cdots$

Example 63. The map $f(x) = \frac{1}{4}x(1-x)$ has a stable fixed point at the origin. Where is its other fixed point? Is it stable?

Example 64. But for the case $f(x) = 2x(1 - x)$ the origin is an unstable fixed point. It has another fixed point at 0.5 which is stable. If we start near $x = 0$ we will be driven away from it towards $x = 0.5$. For example, the orbit of $x_0 = 10^{-6}$ is

0.00100, 0.001998, 0.00398802, 0.00794422, 0.0157622, 0.0310276, 0.0601297, 0.113028, 0.200506, 0.320606, 0.435636, 0.491715, 0.499863, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5 $\cdots$

14.4 Unimodal Maps

Differentiable maps of the unit interval $f : [0, 1] \to [0, 1]$ that satisfy the boundary conditions $f(0) = f(1) = 0$ must have a maximum between 0 and 1. If there is exactly one such maximum they are said to be unimodal. Examples are

$$f(x) = 3x(1 - x), \quad b \sin \pi x, \text{ for } b < 1$$

etc. They have an interesting dynamics. It turns out that the special case

$$f(x) = \mu x(1 - x)$$

for various values of the parameter represents this class of maps very well: many of the properties are **universal**: independent of the particular unimodal map chosen so that a simple example will suffice.

An interpretaion is that this map represents the time evolution of the population of some family of animal (rabbits, maybe). Let x be the number of rabbits as a fraction of the maximum number that can be supported by some carrot patch. If x is too close to the maximum value, the number of rabbits in the next generation will be small: many will die of starvation. But if x is small, the next generation will have a number proportional to x, the proportionality constant

μ being a measure of the fertility of rabbits. [1] We will choose $0 < \mu < 4$ so that the maximum value remains less than one: otherwise the value $f(x)$ might be outside the interval.

A stable fixed point would represent a self-sustaining population. There are at most two fixed points:

$$f(x) = x \implies x = 0,\ 1 - \frac{1}{\mu}$$

Note that

$$f'(0) = \mu, \quad f'\left(1 - \frac{1}{\mu}\right) = 2 - \mu$$

14.4.1 One Stable Fixed Point: $1 > \mu > 0$

When $\mu < 1$ the second fixed point is outside the interval, so it would not be allowed. In this case the fixed point at the origin is stable:

$$f'(0) = \mu < 1$$

Every point on the interval is attracted to the origin. The fertility of our rabbits is not high enough to sustain a stable population.

14.4.2 One Stable and One Unstable Fixed Point: $1 < \mu < 3$

When $1 < \mu < 3$ the fixed point at the origin is unstable while that at $1 - \frac{1}{\mu}$ is stable. A small starting population will grow and reach this stable value after some oscillations. For example when $\mu = 2.9$ this stable value is 0.655172. A population that is close to 1 will get mapped to a small value at the next generation and will end up at this fixed point again.

14.4.3 Stable Orbit of period two : $3 < \mu < 1 + \sqrt{6}$

Interesting things start to happen as we increase μ beyond 3. Both fixed points are now unstable, so it is not possible for the orbits to end in a steady state near them. A periodic orbit of period two occurs when μ is slightly larger than 3. That is, there is a solution to

$$f(f(x)) = x$$

within the interval. For example, when $\mu = 3.1$ the solutions are

$$0,\ 0.558014,\ 0.677419,\ 0.764567$$

[1] This is crude.But then, it is only an attempt at quantitative biology and not physics. At least, it is not economics.

The first and the third are the unstable fixed points of f. we found earlier. The other two values mapped into each other by f and form a periodic orbit of period two.

$$f(x_1) = x_2, \quad f(x_2) = x_1,$$

$$x_1 \approx 0.558014, \quad x_2 = 0.764567$$

Is this stable? That amounts to asking whether $|f_2'| < 1$ at these points. If $f(x_1) = x_2, f(x_2) = x_1$

$$f_2'(x_1) = f'(f(x_1))f'(x_1) = f'(x_2)f'(x_1)$$

Clearly $f_2'(x_2)$is equal to the same thing. So we are asking of the product of the derivatives of f along the points on one period is less than one. Numerically, for $\mu = 3.1$ we get $f_2'(x_1) = 0.59$. This means that the orbit of period two is stable. The population of rabbits ends up alternating between these values forever, for most starting values.

We can understand this analytically. To find the fixed points of f_2 we must solve the quartic equation $f_2(x) = x$. We already know two roots of this equation (since 0, $1 - \frac{1}{\mu}$ are fixed points of f hence of f_2). So we can divide by these (that is, simplify $\frac{f_2(x)-x}{x\left(x-1+\frac{1}{\mu}\right)}$) and reduce the quartic to a quadratic:

$$1 + \mu - x\mu - x\mu^2 + x^2\mu^2 = 0$$

The roots are at

$$x_{1,2} = \frac{1 + \mu \pm \sqrt{-3 - 2\mu + \mu^2}}{2\mu}$$

As long as $\mu > 3$ these roots are real: we get a periodic orbit of period two. We can calculate the derivative at this fixed point as above:

$$f_2'(x_1) = 4 + 2\mu - \mu^2$$

For the period two orbit to be stable we get the condition (put $f_2'(x_1) = -1$ to get the maximum value for stability)

$$\mu < 1 + \sqrt{6} \approx 3.44949$$

14.4.4 Stable Orbit of period four : $3.44949... < \mu < 3.54409..$

Thus, as μ increases further this orbit will become unstable as well: a period four orbit develops. Numerically, we can show that it is stable till $\mu \approx 3.54409..$

14.4.5 $3.54409.. < \mu < \mu_\infty$

After that there is a stable period eight orbit until $\mu \approx 3.5644...$ And so on. Let μ_n be the value at which a period 2^n orbit first appears. They form a convergent sequence :

$$3,\ 3.44949.\ 3.54409,\ 3.5644,\ 3.568759,\ \rightarrow \mu_\infty \approx 3.568856$$

The rate of convergence is geometric:

$$\lim_{n\to\infty} \frac{\mu_n - \mu_{n-1}}{\mu_{n+1} - \mu_n} \equiv \delta = 4.669..$$

A study of another unimodal map such as $f(x) = \mu \sin \pi x$ (restricted to the unit interval) will lead to a similar, but numerically different sequence. While studying this, Feigenbaum made a surprising discovery: although the values at which the period doubles depend on the choice of f, **the rate of convergence is universal :** it is always the same mysterious number 4.669... This reminded him of similar universality in the theory of critical phenomena (such as when a gas turns into a liquid) which had been explained by Wilson using an esoteric theoretical tool called **renormalization**. Feigenbaum then developed a version of renormalization to explain this universality in chaos, leading to the first quantitative theory of chaos. In particular, the universal ratio above turns out to be the eigenvalue of a linear operator.

Chapter 15

Dynamics in One Complex Variable

For a deeper study, see the book by J. Milnor of the same title.

15.1 The Riemann Sphere

The simplest analytic functions are polynomials. They are analytic everywhere on the complex plane. At infinity a polynomial must have infinite value, except in the trivial case where it is a constant. It is convenient to extend the complex plane by including the point at infinity, $\hat{C} = C \cup \{\infty\}$. A rational function is of the form $R(z) = \frac{P(z)}{Q(z)}$ where $P(z)$ and $Q(z)$ are polynomials (without common factors, because they could be cancelled out). Rational functions can be thought of as analytic functions maping $\hat{C}$ to itself: whenever the denominator vanishes, its value is the new point at infinity we added. For example, the function $\frac{1}{z}$ can now be thought of as an analytic function that maps ∞ to zero and zero to infinity.

An important geometric idea is that $\hat{C}$ can be identified with a sphere. More precisely, there is a coordinate system on S^2 that associates to every point p on it a complex number; place the sphere on the plane so that its South pole touches the origin. Dran a straight line from the north pole to $p \in S^2$; continue it until it reaches the plane at some point $z(p)$. This is the co-ordinate of p. Clearly the co-ordinate of the South pole is zero. The equator is mapped to the unit circle. As we get close to the North Pole, the co-ordinate gets larger: the North pole itself is mapped to the point at infinity.

A moments thought will show that this map is invertible: to every point on $\hat{C}$ there is exactly one point on S^2. So the sphere is nothing but the complex plane with the point at infinity added. This point of view on the sphere is named for Riemann, the founder of complex geometry.

Thus Rational functions are complex analytic maps of the sphere to itself.

The study of the dynamics of such maps is an interesting subject. Even simple functions such as $f(z) = z^2 + c$ (for constant c) lead to quite complicated behavior.

15.2 Mobius Transformations

A rotation takes the sphere to itself in an invertible way: each point has a unique inverse. It preserves the distances between points. In complex geometry, there is a larger class of analytic maps that take the sphere to itself invertibly. The rotations are a subest of these. To determine these we ask for the set of rational functions for which the equation

$$R(z) = w$$

has exactly one solution for each w and moreover $R(z) \neq R(z')$ if $z \neq z'$. This is the same as solving the equation

$$P(z) - wQ(z) = 0.$$

The number of solutions is the degree of the polynomial $P(z) - wQ(z)$. This is the larger of the degrees of $P(z)$or $Q(z)$: which is called the degree of the rational function $R(z)$. Thus for example $\frac{1}{z^2+3}$ has degree two. So we see that invertible maps of the sphere correspond to rational functions of degree one; i.e., $P(z)$ and $Q(z)$ are noth linear functions

$$M(z) = \frac{az+b}{cz+d}$$

But, the numerator and denominator should not be just multiples of each other: then $M(z)$ is a constant. To avoid this, we impose

$$ad - bc \neq 0$$

Exercise 65. Check that $ad - bc = 0$ if and only if $R(z)$ is a constant.

In fact by dividing though by a constant we can even choose

$$ad - bc = 1$$

We are interested in iterating maps so let us ask for the composition of two such maps $M_3(z) = M_1(M_2(z))$. It must also be a ratio of linear functions: invertible functions compose to give invertible functions. Calculate away to check that

$$M_i(z) = \frac{a_i z + b_i}{c_i z + d_i}, \quad i = 1, 2, 3$$

have coeffcients related by matrix multiplication:

$$\begin{pmatrix} a_3 & b_3 \\ c_3 & d_3 \end{pmatrix} = \begin{pmatrix} a_1 & b_1 \\ c_1 & d_1 \end{pmatrix} \begin{pmatrix} a_2 & b_2 \\ c_2 & d_2 \end{pmatrix}$$

Thus Mobius transformations correspond to the group of 2×2 matrices of deteminant one, also called $SL(2, C)$. Rotations correspond to the subgroup of unitary matrices. It turns out that the general Mobius transformation is closely related to a Lorentz transformation in space-time. The set of light-rays (null vectors) passing through the origin in Minkowski space is a sphere. A Lorentz transformation will map one light ray to another: this is just a Mobius transformation of the sphere.

If we ask how the ratio of two components of a two coponent vector transform, we get a Mobius transformation

$$\begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} \mapsto \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}$$

$$\frac{\psi_1}{\psi_2} \equiv z \mapsto \frac{az+b}{cz+d}.$$

15.3 Dynamics of a Mobius Transformation

We can now ask for the effect of iterating a Mobius transformation $M(z) = \frac{az+b}{cz+d}$. Given an initial point z_0, we get an infinite sequence

$$z_n = M(z_{n-1})$$

Using matrices we can convert this iteration into a linear operation: multiplying matrices. This is easier to study than rational functions, which are non-linear.

$$z_n = M_n(z_0)$$

where M_n is the nth matrix power :

$$M_n = \begin{pmatrix} a & b \\ c & d \end{pmatrix}^n$$

The simplest case is a diagonal matrix

$$a = d^{-1}, \quad b = c = 0.$$

Then

$$z_n = a^{2n} z_0$$

Thus, any point is eventually mapped to ∞ if $|a| > 1$; or to zero if $|a| < 1$. If $a = e^{i\theta}$ is of magnitude one, we get a rotation around the origin. In this case, every point (other than 0 or ∞) lies on a periodic orbit of period k if the $\frac{\theta}{\pi}$ is a rational number with denominator k .

Recall that the eigenvalues of a matrix are the roos to its characteristic equation

$$\det \begin{pmatrix} a-\lambda & b \\ c & d-\lambda \end{pmatrix} = 0$$

Since the determinant is one, the roots are inverses of each other

$$\lambda_2 = \lambda_1^{-1}.$$

If the roots are distinct $\lambda_1 \neq \lambda_2$(i.e., if this polynomial is not a perfect square) there is always a basis in which the matrix is diagonal.

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = S \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_1^{-1} \end{pmatrix} S^{-1}$$

for some invertible matrix S. Then powers of this matrix are just

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^n = S \begin{pmatrix} \lambda_1^n & 0 \\ 0 & \lambda_1^{-n} \end{pmatrix} S^{-1}$$

Again, every point is driven to infinity or zero if $|\lambda| \neq 1$. If $|\lambda| = 1$ we again get rotations.

But what is the roots are not distinct? For example

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

In this case the Mobius transformation is a translation

$$z \mapsto z + 1$$

Its powers are

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}^n = \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix}$$

so that

$$z \mapsto z + n$$

Jordan's theorem says that any matrix with coincident roots can be brought to the upper triangular form

$$\begin{pmatrix} 1 & \mu \\ 0 & 1 \end{pmatrix}$$

The diagonal entries are equal because the determinant is one; the case where they are both minus one yields the same Mobius transformation.

To conclude, there are three kinds of Mobius transformations

- elliptic: the roots are distinct and of magnitude one: these are equivalent to rotations around the axis connecting the Poles $z \to e^{2i\theta}z$.
- parabolic: the roots are equal. These are equivalent to translations $z \mapsto z + b$.
- hyperbolic: the roots are distinct and of magnitude different from one: these are equivalent to scaling $z \to a^2 z$

Much more interesting dynamics can follow from iterating rational maps of degree greater than one.

15.4 The map $z \mapsto z^2$

Let us look at the simplest map of degree 2

$$z \mapsto z^2.$$

Clearly, any point in the interior of the unit circle will get mapped to the origin after many iterations. Any point outside the unit circle will go off to infinity. On the unit circle itself, we get the doubling map

$$z = e^{2\pi i x}, \quad x \mapsto 2x \text{ mod } 1$$

We saw that the latter is chaotic: irrational values of x have a completely unpredictable orbit. But these are most of the points on the unit circle. Even the periodic orbits (rational numbers with a denominator that is a power of two) have something wild about them: they are all unstable. The unit circle is the union of such unstable periodic points. This is a signature of chaotic behavior.

Thus the complex plane can be divided into two types of points:. Those not on the unit circle, which get mapped predictably; and those on the unit circle whose dynamics is chaotic. We will see that points on the plane can always be classified into two such complementary subsets. The chaotic points belong to the **Julia set** and the rest belong to the **Fatou set**. But in general, the Julia set is not necessarily something simple like the unit circle.

15.5 The map $z \mapsto z^2 + i$

Often, it is a quite intricate fractal. For example, the Julia set of the map $z \mapsto z^2 + i$ is a "dendrite": a set with many branches each of which are branched again. Why does this happen? The key is that at its fixed point (which is unstable), the derivative is not a real number. So in the neighborhood of the fixed point there are orbits that looks like a spiral. Such a set cannot lie in a smooth submanifold of the plane, other than the entire plane. So, whenever we have an unstable fixed point with a derivative that is not real (and we know that the domain of repulsion is not the whole plane) the Julia set will be some kind of fractal. This is very common.

How do we come up with a definition of when an orbit is chaotic and when it is not? i.e., which points belong to the Julia set and which to the Fatou set? Such a precise definition is needed to make further progress.

We will need a notion of convergence of maps to make sense of this idea. Also, a notion of distance on the complex plane that is invariant under the Mobius transformations.

15.6 Metric on the Sphere

In spherical polar co-ordinates the distance between neighboring points on the sphere is

$$ds^2 = d\theta^2 + \sin^2\theta d\phi^2$$

The stereographic co-ordinates are related to this by

$$z = \cot\frac{\theta}{2}e^{i\phi}$$

Rewritten in these co-ordinates the metric is

$$ds^2 = \frac{4|dz|^2}{(1+z\bar{z})^2}.$$

The distance s between two points z_1, z_2 can be calculated from this by integration along a great circle. For our puposes an equally good notion of distance in the length d of the chord that connects the two points. Some geometry will give you the relation between the two notions:

$$d = 2\sin\frac{s}{2} = \frac{2|z_1 - z_2|}{\sqrt{(1+|z_1|^2)(1+|z_2|^2)}}$$

The advantage of these notions of distance over the more familiar $|z_1 - z_2|$ is that they are invariant under rotations. For example, the point at infinity is at a finite distance.

We only care about the topology defined by the distance, which is the same for L and s: i.e., when one is small the other is also. So any sequence that converges in one will converge in the other. It is a matter of convenience which one we use in a computation. They both satisfy the conditions for a metric:

$$d(z_1, z_2) = d(z_2, z_1), \quad d(z_1, z_2) \leq d(z_1, z_3) + d(z_3, z_2)$$

The latter is the **triangle inequality**.

15.7 Convergence

Given a metric, we can define the notion of a convergent sequence : we say that z_n converges to z if for every ϵ there is an N such that $d(z_n z) < \epsilon$ for $n \geq N$.

This is the easy part. The tricky part is that we will also need to know what it means for a sequence of functions to converge to another function, within some domain $U \subset S^2$. We will say that a sequence of functions $f_n : U \to S^2$ converges to a function $\phi : U \to S^2$if, for every ϵ there is an N **that is independent of** $z \in U$ such that the distances between the images $d(f_n(z), \phi(z)) < \epsilon$ for every $n \geq N$. The key condition is that the values $f_n(z)$ converge to $\phi(z)$ **uniformly** in z: i.e., that the number $N$is independent of z.

15.8 Julia and Fatou Sets

In our application, we will look at the case where $f_n(z)$ are iterations of some analytic function $f$on the Riemann sphere. The limiting function may not be rational, but it must be analytic in some domain $U \subset S^2$. If the sequence of iterates f_n contains a subsequence that converges locally uniformly to some analytic function ϕ , we can eventually replace the dynamics of f by that of ϕ. When this happens the dynamics is not chaotic. So we define the **Fatou set** of f to be the largest set on which the iterates f_n have a subsequence that converges locally uniformly to an analytic function ϕ. The complement of the Fatou set is the **Julia set**. On it, the dynamics is chaotic.

Here are some theorems about the Julia set that helps to explain what it is. The proofs are in Milnor's book.

Theorem 66. *Let J be the Julia set of of some analytic function $f : S^2 \to S^2$.Then, $z \in J$, if and only if its image $f(z) \in J$ as well.*

Thus, the Julia set is invariant under the dynamics: it explains the asymptotic behavior of f. That is, the Julia set of f and those of its iterates f_n are the same.

Theorem 67. *Every stable periodic orbit is in the Fatou set. However, every unstable periodic orbit is in the Julia set.*

You see here the essential dichotomy between stable and unstable fixed orbits. We can say more:

Theorem 68. *The basin of attraction of every stable periodic point is in the Fatou set.*

Theorem 69. *The Julia set is the closure of the set of unstable periodic orbits.*

It is useful to have a characterization of the Julia set that allows us to compute it.

Theorem 70. *If z_0 is in the Julia set J of some analytic function f, then the set of preimages of z_0 is everywhere dense in J.*

That is, the Julia set is well-approximated by the set

$$\left\{z \in S^2 : f_n(z) = z_0,\ n \geq 1\right\}$$

where f_n is the nth iterate of f. So to compute a Julia set, we must first find one point z_0 on it. (For example, pick an unstable fixed point .). Then we solve the equation $f(z) = z_0$ to find the first set of preimages. If $f(z)$ is a rational function this amounts to solving some polynomial equation, which has some finite number of solutions. Then we find all the preimages of these points and so on. The number of points grows exponentially. If we plot them, we get an increasingly accurate picture of the Julia set. The point in doing this "backwards evolution" is that it is stable. Precisely because f is unstable near z_0, its inverse is stable: we can reliably calculate the solutions to $f(z) = z_0$ since the errors in z_0 will be damped out in the solutions.

In practice, it might make sense to pick just one pre-image at random at each step. This way, we can search in depth and not be overwhelmed by the exponential growth of pre-images. Other tricks to make nice pictures of Julia sets are described in an appendix of Milnor's book.

Chapter 16

Stability and Curvature

We saw that the geodesics on a Riemannian manifold are an example of a hamiltonian system. Conversely, using the Maupertuis principle, the orbits of any conservative mechanical system without velocity dependent forces can be thought of as geodesics. The curvature of a Riemannian manifold describes the change of geodesics under infinitesimal perturbations. Negative curvature implies an instability and positive curvature a stability. This dynamical notion of stability of orbits is analogous to, but different from, the static stability of an equilibrium point. A system with bounded orbits and negative curvature is chaotic. An example is the system of geodesics on a Riemann surface of genus greater than or equal to two.

16.1 Geodesic Deviation

Recall the geodesic equation of Riemannian geometry

$$\ddot{x}^i + \Gamma^i_{jk}\dot{x}^j\dot{x}^k = 0$$

$$\Gamma^i_{jk} = \frac{1}{2}g^{il}\left[\partial_j g_{lk} + \partial_k g_{lj} - \partial_l g_{jk}\right]$$

Here $\dot{x} = \frac{dx^i}{dt}$ and the "time" t has the geometric meaning of the distance measured along the geodesic (arc length).

Suppose we consider an infinitesimally close geodesic, $x^i(t) + u^i(t)$ for some small u^i. This quantity u^i is a vector field defined along the geodesic: it connects a point in the original geodesic to one on the new geodesic nearby. By taking the infinitesimal variation

$$\ddot{u}^i + \left[u^l\partial_l\Gamma^i_{jk}\right]\dot{x}^j\dot{x}^k + 2\Gamma^i_{jk}\dot{x}^j\dot{u}^k = 0$$

We must rewrite this equation in terms of the covariant derivative of u so that its invariance under change of co-ordinates is made explicit.

$$\frac{Du^i}{dt} = \frac{du^i}{dt} + \Gamma^i_{jk}\dot{x}^j u^k$$

$$\frac{D^2u^i}{dt^2} = \frac{d}{dt}\left[\frac{du^i}{dt} + \Gamma^i_{jk}\dot{x}^j u^k\right] + \Gamma^i_{lm}\dot{x}^l\left[\frac{du^m}{dt} + \Gamma^m_{jk}\dot{x}^j u^k\right]$$

Expanding,

$$\frac{D^2u^i}{dt^2} = \ddot{u}^i + \dot{x}^l\left[\partial_l\Gamma^i_{jk}\right]\dot{x}^j u^k + \Gamma^i_{jk}\ddot{x}^j u^k + 2\Gamma^i_{jk}\dot{x}^j\dot{u}^k + \Gamma^i_{lm}\Gamma^m_{jk}\dot{x}^l\dot{x}^j u^k$$

We then use the geodesic equation to eliminate $\ddot{x}^i$ and its variation to eliminate $\ddot{u}^i$.

$$\frac{D^2u^i}{dt^2} = -\left[u^l\partial_l\Gamma^i_{jk}\right]\dot{x}^j\dot{x}^k + \dot{x}^l\left[\partial_l\Gamma^i_{jk}\right]\dot{x}^j u^k - \Gamma^i_{jk}u^k\left[\Gamma^j_{lm}\dot{x}^l\dot{x}^m\right] + \Gamma^i_{lm}\Gamma^m_{jk}\dot{x}^l\dot{x}^j u^k$$

Collect all the different terms into the **geodesic deviation** equation of Jacobi

$$\frac{D^2u^i}{dt^2} + R^i_{jkl}u^j\dot{x}^k\dot{x}^l = 0$$

where

$$R^i_{jkl} = \partial_j\Gamma^i_{kl} - \partial_k\Gamma^i_{jl} + \Gamma^i_{jm}\Gamma^m_{kl} - \Gamma^i_{km}\Gamma^m_{jl}$$

is the **Riemann tensor** or **curvature tensor**. It describes how nearby geodesics deviate from each other. It is obvious that

$$R^i_{jkl} = -R^i_{kjl}$$

If we lower the index and place it **as the third index** (sorry, this is the convention)

$$R_{jkml} = R^i_{jkl}g_{im}$$

it has additional symmetries

$$R_{jklm} = -R_{jkml}$$

$$R_{jklm} = R_{lmjk}$$

These symmetries are conveniently summarized by the statement that the Riemann tensor is a **bi-quadratic** form. That is,we can define a function of two vectors

$$R(u,v) = R_{jklm}u^j v^k u^l v^m$$

which is quadratic in each argument

$$R(\lambda u, \lambda v) = \lambda^2 R(u, v)$$

and such that

$$R(u, v) = R(v, u), \quad R(u, u) = 0.$$

Conversely, any such function of a pair of vectors defines a tensor with the above symmety properties of the Riemann tensor.

We can now derive an equation for the distance between nearby geodesics:

$$\frac{1}{2} \frac{d^2 |u(t)|^2}{dt^2} = \left| \frac{Du}{dt} \right|^2 - R(\dot{x}, u).$$

Suppose the initial conditions for the Jacobi equation are

$$u(0) = 0, \quad \frac{Du^i}{dt}(0) = w.$$

That is, we consider two geodesics starting at the same point but with slightly different initial velocities.

Then

$$|u(t)|^2 = t^2 |w|^2 - \frac{t^3}{3} R(w, \dot{x}) + \mathrm{O}(t^4).$$

The first term is the rate of divergence of geodesics in Euclidean space. If $R(w, \dot{x}) > 0$, the geodesics will come closer together than they would have in flat space.

If $R(w, \dot{x}) < 0$, the geodesic with tangent vector $\dot{x}$ is unstable with respect to an infinitesimal perturbation in the direction w: the points on it will move farther apart with time. This basic principle is behind many deep theorems of Riemannan geometry.

If the curvature at every point and with respect to any pair of vectors is negative, every geodesic has to move away from every other. If in addition, the diameter of the space is finite, the only way they can continue to move away from each other is to produce chaos: small changes in theinitial conditions lead to unpredictable behavior over long time. Just such a situation is created when we study a Riemann surface of genus greater than one.

16.2 The Laplace Operator

We saw that the geodesics are a hamiltonian system with hamiltonian

$$H = \frac{1}{2} g^{ij} p_i p_j$$

and Poisson brackets

$$\{p_k, x^j\} = \delta_k^j$$

etc. Thus the corresponding quantum theory will have wave functions that are square integrable functions of x on which momentum is represented as a differential opeartor

$$p_k = -i\partial_k$$

The hamiltonian them becomes a second order derivative operator. The second differentiation must be a covariant derivative in order to be independent of co-ordinates. Thus

$$H\psi = -\frac{1}{2}g^{ij}D_i\partial_j\psi$$

This is (up to a factor or a half) the Laplace operator of a Riemannian manifold. Its eigenvalues are important quantities in geometry: we see that they have the meaning as the energy levels of a quantum particle moving on the Riemannian manifold. Thus we should expect that there is a semi-classical formula that approximately relates the lengths of geodesics (the action of classical trajectories) to the eigenvalues of the laplacian (the quantum energy levels). Weyl derived such asymptotics in a famous paper "Can we hear the shape of a drum"? The eigenvalues of the Laplacian also has the meaning of the natural frequencies of vibration in accoustics. He was asking if from the eigenvalues we can reconstruct the underlying geometry. It turns out that we can more or less do this, also there are distinct (non-isometric) Riemannian manifolds with the same spectrum for the laplacian.

A modern version of such an asymptotic formula is due to Gutzwiller. Selberg had derived many years ago a deep formula for the special case of a space of constant curvature (symmetric space) relating the lengths of geodesics to eigenvalues of the Laplacian. See D. A. Hejhal, *The Selberg Trace Formula for* $PSL_2(R)$, Springer NY 1976.

Chapter 17

Fluid Mechanics

17.1 Euler's Equation for an Ideal Fluid

17.2 Lie Algebra of Vector Fields

17.3 Euler Flow as Geodesics

17.4 Arnold's Computation of the Curvture of the Euler Flow

17.5 Dissipative Fluid Flow: Navier Stokes Equations