CLASSICAL MECHANICS AND SYMPLECTIC GEOMETRY

Maxim Jeffs

Version: August 4, 2021

CONTENTS

1 NEWTON'S LAWS OF MOTION

Use the Force, Luke.

Star Wars: A New Hope

Firstly, what is classical mechanics? Classical mechanics is that part of physics that describes the motion of large-scale bodies (much larger than the Planck length) moving slowly (much slower than the speed of light). The paradigm example is the motion of the celestial bodies; this is not only the oldest preoccupation of science, but also has very important practical applications to navigation which continue to the present day. We shall have much to say about this example in this class.

More than this, classical mechanics is a complete mathematical theory of the universe as it appears at such scales, with a rich mathematical structure; it has been studied by almost every great mathematician from Euler to Hilbert. Moreover, a deeper understanding of the structure of classical mechanics reveals unexpected analogies that clarify the otherwise mystifying formalism of quantum mechanics: we shall see the classical analogs of anticommuting variables, the Schrödinger equation, and even the uncertainty principle! Similarly, one can see how even Newtonian gravity can be formulated in terms of 'bending' of space and time. This course will study these structures in a modern way, in the language of *symplectic geometry* with special emphasis on *geometry* and *symmetry*.

We shall mostly be using the beautiful book *Mathematical Methods of Classical Mechanics* by V. Arnol'd. While excellent for giving physical motivation, it is sadly very out of date in many places, and the ideas can be greatly clarified using modern terminology and notation. Be sure you have the second edition, since the most interesting material is lumped in the appendices at the end. We shall also be making reference to the notes by da Silva: all further administrative information may be found on the webpage.

We will begin our discussion with some general physical principles, supported by experimental investigation, and take these as *axioms* for our mathematical study. We shall use these to derive Newton's Laws and the rest of 'college physics', so that everyone is on the same page. Throughout, our guiding principles will always be geometry and symmetry; let us see how these are applied.

FIRST PRINCIPLES

Let us first review some vague *experimentally valid* statements (in appropriate physical regimes), which we will take as *axioms* when building our model below.

Geometry Principle: The universe is a 4-dimensional 'flat' space.

Galilean Relativity: There exist **inertial reference frames** in which all laws of physics are the same; any coordinate system in uniform rectilinear motion with respect to an inertial one is itself inertial. Moreover, inertial reference frames with *any* given relative speed may be found.

For example, a train moving along the tracks is (approximately) an inertial reference frame. So is another train moving along the track at a different speed (uniform rectilinear motion).

Newton's Principle: Knowing all the positions and velocities of a system of particles at a given time is *possible* and uniquely determines their motion for all of time.

AFFINE SPACES

Another of the guiding principles of our discussions will be *general covariance*, the idea that formulations of physical laws should not depend on our choice of coordinate systems, these being artificial structures we place on the world as observers. Our universe does not sit inside some larger space from which it inherits coordinates, it must be treated *intrinsically*. We will have much more to say about this idea later.

Roughly, an *affine space* is a vector space with no uniquely specified origin. This is supposed to express the fact that there is no distinguished 'center of the universe'. More formally, an **affine space** is a *set* A with an *action* of a vector space on the left, such that **translation** at every point is a bijection of the underlying set with the vector space. We can produce in an obvious way an affine space from any vector space and any chosen point in that vector space.

We shall reformulate the first principle as: *the universe is a 4-dimensional affine space* $\mathbb{A}^4$. We say that **time** is a linear map $t : \mathbb{A}^4 \to \mathbb{A}^1$. This allows us to say when events are **simultaneous** and talk about **time intervals**. We define a **spatial slice** to be an affine space $\mathbb{A}^3$ given by the simultaneous events. The universe will come equipped with a *positive definite distance function* on each spatial slice $\mathbb{A}^3$, that is, a bilinear function $d : \mathbb{A}^3 \times \mathbb{A}^3 \to \mathbb{R}$ satisfying $d(a, a) = 0$, and $d(a, b) > 0$ if $a \neq b$. Note that the different spatial slices cannot be canonically identified, and we may only talk about the distance between events that are simultaneous. This will be how we formulate Galileo's principle mathematically: let us see how this is done.

In order to work with such a space, we will have to make it less abstract by choosing a **coordinate system**, that is, a linear identification of $\mathbb{A}^4$ with $\mathbb{R}^3 \times \mathbb{R}$ with time given by projection to the second factor, and the canonical metric on $\mathbb{R}^3$. This is always possible to do, but *not uniquely*. Different such identifications differ by **Galilean transformations** of $\mathbb{A}^4$; we say that these different coordinate systems are in **uniform rectilinear motion**, and one such choice of coordinates specifies the collection of inertial reference frames in our universe.

The objects we consider will all be *point-like* (imagine elementary particles, or planets in space); in a coordinate system they will be represented by a smooth function $x : \mathbb{R} \to \mathbb{R}^3$ called the **world-line**. The **velocity** and **acceleration** of this particle will be defined as the first and second derivatives of x respectively.

The preceding discussion is problematic for many reasons. In order to discuss properties of our point-like object, we had to choose a coordinate system. The formalism of affine spaces is also awkward for this reason: we need to make a non-canonical choice of coordinates in order to do anything, against our principle of general covariance. The language of *manifolds* (to be discussed later in this course) will remedy these deficiencies.

NEWTON'S LAWS

For now we work in a single coordinate system. The precise form of Newton's principle will be that there exists a function $F : \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R}$ so that the world-line of the particle satisfies the differential equation:

$$\boxed{m\ddot{x} = F(x, \dot{x}, t)}$$

for some constant m. Then the existence and uniqueness theorem for ODE will imply that our model satisfies Newton's principle (at least for F a sufficiently nice function). In the case of more than one particle, F may be a function of all their positions and velocities, call them $x_i, \dot{x}_i$.

Point-like objects possess an intrinsic scalar quantity called the **inertial mass**: it is defined to be the proportionality constant in the above equation (i.e. the ratio of the masses of two point-like objects is given by the ratios of their accelerations subject to a given F). We shall encounter a different type of mass later on, called the *gravitational mass*. There is no reason in classical physics to believe that these two masses should be equal.

Note that this equation above is the *definition* of F, called the **force**: it can be determined experimentally (up to a constant) using a *test particle* and measuring the accelerations.

The first of Galileo's principles says that subjecting world-lines of a system to the same Galilean transformation should give world-lines of the same system. In other words, solutions of Netwon's equation must be invariant under Galilean transformations. In the spirit of our guiding ideas above, let's use this to prove some of Newton's Laws.

Example 1. One example of a Galilean transformation is *time-translation*: but invariance under such a transformation **does not imply** that F must be independent of t (despite what is written on p. 9 of Arnol'd's book). We shall assume for the rest of this section that the force F has no explicit dependence on time.

Example 2. Translation invariance of physics also implies *homogeneity of space-time*. Therefore the force F may only depend on $x_i - x_j$ and $\dot{x}_i - \dot{x}_j$, the *relative* positions and veloicities of the point-like objects.

Example 3. The invariance of physics under spatial rotation implies the *isotropy* of space: there is no preferred direction in the universe.

Remark 1. Note that all of these principles apply only to the universe considered *as a whole*. Systems that can be modelled in this way are called **closed systems** or **isolated systems**. However, for example, systems that have external influences they do not affect can be modelled by the use of a time-dependent force on a closed system without violating Galileo's principles.

Newton's First Law: A body at rest will remain at rest, unless acted upon by an external force.

Proof. This will follow from the symmetry arguments above: when our closed system consists of a single point-like object, the force F cannot depend on any of $\dot{x}$ or x. Therefore it must be a constant vector F (independent of the time); the only vector invariant under all rotations is the zero vector. Hence $F \equiv 0$ and so $\ddot{x} = 0$. ∎

Newton's Third Law: Every action has an equal and opposite reaction.

Proving this will be the first problem on this week's problem set.

EXAMPLES

The following are some foundational examples of mechanical systems, also historically the first to be discovered and studied.

Example 4. **Galileo's Law**: A point-like object in free-fall near the surface of the earth obeys the equation:

$$\ddot{x} = -g$$

where x is the coordinate giving the height of the body above the surface of the earth and g is a universal constant. In other words, $F = mg$ is the *gravitational force*.

Example 5. **Newton's Law of gravitation**: suppose x denotes the spatial position of an object near another point-like object of gravitational mass M at the origin; then Newton's Law of gravitation says that:

$$\ddot{x} = -\frac{GMm}{|x|^2}$$

where m is the gravitational mass of the object, and G is a universal constant.

Example 6. **Hooke's Law**: suppose x is the coordinate describing the position from equilibrium of a point-mass attached to a spring. Then Hooke's Law says that:

$$\ddot{x} = -k^2 x$$

where k is some constant associated with the spring. This example is sometimes referred to as **simple harmonic motion**, and we shall return to it frequently.

2 CONSERVATION LAWS

> *Jacobi noted, as mathematics' most fascinating property, that in it one and the same function controls both the presentations of a whole number as a sum of four squares and the real movement of a pendulum.*
>
> Vladimir Arnol'd, *On Teaching Mathematics*

ENERGY

For a wordline $x : \mathbb{R} \to \mathbb{R}^3$ of a point-like object, we define the **kinetic energy** by $K = \frac{1}{2}|\dot{x}|^2$. The precise motivation for this definition will be clear later: intuitively this definition is supposed to capture the amount of energy that the object has as a consequence of its motion. We also define the **momentum** p to be $m\dot{x}$: this is really a terrible definition, as we shall see shortly; we'll have much more to say about momentum soon.

From Newton's second law we know that the motion of this body is determined by a vector field $F : \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R} \to \mathbb{R}$. Given a worldline $x : \mathbb{R} \to \mathbb{R}^3$, we define the **work** along this path to be given by the line integral

$$W = \int_\gamma F \cdot \mathrm{d}x = \int_{t_0}^{t_1} F(x(t), \dot{x}(t), t) \cdot \dot{x}(t)\mathrm{d}t$$

We say that the force is **conservative** is the work W is independent of the choice of path, or equivalently, if the work is zero around any closed loop. This definition should be familiar from multivariable calculus, as should be the following theorem:

THEOREM 1. *A force field F is conservative if and only if there exists a potential $U : \mathbb{R}^3 \to \mathbb{R}$ such that $F = -\nabla U$.*

Here we don't allow the force F to depend on the velocity $\dot{x}$ or time t.

When F is conservative, we call the function U the *potential energy*. When F is conservative, then the potential energy is defined (up to addition of a global constant) by the work taken to move a particle from a given reference point. The sign is purely conventional, chosen so that force points from higher potential energy to lower potential energy: objects fall, rather than float.

Exercise: Look back at the forces described in the previous lecture: which ones were conservative? What were their potential energies? Examples of conservative forces in nature include gravitational and electrostatic forces (but not magnetic ones!)

We call the sum $E = K + U$ the **energy** of the object. The following is almost a tautology:

THEOREM 2. *(Conservation of Energy) If F is a conservative force independent of time, then energy is conserved, that is, E is constant in t.*

Proof. I'd like to take a moment for you to try and prove this first. The proof is very simple:

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \frac{\mathrm{d}K}{\mathrm{d}t} + \frac{\mathrm{d}U}{\mathrm{d}t} = \dot{x} \cdot \ddot{x} - \frac{\mathrm{d}}{\mathrm{d}t}\int_{t_0}^{t} F(x(t))\mathrm{d}t \cdot \dot{x}(t) = \dot{x} \cdot (\ddot{x} - F(x(t))) = 0$$

by Newton's second law. ∎

Despite the almost trivial proof, this conservation law has very non-trivial physical consequences! In fact, it is equivalent to Newton's second law (at least for conservative forces). Let's see how this conservation principle can follow from symmetry considerations:

PROPOSITION 1. *If a field F is invariant under all linear motions of $\mathbb{R}^3$ fixing the origin, then F is conservative.*

Exercise: prove this! In particular, this implies that the gravitational forces described before are conservative.

ANGULAR MOTION

We get another conservation law from invariance under rotation: this is the **angular momentum**, defined by $L = x \times p$. Really this is the angular momentum about $0 \in \mathbb{R}^3$: here $\times$ is the cross product in $\mathbb{R}^3$, so angular momentum has three components, which are the angular momenta around the three axes in $\mathbb{R}^3$ passing through zero.

PROPOSITION 2. *If a field F is invariant under all linear motions of $\mathbb{R}^3$ fixing the origin, then L is conserved.*

Proof. We differentiate:

$$\frac{\mathrm{d}L}{\mathrm{d}t} = m\dot{x} \times \dot{x} + mx \times \ddot{x} = mx \times \ddot{x}$$

which is zero, since x and $\ddot{x}$ are parallel. ■

We also define the **torque** of a force F acting on the body x by $\tau = x \times F$. Then in angular coordinates, Newton's Law takes the form:

PROPOSITION 3. *(Angular Newton's Law) The rate of change of angular momentum is equal to the torque:*

$$\frac{\mathrm{d}L}{\mathrm{d}t} = \tau$$

Note importantly that this is *different* from Newton's Law in Cartesian coordinates: this is in some sense a deficiency of Newtonian mechanics, that the linear and angular momenta seem to be different species of quantities. For example, in angular coordinates in the plane we may write the single component of angular momentum as $L = r^2\dot{\theta}$. Note that this is *not* the same as the expression for the linear momentum. Therefore the proposition gives exactly Kepler's Law: for an orbiting body, the area swept out per unit time is constant. Historically, this preceded the development of Newtonian mechanics and this derivation provided early evidence for its validity. We shall see in the presentations that we can use this conservation law to reduce the orbital motion problem to a 1-dimensional problem.

SYSTEMS OF POINTS

Up until now, we have only treated the case of a single isolated point-like body acted upon by external forces, which break the symmetries of the situation. often, these external forces arise from *interactions* with other bodies. We'll see how we can treat systems of points in the same way, which we'll need for our study of rigid bodies. We'll do this by separating out the *external forces* and the *interaction forces*.

So suppose we have a system of bodies $x_i : \mathbb{R} \to \mathbb{R}^3$ all satisfying Newton's second law $m_i\ddot{x}_i = F_i$ for some forces F_i. We define an **interaction force** between two bodies i, j to be one satisfying Newton's third law, that is, we have $F_i = -F_j$. Then for each F_i, we write it as a sum $F_i^{ext} + \sum_{j\neq i}^{n} F_{ij}^{int}$ of interaction forces

F_{ij}^{int} with other bodies and external forces F_i^{ext}, which are all the forces left after subtracting the interaction forces.

We define the **total momentum** of this system of points by

$$P = \sum_{i=1}^{n} m_i \dot{x}_i$$

and similarly for the other quantities above: we sum over the associated quantities for each of the individual point-like bodies. These become simpler if we make the definition of the **center of mass**:

$$x_c = \frac{\sum_{i=1}^{n} m_i x_i}{\sum_{i=1}^{n} m_i}$$

then we have

PROPOSITION 4. *The total momentum of a system of points is the same as that of a point-like body with mass* $\sum_i m_i$ *at the center of mass* x_c.

The proof is left as an exercise. This means we can treat composite bodies much like point-like bodies for the purposes of mechanics: it allows us to simplify problems by separating out 'bulk' motion of the system from other motion we might be interested in. Furthermore, now we can formulate versions of the principles above for systems of bodies.

THEOREM 3. *(Newton's Second Law) The rate of change of total momentum is equal to the sum of external forces*

$$\frac{\mathrm{d}P}{\mathrm{d}t} = \sum_{i=1}^{n} F_i^{ext}$$

Hence, in a **closed system**, one with no external forces, momentum is conserved.

So, we've seen a fairly general form of Newton's laws, given a complete description of the motion of all bodies: what's there left to say about classical mechanics? Let's consider how we actually use these to solve problems in mechanics. Many of you have probably done this in high school: write down all of the forces acting on all of the bodies; you get a system of simultaneous 2nd-order ordinary differential equations for x_i as a function of t, which you then try to solve. This is hard for two reasons: firstly, the forces are *vectors*: writing them all down quickly becomes tedious. Secondly, even when you have all of the equations, you're very unlikely to ever be able to solve these differential equations explicitly, unless there is some kind of special symmetry. When there's no such special symmetry, we'll still want to say something qualitative about the behaviour of the system, even when we can't obtain actual solutions. The *analytical mechanics* we'll look at next will help us with all of these problems.

3 LAGRANGIAN MECHANICS

The world is the best of all possible worlds, and everything in it is a necessary evil.

F. H. Bradley

I've never really seen a very convincing motivation for the origin of the principle of least action (also called *Hamilton's principle*), so I'll just dive in to the mathematics and we'll see soon why exactly we're doing this. Originally, this was introduced by Maupertuis, who seemed to have been vaguely inspired by the philosophy of Leibniz that ours is the 'best of all possible worlds'. At the very least, it captures the idea that physical motion seems to have a certain kind of 'efficiency'. I can also mumble a fairly unconvincing derivation from path integrals in quantum mechanics if you like.

Given a point-like particle, define the **Lagrangian** to be

$$L = K - U$$

Yes, that's right, the kinetic *minus* the potential energy of the particle. We then define the **action** to be the integral of the Lagrangian over time

$$S = \int_{t_0}^{t_1} L(\dot{x}(t), x(t), t)\mathrm{d}t$$

Note that the Lagrangian here is regarded as a function of three variables, the position, the velocity, and time. Now here's the statement:

THEOREM 4. ***(Principle of Least Action)*** *If the trajectory followed by the point-like particle between $x(t_0)$ and $x(t_1)$ is one that 'extremizes' the action S over all possible smooth paths $y : \mathbb{R} \to \mathbb{R}^3$ with the same endpoints ($y(t_0) = x(t_0)$ and $y(t_1) = x(t_1)$), then it satisfies Newton's Laws of Motion.*

First, we'll need to unpack what this actually means mathematically, and see why 'extremizes' is in quotation marks.

THE CALCULUS OF VARIATIONS

The statement above is some kind of optimization problem, but probably not one of a kind you've ever seen before: instead of minimizing a function over Euclidean space, we're minimizing a *functional*, the action, over an infinite-dimensional space of smooth functions of a certain kind. Just like we can use calculus to give a *necessary* criterion for a solution of the former problem, we can use a kind of calculus, called the *calculus of variations* to give a *necessary* condition for the latter. There's actually a formal mathematical sense in which they are the same: we can just take the 'derivative' of S and set it equal to zero. Rather than developing the whole theory of calculus on infinite-dimensional spaces, we'll take a more *ad hoc* approach:

THEOREM 5. *If a smooth function $x : \mathbb{R} \to \mathbb{R}$ extremizes the functional S over all possible smooth functions $y : \mathbb{R} \to \mathbb{R}$ with the same endpoints: $y(t_0) = x(t_0)$ and $y(t_1) = x(t_1)$, then it must satisfy the* ***Euler-Lagrange equation****:*

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{x}}\right) = \frac{\partial L}{\partial x}$$

A few words about this equation first, since it can be somewhat confusing. Here we regard L simply as a function of three variables $L(a, b, t)$, differentiate with respect to these variables, then substitute in $a = \dot{x}, b = x$, yielding a 2nd-order *ordinary* differential equation for x as a function of t: we'll see an example in a minute. First, the proof.

Proof. So suppose x is such an extremizer: for any smooth function $z : \mathbb{R} \to \mathbb{R}$ with $z(t_0) = 0$ and $z(t_1) = 0$ the function $x + sz$ has the same endpoints. Since S has an extremal value among such functions at $s = 0$, by the usual calculus of one variable, we should therefore have:

$$\left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} S(x + sz) = 0$$

Let's expand what this means by taking the derivative inside the integral:

$$\left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} \int_{t_0}^{t_1} L(\dot{x} + s\dot{z}, x + sz, t)\mathrm{d}t = \int_{t_0}^{t_1} \left(\frac{\partial L}{\partial \dot{x}}\dot{z} + \frac{\partial L}{\partial x}z\right)\mathrm{d}t$$

where we've treated L as an abstract function of three variables as above. Now let's integrate the first term by parts! We get

$$\int_{t_0}^{t_1} \left(\frac{\partial L}{\partial x} - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{x}}\right)\right) z(t)\mathrm{d}t + \left[\frac{\partial L}{\partial \dot{x}}z\right]_{t=t_0}^{t=t_1}$$

Now the second term here is zero by our choice: $z(t_0) = z(t_1) = 0$. So let's take a closer look at the first term: what's inside the brackets depends only on x, not on z. In particular, we must have

$$\int_{t_0}^{t_1} \left(\frac{\partial L}{\partial x} - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{x}}\right)\right) z(t)\mathrm{d}t = 0$$

for *any* choice of smooth function $z : \mathbb{R} \to \mathbb{R}$ satisfying the given boundary conditions. In this case, we actually have

LEMMA 1. ***(Fundamental Lemma of the Calculus of Variations)*** *Suppose $f : \mathbb{R} \to \mathbb{R}$ is a smooth function, and that*

$$\int_{t_0}^{t_1} f(t)g(t)\mathrm{d}t = 0$$

for every smooth function $g : \mathbb{R} \to \mathbb{R}$ with $g(t_0) = g(t_1) = 0$. Then we must have $f \equiv 0$ on $[t_0, t_1]$.

The proof of this Lemma is left as an exercise! Using this Lemma we can therefore conclude that x must satisfy the equation

$$\frac{\partial L}{\partial x} - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{x}}\right) = 0$$

which is exactly the Euler-Lagrange equation. ■

Note that in this proof, we've never appealed to the specific form of the function L: the Euler-Lagrange equation works just as well for any such smooth function we would have chosen. We'll see another such example of a variational principle in physics in the problems: *Fermat's principle of least time* in optics.

However, we should note that this proof shows the Euler-Lagrange equation is a *necessary*, though not sufficient condition for the existence of an extremizer. In fact, the path x could also have been a *stationary point* of the functional S and we would have also reached the same conclusion. Physicists usually ignore this point, though careful physicists sometimes refer instead to the 'stationary action principle'. This is not some mathematician's quibble: this distinction has genuinely important physical consequences. We'll see a specific example of a non-extremizing stationary point with Fermat's principle.

Sufficient conditions for the existence of a minimizer are provided by certain *convexity conditions* on L:

THEOREM 6. ***(Sufficiency)*** *If $L(\dot{x}, x, t)$ is convex in $(\dot{x}, x)$ for every value of t, then every solution to the Euler-Lagrange equation is in fact a minimizer of the action S.*

Exercise: prove this theorem! It's not difficult (see the book by Evans p.453 if you get stuck). There's a more complicated hypothesis in the multivariable case (*polyconvexity*).

This condition in fact often satisfied when L has a term given by kinetic energy, though there are some important physical exceptions to this (such as wave propagation in anisotropic media).

Now here are some questions for you to think about:

- What would happen in the above derivation if we replaced x by a function $x : \mathbb{R} \to \mathbb{R}^n$?
- What would happen in the derivation if L were a function of a collection of different smooth functions x_i and their derivatives $\dot{x}_i$?
- What would happen if instead we had expressed the functional $x : \mathbb{R} \to \mathbb{R}^n$ in polar coordinates? In a different set of coordinates?

Answer: for a collection of $q_i : \mathbb{R} \to \mathbb{R}^n$ the Euler-Lagrange equations become the system of vector ODE

$$\nabla_{q_i} L = \frac{\mathrm{d}}{\mathrm{d}t} \left(\nabla_{\dot{q}_i} L \right)$$

one for each i, where ∇_z denotes the gradient in the z-variables. Notice that these equations have the same form regardless of which coordinates we chose to use to express q_i!

BACK TO MECHANICS

That's enough abstraction for now: let's see the Euler-Lagrange equations in action. Let's take $L = K - U$ and prove the theorem stated at the start of the lecture

Proof. (of Theorem 1) So let $L = \frac{1}{2} m |\dot{x}|^2 - U(x, \dot{x})$: then we calculate:

$$\nabla_{\dot{x}} L = m\dot{x}$$

which is exactly the momentum p; we also have

$$\nabla_x L = \nabla U = -F$$

the force, by definition of the potential. Hence the Euler-Lagrange equations become

$$\frac{\mathrm{d}}{\mathrm{d}t} \left(\nabla_{\dot{q}_i} L \right) - \nabla_{q_i} L = m\ddot{x} - F = 0$$

which is precisely Newton's second law: $F = m\ddot{x}$. ■

Hence we see that from the principle of least action we can recover Newtonian mechanics! (and sometimes vice-versa). But what have we really gained from doing this? In some sense, nothing: when using the least action formulation, we have to translate things back into a differential equation in order to be able to solve things. Really, when physicists write down a Lagrangian, they really want the Euler-Lagrange equation, not the action principle associated to it that is only sometimes equivalent. Differential equations are *local*, whereas action principles are not: they require you to know the endpoint of the motion before you even begin, and for this reason violate important physical principles. Apart from more advanced reasons to prefer Lagrangians, such as the ease of *quantizing* Lagrangian systems using path integrals, there are some advantages however in classical physics.

The first is that L is a scalar, not a vector quantity like F and is therefore very easy to write down, and is easy to transform between coordinate systems: you'll see in the problem set how easy this makes solving otherwise impossible-seeming mechanics problems. Similarly, because L is a scalar, the Euler-Lagrange equations as noted take the same form in whatever coordinate system we wish to use. This is a manifestation of the principle of *general covariance* mentioned earlier. As an exercise, re-derive Newton's Laws in angular coordinates that we saw earlier.

Finally, the Lagrangian formulation makes the relation between symmetries and conservation laws particularly apparent. A simple example is:

COROLLARY 1. ***(Simple Noether's Theorem)*** *If the Lagrangian L does not depend on the variable q, then there is a conserved quantity given by*

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

Proof. This is almost trivial: if L does not depend on q, then $\frac{\partial L}{\partial q} = 0$ and hence

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}}\right) = 0$$

■

In general, we call the quantity $\frac{\partial L}{\partial \dot{q}}$ the **conjugate momentum** of q. Observe that in the Cartesian case above, this is exactly the usual momentum p (and if you did the exercise, you'll see that the conjugate momentum of θ is exactly the angular momentum). Hence we recover both of the conservation of momentum laws we saw in the first two lectures as consequences of this corollary: if $F = 0$, then the linear momenta are conserved, and if F does not depend on θ, then the angular momenta are conserved.

NOETHER'S THEOREM

Noether's theorem provides an important generalization of the Corollary above: if the Lagrangian L doesn't depend on the variable q then this is a kind of *symmetry*, one that is *continuous*. A **continuous symmetry** means that there are smooth functions $q(s), \dot{q}(s)$ such that

$$\frac{\partial}{\partial s} L(q(s), \dot{q}(s), t) = 0$$

this is, the Lagrangian is unchanged as we move along this path. Remember that we continue to consider L as just a function of three variables. An example of such a continuous symmetry is $\theta(s) = 2\pi s$, which is the symmetry associated to L being independent of θ. Sometimes, however, rather than our symmetries being global, like in Corollary 1, they are symmetries only locally or along a particular direction. To make this formal, we call a pair of smooth functions $q(s), \dot{q}(s)$ an **infinitesimal symmetry** if

$$\left.\frac{\partial}{\partial s}\right|_{s=0} L(q(s), \dot{q}(s), t) = 0$$

where now we only require that the derivative be zero at the point $s = 0$. In particular, every continuous symmetry gives an infinitesimal symmetry also; this hence gives us many examples of infinitesimal symmetries. Infinitesimal symmetries play an important role in quantum field theory and we will have more to say about them later in the semester. Associated to an infinitesimal symmetry, we have the **Noether charge** or **conserved current**:

$$Q = \frac{\partial L}{\partial \dot{q}} \cdot \left.\frac{\partial q}{\partial s}\right|_{s=0}$$

which is no longer a function of s. Note the similarity with the quantity in Corollary 1. This allows us to state Noether's theorem:

THEOREM 7. ***(Noether's Theorem)*** *The Noether charge associated to an infinitesimal symmetry is conserved.*

Proof. You could try this for yourself! Let's just calculate:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}} \cdot \frac{\partial q}{\partial s}\right) = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}}\right) \cdot \frac{\partial q}{\partial s} + \left(\frac{\partial L}{\partial \dot{q}}\right) \cdot \frac{\partial \dot{q}}{\partial s}$$

and by the Euler-Lagrange equation this is equal to:

$$\frac{\partial L}{\partial q} \cdot \frac{\partial q}{\partial s} + \frac{\partial L}{\partial \dot{q}} \cdot \frac{\partial \dot{q}}{\partial s} = \frac{\partial L}{\partial s}$$

by the chain rule. Evaluating this at $s = 0$ therefore gives:

$$0 = \left.\frac{\partial L}{\partial s}\right|_{s=0} = \left.\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}} \cdot \frac{\partial q}{\partial s}\right)\right|_{s=0} = \frac{\mathrm{d}Q}{\mathrm{d}t}$$

because Q came from an infinitesimal symmetry. Hence Q is conserved. ■

Note that this entire section carries over verbatim to the multiple-variable case where q and $\dot{q}$ are instead vector-valued. We will have *much much* more to say about this theorem in the future.

4 LEGENDRE TRANSFORM, HAMILTONIAN MECHANICS

'In almost all textbooks, even the best, this principle is presented so that it is impossible to understand' (Jacobi, Lectures on Dynamics). I have chosen not to break with tradition.

Vladimir Arnol'd, *Mathematical Methods of Classical Mechanics*

It has taken me over fifty years to understand the derivation of Hamilton's equations.

Saunders Mac Lane, *Mathematics, Form and Function*

This time we'll discuss how we can transform the Euler-Lagrange equations into a simpler system of first-order ODE, with a much more manifest symmetry, by passing to the *dual problem* using the Legendre transform. There's a more conceptual way of reformulating this transformation which we'll return to later on. This transformation is notoriously difficult to understand (see above), so don't worry if it doesn't make complete sense the first time around.

THE LEGENDRE TRANSFORM

In general, the Legendre transform is a kind of *convex duality*: we say that a smooth function $f : \mathbb{R}^n \to \mathbb{R}$ is **convex** if

$$\frac{\partial^2 f}{\partial x_i \partial x_j}$$

is everywhere positive definite. In one variable, this means that $f'' > 0$ everywhere. These functions arise all the time in optimization problems, since they have unique minima (when they exist). The Legendre transform arises in optimization theory as the solution of a certain dual optimization problem: given the line $y = p \cdot x$ with slope given by $p \in (\mathbb{R}^n)^*$ a linear functional in the dual vector space, find the point x where the vertical distance between the graph $y = f(x)$ and the line $y = p \cdot x$ is maximized. Since f is a convex function, for every such p there is a unique such x (when it exists!) and so we define the **Legendre transform** f^* as a function of $p \in (\mathbb{R}^n)^*$ to be the vertical distance at this point x.

A more analytic way to describe this is by looking at the objective function

$$F(x, p) = p \cdot x - f(x)$$

which is maximized when $\nabla_x F = 0$: this means exactly that $p = \nabla_x f \cdot$. Since $\det D^2 f(x) \neq 0$ everywhere, we can solve for x in terms of p and define

$$f^*(p) = F(x(p), p)$$

This p should be thought of as a 'dual' variable: it gives the slope of the linear functional. Also, $\nabla_x f$ should really be thought of as a linear functional, so that $p = \nabla_x f$ is really a dual variable: we'll have more to say about this later. Moreover, we can observe that the derivative of the Legendre transform is given by:

$$\frac{\partial f^*}{\partial p_i} = x_i + \sum_{j=1}^{n} p_j \frac{\partial x_j}{\partial p_i} - \sum_{j=1}^{n} \frac{\partial f}{\partial x_j} \frac{\partial x_j}{\partial p_i} = x_i$$

that is, we have the symmetric identities $\nabla_x f = p$ and $\nabla_p f^* = x$, which should be clear from the geometric definition given before. That this Legendre transform is a form of *duality* is confirmed by:

THEOREM 8. *The Legendre transform f^* of a convex function f is itself convex, and $(f^*)^*$ on $((\mathbb{R}^n)^*)^* \cong \mathbb{R}^n$ is f itself.*

We won't prove this, but it will be useful context for the next section. It also demonstrates that the Legendre transform in some sense contains the same amount of information as f. Like the calculus of variations, the Legendre transform can also be applied to variational problems in other fields.

HAMILTONIAN MECHANICS

The Euler-Lagrange equations are expressed in terms of a second order ODE for the generalized coordinates q_i as functions of time t. Let's try to re-express these equations in terms of the *conjugate momenta* we saw last time:

$$p_i = \frac{\partial L}{\partial \dot{q}_i}$$

by using the Legendre transform. These p_is are in some sense the 'dual' variables to the $\dot{q}_i$, just like we saw with the linear momentum in Lecture 2. But we want to consider them instead as *independent variables* from the q_i. Now, we know by the implicit function theorem that when

$$\frac{\partial^2 L}{\partial \dot{q}_i \partial \dot{q}_j}$$

is a positive definite matrix at every point, then we can solve for the $\dot{q}_i$ as functions of the p_i and q_i. We generally expect this condition to be satisfied by the kinetic energy. Then we define the **Hamiltonian** to be:

$$H = \sum_{j=1}^{n} p_j \dot{q}_j(p_j) - L(q_i, \dot{q}_i(p_i), t)$$

where this is considered as a function of q_i and p_i by the above. This is exactly the Legendre transform of the Lagrangian L with respect to the variables $\dot{q}_i$, that we saw above.

Now, when we look at the derivatives of H, and apply the Euler-Lagrange equations we see that

$$\frac{\partial H}{\partial q_i} = -\frac{\partial L}{\partial q_i} = -\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}_j}\right) = -\dot{p}_i$$

where this derivative is computed holding p_i constant, so that its possible dependence on q_i is suppressed. Similarly, holding q_i constant we have

$$\frac{\partial H}{\partial p_i} = \dot{q}_i + \sum_{j=1}^{n} p_j \frac{\partial \dot{q}_j}{\partial p_i} - \frac{\partial L}{\partial p_i} = \dot{q}_i + \sum_{j=1}^{n} \frac{\partial \dot{q}_j}{\partial p_i} - \sum_{j=1}^{n} \left(\frac{\partial L}{\partial \dot{q}_j}\frac{\partial \dot{q}_j}{\partial p_i}\right)$$

and applying the definition of the generalized momentum, we get exactly

$$\frac{\partial H}{\partial p_i} = \dot{q}_i$$

which is what we expected from the Legendre transform. Together these form **Hamilton's equations**:

$$\boxed{\frac{\partial H}{\partial p_i} \;=\; \dot{q}_i}$$
$$\boxed{\frac{\partial H}{\partial q_i} = -\dot{p}_i}$$

which are $2n$ first-order ordinary differential equations for q_i, p_i, considered as *independent* variables of time t. Notice the symmetry! The variables q_i and p_i are on 'equal footing': we can for instance swap $(p, q) \mapsto (-q, p)$ and preserve the form of these equations. This symmetry is exactly what will allow us to find a lot of hidden structure in classical mechanics: these equations will be our focus for the rest of this course.

In fact we can go back the other way, to see that these equations are actually *equivalent* to the Euler-Lagrange equations: as we saw above, the Legendre transform of H with respect to the $\dot{q}_i$ is L again, and if Hamilton's equations are satisfied by q_i, p_i, then the Euler-Lagrange equations will be satisfied by the q_i. Exercise!

When q_i are the usual Cartesian coordinates x_i, and we return to the world of mechanics by writing $L = K(\dot{x}) - U(x)$, then we have canonical momentum given by $p_i = m_i \dot{x}_i$ and the Hamiltonian is given by

$$H = \sum_{j=1}^{n} \frac{p_j^2}{m_j} - K(\dot{x}) + U(x) = \sum_{i=1}^{n} \frac{p_j^2}{m_j} - \frac{1}{2} \sum_{i=1}^{n} \frac{p_j^2}{m_j} + U(x) = K + U$$

Therefore, in this case, the Hamiltonian is just the total energy $K+U$! You are warned that this interpretation may not hold in general. In particular, this allows us to prove an important generalization of the conservation of energy:

THEOREM 9. *(Conservation of the Hamiltonian) If H is not a function of t, then the quantity H is conserved under Hamilton's equations.*

Proof. I find this proof particularly beautiful: we just calculate

$$\frac{\mathrm{d}H}{\mathrm{d}t} = \sum_{i=1}^{n} \left(\frac{\partial H}{\partial p_i} \frac{\mathrm{d}p_i}{\mathrm{d}t} + \frac{\partial H}{\partial q_i} \frac{\mathrm{d}q_i}{\mathrm{d}t} \right) + \frac{\partial H}{\partial t} = \sum_{i=1}^{n} (\dot{q}_i \dot{p}_i - \dot{p}_i \dot{q}_i) + 0 = 0$$

Note the difference between the partial and total derivatives. ■

PHASE SPACE

Hamilton's equations have a very natural geometric interpretation in terms of **phase space**: this is the space of all possible pairs (q_i, p_i) of position and momenta. When the motion is unconstrained, this just gives $\mathbb{R}^{2n} = \mathbb{R}^n \times \mathbb{R}^n$, but not in general. For instance, as we saw in the problem session, the phase space of a simple pendulum can be visualized as a cylinder: this a simple example of a *symplectic manifold* with non-trivial topology. We'll have much more to say about constrained mechanics in the next lecture. On this space, Hamilton's equations are first-order: this means that given any point, there is a unique trajectory passing through this point. In fact they can actually be formulated as the flow of a vector field on phase space, as follows.

Given any smooth function $H : \mathbb{R}^{2n} \to \mathbb{R}$ representing a version of the total energy, we have a differential equation given by

$$\dot{z} = -J \nabla H(z)$$

where J is the matrix given in block form by:

$$J = \begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix}$$

which is a more compact form of Hamilton's equations, expressing them as the flow of the **Hamiltonian vector field** $-J\nabla H$: at each point, the particle moves in the direction specified by the direction of the vector field. Notice how the mass has dropped out completely: this was the advantage of using momentum rather than velocity in our phase space. In particular, we can think of Hamilton's equations as some kind of 'gradient flow' where the gradient now is tangent to the level sets of H, so that this gradient flow preserves these level sets.

5 PROBLEMS

FORCE AND ENERGY

Exercise: Look back at the forces described in the first lecture: which ones were conservative? What were their potential energies? Can a force that depends non-trivially on the velocities $\dot{x}$ be conservative?

Exercise: Prove that if a field F is invariant under all linear motions of $\mathbb{R}^3$ fixing the origin, then F is conservative.

(Note that in particular, this implies that the gravitational force from the previous question is conservative.)

Exercise: Prove that the total momentum of a system of particles of masses m_i is the same as that of a point-like body with mass $\sum_i m_i$ at the center of mass x_c.

(This means we can treat composite bodies much like point-like bodies for the purposes of mechanics: it allows us to simplify problems by separating out 'bulk' motion of the system from other motion we might be interested in.)

Exercise: Prove the following generalization of Newton's Second Law: the rate of change of total momentum is equal to the sum of external forces.

(Hence, in a closed system, one with no external forces, momentum is conserved.)

Exercise: Newton's Third Law states that *every action has an equal and opposite reaction*. Formulate this mathematically for an isolated system of two point-like bodies, and prove it using the symmetry principles from the first lecture.

The *principle of similarity* or *unit analysis* is very commonly used to solve problems in physics without having to write down any equations; in this method, one considers how the behaviour of the system changes as one re-scales the various relevant quantities.

Exercise: Consider a simple pendulum consisting of a particle of mass m suspended from a rigid rod of length L near the surface of the Earth. Use the principle of similarity to determine the period of the pendulum.

Exercise: Now use Newton's Laws and the physical principles discussed in lecture 1 to write down the equations obeyed by this pendulum, and solve it when the angle from the original position is small. What is the period?

Exercise: How does the height that an animal can jump depend on its size?

LAGRANGIAN AND HAMILTONIAN MECHANICS

Exercise: Prove that if $L(\dot{x}, x, t)$ is convex in $(\dot{x}, x)$ for every value of t, then every solution to the Euler-Lagrange equation is in fact a minimizer of the action S. (See Evans' *Partial Differential Equations* p.453 if you get stuck.)

Exercise: Now consider the situation of a particle of mass m suspended from a *spring* of length L near the surface of the Earth; use Newton's Laws to derive the equations of motion for the particle. Then use the Lagrangian method to derive the same equations. Which one was easier? What happens to the solutions as the stiffness of the spring tends to ∞?

Exercise: Consider a particle of mass m attached to a spring in the absence of gravity. Use Hooke's Law and the Hamiltonian formalism to calculate the motion of the particle. Draw a picture of the motion in phase space.

Exercise: Use the Euler-Lagrange equations to show that the shortest path between two points is a straight line.

Exercise: Suppose that the Lagrangian L of a system is symmetric under time-translation. What is the associated conserved quantity?

Exercise: Prove that the Legendre transform f^* of a convex function f is itself convex, and $(f^*)^*$ on $((\mathbb{R}^n)^*)^* \cong \mathbb{R}^n$ is f itself.

Exercise: In class we gave a description of the local coordinates on an open subset of a constraint surface X. Suppose two such open subsets intersect: describe how to change from one set of local coordinates to another: first try this in the case of the circle we discussed in class, and then the sphere in $\mathbb{R}^3$; then generalize.

MORE CHALLENGING QUESTIONS

Question: Fermat's principle in optics states that the physical path taken by a ray of light between two points is the one that takes the *least time*. Formulate this as a variational principle, and give an example where the physical path taken by a ray of light actually takes the *greatest time* among nearby paths.

Question: (to think about) Why can we formulate optical and mechanical problems in such a similar way? (this is called the *optical-mechanical analogy*) And for those who have seen some differential geometry: is there a metric one can write down on $\mathbb{R}^n$ so that the shortest path between two points is the physical trajectory followed by a particle?

Question: Consider the Hamiltonian H for a mechanical system with a single generalized coordinate q, and let $S(E)$ be the area enclosed by the curve in phase space given by $H(q) = E$. Show that the period of motion along this curve $H(q) = E$ is given by

$$T = \frac{\mathrm{d}S(E)}{\mathrm{d}E}$$

Now if E_0 is the value of the potential U at a minimum point q_0, show that the period of small oscillations around this critical point is given by

$$\frac{2\pi}{\sqrt{U''(q_0)}}$$

Question: Suppose that H and q are as above; then show that if $U(q) > 0$ everywhere, then solutions to Hamilton's equations exist for all time. Provide an example where this can fail if U is not bounded below.

Question: Suppose $L \subseteq \mathbb{R}^2$ is a simple closed curve in phase space, containing a region of area A; show that there is some Hamiltonian function $H : \mathbb{R}^2 \to \mathbb{R}$ such that the time-1 Hamiltonian flow of H takes L to a circle of area A.

6 CONSTRAINED MECHANICS, SMOOTH MANIFOLDS

I have a Euclidean mind, an earthly mind, and therefore it is not for us to resolve questions that are not of this world.

Dostoyevsky, *The Brothers Karamazov*

We saw something interesting when we talked about the Lagrangian in polar coordinates in the plane. Consider a point-like object of mass m suspended from a rigid rod of length L in the plane, that is free to rotate around the origin. We can describe this system with two polar coordinates, r and θ, the angle the rod makes with the vertical direction. Of course, the variable r here is redundant, since it is constant, always equal to $r = L$ in this case. And this is exactly what we found when we calculated the Euler-Lagrange equations: when we set $\dot{r} = 0$ we obtained only one equation, namely the Euler-Lagrange equation for just the coordinate θ. This is an instance of a more general phenomenon.

We call this kind of example a **holonomic constraint**: given generalized coordinates $q_1, \ldots, q_n \in \mathbb{R}^n$, a holonomic constraint specifies that the motion lies on the subset of $\mathbb{R}^n$ defined by the vanishing of k smooth functions $F_1(q_1, \ldots, q_n), \ldots, F_k(q_1, \ldots, q_n) = 0$, such that the rank of the matrix $\frac{\partial F_k}{\partial q_i}$ is k at every point.

Other examples of holonomic constraints we've seen so far are *rigid bodies*: which are configurations of point masses held the same distance apart by a collection of holonomic constraints. Non-holonomic constraints tend to be more exotic, and we won't discuss them in this course: they involve constraining also the *derivatives* of the q_i, for instance, having a ball roll down a plane without slipping. From now on, a *constraint* will always be holonomic.

How should we think of constraints in the language of Newton's Laws? We should imagine an infinitely strong force that applies to the object whenever it tries to leave the region defined by the constraints. Of course, our formalism so far doesn't allow for infinitely strong forces, so we'll have to define the result as a kind of limit. We'll formalize this as follows.

Let's consider the matrix $\frac{\partial F_k}{\partial q_i}$; since this has full rank k at every point, by the implicit function theorem, we can choose some $q_1, \cdots, q_k$ (without loss of generality) so that on every small enough open set $U \subseteq \mathbb{R}^{n-k}$ there exists a function $f : U \to \mathbb{R}^k$ such that for every i, $F_i(f(q_{k+1}, \ldots, q_n), q_{k+1}, \ldots, q_n) = 0$ for every $(q_{k+1}, \ldots, q_n) \in U$. In other words, we can write the first few coordinates $q_1, \ldots, q_k$ as a smooth function of the $q_{k+1}, \ldots, q_n$ on this small open set. Or that locally the set $X \subseteq \mathbb{R}^n$ defined by $\{F_1 = \cdots = F_k = 0\}$ is simply the graph of $f : U \subseteq \mathbb{R}^{n-k} \to \mathbb{R}^k$. This means that the $q_{k+1}, \ldots, q_n$ locally provide *coordinates* on the set X, since we can smoothly describe any point on X fully by specifying the values of $q_{k+1}, \ldots, q_n$.

To formulate the idea of an infinitely strong force, consider a potential function that is given locally in the coordinates $q_1, \ldots, q_n$ by

$$U(q_1, \ldots, q_n) = K\left((q_1 - f_1)^2 + \cdots + (q_k - f_k)^2\right)$$

for some $K > 0$ constant. We can imagine a potential well will the constraint set X at the bottom, which will get steeper and steeper as we increase K to ∞: this will be our approximation to the infinitely strong constraint force. Let L_K be the Lagrangian with this potential energy, as a function of all of the $q_1, \ldots, q_n$ and let L^* be the *reduced Lagrangian* with only the variables $q_{k+1}, \ldots, q_n$ and the other variables determined by the constraints. Then we have

THEOREM 10. *(Holonomic Constraints) Suppose $q_{1,K}(t), \ldots, q_{n,K}(t)$ is a solution to the Euler-Lagrange equations for L_K with initial conditions on X, tangent to X; then there exists some $\varepsilon > 0$ such that as $K \to \infty$, the solution*

$q_{k+1,K}(t), \ldots, q_{n,K}(t)$ converges in $C^\infty(-\varepsilon, \varepsilon)$ to the solution $q_{k+1}(t), \ldots, q_n(t)$ of the Euler-Lagrange equations for the reduced Lagrangian L^*.

In short, Lagrangian mechanics is *intrinsic* to the constraint set X.

We won't prove this in class, but the idea is to show that the $\dot{q}_{1,K}(t), \ldots, \dot{q}_{k,K}(t)$ converge to zero as $K \to \infty$ using conservation of energy, and then make the observation as we did in the example that in this case the Euler-Lagrange equations for L reduce to those for L^*.

Next let's try and globalize this description: we'll need to introduce the terminology of manifolds.

MANIFOLDS

> *I want to give you some eggs. But if I gave them to you all at once, you would drop them. So, I'm going to make you a basket first.*
>
> John Francis

The theory of manifolds has a double function: it describes both *shape* and *space*. Firstly, we'll see that every holonomic constraint locus X defined above has a natural structure of a smooth manifold. But in some sense all mechanics is constrained mechanics: bodies are constrained to live inside our universe, and we should be able to formulate mechanics intrinsically, without reference to any external space. Then, in the next few lectures we'll discuss how the theory of manifolds forces us to talk about space *intrinsically*, and forces us to do calculus in a way that is coordinate-independent.

DEFINITION 1. *A **smooth manifold** (of **dimension** n) is a set X, along with an **atlas**: a cover by **coordinate charts** (U_i, ϕ_i) for $i \in I$, where U_i is a subset of X and $\phi_i : U_i \to \mathbb{R}^n$ is a bijection with an open subset $\phi_i(U_i)$ of $\mathbb{R}^n$, satisfying the following conditions:*

- *(Compatibility of Charts) If the intersection $U_i \cap U_j \neq \emptyset$ then $\phi_i(U_i \cap U_j) \subseteq \phi_i(U_i)$ and $\phi_j(U_i \cap U_j) \subseteq \phi_j(U_j)$ are open, and the map $\phi_j \circ \phi_i^{-1} : \phi_i(U_i \cap U_j) \to \phi_j(U_i \cap U_j)$ is smooth map between open subsets of $\mathbb{R}^n$ (these are called **transition functions**);*
- *(Covering Property) There is a countable subset $J \subseteq I$ such that $\bigcup_{j \in J} U_j$ is all of X;*

*We call a subset U of X **open** if its image under every ϕ_i is open; we require the further condition that*

- *(Hausdorff property) For every two points $x \neq y$ in X, there exist open sets $U_x, U_y \subseteq X$ so that $x \in U_x, y \in U_y$ and $U_x \cap U_y = \emptyset$.*

In the future we will refer to X simply as a **manifold**.

If you have studied topology before, then you will recognise that this definition gives X the structure of a *topological space*, with the open sets defined as above. With respect to this topology, all of the charts are actually continuous maps by construction. Conversely, if X started off with a topology such that every $U_i \subseteq X$ is open and the ϕ_is are actually homeomorphisms, then this *induced topology* agrees with the original topology on X. We won't bother too much about checking the last two conditions, which say that the topology on X is *second countable* and *Hausdorff*, but they will be important when we come to talk about integration on manifolds later in the course.

Note that we can always add in more charts, so long as they are compatible with the ones we originally have: this will not end up changing the manifold X, so sometimes we will assume that this has been done already and that we have a **maximal atlas**.

Notation: Given a coordinate chart (U, ϕ), we will often denote the components ϕ_i of ϕ by x_i, and denote a point in U simply by $(x_1, \ldots, x_n)$, identifying U with $\phi(U)$. We will often also assume without loss of generality that our chart has been translated so that $\phi(x) = 0$ for some point $x \in U$.

EXAMPLES OF MANIFOLDS

Our first example is almost trivial: any open subset $U \subseteq \mathbb{R}^n$ is a smooth manifold. It has a single chart U with ϕ simply given by the identity map. In particular, $\mathbb{R}^n$ itself is a smooth manifold of dimension n in a natural way. You are warned however that there are also other ways of making $\mathbb{R}^n$ into a manifold, some of which will not be the same! These examples are called 'exotic' and won't appear in this course.

As we saw in our discussion above, all sets X defined by holonomic constraints have a natural structure of a manifold: we call such examples **embedded submanifolds** of $\mathbb{R}^n$, as we saw above. That the transition functions between the coordinates we defined above using the implicit function theorem are smooth is illustrated by the following simple example:

Example 7. (The Circle) Let's consider the example of the circle S^1, the configuration space of the planar pendulum, defined as a set by $x^2 + y^2 - 1 = 0$ in $\mathbb{R}^2$. We define *four* open sets $U_{x\pm} = \{\pm x > 0\}$ and $U_{y\pm} = \{\pm y > 0\}$: the implicit function theorem tells us that on $U_{y\pm}$ we can write y as a function of x: this function is given explicitly by $y = \pm\sqrt{1 - x^2}$: note that since $y \neq 0$ this function is well-defined. Similarly, on $U_{x\pm}$ we can write $x = \pm\sqrt{1 - y^2}$; we can use these functions to give local coordinates in the following way:

- $\phi_{y,+} : U_{y+} \to (-1, 1)$ given by $(x, y) \mapsto x = +\sqrt{1 - y^2}$
- $\phi_{y,-} : U_{y-} \to (-1, 1)$ given by $(x, y) \mapsto x = -\sqrt{1 - y^2}$
- $\phi_{x,+} : U_{x+} \to (-1, 1)$ given by $(x, y) \mapsto y = +\sqrt{1 - x^2}$
- $\phi_{x,-} : U_{x-} \to (-1, 1)$ given by $(x, y) \mapsto y = -\sqrt{1 - x^2}$

One can easily check that these maps are bijections. Let's compute some of these transition functions: on $U_{y+} \cap U_{x+}$, we have a transition function given by $\phi_{x+} \circ \phi_{y+}^{-1} : (-1, 1) \to (-1, 1)$ that takes the form $x \mapsto (x, \sqrt{1 - x^2}) \mapsto \sqrt{1 - x^2}$, which is indeed a smooth map on its domain. A similar argument applies to the other transition functions and hence we see that the circle is a smooth manifold.

One important fact to note is that manifolds have no notion of 'size': the circle of radius 1 is isomorphic as a manifold to the circle of radius 2. This will be an important difference with the theory of symplectic geometry, where *volume* will play an important role.

Actually, it is a theorem of Whitney that all manifolds take the form of embedded submanifolds of $\mathbb{R}^n$, at least locally. But to discuss this, we first need to define a notion of a smooth function between manifolds.

DEFINITION 2. *We say that a map $F : X_1 \to X_2$ between smooth manifolds is **smooth** if for every point $x \in X$, there is a local chart (U, ϕ_1) on X_1 with $x \in U$ and a local chart (V, ϕ_2) on X_2 such that $F(U) \subseteq V$ and $\phi_2 \circ F \circ \phi_1^{-1} : \phi_1(U) \to \phi_2(V)$ is a smooth function (between subsets of $\mathbb{R}^n$!). We call this function $\phi_2 \circ F \circ \phi_1^{-1}$ the **coordinate representation** of the map F.*

Be careful when comparing this to other references! Some books get this definition wrong, by implicitly assuming that F is continuous. This definition actually implies that F is continuous, in the sense that the inverse image of every open subset of X_2 is an open subset of X_1. We say that $F : X_1 \to X_2$ is a **diffeomorphism** or an **isomorphism of manifolds** if it is a smooth map with a smooth inverse.

In particular, when $X_2 = \mathbb{R}$, then it has a single chart given by the identity map, and this definition simplifies to the following:

DEFINITION 3. *A real-valued function $f : X \to \mathbb{R}$ is a **smooth function** if every point $x \in X$, there is a local chart (U, ϕ) on X with $x \in U$ such that $f \circ \phi^{-1} : \phi(U) \to \mathbb{R}$ is a smooth function.*

Example 8. (The Circle Again) There was another way we could have chosen to put coordinates on the circle S^1, namely using angular coordinates. Let's define two sets $U_1 = \{x \neq 1\}$ and $U_2 = \{x \neq -1\}$: on U_1 we have a coordinate function $\phi_1 : U_1 \to (0, 2\pi)$ given by the angle $\theta = \tan^{-1}(y/x)$, well-defined since $x \neq 1$; we have a second coordinate function $\phi_2 : U_2 \to (-\pi, \pi)$ given by $\theta = \tan^{-1}(y/x)$, again similarly well-defined since $x \neq -1$. The transition function between these two charts on $U_1 \cap U_2$ is given by $\mathrm{id} : (0, \pi) \to (0, \pi)$ on the first component ($y > 0$), and by $\theta \mapsto \theta - 2\pi : (\pi, 2\pi) \mapsto (-\pi, 0)$ on the second component ($y < 0$).

Let's now consider the map $f : S^1 \to \mathbb{R}$ given by $(x, y) \mapsto x$, and let's show that this is in fact a smooth function. This will show that the coordinates we introduced above are compatible with these new angular coordinates. Firstly, on chart U_1, the composition $x \circ \phi_1^{-1} : (0, 2\pi) \to \mathbb{R}$ is given by $x = \cos\theta$, which is indeed a smooth function. Similarly, on U_2, the composition $x \circ \phi_2^{-1} : (-\pi, \pi) \to \mathbb{R}$ is given by $x = \cos(\theta - 2\pi)$, also a smooth function. Hence x is a smooth function on all of S^1. These compositions $\cos\theta$ with local coordinates are the *coordinate representations* of the function f, and we will often just write $f(\theta) = \cos\theta$, ignoring the identification that has been made.

Another thing we can do is take **products of manifolds** to get new manifolds: if X_1, X_2 are manifolds, then the Cartesian product of sets $X_1 \times X_2$ has a natural manifold structure: coordinates are given by pairs of coordinates, on sets given by their products. I'll make this a problem on the problem sheet to show that this is a manifold.

For example, the **torus** is the manifold $S^1 \times S^1$. This arises as the configuration space of the double pendulum, but it is also a simple example of a symplectic manifold. There will be a couple of problems about the torus on the problem sheet.

PROJECTIVE SPACES

Let's first consider the 2-sphere S^2, defined as the holonomic constraint manifold $x^2 + y^2 + z^2 = 1$ in $\mathbb{R}^3$. This is the configuration space of a rigid pendulum in 3-space, fixed at the origin. It's also an important example of a symplectic manifold, as we will see. Since this is a holonomic constraint manifold, we could put coordinates on S^2 as above, by defining sets U_x, U_y, U_z: on U_x we have $x \neq 0$ and so we can solve for x as a function of (y, z) via $x = \sqrt{1 - y^2 - z^2}$. This gives a coordinate map $\phi_1 : U_x \to \mathbb{R}^2$ given by $(x, y, z) \mapsto (y, z)$, a bijection onto its image by the theorem, and similarly for the other coordinate sets U_y, U_z. The transition functions between these coordinates are simply given by $\phi_y \circ \phi_x^{-1} : \mathbb{R}^2 \to \mathbb{R}^2$ via $(y, z) \mapsto (\sqrt{1 - y^2 - z^2}, y, z) \mapsto (\sqrt{1 - y^2 - z^2}, z)$, which is indeed a smooth function where it is defined.

However, there's a more convenient way to define coordinates on the 2-sphere that we can write down using complex numbers. Here we're just using the identification of $\mathbb{R}^2$ with $\mathbb{C}$ and using z to denote a complex number $z = x + iy$.

First, let's define the set $\mathbb{P}^1$ to be the set of pairs $(z_1, z_2) \in \mathbb{C}^2$, such that $(z_1, z_2) \neq (0, 0)$, subject to the equivalence relation $(\lambda z_1, \lambda z_2) \sim (z_1, z_2)$ for every $\lambda \in \mathbb{C} \setminus \{0\}$. We write such an equivalence class of pairs by $[z_1 : z_2]$: these are sometimes called **homogeneous coordinates** on the **complex projective line**. As a set, $\mathbb{P}^1 = S^2$, in fact also as a topological space. How do we define coordinates on this space?

Since one of z_1, z_2 is always non-zero, we have a cover of $\mathbb{P}^1$ by two sets, $U_1 = \{z_1 \neq 0\}$ and $U_2 = \{z_2 \neq 0\}$: this is a well-defined subset since $\lambda \neq 0$ in the equivalence relation. Since $z_1 \neq 0$ on U_1, for every equivalence class $[z_1 : z_2]$, we have a representative given by $[z_1/z_1 : z_1/z_2] = [1 : z]$, since $z_1 \neq 0$: here $z \in \mathbb{C}$ may be any complex number. Thus we can define our coordinate function $\phi_1 : U_1 \to \mathbb{C}$ by $[z_1 : z_2] \mapsto z_2/z_1$, which is well-defined and a bijection by the above. Similarly, we have $\phi_2 : U_2 \to \mathbb{C}$ given by $[z_1 : z_2] \mapsto z_1/z_2$. Let's check that these transition functions are indeed smooth: $U_1 \cap U_2$ is the set where $z_1 \neq 0$ and $z_2 \neq 0$: the map $\phi_2 \circ \phi_1^{-1} : \mathbb{C} \setminus \{0\} \to \mathbb{C} \setminus \{0\}$ is given by $z \mapsto [1 : z] \mapsto 1/z$, which is indeed a smooth function away from 0.

How should we think about this construction? We can think of the two sets U_1, U_2 as being given by S^2 minus the north or south pole respectively. The transition function $z \mapsto 1/z$ then tells us to 'glue' these together by flipping the latitude and longitude, exactly like when we consider the sphere.

More generally, we can consider $\mathbb{P}^n$, the **complex projective n-space**: as a set it is given by $\mathbb{C}^{n+1} \setminus \{0\}$, modulo the same equivalence relation $(\lambda z_0, \ldots, \lambda z_n) \sim (z_0, \ldots, z_n)$ for $\lambda \in \mathbb{C} \setminus \{0\}$. Similarly we have homogeneous coordinates written by $[z_0 : \cdots : z_n]$ and we can write down a cover $U_i = \{z_i \neq 0\}$ with coordinate maps given by $[z_0 : \cdots : z_n] \mapsto (z_0/z_i, \ldots, z_n/z_i) \in \mathbb{C}^n$. I'll leave it as an exercise to describe the transition functions and check that they are smooth. Note that this does *not* give us a sphere when $n > 1$, but rather a very important example of a symplectic manifold.

7 THE TANGENT SPACE

Nature is manifold.

Ralph Waldo Emerson

One of the most important features of the theory of manifolds is that it allows us to do *calculus*. To start with, let's first define the notion of a *tangent vector* to a manifold X: if X were a submanifold of $\mathbb{R}^n$, this idea should be intuitively clear. But we can formulate this without reference to any notion of an external space.

DEFINITION 4. *The **tangent space** T_xX of a manifold X at a point $x \in X$ is defined to be the set of **tangent vectors**: equivalence classes $[\gamma]$, where $\gamma : (-\varepsilon, \varepsilon) \to X$ is a smooth map with $\gamma(0) = x$, and two paths γ_1, γ_2 are equivalent if there is some coordinate chart (U, ϕ) with $x \in U$ such that:*

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \phi(\gamma_1(t)) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \phi(\gamma_2(t))$$

as vectors in $\mathbb{R}^n$.

Intuitively, γ is an infinitesimal path along X through x, and we say two such paths are equivalent if they pass through 0 in the same direction, with respect to some coordinate chart.

To see that this is a sane thing to do, as a first example, let's consider the tangent space to $\mathbb{R}^n$ at a point $x \in \mathbb{R}^n$. We claim that $T_p\mathbb{R}^n$ can naturally be identified with $\mathbb{R}^n$: so suppose $v \in \mathbb{R}^n$: then we can find a path $\gamma(t) = p + tv$ such that $\gamma'(0) = v$. Moreover, any $\gamma : (-\varepsilon, \varepsilon) \to \mathbb{R}^n$ with $\gamma(0) = p$ is equivalent to the path $p + t\gamma'(0)$, and so this is a bijection. In particular, this means that $T_p\mathbb{R}^n$ has a natural vector space structure. This paragraph is **very important**: reread it! We will often implicitly identify $T_p\mathbb{R}^n$ with $\mathbb{R}^n$ as vector spaces in what follows and this can quickly get very confusing!

More generally, if $x \in X$ is a point in a manifold, then a local coordinate chart (U, ϕ) around x provides an identification between T_xX and $T_{\phi(x)}\mathbb{R}^n$: given any $v \in T_{\phi(x)}\mathbb{R}^n$, we have the curve $\gamma(t) = \phi^{-1}(\phi(x) + tv)$ in T_xX and by the definition, this is a bijection. Hence T_xX has the structure of a real vector space also.

So far we have only defined the notion of a tangent vector at a single point:

DEFINITION 5. *A **vector field** V on X is a smooth assignment of a tangent vector V_x in T_xX for every $x \in X$. More precisely, for every $x \in X$ and local coordinates (U, ϕ) with $x \in U$, we have an identification of $T_yX \to T_{\phi(y)}\mathbb{R}^n \cong \mathbb{R}^n$ for all $y \in U$ given by ϕ, and the resulting map $V : U \to \mathbb{R}^n$ is required to be smooth.*

These local identifications $T_yX \to T_{\phi(y)}\mathbb{R}^n \cong \mathbb{R}^n$ for all $y \in U$ given by ϕ allow us to write down vector fields on manifolds X simply by writing down a vector field on $\mathbb{R}^n$. You already know how to write down a vector field on a manifold! The only question we need to understand is how the representation of a vector field in coordinates changes if we change our choice of local coordinates: so suppose $x \in U \cap V$ for (U, ϕ) and (V, ψ) local coordinates on X, and suppose $[\gamma] \in T_xX$. Then with respect to ϕ we have one identification of γ with $v \in \mathbb{R}^n$ via $T_xX \to T_{\phi(x)}\mathbb{R}^n \to \mathbb{R}^n$, and with respect to ψ we have another, $w \in \mathbb{R}^n$. Therefore the relation between these is given by:

$$w = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} (\psi \circ \phi^{-1})(\phi(x) + tv))$$

But $\psi \circ \phi^{-1}$ is exactly the transition function, call it F; applying the chain rule therefore gives:

$$w = DF(v)$$

Or in a basis,

$$w_i = \sum_{j=1}^{n} \frac{\partial F^i}{\partial x_j} v_j$$

which is called the **transformation law** for tangent vectors. Sometimes physicists will *define* a tangent vector to be something that transforms in such a way between different coordinate charts. This has the advantage that to specify a vector field on X, we just need to write down a vector field on each $\phi_i(U_i)$ and then check that on overlaps this condition is satisfied. For instance, we can write down a vector field on the circle simply given by $1 \in \mathbb{R}$ in each of the the angular coordinates. Since $DF = \mathrm{id}$, this automatically gives a well-defined vector field on the circle.

A vector field on a manifold X has a naturally associated ordinary differential equation, called the *flow* of the vector field: roughly, we can imagine particles that travel along the directions specified by the vector field at each point. A solution $x : \mathbb{R} \to X$ is a **flow line** of V if it satisfies the equation:

$$x'(t) = V(x(t))$$

where $x'(t)$ is considered as a tangent vector at the point $x(t)$. In coordinates, this is just an autonomous 1st order ODE. The existence and uniqueness theorem for ODE implies that solutions of this equation will exist (at least for short times) and will smoothly depend on the initial condition. This hence allows us to define the *time-t flow* of a vector field, which maps each point $x \in X$ to the result $x(t)$ of solving the ODE above with $x(0) = x$. We summarize this as:

THEOREM 11. *Suppose X is a compact manifold (every sequence has a convergent subsequence) and suppose V is a vector field on X; then there exist some $\varepsilon > 0$ and a smooth map $\Phi : X \times (-\varepsilon, \varepsilon) \to X$ called the* ***flow*** *of the vector field, such that:*

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi(t,x) = V(\Phi(x,t))$$

for all $x \in X$. Moreover, for each fixed t, the map Φ_t is a diffeomorphism of X to itself.

For example, the vector field that rotates around the circle discussed above gives for time t the diffeomorphism of the circle to itself that rotates the circle through an angle t.

COTANGENT VECTORS

The importance of tangent vectors to manifolds is that they allow us to take *derivatives*, to understand how functions on our manifold change in a given direction.

DEFINITION 6. *Suppose $F : X \to Y$ is a smooth map between manifolds: then the* ***derivative*** *of F at x is the linear map $D_xF : T_xX \to T_{F(x)}Y$ which takes a path $\gamma : (-\varepsilon, \varepsilon) \to X$ with $\gamma(0) = x$ in T_xX to the path $F \circ \gamma \in T_{F(x)}Y$.*

The meaning of this should be intuitively clear, but let's check that this actually recovers the original derivative when X, Y are open subsets of $\mathbb{R}^n, \mathbb{R}^m$ respectively. In this case, given $v \in \mathbb{R}^n$, the map $\mathbb{R}^n \cong T_xX \to T_{F(x)}Y \cong \mathbb{R}^m$ is given explicitly by:

$$v \mapsto \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} F(x + tv) = D_vF(x)$$

which is exactly the directional derivative of F in the direction v. In fact, we can always reduce to this case by using the local identifications with open subsets of $\mathbb{R}^n$ given by local coordinates: if $G = \psi \circ F \circ \phi^{-1}$ is the coordinate representation of F, then the map is given in local coordinates exactly by the usual matrix of partial derivatives $DG : \mathbb{R}^n \to \mathbb{R}^m$, at the point x.

Now let's specialize to the case where $f : X \to \mathbb{R}$ is just a smooth real-valued function on a manifold. In this case, for every $x \in X$, we have $d_x f : T_x X \to T_{f(x)}\mathbb{R} \cong \mathbb{R}$. In other words, at every point $d_x f \in (T_x X)^*$, the *dual* of the tangent space.

DEFINITION 7. *If $x \in X$ is a point in a manifold, we write $T_x^* X$ for $(T_x X)^*$ and call it the* ***cotangent space*** *at the point x. Elements of $T_x^* X$ are called* ***cotangent vectors****. In the same way as for vector fields, a smooth assignment α of an element α_x of $T_x^* X$ for every x is called a* ***1-form****-form. If $f : X \to \mathbb{R}$ is a smooth function then the 1-form $\mathrm{d}f = d_x f$ is called the* ***differential*** *of f.*

This definition has the (initially) counterintuitive consequence that the differential of a smooth function at a point is naturally a *covector* rather than a *vector*. For instance, we talked before about how the momentum was naturally a covector: this was because we could write the momentum as $p = d_v L$ for a smooth function L, the Lagrangian. This is important because cotangent vectors transform between coordinates in the *opposite* way to tangent vectors: under the same identifications as above, we have that

$$v_i = \sum_{j=1}^{n} \frac{\partial F^i}{\partial x_j} w_j$$

where v, w have now been switched! This is again sometimes used by physicists as a *definition* of a cotangent vector.

This definition also means that a vector acts as a derivative on smooth functions: given $v \in T_x X$ and a smooth function $f : X \to \mathbb{R}$, we define a value $v(f) \in \mathbb{R}$ by the pairing $d_x f(v)$: we call it the **derivative of f along v**. Moreover, given V a vector field on X and a smooth function $f : X \to \mathbb{R}$, we have smooth function $V(f) : X \to \mathbb{R}$ defined by taking the pairing at every point: $V(f)_x = d_x f(V_x)$. This interpretation of tangent vectors as derivatives motivates the following notation: suppose that (U, ϕ) are local coordinates on X around x; it is common to write the components $\phi^i(x)$ as x_i and assume that $\phi(x) = 0$ (without loss of generality). Then the tangent vector $\gamma(t) = t x_i$ is often written as $\frac{\partial}{\partial x_i}$ since it corresponds to a derivative in the x_i direction in local coordinates. Similarly, the differential of the smooth function x_i is written as $\mathrm{d}x_i$. Then vector fields and 1-forms may be written in local coordinates as:

$$V = \sum_{i=1}^{n} a_i(x) \frac{\partial}{\partial x_i}$$

$$\alpha = \sum_{i=1}^{n} a_i(x) \mathrm{d}x_i$$

where $a_i(x)$ are smooth functions on $\phi(U)$. Given another coordinate chart, (V, ψ), write $y_i = \psi_i$: then the transformation laws for vectors and covectors are given by:

$$\frac{\partial}{\partial y_i} = \sum_{j=1}^{n} \frac{\partial x_j}{\partial y_i} \frac{\partial}{\partial x_j}$$

$$\mathrm{d}y_i = \sum_{j=1}^{n} \frac{\partial y_i}{\partial x_j} \mathrm{d}x_j$$

which are exactly the formulas we obtained earlier. It will often be convenient to write formulas in this notation and elide the distinction between functions on $\phi(U) \subseteq \mathbb{R}^n$ and functions on U.

Let's do a non-trivial example of a 1-form on the circle, denoted $\mathrm{d}\theta$. This notation is misleading, since θ is not a globally defined function on the circle. However, by the transformations above, if we pass to the second coordinate chart, we have $\theta_1 = \theta_2 - 2\pi$ and hence $\mathrm{d}\theta_1 = \mathrm{d}\theta_2$ on the overlap. Hence $\mathrm{d}\theta$ actually gives a well-defined 1-form. This corresponds to the fact that the circle contains non-trivial topological information.

THE TANGENT BUNDLE

There's a more geometric way of talking about vector fields that we'll discuss next: namely, we can organize all of the tangent spaces into a new manifold, called the *tangent bundle*.

DEFINITION 8. *Suppose X is a smooth manifold: define the* ***tangent bundle*** *TX as a set to be the set of pairs (x, v) with $x \in X$ and $v \in T_xX$; define coordinates on TX as follows: suppose that (U, ϕ) is a coordinate chart for X and let T_UX be the pairs (x, v) with $x \in U$; define $\Phi : T_UX \to \phi(U) \times \mathbb{R}^n$ by*

$$\Phi(x, v) = (\phi(x), D_x\phi(v))$$

where $D_x\phi(v) \in T_{\phi(x)}\mathbb{R}^n \cong \mathbb{R}^n$ is our identification with the tangent space above.

We should check that this is indeed a manifold: in fact, we've almost already done so: if $(U, \phi), (V, \psi)$ are two charts and $G = \psi \circ \psi^{-1}$ is the transition function, then we saw that

$$\Psi \circ \Phi^{-1}(x, v) = (G(x), D_aG(v))$$

and since G is a smooth function (by assumption!) so is D_aG. Hence the tangent bundle is indeed a manifold. The reason it is called a *bundle* is because there's a natural map $\pi : TX \to X$, given by $(x, v) \mapsto x$: we can check very easily that this is smooth: in local coordinates it's just given by $(x, v) \mapsto x$ (the same thing!), which is indeed a smooth map. Therefore we can think of TX as a manifold living 'over' X, with the *fiber* T_xX over each point given by a copy of the tangent space at that point, kind of like a hedgehog.

We can do exactly the same thing for the cotangent spaces:

DEFINITION 9. *The* ***cotangent bundle*** *T^*X of a smooth manifold X is given as a set by the pairs (x, α) where $x \in X$ and $\alpha \in T_x^*X$. We define coordinates exactly as above, making this into a smooth manifold with a map $\pi : T_x^*X \to X$.*

Cotangent bundles are *extremely important* for mechanics and symplectic geometry. If X is a holonomic constraint manifold, then T^*X is the **phase space** of X, since all the possible cotangent vectors encode all of the possible momenta, as we saw above. The cotangent bundle of a manifold will always have the structure of a *symplectic manifold* in a very natural way, and will be where we study Hamiltonian mechanics.

Example 9. In fact, in many situations the cotangent bundle is really very concrete: let's revisit a familiar example, $X = S^1$. Then we've see that $D_aG = \mathrm{id}$ for the single transition function, so that actually the cotangent spaces can be globally identified with $\mathbb{R}$: in this case we say that the cotangent bundle is **trivial**. Therefore, we have a diffeomorphism $T^*S^1 \cong S^1 \times \mathbb{R}$, which is precisely the cylinder that we've seen several times.

8 DIFFERENTIAL FORMS

What one fool can do, another can.

Richard Feynman

We finished last time by introducing 1-forms on manifolds and how we can write them in local coordinates and pair them with vectors. But how should we think of a 1-form on a manifold? I like to think of a 1-form as giving some kind of infinitesimal ruler, which measures the *signed length* of a vector along a certain direction.

Now, we want to talk about how we measure area and volume on manifolds. As we've discussed, manifolds don't come with any notion of their *size*: let's see what we need to do to equip our manifolds with extra structure so that we can make such measurements.

Firstly, let's simplify and think about what the properties of the *signed volume* of the parallelogram described by a set of k vectors $v_1, \dots, v_k$ in an abstract real vector space V should be. There are two intuitive properties we should have:

- **Multilinearity**: the volume should be linear in each variable separately: that is: $\mathrm{vol}(v_1, \dots, \lambda v_i + \mu v_i', \dots, v_k) = \lambda \mathrm{vol}(v_1, \dots, v_i, \dots, v_k) + \mu \mathrm{vol}(v_1, \dots, v_i', \dots, v_k)$;
- **Skew-symmetry**: the orientation should change sign whenever two of the vectors are swapped: that is, $\mathrm{vol}(v_1, \dots, v_i, \dots, v_j, \dots, v_k) = -\mathrm{vol}(v_1, \dots, v_j, \dots, v_i, \dots, v_k)$ for $i \neq j$.

We call such maps $\alpha : V^{\times k} \to \mathbb{R}$ satisfying the two conditions above **alternating multilinear maps** or the k**-forms** on V; they form a vector space which we denote by $\bigwedge^k V^*$. Each such k-form on V gives us a way of measuring the signed volume inside the vector space V. Soon, this vector space V will end up being the tangent space T_xX at a point of our manifold X and the form will measure the infinitesimal volume of a set of tangent vectors.

Example 10. Given a basis $e_1, \dots, e_n$ of $\mathbb{R}^n$, and vectors $v_1, \dots, v_k$, then taking the determinant of the projection of the v_i to the k-dimensional subspace spanned by the e_i, that is, $\det(v_i^j)_{1 \leq i,j \leq k}$ where v_i^j is the e_j component of v_i, gives a k-form on $\mathbb{R}^n$.

Our intuition should tell us there's a natural structure of an algebra on the forms $\bigwedge^k V^*$ for all k: given two infinitesimal rulers, we should be able to use them together to produce a measurement of areas. That is, we should have a map $\bigwedge^k V^* \times \bigwedge^\ell V^* \to \bigwedge^{k+\ell} V^*$ which we will call the *wedge product*: this tells us how we convert volume measurements in lower dimensions to ones in higher dimensions. Just as in the previous example, this should be given by some form of determinant operation. We'll begin by defining the wedge product in exactly this way for 1-forms:

DEFINITION 10. *Given 1-forms $\alpha_1, \dots, \alpha_k$ on V, we define their* ***exterior product*** *or* ***wedge product*** *to be the k-form acting on $v_1, \dots, v_k \in V$ by:*

$$(\alpha_1 \wedge \cdots \wedge \alpha_k)(v_1, \dots, v_k) = \det(\alpha_i(v_j))$$

We can indeed verify that this is indeed a k-form from the properties of determinants: they are indeed multilinear and skew-symmetric as functions of their columns.

This wedge product operation we have just defined will clearly be multilinear in the α_is, and also have the property that $\alpha \wedge \alpha = 0$ for any 1-form α: this should make sense since any two vectors that lie along the same ruler cannot have any area.

If we choose a basis e_i for our vector space, we can write x_i for the 1-form given by the dual vector to e_i. Then x_i measures the signed length along the e_i direction, and as in the example above $x_i \wedge x_j$ measures exactly the signed area along of the projection to the e_i, e_j subspace, etc. In fact, these are actually *all* of the k-forms:

PROPOSITION 5. *The vector space $\bigwedge^k V^*$ is spanned by the wedge products of 1-forms. Moreover, there is a basis given by the vectors*

$$x_{i_1} \wedge \cdots \wedge x_{i_k}$$

for $i_1 < \cdots < i_k$.

This proposition isn't to difficult to prove – try it! It says that taking some form of determinant operation is essentially the *only* way to build up a k-form. In particular, we have:

PROPOSITION 6. *Suppose V is an n-dimensional vector space; then every n-form on V is a scalar multiple of $x_1 \wedge \cdots x_n$ where $x_1, \ldots, x_n$ are the dual vectors from a basis of V.*

We call such a form a **volume form** for the obvious reason that it measures volumes in V.

The decomposition result above means that we can define more generally a wedge product of a k-form and an ℓ-form, i.e. a map $\wedge : \bigwedge^k V^* \times \bigwedge^\ell V^* \to \bigwedge^{k+\ell} V^*$ by writing them in a basis consisting of wedge products of 1-forms, and then defining:

$$(\alpha_1 \wedge \cdots \wedge \alpha_k) \wedge (\beta_1 \wedge \cdots \wedge \beta_\ell) \mapsto \alpha_1 \wedge \cdots \wedge \alpha_k \wedge \beta_1 \wedge \cdots \wedge \beta_\ell$$

Then the wedge product of forms automatically satisfies bilinearity and associativity. Moreover, it also has the same antisymmetry of the determinant, that if we swap two of the factors, we get a sign: $\alpha \wedge \beta = (-1)^{k\ell} \beta \wedge \alpha$.

HIC SUNT DRACONES 1. *You are warned that in this and the following sections, there exist different conventions for exterior products and exterior derivatives that are incompatible with the ones used in these notes. Be careful when looking at other references!*

CALCULUS WITH FORMS

Now let's talk about putting these differential forms onto manifolds: it will turn out that they will be exactly the objects that we will end up integrating over manifolds to give us a measure of their volume.

DEFINITION 11. *A **differential k-form** (or sometimes just a k-form) α on a manifold X is a smooth assignment of a k-form $\alpha_x \in \bigwedge^k T_x^* X$ for all $x \in X$. We denote the vector space of differential k-forms on X by $\Omega^k(X)$.*

Let's unpack what it means for α_x to vary smoothly with x; suppose that (U, ϕ) is a local coordinate system on X with $x \in U$, and write $x_i = \phi^i$ for the components of ϕ: we can think of these as smooth functions $U \to \mathbb{R}$. At every $y \in U$, ϕ have an identification $T_y^* X \to T_{\phi(y)} \mathbb{R}^n \cong \mathbb{R}^n$ which is given exactly by taking the basis $d_y x_i \in T_y^* X$. Therefore, by the proposition in the previous section, we have a basis for $\bigwedge^k T_y^* X$ given by

$$\mathrm{d}_y x_{i_1} \wedge \cdots \wedge \mathrm{d}_y x_{i_k}$$

for $i_1 < \cdots < i_k$. Therefore we can write any k-form α on U in coordinates as:

$$\alpha_y = \sum_{i_1 < \cdots < i_k} a_{i_1 \ldots i_k}(y) \mathrm{d}_y x_{i_1} \wedge \cdots \wedge \mathrm{d}_y x_{i_k}$$

and α is smooth precisely when $a_{i_1\dots i_k}(y)$ is a smooth function of y. We will usually write this in shorthand simply as:

$$\alpha = \sum_{i_1 < \cdots < i_k} a_{i_1 \dots i_k} \mathrm{d}x_{i_1} \wedge \cdots \wedge \mathrm{d}x_{i_k}$$

Note that all differential forms on $\mathbb{R}^n$ take precisely this form.

Now, just as we could push tangent vectors forward via the derivative map, the natural thing to do with differential forms is to take the *pullback*:

DEFINITION 12. *Suppose $\alpha \in \Omega^k(Y)$ and suppose $F : X \to Y$ is a smooth map between manifolds; then we define the **pullback** $F^*(\alpha) \in \Omega^k(X)$ of α to be the k-form on X that acts on vectors via:*

$$F^*\alpha_x(v_1, \dots, v_k) = \alpha_{F(x)}(D_xF(v_1), \dots, D_xF(v_k))$$

Let's see how this works in coordinates: suppose that we have a map $F : X \to Y$ between n-dimensional manifolds expressed in local coordinates as $y_i = F_i(x_1, \dots, x_n)$ and a differential n-form

$$\alpha_y = a(y)\mathrm{d}y_1 \wedge \cdots \wedge \mathrm{d}y_n$$

Then $F^*(\alpha)$ will be given by:

$$\begin{aligned} F^*(\alpha)_x(v_1, \dots, v_n) &= a(F(x))\mathrm{d}y_1 \wedge \cdots \wedge \mathrm{d}y_n(D_xF(v_1), \dots, D_xF(v_n)) = (a \circ F)(x) \det(\mathrm{d}y_i(D_xF(v_j))) \\ &= (a \circ F)(x) \det(D_xF_i(v_j)) = (a \circ F)(x) \det\left(\sum_{k=1}^n \frac{\partial F^i}{\partial x_k} v_j^k\right) \\ &= (a \circ F)(x) \det\left(\frac{\partial F^i}{\partial x_k}\right) \det\left(v_j^k\right) \end{aligned}$$

by the multiplicative properties of determinants; this is therefore equal to:

$$F^*(\alpha)_x = (a \circ F)(x) \det(D_xF)\mathrm{d}x_1 \wedge \cdots \wedge \mathrm{d}x_n$$

This formula should be familiar: it is exactly the change of variables rule for multivariable integration, using the *Jacobian determinant*. But, as we have seen, this change-of-variables rule is built into the differential form itself! This is exactly what makes them natural objects to integrate: the integral will naturally be independent of our choice of coordinates.

However, there is one small difference, namely that in the usual change of variables formula for the Riemann integral, the change in the volume factor is actually given by $|\det(D_xF)|$, the *absolute value* of the Jacobian determinant. If $\det(D_xF) > 0$, then we say F is **orientation-preserving**: then the change-of-variables formula will work completely correctly. Thus corresponds to our usual idea of what it means to preserve orientation. This leads us to make the following definition:

DEFINITION 13. *An **orientation** of a manifold X is a choice of charts (U_i, ϕ_i) that cover X and so that the transition maps $\phi_j \circ \phi_i^{-1}$ are orientation-preserving. We say X is **orientable** if such an orientation exists.*

For instance, all of the manifolds we have seen so far have come with an orientation, which is, roughly speaking, a consistent choice of the 'inside' of the manifold. An example of a non-orientable manifold is the infamous *Möbius strip*.

We won't give a definition in these notes of the integral of a differential form, but the rough idea is as follows: for a differential form α supported a single coordinate chart (U, ϕ), we define

$$\int_{(U,\phi)} \alpha = \int_{\mathbb{R}^n} a(x)\mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_n$$

as the usual Riemann integral. Then when we change to another coordinate chart (U, ψ), then we have a change of coordinates $x_i = G_i(y_1, \dots, y_n)$ with $G = \phi \circ \psi^{-1}$ and so we have that

$$\int_{(U,\phi)} \alpha = \int_{\mathbb{R}^n} a(x)\mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_n = \int_{\mathbb{R}^n} (a \circ F)(y)\det(D_yG)\mathrm{d}y_1\mathrm{d}y_2\cdots\mathrm{d}y_n = \int_{(U,\psi)} \alpha$$

since $\det(D_yG) > 0$, and hence this integral is well-defined, that is, independent of our choice of coordinates. We'll summarize this in a theorem:

THEOREM 12. *Given an n-manifold X with an orientation, and α an n-form on X with* ***compact support****, there is a well-defined* ***integral***

$$\int_X \alpha \in \mathbb{R}$$

that is linear in α and satisfies the invariance property that for any diffeomorphism $F : Y \to X$

$$\int_Y F^*(\alpha) = \int_X \alpha$$

DERIVATIVES OF DIFFERENTIAL FORMS

Next, we need to talk about how we *differentiate* a form on a manifold: this is also particularly natural:

DEFINITION 14. *The* ***exterior derivative*** *of a k-form α on X is a $(k+1)$-form $\mathrm{d}\alpha \in \Omega^{k+1}(X)$, defined as follows: if α can be written in a local coordinate chart as*

$$\alpha = \sum_{i_1 < \cdots < i_k} a_{i_1 \dots i_k}(x)\mathrm{d}x_{i_1} \wedge \cdots \wedge \mathrm{d}x_{i_k}$$

then

$$\mathrm{d}\alpha = \sum_{i_1 < \cdots < i_k} \sum_{j=1}^{n} \frac{\partial a_{i_1 \dots i_k}(x)}{\partial x_j} \mathrm{d}x_j \wedge \mathrm{d}x_{i_1} \wedge \cdots \wedge \mathrm{d}x_{i_k}$$

You should check that this is indeed independent of coordinates and hence gives a well-defined map on k-forms. There is a coordinate-independent definition, but it is rarely useful. The idea of the exterior derivative is that it measures some kind of *flux*: this is made precise by *Stokes' theorem* for integrating differential forms (which we will not cover in these notes). In fact, the exterior derivative should already be a familiar operation in $\mathbb{R}^3$:

Example: Let $X = \mathbb{R}^3$ with the usual basis e_1, e_2, e_3: then $\bigwedge^1(\mathbb{R}^3)^*$ and $\bigwedge^2(\mathbb{R}^3)^*$ are 3-dimensional, and using a basis for these spaces allows us to identify $\bigwedge^1(\mathbb{R}^3)^*$ and $\bigwedge^2(\mathbb{R}^3)^*$ with $\mathbb{R}^3$, as well as $\bigwedge^3(\mathbb{R}^3)^*$ with $\mathbb{R}$. Hence we can interpret any 1-form or 2-form on $\mathbb{R}^3$ as a vector field, and any 3-form as a smooth function. With appropriate choices of basis:

- if f is a smooth function on $\mathbb{R}^3$, then $\mathrm{d}f$ corresponds to $\mathrm{grad}(f)$;

- if α is a 1-form on $\mathbb{R}^3$, then $\mathrm{d}\alpha$ corresponds to $\mathrm{curl}(\alpha)$;
- if β is a 2-form on $\mathbb{R}^3$, then $\mathrm{d}\beta$ corresponds to $\mathrm{div}(\beta)$.

You should try and prove this! The exterior derivative also acts compatibly with the wedge product and pullback operations introduced earlier:

PROPOSITION 7. *Let α be a k-form and β be an ℓ-form on Y; let $F : X \to Y$ be a smooth map between manifolds. Then we have*

- $F^*(\mathrm{d}\alpha) = \mathrm{d}F^*(\alpha)$*;*
- $F^*(\alpha \wedge \beta) = F^*(\alpha) \wedge F^*(\beta)$*;*
- $\mathrm{d}(\alpha \wedge \beta) = \mathrm{d}\alpha \wedge \beta + (-1)^k \alpha \wedge \mathrm{d}\beta$*;*
- $\mathrm{d}(\mathrm{d}\alpha) = 0$.

Proof. We'll just prove the last of these identities, you should try to prove the rest. The proof is:

$$\mathrm{d}(\mathrm{d}\alpha) = \sum_{i_1 < \cdots < i_k} \sum_{m=1}^{n} \sum_{j=1}^{n} \frac{a_{i_1 \ldots i_k}(x)}{\partial x_j \partial x_m} \mathrm{d}x_m \wedge \mathrm{d}x_j \wedge \mathrm{d}x_{i_1} \wedge \cdots \wedge \mathrm{d}x_{i_k}$$

and since

$$\frac{a_{i_1 \ldots i_k}(x)}{\partial x_j \partial x_m} = \frac{a_{i_1 \ldots i_k}(x)}{\partial x_m \partial x_j}$$

but $\mathrm{d}x_j \wedge \mathrm{d}x_m = -\mathrm{d}x_m \wedge \mathrm{d}x_j$, this means that the sum must equal zero. ■

This last identity is exactly the generalization of the usual identities $\mathrm{div} \circ \mathrm{curl} = 0$ and $\mathrm{curl} \circ \mathrm{grad} = 0$ from 3-dimensional calculus.

9 PROBLEMS

MANIFOLDS

Exercise: Suppose X_1, X_2 are manifolds: prove that the product set $X_1 \times X_2$ naturally has the structure of a smooth manifold, with smooth maps $\pi_1 : X_1 \times X_2 \to X_1$ and $\pi_2 : X_1 \times X_2 \to X_2$. Moreover, prove that the tangent space $T_{(x,y)}(X_1 \times X_2)$ is isomorphic to $T_x X_1 \times T_y X_2$. Finally, describe these coordinates explicitly in the example of $S^1 \times S^1$, called the *torus*.

Exercise: In the previous problem we introduced the torus as a product manifold. Give a description of the torus as a parametrized surface in $\mathbb{R}^3$, that is, the torus is locally the image of a smooth injective map $r : \mathbb{R}^2 \to \mathbb{R}^3$ with $\partial_x r \times \partial_y r \neq 0$ at every point. Prove that every such parametrized surface has the structure of a smooth manifold.

Exercise: Describe the torus as the result of gluing together the opposite sides of a square. What are the coordinates associated with this description? How do they compare to the coordinates constructed above?

Exercise: Explain why the following two definitions of what it means for a function $f : X \to \mathbb{R}$ on a manifold to be *smooth* are equivalent:

- For every $x \in X$ and for every coordinate chart (U, ϕ) such that $x \in U$, the map $f \circ \phi^{-1} : \phi(U) \to \mathbb{R}$ is smooth;
- For every $x \in X$, there exists some coordinate chart (U, ϕ) such that $x \in U$, the map $f \circ \phi^{-1} : \phi(U) \to \mathbb{R}$ is smooth.

Exercise: The 2-sphere S^2 is defined to be the set $\{x^2 + y^2 + z^2 = 1\} \subseteq \mathbb{R}^3$. Explain how this arises as a holonomic constraint surface in mechanics, and show that is has the structure of a smooth manifold. Also explain how to make S^2 into a manifold using spherical polar coordinates, and show that the resulting manifolds are diffeomorphic.

Exercise: Consider a rigid body in $\mathbb{R}^3$ that does not lie in a single plane. Describe the configuration space of the rigid body as a constraint surface X, and describe it as a set of matrices. Prove that this constraint surface is holonomic, and thus that X is a manifold.

Exercise: Given a smooth map $F : X_1 \to X_2$ between manifolds, and a smooth function $f : X_2 \to \mathbb{R}$, define $F^*(f) : X_1 \to \mathbb{R}$ to be the *pullback* $F^*(f) = f \circ F$. Then show that for any vector field V on X_1, we have the *push-pull formula:* $V(F^*(f)) = DF(V)(f)$.

Exercise: We say a smooth injective map $F : X_1 \to X_2$ of manifolds is an *embedding* or a *submanifold* if DF is injective at every point, and $U \subseteq X_1$ is open if and only if $U = F^{-1}(V)$ for $V \subseteq X_2$ open. Prove that F is an embedding if and only if for every point $x \in X_2$, there exists a coordinate chart (U, ϕ) with $x \in U$ such that $\phi(U \cap F(X_1))$ is equal to $\{x_1 = \cdots = x_{n-k} = 0\} \cap \phi(U)$ for x_i coordinates on $\mathbb{R}^n$ and k the dimension of X_1. Use this to show that all holonomic constraint manifolds X are embedded submanifolds of $\mathbb{R}^n$.

DIFFERENTIAL FORMS

Exercise: Suppose V is an n-dimensional vector space: show that $\bigwedge^n V^*$ is one-dimensional. Given a basis x_i of V, find a basis element for $\bigwedge^n V^*$ expressed in terms of the wedge product of 1-forms.

Exercise: Let α be a k-form and β be an ℓ-form on Y; let $F : X \to Y$ be a smooth map between manifolds. Prove the following identities involving the operations on 1-forms introduced in class:

- $F^*(\mathrm{d}\alpha) = \mathrm{d}F^*(\alpha)$;
- $F^*(\alpha \wedge \beta) = F^*(\alpha) \wedge F^*(\beta)$;
- $\mathrm{d}(\alpha \wedge \beta) = \mathrm{d}\alpha \wedge \beta + (-1)^k \alpha \wedge \mathrm{d}\beta$;
- $\mathrm{d}(\mathrm{d}\alpha) = 0$.

Exercise: Suppose that v, w are vector fields on X and that α is a 1-form: show that $\mathrm{d}\mathcal{L}_v(\alpha) = \mathcal{L}_v \mathrm{d}(\alpha)$ and use this (and other identities from class) to show that $\mathrm{d}\alpha(v, w) = v(\alpha(w)) - w(\alpha(v)) + \alpha([v, w])$.

Exercise: Let $X = \mathbb{R}^3$ with the usual basis e_1, e_2, e_3: show that $\bigwedge^1(\mathbb{R}^3)^*$ and $\bigwedge^2(\mathbb{R}^3)^*$ are 3-dimensional, and give a basis for these spaces. Using such a basis allows us to identify $\bigwedge^1(\mathbb{R}^3)^*$ and $\bigwedge^2(\mathbb{R}^3)^*$ with $\mathbb{R}^3$, as well as $\bigwedge^3(\mathbb{R}^3)^*$ with $\mathbb{R}$, so that we can interpret any 1-form or 2-form on $\mathbb{R}^3$ as a vector field, and any 3-form as a smooth function. Show that with appropriate choices of basis:

- if f is a smooth function on $\mathbb{R}^3$, then $\mathrm{d}f$ corresponds to $\mathrm{grad}(f)$;
- if α is a 1-form on $\mathbb{R}^3$, then $\mathrm{d}\alpha$ corresponds to $\mathrm{curl}(\alpha)$;
- if β is a 2-form on $\mathbb{R}^3$, then $\mathrm{d}\beta$ corresponds to $\mathrm{div}(\beta)$.

Explain the interpretation of the identity $d \circ d = 0$ in terms of calculus in $\mathbb{R}^3$. Given a 3-form, explain how to use these identifications to calculate the integral over a subset of $\mathbb{R}^3$, and explain how to rederive Stokes' theorem for calculus in $\mathbb{R}^3$.

Exercise: Let $X = \mathbb{R}^{2n}$ have coordinates $q_1, \dots, q_n, p_1, \dots, p_n$ and consider the differential form given by

$$\lambda = \sum_{i=1}^{n} p_i \mathrm{d}q_i$$

Prove that:

$$\mathrm{d}\lambda = \sum_{i=1}^{n} \mathrm{d}p_i \wedge \mathrm{d}q_i$$

and that

$$(\mathrm{d}\lambda)^{\wedge n} = n!\ \mathrm{d}p_1 \wedge \dots \wedge \mathrm{d}p_n \wedge \mathrm{d}q_1 \wedge \dots \wedge \mathrm{d}q_n$$

Show that $\mathrm{d}\lambda(v, w)$ is given in the basis $q_1, \dots, q_n, p_1, \dots, p_n$ by the action of the $2n \times 2n$ block matrix:

$$\mathrm{d}\lambda(v, w) = v^T \begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix} w$$

MECHANICS ON MANIFOLDS

Exercise: Define the *cotangent bundle* T^*X of a manifold X to be the set of all pairs (x, α) where $x \in X$ and $\alpha \in T_x^*X$. Prove that the cotangent bundle has the structure of a smooth manifold. Show that T^*S^1 is diffeomorphic to the cylinder $S^1 \times \mathbb{R}$.

Exercise: Use local coordinates on T^*S^1 to write down the Hamiltonian vector field associated with the simple pendulum that we saw in the presentations. Let $\lambda = p\ \mathrm{d}\theta$: show that this defines a 1-form on T^*S^1.

Exercise: Suppose X is a holonomic constraint manifold: explain why the Lagrangian L is a smooth function $TX \to \mathbb{R}$, where TX is the *tangent bundle* of X. Explain why the *vertical* derivative $D_v L$ (i.e. with respect to the $\dot{q}_i$) of the Lagrangian L gives a diffeomorphism $TX \to T^*X$, and explain why the inverse is given by the vertical derivative of the Hamiltonian H, naturally considered as a function $H : T^*X \to \mathbb{R}$.

Exercise: A *Riemannian metric* on a manifold X is a smooth assignment g of an inner product g_x on every tangent space T_xX; formulate what this smooth dependence means and describe the representation of g in local coordinates. Then define the notion of *kinetic energy* on a manifold with a Riemannian metric and, supposing that $U = 0$, explain how physical trajectories are the same as *geodesics* (paths of minimal length) on the manifold X.

Exercise: Suppose $f : X \to \mathbb{R}$ is a smooth function on a compact manifold X with a metric g; define $\nabla_x f \in T_xX$, the *gradient of* f to be the dual of $d_x f \in T_x^*X$ with respect to the inner product g_x. Show that the flow of the vector field $-\nabla f$ takes points of X to the minima of f as $t \to \infty$. Compute $-\nabla f$ when S^2 is the unit sphere in $\mathbb{R}^3$ with Riemannian metric induced from the standard inner product on $\mathbb{R}^3$ and $f : S^2 \to \mathbb{R}$ is the z-coordinate.

Exercise: Suppose now that X is a manifold with a metric g; explain why the isomorphism given by the vertical derivative $D_vL : TX \to T^*X$ of the Lagrangian is exactly the duality between tangent vectors and cotangent vectors given by the metric.

Exercise: Suppose $x : \mathbb{R} \to \mathbb{R}^3$ describes the motion of a particle with potential $U(x)$ in Cartesian coordinates, subject to a holonomic constraint $X = \{G(x) = 0\}$. We define the *constraint force* to be precisely the difference: $R = m\ddot{x} + \frac{\partial U}{\partial x}$. Prove *D'Alembert's principle*, that the work of the constraint force for a *virtual variation*, that is, a vector tangent to the constraint manifold X, is always zero. Conversely, suppose D'Alembert's principle is satisfied: then show that the constraints on the motion are necessarily holonomic. Thus D'Alembert's principle provides an equivalent formulation of the notion of a holonomic constraint.

Exercise: Continuing from the previous problem, we say an *equilibrium point* $x_0 \in X$ is one such that the constant map $x(t) = x_0$ is a physical trajectory. Show that x_0 is an equilibrium point if and only if the force $F = -\frac{\partial U}{\partial x}$ is orthogonal to $T_{x_0}X$. What is the connection to *Lagrange multipliers*? Show that D'Alembert's principle may also be formulated as: every point along x becomes an equilibrium position with respect to the sum of the *inertial force* $-m\ddot{x}$ and the external force F.

Exercise: Prove Noether's Theorem for Lagrangian mechanics on a manifold X: show that the resulting conserved quantity is independent of coordinates.

10 SYMPLECTIC MANIFOLDS

Symplectic (adj.) Placed in or among, or put between, as if ingrained or woven in: specifically noting a bone of the lower jaw of fishes interposed between others.

The Century Dictionary and Cyclopedia

Now that we've talked about manifolds and differential forms, let's review what we've done so far.

We saw that from a mechanical system with holonomic constraints we could naturally produce a manifold X called the **configuration space** or **constraint manifold**. On this manifold, we have some smooth function $U : X \to \mathbb{R}$ called the **potential energy**, which describes the forces acting on the system.

Recall that, given a path $\gamma : \mathbb{R} \to X$, the Lagrangian was a function of both $\gamma(t)$ as well as $\gamma'(t)$, which naturally lives in the tangent space to X. Hence the **Lagrangian** is a smooth function $L : TX \to \mathbb{R}$ on the **tangent bundle** of the configuration space, produced by adding together the **kinetic energy** K and $-U$, pulled back to TX.

Our theorem then said that we could do mechanics intrinsically to X: those paths $\gamma : X \to \mathbb{R}$ such that the action $S = \int L(\gamma(t), \gamma'(t))\mathrm{d}t$ is stationary were described by the Euler-Lagrange equations. We can identify the **cotangent bundle** T^*X with the **phase space**, since the generalized momenta that appeared when we studied the **Legendre transform** were given by

$$p = \mathrm{d}_{\dot{q}} L$$

Because the generalized momenta are defined as a differential of a smooth function L, they are naturally interpreted as a **cotangent vector** in T^*X.

While we could write down the Euler-Lagrange equations in coordinates on TX, there's a more symmetrical and coordinate-free way to write down Hamilton's equations on the cotangent bundle T^*X of the constraint manifold. For this, we'll have to talk about:

SYMPLECTIC MANIFOLDS

Without further ado, here is the long-awaited:

DEFINITION 15. *A* ***symplectic manifold*** *is a smooth $2n$-dimensional manifold X along with a differential 2-form $\omega \in \Omega^2(X)$ called the* ***symplectic form*** *which has the two properties:*

- *ω is* ***closed****: $\mathrm{d}\omega = 0$;*
- *ω is* ***non-degenerate****: if $\omega_x(v, w) = 0$ for all $v \in T_xX$, this implies that $w = 0$.*

Most importantly, a symplectic form allows us to do *Hamiltonian mechanics*:

DEFINITION 16. *Let (X, ω) be a symplectic manifold. A smooth function $H : X \to \mathbb{R}$ is called a* ***Hamiltonian****, and the* ***Hamiltonian vector field*** *X_H of H is defined to be one satisfying the equation:*

$$\boxed{\iota_{X_H}\omega = \mathrm{d}H}$$

Here $\iota_{X_H}\omega = \omega(X_H, \cdot)$ is the 1-form obtained from ω by inserting X_H into the first argument of ω, and we require that it be equal to the 1-form $\mathrm{d}H$. The Hamiltonian dynamics of the Hamiltonian function H are then obtained precisely as the flow of this vector field X_H.

HIC SUNT DRACONES 2. *Sign conventions: this is the most controversial point in symplectic geometry: many authors take instead the convention that $\iota_{X_H}\omega = -\mathrm{d}H$, with the opposite sign in front of the Hamiltonian. The convention I use in these notes follows Arnol'd and McDuff-Salamon, but differs from the one used in da Silva. You have been warned!*

The most important examples of symplectic manifolds are *cotangent bundles*: as discussed above, they exactly represent the phase spaces of mechanical systems.

DEFINITION 17. *Suppose Q is an n-dimensional smooth manifold; let $X = T^*Q$ be the cotangent bundle and $\pi : T^*Q \to Q$ be the projection. We define the **tautological 1-form** λ on X to be:*

$$\lambda_{(q,\alpha)}(v) = \alpha(D_{(q,\alpha)}\pi(v))$$

where $(q, \alpha) \in Q \times T_q^*Q$ and $v \in T_{(q,\alpha)}X$.

To understand this better, let's write this out in local coordinates: we'll imagine that $Q = \mathbb{R}^n$ with coordinates $(q_1, \ldots, q_n)$ and $T^*Q \cong \mathbb{R}^{2n}$ has coordinates $(q_1, \ldots, q_n, p_1, \ldots, p_n)$, such that $\pi : T^*Q \to Q$ is given by $\pi(q_1, \ldots, q_n, p_1, \ldots, p_n) = (q_1, \ldots, q_n)$. Then we have $D\pi : \mathbb{R}^{2n} \to \mathbb{R}^n$ given simply by projection to the first n coordinates, so that:

$$\lambda_{(q,p)}(v_1, \ldots, v_n, v_{n+1}, \ldots, v_{2n}) = \sum_{i=1}^{n} p_i v_i$$

and hence

$$\lambda = \sum_{i=1}^{n} p_i \mathrm{d}q_i$$

DEFINITION 18. *On a cotangent bundle $X = T^*Q$ the **canonical 2-form** ω is defined to be $-\mathrm{d}\lambda$.*

DEFINITION 19. *In general, we say a symplectic manifold is **exact** if the symplectic form ω is the exterior derivative of some 1-form λ, called the **Liouville form**.*

I claim that in fact the canonical 2-form ω will always be a symplectic form. First off, we always have $\mathrm{d}\omega = -\mathrm{d}(\mathrm{d}\lambda) = 0$ since $\mathrm{d} \circ \mathrm{d} = 0$: thus ω is closed. To check that ω is non-degenerate, we can check by working out what ω is in the local coordinates from before: we have

$$\omega = -\mathrm{d}\lambda = -\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial p_i}{\partial p_j} \mathrm{d}p_j \wedge \mathrm{d}q_i = \sum_{i=1}^{n} \mathrm{d}q_i \wedge \mathrm{d}p_i$$

We call this the **canonical symplectic form** on $\mathbb{R}^{2n}$. Since the coefficients of this 2-form are constant, we can just think of it as a matrix acting on $\mathbb{R}^{2n}$, as follows:

$$\omega(e_i, e_j) = \sum_{k=1}^{n} \mathrm{d}q_k \wedge \mathrm{d}p_k(e_i, e_j) = \sum_{k=1}^{n} \det \begin{pmatrix} q_k(e_i) & q_k(e_j) \\ p_k(e_i) & p_k(e_j) \end{pmatrix}$$

so that the entries of this matrix are, in the basis e_i given by $q_1, \ldots, q_n, p_1, \ldots, p_n$:

$$-J = -\begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix}$$

and $\omega(v, w) = -v^T J w$ as a matrix multiplication. In particular, this matrix is *invertible*, so that it has the required non-degeneracy property.

Remark 2. My apologies, but I got this sign wrong in class! You should check for yourself that the signs work out correctly (I didn't!)

Thus we have shown that:

THEOREM 13. *Every cotangent bundle is a symplectic manifold with the canonical symplectic form.*

In particular, $\mathbb{R}^{2n}$ is a symplectic manifold for every n. So is the cylinder T^*S^1 we saw earlier, with symplectic form $\omega = \mathrm{d}\theta \wedge \mathrm{d}p$. But be warned: there can be more than one way to make a cotangent bundle into a symplectic manifold, even for $\mathbb{R}^{2n}$!

Let's now calculate the Hamiltonian flow associated to a smooth function H on the cotangent bundle T^*Q: let's work in the local coordinates above, so that we have:

$$\mathrm{d}H(v) = (\nabla H)^T v = -X_H^T J v$$

In other words, we must have $-X_H^T J = (\nabla H)^T$. Taking the transpose and using the fact that $J^2 = -\mathrm{Id}$, we hence have:

$$X_H = -J\nabla H$$

This was precisely the formula we had for the **symplectic gradient** of the function H! Therefore, the flow of the Hamiltonian vector field of H on the cotangent bundle gives exactly the Hamiltonian dynamics associated to the function H!

Before talking more about mechanics, let's see some more examples of symplectic manifolds that aren't just cotangent bundles.

Example 11. The torus T^2 we saw in Lecture 5 had coordinates (θ_1, θ_2) coming from thinking of T^2 as the product $S^1 \times S^1$. Although neither θ_1 nor θ_2 is a global function, as we saw, their differentials $\mathrm{d}\theta_1$ and $\mathrm{d}\theta_2$ are well-defined global 1-forms. We define $\omega = \mathrm{d}\theta_1 \wedge \mathrm{d}\theta_2$. We can easily see that $\mathrm{d}\omega = 0$ since every 3-form on a 2-manifold is always zero. Furthermore, by exactly the same local calculation as above, we see that ω is non-degenerate. Thus T^2 is a symplectic manifold!

Example 12. Now consider the 2-sphere $S^2 \subseteq \mathbb{R}^3$. If we let θ, r, z be cylindrical polar coordinates on $\mathbb{R}^3$, then these locally define smooth functions on S^2. Just as above, the 2-form $\omega = \mathrm{d}z \wedge \mathrm{d}\theta$ will we a well-defined global 2-form on S^2, and the same arguments will show that S^2 is indeed a symplectic manifold!

Example 13. The next natural class of examples are the *complex projective spaces* along with their **Fubini-Study form**. Actually writing down this symplectic form is quite complicated, so we will reserve it for the problem sheet.

So far all of these examples look surprisingly similar: when we wrote out our symplectic 2-form in coordinates, it always looked the same. A natural question is whether this always happens: is there always some choice of coordinates on a symplectic manifold so that in these coordinates the symplectic form is given by the canonical symplectic form on $\mathbb{R}^{2n}$? The answer to this question is positive, and is given by *Darboux's theorem*. However, first we will prove:

PROPOSITION 8. ***(Linear Darboux's Theorem)*** *Suppose V is a finite-dimensional vector space with an alternating non-degenerate bilinear pairing $\omega : V \times V \to \mathbb{R}$. Then V must be even dimensional, and there exists a* ***symplectic basis*** *of V, that is, one in which ω is given by the matrix J.*

Proof. We'll just sketch the proof here: the argument works by induction. Firstly, let $e_i \in V$ be any non-zero vector: then, by non-degeneracy of ω, there must exist some $e_i' \in V$ such that $\omega(e_i, e_i') = 1$; moreover, e_i' is not a scalar multiple of e_i, since $\omega(e_i, e_i) = -\omega(e_i, e_i) = 0$. Therefore we see that V must have at least two dimensions, and in the basis e_i, e_i', the pairing ω is represented exactly by a 2×2 matrix of the form J. This gives the base case. Let's consider now the symplectic orthogonal: $e_i^\omega = \{v \in V : \omega(v, e_i) = \omega(v, e_i') = 0\}$. By the non-degeneracy of ω, we can show that this is a vector space of dimension $\dim V - 2$, also with a non-degenerate alternating bilinear form. Thus by induction, this too has a symplectic basis. Combining this symplectic basis with e_i, e_i', we are done. ■

COROLLARY 2. *Every symplectic manifold is even-dimensional.*

In order to formulate Darboux's theorem precisely, we need the notion of when two symplectic manifolds are the same. This is given by:

DEFINITION 20. *A* ***symplectomorphism*** *or* ***canonical transformation*** *of symplectic manifolds (X, ω_X), (Y, ω_Y) is a diffeomorphism $F : X \to Y$ such that $F^*(\omega_Y) = \omega_X$.*

DEFINITION 21. *A linear map $A : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ that gives a symplectomorphism of the canonical symplectic form J on $\mathbb{R}^{2n}$, that is:*

$$A^T J A = J$$

is called a ***symplectic transformation****. We denote the group of such transformations by $\mathrm{Sp}(2n)$ and call it the* ***symplectic group****.*

This definition is the origin of the term *symplectic*. Originally, this group was called the *complex group* or the *abelian linear group*; this unfortunate clash of terminology led Weyl in his book *The Classical Groups* to coin the term *symplectic* as a back-formation in Greek of the Latin *complex*. The name has been with us ever since. Amusingly, the only previous use of the word *symplectic* in English was to describe a certain bone found in the jaw of a fish.

Now we can formulate Darboux's theorem precisely:

THEOREM 14. *(**Darboux**) Suppose (X, ω) is a $2n$-dimensional symplectic manifold; then for every $x \in X$ there is a coordinate chart (U, ϕ) with $x \in U$ such that ϕ gives a symplectomorphism between (U, ω) and $\mathbb{R}^{2n}$ with the canonical symplectic form.*

Roughly speaking, Darboux's theorem says that there is no *local* symplectic geometry: all symplectic manifolds locally look the same! This is in stark contrast to Riemannian geometry, where there is a whole suite of different local invariants given by various kinds of *curvature*. This means that there are essentially no interesting local coordinate calculations to be done in symplectic geometry. Whew!

11 SYMPLECTIC MECHANICS

> *Physics is that subset of human experience that can be reduced to coupled harmonic oscillators.*
>
> Michael Peskin

You might reasonably ask what the point of all of developing all this formalism was! Firstly, what is the physical significance of Hamiltonian mechanics on symplectic manifolds that are not cotangent bundles? The answer is given by *symplectic reduction*: these symplectic manifolds naturally arise when studying mechanical systems with symmetries. Explaining the *Marsden-Weinstein theorem*, which says that the quotient of a symplectic manifold by a *Hamiltonian symmetry* is also a symplectic manifold, will be the goal of the next few weeks. For example, the Fubini-Study form on complex projective space will arise in precisely this way. In fact, in a precise sense, all symplectic manifolds are given by a generalized form of symplectic reduction of cotangent bundles (called *Cartan reduction*).

Accepting this for now, we'll start answering the question above by showing how we can use the formalism we have developed for differential forms to give very clean and beautiful proofs of some of the key properties of Hamiltonian mechanics.

POISSON BRACKETS

Let's first try to prove:

THEOREM 15. *(**Conservation of the Hamiltonian**) Suppose $H : X \to \mathbb{R}$ is a smooth function on a symplectic manifold X; then H is constant along the trajectories of the Hamiltonian flow of H.*

Proof. Let $\gamma : \mathbb{R} \to X$ be a flow line of the Hamiltonian vector field X_H: we want to show that

$$\frac{\mathrm{d}}{\mathrm{d}t} H(\gamma(t)) = 0$$

We simply unravel the definitions:

$$\frac{\mathrm{d}}{\mathrm{d}t} H(\gamma(t)) = \mathrm{d}H(\gamma'(t)) = \omega(X_H, \gamma'(t)) = \omega(X_H, X_H) = 0$$

since ω is a 2-form. ■

That was easy! But now we can see how this result can be generalized!

DEFINITION 22. *The **Poisson bracket** of two functions $H, K : X \to \mathbb{R}$ on a symplectic manifold is another smooth function $\{H, K\}$ on X defined by:*

$$\boxed{\{H, K\} = \omega(X_H, X_K)}$$

THEOREM 16. *Suppose that $K : X \to \mathbb{R}$ is a smooth function that Poisson-commutes with the Hamiltonian, that is, $\{K, H\} = 0$. Then K is conserved under the Hamiltonian flow. We say that K is a **integral of motion**.*

Proof. Please take a second to try and prove this! Again, we simply calculate: let $\gamma : \mathbb{R} \to X$ be a flow line of the Hamiltonian vector field X_H:

$$\frac{\mathrm{d}}{\mathrm{d}t} K(\gamma(t)) = \mathrm{d}K(\gamma'(t)) = \omega(X_K, \gamma'(t)) = \omega(X_K, X_H) = \{K, H\} = 0$$

■

In fact, what we have shown in the previous theorem is that for any smooth function $K : X \to \mathbb{R}$, we have $\dot{K} = \{K, H\}$ along the trajectories of H. Thus Poisson brackets control how quantities evolve under Hamiltonian dynamics!

We can do much more with Poisson brackets! First off, we have

PROPOSITION 9. *The collection of all smooth functions on a symplectic manifold X forms a **Lie algebra** with respect to the Poisson bracket: that is, the Poisson bracket satisfies:*

- ***(Bilinearity)*** $\{\lambda f + \mu g, h\} = \lambda\{f, h\} + \mu\{g, h\}$, *for* $\mu, \lambda \in \mathbb{R}$
- ***(Antisymmetry)*** $\{f, g\} = -\{g, f\}$
- ***(Jacobi Identity)*** $\{f, \{g, h\}\} + \{h, \{f, g\}\} + \{g, \{h, f\}\} = 0$

In other words, the Poisson bracket behaves a lot like the *commutator brackets* from quantum mechanics. In fact, the analogy runs deeper than this:

PROPOSITION 10. *Suppose $\mathbb{R}^{2n}$ has the canonical symplectic form $\omega = \sum_{i=1}^{n} \mathrm{d}q_i \wedge \mathrm{d}p_i$; then the Poisson brackets are given by:*

$$\{q_i, p_j\} = \delta_{ij}$$

Proof. We can write the Poisson bracket in local coordinates as follows:

$$\{f, g\} = \omega(X_f, X_g) = -(X_f)^T J(X_g) = -(-J\nabla f)^T J(-J\nabla g) = -(\nabla f)^T J^T J^2(\nabla g) = -(\nabla f)^T J(\nabla g)$$

Since $\nabla q_i, \nabla p_j$ are exactly the standard basis vectors we have that $\{q_i, p_j\}$ are exactly the entries in the lower left-hand Id entry of J, that is, $\{q_i, p_j\} = \delta_{ij}$. ■

So we see that even in classical mechanics, position and momentum don't commute! The subject of *geometric quantization*, which we'll hear about later on in the student presentations, tries to make this connection rigorous: find a means of associating to any symplectic manifold a Hilbert space with observables obeying the same commutation relations (to order $\hbar$).

There's also another natural Lie algebra structure, which exists on any smooth manifold:

DEFINITION 23. *Let X be a manifold and V a vector field on X; the **Lie derivative** $\mathcal{L}_V(\alpha)$ of a differential k-form α on X is defined to be the differential k-form given at a point $x \in X$ by:*

$$(\mathcal{L}_V\alpha)_x = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} (\Phi_t^*\alpha)_x$$

*where Φ_t is the time-t flow of V and the derivative with respect to t is taken inside the vector space $\bigwedge^k T_x^*X$. Similarly, we can define the **Lie bracket** of two vector fields V, W to be the vector field $[V, W]$ given at a point $x \in X$ by:*

$$[V, W]_x = -(\mathcal{L}_V W)_x = -\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} (D_{\Phi_t(x)}\Phi_t^{-1} W)_x$$

where the derivative is taken inside the vector space T_xX.

Note that the minus signs here are taken to make this consistent with our choice of sign in Hamilton's equations. The Lie bracket $[\cdot, \cdot]$ of vector fields makes the vector fields on any manifold (not necessarily symplectic) into an infinite-dimensional Lie algebra. Then we have:

PROPOSITION 11. *If X is a symplectic manifold, then the symplectic gradient gives a homomorphism of Lie algebras from the smooth functions on X to the Lie algebra of vector fields on X: in other words,*

$$X_{\{f,g\}} = [X_f, X_g]$$

for any two smooth functions f, g on X.

Proof. Since differentiation is a linear operation, we have:

$$\omega(\mathcal{L}_{X_f} X_g, \cdot) = \mathcal{L}_{X_f} \left(\omega(X_g, \cdot)\right)$$

where $\omega(X_g, \cdot)$ is considered as a 1-form. Then we have:

$$\mathcal{L}_{X_f} \left(\omega(X_g, \cdot)\right) = \mathcal{L}_{X_f}(\mathrm{d}g) = \mathrm{d}(\mathcal{L}_{X_f} g)$$

by exchanging the derivatives. Then

$$\mathrm{d}(\mathcal{L}_{X_f} g) = \mathrm{d}(\mathrm{d}g(X_f)) = \mathrm{d}\{f, g\}$$

which means exactly that $X_{\{f,g\}} = [X_f, X_g]$. ■

HIC SUNT DRACONES 3. *Note that if we had taken the opposite signs in either Hamilton's equation or the definition of the Lie bracket, this would have been a Lie algebra anti-homomorphism. This I hope belatedly justifies our choice of sign conventions.*

We can even characterize symplectomorphisms in terms of Poisson brackets:

THEOREM 17. *Suppose (X, ω) is a symplectic manifold: then a diffeomorphism $F : X \to X$ is a symplectomorphism if and only if F preserves Poisson brackets, in the sense that $\{f, g\} \circ F = \{f \circ F, g \circ F\}$.*

This is a great exercise, so give it a try! It is one of the rare cases where working in local coordinates can be much easier.

LIOUVILLE'S THEOREM

Now let's try something more difficult, like Liouville's theorem. Let's first begin with a Lemma:

LEMMA 2. *The flow of every Hamiltonian gives a symplectomorphism.*

Let $\Phi_t : X \to X$ be the diffeomorphism of X given by the time-t flow of the vector field X_H of the Hamiltonian $H : X \to \mathbb{R}$. What we want to show is that

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \Phi_t^* \omega = 0$$

But this is exactly the definition of the Lie derivative from above! The main result we will need about the Lie derivative goes under the name

THEOREM 18. *(**Cartan's Magic Formula**) The Lie derivative of a differential form is given by*

$$\mathcal{L}_v\alpha = \mathrm{d}\iota_v\alpha + \iota_v\mathrm{d}\alpha$$

*where ι_v is the **interior product** with a vector field defined above.*

Now we can prove our Lemma:

Proof. Using the definition above, we can recognize the quantity

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0}\Phi_t^*\omega$$

as exactly the Lie derivative $\mathcal{L}_{X_H}\omega$. Applying Cartan's magic formula gives:

$$\mathcal{L}_{X_H}\omega = \iota_{X_H}\mathrm{d}\omega + \mathrm{d}(\iota_{X_H}\omega)$$

Now ω is closed, and, by definition, $\iota_{X_H}\omega = \mathrm{d}H$: thus

$$\mathcal{L}_{X_H}\omega = \mathrm{d}(\mathrm{d}H) = 0$$

since $\mathrm{d}\circ\mathrm{d} = 0$. ■

This may provide some belated justification for why we asked for ω to be closed, since we can use this property to prove:

THEOREM 19. *(**Liouville's theorem**) The flow of a Hamiltonian vector field preserves volumes of subsets of X.*

Firstly, we need to interpret the *volume* of a symplectic manifold correctly using the symplectic form. Given $\omega \in \Omega^2(X)$, we can take the nth wedge power ω^n to get a $2n$-form on X, and define the **symplectic volume** of a subset $U \subseteq X$ to be the integral:

$$\mathrm{vol}(U) = \int_U \omega^{\wedge n}$$

As an exercise, show that in local coordinates on a cotangent bundle where

$$\omega = \sum_{i=1}^n \mathrm{d}q_i \wedge \mathrm{d}p_i$$

then

$$\omega^{\wedge n} = n!\,\mathrm{d}q_1 \wedge \cdots \wedge \mathrm{d}q_n \wedge \mathrm{d}p_1 \wedge \cdots \wedge \mathrm{d}p_n$$

so that the symplectic volume of a subset $U \subseteq \mathbb{R}^{2n}$ is given by

$$\mathrm{vol}(U) = n! \int_U \mathrm{dvol}_{2n}$$

Now we can prove Liouville's theorem:

Proof. Firstly, we use the invariance property of integrals under pullbacks:

$$\int_{\Phi_t(U)} \omega^{\wedge n} = \int_U \Phi_t^*(\omega^{\wedge n})$$

Then we take the derivative and use the linearity of the integral:

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \int_U \Phi_t^*(\omega^{\wedge n}) = \int_U \mathcal{L}_{X_H}\omega^{\wedge n}$$

Finally, using the fact that the Lie derivative acts as a derivation with respect to the wedge product, as well as the previous lemma, we have that:

$$\int_U \mathcal{L}_{X_H}\omega^{\wedge n} = 0$$

which completes the proof. ■

More generally, in fact all symplectomorphisms preserve volumes!

The exact same proof as before implies *Poincaré recurrence*:

THEOREM 20. *(**Poincaré Recurrence**) Suppose X is a symplectic manifold with finite volume and $H : X \to \mathbb{R}$ is a Hamiltonian; let Φ_t be the time-t flow of the Hamiltonian vector field X_H. Then for every point $x \in X$, for every neighbourhood U of x and every $C > 0$, there exists some $y \in U$ and some $t > C$ so that $\Phi_t(y) \in U$.*

12 LAGRANGIAN SUBMANIFOLDS

In mathematics you don't understand things. You just get used to them.

John von Neumann

I feel like I couldn't possibly finish this course without telling you about Lagrangian submanifolds: not only do they provide a link between the classical and quantum worlds, but they are ubiquitous throughout mathematics and have many fascinating properties in their own right. Under *quantization* of a symplectic manifold, they correspond to *quantum states* and as a result display many distinctly quantum properties.

Firstly, what is a *submanifold*? It's kind of exactly what you think it would be, modulo some technical subtleties. Namely, an **(embedded) submanifold** L of smooth manifold X is a subset $L \subseteq X$ such that around every point $p \in L$ there exists some smooth chart (U, ϕ) in which $L \cap U$ is given by the hyperplane $x_1 = x_2 = \cdots = x_k = 0$ inside $\mathbb{R}^n$. Using these coordinates $x_1, \ldots, x_k$ on L gives L itself the structure of a smooth manifold, so that the inclusion map $i : L \to X$ is smooth.

Now, when X is a symplectic manifold with symplectic form ω, there is a certain type of submanifold which play a very important role: not those that are *symplectic* submanifolds, but rather those at the opposite end of the spectrum, where the symplectic form is *zero*. Define a **Lagrangian submanifold** of X to be a submanifold $i : L \to X$ such that $i^*\omega = 0$ and $\dim L = \frac{1}{2}\dim X$. We say L is **exact** if in fact $\omega = \mathrm{d}\lambda$ and $i^*\lambda = \mathrm{d}f$ for some smooth function $f : L \to \mathbb{R}$.

How does this not contradict the non-degeneracy of the symplectic form? Well, a symplectic form can vanish on *subspaces* of a vector space, so long as their dimension is less than or equal to half of the dimension of the vector space. For example, in $\mathbb{R}^{2n}$ with Darboux coordinates $q_1, \ldots, q_n, p_1, \ldots, p_n$, the spaces $\mathbb{R}^n \times \{p\}$ and $\{q\} \times \mathbb{R}^n$ (p, q fixed) are all Lagrangian. In quantum mechanics, you can think of these as the eigenstates of the position and momentum operators.

This (rather confusing) use of the term *Lagrangian* has nothing to do with the Lagrangian function L, but rather with the fact that ω was originally called the *Lagrange bracket*, and so Maslov decided to call a submanifold on which it vanished *Lagrangian*. Here are some examples:

Example: Every smooth curve on a symplectic surface is a Lagrangian submanifold, simply because every 2-form vanishes on a 1-dimensional manifold.

Example: For any smooth manifold Q, the **cotangent fibers** $T_q^*Q \subseteq T^*Q$ are Lagrangian submanifolds of the cotangent bundle with the canonical symplectic form. This is a generalization of the example we saw above.

Example: Another large class of examples is given by *sections*. Given a 1-form α on Q, by definition it smoothly assigns for every $q \in Q$ an element $\alpha_q \in T_q^*Q$: therefore we can associate to α a smooth map $s_\alpha : Q \to T^*Q$ given by $s_\alpha(q) = (q, \alpha_q)$. If $\pi : T^*Q \to Q$ is the projection, then we can observe that $\pi \circ s_\alpha(q) = q$: we call any such map a **section** of the cotangent bundle. It is clear that the data of the 1-form α is equivalent to the data of the section s_α (indeed, some people prefer to *define* 1-forms in this way). Now, by looking at the image of the map s_α, we get a submanifold of Q, called the **graph** of α. We now have:

PROPOSITION 12. *The graph of α is a Lagrangian submanifold of T^*Q with the standard symplectic structure exactly when $\mathrm{d}\alpha = 0$. Moreover, the graph is an exact Lagrangian precisely when $\alpha = \mathrm{d}f$ for $f : Q \to \mathbb{R}$ a smooth function. In particular, when $\alpha = 0$ we call this Lagrangian the* ***zero section****.*

We shall prove this using the following result:

PROPOSITION 13. *Suppose $s_\alpha : X \to T^*X$ is the section associated to a 1-form α on Q; then $s_\alpha^*\lambda = \alpha$.*

In other words, pulling back λ by the section s_α gives exactly the original 1-form α: this is the sense in which the 1-form λ is *tautological*: it is in fact uniquely characterized by this property.

Proof. Please try to do this first! It's just a case of unravelling the definitions: so suppose $v \in T_qQ$ is a tangent vector and let's compute:

$$s_\alpha^*(\lambda)_q(v) = \lambda_{(q,\alpha_q)}(D_q s_\alpha(v)) = \alpha_q(D_{(q,p)}\pi \circ D_q s_\alpha(v)) = \alpha_q(D_{(q,p)}\mathrm{Id}(v)) = \alpha_q(v)$$

where we have used the chain rule along with the fact that $\pi \circ s_\alpha = \mathrm{Id}$. ■

Now we can prove the assertion about graphs by noting that:

$$s_\alpha^*\omega = -\mathrm{d}s_\alpha^*\lambda = -\mathrm{d}\alpha = 0$$

and similarly that

$$s_\alpha^*\lambda = \alpha = \mathrm{d}f$$

In this case we call f a **generating function** for the Lagrangian. The significance of this result comes from the next example:

PROPOSITION 14. *Suppose $F : (X_1, \omega_1) \to (X_2, \omega_2)$ is a smooth map: then F is a symplectomorphim if and only if the graph of F is a Lagrangian submanifold of $(X_1, \omega_1) \times (X_2, -\omega_2)$.*

Proof. If $i : X_1 \to X_1 \times X_2$ represents the graph of F, then $i(x) = (x, F(x))$; then

$$i^*((\omega_1, -\omega_2)) = \omega_1 - F^*\omega_2$$

and this is zero if and only if F is a symplectomorphism. ■

Therefore symplectomorphisms are special cases of Lagrangian submanifolds! In particular, one can recover the original symplectomorphism from the corresponding Lagrangian. Therefore symplectomorphisms can also have generating functions, in the following sense.

Given a symplectomorphism $F : T^*\mathbb{R}^n \to T^*\mathbb{R}^n$, its graph gives a Lagrangian submanifold of $T^*\mathbb{R}^n \times T^*\mathbb{R}^n$, which is symplectomorphic to $T^*\mathbb{R}^{2n}$. Sometimes there exists f, a smooth function on $\mathbb{R}^{2n}$ so that the graph of $\mathrm{d}f$ defines the Lagrangian submanifold of $T^*\mathbb{R}^{2n}$ and thus the symplectomorphism F. We call f a **generating function** for the symplectomorphism: it allows us to specify the entire data of the symplectomorphism in terms of a single smooth function. Notice however that there are actually four distinct linear symplectomorphisms that one could choose $T^*\mathbb{R}^n \times T^*\mathbb{R}^n$ and $T^*\mathbb{R}^{2n}$: these correspond to generating functions of different *types*, as they are called in the physics literature.

One can ask this can be done more generally: Arnol'd's *nearby Lagrangian conjecture* says that if $L \subseteq T^*Q$ is an exact Lagrangian submanifold, then there is some Hamiltonian $H : T^*Q \to \mathbb{R}$ such that the time-1 flow of X_H takes L to the zero section $Q \subseteq T^*Q$. This is a completely open problem, and currently an area of intense research!

Let's change track a bit now, and ask: how do we recover a variational formulation of Hamiltonian mechanics? The answer is subtle, and involves Lagrangian submanifolds in an essential way. Given a symplectic manifold X with an exact symplectic form $\omega = \mathrm{d}\alpha$ for some 1-form α, along with a Hamiltonian H, we can define the **symplectic action** or **Hamiltonian action** of a path $\gamma : [0,1] \to X$ by:

$$S(\gamma) = \int_0^1 \alpha(\dot{\gamma})\mathrm{d}t - \int_0^1 H(\gamma(t))\mathrm{d}t$$

As always, we need to place boundary conditions on γ in order to have a well-defined variational problem: let's suppose that there are two submanifolds L_1, L_2 of X so that we require $\gamma(0) \in L_0$ and $\gamma(1) \in L_1$: examples might include $L_0 = T^*_{q_0}Q, L_1 = T^*_{q_1}Q$ for the usual Dirichlet boundary conditions, or L_0, L_1 are the zero-section for von Neumann boundary conditions.

What is a necessary condition for the symplectic action to be minimized? Following along with our derivation of the Euler-Lagrange equations, let's suppose we have a family of such curves $\gamma_s : [0,1] \to X$ with $s \in \mathbb{R}$ and $\gamma_0 = \gamma$. Then if γ is a critical value of the symplectic action, we will have:

$$\left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} S(\gamma_s) = 0$$

Let's calculate:

$$\begin{aligned} 0 = \left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} S(\gamma_s) &= \int_0^1 \left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} \alpha(\dot{\gamma}_s)\mathrm{d}t - \int_0^1 \left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} H(\gamma_s(t))\mathrm{d}t \\ &= \int_0^1 \alpha\left(\left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} \dot{\gamma}_s\right)\mathrm{d}t - \int_0^1 \mathrm{d}H(V(t))\mathrm{d}t \end{aligned}$$

where $V(t)$ is the vector field along $\gamma(t)$ given by:

$$V(t) = \left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} \gamma_s(t)$$

Now we use the fact that we can change the order of the s and t derivatives to write:

$$\begin{aligned} 0 &= \int_0^1 \alpha_{\gamma(t)}\left(\frac{\mathrm{d}}{\mathrm{d}t}V(t)\right)\mathrm{d}t - \int_0^1 \mathrm{d}H(V(t))\mathrm{d}t \\ &= \int_0^1 \frac{\mathrm{d}}{\mathrm{d}t}\left(\alpha_{\gamma(t)}\left(V(t)\right)\right)\mathrm{d}t + \omega(\dot{\gamma}, V(t))\mathrm{d}t - \int_0^1 \mathrm{d}H(V(t))\mathrm{d}t \end{aligned}$$

Integrating by parts gives:

$$0 = \alpha(V(1)) - \alpha(V(0)) - \int_0^1 \omega(\dot{\gamma}, V(t)) - \mathrm{d}H(V(t))\mathrm{d}t$$

If we neglect the first term for the moment, then since $V(t)$ could have been any smooth vector field along γ, a version of the fundamental lemma of the calculus of variations tells us that if this integral is always zero, then we must have $\omega(\dot{\gamma}, \cdot) = \mathrm{d}H$. In other words, minima of the action must necessarily occur when γ is a flow of the Hamiltonian vector field X_H. This is exactly what we wanted.

However, this was neglecting the first term above! What are $V(1)$ and $V(0)$? Since γ_s must satisfy the boundary conditions for all $s \in \mathbb{R}$, we must have $\gamma_s(0) \in L_0$ and $\gamma_s(1) \in L_1$. Hence it follows that $V(0) \in T_{\gamma(0)}L_0$ and $V(1) \in T_{\gamma(1)}L_1$. The only way we could have $\alpha(V(1)) - \alpha(V(0)) = 0$ for every such γ_s is if α is zero on the tangent spaces to L_0 and L_1. But this means that L_0, L_1 must have $\omega|_{L_0} = 0$ and

$\omega|_{L_1} = 0$. Since we want our spaces of possible boundary conditions to be *maximal*, this means we should have L_0, L_1 be *Lagrangian submanifolds* of X.

Now, since the time-1 flow of the Hamiltonian H gives a symplectomorphism $\Phi^1_H : X \to X$, the graph of Φ^1_X is a Lagrangian submanifold, call it Γ_H inside the product $X \times X$ with symplectic form $(\omega, -\omega)$. Then the solutions to the variational problem above are exactly given by the intersection points of the two Lagrangians: $\Gamma_H \cap (L_0 \times L_1)$.

Versions of this principle exist throughout mathematics and physics: the essence is the fact that *Lagrangians are the natural place to put boundary conditions*. In short, in order to turn a 'global' variational principle into a 'local' differential equation, we need to place our boundary conditions on Lagrangian submanifolds! Moreover *solutions to equations in physics are given by intersections of Lagrangian submanifolds*! This philosophy holds in alarming generality, enough for Weinstein to coin his famous **Lagrangian credo**: **everything is a Lagrangian submanifold**! I think he originally meant everything in symplectic geometry (symplectomorphisms, symplectic reduction, etc.), but I think it's fair to say that it just as well generalizes to **everything in the universe is the intersection of Lagrangian submanifolds**. For instance, under quantization, this Lagrangian intersection corresponds to taking the correlator $\langle L_0 | \mathrm{e}^{tH} | L_1 \rangle$.

This suggested to Weinstein that the natural notion of a symplectic map between two symplectic manifolds X_1, X_2 be generalized to include *all* Lagrangian submanifolds of the product $X_1 \times X_2$: these generalized symplectomorphisms are called **Lagrangian correspondences** and play an important role in the quantum geometry of Lagrangians.

For instance, the set of all Lagrangians inside a symplectic manifold X carries a natural algebraic structure, called a **category**: there is a way to *compose* intersection points between Lagrangians using the pseudoholomorphic curves we will talk about later, called **quantum multiplication**. Using these pseudoholomorphic disks, one can show versions of the uncertainty principle for Lagrangian submanifolds, much like Gromov's non-squeezing theorem. The fact is that these categories of Lagrangians (called **Fukaya categories**) can often be calculated: they turn out to give rise to a kind of *non-commutative algebraic space* called the *mirror* whose points are the original Lagrangian tori inside X. This is a deep area of mathematics called *mirror symmetry* which is currently also an area of active research.

13 PROBLEMS

SYMPLECTIC GEOMETRY IN DIMENSION 2

In general, symplectic manifolds are almost impossible to classify, but the following series of exercises describes how we can classify all symplectic manifolds of dimension 2.

Exercise: Suppose $X \subseteq \mathbb{R}^3$ is an embedded orientable surface in $\mathbb{R}^3$, that is, X is locally given by the image of an injective map $r : \mathbb{R}^2 \to \mathbb{R}^3$ so that X has a global normal vector field $N : X \to \mathbb{R}^3$. Show that there is a 2-form on X given by $\omega_x(v_1, v_2) = N_x \cdot (v_1 \times v_2)$, using the dot and cross products in $\mathbb{R}^3$. Show that ω in fact gives a symplectic structure on X, and hence conclude that every orientable 2-dimensional manifold is a symplectic manifold.

Exercise: Prove that a diffeomorphism $F : X \to Y$ of symplectic manifolds is a symplectomorphism if and only if at every point $x \in X$, there exists a symplectic basis of $T_x X$ and $T_{F(x)} Y$ so that $D_x F$ is represented by a symplectic matrix.

Exercise: Consider the symplectic group $\mathrm{Sp}(2)$ of 2×2 symplectic matrices: give an explicit description of elements of this group.

Exercise: Show that a diffeomorphism $F : X_1 \to X_2$ of 2-dimensional symplectic manifolds is a symplectomorphism if and only if it preserves areas.

Exercise: Write down the canonical 2-form on T^*S^1 in coordinates and use the previous exercises to write down a diffeomorphism of the cylinder T^*S^1 that is a symplectomorphism with respect to the canonical symplectic form. Can you give an example of a diffeomorphism that is *not* a symplectomorphism?

HIGHER-DIMENSIONAL SYMPLECTIC MANIFOLDS

What can we say about classifying symplectic manifolds in higher dimensions?

Exercise: Show that the product of two symplectic manifolds is naturally a symplectic manifold, and describe the symplectic form on the 4-torus $T^4 = (S^1)^4$.

Exercise: Suppose that $\omega = \sum_{i=1}^n \mathrm{d}q_i \wedge \mathrm{d}p_i$ is the standard symplectic form on $\mathbb{R}^{2n}$: show that

$$\omega^{\wedge n} = n!\, \mathrm{d}q_1 \wedge \cdots \wedge \mathrm{d}q_n \wedge \mathrm{d}p_1 \wedge \cdots \wedge \mathrm{d}p_n$$

Exercise: Show that the k-th wedge power of a symplectic form ω on an n-dimensional symplectic manifold X is always non-zero for $k \leq n$. For those who have seen algebraic topology: explain why S^n for $n > 2$ is never a symplectic manifold.

Exercise: Suppose X is an exact symplectic manifold with 1-form λ. Show that X cannot be compact. Hint: introduce the *Liouville vector field* Z defined by $\omega(Z, \cdot) = \lambda$ and mimic the proof of Liouville's theorem to derive a contradiction.

Exercise: Suppose $F : X \to Y$ is a diffeomorphism of smooth manifolds; show that there exists a symplectomorphism $T^*X \to T^*Y$.

Eliashberg's conjecture claims that the converse is true: T^*X and T^*Y are symplectomorphic if and only if X is diffeomorphic to Y. This would imply that classifying symplectic manifolds is as difficult as classifying all smooth manifolds. Now we will see an equivalent way to reformulate this conjecture, which is easier to study, and a very important open problem in symplectic geometry.

LAGRANGIAN SUBMANIFOLDS

Define a **Lagrangian submanifold** L of a symplectic manifold (X, ω) to be an embedding $i : L \to X$ such that $i^*\omega = 0$ and $\dim L = \frac{1}{2} \dim X$. We say L is **exact** if in fact $\omega = \mathrm{d}\lambda$ and $i^*\lambda = \mathrm{d}f$. Such submanifolds play an extremely important role in mechanics and symplectic geometry. Here are some important examples:

Exercise: Show that every simple smooth curve on a symplectic surface is a Lagrangian submanifold.

Exercise: Show that $T_q^*Q \subseteq T^*Q$ and $Q \subseteq T^*Q$ are Lagrangian submanifolds of the cotangent bundle with the standard symplectic form. Show moreover that any 1-form α on Q with $\mathrm{d}\alpha = 0$ defines a Lagrangian submanifold of T^*Q by looking at its *graph*. Show that this graph is *exact* Lagrangian if and only if $\alpha = \mathrm{d}f$.

In this case we call f a *generating function*. Moreover, symplectomorphisms are special cases of Lagrangian submanifolds, and so can also have generating functions:

Exercise: Suppose $F : (X_1, \omega_1) \to (X_2, \omega_2)$ is a smooth map: show that F is a symplectomorphim if and only if the graph of F is a Lagrangian submanifold of $(X_1, \omega_1) \times (X_2, -\omega_2)$.

The *nearby Lagrangian conjecture* says that if $L \subseteq T^*Q$ is an exact Lagrangian submanifold, then there is some Hamiltonian $H : T^*Q \to \mathbb{R}$ such that the time-1 flow of X_H takes L to $Q \subseteq T^*Q$.

Exercise: Explain how Arnol'd's nearby Lagrangian conjecture implies Eliashberg's conjecture, and prove the nearby Lagrangian conjecture for Lagrangians that are C^1-close to Q when $Q = S^1$.

Exercise: Prove that any exact Lagrangian submanifold of T^*S^1 intersects S^1 at at least two points.

This is a special case of *Arnold's conjecture*, which we will discuss later on in the class.

SYMPLECTIC MECHANICS

Exercise: Let (X, ω) be a symplectic manifold: show that a diffeomorphism $F : X \to X$ is a symplectomorphism if and only if it respects Poisson brackets, in the sense that $\{H \circ F, K \circ F\} = \{H, K\} \circ F$ for all smooth functions H, K. Hint: this is one of the few problems where it is easier to work in local coordinates.

Exercise: Suppose $\mathbb{R}^{2n}$ has the canonical symplectic form $\omega = \sum_{i=1}^n \mathrm{d}q_i \wedge \mathrm{d}p_i$; prove that the Poisson brackets are given by:

$$\{q_i, p_j\} = \delta_{ij}$$

Now suppose H is a Hamiltonian on $\mathbb{R}^{2n}$ that depends only on the momentum coordinates p_i: show that momentum is conserved under the flow of the Hamiltonian vector field of H.

Exercise: Show that if X is a symplectic manifold, then the symplectic gradient gives a homomorphism of Lie algebras from the smooth functions on X to the Lie algebra of vector fields on X: in other words,

$$X_{\{f,g\}} = [X_f, X_g]$$

Exercise: Generalize Liouville's Theorem by showing more generally that all symplectomorphisms preserve volumes.

Exercise: Write down the canonical 2-form on T^*S^1 in coordinates and use the previous exercise to write down a diffeomorphism of the cylinder T^*S^1 that is a symplectomorphism with respect to the canonical symplectic form. Can you give an example of a diffeomorphism that is *not* a symplectomorphism? Similarly, give an example of a vector field on T^*S^1 that preserves the symplectic form but does not come from any Hamiltonian.

Exercise: Suppose $F : X \to Y$ is a diffeomorphism of smooth manifolds; show that there exists a symplectomorphism $T^*X \to T^*Y$.

14 LIE GROUPS

Algebraists usually define groups as sets with operations that satisfy a long list of axioms that are difficult to remember. Frankly, I believe it is impossible to understand such a definition.

Vladimir Arnol'd

In the next few lectures, we want to build towards the *Marsden-Weinstein theorem* which will give us a description of Hamiltonian systems with symmetries, generalizing Noether's theorem. Recall that when we discussed Noether's theorem, our definition of a smooth symmetry seemed somewhat artificial. In fact, we saw that all we needed was an infinitesimal symmetry, but again this definition seemed ad hoc. In this lecture, we'll formalize this by discussing *Lie groups* and their *Lie algebras* whose *actions* will precisely give us our smooth and infinitesimal symmetries.

A *Lie group* is essentially just a *smooth group*:

DEFINITION 24. *A **Lie group** is a manifold G along with smooth maps $m : G \times G \to G$ called multiplication and $i : G \to G$ called inversion, as well as a distinguished point $e \in G$, that together satisfy the axioms to be a group: namely, m is associative, e is a left and right identity, and every element has a (two-sided) inverse given by i.*

Just as for usual groups, we can define:

DEFINITION 25. *A (left) **action of a Lie group** G on a manifold X is a smooth map $a : G \times X \to X$, written $(g, x) \mapsto g \cdot x$, that defines an action of the group G on the set X, that is, e acts as the identity map, and $(gh) \cdot x = g \cdot (h \cdot x)$. We say that G gives a **smooth symmetry** of X.*

We will sometimes use $\Phi(g) : X \to X$ to denote the diffeomorphism of X given by the element $g \in G$. In fact, we've already seen many examples of Lie groups and their actions:

Example 14. Consider $\mathbb{R}^n$, with the map $m : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ given by addition, and the map $i : \mathbb{R}^n \to \mathbb{R}^n$ given by negation; let $e = 0$. Then one can easily check that these maps are smooth, and this makes $\mathbb{R}^n$ into a Lie group with its usual additive group structure. Then $\mathbb{R}^n$ acts on itself by *translation*, and this captures the notion that $\mathbb{R}^n$ has a translation-symmetry from the first lecture.

Example 15. We've see the circle S^1 many times before, but in fact it is also a Lie group! The multiplication map m is given by addition of angles from the positive x-axis, modulo 2π: we can check that in the angular coordinates we wrote down, this is in fact a smooth map. Similarly, inversion is just given by negation of this angle, and the identity is the point $(1, 0) \in S^1$ with angle zero. More generally, this also means that the n-torus T^n is a Lie group, as the n-fold product of S^1 also naturally has the structure of a Lie group.

Example 16. For a more non-trivial example of a Lie group, consider the set $\mathrm{GL}_n(\mathbb{R})$ of all $n \times n$ invertible matrices. Using the entries of the matrix as coordinates, we can think of this as sitting as an open subset of $\mathbb{R}^{n^2}$, since the condition $\det(A) \neq 0$ is an open condition. Hence $\mathrm{GL}_n(\mathbb{R})$ is a manifold, and one easily see that the usual group structure given by matrix multiplication is in fact a smooth map.

Example 17. Another similar example we saw when discussing rigid bodies was the **special orthogonal group** $\mathrm{SO}(n)$, the group of rotations of $\mathbb{R}^n$. We can build this as a subset of $\mathrm{GL}_n(\mathbb{R})$ given by those matrices A satisfying $A^T A = \mathrm{Id}$ (orthogonal) and $\det(A) = 1$ (orientation-preserving). One can then check that these conditions do indeed define a smooth manifold, and we know that the multiplication of matrices is a smooth map, making $\mathrm{SO}(n)$ into a Lie group. This group acts on $\mathbb{R}^n$ via rotations: in coordinates, the map a is just given by multiplying a vector in $\mathbb{R}^n$ by a matrix, and so this action is smooth. This captures the intuitive rotation-symmetry of $\mathbb{R}^n$ from the first lecture.

Example 18. Another example is the **symplectic group** $\mathrm{Sp}(2n)$ we saw in Lecture 8: this is again a smooth manifold living as a subgroup of $\mathrm{GL}_n(\mathbb{R})$.

Example 19. To every Lie group G, it has several natural associated actions. The first is that G acts on itself by left multiplication: by definition of a Lie group, this is a smooth map. This captures the intuitive idea that G is 'symmetric'.

Example 20. Also, we can consider the action of G on itself by *conjugation*, namely $g \cdot h = ghg^{-1}$: this is called the **Adjoint action** of G on itself and again this is smooth by definition of a Lie group.

Example 21. Since conjugation fixes the identity element of the group, for every $g \in G$, we can take the differential $D_e\Phi(g) : T_eG \to T_eG$: because of the chain rule, this gives us an action of G on T_eG, the tangent space to the identity, called the **adjoint action** of the group G.

THE LIE ALGEBRA

How do we understand Lie groups and their actions? By differentiating them!

DEFINITION 26. *The **Lie algebra** $\mathfrak{g}$ **of a Lie group** G is the tangent space T_eG at the identity of G.*

Why does this give us an *algebra*? Because it always comes equipped with a binary operation called the *Lie bracket*, defined as follows. Given any $v \in T_eG$, we can form a vector field X_v on the whole of G by using the multiplication map to transport the vector around. If we let $L_g : G \to G$ denote the map given by left multiplication by $g \in G$, then its derivative defines a map $D_eL_g : T_eG \to T_gG$ and we can define:

$$(X_v)_g = D_eL_g(v)$$

This in fact defines a smooth vector field on all of G. Now we define the **Lie bracket** by:

$$[v, w] = [X_v, X_w]_e$$

where $[\cdot, \cdot]$ is the Lie bracket of vector fields on G, and we evaluate this vector field at the identity. This will inherit all of the properties of the Lie bracket of vector fields, namely that $[v, w] = -[w, v]$, so that $\mathfrak{g}$ is always a finite-dimensional Lie algebra. Let's look at some examples:

Example 22. If $G = \mathbb{R}^n$ with respect to addition, then $T_e\mathbb{R}^n \cong \mathbb{R}^n$ and for any $v \in \mathbb{R}^n$, the associated vector field X_v is just the constant vector field with value v. Thus the Lie bracket $[\cdot, \cdot]$ is identically zero, since the Lie bracket of vector fields involves differentiation. This is because $\mathbb{R}^n$ is abelian as an additive group: we similarly have that the bracket on the Lie algebra of T^n vanishes identically.

Example 23. Let G be $\mathrm{GL}_n(\mathbb{R})$: then since G is an open subset of $\mathbb{R}^{n^2}$, we can identify T_eG with $\mathbb{R}^{n^2}$, the space of all $n \times n$ matrices. Then the bracket on the Lie algebra is given by the matrix commutator: if A, B are $n \times n$ matrices, then $[A, B] = AB - BA$. More generally, any compact Lie group that is a a subgroup of $\mathrm{GL}_n(\mathbb{R})$ also has Lie bracket given by the matrix commutator: in fact, *every* compact Lie group is such a subgroup, so we will often simply assume that the Lie bracket is given by the matrix commutator.

Example 24. As we saw above, any Lie group G always acts on $T_eG = \mathfrak{g}$ its Lie algebra via the *adjoint action*. In fact, for $\mathrm{GL}_n(\mathbb{R})$ and its subgroups one can easily see that this action is simply given by conjugation of matrices: $g \cdot A = gAg^{-1}$ for $A \in T_eG$ an $n \times n$ matrix.

How can we use this differentiation to understand group actions? Suppose a Lie group G is acting on a manifold X, and suppose $v \in T_eG$ is an element of the Lie algebra. Then v can be represented by a smooth curve $\gamma : \mathbb{R} \to G$ with $\gamma(0) = e$. Now, we can define a vector field ϕ_γ on X given by:

$$\phi_\gamma(x) = \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} \gamma(t) \cdot x$$

that is, given by the class of the curve $\gamma(t) \cdot x$ through x at time $t = 0$. One can check that the vector field ϕ_γ in fact only depends on v, not the representative γ, so that we have defined a linear map $\phi : \mathfrak{g} \to \mathrm{Vect}(X)$ from the Lie algebra of G to the vector fields on X: we call this **the infinitesimal action** of G.

PROPOSITION 15. *The infinitesimal action $\phi : \mathfrak{g} \to \mathrm{Vect}(X)$ is a homomorphism of Lie algebras: we call this an* ***infinitesimal symmetry*** *of X.*

This is an important check that I've got my sign conventions correct. You should try to prove this in the case where $G = \mathrm{GL}_n(\mathbb{R})$ from above, where the Lie bracket is given by the commutator of matrices. In a precise sense, we can do this without loss of generality, but I'll leave the proof to the problem sheet.

Example 25. If we consider the left action of G on itself, then the infinitesimal action $\mathfrak{g} \to \mathrm{Vect}(G)$ is exactly the map $v \mapsto X_v$ from above. For instance, if we consider the translation action of $\mathbb{R}^n$ on itself, then the vector field associated to $v \in \mathbb{R}^n$ is exactly the constant vector field with value $v \in \mathbb{R}^n$.

QUOTIENT SPACES

Now suppose that G is a Lie group acting on a manifold X: since G gives symmetries of X, there should be a way to eliminate redundant information and pass to a 'reduced space', namely the quotient space X/G: the set of equivalence classes $[x]$ for $x \in X$ with $x \sim g \cdot x$. To the great dismay of differential geometers everywhere, this space X/G need not itself be a smooth manifold and indeed can be a fairly nasty topological space, as the following example shows.

Example 26. Suppose $G = \mathbb{C}^*$ the group of non-zero complex numbers, acting on $\mathbb{C}^2$ by scalar multiplication. Then the point $0 \in \mathbb{C}^2$ is in its own equivalence class; those who know about the quotient topology will know that every open neighbourhood of 0 contains the entire space, and hence X/G fails our Hausdorff requirement for being a smooth manifold: there is really no reasonable way to put a manifold structure on this space. We will see later how this problem may be avoided by use of *symplectic reduction*.

The problems with the previous example are that the group G is *non-compact* and that $0 \in \mathbb{C}^2$ is a *fixed point*. If we fix both of these problems, then we have:

THEOREM 21. ***(Quotient Theorem)*** *Suppose G is a compact Lie group acting freely (no elements $g \in G$ have any fixed points) on a manifold X: then the set X/G has a natural structure of a smooth manifold such that the map $\pi : X \to X/G$ sending a point to its equivalence class is smooth.*

I'm not going to prove this theorem in class: it's a great exercise to try if you're learning differential geometry! This will be a sequence of exercises in the next problem sheet, but I'll sketch the main ideas (how we obtain the charts) in the following few paragraphs. We'll start with a few lemmas:

LEMMA 3. ***(Orbit Lemma)*** *If G is a compact Lie group acting freely on a manifold X, then for every point $x \in X$ the orbit $Gx = \{gx : g \in G\}$ is a closed submanifold of X.*

By **submanifold** we mean that around every point $p \in Gx$ there exists some smooth chart (U, ϕ) in which $Gx \cap U$ is given by the hyperplane $x_1 = x_2 = \cdots = x_k = 0$ inside $\mathbb{R}^n$. We say that a **slice** $S_\varepsilon(p)$ of the action at p is the subset of X given by $\{x_{k+1} = \cdots = x_n = 0\} \cap B_\varepsilon(0)$. Our second lemma concerns these slices:

LEMMA 4. ***(Slice Lemma)*** *Suppose G acts on X as above: then for all $\varepsilon > 0$ sufficiently small and $p \in X$, the map $G \times S_\varepsilon(p) \to X$ given by the group multiplication is a diffeomorphism onto a G-invariant open neighbourhood U of the orbit Gp.*

If U is such a neighbourhood, then we see that U/G is smooth and equal to $S_\varepsilon(p)$: the coordinates near $[p]$ we want to use on the quotient space X/G are precisely given by the slices $S_\varepsilon(p)$, so that the map $X \to X/G$ will necessarily be smooth. The last thing we need to do is check that the transitions between these coordinate patches are smooth. This I will also leave as an exercise.

15 HAMILTONIAN GROUP ACTIONS

Reality is partial to symmetries and slight anachronisms...

Jorge Luis Borges, *The South*

This time we'll continue with our journey to understand symmetries in classical physics. Last time we talked about the notion of a smooth symmetry of a manifold: but what is the right notion of a smooth symmetry of a Hamiltonian dynamical system? Our Hamiltonian systems had extra structure beyond being smooth manifolds: they were equipped with a *symplectic form*. Thus a natural condition might be that the group action respects this structure:

DEFINITION 27. *A Lie group G acting on X a symplectic manifold* ***acts symplectically*** *if every map $\Phi_g : X \to X$ given by $x \mapsto g \cdot x$ is a symplectomorphism.*

This is certainly a necessary condition but it turns out that we'll need a stronger condition than this. Recall that one particular class of symplectomorphisms are those arising as flows of Hamiltonian vector fields. Next, we'll define the notion of a *Hamiltonian group action*, in which every map $\Phi_g : X \to X$ will be given by the flow of a Hamiltonian, but in a manner compatible with the structure of the group. The next few paragraphs should be considered motivational, with the actual definition given below.

Firstly, let's recall the notion of the infinitesimal action of a Lie group G: this was a linear map $\phi : \mathfrak{g} \to \mathrm{Vect}(X)$ given by differentiating the action of the group at a point. If we wanted our action to be given by Hamiltonian flows, then we'd like to assign to every vector field ϕ_v a smooth Hamiltonian function H_v so that ϕ_v is its Hamiltonian vector field. In other words, we'd like a linear map $\mu^* : \mathfrak{g} \to C^\infty(X)$ so that $\phi_v = X_{\mu^*(v)}$ for every $v \in \mathfrak{g}$. Such a map is often called a *comoment map* or *comomentum map* for the group action, for reasons that will become clear shortly. But firstly, we should also require that this assignment of Hamiltonians to elements of $\mathfrak{g}$ be compatible with the group structure in some way! Namely, we should require that μ^* respects the bracket structures on both sides, that is,

$$\mu^*([v, w]) = \{\mu^*(v), \mu^*(w)\}$$

The existence of a comoment map μ^* with this property therefore gives a possible criterion for what it means for the action of G to be *Hamiltonian*.

Unfortunately, since $C^\infty(X)$ is an infinite dimensional space, it is difficult to make sense of what it means for the map μ^* to be *smooth*. So let's look at another (albeit very confusing) way of formulating this criterion: we want a smooth map $\mu : X \to \mathfrak{g}^*$ from X to the *dual* of the Lie algebra: then for every $v \in \mathfrak{g}$, we can take the pairing of $\mu_x \in \mathfrak{g}^*$ and $v \in \mathfrak{g}$ to get $\mu_x(v) \in \mathbb{R}$ at every point $x \in X$ and thus get a smooth function: this should be the Hamiltonian vector field of ϕ_v. Or in symbols:

$$\iota_{\phi_v}\omega = \mathrm{d}\mu(v)$$

DEFINITION 28. *Suppose G is a Lie group acting on a symplectic manifold (X, ω): we say a map $\mu : X \to \mathfrak{g}^*$ is a* ***moment map*** *(or* ***momentum map****) for the action if*

$$\iota_{\phi_v}\omega = \mathrm{d}\mu(v)$$

for every $v \in \mathfrak{g}$.

The terminology *moment map* is more common in modern symplectic geometry and belies its physical origins: the values of the moment map will be exactly the conserved momenta in the conventional sense, as we shall see shortly. Notice again the appearance of the *dual space* when we talk about where momenta live.

Again, for a Hamiltonian group action, we should require some compatibility between the structure of G and the moment map μ. To understand this, we will need to understand how G acts on $\mathfrak{g}^*$: this is called the *coadjoint action* and unfortunately is also very confusing. Recall that we had the *adjoint action* $\mathrm{Ad}_g : \mathfrak{g} \to \mathfrak{g}$ of G on $\mathfrak{g}$ given by differentiating the conjugation action of G on itself. Therefore, we also have a dual map $\mathrm{Ad}_g^* : \mathfrak{g}^* \to \mathfrak{g}^*$: for every $\alpha \in \mathfrak{g}^*$, we have that $\mathrm{Ad}_g^*(\alpha) \in \mathfrak{g}^*$ acts on $v \in \mathfrak{g}$ via

$$\mathrm{Ad}_g^*(\alpha)(v) = \alpha(\mathrm{Ad}_{g^{-1}} v)$$

where we include the g^{-1} so that we get a *left* rather than *right* action on $\mathfrak{g}^*$. This gives an action of G on $\mathfrak{g}^*$ called the **coadjoint action** and the condition we should put on μ is that it preserves the group actions on both sides, namely:

$$\mu(g \cdot x) = \mathrm{Ad}_g^* \mu(x)$$

for every $x \in X$. If our moment map μ has this property, then we say that:

DEFINITION 29. *The action of a Lie group G on a symplectic manifold (X, ω) is said to be **Hamiltonian** (or **Poisson**) if there exists a **moment map** $\mu : X \to \mathfrak{g}^*$ satisfying the following two properties:*

1. *For every $v \in \mathfrak{g}$, we have $\iota_{\phi_v}\omega = \mathrm{d}\mu(v)$;*
2. *(Equivariance) For every $g \in G$, we have $\mu(\Phi_g(x)) = \mathrm{Ad}_g^*\mu(x)$ for every $x \in X$.*

This actually turns out to be equivalent to the earlier definition in terms of the comoment map (exercise!)

EXAMPLES

Now let's see some examples of Hamiltonian group actions:

Example 27. **Translation:** let's consider the action of $G = \mathbb{R}^n$ on $X = \mathbb{R}^{2n}$ with the canonical symplectic form, via translation in the position coordinates:

$$v \cdot (q, p) = (q + v, p)$$

We saw that the infinitesimal action was given by taking $v \in \mathbb{R}^n = T_0\mathbb{R}^n$ to the constant vector field on $\mathbb{R}^{2n}$ with value $(v, 0) \in \mathbb{R}^{2n}$. I claim that in this case the moment map is the map $\mu : \mathbb{R}^{2n} \to \mathbb{R}^n \cong (T_e\mathbb{R}^n)^*$ given by projecting onto the momentum coordinates: $(q, p) \mapsto p$. To see that this indeed gives a moment map, let's take $v = e_i \in \mathbb{R}^n$ a basis vector of the Lie algebra. Then $\mu(e_i)$ is the smooth function given by $p \cdot e_i = p_i$. What is the Hamiltonian vector field of this function? We can just use ordinary calculus on $\mathbb{R}^{2n}$ to calculate $X_H = -J\nabla p_i$ and see that the vector field X_H is exactly the constant vector field with value $(e_i, 0)$ on $\mathbb{R}^{2n}$. Therefore we have a moment map! Finally, since $\mathbb{R}^n$ is an abelian group, its action by conjugation is trivial and we just need to check that $\mu(q + v, p) = \mu(q, p)$, which is clearly true since μ only depends on p. Thus this action is Hamiltonian! As we shall see shortly, this captures the idea that the conserved quantity associated to translation symmetry of $\mathbb{R}^n$ is exactly the linear momentum.

Example 28. **Rotation:** This example will be given as a series of exercises that you should check! Let $G = \mathrm{SO}(3)$ acting on $\mathbb{R}^3$ by rotations. Let's first consider the Lie algebra of $\mathrm{SO}(3)$: you should show that it is given by the 3×3 real skew-symmetric matrices A: those matrices satisfying $A^T = -A$. We can write these out explicitly as:

$$A = \begin{pmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ a_2 & a_1 & 0 \end{pmatrix}$$

and hence identify $A \in \mathfrak{g}$ with the vector $(a_1, a_2, a_3) \in \mathbb{R}^3$. Then you should check that under this identification, the adjoint action of $\mathrm{SO}(3)$ on $\mathfrak{g}$ becomes the usual action of $\mathrm{SO}(3)$ on $\mathbb{R}^3$ by rotation Moreover, under this identification, the commutator bracket of matrices $[A, B] = AB - BA$ is sent to the cross product of vectors in $\mathbb{R}^3$!

But what about the *coadjoint action*? This is where we need the distinction between *active* and *passive* transformations: when I consider an element of $\mathrm{SO}(3)$ acting on $\mathbb{R}^3$, I can either imagine it rotating a vector with respect to a fixed orthogonal coordinate system (the *active* perspective), or rotating an orthogonal coordinate system around a fixed vector (the *passive* perspective). In fact, these two perspectives are dual to each other in a precise sense: once we have chosen a basis for $\mathfrak{g}$ as above, the coadjoint action of $\mathrm{SO}(3)$ on $\mathfrak{g}^*$ is exactly given by the *passive rotations* of $\mathbb{R}^3$ (while the adjoint representation is given by the *active* transformations). You can check this by observing that if a vector v has coordinates (a_1, a_2, a_3) in a given orthogonal basis e_1, e_2, e_3 of $\mathbb{R}^3$, then v has coordinates $T^{-1}(a_1, a_2, a_3)$ with respect to the orthogonal basis Te_1, Te_2, Te_3, for any $T \in \mathrm{SO}(3)$.

Now, the action of $\mathrm{SO}(3)$ lifts to an action on the phase space $X = \mathbb{R}^6$ via $A \cdot (q, p) = (Aq, Ap)$. You should use the fact that the action of $\mathrm{SO}(3)$ on $\mathbb{R}^3$ by rotation, and the fact that the Lie bracket is given by the cross product, to show that the infinitesimal action on $\mathbb{R}^6$ by $v \in \mathbb{R}^3$ is given by the vector field $\phi_v : \mathbb{R}^6 \to \mathbb{R}^6$ equal to $\phi_v(q, p) = (v \times q, v \times p)$.

Now, if we equip $\mathbb{R}^6$ with the canonical symplectic form, I claim that this action is Hamiltonian, with moment map $\mu : \mathbb{R}^6 \to \mathbb{R}^3 \cong (\mathbb{R}^3)^*$ given by

$$\mu(q, p) = q \times p$$

so that for every $v \in \mathbb{R}^3$, we get the smooth function $\mu(a)(q, p) = (q \times p) \cdot v$. Note that this gives us exactly the usual notion of angular momentum in the direction v. To see that this is a moment map, it suffices to use elementary calculus to check that:

$$-J\nabla[(q \times p) \cdot v] = (v \times q, v \times p)$$

for all $p, q, v \in \mathbb{R}^3$. I'll leave verifying the equivariance condition to you.

NOETHER'S THEOREM

Both of the previous examples are special cases of the following, which is the proper formulation of Noether's theorem in the context of symplectic geometry:

THEOREM 22. *(**Noether's Theorem**) Suppose G is a Lie group acting on a smooth manifold Q: then this lifts to an action of G on the cotangent bundle T^*Q, and this action is Hamiltonian with respect to the canonical symplectic form on T^*Q.*

In other words, smooth symmetries of the configuration space become Hamiltonian symmetries of the phase space. Now, in the language of differential forms we can give a suitably beautiful proof of this beautiful theorem. We'll begin with a Lemma:

LEMMA 5. *Firstly, suppose $\Phi : Q \to Q$ is a diffeomorphism of Q: we can lift it to a smooth map $\tilde{\Phi} : T^*Q \to T^*Q$ such that $\tilde{\Phi}^*\lambda = \lambda$.*

Proof. Unfortunately, this proof is going to look like a giant mess of symbols: but it's entirely straightforward. Given $q \in T_q^*Q$, we define

$$\tilde{\Phi}(p, q) = (\Phi(q), (\Phi^{-1})^* p)$$

where here $(\Phi^{-1})^*$ is chosen so that $(\Phi^{-1})^*(p) \in T^*_{\Phi(q)}Q$. This is a lift of the map Φ in the sense that $\pi \circ \tilde{\Phi} = \Phi \circ \pi$ where $\pi : T^*Q \to Q$ is the projection $(q, p) \mapsto q$.

Now, this map has the property that $\tilde{\Phi}^*\lambda = \lambda$: for $v \in T_{(q,p)}(T^*Q)$ we have

$$\left(\tilde{\Phi}^*\lambda\right)_{(q,p)}(v) = \lambda_{\tilde{\Phi}(q,p)}(D_{(q,p)}\tilde{\Phi}(v)) = ((\Phi^{-1})^*p)(D_{\tilde{\Phi}(q,p)}\pi(D_{(q,p)}\tilde{\Phi}(v)))$$

Now, by the fact that $\tilde{\Phi}$ is a lift of Φ, using the chain rule gives

$$((\Phi^{-1})^*p)(D_q\Phi \circ D_{(q,p)}\pi(v)) = p(D_q\Phi^{-1} \circ D_q\Phi \circ D_{(q,p)}\pi(v)) = p(D_{(q,p)}\pi(v))$$

and so we are done. ■

Proof. (of Noether's Theorem) Now if G is a Lie group acting on Q, then we can replace each map $\Phi_g : Q \to Q$ with $\tilde{\Phi}_g : T^*Q \to T^*Q$ and this will give us a lift of the action to T^*Q (this is one reason why we needed the *inverse* of the pullback in the definition above). Now, let $\phi : \mathfrak{g} \to \mathrm{Vect}(T^*Q)$ be the infinitesimal action of G on T^*Q and observe that for every $v \in \mathfrak{g}$, the flow of the vector field $\phi(v)$ through a point $x \in X$ is given by the group action by $\gamma(t) = \mathrm{e}^{tv} \cdot x$, since

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \mathrm{e}^{(t+s)v} \cdot x = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \mathrm{e}^{tv}(\mathrm{e}^{sv} \cdot x) = \phi(v)_{\mathrm{e}^{sv}\cdot x}$$

Therefore, we can write the Lie derivative as:

$$\mathcal{L}_{\phi(v)}\lambda = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \Phi^*_{\mathrm{e}^{tv}}(\lambda) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \lambda = 0$$

Now, if we apply Cartan's magic formula, we get:

$$0 = \mathcal{L}_{\phi(v)}\lambda = \mathrm{d}\lambda(\phi(v)) + \iota_{\phi(v)}\mathrm{d}\lambda$$

Or in other words, $\mathrm{d}(\lambda(\phi(v))) = \iota_{\phi(v)}\omega$, that is, $\lambda(\phi(v))$ is exactly the Hamiltonian function for $\phi(v)$. I leave it to you to check that in fact this also satisfies the equivariance property. ■

You can check that the formula $\lambda(\phi(v))$ does indeed give the conserved momenta from the previous examples (and is, in coordinates, exactly the same expression we obtained for Noether's theorem in Lecture 2). You might now ask how exactly this theorem above gives rise to a conservation law in the usual sense of Noether's theorem. The connection is given by:

PROPOSITION 16. ***(Conservation of Generalized Momentum)*** *Suppose G is a Lie group with a Hamiltonian action on a symplectic manifold (X, ω) with moment map $\mu : X \to \mathfrak{g}^*$. Suppose $H : X \to \mathbb{R}$ is a Hamiltonian function that is G-invariant: then the moment map μ is conserved under the Hamiltonian flow of H.*

Proof. This proof makes a great exercise, so I'd suggest you give it a try first! So what are we trying to prove? Suppose that X_H is the Hamiltonian vector field of H, $\gamma(t)$ is a flow line of X_H, and take $v \in \mathfrak{g}$: we want to look at:

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \mu(v)(\gamma(t)) = d\mu(v)(X_H)$$

by the definition of the differential (we've seen this before, but you can also see this as a special case of Cartan's magic formula). Now by the fact that this action is Hamiltonian and the asymmetry of the symplectic form:

$$d\mu(v)(X_H) = \omega(\phi(v), X_H) = -\omega(X_H, \phi(v)) = -\mathrm{d}H(\phi(v))$$

where we have used the definition of the vector field X_H. Now what is $\mathrm{d}H(\phi(v))$? If $\alpha : \mathbb{R} \to G$ is a path representing v, then this is given by:

$$\mathrm{d}H(\phi(v)) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} H(\alpha(t)\cdot x)$$

and since H is G-invariant:

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} H(\alpha(t)\cdot x) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} H(x) = 0$$

so that hence $\mu(v)$ is constant along flow lines of X_H. ■

Notice that to prove this theorem, we only needed an *infinitesimal* Hamiltonian action of G on X, that is, a Lie algebra homomorphism $\phi : \mathfrak{g} \to \mathrm{Vect}(X)$ that had a moment map! This makes precise the sense in which Noether's theorem is true for *infinitesimal symmetries* of the Hamiltonian: those for which $\mathrm{d}H(\phi(v)) = 0$ for all $v \in \mathfrak{g}$.

SYMPLECTIC REDUCTION

We didn't come all this way just to recover Noether's theorem! Recall that when we first introduced the Hamiltonian formalism, one of the key advantages was that it had *more symmetries* than the Newtonian or Lagrangian formulations: this extra freedom to interchange position and momentum is formalized in the fact that there are (Hamiltonian) Lie group actions on T^*Q that do not arise simply from Lie group actions on the configuration space Q! Let's look at an important example:

Example 29. Let $G = S^1$, which we will identify with the unit complex numbers, and let $X = \mathbb{R}^{2n}$, which we will identify with $\mathbb{C}^n$, and equip it with the standard symplectic form ω, which we may write in polar coordinates as:

$$\omega = \sum_{i=1}^{n} r_i \mathrm{d}r_i \wedge \mathrm{d}\theta_i$$

Now S^1 acts on $\mathbb{C}^n$ via rotation of the arguments: $t(z_1, \ldots, z_n) = (z_1 \mathrm{e}^{it}, \ldots, z_n \mathrm{e}^{it})$. From the form of the symplectic form in polar coordinates, we can clearly see that this action is symplectic: we claim furthermore that it is actually Hamiltonian, with moment map $\mu : X \to \mathbb{R} \cong T_1^* S^1$ given by

$$\mu(z_1, \ldots, z_n) = -\frac{1}{2}\sum_{i=1}^{n} |z_i|^2 + C$$

for any constant C. Let's check this by first working out the infinitesimal action: for $v \in \mathbb{R}$ the unit vector we have:

$$\phi(v)(z_1, \ldots, z_n) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} (r_1, \ldots, r_n, \theta_1 + t, \ldots \theta_n + t) = (0, \ldots, 0, 1, \ldots, 1)$$

the constant vector field with all 1s in the angular components. On the other hand, we have

$$\mathrm{d}\mu(v) = \mathrm{d}\left(-\frac{1}{2}\sum_{i=1}^{n} r_i^2 + C\right) = -\sum_{i=1}^{n} r_i \mathrm{d}r_i = \iota_{\phi(v)}\left(\sum_{i=1}^{n} r_i \mathrm{d}r_i \wedge \mathrm{d}\theta_i\right)$$

so that μ does indeed give a moment map for this S^1 action.

Now let's suppose we are studying Hamiltonian mechanics on $\mathbb{R}^{2n}$ and our Hamiltonian is symmetric under the action of $G = S^1$. Then by the conservation of the moment map, we know that the Hamiltonian dynamics of H is constrained to lie on a level set of μ, say $\mu^{-1}(0)$. Moreover, since H is S^1-invariant, the dynamics of H should in some sense take place on the quotient space $\mu^{-1}(0)/S^1$. But, we already know what this quotient space is! Firstly, if we take $C = 1/2$, then $\mu^{-1}(0) = S^{2n-1} \subseteq \mathbb{R}^{2n}$, the unit sphere! Moreover, we have that:

$$\{z \in \mathbb{C}^n : \ |z| = 1\} \,/\, \left\{z_1 \sim z_2 \text{ iff } z_1 = \mathrm{e}^{it} z_2\right\} = \{z \in \mathbb{C}^n : |z| \neq 0\} \,/\, \{z_1 \sim z_2 \text{ iff } z_1 = \lambda z_2 \text{ for } \lambda \in \mathbb{C}^*\}$$

since we can use the modulus of $\lambda \in \mathbb{C}^*$ to scale $z \in \mathbb{C}^n \setminus \{0\}$ so that it has $|z| = 1$. But the latter set was exactly our definition of the complex projective space $\mathbb{CP}^{n-1}$, which we saw was a symplectic manifold! This is the first example of what we will see next time: the *Marsden-Weinstein theorem*, which states that the *reduced space* $\mu^{-1}(0)/G$ is also a symplectic manifold, and captures the original Hamiltonian dynamics.

16 MARSDEN-WEINSTEIN THEOREM

Symmetry, as wide or as narrow as you may define its meaning, is one idea by which man through the ages has tried to comprehend and create order, beauty and perfection.

Hermann Weyl, *Symmetry*

Without further ado, here is the main theorem of the course:

THEOREM 23. ***(Marsden-Weinstein Theorem)*** *Suppose G is a compact Lie group with a Hamiltonian action on a symplectic manifold (X, ω), with moment map $\mu : X \to \mathfrak{g}^*$. Suppose that G acts freely on $\mu^{-1}(0)$. Then:*

1. *The quotient space $\mu^{-1}(0)/G$ is a smooth symplectic manifold with a symplectic form ω_r, called the* ***symplectic reduction*** *or* ***Marsden-Weinstein quotient****, such that if $\pi : \mu^{-1}(0) \to \mu^{-1}(0)/G$ is the projection, then $\pi^*\omega_r = \omega$.*
2. *If a smooth function $H : X \to \mathbb{R}$ is invariant under the action of G, then H descends to a smooth function $H : \mu^{-1}(0)/G \to \mathbb{R}$ whose Hamiltonian vector field with respect to the reduced symplectic structure ω_r is equal to $D\pi(X_H)$ for X_H the Hamiltonian vector field of H on X.*

Remark 3. Note that this theorem goes under many different names, for instance, the *symplectic reduction theorem*, the *Marsden-Weinstein-Meyer-Ratiu theorem* etc. As Arnol'd notes, the ideas of this theorem go all the way back to work of Poisson, Jacobi and Poincaré, who called it *reduction of order* or *elimination of the nodes*, though a modern formulation in terms of manifolds first appeared in a paper of Smale in 1970. You know it's an important theorem when it's been reproved numerous times over 200 years.

Despite the fact that this is an important theorem with many applications, it is in fact quite easy to prove: I'll sketch the main ideas here an leave the rest as a series of exercises.

Firstly, we need to show that this space $\mu^{-1}(0)/G$ actually makes sense as a manifold! We should first show that the group action preserves $\mu^{-1}(0)$! This follows from the equivariance: if $x \in \mu^{-1}(0)$, then for every $g \in G$, if we look at $\Phi_g(x)$, then we have

$$\mu(\Phi_g x) = \mathrm{Ad}_g^* \mu(x) = \mathrm{Ad}_g^*(0) = 0$$

since the coadjoint action was by linear maps (just the differentials of the adjoint action). Hence $\Phi_g(x) \in \mu^{-1}(0)$ also.

Next, we should show that the group orbits $Gx = \{gx : g \in G\}$ are actually smooth manifolds! This will follow from the following:

LEMMA 6. ***(Orbit Lemma)*** *If G is a compact Lie group acting freely on any smooth manifold X, then for every point $x \in X$ the orbit $Gx = \{gx : g \in G\}$ is a closed submanifold of X: namely the map $G \to X$ given by $g \mapsto g \cdot x$ is injective and has injective differential at every point.*

Proof. Recall that a group action is *free* if the maps $\Phi_g : X \to X$ have no fixed points unless g is the identity. From this, it is clear that the map is injective. To see that the differential is injective, we first consider the differential at the identity $e \in G$: but this is exactly the infinitesimal action at x, that is, the map $\phi(x) : \mathfrak{g} \to T_xX$. We want to show that this map is injective, so suppose we have some $v \in \mathfrak{g}$ for which $\phi(x)(v) = 0$.

If we temporarily assume that G is a matrix Lie group, so that for every $v \in \mathfrak{g}$ we have a path through the identity representing v given by $\gamma(t) = \mathrm{e}^{tv}$, then we can calculate that for all $s \in \mathbb{R}$ we have

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \mathrm{e}^{(t+s)v} \cdot x = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \mathrm{e}^{sv} \cdot (\mathrm{e}^{tv} \cdot x) = D\Phi_{\mathrm{e}^{sv}} \left(\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \mathrm{e}^{tv} \cdot x \right) = D\Phi_{\mathrm{e}^{sv}}(0) = 0$$

Therefore we see that $\mathrm{e}^{tv} \cdot x = x$ for all $t \in \mathbb{R}$: but this contradicts the fact that the action was free unless $\mathrm{e}^{tv} = \mathrm{Id}$ for all t: in this case, e^{tv} represents the constant path and so $v = 0$. Now, for a general Lie group, we can in fact define an *exponential map* $\exp : \mathfrak{g} \to G$ with the same property that e^{tv} at 0 represents v and $\mathrm{e}^{(t+s)v} = \mathrm{e}^{tv}\mathrm{e}^{sv}$ for all $s, t \in \mathbb{R}$. ■

Being a **submanifold** means that around every point $p \in Gx$ there exists some smooth chart (U, ϕ) in which $Gx \cap U$ is given by the hyperplane $x_1 = x_2 = \cdots = x_k = 0$ inside $\mathbb{R}^n$. We say that a **slice** $S_\varepsilon(p)$ of the action at p is the subset of X given by $\{x_{k+1} = \cdots = x_n = 0\} \cap B_\varepsilon(0)$. The *quotient theorem* says that these give local coordinates on the quotient space X/G:

THEOREM 24. *(**Quotient Theorem**) Suppose G is a compact Lie group acting freely on a manifold X: then the set X/G has a natural structure of a smooth manifold with local coordinates given by the slices of the action, such that the map $\pi : X \to X/G$ sending a point to its equivalence class is smooth.*

But to apply the quotient theorem, we need to show that $\mu^{-1}(0)$ is actually a smooth manifold! This will in fact always be true when G acts freely on $\mu^{-1}(0)$:

PROPOSITION 17. *For every $x \in \mu^{-1}(0)$, we have that the kernel of the map $D_x\mu : T_xX \to T_0\mathfrak{g}^* \cong \mathfrak{g}^*$ is given by the symplectic orthogonal to the image of the infinitesimal action at x: that is*

$$\ker D_x\mu = \{w \in T_xX : \ \omega(w, \phi(v)) = 0 \ \textit{for all} \ \ v \in \mathfrak{g}\}$$

Furthermore, $D_x\mu$ is surjective.

Proof. Both of these statements will follow from the basic relation:

$$\omega_x(\phi(v), w) = (D_x\mu(w), v)$$

expressing the fact that μ is a moment map, where $(\cdot, \cdot)$ is the pairing $\mathfrak{g}^* \times \mathfrak{g} \to \mathbb{R}$. Firstly, we have that w is in the kernel of $D_x\mu$ if and only if $(D_x\mu(w), v) = 0$ for all $v \in \mathfrak{g}$, if and only if $\omega_x(\phi(v), w) = 0$ for all $v \in \mathfrak{g}$: but this is exactly the symplectic orthogonal to the image of the infinitesimal action.

Similarly, $D_x\mu$ is surjective if and only if $(D_x\mu(w), v) = 0$ for all $w \in T_xX$ implies that $v = 0$. But, we know that $(D_x\mu(w), v) = 0$ implies $\omega_x(\phi(v), w) = 0$ for all $w \in T_xX$ and hence $v = 0$ since ϕ is injective and ω_x is non-degenerate. ■

Now, the surjectivity of $D_x\mu$ means that $\mu^{-1}(0)$ is a smooth manifold, by a version of the implicit function theorem:

THEOREM 25. *(**Transversality Theorem**) Suppose $F : X \to \mathbb{R}^n$ is a smooth function with $D_xF : T_xX \to \mathbb{R}^n$ surjective at every point $x \in F^{-1}(c)$. Then $F^{-1}(c)$ is a smooth manifold.*

The next thing to do is consider the tangent space to $\mu^{-1}(0)/G$:

LEMMA 7. *1. the tangent space at $x \in X$ to a G-orbit Gx is given by the image of the infinitesimal action ϕ at x;*

2. *the tangent space to $\mu^{-1}(0)$ at a point $x \in \mu^{-1}(0)$ is given by the kernel of $D_x\mu$;*

3. *the tangent space to $[x] \in \mu^{-1}(0)/G$ is naturally identified with the quotient* $\ker D_x\mu/T_x(Gx)$.

I'll leave the proof of this lemma as an exercise. Now, the Marsden-Weinstein theorem will follow from a simple linear algebra result:

LEMMA 8. *If V is a vector space with a symplectic form ω, and W is a subspace on which the symplectic form vanishes, then W^ω/W has a natural symplectic structure, where W^ω is the symplectic orthogonal to W.*

Proof. Define ω_r on W^ω/W by $\omega_r([v],[w]) = \omega(v,w)$. The fact that ω is zero on W will mean that ω_r is well-defined and non-degenerate, while antisymmetry and bilinearity are apparent. ∎

Proof. (*of the Marsden-Weinstein Theorem*) We can use this lemma to equip $T_{[x]}\mu^{-1}(0)/G$ with a reduced symplectic form ω_r: we identify $T_{[x]}\mu^{-1}(0)/G = \ker D_x\mu/T_x(Gx)$, and recall that we showed that $\ker D_x\mu$ was the symplectic orthogonal $(T_x(Gx))^{\omega_x}$. This defines a symplectic form ω_r on $\mu^{-1}(0)/G$ and once we have checked that $\mathrm{d}\omega_r = 0$ we will have shown that $\mu^{-1}(0)/G$ is a symplectic manifold!

The differential of the map $\pi : \mu^{-1}(0) \to \mu^{-1}(0)/G$ can be identified with the quotient map $\ker D_x\mu \to T_{[x]}\mu^{-1}(0)/G$: then $\pi^*\omega_r = \omega$ by definition of ω_r. Since the exterior derivative d commutes with pullbacks and π^* is injective, this implies that $\mathrm{d}\omega_r = 0$, as desired.

Proving part (2) of the theorem is then straightforward: if $\tilde{H} : \mu^{-1}(0)/G \to \mathbb{R}$ denotes the smooth function induced on $\mu^{-1}(0)/G$ from the G-invariant smooth function H on X, then we have that:

$$\mathrm{d}H = \mathrm{d}(\pi^*\tilde{H}) = \pi^*\mathrm{d}\tilde{H}$$

and similarly,

$$\pi^*(\iota_{D\pi(X_H)}\omega_r) = \iota_{X_H}\pi^*\omega_r = \iota_{X_H}\omega$$

where we think of $D\pi$ as the quotient map. Thus $\pi^*(d\tilde{H}) = \pi^*(\iota_{D\pi(X_H)}\omega_r)$. Since π^* is injective, we hence have that $D\pi(X_H)$ is the Hamiltonian vector field of $\tilde{H}$ and we are done. ∎

APPLICATIONS TO RIGID BODIES

FREE RIGID BODIES

Before we discuss how to use the Marsden-Weinstein theorem to prove Poinsot's theorem, let's first talk about some generalities on rigid bodies.

Recall that a *rigid body* was defined to be a configuration of point-like bodies x_i in $\mathbb{R}^3$ with the system of holonomic constraints of the form $|x_i - x_j| = C$: they are required to remain the same distance apart. For the purposes of the following discussion, let's assume that all of the points in our rigid body don't lie in a plane. The configuration space Q of our rigid body is cut out from some $\mathbb{R}^{3n}$ by this system of equations, but this isn't exactly a very useful description, so let's try and put coordinates on this space by hand that parametrize all of the configurations of a rigid body:

PROPOSITION 18. *The configuration space of a rigid body (not lying in a plane) is diffeomorphic to* $\mathrm{SO}(3) \times \mathbb{R}^3$, *where* $\mathrm{SO}(3)$ *is the manifold of orthogonal* 3×3 *matrices with determinant* 1.

First off, we can always translate a rigid body in space: these translations are parametrized by a copy of $\mathbb{R}^3$ given by the position of a point p in the body. Now, if we fix this point p at the origin, then we can also consider rotations of the rigid body about the point p. We can imagine the rigid body coming equipped with three orthonormal vectors e_1, e_2, e_3 living inside the rigid body: the positions of these three vectors in space completely specifies the rotation of the rigid body. Moreover, for any two orthonormal bases for $\mathbb{R}^3$, there exists a unique orthogonal matrix A relating them. Therefore we identify the configuration space of the rigid body with $Q = \mathbb{R}^3 \times \mathrm{SO}(3)$ (for those in the know, as a *group*, this is actually a *semidirect product* rather than a usual product, but as a *manifold*, these two are the same.) The dynamics of a rigid body fixed at a point is therefore described by paths in the space $\mathrm{SO}(3)$.

DEFINITION 30. *Given a path $\gamma : \mathbb{R} \to \mathrm{SO}(3)$, the **angular velocity** ω_c at time t (in the **body frame**) is the element of the Lie algebra $\mathfrak{g}$ of $\mathrm{SO}(3)$ given by:*

$$\omega_c = DL_{\gamma(t)^{-1}}(\gamma'(t))$$

In other words, we translate the path back to the identity using left multiplication, and then take the derivative. Elements of $\mathfrak{g}$ will be referred to as angular velocities. The term **body frame** is included because we could have instead used *right multiplication* to translate back to the identity: this would give us the angular momentum in the **space frame**. This distinction corresponds to whether we consider a rotation as rotating the body or the ambient space around it.

Let's consider just the **free rigid body**, that is, the dynamics in the absence of any external forces: then the Hamiltonian H will simply be given by the kinetic energy (expressed in terms of a suitable momentum). This kinetic energy should depend only on the angular velocity, and not on the position in space: therefore there should be a positive definite inner product $\langle \cdot, \cdot \rangle$ on $\mathfrak{g}$ so that the kinetic energy T at a point $g \in \mathrm{SO}(3)$ is given on $v1, v_2 \in T_g\mathrm{SO}(3)$ by:

$$T_g(v_1, v_2) = \langle DL_{g^{-1}}(v_1), DL_{g^{-1}}(v_2) \rangle$$

This inner product $\langle \cdot, \cdot \rangle$ contains all of the information about which specific rigid body we are considering: we won't actually need to know anything more about it!

DEFINITION 31. *The **moment of inertia tensor** A is the identification $A : \mathfrak{g} \to \mathfrak{g}^*$ given by the positive definite inner product $\langle \cdot, \cdot \rangle$. Given an angular velocity $\omega_c \in \mathfrak{g}$, the image $A(\omega_c)$ in $\mathfrak{g}^*$ is called the **angular momentum** (in the **body frame**). We call elements of $\mathfrak{g}^*$ **angular momenta**.*

In concrete terms, the moment of inertia is a 3×3 matrix which tells us how to convert velocities (tangent vectors) into momenta (cotangent vectors), just like the usual mass relates linear velocity to linear momentum. It measures the difficulty of rotating a body around different axes, and is easily determined.

Now, the Hamiltonian H will be a function on the phase space $T^*\mathrm{SO}(3)$ given at a point $g \in \mathrm{SO}(3)$ on $p_1, p_2 \in T_g^*\mathrm{SO}(3)$ by:

$$H_g(p_1, p_2) = \langle A(L_g^* p_1), A(L_g^* p_2) \rangle$$

In other words, we use the identification of $T_g^*\mathrm{SO}(3)$ with $T_g\mathrm{SO}(3)$ given by the moment of inertia to rewrite the kinetic energy as a function of the momentum in terms of the inner product on $\mathfrak{g}$.

SYMPLECTIC REDUCTION IN ACTION

In the above, I jumped to describing the motion of a rigid body fixed at a point using Hamiltonian mechanics. Why can we do this? Let's show rigorously that we can *decouple* the translational and rotational motions: we can do this using symplectic reduction by the action of $\mathbb{R}^3$ on X by translation, which is a symmetry of the Hamiltonian.

We have seen before how $\mathbb{R}^3$ acts on $\mathbb{R}^6$ with moment map given by $\mu : \mathbb{R}^6 \to \mathbb{R}^3$ taking the momentum coordinates, and we just define the action to be trivial on the $T^*\mathrm{SO}(3)$ factor. This will again give us a Hamiltonian action. Now, if we look at the level set $\mu^{-1}(0)$, this will be equal to $\mathbb{R}^3 \times T^*\mathrm{SO}(3)$; the action of $\mathbb{R}^3$ on this $\mu^{-1}(0)$ is just translation on the $\mathbb{R}^3$ factor: but under this action, all points in $\mathbb{R}^3$ are equivalent, since any two vectors in $\mathbb{R}^3$ differ by a translation. Therefore, $\mu^{-1}(0)/\mathbb{R}^3 \cong T^*\mathrm{SO}(3)$, and the Marsden-Weinstein theorem says that the dynamics on H on this space are the same as the original dynamics on X. It's not hard to see that in fact the reduced symplectic structure we get from the theorem coincides with the canonical symplectic structure on the cotangent bundle.

Now, let's consider another symmetry of this situation: $\mathrm{SO}(3)$ acts itself by left multiplication, and this lifts to give an action of $\mathrm{SO}(3)$ on $T^*\mathrm{SO}(3)$ as we saw before, which gives us another symmetry of the Hamiltonian. One can check that the moment map given by Noether's theorem for this action has the explicit form $\mu : T^*\mathrm{SO}(3) \to T_e^*\mathrm{SO}(3)$ given by $\mu(g,p) = R_g^* p$. This moment map gives the three components of angular momentum for this rigid body, and the moment map says that to find the angular momentum in some position, we first rotate back to the original position of the rigid body, and then calculate the angular momentum there.

Now, we want to consider some non-zero value $v \in T_e^*\mathrm{SO}(3) \cong \mathbb{R}^3$ of the angular momentum, so that our rigid body will actually be rotating! However, taking $\mu^{-1}(v)$ doesn't fall under the scope of our previous discussion of symplectic reduction: but we can apply the theory anyway! First off, note that $\mu^{-1}(v)$ isn't preserved by the whole group $\mathrm{SO}(3)$. Recall that the coadjoint action of $\mathrm{SO}(3)$ on $T_e^*\mathrm{SO}(3) \cong \mathbb{R}^3$ was exactly given by rotation: if we assume that $|v| = R$, then the adjoint action of $\mathrm{SO}(3)$ on v sweeps out the whole sphere S^2. Then by the previous argument, the action of $\mathrm{SO}(3)$ on $T^*\mathrm{SO}(3)$ actually preserves $\mu^{-1}(S^2)$! Therefore, we can look at the symplectic reduction with respect to $\mathrm{SO}(3)$ on $\mu^{-1}(S^2)$, namely the quotient $\mu^{-1}(S^2)/G$. But, since the right action R_g* preserves the lengths of vectors $v \in \mathbb{R}^3$, we just have that $\mu^{-1}(S^2) = \{(g,p) :\ g \in G,\ p \in S^2 \subseteq T_e^*\mathrm{SO}(3)\}$. Therefore the quotient is exactly S^2! We have seen before that this has a symplectic structure, and again, one can show that this symplectic structure agrees with the one coming from the Marsden-Weinstein theorem. We can think of this S^2 as parametrizing the possible axes of rotation of the rigid body, and the Marsden-Weinstein theorem says that the Hamiltonian dynamics of the rigid body are entirely captured by the dynamics of the axis of rotation with respect to the restricted Hamiltonian.

Now, let's think about the restricted Hamiltonian: it is some quadratic function of the momenta in $\mathbb{R}^3 \cong T_e^*\mathrm{SO}(3)$ that we restrict to S^2. Now, you should recall that for any (non-degenerate) quadratic function on $\mathbb{R}^3$, there exists an orthonormal basis v_1, v_2, v_3 in which H has the form $H = \lambda_1 p_1^2 + \lambda_2 p_2^2 + \lambda_3 p_3^2$ for $\lambda_i > 0$. We call the v_is the **principal axes** around which the rigid body can rotate; the numbers λ_i measure how difficult it is to get the body to rotate around the axis v_i, and we call it the **moment of inertia** along v_i. Let's assume for simplicity that $\lambda_1 > \lambda_2 > \lambda_3$: then we call v_1 the **major axis**, v_3 the **minor axis**, and v_2 the **intermediate axis**. This brings us to the following:

THEOREM 26. ***(Poinsot's Theorem/Intermediate Axis Theorem/Tennis Racket Theorem)*** *The rotations of a rigid body around the major and minor axes are stable, but rotations around the intermediate axis are unstable.*

We'll see what we mean by the terminology *stable*: roughly it means that a small perturbation of the trajectory will stay in a small neighbourhood of the original trajectory for all time.

Proof. This is now reduced simply to a question about Hamiltonian dynamics on $S^2 \subseteq \mathbb{R}^3$ with respect to the Hamiltonian $H = \lambda_1 p_1^2 + \lambda_2 p_2^2 + \lambda_3 p_3^2$. But since the Hamiltonian H must be conserved under any dynamical system, the trajectories for X_H must lie on the intersection of S^2 (of radius R), and the ellipsoid $H = \lambda_1 p_1^2 + \lambda_2 p_2^2 + \lambda_3 p_3^2 = S$. Let's imagine what the intersection looks like as we increase R from 0 to

∞. At first, the sphere intersects the ellipsoid only near the shortest axis, in small circles; similarly, when R is large, the sphere only intersects the ellipsoid near the longest axis along small circles. Therefore we see that small perturbations in the value of the energy or the angular momentum mean that the body continues to move in small circles: we call this behaviour **precession**. Finally, in between these two axes, the sphere will intersect the ellipsoid along long arcs from one side of the sphere to the other (see figure), meaning that the axis of rotation will flip from one side of the sphere to the other given a small perturbation in either the energy or the angular momentum: this is what we mean by the trajectories being unstable. ■

This theorem corresponds to a real physical effect you can observe with rotating objects, called the Dzhanibekov effect: it's best observed in space, where the assumptions of our theorem approximately hold, and you can see the spooky behaviour of the body as it flips back and forth between rotating around two different axes. Notice how we managed to show the existence of this behaviour without needing to write down a single differential equation!

One doesn't have to be in space to see the power of symplectic reduction: if we instead consider a rigid body (fixed at one point) in an axially symmetric force field (such as gravity near the Earth's surface), then one still has an S^1 symmetry of the rigid body, given by rotating around this axis. The reduced space will now be T^*S^2 and one can similarly study the stationary rotations using tools from Morse theory: in fact, there will always be at least two stationary rotations!

17 ARNOL'D-LIOUVILLE THEOREM

Nothing is so stifling as symmetry. Symmetry is boredom, the quintessence of mourning. Despair yawns.

Victor Hugo, *Les Misérables*

One of our main motivations for introducing the Hamiltonian formalism was that it had a much greater number of symmetries! More symmetries means more coordinate changes and hence more opportunities to write our problem in coordinates where the problem is particularly simple. These will be the *action-angle coordinates*, and the *Arnol'd-Liouville theorem* we discuss today will be another result that describes the geometry associated to conserved quantities, much like the Marsden-Weinstein theorem.

Firstly, let's see an example where the choice of the right coordinates can put a problem into an extremely simple form:

Example: Consider the harmonic oscillator problem, which was given by the Hamiltonian $H = p^2 + \omega^2 q^2$ on the phase space $\mathbb{R}^2$: here ω is the angular frequency of the oscillation. Now, let's transform this problem into angular coordinates, given by $q = \sqrt{I/\omega}\sin\theta$ and $p = \sqrt{I\omega}\cos\theta$. We should check that this actually gives us a symplectomorphism, by calculating $\omega = \mathrm{d}I \wedge \mathrm{d}\theta$ in these coordinates. Now, in these coordinates, we will have that $H = I\omega\cos^2\theta + I\omega\sin^2\theta = \omega I$. Therefore we have reduced to a new problem in which H is only a function of one of the variables! This means that the form taken by Hamilton's equations in these coordinates is very simple:

$$\dot\theta = \frac{\partial H}{\partial I} = \omega$$
$$\dot I = \frac{\partial H}{\partial \theta} = 0$$

Therefore, I remains constant and $\theta(t) = \omega t$, so that the motion is simply given by a straight line in these coordinates. Geometrically, I determines which ellipse in the plane on which the movement will lie, and θ traces out the ellipse at constant speed. We have succeeded in reducing our problem to one that is as trivial as we could hope for! This is the simplest example of what we will call *action-angle coordinates*: here θ is the angle and I the action.

COMPLETELY INTEGRABLE SYSTEMS

The question we want to answer in this lecture is: when can we put a Hamiltonian system into such a simple form? First, we need to define exactly what the simplest form we could hope for in a Hamiltonian system:

Example 30. Let X be the manifold $T^n \times \mathbb{R}^n$ with coordinates $\theta_1, \dots, \theta_n$ (**angle coordinates**) on T^n and coordinates $I_1, \dots, I_n$ on $\mathbb{R}^n$ (**action coordinates**). This comes with a natural symplectic form given by $\omega = \sum_{i=1}^n \mathrm{d}I_i \wedge \mathrm{d}\theta_i$, which makes the action-angle coordinates into Darboux coordinates. Moreover, let the Hamiltonian H be a function only of the $I_1, \dots, I_n$. Then Hamilton's equations for H preserve $I_1, \dots, I_n$ and the angular coordinates θ_i rotate around with the constant speed $\theta_i = \frac{\partial H}{\partial I_i}$. This is the local model for a *Lagrangian torus fibration*.

DEFINITION 32. *We say that a symplectic manifold X with a Hamiltonian $H : X \to \mathbb{R}$ has* ***action-angle coordinates*** *if it is locally symplectomorphic to the previous example.*

Now, not every Hamiltonian H on a symplectic manifold X can be locally symplectomorphic to the above example, but there are some obvious *necessary conditions*. First off, there must be n functions $F_1, \dots, F_n : X \to$

$\mathbb{R}$ that are all conserved under Hamiltonian flow of H which can correspond to the I_is: moreover, they must be **functionally independent**, meaning informally that one can't be expressed in terms of the others, and formally that $d_pF_i \in T_p^*X$ are linearly independent at each point $p \in X$. Also, we should be able to express H in terms of the $F_1, \ldots, F_n$, and so since H itself is a conserved quantity, then without loss of generality say that $H = F_1$. In short, the Hamiltonian system must have a lot of conserved quantities, but still fewer than the $2n$ conserved quantities that you might expect would be necessary to completely determine the motion! There's another condition that we'll need as well, coming from the symplectic structure.

Let's recall the *Poisson brackets* on a symplectic manifold: recall that if $f, g : X \to \mathbb{R}$ were two smooth functions, we defined the Poisson bracket $\{f, g\}$ to be the smooth function given at a point $x \in X$ by the pairing:

$$\{f, g\}(x) = \omega_x(X_f, X_g)$$

where X_f, X_g are the Hamiltonian vector fields of f, g with respect to the symplectic form ω. Recall that we also showed that under the Hamiltonian flow of X_H:

$$\dot{f} = \{f, H\}$$

and therefore f is a conserved quantity of the Hamiltonian flow of H exactly when $\{f, H\} = 0$. Moreover, if p_i, q_i are coordinates on $\mathbb{R}^{2n}$ with the standard symplectic form $\omega = \sum_{i=1}^n \mathrm{d}p_i \wedge \mathrm{d}q_i$, then in fact we have the commutation relations given by:

$$\{p_i, p_j\} = 0$$

for every i, j: that is, all of the momentum coordinates commute with each other.

Now, in our example above, the action coordinates I_i played the role of momentum coordinates in a Darboux chart on X and hence we must have that $\{I_i, I_j\} = 0$ for all i, j: we saw that the I_is are **mutually commuting** conserved quantities. If we assume our symplectic manifold X with Hamiltonian H can be expressed locally in this form, this means that we must have the conserved quantities F_i mutually commuting also, since symplectomorphisms preserve the Poisson brackets. In other words, we should have:

DEFINITION 33. *A* ***completely integrable system*** *on a $2n$-dimensional symplectic manifold X is a collection of n functionally independent smooth functions $H = F_1, F_2, \ldots, F_n : X \to \mathbb{R}$ such that $\{F_i, F_j\} = 0$ for all i, j.*

The intuition is that an integrable system has many 'compatible' conserved quantities: every such quantity F_i is conserved under the flow of any other, so we could symmetrically consider any of the F_is as being the Hamiltonian. This was our *necessary* condition for having action-angle coordinates. And in fact, the Arnol'd-Liouville theorem will tell us that having enough conserved quantities is actually **sufficient**, and even more:

THEOREM 27. *(**Arnol'd-Liouville Theorem**) Suppose $F : X \to \mathbb{R}^n$ (with components $F_1, \ldots, F_n$) forms a completely integrable system on a $2n$-dimensional compact symplectic manifold X. Then for any value $c \in \mathbb{R}^n$, $F^{-1}(c)$ is a smooth manifold and if it is connected then there is an open neighbourhood U of $c \in \mathbb{R}^n$, a diffeomorphism $\psi : U \to \mathbb{R}^n$, and a symplectomorphism $\Psi : F^{-1}(U) \to T^n \times \mathbb{R}^n$ such that $\psi \circ F = I \circ \Psi$: that is, the following diagram commutes:*

$$\begin{tikzcd} F^{-1}(U) \arrow[r, "\Psi"] \arrow[d, "F"] & T^n \times \mathbb{R}^n \arrow[d, "I"] \\ U \arrow[r, "\psi"] & \mathbb{R}^n \end{tikzcd}$$

We say that in this case, X carries a **Lagrangian torus fibration**.

Slogan for the Arnol'd-Liouville Theorem:

completely integrable systems $\leftrightarrow$ Lagrangian torus fibrations

We imagine the tori T^n living over $\mathbb{R}^n$ as the *fibers* of the map F, as in the image below. Note that for any of the tori T^n given by $I = c$, the restriction of the symplectic form $\omega = \sum_{i=1}^n \mathrm{d}I_i \wedge \mathrm{d}\theta_i$ is zero: this is why we call these tori *Lagrangian*. This (rather confusing) usage has nothing to do with the Lagrangian L, but rather with the fact that ω was originally called the *Lagrange bracket*, and so Maslov decided to call a submanifold on which it vanished *Lagrangian*. The existence of a Lagrangian torus fibration imposes some very strong conditions on the topology and geometry of the manifold X! It is far from true that every symplectic manifold possesses a Lagrangian torus fibration. The search for so-called *special* Lagrangian torus fibrations is an active area of research with applications to string theory. The points where $F^{-1}(c)$ fails to be a smooth manifold will be where the condition that F_i are functionally independent fails to hold, and are called the **separatrices**: we will give an example of these at the end of this lecture.

Historically, the procession of ideas was the opposite to that described above: completely integrable systems appear everywhere in physics, and it was discovered that they could all be described in terms of their *Liouville tori* or *invariant tori* (the fibers $F^{-1}(c)$).

The proof of this theorem is again not so difficult: I'll break it into a series of lemmas and leave some of them to you:

LEMMA 9. *If $F_1, F_2, \ldots, F_n : X \to \mathbb{R}$ is a completely integrable system on X, then the Hamiltonian vector fields X_{F_i} all commute under the Lie bracket and the symplectic form, and their flows preserve the fibers $F^{-1}(c)$.*

LEMMA 10. *Let X be any n-dimensional compact manifold with $X_1, \ldots, X_n$ mutally commuting vector fields, that is, $[X_i, X_j] = 0$ for all i, j, that are linearly independent at every point, then X is diffeomorphic to a torus T^n.*

The explicit map is given as follows: fix a point $x \in X$ and define a map $\Phi : \mathbb{R}^n \to X$ given on $(\theta_1, \ldots, \theta_n)$ by $\Phi_{X_1}^{\theta_1} \circ \cdots \circ \Phi_{X_n}^{\theta_n}(x)$: that is, flow by X_n by time θ_n, then flow by X_{n-1} by time θ_{n-1} etc.; this will in fact be independent of the order in which we take the flows of these vector fields, since they all commute. One can show that this map must be surjective, and when X is compact, each of the flows X_1 must actually be periodic, so that we get a diffeomorphism $\mathbb{R}^n/\mathbb{Z}^n \cong T^n \to X$: these coordinates θ_i will exactly be our angle coordinates.

From these two lemmas, we see that the fibers of our integrable system $F : X \to \mathbb{R}^n$ are automatically Lagrangian tori! We're halfway to proving the theorem! What we're missing now is just the *action* coordinates, since the F_i won't in general give a Darboux chart when used with the θ_is. We need to define some coordinate transformation ψ on $\mathbb{R}^n$ that will give us the canonical partners of the θ_i. We can do this as follows: write γ_i for the loop in $F^{-1}(c)$ traced out by the flow of the vector field X_{F_i}; and write $\gamma_i(b)$ for the corresponding loop in the fiber $F^{-1}(b)$ for $b \in U$ a neighbourhood of c. Then define the **action** $I_i : U \to \mathbb{R}$ by the integral:

$$I_i(b) = \int_{\Gamma_b} \omega$$

where Γ_b is any smooth cylinder in X with boundary given by γ_i and $\gamma_i(b)$. If $\omega = \mathrm{d}\lambda$ (which we can always assume to be the case locally), then we will have

$$I_i(b) = \int_{\gamma_i(b)} \lambda - \int_{\gamma_i} \lambda$$

by Stokes' theorem, and so $I_i(b)$ is independent of our choice of cylinder between γ_i and $\gamma_i(b)$. Lastly, we just need to show that this does indeed give the canonical coordinates conjugate to θ_i:

LEMMA 11. *The map defined by* $b \mapsto I_i(b)$ *defines a diffeomorphism* $U \to \mathbb{R}^n$ *and in these new coordinates, the symplectic form on* X *may be expressed as*

$$\omega = \sum_{i=1}^{n} \mathrm{d}I_i \wedge \mathrm{d}\theta_i$$

We'll see later where the term *action* comes from in the name of these coordinates.

Example 31. Actually every Hamiltonian function H on a 2-dimensional symplectic manifold defines a completely integrable system, since H is always conserved under the flow of X_H and $\{H, H\} = 0$ by antisymmetry: the only problems will occur at the separatrices where $\mathrm{d}H = 0$. For instance, we can look at the simple pendulum we saw earlier, where $X = S^1 \times \mathbb{R}$ and $H = p^2 + (1 - \cos\theta)$. Then the Lagrangian torus fibration on X is given exactly by the picture we have seen of circles on the cylinder.

18 THE HAMILTON-JACOBI EQUATION

19 PROBLEMS

LIE GROUP ACTIONS

Exercise: Rotations of $\mathbb{R}^3$:

1. Let $G = \mathrm{SO}(3)$ act on $\mathbb{R}^3$ by rotations. Show that the Lie algebra of G is given by the 3×3 real skew-symmetric matrices A. Show that, after identifying this space with $\mathbb{R}^3$, the adjoint action of $\mathrm{SO}(3)$ on $\mathfrak{g}$ becomes the usual action of $\mathrm{SO}(3)$ on $\mathbb{R}^3$ by rotation, and that under this identification, the commutator bracket of matrices $[A, B] = AB - BA$ is sent to the cross product of vectors in $\mathbb{R}^3$;
2. Show that the infinitesimal action on $\mathbb{R}^3$ by $v \in \mathbb{R}^3$ is given by the vector field $\phi_v : \mathbb{R}^3 \to \mathbb{R}^3$ equal to $\phi_v(q) = v \times q$;
3. Show that if we lift the action of G to $T^*\mathbb{R}^3 \cong \mathbb{R}^6$, then the moment map $\mu : \mathbb{R}^6 \to \mathbb{R}^3 \cong (\mathbb{R}^3)^*$ coming from Noether's theorem is given by

$$\mu(q, p) = q \times p$$

and show that this makes rotation into a Hamiltonian action.

Exercise: More generally, check that the lift of an action of G on Q to an action of G on T^*Q constructed in Noether's theorem satisfies the equivariance condition for a Hamiltonian action.

Exercise: Suppose $\mu : X \to \mathfrak{g}^*$ is a moment map for a Lie group action; show that the map $\mu^* : \mathfrak{g} \to C^\infty(X)$ defined by $v \mapsto \mu(v)$ satisfies $\mu^*([v, w]) = \{\mu^*(v), \mu^*(w)\}$ if and only if μ satisfies the equivariance condition $\mu(\Phi_g(x)) = \mathrm{Ad}_g^*\mu(x)$.

Exercise: Suppose that G is a Lie group acting on a smooth manifold X with Lie bracket given by the commutator of matrices; show that the infinitesimal action $\phi : \mathfrak{g} \to \mathrm{Vect}(X)$ is a homomorphism of Lie algebras.

The following series of exercises asks you to prove the **quotient theorem** for actions of Lie groups, which is a great exercise if you're learning differential geometry!

THEOREM 28. *(**Quotient Theorem**) Suppose G is a compact Lie group acting freely (no elements $g \in G$ have any fixed points) on a manifold X: then the set of equivalence classes X/G has a natural structure of a smooth manifold such that the map $\pi : X \to X/G$ sending a point to its equivalence class is smooth.*

We can prove this by showing that the slices defined above give local coordinates on the quotient:

Exercise: (**Slice Lemma**) Suppose G acts on X as above: then for all $\varepsilon > 0$ sufficiently small and $p \in X$, the map $G \times S_\varepsilon(p) \to X$ given by the group multiplication is a diffeomorphism onto a G-invariant open neighbourhood U of the orbit Gp.

If U is such a neighbourhood, then we see that U/G is smooth and equal to $S_\varepsilon(p)$: the coordinates near $[p]$ we want to use on the quotient space X/G are precisely given by the slices $S_\varepsilon(p)$, so that the map $X \to X/G$ will necessarily be smooth.

Exercise: Finally, check that the transitions between these coordinate patches are smooth. Moreover, check that X/G has the Hausdorff property.

MARSDEN-WEINSTEIN THEOREM

Suppose G is a compact Lie group with a Hamiltonian action on a symplectic manifold (X, ω), with moment map $\mu : X \to \mathfrak{g}^*$. Suppose that G acts freely on $\mu^{-1}(0)$; here's the proof of the Marsden-Weinstein theorem broken down into steps:

Exercise: Show that G actually does act on $\mu^{-1}(0)$!

Exercise: (**Orbit Lemma**) If G is a compact Lie group acting freely on any smooth manifold X, then for every point $x \in X$ the orbit $Gx = \{gx : g \in G\}$ is a closed submanifold of X: namely the map $G \to X$ given by $g \mapsto g \cdot x$ is injective and has injective differential at every point.

Being a **submanifold** means that around every point $p \in Gx$ there exists some smooth chart (U, ϕ) in which $Gx \cap U$ is given by the hyperplane $x_1 = x_2 = \cdots = x_k = 0$ inside $\mathbb{R}^n$. We say that a **slice** $S_\varepsilon(p)$ of the action at p is the subset of X given by $\{x_{k+1} = \cdots = x_n = 0\} \cap B_\varepsilon(0)$. The **quotient theorem** says that these give local coordinates on the quotient space X/G.

Exercise: To apply the quotient theorem, we need to show that $\mu^{-1}(0)$ is actually a smooth manifold. Show that show that $\ker D_x\mu$ is the symplectic orthogonal of $T_x(Gx)$ and thus show that $D_x\mu$ is surjective for every $x \in \mu^{-1}(0)$.

The surjectivity of $D_x\mu$ means that $\mu^{-1}(0)$ is a smooth manifold, by a version of the implicit function theorem. The next thing to do is consider the tangent space to $\mu^{-1}(0)/G$:

Exercise: Show that:

1. the tangent space at $x \in X$ to a G-orbit Gx is given by the image of the infinitesimal action ϕ at x;
2. the tangent space to $\mu^{-1}(0)$ at a point $x \in \mu^{-1}(0)$ is given by the kernel of $D_x\mu$;
3. the tangent space to $[x] \in \mu^{-1}(0)/G$ is naturally identified with the quotient $\ker D_x\mu / T_x(Gx)$.

Exercise: Show that if V is a symplectic vector space, and W is a subspace on which the symplectic form vanishes, then W^ω/W has a natural symplectic structure, where W^ω is the symplectic orthogonal.

Exercise: Use the previous two exercises to conclude that $\mu^{-1}(0)/G$ is a symplectic manifold!

ARNOL'D-LIOUVILLE THEOREM

The Arnol'd-Liouville theorem takes a particularly simple form in dimension 2:

Exercise: Show that any Hamiltonian H on a 2-dimensional symplectic manifold X defines a completely integrable system away from the points where $\mathrm{d}H = 0$.

Exercise: Draw a picture of the Lagrangian torus fibration in the case of the simple pendulum, when $X = S^1 \times \mathbb{R}$ with coordinates (θ, p) and $H = p^2 + (1 - \cos\theta)$.

Exercise: Suppose now that H is a Hamiltonian on $\mathbb{R}^2$ such that all of the level sets $H^{-1}(c)$ are smooth and bounded. What is the associated action coordinate on $\mathbb{R}^2$?

Exercise: Recall that in problem sheet 1 you proved that if $A(s)$ denotes the area enclosed by the curve $H = s$, then $\frac{\mathrm{d}A}{\mathrm{d}s}$ is the period of motion along the curve. Now deduce this from the Arnol'd-Liouville theorem.

Here are some of the lemmas we used in proving the Arnol'd-Liouville theorem:

Exercise: If $F_1, F_2, \ldots, F_n : X \to \mathbb{R}$ is a completely integrable system on X, show that:

1. $\omega(X_{F_i}, X_{F_j}) = 0$ for all i, j;
2. $[X_{F_i}, X_{F_j}] = 0$ for all i, j;
3. the flow of X_{F_i} applied to a point $x \in F^{-1}(c)$ remains in $F^{-1}(c)$ for all times;
4. X_{F_i} are a basis for the tangent space to $F^{-1}(c)$ at every point.

Exercise: Let X be any n-dimensional compact manifold with $X_1, \ldots, X_n$ mutally commuting vector fields, that is, $[X_i, X_j] = 0$ for all i, j, that are linearly independent at every point, show that X is diffeomorphic to a torus T^n using the hint given in the lecture.

20 PSEUDOHOLOMORPHIC CURVES

Everything is impossible, until you understand it. Then it becomes trivial.

Ernest Rutherford

One of the key tools in modern symplectic topology is the study of certain surfaces inside symplectic manifolds (called *pseudoholomorphic curves* because they have one complex dimension): their study, initiated by Gromov, allows us to generalize many powerful results from complex analysis and geometry into the setting of symplectic geometry. In the next lecture, we will use pseudoholomorphic curves to prove Gromov's celebrated *non-squeezing theorem*, which says, roughly, that one cannot symplectically embed a ball into a cylinder of a smaller radius.

ALMOST-COMPLEX STRUCTURES

Firstly, let's ask a simple question: given a linear map $A : \mathbb{R}^2 \to \mathbb{R}^2$, we can also consider it as a map $\tilde{A} : \mathbb{C} \to \mathbb{C}$ via $\tilde{A}(x+iy) = A_1(x,y) + iA_2(x,y)$: when is this map a linear map over the complex numbers? Well, A is already linear with respect to the real numbers ($\tilde{A}(\lambda v) = \lambda \tilde{A}(v)$ for all $\lambda \in \mathbb{R}$), so in order for it to be complex linear, $\tilde{A}$ simply needs to commute with multiplication by i, in the sense that $\tilde{A}(iv) = i\tilde{A}(v)$ for all $v \in \mathbb{C}$. More generally, an $\mathbb{R}$-linear map $A : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ gives a $\mathbb{C}$-linear map $\tilde{A} : \mathbb{C}^n \to \mathbb{C}^n$ exactly when $\tilde{A}$ commutes with multiplication by i. How do we express this condition in terms of the original real matrix A? Well, if we identify $\mathbb{C}^n$ with $\mathbb{R}^{2n}$ with via the map $(z_1, \ldots, z_n) \mapsto (\mathrm{Re}(z_1), \ldots, \mathrm{Re}(z_n), \mathrm{Im}(z_1), \ldots, \mathrm{Im}(z_n))$, then multiplication by i becomes matrix multiplication by the block matrix $J : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ given by

$$J = \begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix}$$

and $\tilde{A}$ commutes with i if and only if $AJ = JA$. We summarize this as follows:

PROPOSITION 19. *If we identify $\mathbb{C}^n$ with $\mathbb{R}^{2n}$ via the map $z \mapsto (\mathrm{Re}(z), \mathrm{Im}(z))$, then a matrix $A : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ gives a complex linear map $\tilde{A} : \mathbb{C}^n \to \mathbb{C}^n$ if and only if it commutes with J, in the sense that $AJ = JA$.*

This J is exactly our friendly matrix from before – this relationship hints at the beginning of a close relationship between the symplectic and complex words. We call J the **standard complex structure** on $\mathbb{R}^{2n}$. Note that it captures all of the information of the complex structure on $\mathbb{R}^{2n}$: given a matrix J such that $J^2 = -I$, we can use it to tell us how i acts, and give an identification $\mathbb{R}^{2n} \to \mathbb{C}^n$:

PROPOSITION 20. *Given any linear map $J : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ such that $J^2 = -1$, there is an $\mathbb{R}$-linear isomorphism $\phi : \mathbb{R}^{2n} \to \mathbb{C}^n$ such that $\phi(Jv) = i\phi(v)$.*

Now, suppose $f : \mathbb{R}^2 \to \mathbb{R}^2$ is a smooth map, not necessarily linear; but we would still like to impose some kind of compatibility between f and the complex structure on $\mathbb{R}^2 \cong \mathbb{C}$. We will call f **holomorphic** if the differential $D_x f$ is complex-linear at every point. By the above criterion, this means that $D_x f$ must commute with the matrix J. If we write the components of f as $f = (u, v)$, then this condition is given explicitly by:

$$\begin{pmatrix} -\frac{\partial v}{\partial x} & -\frac{\partial v}{\partial y} \\ \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{pmatrix} = \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} \frac{\partial u}{\partial y} & -\frac{\partial u}{\partial x} \\ \frac{\partial v}{\partial y} & -\frac{\partial v}{\partial x} \end{pmatrix}$$

hence we derive the **Cauchy-Riemann equations**:

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}$$
$$\frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y}$$

which may or may not be familiar from complex analysis. This is a system of two linear partial differential equations for the components of f, whose solutions have very remarkable properties. More generally, we define:

DEFINITION 34. *A smooth map $f : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ is **holomorphic** if f satisfies the **Cauchy-Riemann equation**: $D_x f \circ J - J \circ D_x f = 0$ for all points $x \in \mathbb{R}^{2n}$.*

Now, we want to do geometry, so the natural question is, how do we generalize this to manifolds? It turns out there are (at least two) ways to do this. The first one is to try and generalize the notion of the operator J:

DEFINITION 35. *An **almost-complex structure** J on a smooth manifold X is a smooth map $J : TX \to TX$ such that $J_x : T_xX \to T_xX$ is linear and satisfies $J_x^2 = -1$.*

In other words, J_x will formally give the tangent space T_xX at every point $x \in X$ the structure of a complex vector space, and J_x does this smoothly. Note that this is a *choice* of extra data on the manifold. Then we can define as before:

DEFINITION 36. *A smooth map $f : X_1 \to X_2$ between smooth manifolds X_1, X_2 with almost-complex structures J_1, J_2 is said to be **pseudoholomorphic** (or $J_1 - J_2$ **holomorphic**) if the differential $D_x f : T_xX_1 \to T_{f(x)}X_2$ satisfies the Cauchy-Riemann equation: $D_x f \circ J_1 - J_2 \circ D_x f = 0$.*

It is important to note that whether a map is pseudoholomorphic *depends* on which almost-complex structures we are talking about! As we will see shortly, there will often be *many* almost-complex structures and we will be required to chose from among them.

On the other hand, there is a stricter notion of a **complex manifold**: we go through the definition of a manifold, but instead require all of the transition functions to be not just smooth but *holomorphic* (as maps $\mathbb{R}^{2n} \to \mathbb{R}^{2n}$ in the previous sense). We say that these manifolds have **holomorphic charts**: local coordinates $z_1, \ldots, z_n$ that give an identification with $\mathbb{C}^n$. For instance, all open subsets of $\mathbb{C}^n$ are complex manifolds; so are the complex projective spaces $\mathbb{P}^n$ we saw before: you can check that their transition functions we defined are in fact holomorphic maps. Complex manifolds always have almost-complex structures, arising from multiplication by i: we call almost-complex structures arising in this way **integrable**. However, this notion is much stronger: complex manifolds are rare and have very special properties, some of which we will hear about in the presentations.

SYMPLECTIC MANIFOLDS

Let's now return to the setting of symplectic geometry, and see how we can import some of these complex techniques. Firstly, we should require some compatibility between the symplectic structure and the almost-complex structure:

DEFINITION 37. *Suppose X is a symplectic manifold with symplectic form ω, and almost-complex structure J: we say J is **compatible** with ω if $\omega(Jv, Jw) = \omega(v, w)$ for all $v, w \in T_xX$ and $\omega(v, Jv) > 0$ for all non-zero vectors $v \in T_xX$.*

For example, the standard complex structure J on $\mathbb{R}^{2n}$ is compatible with the standard symplectic form ω, since $\omega(v, w) = -v^T J w$. Since $J^2 = -I$, thus $\omega(Jv, v) = v^T v$ which is exactly the Euclidean length squared $|v|^2$ of the vector. We say that $\omega(\cdot, J\cdot)$ is a **compatible metric** on the symplectic manifold X, since we can use it to measure the lengths of tangent vectors.

A natural question to ask at this point would be: do these actually exist? And if so, are they unique? These are answered by the following theorem of Gromov.

THEOREM 29. *(**Gromov**) Suppose X is a symplectic manifold: then the space of compatible almost-complex structures is a non-empty infinite dimensional contractible space.*

You should interpret the three points of this theorem as saying:

1. compatible almost-complex structures exist;
2. there are *lots of them*;
3. they aren't unique, but any two choices can be connected by a smooth path.

Hence we are guaranteed that we can always add the extra structure of an almost-complex structure to our symplectic manifold, and our resulting constructions will be unique (in a suitably homotopical sense).

By Darboux's theorem this result more or less follows from the corresponding statement for symplectic vector spaces, which you might enjoy trying to prove.

PSEUDOHOLOMORPHIC CURVES

There's one particular type of pseudoholomorphic map we'll be interested in the most:

DEFINITION 38. *Suppose X is a symplectic manifold and J is a compatible almost-complex structure: we say a smooth map $u : \mathbb{C} \to X$ is a **pseudoholomorphic curve** (or a J-**holomorphic curve**) if u is pseudoholomorphic with respect to the complex structure i on $\mathbb{C}$ and the almost-complex structure J on X: in other words:*

$$Du \circ i - J \circ Du = 0$$

We can make the definition more generally when $\mathbb{C}$ is replaced by any almost-complex manifold Σ of dimension 2. Such maps have very similar properties to the holomorphic functions you might have studied in complex analysis, as well as close connections with the theory of *minimal surfaces*.

DEFINITION 39. *Let X be a symplectic manifold with symplectic form ω and almost-complex structure J, and suppose $u : \Sigma \to X$ is a smooth map. We define the **energy** of u to be:*

$$E(u) = \int_\Sigma |Du|^2 \mathrm{d}x \wedge \mathrm{d}y$$

where $z = x + iy$ are complex coordinates on Σ and $|v|^2 = \omega(v, Jv)$ is the length.

Note that this really depends on our choice of almost-complex structure J, so I'll sometimes write this as $E_J(u)$ when I want to make this apparent. The terminology comes from physics, where this is considered the 'action' or 'energy' of the worldsheet of a string. While the functional E isn't quite the surface area of u (it's missing a square root), it is closely related to it:

DEFINITION 40. *A smooth map $u : \Sigma \to X$ is called a **minimal surface** if it has the least value of $E(u)$ among smooth maps $\Sigma \to X$.*

By formulating this as a calculus of variations problem, one can in fact show that minimal surfaces must have (at least locally) the least surface area among nearby maps.

Given such a curve $u : \Sigma \to X$ define the **complex anti-linear part** of the derivative Du to be given by

$$\bar{\partial}_J u = \frac{1}{2}(Du \circ i - J \circ Du)$$

and the **complex linear part** to be:

$$\partial_J u = \frac{1}{2}(Du \circ i + J \circ Du)$$

so that $Du = \bar{\partial}_J u + \partial_J u$, and $\bar{\partial}_J u = 0$ if and only if u is J-holomorphic. Moreover, $\bar{\partial}_J u \circ i = -J \circ \bar{\partial}_J u$, that is, it is *antilinear* with respect to the complex structures. Then we have:

PROPOSITION 21. ***(Energy Identity)*** *Let X be a symplectic manifold with symplectic form ω and almost-complex structure J, and suppose $u : \Sigma \to X$ is a smooth map. Then we can write:*

$$\int_\Sigma u^*\omega = \int_\Sigma |\partial_J u|^2 \mathrm{d}x \wedge \mathrm{d}y - \int_\Sigma |\bar{\partial}_J u|^2 \mathrm{d}x \wedge \mathrm{d}y$$

and

$$E(u) = \int_\Sigma |\partial_J u|^2 \mathrm{d}x \wedge \mathrm{d}y + \int_\Sigma |\bar{\partial}_J u|^2 \mathrm{d}x \wedge \mathrm{d}y$$

From this it follows that:

COROLLARY 3. *If $u : \Sigma \to X$ is a J-holomorphic map, and $v : \Sigma \to X$ is a homotopic smooth map, that is, there is a smooth map $w : \Sigma \times [0,1] \to X$ such that $w|_0 = v$ and $w|_1 = u$, then $E(u) \leq E(v)$.*

Proof. Since $\mathrm{d}\omega = 0$, by Stokes' theorem we have that:

$$0 = \int_{\Sigma \times [0,1]} \mathrm{d}w^*(\omega) = \int_\Sigma u^*\omega - \int_\Sigma v^*\omega$$

On the other hand, since u is J-holomorphic, $\bar{\partial}_J u = 0$ and so

$$E(u) = \int_\Sigma |\partial_J u|^2 \mathrm{d}x \wedge \mathrm{d}y = \int_\Sigma u^*\omega$$

while

$$E(v) = \int_\Sigma |\partial_J v|^2 \mathrm{d}x \wedge \mathrm{d}y + \int_\Sigma |\bar{\partial}_J v|^2 \mathrm{d}x \wedge \mathrm{d}y \geq \int_\Sigma v^*\omega = E(u)$$

which completes the proof. ■

This statement means exactly that u is a kind of minimal surface: this gives us one reason to care about holomorphic curves: they provide a way of finding minimal surfaces.

So why holomorphic curves? Unlike minimal surfaces, pseudoholomorphic curves have good theories of *compactness* and *deformations*: typically, the set of all pseudoholomorphic curves satisfying a given condition will not only be a set, but a compact finite-dimensional smooth manifold, called the **moduli space**. Moreover, these moduli spaces will be, in a suitable sense, independent of the choice of almost-complex structure J and so give us invariants of symplectic manifolds. Next time, we shall see an example of how this can be applied to prove Gromov's non-squeezing theorem.

21 SYMPLECTIC TOPOLOGY

Since the first edition of this book had appeared in 1974, the content of this Appendix has grown into a new branch of mathematics: symplectic topology. To describe this development (triggered by the conjectures in this Appendix, which still remain, for general manifolds, neither proved, nor disproved) one would need a book longer than the present one.

V. Arnol'd, *Mathematical Methods of Classical Mechanics*

GROMOV NON-SQUEEZING

Gromov's non-squeezing theorem is often regarded as the first theorem in a new field: *symplectic topology*, which uses the theory of pseudoholomorphic curves to study the surprising topological 'rigidity' of symplectic geometry. Gromov's non-squeezing theorem says heuristically that balls in phase space can't be 'squeezed' by symplectomorphisms; as such it can be regarded as a classical analog of Heisenberg's uncertainty principle, purely within classical mechanics (there is actually a precise sense in which this is true). More precisely, if we let $B^{2n}(r)$ denote the closed ball of radius r inside $\mathbb{R}^{2n}$: the statement of Gromov's non-squeezing theorem is:

THEOREM 30. *(**Gromov Non-Squeezing**) There exists a symplectic embedding $B^{2n}(r) \to B^2(R) \times \mathbb{R}^{2n-2}$ if and only if $r \leq R$.*

Here by symplectic embedding, we mean a map that is a symplectomorphism onto its image.

Why exactly is this result so surprising? Recall that by Liouville's theorem, every symplectomorphism must preserve areas, and it is clear that we can always find an *area-preserving* map from $B^{2n}(r) \to B^2(R) \times \mathbb{R}^{2n-2}$ for any $r, R > 0$ by suitably squeezing the ball. Even more, it is known that a volume-preserving map can be approximated arbitrarily well by symplectomorphisms. But Gromov's theorem shows the existence of a purely symplectic phenomenon: an unexpected obstruction, called a *symplectic capacity*, to the existence of symplectic maps.

There now exist many proofs of this fundamental result, none of them elementary; the proof we'll outline here is closest in spirit to Gromov's original proof using pseudoholomorphic curves and their relation to minimal surfaces. Of course, we won't be able to do much more than just sketch the main ideas here.

The key we will want to use is the following classical result:

THEOREM 31. *(**Isoperimetric Inequality**) Suppose $u : D^2 \to B^{2n}(r)$ is a minimal surface with respect to the standard symplectic form and almost-complex structure ω and J on $\mathbb{R}^{2n}$ such that u passes through 0 and $u(\partial D^2) \subseteq \partial B^{2n}(r)$. Then*

$$E_J(u) \geq \pi r^2$$

In other words, we can't have any less surface area than a circle, no matter how hard we try.

Now, here's the plan of the argument: suppose we have a symplectic embedding $\phi : B^{2n}(r) \to B^2(R) \times \mathbb{R}^{2n-2}$: then in $B^2(R) \times \mathbb{R}^{2n-2}$ we have a disk u through $\phi(0)$ given by $\{\phi(0) + (x, 0) : x \in B^2(R)\}$, and this will be holomorphic with respect to the standard complex structure J on $\mathbb{R}^{2n}$; moreover we have $E(u \cap$

$\phi(B^{2n}(R)) \leq \pi R^2$. Then the inverse image $\phi^{-1}(u)$ is a disk in $B^{2n}(r)$ passing through zero, with boundary on $\partial B^{2n}(r)$. Then if we could show that $\phi^{-1}(u)$ is still holomorphic with respect to the standard almost-complex structure on $\mathbb{R}^{2n}$, by the above theorem we would have:

$$\pi r^2 \leq E_J(\phi^{-1}(u)) \leq E_J(u) = \pi R^2.$$

and so $r \leq R$ and we would be done.

Problem is, there's no reason to expect $\phi^{-1}(u)$ to be J-holomorphic with respect to the standard almost-complex structure J on $\mathbb{R}^{2n}$, since an arbitrary symplectomorphism ϕ has no need to preserve the almost-complex structure or the area $E_J(u)$ with respect to the metric $\omega(\cdot, J\cdot)$.

What we will do instead is *push forward* the standard almost-complex structure on $\mathbb{R}^{2n}$ via ϕ: we can always extend this smoothly to a new almost-complex structure $\tilde{J}$ on the cylinder $B^2(R) \times \mathbb{R}^{2n-2}$ such that ϕ is now $J - \tilde{J}$ holomorphic. Then we will appeal to the result:

THEOREM 32. *For any almost-complex structure $\tilde{J}$ on $B^2(R) \times \mathbb{R}^{2n-2}$ and any interior point $z \in B^2(R) \times \mathbb{R}^{2n-2}$, there exists a $\tilde{J}$-holomorphic disk passing through z with boundary in the boundary of the cylinder.*

Now, since this curve u is $\tilde{J}$-holomorphic, we have:

$$E_{\tilde{J}}(u) = \int_{D^2} u^*\omega = \pi R^2$$

since the symplectic form ω is unchanged, and the integral $\int_{D^2} u^*\omega$ is an invariant of the choice of u within the same homotopy class.

Now, the map ϕ is $J - \tilde{J}$-holomorphic by construction, so when we take $\phi^{-1}(u)$ this curve will be J-holomorphic and will satisfy $E_J(\phi^{-1}(u)) \leq E_{\tilde{J}}(u)$. Now we can apply the isoperimetric inequality to conclude that:

$$\pi r^2 \leq E_J(\phi^{-1}(u)) \leq E_{\tilde{J}}(u) = \pi R^2.$$

as desired.

This still leaves unanswered the question: where does the $\tilde{J}$-holomorphic curve in the theorem come from? The idea is to use the deformation theory of holomorphic curves!

By Gromov's theorem on almost-complex structures, we know that we can connect J and $\tilde{J}$ by a path of compatible almost-complex structures, call them J_t with $t \in [0,1]$ and $J_0 = J, J_1 = \tilde{J}$. Then we want to consider the following *moduli space*:

$$\mathcal{M}(J_t, [u]) = \{(t, u_t) : \; t \in [0,1], u_t \text{ is } J_t\text{-holomorphic, passes through } z \text{ and is homotopic to } u\}$$

Strictly speaking, we really want to look at curves u_t living inside a compactification $S^2 \times T^{2n-2}$ of $B^2(R) \times \mathbb{R}^{2n-2}$. The consequence of pseudoholomorphic curves having good compactness and deformation properties means that:

THEOREM 33. *(Gromov) For a suitably 'generic' path J_t, the set $\mathcal{M}(J_t, [u])$ has the structure of a 1-dimensional compact manifold with boundary.*

How do we use this to prove the existence theorem above? Well, we already know that there is the disk u_0 in the standard almost-complex structure that we found above; moreover, one can show that this is in fact the unique such J-holomorphic curve. Now, the boundary of $\mathcal{M}(J_t, [u])$ is given exactly by the union of

$\mathcal{M}(J,[u])$ and $\mathcal{M}(\tilde{J},[u])$, the moduli spaces of such pseudoholomorphic curves with fixed almost-complex structures. Since a 1-dimensional compact manifold with boundary must have an even number of boundary points, and we know $\mathcal{M}(J,[u]) = \{u_0\}$, hence we know that $\mathcal{M}(\tilde{J},[u])$ is non-empty!

Therefore we've now reduced the difficulty to showing that $\mathcal{M}(J_t,[u])$ is in fact a smooth compact manifold: this is a difficult statement, and I want to describe some of the complications. The idea of showing that $\mathcal{M}(J_t,[u])$ is a smooth manifold is to consider the Cauchy-Riemann operator $\bar{\partial}_J$ as a smooth function on some infinite-dimensional space of maps $D^2 \to M$, and then to try and apply a version of the implicit function theorem to show that the set $(\bar{\partial}_J)^{-1}(0)$ is a smooth manifold whenever the derivative $D(\bar{\partial}_J)$ (in an appropriate sense) is surjective. The first problem with this is the failure of the implicit function theorem in general infinite-dimensional spaces: one needs to pass to suitable *Banach completions* in order for such a statement to hold. The second problem is that, even once we have done this, the derivative operator $D(\bar{\partial}_J)$ will not always be surjective, but only surjective for a suitably 'generic' choice of the path J_t. This is called the problem of **transversality** and plagues the entire subject of symplectic topology.

The second major difficulty is compactness of this manifold: as we continue to deform J-holomorphic curves, we don't expect them to necessarily remain smooth: their energy can concentrate at single points, allowing parts of the curve to 'bubble off' to form a new singular curve, as illustrated in the figure. This process could in principle continue indefinitely! Thus we shouldn't expect our moduli spaces of pseudoholomorphic curves to be compact, unless we include configurations with 'sphere bubbles'. The key result, called **Gromov compactness**, says that when the energy of the pseudohomolorphic curves is bounded, then if we add in these limiting points to our moduli space, it will become compact.

SYMPLECTIC CAPACITIES

One corollary of Gromov's non-squeezing theorem is that it implies the existence of a *symplectic capacity* of open subsets of $\mathbb{R}^{2n}$:

DEFINITION 41. *A* ***symplectic capacity*** *Z is a map from open subsets of $\mathbb{R}^{2n}$ to $[0,\infty]$ such that:*

- *(**Monotonicity**) If we have a symplectic embedding $U_1 \to U_2$, then $Z(U_1) \leq Z(U_2)$;*
- *(**Conformality**) If $\lambda > 0$ is a constant, then $Z(\lambda U) = \lambda Z(U)$;*
- *(**Non-triviality**) We have $Z(B^{2n}(1)) > 0$ and $Z(B^2(1) \times \mathbb{R}^{2n-2}) < \infty$.*

DEFINITION 42. *The* ***Gromov width*** *Z_G of an open subset of $\mathbb{R}^{2n}$ is the radius of the largest ball $B^{2n}(r)$ that can be symplectically embedded into U:*

$$Z_G(U) = \sup_{r>0} \left\{ \pi r^2 : \ B^{2n}(r) \textit{ embeds symplectically into } U \right\}$$

The first two axioms of a symplectic capacity are obviously satisfied by the Gromov width, and the last is actually equivalent to Gromov's non-squeezing theorem. Let's see one consequence of the existence of a symplectic capacity:

THEOREM 34. *(Eliashberg) A smooth map $\phi : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ is a symplectomorphism (or an antisymplectomorphism) if and only if ϕ preserves the symplectic capacities of all ellipsoids.*

The significance of this theorem is that implies the existence of *symplectic topology*: preserving the symplectic capacities of ellipsoids is a purely *topological* property: it makes sense for any continuous map, and it is preserved under uniform limits of continuous functions! Thus it makes sense to talk about whether an arbitrary continuous map is a symplectomorphism! It also implies that symplectomorphisms are *rigid*: they are closed under uniform limits!

Proof. (Sketch) One direction is evident: to prove the converse, suppose ϕ preserves the symplectic capacities of all ellipsoids, and let $x \in \mathbb{R}^{2n}$ be any point. We may without loss of generality apply translations (which are symplectomorphisms) to $\mathbb{R}^{2n}$ so that $x = 0$ and $\phi(0) = 0$. Then, since the condition of preserving capacities of ellipsoids is preserved under uniform limits, then the function given by:

$$\Phi(x) = \lim_{t \to 0} \frac{1}{t} \phi(tx)$$

must also preserve capacities of ellipsoids. But this Φ is exactly the derivative $D_0\phi$. Now it is simply a matter of linear algebra to check that a linear map Φ is symplectic if and only if Φ preserves symplectic capacities of ellipsoids centered at 0. ■

ARNOL'D'S CONJECTURE

Finally, I want to end with another question coming from Hamiltonian dynamics, which also motivated the development of symplectic topology and which (in its original form) remains an open problem to this day. To begin, let's recall Poincaré's Last Theorem, which said that:

THEOREM 35. *(**Poincaré**) Suppose $f : A \to A$ is an area-preserving diffeomorphism of the annulus A, that twists the two boundaries in opposite directions. Then f has at least two fixed points.*

There was for a long time the question of how this result could be generalized to higher dimensions, since the proof (eventually provided by the elder Birkhoff) was very convoluted. It was observed by Arnol'd that this theorem was implied by:

THEOREM 36. *(Arnol'd) If $\phi : T^2 \to T^2$ is symplectomorphism of the 2-torus preserving the center of gravity, then ϕ has at least four fixed points (with multiplicity).*

The idea is to think of gluing two copies of the annulus together along their boundaries: the fact that f twists the boundaries in opposite directions is exactly what allows us to ensure that the center of gravity is preserved, in the sense that, if $\phi(\theta_1, \theta_2) = (\theta_1 + f(\theta_1, \theta_2), \theta_2 + g(\theta_2, \theta_2))$, then the average values of f, g are zero. Then the four fixed points yield two fixed points on each annulus.

This suggested to Arnol'd that the correct formulation of Poincaré's Last Theorem in higher dimensions would be in terms of **Hamiltonian diffeomorphisms**: those symplectomorphisms given by the flow of a time-dependent Hamiltonian function H_t. In particular, these are exactly the symplectomorphisms of T^2 preserving the center of gravity, and are also precisely the diffeomorphisms of interest in the study of dynamical systems. Arnol'd conjectured that:

Arnol'd's Conjecture: *If X is a compact symplectic manifold, then every Hamiltonian diffeomorphism $\Phi : X \to X$ has at least as many fixed points as a smooth function on X has critical points.*

The latter value, the minimum number of critical points of a smooth function, is studied in **Morse theory**: it has a lower bound coming purely from the topology of X, namely the rank of the cohomology of X. In fact, there is also a lower bound on the number of fixed points of a diffeomorphism $X \to X$, given by the **Euler characteristic** of X, another topological quantity that is typically much smaller than the rank of the cohomology. Thus Arnold's conjecture gives us a relation between topology and symplectic geometry: it states that Hamiltonian diffeomorphisms are very special or 'rigid' topologically, in the sense that they have a lot more fixed points than one would expect!

Shortly after Gromov's introduction of pseudoholomorphic curves, a version of this conjecture was proved by Floer using similar techniques to define his **Floer homology** for Hamiltonian diffeomorphisms. The topological version of the conjecture was only proved by Fukaya-Ono in 1999 and is considered one of the deepest results in symplectic topology. The original version of the conjecture remains open and continues to inspire the development of symplectic topology.

22 CELESTIAL MECHANICS

To be added.

23 RIGID BODY MECHANICS

To be added.

24 GEOMETRIC QUANTIZATION

To be added.

25 KOLMOGOROV-ARNOL'D-MOSER THEORY

To be added.

26 SYMPLECTIC TORIC MANIFOLDS

To be added.

27 CONTACT MANIFOLDS

To be added.

28 KÄHLER MANIFOLDS

To be added.

INDEX

REFERENCES