Symplectic Geometry and Integrable Systems (MATH 538-003)
Lecture notes
Anton Izosimov
Last updated: February 26, 2020

Contents

1 Introduction to planar billiards 2
2 Invariant area form of the billiard map and Poincaré’s recurrence theorem 4
3 Billiards in disks 8
4 Billiards in ellipses 9
5 Geometric meaning of the Joachimsthal integral 12
6 The phase portrait of the billiard in an ellipse 15
7 Liouville integrable maps in 2D 17
8 Proof of the Arnold-Liouville theorem in 2D and Poncelet’s closure theorem 19
9 Algebraic integrability of elliptic billiards I 21
10 Algebraic integrability of elliptic billiards II 22
11 Canonical Hamilton’s equations 26
12 Symplectic vector spaces and symplectic manifolds 28
13 Vector fields and flows I 31
14 Vector fields and flows II 33
15 Hamiltonian vector fields 36
16 Poisson brackets 41
17 Liouville integrable systems on symplectic manifolds 43
Lecture 1: Introduction to planar billiards

Let \( D \subset \mathbb{R}^2 \) be a compact planar domain bounded by a smooth curve \( \partial D \). The dynamics of a billiard ball (which we assume to be a point mass) in \( D \) is defined as follows: the ball is moving inside \( D \) along a straight line with constant velocity, until it hits the boundary. Upon hitting the boundary, the ball is reflected from it according to the law *the angle of reflection is equal to the angle of incidence*, after which it continues moving along a straight line with constant velocity until it hits the boundary again. Then it is again reflected according to the same law, and so on. See Figure 1.

We would like to understand the dynamics of the ball, i.e. find, if possible, its position and velocity as a function of time, as well of the initial position and velocity. To begin with, we need to describe the phase space of the ball, that is the space of all its possible positions and velocities. There are two possible approaches to defining this space for the billiard system. The first one is to consider all possible positions, including those strictly inside the domain \( D \). However, in this approach, the dynamics is trivial on most of the phase space, because when the ball is strictly inside \( D \), it is just moving along a straight line. The second approach is to only look at the positions of the ball at the boundary \( \partial D \). Knowing the current position of the ball at the boundary, and also its current velocity, we can easily recover the subsequent dynamics of the ball, at least until it hits the boundary next time. So, we can only trace subsequent positions of the ball at the boundary, while all other positions are easily recovered. Also notice that since the magnitude of the velocity vector is preserved and does not affect the dynamics at all, we can assume that the velocity is given by a unit vector. Finally, notice that there is two kind of possible velocity vectors of the ball at the boundary: outward velocities, and inward velocities. Outward velocity means that the ball has just hit the boundary, and has not yet been reflected, while inward velocities are velocities after reflection.

**Remark 1.1.** Let \( \gamma(t) \) be a parametrization of \( \partial D \), corresponding to its counter-clockwise orientation. We say that a vector \( v \) at \( x = \gamma(t_0) \) is inward if the orientation of the frame \( (\gamma'(t_0), v) \) is positive, and outward if the orientation of the frame \( (\gamma'(t_0), v) \) is negative, see Figure 2. Also note that a velocity vector which is neither inward, nor outward, must be tangent to the boundary. Such velocities are only possible for billiards in non-convex domains, see Figure 3.

Since inward velocities and outward velocities are connected by the reflection law, it suffices to consider just one type of velocity vectors, for example the inward ones. So, we define the phase space of a billiard ball in \( D \) as the space of inward unit vectors attached at points of \( \partial D \). In other words, an element of the phase space is a pair \( (x, v) \), where \( x \in \partial D \), and \( v \in T_x D \) is an inward tangent vector. Note that since \( \partial D \) is homeomorphic to a circle, while an inward tangent vector \( v \) is characterized by the angle \( \alpha \in (0, \pi) \) it makes with the positive direction of \( \partial D \) (see Figure 4), our phase space can be viewed as the cylinder \( S^1 \times (0, \pi) \). As coordinates on this cylinder, we can take the angle \( \alpha \in (0, \pi) \), as well as any coordinate on \( S^1 = \partial D \). In what follows, as coordinate on \( \partial D \) we will use the arc-length parameter \( t \). So, the phase

![Figure 1: Trajectory of a billiard ball in a planar domain D.](image-url)
Figure 2: $v$ is an inward velocity vector, and $w$ is an outward velocity vector.

Figure 3: In a non-convex domain $D$, the velocity of a billiard ball may be tangent to the boundary.

space of the billiard in $D$ is a cylinder with coordinates $(t, \alpha)$, where $t$ is a real number defined up to an integer multiple of the total length $L$ of $\partial D$, and $\alpha \in (0, \pi)$.

Let $M$ be the phase cylinder of the billiard in $D$. Define the billiard map $T: M \to M$ as follows: for an outward velocity vector $v$ at a boundary point $x \in \partial D$, its image under $T$ is the velocity of the billiard ball with initial velocity $v$ immediately after it is reflected from the boundary for the first time, see Figure 5. Then the iterate $T^k$ computes the velocity of the ball after it hits the boundary for the $k$'th time (see Figure 6). So, the dynamics of the billiard ball is pretty much determined by the behavior of the billiard map $T: M \to M$ and its iterates. For this reason, most (if not all) people studying billiards study properties of this map.

The first thing we would like to understand about this map $T$ is how regular it is. This very much depends on the properties of the domain $D$ and its boundary $\partial D$. In particular, if $D$ is not convex, then $T$ does not even have to be continuous. For example, in Figure 7, initially close vectors $v$ and $w$ are mapped by $T$ to vectors that are far away from each other.

For convex domains with smooth boundary, a routine application of the implicit mapping theorem show that $T$ is continuous, and, moreover, a diffeomorphism of the open phase cylinder $M = \partial D \times (0, \pi)$.

Figure 4: An element of the phase space for the billiard in $D$ is a point $x \in D$ and an inward tangent vector at $v \in T_xD$. The latter vector can be characterized by the angle $\alpha \in (0, \pi)$ it makes with the positive direction of $\partial D$. 
to itself. If, moreover, the boundary of $D$ has strictly positive curvature, then we have the following result:

**Theorem 1.2** (see e.g. [2]). If the boundary of the domain $D$ is smooth and has strictly positive curvature, then the corresponding billiard map $T: M \rightarrow M$ extends to a diffeomorphism of the closed phase cylinder $\partial D \times [0, \pi]$ to itself.

In other words, for sufficiently nice boundaries, the billiard map stays well-defined and smooth when the initial velocity vector tends to a vector parallel to the boundary.

**Lecture 2: Invariant area form of the billiard map and Poincaré’s recurrence theorem**

Recall that the phase space $M$ of a billiard in a domain $D$ is the space of pairs $(x, v)$, where $x \in \partial D$ is a point at the boundary of $D$, and $v \in T_x D$ is an inward unit tangent vector at $x$. Since $\partial D$ is homeomorphic to a circle, and the space of unit inward tangent vectors at each point is homeomorphic to an open interval, the phase space $M$ is topologically a cylinder. As coordinates on $M$, we take the arc
length parameter \( t \) on \( \partial D \), and the angle \( \alpha \) which the inward vector \( v \) makes with the positive direction of \( \partial D \). Note that the coordinate \( t \) is defined modulo the total length \( L \) of \( \partial D \), while \( \alpha \) is well-defined real number in \((0, \pi)\). Consider the following 2-form on \( M \):

\[
\omega = \sin(\alpha) d\alpha \wedge dt.
\]

Note that even though \( t \) is not quite a function (it is defined up to an additive constant \( L \), so it is in a fact a map to a circle), its differential \( dt \) is a well-defined 1-form. So, the 2-form \( \omega \) is well-defined as well. Furthermore, we have \( d\alpha \wedge dt \neq 0 \), because \( \alpha \) and \( t \) are, by construction, coordinates on \( M \). Finally, notice that \( \sin(\alpha) \neq 0 \), since \( \alpha \in (0, \pi) \). So, \( \omega \) is a non-vanishing 2-form on a 2-manifold \( M \), i.e. it is an area form (a particular case of a volume form, that is a non-vanishing \( n \)-form on an \( n \)-dimensional manifold).

**Theorem 2.1.** The area form \( \omega \) on \( M \) is preserved by the billiard map \( T \): \( T^* \omega = \omega \). In other words, for any domain \( U \subset M \), we have

\[
\int_U \omega = \int_{T(U)} \omega,
\]

provided that at least one of these integrals is well-defined.

The proof is based on the following lemma:

**Lemma 2.2.** Let \( \gamma \) be an arc length parametrized curve in \( \mathbb{R}^2 \). Let also \( ||\gamma(t') - \gamma(t)|| \) be the Euclidian distance between the points \( \gamma(t) \) and \( \gamma(t') \). Finally, let \( \alpha \) and \( \alpha' \) be the angles between the chord joining \( \gamma(t) \) and \( \gamma(t') \) and the arc of \( \gamma \) connecting those two points (see Figure 8). Then

\[
\frac{\partial}{\partial t} ||\gamma(t') - \gamma(t)|| = -\cos(\alpha), \quad \frac{\partial}{\partial t'} ||\gamma(t') - \gamma(t)|| = \cos(\alpha').
\]  

(1)

**Proof of the lemma.** The expression \( ||\gamma(t') - \gamma(t)|| \) is a symmetric function in \( t \) and \( t' \), so its partial derivatives with respect to these variables should be the same, up to replacing \( t \) with \( t' \). The reason for different signs of the partial derivatives in (1) is the asymmetric definition of the angles \( \alpha \) and \( \alpha' \): \( \alpha \) is the angle between the chord \( \gamma(t)\gamma(t') \) and the positive direction of \( \gamma \), while \( \alpha' \) is the angle between the same chord and the negative direction of \( \gamma \) (see Figure 8). So, it suffices to establish the second of formulas (1).

The first one then follows by symmetry.

To prove second of formulas (1), fix \( t \) and consider the function \( g(x) = ||x - \gamma(t)|| \), where \( x \in \mathbb{R}^2 \). The level set of this function through the point \( \gamma(t') \) is a circle centered at \( \gamma(t) \). Therefore, the gradient of \( g(x) \) at \( \gamma(t') \) is orthogonal to that circle and hence collinear to the vector \( \gamma(t') - \gamma(t) \), see Figure 8 (this can also be seen from the definition of the gradient as the direction of fastest increase). Furthermore, the derivative of the function \( g(x) = ||x - \gamma(t)|| \) at \( x = \gamma(t') \) in the direction \( \gamma(t') - \gamma(t) \) is clearly equal to

\[1\] Here we distinguish between the derivative in the direction \( v \), and the derivative along \( v \). By definition, the derivative in the direction \( v \) is the derivative along \( v/||v|| \).
1, so the gradient of \( g(x) \) at \( \gamma(t') \) is a unit vector positively collinear to \( \gamma(t') - \gamma(t) \) (here we use that the magnitude of the gradient is equal to the derivative of the function in the direction of the gradient, and also that a function must increase, not decrease in the direction of its gradient).

Now we have that \( ||\gamma(t') - \gamma(t)|| = g(\gamma(t')) \), so the partial derivative of \( ||\gamma(t') - \gamma(t)|| \) with respect to \( t' \) is equal to the derivative of the function \( g(x) \) along the velocity vector \( \gamma'(t') \) of the curve \( \gamma \) at the point \( \gamma(t') \). Therefore, we have

\[
\frac{\partial}{\partial t'} ||\gamma(t') - \gamma(t)|| = \langle \text{grad} g(\gamma(t')), \gamma'(t') \rangle,
\]

where \( \text{grad} g(\gamma(t')) \) is the gradient of \( g(x) \) at \( x = \gamma(t') \). Furthermore, we already saw that \( \text{grad} g(\gamma(t')) \) is a unit vector, while \( \gamma'(t') \) is a unit vector because the curve \( \gamma \) is arc length parametrized. So, the inner product between these two vectors is equal to the cosine of the angle between them. To complete the proof, it suffices to notice that this angle is precisely \( \alpha' \) (see Figure 9).

**Proof of Theorem 2.7.** Let \( (t', \alpha') \) be the coordinates of the image of the point \( (t, \alpha) \) under the billiard map \( T \). Then \( t', \alpha' \) are smooth functions of \( t, \alpha \), and hence smooth functions on \( M \). Let also \( \gamma \) be the arc length parametrized boundary of the billiard table \( D \). Then \( \alpha' \) can be defined as the angle between the chord \( \gamma(t)\gamma(t') \) and the negative direction of the curve \( \gamma \), see Figure 10.

Now, consider the function \( f = ||\gamma(t') - \gamma(t)|| \). Since \( t \) and \( t' \) are smooth functions on \( M \), so is \( f \). Furthermore, by Lemma 2.2 we have

\[
df = \cos(\alpha')dt' - \cos(\alpha)dt.
\]

Consider now the 1-form \( \xi \) on \( M \) given by

\[
\xi = -\cos(\alpha)dt.
\]

Notice that

\[
d\xi = d(-\cos(\alpha)) \wedge dt = \sin(\alpha)d\alpha \wedge dt = \omega.
\]

Furthermore, by (2) we have

\[
df = \xi - T^*\xi.
\]

Any \( f \) satisfying this equation for some \( \xi \) such that \( d\xi = \omega \) is called a generating function of the map \( T \). A map possessing a generating function is automatically area-preserving. Indeed, taking the differential of both sides in (3), we get

\[
0 = d\xi - dT^*\xi = d\xi - T^*d\xi = \omega - T^*\omega,
\]

so

\[
T^*\omega = \omega,
\]

as desired. Thus, the billiard map \( T \) has a generating function given by \( ||\gamma(t') - \gamma(t)|| \) and hence is area-preserving, q.e.d.

![Figure 9: The gradient of the function \( g(x) = ||x - \gamma(t)|| \) at the point \( x = \gamma(t') \).](image-url)
Figure 10: The billiard map $T$ takes the point $(t, \alpha)$ to the point $(t', \alpha')$.

**Theorem 2.3** (Poincaré’s recurrence theorem for billiards). The billiard map $T : M \to M$ has the following recurrence property: for any open subset $U \subset M$, almost all points in $U$ will eventually return, under the action of iterations of $T$, to the set $U$. In other words, for almost any $v \in U$ there exists a natural number $n > 0$ such that $T^n(v) \in U$.

**Remark 2.4.** As one can see from the proof, this theorem is true for any volume-preserving map of an $n$-dimensional manifold, provided that the total volume of the manifold is finite. Furthermore, the same result holds for any measure-preserving map of any measure space, again provided that the measure of the whole space is finite. This is the way this theorem is usually stated.

**Proof of Theorem 2.3.** Let $V \subset U$ be the set of points which do not return to $U$ under the iterations of $T$. This means that $V$ consists of points $v \in U$ such that $T^n(v) \notin U$ for any natural number $n > 0$. We need to show that the set $V$ has area 0. To that end, notice that for any natural $n > 0$, the set $T^n(V)$ is disjoint from $V$. Indeed, by definition of $V$, the set $T^n(V)$ does not intersect $U$ and hence $V$. So, we have that $T^n(V) \cap V$ is an empty set. Furthermore, since $T$ is invertible, it follows that for any integer $l \geq 0$ the set $T^{n+l}(V) \cap T^l(V)$ is also empty (if $v \in T^{n+l}(V) \cap T^l(V)$, then it must be that $T^{-l}(v) \in T^n(V) \cap V$, which is impossible). And since $n > 0$ and $l \geq 0$ are arbitrary, it follows that the sets

$$V, T(V), T^2(V), \ldots$$

(4)

are all pairwise disjoint. So, their total area cannot exceed the total area of the phase cylinder $M$, which is

$$\int_0^L \int_0^\pi \sin(\alpha) d\alpha dt = 2L,$$

i.e. twice the perimeter of the billiard table. At the same time, since $T$ is area-preserving, all the sets have the same area and thus may have finite total area only if each of them has area zero. Thus, the theorem is proved.

**Remark 2.5.** There is one step that we skipped in the proof: one actually needs to explain why the set $V$ has well-defined area at all. This can be done by showing that $V$ is a difference of two open sets and hence measurable.

**Example 2.6.** Consider a billiard table $D$ of an arbitrary shape, and let $I \subset \partial D$ be a (possibly very small) open subset of the boundary. Consider billiard trajectories that start at points of $I$ and make an angle with $\partial D$ which is between $89^\circ$ and $91^\circ$. Then, according to Poincaré’s theorem (applied to the subset $U = I \times (89^\circ, 91^\circ) \subset M$), almost all of these trajectories will eventually hit $I$ again, and, moreover, the angle of incidence will again be between $89^\circ$ and $91^\circ$. Of course, it may (and, generally speaking, will) take a long long time for this event to occur: Poincaré’s recurrence theorem is saying that for almost all
initial data $v$ in our open subset $U$, there is $n > 0$ such that $T^n(v) \in U$, but it is saying nothing about what the value of $n$ is.

**Lecture 3: Billiards in disks**

We now turn to studying concrete examples of billiards. We will start with the simplest case of billiards in disks. Clearly, it is sufficient to consider the case when $D$ is a unit disk: billiards in disks of different radii behave in the same way. Let $A$ be a point in the unit circle. Consider the billiard trajectory in the unit disk which starts at $A$ and makes an angle $\alpha$ with the positively oriented unit circle (see Figure 11). Then this trajectory will meet the unit circle again at a point $B$, and the angle of incidence is also equal to $\alpha$ (because a chord intersects a circle at the same angle at both intersection points). Therefore, after the trajectory is reflected at $B$, it will again make an angle $\alpha$ with the positive direction of the unit circle. This means that billiard map $T$ does not affect the $\alpha$-coordinate on the phase cylinder $M$ at all: $\alpha' = \alpha$. Further, denote the $t$-coordinate of $A$ by $t$, and the $t$-coordinate of $B$ by $t'$. Then the increment $t' - t$ is equal to the measure of the arc of the circle that goes from $A$ to $B$ (in the counter-clockwise direction). This measure is equal to the angle $\angle AOB$ (where $O$ is the center of the circle), which, in turn, is equal to $2\alpha$ (by the tangent-chord theorem combined with the inscribed angle theorem). Therefore, in coordinates, the billiard map $T$ in the unit disk is given by

$$
\begin{cases}
  t' = t + 2\alpha, \\
  \alpha' = \alpha,
\end{cases}
$$

where addition in the formula for $t'$ is understood modulo the total length of the unit circle, which is $2\pi$. This is a linear map (at least if we forget about the periodicity of the $t$-coordinate), and the corresponding dynamics (i.e. the behavior of iterates of this map) is easy to understand. Namely, we have the following explicit formulas for the iterate $T^k$:

$$
\begin{cases}
  t' = t + 2n\alpha, \\
  \alpha' = \alpha,
\end{cases}
$$

where, again, addition in the formula for $t'$ is understood modulo $2\pi$. We, however, would like to understand what the dynamics looks like qualitatively. To that end, notice that since $\alpha$ is preserved by the map, we can just fix it and study the dynamics, of the $t$ variable, which is given by

$$
t \mapsto t + 2\alpha.
$$

![Figure 11: Billiard in a disk.](image-url)
Recall that $t$ can be thought of as a point in a circle, $t \in \mathbb{R}/2\pi\mathbb{Z}$. The map (5) is then a rotation of that circle by angle $2\alpha$. This rotation behaves differently depending on whether the number $\alpha/\pi$ is rational or irrational. If $\alpha/\pi = m/n$ is rational, then $2n\alpha = 2m\pi$, and thus the $n$'th power of the map (5) is the identity map. So, for $\alpha/\pi \in \mathbb{Q}$, the map (5) is periodic. In terms of the billiard, this means that any trajectory with such $\alpha$ eventually closes up. Figure 12 shows two 5-periodic trajectories corresponding to $\alpha = \pi/5$ and $\alpha = 2\pi/5$.

Now, consider the case $\alpha/\pi \notin \mathbb{Q}$. In this case, we can still approximate $\alpha/\pi$ with a rational, with arbitrary precision (the denominator of that rational may be large, though). So, even though the dynamics in this case is not periodic, it is close to periodic. Such dynamics is known as almost periodic or quasi-periodic (the periodic case can also be considered a particular case of the quasi-periodic one). Note that in this case $t + 2n\alpha \neq t + 2m\alpha$ modulo $2\pi$ unless $m = n$, which means that the orbit of any $t \in S^1$ under iterations of the map (5) is infinite. One can in fact quite easily show that any orbit is dense in the circle. Figure 13 shows what a piece of a non-periodic billiard trajectory in a disk, corresponding to $\alpha/\pi \notin \mathbb{Q}$, looks like. We in fact generated this as a periodic trajectory with large period. In practice, one does not really see the difference between quasiperiodic dynamics and periodic dynamics with large period.

We conclude this discussion by noticing that Poincaré recurrence clearly holds for the billiard in a disk (as it should by Theorem 2.3). Moreover, since the dynamics is quasi-periodic, every point eventually returns to any its small neighborhood.

**Lecture 4: Billiards in ellipses**

We now turn our attention to billiards in ellipses. Most of the results we will obtain also apply to hyperbolas. The case of a hyperbola is, however, somewhat more sophisticated, because, in contrast to ellipses, hyperbolas do not bound compact domains. So we will stick with ellipses. That being said,
hyperbolas will still be appearing from time to time in our study.

Consider an ellipse given in a Cartesian coordinate system \((x_1, x_2)\) by

\[
\frac{x_1^2}{a_1^2} + \frac{x_2^2}{a_2^2} = 1.
\]

As known from linear algebra, any ellipse can be written in this form in a suitable orthogonal coordinate system.

Given a point \((x, v)\) in the phase cylinder \(M\) for the billiard in the ellipse, consider the function

\[
J = -\left(\frac{x_1 v_1}{a_1^2} + \frac{x_2 v_2}{a_2^2}\right),
\]

(6)

where \((x_1, x_2)\) are components of \(x\), and \((v_1, v_2)\) are components of \(v\).

**Proposition 4.1.** The function \(J\) is invariant under the billiard map \(T: M \to M\), i.e. \(J(T(x, v)) = J(x, v)\).

**Remark 4.2.** A function invariant under a map is called a conserved quantity, a first integral, or just an integral of the map. This particular function \(J\) is known as the Joachimsthal integral. Note that this integral can also be rewritten as

\[
J = -\frac{1}{2} \langle v, \nabla f(x) \rangle,
\]

where

\[
f = \frac{x_1^2}{a_1^2} + \frac{x_2^2}{a_2^2}
\]

is the defining function of the ellipse. This explains the negative sign in the definition: since \(\nabla f(x)\) is an outward normal to the ellipse, and \(v\) is an inward vector, the angle between \(\nabla f(x)\) and \(v\) must be obtuse, and \(\langle v, \nabla f(x) \rangle < 0\). Therefore, our function \(J\) is always positive. Of course, the function

\[
-J = \frac{x_1 v_1}{a_1^2} + \frac{x_2 v_2}{a_2^2}
\]

is also preserved by the billiard map, it is just somewhat more pleasant to work with positive functions than with negative ones.

**Remark 4.3.** Also notice that in the case of a unit circle the norm of the vector \(\nabla f(x)\) is equal to 2, so

\[
J = -\frac{1}{2} \langle v, \nabla f(x) \rangle = -\cos(\alpha + \pi/2) = \sin(\alpha),
\]

where we also used that \(v\) is a unit vector, and that its angle with the outward normal \(\nabla f(x)\) is equal to \(\alpha + \pi/2\), with \(\alpha\) being the angle between \(v\) and the positive direction of the unit circle. So, in the circle case the conserved quantity \(J\) essentially coincides with the conserved quantity \(\alpha\) which we found in the previous lecture (more precisely, \(J = \sin(\alpha)\), but saying that \(\alpha\) is preserved is more or less the same as saying that \(\sin(\alpha)\) is preserved). So, we already know that Proposition 4.1 is true for the circle.

**Proof of Proposition 4.1.** Let \((x, v) \in M\) be a point in the phase cylinder, and let \((x', v') = T(x, v)\) be its image under the billiard map. Then, by definition of the billiard map, the vectors \(v\) and \(v'\) make the same angle with the (tangent line to the) ellipse at \(x'\). Since those are unit vectors, this is equivalent to saying that the vector \(v + v'\) is tangent to the ellipse at \(x'\), see Figure 14. Using also that the gradient \(\nabla f(x')\) is orthogonal to the ellipse at \(x'\), this gives

\[
\langle v + v', \nabla f(x') \rangle = 0,
\]

10
so
\[ J(x', v') = -\frac{1}{2} \langle v', \text{grad } f(x') \rangle = \frac{1}{2} \langle v, \text{grad } f(x') \rangle. \]

We need to show that this is equal to
\[ J(x, v) = -\frac{1}{2} \langle v, \text{grad } f(x) \rangle, \]
which is equivalent to proving that
\[ \langle v, \text{grad } f(x) + \text{grad } f(x') \rangle = 0. \]

Also notice that \( v \) is collinear to \( x' - x \) (by definition of the billiard map), so we need to show that
\[ \langle x - x', \text{grad } f(x) + \text{grad } f(x') \rangle = 0. \tag{7} \]

To show this, we will use that \( f \) is a homogeneous quadratic function. Write \( f \) in the form \( f = \langle x, Ax \rangle \), where
\[ A = \begin{pmatrix} \frac{1}{a_1^2} & 0 \\ 0 & \frac{1}{a_2^2} \end{pmatrix}. \]

Notice that \( \text{grad } f(x) = 2Ax \). So, the left hand side of (7) can be rewritten as
\[ 2\langle x - x', Ax + Ax' \rangle = 2 \left( \langle x, Ax \rangle - \langle x', Ax' \rangle + \langle x, Ax' \rangle - \langle x', Ax \rangle \right). \]

But \( \langle x, Ax \rangle = f(x) = 1 \), since \( x \) lies on the ellipse. Likewise, \( \langle x', Ax' \rangle = 1 \). Finally, notice that \( \langle x, Ax' \rangle = \langle x', Ax \rangle \) since \( A \) is symmetric. So, we conclude that (7) indeed holds, as desired.

We now want to find a geometric interpretation of the Joachimsthal integral \( J \). To that end, we will need a more geometric definition of an ellipse:

**Definition 4.4.** Let \( f_1, f_2 \) be two points in the Euclidian plane \( \mathbb{R}^2 \), and let \( l > 0 \) be a positive real number. Then the set of points \( \{ x \in \mathbb{R}^2 \mid \|xf_1\| + \|xf_2\| = l \} \) is called an ellipse with foci \( f_1, f_2 \). Here \( \|xf_i\| \) stands for the Euclidian distance between \( x \) and \( f_i \). Similarly, the set of points \( \{ x \in \mathbb{R}^2 \mid \|xf_1\| - \|xf_2\| = l \} \) is called a hyperbola with foci \( f_1, f_2 \).

We now want to obtain an analytic description of ellipses and hyperbolas with given foci \( f_1, f_2 \). We will assume that \( f_1 = (-a, 0) \), \( f_2 = (a, 0) \). By doing so we do not loose any generality because any pair of points has such coordinates in a suitable orthogonal coordinate system.
Proposition 4.5. The equation of an ellipse/hyperbola with foci \( f_1 = (-a, 0), f_2 = (a, 0) \) and given parameter \( l > 0 \) is
\[
\frac{x_1^2}{l^2/4} + \frac{x_2^2}{l^2/4 - a^2} = 1. \tag{8}
\]

Remark 4.6. The reader may be confused by the fact that an ellipse and a hyperbola with the same parameters are described by the same equation. But there is no contradiction here. The point is that by the triangle inequality we must always have \( l > 2a \) for an ellipse and \( l < 2a \) for a hyperbola. So for any \( l \) there may exist either an ellipse, or a hyperbola with such \( l \), but not both at a time.

The proof of Proposition 4.5 is a straightforward algebraic verification of the equivalence between equation (8) and Definition 4.4.

Lecture 5: Geometric meaning of the Joachimsthal integral

In this lecture we will reveal the geometric meaning of the Joachimsthal integral \( \int \) for the billiard map in an ellipse. Consider once again an ellipse given by
\[
\frac{x_1^2}{a_1^2} + \frac{x_2^2}{a_2^2} = 1. \tag{9}
\]
Without loss of generality, we may assume that this ellipse is “horizontal”, i.e. \( a_1 \geq a_2 \) (this can be always achieved by rotating the coordinate system if necessary). Furthermore, since we already studied the circle case in detail, we will assume that our ellipse is not a circle, which means that \( a_1 > a_2 \).

Proposition 5.1. The foci of the ellipse (9) are the points \((\pm a, 0)\), where \( a = \sqrt{a_1^2 - a_2^2} \).

Remark 5.2. We assumed that \( a_1 \geq a_2 \) to have foci on the horizontal axis. For \( a_1 \leq a_2 \), the foci are on the vertical axis (see Figure 15).

Proof of Proposition 5.1. Equation (9) coincides with (8) if we set \( a = \sqrt{a_1^2 - a_2^2} \) and \( L = 2a_1 \).

It follows that the equations of conics (i.e. ellipses and hyperbolas) which are confocal with (i.e. have the same foci as) our ellipse (9) is
\[
\frac{x_1^2}{l^2/4} + \frac{x_2^2}{l^2/4 - a_1^2 + a_2^2} = 1.
\]
This can be written in a more symmetric way if we define \( \lambda = l^2/4 - a_1^2 \). Then the above equation becomes
\[
\frac{x_1^2}{a_1^2 - \lambda} + \frac{x_2^2}{a_2^2 - \lambda} = 1. \tag{10}
\]
This is a standard equation of a *confocal family*. For an arbitrary value of $\lambda \neq a_1^2, a_2^2$, this equation defines a conic, and all these conics have the same foci as the ellipse (9). And conversely, any conic confocal with the ellipse (9) is contained in the family (10). Figure 16 shows an example of a confocal family.

We will now be interested in the following question: given a line in $\mathbb{R}^2$, how many, if any, conics from the family (10) are tangent to that line? The following lemma says that there is almost always exactly one, except for a couple of cases when there is none:

**Lemma 5.3.** Consider the line $x + tv$ through the point $x = (x_1, x_2)$ with direction $v = (v_1, v_2)$. Then

1. This line is tangent to at most one conic from the family (10).

2. If such a tangent conic exists, then its parameter $\lambda$ which distinguishes it in the family (10) is given by

   $$
   \lambda = \frac{a_1^2 v_2^2 + a_2^2 v_1^2 - (x_1 v_2 - x_2 v_1)^2}{v_1^2 + v_2^2}.
   $$

3. Such tangent conic does not exist in the following two cases:

   (a) The number $\lambda$ given by formula (11) is equal to $a_1^2$, in which case the line $x + tv$ coincides with the minor axis of the ellipse (9).

   (b) The number $\lambda$ given by formula (11) is equal to $a_2^2$, in which case the line $x + tv$ is passing through one of the foci of the ellipse (9).

**Proof.** The line $x + tv$ is tangent to the conic (10) when the equation

$$
\frac{(x_1 + tv_1)^2}{a_1^2 - \lambda} + \frac{(x_2 + tv_2)^2}{a_2^2 - \lambda} = 1
$$

for their intersection points has exactly one solution in terms of $t$; that is, when its discriminant is equal to 0. Equating the discriminant to 0 and solving for $\lambda$, we get formula (11). The expression on the right-hand side of (11), however, may be equal to $a_1^2$ or $a_2^2$, which does not correspond to any conic in the family (10). In that case, there is no conic in the family (10) tangent to the line $x + tv$ in all other cases, such a conic exists, is unique, and corresponds to $\lambda$ given by (11).

To complete the proof it now suffices to obtain a geometric interpretation of the cases $\lambda = a_1^2$ and $\lambda = a_2^2$. First assume that $\lambda = a_1^2$. Then (11) gives

$$
(a_1^2 - a_2^2) v_1^2 = -(x_1 v_2 - x_2 v_1)^2.
$$
Notice the left-hand side of this equation is non-negative (since \( a_1 > a_2 > 0 \)), while the right-hand side is non-positive. Therefore, this equation holds if and only if both sides are equal to 0, which is equivalent to \( v_1 = 0 \) and \( x_1 = 0 \). But these two conditions together hold precisely when the line \( x + tv \) coincides with the vertical coordinate axis or, which is the same, with the minor axis of the ellipse \( (9) \).

Similarly, if \( \lambda = a_2^2 \), then

\[
(a_2^2 - a_1^2)v_2^2 = -(x_1v_2 - x_2v_1)^2,
\]

which is equivalent to

\[
x_1v_2 - x_2v_1 = \pm av_2,
\]

with \( a = \sqrt{a_1^2 - a_2^2} \). The geometric meaning of this equation is that the vectors \((x_1, x_2) - (\pm a, 0)\) and \((v_1, v_2)\) are collinear, which is the same as to say that the line \( x + tv \) is passing through the one of the points \((a, 0), (-a, 0)\). But those points are precisely the foci of the ellipse \( (9) \), hence the result.

We now show that Joachimsthal integral computed at a point \((x, v)\) of the phase cylinder for the billiard in an ellipse is closely related to the parameter \( \lambda \) of the confocal conic to which the line \( x + tv \) is tangent:

**Proposition 5.4.** The quantity \( \lambda \) given by formula \((11) \), regarded as a function on the phase cylinder \( M \) of the billiard in the ellipse \((9) \), is related to the Joachimsthal integral \( J \) by the formula

\[
\lambda = a_1^2a_2^2J^2.
\]

**Proof.** For a unit vector \( v \), formula \((11) \) becomes

\[
\lambda = a_1^2v_2^2 + a_2^2v_1^2 - (x_1v_2 - x_2v_1)^2 = (a_1^2 - x_1^2)v_2^2 + (a_2^2 - x_2^2)v_1^2 + 2x_1x_2v_1v_2.
\]

Furthermore, since the point \((x_1, x_2)\) lies on the ellipse \((9) \), we have

\[
\frac{x_1^2}{a_1^2} + \frac{x_2^2}{a_2^2} = 1 \quad \Rightarrow \quad \frac{x_1^2}{a_1^2} + \frac{x_2^2}{a_2^2} = 0 \quad \Rightarrow \quad a_1^2 - x_1^2 = \frac{a_2^2x_2^2}{a_2^2}.
\]

Similarly, we have

\[
a_2^2 - x_2^2 = \frac{a_2^2x_1^2}{a_1^2},
\]

so \((13) \) can be written as

\[
\lambda = \frac{a_2^2x_2^2}{a_2^2} + \frac{a_2^2x_1^2}{a_1^2} + 2x_1x_2v_1v_2 = a_1^2a_2^2 \left( \frac{x_1v_1}{a_1^2} + \frac{x_2v_2}{a_2^2} \right)^2 = a_1^2a_2^2J^2,
\]

as desired. \(\square\)

**Corollary 5.5.**

1. Assume that a segment of a billiard trajectory in the ellipse \((9) \) is tangent to some confocal conic \((10) \). Then all segments of that trajectory are tangent to that conic.

2. Assume that a segment of a billiard trajectory in the ellipse \((9) \) belongs to the minor axis. Then all segments belong to the minor axis.

3. Assume that a segment of a billiard trajectory in the ellipse \((9) \) is passing through one of the foci. Then all segments are passing through one of the foci.

**Proof.** From \((12) \) and preservation of \( J \) it follows that the function \( \lambda \) given by \((11) \) is preserved by the billiard map. The above three statements are the particular cases of this result corresponding to, respectively, generic values of \( \lambda, \lambda = a_1^2, \) and \( \lambda = a_2^2 \). \(\square\)
Remark 5.6. In view of Proposition 5.4 Corollary 5.5 is equivalent to preservation of the Joachimsthal integral and thus can be viewed as a geometric form of the latter. Furthermore, all statements of Corollary 5.5 can be obtained geometrically. The second statement is particularly straightforward: since the minor axis is orthogonal to the ellipse, a billiard ball moving along that axis will continue doing so after any number of reflections (see Figure 17). Of course, the same is true for the major axis, but the major axis is not distinguished by any specific value of $\lambda$: its $\lambda$ is the same as for any other line passing through one of the foci.

Also note that the third statement of the corollary can be strengthened as follows: if a segment of a billiard trajectory in the ellipse is passing through one of the foci, then the next segment must pass through the other focus. Indeed, a generic billiard trajectory through one of the foci is not orthogonal to the ellipse (see Figure 18) and hence cannot be reflected to the same focus. So, almost all trajectories through one of the foci get reflected to the other focus, and by continuity it must be true for all such trajectories.

The fact that a billiard trajectory through one of the foci is reflected to the other focus is known as the \textit{optical property of the ellipse}. It can be reformulated by saying that all light rays starting at one of the foci get reflected, by an elliptic mirror, to the other focus. The optical property can be proved quite easily without using the Joachimsthal integral, see e.g. [4, Lemma 4.2]. See also [4, Theorem 4.4] for an independent geometric proof of the first result of Corollary 5.5.

Lecture 6: The phase portrait of the billiard in an ellipse

The phase cylinder $M$ of the billiard in an ellipse

$$\frac{x_1^2}{a_1^2} + \frac{x_2^2}{a_2^2} = 1. \tag{14}$$

is foliated into level sets of the Joachimsthal integral

$$J = -\left( \frac{x_1v_1}{a_1^2} + \frac{x_2v_2}{a_2^2} \right). \tag{15}$$

Each of these level sets is preserved by the billiard map $T$. Furthermore, according to the previous lecture each of these level sets can be seen as the set of velocity vectors $(x, v) \in M$ corresponding to billiard

Figure 17: A billiard trajectory along the minor axis of an ellipse.

Figure 18: A segment of a billiard trajectory through a focus of an ellipse.
trajectories tangent to one and the same conic confocal with the ellipse \([14]\). This confocal conic is explicitly given by
\[ \frac{x_1^2}{a_1^2} - \lambda + \frac{x_2^2}{a_2^2} = 1, \tag{16} \]
where the constant \(\lambda\) is related to the value of the Joachimsthal integral by the formula
\[ \lambda = a_1^2 a_2^2 J^2. \tag{17} \]

We would now like to understand the structure of level sets of the function \(J\). To that end, we rewrite this function in terms of coordinates on \(M\). As one of the coordinates, we take the angle \(\alpha\) (the angle between the velocity vector \(v\) and the positively oriented ellipse), and as a second coordinate we take the parameter \(t\) on the ellipse \([14]\) corresponding to the parametrization
\[ x_1 = a_1 \cos(t), \quad x_2 = a_2 \sin(t). \]

Remark 6.1. Note that this is not an arc length parameter, and in fact one cannot write the arc length parametrization of an ellipse in terms of elementary functions.

Now, as in Lecture 4 we rewrite \([15]\) as
\[ J = -\frac{1}{2} \langle v, \nabla f(x) \rangle, \tag{18} \]
where
\[ f(x) = \frac{x_1^2}{a_1^2} + \frac{x_2^2}{a_2^2}. \]

Further, since \(v\) is a unit vector, and \(\nabla f(x)\) is an outward normal to the ellipse, \([18]\) can be rewritten as
\[ J = -\frac{1}{2} ||\nabla f(x)|| \cos(\alpha + \pi/2) = \frac{1}{2} ||\nabla f(x)|| \sin \alpha = ||\nabla (f(x)/2)|| \sin \alpha = \sqrt{\frac{x_1^2}{a_1^2} + \frac{x_2^2}{a_2^2}} \sin \alpha, \]
so in \((t, \alpha)\) coordinates we have
\[ J = \sqrt{\cos^2 t / a_1^2 + \sin^2 t / a_2^2} \sin \alpha. \]

The level sets of this function are shown in Figure 19. This picture should be thought of as living on a cylinder, which is obtained by identifying the opposite sides \(t = 0\) and \(t = 2\pi\) of the colored rectangle. The solid dots are located at critical points \((0, \pi/2), (\pi/2, \pi/2), (\pi, \pi/2), (3\pi/2, \pi/2)\) of the function \(J\). They correspond to (initial velocity vectors of) billiard trajectories aligned with one of the axes of the ellipse. The colors are used to show two regions with qualitatively different behavior of billiard trajectories, as explained below.

Consider first the blue region. On the exterior boundaries of this region (given by \(\alpha = 0\) and \(\alpha = \pi\)) we have \(J = 0\), while the value of \(J\) on the interior boundary is equal to that at the point \((0, \pi/2)\) and hence is \(1/a_1\). Therefore, by formula \([17]\), we have \(\lambda = 0\) at the exterior boundary and \(\lambda = a_2^2\) at the interior boundary. And since there is no critical points in the interior of the region, in that interior we must have \(0 < \lambda < a_2^2\). But then it follows from equation \([16]\) that confocal conics tangent to billiard rays with initial conditions in the (interior of the) blue region are ellipses. Another important feature of the blue domain is that even though each level set of \(J\) in that domain consists of two connected components (corresponding
Figure 19: The phase portrait of the billiard in an ellipse.

to $\alpha < \pi/2$ and $\alpha > \pi/2$), the billiard map preserves each of those components individually. Indeed, if we take an initial velocity vector $v$ with angle $\alpha$ close to 0, then its image under the billiard map will also have $\alpha$ close to 0. Therefore, components of level sets of $J$ close to the lower boundary $\alpha = 0$ cannot be mapped to components close to the upper boundary $\alpha = \pi$. But then it follows by continuity that all components in the $\alpha < \pi/2$ region are mapped to themselves, and the same is true for the $\alpha > \pi/2$ region.

As for the green region, a similar analysis shows that it corresponds to trajectories tangent to confocal hyperbolas (except for the critical points $(\pi/2, \pi/2)$, $(3\pi/2, \pi/2)$ whose corresponding trajectories are aligned with the minor axis and hence are not tangent to any confocal conic). Another difference with the blue region is that for the green one two connected components of each $J$-level are interchanged by the billiard map. Indeed, this is obviously true for the critical points $(\pi/2, \pi/2)$, $(3\pi/2, \pi/2)$ and thus is true for all components by continuity.

The curves separating the regions of different color are separatrices. Since at those curves we have $\lambda = a^2/2$, they correspond to billiard trajectories through the foci of the ellipse. Consider, for example, the upper-left separatrix. On this separatrix we have $t < \pi$ and $\alpha > \pi/2$, which means that the corresponding velocity vectors $v$ are attached at the upper half of the ellipse and make an obtuse angle with its positive direction. From this it is easy to see that such vectors give rise to billiard trajectories whose initial segments pass through the right focus. Similarly, the lower-right separatrix also corresponds to the right focus, while the lower-left and the upper-right ones correspond to the right focus.

Lecture 7: Liouville integrable maps in 2D

The billiard map for an ellipse preserves an area form and a function. In this lecture, we will discuss this situation in an abstract setting:

**Definition 7.1.** Let $M$ be a 2-dimensional manifold, and let $T: M \to M$ be an area-preserving diffeomorphism[^1] (i.e. there is an area form $\omega$ on $M$ such that $T^*\omega = \omega$). Assume also that $T$ has a first

[^1]: If $M$ is compact, then any area-preserving smooth map of $M$ to itself is automatically a diffeomorphism. This is, however,
integral (i.e. a smooth function \( f: M \to \mathbb{R} \) such that \( T^*f = f \)) whose differential does not vanish almost everywhere on \( M \). Then the map \( T \) is called Liouville integrable, completely integrable, or just integrable.

Here “almost everywhere” means “on an open dense subset”. Note that the set where \( df \neq 0 \) is automatically open, so what is really required is that this set is dense. This is equivalent to saying that there exist no open subset \( U \subset M \) such that the restriction of \( f \) to \( U \) is constant. In particular, this prohibits the situation when \( f \) is a constant function. Constant functions are first integrals for any map, so there is absolutely no reason to call the above map \( T \) integrable if it possesses a constant first integral.

One of the main general results on integrable systems is that integrability implies quasi-periodic dynamics. More precisely, under some natural assumptions, integrability implies quasi-periodic dynamics on regular level sets of \( f \). A level set of a smooth function \( f: M \to \mathbb{R} \) is called regular if it contains no critical points of \( f \), i.e. zeros of \( df \). Our assumption that \( df \neq 0 \) almost everywhere, combined with Sard’s lemma, implies that the union of regular level sets is dense in \( M \). So, integrability guarantees that the dynamic is relatively simple everywhere except possibly a nowhere dense set.

**Theorem 7.2** (Arnold-Liouville theorem for maps in 2D). Let \( M \) be a 2-dimensional manifold, and let \( T: M \to M \) be an integrable area-preserving diffeomorphism with first integral \( f: M \to \mathbb{R} \). For \( a \in \mathbb{R} \), let \( M_a = f^{-1}(a) \) be the corresponding level set of \( f \). Assume that \( M_a \) is regular, compact, and connected. Then \( M_a \) is diffeomorphic to a circle, and the dynamics of \( T \) restricted to \( M_a \) is quasiperiodic. In other words, there is a periodic coordinate \( \phi: M_a \to \mathbb{R}/2\pi\mathbb{Z} \) such that in terms of \( \phi \) the map \( T|_{M_a} \) is a translation \( \phi \mapsto \phi + c \).

**Remark 7.3.** The standard version of the Arnold-Liouville theorem is for differential equations (vector fields) instead of maps. We will discuss that version later on in the course. The version for maps is due to A. Veselov [5].

**Remark 7.4.** The fact that \( M_a \) is a circle is trivial, since any compact connected 1-dimensional manifold is (diffeomorphic to) a circle. The reason we emphasize this statement is to make connections with the multidimensional version of this theorem where a circle is replaced by a torus. In that multidimensional setting, the statement becomes not trivial at all: tori are very far from being the only compact connected manifolds in any dimension \( d > 1 \).

**Remark 7.5.** Instead of requiring that \( M_a \) is connected, we can assume that \( T \) preserves the given connected component \( M^0_a \) of \( M_a \). Then the conclusion of the theorem stays the same, with \( M_a \) replaced by \( M^0_a \). Furthermore, even if \( M^0_a \) is not preserved by \( T \), there still exists \( k > 1 \) such that \( M^0_a \) is preserved by \( T^k \) (because \( M_a \) is compact and thus has only finitely many connected components). So, we can always apply the theorem for some iteration of \( T \).

Also note that the compactness assumption is not really essential in our 2D case either, but it does become important in the multidimensional setting. Furthermore, it simplifies things even in 2D, so we include it.

To prove Theorem 7.2 we introduce the notion of a Hamiltonian vector field (which is also very important on its own right). Let \( M \) be a 2D manifold equipped with an area form \( \omega \). Then, for any \( x \in M \), \( \omega \) is a non-degenerate skew-symmetric bilinear form on \( T_xM \) and thus can be viewed as an invertible linear map \( \omega: T_xM \to T_x^*M \). Therefore, there is a well-defined linear map \( \omega^{-1}: T_x^*M \to T_xM \).

**Definition 7.6.** Let \( M \) be a 2-dimensional manifold \( M \) equipped with an area form \( \omega \). Let also \( f: M \to \mathbb{R} \) be a smooth function. Then the associated Hamiltonian vector field \( X_f \) is defined by \( X_f = \omega^{-1}(df) \). The function \( f \) itself is called the Hamiltonian corresponding to the field \( X_f \).

---

not true in the non-compact case. For example, \((x, y) \mapsto (x + 1, y)\) is an area-preserving map of the right-half plane to itself (with respect to the standard area form \( dx \wedge dy \)), but it is not surjective.
Remark 7.7. A different notation for the Hamiltonian vector field associated with \( f \) is \( sgrad f \), the skew gradient of \( f \). Its definition is indeed similar to that of a gradient, with the metric replaced by the area form.

Example 7.8. Assume that \( \omega = dx \wedge dy \). Then the components of \( X_f \) are \( \frac{\partial f}{\partial y} \) and \( -\frac{\partial f}{\partial x} \).

Proposition 7.9. The derivative of \( f \) along the corresponding Hamiltonian vector field \( X_f \) is 0.

Proof. This derivative is equal to \( \langle df, X_f \rangle \), where \( \langle , \rangle \) is the pairing between covectors and vectors. But \( \langle df, X_f \rangle = \langle \omega(X_f), X_f \rangle = \omega(X_f, X_f) = 0 \), as desired.

Lecture 8: Proof of the Arnold-Liouville theorem in 2D and Poncelet’s closure theorem

We are now in a position to prove Theorem 7.2. First note that the regular level set \( M_a = f^{-1}(a) \) of \( f \) is automatically a circle simply because it is a compact connected 1-dimension manifold. So it suffices to construct a periodic coordinate \( \phi: M_a \to \mathbb{R}/2\pi\mathbb{Z} \) such that in terms of \( \phi \) the map \( T|_{M_a} \) is a translation \( \phi \mapsto \phi + c \). To that end, consider the Hamiltonian vector field \( X_f \) associated with \( f \). By Proposition 7.9, the vector field \( X_f \) is tangent to the circle \( M_a \) and, therefore, can be restricted to it. Also note that \( X_f \) does not vanish in \( M_a \) since \( X_f = -\omega^{-1}(df) \) vanishes only at those points where \( df = 0 \), and there is no such points in \( M_a \). So, \( X_f \) is a non-vanishing vector field on a circle. But then there must exist a periodic coordinate \( \psi: M_a \to \mathbb{R}/L\mathbb{Z} \) on \( M_a \) such that \( X_f|_{M_a} = \partial/\partial\psi \). Locally the existence of such a coordinate is a particular case of the general rectification theorem for vector fields, but in the case of the circle this is also true globally, and can be shown as follows. Let \( s \) be any periodic coordinate on the circle \( M_a \). Then \( X_f \), written in terms of \( s \), reads \( v(s)\partial/\partial s \). We want to find a new coordinate \( \psi \) such that

\[
\frac{\partial}{\partial \psi} = v(s)\frac{\partial}{\partial s}.
\]

The latter expression can be rewritten as

\[
v(s)\frac{\partial}{\partial s} = v(s)\frac{d\psi}{ds} \frac{\partial}{\partial \psi},
\]

so

\[
\frac{\partial}{\partial \psi} = v(s)\frac{d\psi}{ds} \frac{\partial}{\partial \psi},
\]

which means that \( \psi \) can be defined by

\[
\psi = \int \frac{ds}{v(s)}.
\]

The latter expression has a simple physical meaning. Define the distance traveled as the net change in the \( s \) coordinate. Then \( ds \) is an infinitesimal displacement, and \( ds/v(s) \) is the time it takes to travel the infinitesimal distance \( ds \) with speed \( v(s) \). So, the above integral can be interpreted as the total time it takes to get from a fixed point on the circle to a given one, when traveling with velocity prescribed by our vector field \( v(s)\partial/\partial s \). It is not a surprise that \( \psi \) has a meaning of time: since we want the velocity written in terms of the new coordinate to be equal to 1, the displacement and time should be the same.

The above formula for \( \psi \) also suggests a coordinate-free construction of that function. Notice that the form \( \alpha = ds/v(s) \) takes the value 1 on the vector \( X_f = v(s)\partial/\partial s \). So, this form can be defined in a
coordinate-free fashion as the unique 1-form $\alpha$ such that at every point of $M_a$ we have $\langle \alpha, X_f \rangle = 1$. In other words, at every point $x \in M_a$, $\{\alpha|_x\}$ is the basis of $T^*_xM_a$ dual to the basis $\{X_f|_x\}$ of $T_xM_a$. Then the function $\psi$ can be constructed as

$$\psi(x) := \int_{x_0}^x \alpha,$$

where $x_0 \in M_a$ is any fixed point. This integral is well-defined up to the integral of $\alpha$ over the whole circle, so it is indeed a periodic coordinate (one also needs to check that $d\psi \neq 0$, but this follows from $d\psi = \alpha$). Furthermore, it is immediate from this construction that $X_f = \partial/\partial \psi$, because $\langle d\psi, X_f \rangle = \langle \alpha, X_f \rangle = 1$.

We are now in the following situation. On the circle $M_a$, we have a constant vector field $X_f = \partial/\partial \psi$. This vector field is preserved by the map $T$, because $T$ preserves both the Hamiltonian $f$ and the area form, and $X_f$ is constructed out of these two ingredients. But the only maps that preserve a constant (non-zero) vector field on a circle are translations. The easiest way to see that is once again consider the 1-form $d\psi$ dual to the vector-field $X_f$. Since $X_f$ is preserved by $T$, so is $d\psi$. But this means that

$$dT^*\psi = T^*d\psi = d\psi \iff d(T^*\psi - \psi) = 0 \iff T^*\psi = \psi + c,$$

where $c$ is a constant. But this means that in terms of $\psi$ the map $T$ is $\psi \mapsto \psi + c$. This is almost what we need, the only problem is that $\psi$ is $L$-periodic instead of $2\pi$-periodic. This can be fixed by defining a new coordinate $\phi = 2\pi\psi/L$. In terms of this new coordinate, the map $T$ is still a translation, by $2\pi c/L$. Thus, the Arnold-Liouville theorem for maps in 2D is proved.

**Corollary 8.1.** Let $M$ be a 2-dimensional manifold, and let $T: M \to M$ be an integrable area-preserving diffeomorphism with first integral $f: M \to \mathbb{R}$. For $a \in \mathbb{R}$, let $M_a = f^{-1}(a)$ be the corresponding level set of $f$. Assume that $M_a$ is regular, compact, and connected. Assume also that there is a natural number $k \geq 1$ and a point $x \in M_a$ such that $T^k(x) = x$. Then $T^k(y) = y$ for all points $y \in M_a$. In other words, if $T|_{M_a}$ has a periodic point, then all points in $M_a$ are periodic with the same period.

**Proof.** By the Arnold-Liouville theorem, there is a $2\pi$-periodic coordinate $\phi$ on $M_a$ such that, in terms of this coordinate, the map $T|_{M_a}$ reads $\phi \mapsto \phi + c$. Then the map $T^k|_{M_a}$ reads $\phi \mapsto \phi + kc$, and since $T^k(x) = x$, we must have that $kc$ is an integer multiple of $2\pi$. But this means that $T^k|_{M_a}$ is the identity map, as desired. \qed

**Remark 8.2.** As can be seen from the proof of the Arnold-Liouville theorem, the coordinate $\phi$ in which $T|_{M_a}$ is a rotation can be defined by an explicit formula involving integration. Furthermore, in terms of $\phi$ the dynamics of $T$, i.e. its iterate $T^n$, takes the form $\phi \mapsto \phi + nc$, where $c$ can be computed as $\phi(T(x)) - \phi(x)$ for any $x \in M_a$. So, to compute $T^n(x)$, one needs to do the following. First, find the $\phi$-coordinates of $x$ and $T(x)$ by performing integration. Then compute $c$ and hence the $\phi$-coordinate of $T^n(x)$. Finally, one needs to invert the formula expressing $\phi$, in order to find $T^n(x)$ in original coordinates. So, an integrable system can be explicitly “solved” in terms of integration and computing inverse functions. A possibility to give a solution using just these operations is known as “solvability by quadratures”. Quadrature is a historical term which means the process of determining area, i.e. integration. Thus, it follows from (the proof of) the Arnold-Liouville theorem that integrable maps in 2D are solvable by quadratures. This result is in fact usually included as one of the statements of the theorem.

From Corollary 8.1 applied to elliptic billiards, we get the following geometric result, known as Poncelet’s closure theorem:

**Theorem 8.3.** Let $E_1, E_2$ be two ellipses. Assume that $E_2$ lies in the interior of the domain bounded by $E_1$. Assume also that there exists an $n$-gon inscribed in $E_1$ and circumscribed about $E_2$. Then there exists infinitely many such $n$-gons, and every point on $E_1$ is a vertex of such an $n$-gon (see Figure 20 illustrating the case $n = 3$).
Figure 20: Poncelet triangles inscribed in $E_1$ and circumscribed about $E_2$.

**Proof.** We will first prove the theorem under the assumption that $E_1$ and $E_2$ are confocal. In this case, an $n$-gon inscribed in $E_1$ and circumscribed about $E_2$ is a closed billiard trajectory in $E_1$. Indeed, any side of such an $n$-gon can be viewed as a segment of a billiard trajectory. It is tangent to $E_2$, so by Corollary 5.5 all segments of the corresponding trajectory are tangent to $E_2$. So both the given polygon and the billiard trajectory are inscribed in $E_1$ and circumscribed about $E_2$, and since they have a common segment, they must coincide.

Now, it suffices to prove that any other billiard trajectory in $E_1$ which is tangent to $E_2$ is closed with the same period. But this follows from Corollary 8.1 because by Proposition 5.4 trajectories tangent to the same confocal conic belong to the same level set of the Joachimsthal integral. Note also that even though this level set of the Joachimsthal integral is not connected, it lies in the blue region in Figure 19 which means that each connected component of the level set is preserved by the billiard map. Therefore, we can simply apply Corollary 8.1 with $M$ being the upper part of the blue region (it is also not too hard to modify this argument to work for the green region as well).

To prove the theorem for arbitrary, not necessarily confocal, ellipses $E_1, E_2$, one can use a projective transformation taking them to confocal ones. For details, see e.g. [3, Section 3].

**Lecture 9: Algebraic integrability of elliptic billiards I**

We will now discuss the concept of algebraic integrability. Algebraic integrability is a different approach to establishing quasiperiodic dynamics, based on complex geometry. The advantage of this approach, as compared to the one provided by the Arnold-Liouville theorem, is that it explains the geometric meaning of the coordinate $\phi$ in which an integrable map is a rotation. It turns out that in order to reveal this meaning, one needs to complexify the phase space as well as the level sets of the first integral. We will demonstrate this approach by showing how it works in the elliptic billiard example. Our exposition is a variation on the proof of Poncelet’s closure theorem by P. Griffiths and J. Harris [4].

Recall that the phase space $M$ of the billiard in an ellipse $E_1$ consists of pairs $(x, v)$, where $x \in E_1$ is a point on the ellipse, and $v \in T_x \mathbb{R}^2$ is an inward tangent vector at $x$. Clearly, for fixed $x$, the vector $v$ is uniquely determined by the non-oriented direction that it defines. Moreover, almost all directions are realized in this way, except for the direction of the tangent line at $x$. So, one can regard $v$ as an element of $\mathbb{RP}^1$ which is allowed to take any value except for one. For the sequel, it will be convenient to allow that value corresponding to the tangent line as well. This turns the phase space into the direct product $E_1 \times \mathbb{RP}^1$, i.e. a torus. In terms of the phase cylinder $M$, adding tangent directions corresponds to identifying the opposite boundaries of $M$. One can easily see that both the billiard map, and its first integral are well-defined on this toric version of the phase space (for vectors $v$ tangent to the ellipse, we set $J = 0$ and define the billiard map to be the identity).
The complexified version of the phase space can now be defined as the space of pairs \((x,v)\), where \(x\) belongs to the complex ellipse, and \(v \in \mathbb{CP}^1\) is a direction. A complex ellipse is defined in the complex plane \(\mathbb{C}^2\) by the same equation as its real counterpart, i.e.

\[
\frac{x_1^2}{a_1^2} + \frac{x_2^2}{a_2^2} = 1.
\]  

(19)

However, this set is not compact, which is inconvenient. In order to compactify it, we will consider its closure inside the complex projective plane \(\mathbb{CP}^2\). This closure is obtained by “homogenizing” the equation \((19)\), i.e. substituting \(x_i = z_i/\omega_0\), and multiplying by a suitable power of \(\omega_0\) to make the equation polynomial. This gives

\[
\frac{z_1^2}{a_1^2} + \frac{z_2^2}{a_2^2} = z_0^2.
\]

(20)

We will denote the set of points in \(\mathbb{CP}^2\) whose homogeneous coordinates satisfy this equation in the same way as the real ellipse, i.e. \(E_1\). It can be viewed as a natural complexification of the real ellipse. The complexified billiard phase space is thus \(E_1 \times \mathbb{CP}^1\). Furthermore, the complex conic \(E_1\) itself can in fact be identified with \(\mathbb{CP}^1\). In order to do that, one fixes any point \(x_0 \in E_1\) and assigns to every \(x \in E_1\) the direction of the line through \(x_0\) and \(x\) (where the direction assigned to \(x_0\) itself is the limiting position of the chord \(x_0x\) as \(x \rightarrow x_0\), i.e. the tangent line to \(E_1\) at \(x_0\)). This gives a one-to-one correspondence \(E_1 \simeq \mathbb{CP}^1\), which is moreover a biholomorphic map between complex manifolds. Thus, the complexified billiard phase space is complex diffeomorphic to \(\mathbb{CP}^1 \times \mathbb{CP}^1\), while from the real point of view it is the product of spheres \(S^2 \times S^2\).

We now complexify a regular level set \(M_a\) of the Joachimsthal integral \(J\). To that end, recall that by Proposition 5.4 such a level set can be described as the set of segments of billiard trajectories which are tangent to a given confocal conic \(E_2\). In the complex setting, we use this property as the definition of \(M_a\): \(M_a\) is the set of pairs \((x,v) \in E_1 \times \mathbb{CP}^1\) such that the (complex) line through \(x\) with direction \(v\) is tangent to a fixed complex conic \(E_2\) confocal with \(E_1\) (confocality in the complex projective setting can be defined by the homogenized version of the equation \((10)\)). This tangency condition is algebraic, so \(M_a\) is defined as a subset of \(\mathbb{CP}^1 \times \mathbb{CP}^1\) satisfying an algebraic equation. In any affine chart in \(\mathbb{CP}^1 \times \mathbb{CP}^1\), this condition can be written in the form \(f(z,\omega) = 0\), and for generic conics \(E_1, E_2\), the zero level set of \(f\) is regular. Therefore, \(M_a\) is a one-dimensional complex (and hence two-dimensional real) submanifold of \(\mathbb{CP}^1 \times \mathbb{CP}^1\). In the next lecture, we will investigate the topology of that manifold.

Lecture 10: Algebraic integrability of elliptic billiards II

In the previous lecture we defined the complexified level set \(M_a\) of the Joachimsthal integral as the set of pairs \((x,v) \in \mathbb{CP}^2 \times \mathbb{CP}^1\) such that \(x\) belongs to a fixed conic \(E_1 \subset \mathbb{CP}^2\), and the line through \(x\) in direction \(v\) is tangent to another fixed conic \(E_2\) confocal with \(E_1\). For generic conics \(E_1, E_2\), the set \(M_a\) is a smooth manifold of complex dimension one and real dimension two. Also note that \(M_a\) is compact as a closed subset of a compact manifold. Furthermore, we claim that \(M_a\) is connected. To prove that consider the projection \(\pi: M_a \rightarrow E_1 \simeq \mathbb{CP}^1\) given by \((x,v) \mapsto x\). The number of preimages of a point

3In the complex setting, there is no difference between ellipses and hyperbolas, because one can always change the sign of the second term in the left-hand side of \((19)\) by means of a change of variables \(x_2 \mapsto \sqrt{-1}x_2\). So, in what follows, instead of talking about complex ellipses and hyperbolas, we will just use the word conics.

4Note that equation \((20)\) describes ellipses, hyperbolas, and also parabolas. Indeed, a parabola is defined by the equation \(x_2 = x_1^2\), whose homogeneous form is \(z_0z_2 = z_1^2\). By making a change of variables \(z_0 \mapsto z_0 - z_2, z_2 \mapsto z_0 + z_2\), we can rewrite this equation in the form \((20)\). The same is actually true in the real projective setting: in \(\mathbb{RP}^2\) ellipses, hyperbolas, and parabolas are all projectively equivalent to each other.
Figure 21: The complexified level set of the first integral is a double covering of $\mathbb{CP}^1$ ramified at four points.

$x \in \mathcal{E}_1$ under this projection is the number of tangent lines to $\mathcal{E}_2$ through $x$. One can easily show that the number of tangent lines to a given conic $\mathcal{E} \subset \mathbb{CP}^2$ through a given point $x \in \mathbb{CP}^2$ is exactly two, unless $x \in \mathcal{E}$. In the latter case, there is only one tangent line, that is the tangent to $\mathcal{E}$ at $x$. Thus $\pi^{-1}(x)$ consists of two points if $x \notin \mathcal{E}_2$, and one point if $x \in \mathcal{E}_2$. But $x \in \mathcal{E}_2$ if and only if $x$ is an intersection point of $\mathcal{E}_1$, $\mathcal{E}_2$, and for generic conics in $\mathbb{CP}^2$ there are exactly four such intersection points. So, there exist four points $x_1, x_2, x_3, x_4 \in \mathcal{E}_1$ such that $|\pi^{-1}(x_i)| = 1$, while for points $x \neq x_i$ we have $|\pi^{-1}(x)| = 2$. Since the mapping $\pi: M_a \to \mathbb{CP}^1$ is 2-to-1 everywhere except for four points, one says that $M_a$ is a double covering of $\mathbb{CP}^1$ ramified at four points. The mapping $\pi$ is schematically shown in Figure 21 (of course, $M_a$ does not really have self-intersection points, because it is smooth; a better model for the behavior of the map $\pi$ near the preimages of the points $x_i$ is given by the map $z \mapsto z^2$ near the origin; one can indeed show that $\pi$ takes this form in an appropriate coordinate system near $\pi^{-1}(x_i)$). It is clear from this figure that $M_a$ is connected: any two points in $M_a$ can be connected by a path through one of the ramification points $\pi^{-1}(x_i)$.

Thus, $M_a$ is a compact connected manifold of real dimension two. Furthermore, $M_a$ is orientable. Indeed, $M_a$ is a complex manifold and thus has holomorphic transition functions. But any holomorphic function viewed as a map from $\mathbb{R}^2 \to \mathbb{R}^2$ has positive Jacobian.

Now recall that any compact connected orientable surface is diffeomorphic to a “sphere with $g$ handles”. The number $g$ is known as the genus of the surface. For example, a surface of genus zero is a sphere, a surface of genus one is a torus, a surface of genus two is a pretzel etc. The genus is closely related to the notion of Euler characteristic. The latter can be computed by considering an arbitrary triangulation of the surface. For any triangulation, let $v$ be the number of vertices, $e$ be the number of edges, and $f$ be the number of faces. Then the Euler characteristic is defined as $\chi = v - e + f$; this number does not depend on the triangulation. The Euler characteristic $\chi$ and the genus $g$ are related by the formula $\chi = 2 - 2g$.

We will now compute the Euler characteristic, and hence the genus, of the surface $M_a$ using that it is a double covering of $\mathbb{CP}^1$ ramified at four points. To that end, connect each of the points $x_1, x_2, x_3, x_4 \in \mathbb{CP}^1$ to each other by an edge. This can be done without self-intersection, and gives a triangulation of $\mathbb{CP}^1$ which combinatorially look like a tetrahedron. It has four vertices, six edges, and four faces. Note that $4 - 6 + 4 = 2$, which agrees with the fact that the Euler characteristic of $\mathbb{CP}^1 \simeq S^2$ is equal to 2. We now lift this triangulation to $M_a$ by considering the preimages of vertices, edges, and faces under the mapping $\pi$. This gives a triangulation of $M_a$. The number of vertices of that triangulation is still four, because each of the points $x_i$ has exactly one preimage under $\pi$. However, the number of edges and faces doubles,
because points \( x \neq x_i \) have two preimages. So, we obtain a triangulation of \( M_a \) with four vertices, twelve edges, and eight faces. Thus, the Euler characteristic of \( M_a \) is \( \chi = 4 - 12 + 8 = 0 \), which means that \( g = 0 \), i.e. \( M_a \) is a torus.

We are now almost ready to describe the behavior of the complexified billiard map restricted to \( M_a \). As a last preparatory step, we will need a description of holomorphic maps from a complex torus to itself. This description is based on the following fundamental result in complex geometry:

**Theorem 10.1** (Uniformization theorem in genus one). Let \( X \) be a Riemann surface of genus one, i.e. a complex one-dimensional manifold which is real diffeomorphic to a torus. Then \( X \) is complex diffeomorphic to a surface of the form \( \mathbb{C}/L \), where \( L \subset \mathbb{C} \) is a full rank lattice, i.e. the set of integral linear combinations of two complex numbers which are linearly independent over \( \mathbb{R} \).

We can now prove that there are very few holomorphic maps from a Riemann surface of genus one to itself (in fact, for higher genera there are even less):

**Proposition 10.2.** Assume that \( \phi : \mathbb{C}/L \to \mathbb{C}/L \) is a holomorphic map from a Riemann surface of genus one to itself. Then \( \phi \) is linear, i.e. \( \phi(z) = \alpha z + \beta \).

**Proof.** The mapping \( \phi \) can be viewed as a multivalued holomorphic function on \( \mathbb{C}/L \), which is defined up to addition of complex numbers \( z \in L \). Therefore, the derivative of \( \phi \) is a genuine single-valued holomorphic function. But by the maximum principle there can be no non-constant holomorphic functions on a compact complex manifold. Thus \( \phi' \) is constant, and \( \phi \) is linear, as desired.

**Remark 10.3.** Not every function of the form \( \alpha z + \beta \) defines a holomorphic map \( \mathbb{C}/L \to \mathbb{C}/L \), because multiplication by \( \alpha \) may not commute with the action of \( L \) on \( \mathbb{C} \) and hence may not descend to the quotient \( \mathbb{C}/L \). In fact one can show that for generic lattices \( L \) the only bijective holomorphic maps \( \mathbb{C}/L \to \mathbb{C}/L \) are of the form \( z \mapsto \pm z + \beta \).

We could in principle use this remark to prove that the restriction of the complexified billiard map to \( M_a \) is a translation \( z \mapsto z + \beta \), at least if the corresponding lattice \( L \) is generic. Indeed, the billiard map cannot be of the form \( z \mapsto -z + \beta \), because the latter map is an involution, i.e. its square is the identity. However, the square of the billiard map is not identity for generic \( M_a \), which means that the billiard map is of the form \( z \mapsto z + \beta \) for almost all, and hence for all (by continuity), non-singular level sets \( M_a \). Furthermore, this argument can be modified to work even for non-generic lattices, because even for such lattices any automorphism of \( \mathbb{C}/L \) which is not a translation has finite order. However, we will not pursue this approach, because in our setting it is easier to use the explicit description of the billiard map to show that it is a translation.

**Proposition 10.4.** The billiard map \( T \) for the ellipse, restricted to the complexified level set \( M_a \) of the first integral, is a translation relative to the group structure on \( M_a \) coming from the identification \( M_a \cong \mathbb{C}/L \).

**Remark 10.5.** The group structure on \( M_a \) is not completely unique, but it can be made unique by specifying a point \( O \in M_a \) which is identified with the origin in \( \mathbb{C}/L \). Indeed, if \( f,g \) are two different biholomorphic maps from \( M_a \) to \( \mathbb{C}/L \) which take \( O \) to 0, then \( f \circ g^{-1} \) is a biholomorphic automorphism of \( \mathbb{C}/L \) which preserves 0. Therefore, from the above classification of holomorphic maps of \( \mathbb{C}/L \) to itself it follows that \( f \circ g^{-1} \) is of the form \( z \mapsto a z \). But any such map is a homomorphism of \( \mathbb{C}/L \) to itself, which guarantees that the pull-backs of the group structure from \( \mathbb{C}/L \) to \( M_a \) by means of \( f \) and \( g \) coincide.\(^5\)

Thus, any (compact and connected) Riemann surface of genus 1 with a distinguished point has a canonical

\(^5\)We define a pull-back of the group structure by \( f : M_a \to \mathbb{C}/L \) as the unique group structure on \( M_a \) such that \( f \) is an isomorphism relative to that structure. Explicitly, the addition in this group structure is defined by \( x + y = f^{-1}(f(x) + f(y)) \).
group structure. A pair consisting of a Riemann surface of genus 1 and a distinguished point on it is called an elliptic curve. Sometimes this term is also used as a synonym for Riemann surface of genus 1 (without a distinguished point). Also note that the notion of a translation does not depend on the choice of the distinguished point (similarly to how the notion of translation in the affine space does not depend on the choice of the origin).

**Proof of Proposition 10.4.** We can think of $M_a$ as the space of pairs $(x,l)$, where $x$ is a point in the conic $E_1$, and $l$ is a line through $x$ tangent to $E_2$. The billiard map $T: M_a \to M_a$ takes the pair $(x,l)$ to the pair $(x',l')$, where $x'$ is the second intersection point of $l$ with $E_1$ (along with $x$), and $l'$ is the second tangent line from $x'$ to $E_2$ (along with $l$). Therefore, the billiard map can be written as a composition $\sigma_2 \circ \sigma_1$, where the maps $\sigma_1, \sigma_2: M_a \to M_a$ are defined by $\sigma_1(x,l) = (x',l)$, $\sigma_2(x',l) = (x',l')$. In other words, $\sigma_1$ interchanges the endpoints of a given tangent line to $E_2$, while $\sigma_2$ swaps the tangent lines to $E_2$ passing through a fixed point. In particular, both maps $\sigma_1$ and $\sigma_2$ are involutions, i.e. $\sigma_1^2 = \sigma_2^2$ is the identity map. Also note that both $\sigma_1$ and $\sigma_2$ can be computed by solving a quadratic equation. For $\sigma_1$, this is the equation determining the intersection points of a line and a conic. For $\sigma_2$, this is the equation determining the tangent lines to a conic from a given point. In both cases, one of the roots of the corresponding quadratic equation is known, so the second root can be found by the Vieta theorem. This guarantees that the maps $\sigma_1, \sigma_2: M_a \to M_a$ are holomorphic. So, the billiard map $T = \sigma_2 \circ \sigma_1$ is holomorphic as well. Thus, we can apply our classification of holomorphic maps $\mathbb{C}/L \to \mathbb{C}/L$ to $T$. It is, however, easier to first use that classification to describe the maps $\sigma_1, \sigma_2$. Assume that $\sigma: \mathbb{C}/L \to \mathbb{C}/L$ is a holomorphic involution. Then $\sigma(z) = az + b$, and since it is an involution, we must have $a = \pm 1$. Furthermore, if $a = +1$, then $\sigma(z) = az + b = z + b$ is an involution if and only if $2b \in L$. To distinguish between involutions $\sigma(z) = -z + b$ and $\sigma(z) = z + b$, where $2b \in L$, one can count the fixed points of $\sigma$. In the case $\sigma(z) = z + b$ case there is no fixed points unless $b \in L$, in which case $\sigma$ is the identity. In the case $\sigma(z) = -z + b$ there is at least one fixed point, $\frac{1}{2}b$ (in fact, one can show that there are four fixed points). So, any non-trivial involution $\sigma: \mathbb{C}/L \to \mathbb{C}/L$ with fixed points must be of the form $z \mapsto b - z$. Now observe that the involutions $\sigma_{1,2}$ have fixed points. For $\sigma_1$, the fixed points are of the form $(x,l)$, where $x \in E_1 \cap E_2$, and $l$ is the tangent line to $E_2$ at $x$. For $\sigma_1$, the fixed points are of the form $(x,l)$, where $l$ has a property of being tangent to both $E_1$ and $E_2$. Thus, we have $\sigma_1(z) = b_1 - z$, $\sigma_2(z) = b_2 - z$. So, $T(z) = \sigma_2(\sigma_1(z)) = z + b_2 - b_1$, as desired. □

**Remark 10.6.** As we know from Lecture 6, the real part of $M_a$ consists of two circles. We thus have two closed curves on the torus $M_a$. According to Lecture 6, the billiard map may either preserve each of these circles (if we are in the blue region) or interchange them (if we are in the green region). This situation can be understood by looking at the following model example (in fact, one can prove that this is exactly what is happening). Consider the lattice $L$ spanned by 1 and $\tau \sqrt{-1}$, where $\tau \neq 0$ is real. Then this lattice is invariant under complex conjugation. Therefore, one has a well-defined complex conjugation operation on $\mathbb{C}/L$ (in this case one also says that the complex curve $\mathbb{C}/L$ is endowed with a real structure). The set of real points in $\mathbb{C}/L$ consists of points that are invariant under complex conjugation. This set is a union of two circles, one of which is the image of the real axis in $\mathbb{C}$, and the second one consists of numbers whose imaginary part is $\frac{1}{2}\tau \sqrt{-1}$. Then we have two kinds of translations in $\mathbb{C}/L$ which preserve the set of the real points: translation by a real number, and translation by a number whose imaginary part is $\frac{1}{2}\tau \sqrt{-1}$. Translations of the first type clearly preserve the components of the real part of $\mathbb{C}/L$, while translations of the second type interchange them.

**Remark 10.7.** Both the Arnold-Liouville theorem, and the complex geometric approach provide a coordinate $\phi$ defined modulo $2\pi$ and such that the billiard map written in this coordinate is a translation. We claim that the coordinates coming from these two approaches coincide. Indeed, assume that $\phi, \psi$ are two such coordinates. Then the billiard map must preserve two non-vanishing 1-forms $d\phi$ and $d\psi$. The ratio
of these 1-forms is then a well-defined smooth function on the circle invariant under the billiard map. But if the billiard map is a rotation by an angle which is not a rational multiple of $2\pi$, then its orbits are dense and it does not admit non-constant invariant functions. So $d\psi$ should be a constant multiple of $d\phi$, and since have integral $2\pi$ over the whole circle, they must coincide. Therefore, $\phi$ and $\psi$ are the same up to an additive constant. Note this is also true for level sets on which the rotation angle is a rational multiple of $2\pi$, by continuity.

**Exercises.**

1. Explain why any line and any conic in the complex projective plane either intersect at two distinct points, or are tangent to each other.

2. This exercise suggests a useful method or proving results in projective geometry, based on projective duality. Recall that the dual projective plane is defined as the set of straight lines in the given projective plane. More concretely, every line $l$ is defined in homogeneous coordinates by an equation of the form $a_0 z_0 + a_1 z_1 + a_2 z_2 = 0$, and $a_0, a_1, a_2$ are, by definition, the homogeneous coordinates of $l$ in the dual projective plane. Show that straight lines passing through a given point form a straight line in the dual projective plane, and all lines in the dual plane are of this form (thus points of the dual plane can be identified with lines in the initial one, and vice versa). Then show that the set of tangents to a given conic forms a conic in the dual projective plane (the dual conic). Using this and the result of the previous exercise, prove that there always exist two distinct tangent lines to a given conic through a given point, provided that the point does not lie on the conic.

3. Consider the level set of the Joachimsthal integral that corresponds to trajectories through the foci of the ellipse (this set is the boundary between two different colors in Figure 19). Show that the complexification of that level set is homeomorphic to two spheres attached to each other at two points.

4. Prove that each of the four pieces obtained by removing critical points from that level set admits a coordinate relative to which the billiard map $T$ is a translation. Show that for any $x$ in any of these four pieces the sequence $T^n(x)$ converges to a critical point. What geometric behaviour of the billiard trajectories does this correspond to?

5. For the billiard in a circle $C$ define the complexified level set of the first integral as the set of pairs $(x, l)$, where $x$ is a point in the circle $C$ and $l$ is a line through $x$ tangent to a fixed circle concentric with $C$. Prove that in this case each such level set is homeomorphic to two spheres attached to each other at two points.

6. Let $p(x)$ be a cubic polynomial without multiple roots. Prove that the closure of the complex curve $y^2 = p(x)$ in $\mathbb{CP}^2$ is diffeomorphic to a torus.

**Lecture 11: Canonical Hamilton’s equations**

In this lecture we discuss Hamilton’s canonical formalism, which can be viewed as an ancestor of modern symplectic geometry. Let $q = (q_1, \ldots, q_n)$ be a point of mass $m$ in $\mathbb{R}^n$ moving in the potential force field with potential $U(q) = U(q_1, \ldots, q_n)$. Then, by definition of the potential, the force field is given by $F = -\nabla U$, and Newton’s second law gives

$$m\ddot{q}_i = -\frac{\partial U}{\partial q_i}.$$
where prime stands for the time derivative. We want to rewrite these equations as first order equations. To that end, introduce the momenta $p_i = m q_i'$. Then we get

$$
\begin{align*}
  p_i' &= -\frac{\partial U}{\partial q_i}, \\
  q_i' &= \frac{1}{m} p_i.
\end{align*}
$$

This can be further rewritten by introducing the function

$$
H = U + \frac{m}{2} \sum_{i=1}^{n} (q_i')^2 = U + \frac{1}{2m} \sum_{i=1}^{n} (p_i)^2,
$$

which has the physical meaning of total energy and is also known as the Hamiltonian. In terms of this function, the above system can be rewritten as

$$
\begin{align*}
  p_i' &= -\frac{\partial H}{\partial q_i}, \\
  q_i' &= \frac{\partial H}{\partial p_i}.
\end{align*}
$$

These equations are known as canonical Hamilton's equations. In vector form, these equations can be rewritten as

$$
x' = \pi dH(x),
$$

where $x$ is the column vector $(p, q)^t = (p_1, \ldots, p_n, q_1, \ldots, q_n)^t$, and $\pi$ is the matrix

$$
\pi = \begin{pmatrix}
  0 & -\text{Id} \\
  \text{Id} & 0
\end{pmatrix}.
$$

Note that $\pi$ takes a covector $dH(x)$ to a vector $x'$. So, for every $x \in \mathbb{R}^n$, the matrix $\pi$ should be thought of as a map from the cotangent space at $x$ to the tangent space at $x$. We can, therefore, also regard $\pi$ as a bilinear form on cotangent vectors. Explicitly, this form can be written as

$$
\pi = -\sum_{i=1}^{n} \frac{\partial}{\partial p_i} \wedge \frac{\partial}{\partial q_i}.
$$

This form is skew-symmetric. Bilinear skew-symmetric forms on the dual space are also known as bivectors (we can also say that $\pi$ is a bivector field, because it is a bivector defined on every tangent space). This particular bivector $\pi$ entering Hamilton’s equation is called the Poisson tensor.

One can also rewrite Hamilton’s equation in terms of the symplectic structure. The latter is defined as $\omega = \pi^{-1}$. It is a map from the tangent space to the cotangent space, and thus can be viewed as a 2-form. By inverting the matrix of $\pi$ we get

$$
\omega = \begin{pmatrix}
  0 & \text{Id} \\
  -\text{Id} & 0
\end{pmatrix},
$$

so

$$
\omega = \sum_{i=1}^{n} dp_i \wedge dq_i.
$$

Note, in particular, that this form is closed. This property is very important and is included as one of the requirements in the abstract definition of a symplectic structure: a symplectic structure on a manifold is a closed non-degenerate skew-symmetric 2-form. We will be discussing the notion of an abstract symplectic structure, as well as of an abstract Poisson structure, in the next few lectures. We will also return to the above example of motion in the potential force field and see that this corresponds to a canonical symplectic structure on the cotangent bundle.
Lecture 12: Symplectic vector spaces and symplectic manifolds

Definition 12.1. A symplectic vector space is a vector space $V$ endowed with a symplectic form $\omega$, i.e. a non-degenerate skew-symmetric bilinear form.

Example 12.2. There exist no symplectic structures on one-dimensional vector spaces. Indeed, let $V$ be a one-dimensional vector space, and let $v$ be a basis vector in $V$. Then by skew-symmetry we have $\omega(v, v) = 0$, so $\omega$ is identically zero and hence degenerate.

Example 12.3. Any two-dimensional vector space $V$ admits a symplectic structure. Such a structure is uniquely defined by $\omega(v_1, v_2) = 1$, where $v_1, v_2$ is any fixed basis in $V$. By skew-symmetry we also have $\omega(v_2, v_1) = -1, \omega(v_1, v_1) = 1, \omega(v_2, v_2) = 0$, so the matrix of $\omega$ in the basis $v_1, v_2$ is

$$\Omega = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$ 

More generally, one has the following:

Proposition 12.4. A finite-dimensional vector space $V$ admits a symplectic structure if and only if $V$ is even-dimensional.

Proof. Let $V$ be a symplectic vector space, and let $\Omega$ be the matrix of the symplectic form written in some basis. Then $\Omega$ is skew-symmetric, so

$$\det \Omega = \det(-\Omega^t) = (-1)^{\dim V} \det \Omega^t = (-1)^{\dim V} \det \Omega.$$ 

From the non-degeneracy of the symplectic structure we have that $\det \Omega \neq 0$, so it follows that $(-1)^{\dim V} = 1$, which means that $V$ is even-dimensional. Conversely, if $V$ is even-dimensional, then we can define a symplectic structure $\omega$ by taking a basis and considering a bilinear form which in this basis is given by the following block-diagonal matrix:

$$\Omega = \begin{pmatrix} 0 & 1 & & & \\ -1 & 0 & & & \\ & 0 & 1 & & \\ & -1 & 0 & & \\ & & & & \ddots \end{pmatrix}.$$ 

(21)

It turns out that the above example of a symplectic form on an even-dimensional vector space is in fact universal in the following sense:

Theorem 12.5 (Linear Darboux theorem). In any finite-dimensional symplectic vector space there exists a basis in which the matrix of the symplectic form is (21).

Remark 12.6. By rearranging basis vectors $v_1, \ldots, v_{2n}$ as $v_1, v_3, \ldots, v_{2n-1}, v_2, v_4, \ldots, v_{2n}$ we can also rewrite the above matrix as

$$\begin{pmatrix} 0 & \text{Id} \\ -\text{Id} & 0 \end{pmatrix},$$

(22)

which is an already familiar to us symplectic structure of canonical Hamilton’s equations.
This theorem can be proved by induction, based on the notion of a symplectic orthogonal complement, or, which is the same, skew-orthogonal complement. If \( W \) is a subspace of a symplectic vector space \((V, \omega)\), then its orthogonal complement is defined by

\[ W^\perp = \{ v \in V \mid \omega(v, w) = 0 \forall w \in W \}. \]

In contrast to the orthogonal complement in an inner product space, the symplectic orthogonal complement may not be an actual complement, i.e. in general it is not true that \( V = W \oplus W^\perp \). The reason for that is that \( W \) and \( W^\perp \) may have non-trivial intersection. For example, any vector is symplectic orthogonal to itself, so any one-dimensional subspace is contained in its own symplectic orthogonal complement. However, even though \( W \) and \( W^\perp \) are in general not complementary to each other, their dimensions always are:

**Lemma 12.7.** For any subspace \( W \) of a symplectic vector space \( V \), we have \( \dim W^\perp = \dim V - \dim W \).

**Proof.** Consider the symplectic structure \( \omega \) as a linear map \( V \to V^* \). Then \( \omega(W) \) consists of those covectors which annihilate \( W \). So, by linear algebra we have \( \dim \omega(W) = \dim V - \dim W \). But \( \omega \) is non-degenerate, so \( \dim \omega(W) = \dim W \), which proves the desired equality. \( \square \)

**Exercise 12.8.** Deduce the linear Darboux theorem from this lemma.

The following subspaces in a symplectic vector space are of particular interest:

**Definition 12.9.** A subspace \( W \) of a symplectic vector space \((V, \omega)\) is called isotropic if the restriction of \( \omega \) to \( W \) is zero, i.e. if \( \omega(w_1, w_2) = 0 \) for any \( w_1, w_2 \in W \).

**Example 12.10.** Any one-dimensional subspace is isotropic.

**Example 12.11.** Assume that the symplectic form is given, in some basis \( v_1, \ldots, v_{2n} \), by the matrix \((22)\). Then the subspace spanned by any subset of \( v_1, \ldots, v_n \) is isotropic, and so is the subspace spanned by any subset of \( v_{n+1}, \ldots, v_{2n} \).

**Proposition 12.12.** The dimension of an isotropic subspace of a symplectic vector space \( V \) may not exceed \( \frac{1}{2} \dim V \).

**Proof.** A subspace \( W \) is isotropic if and only if it is contained in its orthogonal complement \( W^\perp \). So, if \( W \) is isotropic, then

\[ \dim W \leq \dim W^\perp = \dim V - \dim W, \]

which implies

\[ 2 \dim W \leq \dim V, \]

q.e.d. \( \square \)

**Definition 12.13.** An isotropic subspace \( W \) of a symplectic vector space \( V \) is called Lagrangian if it has maximal possible dimension \( \frac{1}{2} \dim V \).

Lagrangian subspaces are also maximal isotropic in the sense that any isotropic subspace can be extended to a Lagrangian subspace. Indeed, if we have a non-Lagrangian isotropic subspace \( W \), then by dimension reasons \( W \) is strictly contained in \( W^\perp \), which implies that there exist vectors not in \( W \) which are orthogonal to all of \( W \). Adding any such vector to \( W \) increases its dimension, and by doing so several times we will finally make \( W \) Lagrangian.

**Example 12.14.** Any one-dimensional subspace of a two-dimensional symplectic space is Lagrangian.
Example 12.15. Assume that the symplectic form is given, in some basis \( v_1, \ldots, v_{2n} \) by the matrix (22). Then the subspace spanned by \( v_1, \ldots, v_n \) is Lagrangian, and so is the subspace spanned by \( v_{n+1}, \ldots, v_{2n} \).

We now turn our attention to symplectic manifolds.

Definition 12.16. A symplectic manifold is a manifold \( M \) endowed with a symplectic form (or symplectic structure) \( \omega \), i.e. a closed non-degenerate differential 2-form.

Since the tangent space of a symplectic manifold at every point is a symplectic vector space, it follows that there exist no odd-dimensional symplectic manifolds. So, the simplest example of a symplectic manifold is a two-dimensional one:

Proposition 12.17. Any orientable two-dimensional manifold admits a symplectic structure.

Proof. The closedness condition of a 2-form on a 2-manifold is vacuous, so a symplectic form in 2D is the same as an area form, i.e. a non-vanishing differential 2-form. Such a form can be constructed, for example, using a partition of unity subordinate to any oriented atlas (i.e. an atlas whose transition maps are orientation-preserving), or using a Riemannian metric. □

Clearly, orientability condition cannot be omitted, because an area form defines an orientation. So, there is no symplectic structures on non-orientable 2D manifolds. More generally, we have the following:

Proposition 12.18. Any symplectic manifold is orientable.

Proof. Let \( M \) be a \( 2n \)-dimensional manifold with symplectic structure \( \omega \). Then from non-degeneracy of \( \omega \) it follows that the \( 2n \)-form

\[
\wedge^n \omega = \underbrace{\omega \wedge \cdots \wedge \omega}_{n \text{ times}}
\]

is non-vanishing and hence a volume form. Therefore, \( M \) is orientable. □

One of the most important examples of a symplectic manifold is the cotangent bundle:

Proposition 12.19. The cotangent bundle of any manifold \( M \) has a canonical symplectic structure.

Proof. We will first construct a 1-form \( \lambda \) on \( T^*M \), and then define a symplectic structure \( \omega \) by \( \omega = d\lambda \). To define a 1-form we should describe its value on any tangent vector \( v \). Assume that \( v \in T_{(q,p)}T^*M \), where \( (q,p) \) is a point in \( T^*M \), i.e. \( q \in M \) and \( p \in T^*_qM \). To define \( \lambda(v) \), consider the projection \( \pi: T^*M \to M \) given by \( (q,p) \mapsto q \). Then \( \lambda(v) \) is defined by

\[
\lambda(v) = p(d\pi(v)).
\]

This is known as the tautological 1-form on the cotangent bundle, or Liouville 1-form. Let us compute this form in coordinates. Take any local chart \( (q_1, \ldots, q_n) \) on \( M \). Then this chart also defines coordinates on every cotangent space (which is within this chart). Those coordinates \( (p_1, \ldots, p_n) \) are, by definition, coordinates of a covector in the basis \( dq_1, \ldots, dq_n \). Then \( (p_1, \ldots, p_n, q_1, \ldots, q_n) \) form a local coordinate system on \( T^*M \). In these coordinates, the projection \( \pi \) is given by

\[
(p_1, \ldots, p_n, q_1, \ldots, q_n) \mapsto (q_1, \ldots, q_n),
\]

so the differential of the projection maps \( \partial/\partial p_i \) to 0 and \( \partial/\partial q_i \) to \( \partial/\partial q_i \). We now take any \( v \in T_{(q,p)}T^*M \). This vector can be written as

\[
v = \sum_{i=1}^n dp_i(v) \frac{\partial}{\partial p_i} + \sum_{i=1}^n dq_i(v) \frac{\partial}{\partial q_i}.
\]
Therefore,
\[ d\pi(v) = \sum_{i=1}^{n} dq_i(v) \frac{\partial}{\partial q_i}, \]
and
\[ \lambda(v) = p(d\pi(v)) = p \left( \sum_{i=1}^{n} dq_i(v) \frac{\partial}{\partial q_i} \right) = p_i(v) dq_i(v), \]
so in \((p,q)\) coordinates the form \(\lambda\) reads
\[ \lambda = \sum_{i=1}^{n} p_i dq_i. \]

We now define the form \(\omega\) by
\[ \omega = d\lambda = \sum_{i=1}^{n} dp_i \wedge dq_i. \]

This 2-form is automatically closed (since it is exact) and from the coordinate representation it is clear that \(\omega\) is non-degenerate. Therefore, \(\omega\) is indeed a symplectic structure. It is canonical in a sense that its construction can be performed without coordinates and does not require any additional structures on \(M\).

\[ \square \]

**Corollary 12.20.** The cotangent bundle of any manifold has a canonical volume form and is orientable.

**Proof.** The volume form is defined as \(\wedge^n \omega\).

\[ \square \]

**Lecture 13: Vector fields and flows I**

Let \(M\) be a manifold, and let \(v\) be a vector field on \(M\).

**Definition 13.1.** A parametrized smooth curve \(\gamma(t)\) is called an integral trajectory of \(v\) if
\[ \frac{d\gamma}{dt} = v(\gamma(t)) \]
for any \(t\) for which \(\gamma(t)\) is defined.

In other words, at any point of an integral trajectory its velocity vector coincides with the value of the vector field \(v\) at that point.

**Proposition 13.2.** For any \(x \in M\) there exists a unique integral trajectory \(\gamma(t)\) of \(v\) defined for \(t \in (-\varepsilon, \varepsilon)\), where \(\varepsilon > 0\), such that \(\gamma(0) = x\).

**Proof.** Let \(x_1, \ldots, x_n\) be local coordinates near the point \(x\). Then the sought integral trajectory can be written in these coordinates as a vector function \(x_1(t), \ldots, x_n(t)\), and the condition on that function to define an integral trajectory of \(v\) is equivalent to the system of ODEs
\[ \dot{x}_i = v_i(x_1, \ldots, x_n), \]
where \(v_i\)'s are components of the vector field \(v\) in coordinates \(x_1, \ldots, x_n\). Furthermore, the requirement \(\gamma(0) = x\) specifies an initial condition for this system of ODEs, so the desired statement follows from the existence and uniqueness theorem for ODEs.

\[ \square \]
As can be seen from this proof, vector fields can be thought of as systems of ODEs on manifolds. “Solving” a vector field means finding its integral trajectories. In local coordinates this is equivalent to solving actual systems of ODEs.

**Definition 13.3.** A vector field \( v \) is called complete if any its integral trajectory \( \gamma(t) \) is defined for any \( t \in \mathbb{R} \).

As we know from the theory of ODEs, not every vector field is complete. For example, the vector field defined by the differential equation \( \dot{x} = x^2 \) (that is \( v = x^2 \partial/\partial x \)) on \( \mathbb{R} \) is not complete. Indeed, its integral trajectories are given by \( x(t) = 1/(c - t) \) and hence are not defined for every value of \( t \).

**Theorem 13.4.** Any vector field on a compact manifold is complete.

The proof is quite straightforward and can be found in most differential geometry textbooks.

**Definition 13.5.** Let \( v \) be a complete vector field on \( M \). The flow of \( v \) is a family of transformations \( \phi_t : M \to M \) defined as follows. Let \( \gamma(t) \) be an integral trajectory of \( v \) with \( \gamma(0) = x \). Then \( \phi_t(x) = \gamma(t) \).

One can also define the flow of an incomplete vector field, but that is, in general, only defined locally (i.e. in a neighbourhood of a given point \( x \in M \)) and only for \( t \in (-\varepsilon, \varepsilon) \). The definition of the flow \( \phi_t \) can also be expressed by saying that

\[
\frac{d}{dt}\phi_t(x) = v(\phi_t(x)).
\]

In order to be equivalent to the above definition, this equation needs to be supplemented by the condition \( \phi_0 = id \). This equation is also sometimes written as

\[
\frac{d}{dt}\phi_t = v \circ \phi_t.
\]

**Proposition 13.6.** The flow \( \phi_t \) is a diffeomorphism for every \( t \).

**Proof.** Smoothness of \( \phi_t \) follows from smooth dependence of solutions of ODEs on initial conditions. Furthermore, \( \phi_t^{-1} \) is also smooth since \( \phi_t^{-1} = \phi_{-t} \).

Along with the property \( \phi_t^{-1} = \phi_{-t} \), we also have that \( \phi_t \circ \phi_s = \phi_{s+t} \) and \( \phi_0 = id \). In other words, the diffeomorphisms \( \phi_t \) (considered for all values of \( t \)) form a group under composition. This group is isomorphic to \( \mathbb{R} \) with isomorphism given by \( t \mapsto \phi_t \). Due to these properties, the flows of vector fields are also called 1-parametric groups of diffeomorphisms.

We now want to discuss how flows of vector fields act on different geometric objects on \( M \), in particular on tensor fields. Let \( \xi \) be a tensor field. Then one can consider the pull-back \( \phi_t^* \xi \). It is again a tensor field, of the same type as \( \xi \). This in particular allows us to compute the derivative \( d/dt (\phi_t^* \xi) \). This derivative is defined point-wisely: for any \( x \) we evaluate the tensor field \( \phi_t^* \xi \) at that particular point \( x \). That gives a family of tensors at \( x \) depending on the parameter \( t \). This family can be differentiate with respect to \( t \), which produces a new tensor of the same type.

**Definition 13.7.** The Lie derivative of a tensor field \( \xi \) with respect to the vector field \( v \) is defined by

\[
L_v \xi = \frac{d}{dt} \bigg|_{t=0} (\phi_t^* \xi),
\]

where \( \phi_t \) is the flow of \( v \).

The Lie derivative \( L_v \xi \) is a tensor field of the same type as \( \xi \).
**Proposition 13.8.** The Lie derivative $L_v \xi$ vanishes if and only if $\xi$ is invariant under the flow, i.e.

$$
\phi_t^* \xi = \xi. \tag{23}
$$

**Proof.** Assume that $\xi$ is invariant under the flow. Then differentiating (23) with respect to $t$ at $t = 0$ gives $L_v \xi = 0$. Conversely, assume that $L_v \xi = 0$. To establish (23), we show that

$$
\frac{d}{dt} (\phi_t^* \xi) = 0.
$$

for every $t$. We have

$$
\frac{d}{dt} \bigg|_{t=0} (\phi_t^* \xi) = \frac{d}{dt} \bigg|_{t=0} (\phi_{t+0}^* \xi).
$$

Using the group property of the flow, this can be further rewritten as

$$
\frac{d}{dt} \bigg|_{t=0} (\phi_t^* \phi_t^* \xi) = \phi_{t+0}^* \left( \frac{d}{dt} \bigg|_{t=0} (\phi_t^* \xi) \right) = \phi_{t+0}^* (L_v \xi) = 0,
$$

as desired. \hfill \Box

**Lecture 14: Vector fields and flows II**

In the case when $\xi$ itself is a vector field $\xi = w$, the condition $L_v w = 0$ is also equivalent to the fact that the flows of $v$ and $w$ commute:

**Proposition 14.1.** Let $v, w$ be vector fields on a manifold $M$. Then $L_v w = 0$ if and only if the flow $\{\phi_t \mid t \in \mathbb{R}\}$ of $v$ commutes with the flow $\{\psi_t \mid t \in \mathbb{R}\}$ of $w$, which means that

$$
\phi_t \circ \psi_{t'} = \psi_{t'} \circ \phi_t \quad \forall t, t' \in \mathbb{R}.
$$

**Proof sketch.** As we know, $L_v w = 0$ is equivalent to the fact that $\phi_t^* w = w$ for any $t \in \mathbb{R}$. The latter in turn is equivalent to $\phi_t$ preserving integral trajectories of $w$, i.e. if $\gamma(t)$ is an integral trajectory of $w$, then so is $\phi_t(\gamma(t))$. Let us, for example, show that if $\phi_t^* w = w$ for any $t \in \mathbb{R}$ then $\phi_t$ preserves integral trajectories. Let $\gamma(t)$ be an integral trajectory of $w$. We want to show that $\phi_t(\gamma(t))$ is also an integral trajectory. To that end, observe that the derivative

$$
\frac{d}{dt} \phi_t(\gamma(t))
$$

is equal to the push-forward of the tangent vector $\gamma'(t) = w(\gamma(t))$ by the map $\phi_t$, which is the same as the pull-back by $\gamma_{-t}$. Also using that $\phi_t^* w = w$ for any $t$, we get

$$
\frac{d}{dt} \phi_t(\gamma(t)) = (\phi_t^* w)(\gamma(t)) = w(\gamma(t)),
$$

which means that $\phi_t(\gamma(t))$ is indeed an integral trajectory of $w$.

It now suffices to show that $\phi_t$ preserves integral trajectories of $w$ if and only if it commutes with the flow of $w$. This is almost a tautology. Indeed, equation $\phi_t \circ \psi_{t'} = \psi_{t'} \circ \phi_t$ means that

$$
\phi_t(\psi_{t'}(x)) = \psi_{t'}(\phi_t(x)) \quad \forall x \in M. \tag{24}
$$

But $\psi_{t'}(x)$ considered as a function of $t'$ (for fixed $x$) is an integral trajectory of $w$, and all integral trajectories of $w$ are, by definition, of this form. Similarly, $\psi_{t'}(\phi_t(x))$ is also an integral trajectory of $w$ (for fixed $x$ and $t$). So, if the flows commute, then $\phi_t$ maps integral trajectories of $w$ to integral trajectories of $w$. Conversely, if $\phi_t$ maps integral trajectories of $w$ to integral trajectories of $w$, then it must map a trajectory starting at $x$ to the trajectory starting at $\phi_t(x)$. But these trajectories are given by $\psi_{t'}(x)$ and $\psi_{t'}(\phi_t(x))$, which gives (24). \hfill \Box
We will now discuss how to compute the Lie derivative $L_v \xi$ for particular types of tensors. We start with functions:

**Proposition 14.2.** If $f$ is a smooth function, then $L_v f$ is equal to the directional derivative of $f$ along $v$:

$$L_v f = df(v). \quad (25)$$

**Proof.** We have

$$(L_v f)(x) = \frac{d}{dt} \bigg|_{t=0} (\phi_t^* f)(x) = \frac{d}{dt} \bigg|_{t=0} f(\phi_t(x)) = df \left( \frac{d}{dt} \bigg|_{t=0} \phi_t(x) \right) = df(v(x)), \quad \text{q.e.d.}$$

**Corollary 14.3.** For a function $f$ and vector field $v$, the following conditions are equivalent:

1. The directional (equivalently, Lie) derivative of $f$ along $v$ vanishes: $L_v f = df(v) = 0$.
2. The function $f$ is preserved by the flow of $v$.
3. The function $f$ is constant along integral trajectories of $v$.

**Proof.** The first two conditions are equivalent by Proposition 13.8. The last two conditions are equivalent because

$$(\phi_t^* f)(x) = f(\phi_t(x)),$$

so $f$ is preserved by the flow if and only if it is constant along curves of the form $\phi_t(x)$ (for fixed $x$). But such curves are, by definition, exactly the integral trajectories of $v$. \qed

Formula (26) can also be rewritten as

$$L_v f = i_v df, \quad (26)$$

where the operation $i_v$ is defined for an arbitrary $k$-form as follows: if $\xi$ is a $k$-form, then $i_v \xi$ is a $(k - 1)$-form given by

$$i_v \xi(w_1, \ldots, w_{k-1}) = \xi(v, w_1, \ldots, w_{k-1}).$$

For $k$-forms the formula for the Lie derivative is similar to (26) but has one more additional term:

**Proposition 14.4** (Cartan’s magic formula). Let $\xi$ be a differential form, and $v$ be a vector field. Then

$$L_v \xi = i_v d\xi + d i_v \xi.$$

**In other words, for differential forms we have**

$$L_v = i_v d + di_v.$$

**Proof sketch.** Using the fact that pull-backs commute with wedge products, one establishes the product rule for the Lie derivative: $L_v (\alpha \wedge \beta) = (L_v \alpha) \wedge \beta + \alpha \wedge (L_v \beta)$. The right hand side of Cartan’s formula can also be shown to obey this rule, which reduces the proof to the case of $\xi = dx_i$, where $x_i$ is a local coordinate. But since pull-backs also commute with the exterior derivative, it follows that $L_v d = d L_v$, and thus

$$L_v (dx_i) = d(L_v x_i) = d(i_v dx_i),$$

which coincides with Cartan’s formula in the case $\xi = dx_i$. \qed

34
We now establish a formula for the Lie derivative of a vector field. To that end, we will need the notion of the commutator (Lie bracket) of vector fields:

**Definition 14.5.** The Lie bracket of two vector fields $v, w$ is the unique vector field $[v, w]$ such that

$$L_{[v,w]}f = L_v L_w f - L_w L_v f$$

for any function $f$.

**Proposition 14.6.** Such a vector field $[v, w]$ exists and it is indeed unique.

*Proof.* If there were two such vector fields $u$ and $u'$ then we would have $L_u f = L_{u'} f$ for any $f$, i.e.

$$df(u - u') = 0.$$  

But any cotangent vector at any point is the differential of a suitable function, so all cotangent vectors vanish on $u - u'$, which means that $u - u' = 0$ and $u = u'$. This proves uniqueness. As for existence, it is sufficient to verify it locally, because then by the uniqueness part so-defined local vector fields can be patched into a global one. In local coordinates we have

$$L_v L_w f - L_w L_v f = \sum_{i,j} v_i \frac{\partial}{\partial x_i} \left( w_j \frac{\partial f}{\partial x_j} \right) - w_i \frac{\partial}{\partial x_i} \left( v_j \frac{\partial f}{\partial x_j} \right).$$

All terms containing second derivatives cancel out, and we end up with

$$L_v L_w f - L_w L_v f = \sum_{i,j} v_i \frac{\partial}{\partial x_i} w_j \frac{\partial f}{\partial x_j} - w_i \frac{\partial}{\partial x_i} v_j \frac{\partial f}{\partial x_j},$$

which is the derivative of $f$ along the vector field

$$\sum_{i,j} \left( v_i \frac{\partial}{\partial x_i} w_j - w_i \frac{\partial}{\partial x_i} v_j \right) \frac{\partial}{\partial x_j}.$$ 

So, the vector field $[v, w]$ indeed exists and it is given in local coordinates by the formula

$$[v, w] = \sum_{i,j} \left( v_i \frac{\partial}{\partial x_i} w_j - w_i \frac{\partial}{\partial x_i} v_j \right) \frac{\partial}{\partial x_j}. \quad \Box$$

**Example 14.7.** In $\mathbb{R}^2$ we have

$$\left[ \frac{\partial}{\partial x}, \frac{\partial}{\partial y} \right] = 0,$$

since

$$\frac{\partial}{\partial x} \frac{\partial}{\partial y} f = \frac{\partial}{\partial y} \frac{\partial}{\partial x} f$$

for any function $f$.

**Proposition 14.8.** For any vector fields $v, w$ we have

$$L_v w = [v, w].$$
Proof sketch. From the definition of the Lie derivative it easily follows that for any vector field \( v \) and any 1-form \( \alpha \) we have
\[
L_v(i_w\alpha) = i_{L_vw}\alpha + i_w(L_v\alpha),
\]
which can also be rewritten as
\[
i_{L_vw}\alpha = L_v(i_w\alpha) - i_w(L_v\alpha).
\]
Taking \( \alpha = df \) we get
\[
i_{L_vw}df = L_v(i_wdf) - i_w(L_vdf) = L_v(L_wf) - i_w(L_vdf).
\]
Using Cartan’s magic formula, the latter term can be rewritten as
\[
i_w(di_vdf) = L_wL_vdf,
\]
so
\[
i_{L_vw}df = L_v(L_wf) - L_w(L_vdf),
\]
which means that \( L_vw = [v,w] \), as desired.

Corollary 14.9. \([v,w] = 0 \) if and only the flows of \( v \) and \( w \) commute.

Lecture 15: Hamiltonian vector fields

We now discuss the notion of a Hamiltonian vector field on a symplectic manifold (cf. Lecture 7).

Definition 15.1. A vector field \( v \) on a symplectic manifold \((M,\omega)\) is a Hamiltonian vector field if there exists a smooth function \( H \) on \( M \) (called the Hamiltonian of the vector field \( v \)) such that
\[
i_v\omega = dH.
\]

 Explicitly, the Hamiltonian vector field with Hamiltonian \( H \) is given by \( \omega^{-1}dH \), where \( \omega \) is regarded as an operator from the tangent space to the cotangent space. The Hamiltonian vector field corresponding to the function \( H \) is denoted by \( X_H \) or \( \text{sgrad } H \). Thus, we have
\[
X_H = \text{sgrad } H = \omega^{-1}dH.
\]

Proposition 15.2. The flow of the Hamiltonian vector field \( X_H = \omega^{-1}dH \) preserves both the function \( H \) and the symplectic form \( \omega \).

Proof. We have
\[
L_{X_H}H = i_{X_H}dH = i_{X_H}i_{X_H}\omega = \omega(X_H,X_H) = 0,
\]
so \( H \) is indeed preserved. Further, by Cartan’s magic formula, we get
\[
L_{X_H}\omega = i_{X_H}d\omega + di_{X_H}\omega,
\]
but \( \omega \) is closed so this rewrites as
\[
di_{X_H}\omega = d(dH) = 0,
\]
which means that the symplectic form \( \omega \) is preserved as well.

Corollary 15.3. The flow of a Hamiltonian vector field consists of symplectic diffeomorphisms: \( \phi_t^*\omega = \omega \).
In fact, preservation of the symplectic form $\omega$ is almost equivalent to the Hamiltonian property. Namely, one has the following.

**Proposition 15.4.** Let $v$ be a vector field on a symplectic manifold $(M, \omega)$. Then $L_v \omega = 0$ if and only if $v$ is locally Hamiltonian, which means that it can be written as $X_H$ in a neighbourhood of any point in $M$. Furthermore, if $H^1(M, \mathbb{R}) = 0$, then $L_v \omega = 0$ if and only if $v$ is Hamiltonian.

**Proof.** We have $L_v \omega = d(i_v \omega)$, so $L_v \omega = 0$ if and only if the 1-form $i_v \omega$ is closed. But closed forms can be equivalently characterized as locally exact forms, so $L_v \omega = 0$ if and only if $i_v \omega = dH$ locally, which means that $v = X_H$. Furthermore, if $H^1(M, \mathbb{R}) = 0$, then closed 1-forms and exact 1-forms are the same, so in that case $L_v \omega = 0$ if and only if $v$ is Hamiltonian.

Vector fields $v$ such that $L_v \omega = 0$ are known as **symplectic**. They can be equivalently characterized as vector fields whose flows consist of symplectic diffeomorphisms. The above proposition shows that a vector field is symplectic if and only if it is locally Hamiltonian. It is easy to construct an example of a vector field which is symplectic but not globally Hamiltonian: take a closed non-exact 1-form $\alpha$ and define $v = \omega - \frac{1}{\omega} \alpha$. For example, one can take $\alpha$ to be the differential of the polar angle function in $\mathbb{R}^2$ without the origin.

From Proposition [15.2] we also get the following.

**Corollary 15.5** (Liouville’s theorem). Any Hamiltonian vector field preserves the volume form $\Omega = \^n \omega$.

**Proof.** Since the flow $\phi_t$ of a Hamiltonian vector field preserves $\omega$, it also preserves $\^n \omega$. Alternatively, one can use the product rule for the Lie derivative:

$$L_{X_H} (\^n \omega) = (L_{X_H} \omega) \wedge \omega \wedge \cdots \wedge \omega + \omega \wedge (L_{X_H} \omega) \wedge \omega \wedge \cdots \wedge \omega + \cdots = 0.$$ 

We now discuss examples of Hamiltonian vector fields.

**Example 15.6.** Let $(M, \omega)$ be a 2-dimensional symplectic manifold (equivalently, a surface with an area form). Let also $H$ be a function on $M$. Then the level sets of $H$ are 1-dimensional. The Hamiltonian vector field $X_H$ is tangent to those level sets (by Proposition [15.2]), which means that in the 2D case the trajectories of $X_H$ are almost the same as level sets of $H$. More precisely, we have the following:

**Proposition 15.7.**

1. Each connected component of each regular level set of $H$ is a trajectory of $X_H$.

2. For an arbitrary level set $H^{-1}(c) \subset M$, each critical point $x \in H^{-1}(c)$ is a 1-point trajectory (an equilibrium point) of $X_H$, and each connected component of the set of non-singular points in $H^{-1}(c)$ is a trajectory of $X_H$.

**Proof.** The first statement is a particular case of the second one, so we only prove the second statement. By definition, we have $X_H = \omega^{-1} dH$, so critical points of $H$ are indeed equilibrium points of $X_H$. Consider now the set $N$ of non-critical points in the given level set of $H$. At such points we have $X_H \neq 0$, so the set $N$ is a disjoint union of non-trivial (i.e. non-stationary) trajectories of $X_H$. Furthermore, since both $N$ and the trajectories are 1-dimensional manifolds, each trajectory is open $N$. Also, each trajectory is by definition connected. So, $N$ is a disjoint union of open connected subsets, which means that these subsets (the trajectories of $X_H$) are connected components of $N$, q.e.d.
In $\mathbb{R}^2$ (as well as in the sphere $S^2$), the class of Hamiltonian vector fields coincides with symplectic (=area-preserving) vector fields, by Proposition 15.4. If we endow $\mathbb{R}^2$ we the standard area form $dx \wedge dy$, then the area-preserving condition for a vector field $v = (v_x, v_y)$ is

$$L_v(dx \wedge dy) = 0.$$ 

Explicitly, we have

$$L_v(dx \wedge dy) = d(v_x dy - v_y dx) = \left( \frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y} \right) dx \wedge dy.$$ 

So, a vector field in $\mathbb{R}^2$ is Hamiltonian if and only if it has zero divergence

$$\text{div } v = \frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y}.$$ 

In particular, any divergence-free (=area-preserving) vector field in $\mathbb{R}^2$ automatically has a first integral, namely the Hamiltonian function. For more complicated surfaces this does not have to be so, as shown by the following example. Consider the torus $T^2$ with coordinates $\phi, \psi$, defined modulo $2\pi$. Then the vector field

$$\begin{cases} \dot{\phi} = 1, \\ \dot{\psi} = \alpha, \end{cases}$$

where $\alpha$ is constant, has zero divergence and thus preserves the area form $d\phi \wedge d\psi$. However, if $\alpha$ is irrational, then this vector field does not have any first integrals. Indeed, consider the trajectory of this vector field (i.e. a solution of the above system of ODEs) with initial condition $\phi = 0$, $\psi = 0$. It is explicitly given by

$$\begin{cases} \phi(t) = t \text{ mod } 2\pi, \\ \psi(t) = \alpha t \text{ mod } 2\pi. \end{cases}$$

Consider also the circle $\phi = c$. The above trajectory intersects this circle when $t = c + 2\pi k$. The corresponding values of $\psi$ are $\alpha(c + 2\pi k) = \alpha c + 2\alpha \pi k$, which means that consecutive intersection points are obtained from each other by means of a rotation with the angle $2\alpha \pi$. This angle is not a rational multiple of $\pi$, so the intersection points are dense in the circle $\phi = c$ (see Lecture 14). But since our trajectory has a dense intersection with every circle $\phi = c$, and the union of such circles is the whole torus, it follows that the trajectory is dense in the torus. So, any continuous function which is constant along the trajectory must be constant on the whole torus. But that means that there are no non-constant first integrals, as desired.

The trajectories of the above vector field are known as irrational windings of the torus. One can similarly show that each trajectory is dense. This can be also derived from the fact that rotations (translations) of the torus take trajectories to trajectories, because the vector field defining the irrational winding commutes (has zero Lie bracket) with $\partial/\partial \phi$ and $\partial/\partial \psi$.

**Example 15.8.** Our next example is natural mechanical systems. A mechanical system is called natural if all exterior forces are potential. An example is a motion of an object under gravity.

The configuration space $M$ of a mechanical system is the space of all its possible positions, endowed with the topology given by the natural notion of “closeness” of two positions. This topology can be rigorously defined as the initial topology with respect to the maps $\phi_i : M \to \mathbb{R}^3$ which compute the position of the $i$’th particle of the system for a given position of the whole system. In more down-to-earth
terms, if a mechanical system consists of \( n \) particles, then to each configuration we can assign a point in \( \mathbb{R}^{3n} \) whose coordinates are positions of all the \( n \) particles. The configuration space is then defined as the subset of \( \mathbb{R}^{3n} \) defined by the constraints on the positions of particles. Furthermore, one can extend this definition to the case of infinitely many particles and describe the configuration space of a system as a subset of an infinite-dimensional space. However, this does not seem to be a good definition from a practical point of view. In practice, it is not necessary to consider positions of all particles of the system to describe the configuration space. It is sufficient to consider a subset of particles with the property that their positions uniquely determine the position of the whole system. As an example, consider the pendulum, i.e. a rigid rod fixed at one of its ends. Although it consists of infinitely many particles, its position is uniquely described by the position of its free endpoint. So, the configuration space of the pendulum is a circle, while the configuration space of the spherical pendulum is the sphere \( S^2 \).

In good examples, the configuration space \( M \) is a manifold. Its dimension is called the number of degrees of freedom of the system. It can be thought of as a number of parameters needed to specify a position of the system. Any motion of a system is described by a curve in its configuration space. An infinitesimal motion is thus a tangent vector to \( M \). The space of all tangent vectors, i.e. the tangent bundle \( TM \) of the configuration manifold \( M \), is called the phase space. Each infinitesimal motion \( v \in TM \) of the system determines the velocity of each individual particle. These velocities can be computed by applying the differentials of the above “position calculating” maps \( \phi_i \) to \( v \). Since the differential is a linear map, the velocity of each particle is a linear function of the velocity \( v \in TM \) of the system.

The kinetic energy of the system is defined by

\[
K = \frac{1}{2} \sum m_i v_i^2,
\]

where the sum is taken over all particles, \( m_i \) is the mass of the \( i \)'th particle, and \( v_i \) is its velocity. In the case of infinitely many particles, this is replaced by the integral

\[
K = \frac{1}{2} \int v^2 \, dm.
\]

Since the velocity of each particle is a linear function of the velocity \( v \) of the system, \( K \) can be viewed as a quadratic form on tangent vectors to \( M \). It is positive definite, so it defines an inner product on each tangent space. It also depends smoothly on the point in \( M \). In other words, the kinetic energy is a Riemannian metric on the configuration space.

For convenience purposes, as a Riemannian metric on \( M \) we take twice the kinetic energy:

\[
\langle v, v \rangle = 2K(v).
\]

This metric defines a map from the tangent bundle \( TM \) to the cotangent bundle \( T^*M \) given by \( v \mapsto \langle v, * \rangle \). This mapping is a particular case of a Legendre transform in mechanics. The image of the velocity \( v \) under the Legendre transform is denoted by \( p \) and is called the momentum. The space of momenta is the cotangent bundle \( T^*M \). Since it is identified with \( TM \) by means of the Legendre transform, it is also called the phase space (sometimes only \( T^*M \), but not \( TM \), is called the phase space).

We also have the potential energy function \( U \) on \( M \), defined by summing up (or integrating) the potential over all particles. The total energy of the system is \( H = K + U \). It is a function on the tangent bundle \( TM \). We can also transport this function to the cotangent bundle using the Legendre transform. The resulting function, also denoted \( H \), is given by

\[
H = \frac{1}{2} \langle p, p \rangle + U(q).
\]
Here $p \in T^*_q M$. The inner product $\langle p, p \rangle$ on momenta is defined by inverting the inner product on velocities (i.e. inverting the corresponding map $TM \to T^*M$ and interpreting the resulting map $T^*M \to TM$ as a bilinear form on cotangent vectors). The physical meaning of this function is still total energy, but expressed in terms of momenta.

**Theorem 15.9.** The motion of a natural mechanical system is described by a Hamiltonian vector field on the phase $T^*M$. The corresponding Hamiltonian function is the total energy $H$, while the symplectic structure is the canonical symplectic structure on the cotangent bundle.

For the motion of a single particle in $\mathbb{R}^n$, we derived this result in Lecture 11.

**Example 15.10.** One more class of examples of Hamiltonian systems is provided by geodesic flows. Let $M$ be a Riemannian manifold, i.e. a manifold endowed with an inner product on each tangent space which depends smoothly on the point. Then one can define the length of a curve $\gamma(t)$, $t_1 \leq t \leq t_2$ on $M$ by

$$\int_{t_1}^{t_2} \sqrt{\langle \gamma'(t), \gamma'(t) \rangle} \, dt.$$ 

The length of a curve clearly does not depend on a parametrization.

Geodesics on $M$ are defined as curves $\gamma(t)$ with the following property: for any real numbers $t_1, t_2$ sufficiently close to each other, the arc of $\gamma(t)$ corresponding to $t_1 \leq t \leq t_2$ is the shortest curve joining the points $\gamma(t_1), \gamma(t_2)$. For example, geodesics in $\mathbb{R}^2$ are straight lines, while geodesics on the sphere $S^2$ are great circles, i.e. sections of the sphere by 2-planes passing through its center. Note that this definition of geodesics does not assume any specific parametrization. A more precise definition of geodesics includes an additional requirement $\langle \gamma'(t), \gamma'(t) \rangle = \text{const}$, i.e. the velocity vector of the curve has constant magnitude.

For surfaces in $\mathbb{R}^n$ (with Riemannian metric induced by the inner product in the ambient space), geodesics can be alternatively defined as curves whose acceleration $\gamma''(t)$ is orthogonal to the surface.

Geodesics are described by a second order differential equation on $M$, i.e. a vector field on the tangent bundle $TM$. However, it turns out to be more convenient to write the geodesic equation on the cotangent bundle $T^*M$. To that end, one identifies the tangent and cotangent bundles using the Riemannian metric (which is the same as the Legendre transform in Example 15.8). This gives a vector field on $T^*M$, known as the geodesic flow (this term may also refer to the flow of that vector field).

**Theorem 15.11.** The geodesic flow is a Hamiltonian vector field on $T^*M$. The corresponding Hamiltonian function is the “kinetic energy”

$$K = \frac{1}{2} \langle p, p \rangle,$$

while the symplectic structure is the canonical symplectic structure on the cotangent bundle.

As in the previous example, the inner product $\langle p, p \rangle$ on momenta is defined by inverting the inner product on velocities.

By comparing this with the previous example, we see that the motion of any free mechanical system (which means no external forces) is given by geodesics on the configuration space. The corresponding metric is twice the kinetic energy of the system. For example, the spherical pendulum with no gravity moves along great circles (the kinetic energy of the pendulum coincides, up to a constant factor, with the standard metric on the sphere; this is because both of these metrics are rotationally invariant). Thus, free physical systems have a geometric interpretation. Conversely, geodesics on a surface embedded in the Euclidian space has a physical interpretation: they are trajectories of a particle that is constrained to a given surface. This in particular follows from the definition of geodesics as curves whose acceleration is orthogonal to the surface. Indeed, if a particle is moving freely on the surface, then the only force acting on it is the reaction force which is orthogonal to the surface. So, by Newton’s second law, the acceleration
is orthogonal to the surface as well, which means that trajectories of the particle are indeed geodesics (and the other way around).

**Lecture 16: Poisson brackets**

**Definition 16.1.** Let $M$ be a symplectic manifold with symplectic form $\omega$. The *Poisson bracket* of two smooth functions $f, g$ on $M$ is a smooth function $\{f, g\}$ defined by

$$\{f, g\} = -\omega(X_f, X_g).$$

Using that the 1-form $\omega(X_f, \ldots)$ is, by definition of $X_f$, equal to $df$, we can rewrite the formula for the Poisson bracket as

$$\{f, g\} = -\omega(X_f, X_g) = \omega(X_g, X_f) = dg(X_f) = L_{X_f}g.$$

Also, we have

$$dg(X_f) = dq(\omega^{-1}(df)) = \omega^{-1}(df, dg),$$

where in the latter formula $\omega^{-1}$ is interpreted as a form on cotangent vectors. This form is known as the *Poisson tensor*. We will denote it by $\pi$. To summarize, we have the following equivalent definitions of the Poisson bracket:

$$\{f, g\} = -\omega(X_f, X_g) = L_{X_f}g = \omega^{-1}(df, dg).$$

The formula $\{f, g\} = L_{X_f}g$ provides an alternative formulation of Hamiltonian dynamics: the evolution of any smooth function $g$ along the flow of $X_f$ is given by

$$\frac{dg}{dt} = \{f, g\}.$$

Here the derivative in the left hand side is the derivative along the trajectories of $X_f$, i.e. it is the time derivative of the function $g(x(t))$, where the point $x(t)$ moves along the vector field $X_f$. In particular, we have the following.

**Proposition 16.2.** A function $g$ is a first integral of the Hamiltonian system $X_f$ (i.e. constant along trajectories of $X_f$) if and only if $\{f, g\} = 0$.

When functions $f, g$ satisfy $\{f, g\} = 0$, one says that they are *in involution*.

We now describe algebraic properties of Poisson brackets.

**Proposition 16.3.** The Poisson bracket operation on a symplectic manifold has the following properties:

1. It is skew-symmetric, i.e. $\{g, f\} = -\{f, g\}$.

2. It is bilinear over real numbers. For example, linearity in the first argument means that $\{af + bg, h\} = a\{f, h\} + b\{g, h\}$, where $f, g, h$ are functions, and $a, b$ are real numbers.

3. It satisfies the product rule in both arguments, e.g. $\{fg, h\} = f\{g, h\} + g\{f, h\}$.

4. It satisfies the *Jacobi identity*

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

Note that some authors define the Poisson bracket with an opposite sign.
Proof. The only property that does not follow right from the definition is the Jacobi identity. To prove the latter, we first rewrite it, using skew-symmetry, as

\[ \{ \{ f, g \}, h \} = \{ f, \{ g, h \} \} - \{ g, \{ f, h \} \}. \]

Using the property \( \{ f, g \} = L_{X_f} g \), this can be further rewritten as

\[ L_{X_{\{ f, g \}}} h = L_{X_f} L_{X_g} h - L_{X_g} L_{X_f} h. \]

But

\[ L_{X_f} L_{X_g} h - L_{X_g} L_{X_f} h = L_{[X_f, X_g]} h, \]

so the Jacobi identity rewrites as

\[ X_{\{ f, g \}} = [X_f, X_g]. \quad (28) \]

To prove this identity, we will use the product rule for the Lie derivative: for any vector fields \( v, w \) and any \( k \)-form \( \xi \) we have

\[ L_v (\xi(w)) = (L_v \xi)(w) + \xi(L_v w), \]

where \( \xi(w) = i_w \alpha \). Using that \( L_v i_w = i_w L_v \), this can also be written as

\[ L_v i_w = i_w L_v + i_{[v, w]}, \quad (29) \]

where both sides should be understood as operations on differential forms. Expressing \( i_{[v, w]} \), we also find that

\[ i_{[v, w]} = L_v i_w - i_w L_v, \]

which is sometimes abbreviated as

\[ i_{[v, w]} = [L_v, i_w]. \]

This product rule follows from the product rule for the usual derivative, because the Lie derivative is defined as the time derivative of \( \phi_t^* (\ldots) \).

Using (29), we find that

\[ i_{[X_f, X_g]} = L_{X_f} i_{X_g} \omega - i_{X_g} L_{X_f} \omega. \]

But since Hamiltonian vector fields preserve the symplectic form, we have \( L_{X_f} \omega = 0 \), so

\[ i_{[X_f, X_g]} \omega = L_{X_f} i_{X_g} \omega = L_{X_f} dg = di_{X_f} dg = d(L_{X_f} g) = d\{ f, g \}, \]

which is equivalent to (28). Thus, (28) and the Jacobi identity are proved.

Corollary 16.4. If \( g, h \) are first integrals of \( X_f \), then so is their Poisson bracket \( \{ g, h \} \).

Proof. Since \( g \) and \( h \) are first integrals, we have \( \{ f, g \} = 0 \) and \( \{ f, h \} = 0 \), so the first and the last term on the left hand side of (27) vanish. Therefore, the second term also vanishes, i.e. \( \{ \{ g, h \}, f \} = 0 \), which is equivalent to saying that \( \{ g, h \} \) is a first integral of \( X_f \).

Corollary 16.5. If \( g \) is a first integral of \( X_f \), then \( X_g \) is a symmetry of \( X_f \), i.e. \( [X_f, X_g] = 0 \).

Proof. Indeed, from (28) we have that \( [X_f, X_g] = 0 \) whenever \( \{ f, g \} = 0 \).
Lecture 17: Liouville integrable systems on symplectic manifolds

**Definition 17.1** (Continuous time Liouville integrable system). Let \( X_H \) be a Hamiltonian vector field on a \( 2n \)-dimensional symplectic manifold \((M, \omega)\). Then \( X_H \) is called completely integrable (or Liouville integrable, or just integrable) if it has \( n = \frac{1}{2} \dim M \) first integrals \( F_1, \ldots, F_n \) which are:

1. **Functionally independent**, which means that the differentials \( dF_1, \ldots, dF_n \) are linearly independent almost everywhere on \( M \).
2. **In involution**, i.e. \( \{F_i, F_j\} = 0 \) for all \( i, j = 1, \ldots, n \).

Let us comment on these conditions. First of all, since the Hamiltonian \( H \) itself is a first integral of \( X_H \), one can always assume that \( F_1 = H \). If this is the case, then the condition \( \{F_i, F_i\} = 0 \) is satisfied automatically. In particular, the \( \{F_i, F_j\} = 0 \) condition is vacuous for \( 2n \leq 4 \). If \( 2n \geq 6 \), then this condition becomes non-trivial and it is equivalent to requiring that each of the functions \( F_1, \ldots, F_n \) is a first integral for each of the Hamiltonian vector fields \( X_{F_1}, \ldots, X_{F_n} \).

The condition that the differentials of \( F_1, \ldots, F_n \) are generically independent roughly speaking means that none of the \( F_i \)'s can be expressed as a function of the other ones. Indeed, if, say, \( F_1 = G(F_2, \ldots, F_n) \), then

\[
dF_1 = \sum_{i=2}^{n} \frac{\partial G}{\partial x_i} (F_2, \ldots, F_n) dF_i,
\]

so the differentials of \( F_1, \ldots, F_n \) are linearly dependent. Conversely, it is not hard to show that linear dependence of differentials implies that one of the \( F_i \)'s can be locally expressed as a function of the other ones.

One can also reformulate the functional independence condition in terms of the **moment map**, that is the map \( F: M \to \mathbb{R}^n \) given by \( F(x) = (F_1(x), \ldots, F_n(x)) \). Clearly, the differentials of \( F_1, \ldots, F_n \) are linearly independent if and only if the differential of the moment map has maximal rank. So, \( dF_1, \ldots, dF_n \) are linearly independent almost everywhere on \( M \) if and only if the moment map is generically a submersion. In particular, generic level sets of the moment map of an integrable system on \( M \) are half-dimensional submanifolds of \( M \).

Finally, notice that since \( dF_1, \ldots, dF_n \) are linearly independent almost everywhere, the same is true for the Hamiltonian vector fields \( X_{F_1}, \ldots, X_{F_n} \). Furthermore, \( \omega(X_{F_i}, X_{F_j}) = -\{F_i, F_j\} = 0 \), so \( X_{F_1}, \ldots, X_{F_n} \) span an isotropic subspace in every tangent space of \( M \). At generic points this subspace has dimension \( n = \frac{1}{2} \dim M \) and is, therefore, a Lagrangian (= maximal isotropic) subspace. This in particular means that \( n = \frac{1}{2} \dim M \) is the maximal number of independent first integrals in involution. Indeed, if we had more, then their Hamiltonian vector fields would span an isotropic subspace of dimension \( > \frac{1}{2} \dim M \), which is not possible. So, completely integrable systems are, in a sense, “maximally integrable”: they have a maximal possible number of first integrals in involution. There do exist systems which have more integrals, but in this case some of the integrals must have non-zero pairwise Poisson brackets. Such systems are known as **superintegrable**, or **degenerately integrable** (we will discuss a precise definition later on).

In the discrete setting, the definition of Liouville integrability is the same, with a map \( T \) instead of a vector field \( X_H \). The Hamiltonian property of \( X_H \) is replaced by the assumption that the map \( T \) preserves the symplectic structure: \( T^* \omega = \omega \). This is indeed a natural analog of the Hamiltonian property, because, as we know, Hamiltonian vector fields can be viewed as infinitesimal symplectic diffeomorphisms, and, conversely, any vector field whose flow preserves the symplectic structure is locally Hamiltonian.

**Definition 17.2** (Discrete time Liouville integrable system). Let \( T: M \to M \) be a symplectic diffeomorphism of a \( 2n \)-dimensional symplectic manifold \((M, \omega)\) (i.e. \( T^* \omega = \omega \)). Then \( T \) is called completely
integrable (or Liouville integrable, or just integrable) if it has \( n = \frac{1}{2} \dim M \) functionally independent first integrals \( F_1, \ldots, F_n \) in involution.

This definition in particular means that to every discrete time integrable system one can associate a continuous time system: as a Hamiltonian of such a system, one can take any of the integrals \( F_i \). Conversely, one can obtain a discrete integrable system out of a continuous one by taking the time 1 shift along the trajectories of \( X_H \).

Also note that while any continuous time Hamiltonian system on a 2D manifold is automatically integrable (because the Hamiltonian function is a first integral), there is no reason for the corresponding statement in the discrete setting to be true. In other words, an area-preserving map of a surface to itself does not have to be integrable.

The main structural result about integrable systems is the \textit{Arnold-Liouville theorem}. We state it simultaneously in the continuous and discrete settings. The continuous version is by now considered a classical result, while the discrete version is relatively recent and belongs to A. Veselov [5].

\begin{theorem}[Arnold-Liouville theorem] Consider a continuous or discrete time integrable system on a \( 2n \)-dimensional symplectic manifold \((M, \omega)\). Let \( F_1, \ldots, F_n \) be its independent first integrals in involution, and let \( F = (F_1, \ldots, F_n) \) be the corresponding moment map. Consider a level set \( M_c = F^{-1}(c) \) of the moment map (i.e. a joint level set \( \{ x \in M \mid F_i(x) = c_i \} \)), and assume that \( M_c \) is regular (i.e. \( dF \) has maximal rank on \( M_c \)), compact, and connected. Then:

1. The submanifold \( M_c \) is diffeomorphic to an \( n \)-dimensional torus.
2. The dynamics on \( M_c \) is quasi-periodic. In other words, there exist \( 2\pi \)-periodic coordinates \( \phi_1, \ldots, \phi_n \) on the torus \( M_c \) such that the continuous dynamics on \( M_c \) is given by

\( 
\dot{\phi}_i = k_i, 
\)

while the discrete dynamics is given by

\( 
\phi_i \mapsto \phi_i + l_i. 
\)

3. The restriction of the symplectic structure \( \omega \) to \( M_c \) vanishes (half-dimensional submanifolds with this property are called \textit{Lagrangian submanifolds}; a submanifold is Lagrangian if and only if its tangent space at every point is Lagrangian). Furthermore, one can choose the angle coordinates \( \phi_1, \ldots, \phi_n \) on the torus \( M_c \) as well as on nearby level sets, and also choose new first integrals \( s_i = s_i(F_1, \ldots, F_n) \) such that

\( 
\omega = \sum_{i=1}^{n} ds_i \wedge d\phi_i. 
\)

The variables \( s_i \) are known as \textit{action variables}, while \( \phi_i \)'s are \textit{angle variables}. Together they are called \textit{action-angle variables}.

4. The dynamics on \( M_c \) (i.e. the integral trajectories in the continuous case and iterates of the map in the discrete case) can be explicitly described by quadratures.

The connectedness assumption of the theorem is not crucial. If the level set \( M_c \) is disconnected, then each of the connected components of \( M_c \) is a torus, while the rest of the theorem stays intact. In the discrete setting one needs an additional assumption that each connected component is invariant under the map \( T \), otherwise one cannot restrict the map to \( M_c \). In any case, this assumption is always satisfied for a sufficiently high power of \( T \). The compactness assumption can also be weakened, as we will see further.
As a corollary of the Arnold-Liouville theorem, the phase space of an integrable system (i.e. the symplectic manifold where the system is defined) is almost everywhere foliated by Lagrangian tori, and the dynamics on these tori is quasi-periodic.

**Exercises.**

1. Let \( f, g \) be two functions in \( \mathbb{R}^2 \) such that \( df \) and \( dg \) are linearly dependent at every point. Show that there exists an open subset \( U \subset \mathbb{R}^2 \) in which either \( f \) is a function of \( g \), or \( g \) is a function of \( f \).

2. Let \( \dot{\phi}_i = k_i \) be a vector field on a torus. Prove that the trajectories of this vector field are periodic if and only if the vector \( k = (k_1, \ldots, k_n) \) can be expressed as \( k = \lambda z \), where \( \lambda \in \mathbb{R} \), and \( z \in \mathbb{Z}^n \) is a vector with integer components.

3. What is the condition on \( l_i \)'s under which the map \( \phi_i \mapsto \phi_i + l_i \) of the torus to itself has periodic orbits?

4. Let \( s_i \) be one of the action variables. Prove that the trajectories of the vector field \( X_{s_i} \) are \( 2\pi \)-periodic (hint: compute the Poisson bracket \( \{ s_i, \phi_j \} \)).

5. Assume that \( M_c \) is compact but possibly disconnected. Prove that in the discrete case there exists \( k > 0 \) such that \( T^k \) maps every connected component of \( M_c \) to itself.

**References**


