RAJAN MEHTA

GEOMETRY AND MECHANICS

VERSION: DECEMBER 4, 2019

Contents

1 *Symmetries of space and time*

There are no landmarks in space; one portion of space is exactly like every other portion, so that we cannot tell where we are...we may compute our rate of motion with respect to the neighboring bodies, but we do not know how these bodies may be moving in space. — James Clerk Maxwell

The universe around us appears to be 3-dimensional[1]. In mathematics, when we think of 3-dimensional space, it has a special point (the origin) and special directions (the x-, y-, and z-axes). However, the universe does not have any special points or directions. It's just space.

To facilitate a quantitative study of events that occur in space, we need to *choose* an origin and directions for the axes. This choice is called a *reference frame*. After choosing a reference frame, you can use coordinates to describe where events occur, describe relationships between events in terms of equations, and give quantitative answers to questions about the events.

But what if two different people answered the same questions using different reference frames? Their numerical answers might disagree, but there should be some precise sense in which their answers agree. To address this issue, we need to be able to take data collected in one reference frame and convert it so it can be compared with data collected in a different reference frame[2].

As a simple example, suppose someone stands on the roof of a 20 m building, throws a ball into the air, and, using the roof as height zero, measures that the ball reaches a maximum height of 5 m. Another person, using the ground as height zero, measures the maximum height of the ball to be 25 m. Even though the numbers are different, each number is correct *in the appropriate reference frame*. By adding or subtracting the height of the building, you can translate from one reference frame to another.

In this chapter, we will learn more about reference frames and the mathematical framework for translating between them.

[1] This may not be true! According to string theory (which remains speculative), there are extra dimensions that are too small for us to see. Although our focus will be on 3-dimensional space, we will study mathematical frameworks that can work in any number of dimensions.

[2] It might help to think of this as being analogous to translating from one language to another.

1.1 *Frames and transformations in the plane*

Reference frames in the plane

To build some intuition for how reference frames work, let's start with the simple case of 2-dimensional space, i.e. the plane[3]. We want to consider different ways to draw axes. To simplify the discussion, we can agree to the following rules for drawing the axes[4]:

1. The units for the axes should be some predetermined length (say, meters).
2. The x- and y-axes should be chosen to be perpendicular to each other.
3. The x- and y-axes should be chosen so that the y-axis is counterclockwise from the x-axis.

To describe a reference frame satisfying these rules, it is enough to say where the origin is and what direction the x-axis is pointing in. It may help to think of a particular reference frame as being the perspective of an imaginary observer who is standing at the origin and facing the x-axis.

Now, suppose that we are looking at an object that is located at some point p (see Figure 1.1). Two people with different reference frames will obtain different coordinates to describe the location of p, but there should be equations that relate the coordinates in one reference frame to those in the other. Collectively, these equations make up what we will call a *transformation*[5] of the plane; in particular, the transformations of the plane relating reference frames satisfying the above rules are called *special Euclidean transformations*[6].

[3] Imagine the plane as an infinitely large sheet of completely blank paper.

[4] Later, we will be able to consider the possibility of lifting these rules.

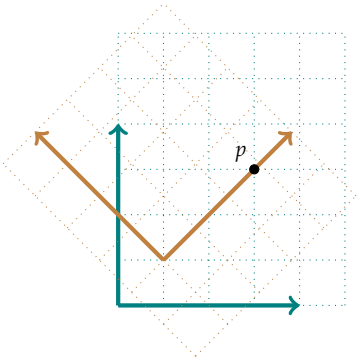

Figure 1.1: Two different reference frames.

[5] In general, a transformation is a map from a space to itself.

[6] An explanation of this term will come a little bit later.

Formulas for special Euclidean transformations

We would like to find a description of all the special Euclidean transformations. Let's start with the following two simple cases:

- *Translations* shift the points in the plane over by some fixed amount. They take the form

$$T_{(a,b)} : (x, y) \mapsto (x + a, y + b),$$

  where a and b can be any real numbers. Translations change between the perspectives of observers who are located at different points but facing the same direction; see Figure 1.2.

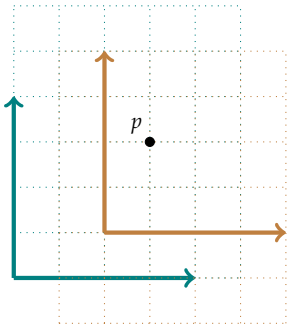

Figure 1.2: Reference frames related by a translation.

- *Rotations* rotate the points in the plane around the origin. They take the form

$$R_\theta : (x, y) \mapsto (x \cos\theta - y \sin\theta, x \sin\theta + y \cos\theta),$$

where θ is the angle of counterclockwise rotation. Rotations change between the perspectives of observers who are located at the same point but facing different directions; see Figure 1.3.

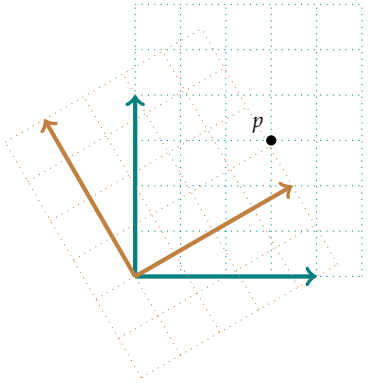

Figure 1.3: Reference frames related by a rotation.

By thinking about observers, you should be able to convince yourself that any two reference frames can be related by a combination of a rotation and a translation. Thus, every special Euclidean transformation can be expressed in the form

$$M_{\theta,(a,b)} : (x, y) \mapsto (x\cos\theta - y\sin\theta + a, x\sin\theta + y\cos\theta + b) \tag{1.1}$$

for some θ, a, and b.

We can make expression (1.1) look nicer by using vector notation. Let $\mathbf{x}$ denote the vector (x, y), let $\mathbf{a}$ denote the vector (a, b), and let[7]

$$U_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}. \tag{1.2}$$

A special Euclidean transformation involves rotating by some angle θ and then translating by some vector $\mathbf{a}$. Algebraically, this means we multiply $\mathbf{x}$ by U_θ and then add $\mathbf{a}$, so the general formula is

$$M_{\theta,\mathbf{a}} : \mathbf{x} \mapsto U_\theta\mathbf{x} + \mathbf{a}. \tag{1.3}$$

In fact, this is the same formula as (1.1), just written in a more compact form.

[7] You probably have seen the matrix U_θ before. It's called a *rotation matrix* because multiplication by U_θ has the effect of rotating vectors counterclockwise by the angle θ.

Groups and symmetry

Observe that the special Euclidean transformations satisfy the following properties[8]:

1. *Closure*: The composition of two special Euclidean transformations is another special Euclidean transformation.
2. *Invertibility*: Special Euclidean transformations can be reversed.

[8] The easiest way to make this observation is to think in terms of translating between the perspectives of observers.

The above properties imply that the special Euclidean transformations form something called a *group*. Those of you who have taken a course in abstract algebra will be familiar with the term. For those of you who haven't, it will be good enough to think of the term *group* as being a shorthand way of saying that a particular set of transformations satisfies the above two properties.

In 1872, Felix Klein[9] gave a speech where he set forth one of the most influential ideas in modern mathematics, now known as the *Erlangen program*. At its essense was the idea that you could study a geometric structure by considering the group of transformations that preserve the structure, i.e. the group of *symmetries*[10].

[9] This is the guy who the Klein bottle is named after.

[10] Here is an illustration in a simple case. An equilateral triangle has 6 symmetries (3 rotations, including the trivial "rotation by 0", and 3 reflections), whereas a square has 8 symmetries. Since they have different numbers of symmetries, we can conclude (in case you weren't sure) that a triangle is different from a square.

But the Erlangen program can also work in the reverse direction. If we have a group of transformations, we can ask, "What is the thing that these transformations are symmetries of?" Usually, the answer takes the form of some kind of geometric information. This information effectively defines the group[11].

Let's do this with the special Euclidean group. First, observe that the distance between any two points is preserved by all special Euclidean transformations. This follows from the fact that distance is preserved by translations and by rotations.

Angles are also preserved, but this can be seen as a consequence of preservation of distance[12], so this doesn't tell us anything new.

But there is another geometric structure that the special Euclidean transformations preserve: *orientation*. In the plane, you can think of orientation as being the distinction between clockwise and counterclockwise. That is, the transformations allow space to be rotated and shifted, but never reflected or flipped[13].

I claim that the special Euclidean transformations are exactly the transformations of the plane that preserve distance and orientation[14]. The above discussion explains why all the special Euclidean transformations preserve distance and orientation, but it remains to conversely show that any transformation of the plane that preserves distance and orientation is of the form (1.3). You will be guided through a proof in the exercises.

[11] Imagine that the transformations are saying, "We know how we're defined as formulas, but we want to know what our purpose in life is."

[12] Recall the "side-side-side" rule from high school geometry.

[13] Take a sheet of paper and draw a big counterclockwise arrow ↺ on it, using a pen or marker that's dark enough to see on the other side of the paper. When you rotate or move the paper around, the arrow remains counterclockwise. But when you flip the paper over or look at it in the mirror, the arrow becomes clockwise. That's orientation.

[14] This is the explanation for the term *special Euclidean*; *Euclidean* means distance-preserving, and the modifier *special* indicates that we are only considering the Euclidean transformations that are orientation-preserving.

The big picture

Let's step back and summarize what has happened in this section. We started by considering reference frames in the plane. Then we described the transformations that translate between reference frames. Then we recognized these transformations as being the ones that preserve distance and orientation. The conclusion is that the underlying geometry of the reference frame problem is the usual Euclidean geometry (which gives the notion of distance), together with orientation.

Exercises

1. In school, you learned that people used to think that the earth is the center of the solar system, and that Copernicus discovered that the sun is the center of the solar system. Let's think more deeply about this. What exactly did Copernicus discover? What do we mean when we say that the sun is the center of the solar system?

2. (a) Give a conceptual explanation, in terms of transformations, for why $U_\theta U_\varphi$ should be equal to $U_{\theta+\varphi}$ for all angles θ and φ.

(b) Do the matrix multiplication to verify that $U_\theta U_\varphi = U_{\theta+\varphi}$.

3. (a) Find the "group law" for the special Euclidean transformations in the form (1.3). That is, compute the composition $M_{\theta,\mathbf{a}} \circ M_{\varphi,\mathbf{b}}$ and find the angle of rotation and the translation vector that you would require to do the composition in one step.

   (b) Find a formula that expresses the inverse of $M_{\theta,\mathbf{a}}$ in the form (1.3). [Hint: Use your result from (a). Solve the equation $M_{\theta,\mathbf{a}} \circ M_{\varphi,\mathbf{b}} = M_{0,0}$ for φ and $\mathbf{b}$ in terms of θ and $\mathbf{a}$.]

4. It's important to remember that the results in this section rely in a significant way on the ground rules that we used for constructing reference frames.

   (a) How would the results change if we did not include the last ground rule for choosing reference frames, that the y-axis should be counterclockwise from the x-axis? What would the resulting group of transformations be? What geometric information would the transformations preserve?

   (b) Try to grapple with the question of how the results would change if we didn't include *any* of the ground rules for choosing reference frames.

5. (a) Show that the matrix U_θ in equation (1.2) is orthogonal[15].

   (b) Show that $\det U_\theta = 1$.

   (c) Show that, conversely, if a 2×2 matrix U is orthogonal and such that $\det U = 1$, then U is equal to U_θ for some angle θ.

[15] see the exercises in Appendix A.

6. Suppose that $g : \mathbb{R}^2 \to \mathbb{R}^2$ is a map that fixes the origin (i.e. $g(0,0) = (0,0)$) and preserves distances between points (i.e. $\|g(\mathbf{x}) - g(\mathbf{y})\| = \|\mathbf{x} - \mathbf{y}\|$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$).

   (a) Show that g preserves dot products, i.e. $g(\mathbf{x}) \cdot g(\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$.

   (b) Show that $\{g(\mathbf{e}_1), g(\mathbf{e}_2)\}$ is an orthonormal basis.

   (c) Show that, if $\mathbf{x} = (x, y)$, then $g(\mathbf{x}) = xg(\mathbf{e}_1) + yg(\mathbf{e}_2)$. Explain why it follows that $g(\mathbf{x}) = U\mathbf{x}$, where the columns of U are $g(\mathbf{e}_1)$ and $g(\mathbf{e}_2)$.

   (d) Explain why, if we furthermore assume that g preserves orientation, then it must be that $\det U = 1$.

   (e) Suppose that $f : \mathbb{R}^2 \to \mathbb{R}^2$ is a map that preserves distances between points and orientation. Show that f is a special Euclidean transformation. [Hint: Let g be given by $g(\mathbf{x}) = f(\mathbf{x}) - f(0,0)$.]

1.2 *The rotation group $SO(3)$*

Special Euclidean transformations of $\mathbb{R}^3$

Now let's consider special Euclidean transformations of $\mathbb{R}^3$. Remarkably, if we use vector and matrix notation, most of the mathematics is exactly the same as in the 2-dimensional case[16].

By definition, a transformation $f : \mathbb{R}^3 \to \mathbb{R}^3$ is a *Euclidean transformation* if f preserves distances[17]. A proof virtually identical to Exercise 6 in Section 1.1 shows that f is a Euclidean transformation if and only if it can be written in the form

$$f : \mathbf{x} \mapsto U\mathbf{x} + \mathbf{a},$$

where U is an orthogonal 3×3 matrix and $\mathbf{a}$ is a vector.

We then furthermore say that f is a *special Euclidean transformation* if $\det U = 1$. This condition should be interpreted as meaning that f preserves orientation.

Let's go into more detail about what orientation means in $\mathbb{R}^3$. Recall that, in the plane, orientation is the distinction between clockwise and counterclockwise. In 3 dimensions, orientation can be visualized as the distinction between right hands and left hands[18].

As a matter of convention, we draw the three axes in $\mathbb{R}^3$ according to the "right-hand rule": if you point the fingers of your right hand toward the x-axis and curl them toward the y-axis, your thumb will point toward the z-axis. Try to convince yourself that, if two different people draw axes according to this rule, they can translate between their reference frames using a special Euclidean transformation.

The rotation group

The orthogonal 3×3 matrices with determinant 1 form a group that is denoted $SO(3)$. Mathematically, the elements of $SO(3)$ are the special Euclidean transformations that fix the origin. Geometrically, they are the transformations that rotate points around the origin; as a result, $SO(3)$ is sometimes referred to as the *rotation group* of $\mathbb{R}^3$.

In the plane, it is fairly easy to describe all of the rotations, since the only thing we need to specify is the angle of rotation θ. However, in 3-dimensional space, you can rotate in an infinite number of directions, and the interaction between rotations in different directions is quite complicated, as you will see in the exercises.

Nonetheless, the rotation group is very important[19], so we should try to understand it however we can. One way to do so is to look at infinitesimals.

[16] In fact, we could just as easily consider special Euclidean transformations of $\mathbb{R}^n$ for arbitrary n.

[17] in other words, $\|f(\mathbf{x}) - f(\mathbf{y})\| = \|\mathbf{x} - \mathbf{y}\|$

[18] Try moving your right hand around, rotating it in different ways. No matter what you do, you can always tell that it's a right hand. But in the mirror, it looks like a left hand, because the mirror reverses orientation.

[19] It becomes even more important in the study of quantum mechanics.

Angular velocities

Consider a velocity vector in the space of rotations. That is, rather than thinking of the discrete change that takes place when you perform a rotation, imagine the continuous change that takes place as you rotate. At every instant of the process, you might say that you are rotating in a particular direction with a particular speed. What you are describing is an infinitesimal rotation.

Recall that, in calculus, you learned to interpret velocity as an infinitesimal displacement. This is the angular version of the same story. We are considering *angular velocities* and interpreting them as infinitesimal rotations.

To make the idea precise, consider a t-dependent path of rotations $U(t)$, describing a continuous rotation. We assume that $U(0) = I$, which means that at $t = 0$ no rotation has taken place yet. By definition, $U'(0)$ is an angular velocity.

The set of all angular velocities is denoted $\mathfrak{so}(3)$. So what do the elements of $\mathfrak{so}(3)$ look like? We know that $U(t)$ is orthogonal for all t, so it satisfies the equation

$$U(t)^\intercal U(t) = I$$

for all t. If we differentiate[20] both sides of this equation with respect to t, we get

$$U'(t)^\intercal U(t) + U(t)^\intercal U'(t) = 0,$$

and if we then evaluate at $t = 0$, we get

$$U'(0)^\intercal + U'(0) = 0. \tag{1.4}$$

Equation (1.4) says that the angular velocity $U'(0)$ is *skew-symmetric*, i.e. it is equal to the negative of its own transpose[21]. This allows us to explicitly describe them; every skew-symmetric matrix is of the form

$$A = \begin{bmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{bmatrix}$$

for some $a, b, c \in \mathbb{R}$. So this is what angular velocities look like.

Here are some useful properties that $\mathfrak{so}(3)$ has:

1. It is a vector space. Specifically, if you add two elements of $\mathfrak{so}(3)$ together, you get another element of $\mathfrak{so}(3)$. And if you multiply an element of $\mathfrak{so}(3)$ by a scalar, you get another element of $\mathfrak{so}(3)$. Physically, this makes sense, because it should be possible to add angular velocities, in a similar way to how you can add ordinary velocities.

[20] We're using the product rule on a product of matrices! You should make sure that you believe this is valid.

[21] In particular, all the entries along the diagonal must be 0.

2. If A and B are elements of $\mathfrak{so}(3)$, then $AB - BA$ is again in $\mathfrak{so}(3)$.

   This operation plays a hugely important role in both math and physics, so there is a special notation given to it:

   $$[A, B] := AB - BA.$$

   Some of the properties of this *commutator bracket*[22] are in the exercises.

[22] This operation is also sometimes referred to as the *Lie bracket*, pronounced "lee".

3. Every element A of $\mathfrak{so}(3)$ satisfies the property $\det A = 0$. This follows from the calculation $\det A = \det(A^T) = \det(-A) = (-1)^3 \det A = -\det A$.

   The fact that the determinant is 0 tells us that 0 is an eigenvalue. The physical explanation is that, if A is an infinitesimal rotation, then the axis of rotation is held fixed. The 0-eigenvectors of A are exactly the vectors that form the axis of rotation.

As a vector space, $\mathfrak{so}(3)$ is 3-dimensional. One reasonable choice of basis is

$$L_x = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}, \quad L_y = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}, \quad L_z = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Physically, L_x is the counterclockwise infinitesimal rotation around the x-axis with angular speed 1. Similarly, L_y and L_z are infinitesimal rotations around the y- and z-axes, respectively.

Matrix exponentials

There's one place above where I made a claim that wasn't fully justified. We saw that angular velocities are always skew-symmetric matrices, but we didn't prove that *every* skew-symmetric matrix is an angular velocity. In other words, we need to show that, for every skew-symmetric matrix A, there exists some path of rotations $U(t)$ such that $U(0) = I$ and $U'(0) = A$.

To prove this, we'll use *matrix exponentials*, defined as follows. For any square matrix A, consider the power series[23]

[23] You'll prove in the exercises that this series converges for all t.

$$e^{At} := I + At + \frac{A^2}{2!}t^2 + \frac{A^3}{3!}t^3 + \dots$$

If we set $U(t) = e^{At}$, then $U(0) = I$, and we can calculate that $U'(t) = Ae^{At}$, so $U'(0) = A$.

I claim that, if A is skew-symmetric, then $U(t)$ is orthogonal for all t. The quickest way to prove it is to observe that, because of the skew-symmetry identity $A^\intercal = -A$,

$$U(t)^\intercal = (e^{At})^\intercal = I - At + \frac{A^2}{2!}t^2 - \frac{A^3}{3!}t^3 + \dots = e^{-At}.$$

The result then follows from the formal[24] identity $e^{-At}e^{At} = I$.

Once we know that $U(t)$ is orthogonal for all t, then we can deduce that $\det(U(t)) = 1$ for all t. This is because the determinant of an orthogonal matrix is always ± 1, and by continuity the determinant can't jump from 1 to -1.

Beware! For matrix exponentials, $e^A e^B$ is generally not equal to e^{A+B}. We can see this already by calculating the first few terms of the power series for each expression:

$$\begin{aligned} e^A e^B &= (I + A + \frac{1}{2}A^2 + \dots)(I + B + \frac{1}{2}B^2 + \dots) \\ &= I + A + B + \frac{1}{2}A^2 + AB + \frac{1}{2}B^2 + \dots, \end{aligned}$$

whereas

$$\begin{aligned} e^{A+B} &= I + A + B + \frac{1}{2}(A+B)^2 + \dots \\ &= I + A + B + \frac{1}{2}A^2 + \frac{1}{2}(AB + BA) + \frac{1}{2}B^2 + \dots \end{aligned}$$

Observe that, at least as far as these calculations go, the two expressions *would* be equal if it were true that $AB = BA$. This continues to be the case for higher-order terms. Thus, the usual addition rule for exponentials fails to hold *precisely because of the fact that matrix multiplication is not commutative*.

It turns out that you can compensate by adding correction terms involving the commutator bracket. As a first correction, observe that

$$e^{A+B+\frac{1}{2}[A,B]} = I + A + B + \frac{1}{2}A^2 + AB + \frac{1}{2}B^2 + \dots,$$

so we can see that $e^A e^B \approx e^{A+B+\frac{1}{2}[A,B]}$ to second-order.

There is a theorem asserting that this process can be continued indefinitely, so that you can write $e^A e^B = e^C$, where C is an infinite sum whose terms are iterated commutator brackets of A and B. The first few terms are:

$$C = A + B + \frac{1}{2}[A,B] + \frac{1}{12}([A,[A,B]] + [B,[B,A]]) + \dots$$

This is known as the *Baker-Campbell-Hausdorff formula*. We won't prove it here, but it may be helpful for you to be aware of its existence. In the present situation, it tells us that, if we rotate by some angle around some axis, and then rotate by some other angle around some other axis, then the total effect can also be described as a rotation around some axis by some angle.

[24] Here, "formal" refers to the fact that it is just an identity of power series, not due to some law of exponents. This is important because, as we are about to see, the usual laws of exponents don't actually hold for matrices!

Exercises

1. One way that the rotation group is complicated is that it is non-commutative. Use a real-life object to demonstrate that, if you

rotate about the x-axis and then rotate about the y-axis, the cumulative effect is different than if you rotate about the y-axis and then rotate about the x-axis.

2. Another way that the rotation group is complicated is that the group law is difficult to describe geometrically. Suppose I rotate by $\pi/6$ about the x-axis and then rotate by $\pi/4$ about the y-axis. Try to describe the resulting rotation. [Note: the goal isn't really to find the answer but to appreciate the difficulty of the problem.]

3. Let A be an arbitrary 3×3 matrix, and let M be the maximum absolute value of all of the entries of A.

   (a) Show that the absolute value of every entry of A^n is less than or equal to $3^{n-1}M^n$. [Hint: Use induction.]

   (b) Show that the series $\sum_{n=0}^{\infty} \frac{3^{n-1}M^n}{n!}$ converges regardless of what M is.

   (c) Explain how it follows that the series for e^A converges.

4. (a) Show that, for any scalar a,

$$e^{aL_x} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(a) & -\sin(a) \\ 0 & \sin(a) & \cos(a) \end{bmatrix}.$$

   (b) Find e^{aL_y} and e^{aL_z}.

5. Show that the commutator bracket satisfies the following properties:

   (a) $[A, B + C] = [A, B] + [A, C]$.

   (b) $[A, rB] = r[A, B]$ for any scalar r.

   (c) $[A, B] = -[B, A]$.

   (d) $[A, [B, C]] = [[A, B], C] + [B, [A, C]]$. (This is a famous equation called the *Jacobi identity*.)

6. Physicists actually use vectors, not matrices, to describe angular velocities. This problem explains how their approach is related to ours. The choice of a basis for $\mathfrak{so}(3)$ gives us a way to identify vectors in $\mathbb{R}^3$ with elements in $\mathfrak{so}(3)$. Specifically, we can define a map $\beta : \mathbb{R}^3 \to \mathfrak{so}(3)$ by

$$\beta : (a, b, c) \mapsto aL_x + bL_y + cL_z = \begin{bmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{bmatrix}.$$

(a) Show that, for every vector $\mathbf{x}$,

$$\beta(\mathbf{x})\mathbf{x} = 0.$$

This means that $\mathbf{x}$ is a 0-eigenvector for $\beta(\mathbf{x})$, and therefore $\mathbf{x}$ points in the direction of the axis of rotation of $\beta(\mathbf{x})$.

(b) Show that, for any vectors $\mathbf{x}$ and $\mathbf{y}$,

$$[\beta(\mathbf{x}), \beta(\mathbf{y})] = \beta(\mathbf{x} \times \mathbf{y}).$$

In other words, the map β relates the commutator bracket to the cross product[25].

[25] This is why the cross product appears so much in formulas involving angular motion!

7. Show that $e^{At}e^{Bt}e^{-At}e^{-Bt} \approx I + t^2[A, B]$ to second-order in t. This fact can be interpreted as saying that the commutator bracket really does measure the infinitesimal failure of A and B to commute as infinitesimal transformations.

1.3 *Vector fields and flows*

Vector fields

Recall that the special Euclidean transformations are of the form

$$f : \mathbf{x} \mapsto U\mathbf{x} + \mathbf{a},$$

where U is a rotation matrix and $\mathbf{a}$ is a vector. In the last section, we focused on the rotation component of the transformation, and now we would like to bring the translations in as well.

The translations aren't very interesting in and of themselves, but the translations and rotations interact with each other in a nontrivial way. This is because, in general, if you translate and then rotate, the end result will be different than if you rotate and then translate.

Our approach to studying rotations has made heavy use of the fact that rotations can be described by matrices. Bringing translations into the picture requires a more general framework, which *vector fields* provide.

Recall that a vector field is something that assigns a vector to every point in space. In two dimensions, a vector field is of the form $\mathbf{F} = (P(x, y), Q(x, y))$. In three dimensions, a vector field would take the form $\mathbf{F} = (P(x, y, z), Q(x, y, z), R(x, y, z))$.

You can visually represent a vector field by drawing, at each point, an arrow indicating the vector that is the image of that point; see Figure 1.4.

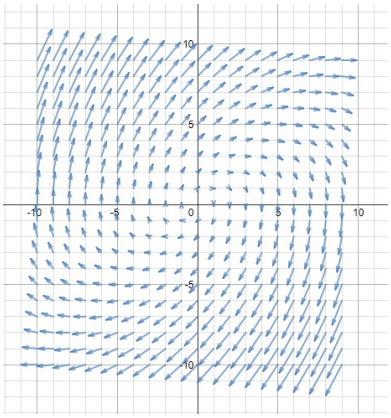

Figure 1.4: The vector field $(y, y - x)$.

Flows of vector fields

Given a vector field $\mathbf{F} = (P(x,y), Q(x,y))$, we can use the component functions to write down the following system of first-order differential equations in two variables[26]:

$$\frac{dx}{dt} = P(x,y), \qquad \frac{dy}{dt} = Q(x,y). \tag{1.5}$$

A solution to such a system is a pair of functions $x(t)$ and $y(t)$ that satisfy the above equations. Frequently, the problem will come with a set of initial values $x(0) = x_0$, $y(0) = y_0$ that your solutions should also satisfy.

Geometrically, a solution $(x(t), y(t))$ to a system of differential equations (1.5) can be viewed as a parametrized path for which the velocity vector at each point is exactly the vector given by the vector field $\mathbf{F}$ at that point. In other words, a solution is a path that "flows" according to the arrows. The initial conditions, if given, specify the starting point of the path. See Figure 1.5.

There is a theorem that says that, if $P(x,y)$ and $Q(x,y)$ have continuous partial derivatives in the vicinity of the initial point (x_0, y_0), then there is a unique solution $(x(t), y(t))$, defined for at least some short amount of time. In the cases that we are mostly interested in, the solutions are actually defined for all times[27] at every point; in this case the vector field is called *complete*.

Let $\mathbf{F}$ be a complete vector field. Given any solution $(x(t), y(t))$ to (1.5), we can ask what its initial values $x_0 = x(0)$ and $y_0 = y(0)$ are. Conversely, given any choice of initial values, we can obtain a unique solution. This means that the *general solution*, i.e. the set of all solutions, can be parametrized by the initial values[28]. When we write it this way, the solution has x_0, y_0, and t as unknowns.

Normally, we treat x_0 and y_0 as parameters, whereas we treat t as the independent variable. But this distinction is just a matter of perspective, and we could just as easily treat the initial values as variables. However, we're used to using notation like x_0 and y_0 to represent parameters, so it would be better to drop the subscripts and just write x and y, which "feel" like variables.

The result is a time-dependent "flow map" $\phi_t : \mathbb{R}^2 \to \mathbb{R}^2$, defined by the rule that $\phi_t(x,y)$ is the point where you end up if you start at (x,y) and then "flow" along the arrows of the vector field for time t. To be more precise, ϕ_t is defined as the unique time-dependent map from $\mathbb{R}^2$ to $\mathbb{R}^2$ satisfying the differential equation[29]

$$\frac{d}{dt}\left[\phi_t(x,y)\right] = (P(\phi_t(x,y)), Q(\phi_t(x,y))) \tag{1.6}$$

and the initial condition $\phi_0(x,y) = (x,y)$. Essentially, ϕ_t solves the system of differential equations for *all* initial values simultaneously.

[26] In this and the next few sections, we will focus on the two-dimensional case, but this is just for notational simplicity. The general picture is similar in three dimensions or even n dimensions.

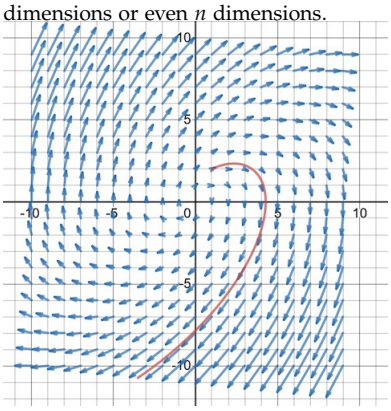

Figure 1.5: The vector field $(y, y-x)$ with the solution with initial condition $(x_0, y_0) = (1,2)$.

[27] This includes negative times.

[28] As a simple one-variable example of this, consider the equation $\frac{dx}{dt} = x$. The general solution can be written as $x(t) = x_0 e^t$.

[29] In the above one-variable example, the flow map is $\phi_t(x) = xe^t$. It's basically the same thing as the general solution, except we are using different notation for the map ($\phi_t(x)$ instead of $x(t)$), and we replaced x_0 with x.

The geometric interpretation is that ϕ_t is a time-dependent path of transformations of the plane, and the vector field $\mathbf{F}$ is the infinitesimal transformation that generates ϕ_t. You can imagine all the points of space moving around, with $\mathbf{F}$ describing the velocities of the points, and ϕ_t describing the actual change of position over time[30].

[30] The situation is analogous to the relationship between infinitesimal rotations and actual rotations, as discussed in Section 1.2. In fact, we will soon see that it is more than just an analogy.

Constant vector fields

A vector field $\mathbf{F} = (P, Q)$ is called *constant* if the functions $P(x, y)$ and $Q(x, y)$ are constant functions. Thus, any constant vector field is of the form $(P, Q) = (a, b) = \mathbf{a}$ for some vector $\mathbf{a}$. Geometrically, a vector field is constant if the vector is the same at every point; see Figure 1.6.

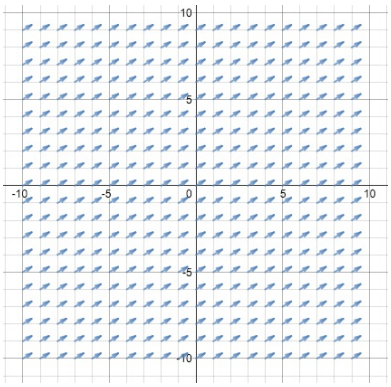

Figure 1.6: A constant vector field.

The system of differential equations corresponding to a constant vector field is

$$\frac{dx}{dt} = a, \qquad \frac{dy}{dt} = b. \tag{1.7}$$

The solution to this system with initial point (x_0, y_0) is

$$x(t) = x_0 + at, \qquad y(t) = y_0 + bt.$$

Thus, the corresponding flow ϕ_t is given by

$$\phi_t(x, y) = (x + at, y + bt),$$

or, in vector form,

$$\phi_t(\mathbf{x}) = \mathbf{x} + t\mathbf{a}.$$

So the time t flow of the constant vector field $\mathbf{a}$ is simply a translation by $t\mathbf{a}$. In other words, constant vector fields are infinitesimal translations.

Linear vector fields

A vector field $\mathbf{F} = (P, Q)$ is called *linear* if the functions $P(x, y)$ and $Q(x, y)$ are linear functions, so they take the form

$$P(x, y) = ax + by, \qquad Q(x, y) = cx + dy,$$

for some constants a, b, c, d. The corresponding system of differential equations is

$$\frac{dx}{dt} = ax + by,$$
$$\frac{dy}{dt} = cx + dy.$$

It is convenient to express this system in vector form as

$$\frac{d\mathbf{x}}{dt} = A\mathbf{x}, \tag{1.8}$$

where

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

So we can see that there is a correspondence between linear vector fields and square matrices. Since it will come up a lot, let's refer to a vector field as being *skew-symmetric linear* if it is a linear vector field that corresponds to a skew-symmetric matrix.

In introductory courses on linear algebra and differential equations, students are taught to solve systems like (1.8) using eigenvalues and eigenvectors. But you can use the matrix exponential to give a simple formal description of the solution[31]:

$$\mathbf{x}(t) = e^{At}\mathbf{x}_0 = \left(I + At + \frac{A^2}{2!}t^2 + \dots \right) \mathbf{x}_0. \tag{1.9}$$

[31] In the exercises, you'll verify that this is indeed the solution.

So the corresponding flow ϕ_t is given by

$$\phi_t(\mathbf{x}) = e^{At}\mathbf{x}.$$

In particular, if A is skew-symmetric, then the flow is a rotation.

Remark: It may seem that the matrix exponential saves you from having to calculate eigenvalues and eigenvectors when solving systems like (1.8). But remember that the matrix exponential is really just a symbol that represents an infinite series. To actually calculate the sum of the series in (1.9), you would generally need to use the eigenvalues and eigenvectors (see the exercises). So you should think of the matrix exponential as just being a different way of thinking about the problem, and not really a different method of solution.

The big picture

Complete vector fields can be thought of as infinitesimal transformations. The flow then describes the actual transformations that occur over time. We saw that constant vector fields are infinitesimal translations, and that skew-symmetric linear vector fields are infinitesimal rotations. Thus, complete vector fields give us a sufficiently general framework to handle all infinitesimal special Euclidean transformations.

Exercises

1. Find the flows $\phi_t(x, y)$ of each of the following vector fields in two different ways: (i) by directly solving the systems of differential equations, (ii) by computing the matrix exponentials.

   (a) $\mathbf{F} = (x, y)$.

   (b) $\mathbf{F} = (y, 0)$.

2. Consider the vector field $\mathbf{F}(x, y) = (x^2, y^2)$. This might seem innocuous (the functions are polynomial!), but actually this vector field is *not* complete. Solve the corresponding system of differential equations and explain how it fails to be complete.

3. (a) Show that, for any square matrix A,

$$\frac{d}{dt}[e^{At}] = Ae^{At}.$$

   (b) Show that $\mathbf{x}(t) = e^{At}\mathbf{x}_0$ is the solution to the differential equation (1.8) with initial value $\mathbf{x}(0) = \mathbf{x}_0$.

   (c) Explain how it follows from (b) that $\phi_t(\mathbf{x}) = e^{At}\mathbf{x}$ is the flow of the linear vector field corresponding to A.

4. (a) Let ϕ_t be the flow of a vector field (and assume that the flow is globally defined for all t). Show that, for all values of s and t,

$$\phi_s \circ \phi_t = \phi_{s+t}.$$

   (b) Show that the time $-t$ flow of a vector field $\mathbf{F}$ is equal to the time t flow of $-\mathbf{F}$.

5. Suppose that A is a 2×2 matrix that has two distinct eigenvalues λ_1, λ_2. Let $\mathbf{v}_1, \mathbf{v}_2$ be eigenvectors that correspond to each eigenvalue. Recall from linear algebra that, if you form the matrix $P = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix}$ (the eigenvectors are the columns of P), then $P^{-1}AP = D$, where

$$D = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}.$$

   This process is called *diagonalization*.

   (a) Show that

$$e^D = \begin{bmatrix} e^{\lambda_1} & 0 \\ 0 & e^{\lambda_2} \end{bmatrix}.$$

   (b) Show that $A^k = PD^kP^{-1}$ for all positive integers k.

   (c) Show that $e^A = Pe^DP^{-1}$.

   (d) Use the result of part (c) to calculate e^A, where

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}.$$

   (e) Calculate e^{At}, where A is the same matrix as in part (d).

1.4 *One-parameter groups of transformations*

In Section 1.3, we saw that, given a complete vector field in $\mathbb{R}^n$, we can obtain[32] a flow ϕ_t, which is a time-dependent transformation of $\mathbb{R}^n$. As part of the definition, the flow satisfies $\phi_0(\mathbf{x}) = (\mathbf{x})$. Additionally, in the exercises, you saw that the flows satisfy the property $\phi_s \circ \phi_t = \phi_{s+t}$ for all values of s and t. Based on these special properties, we can make the following definition.

[32] At this point we'll use vector notation, which allows us to work in any dimension.

Definition: A *one-parameter group* of transformations is a time-dependent transformation ϕ_t such that ϕ_0 is the identity map and $\phi_s \circ \phi_t = \phi_{s+t}$ for all values of s and t.

We have already seen several examples of one-parameter groups, including the following:

- Translations in the direction of a vector $\mathbf{a}$:

$$\phi_t : \mathbf{x} \mapsto \mathbf{x} + t\mathbf{a}.$$

- Rotations of the plane:

$$\phi_t : \mathbf{x} \mapsto U_t\mathbf{x},$$

  where U_t is given by (1.2), using t in place of θ.

- Dilations:

$$\phi_t : \mathbf{x} \mapsto e^t\mathbf{x}$$

Note that the last two are special cases of the more general fact that, for any square matrix A, the map $\phi_t : \mathbf{x} \mapsto e^{At}\mathbf{x}$ is a one-parameter group. Furthermore, all of the above examples are special cases of the even more general fact[33] that *every flow is a one-parameter group*. So, given a complete vector field, we can find its flow, which is a one-parameter group.

[33] We've already proven this! It was actually the motivation for the definition!

Conversely, given a one-parameter group ϕ_t, we can find a vector field by[34] differentiating with respect to t at $t = 0$:

[34] This equation is what you get get when you set $t = 0$ in (1.6).

$$\mathbf{F}(\mathbf{x}) = \frac{d}{dt}|_{t=0}[\phi_t(\mathbf{x})].$$

I claim that ϕ_t is the flow of $\mathbf{F}$. In other words, *every one-parameter group is a flow.*

To prove the claim, let's start with the equation

$$\phi_{s+t}(\mathbf{x}) = \phi_s(\phi_t(\mathbf{x}))$$

and differentiate both sides with respect to s:

$$\frac{d}{ds}\left[\phi_{s+t}(\mathbf{x})\right] = \frac{d}{ds}\left[\phi_s(\phi_t(\mathbf{x}))\right]. \tag{1.10}$$

Because of the chain rule and the fact that $\frac{d}{ds}[s+t] = \frac{d}{dt}[s+t] = 1$, it's actually true that

$$\frac{d}{ds}\left[\phi_{s+t}(\mathbf{x})\right] = \frac{d}{dt}\left[\phi_{s+t}(\mathbf{x})\right].$$

Putting this into (1.10), we get

$$\frac{d}{dt}\left[\phi_{s+t}(\mathbf{x})\right] = \frac{d}{ds}\left[\phi_s(\phi_t(\mathbf{x}))\right].$$

If we set $s = 0$ on both sides, we get

$$\begin{aligned}\frac{d}{dt}\left[\phi_t(\mathbf{x})\right] &= \frac{d}{ds}|_{s=0}\left[\phi_s(\phi_t(\mathbf{x}))\right] \\ &= \mathbf{F}(\phi_t(\mathbf{x})),\end{aligned}$$

which is exactly the equation (1.6) that defines the flow.

The conclusion is that there is a correspondence between one-parameter groups and complete vector fields. Given a complete vector field, you can[35] find its flow, which is a one-parameter group. Given a one-parameter group, you can differentiate at $t = 0$ to get the vector field for which it is flow. As a shorthand terminology for this relationship, we say that a vector field $\mathbf{F}$ *generates* the one-parameter group ϕ_t if ϕ_t is the flow of $\mathbf{F}$.

[35] At least in principle; finding the flow requires solving differential equations, which can be difficult in practice.

Exercises

1. Let $\phi_t : \mathbf{x} \mapsto \mathbf{x} + t\mathbf{a}$ for some vector $\mathbf{a}$.
   (a) Directly verify that ϕ_t is a one-parameter group.
   (b) Differentiate at $t = 0$ to find the vector field that generates the one-parameter group.
2. Let $\phi_t : \mathbf{x} \mapsto (1+t)\mathbf{x}$.
   (a) Give a geometric description of how ϕ_t transforms space as t increases.
   (b) From the geometric description, you might think ϕ_t is a one-parameter group. Do the calculations to show that is isn't.
3. Let ϕ_t be a time-dependent special Euclidean transformation of $\mathbb{R}^n$ (not necessarily a one-parameter group). In other words,

   $$\phi_t : \mathbf{x} \mapsto U_t\mathbf{x} + \mathbf{a}_t,$$

   where U_t is a time-dependent rotation matrix and $\mathbf{a}_t$ is a time-dependent vector.
   (a) Show that the vector field $\mathbf{F} = \frac{d}{dt}|_{t=0}[\phi_t]$ is a sum of a skew-symmetric linear vector field and a constant vector field.

(b) If ϕ_t is a one-parameter group, then it is the flow of the vector field in (a). In this case, find formulas for U_t and $\mathbf{a}_t$ in terms of the matrix A and vector $\mathbf{b}$ corresponding to the linear and constant parts of the vector field, respectively.

1.5 *Vector fields as derivations*

In calculus, there are two different but closely related ways to think about the derivative of a function $f(x)$. One is that, at a specific x-value, you can consider the rate of change of $f(x)$ with respect to a change in x. This rate can be geometrically interpreted as the slope of the tangent line to the graph of $f(x)$ at that specific x-value. From this point of view, the derivative is a number. Let's call this the *pointwise* view.

The other point of view is that you can consider the rate of change at *every* x-value simultaneously. Then you would view the derivative of $f(x)$ as a function $f'(x)$. Let's call this the *global* view.

The pointwise view is more intuitive and better illuminates the conceptual meaning of the derivative. But the global view is more powerful, allowing you to employ the various rules of calculus (such as the chain rule, product rule, etc.) to calculate derivatives fairly easily.

For a function of more than one variable, you instead have *directional derivatives*. In a typical multivariable calculus course, directional derivatives are usually only taught from the pointwise view. At a specific point, you choose a direction and then consider the rate of change of the function in that direction. Since you can easily do the calculations using partial derivatives, it may not seem like the global view is necessary. But we will see that the global view is quite interesting.

Directional derivatives

Recall that, for a function $f(x,y)$, the directional derivative at a point (x_0,y_0) in the direction[36] of a vector (a,b) is

$$D_{(a,b)}f(x_0,y_0) := (a,b)\cdot\nabla f(x_0,y_0) = a\frac{\partial f}{\partial x}(x_0,y_0) + b\frac{\partial f}{\partial y}(x_0,y_0).$$

To pass to the global view, we will want to compute a directional derivative at every point simultaneously. To do this, we will need a vector at every point. In other words, we need a vector field.

Let $\mathbf{F} = (P(x,y),Q(x,y))$ be a vector field. Given a function $f(x,y)$, we can take the directional derivative of f at each point (x,y) with

[36] In multivariable calculus, the definition of directional derivative usually includes the requirement that the vector (a,b) be a unit vector, so that the only information contained in the vector is the direction. For our purposes, we will want to allow (a,b) to be *any* vector. As a result, the directional derivative will depend on the magnitude, in addition to the direction.

respect to the vector $\mathbf{F}(x,y)$. The result is another function of x and y:

$$D_{\mathbf{F}} f := \mathbf{F} \cdot \nabla f = P(x,y)\frac{\partial f}{\partial x} + Q(x,y)\frac{\partial f}{\partial y}.$$

You could argue that the directional derivative operation is so fundamental that you shouldn't really distinguish between a vector field and its associated directional derivative operator. So, instead of writing $\mathbf{F} = (P(x,y), Q(x,y))$, you would write[37]

$$\mathbf{F} = P(x,y)\frac{\partial}{\partial x} + Q(x,y)\frac{\partial}{\partial y}.$$

[37] This might remind you of the $\mathbf{i}, \mathbf{j}, \mathbf{k}$ notation that is frequently used in calculus and physics classes.

From here on, we will adopt this notation for vector fields, and we will use $\mathbf{F}(f)$ to denote the directional derivative of a function f with respect to $\mathbf{F}$.

The directional derivative operator satisfies the product rule[38]

$$\mathbf{F}(fg) = f\mathbf{F}(g) + g\mathbf{F}(f) \tag{1.11}$$

[38] This is an exercise, but it's basically a consequence of the fact that the partial derivative operators satisfy the product rule.

for any functions f and g. Operators that satisfy the product rule are called *derivations*. Since we are identifying vector fields with their directional derivative operators, we can then say that all vector fields are derivations.

It can conversely be shown[39] that every derivation is a vector field. So the conclusion is that vector fields and derivations are essentially the same thing.

[39] In the exercises, you will work through a weaker version of the proof, which avoids a technical step that you would need for the full proof.

Exercises

1. Using derivation notation, write out the vector fields that generate each of the following one-parameter groups of transformations of $\mathbb{R}^3$:
   (a) translation in the x-direction: $\phi_t(x,y,z) = (x+t, y, z)$
   (b) rotation about the x-axis: $\phi_t(x,y,z) = (x, \cos(t)y + \sin(t)z, -\sin(t)y + \cos(t)z)$
2. Show that the directional derivative operator satisfies the product rule (1.11).
3. Suppose that D is an operator on functions that satisfies the product rule (in other words, D is a derivation). Show that there exists a vector field $\mathbf{F}$ such that $\mathbf{F}$ and D agree on polynomial functions. [Hint: Let $P := D(x)$ and $Q := D(y)$. Then show that $\mathbf{F} := P\frac{\partial}{\partial x} + Q\frac{\partial}{\partial y}$ agrees with D when acting on any monomial $f(x,y) = x^m y^n$.]
4. Show that, if $\mathbf{F}$ and $\mathbf{G}$ are derivations, then $\mathbf{F} \circ \mathbf{G} - \mathbf{G} \circ \mathbf{F}$ is a derivation.

5. Suppose that $\mathbf{F}$ is a vector field that generates the flow ϕ_t. Show that $\mathbf{F}(f) = 0$ if and only if $f(\phi_t(\mathbf{x})) = f(\mathbf{x})$ for all $\mathbf{x}$. This is a very important relationship that we will keep coming back to, so try to gain an intuitive feel for the result as well.

6. Consider the function $f(x, y, z) = x^2 + y^2 + z^2$. Find several examples of vector fields $\mathbf{F}$ such that $\mathbf{F}(f) = 0$. [Hint: it may help to think about what kind of transformations don't affect f.]

7. Consider the function $g(x, y, z) = z$. Find several examples of vector fields $\mathbf{F}$ such that $\mathbf{F}(g) = 0$.

1.6 Time according to Galileo

Now that we have a decent understanding of the geometry of space, we can introduce time into the picture. At first, we will consider the Galilean view, which probably agrees with your intuitive understanding of how time works. In this view, the passing of time can be unambiguously measured. Specifically, for any two events, any two observers should agree on which event occurred first as well as how much time elapsed between the first and second event.

Later, we will learn about Einstein's special relativity, in which the passage of time is no longer absolute.

The law of inertia

Think back to Section 1.1, where we considered the possibility of two different observers choosing different reference frames. Let's try to modify that discussion to include the concept of time. There are two issues that emerge. One is that, as part of choosing a reference frame, we need to choose a way to measure time. The other is that there is now a possibility that different reference frames may actually be moving relative to each other over time.

The latter issue is a bit tricky from the physical point of view, since there are many different possible ways to move, and if we allow reference frames to move arbitrarily, then we can get bizarre results. For example, suppose that four people are observing a rock that is lying still on the ground. Person A observes the rock as being motionless. Person B observes from a steadily-moving bicycle and says that the rock is moving with constant velocity. Person C observes while on a pogo stick and says that the rock is bouncing up and down. Person D observes while spinning around and says that the rock is moving in a circle.

If all of these perspectives were allowed, then it would be impossible to say anything meaningful about the motion of the rock. We

could literally choose reference frames that make the rock appear to do anything. This problem is resolved by Newton's First Law, which is a refined form of Galileo's *law of inertia*. It says that an object at rest will remain at rest unless acted on by a force, and an object in motion will maintain a constant velocity unless acted on by a force.

You've probably internalized the law of inertia to the point that it seems trivial to you. But there is something nontrivial there. Newton's First Law asserts that there *objectively exist ideal reference frames* in which the law of inertia holds. The term for such a frame, which we will henceforth use, is *inertial frame*.

Since the rock is not being acted on by a force[40], it should, in an inertial frame, either remain at rest or move with constant velocity. Person A and Person B have inertial frames, but Person C and Person D do not.

[40] Okay, the rock is being acted on by forces; gravity is pushing it down, and the earth is pushing it up. But the net force is zero.

The special Galilean group

Let's return to the problem of choosing reference frames. In addition to the three spatial ground rules established in Section 1.1, we now add the following:

4. The unit for time will be some predetermined length of time (for example, seconds).

5. The reference frame will be chosen to be inertial.

Next, we consider the transformations that would allow us to change between two different reference frames. Keep in mind that these transformations take place in *spacetime*, which is one dimension higher than in space[41]. The special Euclidean transformations of space remain as possibilities, but the following transformations involving time are also now possible:

[41] The formulas below are for two space dimensions, but the situation is similar in three or more dimensions.

- *Time-translations* shift the time values by a constant. They are of the form $(x, y, t) \mapsto (x, y, t + c)$. Time-translations are used to shift between the perspective of observers who choose different times for $t = 0$.

- *Galilean boosts* arise from the fact that two observers may be in motion relative to each other. Because both reference frames are assumed to be inertial, their relative velocity must be constant. The Galilean boosts are of the form $(x, y, t) \mapsto (x + v_x t, y + v_y t, t)$, where v_x and v_y are the components of the relative velocity.

When combined, the time-translations, spatial translations, rotations, and Galilean boosts are sufficient to relate any two reference frames that are chosen according to the above rules. The resulting

transformations are called *special Galilean transformations*. The general form of a special Galilean transformation is[42]

$$(\mathbf{x}, t) \mapsto (U\mathbf{x} + \mathbf{v}t + \mathbf{a}, t + c), \tag{1.12}$$

where U is a rotation matrix, $\mathbf{v}$ and $\mathbf{a}$ are vectors, and c is a constant.

[42] This vector form is valid in any number of dimensions.

Geometric invariants

As we did with the spatial transformations, we could consider what quantities are preserved by special Galilean transformations, thereby discovering the underlying geometric structure of spacetime according to Galileo and Newton. You can verify that the following are preserved by all special Galilean transformations.

- The difference in time $t_2 - t_1$ between any two points (x_1, y_1, t_1), (x_2, y_2, t_2) is preserved.
- For points with *the same time value* (x_1, y_1, t), (x_2, y_2, t), the distance $\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$ is preserved.
- Orientation of space is preserved at each time value.

But, as you will see in the exercises, there is even more. The point is that, even though the Galilean transformations seem fairly simple, the assertion that they form the "correct" group of transformations implies a not-so-simple geometry of space and time.

Infinitesimal Galilean transformations

The infinitesimal Galilean transformations can be described in terms of vector fields on spacetime. In two space dimensions, we introduce the following notation for some important infinitesimal transformations:

- infinitesimal translations $T_x = \frac{\partial}{\partial x}$ and $T_y = \frac{\partial}{\partial y}$,
- infinitesimal rotation $L = -y\frac{\partial}{\partial x} + x\frac{\partial}{\partial y}$,
- infinitesimal time-translation $H = \frac{\partial}{\partial t}$,
- infinitesimal Galilean boosts $C_x = t\frac{\partial}{\partial x}$ and $C_y = t\frac{\partial}{\partial y}$.

These six vector fields form a basis for all of the infinitesimal Galilean transformations.

Galilean relativity

The Galilean principle of relativity says that the universal laws of motion should be the same in every inertial frame. Mathematically, this means that the equations that describe the laws of motion should not change under Galilean transformations. We will explore this idea more carefully later, but for now consider the following examples:

- Newton's Law of Gravity says that the force of attraction between two objects depends only on the distance between the objects[43]. The force is assumed to be instantaneous, so it can be calculated at each time value independently. Since the Galilean transformations preserve the distance between two points at any given time, we can see that Newton's Law of Gravity does not change under Galilean transformations and is therefore compatible with Galilean relativity.

[43] and their masses, which aren't affected by any transformations.

- Coulomb's Law says that the electrostatic force of attraction/repulsion between two charged particles only depends on the distance between the particles[44]. As with Newton's Law, it is similarly compatible with Galilean relativity. However, it should be stressed that Coulomb's Law is only valid for particles that are not moving in relation to each other. When they are moving, we need to use more general laws that are *not* compatible with Galilean relativity. The discovery of these laws was the first step toward the development of special relativity.

[44] and their charges, which aren't affected by any transformations.

Exercises

1. Recall that special Galilean transformations preserve the following geometric properties: (1) the time difference $t_2 - t_1$ between two points, (2) the distance $\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$ between two points with the same time value, and (3) orientation at each time value.

   (a) Give an example of a transformation of spacetime that preserves all three of the above properties but is *not* a special Galilean transformation. [Hint: try something that is similar to a Galilean boost but with nonzero acceleration.] This demonstrates that these properties are not sufficient to completely characterize the special Galilean transformations.

   (b) Consider three spacetime points $(\mathbf{x}_1, t_1), (\mathbf{x}_2, t_2), (\mathbf{x}_3, t_3)$. Show that the quantity

   $$\|(t_2 - t_1)(\mathbf{x}_3 - \mathbf{x}_1) - (t_3 - t_1)(\mathbf{x}_2 - \mathbf{x}_1)\|$$

   is preserved by all Galilean transformations.

(c) Show that, if f is a transformation of spacetime that preserves time difference, orientation, and the quantity in part (b), then f is a special Galilean transformation.

(d) Find a geometric or physical interpretation of the quantity in (b). [Note: I'm not actually sure there's a good answer to this. A possible clue is that it looks vaguely similar to terms that appear in cross products.]

2. Find the flows of each of the vector fields H, C_x, and C_y. [Note; Since t is now being used as a variable in spacetime, you should use something else (such as s) as the flow parameter.]

3. (a) If you ignore the space and time translations, then the remaining part of a special Galilean transformation takes the form $(\mathbf{x}, t) \mapsto (U\mathbf{x} + \mathbf{v}t, t)$. This is linear in both $\mathbf{x}$ and t, so if we view $(\mathbf{x}, t)$ as an $(n+1)$-dimensional vector, then we can describe the map by an $(n+1) \times (n+1)$ matrix. What is this matrix? (It may help to first do it in the case $n = 2$.)

(b) Sometimes people use a fancy trick that allows you to view all the special Galilean transformations, including translations, as matrices. The idea is to introduce some additional parameter λ and view $(\mathbf{x}, t, \lambda)$ as an $(n+2)$-dimensional vector. Then the map $(\mathbf{x}, t, \lambda) \mapsto (U\mathbf{x} + \mathbf{v}t + \mathbf{a}\lambda, t + c\lambda, \lambda)$ is linear and can be described by an $(n+2) \times (n+2)$ matrix. Describe the corresponding matrix and show that multiplication of matrices correctly corresponds to composition of transformations.

4. One way to handle working in a non-inertial frame is to introduce so-called *fictional forces* to account for the acceleration of objects due to the non-inertial motion of the frame. Consider the situation of the rock lying on the ground. For each of the following observers, give a qualitative description of the fictional force that they would need to introduce.

(a) Person C, who is observing while on a pogo stick.

(b) Person D, who is observing while spinning around.

1.7 *Special relativity*

The invariance of the speed of light

At the turn of the 20th century, the properties and behavior of light were becoming well-understood. It had been known for some time[45] that the speed of light was finite, approximately 300,000,000 meters per second[46]. Maxwell's equations showed that electromagnetic

[45] Galileo suspected that the speed of light was finite, but he didn't successfully measure it. The first measurement of the speed of light was in 1676 by Olaus Römer. His measurement was not very accurate, but by the late 1800's, several very accurate measurements had been made.

[46] The exact value is 299,792,458 meters per second. This is actually true by fiat, since in 1983 the meter was redefined to be the distance traveled by light in 1/299,792,458 seconds.

waves move at a similar speed, leading him to suggest that light itself is an electromagnetic wave. And an 1887 experiment by Michelson and Morley seemed to suggest that the speed of light is a universal constant, *irrespective of the reference frame*.

If you've never encountered this last point before, you should find it pretty unbelievable[47], since it means that light doesn't behave like ordinary matter. Imagine that a lion is running toward you, and you are running away from the lion. The lion's speed relative to you will be slower than the lion's speed relative to a stationary observer[48].

If someone shines a laser at you, the situation is totally different. No matter how fast you run away from the beam, it will always approach you at 299,792,458 meters per second. You can't outrun light.

If we accept this behavior of light as fact, it creates a mathematical problem, since the speed of anything, including light, will change under a Galilean boost. So we are inevitably led to the conclusion that we're changing reference frames wrong. We need to start over and look for a new set of transformations that do not change the speed of light.

[47] In fact Michelson himself didn't believe it at first, thinking that something was wrong with the experiment.

[48] If you are Usain Bolt, the lion's speed relative to you may even be zero.

Worldlines

We are used to describing the motion of an object by a function $\mathbf{x}(t)$, giving the position of the object as a function of time. But there is another way to describe motion, using a curve in spacetime called a *worldline*. The worldline of an object consists of all the points $(\mathbf{x}(t), t)$, where t ranges through all possible time values.

Of course, it can be difficult to visualize a curve in 4-dimensional spacetime, but you can actually learn a lot from the simple version where space is one-dimensional. In this case, spacetime is the two-dimensional xt-plane, and the worldline of an object can actually be drawn as a curve in the plane. In fact, the worldline is just the graph of the position function $x(t)$, but drawn with the axes flipped[49].

[49] Actually, physicists usually put the t variable first, in which case two-dimensional spacetime would be the tx-plane, and the worldline would exactly be the graph of the position function.

Minkowski distance

Let's continue to work in one-dimensional space for simplicity. As is commonplace, we will use c to denote the speed of light.

Suppose that (x_1, t_1) and (x_2, t_2) are two points on the worldline of a pulse of light. Because light has the constant speed c, it will be the case that the change in position is $\pm c$ times the change in time[50]:

$$x_2 - x_1 = \pm c(t_2 - t_1).$$

[50] The sign depends on which direction the light is moving in.

The sign ambiguity is annoying, but we can fix it by squaring both

sides to get

$$(x_2 - x_1)^2 = c^2(t_2 - t_1)^2.$$

And we might as well move all the terms to one side to get

$$(x_2 - x_1)^2 - c^2(t_2 - t_1)^2 = 0. \tag{1.13}$$

Since the speed of light should be constant in all reference frames, we conclude that any coordinate transformation should preserve[51] the equation (1.13).

Now consider two arbitrary spacetime points (x_1, t_1) and (x_2, t_2). The square root of the quantity on the left side of (1.13) is called the *Minkowski distance*[52] between the points. When space is three-dimensional, the Minkowski distance between (x_1, y_1, z_1, t_1) and (x_2, y_2, z_2, t_2) is

$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2 - c^2(t_2 - t_1)^2}.$$

Or, in vector form, the Minkowski distance between $(\mathbf{x}_1, t_1)$ and $(\mathbf{x}_2, t_2)$ is

$$\sqrt{\|\mathbf{x}_2 - \mathbf{x}_1\|^2 - c^2(t_2 - t_1)^2}. \tag{1.14}$$

If $t_1 = t_2$, then the Minkowski distance coincides with the actual spatial distance between the two points. But in general, the Minkowski distance is strange, and rather than trying to understand it physically, you should probably think of it as a mathematical quantity that we naturally arrived at by thinking about the speed of light.

One bizarre aspect of the Minkowski distance is that it is possible for it to be imaginary[53]. Of course, if we're just thinking of it as a mathematical quantity, then that's not a problem[54]. But a pro tip is that you can generally avoid dealing with imaginary numbers by eliminating the square root whenever possible, instead doing all calculations using the square of the Minkowski distance.

[51] (What we mean by preserving the equation is that, if (x_1, t_1) and (x_2, t_2) are points in spacetime that satisfy (1.13), then their images under the transformation should also satisfy (1.13).

[52] Warning: Many authors use opposite signs in their definition of Minkowski distance, so be careful about this when using formulas from other resources.

[53] For example, consider two spacetime points with the same position but different times.

[54] There is actually a meaningful physical interpretation for whether the Minkowski distance is real vs. imaginary. You'll explore this in the exercises.

Lorentz boosts

We can interpret (1.13) as saying that any two points on the worldline of a pulse of light have Minkowski distance 0. This property will be maintained by any transformation of spacetime that preserves Minkowski distance, and therefore such transformations will preserve the speed of light. This leads us to the mathematical starting point of relativity: the "correct" set of transformations are those that preserve Minkowski distance. The set of such transformations is called the *Poincaré group*.

What do the elements of the Poincaré group look like? Well, there is some overlap with the Galilean group. In particular, you can check that the following transformations preserve Minkowski distance:

- translations
- time-translations
- rotations of space

The Galilean boosts, however, do not preserve Minkowski distance, so they are *not* included in the Poincaré group. Instead, there is another class of transformations, known as *Lorentz boosts*.

The Lorentz boost in the x direction with velocity v is of the form

$$(x, y, z, t) \mapsto \left(\gamma(x - vt), y, z, \gamma(t - vx/c^2)\right), \tag{1.15}$$

where $\gamma = \frac{1}{\sqrt{1-v^2/c^2}}$. There similarly exist Lorentz boosts in the y and z directions.

I claim that every transformation[55] that preserves Minkowski distance is a combination of translations, time-translations, rotations, and Lorentz boosts. A detailed proof of this claim would be beyond our scope, but in the next section we will prove an infinitesimal version.

[55] ignoring "discrete" transformations like reflections.

The principle of special relativity

Let's start with a bit of history and context[56]. As mentioned previously, the roots of special relativity can be traced back to Maxwell's theory of electromagnetism, which was published in 1865, and the Michelson-Morley experiment of 1887. It was known that neither Maxwell's equations nor the Michelson-Morley result are compatible with Galilean boosts, so there was an effort to make sense of all these seemingly contradictory ideas.

[56] This isn't intended to be a complete history. The goal is to get a general sense of how the ideas of relativity appeared.

Some glimpses of special relativity had emerged before Einstein. In 1889, George FitzGerald tried to explain the Michelson-Morley experiment by making the audacious suggestion that objects become shorter as their speed increases[57]. By early 1905, the Poincaré group had been discovered, and it was known to be compatible with both Maxwell's theory and the Michelson-Morley result.

[57] In 1892, Hendrik Lorentz, unaware of FitzGerald's paper, made the same suggestion. Unfortunately for FitzGerald, this phenomenon is now known as *Lorentz contraction*.

Still, it wasn't quite clear what role the Lorentz boosts should play in fundamental physics. People were still under the belief that there was an "ether" in which electromagnetic waves traveled. You could set an extra ground rule that your reference frame should be motionless with respect to the ether, which would eliminate the need for Galilean boosts; however, it would also eliminate the need for Lorentz boosts.

In late 1905, Einstein published a paper asserting that we could completely discard the idea of the ether and *interpret the Lorentz transformations as changes of reference frame between observers that are in*

relative motion to each other[58]. Just as Galilean relativity is the principle that the laws of physics should not change under Galilean transformations, Einstein's special relativity is the principle that the laws of physics should not change under the transformations in the Poincaré group.

In 1907 and 1909, Hermann Minkowski connected all the dots in a truly elegant way, introducing[59] the Minkowski metric and showing that it provides the underlying geometry of the Poincaré group.

[58] The v in (1.15) is the relative velocity between the two frames.

[59] Actually, the Minkowski metric had previously been studied as part of the Erlangen Program, but only as an abstract mathematical structure.

Some consequences of special relativity

As a mathematician, I'm tempted to just take the beautiful theory of special relativity and run with it. But physicists are obligated to compare their theories with what we can observe in the real world. Thus, it is important to remember that the Galilean theory seems to be correct for things that we experience in our quotidian lives. If special relativity is correct, then it should closely resemble the Galilean theory in these situations.

Luckily, this is the case. At speeds that are negligible in comparison to the speed of light, v/c is close to 0. As v/c approaches 0, the formula for a Lorentz boost (1.15) approaches the formula for a Galilean boost, since γ approaches 1 and vx/c^2 approaches 0.

In cases where v/c is measurably larger than 0, then special relativity makes some interesting and counterintuitive predictions about the universe. Here are a couple of them[60].

[60] It should be emphasized that the following results are purely mathematical consequences of the assumption that Lorentz boosts are the correct way to transform between moving reference frames.

- *The cosmic speed limit:* When $v \geq c$, the Lorentz boost formula (1.15) fails to be defined, because γ isn't defined. This is the mathematical manifestation of the fact that, in special relativity, it is not possible to travel faster than the speed of light.

  Of course, light travels *at* the speed of light, but the Lorentz boosts don't allow us to directly transform into the reference frame of light. Instead, we can look at what happens in the limit as v approaches c.

- *Lorentz contraction:* We will work in one spatial dimension[61]. Consider a rod of length L, with one end at $x = 0$ and the other end at $x = L$.

  And suppose that, in our reference frame, the rod is not moving. Then the worldlines of the ends of the rod can be drawn as vertical lines at $x = 0$ and $x = L$ in the xt-plane. The points on the worldlines are of the form $(0, a)$ and (L, a), where a ranges over all values.

  Now consider what happens when you transform spacetime by a

[61] This is just for notational simplicity; note that the Lorentz boost (1.15) in the x direction keeps y and z fixed.

Figure 1.7: Worldlines of the ends of a motionless rod of length L

Lorentz boost

$$(x,t) \mapsto (\gamma(x-vt), \gamma(t-vx/c^2)). \tag{1.16}$$

Under this transformation, the points $(0,a)$ get sent to $(-\gamma va, \gamma a)$, so these are the points in the new worldline. Observe that, for these points, the x-coordinate is always $-v$ times the t-coordinate. Therefore, the equation of the new worldline is $x = -vt$, or equivalently,

$$t = -\frac{1}{v}x. \tag{1.17}$$

Similarly, the points (L,a) get sent to $(\gamma(L-va), \gamma(a-va/c^2))$. The relationship between the x- and t-coordinates isn't so obvious, but with some algebraic jujitsu you can obtain the equation

$$t = -\frac{1}{v}x + \gamma L\left(\frac{1}{v} - \frac{v}{c^2}\right). \tag{1.18}$$

The equations for the new worldlines both have slope $-1/v$, which tells us that the rod now appears to be moving with velocity $-v$. This is what we expect, since it's the perspective of someone moving with velocity v.

Figure 1.8: Worldlines of the ends of the rod, transformed by a Lorentz boost

But here's the crazy part. Suppose an observer in the new reference frame observed the rod at time $t = 0$. One end of the rod is described by equation (1.17), so at $t = 0$ it is located at $x = 0$. The other end of the rod is described by equation (1.18), so at $t = 0$ it is located at $x = v\gamma L\left(\frac{1}{v} - \frac{v}{c^2}\right)$, which can be simplified[62] to $L\sqrt{1-\frac{v^2}{c^2}}$. So, to this observer, the length of the rod is $L\sqrt{1-\frac{v^2}{c^2}}$. This is less than L, since the scaling factor $\sqrt{1-\frac{v^2}{c^2}}$ is less than 1. The conclusion is that, in a frame where the rod is in motion, it appears shorter!

[62] Remember that $\gamma = \frac{1}{\sqrt{1-v^2/c^2}}$.

Exercises

1. (a) Let (x_0, t_0) be a point in spacetime, with one space dimension. Find an equation for the set of all points (x,t) whose Minkowski distance from (x_0, t_0) is zero. Give a qualitative geometric description of this set of points.

   (b) Do part (a) again, but with two space dimensions. Then think about three spatial dimensions.

2. What does it mean physically when the Minkowski distance between two points is real vs. imaginary? [Hint: Think about how fast you would need to travel to get from one point to the other.]

3. In one space dimension, verify that the Lorentz boost (1.16) preserves Minkowski distance.

4. Consider two points (x_0, t_1) and (x_0, t_2) with the same x-value but different time values. Apply the Lorentz boost (1.16) to these points, and observe that the difference in time between the two points is actually larger in the new reference frame. In other words, for an observer who is moving quickly, time passes more slowly. This effect is known as *time dilation*.

5. If one thing happens and then something else happens later, we might try to ascribe a causal relationship between the events, saying that the first event caused the second event to occur.

   (a) Consider a point (x_1, t_1) whose Minkowski distance from $(0, 0)$ is real and positive. Also assume that $t_1 > 0$, meaning that an event at (x_1, t_1) occurs *after* an event at $(0, 0)$. Show that there exists a Lorentz boost that maps (x_1, t_1) to a point with a negative time value. Physically, this means that different observers will disagree on which event occurred first. Therefore it is impossible for there to be a causal relationship between the events.

   (b) Show that, if (x_1, t_1) has imaginary Minkowski distance from $(0, 0)$, and if $t_1 > 0$, then every Lorentz boost maps (x_1, t_1) to a point with a positive time value. Thus, in this case, it is objectively true that an event at $(0, 0)$ occurs before an event at (x_1, t_1). It is then possible (though far from guaranteed) for a causal relationship to exist.

1.8 The Lorentz group

Recall that, when studying the geometry of space, we first decided that the transformations that preserve the geometry[63] are combinations of rotations and translations. Then we put the translations aside for a while and just focused on the rotations. We're now going to do the same thing for the Poincaré group.

[63] i.e. distance and orientation

The *Lorentz group* is defined to be the group of transformations that preserve Minkowski distance and fix the origin. The Lorentz group is a subgroup of the Poincaré group[64], consisting of the transformations in the Poincaré group that do not have a translational component. Ultimately, our goal is to understand the Poincaré group, but our plan will be to first understand the Lorentz group, and then bring translations back into the picture.

[64] A *subgroup* is a subset of a group that is also a group.

The Minkowski inner product

First let's introduce some new operations[65]. In what follows, I will use the capital letter $\mathbf{X}$ to denote a spacetime vector (so $\mathbf{X} = (\mathbf{x}, t)$). Although we are primarily thinking of three space dimensions, the vector notation allows for arbitrary n space dimensions.

[65] Although new, much of this should feel vaguely familiar. There are close parallels between Minkowski geometry and Euclidean geometry.

The *Minkowski norm* $\|\mathbf{X}\|_{\mathcal{M}}$ is defined to be the Minkowski distance between $\mathbf{X}$ and the origin. So, from (1.14), we have

$$\|\mathbf{X}\|_{\mathcal{M}} = \sqrt{\|\mathbf{x}\|^2 - c^2t^2}.$$

Intuitively, you should think of the Minkowski norm as measuring the length[66] of the vector $\mathbf{X}$. Since the transformations in the Lorentz group preserve Minkowski distance and fix the origin, they also preserve the Minkowski norm.

[66] But beware that the Minkowski norm can be zero for nonzero vectors, or even imaginary.

Given two vectors $\mathbf{X_1} = (\mathbf{x}_1, t_1), \mathbf{X_2} = (\mathbf{x}_2, t_2)$, their *Minkowski inner product* $\langle \mathbf{X}_1, \mathbf{X}_2 \rangle_{\mathcal{M}}$ is defined as

$$\langle \mathbf{X}_1, \mathbf{X}_2 \rangle_{\mathcal{M}} = \mathbf{x}_1 \cdot \mathbf{x}_2 - c^2 t_1 t_2.$$

There is a close analogy to the dot product, which we will continue to develop.

Just as the dot product is related to the length of a vector, the Minkowski inner product is related to the Minkowski norm. Specifically,

$$\|\mathbf{X}\|_{\mathcal{M}} = \sqrt{\langle \mathbf{X}, \mathbf{X} \rangle_{\mathcal{M}}}.$$

A less obvious but useful fact is that, conversely, the Minkowski inner product can be expressed in terms of the Minkowski norm[67]:

$$\langle \mathbf{X}_1, \mathbf{X}_2 \rangle_{\mathcal{M}} = \frac{1}{2} \left(\|\mathbf{X}_1 + \mathbf{X}_2\|^2_{\mathcal{M}} - \|\mathbf{X}_1\|^2_{\mathcal{M}} - \|\mathbf{X}_2\|^2_{\mathcal{M}} \right). \tag{1.19}$$

[67] Compare with Appendix A, Exercise 5. Because the Minkowski inner product is distributive, the proofs are identical.

Using (1.19), we can follow the same outline as Exercise 6 in Section 1.1 to prove that the Lorentz transformations must be linear. Thus every Lorentz tranformation is of the form $\mathbf{X} \mapsto U\mathbf{X}$ for some $(n+1) \times (n+1)$ matrix U. Preservation of the Minkowski inner product can then be expressed as

$$\langle U\mathbf{X}_1, U\mathbf{X}_2 \rangle_{\mathcal{M}} = \langle \mathbf{X}_1, \mathbf{X}_2 \rangle_{\mathcal{M}} \tag{1.20}$$

for all $\mathbf{X}_1, \mathbf{X}_2$.

In the exercises[68], you will show that U satisfies (1.20) if and only if $U^\intercal \eta U = \eta$, where (in the case $n = 3$)

[68] Note the similarity with Exercise 6 in Appendix A.

$$\eta = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -c^2 \end{bmatrix}. \tag{1.21}$$

You will also show that $U^\intercal \eta U = \eta$ if and only if the columns of U are "Minkowski orthonormal", meaning that

1. The Minkowski inner product of any two different columns is zero,
2. The Minkowski norm of each of the first three columns is 1,
3. The Minkowski norm of the last column is ic.

You'll also write out the Lorentz boosts as matrices and check that they satisfy one of the above equivalent conditions, thereby showing that the Lorentz boosts preserve the Minkowski distance.

Orientation and the direction of time

In this section, we'll focus on the (presumably) real-life case of three space dimensions, although the results can easily be adapted to other dimensions.

So far, we have found that every Lorentz transformations is given by a 4×4 matrix U such that $U^\intercal \eta U = \eta$, where η is the matrix in (1.21). If we take the determinant of both sides of this equation, we can cancel $\det \eta$ and obtain the equation $(\det U)^2 = 1$. Thus it is necessary that $\det U = \pm 1$. Some simple examples of Lorentz transformations with determinant -1 are

$$\begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \text{ and } \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}.$$

The first of these is a reflection of space, and the second one has the effect of reversing time. Neither of these describes the realistic perspective of an observer[69], so it makes sense to require the condition $\det U = 1$. This is an analogue of the orientation-preserving requirement for rotations.

[69] Maybe the observer is using a mirror? Or watching a video that's playing backward?

But now consider the matrix

$$\begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}.$$

This one reflects space *and* reverses time, which is still not very realistic. However, this matrix has determinant 1, so we haven't ruled it out yet. To rule it out, we need impose another requirement, but it will take a little bit of work to discover it.

Let U be a Lorentz transformation, and write

$$U = \begin{bmatrix} u_{11} & u_{12} & u_{13} & u_{14} \\ u_{21} & u_{22} & u_{23} & u_{24} \\ u_{31} & u_{32} & u_{33} & u_{34} \\ u_{41} & u_{42} & u_{43} & u_{44} \end{bmatrix}.$$

In principle, you could fully expand out the equation $U^\intercal \eta U = \eta$ in terms of the entries u_{ij}, but it would be a lot of work. However, it's not too bad to just calculate the lower right entry, and the resulting equation is

$$(u_{14})^2 + (u_{24})^2 + (u_{34})^2 - (u_{44})^2 c^2 = -c^2.$$

Moving the last term on the left to the right, we get

$$(u_{14})^2 + (u_{24})^2 + (u_{34})^2 = ((u_{44})^2 - 1)c^2.$$

The left side must be greater than or equal to 0, so the right side must be as well. This tells us that $(u_{44})^2 - 1 \geq 0$, so either $u_{44} \geq 1$ or $u_{44} \leq -1$.

If $u_{44} \leq -1$, then the transformation reverses time. As with transformations that reverse orientation, this is a situation we would like to avoid, so we will add the extra requirement that $u_{44} \geq 1$.

To summarize, the transformations that we are interested in are 4×4 matrices such that

- $U^\intercal \eta U = \eta$,
- $\det U = 1$,
- $u_{44} \geq 1$.

Such transformations are called *proper orthochronous Lorentz transformations*, but that's some pretty unwieldy terminology, so from here on out we will simply call them Lorentz transformations, with the modifiers "proper orthochronous" being implied.

Infinitesimal Lorentz transformations

Finding an explicit description of all solutions to the equation $U^\intercal \eta U = \eta$ is difficult, so it will be easier to instead consider the infinitesimal Lorentz transformations, in a similar way to how we considered infinitesimal rotations in Section 1.2.

Consider an s-dependent path[70] of Lorentz transformations $U(s)$ with initial point $U(0) = I$. If we differentiate the equation

$$U(s)^\intercal \eta U(s) = \eta,$$

[70] We're using s as the parameter now because t is a variable in spacetime.

with respect to s, we get

$$U'(s)^\intercal \eta U(s) + U(s)^\intercal \eta U'(s) = 0.$$

Evaluating at $s = 0$ and writing $A = U'(0)$, we get the infinitesimal condition

$$A^\intercal \eta + \eta A = 0. \tag{1.22}$$

Conversely, you can verify that, if A satisfies (1.22), then e^{As} is a Lorentz transformation for all s. Thus, (1.22) is *the* equation that characterizes infinitesimal Lorentz transformations.

Equation (1.22) is linear in the entries of A, so it can be completely solved fairly easily. In fact, there is a slightly fancy way that allows us to solve it with only a few calculations. Recognizing that $\eta^\intercal = \eta$, we can rewrite (1.22) as

$$(\eta A)^T = -\eta A,$$

which says that ηA is skew-symmetric. Therefore ηA must be of the form

$$\eta A = \begin{bmatrix} 0 & a_1 & a_2 & b_1 \\ -a_1 & 0 & a_3 & b_2 \\ -a_2 & -a_3 & 0 & b_3 \\ -b_1 & -b_2 & -b_3 & 0 \end{bmatrix},$$

and we can multiply by η^{-1} to solve for A:

$$A = \begin{bmatrix} 0 & a_1 & a_2 & b_1 \\ -a_1 & 0 & a_3 & b_2 \\ -a_2 & -a_3 & 0 & b_3 \\ b_1/c^2 & b_2/c^2 & b_3/c^2 & 0 \end{bmatrix}.$$

There are six free variables in A, so the space of solutions is six dimensional. A good choice of basis is

$$L_x = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad L_y = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad L_z = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

$$B_x = \begin{bmatrix} 0 & 0 & 0 & c \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \frac{1}{c} & 0 & 0 & 0 \end{bmatrix}, \quad B_y = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & c \\ 0 & 0 & 0 & 0 \\ 0 & \frac{1}{c} & 0 & 0 \end{bmatrix}, \quad B_z = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & c \\ 0 & 0 & \frac{1}{c} & 0 \end{bmatrix}.$$

The first three generate spatial rotations around each axis[71] You might expect the last three to generate Lorentz boosts in each direction, and they do, but it takes a little bit of work to see it. To do so, we need to exponentiate the matrices and see that their exponentials are Lorentz boosts.

[71] These are just the infinitesimal rotation matrices with an extra row and column of all zeros.

For simplicity, I'll actually do the calculation in one spatial dimension, which contains all of the essential steps of the calculation in three spatial dimensions. Consider the matrix

$$\beta = \begin{bmatrix} 0 & c \\ \frac{1}{c} & 0 \end{bmatrix}.$$

Observe that $\beta^2 = I$. Therefore $\beta^3 = \beta$, $\beta^4 = I$, etc. It follows that

$$\begin{aligned} e^{\beta s} &= I + \beta s + \frac{s^2}{2!} I + \frac{s^3}{3!}\beta + \frac{s^4}{4!} I + \dots \\ &= \left(1 + \frac{s^2}{2!} + \frac{s^4}{4!} + \dots\right) I + \left(s + \frac{s^3}{3!} + \frac{s^5}{5!} + \dots\right)\beta. \end{aligned}$$

The power series in parentheses look similar to the series for cosine and sine, except that the signs don't alternate. These are in fact the power series for *hyperbolic* cosine and sine. Thus,

$$e^{\beta s} = \cosh(s) I + \sinh(s)\beta = \begin{bmatrix} \cosh(s) & c\sinh(s) \\ \frac{1}{c}\sinh(s) & \cosh(s) \end{bmatrix}.$$

This is pretty cool[72], but it doesn't look anything like the formula for a Lorentz boost. To make it look more familiar, make the substitutions

$$\cosh(s) = \gamma = \frac{1}{\sqrt{1 - v^2/c^2}}, \quad \sinh(s) = -\gamma v/c = -\frac{v}{c\sqrt{1 - v^2/c^2}}.$$

This substitution is possible because it satisfies[73] the condition $\cosh^2(s) - \sinh^2(s) = 1$. We then have

$$e^{\beta s} = \begin{bmatrix} \gamma & -\gamma v \\ -\frac{\gamma v}{c^2} & \gamma \end{bmatrix}, \tag{1.23}$$

which is indeed the correct matrix for a Lorentz boost in one spatial dimension.

We have seen that the infinitesimal spatial rotations and infinitesimal Lorentz boosts form a basis for the infinitesimal Lorentz transformations. This means that, at least at the infinitesimal level, the Lorentz transformations are combinations of spatial rotations and Lorentz boosts. Since the transformations in the Poincaré group are combinations of Lorentz transformations and translations (both of space and time), we can conclude that the transformations in the Poincaré group are combinations of spatial rotations, Lorentz boosts, and translations.

[72] It's like a hyperbolic "rotation" matrix.

[73] This is similar to something you did in Exercise 5 of Section 1.1, where you noted that if $a^2 + b^2 = 1$, then there must be a θ for which $a = \cos\theta$ and $b = \sin\theta$.

Exercises

1. (a) Show that the Minkowski inner product can be expressed in terms of matrix multiplication via the equation

$$\langle \mathbf{X}_1, \mathbf{X}_2 \rangle_{\mathcal{M}} = \mathbf{X}_1^{\mathsf{T}} \eta \mathbf{X}_2,$$

where η is given in (1.21).

(b) Show that a 4×4 matrix U satisfies (1.20) for all $\mathbf{X}_1, \mathbf{X}_2$ if and only if $U^\intercal \eta U = \eta$.

(c) Show that $U^\intercal \eta U = \eta$ if and only if the columns of U are Minkowski orthonormal.

2. (a) Write down the matrix that corresponds to the Lorentz boost (1.15).

(b) Show that the Lorentz boost (1.15) preserves the Minkowski distance. [Hint: the easiest way to do this will be to check one of the conditions involving the matrix.]

3. Recall from (1.23) that, in one spatial dimension, the matrix form of a Lorentz boost with velocity v is[74]

$$U_v = \begin{bmatrix} \gamma & -\gamma v \\ -\frac{\gamma v}{c^2} & \gamma \end{bmatrix}.$$

Use this to find the group law for Lorentz boosts; in other words, calculate $U_{v_1} U_{v_2}$ and then find the value of v_3 such that $U_{v_1} U_{v_2} = U_{v_3}$. The formula you obtain describes how you are actually supposed to add velocity vectors in special relativity.

4. Now that we've studied the part of the Poincaré group consisting of transformations that fix the origin, we can use vector fields to describe infinitesimal transformations in the full Poincaré group. Write down vector fields that correspond to each of the matrices $L_x, L_y, L_z, B_x, B_y, B_z$, as well as each of the four translations. Together, these ten vector fields form a basis for the infinitesimal Poincaré group.

5. Consider the following "inner product" defined on $\mathbb{R}^2$:

$$\langle (x_1, y_1), (x_2, y_2) \rangle = x_1 y_2 + x_2 y_1.$$

Figure out as much as you can about the transformations that preserve this inner product and their infinitesimal counterparts.

[74] Don't forget that γ depends on v!

2 *Hamiltonian Mechanics*

Hamilton seemed to have some remarkable insight into what was important — one of the most remarkable insights, I suppose, which a mathematician has ever had. — Paul Dirac

So far, we haven't done any mechanics. We have only discussed the structure of space and time without considering how things move when acted on by forces. In this chapter, we will introduce one geometric formalism for mechanics, due to William Rowan Hamilton[1]. The key idea was to treat an object's momentum as an additional variable, independent of its position. This allowed him to express the equations of motion in a form that contained a great deal of symmetry, including transformations that mix position and momentum variables.

We will begin with a brief introduction to Newtonian mechanics and conservation laws. Then we will introduce Hamilton's equations and discover that we can actually rewrite the equations in terms of a dot-product-esque operation. We can think of this operation as being the geometric structure that underlies the Hamiltonian approach; any transformation that preserves the operation will preserve the mechanics of the system. This allows us to define *symmetries* of the system.

Finally, we will show that there is a correspondence between one-parameter families of symmetries and conserved quantities, thus obtaining a version of Noether's Theorem.

It's worth remarking that one of the main motivations for learning about Hamiltonian mechanics is as a precursor to quantum mechanics. The original framework for quantum mechanics begins with a Hamiltonian system and then uses certain rules to convert it into a quantum mechanical system[2]. This is still the way that students usually learn the subject.

[1] Hamilton is also the inventor of *quaternions*, which played an important role in the development of vector calculus.

[2] This is what Dirac was referring to in the above quote. Hamilton developed his framework in 1834, approximately 80 years before the development of quantum mechanics.

2.1 Newtonian mechanics

Newton's Second Law

Suppose you have an object moving around in n-dimensional space. The trajectory of the object can be described by a map $\mathbf{x} : \mathbb{R} \to \mathbb{R}^n$, where $\mathbf{x}(t)$ is the position of the object at time t. We can then define the velocity $\mathbf{v}(t) = \mathbf{x}'(t)$ and the acceleration $\mathbf{a}(t) = \mathbf{x}''(t)$.

Newton's Second Law says that force equals mass times acceleration. Let's unravel what this means precisely. First of all, at any given point in time, the force $\mathbf{F}$ is described by a vector. Since the force could possibly depend on many things which could be changing over time[3], we will also treat it as a function of time $\mathbf{F}(t)$. So Newton's Second Law is an equation of vectors:

[3] Later, we will make more specific assumptions about $\mathbf{F}$, but Newton's Laws are valid even for forces that are very erratic and bizarre.

$$\mathbf{F}(t) = m\mathbf{a}(t).$$

When we rewrite it as

$$\mathbf{F}(t) = m\mathbf{x}''(t), \tag{2.1}$$

we recognize it as a second-order differential equation. Given $\mathbf{F}$, m, an initial position $\mathbf{x}_0$, and an initial velocity $\mathbf{v}_0$, we can,at least in principle, solve this equation to determine $\mathbf{x}(t)$. Even if you've never taken physics, you almost certainly have solved this equation in your calculus classes for some simple examples such as projectiles and springs.

Momentum, energy, and conservation of energy

If you integrate force in different ways, you get other quantities that are interesting. For example, consider the integral of force over time[4]:

[4] Here we are assuming that $\mathbf{x}(t)$ is a solution of (2.1).

$$\int_{t_0}^{t_1} \mathbf{F}(t)dt = \int_{t_0}^{t_1} m\mathbf{x}''(t)dt = \left[m\mathbf{x}'(t)\right]_{t_0}^{t_1} = \left[m\mathbf{v}(t)\right]_{t_0}^{t_1}. \tag{2.2}$$

The quantity $m\mathbf{v}(t)$ that appears on the right of (2.2) is the *momentum*, denoted as $\mathbf{p}(t)$:

$$\mathbf{p}(t) = m\mathbf{v}(t).$$

Equation (2.2) says that the integral of force over time is equal to the change in momentum. Or, using the Fundamental Theorem of Calculus, we can equivalently say that $\mathbf{F}(t) = \mathbf{p}'(t)$.

Now let's do something different and consider the line integral[5] of

[5] You would have encountered line integrals in multivariable calculus. Many books denote them as $\int \mathbf{F} \cdot d\mathbf{r}$.

force over the path $\mathbf{x}(t)$:

$$\begin{aligned}\int_{t_0}^{t_1} \mathbf{F}(t)\cdot\mathbf{x}'(t)dt &= \int_{t_0}^{t_1} m\mathbf{x}''(t)\cdot\mathbf{x}'(t)dt \\ &= \left[\frac{1}{2}m\mathbf{x}'(t)\cdot\mathbf{x}'(t)\right]_{t_0}^{t_1} \\ &= \left[\frac{1}{2}m\|\mathbf{x}'(t)\|^2\right]_{t_0}^{t_1} \\ &= \left[\frac{1}{2}m\|\mathbf{v}(t)\|^2\right]_{t_0}^{t_1}\end{aligned} \tag{2.3}$$

The quantity $\frac{1}{2}m\|\mathbf{v}(t)\|^2$ is the *kinetic energy*, denoted as $T(t)$:

$$T(t) = \frac{1}{2}m\|\mathbf{v}(t)\|^2.$$

There are two good ways to think of equation (2.3). One is that the integral over time of the dot product of force and velocity is equal to the change in kinetic energy[6]. Using the Fundamental Theorem of Calculus, we can equivalently say that $\mathbf{F}\cdot\mathbf{v}(t) = T'(t)$.

The other possibility is to try to apply the Fundamental Theorem of Calculus (for line integrals) to (2.3). However, to do so we need to assume that $\mathbf{F}$ is a *conservative* vector field, i.e. that $\mathbf{F}(t) = -\nabla V(\mathbf{x}(t))$ for some function V on $\mathbb{R}^n$. In this case, V is called the *potential energy*. Then the FTC tells us that

$$\int_{t_0}^{t_1} \mathbf{F}(t)\cdot\mathbf{x}'(t)dt = -\int_{t_0}^{t_1} \nabla V(\mathbf{x}(t))\cdot\mathbf{x}'(t)dt = -\left[V(\mathbf{x}(t))\right]_{t_0}^{t_1}. \tag{2.4}$$

Putting (2.3) and (2.4) together, we deduce that

$$\left[T(t)\right]_{t_0}^{t_1} = -\left[V(\mathbf{x}(t))\right]_{t_0}^{t_1},$$

so

$$\left[T(t) + V(\mathbf{x}(t))\right]_{t_0}^{t_1} = 0.$$

In other words, the *total energy* $E(t) = T(t) + V(\mathbf{x}(t))$ does not change over time. This is the law of *conservation of energy*.

Remarks:

- We will henceforth always assume that we are dealing with forces that are conservative. In this case, it's probably better to write $\mathbf{F}(\mathbf{x}(t))$ to emphasize the fact that $\mathbf{F}$ only depends on the position $\mathbf{x}$, and that the time-dependence of $\mathbf{F}$ is only a consequence of the fact that $\mathbf{x}$ depends on t.
- The potential energy function V is only determined up to a constant, so the total energy E is also only determined up to a constant.

[6] This change is usually referred to as the *work* done by the force. In fact, this was probably the physical motivation that was given in your multivariable calculus class.

Conservation laws at work: some 1D examples

Even though we don't live in a one-dimensional universe, it's often possible to study a physical system as if it is one-dimensional. For example, when you throw a ball in the air, you can just consider the height. Or, if you are looking at the motion of a spring, then you can just consider the distance that the spring is stretched or compressed.

In the one-dimensional case, all of the quantities that were vectors in the previous section are one-dimensional vectors, so we can treat them as numbers. This simplifies notation somewhat; for example, the dot product of one-dimensional vectors is the same thing as multiplication of numbers. But also, in the one-dimensional case, the equation $\mathbf{F}(\mathbf{x}(t)) = -\nabla V(\mathbf{x}(t))$ becomes $F(x(t)) = -V'(x(t))$. So, as long as we know that $F(x(t))$ only depends on the position $x(t)$, we can obtain the potential V by just integrating:

$$V(x) = -\int F(x)dx.$$

Because this is an indefinite integral, there is a "$+C$" ambiguity. You should remember that potential energy isn't an absolute physically-measurable quantity, but just a useful mathematical device. For physically meaningful calculations, the constant shouldn't matter, and you can choose it to be anything you want.

Example 2.1.1 (Gravity for the earthbound). Consider the case of gravity near the surface of the earth. Let $x(t)$ denote the height of an object with mass m. You are surely well aware that things fall with constant acceleration $-g$. Since force is mass times acceleration, the force due to gravity is $F = -mg$, so the potential energy is $V = -\int Fdx = mgx$.

Of course, in this case you could easily solve the equation $F = ma$ to find the trajectory of the object, but it's worth noting that there are some questions you can answer *without* solving any differential equations, only using conservation of energy. For example, suppose I launch a ball from the ground with vertical velocity v_0. How high will the ball go? Well, at launch, the ball has kinetic energy $\frac{1}{2}mv_0^2$ and zero potential energy, and at the max height $x_{\max}$ it has zero kinetic energy and potential energy $mgx_{\max}$. Using conservation of energy, we get

$$\frac{1}{2}mv_0^2 = mgx_{\max},$$

so $x_{\max} = \frac{v_0^2}{2g}$.

Example 2.1.2 (Gravity for space-travelers). If you are going to very high altitudes, then you shouldn't consider gravitational force as being constant. Rather, you should use Newton's Law of Gravity,

which says that, if x is the distance from the center of the earth, then $F = -\frac{m\mu}{x^2}$ for some[7] constant μ.

The potential energy is $V = -\int F dx = -\frac{m\mu}{x}$. The negative sign might seem strange, but it's actually not a problem, since potential energy is just a useful mathematical device. Observe that, as x increases, V increases[8]. When x is very large, then V is close to zero.

If I launch a ball from the ground with vertical velocity v_0, then at launch the kinetic energy is $\frac{1}{2}mv_0^2$ and the potential energy is $-\frac{m\mu}{R}$, where R is the radius of the earth. If the ball were to reach a max height, then at that point it would have zero kinetic energy and potential energy $-\frac{m\mu}{x_{\max}}$, so the total energy would be negative.

However, if the initial velocity is such that $v_0^2 \geq \frac{2\mu}{R}$, then the total energy is positive, so there could be no max height. In this case, the ball would go off into outer space and never return. The velocity $v_0 = \sqrt{2\mu/R}$ where this occurs is called *escape velocity*.

[7] For those of you familiar with this stuff: $\mu = GM$, where G is the universal gravitational constant and M is the mass of the earth.

[8] in the sense of becoming less negative.

Some comments about conservation laws

In general, solving differential equations is hard[9]. The differential equations you solve in your calculus classes are very simple examples[10], and when you consider situations that are a little bit more complex, you are likely to find yourself with a differential equation that is impossible to solve.

Conservation laws can allow you to simplify, and in some cases, completely solve a system, using only algebra. If your goal is to avoid solving differential equations, then conservation laws are your friends.

In the above one-dimensional examples, we made use of conservation of energy. For other systems, you may need other conservation laws. For example, to derive the properties of planetary motion, you will also need conservation of angular momentum. In fact, the reason we were able to treat real-life problems as one-dimensional was because we had already internalized various conservation laws. Example 2.1.1 was possible because momentum is conserved in the horizontal directions. Example 2.1.2 was possible because angular momentum is conserved.

It is possible to derive conservation laws from Newton's Second Law on a case-by-case basis. For example, consider angular momentum, which is denoted $\mathbf{J}$ and defined as

$$\mathbf{J}(t) := \mathbf{x}(t) \times \mathbf{p}(t).$$

What it means for angular momentum to be conserved is that $\mathbf{J}(t)$ is constant. Of course, this is true if and only if $\mathbf{J}'(t) = 0$. So the idea is to calculate $\mathbf{J}'(t)$ and see what needs to be true in order for

[9] Don't forget this!

[10] Luckily, the simple examples have important applications modeling things like exponential growth and harmonic motion.

everything to cancel. Let's do it[11]:

$$\mathbf{J}'(t) = \mathbf{x}'(t) \times \mathbf{p}(t) + \mathbf{x}(t) \times \mathbf{p}'(t).$$

In the first term on the right, substitute $\mathbf{p}(t) = m\mathbf{x}'(t)$, and in the second term, substitute $\mathbf{F} = \mathbf{p}'(t)$, which we derived from (2.2) using Newton's Second Law. This gives us

$$\mathbf{J}'(t) = \mathbf{x}'(t) \times m\mathbf{x}'(t) + \mathbf{x}(t) \times \mathbf{F}.$$

The first term $\mathbf{x}'(t) \times m\mathbf{x}'(t)$ is zero because the cross product of parallel vectors is always zero. Similarly, if we know that $\mathbf{F}$ is parallel to $\mathbf{x}(t)$, then the second term is zero[12]. Thus we can conclude that angular momentum is conserved whenever the force is always parallel to the position vector.

So that works, and we can reverse-engineer an intuitive explanation. When the force points is parallel to the position vector, then the force is pointing in a "radial" direction, either directly toward or away from the origin, and not in an "angular" direction. It seems reasonable to guess that such a force would not cause angular momentum to change.

But there are two downsides to this approach. One is that the calculation is very ad-hoc. We needed to make specific substitutions and use a special property of the cross product to obtain the above result. The other downside is that, to get to this point, we needed to already have a suspicion that angular momentum *might* be conserved; otherwise, we wouldn't have had any reason to do the calculation in the first place. What if I asked you to find other conserved quantities? Where would you start? The Hamiltonian approach will resolve both of these issues.

[11] You may want to convince yourself that the product rule works for cross products.

[12] This is the case for planetary motion according to Newton's Law of Gravity, since the force points toward the origin, in exactly the opposite direction to $\mathbf{x}(t)$.

Exercises

1. Suppose I launch a ball from the ground[13] with vertical velocity v_0. How fast is the ball moving when the ball is halfway up (i.e. when $x = x_{\max}/2$)?

2. An object attached to a spring behaves approximately as if it is under a force given by *Hooke's Law*: $F = -kx$, where k is a constant and x is the position of the object relative to equilibrium.

   (a) What is the potential energy V corresponding to the force F?

   (b) Note that the formula for the total energy E only involves $x(t)$ and $x'(t)$, not $x''(t)$. Starting with the formula for E, solve for $x'(t)$ to obtain a first-order differential equation. You don't need to solve the differential equation[14].

[13] Assume that the initial velocity is small enough that the "earthbound" model for gravity is valid.

[14] In principle, you can then use separation of variables to solve for x, but the integral is a bit gnarly to do by hand; it requires a trig substitution. But if you want, you can ask Wolfram Alpha to do it for you.

3. It could be possible for there to be a force such that angular momentum is not conserved in its entirety, but only one component of angular momentum is conserved. Find a condition on $\mathbf{F}$ that implies that the x-component of angular momentum is conserved.

4. (a) Find a condition on $\mathbf{F}$ that implies that the x-component of momentum is conserved.

   (b) What can you way about $\mathbf{F}$ if *all* the components of momentum are conserved?

2.2 *Hamilton's equations*

Rewriting second-order equations as first-order equations

As we have seen, Newton's Second Law $\mathbf{F} = m\mathbf{a}$ is a second-order differential equation. But there is a trick that allows you to convert a second-order equation into a system of first-order equations. I will describe the trick below, but first let's think about why you would want to do this.

One reason is that solving differential equations is hard[15]. For systems of first-order equations, there are numerical methods that can give you very good approximate solutions, so this trick allows you to apply those methods to second-order equations as well.

[15] Did you forget this?

Another reason, more relevant for us, is that systems of first-order equations can be seen as vector fields[16]. This gives us a geometric interpretation of the equations, which can be helpful in studying them.

[16] See Section 1.3.

The way the trick works is that you introduce the velocity $\mathbf{v}(t)$ as if it is an independent quantity, not *a priori* related to the position $\mathbf{x}(t)$. Then Newton's Second Law can be expressed as $\mathbf{F}(\mathbf{x}(t)) = m\mathbf{v}'(t)$. And then we add the relationship $\mathbf{v}(t) = \mathbf{x}'(t)$ as an additional equation. Together, this gives us a system of first-order equations:

$$\begin{aligned} \mathbf{x}'(t) &= \mathbf{v}(t), \\ \mathbf{v}'(t) &= \frac{1}{m}\mathbf{F}(\mathbf{x}(t)). \end{aligned} \tag{2.5}$$

The cost of this trick is that it doubles the number of variables. Counting each vector component individually, we started with a system of n second-order equations, and we ended up with a system of $2n$ first-order equations. But we'll see that it's worth the cost.

Hamilton's approach

To arrive at Hamilton's equations, we actually need to rewrite the system (2.5) in terms of position and momentum, rather than position

and velocity. By substituting $\mathbf{p}(t) = m\mathbf{v}(t)$ into (2.5), we get

$$\begin{aligned} \mathbf{x}'(t) &= \frac{1}{m}\mathbf{p}(t), \\ \mathbf{p}'(t) &= \mathbf{F}(\mathbf{x}(t)). \end{aligned} \tag{2.6}$$

This might not appear to be much of an improvement over the previous system, but we'll soon see that it makes a surprising difference.

For now, let's keep things simple and pretend we're in the one-dimensional situation. In terms of p, the kinetic energy of the object is $p^2/2m$. We will again assume that the force is conservative[17], so there is a potential energy function $V(x)$ such that $F(x) = -V'(x)$.

[17] In the one-dimensional case, the force is automatically conservative once we know that it only depends on the position x. But in higher dimensions, the equation $F = -\nabla V$ is a significant assumption.

Let $H(x, p)$ denote the total energy, viewed as a function of the variables x and p:

$$H(x, p) = \frac{p^2}{2m} + V(x).$$

This is called the *Hamiltonian* function.

Observe that $\frac{\partial H}{\partial p} = \frac{p}{m}$ and $\frac{\partial H}{\partial x} = V'(x) = -F$. So the system (2.6) can be written in the following intriguing form:

$$\begin{aligned} x'(t) &= \frac{\partial H}{\partial p}, \\ p'(t) &= -\frac{\partial H}{\partial x}. \end{aligned} \tag{2.7}$$

These are known as *Hamilton's equations*. In the general n-dimensional case, you can write out the components of position and momentum, $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{p} = (p_1, \ldots, p_n)$, and Hamilton's equations are

$$\begin{aligned} x_i'(t) &= \frac{\partial H}{\partial p_i}, \\ p_i'(t) &= -\frac{\partial H}{\partial x_i}. \end{aligned} \tag{2.8}$$

Hamilton's equations and vector fields

The $2n$-dimensional space with coordinates $(x_1, \ldots, x_n, p_1, \ldots, p_n)$ is called *phase space*. So a point in phase space contains the information about the position *and* the momentum of the physical system. It's worth noting that the realistic examples aren't limited to the cases where $n \leq 3$. For example, suppose that you are studying the physical system consisting of the sun, the earth, and the moon. Then you would probably choose a reference frame with the sun at the origin, use three variables to describe the position of the earth and another three variables to describe the position of the moon. So in this case, you would have $n = 6$.

For simplicity[18], we'll stick to the case $n = 1$ here and in some

[18] Mostly so we are able to actually draw pictures of what's going on.

future sections, but at a theoretical level everything we'll do will work in any number of dimensions.

Recall that a system of first-order equations such as (2.7) can be visualized as a vector field on the xp-plane (aka phase space). By plotting this vector field, we can get an idea of how the system behaves. The flow of the vector field determines how the position and momentum change over time.

For example, in the case of gravity near the surface of the earth (see Example 2.1.1), the Hamiltonian function is

$$H = \frac{p^2}{2m} + mgx,$$

so Hamilton's equations are

$$\begin{aligned} x' &= \frac{p}{m}, \\ p' &= -mg. \end{aligned}$$

This system corresponds to the vector field $(p/m, -mg)$. Figure 2.1 shows this vector field with $m = 1$ and $g = 9.8$. Suppose I throw a 1 kg ball straight up with initial momentum 9 kg m/s and initial height 0. What will the max height of the ball be? You can get a rough estimate by starting at the point $(0, 9)$ on the graph and following the arrows until you get to the x-axis. To me, it looks like this occurs somewhere between $x = 4$ and $x = 5$; the actual value is about 4.1.

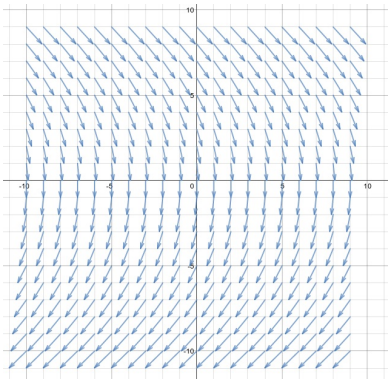

Figure 2.1: The vector field $(p, -9.8)$.

The general formula for the vector field corresponding to Hamilton's equations (in the case $n = 1$) is

$$V_H = \left(\frac{\partial H}{\partial p}, -\frac{\partial H}{\partial x} \right).$$

For arbitrary n, the formula is[19]

$$V_H = \left(\frac{\partial H}{\partial p_1}, \ldots, \frac{\partial H}{\partial p_n}, -\frac{\partial H}{\partial x_1}, \ldots, -\frac{\partial H}{\partial x_n} \right).$$

The vector field V_H should be reminiscent of the gradient of H, in the sense that it's a vector field made up out of the partial derivatives. But the partial derivatives are in the wrong order and there are minus signs, so it isn't actually the gradient. The key to understanding Hamiltonian mechanics lies in understanding V_H as a "sibling" to the gradient[20].

[19] This formula comes from the right side of (2.8).

[20] Siblings are similar in some ways, but different in others!

Exercises

1. Consider an object of mass m attached to a spring. (Recall that in this situation, the force is $F = -kx$).

(a) Write out Hamilton's equations and graph the corresponding vector field for some arbitrary values of m and k. (You may need to try a few different sets of values to get the most useful-looking graph.) From looking at the graph, make an educated guess about what type of curve the solution curves are.

(b) In the case where $m = 1$ and $k = 1$, the vector field should look familiar. What is it? What are the solution curves in this case?

(c) Because of conservation of energy, we know that the Hamiltonian function H should stay constant along every solution curve. What types of curves are described by equations of the form $H = c$ in this case?[21]

[21] The moral here is that, at least in this case, conservation of energy allows you to find the solution curves without solving differential equations.

2. Suppose that an object of mass m is under a repulsive linear force $F = kx$ for some constant k.

(a) Write out Hamilton's equations and graph the corresponding vector field for some arbitrary values of m and k. From looking at the graph, make an educated guess about what type of curve the solution curves are.

(b) What types of curves are described by equations of the form $H = c$ in this case?

3. (a) Consider a planet orbiting a star. In an inertial reference frame with the star at the origin, the force is $\mathbf{F} = -\frac{GmM}{\|\mathbf{x}\|^3}\mathbf{x}$, where $\mathbf{x}$ is the three-dimensional vector that describes the position of the planet, G is Newton's gravitation constant, M is the mass of the star, and m is the mass of the planet. Write down Hamilton's equations in this case.

(b) Now consider two planets, with respective masses m_1 and m_2, orbiting a star. For each planet, the force attracting them to the star is given by a formula like in part (a). But there is also a gravitational force attracting the two planets to each other. Write down Hamilton's equations in this case. [Hint: forces are additive, so potential energies are additive as well. Calculate the potential energy for each planet-star and planet-planet pair, and sum them to get the total potential energy of the system.]

2.3 *The geometry of Hamilton's equations*

The gradient and the dot product

Recall that, for a function $f(x, y)$, the gradient ∇f is a vector field:

$$\nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right).$$

So we can think of ∇ as being an operation that takes a function as input and produces a vector field as output.

But also recall that vector fields can act as derivations on functions. Using this point of view, we could use "derivation notation" and write

$$\nabla f = \frac{\partial f}{\partial x}\frac{\partial}{\partial x} + \frac{\partial f}{\partial y}\frac{\partial}{\partial y}.$$

Given another function $g(x, y)$, the action of ∇f on g as a derivation is

$$\nabla f(g) = \frac{\partial f}{\partial x}\frac{\partial g}{\partial x} + \frac{\partial f}{\partial y}\frac{\partial g}{\partial y} = \nabla f \cdot \nabla g. \tag{2.9}$$

This demonstrates that the gradient is closely connected to the dot product, which in turn[22] is closely connected to the geometry of the plane.

[22] Remember that the dot product allows you to compute quantities like length, distance, and angles.

The dot product is symmetric, so $\nabla f \cdot \nabla g = \nabla g \cdot \nabla f$. Combining this fact with (2.9) gives us an interesting result: $\nabla f(g) = \nabla g(f)$. That is, the directional derivative of g in the direction of ∇f is equal to the directional derivative of f in the direction of ∇g.

The symplectic pairing

Let's now consider the not-exactly-the-gradient operation suggested by Hamilton's equations. Given a function $f(x, p)$ on phase space, define the *Hamiltonian vector field* of f to be

$$V_f := \left(\frac{\partial f}{\partial p}, -\frac{\partial f}{\partial x}\right). \tag{2.10}$$

Like the gradient, we can think of V as being an operation that takes a function as input and produces a vector field as output.

Using derivation notation, we can also write

$$V_f = \frac{\partial f}{\partial p}\frac{\partial}{\partial x} - \frac{\partial f}{\partial x}\frac{\partial}{\partial p}.$$

Given another function $g(x, p)$, the action of V_f on g as a derivation is

$$V_f(g) = \frac{\partial f}{\partial p}\frac{\partial g}{\partial x} - \frac{\partial f}{\partial x}\frac{\partial g}{\partial p}. \tag{2.11}$$

In analogy with (2.9), we would like to be able to interpret (2.11) as something like a dot product of V_f and V_g. However, it isn't exactly the dot product, so this leads us to define a new not-exactly-the-dot-product operation for vectors in phase space:

$$\omega((a_1, a_2), (b_1, b_2)) = a_2 b_1 - a_1 b_2. \tag{2.12}$$

The operation ω in (2.12) is called the *symplectic pairing*. With it, we can enhance (2.11) to make it look very similar to (2.9):

$$V_f(g) = \frac{\partial f}{\partial p}\frac{\partial g}{\partial x} - \frac{\partial f}{\partial x}\frac{\partial g}{\partial p} = \omega(V_f, V_g). \tag{2.13}$$

Just as (2.9) reveals a close relationship between the gradient and the dot product, (2.13) reveals a close relationship between Hamiltonian vector fields and the symplectic pairing. Just as the dot product is the key to geometry of the plane, the symplectic pairing is the key to geometry of phase space.

The symplectic pairing is skew-symmetric[23]:

$$\omega(\mathbf{a}, \mathbf{b}) = -\omega(\mathbf{b}, \mathbf{a}).$$

[23] This is one of the main ways that the symplectic pairing is different from the dot product.

Combining this fact with (2.13) gives us the result

$$V_f(g) = -V_g(f). \tag{2.14}$$

That is, the directional derivative of g in the direction of V_f is *minus* the directional derivative of f in the direction of V_g. We will better understand the significance of this fact in the next section.

Because you'll need it later, here is the formula for Hamiltonian vector fields in n spatial dimensions:

$$V_f := \left(\frac{\partial f}{\partial p_1}, \dots \frac{\partial f}{\partial p_n}, -\frac{\partial f}{\partial x_1}, \dots, -\frac{\partial f}{\partial x_n} \right). \tag{2.15}$$

Then

$$V_f(g) = \sum_{i=1}^{n} \frac{\partial f}{\partial p_i}\frac{\partial g}{\partial x_i} - \frac{\partial f}{\partial x_i}\frac{\partial g}{\partial p_i}.$$

The general formula for the symplectic pairing is

$$\omega((a_1, \dots, a_{2n}), (b_1, \dots, b_{2n})) = a_{n+1}b_1 + \cdots + a_{2n}b_n - a_1 b_{n+1} - \cdots - a_n b_{2n}.$$

This formula is a bit unwieldy, so it's convenient to rewrite it in matrix form. Define the $2n \times 2n$ matrix[24]

$$\Omega := \begin{bmatrix} 0 & -I \\ I & 0 \end{bmatrix} \tag{2.16}$$

[24] This is written in block form, where each block is $n \times n$.

Then we have[25]

$$\omega(\mathbf{a}, \mathbf{b}) = \mathbf{a}^\mathsf{T} \Omega \mathbf{b}. \tag{2.17}$$

[25] For example, when $n = 1$ we have

$$\omega((a_1, a_2), (b_1, b_2)) = \begin{bmatrix} a_1 & a_2 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}.$$

Try expanding this out and check that it's true!

Conservation laws in Hamiltonian mechanics

Let's remember how the math in the previous section relates to mechanics. The setup of Hamiltonian mechanics is that you have a special function H, called the Hamiltonian function, which describes the energy of the system as a function of the phase space variables $(x_1, \dots, x_n, p_1, \dots, p_n)$. You can then find the vector field V_H, and the flows of this vector field[26] describe how the system changes over time.

[26] Yes, this vector field is the Hamiltonian vector field of the Hamiltonian function. Don't blame me for the terminology.

To be more specific, suppose (in the case $n = 1$) that we know the initial position x_0 and initial momentum p_0 of an object, and we want

to know its position and momentum at some later time t. To find it, we would start at (x_0, p_0) and flow along V_H for time t. This could involve solving the system of differential equations corresponding to V_H, using (x_0, p_0) as the initial values. Alternatively, we could find the flow $\phi_t(x, p)$ of the vector field V_H, and then $\phi_t(x_0, p_0)$ would tell us the position and momentum at time t.

In general, a function f of phase space can be thought of as some quantity that you could measure or calculate about the system; for example, $f(x, p) = x$ is the position and $f(x, p) = \frac{p^2}{2m}$ is the kinetic energy. Note that any function is allowed[27]. For example, $f(x, p) = xp$ is a perfectly good quantity, even though it doesn't have a commonly used name.

[27] The math even allows for physically meaningless functions such as $f(x, p) = x + p$. Do you know why this is physically meaningless?

Consider $V_H(f)$, i.e. the directional derivative of f in the direction of V_H. Since the flow of V_H describes how the system changes over time, it follows that $V_H(f)$ tells us how f is changing over time. Specifically,

$$\begin{aligned}\frac{d}{dt}\left[f(x(t), p(t))\right] &= \frac{\partial f}{\partial x}x'(t) + \frac{\partial f}{\partial p}p'(t) \text{ by the chain rule} \\ &= \frac{\partial f}{\partial x}\frac{\partial H}{\partial p} - \frac{\partial f}{\partial p}\frac{\partial H}{\partial x} \text{ by Hamilton's equations} \\ &= V_H(f).\end{aligned} \tag{2.18}$$

In light of equation (2.18), we can now state a general conservation law for Hamiltonian mechanics: If $V_H(f) = 0$, then the quantity measured by f is not changing with respect to time, i.e. it is *conserved*. Furthermore, this statement becomes "if and only if" if we stipulate that conservation means the quantity is conserved for all possible states of the system.

In particular, the skew-symmetry property (2.14) implies that $V_H(H) = 0$, so we can conclude that energy is conserved in all cases where Hamiltonian mechanics applies; *conservation of energy is "built in" to the theory.*

As a concrete example, consider again the case of gravity near the surface of the earth, but now with all three spatial coordinates x, y, z, where z denotes the height above the ground. We'll use p_x, p_y, p_z to denote the corresponding components of momentum. The Hamiltonian is then

$$H = \frac{1}{2m}(p_x^2 + p_y^2 + p_z^2) + mgz, \tag{2.19}$$

and the corresponding Hamiltonian vector field is

$$V_H = \frac{p_x}{m}\frac{\partial}{\partial x} + \frac{p_y}{m}\frac{\partial}{\partial y} + \frac{p_z}{m}\frac{\partial}{\partial z} - mg\frac{\partial}{\partial p_z}.$$

Since V_H doesn't have $\frac{\partial}{\partial p_x}$ or $\frac{\partial}{\partial p_y}$ terms we can immediately see that $V_H(p_x) = 0$ and $V_H(p_y) = 0$, which tells us that momentum is

conserved in the x and y directions. But in this case there are other less-obvious conserved quantities. For example, you can check that $V_H(xp_y - yp_x) = 0$. This shows that the function $xp_y - yp_x$, which is the z-component of angular momentum, is conserved.

Let's summarize where we are so far. The situation is now better than it was when we were deriving conservation laws directly from Newton's Law. Whereas the calculations there were ad-hoc, depending on the quantity we were considering, we now have one simple equation that we can use to check whether a quantity is conserved. But we still don't have a good way to *discover* conserved quantities, other than by trial and error. In the above example, how did I know that the z-component of angular momentum would work? Was it just a lucky guess? We will turn our attention to this issue next.

Symmetries and conservation laws

At this point, you might think that V_H tells us every thing we need to know about the dynamics of a physical system. But we defined the notion of Hamiltonian vector field V_f more generally, where f could be any function. Why did we do that? Does V_f have any meaning when f is another function besides H? Let's look at some simple (but important) examples:

- In one spatial dimension, consider the momentum function p. If you plug $f(x,p) = p$ into (2.10), then you get the constant vector field $V_p = (1,0)$. The flows of V_p are translations in the x direction, so we say that V_p *generates* translations in the x direction. More generally, in higher dimensions, V_{p_i} is the vector field that generates translations in the x_i direction.

- When there are two spatial dimensions, then phase space has coordinates (x_1, x_2, p_1, p_2), and the formula for the Hamiltonian vector field is

$$V_f = \left(\frac{\partial f}{\partial p_1}, \frac{\partial f}{\partial p_2}, -\frac{\partial f}{\partial x_1}, -\frac{\partial f}{\partial x_2}\right).$$

  The angular momentum function is $J = x_1p_2 - x_2p_1$, and its Hamiltonian vector field is

$$V_J = (-x_2, x_1, -p_2, p_1).$$

  This is a linear vector field that generates rotations of the x_1x_2-plane while simultaneously rotating the p_1p_2-plane.

- In three spatial dimensions, then the first component of angular momentum is $J_1 = x_2p_3 - x_3p_2$. Its Hamiltonian vector field $V_{J_1} = (0, -x_3, x_2, 0, -p_3, p_2)$ simultaneously generates rotations

of $x_1x_2x_3$-space around the x_1-axis and rotations of $p_1p_2p_3$-space around the p_1-axis. The other components of angular momentum similarly generate rotations around the other axes.

These examples demonstrate that, in a similar way to how H generates transformations that describe how the system changes over time, other functions generate other transformations that may have useful geometric interpretations.

In the last section, we saw that the quantity measured by a function f is conserved if and only if $V_H(f) = 0$. Using the skew-symmetry property (2.14), we can equivalently say that f is conserved if and only if $V_f(H) = 0$. The equation $V_f(H) = 0$ says that the rate of change of H is 0 in the direction of the vector field V_f, which is equivalent to saying that the value of H should be constant along the flows of V_f. In other words, H should be preserved by the transformations[28] generated by f.

[28] See Exercise 5 in Section 1.5.

The chain of ideas we just went through is really important, so let me rewrite it in a more linear form:

$$\begin{aligned} f \text{ is conserved} &\iff V_H(f) = 0 \text{ by (2.18)} \\ &\iff V_f(H) = 0 \text{ by (2.14)} \\ &\iff H \text{ is preserved by the flows of } V_f \end{aligned}$$

Note that the key step in this argument utilizes the skew-symmetry property (2.14).

You can start to see a version of Noether's Theorem coming into focus here. Any conserved quantity f corresponds to a one-parameter group (the flow of V_f) that preserves H. Looking back at the example of gravity near the surface of the earth, where the formula for H is (2.19), we can see this correspondence in the following instances:

- p_x is conserved because the flows of V_{p_x} are translations in the x direction, and H is preserved by such translations because H does not depend on x. Similarly, p_y is conserved because H does not depend on y.
- The z-component of angular momentum, $J_z = xp_y - yp_x$, is conserved because the flows of V_{J_z} simultaneously rotate xyz-space about the z-axis and $p_xp_yp_z$-space about the p_z-axis, and H does not change under such a transformation.

This is cool! Conserved quantities correspond to symmetries! But there still remains a difficulty. Let's think about how you would generally want to use the result in practice. Typically, you would know the Hamiltonian H, and by studying its formula, you might

recognize that H is preserved by some one-parameter group of transformations. For example, you can look at the Hamiltonian (2.19), and you can see that it depends on exactly two quantities: the magnitude of the momentum vector (p_x, p_y, p_z) and the height z. You can then begin to imagine different types of transformations that preserve both of those quantities. Maybe you could rotate the momentum vector and/or do something that changes the x and y values but keeps z unchanged.

But the catch is that, to apply the above correspondence, you need to find functions whose Hamiltonian vector fields generate the transformations that you're interested in. To be concrete, let's say that you wanted to consider the set of transformations that rotate momentum vectors around the p_z-axis while keeping the position variables fixed. That seems to be a perfectly good 1-parameter group of transformations that preserve the Hamiltonian, so how can we find an f such that V_f generates them? Wait, do we even know that such an f exists? We'll figure out how to answer these questions in the next section.

Exercises

1. In this problem, you'll investigate one important way that the Hamiltonian vector field is different from the gradient.

   (a) Use (2.9) to show that $\nabla f(f)$ is positive when $\nabla f \neq 0$. This fact can be interpreted as saying that ∇f always points in a direction where f is increasing (which is something you should remember from multivariable calculus).

   (b) What can you say about the sign of $V_f(f)$? What does your answer mean in terms of how f is changing in the direction of V_f?

   (c) Pick a couple of functions $f(x, p)$ and, for each one, use a computer to plot the vector field V_f along with a few level curves of f. Observe that the vector field and the level curves are geometrically related in a way that agrees with your answer to (b).

2. (a) We'll say that a $2n \times 2n$ matrix A preserves the symplectic pairing if $\omega(A\mathbf{a}, A\mathbf{b}) = \omega(\mathbf{a}, \mathbf{b})$ for all vectors $\mathbf{a}$ and $\mathbf{b}$. Show that A preserves the symplectic pairing if and only if $A^\intercal \Omega A = \Omega$. [Hint: use (2.17).]

   (b) Suppose that A is of the block form

$$A = \begin{bmatrix} B & 0 \\ 0 & (B^{-1})^\intercal \end{bmatrix},$$

where B is any invertible $n \times n$ matrix. Show that A preserves the symplectic pairing. In particular, observe that, if B is an orthogonal matrix, then A looks particularly nice because $(B^{-1})^\intercal = B$.

3. (a) Show that, if A preserves the symplectic pairing, then[29] $\det A = \pm 1$.

   (b) Check that, in the case $n = 1$, A preserves the symplectic pairing if and only if $\det A = 1$.

   (c) Find an example of a 4×4 matrix A such that $\det A = 1$ but A does *not* preserve the symplectic pairing. This demonstrates that the converse of the statement in part (a) is not true when $n > 1$.

4. For the case of a planet orbiting a star (see problem 3 in the previous section), show that all components of angular momentum are conserved.

[29] It actually turns out that $\det A = 1$, but that is more difficult to prove.

2.4 *Symplectic transformations*

In the last section, we found that there is a correspondence between symmetries and conserved quantities, but there is something that we still need to figure out. The symmetries that appear in the correspondence are the flows of Hamiltonian vector fields V_f. To understand what a "symmetry" really is in this context, we need to find a way to characterize the Hamiltonian vector fields and their flows. This would allow us to determine whether or not any given one-parameter group of transformations fits into this correspondence. So, basically, we want to answer the following questions:

1. Which vector fields are Hamiltonian? Specifically, how can we tell if some given vector field $\mathbf{F}$ is equal to V_f for some f?

2. What special properties do the flows of the vector fields V_f have? Specifically, given a one-parameter group of transformations, is there a geometric condition we can check to determine if it is the flow of a Hamiltonian vector field?

Which vector fields are conservative?

Following our analogy between gradients and Hamiltonian vector fields, let's start by asking a similar question to Question 1 above, but for gradients.

Recall that, given a function $f(x_1, \ldots, x_n)$, the gradient ∇f is defined as

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n} \right).$$

Given a vector field $\mathbf{F} = (F_1, \ldots, F_n)$, how can we tell if $\mathbf{F}$ is *conservative*[30], i.e. if $\mathbf{F} = \nabla f$ for some function f? For $n = 2$ and $n = 3$, this question is usually answered in a multivariable calculus class, so hopefully you will find some familiarity in this question.

[30] This terminology actually arises from that fact that conservative vector fields are those that describe forces where energy is conserved.

It's fairly simple to come up with some necessary conditions for $\mathbf{F}$ to be conservative. If $\mathbf{F}$ is conservative, then, by definition, there is a function f such that

$$F_i = \frac{\partial f}{\partial x_i}$$

for all i. But then

$$\frac{\partial F_i}{\partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i} = \frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial F_j}{\partial x_i}$$

for all i and j. So we can conclude that, if $\mathbf{F}$ is conservative, then

$$\frac{\partial F_i}{\partial x_j} = \frac{\partial F_j}{\partial x_i} \tag{2.20}$$

for all i and j. This is a necessary condition, meaning that, if you had a vector field $\mathbf{F}$ that didn't satisfy (2.20) for some i and j, then you would know for sure that $\mathbf{F}$ is *not* conservative.

Since we are asking for (2.20) to be true for all i and j, it's convenient to formulate the set of all equations (2.20) as a single matrix equation. One way to do this is to consider the total derivative of $\mathbf{F}$:

$$D_{\mathbf{F}} = \begin{bmatrix} \frac{\partial F_1}{\partial x_1} & \cdots & \frac{\partial F_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial F_n}{\partial x_1} & \cdots & \frac{\partial F_n}{\partial x_n} \end{bmatrix}.$$

Observe that the single equation $D_{\mathbf{F}} = D_{\mathbf{F}}^{\mathsf{T}}$ incorporates *all* of the equations in (2.20). Thus we have the following result:

$$\mathbf{F} \text{ is conservative} \implies D_{\mathbf{F}} \text{ is symmetric.} \tag{2.21}$$

In multivariable calculus, you may have learned (in the case $n = 3$) that, if $\mathbf{F}$ is conservative, then $\operatorname{curl} \mathbf{F} = 0$. To see the relationship between $D_{\mathbf{F}}$ and $\operatorname{curl} \mathbf{F}$, we just need to rewrite the equation $D_{\mathbf{F}} = D_{\mathbf{F}}^{\mathsf{T}}$ as $D_{\mathbf{F}} - D_{\mathbf{F}}^{\mathsf{T}} = 0$. Then, in the case $n = 3$, the left side of this equation is

$$D_{\mathbf{F}} - D_{\mathbf{F}}^{\mathsf{T}} = \begin{bmatrix} 0 & \frac{\partial F_1}{\partial x_2} - \frac{\partial F_2}{\partial x_1} & \frac{\partial F_1}{\partial x_3} - \frac{\partial F_3}{\partial x_1} \\ \frac{\partial F_2}{\partial x_1} - \frac{\partial F_1}{\partial x_2} & 0 & \frac{\partial F_2}{\partial x_3} - \frac{\partial F_3}{\partial x_2} \\ \frac{\partial F_3}{\partial x_1} - \frac{\partial F_1}{\partial x_3} & \frac{\partial F_3}{\partial x_2} - \frac{\partial F_2}{\partial x_3} & 0 \end{bmatrix}.$$

The entries of this matrix are exactly (up to sign) the components[31] of $\operatorname{curl} \mathbf{F}$. Thus, the statement you learned in multivariable calculus is just the $n = 3$ case of the more general statement (2.21).

[31] In the exercises, you'll see that there's a more precise way to state the relationship.

There is a theorem that says that the converse of (2.21) is true if **F** is defined in a region that is "simply connected", which roughly means that the region does not have any holes that would prevent a loop from being shrunk down to a point. For our purposes, it is sufficient to consider vector fields that are defined everywhere. Thus we have the following result: If **F** is defined everywhere and $D_{\mathbf{F}}$ is symmetric, then **F** is conservative.

Which vector fields are Hamiltonian?

Let's now consider a vector field $\mathcal{F} = (F_1, \ldots, F_n, G_1, \ldots, G_n)$ on phase space, and let's ask how we can tell if $\mathcal{F} = V_f$ for some function $f(x_1, \ldots, x_n, p_1, \ldots, p_n)$, i.e. if $\mathcal{F}$ is Hamiltonian. (See (2.15) for the formula for V_f.) In a similar way to how we found necessary conditions for a vector field to be conservative, we can find the following necessary conditions for $\mathcal{F}$ to be Hamiltonian:

$$\frac{\partial F_i}{\partial p_j} = \frac{\partial F_j}{\partial p_i}, \qquad \frac{\partial F_i}{\partial x_j} = -\frac{\partial G_j}{\partial p_i}, \qquad \frac{\partial G_i}{\partial x_j} = \frac{\partial G_j}{\partial x_i} \tag{2.22}$$

for all i and j. Similarly to what we did in the previous section, we would like to formulate the set of equations (2.22) as a single matrix equation. Consider the total derivative

$$D_{\mathcal{F}} = \begin{bmatrix} \frac{\partial F_1}{\partial x_1} & \cdots & \frac{\partial F_1}{\partial x_n} & \frac{\partial F_1}{\partial p_1} & \cdots & \frac{\partial F_1}{\partial p_n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial F_n}{\partial x_1} & \cdots & \frac{\partial F_n}{\partial x_n} & \frac{\partial F_n}{\partial p_1} & \cdots & \frac{\partial F_n}{\partial p_n} \\ \frac{\partial G_1}{\partial x_1} & \cdots & \frac{\partial G_1}{\partial x_n} & \frac{\partial G_1}{\partial p_1} & \cdots & \frac{\partial G_1}{\partial p_n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial G_n}{\partial x_1} & \cdots & \frac{\partial G_n}{\partial x_n} & \frac{\partial G_n}{\partial p_1} & \cdots & \frac{\partial G_n}{\partial p_n} \end{bmatrix}$$

If you multiply $D_{\mathcal{F}}$ by $\Omega = \begin{bmatrix} 0 & -I \\ I & 0 \end{bmatrix}$, you get

$$\Omega D_{\mathcal{F}} = \begin{bmatrix} -\frac{\partial G_1}{\partial x_1} & \cdots & -\frac{\partial G_1}{\partial x_n} & -\frac{\partial G_1}{\partial p_1} & \cdots & -\frac{\partial G_1}{\partial p_n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ -\frac{\partial G_n}{\partial x_1} & \cdots & -\frac{\partial G_n}{\partial x_n} & -\frac{\partial G_n}{\partial p_1} & \cdots & -\frac{\partial G_n}{\partial p_n} \\ \frac{\partial F_1}{\partial x_1} & \cdots & \frac{\partial F_1}{\partial x_n} & \frac{\partial F_1}{\partial p_1} & \cdots & \frac{\partial F_1}{\partial p_n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial F_n}{\partial x_1} & \cdots & \frac{\partial F_n}{\partial x_n} & \frac{\partial F_n}{\partial p_1} & \cdots & \frac{\partial F_n}{\partial p_n} \end{bmatrix}$$

Observe that the single equation $\Omega D_{\mathcal{F}} = (\Omega D_{\mathcal{F}})^T$ incorporates *all* of the equations in (2.22). Thus we have the following result:

$$\mathcal{F} \text{ is Hamiltonian} \implies \Omega D_{\mathcal{F}} \text{ is symmetric.} \tag{2.23}$$

In the exercises, you'll furthermore show that the converse of (2.23) is true if $\mathcal{F}$ is defined everywhere.

Which transformations preserve the dot product?

At the beginning of the course, we studied transformations of $\mathbb{R}^n$ that preserve the distance between points, and we found that such transformations are just combinations of rotations and translations[32].

We're now going to study the same problem, but phrased in a way that emphasizes calculus, rather than algebra. Of course, we're going to end up with the same answer as before, but the techniques used here will be useful in solving the analogous problem in phase space.

Recall from multivariable calculus that the arclength of a curve $\gamma(t)$ in $\mathbb{R}^n$ is[33]

$$\int_{t_0}^{t_1} \|\gamma'(t)\| dt. \tag{2.24}$$

If we think of the parametrized curve $\gamma(t)$ as describing the motion of a particle, then the formula (2.24) says that the distance traveled by the particle is equal to the integral of its speed.

Formula (2.24) tells us that we can think of distance as being a consequence of infinitesimal considerations. In other words, if we didn't know the formula for the distance between two points, we could *define* the distance between two points $\mathbf{x}_0$ and $\mathbf{x}_1$ as the arclength of the shortest curve from $\mathbf{x}_0$ to $\mathbf{x}_1$. Then we could, at least in principle, derive the distance formula[34] from (2.24).

To summarize what we have just learned, the notion of distance can be derived from the arclength formula. Conversely, in (2.24) we can see that arclength is calculated in terms of the length of the velocity vectors $\gamma'(t)$, which in turn can be defined as limits of quantities involving distance. So distance, arclength, and vector length form a circle of concepts that are "equivalent", in the sense that any one of them can be defined in terms of any other. But recall[35] that vector length and the dot product can also be defined in terms of each other, so the dot product can also be included in the circle of concepts.

Consider a transformation $\phi : \mathbb{R}^n \to \mathbb{R}^n$. For the purpose of calculations, it will help to write $\phi(\mathbf{x}) = (\phi_1(\mathbf{x}), \ldots, \phi_n(\mathbf{x}))$. How can we determine whether ϕ preserves distances? Based on the above discussion, one way to determine it is to ask whether ϕ preserves dot products.

So what exactly does it mean for ϕ to preserve dot products? To figure this out, it is helpful to think of vectors as being velocities of particles. Let $\gamma_1(t)$ and $\gamma_2(t)$ be two paths that start at the same point: $\gamma_1(0) = \gamma_2(0) = \mathbf{x}_0$. You could think of γ_1 and γ_2 as describing two possible trajectories that a particle could take, starting from the

[32] Reflections also preserve distance, but we excluded those by requiring the transformations to preserve orientation as well.

[33] If this equation doesn't look familiar to you, it might help to consider the case $n = 2$, where you could write $\gamma(t) = (x(t), y(t))$. In this case, (2.24) becomes

$$\int_{t_0}^{t_1} \sqrt{x'(t)^2 + y'(t)^2} dt.$$

[34] Actually, thinking of things this way raises questions about things you've probably been taking for granted your whole life. For example, how do we know that the shortest curve between two points is the straight line? Later in the course, you'll actually prove this and use it to derive the distance formula.

[35] See Exercise 5 in Appendix A.

initial position $\mathbf{x}_0$. At $t = 0$, the two corresponding velocity vectors are $\gamma_1'(0)$ and $\gamma_2'(0)$. So, on the one hand, we could calculate the dot product $\gamma_1'(0) \cdot \gamma_2'(0)$. On the other hand, we could consider the images of the paths under the transformation ϕ,

$$\phi(\gamma_1(t)) \text{ and } \phi(\gamma_2(t)),$$

find their initial velocity vectors

$$\frac{d}{dt}|_{t=0} \left[\phi(\gamma_1(t))\right] \text{ and } \frac{d}{dt}|_{t=0} \left[\phi(\gamma_2(t))\right],$$

and take the dot product of those vectors. So when we ask for ϕ to preserve dot products, we are asking for the result of both calculations to be the same:

$$\gamma_1'(0) \cdot \gamma_2'(0) = \frac{d}{dt}|_{t=0} \left[\phi(\gamma_1(t))\right] \cdot \frac{d}{dt}|_{t=0} \left[\phi(\gamma_2(t))\right] \tag{2.25}$$

for all paths $\gamma_1(t)$, $\gamma_2(t)$. This might seem like a scary condition to check, but keep reading; it will turn out to not be so bad. We just need to figure out what those derivatives on the right side of (2.25) are.

Let $\gamma(t)$ be a path. Then $\phi(\gamma(t)) = (\phi_1(\gamma(t)), \ldots, \phi_n(\gamma(t)))$, so

$$\frac{d}{dt} \left[\phi(\gamma(t))\right] = \left(\frac{d}{dt} \left[\phi_1(\gamma(t))\right], \ldots, \frac{d}{dt} \left[\phi_n(\gamma(t))\right]\right).$$

By the Chain Rule, we can write this as

$$\left(\nabla\phi_1 \cdot \gamma'(t), \ldots, \nabla\phi_n \cdot \gamma'(t)\right). \tag{2.26}$$

Since every component of this vector is a dot product with $\gamma'(t)$, we can rewrite it as a matrix multiplication. Consider the matrix

$$\begin{bmatrix} \frac{\partial\phi_1}{\partial x_1} & \cdots & \frac{\partial\phi_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial\phi_n}{\partial x_1} & \cdots & \frac{\partial\phi_n}{\partial x_n} \end{bmatrix}, \tag{2.27}$$

whose rows are the gradients of $\nabla\phi_i$. If you multiply the matrix (2.27) by $\gamma'(t)$, you get (2.26). But wait, we've seen this before! The matrix in (2.27) is the total derivative D_ϕ of ϕ. So putting things together, we see that

$$\frac{d}{dt} \left[\phi(\gamma(t))\right] = D_\phi \gamma'(t),$$

so

$$\frac{d}{dt}|_{t=0} \left[\phi(\gamma(t))\right] = D_\phi \gamma'(0).$$

We can thus rewrite condition (2.25) as

$$\gamma_1'(0) \cdot \gamma_2'(0) = (D_\phi \gamma_1'(0)) \cdot (D_\phi \gamma_2'(0)).$$

But since the vectors $\gamma_1'(0)$ and $\gamma_2'(0)$ can be any vectors, we can replace them with $\mathbf{v}$ and $\mathbf{w}$ to get the condition[36]

[36] See, I told you it would get better.

$$\mathbf{v} \cdot \mathbf{w} = (D_\phi \mathbf{v}) \cdot (D_\phi \mathbf{w}). \tag{2.28}$$

Remark: the notation in 2.28 hides the important fact that D_ϕ generally depends on $\mathbf{x}$ and needs to be evaluated at the initial point where the vectors are located. But this initial point can be anywhere, so our requirement is that (2.28) should be true for all $\mathbf{v}$, $\mathbf{w}$, and $\mathbf{x}$.

But we can go even further, since (2.28) is true for all $\mathbf{v}$ and $\mathbf{w}$ if and only if[37]

[37] See Exercise 6 in Appendix A.

$$D_\phi^T D_\phi = I. \tag{2.29}$$

So we finally conclude that ϕ preserves the dot product (and hence preserves distance) if and only if D_ϕ is orthogonal for all $\mathbf{x}$.

This result is consistent with what we know already, which is that the distance-preserving transformations are of the form $\phi(\mathbf{x}) = U\mathbf{x} + \mathbf{a}$, where U is an orthogonal matrix and $\mathbf{a}$ is a vector. In this case, $D_\phi = U$, so D_ϕ is indeed orthogonal. To connect all the dots, we would also need to prove that, if D_ϕ is orthogonal for all $\mathbf{x}$, then actually D_ϕ is constant, and therefore ϕ is of the form $\phi(\mathbf{x}) = U\mathbf{x} + \mathbf{a}$. However, while this is certainly true, it isn't so easy to prove.

Which transformations preserve the symplectic pairing?

Recall that in Section 2.3, we discovered the symplectic pairing $\omega(\mathbf{a}, \mathbf{b})$ as being related to Hamiltonian vector fields in the same way that the dot product is related to gradients. In analogy with the previous section, we now ask which transformations preserve the symplectic pairing.

Suppose that ϕ is a transformation of phase space $\mathbb{R}^{2n}$. We can define what it means for ϕ to preserve the symplectic pairing in a similar way to (2.25):

$$\omega(\gamma_1'(0), \gamma_2'(0)) = \omega\left(\frac{d}{dt}|_{t=0}\left[\phi(\gamma_1(t))\right], \frac{d}{dt}|_{t=0}\left[\phi(\gamma_2(t))\right]\right)$$

for all paths $\gamma_1(t)$, $\gamma_2(t)$. But the calculations between (2.25) and (2.28) apply equally well in this setting, so we can rewrite the above condition in a similar way to (2.28) as

$$\omega(\mathbf{a}, \mathbf{b}) = \omega(D_\phi \mathbf{a}, D_\phi \mathbf{b}) \tag{2.30}$$

for all $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{x}$.

Using (2.17)[38], we can go further and conclude that ϕ preserves the symplectic pairing if and only if

[38] Also see Exercise 2 in Section 2.3.

$$D_\phi^T \Omega D_\phi = \Omega. \tag{2.31}$$

If this is the case, we will say that ϕ is a *symplectic transformation* or, more concisely, a *symplectomorphism*[39].

[39] The latter is the actual commonly-used term.

Here are some examples of symplectomorphisms.

- *Translations:* Since a translation just adds constants to the coordinates, its Jacobian matrix is equal to the identity matrix, which clearly satisfies (2.31).

- *Linear symplectomorphisms:* Suppose that ϕ is linear, so that it has the form $\phi(\mathbf{x}, \mathbf{p}) = A \begin{bmatrix} \mathbf{x} \\ \mathbf{p} \end{bmatrix}$, where A is a $2n \times 2n$ matrix. Then $D_\phi = A$, so ϕ is a symplectomorphism whenever $A^\intercal \Omega A = \Omega$.

  You looked at matrices satisfying this condition in Exercise 2 in Section 2.3. A particularly relevant example is where U is an $n \times n$ rotation matrix. Then

  $$A = \begin{bmatrix} U & 0 \\ 0 & U \end{bmatrix}$$

  satisfies the condition $A^\intercal \Omega A = \Omega$. The corresponding transformation is $\phi(\mathbf{x}, \mathbf{p}) = (U\mathbf{x}, U\mathbf{p})$, which means that ϕ rotates position space and simultaneously rotates momentum space in exactly the same way.

  More generally, if B is any invertible $n \times n$ matrix, then

  $$A = \begin{bmatrix} B & 0 \\ 0 & (B^{-1})^\intercal \end{bmatrix}$$

  satisfies the condition $A^\intercal \Omega A = \Omega$. In this case, position space and momentum space are transformed in different, but closely related, ways.

  There also exist examples where A is not block-diagonal, in which case the transformation mixes the position and momentum variables. For example, in Exercise 3 in Section 2.3, you showed that, when $n = 1$, any 2×2 matrix with determinant 1 preserves the symplectic pairing.

- *Nonlinear symplectomorphisms:* In contrast to the case of distance-preserving transformations, there exist many examples of nonlinear symplectomorphisms. Here is one example[40]; in the exercises, you will see another.

  Let $\phi(x, p) = (x, p + f(x))$, where f can be any function. Then

  $$D_\phi = \begin{bmatrix} 1 & 0 \\ f'(x) & 1 \end{bmatrix}.$$

  You can then check[41] that D_ϕ satisfies (2.31).

[40] This example is in the case $n = 1$, but it can be generalized to arbitary n.

[41] An easier check is to observe that $\det(D_\phi) = 1$, but beware that this is only a sufficient condition in the case $n = 1$; see Exercise 3 in Section 2.3.

Exercises

1. Let $\mathbf{F}$ be a vector field on $\mathbb{R}^3$. Show that $D_{\mathbf{F}} - D_{\mathbf{F}}^{\mathrm{T}} T = \beta(\operatorname{curl} \mathbf{F})$, where β is the map defined in Exercise 6 of Section 1.2.

2. (a) Show that, for a vector field $\mathcal{F}$ on phase space, the operations of multiplying by Ω and taking the total derivative commute with each other. That is, show that $D_{\Omega\mathcal{F}} = \Omega D_{\mathcal{F}}$.

   (b) Show that, if $f(x_1, \ldots, x_n, p_1, \ldots, p_n)$ is a function on phase space, then $\nabla f = \Omega V_f$.

   (c) Show that, if $\mathcal{F}$ is defined in a "nice" region, then the converse of (2.23) is true. [Hint: use the results of parts (a) and (b), together with the fact that, if $\mathcal{F}$ is defined in a "nice" region, then the converse of (2.21) is true.]

3. Consider a phase space transformation of the form $\phi(\mathbf{x}, \mathbf{p}) = (\mathbf{x}, \mathbf{p} + \mathbf{F}(\mathbf{x}))$, where $\mathbf{F}$ is a vector field on position space. Show that ϕ is a symplectomorphism if and only if $D_{\mathbf{F}}$ is symmetric.

4. Consider a linear transformation $\phi(\mathbf{x}, \mathbf{p}) = A \begin{bmatrix} \mathbf{x} \\ \mathbf{p} \end{bmatrix}$, where

$$A = \begin{bmatrix} I & C \\ 0 & I \end{bmatrix}.$$

   Find a condition on C that holds if and only if ϕ is a symplectomorphism.

5. Let $\phi(x, p) = (f(x), \frac{p}{f'(x)})$, where f can be any function such that $f'(x) \neq 0$ for all x. Compute D_ϕ, and show that ϕ is a symplectomorphism.

2.5 *Infinitesimal symplectomorphisms*

Which vector fields preserve the dot product?

Recall[42] that, if $\mathbf{F}$ is a vector field on on $\mathbb{R}^n$, then its flow ϕ_s is a one-parameter group of transformations of $\mathbb{R}^n$. We then think of $\mathbf{F}$ as being an infinitesimal transformation that generates the actual transformations ϕ_s.

[42] See Section 1.4.

We will say that a vector field $\mathbf{F}$ preserves the dot product if its flow consists of transformations that preserve the dot product. We already know what these vector fields are — they are the *Euclidean vector fields*, i.e. combinations of infinitesimal rotations and infinitesimal translations. But we can gain additional insights by revisiting the problem using the results of Section 2.4.

We would like to know what information about a vector field $\mathbf{F}$ can be derived from the assumption that its flow ϕ_s satisfies (2.29) for all s. We first make the following observation:

$$D_{\mathbf{F}} = \frac{d}{ds}|_{s=0} \left[D_{\phi_s} \right]. \tag{2.32}$$

The proof of (2.32) basically comes down to the fact that partial derivatives commute. To get $\mathbf{F}$, we take the partial derivative of ϕ_s with respect to s, and to get $D_{\mathbf{F}}$, we then take partial derivatives with respect to x_i. On the other hand, to get the right side of (2.32), you first take partial derivatives with respect to x_i, and then take the partial derivative with respect to s.

We can now do a calculation that should feel vaguely familiar to you. If $\mathbf{F}$ preserves the dot product, then ϕ_s satisfies (2.29):

$$D_{\phi_s}^{\intercal} D_{\phi_s} = I$$

for all s. Taking the derivative of both sides with respect to s, we get

$$\frac{d}{ds}\left[D_{\phi_s}^{\intercal} \right] D_{\phi_s} + D_{\phi_s}^{\intercal} \frac{d}{ds}\left[D_{\phi_s} \right] = 0.$$

Setting $s = 0$ and using the fact that ϕ_0 is the identity map, we get

$$\frac{d}{ds}|_{s=0}\left[D_{\phi_s}^{\intercal} \right] + \frac{d}{ds}|_{s=0}\left[D_{\phi_s} \right] = 0.$$

Using (2.32) and the fact that the transpose operation commutes with taking derivatives, we finally get

$$D_{\mathbf{F}}^{\intercal} + D_{\mathbf{F}} = 0. \tag{2.33}$$

The conclusion is that, if $\mathbf{F}$ preserves the dot product, then $D_{\mathbf{F}}$ is skew-symmetric. And actually, this is "if and only if", since you can show[43] that if $D_{\mathbf{F}}$ is skew-symmetric, the flow of $\mathbf{F}$ will satisfy (2.29).

[43] We'll leave this as an exercise.

The equation (2.33) is a first-order partial differential equation. In general, finding general solutions to partial differential equations can be difficult, but in this case it is possible to do so. I'll do the two-dimensional case here. Let $\mathbf{F}(x, y) = (F_1(x, y), F_2(x, y))$. Then

$$D_{\mathbf{F}} = \begin{bmatrix} \frac{\partial F_1}{\partial x} & \frac{\partial F_1}{\partial y} \\ \frac{\partial F_2}{\partial x} & \frac{\partial F_2}{\partial y} \end{bmatrix}.$$

In this case, equation (2.33) tells us that $\frac{\partial F_1}{\partial x} = 0$ (so F_1 only depends on y), $\frac{\partial F_2}{\partial y} = 0$ (so F_2 only depends on x), and

$$\frac{\partial F_2}{\partial x} = -\frac{\partial F_1}{\partial y}. \tag{2.34}$$

The left side of (2.34) doesn't depend on x and the right side doesn't depend on y. But they're equal, so they don't depend on x or y,

and therefore must be constant. Thus there is a constant c such that $\frac{\partial F_2}{\partial x} = c$ and $\frac{\partial F_1}{\partial y} = -c$.

We can finish solving by integration:

$$F_1 = \int \frac{\partial F_1}{\partial y} dy = -\int c dy = -cy + a,$$

$$F_2 = \int \frac{\partial F_2}{\partial x} dx = \int c dx = cx + b,$$

so the general solution is

$$\mathbf{F} = (-cy + a, cx + b),$$

or, in derivation notation,

$$\mathbf{F} = (-cy + a)\frac{\partial}{\partial x} + (cx + b)\frac{\partial}{\partial y}.$$

In terms of the vector fields $T_x = \frac{\partial}{\partial x}$, $T_y = \frac{\partial}{\partial y}$, and $L = -y\frac{\partial}{\partial x} + x\frac{\partial}{\partial y}$ that generate translations and rotation, we can rewrite the solution as

$$\mathbf{F} = aT_x + bT_y + cL,$$

showing that the general solution is an arbitrary linear combination of T_x, T_y, and L.

Which vector fields preserve the symplectic pairing?

Now let's perform a similar analysis to find out what we can say about a vector field $\mathcal{F}$ on $\mathbb{R}^{2n}$ if its flow ϕ_s consists of symplectomorphisms. If this is the case, then ϕ_s satisfies (2.31):

$$D_{\phi_s}^{\mathsf{T}} \Omega D_{\phi_s} = \Omega$$

for all s. By calculations that are very similar to what we did in the previous section, we can deduce that

$$D_{\mathcal{F}}^{\mathsf{T}} \Omega + \Omega D_{\mathcal{F}} = 0. \tag{2.35}$$

Using the fact that $\Omega^{\mathsf{T}} = -\Omega$, we can rewrite (2.35) as

$$\Omega D_{\mathcal{F}} = (\Omega D_{\mathcal{F}})^{\mathsf{T}}.$$

The conclusion is that, if $\mathcal{F}$ generates symplectomorphisms, then $\Omega D_{\mathcal{F}}$ is symmetric. As above, you can show that this is "if and only if".

Look back at (2.23). It tells us that, if $\mathcal{F}$ is Hamiltonian, then $\Omega D_{\mathcal{F}}$ is symmetric. Combining this with the result we just found, we can conclude that, if $\mathcal{F}$ is Hamiltonian, then its flow consists of symplectomorphisms. Furthermore, under the assumption that $\mathcal{F}$ is defined everywhere, then the converse of (2.23) is also true. This finally gives us an answer to Question 2 from the beginning of Section 2.4: a one-parameter group ϕ_s consists of symplectomorphisms if and only if it is the flow of a Hamiltonian vector field.

The Big Picture: Noether's Theorem for Hamiltonian mechanics

We'll now collect the various ideas we've been discussing to give a complete formulation of Noether's Theorem for Hamiltonian mechanics. Let's start by recalling what data makes up a Hamiltonian system. Starting from a physical system, you construct a phase space $\mathbb{R}^{2n}$ and the Hamiltonian function H. Phase space comes automatically equipped with two operations[44]:

1. the operation V, which takes functions f to their Hamiltonian vector fields V_f,
2. the symplectic pairing ω.

[44] Note that these two operations are closely related. You can use (2.13) to show that V can be expressed in terms of ω, and vice versa.

The flow of V_H describes the dynamics of the system. Specifically, if ϕ_s is the flow of V_H, then $\phi_s(\mathbf{x}, \mathbf{p})$ describes the state of the system at time s, if the system is in state $(\mathbf{x}, \mathbf{p})$ at time 0.

A *conserved quantity* is defined to be a function $f(\mathbf{x}, \mathbf{p})$ such that $V_H(f) = 0$. Geometrically, this means that the value of f does not change as you flow[45] along V_H. Physically, it means that system changes in a way such that the value of f is constant with respect to time.

[45] See Exercise 5 in Section 1.5.

A *symmetry* is defined to be a one-parameter group of symplectomorphisms ϕ_s which also preserve H, in the sense that $H(\phi_s(\mathbf{x}, \mathbf{p})) = H(\mathbf{x}, \mathbf{p})$ for all s. The key point to observe here is that we are requiring ϕ_s to preserve *all* of the structures that are relevant to Hamiltonian mechanics: being a symplectomorphism means that it preserves ω; since V can be derived from ω, we are implicitly requiring ϕ_s to preserve V; and we are also requiring ϕ_s to preserve H.

In Section 2.3, we saw that, if f is a conserved quantity, then the flows of V_f preserve H. Additionally, in the last section, we saw that the flows of Hamiltonian vector fields are symplectomorphisms. Combining these results, we see that, if f is a conserved quantity, then the flows of V_f form a symmetry. Note that $V_{f+c} = V_f$ for any constant c, so the conserved quantities f and $f + c$ both lead to the same symmetry.

Conversely, we saw in the last section[46] that the vector field $\mathcal{F}$ associated to a one-parameter group ϕ_s of symplectomorphisms is Hamiltonian, so $\mathcal{F} = V_f$ for some function f. Additionally, if ϕ_s preserves H for all s, then $\mathcal{F}(H) = 0$. Using the results of Section 2.3, we can then conclude that f is a conserved quantity. Note that, because f is obtained by integration, it is only well-defined up to a constant.

[46] This was under the assumption that $\mathcal{F}$ is defined on all of $\mathbb{R}^{2n}$.

Putting everything together, we can now state Noether's Theorem for Hamiltonian mechanics: *There is a one-to-one correspondence between conserved quantities (up to constants) and symmetries.*

Let's return to the example of gravity near the surface of the earth in three spatial coordinates. In this case, the phase space variables are (x, y, z, p_x, p_y, p_z), and the Hamiltonian (2.19) can be rewritten as

$$H = \frac{\|\mathbf{p}\|^2}{2m} + mgz.$$

From this, we see that H only depends on two quantities, the magnitude of $\mathbf{p}$ and the height z. So, if we are looking for symmetries, we could look for one-parameter groups of symplectomorphisms that preserve[47] both $\|\mathbf{p}\|$ and z.

Here are some relatively easy-to-find symmetries of this system[48]:

- Any translation is a symplectomorphism. But, since we want to preserve $\|\mathbf{p}\|$, we can't translate momentum space at all. We also want to preserve z, so we shouldn't translate in that direction. But translations in the x and y directions are symmetries. They correspond to the conserved quantities p_x and p_y.

- If B is a 3×3 rotation matrix, then $\phi(\mathbf{x}, \mathbf{p}) = (B\mathbf{x}, B\mathbf{p})$ is a symplectomorphism. Because B is a rotation, it will preserve the magnitude of $\mathbf{p}$. But, in order to preserve z, we can only allow B to rotate around the z-axis. The rotations around the z-axis form a symmetry that corresponds to the conserved quantity J_z (i.e. the z component of angular momentum).

Can you find other examples of symmetries in this example? They exist, but they aren't as simple to describe geometrically. For example, we know that H itself is a conserved quantity. The corresponding symmetry is the family of transformations that describe how the system changes over time:

$$\phi_s(x, y, z, p_x, p_y, p_z) = \left(x + \frac{p_x}{m}s, y + \frac{p_y}{m}s, z + \frac{p_z}{m}s - \frac{g}{2}s^2, p_x, p_y, p_z - mgs\right).$$

Exercises

1. Let $\mathbf{F}$ be a vector field on $\mathbb{R}^n$. Show that, if $D_{\mathbf{F}}$ is skew-symmetric, then the flow of $\mathbf{F}$ satisfies (2.29).

2. In three dimensions, solve the equation (2.33). Observe that the general solution is an arbitrary linear combination of the six vector fields in Section 6.2.

3. Let g be a function on $\mathbb{R}^n$, and let ϕ_s be a family of transformations of $\mathbb{R}^n$ that preserve g, in the sense that $g(\phi_s(\mathbf{x})) = g(\mathbf{x})$ for all $\mathbf{x}$ and all s. Let $\mathbf{F} = \frac{d}{ds}|_{s=0}[\phi_s]$ be the vector field associated to the family ϕ_s. Show that $\mathbf{F}(g) = 0$.

[47] There could be yet other possibilities. It could be that $\|\mathbf{p}\|$ and z change but the changes cancel out when you calculate H. But requiring both $\|\mathbf{p}\|$ and z to be preserved keeps things more simple.

[48] It may help to refer to Section 2.4 to see the examples of symplectomorphisms we found there.

4. Let f and g be arbitrary functions on $\mathbb{R}^{2n}$.

   (a) Show that the Hamiltonian vector field operation satisfies the properties $V_{f+g} = V_f + V_g$ and $V_{fg} = fV_g + gV_f$.

   (b) Use part (a) to deduce that, if $V_f(H) = 0$ and $V_g(H) = 0$, then $V_{f+g}(H) = 0$ and $V_{fg}(H) = 0$.

   (c) Give a simple interpretation of the results in part (b) in terms of conserved quantities.

5. The (Galilean) principle of relativity basically says that *all* laws of physics should be symmetric under the Galilean transformations. In light of Noether's Theorem, this implies that the associated quantities (momentum, angular momentum, etc) should always be conserved. But on the other hand, we've often encountered systems where not all of the usual quantities are conserved. For example, in our model of near-earth gravity, momentum in the vertical direction is obviously not conserved; if you throw a ball up, it will eventually fall down.

   The way to reconcile these things is to recognize that the system under consideration is not actually taking into account all of the interactions. In this case, the issue is that the force of gravity (and, more generally, any force) should always come in equal and opposite pairs. In this case, gravity is pulling the ball down, but it's also *pulling the earth up*.

   Here is one way to model near-earth gravity to take the motion of the earth into account. Let x_B be the height of the ball (relative to some arbitrary point in space), and let x_E be the height of the surface of the earth (relative to the same arbitrary point). Set the potential energy to be $V = mg(x_B - x_E)$, where m is the mass of the ball.

   Use the Hamiltonian formalism to analyze this system:

   (a) Write out the Hamiltonian H for this system, write out Hamilton's equations, and solve them.

   (b) Show that this system has a translation symmetry and find the corresponding conserved quantity.

   (c) Explain why, even though it is not strictly constant, x_E does not change very much. (This is why the usual model of near-earth gravity is a reasonable approximation).

2.6 *Some interesting consequences of the Hamiltonian approach*

Coordinate changes

Let ϕ be a symplectomorphism of phase space. We know that, if ϕ preserves the Hamiltonian function H, then ϕ preserves all of the structures that are relevant to Hamiltonian mechanics. But what if ϕ doesn't preserve H? Then we can still do something potentially useful with it.

Recall that, when we say that ϕ preserves H, we mean that $H(\phi(\mathbf{x}, \mathbf{p})) = H(\mathbf{x}, \mathbf{p})$. If don't assume that ϕ preserves H, we can instead define a new function G by the formula

$$G(\mathbf{x}, \mathbf{p}) = H(\phi(\mathbf{x}, \mathbf{p})). \tag{2.36}$$

Then V_G will be[49] related to V_H by the formula $V_H|_{\phi(\mathbf{x},\mathbf{p})} = D_\phi V_G|_{(\mathbf{x},\mathbf{p})}$. In other words, if H and G are related by ϕ, then V_H and V_G are related by ϕ as well[50]. In practice, this means that we can change coordinates using any symplectomorphism, and as long as we change the Hamiltonian function accordingly, then the Hamiltonian method still gives the correct dynamics of the system. This is probably best illustrated by an example, which we will now do.

[49] This is an exercise.

[50] Remember that D_ϕ tells us how vectors are transformed by ϕ.

Example 2.6.1 (The pendulum). Consider a pendulum of length ℓ and mass m. All of the mass is assumed to be at the end of the pendulum. The position of the end of the pendulum is a point (x, y) on the plane, so we would have a 4-dimensional phase space with coordinates (x, y, p_x, p_y). The variable y represents the height, so the potential energy associated to gravity is mgy. The Hamiltonian function is then

$$H = \frac{p_x^2 + p_y^2}{2m} + mgy.$$

Of course the pendulum isn't free to move in the plane; the end of the pendulum is constrained to move on a circle of radius ℓ. So we need to implement the constraint $x^2 + y^2 = \ell$. One way to do this is to first change to polar coordinates, where the constraint is simply $r = \ell$.

To switch to polar coordinates, we'll use the equations[51]

$$x = r \sin\theta, \qquad y = -r\cos\theta.$$

[51] These equations are slightly different than what you're used to. They're chosen so that $\theta = 0$ corresponds to when the pendulum is hanging straight down.

These equations define a map ψ from the (r, θ)-plane to the (x, y)-plane:

$$\psi(r, \theta) = (r\sin\theta, -r\cos\theta).$$

This map can be extended to a symplectomorphism ϕ of phase space by setting

$$p_x = (\sin\theta)p_r + \frac{1}{r}(\cos\theta)p_\theta, \quad p_y = -(\cos\theta)p_r + \frac{1}{r}(\sin\theta)p_\theta, \tag{2.37}$$

so

$$\phi(x, y, p_x, p_y) = \left(r\sin\theta, -r\cos\theta, (\sin\theta)p_r + \frac{1}{r}(\cos\theta)p_\theta, -(\cos\theta)p_r + \frac{1}{r}(\sin\theta)p_\theta \right).$$

Equations (2.37) come from the general rule[52] that, if ψ is a transformation of position space, then the momentum coordinates transform as

$$\mathbf{p} \mapsto (D_\psi^{-1})^T \mathbf{p}',$$

where $\mathbf{p}'$ is the "new" momentum vector (in this case (p_r, p_θ)).

If you use equations (2.37) and simplify, you'll get $p_x^2 + p_y^2 = p_r^2 + \frac{1}{r^2}p_\theta^2$. Thus, under the map ϕ, the Hamiltonian function H transforms into

$$G = \frac{p_r^2 + \frac{1}{r^2}p_\theta^2}{2m} - mgr\cos\theta.$$

At this point, we can introduce the constraint $r = L$ together with[53] $p_r = 0$, giving us

$$G = \frac{p_\theta^2}{2m\ell^2} - mg\ell\cos\theta.$$

Note that, after introducing the constraints, G has become a function on the 2-variable phase space (θ, p_θ). The associated Hamiltonian vector field is

$$V_G = \left(\frac{\partial G}{\partial p_\theta}, -\frac{\partial G}{\partial \theta} \right) = \left(\frac{p_\theta}{m\ell^2}, mg\ell\sin\theta \right).$$

In Figure 2.2, this vector field is plotted with the constants chosen so you can clearly see the dynamics, and with a couple of the level curves of G. You can see that, in some cases, the trajectory is a vaguely circular loop; this occurs when the energy of the pendulum is relatively small, so that the pendulum oscillates. But, in some cases, the trajectory looks more like a wave, this occurs when the energy of the pendulum is high enough to make it go around in complete circles.

Remark: Note that G and V_G are both periodic with respect to θ. This makes sense, since θ is physically only well-defined up to multiples of 2π. Geometrically, this means that, in this example, the phase space should really be thought of as a cylinder rather than a plane. Imagine cutting out Figure 2.2 and rolling it into a cylinder so that the periods overlap.

This example demonstrates that, if you are trying to formulate a general mathematical theory of phase spaces, you can't just study $\mathbb{R}^{2n}$. Instead, you would study "symplectic manifolds" which look like ordinary phase space $\mathbb{R}^{2n}$ when you zoom in very close, but that might look like cylinders, spheres, or even more complicated things when you zoom out. This is something that you would encounter in a graduate-level course on the subject.

[52] You'll show that this rule works in the exercises.

[53] Note that the pendulum will have no momentum in the r direction.

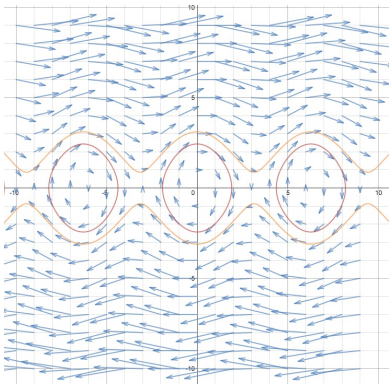

Figure 2.2: The vector field V_G with $m = 1$ and $\ell = 0.6$, along with the level curves $G = 2$ and $G = 7$.

Liouville's Theorem

Recall that, if ϕ is a symplectomorphism of phase space, then its total derivative D_ϕ satisfies the equation $D_\phi^T \Omega D_\phi = \Omega$. From this equation, we can deduce[54] that $\det D_\phi = \pm 1$.

[54] In fact, it's possible to show that $\det D_\phi = 1$, but proving this requires some advanced techniques.

Let H be the Hamiltonian function, and let ϕ_t be the flow of V_H. Recall that ϕ_t describes how the physical system evolves over time. Specifically, if a physical system has position and momentum $(\mathbf{x}, \mathbf{p})$ at time 0, then physically, $\phi_t(\mathbf{x}, \mathbf{p})$ tells us the position and momentum of the system at time t.

Since ϕ_t is a symplectomorphism for all t, we know[55] that $\det D_{\phi_t} = 1$. The determinant of the Jacobian of a transformation tells us (infinitesimally) how volumes change under the transformation, so the fact that $\det D_{\phi_t} = 1$ tells us that ϕ_t preserves volumes[56]. This is *Liouville's Theorem*.

[55] We know this without even using the advanced techniques mentioned above, because ϕ_0 is the identity map and ϕ_t varies continuously with respect to t.

[56] Note that I am using "volume" in the general sense of $2n$-dimensional volume, which is actually area when $n = 1$.

Liouville's Theorem is an interesting fact that you wouldn't even be able to make sense of in Newtonian mechanics. If you consider some region R of phase space, the points in R will move over time to a new region R_t, consisting of all the points $\phi_t(\mathbf{x}, \mathbf{p})$ where $(\mathbf{x}, \mathbf{p}) \in R$. Liouville's Theorem tells us that the volume of R_t is constant with respect to t.

One way to interpret Liouville's Theorem physically is in terms of statistical mechanics. Suppose you have a cloud of particles that all have different positions and momenta. You could (approximately) describe the cloud by a function $f(x, p)$ (in the case $n = 1$) that gives the density of the particles in phase space. Then, given a region R in phase space, the number of particles with position and momentum in that region would be (approximately) equal to

$$\iint_R f(x,p)\,dx\,dp.$$

Over time, the distribution of particles will change, so we get a family of functions f_t, determined by the equation

$$f_t(\phi_t(x,p)) = f(x,p). \tag{2.38}$$

Now fix some value of t, and write $\phi_t = (\phi_1(x,p), \phi_2(x,p))$. We can view ϕ_t as being a change of variables from (x, p) to some new variables (u, q), where $u = \phi_1(x,p)$ and $q = \phi_2(x,p)$. Then, for this value of t, equation (2.38) can be written simply as $f_t(u,q) = f(x,p)$. The number of particles in the region R_t at time t is

$$\iint_{R_t} f_t(u,q)\,du\,dq.$$

By the change of variables formula for double integrals, this is equal to

$$\iint_R f(x,p)|\det D_{\phi_t}|\,dx\,dp.$$

But Liouville's Theorem tells us $|\det D_{\phi_t}| = 1$, so this becomes

$$\iint_R f(x,p)dxdp.$$

Thus, the number of particles in R_t at time t is equal to the number of particles in R at time 0.

Exercises

1. Let ϕ be a symplectomorphism of phase space. Let $G(\mathbf{x},\mathbf{p})$ be defined by the formula (2.36). Show that $V_H|_{\phi(\mathbf{x},\mathbf{p})} = D_\phi V_G|_{(\mathbf{x},\mathbf{p})}$.

2. Show that, if $\psi : \mathbb{R}^2 \to \mathbb{R}^2$ is any transformation of position space, then $\phi(\mathbf{x},\mathbf{p}) = (\psi(\mathbf{x}), (D_\psi^{-1})^T\mathbf{p})$ is a symplectomorphism of phase space.

3. If a pendulum is rotating in a third dimension (imagine a grandfather clock on a turntable) with angular velocity ω, then, in the pendulum's reference frame, we would see a centrifugal force[57], making the total force $\mathbf{F} = (m\omega^2 x, mg)$.

   (a) Observe that $\mathbf{F}$ is conservative, and find the corresponding potential energy V.

   (b) Write down the Hamiltonian function H in this case.

   (c) Find the function G that H transforms into when you switch to polar coordinates and introduce the constraints $r = L$ and $p_r = 0$.

   (d) Plot the vector field V_G along with some level curves of G. (You probably should play with the constants m, ω, and L until you get a nice-looking graph.) Make some qualitative observations about the dynamics of the system.

4. When $n = 1$, the converse of Liouville's theorem is true. Specifically, if a family ϕ_t of transformations of $\mathbb{R}^2$ preserves area, then the vector field associated to ϕ_t is Hamiltonian. (Note that the analogous statement is *not* true for $n > 1$; if you're interested in knowing more, look up the *non-squeezing theorem*.) Consider the family of transformations given by

$$\phi_t(x,p) = (e^t x, e^{-t} p).$$

   (a) Verify that ϕ_t is a one-parameter group.

   (b) Give a geometric description of how ϕ_t transforms phase space, and explain why it preserves area.

   (c) Find the vector field $\mathcal{F}$ for which ϕ_t is the flow.

   (d) Find a function $f(x,p)$ such that $V_f = \mathcal{F}$.

[57] You don't need to know anything about centrifugal force to do this problem; just use this formula for $\mathbf{F}$ and do the math.

3 Lagrangian Mechanics

The miracle of the appropriateness of the language of mathematics for the formulation of the laws of physics is a wonderful gift which we neither understand nor deserve. — Eugene Wigner

3.1 The action principle

The action

In Hamiltonian mechanics, we consider phase space $\mathbb{R}^{2n}$, where the first n variables describe the position of the object(s) in the system, and the second n variables describe the momentum. In Lagrangian mechanics, we instead consider the *velocity space*, which is also $2n$-dimensional, but where the second n variables describe the velocity. So a point in velocity space is of the form $(\mathbf{x}, \mathbf{v})$. It's helpful to visualize $(\mathbf{x}, \mathbf{v})$ as a vector $\mathbf{v}$ located at the point $\mathbf{x}$.

What's the difference between phase space and velocity space? They're both just $\mathbb{R}^{2n}$, but the difference is in how the variables transform under a change of the $\mathbf{x}$ variables. Under a transformation ψ of position space, the velocity vectors transform under the total derivative:

$$\mathbf{v} \mapsto D_\psi \mathbf{v}, \tag{3.1}$$

whereas the momentum vectors transform under the inverse transpose of the total derivative[1]:

$$\mathbf{p} \mapsto (D_\psi^{-1})^T \mathbf{p}. \tag{3.2}$$

The *Lagrangian* is a function on velocity space:

$$L(\mathbf{x}, \mathbf{v}) = \text{Kinetic} - \text{Potential} = \frac{m\|\mathbf{v}\|^2}{2} - V(\mathbf{x}).$$

Let $\gamma(t) : [t_0, t_1] \to \mathbb{R}^n$ be a parametrized path in position space describing a possible trajectory of an object. Then the velocity of the object is $\gamma'(t)$. Then we can plug in $\gamma(t)$ for $\mathbf{x}$ and $\gamma'(t)$ for $\mathbf{v}$ to get L as a function of t.

[1] See Exercise 2 in Section 2.6. The difference between equations (3.1) and (3.2) indicate a difference between velocity and momentum that's more substantive than you may have expected. The moral is that momentum isn't actually mass times velocity, but it only looks that way in Cartesian coordinates. We'll elaborate on this point later.

The *action* of the path γ is defined as the integral of L:

$$S(\gamma) = \int_{t_0}^{t_1} L(\gamma(t), \gamma'(t))dt. \tag{3.3}$$

This is a definite integral, so $S(\gamma)$ is a number[2]. Since S takes a path γ as input and produces a number as its output, we can view S as a function on *the space of all paths*. Keep in mind that the space of all paths is really big — it's infinite-dimensional.

[2] What does this number mean physically, you ask? It might help to remember that energy is just a mathematical construct that was originally discovered as "a quantity that is often conserved". You've probably encountered the concept of energy enough that you feel comfortable with it. Action is also a mathematical construct. It may be new to you, but you'll soon get to know it better.

The principle of least action

In spirit, the principle of least action arises from the idea that nature does things as efficiently as possible[3]. In the present context, this means that nature tries to minimize the action. Let's make this a little bit more precise. (We'll make it fully precise in the next section.)

[3] If you'd prefer a philosophy that's less new-agey, I'd say that basically any quantitative problem can be characterized as an optimization problem.

The type of problem that the action principle addresses is a *boundary value problem*, where you know the values of $\gamma(t)$ at the endpoints t_0 and t_1, and you want to determine what happens at all the t-values in the middle. So let $\mathbf{x}_0$ be the initial position, and let $\mathbf{x}_1$ be the terminal position. We want to restrict our attention to paths with these initial and terminal positions, i.e. such that

$$\gamma(t_0) = \mathbf{x}_0, \qquad \gamma(t_1) = \mathbf{x}_1. \tag{3.4}$$

From now on, let's suppose that the boundary data t_0, t_1, $\mathbf{x}_0$, and $\mathbf{x}_1$ are fixed, and that, whenever we talk about "paths", we are actually talking about paths that satisfy the boundary values (3.4).

The *principle of least action* (or, more succinctly, the *action principle*) says that the paths that actually occur are those that are critical points[4] of the action function S. There are technical issues here[5], but this is a fantastic idea. The simplest principle of calculus, that you can optimize a function by looking for critical points, turns out to be a fundamental principle of nature.

[4] So, yes, the term "least action" is a bit of a misnomer. It could be a local minimum but not a global minimum, and it could even be a critical point that isn't a local minimum. For this reason, you could argue that the term "action principle" is the better option.

[5] E.g., how do you find the critical points when there aren't a finite number of variables?

Coordinate changes and constraints

For the time being, let's take for granted that the principle of least action can be made rigorous, in that there is a way to define critical points of S, and that there is an effective way to solve for them. And then let's consider what advantages the Lagrangian perspective might have over the Newtonian and Hamiltonian perspectives.

One thing is that it is very geometrically and conceptually vivid. You can imagine a whole bunch of different paths[6] satisfying (3.4) and know that there's an objective way to identify certain paths as being better than the others.

[6] If there's only one space dimension, then you can graph the paths in the tx-plane.

More concretely, an important advantage is that the mechanics of the system are completely encoded in the function L. This makes it very easy to perform changes of coordinates and to impose constraints. To illustrate this, let's look back at the example of the pendulum of length ℓ, which we considered in Section 2.6 from the Hamiltonian perspective, and let's now see how it would work in the Lagrangian formalism.

In this case, the position of the pendulum is described by coordinates (x, y), so the coordinates on velocity space are (x, y, v_x, v_y). The potential energy is mgy, so the Lagrangian is

$$L = \text{Kinetic} - \text{Potential} = \frac{m}{2}(v_x^2 + v_y^2) - mgy.$$

The question then becomes how to impose the constraint $\|\gamma(t)\| = \ell$. One possibility is to do constrained optimization (i.e. use Lagrange multipliers).

But I'm going to guess that you don't want to use Lagrange multipliers[7]. Another possible approach is to do something similar to what we did in Section 2.6, and first switch to polar coordinates. We'll again use the equations

$$x = r\sin\theta, \qquad\qquad y = -r\cos\theta$$

and view these equations as defining a map ψ from the (r, θ)-plane to the (x, y)-plane:

$$\psi(r, \theta) = (r\sin\theta, -r\cos\theta).$$

Then

$$D_\psi = \begin{bmatrix} \sin\theta & r\cos\theta \\ -\cos\theta & r\sin\theta \end{bmatrix},$$

so, according to the rule (3.1) for how velocities transform, we would have

$$v_x = (\sin\theta)v_r + r(\cos\theta)v_\theta, \qquad v_y = -(\cos\theta)v_r + r(\sin\theta)v_\theta.$$

If you use these equations to write out $v_x^2 + v_y^2$ and simplify, you'll get $v_x^2 + v_y^2 = v_r^2 + r^2 v_\theta^2$. So, under this transformation, the Lagrangian becomes

$$L = \frac{m}{2}(v_r^2 + r^2 v_\theta^2) + mgr\cos\theta.$$

We then impose the constraints $r = \ell$ and $v_r = 0$ to get

$$L = \frac{m\ell^2 v_\theta^2}{2} + mg\ell\cos\theta. \tag{3.5}$$

We could now use this Lagrangian to define the action S on the space of paths in the one-dimensional space with coordinate θ and solve for critical points of this action. We'll do this later[8].

[7] Although, it actually is a pretty good way to solve the problem.

[8] Well, we won't actually solve, but we'll use the action to derive a differential equation that the critical points satisfy.

When it was first discovered, some of the big successes of Lagrangian mechanics came from its ability to accommodate coordinate changes and constraints. This allowed people to solve various mechanical systems, such as the double pendulum, for which Newton's equations were intractable.

Exercises

1. The action $S(\gamma)$ isn't just a number; it has units.

   (a) Explain why $S(\gamma)$ has units of (energy)(time).

   (b) Show that (energy)(time) = (position)(momentum).

2. To do Lagrangian mechanics, we'll need to learn mathematical tools for minimizing functions on spaces of paths. These tools can apply to minimization problems that aren't of a mechanical nature as well, and perhaps our intuition for these problems can help us gain intuition for the action principle.

   (a) You are flying from Boston to London. The flight should (approximately) follow a path that minimizes the distance. But the in-flight display shows a path on the map that seems to curve. Why does this happen?

   (b) You hold up two ends of a string and look at how it hangs. What quantity is being minimized?

3. Suppose we throw a ball in the air. Let $\gamma(t)$ describe the height of the ball as a function of time. Suppose the initial and terminal heights are $\gamma(t_0) = x_0$ and $\gamma(t_1) = x_1$.

   (a) Use what you know from basic calculus or physics to find the formula for $\gamma(t)$.

   (b) Plug your answer from (a) into the Lagrangian to get the function $L(\gamma(t), \gamma'(t))$. Pick some arbitrary values for all of the constants and graph $L(\gamma(t), \gamma'(t))$ as a function of t.

   (c) The trajectory of the ball roughly breaks up into three "epochs". In the first epoch, the ball has low height and high speed. In the second epoch, it has high height and low speed. And in the third epoch, it has low height and high speed again. Rank the epochs in terms of how high $L(\gamma(t), \gamma'(t))$ is. Which epoch lasts the longest?

4. A *spherical pendulum* is a pendulum that is free to swing in all directions, not just back and forth. Consider a spherical pendulum of length ℓ, with mass m completely concentrated at the end of the pendulum.

(a) In terms of Cartesian coordinates (x, y, z) and their velocities, write down the Lagrangian for the spherical pendulum.

(b) Use the spherical coordinate transformation $(\rho, \theta, \varphi) \mapsto (\rho\cos\theta\sin\varphi, \rho\sin\theta\sin\varphi, -\rho\cos\varphi)$ to rewrite the Lagrangian.

(c) Impose the constraints $\rho = \ell$ and $v_\rho = 0$ to get the Lagrangian in terms of the angular coordinates and their velocities.

3.2 *The calculus of variations*

The definition of critical points

Let $f(\mathbf{x})$ be a function of n variables. By definition, a *critical point*[9] of f is a point $\mathbf{x}$ where $\nabla f = 0$. You are probably comfortable with this definition, since it closely resembles basic calculus, with ∇f being the multi-dimensional analogue of the derivative.

[9] Points where f is not differentiable are also considered critical points, but we'll avoid this issue by assuming f is differentiable.

But there's another way to think of critical points that is more geometrically intuitive, namely that a critical point is a point where f does not change with respect to an infinitesimal change in $\mathbf{x}$. In the one-variable case, the derivative $f'(x)$ is the rate of change, so $f'(x) = 0$ means that f is not changing. But in more than one variable, we need to consider directional derivatives and require that the directional derivative of f is zero in every direction.

The directional derivative in the direction of a vector $\mathbf{e}$ can be defined as a limit:

$$D_{\mathbf{e}} f(\mathbf{x}) = \lim_{h\to 0} \frac{f(\mathbf{x} + h\mathbf{e}) - f(\mathbf{x})}{h}.$$

If $\mathbf{x}$ is a point for which this limit is 0 for all $\mathbf{e}$, then $\mathbf{x}$ is a critical point of f. Note that directional derivatives are closely related to the gradient, via the formula

$$D_{\mathbf{e}} f(\mathbf{x}) = \nabla f|_{\mathbf{x}} \cdot \mathbf{e}. \tag{3.6}$$

From this, it follows that the definition in terms of the gradient is equivalent to the definition in terms of directional derivatives.

Let's now consider a function S on the space of paths. Since S is infinite-dimensional, there's no sensible way to compute a gradient. But we can still make sense of directional derivatives, as follows. Let $\gamma(t)$ be a path. An infinitesimal change in γ is given by another path $\varepsilon(t)$, which we interpret as describing the directions in which the points of γ are moving.

We can then define the directional derivative of S at γ in the direction of ε:

$$D_\varepsilon S(\gamma) = \lim_{h\to 0} \frac{S(\gamma + h\varepsilon) - S(\gamma)}{h}. \tag{3.7}$$

We will say that γ is a critical point of S if the directional derivative is zero for all ε. Note that, because we are restricting our attention to paths γ that satisfy the boundary values (3.4), we require that $\varepsilon(t_0) = \varepsilon(t_1) = 0$. This ensures that, as γ moves in the direction of ε, the boundary values don't change. If γ is a path such that $D_\varepsilon S(\gamma) = 0$ for all such ε, then γ is a critical point.

The Euler-Lagrange equations

Now that we have a well-definition of critical points of S, the next question is how to find the critical points. We're not actually going to calculate limits[10], are we? The *calculus of variations* provides us with a method for finding the critical points. We will now derive this method.

[10] Actually, in some cases, it's not that bad to calculate the limits.

The first step is to use the fact that S is of the form (3.3) for some function $L(\mathbf{x}, \mathbf{v})$. Putting this into (3.7), we get

$$D_\varepsilon S(\gamma) = \lim_{h\to 0} \frac{1}{h} \int_{t_0}^{t_1} L\left(\gamma(t) + h\varepsilon(t), \gamma'(t) + h\varepsilon'(t)\right) - L\left(\gamma(t), \gamma'(t)\right) dt.$$

We can move the limit inside the integral[11] to get

$$D_\varepsilon S(\gamma) = \int_{t_0}^{t_1} \lim_{h\to 0} \frac{L\left(\gamma(t) + h\varepsilon(t), \gamma'(t) + h\varepsilon'(t)\right) - L\left(\gamma(t), \gamma'(t)\right)}{h} dt. \tag{3.8}$$

[11] A theorem of analysis says that this is okay under hypotheses which are satisfied, for example, when the integrand has continuous partial derivatives with respect to h and t.

Observe that, for any particular value of t, the limit inside the integral is just a directional derivative of L:

$$\lim_{h\to 0} \frac{L\left(\gamma(t) + h\varepsilon(t), \gamma'(t) + h\varepsilon'(t)\right) - L\left(\gamma(t), \gamma'(t)\right)}{h} = D_{(\varepsilon(t), \varepsilon'(t))} L\left(\gamma(t), \gamma'(t)\right).$$

This is great, because we've now reduced the infinite-dimensional problem of calculating the directional derivatives of S to the finite-dimensional problem of calculating directional derivatives of L. Using (3.6), we have

$$D_{(\varepsilon(t), \varepsilon'(t))} L(\gamma(t), \gamma'(t)) = \nabla L|_{(\gamma(t), \gamma'(t))} \cdot (\varepsilon(t), \varepsilon'(t)).$$

Putting this into (3.8), we get

$$D_\varepsilon S(\gamma) = \int_{t_0}^{t_1} \nabla L|_{(\gamma(t), \gamma'(t))} \cdot (\varepsilon(t), \varepsilon'(t)) dt. \tag{3.9}$$

Now let's write everything out in coordinates. The coordinates on velocity space are $(x_1, \ldots, x_n, v_1, \ldots, v_n)$, and L is a function of those variables. The paths $\gamma(t)$ and $\varepsilon(t)$ are in position space, so they can be written out as $\gamma(t) = (\gamma_1(t), \ldots, \gamma_n(t))$ and $\varepsilon(t) = (\varepsilon_1(t), \ldots, \varepsilon_n(t))$. We can then expand out the dot product in (3.9) to get[12]

[12] Keep in mind that in this formula, the partial derivatives of L are supposed to be evaluated at $(\gamma(t), \gamma'(t))$. We'll continue suppressing this in the notation in what follows.

$$\begin{aligned} D_\varepsilon S(\gamma) &= \int_{t_0}^{t_1} \frac{\partial L}{\partial x_1}\varepsilon_1(t) + \cdots + \frac{\partial L}{\partial x_n}\varepsilon_n(t) + \frac{\partial L}{\partial v_1}\varepsilon_1'(t) + \cdots + \frac{\partial L}{\partial v_n}\varepsilon_n'(t)dt \\ &= \sum_{i=1}^{n}\left(\int_{t_0}^{t_1} \frac{\partial L}{\partial x_i}\varepsilon_i(t)dt + \int_{t_0}^{t_1} \frac{\partial L}{\partial v_i}\varepsilon_i'(t)dt\right). \end{aligned} \tag{3.10}$$

Let's focus on the terms of the form

$$\int_{t_0}^{t_1} \frac{\partial L}{\partial v_i}\varepsilon_i'(t)dt.$$

We can do integration by parts on these. Set

$$u = \frac{\partial L}{\partial v_i}, \qquad dv = \varepsilon_i'(t)dt,$$

so

$$du = \frac{d}{dt}\left[\frac{\partial L}{\partial v_i}\right], \qquad v = \varepsilon_i(t).$$

So by integration by parts,

$$\begin{aligned} \int_{t_0}^{t_1} \frac{\partial L}{\partial v_i}\varepsilon_i'(t)dt &= \left[\frac{\partial L}{\partial v_i}\varepsilon_i(t)\right]_{t_0}^{t_1} - \int_{t_0}^{t_1} \frac{d}{dt}\left[\frac{\partial L}{\partial v_i}\right]\varepsilon_i(t)dt \\ &= -\int_{t_0}^{t_1} \frac{d}{dt}\left[\frac{\partial L}{\partial v_i}\right]\varepsilon_i(t)dt, \end{aligned} \tag{3.11}$$

where in the last step we used the assumption that $\varepsilon(t_0) = \varepsilon(t_1) = 0$. Putting (3.11) into (3.10), we get

$$\begin{aligned} D_\varepsilon S(\gamma) &= \sum_{i=1}^{n}\left(\int_{t_0}^{t_1} \frac{\partial L}{\partial x_i}\varepsilon_i(t)dt - \int_{t_0}^{t_1} \frac{d}{dt}\left[\frac{\partial L}{\partial v_i}\right]\varepsilon_i(t)dt\right) \\ &= \sum_{i=1}^{n}\left(\int_{t_0}^{t_1}\left(\frac{\partial L}{\partial x_i} - \frac{d}{dt}\left[\frac{\partial L}{\partial v_i}\right]\right)\varepsilon_i(t)dt\right). \end{aligned}$$

In order for γ to be a critical point, we need $D_\varepsilon S(\gamma)$ to be zero *for all* ε. In order for this to be true, it is necessary that

$$\frac{\partial L}{\partial x_i} - \frac{d}{dt}\left[\frac{\partial L}{\partial v_i}\right] = 0 \tag{3.12}$$

for all i and all t. Otherwise, it would be possible to choose ε_i to be zero at most t-values and "pop up" for a bit in the vicinity of a t-value where (3.12) does not hold, giving a nonzero integral.

The equations (3.12) are called the *Euler-Lagrange equations*. What we have shown is that a path γ is a critical point of S if and only if γ satisfies the Euler-Lagrange equations. Here are a couple of simple examples where the Euler-Lagrange equations are put to use.

Example 3.2.1 (Gravity near earth). Consider the case of gravity near earth, with one spatial dimension x representing the height. Then

$$L(x, v) = \frac{mv^2}{2} - mgx,$$

so

$$\frac{\partial L}{\partial x} = -mg, \qquad \frac{\partial L}{\partial v} = mv.$$

When evaluated at $(\gamma(t), \gamma'(t))$, the partial derivatives of L become

$$\frac{\partial L}{\partial x} = -mg, \qquad \frac{\partial L}{\partial v} = m\gamma'(t).$$

Since $n = 1$ in this case, there is only one Euler-Lagrange equation

$$-mg - \frac{d}{dt}[m\gamma'(t)] = 0,$$

which can be simplified to

$$-g = \gamma''(t).$$

This is Newton's Law!

Example 3.2.2 (The shortest distance between two points). The calculus of variations is valid for any function on the space of paths that can be expressed in the form (3.3), even if it doesn't come from a mechanics problem. For example, in two spatial dimensions (x, y), consider the Lagrangian

$$L(x, y, v_x, v_y) = \sqrt{v_x^2 + v_y^2}. \tag{3.13}$$

The corresponding action

$$S(\gamma) = \int_{t_0}^{t_1} \sqrt{\gamma_1'(t)^2 + \gamma_2'(t)^2}\, dt$$

calculates the arclength of γ. Therefore, the Euler-Lagrange equations will allow us to determine which paths γ minimize arclength[13].

[13] Of course, the paths that minimize arclength are straight lines, but I bet you've never seen a proof!

The partial derivatives of L are

$$\frac{\partial L}{\partial x} = \frac{\partial L}{\partial y} = 0, \qquad \frac{\partial L}{\partial v_x} = \frac{v_x}{\sqrt{v_x^2 + v_y^2}}, \qquad \frac{\partial L}{\partial v_y} = \frac{v_y}{\sqrt{v_x^2 + v_y^2}}.$$

When evaluated at $(\gamma_1(t), \gamma_2(t), \gamma_1'(t), \gamma_2'(t))$, these become

$$\frac{\partial L}{\partial x} = \frac{\partial L}{\partial y} = 0, \qquad \frac{\partial L}{\partial v_x} = \frac{\gamma_1'(t)}{\|\gamma'(t)\|}, \qquad \frac{\partial L}{\partial v_y} = \frac{\gamma_2'(t)}{\|\gamma'(t)\|}.$$

The Euler-Lagrange equations are then

$$\frac{d}{dt}\left[\frac{\gamma_1'(t)}{\|\gamma'(t)\|}\right] = 0, \qquad \frac{d}{dt}\left[\frac{\gamma_2'(t)}{\|\gamma'(t)\|}\right] = 0. \tag{3.14}$$

Note that arclength doesn't depend on the parametrization, so without loss of generality, we can assume that the paths are parametrized to unit speed, i.e. $\|\gamma'(t)\| = 1$. This allows us to simplify (3.14) to

$$\gamma_1''(t) = 0, \qquad \gamma_2''(t) = 0,$$

from which we conclude that $\gamma'(t)$ is constant, so $\gamma(t)$ is a straight line.

Exercises

1. In Section 3.1, we saw that the pendulum can be described by a one-dimensional position space with coordinate θ, where the Lagrangian is (3.5). In this case, write out the Euler-Lagrange equation and explain how it is equivalent to Newton's Law $F = ma$.

2. For simplicity, consider the case of an object moving in a one-dimensional position space (although the results of this problem are true in any number of dimensions), under a force with potential energy $V(x)$. In this case, the Lagrangian is $L = \frac{mv^2}{2} - V(x)$. Write out the Euler-Lagrange equation in this case, and explain why it is the correct equation of motion.

3. Again for simplicity, consider the case of a Lagrangian system in a one-dimensional position space.

   (a) Show that, if $\gamma(t)$ satisfies the Euler-Lagrange equation, then it satisfies the property that $B(t) := \gamma'(t)\frac{\partial L}{\partial v} - L(\gamma(t), \gamma'(t))$ is constant. Note that here, $\frac{\partial L}{\partial v}$ is supposed to be evaluated at $(\gamma(t), \gamma'(t))$. [Hint: frequently the easiest way to prove that something is constant is to show that its derivative is 0.]

   (b) For an object acted on by a conservative force with potential energy $V(x)$, the Lagrangian is $L = \frac{mv^2}{2} - V(x)$. In this case, calculate $B(t)$ and give a simple physical interpretation of it.

4. The coordinates on the standard map of the earth represent latitude and longitude, which are essentially the angular components of spherical coordinates. Specifically, let θ represent longitude, and let ϕ represent latitude, where $-\pi/2 \leq \phi \leq \pi/2$. Note that θ "wraps around", so for our purposes there's no need to worry about bounds[14]. Then a path on the earth can be represented by a parametrized curve $\gamma(t) = (\gamma_1(t), \gamma_2(t))$ on the (θ, ϕ)-plane.

   In units where the radius of the earth is 1, the arclength of a path is given by the formula

   $$S(\gamma) = \int_{t_0}^{t_1} \sqrt{\cos^2(\gamma_2(t))\gamma_1'(t)^2 + \gamma_2'(t)^2}\, dt.$$

   Use the calculus of variations to find differential equations satisfied by arclength-minimizing paths.

[14] An increase in θ represents eastward movement, and it's totally possible to go east by any amount, possibly circling the earth a number of times. But the bounds for ϕ are important, because if you travel north or south you eventually reach a pole.

3.3 *Noether's Theorem (version 1)*

We are finally ready to prove a preliminary version of Noether's Theorem, which roughly says that symmetries lead to conserved quantities. Later, we will prove a more general result.

Derivation of Noether's Theorem

Suppose you have a Lagrangian system. Let ψ_s be a one-parameter group of transformations of position space, with associated vector field $\mathbf{G}$.

Using the rule (3.1) for how velocity vectors transform, we can extend ψ_s to a one-parameter group of transformations $\tilde{\psi}_s$ of velocity space, given by

$$\tilde{\psi}_s(\mathbf{x}, \mathbf{v}) = (\psi_s(\mathbf{x}), D_{\psi_s}\mathbf{v}).$$

We will say that ψ_s is a *symmetry* if $\tilde{\psi}_s$ preserves L, in the sense that

$$L(\tilde{\psi}_s(\mathbf{x}, \mathbf{v})) = L(\mathbf{x}, \mathbf{v}) \tag{3.15}$$

for all s, $\mathbf{x}$, and $\mathbf{v}$.

Let $\tilde{\mathbf{G}}$ denote the vector field that generates $\tilde{\psi}_s$. Using the result from Exercise 5 in Section 1.5, we see that, if ψ_s is a symmetry, then $\tilde{\mathbf{G}}(L) = 0$. Let's explore the implications of this fact.

Let $\mathbf{G}$ denote the vector field that generates ψ_s. Using (2.32), we can obtain a simple formula for $\tilde{\mathbf{G}}$ in terms of $\mathbf{G}$:

$$\tilde{\mathbf{G}} = (\mathbf{G}, D_{\mathbf{G}}\mathbf{v}).$$

If we write out $\mathbf{G}$ in coordinates as $\mathbf{G} = (G_1, \ldots, G_n)$, then we get

$$D_{\mathbf{G}}\mathbf{v} = \begin{bmatrix} \frac{\partial G_1}{\partial x_1} & \cdots & \frac{\partial G_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial G_n}{\partial x_1} & \cdots & \frac{\partial G_n}{\partial x_n} \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} \nabla G_1 \cdot \mathbf{v} \\ \vdots \\ \nabla G_n \cdot \mathbf{v} \end{bmatrix},$$

so we obtain an expression for $\tilde{\mathbf{G}}$ in coordinates:

$$\tilde{\mathbf{G}} = (G_1, \ldots, G_n, \nabla G_1 \cdot \mathbf{v}, \ldots, \nabla G_n \cdot \mathbf{v}).$$

We can then use this to expand out the equation $\tilde{G}(L) = 0$ as

$$\sum_{i=1}^{n} \left(G_i \frac{\partial L}{\partial x_i} + (\nabla G_i \cdot \mathbf{v}) \frac{\partial L}{\partial v_i} \right) = 0. \tag{3.16}$$

Now assume that the left side of (3.16) is being evaluated at $(\gamma(t), \gamma'(t))$, where γ is assumed to satisfy the Euler-Lagrange equations (3.12). Then the above equation becomes

$$\sum_{i=1}^{n} \left(G_i \frac{d}{dt}\left[\frac{\partial L}{\partial v_i}\right] + (\nabla G_i \cdot \gamma'(t)) \frac{\partial L}{\partial v_i} \right) = 0, \tag{3.17}$$

where all of the functions[15] are now being evaluated at $\mathbf{x} = \gamma(t)$, $\mathbf{v} = \gamma'(t)$.

[15] G_i and its partial derivatives, and the partial derivatives of L

The chain rule tells us that

$$\frac{d}{dt}[G_i] = \nabla G_i \cdot \gamma'(t).$$

Putting this into (3.17) and using the product rule, we get

$$\begin{aligned} 0 &= \sum_{i=1}^{n} \left(G_i \frac{d}{dt}\left[\frac{\partial L}{\partial v_i}\right] + \frac{d}{dt}\left[G_i\right]\frac{\partial L}{\partial v_i}\right) \\ &= \sum_{i=1}^{n} \frac{d}{dt}\left[G_i \frac{\partial L}{\partial v_i}\right] \\ &= \frac{d}{dt}\left[\sum_{i=1}^{n} G_i \frac{\partial L}{\partial v_i}\right], \end{aligned}$$

so the quantity $\sum_{i=1}^{n} G_i \frac{\partial L}{\partial v_i}$ is constant with respect to time.

To summarize the above discussion: If ψ_s is a symmetry of a Lagrangian system, and if $\mathbf{G}$ is the vector field that generates ψ_s, then

$$\sum_{i=1}^{n} G_i \frac{\partial L}{\partial v_i} \tag{3.18}$$

is constant for any path γ that satisfies the Euler-Lagrange equations. This is *Noether's Theorem*[16].

[16] It took some work to get here, but the result is very concise!

Noether's Theorem in examples

Before applying Noether's Theorem to specific examples of Lagrangian systems, let's first consider the most important examples of possible symmetries and see what the corresponding conserved quantities are. For simplicity of notation, I'll assume a 2-dimensional position space. But the results are similar in 3 or more dimensions.

- *Translations*: Translations in the x direction[17] can be described by the one-parameter group $\psi_s(x, y) = (x + s, y)$. The total derivative of ψ_s is the identity matrix, so the extension to velocity space is $\tilde{\psi}_s(x, y, v_x, v_y) = (x + s, y, v_x, v_y)$. Thus, anytime L does not depend on x, translations in the x direction will form a symmetry.

  The vector field that generates ψ_s is $\mathbf{G} = (1, 0)$, so, if x-translations form a symmetry, the corresponding conserved quantity (3.18) is $\frac{\partial L}{\partial v_x}$.

  In the standard case of an object being acted on by a conservative force, where the Lagrangian is of the form

  $$L = \frac{m\|\mathbf{v}\|^2}{2} - V(\mathbf{x}), \tag{3.19}$$

  the conserved quantity associated to x-translation is mv_x, i.e. the x-component of momentum.

[17] Everything here is similar for translations in the y or z direction.

- *Rotations:* In two dimensions, rotations of the plane can be described by the one-parameter group $\psi_s(\mathbf{x}) = U_s\mathbf{x}$, where U_s is the

rotation matrix (1.2) (with s in place of θ). The total derivative of ψ_s is U_s, so the extension to velocity space is

$$\tilde{\psi}_s(\mathbf{x}, \mathbf{v}) = (U_s\mathbf{x}, U_s\mathbf{v}).$$

In other words, the position and velocity vectors are rotated in exactly the same amount. If L does not change under such a transformation[18], then rotations will form a symmetry.

Recall that the vector field that generates ψ_s is $\mathbf{G} = (-y, x)$. If rotations form a symmetry, then corresponding conserved quantity is $-y\frac{\partial L}{\partial v_x} + x\frac{\partial L}{\partial v_y}$.

In the case where the Lagrangian is of the form (3.19), the conserved quantity associated to rotation is $m(xv_y - yv_x)$, which is angular momentum.

[18] In particular, this would be the case if L only depends on the magnitudes of $\mathbf{x}$ and $\mathbf{v}$.

Example 3.3.1 (Gravity near earth). Consider near-earth gravity in three spatial dimensions, where the Lagrangian is

$$L = \frac{m(v_x^2 + v_y^2 + v_z^2)}{2} - mgz.$$

Since L does not depend on x or y, translations in the x and y directions form symmetries. Additionally, since L only depends on z and the length of the velocity vector $\mathbf{v}$, rotation about the z-axis forms a symmetry[19]. For each of these symmetries, we can recall the generating vector field and then use (3.18) to find the associated conserved quantity. The following table describes the results.

[19] As in the above discussion, a one-parameter group of rotations $\psi_s(\mathbf{x}) = U_s\mathbf{x}$ extends to velocity space as $\tilde{\psi}_s(\mathbf{x}, \mathbf{v}) = (U_s\mathbf{x}, U_s\mathbf{v})$. So if ψ_s rotates $\mathbf{x}$ about the z-axis, then $\tilde{\psi}_s$ rotates both $\mathbf{x}$ and $\mathbf{v}$ about the z-axis.

symmetry	vector field	conserved quantity
x-translation	$(1, 0, 0)$	$\frac{\partial L}{\partial v_x} = mv_x$
y-translation	$(0, 1, 0)$	$\frac{\partial L}{\partial v_y} = mv_y$
xy-rotation	$(-y, x, 0)$	$-y\frac{\partial L}{\partial v_x} + x\frac{\partial L}{\partial v_y} = -mv_xy + mv_yx$

Thus, the x- and y-components of momentum and the z-component of angular momentum are all conserved.

It's useful to have these conserved quantities — they explain why it's okay to consider horizontal motion separately from vertical motion. But it's a bit disappointing that none of the above conserved quantities involve z. For this, we'll need the more general version of Noether's Theorem that will come later.

Example 3.3.2 (The shortest distance between two points). Recall from Example 3.2.2 the Lagrangian

$$L = \sqrt{v_x^2 + v_y^2}$$

for which the corresponding action computes the arclength of a path in the plane. Since L does not depend on x or y, translations

in the x and y directions form symmetries. Additionally, since L only depends on the length of $\mathbf{v}$, rotation forms a symmetry[20]. The following table finds the associated conserved quantities.

symmetry	vector field	conserved quantity
x-translation	$(1,0)$	$\frac{\partial L}{\partial v_x} = \frac{v_x}{\sqrt{v_x^2+v_y^2}}$
y-translation	$(0,1)$	$\frac{\partial L}{\partial v_y} = \frac{v_y}{\sqrt{v_x^2+v_y^2}}$
rotation	$(-y,x)$	$y\frac{\partial L}{\partial v_x} + x\frac{\partial L}{\partial v_y} = \frac{-yv_x+xv_y}{\sqrt{v_x^2+v_y^2}}$

[20] These symmetries have a simple intuitive explanation: if you move a path around by translations and rotations, the arclength doesn't change.

According to Noether's Theorem, any path that minimizes arclength must be such that all three of the conserved quantities are constant. By dividing the second quantity by the first, we deduce that v_y/v_x is constant[21]. So let's write $v_y/v_x = a$.

Next, let's divide the third quantity by the first and call the result b:

$$b = -y + \frac{xv_y}{v_x} = -y + ax.$$

Solving for y, we get

$$y = ax - b,$$

which is a general equation of a line[22].

[21] At this point, you could say that this means the slope is constant, and therefore it must be a line. But in doing so, you would be solving a differential equation, albeit a very simple one. I want to demonstrate that, when you have enough conserved quantities, you don't need to solve *any* differential equations at all.

[22] Actually, this doesn't include vertical lines. What caused us to lose those solutions?

Compare what we did here with what we did in Example 3.2.2. The difference is that there, we needed to solve some differential equations at the end[23], whereas here, we were able to use conserved quantities to solve the problem using nothing but algebra.

[23] Admittedly, they weren't difficult differential equations, but this is a matter of principle.

Example 3.3.3 (Hyperbolic distance). Consider the Lagrangian

$$L = \frac{1}{y}\sqrt{v_x^2 + v_y^2}, \tag{3.20}$$

where position space is the *upper half-plane*, consisting of points (x,y) such that $y > 0$. The corresponding action calculates arclengths in hyperbolic geometry[24] Thus, the solutions to the Euler-Lagrange equations should describe paths that minimize hyperbolic arclength.

[24] If you aren't familiar with hyperbolic geometry, you can take this as the definition.

In the exercises, you'll find the Euler-Lagrange equations for this Lagrangian. However, you'll see that the Euler-Lagrange equations are nonlinear and not so easy to solve directly. Instead, we can look for symmetries and use the corresponding conserved quantities to help us find the solutions.

Since L does not depend on x, translations in the x direction form a symmetry. Noether's theorem then tells us that

$$\frac{\partial L}{\partial v_x} = \frac{v_x}{y\sqrt{v_x^2 + v_y^2}} \tag{3.21}$$

is conserved.

Another symmetry in this case is the one-parameter group of *dilations* $\psi_s(x,y) = (e^s x, e^s y)$. These transformations are linear, so they act on velocity in the same way:

$$\tilde{\psi}_s(x,y,v_x,v_y) = (e^s x, e^s y, e^s v_x, e^s v_y).$$

If you put $\tilde{\psi}_s(x,y,v_x,v_y)$ into L, all the factors of e^s cancel, which shows that dilations indeed form a symmetry.

The vector field that generates ψ_s is $\mathbf{G} = (x,y)$, so the corresponding conserved quantity is

$$x\frac{\partial L}{\partial v_x} + y\frac{\partial L}{\partial v_y} = \frac{xv_x + yv_y}{y\sqrt{v_x^2 + v_y^2}}. \tag{3.22}$$

Dividing (3.22) by (3.21), we see that

$$x + y\frac{v_y}{v_x} = a$$

for some constant a. Since $\frac{v_y}{v_x}$ is the same thing as the slope $\frac{dy}{dx}$, we get a differential equation that we can solve by separation of variables[25]:

$$\begin{aligned} x + y\frac{dy}{dx} &= a, \\ y\frac{dy}{dx} &= -(x-a), \\ \int y\,dy &= -\int (x-a)\,dx, \\ \frac{1}{2}y^2 &= -\frac{1}{2}(x-a)^2 + c, \\ y^2 + (x-a)^2 &= 2c. \end{aligned}$$

This is the equation of a circle centered at $(a,0)$. So we conclude that the paths that minimize hyperbolic distance are circles centered anywhere on the x-axis[26].

[25] If you're disappointed that we're solving a differential equation, you'll be cheered to learn that you'll find the solutions algebraically in the exercises.

[26] Actually, since the domain is the upper half-plane, they are semicircles.

Exercises

1. Throughout this problem, consider a Lagrangian system where L is of the form (3.19), with n space dimensions.

   (a) Consider the one-parameter group of translations in the direction of a vector $\mathbf{a}$:

   $$\psi_s(\mathbf{x}) = \mathbf{x} + s\mathbf{a}.$$

   Show that, if the potential energy V is preserved by ψ_s, then the quantity $m\mathbf{v}\cdot\mathbf{a}$ is conserved. This quantity can be interpreted as the momentum in the direction of $\mathbf{a}$.

(b) More generally, suppose ψ_s is a symmetry, and that $\mathbf{G}$ is the vector field that generates ψ_s. Show that the quantity $m\mathbf{v} \cdot \mathbf{G}$ is conserved.

2. Show that the conserved quantities found in Example 3.3.1 are sufficient to prove that, when you throw a ball, the x- and y-values satisfy a linear relationship of the form $ax + by = c$.

3. Consider the Lagrangian (3.20) that corresponds to hyperbolic arclength.

   (a) Write out the corresponding Euler-Lagrange equations and simplify them as much as you can. [Hint: It will simplify the calculations if you assume, without loss of generality, that the paths γ are parametrized to unit speed, i.e. $\|\gamma'(t)\| = 1$.]

   (b) Use the fact that the quantities (3.21) and (3.22) are conserved to *algebraically* show that the solutions to the Euler-Lagrange equations are circles centered anywhere on the x-axis. [Hint: Set (3.21) and (3.22) equal to constants, then solve out for v_x and v_y to obtain an equation that only involves x and y.]

   (c) Write down a formula for a path $\gamma(t)$ that parametrizes the semicircle centered at a with radius R. Show that $\gamma(t)$ satisfies the Euler-Lagrange equations. [Note: depending on whether you used the hint in part (a), you may need to normalize your parametrization so that it has unit speed.]

4. Let $\mathbf{F} = (F_1, \ldots, F_n)$ be a vector field in $\mathbb{R}^n$. (The F_i are functions that depend on the position variables x_i.) Consider the Lagrangian

$$L = F_1 v_1 + \ldots F_n v_n.$$

   (a) Show that, for a path $\gamma(t)$, the action $S(\gamma)$ is equal to the line integral of $\mathbf{F}$ along γ:

$$S(\gamma) = \int_{t_0}^{t_1} \mathbf{F}(\gamma(t)) \cdot \gamma'(t) dt = \int_\gamma \mathbf{F} \cdot d\gamma.$$

   (b) Write out the Euler-Lagrange equations in this case, and observe that, if $D_\mathbf{F}$ is symmetric, then the Euler-Lagrange equations are automatically satisfied *for all* paths γ. (If you need a reminder on $D_\mathbf{F}$, see Section 2.4.)

   (c) Use the result from (b) to explain why when $D_\mathbf{F}$ is symmetric, the line integral of $\mathbf{F}$ along γ does not change when γ is continuously deformed[27].

[27] This result is used to prove the converse of (2.21) when $\mathbf{F}$ is defined everywhere. The idea is that you can set $f(\mathbf{x}) = \int_\gamma \mathbf{F} \cdot d\gamma$, where γ is *any* path from 0 to $\mathbf{x}$. This definition is well-defined since any path can be continuously deformed into any other path without changing the integral. You can then use the FTC to show that $\nabla f = \mathbf{F}$.

3.4 *Some refinements of the Lagrangian approach*

Generalized momentum

In Section 3.3, we saw that, if a Lagrangian L is symmetric under translation in the x_i direction, then the corresponding conserved quantity is $\frac{\partial L}{\partial v_i}$. In the standard physical situation, where L is of the form (3.19), this conserved quantity is equal to the ith component of momentum, mv_i.

But in other cases, such as the Lagrangian (3.13) for arclength, the conserved quantity associated to translation-symmetry[28]. looks very different from the momentum formula. And, since this situation doesn't even describe the motion of a physical object, the concept of momentum doesn't even make sense.

[28] See Example 3.3.2

Nonetheless, it would be nice to be able to say, without caveats, that, in Lagrangian mechanics, translation symmetries correspond to conservation of momentum. We can actually do this by changing the definition of momentum[29]. *Generalized momentum* is the vector whose components are defined as

[29] In math, we get to define things however we want. This is the same cavalier spirit that led to the creation of imaginary numbers.

$$p_i = \frac{\partial L}{\partial v_i}. \tag{3.23}$$

This definition is justified by the fact that, in the cases where we know what momentum is, the generalized momentum is the same thing as momentum. But (3.23) makes sense in *any* Lagrangian system.

The formula (3.23) helps to explain why momentum transforms differently from velocity under coordinate changes. Momentum isn't actually mass times velocity by definition; it's defined by (3.23), and it just happens to be mass times velocity in the usual coordinate system.

The Beltrami identity

In Exercise 3 of Section 3.2, you showed (for a one-dimensional position space) that, if $\gamma(t)$ satisfies the Euler-Lagrange equation, then $B(t) = \gamma'(t)\frac{\partial L}{\partial v} - L(\gamma(t), \gamma'(t))$ is constant. We could express this fact concisely by saying that $B = v\frac{\partial L}{\partial v} - L$ is conserved.

In n dimensions, it can be similarly shown that

$$B = \left(\sum_{i=1}^{n} v_i \frac{\partial L}{\partial v_i}\right) - L \tag{3.24}$$

is conserved. This fact is called the *Beltrami identity*. But, in the standard physical case where L is of the form (3.19), the right side of (3.24) is the total energy. In other words, the Beltrami identity is the

manifestation, in Lagrangian mechanics, of the law of conservation of energy.

As with momentum, there isn't any sensible notion of energy for Lagrangian systems that are nonphysical. But who's going to stop us if we *define* the quantity B in (3.24) to be energy? In the next section, we will see that conservation of energy can be associated to the symmetry of time-translation.

In more advanced courses on mechanics, the Lagrangian perspective is seen as being more fundamental than the Hamiltonian one. For any given physical system, you're always supposed to start by writing down a Lagrangian L for which the Euler-Lagrange equations are the correct equations of motion. Then, if you want to pass to the Hamiltonian perspective, you would use (3.23) to rewrite (3.24) completely in terms of position and momentum, and the result would be the Hamiltonian function H that you would use for the system. We aren't going to explore this any further, but hopefully it provides more context for why we would want to define momentum and energy using (3.23) and (3.24). The philosophy is that everything comes from the Lagrangian.

Noether's Theorem (version 2)

We're now going to prove a more general version of Noether's Theorem than the one we saw in Section 3.3. As there, we consider a one-parameter group ψ_s of transformations of position space, and the extension $\tilde{\psi}_s$ to velocity space.

We will say that ψ_s is an *f-symmetry* for some s-dependent function $f_s(\mathbf{x}, t)$ if

$$L(\tilde{\psi}_s(\gamma(t), \gamma'(t))) = L(\gamma(t), \gamma'(t)) + \frac{d}{dt}\left[f_s(\gamma(t), t)\right] \tag{3.25}$$

for all s and all paths $\gamma(t)$. Note that f_s is allowed to explicitly depend on t.

To understand why this is a reasonable definition, let's look at how the action is affected by an f-symmetry. If you transform a path by ψ_s and then compute the action of the transformed path, you get

$$S(\psi_s(\gamma)) = \int_{t_0}^{t_1} L(\tilde{\psi}_s(\gamma(t), \gamma'(t)))dt.$$

If ψ_s is an f-symmetry, then we can sub in (3.25) to get

$$\begin{aligned} S(\psi_s(\gamma)) &= \int_{t_0}^{t_1} L(\gamma(t), \gamma'(t))dt + \int_{t_0}^{t_1} \frac{d}{dt}\left[f_s(\gamma(t), t)\right] dt \\ &= S(\gamma) + f_s(\gamma(t_1), t_1) - f_s(\gamma(t_0), t_0) \\ &= S(\gamma) + f_s(\mathbf{x}_1, t_1) - f_s(\mathbf{x}_0, t_0). \end{aligned}$$

Recall that the equations of motion come from fixing the boundary values $\gamma(t_0) = \mathbf{x}_0$ and $\gamma(t_1) = \mathbf{x}_1$ and then looking for critical points of the action. After we fix the boundary values, the terms $f_s(\mathbf{x}_1, t_1) - f_s(\mathbf{x}_0, t_0)$ become constant, so they don't affect the critical points. The conclusion is that f-symmetries transform space in a way that doesn't affect the Euler-Lagrange equations, making them a reasonable notion of symmetry.

But the real justification of the definition comes from the fact that conserved quantities can be obtained from f-symmetries[30]. To do this, take the derivative of both sides of (3.25) with respect to s and then evaluate at $s = 0$. Using the chain rule, the left side becomes

$$\begin{aligned}\frac{d}{ds}|_{s=0}\left[L(\tilde{\psi}_s(\gamma(t), \gamma'(t)))\right] &= \nabla L \cdot \left(\frac{d}{ds}|_{s=0}\left[\tilde{\psi}_s(\gamma(t), \gamma'(t))\right]\right) \\ &= (\nabla L \cdot \tilde{G}) \\ &= \tilde{G}(L),\end{aligned}$$

where $\tilde{G} = \frac{d}{ds}|_{s=0}\left[\tilde{\psi}_s(\mathbf{x}, \mathbf{v})\right]$ is the vector field that generates $\tilde{\psi}_s$, and where it is understood that everything is evaluated at $\mathbf{x} = gamma(t)$, $\mathbf{v} = \gamma'(t)$. On the right side of (3.25), the term $L(\gamma(t), \gamma'(t))$ doesn't depend on s, so its derivative is 0. This gives us the equation

$$\tilde{G}(L) = \frac{d}{dt}\left[\frac{d}{ds}|_{s=0}[f_s]\right]. \tag{3.26}$$

In Section 3.3, we saw that, when you expand $\tilde{G}(L)$ in coordinates and use the Euler-Lagrange equations, then you can rewrite it as

$$\tilde{G}(L) = \sum_{i=1}^{n} \frac{d}{dt}\left[G_i \frac{\partial L}{\partial v_i}\right].$$

Putting this into (3.26), we get

$$\sum_{i=1}^{n} \frac{d}{dt}\left[G_i \frac{\partial L}{\partial v_i}\right] = \frac{d}{dt}\left[\frac{d}{ds}|_{s=0}[f_s]\right].$$

Therefore,

$$\frac{d}{dt}\left[\sum_{i=1}^{n} G_i \frac{\partial L}{\partial v_i} - \frac{d}{ds}|_{s=0}[f_s]\right] = 0,$$

so the quantity

$$\sum_{i=1}^{n} G_i \frac{\partial L}{\partial v_i} - \frac{d}{ds}|_{s=0}[f_s] \tag{3.27}$$

is conserved.

Note that we can view the version of Noether's Theorem from Section 3.3 as a special case of this one, where $f_s = 0$. You'll see a simple example where $f_s \neq 0$ in the exercises. But the most interesting example will appear from Galilean boosts, in the next section.

[30] At this point, we're almost treating Noether's Theorem as an axiom, saying that a transformation deserves to be called a symmetry if you can get a conserved quantity from it.

Exercises

1. Consider near-earth gravity, where the Lagrangian is

$$L = \frac{m\|\mathbf{v}\|^2}{2} - mgz.$$

   (a) Consider the one-parameter group $\psi_s(x, y, z) = (x, y, z + s)$ that translates in the z-direction. Show that ψ_s forms an f-symmetry with $f_s(\mathbf{x}, t) = -mgts$.

   (b) Find the conserved quantity associated to the f-symmetry in part (a). (The answer should be something that "feels" right.)

   (c) Is rotation about the x-axis an f-symmetry? If so, find the associated conserved quantity.

2. Consider the Lagrangian $L = \sqrt{v_x^2 + v_y^2}$, for which the corresponding action calculates arclength.

   (a) Find formulas for the generalized momenta p_x and p_y in this case.

   (b) Explain why $p_x = \cos\theta$ and $p_y = \sin\theta$, where θ is the polar angle of the velocity vector. So in this case, conservation of momentum implies that the direction of the velocity vector is constant.

   (c) Find a formula for the energy in this case. The answer might seem a bit disappointing, but there's something useful in it. It tells us that it isn't possible to study arclength from the Hamiltonian perspective.

3.5 *Lagrangian mechanics in spacetime*

Spacetime paths and velocity spacetime

So far, we have not been able to incorporate symmetries of spacetime (specifically, time-translation and Galiliean boosts) in the framework of Noether's Theorem. We will now do so by considering Lagrangian mechanics in spacetime.

The idea is simply that we will now let the $(n+1)$-dimensional spacetime with coordinates $(\mathbf{x}, t)$ play the same role that n-dimensional space previously played. We will consider paths $\alpha(\tau)$ in spacetime, where τ is playing the role that t previously played, as a parameter for the paths[31]. It will sometimes be helpful to separate a path α into its space and time components as $\alpha(\tau) = (\alpha_{\mathbf{x}}(\tau), \alpha_t(\tau))$.

[31] But keep in mind that τ is just an auxilliary variable without any physical meaning.

A path $\gamma(t)$ in space can be identified with a spacetime path

$$\hat{\gamma}(\tau) = (\gamma(\tau), \tau), \tag{3.28}$$

which traces out the worldline of $\gamma(t)$. As a graph of a map, it satisfies something like a "vertical line test", where the path only has one point at each t-value. In contrast, the general spacetime paths α aren't constrained in this way. They can go forward and backward in time and do all kinds of crazy stuff.

Instead of velocity space, we now have a velocity spacetime, with coordinates $(\mathbf{x}, t, \mathbf{u}, \lambda)$, whose points we think of as vectors in spacetime. Note that we are now using $\mathbf{u}$, not $\mathbf{v}$, to denote the velocity associated to $\mathbf{x}$. The reason is that $\mathbf{u}$ plays the role of the rate of change of $\mathbf{x}$ with respect to τ, which is not time. The physically observable velocity is[32]

$$\mathbf{v} = \frac{1}{\lambda}\mathbf{u}.$$

[32] When you plug in paths, this equation becomes $d\mathbf{x}/dt = \frac{d\mathbf{x}/d\tau}{dt/d\tau}$.

Given a spacetime path $\alpha(\tau) = (\alpha_{\mathbf{x}}(\tau), \alpha_t(\tau))$, we get a corresponding path in velocity spacetime:

$$(\alpha(\tau), \alpha'(\tau)) = (\alpha_{\mathbf{x}}(\tau), \alpha_t(\tau), \alpha'_{\mathbf{x}}(\tau), \alpha'_t(\tau)).$$

In particular, if $\gamma(t)$ is a path in space, then the spacetime path (3.28) leads to the velocity spacetime path

$$(\hat{\gamma}(\tau), \hat{\gamma}'(\tau)) = (\gamma(\tau), \tau, \gamma'(\tau), 1). \tag{3.29}$$

Spacetime actions

We will now consider Lagrangians $\mathbb{L}$ which are functions on velocity spacetime. For any path α, we can evaluate a Lagrangian at $(\alpha(\tau), \alpha'(\tau))$ and integrate to get the corresponding action:

$$\mathbb{S}(\alpha) = \int_{\tau_0}^{\tau_1} \mathbb{L}(\alpha_{\mathbf{x}}(\tau), \alpha_t(\tau), \alpha'_{\mathbf{x}}(\tau), \alpha'_t(\tau))d\tau.$$

Note that, mathematically, there's absolutely no difference between $(n+1)$-dimensional spacetime and $(n+1)$-dimensional space; the only difference is that the variables are labeled differently. Thus, all of the mathematical results from before — the calculus of variations, the Euler-Lagrange equations, and both versions of Noether's Theorem — remain valid in this setting.

What is not so clear yet is whether it is possible to construct spacetime Lagrangians that are physically meaningful, i.e. for which the Euler-Lagrange equations actually describe the equations of motion for physical systems. But we can, as follows. Given a Lagrangian $L(\mathbf{x}, \mathbf{v})$ of the kind that we studied in the previous section, we can construct a Lagrangian $\hat{L}$ on velocity spacetime, defined by the formula

$$\hat{L}(\mathbf{x}, t, \mathbf{u}, \lambda) = \lambda L\left(\mathbf{x}, \frac{1}{\lambda}\mathbf{u}\right). \tag{3.30}$$

The associated action is

$$\hat{S}(\alpha) = \int_{\tau_0}^{\tau_1} \alpha_t'(\tau) L\left(\alpha_{\mathbf{x}}(\tau), \frac{1}{\alpha_t'(\tau)}\alpha_{\mathbf{x}}'(\tau)\right) d\tau. \tag{3.31}$$

The action (3.31) satisfies the following important properties:

1. $\hat{S}(\hat{\gamma}) = S(\gamma)$ for any ordinary path $\gamma(t)$. To see this, we would simply plug (3.28) into (3.31). Specifically, we would substitute $\alpha_{\mathbf{x}}(\tau) = \gamma(\tau)$, $\alpha_t(\tau) = \tau$, $\alpha_{\mathbf{x}}'(\tau) = \gamma'(\tau)$, and $\alpha_t'(\tau) = 1$, to get

$$\hat{S}(\hat{\gamma}) = \int_{\tau_0}^{\tau_1} L\left(\gamma(\tau), \gamma'(\tau)\right) d\tau = S(\gamma).$$

2. $\hat{S}$ is parametrization-independent[33], meaning that it only depends on the spacetime path α itself and not on the way the path is parametrized in terms of τ. This is important because a path in spacetime already includes all of the physical information about an object's trajectory, and the τ is a meaningless parameter.

[33] I won't give a detailed proof, but it basically involves showing that, if you make a substitution $\tau = f(\mu)$ in (3.31), you end up with a formula that looks exactly the same as before, but with μ in the place of τ. This is similar to the proof in multivariable calculus that the line integral of a vector field is parametrization-independent.

The above properties are mostly of a conceptual nature; they help to explain why (3.30) is the correct formula to use for the spacetime Lagrangian. Mathematically, the most relevant property is the following: $\hat{\gamma}$ satisfies the Euler-Lagrange equations for $\hat{L}$ if and only if γ satisfies the Euler-Lagrange equations for L. In other words, the equations of motion that you get from $\hat{L}$ in spacetime are consistent with the equations of motion that you get from L in space. Here is an outline of the proof of this assertion:

- First we compute, using (3.30) and the chain rule that

$$\begin{aligned}\frac{\partial \hat{L}}{\partial x_i} &= \lambda \frac{\partial}{\partial x_i}\left[L\left(\mathbf{x}, \frac{\mathbf{u}}{\lambda}\right)\right] \\ &= \lambda \frac{\partial L}{\partial x_i}\end{aligned}$$

  and that

$$\begin{aligned}\frac{\partial \hat{L}}{\partial u_i} &= \lambda \frac{\partial}{\partial u_i}\left[L\left(\mathbf{x}, \frac{\mathbf{u}}{\lambda}\right)\right] \\ &= \lambda \frac{\partial L}{\partial v_i}\left(\frac{1}{\lambda}\right) \\ &= \frac{\partial L}{\partial v_i}\end{aligned}$$

  with the understanding that the partial derivatives of L are evaluated at $(\mathbf{x}, \frac{\mathbf{u}}{\lambda})$.

- The first n Euler-Lagrange equations for $\hat{L}$ are

$$\frac{\partial \hat{L}}{\partial x_i} = \frac{\partial}{\partial \tau}\left[\frac{\partial \hat{L}}{\partial u_i}\right].$$

Substituting the above calculations for the partial derivatives, we get

$$\lambda \frac{\partial L}{\partial x_i} = \frac{\partial}{\partial \tau}\left[\frac{\partial L}{\partial v_i}\right].$$

- To check whether a path $\hat{\gamma}$ satisfies the above equations, we would substitute (3.29) into both sides and see if they are equal. But (3.29) is of a special form where $\lambda = 1$. So, for paths of this special form, the above equations are equivalent to

$$\frac{\partial L}{\partial x_i} = \frac{\partial}{\partial \tau}\left[\frac{\partial L}{\partial v_i}\right],$$

  which are exactly the Euler-Lagrange equations for L (with τ playing the role that t previously played).

- We're not done yet, since there's one more Euler-Lagrange equation for $\hat{L}$:

$$\frac{\partial \hat{L}}{\partial t} = \frac{\partial}{\partial \tau}\left[\frac{\partial \hat{L}}{\partial \lambda}\right].$$

  In the exercises, you'll show that, for paths of the form (3.29), this equation is equivalent to the Beltrami identity, which we know is satisfied by paths that satisfy the Euler-Lagrange equations for L.

Applying Noether's Theorem to spacetime paths

Since Noether's Theorem is a purely mathematical result, it applies equally well in n-dimensional space and $(n+1)$-dimensional spacetime. By working in spacetime, we are able to consider the full set of Galilean transformations in our search for symmetries. To be more precise, if ϕ_s is a one-parameter group of transformations of spacetime that form a symmetry[34] for $\hat{L}$, then Noether's Theorem applies and gives us a conserved quantity Q, which would be a function on velocity spacetime. If γ is a path that satisfies the Euler-Lagrange equations for L, then $\hat{\gamma}$ satisfies the Euler-Lagrange equations for $\hat{L}$, and therefore Q is indeed conserved along $\hat{\gamma}$. In this manner, we can use spacetime symmetries to obtain conserved quantities.

[34] Keep in mind that "symmetry" could mean "f-symmetry".

Here are the important examples that we gain from this perspective:

- *Time-translation*: The formula for the one-parameter group of time-translations is

$$\phi_s(\mathbf{x}, t) = (\mathbf{x}, t + s).$$

  From the formula (3.30), we can see that $\hat{L}$ doesn't depend on t at all, so regardless of what L is, time translation is a symmetry.

The vector field that generates time-translations is $\mathbf{G} = (0, \ldots, 0, 1)$, so the associated conserved quantity is $\frac{\partial \hat{L}}{\partial \lambda}$. In the exercises, you'll show that

$$\frac{\partial \hat{L}}{\partial \lambda} = L - \frac{1}{\lambda} \sum_{i=1}^{n} \frac{\partial L}{\partial v_i} u_i,$$

which, when applied to paths of the form $\hat{\gamma}$, becomes[35]

$$\frac{\partial \hat{L}}{\partial \lambda} = L - \sum_{i=1}^{n} \frac{\partial L}{\partial v_i} v_i,$$

[35] Recall that, for such paths, $\lambda = 1$ and $u_i = v_i$.

This is exactly the minus of the energy B, as defined in (3.24). So the conserved quantity associated to time-translation is energy.

Of course, we already knew that energy was conserved from the Beltrami identity, but it's nice to see that there is a symmetry from which conservation of energy can be deduced.

- *Galilean boosts*: The one-parameter groups of Galilean boosts are of the form $\phi_s(\mathbf{x}, t) = (\mathbf{x} - \mathbf{a}st, t)$, where $\mathbf{a}$ can be any vector indicating the direction (and magnitude) of the boost. The vector field that generates these boosts is $\mathbf{G} = (-\mathbf{a}t, 0)$.

  To keep things as simple as possible, let's consider the case of a free particle[36]. In this case, the Lagrangian is $L = \frac{m\|\mathbf{v}\|^2}{2}$. Then the associated spacetime Lagrangian is

  $$\hat{L} = \lambda \frac{m\|\frac{1}{\lambda}\mathbf{u}\|^2}{2} = \frac{m\|\mathbf{u}\|^2}{2\lambda}.$$

  [36] A free particle is a particle that is not being acted on by any force. In other words, this is the case where the potential energy is 0.

  Since ϕ_s is a linear transformation, it will act on the velocities in the same way as the spacetime variables, so the associated transformation of velocity spacetime is

  $$\tilde{\phi}_s(\mathbf{x}, t, \mathbf{u}, \lambda) = (\mathbf{x} - \mathbf{a}s\lambda, t, \mathbf{u} - \mathbf{a}s\lambda, \lambda).$$

  Putting the above two equations together, we get

  $$\begin{aligned} \hat{L}(\tilde{\phi}_s(\mathbf{x}, t, \mathbf{u}, \lambda)) &= \frac{m\|\mathbf{u} - \mathbf{a}s\lambda\|^2}{2\lambda} \\ &= \frac{m\|\mathbf{u}\|^2}{2\lambda} - ms(\mathbf{u} \cdot \mathbf{a}) + \frac{ms^2\lambda\|\mathbf{a}\|^2}{2}. \end{aligned}$$

  This is *not* equal to $\hat{L}(\mathbf{x}, t, \mathbf{u}, \lambda)$, so you might not think that ϕ_s is a symmetry. But it is an f-symmetry. This is because the difference $\hat{L}(\tilde{\psi}_s(\mathbf{x}, t, \mathbf{u}, \lambda)) - \hat{L}(\mathbf{x}, t, \mathbf{u}, \lambda)$ is

  $$-ms(\mathbf{u} \cdot \mathbf{a}) + \frac{ms^2\lambda\|\mathbf{a}\|^2}{2},$$

  which, when we evaluate on a path, is[37]

  [37] When you evaluate on a path, $\mathbf{u} = \frac{d\mathbf{x}}{d\tau}$ and $\lambda = \frac{dt}{d\tau}$.

$$\frac{d}{d\tau}\left[-ms(\mathbf{x}\cdot\mathbf{a})+\frac{ms^2t\|\mathbf{a}\|^2}{2}\right].$$

So ψ_s is an f-symmetry with

$$f_s(\mathbf{x},t)=-ms(\mathbf{x}\cdot\mathbf{a})+\frac{ms^2t\|\mathbf{a}\|^2}{2}.$$

Using $\mathbf{G}=(-\mathbf{a}t,0)$, the corresponding conserved quantity (see (3.27)) is

$$\sum_{i=1}^{n}\left(-a_it\frac{\partial\hat{L}}{\partial u_i}\right)+m\mathbf{x}\cdot\mathbf{a}.$$

Using the generalized momenta $p_i=\frac{\partial\hat{L}}{\partial u_i}$, we can rewrite the conserved quantity as

$$\begin{aligned}&-t\sum_{i=1}^{n}a_ip_i+m\mathbf{x}\cdot\mathbf{a}\\&=-t\mathbf{a}\cdot\mathbf{p}+m\mathbf{x}\cdot\mathbf{a}\\&=(m\mathbf{x}-t\mathbf{p})\cdot\mathbf{a}.\end{aligned}$$

Since the vector $\mathbf{a}$ is arbitrary, we can conclude that the vector quantity $m\mathbf{x}-t\mathbf{p}$ is conserved. The quantity $m\mathbf{x}$ is known as the *center of mass*. So the conservation of $m\mathbf{x}-t\mathbf{p}$ tells us that the center of mass $m\mathbf{x}$ is moving at a constant speed, equal to the momentum.

Exercises

1. Sometimes math is, as Wigner put it, "unreasonably effective," in the sense that seemingly-weird mathematical phenomena can sometimes turn out to describe things that are physically real. Spacetime paths that go backward in time might sound crazy, but Feynman interpreted such paths as antiparticles.

   Imagine a path in $(1+1)$-dimensional spacetime of the form $\alpha(\tau)=(\cos\tau,\sin\tau)$. Draw a picture of this path and use Feynman's interpretation to tell the story about what the path describes.

2. Let $L(\mathbf{x},\mathbf{v})$ be a Lagrangian on velocity space, and let $\hat{L}$ be the Lagrangian on velocity spacetime defined in (3.30).

   (a) Show that $\frac{\partial\hat{L}}{\partial\lambda}=L-\frac{1}{\lambda}\sum_{i=1}^{n}\frac{\partial L}{\partial v_i}u_i$, with the understanding that the partial derivatives of L are evaluated at $(\mathbf{x},\frac{\mathbf{u}}{\lambda})$.

   (b) What is $\frac{\partial\hat{L}}{\partial t}$?

   (c) Let $\gamma(t)$ be a path. Show that $\hat{\gamma}$ satisfies the Euler-Lagrange equation

   $$\frac{\partial\hat{L}}{\partial t}=\frac{\partial}{\partial\tau}\left[\frac{\partial\hat{L}}{\partial\lambda}\right],$$

   if and only if γ satisfies the Beltrami identity.

3. For a Lagrangian on velocity space $L(\mathbf{x}, \mathbf{v})$, we define the generalized momenta as $p_i = \frac{\partial L}{\partial v_i}$. For the corresponding spacetime Lagrangian $\hat{L}$, the generalized momenta are defined as $\frac{\partial \hat{L}}{\partial u_i}$ and $\frac{\partial \hat{L}}{\partial \lambda}$. Compare $\frac{\partial \hat{L}}{\partial u_i}$ with $\frac{\partial L}{\partial v_i}$. Are they equal? If not, in what sense are they consistent?

4. Consider the case of near-earth gravity in one dimension x representing height, so that the Lagrangian is $L = \frac{mv^2}{2} - mgx$.

   (a) Write out the associated spacetime Lagrangian.

   (b) Show that this system has a Galilean boost symmetry, and find the corresponding conserved quantity.

3.6 *Relativistic actions*

In this section, we will see some examples of actions which are symmetric under the Poincaré group and are therefore compatible with special relativity. I'll emphasize that the Lagrangians that appear here don't arise from spatial Lagrangians, like the ones in the previous section did. This is because in special relativity, space and time are inextricably intertwined.

The relativistic free particle

Consider a spacetime path $\alpha(\tau) = (\alpha_x(\tau), \alpha_t(\tau))$ in 2-dimensional spacetime[38]. The *proper time* of α is defined as

$$\begin{aligned} \pi(\alpha) &:= \int_{\tau_0}^{\tau_1} \sqrt{\alpha_t'(\tau)^2 - \frac{1}{c^2}\alpha_x'(\tau)^2}\, d\tau \\ &= \int_{\tau_0}^{\tau_1} \sqrt{-\frac{1}{c^2}\|\alpha'(\tau)\|_{\mathcal{M}}^2}\, d\tau \\ &= \frac{1}{c}\int_{\tau_0}^{\tau_1} \sqrt{-\|\alpha'(\tau)\|_{\mathcal{M}}^2}\, d\tau. \end{aligned} \tag{3.32}$$

[38] For simplicity, we will work in one spatial dimension, although where possible I'll use coordinate-free formulas that work in any number of dimensions.

Recall that, in special relativity, there is a time-dilation phenomenon, where different observers might measure the passage of time differently. Proper time is the change in time, as calculated by the object itself[39], as it travels along the worldline α.

[39] The exercises will guide you through a derivation of the formula.

Note that proper time is only sensible for paths that never exceed the speed of light, so that $\|\alpha'(\tau)\|_{\mathcal{M}}^2 < 0$. Since special relativity prohibits faster-than-light travel anyway, it's totally fine for us to only allow paths of this type.

The relativistic action for a free particle is defined by

$$\mathbb{S}(\alpha) = -mc\int_{\tau_0}^{\tau_1} \sqrt{-\|\alpha'(\tau)\|_{\mathcal{M}}^2}\, d\tau = -mc^2\pi(\alpha), \tag{3.33}$$

so the corresponding Lagrangian, in terms of the velocity spacetime variables (x, t, u_x, λ), is

$$\mathbb{L} = -mc\sqrt{c^2\lambda^2 - u_x^2}. \tag{3.34}$$

Why is this formula reasonable? Here are some justifications for it:

1. Recall that action has units of energy $\times$ time. Proper time $\pi(\alpha)$ has units of time, and mc^2 has units of energy. The famous equation $E = mc^2$ actually says that mc^2 is the (kinetic) energy contained in an object with mass m *in a frame where the object is at rest*. Similarly, the proper time measures the passage of time in a frame where the object is at rest. So the action (3.33) is literally just[40] the product of the "rest energy" with the "rest time". So, not only does this action have the correct units, but it seems to be constructed in a sensible way.

[40] up to the sign, which is explained below

2. The action is symmetric under the Poincaré group, in the sense that, if ϕ is a Poincaré transformation, then $\mathbb{S}(\phi(\alpha)) = \mathbb{S}(\alpha)$. This is an immediate consequence of the fact that $\mathbb{S}$ can be completely expressed in terms of the Minkowski norm.

3. At low velocities, the Lagrangian (3.34) is consistent with the Lagrangian for the nonrelativistic free particle[41]. This compatibility is the reason for the minus sign in the formula.

[41] More on this in the exercises.

The generalized momentum for this system is

$$p_x = \frac{\partial \mathbb{L}}{\partial u_x} = \frac{mu_x}{\sqrt{\lambda^2 - \frac{u_x^2}{c^2}}}.$$

and the energy is

$$B = \frac{\partial \mathbb{L}}{\partial \lambda} = \frac{-mc^2\lambda}{\sqrt{\lambda^2 - \frac{u_x^2}{c^2}}}.$$

Since $\mathbb{L}$ doesn't depend on x, y, z, or t, the Euler-Lagrange equations are

$$0 = \frac{d}{d\tau}p_x, \qquad\qquad 0 = \frac{d}{d\tau}B, \tag{3.35}$$

from which you can deduce that the solutions are the paths $\alpha(\tau)$ which are straight lines in spacetime.

In the relativistic situation, the most interesting symmetries are the Lorentz boosts. A Lorentz boost is generated by the vector field $(ct, x/c)$, so the corresponding conserved quantity is

$$ctp_x + \frac{xB}{c} = \frac{mc(u_x t - \lambda x)}{\sqrt{\lambda^2 - \frac{u_x^2}{c^2}}}.$$

If we make the substitution $v_x = \frac{u_x}{\lambda}$, then we can rewrite the conserved quantity as

$$\frac{mc(v_x t - x)}{\sqrt{1 - \frac{v_x^2}{c^2}}}.$$

Compare this with the calculation for Galilean boosts in Section 3.5.

The charged particle

Recall that the development of special relativity was closely tied to Maxwell's theory of electromagnetism. The fact that Maxwell's theory is symmetric under the Poincaré transformations means that electromagnetic forces should be modeled by relativistic actions.

An electromagnetic field can be described by a *magnetic potential* $\mathbf{A}$, which is a three-dimensional vector field that depends on time[42] and an *electric potential* ϕ, which is a function on spacetime[43].

For a charged particle with charge q moving in an electromagnetic field, the Lagrangian is

$$L = -mc\sqrt{c^2\lambda^2 - u_x^2 - u_y^2 - u_z^2} + q\left(c^2\phi\lambda - \mathbf{A}\cdot\mathbf{u}\right).$$

If we combine the electric and magnetic potentials to create a 4-dimensional vector field $\hat{\mathbf{A}} = (A_1, A_2, A_2, \phi)$, then we can express the corresponding action as

$$S(\alpha) = \int_{\tau_0}^{\tau_1} \left(-mc\sqrt{-\|\alpha'(\tau)\|_{\mathcal{M}}^2} + q\langle\hat{\mathbf{A}}, \alpha'(\tau)\rangle_{\mathcal{M}}\right) d\tau.$$

You can see the relativistic nature of the theory in the fact that the action is completely expressed in terms of Minkowski inner products.

Note that, in the case where $q = 0$, this action reduces to that of the free relativistic particle; this makes sense, because if the particle has no charge, then its motion should be unaffected by the electromagnetic field. From this action, one can derive the Lorentz force law: $F = q(\mathbf{E} + \mathbf{v}\times\mathbf{B})$.

Electromagnetism

The Lagrangian formalism could be described very generally as a situation where you have a space of "all things that might possibly happen" and an action, which is a function on the space of "all things that might possibly happen". The critical points of the action are then the "things that actually happen". This framework is versatile enough to incorporate every known interaction in classical physics, including general relativity, electromagnetism, and even speculative ideas such as string theory.

[42] In other words, $\mathbf{A} = (A_1, A_2, A_3)$, where the functions A_i depend on x, y, z, and t. Note that we are now using three spatial dimensions, since Maxwell's equations use vector calculus operations (such as curl) that only make sense in three dimensions.

[43] For those of you who are familiar with electromagnetism, the magnetic and electric fields are related to the potentials by the formulas

$$\mathbf{B} = \operatorname{curl}\mathbf{A},$$
$$\mathbf{E} = -\nabla\phi - \frac{1}{c}\frac{\partial\mathbf{A}}{\partial t}.$$

This relationship is discussed a bit more below.

To get an idea of how this works, we'll start with a quick introduction to Maxwell's equations before describing the Lagrangian formulation of electromagnetism. This is only intended to give you an idea of what lies beyond, so you should expect it to raise more questions than it answers.

In a certain nice choice of units, Maxwell's equations are

$$\operatorname{div} \mathbf{B} = 0, \tag{3.36}$$

$$\operatorname{curl} \mathbf{E} + \frac{1}{c}\frac{\partial \mathbf{B}}{\partial t} = 0, \tag{3.37}$$

$$\operatorname{div} \mathbf{E} = \rho, \tag{3.38}$$

$$\operatorname{curl} \mathbf{B} - \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t} = \frac{1}{c}\mathbf{J}. \tag{3.39}$$

Here, ρ is a function of x, y, and z, representing the charge density, $\mathbf{J}$ is a vector field in $\mathbb{R}^3$ representing the electric current, and $\mathbf{E}$ (the electric field) and $\mathbf{B}$ (the magnetic field) are both vector fields in $\mathbb{R}^3$ which also depend on time. We think of ρ and $\mathbf{J}$ as being fixed parameters of the system, and we think of $\mathbf{E}$ and $\mathbf{B}$ as being the "variables" that we are trying to solve for.

There is a theorem from multivariable calculus that says that, if a vector field $\mathbf{B}$ is defined everywhere, then $\operatorname{div} \mathbf{B} = 0$ if and only if $\mathbf{B} = \operatorname{curl} \mathbf{A}$ for some vector field $\mathbf{A}$ (the magnetic potential). If we put this into equation (3.37), then we get

$$\operatorname{curl} \mathbf{E} + \frac{1}{c}\operatorname{curl}\left(\frac{\partial \mathbf{A}}{\partial t}\right) = 0,$$

$$\operatorname{curl}\left(\mathbf{E} + \frac{1}{c}\frac{\partial \mathbf{A}}{\partial t}\right) = 0,$$

and therefore

$$\mathbf{E} + \frac{1}{c}\frac{\partial \mathbf{A}}{\partial t} = -\nabla\phi$$

for some function ϕ (the electric potential).

The idea is that we can think of $\mathbf{A}$ and ϕ as being fundamental objects from which the electric and magnetic fields can be derived by taking

$$\mathbf{B} = \operatorname{curl} \mathbf{A},$$

$$\mathbf{E} = -\nabla\phi - \frac{1}{c}\frac{\partial \mathbf{A}}{\partial t}.$$

Then equations (3.36) and (3.37) hold automatically, and only equations (3.38) and (3.39) remain.

The Lagrangian for the theory is

$$\begin{aligned} L(\mathbf{A}, \phi) &= \frac{1}{2}\left(\|\mathbf{E}\|^2 - \|\mathbf{B}\|^2\right) - \rho\phi + \frac{1}{c}\mathbf{J}\cdot\mathbf{A} \\ &= \frac{1}{2}\left(\left\|\nabla\phi + \frac{1}{c}\frac{\partial \mathbf{A}}{\partial t}\right\|^2 - \|\operatorname{curl}\mathbf{A}\|^2\right) - \rho\phi + \frac{1}{c}\mathbf{J}\cdot\mathbf{A}, \end{aligned}$$

and since the fields are everywhere in spacetime (they're not just a path), the action is computed by integrating over all of spacetime:

$$S(\mathbf{A}, \phi) = \int_{\mathbb{R}^4} L(\mathbf{A}, \phi) dx\, dy\, dz\, dt.$$

The Euler-Lagrange equations were derived for paths, so we can't use them here. But we can still use the calculus of variations to find critical points of the action. The result ends up being that the critical points are the fields that satisfy (3.38) and (3.39); in other words, Maxwell's equations can be derived from the action principle.

When you look at Maxwell's equations, it's not at all obvious that they are symmetric under Poincaré transformations[44]. But it's possible to rewrite the Lagrangian in a way that only uses Minkowski inner products, making the symmetry an immediate consequence.

[44] Indeed, it took a couple of decades for people to work this out.

Exercises

1. Consider a vector (x_0, t_0) in spacetime (with just one spatial dimension, for simplicity), such that $x_0/t_0 < c$. Show that the Lorentz boost (1.16) with $v = x_0/t_0$ maps (x_0, t_0) to $(0, \sqrt{t_0^2 - x_0^2/c^2})$. (This is the explanation for the formula for proper time; in a frame where the vector has no spatial component, its length in the time direction is $\sqrt{t_0^2 - x_0^2/c^2}$, so we could call this quantity the "proper time" of the vector. The proper time of a path is then obtained by integrating the proper times of its velocity vectors.)

2. (a) Make the substitution $v_x = \frac{u_x}{\lambda}$ into (3.34), and show that, when v_x/c is small,

$$\mathbb{L} \approx \lambda \left(\frac{mv^2}{2} - mc^2 \right).$$

   [Hint: Use the first-order Taylor approximation $\sqrt{1-x} \approx 1 - \frac{x}{2}$.] This is the spacetime Lagrangian that corresponds to the spatial Lagrangian $L = \frac{mv^2}{2} - mc^2$, which is the nonrelativistic Lagrangian for an object with constant potential energy mc^2.

   (b) The Lagrangian for the nonrelativistic free particle is $L = \frac{mv^2}{2}$, but the result from (a) has the extra term $-mc^2$. Explain why this extra term doesn't matter[45].

[45] At least it doesn't matter in the nonrelativistic world. In relativity, the fact that an object at rest has potential energy mc^2 is very important!

A Some linear algebra and multivariable calculus

This appendix contains a very brief review of facts from linear algebra and multivariable calculus that are used in this book. It is written under the assumption that you already have some familiarity with the material[1].

For the sake of simplicity, most of the formulas here will be written in 2-dimensional space, which is sufficient to illustrate the results without the distraction of indices. However, all of the results are, with appropriate modifications, true in other dimensions, and it's a good exercise to think about what those modifications might be. Where possible, we will give dimension-neutral formulas using vector and matrix notation.

[1] For example, it is assumed that you know how to multiply matrices and compute partial derivatives.

A.1 Linear maps

A *linear map* $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ is a map of the form $\phi(x, y) = (ax + by, cx + dy)$, where a, b, c, d are constants. It is convenient to use vector notation, where we can interpret ϕ as multiplication by a matrix:

$$\phi\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} ax + by \\ cx + dy \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

More generally, a linear map $\phi : \mathbb{R}^n \to \mathbb{R}^m$ is a map of the form $\phi(\mathbf{x}) = A\mathbf{x}$, where A is an $m \times n$ matrix[2]. Thus there is a correspondence between matrices and linear maps; if you have one then you have the other. This is an important point[3].

[2] An $m \times n$ matrix has m rows and n columns.

[3] In fact, it can sometimes be convenient to blur the distinction and say that a linear map is the same thing as a matrix.

Suppose we have linear maps $\phi_1 : \mathbb{R}^m \to \mathbb{R}^k$ and $\phi_2 : \mathbb{R}^n \to \mathbb{R}^m$. Then there is a $k \times m$ matrix A_1 and an $m \times n$ matrix A_2 such that $\phi_1(\mathbf{x}) = A_1\mathbf{x}$ and $\phi_2(\mathbf{x}) = A_2\mathbf{x}$. Since matrix multiplication is associative, we see that

$$\begin{aligned} (\phi_1 \circ \phi_2)(\mathbf{x}) &= \phi_1(\phi_2(\mathbf{x})) \\ &= A_1(A_2\mathbf{x}) \\ &= (A_1A_2)\mathbf{x}. \end{aligned}$$

This means that, under the correspondence between matrices and linear maps, matrix multiplication corresponds to composition of maps[4].

[4] If you've ever wondered why matrix multiplication is what it is, this is the answer.

A.2 Differentiable maps

Consider a function of one variable $f(x)$. We can think of f as a map from $\mathbb{R}$ to $\mathbb{R}$. If f is a linear map, then it must be of the form $f(x) = ax$ for some constant a. Of course, there are many more functions, such as $f(x) = x^2$ or $f(x) = e^x$, that are interesting but nonlinear. One of the main themes of calculus is that, if a function is nice enough, then it can be approximated (up to a shift) by a linear function in the vicinity of any particular point. Specifically, if f is differentiable at x_0, then, when x is near x_0, we have the tangent line approximation

$$f(x) - f(x_0) \approx f'(x_0)(x - x_0).$$

Note that the right side is a linear function of $x - x_0$.

A map $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ can be described by a pair of coordinate functions $u(x, y)$ and $v(x, y)$:

$$\phi(x, y) = (u(x, y), v(x, y)).$$

We can think of ϕ as being a map from the xy-plane to the uv-plane; given an input point (x, y), we can calculate an output point (u, v). Here, there is a notion of differentiability that is similar to the 1-variable situation[5]. Specifically, ϕ is differentiable at $\mathbf{x}_0 = (x_0, y_0)$ if there exists a matrix D_ϕ such that

$$\phi(\mathbf{x}) - \phi(\mathbf{x}_0) \approx D_\phi(\mathbf{x} - \mathbf{x}_0)$$

when $\mathbf{x}$ is near $\mathbf{x}_0$.

[5] I'm going to avoid giving a precise definition of differentiability, leaving it for a course in analysis or advanced calculus. All of the maps that we will encounter in this book will be differentiable everywhere they are defined.

If ϕ is differentiable, then the matrix D_ϕ is made up out of the partial derivatives of the coordinate functions:

$$D_\phi = \begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{bmatrix}.$$

For this reason, D_ϕ is called the *total derivative* of ϕ.

A.3 The dot product

A 2-dimensional vector $\mathbf{a}$ should usually be thought of as a matrix with one column[6]

$$\mathbf{a} = \begin{bmatrix} a \\ b \end{bmatrix}.$$

[6] Sometimes we will instead write $\mathbf{a} = (a, b)$, but the only reason to do so is to save vertical space on the page.

Suppose now we have two vectors

$$\mathbf{a}_1 = \begin{bmatrix} a_1 \\ b_1 \end{bmatrix}, \qquad \mathbf{a}_2 = \begin{bmatrix} a_2 \\ b_2 \end{bmatrix}.$$

The dot product of $\mathbf{a}$ and $\mathbf{b}$ is defined as

$$\mathbf{a} \cdot \mathbf{b} = a_1 a_2 + b_1 b_2.$$

It's useful to think of the dot product as being the key that unlocks all of the geometric information about vectors. Specifically, the length of $\mathbf{a}$ is

$$\|\mathbf{a}\| = \sqrt{a^2 + b^2} = \sqrt{\mathbf{a} \cdot \mathbf{a}},$$

and, assuming $\mathbf{a}_1$ and $\mathbf{a}_2$ are nonzero, the angle θ between them can be found by the formula[7]

$$\cos\theta = \frac{\mathbf{a}_1 \cdot \mathbf{a}_2}{\|\mathbf{a}_1\|\|\mathbf{a}_2\|}.$$

[7] These equations also hold in 3 dimensions. In dimensions greater than 3, we don't have any pre-existing notion of length and angle, so we can take these equations to be the *definitions* of length and angle.

In particular, $\mathbf{a}_1$ and $\mathbf{a}_2$ are *orthogonal* to each other when $\mathbf{a}_1 \cdot \mathbf{a}_2 = 0$.

Recall that the *transpose* operation on matrices exchanges the rows and columns. For example, if

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix},$$

then the transpose of A is

$$A^\intercal = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}.$$

In particular, the transpose of a vector (which should be thought of as a matrix with one column) is a matrix with one row.

If $\mathbf{a}_1$ and $\mathbf{a}_2$ are vectors as above, then $\mathbf{a}_1^\intercal \mathbf{a}_2$ is a 1×1 matrix, whose only entry is the dot product:

$$\mathbf{a}_1^\intercal \mathbf{a}_2 = \begin{bmatrix} a_1 & b_1 \end{bmatrix} \begin{bmatrix} a_2 \\ b_2 \end{bmatrix} = a_1 a_2 + b_1 b_2 = \mathbf{a}_1 \cdot \mathbf{a}_2.$$

A.4 Orthonormal bases

A pair of vectors $\{\mathbf{u}_1, \mathbf{u}_2\}$ form an *orthonormal basis* of $\mathbb{R}^2$ if $\|\mathbf{u}_1\| = 1$, $\|\mathbf{u}_2\| = 1$, and $\mathbf{u}_1 \cdot \mathbf{u}_2 = 0$. In other words, $\mathbf{u}$ and $\mathbf{v}$ are unit vectors that are orthogonal to each other[8].

[8] More generally, an orthonormal basis of $\mathbb{R}^n$ is a set of n unit vectors that are all orthogonal to each other.

A simple yet important example is the *standard basis* $\{\mathbf{e}_1, \mathbf{e}_2\}$, where

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \qquad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Suppose I have a vector $\mathbf{a} = (a, b)$. By taking dot products of $\mathbf{a}$ with the standard basis vectors, we can extract the coordinates of $\mathbf{a}$:

$$\mathbf{a} \cdot \mathbf{e}_1 = a, \qquad \mathbf{a} \cdot \mathbf{e}_2 = b.$$

On the other hand, we can write $\mathbf{a}$ as a linear combination of the standard basis vectors:

$$\mathbf{a} = a\mathbf{e}_1 + b\mathbf{e}_2.$$

Putting the above facts together, we obtain the following expansion[9], which is valid for any vector $\mathbf{a}$:

$$\mathbf{a} = (\mathbf{a} \cdot \mathbf{e}_1)\mathbf{e}_1 + (\mathbf{a} \cdot \mathbf{e}_2)\mathbf{e}_2.$$

[9] This expansion explicitly describes $\mathbf{a}$ as a linear combination of the standard basis vectors, where the coefficients are computed using only dot products.

Amazingly, a similar result holds for *any* orthonormal basis. Specifically, if $\{\mathbf{u}_1, \mathbf{u}_2\}$ is an orthonormal basis, then[10]

$$\mathbf{a} = (\mathbf{a} \cdot \mathbf{u}_1)\mathbf{u}_1 + (\mathbf{a} \cdot \mathbf{u}_2)\mathbf{u}_2 \tag{A.1}$$

[10] You will derive this formula in the exercises.

for any vector $\mathbf{a}$.

Exercises

1. You are walking around the Math department and overhear someone say that "a linear function is just a row vector". Explain this statement with more details and precision. [Note: The word *function* is generally used for maps to a set of numbers, such as $\mathbb{R}$.]

2. Suppose that $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ is a map. Show that ϕ is linear if and only if it has the property that $\phi(\mathbf{x} + \mathbf{y}) = \phi(\mathbf{x}) + \phi(\mathbf{y})$ for all vectors $\mathbf{x}, \mathbf{y}$.

3. Suppose that $\phi : \mathbb{R}^n \to \mathbb{R}^m$ is a linear map, and let A be the matrix such that $\phi(\mathbf{x}) = A\mathbf{x}$. Verify that $D_\phi = A$.

4. Suppose that A is an $m \times n$ matrix, $\mathbf{a}$ is an m-dimensional vector, and that $\mathbf{b}$ is an n-dimensional vector. Show that

   $$\mathbf{a} \cdot (A\mathbf{b}) = (A^\intercal \mathbf{a}) \cdot \mathbf{b}.$$

   [Hint: Recall that the transpose operation exchanges the order of multiplication, i.e. $(AB)^\intercal = B^\intercal A^\intercal$.]

5. Show that, for all vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$,

   $$\mathbf{a} \cdot \mathbf{b} = \frac{1}{2}\left(\|\mathbf{a} + \mathbf{b}\|^2 - \|\mathbf{a}\|^2 - \|\mathbf{b}\|^2\right).$$

   Thus, the dot product can be completely expressed in terms of lengths of vectors.

6. Let U be an $n \times n$ matrix. As a definition, we say that U is *orthogonal* when $U^\intercal = U^{-1}$.

   (a) Show that U is orthogonal if and only if the columns of U form an orthonormal basis.

   (b) Show that U is orthogonal if and only if it preserves the dot product, in the sense that $(U\mathbf{a}) \cdot (U\mathbf{b}) = \mathbf{a} \cdot \mathbf{b}$.

   (c) Suppose that U is orthogonal. Then $UU^\intercal \mathbf{a} = \mathbf{a}$. Use this to derive the expansion (A.1).