

**MECHANICS FOR MATHEMATICIANS**  
**August 29, 2020**

JARED WUNSCH

CONTENTS

1. Preface	2
2. Introduction: Newton's Law(s)	3
3. ODE	5
3.1. Solving an ODE: gravity	6
3.2. The harmonic oscillator	7
3.3. General second-order equations	8
3.4. A note on solving ODEs	11
3.5. Back to systems	12
3.6. Newton's second law	14
4. One-dimensional force laws	15
4.1. Motion in a potential; energy conservation	15
4.2. Features of one-dimensional potential motion	18
5. The phase plane	19
5.1. Behavior of the motion near critical points	22
5.2. Closed phase curves and periodic motion	24
6. Motion in a potential in higher dimensions	26
6.1. Conservative Potentials	34
6.2. Consequences: bound and unbound motion	36
6.3. Example: Lissajous curves	37
7. The principle of least action	41
7.1. Introduction: why potentials?	41
7.2. The principle of least action	42
7.3. The calculus of variations	43
7.4. Other applications of the calculus of variations	48
7.5. Conservation Laws I	51
7.6. Why Lagrangian mechanics?	53
7.7. Change of coordinates	55
7.8. Motion with constraints	57
7.9. Conservation Laws II	63
8. Two particle interactions	70
9. Radial potentials	72
10. The Kepler problem	78
11. Hamiltonian mechanics	82
11.1. Hamilton's equations of motion	82

11.2. Vector fields and flows	87
11.3. Poincaré recurrence	94
11.4. Poisson brackets	94
11.5. Symplectic structure	96
Appendix A. Separable first order differential equations	98
Appendix B. Constant coefficient second order equations	98

## 1. PREFACE

These lecture notes were developed for Math 327, *Mechanics for Mathematicians*, a one quarter course at Northwestern University. I developed this course to try to address the problem that many of our most talented students of mathematics were finishing their undergraduate careers having learned no physics whatsoever. Physics as a source of both the motivating problems in mathematics and the intuition for how those problems might be solved has always seemed to me an essential aspect of undergraduate preparation. To study calculus and not understand how Newton developed *and then employed* it to show that Kepler's laws of motion can be derived from the inverse square law of gravitation seems tragic: why develop all this fancy machinery if *not* to understand how it applies to tell one of the greatest adventure stories in intellectual history?

Math 327 was intended to make a beeline for the mathematically beautiful and significant aspect of mechanics, with an emphasis on Lagrangian and Hamiltonian formalisms, while short-cutting all the complications of practical physics. I might in fact prefer that our students take several years' worth of courses in honest-to-god physics and learn all the lovely things I have been forced by limitations of time to omit, but this is supposed to be enough physics to get them started, and to enable them to learn more on their own. In particular, I have been very influenced in writing these notes by two masterpieces of the Russian school: Landau and Lifshitz's *Mechanics*, and Arnol'd's *Mathematical Methods of Classical Mechanics*. These notes are, in effect, training for reading those (rather sophisticated) books as a next step. I fantasize about a continuation into quantum mechanics, but have no immediate plans to write one.

Prerequisites are a firm grasp of linear algebra and multivariable calculus. At various points I have hinted at a rigorous treatment only possible with the application of more advanced ideas from a real analysis course, but those subtleties are optional. There are some sections early on about conservative force laws that should be skipped if students have a really solid recollection of these parts of multivariable calculus. My experience has been that in the absence of connections to physics, students have in fact found these beautiful results to be quite forgettable at a first acquaintance, and it has been useful to review them. Likewise, the material on ODE should certainly be omitted if students have studied that subject; I have generally found, though, that

our strongest pure math students regard the traditionally cookbook courses that we offer in ODE as being beneath their dignity, so this course also serves to explain to them why they should care about it.

The notes work well as a one quarter course if one is brisk in going through the material on ODE and vector calculus in the early going. I hope at a later date to add material on rigid body motion and on classical field theories, but have no means of fitting these into my one quarter course as it stands—they would be useful to those with a semester to spend. I have included lots of problems, since being able to compute is, in the end, the main take-away, even for a mathematicians’ course in physics.

These lecture notes have benefited considerably from the TA’s for Math 327, Owen Gwilliam (2010), Eric Potash (2012), Spencer Liang (2014), Kyle Casey (2018), Mengxuan Yang (2020). I am grateful to them all.

## 2. INTRODUCTION: NEWTON’S LAW(S)

In Newtonian physics, a particle of mass  $m$  moves through three-dimensional space according to the law

$$(1) \quad \mathbf{F} = m\mathbf{a}$$

where  $\mathbf{F}$  is the force acting on the object and

$$\mathbf{a} = \ddot{\mathbf{x}}(t)$$

is the second derivative<sup>1</sup> of the position  $\mathbf{x}(t)$ . Here  $\mathbf{x}(t)$  tells us the location of the particle at time  $t$  and  $\mathbf{F}$  is allowed to depend both on  $\mathbf{x}$  and on  $t$  (and even on  $m$ ). The force, the left-hand side of this equation, is a little subtle to “define.” Indeed, maybe the easiest way to think of getting at what the force is is to do repeated physical experiments to measure  $\mathbf{a}$ , the acceleration of various particles launched on various trajectories; then we can try and deduce what  $\mathbf{F}$  must have been in order to satisfy Newton’s law. To take a particularly famous instance, Tycho Brahe’s observations of planetary motion, as interpreted after his death by Johannes Kepler, famously led Isaac Newton to deduce what the force law of gravity must be. In this example as in others, once we’ve understood the force law in a given physical setting, we can make predictions about further experiments, as we shall see.

Here are some examples of force laws, derived from experiments and observations over the centuries.

- (1) A particle in a uniform gravitational field. On earth’s surface, the force of gravity is very well approximated by  $\mathbf{F} = -mg\mathbf{e}_3$ , where  $\mathbf{e}_3$  is the unit vector in the  $x_3$  (“vertical”) direction, and  $g$  is a scalar constant, approximately  $9.8m/s^2$ . Thus, the  $m$ ’s cancel from Newton’s law above, and it reads

$$\ddot{\mathbf{x}}(t) = -g\mathbf{e}_3.$$

---

<sup>1</sup>We will consistently use the “dot” notation for time derivatives, i.e.  $\dot{f} = df/dt$ .

This single equation is of course really shorthand for the system of three equations

$$\ddot{x}_1 = 0, \quad \ddot{x}_2 = 0, \quad \ddot{x}_3 = -g.$$

As the first two of the equations in this system are rather trivial, we may wish to simply write  $x = x_3$  and consider the one-dimensional problem

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

as our simplest example of a mechanical system.

- (2) A mass on a spring. Here again, for simplicity, we'll take our mass to live at point in  $\mathbb{R}^1$ : maybe it's constrained to slide along a rod. The force law is known as "Hooke's law," and reads

$$F = -kx,$$

where the constant  $k$  depends on how stretchy our spring is (and so is often called the "spring constant"). Newton's law reads

$$m\ddot{x}(t) = -kx(t).$$

- (3) A charged particle, with charge  $q$ , in a magnetic field  $\mathbf{B}$ , experiences a force

$$\mathbf{F} = q\mathbf{v} \times \mathbf{B}.$$

Here  $\mathbf{v} = \dot{\mathbf{x}}$  is the velocity of the particle, and  $\mathbf{B}$  is a fixed vector field.

- (4) Newton's law of gravitation: the force on a celestial body (e.g., a planet) of mass  $m$  at position  $\mathbf{x}$  attracted another celestial body of mass  $M$  (e.g., the sun) at position  $\mathbf{y}$  is

$$-\frac{GMm}{|\mathbf{x} - \mathbf{y}|^2} \widehat{(\mathbf{x} - \mathbf{y})}.$$

Here  $\widehat{\mathbf{x}} = \mathbf{x}/|\mathbf{x}|$  denotes the unit vector in the direction of  $\mathbf{x}$ , and  $G = 6.67 \times 10^{-11} m^3 kg^{-1} s^{-2}$  is Newton's gravitational constant, a fundamental physical constant.<sup>2</sup> Newton's conjecture of this physical law, and his use of the nascent tools of calculus to show that it implies Kepler's laws of planetary motion, which Kepler had formulated based on Tycho's empirical observations, are one of the great triumphs of human thought.

In the case of the system of a sun and a planet, it turns out to be a reasonable assumption to place the sun at the origin of coordinates, and assume that its position is fixed,<sup>3</sup> Newton's law tells us that the

---

<sup>2</sup>Unlike  $g$ , which reflects the mass of the earth,  $G$  is really a basic constant built into the law of gravitation—it has nothing to do with the accident of what solar system we happen to live in.

<sup>3</sup>Indeed, we can always reduce the motion of a two-body system to this case, with no loss in generality or accuracy, as we'll see later.

position  $\mathbf{x}(t)$  of the planet satisfies

$$m\ddot{\mathbf{x}}(t) = -\frac{GMm}{|\mathbf{x}|^2}\hat{\mathbf{x}}.$$

In the interest of focusing on mathematics, rather than physics, we will drop most physical constants from our problems. (In many cases this can be accomplished by a change of units.) So we will boil down the equations listed to simpler versions for mathematical analysis. We'll sometimes keep some of these parameters around, though, as it can be instructive to keep track of the dependence of the solutions on the essential parameters of the problem, and to understand limits in which these parameters may become small or large.

To recap, here are our examples above, stripped down to their mathematical essences:

(1) Particle in a uniform gravitational field. Setting  $g = 1$  gives

$$(2) \quad \ddot{\mathbf{x}}(t) = -\mathbf{e}_3,$$

or, in the one-dimensional version,

$$(3) \quad \ddot{x}(t) = -1.$$

(2) A mass on a spring. We simply set  $k/m = \omega^2$ ; then without any loss of generality, the equation is

$$(4) \quad \ddot{x}(t) = -\omega^2 x(t).$$

We'll keep the parameter  $\omega^2$  around, as it's somewhat instructive.

(3) Charged particle in magnetic field: we let  $q = m = 1$ , so

$$(5) \quad \ddot{\mathbf{x}} = \dot{\mathbf{x}} \times \mathbf{B}.$$

In the special case when  $\mathbf{B}$  is the constant vector field  $\mathbf{B} = \mathbf{e}_3$ , we get the equations

$$(6) \quad \ddot{\mathbf{x}} = \dot{\mathbf{x}} \times \mathbf{e}_3 = \begin{pmatrix} \dot{x}_2 \\ -\dot{x}_1 \\ 0 \end{pmatrix}$$

(4) Newton's law of gravitation: if  $GMm = 1$  and  $\mathbf{y} = 0$  is fixed,

$$(7) \quad \ddot{\mathbf{x}}(t) = -\frac{1}{|\mathbf{x}|^2}\hat{\mathbf{x}}.$$

### 3. ODE

All of these equations (2)–(7) have one important feature in common: they are *second-order ordinary differential equations* for the unknown function(s)  $\mathbf{x}$ . This means that they can be written in the form

$$(8) \quad \ddot{\mathbf{x}} = \text{a function of } \dot{\mathbf{x}}, \mathbf{x}, t.$$

In general, an ordinary differential equation is a set of equations for a function or functions of a single variable<sup>4</sup> involving derivatives of the function, the function itself, and the independent variable (which is  $t$  here). The *order* of the equation is the largest number of derivatives involved.

In our situation, since  $\mathbf{x}$  may have three components, you should in general think of our equation (8) as unpacking to give a *system* of three such equations

$$(9) \quad \begin{aligned} \ddot{x}_1 &= \text{a function of } \dot{x}_1, \dot{x}_2, \dot{x}_3, x_1, x_2, x_3, t \\ \ddot{x}_2 &= \text{a function of } \dot{x}_1, \dot{x}_2, \dot{x}_3, x_1, x_2, x_3, t \\ \ddot{x}_3 &= \text{a function of } \dot{x}_1, \dot{x}_2, \dot{x}_3, x_1, x_2, x_3, t. \end{aligned}$$

In the case where we've chosen to write our equations in just one dimension, such as (3), then we get not a system but a single equation

$$\ddot{x} = \text{a function of } \dot{x}, x, t.$$

Why do the equations we wrote down above all have this same form? Well, if the force  $\mathbf{F}$  depends on  $\dot{\mathbf{x}}$ ,  $\mathbf{x}$ , and  $t$ , then the form of Newton's second law (1), when we unpack it in coordinates, is exactly that of (9). In other words, Newton's law is telling us that mechanical systems are governed by second order ODE's. A general set of equations like (9) is pretty awful, but the examples we've seen above will turn out to be beautiful and special, as we'll soon see.

**3.1. Solving an ODE: gravity.** Let us now try *solving* some differential equations; we will begin by working in the 1-dimensional case for simplicity. Solving an ODE means writing down all possible functions  $x(t)$  that satisfy the equation. As our first example, consider 1-dimensional terrestrial gravity,

$$\ddot{x} = -1.$$

This is an ODE that we know how to solve from calculus class: just take an antiderivative, twice! We successively find, keeping careful track of constants of integration:

$$\dot{x}(t) = -t + A$$

and then

$$(10) \quad x(t) = -t^2/2 + At + B.$$

Note that this equation had *many* solutions: we get to choose any values of  $A, B$  and the resulting quadratic function solved the equation. The fact that we get to choose two constants in specifying a solution will turn out to be an essential feature of second order equations.

Having found all the solutions to our ODE now enables us to solve the main physical problem motivating our equation: the *initial value problem*.

---

<sup>4</sup>This is the "ordinary" part, as we only have "ordinary" rather than "partial" derivatives. In a *partial differential equation* we would have several independent variables instead of just  $t$ , and the equation would involve partial derivatives with respect to those variables.

Supposing we specify the initial state of our mass on a spring. How do we find where it is at later time?

First, we have to think a little about what it means to specify an initial state. Certainly we want to know the location of the mass, which is to say we should specify  $x(0) = x_0$ . Next, we need to know how fast it's moving to start with:  $\dot{x}(0) = v_0$ . Now we should try and find a solution that has these initial conditions. By (10) we must have

$$(11) \quad x_0 = -0^2/2 + A \cdot 0 + B,$$

hence we read off  $B = x_0$ . Differentiating (10) yields in general the equation

$$\dot{x}(t) = -t + A$$

hence plugging in  $t = 0$  yields

$$v_0 = -0 + A$$

i.e.,  $A = v_0$ . Thus we finally arrive at the solution

$$x(t) = -t^2/2 + v_0 t + x_0.$$

We have thus solved the “initial value problem”: we found a (indeed, *the*) solution with the initial data  $(x_0, v_0)$ .

**3.2. The harmonic oscillator.** Now let us consider the equation for a mass on a spring, in the special case where  $\omega = 1$  :

$$(12) \quad \ddot{x}(t) = -x(t).$$

We have an equation for an unknown function  $x(t)$ , and the equation specifies that its second derivative is minus it. You probably already know a function solving this equation:  $\sin t$  has this property. So, indeed, does  $\cos t$ . In fact, so does any *linear combination*

$$(13) \quad A \sin t + B \cos t$$

for  $A, B \in \mathbb{R}$  (or even in  $A, B \in \mathbb{C}$  if we want to be that general about our solutions—but you might find the idea of a complex number representing position a little alarming!).

In fact, it turns out that *every* function solving (12) can be written in the form (13).

*Exercise 3.1.* Prove directly that any solution to (12) is of the form (13) as follows:

- (1) Multiply the equation by  $\dot{x}$  on both sides and integrate.
- (2) Write the resulting equation in the form

$$(14) \quad \dot{x} = \sqrt{C^2 - x^2}.$$

- (3) Integrate the resulting first-order differential equation to find that

$$x(t) = C \sin(t + \delta)$$

for some constant  $\delta$ . If you've never studied first-order ODE's of the special form (14) (“separable equations”), see Appendix A.

- (4) Use the addition formula for sin to write the answer in the form (13) (with the constants  $A, B$  depending on  $\delta, C$ ).

Again, we want to use the general solution of the ODE to solve the initial value problem. Differentiating (13) gives

$$\dot{x}(t) = A \cos t - B \sin t,$$

so plugging in  $t = 0$ , we find that the two conditions we should specify are

$$A \cdot 0 + B \cdot 1 = x_0,$$

$$A \cdot 1 - B \cdot 0 = v_0;$$

in other words, we should take  $B = x_0$  and  $A = v_0$ , and our solution turns out to be

$$x(t) = v_0 \sin t + x_0 \cos t.$$

The differential equation we just solved is usually known as the “harmonic oscillator” and is ubiquitous in physics and mathematics. It is very special. But some of the features of the solutions to the equation turn out to hold true no matter what equation we consider.

*Exercise 3.2.* It is instructive to note that we can write the solutions to (12) in another way: for any  $P, Q \in \mathbb{C}$ , the function

$$Pe^{it} + Qe^{-it}$$

is a solution to the equation. Explain how this is consistent with our description of the solutions above. **HINT:** It’s helpful to think in the language of linear algebra: show that we are exhibiting two different bases for the same vector space.

Also show that if we have a complex-valued solution to (12), then the real and imaginary parts of the solution solve the equation separately.

**3.3. General second-order equations.** Supposing we’re given some more complicated second order equation to solve, such as the fairly unphysical equation

$$(15) \quad \ddot{x}(t) = (1+t)^2 e^{2x}.$$

What can we say about solutions to this equation? The central result in the theory of ODEs is that we can still solve the initial value problem for this equation, and indeed that the solution with given initial data is unique. In other words, specifying, say,

$$x(0) = 1, \quad \dot{x}(0) = 2$$

specifies a unique solution to (15). Here is the general statement of the theorem, sometimes called the Picard-Lindelöf theorem:<sup>5</sup>

---

<sup>5</sup>The hypothesis of differentiability on  $G$  in the statement of the theorem can in fact be somewhat weakened.

**Theorem 3.1.** *Consider the ODE*

$$(16) \quad \ddot{x} = G(t, x, \dot{x})$$

for  $x \in \mathbb{R}$ , where  $G(t, x, v)$  is a differentiable function of its arguments. For any  $x_0, v_0 \in \mathbb{R}$  there exists  $T > 0$  such that there is a function  $x(t)$  defined for  $t \in [-T, T]$  solving (16) and satisfying

$$x(0) = x_0, \quad \dot{x}(0) = v_0;$$

this solution is unique.

We will not prove this theorem, as the subject of these notes is physics, not ODEs, but let's try and see why this theorem ought to be true. We'll do this in two ways.

METHOD 1: Solve the equation as a power series

$$x(t) = a_0 + a_1 t + a_2 \frac{t^2}{2!} + \dots$$

How can we find the coefficients  $a_j$ ? Well, first we notice that our initial conditions give us the first two for free! We must have  $a_0 = x(0) = 1$ , and  $a_1 = \dot{x}(0) = 2$ . What can we now do about  $a_3$ , though? Well, this is where the ODE comes in! Our ODE (15) tells us that

$$\ddot{x}(0) = (1 + 0)^2 e^{2x(0)} = e^2.$$

This of course tells us that  $a_2 = e^2$ , so we've moved one more step along. How about  $a_3 = \ddot{x}(0)$ ? Well, we do have one sure-fire way to learn about the third derivative of  $x$ : just take a derivative of the equation (15). In doing this, we have to remember to use the chain rule: the right side involves  $x$ , which is itself a function of  $t$ . So we get the new equation

$$\ddot{x} = 2(1+t)e^{2x} + 2(1+t)^2 e^{2x} \dot{x}.$$

Setting  $t = 0$  now does tell us what we wanted:

$$a_3 = \ddot{x}(0) = 2e^{2x(0)} + 2e^{2x(0)} \dot{x}(0) = 2 \cdot e^2 + 2 \cdot e^2 \cdot 2.$$

We will not pursue this masochism, but it should be clear that we continue in this vein: at each step, we take another derivative of the differential equation; then setting  $t = 0$  allows us to solve for one more derivative of  $x$  at  $t = 0$ , hence for one more coefficient in its Taylor series. So the solution with these initial conditions seems to exist and to be unique.

What are the drawbacks of this argument? Well, just the usual caveats about Taylor series. The main one is that it is a somewhat intricate business to check that the Taylor series we get has a positive radius of convergence! It turns out that for an equation with like (15) where the right hand side isn't too awful, it does, but that's a job for another set of lecture notes.

*Exercise 3.3.* Use this method to find the Taylor series for the solution to the initial value problem

$$\ddot{x} = -x, \quad x(0) = 0, \quad \dot{x}(0) = 1.$$

Verify that this agrees with the exact answer that we computed above!

**METHOD 2:** A method that is perhaps more honest from the practical point of view is to try and give a different recipe for an approximate solution to our equation, one which would be fairly simple and precise to implement on a computer. This particular method is called *Euler's method*, but many more sophisticated variants exist as well.

What we do is to keep track of both  $x(t)$  and  $v(t) = \dot{x}(t)$  at each time. We fix a small time-step  $\Delta t$ , and we try to find approximate values of both of these quantities at times  $0, \Delta t, 2\Delta t, \dots$  based on their initial values, and simple linear approximation.

Say we're trying to solve the initial value problem

$$\ddot{x} = G(x, \dot{x}, t), \quad x(0) = x_0, \quad \dot{x}(0) = v_0.$$

It's handy to rewrite this equation in terms of  $x(t)$  and  $v(t) = \dot{x}(t)$ : we now get a system of equations

$$(17) \quad \begin{aligned} \dot{x} &= v, \\ \dot{v} &= G(x, v, t) \end{aligned}$$

where the first equation is simply the definition of  $v$  and the second is our ODE rewritten in these new variables. This process converts a single second order equations to a pair of first order ones: no more second derivatives appear!

Now let's use (17) to find our approximate solution. Using a linear approximation, we note that a solution would satisfy

$$\begin{aligned} x(\Delta t) &\cong x(0) + \dot{x}(0)\Delta t = x_0 + v_0\Delta t \\ v(\Delta t) &\cong v(0) + \dot{v}(0)\Delta t = v_0 + G(x_0, v_0, 0)\Delta t. \end{aligned}$$

In other words, we can compute approximations of  $x$  and  $v$  at time  $\Delta t$  by knowing their values at  $t = 0$ —we have of course used (17) to write  $\dot{v}(0) = \ddot{x}(0)$  in terms of  $x(0)$  and  $v(0)$ .

Now we can continue the process. We can compute  $x(2\Delta t)$  by a linear approximation starting at  $\Delta t$ :

$$\begin{aligned} x(2\Delta t) &\cong x(\Delta t) + v(\Delta t)\Delta t \\ v(2\Delta t) &\cong v(\Delta t) + \dot{v}(\Delta t)\Delta t = v(\Delta t) + G(x(\Delta t), v(\Delta t), \Delta t)\Delta t. \end{aligned}$$

We can keep this process running over as many time intervals as we want: if we already have computed an approximation for  $x(k\Delta t), v(k\Delta t)$  then we may approximate

$$\begin{aligned} x((k+1)\Delta t) &\cong x(k\Delta t) + v(k\Delta t)\Delta t \\ v((k+1)\Delta t) &\cong v(k\Delta t) + \dot{v}(k\Delta t)\Delta t = v(k\Delta t) + G(x(k\Delta t), v(k\Delta t), k\Delta t)\Delta t. \end{aligned}$$

*Exercise 3.4.* Consider the initial value problem

$$\ddot{x} = -x, \quad x(0) = 0, \quad \dot{x}(0) = 1.$$

Setting  $\Delta t = \pi/5$ , find an approximate value for  $x(\pi)$  by this procedure. What would the exact answer have been?

This procedure is, as you'll see by doing the exercise, messier to compute with than our nice Taylor series prescription above. On the other hand, it turns out to be a much better procedure to implement on a computer. Moreover, a fine way to prove the theorem about existence and uniqueness to initial value problems is to show that as we shrink the step size  $\Delta t$  toward zero, the approximate solutions that we get will converge to an exact solution to our initial value problem.

*Exercise 3.5.* Show that the constant  $T$  in the Picard-Lindelöf theorem may indeed need to be a finite number by considering the simple ODE

$$\dot{x} = x^2.$$

(See Appendix A if you need help solving this equation.)

**3.4. A note on solving ODEs.** The good news, as we've seen, is that solutions to ODEs are guaranteed to exist. But there's also a piece of news that you might think of as bad: you're unlikely to be able to write down solutions to them in explicit formulae. Yes, we've seen some examples where you can, and we'll see a few more, but the reality is that if you write down a typical ODE, while you know that solutions must exist there is probably not a formula in terms of the functions that you know and love (mostly trigonometric, exponential, and polynomial) that gives the answer.

But the good news is that the bad news isn't so bad. Just because you don't have a formula for a function doesn't mean it's not a perfectly good function. Indeed, lots of important functions in math are *defined* as solutions to ODEs. We can study these functions without necessarily having names for them, and a lot of the tools will develop in this class are exactly geared toward understanding qualitative behavior of motion where we don't have explicit formulas available.

To take one rather familiar example, we could define a function  $f$  to be the solution to the initial value problem

$$f'' + \omega^2 f = 0, \quad f(0) = 0, \quad f'(0) = 1.$$

We could study it based on this definition and by doing a little work, we could find that it enjoyed many marvelous and interesting properties (for instance, it's  $2\pi/\omega$  periodic, and  $f^2 + (f')^2 = 1$ ). Of course, this is a familiar function after all, known as  $\sin$ , but you could have taken this ODE to be its *definition*. Similarly, the solutions to

$$r^2 g''(r) + r g'(r) + (r^2 - \omega^2) g(r) = 0$$

have lots of wonderful properties and indeed show up a lot in physics. These are known as *Bessel functions*, and while there isn't a button for them on your pocket calculator, they're just as mathematically honest as  $\sin$  and  $\cos$ . It's probably just the existence of the calculator button that's responsible

for most people’s sense that  $\sin$  and  $\cos$  are familiar and concrete objects, while Bessel functions and the like are somehow mysterious and remote.

**3.5. Back to systems.** What about solving systems of equations?—after all, this is the problem we started with. It turns out that the same story holds as in single equations, only if we have, say, a system of three second ODEs for  $x_1, x_2, x_3$ , we need to specify as our initial data three sets of initial positions and velocities:

$$x_1(0), x_2(0), x_3(0), \dot{x}_1(0), \dot{x}_2(0), \dot{x}_3(0).$$

Both of the “proofs” given above that a solution exists and is unique go through in this higher-dimensional setting, hence the Picard-Lindelöf Theorem holds for ODEs

$$\ddot{\mathbf{x}}G(t, \mathbf{x}, \dot{\mathbf{x}})$$

in any number of dimensions.

*Exercise 3.6.* Returning to single equations: find all solution to the particle in a constant gravitational field in one dimension:

$$\ddot{x} = -1.$$

Find the solution that matches the initial conditions  $x(0) = 0, \dot{x}(0) = 5$  corresponding to a ball thrown vertically into the air from the ground with initial velocity 5. At what time does the ball attain its maximum height?

*Exercise 3.7.* Now in three-dimensions, find all solutions to the particle in a constant gravitational field

$$\ddot{\mathbf{x}} = -\mathbf{e}_3.$$

Find the solution that matches the initial conditions

$$\mathbf{x}(0) = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad \dot{\mathbf{x}}(0) = \begin{pmatrix} 1 \\ 2 \\ 5 \end{pmatrix}$$

corresponding to a ball thrown vertically into the air from the origin with vertical speed  $t$  and horizontal velocity  $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ .

Where does the ball hit the “ground,” i.e. the plane  $x_3 = 0$ ?

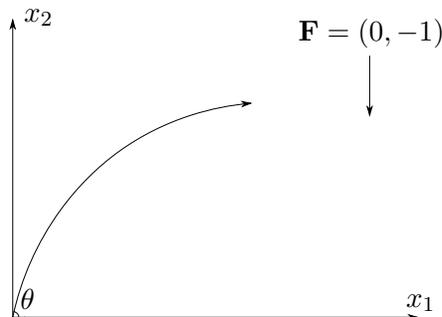
*Exercise 3.8.* In two dimensions, find the general solution to the system of harmonic oscillators

$$(18) \quad \ddot{\mathbf{x}} = \begin{pmatrix} -x_1 \\ -3x_2 \end{pmatrix}$$

Which of these solutions to the system are *periodic*?

*Exercise 3.9.* Consider a unit mass shot from a cannon at the origin of coordinates  $(0, 0) \in \mathbb{R}^2$ ; we will take gravity to act by  $\mathbf{F} = (0, -1)$ . Suppose the particle is shot at unit speed with angle  $\theta$  to the  $x_1$ -axis, hence  $\dot{\mathbf{x}}(0) = (\cos \theta, \sin \theta)$ . Find the point where the object hits the “ground” (i.e. the axis

$x_2 = 0$ ) as a function of  $\theta$ . What  $\theta$  should we choose to make the projectile go as far as possible?



*Exercise 3.10.* Consider a unit mass on a spring with spring constant 1 *acted on by friction*. We model the force of friction to be proportional to minus the velocity, hence from Newton's second law we obtain the ODE

$$\ddot{x}(t) = -\gamma\dot{x}(t) - x(t)$$

for the displacement  $x(t)$  of the mass at time  $t$ ; here  $\gamma$  is a (positive) constant measuring how strong the force of friction is.

For help in solving this ODE, you might like to consult Appendix B.

- (1) Assume that  $\gamma > 2$ . Find all possible solutions to this ODE, and sketch a typical solution.
- (2) Now do the same when  $\gamma < 2$ . Once again sketch a typical solution.
- (3) Now do the special case  $\gamma = 2$ .
- (4) Find the solution to the problem corresponding to the initial condition  $x(0) = 5$ ,  $\dot{x}(0) = 0$  in each of the three scenarios above. If this damped oscillator represents a screen door on a closing mechanism (a spring to pull the door shut and a pneumatic device to add friction) how do you think you ought to choose  $\gamma$  to make the thing work best?

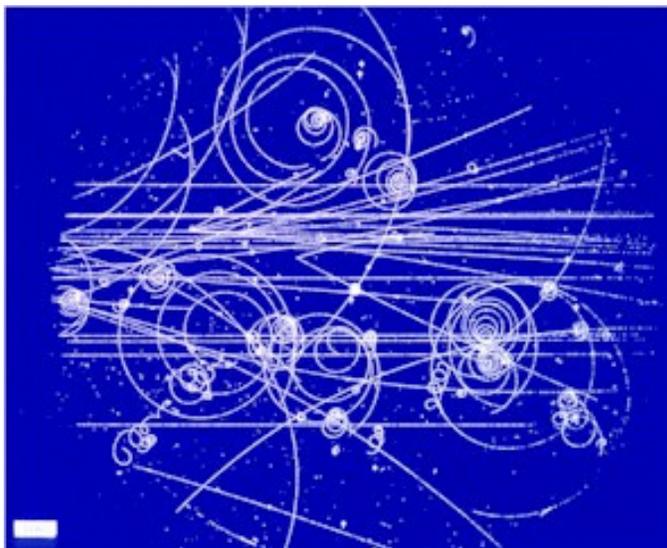
*Exercise 3.11.* Recall that a charged particle with charge  $q$  (positive or negative) moving in a magnetic field  $\mathbf{B}$  experiences a force (the “Lorentz force”)  $\mathbf{F} = q\dot{\mathbf{x}} \times \mathbf{B}$ . Here  $\mathbf{B}$  is a vector field, i.e. a vector at each point in space representing the “magnetic field” and we are taking the cross product. (Note: here is a force law that doesn't depend on  $\mathbf{x}$  but rather on  $\dot{\mathbf{x}}$ !)

Consider the simplest case where  $\mathbf{B} = (0, 0, 1)$  everywhere in space,  $q = 1$  and the mass  $m = 1$ . Thus, the position of the particle satisfies

$$\ddot{\mathbf{x}} = \dot{\mathbf{x}} \times (0, 0, 1).$$

Assume that the initial location of the particle is  $\mathbf{x}(0) = (0, 0, 0)$  and the initial velocity is  $\dot{\mathbf{x}}(0) = (1, 0, 0)$ .

- (1) Write down a system of equations for the components of  $\mathbf{x}(t)$ , i.e. equations for  $\ddot{x}_1$ ,  $\ddot{x}_2$ ,  $\ddot{x}_3$ . Note that unlike, for instance, the system



(18), these equations are not independent of one another: e.g. the equation for  $\ddot{x}_1$  will involve  $\dot{x}_2$ .

- (2) To make your life easier, consider the components of the velocity  $v_i = \dot{x}_i$  for  $i = 1, 2, 3$ . Show that your equations turn into three first order differential equations for the velocity components:  $\dot{v}_1 = v_2$ ,  $\dot{v}_2 = -v_1$ ,  $\dot{v}_3 = 0$ .
- (3) Solve these equations by differentiating them and substituting one into another, then recognizing the harmonic oscillator equations. Use your initial conditions to find an explicit formula for  $\dot{\mathbf{x}}(t)$  and then integrate to get  $\mathbf{x}(t)$ . Sketch the solution.

Now explain Figure 3, which is a photo from a “bubble chamber,” of charged particles moving in a magnetic field pointing perpendicular to the plane of the picture.

**3.6. Newton’s second law.** In ending our discussion of ODE, I want to try and address the question: What does Newton’s second law of motion

$$\mathbf{F} = m\mathbf{a}$$

really mean as a fundamental law of nature? After all, whatever trajectory a particle takes, we can always compute  $\mathbf{a}$  and retroactively *define*  $\mathbf{F}$  to be this quantity. This seems to have no content unless we have a good definition for  $\mathbf{F}$ .

One big piece of content comes in the implicit assumption that  $\mathbf{F}$  actually only depends on  $t$ ,  $\mathbf{x}$  and  $\mathbf{v} = \dot{\mathbf{x}}$ . In particular, it shouldn’t depend on  $\mathbf{a} = \ddot{\mathbf{x}}$  or on higher derivatives: if it did, the equation  $\mathbf{F} = m\mathbf{a}$  wouldn’t mean very much, after all—it might just read  $m\mathbf{a} = m\mathbf{a}$ , for instance.

Thus, the content of Newton’s second law is that we can write  $\ddot{\mathbf{x}}$  in terms of  $t$ ,  $\mathbf{x}$ , and  $\dot{\mathbf{x}}$ , which is simply to say: *the position  $\mathbf{x}$  satisfies a second*

*order ODE*. This is a statement with real consequences. Since the initial value problem for a second-order ODE with initial position and velocity has a unique solution, Newton's second law therefore tells us that knowledge of a particle's position and momentum at a given moment are sufficient to determine its position at all later times. Maybe there exists some other universe in which the equations of motion are 17<sup>th</sup> order and therefore one would have to know the first 16 derivatives of the position to predict the future, but in our world, second-order equations seem to be the fundamental ones.

The second law implicitly tells us a lot else about mass and force and how they behave, but we will set aside these more physical, rather than mathematical, aspects of the law in order to get as quickly as possible to reformulating mechanics, so as to eliminate the pesky notion of force altogether.

#### 4. ONE-DIMENSIONAL FORCE LAWS

**4.1. Motion in a potential; energy conservation.** A frequent feature of force laws, shared by all but one of our motivating examples (2)–(7), is that the force  $\mathbf{F}(\mathbf{x}, \dot{\mathbf{x}}, t)$  may depend only on the position  $\mathbf{x}$  and not on the velocity  $\dot{\mathbf{x}}$  or on the time  $t$ . Let us explore this set-up in the case of one-dimensional problems, so that if  $m = 1$ , Newton's Second Law reads

$$(19) \quad \ddot{x} = F(x).$$

Our harmonic-oscillator example fits into this framework with  $F(x) = -\omega^2 x$ ; so does terrestrial gravity (written here with  $g = 1$ ):  $F(x) = -1$ .

It turns out that these equations have some special structure that makes them highly tractable. We can mimic the technique we used in Exercise 3.1, multiplying both sides of the equation by  $\dot{x}$ . This gives

$$(20) \quad \ddot{x}\dot{x} = F(x)\dot{x}.$$

Now we can recognize the left-hand side as the time derivative of  $(\dot{x})^2/2$ , by the chain rule. We would like to recognize the right side as a derivative, too; to do that, we need a name for the (minus) the anti-derivative of  $F(x)$ : let  $V(x)$  be the function of  $x$  such that<sup>6</sup>  $V'(x) = -F(x)$ . ( $V$  is then only determined up to an arbitrary additive constant, but this choice of constant will not matter for any of our reasoning.) Then the right side is just  $(d/dt)V(x)$ , where the chain rule is responsible for the factor of  $\dot{x} = dx/dt$ . Thus, we can rewrite (19) as

$$(21) \quad \frac{d}{dt} \left( \frac{1}{2} \dot{x}^2 + V(x) \right) = 0.$$

Integrating both sides, we just get

$$\frac{1}{2} \dot{x}^2 + V(x) = E,$$

---

<sup>6</sup>Our choice to give a name to minus the anti-derivative rather than the antiderivative itself is arbitrary: it happens that this is the conventional way to write things.

where  $E$  is a constant, independent of time. *We have just derived our first conservation law*, which is to say, we have found a quantity that is constant in time as our physical system evolves. This particular quantity is usually known as *energy*, hence the choice of the letter  $E$  for its value. On the left side, the term

$$\frac{1}{2}\dot{x}^2$$

is usually called *kinetic energy*, as it derives from the motion of the particle (“kinetic” derives from the Greek verb “to move”); the term  $V(x)$  is called *potential energy*. In the example of the harmonic oscillator  $\ddot{x} = -\omega^2 x$ , we have  $F(x) = -\omega^2 x$  hence  $V(x) = (1/2)\omega^2 x^2$ . In the example of terrestrial gravity, we have  $F(x) = -1$ , hence  $V(x) = x$ . Thus, for the harmonic oscillator

$$\frac{1}{2}\dot{x}^2 + \frac{\omega^2}{2}x^2$$

is a constant, while for the particle under the influence of terrestrial gravity, it is

$$\frac{1}{2}\dot{x}^2 + x$$

that is conserved.

For a one-dimensional system like these, this observation of energy conservation is very powerful: powerful enough that it enables us to completely solve the system, at least up to the problem of doing a (potentially hard) integral. In a physics problem, this represents victory: if you can reduce the problem to doing an integral, then at worst, you can just do the integral numerically, on a computer, and get as good an approximation to the exact answer as you like, and very efficiently too.

Let’s go through this process for the particle in a gravitational field. We have a “conservation law”

$$\frac{1}{2}\dot{x}^2 + x = E,$$

and we can solve for  $\dot{x}$  to get

$$\frac{dx}{dt} = \pm\sqrt{2E - 2x},$$

i.e., solving this separable equation (see Appendix A) we obtain (fixing a choice of the sign of the square root)

$$\int dt = \int \frac{dx}{(2E - 2x)^{1/2}} = -(2E - 2x)^{1/2} + \text{const.}$$

Thus, for some constant  $C$ ,

$$\sqrt{2E - 2x} = -t + C$$

i.e.,

$$x = E - \frac{1}{2}(t - C)^2.$$

This gives you a family of solutions to the original ODE with two undetermined constants,  $E$  and  $C$ ; with this in hand, you can solve the initial value problem. You might reasonably feel that this was an unnecessarily complicated way to solve an easy problem, but the good news is that you can solve hard problems this way too!

*Exercise 4.1.* Solve the initial value problem this way, finding  $E$  and  $C$  so as to arrange

$$x(0) = 1, \quad \dot{x}(0) = 3.$$

Remember also how this procedure went with the harmonic oscillator. In our new language, we know that

$$E = \frac{1}{2}\dot{x}^2 + \frac{\omega^2}{2}x^2$$

is constant. So we can write

$$\frac{dx}{dt} = \pm\sqrt{E - \omega^2 x^2},$$

i.e. integrating yields

$$t + C = \pm \int \sqrt{E - \omega^2 x^2} dx.$$

We can do this integral, and then solve for  $x$  in terms of  $t$  to find all possible solutions to the system.

*Exercise 4.2.* Do this (again), writing your solution in terms of the constants  $E$  and  $C$ , this time.

Note that once we have the force law  $F(x)$  we immediately compute the *potential*  $V(x)$  with  $V' = -F$  and then forget about the force. Thus, the potential is in fact the crucial thing to specify for a 1D problem of this type, and once we know it, and hence the conserved energy, our problem is one of doing an integral. So often it's easier just to specify the potential!

*Exercise 4.3.* The “real pendulum” consists of a mass on a rod in the plane, making an angle of  $\theta$  with the downward vertical. It has a potential function<sup>7</sup>  $V(\theta) = -\cos\theta$ , and the conserved energy is

$$E = \frac{1}{2}\dot{\theta}^2 - \cos\theta.$$

Write  $t$  as a function of  $\theta$  in terms of a definite integral. This is one that you can't do in closed form, but it's a famous kind of integral—an “elliptic integral” that's been much studied and tabulated.

---

<sup>7</sup> $V$  is simply given by the restriction of the gravitational potential  $V(x_1, x_2) = x_2 - 1$  to the circle  $x_1^2 + x_2^2 = 1$ —we'll discuss this situation properly when we talk about constraints.

**4.2. Features of one-dimensional potential motion.** For a particle moving in a one-dimensional potential, the method we've just described allows us in principle to solve the motion completely, but at the cost of having to do a potentially nasty integral and then taking the inverse function to the result. But with much less work, we can still learn a lot about the motion. Let's take the example of the harmonic oscillator. Say we have the motion in the potential

$$V(x) = \frac{1}{2}x^2$$

and we're told the initial position and velocity

$$x(0) = 1, \quad \dot{x}(0) = 3.$$

We know that the energy

$$E = \frac{1}{2}\dot{x}^2 + \frac{1}{2}x^2$$

is conserved, and the initial data allows us to compute that for this particle,  $E = 1/2 + 9/2 = 5$ . Without doing any work, we can find various nice qualitative features of the motion. For example, intuitively, the mass begins moving to the right, and keeps moving to the right (increasing  $x$ ) until it reaches some maximum displacement, and the spring starts yanking it back to the left ( $x$  starts to decrease). What is the maximum displacement, i.e. the maximum value achieved by  $x$ ? Well, we know that  $E = 5$  and  $E$  is a sum of two nonnegative quantities (both are squares). Thus, for  $x$  to be as big as possible, leaving  $E = 5$ , we require  $\dot{x} = 0$ . To put it another way: when  $x(t)$  is maximum, we of course have  $\dot{x} = 0$  as you learned in calculus. So when the spring is maximally stretched,

$$\frac{1}{2}x^2 = 5$$

i.e.,

$$x = \sqrt{10}.$$

By the same reasoning, you can see that  $x$  then decreases to a minimum of  $-\sqrt{10}$  before once more increasing to  $+\sqrt{10}$ , and so on.

You can also figure out how fast the mass gets going by the same line of reasoning:  $\dot{x}$  will be at its maximum, subject to the constraint

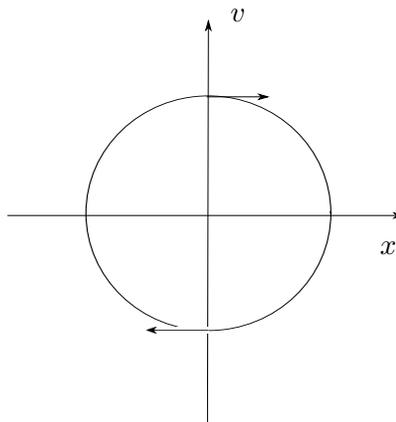
$$\frac{1}{2}\dot{x}^2 + \frac{1}{2}x^2 = 5$$

when the second term is zero. Thus, the maximum speed, i.e., the maximum of  $|\dot{x}|$ , is also  $\sqrt{10}$ .

*Exercise 4.4.* For the real pendulum, we have

$$E = \frac{1}{2}\dot{\theta}^2 - \cos \theta.$$

If  $\theta(0) = 0$ ,  $\dot{\theta}(0) = 1$ , find the maximum value of  $\theta$ . What if  $\theta(0) = 0$ ,  $\dot{\theta}(0) = 2$ ?



## 5. THE PHASE PLANE

Our discussion of the qualitative features of one-dimensional potential motion may seem clearer if we draw some pictures. We can represent the solution to our ODEs most clearly if we plot the values of the position  $x(t)$  and the velocity  $v(t) = \dot{x}(t)$  in the plane at each time. As time evolves, the values of these variables then trace out a curve in the plane. The plane with variables  $(x, v)$  in which we plot the trajectories is often called the *phase plane* and these curves are *phase curves*.

To begin with, let us try the harmonic oscillator with  $V(x) = x^2/2$ . Then our conservation law  $E = \dot{x}^2/2 + x^2/2 = (1/2)(v^2 + x^2)$  shows us that as time evolves, the value of  $v^2 + x^2$  remains constant. The curves  $v^2 + x^2 = \text{constant}$  in  $\mathbb{R}^2$  are of course very familiar ones: they are circles. Thus, the phase curve traces out a circle as time evolves. When  $v \geq 0$ , we have  $\dot{x} = v \geq 0$ , hence when  $v > 0$  the motion along the phase curve is toward the right and when  $v < 0$  it is to the left, so it is not hard to see that the motion around these circles in the  $(x, v)$  plane is clockwise.

We know, from our explicit solution of the harmonic oscillator, that the particle returns to its original state at time  $t = 2\pi$ —in other words, this is how long it takes to traverse each circle in the phase plane. It is a highly unusual feature of the harmonic oscillator that each of these circles is traversed in exactly the same length of time! More typically, the period is different along different phase curves, as you will see in the following exercise.

*Exercise 5.1.* Consider the “quartic oscillator,” i.e. the motion of a particle in the potential

$$V(x) = x^4.$$

- (1) Draw a phase plane diagram for the motion, and describe in words how the particle moves.
- (2) Find the maximum and minimum values of  $x$  attained along the motion, as functions of  $E$ , the energy.
- (3) Give a formula for how long it takes the particle to travel from the minimum to the maximum value; use this to find the *period*  $\tau$  of the motion, as a function of  $E$ . *Note that your answer will be in the form of an integral which you will probably not be able to evaluate explicitly.*
- (4) Make a change of variable  $u = x/(2E)^{1/4}$  in your integral to rewrite the period as a power of  $E$  times a (somewhat nasty) constant.

Now let us examine motion in a different potential; we will use the potential from the “real pendulum” in Exercise 4.3, where we write our position variable as  $\theta$  (as it is an angle of displacement) and the potential is  $V(\theta) = -\cos\theta$ . Thus the conserved energy is

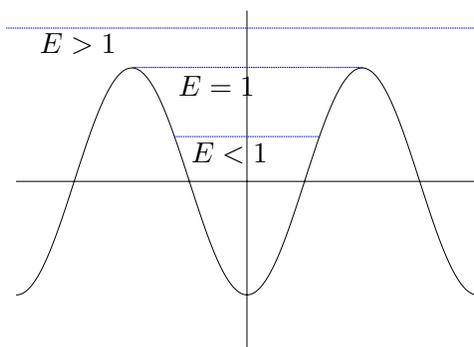
$$\frac{1}{2}\dot{\theta}^2 - \cos\theta = E.$$

Let us let  $\omega$  denote the “angular velocity”  $\dot{\theta}$ . The motion of this system in the  $\theta, \omega$  plane is thus along the curves

$$\frac{1}{2}\omega^2 - \cos\theta = E.$$

These curves, unlike those in the oscillator example, are interestingly varied as  $E$  varies. To see why, let us graph the potential  $-\cos\theta$  and recall that the maxima and minima that  $\theta$  can attain along an orbit are at solutions of  $-\cos\theta = E$  (see Figure 5). If  $|E| \leq 1$ , then there are indeed solutions to this equation, but if  $|E| > 1$ , there are none. Correspondingly, if the energy exceeds 1 then the phase curves go off to infinity, while if  $|E| < 1$ , they stay bounded. Thus the motion for  $E < 1$  oscillates back and forth between the maximum and minimum values of  $\theta$ —the pendulum oscillates like the pendulum of a clock. On the other hand, for  $E > 1$ , the pendulum has enough energy to go “over the top” and it simply spins around and around. In the critical case  $E = 1$ , the system has just enough energy to approach the potential maximum at  $\theta = \pi$  (i.e. straight up) as  $t \rightarrow \infty$ , but not to get past it.

Now let us graph the corresponding trajectories in the phase plane. For  $E < 1$ , we will have closed curves, traversed repeatedly in a clockwise direction; for  $E > 1$  the curves will extend to  $\theta \rightarrow \pm\infty$ . It is exceptionally instructive to note what happens when  $E$  is near  $-1$ , its minimum value for



closed trajectories. Let us write  $E = -1 + \epsilon$ . We have

$$\frac{1}{2}\omega^2 - \cos \theta = -1 + \epsilon$$

so since the first term is positive,

$$-\cos \theta \leq -1 + \epsilon$$

i.e.,

$$\cos \theta \geq 1 - \epsilon$$

Thus,  $\theta$  must be quite near zero (or some multiple of  $2\pi$  anyway). If we take advantage of this knowledge to Taylor expand  $\cos$  near  $\theta = 0$  we find that our original energy conservation equation reads

$$\frac{1}{2}\omega^2 - 1 + \frac{\theta^2}{2} \cong -1 + \epsilon,$$

i.e.,

$$\frac{1}{2}\omega^2 + \frac{\theta^2}{2} \cong +\epsilon,$$

But now this is just the equation for energy conservation of a harmonic oscillator with potential  $\frac{\theta^2}{2}$ ! This is a problem that we understand how to solve exactly: the curves traversed in the phase plane are simply circles and the solutions are in fact of the form  $\theta(t) = A \cos t + B \sin t$ .

Thus the level curves for our original problem that are near  $2\pi\mathbb{Z}$ , corresponding to  $E$  a little larger than  $-1$ , are nearly concentric circles centered at points in  $2\pi\mathbb{Z}$ . By contrast if  $E \gg 0$ , we have

$$\frac{1}{2}\omega^2 - \cos \theta = E.$$

hence we may simply solve for  $\omega$  in terms of  $\theta$ : we have

$$\omega = \pm\sqrt{2E + 2\cos \theta}$$

These periodic curves are relatively easy to graph.

Overall, a plot of the phase curves for the system is pictured in Figure 5. Note the special phase curves through the points  $(\theta = k\pi, \omega = 0)$ : these are the special curves on which  $E = \pm 1$ , and are the only level curves of

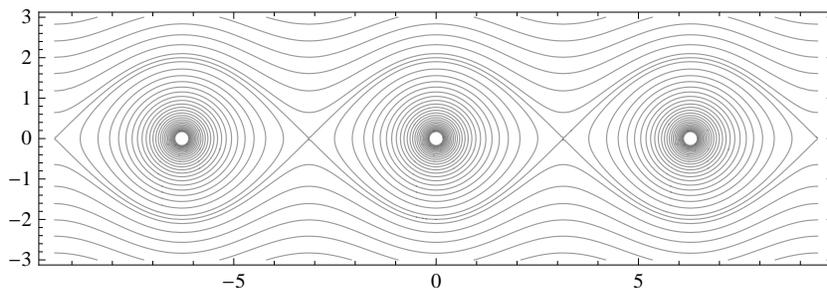


FIGURE 1. The phase plane for the real pendulum.

$(1/2)\omega^2 - \cos \theta$  that fail to be smooth curves, owing to the x-shape you see at these special points where the pendulum is upward-pointing vertical ( $k$  odd) and the phase curves corresponding to the pendulum at equilibrium ( $k$  even) that are single points.

**5.1. Behavior of the motion near critical points.** Our phase plane analysis of the real pendulum involved one exceptionally important feature: the analysis of the behavior of the *equilibrium solutions*.

*Definition 5.1.* An equilibrium solution of a system (in any number of dimensions) is one of the form  $\mathbf{x}(t) = \text{constant}$ .

In our one-dimension potential motion

$$\ddot{x} = -V'(x),$$

It is easy to see that if  $V'(x_0) = 0$  is a solution then the constant solution  $x(t) = x_0$  is an equilibrium solution; conversely, if an equilibrium solution at  $x_0$  exists, we have  $\ddot{x} = 0$  for this solution, hence  $V'(x_0) = 0$ .

Thus, *equilibrium solutions correspond to critical points of the potential*.

Equilibria may have a very different character depending on whether they are *stable* or *unstable*.

*Definition 5.2.* An equilibrium solution at  $\mathbf{x} = \mathbf{x}_0$  is called *stable* if for any neighborhood  $U$  of  $(\mathbf{x}_0, \mathbf{v} = 0)$  in the “phase space” there exists a smaller neighborhood set  $U' \subset U$  such that trajectories starting in  $U'$  stay in  $U$  for all  $t \in \mathbb{R}$ .

An equilibrium point is called *unstable* if it is not stable.

This definition (whose fine points we will not spend too much time on) just says that if we perturb a particle at a stable equilibrium, changing its position and momentum ever so slightly, it will stay nearby for all time. In the real world, we don’t usually expect to see particles in a state of unstable equilibrium as their position and momentum would have to be specified with infinite precision in order to keep them in equilibrium, and any small air current blowing through our laboratory would be ruinous. For instance,

we will see shortly that a pendulum standing straight up is an example of an unstable equilibrium. This is hard to achieve!

To analyze equilibria of one-dimensional potentials, we will use the Taylor series approximation of  $V$  as we did in the above example. If  $x(t) = x_0$  is an equilibrium solution of

$$\ddot{x} = -V'(x)$$

we of course must have  $V'(x_0) = 0$ . The motion for  $x$  close to  $x_0$  is well approximated if we replace  $V$  by its second order Taylor series expansion

$$\begin{aligned} V(x) &\cong V(x_0) + V'(x_0)(x - x_0) + \frac{1}{2}V''(x_0)(x - x_0)^2 \\ &= V(x_0) + \frac{1}{2}V''(x_0)(x - x_0)^2 \equiv V_{\text{quad}}. \end{aligned}$$

Let's let  $x - x_0 = y$  be a new variable; note that we still have  $\dot{y} = v$ . Then

$$V_{\text{quad}} = V(x_0) + \frac{1}{2}V''(x_0)y^2 = c \pm \frac{1}{2}\omega^2 y^2$$

where  $c = V(x_0)$  and  $\pm\omega^2 = V''(x_0)$ —we choose the sign according to whether this quantity is positive or negative.

If  $V''(x_0) > 0$ , the sign is  $+$ , and we note that  $V_{\text{quad}}$  is again the potential for our old friend, the harmonic oscillator (shifted by a constant, which makes no difference to the motion). Level sets of

$$E = \frac{v^2}{2} + V_{\text{quad}}(y) = \frac{1}{2}v^2 + c + \frac{\omega^2}{2}y^2$$

are thus *ellipses* in the  $(y, v)$  plane, except for the special case  $E = c$  which defines the single point  $y = v = 0$ . This is our equilibrium solution  $x = x_0$ , of course. Orbits starting near our equilibrium solution thus stay near it, as they are constrained to stay on these small ellipses.

On the other hand, if  $V''(x_0) < 0$ , we have

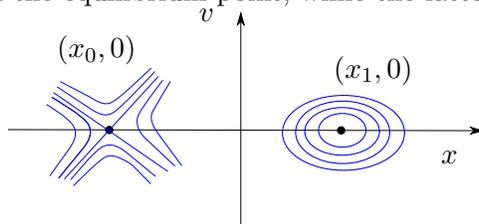
$$E = \frac{v^2}{2} + V_{\text{quad}}(y) = \frac{1}{2}v^2 + c - \frac{\omega^2}{2}y^2.$$

The level sets of this energy function are *hyperbolas*, again except for the exceptional case when  $E = c$  corresponding to our equilibrium solution. None of these hyperbolas is completely contained in a small neighborhood of the point  $(y = 0, v = 0)$ , so it is not hard to see that the flow must move out of any neighborhood of this point as  $t$  increases or decreases. Thus, we have established the following:

**Theorem 5.3.** *For a one-dimensional potential  $V(x)$ , an equilibrium solution at  $x_0$  is stable if  $V''(x_0) > 0$  and unstable if  $V''(x_0) < 0$ .*

*Exercise 5.2.* Pursue our discussion of stable equilibrium solutions a little further as follows: Let  $V(x)$  be a potential with a local minimum at  $x = 0$ .

FIGURE 2. Phase plane diagram with unstable equilibrium at  $x = x_0$  and stable equilibrium at  $x = x_1$ . Note that the former curves are not completely contained in a small neighborhood of the equilibrium point, while the latter are.

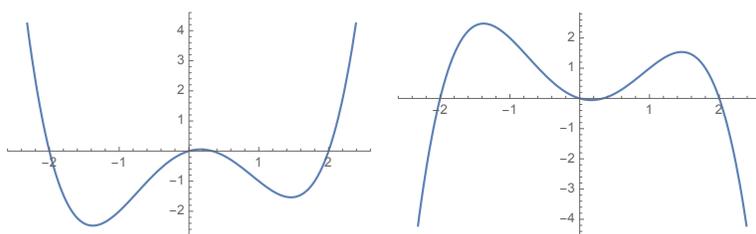


Show that trajectories for a unit-mass particle near the equilibrium solution  $x = 0$  oscillate with period approximately  $2\pi/\sqrt{V''(0)}$ .

*Note that this exercise explains the universal appearance of the harmonic oscillator throughout physics: it describes the first approximation to any system near a nondegenerate stable equilibrium (nondegenerate means the strict inequality  $V'' > 0$ ).*

*Exercise 5.3.* Now take the potential  $V(x) = -\omega^2 x^2$ , our model form for an unstable equilibrium. Solve the equations of motion to show that for *almost* any initial conditions of the particle near  $x = 0, v = 0$ , the particle moves away from  $x = 0$  (indeed, rather quickly) as times runs forward or backward. Show, however, that there are certain, rather special, initial conditions for which the particle fails to escape the equilibrium either for positive or negative  $t$ . What happens along these trajectories?

*Exercise 5.4.* Draw a phase plane diagram for motion in each of the following two potentials.



Make sure to label equilibrium points and draw the motion near these points with some care. What happens along the trajectories that escape to infinity?

**5.2. Closed phase curves and periodic motion.** In this section we describe a feature of one-dimensional potential motion that was at least intuitively clear in the case of the real pendulum, but is true in much greater generality. Recall that in the case of the real pendulum, the phase curves with  $E < 1$  were smooth, closed curves. We noted that the motion corresponded to the pendulum oscillating back and forth between two extrema,

and in particular the motion was *periodic*. This is true much more generally whenever phase curves are closed curves in the plane, as long as they do not look like the funny and rather special closed curves at  $E = 1$ , which contain equilibrium points.

**Theorem 5.4.** *Let  $V(x) \in \mathcal{C}^1(\mathbb{R})$  be a potential and consider the phase curve*

$$\Gamma = \{(x, v) : v^2/2 + V(x) = E\}.$$

*Assume that  $\Gamma$  is a closed curve. Suppose that there is no point on  $\Gamma$  on which simultaneously  $v = 0$  and  $V'(x) = 0$ . Then if the trajectory  $x(t)$  satisfies  $\dot{x}^2/2 + V(x) = E$ , the motion is periodic, i.e., for some  $\tau > 0$  we have*

$$x(t + \tau) = x(t)$$

*for all  $t$ .*

We now sketch the proof of this theorem; a fully rigorous proof is beyond the scope of these notes, but this one will be close to rigorous.

The main thing to check is that the trajectory must go all the way around the phase curve, and cannot get “stuck” partway along it. To this end, we let  $s$  denote the “speed” of the trajectory in the  $(x, v)$  plane given by

$$s = (\dot{x}^2 + \dot{v}^2)^{1/2}.$$

(Note that this has nothing much to do with the speed of the physical particle, which is simply  $|v|$ .) The speed  $s$  can never be 0 since if that happened, we would have  $\dot{x} = \dot{v} = 0$ . Since  $\dot{x} = v$  and  $\dot{v} = \ddot{x} = -V'(x)$  this could only happen at a point at which  $v = 0$ ,  $V'(x) = 0$  and by hypothesis, none of these lie on  $\Gamma$ . Therefore, since the curve is closed<sup>8</sup> there is a minimum value  $s_0$  for  $s$ . Now we also note that on  $v > 0$ ,  $\dot{x} > 0$  and on  $v < 0$ ,  $\dot{x} < 0$  so that the trajectory traces out the phase curve clockwise, with at least speed  $s_0$ . Thus if  $L$  denotes the length of the curve  $\Gamma$ , the trajectory starting at  $(x_0, v_0)$  must return to the point  $(x_0, v_0)$  in some time no greater than  $L/s_0$ . Let  $\tau$  denote the smallest positive time such that

$$(x(\tau), v(\tau)) = (x_0, v_0)$$

Now consider the function  $y(t) = x(\tau + t)$ . We easily compute that  $y(0) = x(\tau) = x_0$  and  $\dot{y}(0) = \dot{x}(\tau) = v_0$ . Also, of course since  $\ddot{x} = -V'(x)$ , we also have

$$\ddot{y} = -V'(y).$$

Thus,  $y(t)$  satisfies the same initial value problem as  $x(t)$ . Hence by uniqueness of solutions to ODE with given initial conditions,

$$y(t) = x(t) \text{ for all } t,$$

i.e.,

$$x(t + \tau) = x(t)$$

---

<sup>8</sup>Technically the big point here is that  $\Gamma$  is a compact set and  $s$  is a continuous function on it, everywhere positive.

for all  $t$ . □

*Exercise 5.5.* For each  $E \in \mathbb{R}$  let  $\Gamma(E)$  denote the level set  $(1/2)v^2 + V(x) = E$  in the phase plane. Assume that these are all closed curves. For each  $E$ , let  $A(E)$  denote the area enclosed by  $\Gamma(E)$ . Let  $\tau(E)$  denote the period of the motion on  $\Gamma(E)$ . Show that

$$\tau(E) = \frac{dA}{dE}.$$

## 6. MOTION IN A POTENTIAL IN HIGHER DIMENSIONS

The real world has (at least) three space dimensions, so we now turn to the study of higher-dimensional motion; we may as well allow the more general setting  $\mathbf{x} \in \mathbb{R}^n$  at this point,<sup>9</sup> although we will return to the special case of 3 dimensions in a little while.

We will continue to pursue the case of a force law that depends on the location of the particle, but not on velocity or time, so that we have a system of ODEs that can be written in vector notation as.

$$(22) \quad \ddot{\mathbf{x}}(t) = \mathbf{F}(\mathbf{x}(t)).$$

In the 1-dimensional case, we made great strides at this point and taking an anti-derivative, writing the force as  $-V'(x)$ , then integrating and finding a conserved energy. What, you should wonder, is the analogous procedure in higher dimensions? What does it even mean to write the force, which is a *vector*, as a derivative?

One compelling answer to this is that it might be sensible to hope that the vector field  $\mathbf{F}(\mathbf{x})$  is (minus) the *gradient* of a scalar function  $V(\mathbf{x})$ :

$$\mathbf{F}(\mathbf{x}) = -\nabla V(\mathbf{x}).$$

In  $\mathbb{R}^1$ , the existence of such an antiderivative is guaranteed by the Fundamental Theorem of Calculus: we can always produce our potential  $V$  from  $F$ . But in higher dimensions this is not always possible. For instance, consider, in  $\mathbb{R}^2$ , the vector field

$$\mathbf{F}(x_1, x_2) = (-x_2, x_1).$$

If we have  $\mathbf{F} = -\nabla V$ , then  $V$  solves the system of partial differential equations

$$\frac{\partial V}{\partial x_1} = x_2, \quad \frac{\partial V}{\partial x_2} = -x_1$$

Let us apply  $\partial/\partial x_2$  to the first of these equation and  $\partial/\partial x_1$  to the second. This gives

$$\frac{\partial^2 V}{\partial x_1 \partial x_2} = 1, \quad \frac{\partial^2 V}{\partial x_2 \partial x_1} = -1.$$

---

<sup>9</sup>There are good reasons for allowing dimensions higher than 3. One compelling one is that if we have a *system* of  $N$  particles in  $\mathbb{R}^3$ , interacting with each other and with some externally-imposed force, we can collect all the position vectors into a single vector  $\mathbf{x} \in \mathbb{R}^{3N}$  and write our second-order ODE for this vector.

But remember from calculus class that for any function that is in  $\mathcal{C}^2$ , i.e., has two derivatives everywhere, varying continuously in space (as the functions  $\pm 1$  assuredly do), the second mixed partials commute, i.e., should satisfy

$$\frac{\partial^2 V}{\partial x_1 \partial x_2} = \frac{\partial^2 V}{\partial x_2 \partial x_1}.$$

So no potential function can exist for the force field  $\mathbf{F} = (-x_2, x_1)$ .

In general, how can we tell, given a vector field  $\mathbf{F}$ , whether  $\mathbf{F} = -\nabla V$  for some function  $V$ ? We will soon answer this in 2 and 3 dimensions (but not in general). We will begin, though, by noting that the procedure we used on the 2-dimensional example above, to check that it was indeed not a gradient, has a nice generalization to  $\mathbb{R}^3$ . To wit, given a vector field  $\mathbf{F}$  in  $\mathbb{R}^3$  consider the *curl* of the vector field

$$\nabla \times \mathbf{F}.$$

Recall that you can compute this by formally taking the determinant

$$\begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ (\partial/\partial x_1) & (\partial/\partial x_2) & (\partial/\partial x_3) \\ F_1 & F_2 & F_3 \end{vmatrix} = \begin{pmatrix} \frac{\partial F_2}{\partial x_3} - \frac{\partial F_3}{\partial x_2} \\ \frac{\partial F_3}{\partial x_1} - \frac{\partial F_1}{\partial x_3} \\ \frac{\partial F_1}{\partial x_2} - \frac{\partial F_2}{\partial x_1} \end{pmatrix}$$

The following is an essential fact of vector calculus:

**Proposition 6.1.** *For any function  $V \in \mathcal{C}^2$ ,*

$$\nabla \times (\nabla V) = 0.$$

*Exercise 6.1.* Prove the proposition.

As a consequence, we do have an easy way to test when a vector field is *not* a gradient!

**Corollary 6.2.** *Let  $\mathbf{F}$  be a vector field in  $\mathbb{R}^3$ . If  $\nabla \times \mathbf{F} \neq 0$ , then there does not exist a potential function  $V \in \mathcal{C}^2$  with  $\mathbf{F} = -\nabla V$ .*

What about the converse? Is the condition of having vanishing curl necessary and sufficient to be a gradient? The answer turns out to be an interestingly qualified yes. To see how this works, we need to recall some further vector calculus concepts. In particular, since undoing the operation of derivative involves integrating, we will need to think about undoing the gradient by some kind of integral applied to a vector field.

Let  $\gamma(s)$  be a parametrized curve in  $\mathbb{R}^n$ , i.e.

$$\gamma(s) = \begin{pmatrix} x_1(s) \\ \vdots \\ x_n(s) \end{pmatrix}$$

is a map from some interval  $[a, b] \subset \mathbb{R}$  to  $\mathbb{R}^n$ . We will assume that all the  $x_i$  are infinitely differentiable; we then say  $\gamma(s) \in \mathcal{C}^\infty$ .

*Definition 6.3.* Let  $\gamma(t) \in \mathcal{C}^\infty$  be a parametrized curve, and let  $\mathbf{v}(\mathbf{x})$  be a smooth vector field on  $\mathbb{R}^n$ . The *line integral*

$$(23) \quad \int_{\gamma} \mathbf{v} \cdot d\mathbf{x}$$

is defined as

$$(24) \quad \int_a^b \mathbf{v}(\gamma(s)) \cdot \gamma'(s) ds.$$

Here

$$\gamma'(s) = \begin{pmatrix} x_1'(s) \\ \vdots \\ x_n'(s) \end{pmatrix}$$

is the tangent vector to the curve. This definition thus defines the line integral in terms of ordinary Riemannian integration of scalar functions.

**Proposition 6.4.** *The definition of the line integral does not depend on the parametrization of  $\gamma$ , but only on its orientation, i.e., the direction in which it is traversed. In other words if  $s = f(s')$  with  $f$  smooth and increasing, and  $\tilde{\gamma}(s') = \gamma(f(s'))$  is a “new” curve defined for  $s' \in [f^{-1}(a), f^{-1}(b)]$  then*

$$\int_{\gamma} \mathbf{v} \cdot d\mathbf{x} = \int_{\tilde{\gamma}} \mathbf{v} \cdot d\mathbf{x}$$

*Exercise 6.2.* Prove the proposition. HINT: It’s just the chain rule!

*Definition 6.5.* A vector field  $\mathbf{v}$  defined on a connected<sup>10</sup> open set  $\Omega \subset \mathbb{R}^n$  is *conservative* if  $\int \mathbf{v} \cdot d\mathbf{x}$  is independent of path, i.e. if whenever  $\gamma$  and  $\rho$  are paths with the same starting and ending points, we have

$$\int_{\gamma} \mathbf{v} \cdot d\mathbf{x} = \int_{\rho} \mathbf{v} \cdot d\mathbf{x}$$

This notion of conservativity turns out to be equivalent to being a gradient:

**Theorem 6.6.** *Let  $\mathbf{v}$  be a smooth vector field defined in a connected open set  $\Omega \subset \mathbb{R}^n$  that is conservative. Then there exists a function  $g(\mathbf{x})$  with  $\nabla g = \mathbf{v}$ . Conversely, any  $\mathbf{v}$  that is given by the gradient of some scalar function  $g$  is a conservative vector field.*

Thus, being conservative is the same as having an anti-derivative in exactly the sense that we want.

---

<sup>10</sup>We assume for simplicity that  $\Omega$  is *connected*, which means, for our purposes, that any two points may be connected by a path  $\gamma$ .

*Proof.* Assume that  $\Omega$  is a connected set (and work in a single connected component if it is not.) Fix a point  $\mathbf{x}_0 \in \Omega$ . For any  $\mathbf{y} \in \Omega$ , let  $\gamma_{\mathbf{y}}$  denote a smooth oriented curve starting at  $\mathbf{x}_0$  and ending at  $\mathbf{y}$ . Define

$$g(\mathbf{y}) = \int_{\gamma_{\mathbf{y}}} \mathbf{v} \cdot d\mathbf{x}.$$

This is well-defined, regardless of which path we choose for our  $\gamma_{\mathbf{y}}$  by our assumption on path independence.

We now compute the partial derivatives of  $g$  at  $\mathbf{y} = (y_1, \dots, y_n)$ , starting with  $\partial g / \partial x_1$ . We first compute

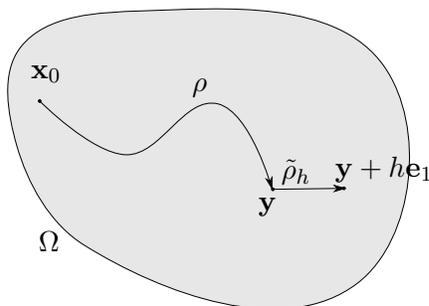
$$g(\mathbf{y} + h\mathbf{e}_1)$$

for each  $h$  by defining a family of oriented curves  $\gamma_h$  by letting each  $\gamma_h$  be the concatenation of a fixed path  $\rho$  from  $\mathbf{x}_0$  to  $\mathbf{y}$  followed by the straight line path from  $\mathbf{y}$  to  $\mathbf{y} + h\mathbf{e}_1$  given by

$$\mathbf{y} + s h \mathbf{e}_1, \quad s \in [0, 1]$$

and denoted  $\tilde{\rho}_h$ . Then

FIGURE 3. The curve  $\gamma$ , consisting of the curve  $\rho$  from  $\mathbf{x}_0$  to  $\mathbf{y}$ , followed by a straight line.



$$\begin{aligned} g(\mathbf{y} + h\mathbf{e}_1) - g(\mathbf{y}) &= \int_{\gamma_h} \mathbf{v} \cdot d\mathbf{x} - \int_{\rho} \mathbf{v} \cdot d\mathbf{x} \\ &= \int_{\tilde{\rho}_h} \mathbf{v} \cdot d\mathbf{x} \\ &= \int_0^1 \mathbf{v}(\mathbf{y} + s h \mathbf{e}_1) \cdot h \mathbf{e}_1 ds \\ &= \int_0^h v_1(\mathbf{y} + s' \mathbf{e}_1) ds' \end{aligned}$$

Hence as  $h \rightarrow 0$ ,

$$\frac{1}{h}(g(\mathbf{y} + h\mathbf{e}_1) - g(\mathbf{y})) = \frac{1}{h} \int_0^h v_1(\mathbf{y} + s'\mathbf{e}_1) ds' \rightarrow v_1(\mathbf{y})$$

by the Fundamental Theorem of Calculus. Thus, we have established that  $\partial g/\partial x_1 = v_1$ , and the computation of the other components follows in the same manner. This proves that a conservative vector field is a gradient.

The converse is very easy. If  $\mathbf{v} = \nabla g$ , then path independence follows from the Fundamental Theorem of Calculus (Exercise 6.3 below.)  $\square$

*Exercise 6.3.* Use the Fundamental Theorem of Calculus to show that a vector field that is a gradient is path independent.

We have thus reformulated our problem of seeing when a vector field is a gradient to checking that it is conservative, i.e., that its integral is path independent. It may seem that this does not simplify the problem at all: how might we know that the integral of a vector field doesn't depend on the path chosen? A very special answer to this question is available in  $\mathbb{R}^3$ , where a key theorem from vector calculus allows us to see that there are some circumstances in which the integral only depends on the starting and ending points for the path  $\gamma$ . This result is Stokes's Theorem.

To state Stokes's Theorem, recall that a surface in  $\mathbb{R}^3$  is said to be *oriented* if we equip it with a smoothly varying choice of unit normal vector at each point.<sup>11</sup> Let  $\mathbf{v}$  be smooth vector field in  $\mathbb{R}^3$ . Let  $\Sigma \subset \mathbb{R}^3$  denote an oriented surface with boundary, and let  $\partial\Sigma$  denote the boundary of the surface. If we let  $\mathbf{N}$  denote the oriented normal to  $\Sigma$ , we choose an orientation of  $\partial\Sigma$ , i.e., direction of the unit tangent vector  $\mathbf{t}$  of  $\partial\Sigma$ , in such a way that the cross product  $\mathbf{N} \times \mathbf{t}$ , which perpendicular to both, points *into* the surface.<sup>12</sup>

**Theorem 6.7.** For any  $C^\infty$  vector field  $\mathbf{v}$  on  $\mathbb{R}^3$ ,

$$\int_{\Sigma} (\nabla \times \mathbf{v}) \cdot \mathbf{N} dS = \int_{\partial\Sigma} \mathbf{v} \cdot d\mathbf{x}.$$

The left side is a *surface integral* with respect to the surface element on  $\Sigma$ . The theorem thus says that the circulation of  $\mathbf{v}$  around the boundary of  $\Sigma$  equals the integral of its curl on the interior. It justifies our intuition that the curl of a vector field measures its infinitesimal circulation at one point.

---

<sup>11</sup>Not every surface can be oriented: think about the Möbius band. If a surface can in fact be oriented, then there are two possible choices for the orientation corresponding to the two sides of the surface.

<sup>12</sup>This condition on the compatibility of the orientations can be rephrased in a more intuitive way as follows: we can regard  $\mathbf{N}$  as giving us a notion of *above* and *below*  $\Sigma$ , with the former being points of the form  $\mathbf{x} + \epsilon\mathbf{N}$  with  $\mathbf{x} \in \Sigma$  and  $\epsilon > 0$ . Then if we traverse  $\partial\Sigma$  in the direction of its orientation, with our heads *above*  $\Sigma$ , the surface  $\Sigma$  is on our left.

**Corollary 6.8.** *Let  $\mathbf{v}$  be a  $C^\infty$  vector field on  $\mathbb{R}^3$  with  $\nabla \times \mathbf{v} = 0$ . Then if  $\gamma_1, \gamma_2$  are two oriented curves that start and end at the same points,*

$$\int_{\gamma_1} \mathbf{v} \cdot d\mathbf{x} = \int_{\gamma_2} \mathbf{v} \cdot d\mathbf{x}.$$

*In other words, the integral is path-independent and  $\mathbf{v}$  is conservative.*

*Proof.* (Sketch.) Given such  $\gamma_i$ , there exists an oriented surface  $\Sigma$  whose boundary is  $\gamma_1 - \gamma_2$ , meaning  $\gamma_1$  traversed forwards (in the direction of its orientation) followed by  $\gamma_2$  traversed backwards (in the opposite direction to its orientation). Then applying Stokes's theorem gives

$$0 = \int_{\gamma_1 - \gamma_2} \mathbf{v} \cdot d\mathbf{x} = \int_{\gamma_1} \mathbf{v} \cdot d\mathbf{x} - \int_{\gamma_2} \mathbf{v} \cdot d\mathbf{x}.$$

(The existence of the surface in question is a bit of a cheat here. For any two reasonable curves  $\gamma_i$ , it is not hard to draw one, but if both curves are knotted as well as entangled with one another, the surface is a little less obvious. We will leave a full and rigorous proof of this result for the student to read elsewhere.)  $\square$

*Exercise 6.4.* Define a vector field on  $\mathbb{R}^3$  by

$$\mathbf{v}(\mathbf{x}) = \begin{pmatrix} x_2/(x_1^2 + x_2^2) \\ -x_1/(x_1^2 + x_2^2) \\ 0 \end{pmatrix}.$$

- (1) Compute  $\nabla \times \mathbf{v}$ .
- (2) Compute  $\int_{\gamma} \mathbf{v} \cdot d\mathbf{x}$  where  $\gamma$  is the circle in the  $x_1 - x_2$  plane

$$\gamma(s) = \begin{pmatrix} \cos s \\ \sin s \\ 0 \end{pmatrix}.$$

- (3) Why aren't the first two parts of this problem a contradiction to Stokes's theorem?

While we are at it, it is instructive to analyze a related problem in vector calculus. We found that in  $\mathbb{R}^3$ , the curl of a gradient of any function is always zero. What if we start with a vector field and take its curl: is there some operation we can do on *that* that always gives zero? We only have one other arrow in our vector calculus quiver, only one other operation on vector fields: the divergence. And indeed, the divergence of a curl is always zero!

**Proposition 6.9.** *For any  $C^2$  vector field  $\mathbf{v}$  on  $\mathbb{R}^3$ ,*

$$\nabla \cdot (\nabla \times \mathbf{v}) = 0.$$

*Exercise 6.5.* Prove the proposition.

Just as with the curl of a gradient, this leads to a natural next question: if the divergence of a vector field  $\mathbf{w}$  is zero, must there exist a vector field  $\mathbf{v}$  with  $\mathbf{w} = \nabla \times \mathbf{v}$ ? As before, it turns out that the answer is yes, provided we are careful to make sure that  $\mathbf{w}$  is defined and divergence-free on all of  $\mathbb{R}^3$ :

**Theorem 6.10.** *Let  $\mathbf{w}$  be a  $C^1$  vector field on  $\mathbb{R}^3$  with  $\nabla \cdot \mathbf{w} = 0$ . Then there exists a vector field  $\mathbf{v}$  such that*

$$\nabla \times \mathbf{v} = \mathbf{w}.$$

This theorem is trickier to prove than the analogous result about conservative vector fields, so we will omit a proof: consult a serious vector calculus text if you are interested in seeing one. The basic idea of the proof is the same: compute  $\mathbf{v}$  as some kind of an integral of  $\mathbf{w}$ , but the integral is not just a line integral in this case!

As mentioned above, this result, too, fails if  $\mathbf{w}$  is not well-defined on all of  $\mathbb{R}^3$ . Indeed, the following example is especially significant in light of the computations which we will soon do in Newtonian gravity (Exercise 6.9 below).

*Exercise 6.6.* Let

$$\mathbf{w}(\mathbf{x}) = \frac{1}{|\mathbf{x}|}.$$

- (1) Show that  $\nabla \cdot \mathbf{w} = 0$  for  $\mathbf{x} \neq 0$ .
- (2) Compute

$$\iint_{S^2} \mathbf{N} \cdot \mathbf{w} \, dS$$

where  $S^2$  is the unit sphere in  $\mathbb{R}^3$  and  $\mathbf{N}$  is the outward-pointing unit normal.

- (3) Using Stokes's theorem, show that if there were to exist a vector field  $\mathbf{v}$  with  $\nabla \cdot \mathbf{v} = \mathbf{w}$  then we would have

$$\iint_{S^2} \mathbf{N} \cdot \mathbf{w} \, dS = 0,$$

contradicting your computation in the previous step.

It is quite helpful to summarize our vector calculus results in the following handy table of mappings, valid on  $\mathbb{R}^3$ :

$$\text{functions} \xrightarrow{\text{grad}} \text{vector fields} \xrightarrow{\text{curl}} \text{vector fields} \xrightarrow{\text{div}} \text{functions} \longrightarrow 0,$$

where for clarity we have used the notation grad, curl and div for the operations  $f \mapsto \nabla f$ ,  $\mathbf{v} \mapsto \nabla \times \mathbf{v}$ , and  $\mathbf{v} \mapsto \nabla \cdot \mathbf{v}$  respectively. The arrow at far right denotes the map from the space of all functions to the zero function. This is of course not so interesting, but as you will see it makes our diagram a little more informative.

The information contained in this very useful diagram is as follows: taking any two steps in a row along these maps gives zero. This reminds us that

$\text{curl} \circ \text{grad}$  and  $\text{div} \circ \text{curl}$  are both zero. To put it more stylishly, the range of each map lies in the nullspace of the next. Conversely, we have also found that any object that gets mapped to zero under one of these maps arises in the image of the map just to the left, i.e., it is in fact the case that *the range of each map equals the nullspace of the next*. So this reminds us that any vector field with zero curl is a gradient, and any vector field with zero divergence is a curl. It also gives us one more assertion, which is easier to check: reading the last pair of arrows, it asserts that *any function at all can be written as the divergence of a vector field*.

*Exercise 6.7.* Show that any function can be written as the divergence of a vector field. HINT: Let

$$v_1(\mathbf{x}) = \int_0^x f(s, x_2, x_3) ds$$

be an anti-derivative in the  $x_1$  variable, and try

$$\mathbf{v}(\mathbf{x}) = (v_1(\mathbf{x}), 0, 0).$$

*Exercise 6.8.* Show that if we add a zero at the beginning as well:

$$0 \longrightarrow \text{functions} \xrightarrow{\text{grad}} \text{vector fields} \xrightarrow{\text{curl}} \text{vector fields} \xrightarrow{\text{div}} \text{functions} \longrightarrow 0,$$

then the assertion that anything mapped to zero is in the range of the previous map is not quite true at the left end of the diagram. What, in fact, are all the functions that have zero gradient?

The fascinating thing about this diagram is that instead of considering functions and vector fields on  $\mathbb{R}^3$ , we consider instead functions and vector fields that are defined only on some subset  $\Omega \subset \mathbb{R}^3$  then some of the assertions above break down. It is still true that moving twice to the right always gives zero. However in general, there *are* curl-free vector fields that are not gradients, and divergence-free vector fields that are not curls. Instead of considering these failures as a problem, we should glory in them, as they turn out to give measurements of various aspects of how  $\Omega$  is shaped—in particular, of its *topology*. The refined study of this sequence of maps for more general regions in  $\mathbb{R}^3$  and (suitably generalized) in any number of dimensions, is the subject of *de Rham cohomology*, a very powerful tool in the topologist's kit.

Finally, we end this section with some discussion of what happens in higher dimensions. In  $\mathbb{R}^2$  and  $\mathbb{R}^3$ , we found that a vector field is conservative if and only if it has vanishing curl. Computing the components of this, say in  $\mathbb{R}^3$ , we see that the requirement is that

$$\frac{\partial v_1}{\partial x_2} = \frac{\partial v_2}{\partial x_1}, \quad \frac{\partial v_3}{\partial x_2} = \frac{\partial v_2}{\partial x_3}, \quad \frac{\partial v_1}{\partial x_3} = \frac{\partial v_3}{\partial x_1}.$$

Note that this list exhausts all possibilities of pairs  $i, j$  and just says that  $\partial v_i / \partial x_j = \partial v_j / \partial x_i$  for each  $i, j$ . (The quantity of course vanishes automatically if  $i = j$ .)

More generally, we easily see that if  $\mathbf{v} = \nabla f$  in  $\mathbb{R}^n$  then we have

$$v_i = \frac{\partial f}{\partial x_i}.$$

Applying  $\partial/\partial x_j$  to both sides reveals that

$$\frac{\partial v_i}{\partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}.$$

Switching the roles of  $i$  and  $j$  reveals also that

$$\frac{\partial v_j}{\partial x_i} = \frac{\partial^2 f}{\partial x_i \partial x_j}.$$

Now we recall from vector calculus that for any  $f \in \mathcal{C}^2(\mathbb{R}^n)$  we have

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

hence it must be the case that for each  $i, j$ ,

$$\frac{\partial v_i}{\partial x_j} = \frac{\partial v_j}{\partial x_i},$$

just as in  $\mathbb{R}^3$ . It turns out that just as in  $\mathbb{R}^3$ , this condition is necessary as well as sufficient, so we again have a characterization of conservative vector fields:

**Theorem 6.11.** *Let  $\mathbf{v} \in \mathcal{C}^1(\mathbb{R}^n)$ . Then  $\mathbf{v}$  is conservative if and only if*

$$\frac{\partial v_i}{\partial x_j} = \frac{\partial v_j}{\partial x_i}$$

for each  $i, j = 1, \dots, n$ .

The proof of this theorem is beyond the scope of these notes.

### 6.1. Conservative Potentials.

Now we return to physics. Say we are given a force field  $\mathbf{F}(\mathbf{x})$  on  $\mathbb{R}^n$  that we know (perhaps by computing its curl, if we are in  $\mathbb{R}^3$ ) to be conservative. Then we may write

$$\mathbf{F} = -\nabla V$$

for some  $V \in \mathcal{C}^\infty(\mathbb{R}^3)$ . Again, Newton's Second Law reads

$$\ddot{\mathbf{x}} = -\nabla V.$$

Taking the dot product of both sides by  $\dot{\mathbf{x}}$  yields

$$\ddot{\mathbf{x}} \cdot \dot{\mathbf{x}} = -\nabla V \cdot \dot{\mathbf{x}},$$

and we may employ the chain rule to recognize the left side as  $(d/dt)(\dot{\mathbf{x}})^2/2$  and the right as  $-(d/dt)V(\mathbf{x})$ . Rearranging thus gives

$$\frac{d}{dt} \left( \frac{1}{2} \dot{\mathbf{x}}^2 + V(\mathbf{x}) \right) = 0$$

hence

$$E \equiv \frac{1}{2} \dot{\mathbf{x}}^2 + V(\mathbf{x})$$

is *conserved*, just as in one dimension.

For example, if we take a three-dimensional harmonic oscillator with

$$\mathbf{F}(\mathbf{x}) = \begin{pmatrix} -k_1 x_1 \\ -k_2 x_2 \\ -k_3 x_3 \end{pmatrix}$$

we can easily find the potential  $V$  for which  $\mathbf{F} = -\nabla V$  just by solving three very simple PDEs: we want

$$\frac{\partial V}{\partial x_1} = k_1 x_1, \quad \frac{\partial V}{\partial x_2} = k_2 x_2, \quad \frac{\partial V}{\partial x_3} = k_3 x_3.$$

We can simply integrate these equations: integrating the first by taking  $\int_0^{x_1} \bullet(s, x_2, x_3) ds$  of both sides shows that

$$V(x_1, x_2, x_3) = \frac{1}{2} k_1 x_1^2 + C_1(x_2, x_3)$$

where the only subtlety is that  $C_1(x_2, x_3)$ , the “constant of integration,” depends on  $x_2, x_3$ . Then iterating the second equation shows that  $C_1(x_2, x_3) = \frac{1}{2} k_2 x_2^2 + C_2(x_3)$  and finally iterating the third tells us the  $C_2(x_3) = \frac{1}{2} k_3 x_3^2 + C_3$  for  $C_3$  now really a constant. Choosing  $C_3 = 0$  gives a solution

$$V(\mathbf{x}) = \frac{1}{2} (k_1 x_1^2 + k_2 x_2^2 + k_3 x_3^2),$$

hence

$$E = \frac{1}{2} \dot{\mathbf{x}}^2 + \frac{1}{2} (k_1 x_1^2 + k_2 x_2^2 + k_3 x_3^2)$$

is preserved under time-evolution of the system.

A somewhat more interesting example is that of Newtonian gravity, which we leave as a (very important) exercise:

*Exercise 6.9.* Check that the force law for Newtonian gravity,

$$\mathbf{F}(\mathbf{x}) = -\frac{1}{|\mathbf{x}|^2} \hat{\mathbf{x}} \equiv \mathbf{F}$$

is conservative. Do this first by computing its curl, and then by explicitly checking that

$$\mathbf{F} = -\nabla V$$

with

$$V(\mathbf{x}) = -\frac{1}{|\mathbf{x}|}$$

You may find the computations less onerous if you first check by computing in components that

$$\nabla |\mathbf{x}| = \frac{\mathbf{x}}{|\mathbf{x}|}$$

and then employ the chain rule.

Our knowledge of energy conservation for potentials like these in higher dimensions is a very powerful tool, but unlike the one dimensional case, it does not solve the problem for us. For motion in  $\mathbb{R}^3$ , for instance, the analog of the phase plane involves 6 variables: three of position and three of velocity. The fact that the energy is conserved gives us one equation satisfied by  $x_1, x_2, x_3, v_1, v_2, v_3$ . We can solve this equation for, say  $v_1$  in terms of the other 5 variables, but this still leaves 5 free variables in which the motion occurs. In the 1 dimensional case, we were in a plane, and the set satisfying energy conservation was a curve, hence we knew the particle's exact trajectory in the phase plane. But in higher dimensions we are still a long way from solving for the motion: we merely know it lies in a five-dimensional "submanifold" of  $\mathbb{R}^6$ !

**6.2. Consequences: bound and unbound motion.** One interesting way that we can use energy conservation is to see, in examples like that of Newtonian gravity, whether a particle's trajectory is trapped in some bounded set by energy considerations, or whether it has a chance at heading off to infinity.<sup>13</sup> Remember that for Newtonian gravity, the conserved energy is

$$E = \frac{1}{2}|\dot{\mathbf{x}}|^2 - \frac{1}{|\mathbf{x}|}.$$

If, for example, we know that at  $t = 0$ ,  $\mathbf{x} = (1, 0, 0)$  and  $\dot{\mathbf{x}} = (1/2, 1/2, 0)$ , then we easily compute that  $E = -3/4$ . On the other hand, we certainly know that now matter the position and velocity of our particle,

$$E \geq -\frac{1}{|\mathbf{x}|},$$

hence along our trajectory we must always have

$$-3/4 \geq -\frac{1}{|\mathbf{x}|},$$

i.e.,  $|\mathbf{x}| \leq 4/3$ . This limits the range of the motion. By contrast, if we choose initial conditions for which  $E \geq 0$ , then the energy does not form any barrier having the particle escape to infinity. Indeed, it is not hard to check that if  $E > 0$  and if  $\dot{\mathbf{x}}$  and  $\mathbf{x}$  are positive scalar multiples of one another, then the particle does indeed escape to infinity (i.e.  $|\mathbf{x}|$  is unbounded as  $t \rightarrow \infty$ .)

*Exercise 6.10.* Verify that for any  $E > 0$  there are trajectories with energy  $E$  with  $\dot{\mathbf{x}}$  equal to a scalar times  $\mathbf{x}$  that escape to infinity as  $t \rightarrow \infty$ . Do it by noting that without loss of generality, this means that  $\mathbf{x} = (x_1, 0, 0)$  and  $\dot{\mathbf{x}} = (\dot{x}_1, 0, 0)$ . This reduces the problem to a one-dimensional problem for  $x_1$  with conserved energy  $E = (1/2)\dot{x}_1^2 - 1/x_1$ ; solve it, as well as graphing the level sets of this energy in the phase plane for  $x_1$ .

<sup>13</sup>This is analogous to what we explored with the real pendulum, but in that case, our version of "going to infinity" was merely that the angle kept increasing as  $t \rightarrow \infty$ , which simply meant that the pendulum kept going round and round instead of oscillating back and forth.

We will study issue further in §10, when we address the Kepler problem in more detail.

**6.3. Example: Lissajous curves.** To give you a small taste of the rich phenomena that occur in higher dimensional motion, we now look at one of the simplest possible examples: a two-dimensional oscillator with force

$$\mathbf{F} = \begin{pmatrix} -\omega_1^2 x_1 \\ -\omega_2^2 x_2 \end{pmatrix}.$$

As discussed above, this force field is associated to the potential

$$V(x_1, x_2) = \frac{1}{2}(\omega_1^2 x_1^2 + \omega_2^2 x_2^2).$$

We can solve the equations of motion explicitly in this example, as they are uncoupled from one another and become simply two second order ODEs in one dimension:

$$\ddot{x}_1 = -\omega_1^2 x_1, \quad \ddot{x}_2 = -\omega_2^2 x_2.$$

We recognize these as harmonic oscillator equations, and hence can solve:

$$(25) \quad x_1 = A \cos \omega_1 t + B \sin \omega_1 t, \quad x_2 = C \cos \omega_2 t + D \sin \omega_2 t.$$

Even these very simple trajectories in  $\mathbb{R}^2$  are quite beautiful and intriguing. In studying them, we first note that there is a big difference between the cases when the “frequencies” have a rational relationship:

$$\omega_1/\omega_2 = \frac{p}{q} \in \mathbb{Q},$$

and when the ratio is irrational. In the former case, we note that if we set  $\tau = 2\pi q/\omega_2$ , then both  $\omega_1\tau$  and  $\omega_2\tau$  are integer multiples of  $2\pi$ . Hence, for all  $A, B, C, D$  and every  $t \in \mathbb{R}$ ,

$$\begin{aligned} \mathbf{x}(t + \tau) &= \begin{pmatrix} A \cos \omega_1(t + \tau) + B \sin \omega_1(t + \tau) \\ C \cos \omega_2(t + \tau) + D \sin \omega_2(t + \tau) \end{pmatrix} \\ &= \begin{pmatrix} A \cos \omega_1 t + B \sin \omega_1 t \\ C \cos \omega_2 t + D \sin \omega_2 t. \end{pmatrix} \end{aligned}$$

i.e., the motion is *periodic* with period  $\tau$ . By contrast, if  $\omega_1/\omega_2 \notin \mathbb{Q}$ , we claim that the motion never repeats itself. A proof of this is the content of the following exercise.

*Exercise 6.11.* Show that if  $\omega_1/\omega_2 \notin \mathbb{Q}$ , there does not exist any  $\tau$  such that

$$(26) \quad \mathbf{x}(t + \tau) = \mathbf{x}(t)$$

for every  $A, B, C, D$  and  $t$  HINT: Try taking particular values for  $A, B, C, D$  to show first that  $\omega_1\tau \in 2\pi\mathbb{Z}$  and then  $\omega_2\tau \in 2\pi\mathbb{Z}$ .

Note, however, that certain choices of  $A, B, C, D$  will give rise to periodic trajectories. What are they?

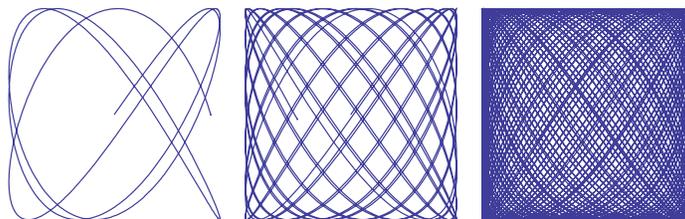


FIGURE 4. Plot of  $(\sin t, \sin \sqrt{2}t)$  for  $t$  running from 0 to 20, 100, and 500.

Note that we can always write

$$(27) \quad \begin{aligned} x_1(t) &= A \cos \omega_1 t + B \sin \omega_1 t = R_1 \sin \omega_1(t + \delta_1) \\ x_2(t) &= C \cos \omega_2 t + D \sin \omega_2 t = R_2 \sin \omega_2(t + \delta_2) \end{aligned}$$

(see exercise below). This simplifies our description of the motion somewhat, and enables us to see that we will always have

$$x_1 \in [-R_1, R_1], \quad x_2 \in [-R_2, R_2].$$

*Exercise 6.12.* Show for any choice of  $A, B$ , we can always write  $A \cos x + B \sin x = R \sin(x + \delta)$  for some  $R > 0$  and  $\delta \in [0, 2\pi)$ . HINT: Expand the right side using the addition formula for sin and then try to solve for  $R, \delta$  in terms of  $A, B$ .

*Exercise 6.13.* In this exercise we think a bit harder about why a Lissajous curve should be contained in a rectangle.

- (1) Show that energy conservation implies that the trajectory  $(x_1(t), x_2(t))$  is contained in the interior of an ellipse in  $\mathbb{R}^2$ .
- (2) Show that in the special case of the harmonic oscillator, the two separate energies in each coordinate

$$E_i = \frac{\dot{x}_i^2}{2} + \omega_i^2 \frac{x_i^2}{2}$$

are conserved (and  $E_1 + E_2 = E$ ).

- (3) Show that the trajectory stays in the *rectangle*

$$\left\{ |x_1| < \frac{\sqrt{2E_1}}{\omega_1}, \quad |x_2| < \frac{\sqrt{2E_2}}{\omega_2} \right\}$$

- (4) Write the energies  $E_i$  in terms of the coefficients  $R_i$  in (27).
- (5) What is the relationship of this rectangle to the ellipse from the first part? Does one fit inside the other?

It turns out that in the case  $\omega_1/\omega_2 \notin \mathbb{Q}$ , most of the trajectories in fact are *dense* in the square  $[-R_1, R_1] \times [-R_2, R_2]$ . For a trajectory to be dense means that it passes through any open set.

**Theorem 6.12.** *Assume that  $\omega_1/\omega_2 \notin \mathbb{Q}$ . Then the trajectory (27) is dense in  $[-R_1, R_1] \times [-R_2, R_2]$ .*

In the proof, we will use a convenient notation for the “fractional part” of a number: if we have  $x = n + \alpha$  with  $n$  an integer and  $\alpha \in [0, 1)$  then we write

$$[x] = \alpha.$$

Note, crucially, that periodicity of  $\sin$  means that for  $s \in [0, 2\pi)$  and  $k \in \mathbb{Z}$ ,  $\sin(s + 2\pi k) = \sin s$ , hence writing  $t = s + 2\pi$  gives

$$\sin t = \sin(2\pi[t/2\pi]).$$

*Proof.* We sketch the proof, leaving some of the details to the student. (Some knowledge of real analysis is helpful here.)

Let

$$F : [0, 1] \times [0, 1] \rightarrow [-R_1, R_1] \times [-R_2, R_2]$$

be given by

$$F \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} R_1 \sin(2\pi y_1) \\ R_2 \sin(2\pi y_2) \end{pmatrix}.$$

Then the orbit in question is just

$$F \begin{pmatrix} [\omega_1(t + \delta_1)/(2\pi)] \\ [\omega_2(t + \delta_2)/(2\pi)] \end{pmatrix}$$

in order to show that the orbit gets arbitrarily close to every point in the range of  $F$ , it suffices to show that the input to  $F$  gets as close as you like to any point in the square  $[0, 1] \times [0, 1]$ . In particular, it suffices for each  $N$  to chop the square into smaller squares of sidelength  $1/N$ , and to show that the input gets into every one of those subsquares, i.e., we now want to show that for every  $N \in \mathbb{N}$ , and for every  $j, k \in \{-N, \dots, N - 1\}$ , there exists  $t$  such that

$$[\omega_1(t + \delta_1)/(2\pi)] \in [j/N, (j + 1)/N], \quad [\omega_2(t + \delta_2)/(2\pi)] \in [k/N, (k + 1)/N].$$

This is true if and only if it is also true with  $\delta_1 = \delta_2 = 0$ , since if  $\omega_1(t + \delta_1)/(2\pi)$  lands in every such interval, so does  $\omega_1 t/(2\pi)$  and vice-versa. So we deal only with the case  $\delta_i = 0$ , i.e. we now want for every  $N, j, k$

$$[\omega_1 t/(2\pi)] \in [j/N, (j + 1)/N], \quad [\omega_2 t/(2\pi)] \in [k/N, (k + 1)/N].$$

for some  $t$ .

Now choose  $t_0$  such that, say  $[\omega_2 t_0/(2\pi)] = (k + 1/2)/N$ . Then consider the sequence

$$t_\ell \equiv t_0 + 2\pi\ell/\omega_2, \quad \ell \in \mathbb{N}$$

of further times at which this also occurs. Then we examine

$$[\omega_1 t_\ell/(2\pi)] = [\omega_1 t_0/(2\pi) + \ell(\omega_1/\omega_2)].$$

Since  $\omega_1/\omega_2 \notin \mathbb{Q}$ , Lemma 6.13 (to be proved below), tells us that this sequence is dense, hence we may simultaneously achieve both the precise equation

$$[\omega_2 t_\ell / (2\pi)] = (k + 1/2)/N$$

as well as

$$[\omega_1 t_\ell / (2\pi)] \in [j/N, (j + 1)/N].$$

□

The following lemma was the crucial ingredient in our density proof.

**Lemma 6.13.** *Let  $\omega \notin \mathbb{Q}$ . Then the numbers  $[n\omega]$ ,  $n \in \mathbb{N}$ , are dense in  $[0, 1]$ .*

*Proof.* Subdivide  $[0, 1]$  into  $[j/N, (j + 1)/N]$ . It suffices to show that some  $[n\omega]$ ,  $n \in \mathbb{N}$  lies in each of these intervals. To see this, note that there are  $N$  such intervals, so if we examine the  $N + 1$  numbers

$$[\omega], [2\omega], [3\omega], \dots, [(N + 1)\omega]$$

then by the “pigeonhole principle”<sup>14</sup> two of these numbers  $[m_1\omega]$  and  $[m_2\omega]$  must lie in some interval  $[j/N, (j + 1)/N]$ , i.e. we must have

$$m_1\omega - \ell_1, m_2\omega - \ell_2 \in [j/N, (j + 1)/N].$$

for  $\ell_1, \ell_2 \in \mathbb{Z}$ . Let us say  $m_2 > m_1$ . Then taking the difference of these numbers we find

$$(m_2 - m_1)\omega + (\ell_1 - \ell_2) = \rho \in [0, 1/N]$$

Now consider the numbers

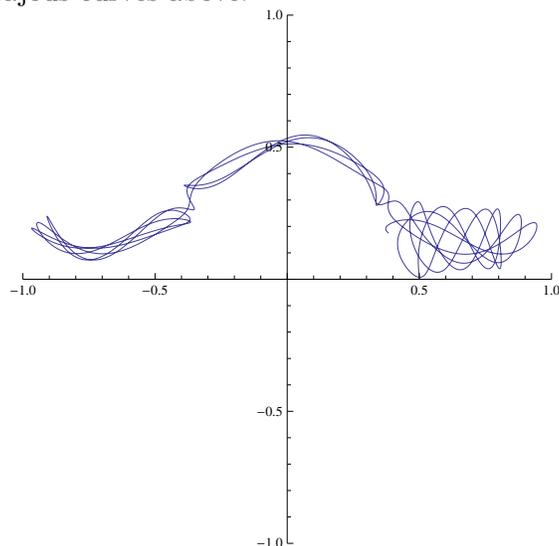
$$k\rho = i(m_2 - m_1)\omega + k(\ell_1 - \ell_2) \quad k \in \mathbb{N}.$$

The difference between successive numbers in the sequence is less than  $1/N$ , hence the fractional parts must land in *every* interval  $[j/N, (j + 1)/N]$ . This solves our problem, since  $[k\rho] = [k(m_2 - m_1)\omega]$ . □

Our explorations of Lissajous curves illustrate that the motion of particles in a multi-dimensional oscillator potential is a rich and interesting subject. It turns out that this is just about the simplest and most analyzable potential we can write down: as we will see in our later exploration of integrable systems, the multi-dimensional oscillator is the model for this class of problems that we can consider explicitly solvable by integration. How much more intricate, then, must be the motion in a more general two-dimensional potential! Figure 6.3 shows the trajectory of a particle in  $\mathbb{R}^2$  moving in the potential  $V(x_1, x_2) = x_1^2 + x_2^2 + \cos(4x_1)\sin(10x_2) + \sin(x_1/10)\sin(x_2/3)$ . You are primarily supposed to notice merely that the resulting curve is complicated and difficult to describe!

<sup>14</sup>If we have  $N + 1$  pigeons living in  $N$  pigeonholes, then at least one pigeonhole must hold at least two pigeons. This is simple, yet surprisingly powerful idea.

FIGURE 5. A two dimensional non-integrable system. The potential is  $x_1^2 + x_2^2 + \cos(4x_1) \sin(10x_2) + \sin(x_1/10) \sin(x_2/3)$ , and the trajectory is plotted in the  $x_1, x_2$  plane, just as with the Lissajous curves above.



## 7. THE PRINCIPLE OF LEAST ACTION

**7.1. Introduction: why potentials?** As we have seen in the examples we have dealt with so far, when dealing with motion in a conservative force field, it can be considerably simpler to think about the *potential*  $V(\mathbf{x})$  rather than directly about the force  $\mathbf{F} = -\nabla V$ . These advantages will become even more pronounced when we deal with motion with *constraints* in a later section. A simple example is that of the real pendulum, i.e., a mass held at the end of a rigid rod in  $\mathbb{R}^2$  and acted upon by the force of gravity  $\mathbf{F}_g = (0, -1)$ . What we mean when we say that the particle is attached to a rigid rod is that there is *another* force  $\mathbf{F}_r$  exerted on the particle by the rod at all times sufficient to keep the particle on the circle  $x_1^2 + x_2^2 = 1$ ; moreover, the rod can only exert this force in the radial direction (i.e., along the rod itself) i.e.  $\mathbf{F}_r$  is a scalar multiple of the position vector of the particle. From this information, we can draw a “force diagram” and figure out what  $\mathbf{F}_r$  must be, and thus solve the motion. One learns to do this in a real physics class. But how much easier it is to simply cite the principle that the potential  $V$  associated to the system is the same as that of the particle in  $\mathbb{R}^2$  acting by gravity, while the kinetic energy is simply  $(1/2)$  times the squared norm of the velocity, as always!

The analysis of systems like the real pendulum, in which we have constraints on the motion, is one example of a situation where energy is easier to deal with than force. We therefore now turn to a way of formulating the

basic principles of mechanics that bypasses Newton's second law, and more generally, the whole notion of force, in favor of energy.

**7.2. The principle of least action.** For now, let us consider a particle or particles moving in  $\mathbb{R}^n$ ; we have seen that the quantity

$$T \equiv \frac{1}{2} |\dot{\mathbf{x}}|^2,$$

the “kinetic energy,” arose in the equations for energy conservation. Let us suppose that we are also given a second function,  $V(\mathbf{x})$ , the potential for the motion. We of course know already how to write down the equations of motion for this system, but we turn to a new way of arriving at those equations. Consider the quantity

$$L(\mathbf{x}, \dot{\mathbf{x}}) = T - V = \frac{1}{2} |\dot{\mathbf{x}}|^2 - V(\mathbf{x}).$$

$L$  is called the *Lagrangian* of the system. Note that it differs from the energy in a crucial way: it is the *difference* of kinetic and potential energy rather than the sum.

Now say we send a particle along a trajectory  $\gamma(t)$  between time  $t = a$  and  $t = b$ . This need not be the physical trajectory of the particle moving in the potential: if  $V$  is some potential created in our laboratory apparatus, we may move the particle along  $\gamma(t)$  by sticking our hand into the apparatus and forcing it along in some very artificial way. Associated to this trajectory is a quantity called the *action*, denoted  $S$ :

$$S(\gamma(t)) = \int_a^b L(\gamma(t), \dot{\gamma}(t)) dt.$$

For example for the one-dimensional harmonic oscillator,

$$L = \frac{1}{2} \dot{x}^2 - \frac{1}{2} x^2$$

and if  $\gamma(t)$  is the path

$$t \mapsto t + t^2, \quad t \in [0, 1]$$

then

$$L(\gamma(t), \dot{\gamma}(t)) = \frac{1}{2}(1 + 2t)^2 - \frac{1}{2}(t + t^2)^2$$

and

$$S(\gamma(t)) = \int_0^1 \frac{1}{2}(1 + 2t)^2 - \frac{1}{2}(t + t^2)^2 dt,$$

whatever that might be.

You should think of  $S$  as a machine that eats a path  $\gamma(t)$  and spits out a number associated to it. It's a kind of function, but on a huge space: the space of all paths. To remind ourselves that this is a function on such a big space, we call such an object a *functional*.

Now we are ready to reformulate classical mechanics. This formulation is called the “Principle of Least Action” but you will see that this is really

a misnomer. We will spend quite some time explaining exactly what this definition means, so don't panic.

*Principle of Least Action.* The trajectory followed by the particle moving in the system with action

$$S(\gamma) = \int L(\gamma, \dot{\gamma}) dt$$

is one at which the action functional  $S$  is *stationary* with respect to variations preserving the endpoints.

The principle should of course really be called the *principle of stationary action* but that does not sound quite as elegant. The distinction being made is the same as we make in calculus between a *minimum* of a function and a *stationary point* where the derivative vanishes, which might be a (local) minimum, or maximum, or neither.

*Exercise 7.1.* Consider the solution to the one-dimensional harmonic oscillator

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

with  $x(0) = 1$  and  $x(\pi) = -1$  given by  $x(t) = \cos t$ . For this trajectory, compute

$$\int_0^\pi T - V dt$$

where  $T$  is kinetic energy and  $V$  is potential energy. Compute the same quantity for the (nonphysical) trajectory  $x(t) = 1 - 2t/\pi$  which has the same location at  $t = 0$  and  $t = \pi$ . Compare the two answers you get and explain.

**7.3. The calculus of variations.** The calculus of variations is the generalization of multi-variable calculus that deals with functionals like the  $S(\gamma)$  that we have defined, rather than functions of finitely many variables. It is applicable to a wide variety of problems in addition to those that are directly relevant to classical mechanics, so as long as we are at it, we will develop it in a little more generality. Thus, we will not make any assumptions about the form of  $S(\gamma)$  for now, but rather we will assume merely that it is a functional returning a number for each path  $\gamma(t)$  from  $\mathbf{x}_0$  to  $\mathbf{x}_1$  ( $t \in [0, T]$ ).

Let us now explain what is meant by being stationary with respect to variations preserving the endpoints. Remember that we think of  $S$  as eating a path and spitting out a number. We will restrict our attention to paths that start at  $t = 0$  (say) at some particular point  $\mathbf{x}_0$  and end at time  $t = T$  at some other point  $\mathbf{x}_1$ . It is among these paths that  $S$  is supposed to be stationary.

*Definition 7.1.* Let  $\mathcal{X}$  denote the space of all  $C^\infty$  paths  $\gamma(t)$  with

$$\gamma(0) = \mathbf{x}_0, \quad \gamma(T) = \mathbf{x}_1.$$

Now what do we mean by stationary? We think back to multi-variable calculus to understand this better. Remember that at a local extremum of a function  $f(\mathbf{y})$  on  $\mathbb{R}^N$ , we have the necessary condition  $\nabla f(\mathbf{y}) = 0$ . We are going to try to replace the function  $f$  by the functional  $S$  and find an analog of this condition, but the gradient of a functional on paths is a hard notion to swallow. So notice that we can reformulate the stationarity condition

$$\nabla f(\mathbf{y}) = 0$$

as follows: let  $\mathbf{v} \in \mathbb{R}^N$  be any vector. We compute by the chain rule

$$(28) \quad \left. \frac{d}{ds} \right|_{s=0} f(\mathbf{y} + s\mathbf{v}) = \nabla f(\mathbf{y}) \cdot \mathbf{v}.$$

Thus

$$\nabla f(\mathbf{y}) = 0$$

is equivalent to the condition

$$\left. \frac{d}{ds} \right|_{s=0} f(\mathbf{y} + s\mathbf{v}) = 0 \text{ for all } \mathbf{v} \in \mathbb{R}^N.$$

In words, this is just saying that the gradient is zero if and only if all the *directional derivatives* are zero: the left-hand side of (28) is nothing but the directional derivative in direction  $\mathbf{v}$ .

Now this directional derivative notion *is* one that we can generalize to our functional  $S$ .  $S$  acts on the space  $\mathcal{X}$  of all paths with fixed values at  $t = 0, T$ . How do we move a little in this space? Any two such paths  $\gamma_0, \gamma_1$  have the property that if  $\varphi = \gamma_1 - \gamma_0$  then

$$(29) \quad \varphi(0) = \varphi(T) = 0.$$

Conversely, given a path  $\gamma \in \mathcal{X}$  if  $\varphi$  satisfies (29) then  $\gamma + \varphi \in \mathcal{X}$ . Indeed, the same is true of

$$\gamma + s\varphi$$

for any  $s \in \mathbb{R}$ .

*Definition 7.2.* Let  $\mathcal{T}$  denote the space of all  $C^\infty$  paths  $\gamma(t)$  with

$$\gamma(0) = 0, \quad \gamma(T) = 0.$$

So here is what we do: we think of the  $\varphi \in \mathcal{T}$  as analogous to our vectors  $\mathbf{v}$  in which we can take directional derivatives.<sup>15</sup>

*Definition 7.3.* We say that the functional  $S$  is stationary at  $\gamma \in \mathcal{X}$  if

$$\left. \frac{d}{ds} \right|_{s=0} S(\gamma + s\varphi) = 0$$

for all  $\varphi \in \mathcal{T}$ .

---

<sup>15</sup>Technically, the space  $\mathcal{X}$  is an *affine space*.  $\mathcal{T}$  is its tangent space, and is a vector space (closed under addition and scalar multiplication).

What good is this definition? The analogous definition for functions on  $\mathbb{R}^N$  was useful in multivariable calculus because it was a useful test for trying to locate maxima and minima (“extrema”) of functions. It turns out the same is true here, as we now show:

Say we know that a functional  $S$  takes on a minimum at  $\gamma$  among all paths in  $\mathcal{X}$ , i.e. for all  $\gamma' \in \mathcal{X}$ ,

$$S(\gamma) \leq S(\gamma').$$

Then in particular, we certainly know that for all  $\varphi \in \mathcal{T}$ , and for all  $s \in \mathbb{R}$ ,

$$S(\gamma) \leq S(\gamma + s\varphi).$$

Since the two sides of this equation are equal when  $s = 0$ , the  $s$ -derivative of the right side must vanish when  $s = 0$ —otherwise, either for small enough  $s > 0$  (if the derivative is negative) or for  $s < 0$  (if the derivative is positive) the right side would be smaller. Thus we must have

$$\left. \frac{d}{ds} S(\gamma + s\varphi) \right|_{s=0} = 0,$$

hence we have established that stationarity of  $S$  plays the same role as it does in multivariable calculus, as a necessary condition for an extremum:

**Proposition 7.4.** *A necessary condition for a functional  $S$  to have an extremum at  $\gamma \in \mathcal{X}$  is that it must be stationary at  $\gamma$ .*

The next natural question is: given  $S$ , how might we find an extremum in practice? At this point we do need to assume *something* about the form of  $S(\gamma)$ : we will assume that it is given by an integral of some function of  $\gamma$ ,  $\dot{\gamma}$ , and  $t$ , i.e. we assume that there exists  $L(\mathbf{x}, \mathbf{v}, t)$  such that

$$(30) \quad S(\gamma) = \int_0^T L(\gamma(t), \dot{\gamma}(t), t) dt.$$

This certainly includes the examples we are interested in for Newtonian mechanics, where  $L$  has the special form

$$L(\mathbf{x}, \mathbf{v}, t) = \frac{1}{2} |\mathbf{v}|^2 + V(\mathbf{x})$$

and (at least in the examples we have considered) has no explicit  $t$ -dependence. We have named the dummy variables in  $\mathbb{R}^n$  on which  $L$  depends  $\mathbf{x}$  and  $\mathbf{v}$  so as not to get confused with the specific inputs  $\gamma(t)$  and  $\dot{\gamma}(t)$  that occur when we evaluate the integral; this will reduce confusion a little later when we need to consider partial derivatives of  $L$ . All in all,  $L$  is a function of  $2n + 1$  variables.

Now let us try and compute the stationary points of  $S$  of the form (30). The condition to be stationary at  $\gamma \in \mathcal{X}$  is that for all  $\varphi \in \mathcal{T}$ , we have

$$\left. \frac{d}{ds} S(\gamma + s\varphi) \right|_{s=0} = 0,$$

i.e., that

$$\frac{d}{ds}\Big|_{s=0} \int_0^T L(\gamma(t) + s\varphi(t), \dot{\gamma}(t) + s\dot{\varphi}(t), t) dt = 0.$$

This is easy enough to compute; since  $s$  is a parameter in the integral, and every function involved is assumed to be infinitely differentiable, we may simply bring the  $s$  derivative under the integral sign: we need to compute

$$\int_0^T \frac{d}{ds}\Big|_{s=0} L(\gamma(t) + s\varphi(t), \dot{\gamma}(t) + s\dot{\varphi}(t), t) dt = 0.$$

Now by the chain rule,

$$\frac{d}{ds}\Big|_{s=0} L(\gamma(t) + s\varphi(t), \dot{\gamma}(t) + s\dot{\varphi}(t), t) = \sum_{j=1}^n \frac{\partial L}{\partial x_j}\Big|_{(\gamma(t), \dot{\gamma}(t), t)} \varphi_j(t) + \frac{\partial L}{\partial v_j}\Big|_{(\gamma(t), \dot{\gamma}(t), t)} \dot{\varphi}_j(t);$$

here the  $x_j$  partial derivatives of  $L$  refer to derivatives in the first  $n$  “slots,” and the  $v_j$  in the next  $n$  slots, in keeping with our notation  $L = L(\mathbf{x}, \mathbf{v}, t)$ . Make sure you understand our use of the chain rule here: it is easy to get confused with so many functions involved.

To simplify notation, we will use subscripts to indicate partial derivatives, hence write

$$L_{x_j}(\gamma(t), \dot{\gamma}(t), t) \equiv \frac{\partial L}{\partial x_j}\Big|_{(\gamma(t), \dot{\gamma}(t), t)}, \quad L_{v_j}(\gamma(t), \dot{\gamma}(t), t) \equiv \frac{\partial L}{\partial v_j}\Big|_{(\gamma(t), \dot{\gamma}(t), t)}.$$

Then we have established

$$\frac{d}{ds}\Big|_{s=0} S(\gamma + s\varphi) = \sum \int_0^T L_{x_j}(\gamma(t), \dot{\gamma}(t), t) \varphi_j(t) + L_{v_j}(\gamma(t), \dot{\gamma}(t), t) \dot{\varphi}_j(t) dt.$$

For reasons that will be clear very shortly, we like the form of the first term of the integrand, in which  $\varphi_j$  appears with no derivatives in it, and we do not so much like the second, where  $\varphi_j$  appears with a time derivative. Fortunately, there is a very nice way of making the second term look more like the first: integration by parts. If we have two functions  $f(t)$  and  $g(t)$  and we know that  $g(0) = g(T) = 0$ , then integration by parts tells us that

$$\int_0^T f(t) \dot{g}(t) dt = - \int_0^T \dot{f}(t) g(t) dt.$$

Note that the vanishing of  $g$  at the endpoints is crucial in order that no boundary terms appear. Now let

$$f(t) = L_{v_j}(\gamma(t), \dot{\gamma}(t), t)$$

and

$$g(t) = \varphi_j(t).$$

Since  $\varphi \in \mathcal{T}$ , all of its components do indeed vanish at the endpoints. So we have

$$\int L_{v_j}(\gamma(t), \dot{\gamma}(t), t) \dot{\varphi}_j(t) dt = - \int_0^T \frac{d}{dt} [L_{v_j}(\gamma(t), \dot{\gamma}(t), t)] \varphi_j(t).$$

Note that if we want to evaluate

$$\frac{d}{dt}L_{v_j}(\gamma(t), \dot{\gamma}(t), t),$$

we must apply the chain rule somewhat carefully, as the derivative involves both the direct dependence on  $t$  through the dependence of  $L(\bullet, \bullet, t)$  on its third variable, as well as the dependence via  $\gamma(t), \dot{\gamma}(t)$  in the first  $2n$  arguments.<sup>16</sup> In any case, we now have

$$\begin{aligned} \frac{d}{ds}\Big|_{s=0}S(\gamma + s\varphi) &= \sum \int_0^T L_{x_j}(\gamma(t), \dot{\gamma}(t), t)\varphi_j(t) - \frac{d}{dt}L_{v_j}(\gamma(t), \dot{\gamma}(t), t)\varphi_j(t) dt \\ &= \sum \int_0^T \left( L_{x_j}(\gamma(t), \dot{\gamma}(t), t) - \frac{d}{dt}L_{v_j}(\gamma(t), \dot{\gamma}(t), t) \right) \varphi_j(t) dt. \end{aligned}$$

Remember: *this thing is supposed to be zero for every choice of  $\varphi \in \mathcal{T}$* . How can this be? If for any  $j = 1, \dots, n$ ,

$$L_{x_j}(\gamma(t), \dot{\gamma}(t), t) - \frac{d}{dt}L_{v_j}(\gamma(t), \dot{\gamma}(t), t)$$

is nonzero—say, positive—for some  $t = t_0$ , then I could choose  $\varphi_j$  to be a little bump function that's positive near  $t = t_0$  and zero otherwise; and choose all the rest of the components of  $\varphi$  to simply be zero. Then I would have achieved

$$\sum \int_0^T \left( L_{x_j}(\gamma(t), \dot{\gamma}(t), t) - \frac{d}{dt}L_{v_j}(\gamma(t), \dot{\gamma}(t), t) \right) \varphi_j(t) dt > 0.$$

In other word, the only way to have

$$\frac{d}{ds}\Big|_{s=0}S(\gamma + s\varphi) = 0$$

for all  $\varphi \in \mathcal{T}$  is if we have

$$(32) \quad \begin{aligned} L_{x_j}(\gamma(t), \dot{\gamma}(t), t) &= \frac{d}{dt}L_{v_j}(\gamma(t), \dot{\gamma}(t), t) \\ &\text{for all } t \in [0, T] \text{ and } j = 1, \dots, n. \end{aligned}$$

The system of equations (32) are important enough to have a name: they are called the *Euler-Lagrange* equations associated to the Lagrangian  $L$ . We have established:

<sup>16</sup>This is a little clearer if we write  $L = L(\mathbf{x}, \mathbf{v}, r)$ , calling the last dummy variable  $r$  instead of  $t$ . Then in its full glory,

$$(31) \quad \frac{d}{dt}L_{v_j}(\gamma(t), \dot{\gamma}(t), t) = L_r(\gamma(t), \dot{\gamma}(t), t) + \sum_k L_{v_j x_k}(\gamma(t), \dot{\gamma}(t), t)\dot{\gamma}_k(t) + L_{v_j v_k}(\gamma(t), \dot{\gamma}(t), t)\ddot{\gamma}_k(t).$$

We will following more usual notation, though, and write  $L_t$  instead of  $L_r$ , at a slight risk of confusion. Just remember that this notation means a partial derivative with respect to the last variable, and do not confuse it with the total derivative with respect to  $t$ ,  $(d/dt)L(\gamma(t), \dot{\gamma}(t), t)$  which contains other terms dictated by the chain rule.

**Theorem 7.5.** *The functional  $S(\gamma)$  given by (30) is stationary at  $\gamma \in \mathcal{X}$  if and only if the Euler-Lagrange equations (32) hold.*

Let us examine the equations (32) in our favorite specific example, the one-dimensional harmonic oscillator with

$$L = \frac{1}{2}v^2 - \frac{\omega^2}{2}x^2.$$

Then  $L_x = -\omega^2x$  and  $L_v = v$ , hence (32) reads:

$$\frac{d}{dt}\dot{\gamma} = -\omega^2\gamma,$$

i.e.,

$$\ddot{\gamma} = -\omega^2\gamma.$$

You might recognize this as simply Newton's law of motion for the one-dimensional oscillator.

This observation is indeed much more general: if we return to our Lagrangians associated to (time-independent) potentials

$$L = \frac{1}{2}|\mathbf{v}|^2 - V(\mathbf{x}),$$

then  $L_{v_j} = v_j$ ,  $L_{x_j} = -\partial V/\partial x_j$  and the Euler-Lagrange equations read

$$\frac{d}{dt}\dot{v}_j = -\frac{\partial V}{\partial x_j},$$

i.e.,

$$\ddot{\mathbf{x}} = -\nabla V \equiv \mathbf{F}.$$

So we always recover Newton's laws of motion. Thus, we have *derived* the principle of least/stationary action for motion in a potential: Newton's law of motion is *equivalent* to this principle. Henceforth, when we discuss more general physical systems, we will aim to work backwards: we will forget all about the notion of force, and try to begin by writing down a Lagrangian, and invoking the principle of least action.

**7.4. Other applications of the calculus of variations.** In this section, we explore some other uses of the calculus of variations besides simply reformulating Newton's law of motion for a particle in a potential.

**7.4.1. Straight lines.** What's the shortest distance between two points in the plane? Well, you know the answer to this one, but let's check our techniques out on this problem. Fix points  $\mathbf{y}^0$  and  $\mathbf{y}^1$  in  $\mathbb{R}^2$ . Consider all possible paths connecting them:  $\gamma$  with  $\gamma(0) = \mathbf{y}^0$ ,  $\gamma(1) = \mathbf{y}^1$ . Recall from calculus class that the *length* of the path is just

$$\int_0^1 |\dot{\gamma}(t)| dt.$$

Hence the Lagrangian here is just

$$L = |\mathbf{v}| = \sqrt{v_1^2 + v_2^2}.$$

Then the Euler–Lagrange equations read

$$0 = \frac{d}{dt} \frac{v_j}{\sqrt{v_1^2 + v_2^2}}, \quad j = 1, 2.$$

Thus we find that  $\frac{v_j}{\sqrt{v_1^2 + v_2^2}} = w_j$  are constant for  $j = 1, 2$ . In other words, the unit vector in the direction of the velocity

$$\hat{\mathbf{v}} \equiv \frac{1}{|\mathbf{v}|} \mathbf{v} \equiv \mathbf{w}$$

is constant. Interestingly, this doesn't tell us everything we might like to know about a trajectory: we have not completely solved for  $\mathbf{v}$ . But on the other hand, it does imply, by integrating, that

$$\begin{aligned} \mathbf{x}(t) &= \int_0^t \mathbf{v}(t) dt + \mathbf{x}(0) \\ &= \int_0^t |\mathbf{v}(t)| \mathbf{w} dt + \mathbf{y}^0. \end{aligned}$$

Thus  $\mathbf{x}(t) - \mathbf{y}^0$  is an integral of a variable multiple of the constant vector  $\mathbf{w}$ , hence lies in the line through the origin in direction  $\mathbf{w}$ . Hence  $\mathbf{x}$  itself lies in the line in direction  $\mathbf{w}$  through the point  $\mathbf{y}^0$ . Since we have asked that  $\mathbf{x}(0) = \mathbf{y}^1$ , the point  $\mathbf{y}^1$  is also on this line, hence in fact the unit vector  $\mathbf{w}$  must be in the direction  $\mathbf{y}^1 - \mathbf{y}^0$ , i.e.,

$$\mathbf{w} = \frac{1}{|\mathbf{y}^1 - \mathbf{y}^0|} (\mathbf{y}^1 - \mathbf{y}^0).$$

This shows that the motion must lie on the straight line connecting  $\mathbf{y}^0$  to  $\mathbf{y}^1$ .

You might think what we've achieved is unsatisfactory, as we still haven't actually found  $\mathbf{x}$ ! But note that there are in fact *many* solutions to this problem. This has to do with the fact that the length of the trajectory is independent of how the trajectory is parametrized: we could move along this straight line initially very fast, and then slow down, or go slow then fast, or move at constant speed, and the length that we traverse is of course always the same. So the ambiguity in our answer—all we know is that the *direction* stays constant—in fact reflects a real-life ambiguity. These kind of ambiguities don't show up in the examples we consider from physics, where the action does not have this indeterminacy to it, and trajectories satisfying the Euler–Lagrange equations (with initial conditions specified) are unique.

**7.4.2. Minimal surface of rotation.** Consider the surface generated by rotating the curve  $x_3 = f(x_1)$  between  $x_1 = -1$  and  $x_1 = 1$  about the  $x_1$ -axis. This yields a surface whose area, as we learn in calculus class, is

$$\int_{-1}^1 2\pi f(s) \sqrt{1 + f'(s)^2} ds$$

We now pose the question: what is the surface of this type having least area, subject to the constraints  $f(\pm 1) = R$ , i.e. having as boundary the two circles of radius  $R$  perpendicular to the  $x_1$  axis and centered at  $(\pm 1, 0, 0)$ ? We will not be able to address the question of minimality rigorously, but we note that if a minimal such surface exists, the corresponding function  $f$  must be a stationary point for the action integral

$$S[f] = \int_{-1}^1 2\pi f(s) \sqrt{1 + f'(s)^2} ds \equiv \int_{-1}^1 L(f(s), f'(s), s) ds$$

within the space of functions  $f$  having the value  $R$  at  $s = \pm 1$ . Thus, by Theorem 7.5,  $f$  must satisfy the Euler-Lagrange equations for the Lagrangian

$$L(f, f', s) = 2\pi f \sqrt{1 + (f')^2}.$$

The Euler-Lagrange equations are thus

$$(33) \quad (1 + f'(s)^2)^{1/2} = \frac{d}{ds} \left( \frac{f(s)f'(s)}{(1 + f'(s)^2)^{1/2}} \right).$$

This is somewhat dreadful looking.

**7.4.3. Lorentz force.** In the following extended exercise, we show that the Lagrangian formulation of mechanics can be applied to the example of Lorentz force

$$\mathbf{F} = q\mathbf{v} \times \mathbf{B}$$

exerted by a magnetic field  $\mathbf{B}$  on a charged particle with charge  $q$ , even though this example is not motion in a potential.

*Exercise 7.2.* We begin by noting that in electromagnetism, we always have

$$\nabla \cdot \mathbf{B} = 0$$

(“no magnetic monopoles”). Thus (assuming  $\mathbf{B}$  is defined on all of  $\mathbb{R}^3$ ), by Theorem 6.10 there exists a vector field  $\mathbf{A}$  called a “magnetic potential” with

$$\mathbf{B} = \nabla \times \mathbf{A}.$$

Hence in terms of  $\mathbf{A}$  our force law reads

$$\mathbf{F} = q\mathbf{v} \times (\nabla \times \mathbf{A}).$$

(1) Show, by brute force, that for any vector fields  $\mathbf{a}$  and  $\mathbf{b}$  on  $\mathbb{R}^3$ ,

$$\mathbf{a} \times (\nabla \times \mathbf{b}) = \begin{pmatrix} \mathbf{a} \cdot \partial_1 \mathbf{b} \\ \mathbf{a} \cdot \partial_2 \mathbf{b} \\ \mathbf{a} \cdot \partial_3 \mathbf{b} \end{pmatrix} - (\mathbf{a} \cdot \nabla) \mathbf{b},$$

which you might prefer to write as

$$\nabla^{\mathbf{b}}(\mathbf{a} \cdot \mathbf{b}) - (\mathbf{a} \cdot \nabla) \mathbf{b},$$

with the  $\nabla^{\mathbf{b}}$  denoting a derivative only applied to the factor  $\mathbf{b}$ , i.e.,

$$\nabla^{\mathbf{b}}(\mathbf{a} \cdot \mathbf{b})(\mathbf{x}) = \nabla_{\mathbf{y}}(\mathbf{a}(\mathbf{x}) \cdot \mathbf{b}(\mathbf{y}))|_{\mathbf{y}=\mathbf{x}}.$$

- (2) Use this identity to rewrite the Lorentz force law as

$$\mathbf{F} = q(\nabla(\mathbf{v} \cdot \mathbf{A}) - (\mathbf{v} \cdot \nabla)\mathbf{A})$$

where this time we will simply remember that the  $\nabla$  in the second term is interpreted as having no effect on  $\mathbf{v}(t)$  (which is, after all, not a vector field defined on all of  $\mathbb{R}^3$  anyway—it's just a path, so taking its gradient would not be very sensible).

- (3) Show, using the chain rule, that if  $\mathbf{x}$  moves along a curve  $\mathbf{x}(t)$  with velocity  $\mathbf{v}(t) = \dot{\mathbf{x}}(t)$  then

$$\frac{d}{dt}\mathbf{A}(\mathbf{x}(t)) = (\mathbf{v} \cdot \nabla)\mathbf{A}.$$

- (4) Now take the beautifully simple Lagrangian

$$L(\mathbf{x}, \mathbf{v}, t) = \frac{1}{2}m|\mathbf{v}|^2 + q\mathbf{v} \cdot \mathbf{A}(\mathbf{x}).$$

Compute the Euler-Lagrange equations for this Lagrangian, and show that they yield the Lorentz force law.

**7.5. Conservation Laws I.** In this section, we illustrate how the form of the Lagrangian can be used to read off certain conserved quantities in a mechanical system.

We begin by making a simple observation: if we happen to have a coordinate  $x_j$  which does not appear explicitly in the form of the Lagrangian, i.e. if we have  $L_{x_j} = 0$ , then the corresponding Euler-Lagrange equation reads:

$$\frac{d}{dt}L_{v_j}(\gamma, \dot{\gamma}, t) = 0.$$

In other words, the quantity

$$p_j \equiv L_{v_j}(\gamma, \dot{\gamma}, t)$$

is constant under the time-evolution, i.e., is a conserved quantity. This often arises in mechanical systems. For instance, if we have a potential  $V(\mathbf{x})$  in  $\mathbb{R}^3$  that happens to be independent of the first variable, so that  $V = V(x_2, x_3)$ , then the above observations tell us that for solutions of the equations of motion for the Lagrangian

$$\frac{1}{2}|\dot{\mathbf{x}}|^2 - V(x_2, x_3),$$

the quantity

$$p_1 = \dot{x}_1$$

is constant. In particular, if  $V = 0$ , the same reasoning holds in every coordinate, and we have  $p_i = \dot{x}_i$  all conserved for  $i = 1, 2, 3$ . Thus, we have derived the law of *conservation of momentum*.

We record our simple observations about conserved quantities as a proposition, as we will be returning to this theme repeatedly:

**Proposition 7.6.** *If  $L_{x_j} = 0$  then  $L_{v_j}$  is a conserved quantity.*

There is a similar but more intricate analysis in the case when  $L$  is time-independent, i.e.  $L_t$ , the partial derivative of  $L$  with respect to its explicit dependence on time in the  $2n+1$ 'st variable, is zero. To see this, we note that while  $L$  is not a conserved quantity, we may compute  $dL/dt$  and, provided  $L$  is time independent, obtain by the chain rule

$$(34) \quad \frac{d}{dt}(L(\boldsymbol{\gamma}, \dot{\boldsymbol{\gamma}})) = \sum_j L_{x_j}(\boldsymbol{\gamma}, \dot{\boldsymbol{\gamma}})\dot{\gamma}_j + L_{v_j}(\boldsymbol{\gamma}, \dot{\boldsymbol{\gamma}})\ddot{\gamma}_j.$$

(If  $L$  depended directly on  $t$  as well, this would of course contribute an extra term.) On the other hand, the Euler-Lagrange equations can be used to describe the first terms in the sum: they tell us that for each  $j$

$$L_{x_j}(\boldsymbol{\gamma}, \dot{\boldsymbol{\gamma}}) = \frac{d}{dt}L_{v_j}(\boldsymbol{\gamma}, \dot{\boldsymbol{\gamma}})$$

hence multiplying by  $\dot{\gamma}_j$  and summing yields

$$\sum L_{x_j}(\boldsymbol{\gamma}, \dot{\boldsymbol{\gamma}})\dot{\gamma}_j = \sum \dot{\gamma}_j \frac{d}{dt}L_{v_j}(\boldsymbol{\gamma}, \dot{\boldsymbol{\gamma}}).$$

Substituting this into (34) yields

$$(35) \quad \frac{d}{dt}(L(\boldsymbol{\gamma}, \dot{\boldsymbol{\gamma}})) = \sum_j \dot{\gamma}_j \frac{d}{dt}L_{v_j}(\boldsymbol{\gamma}, \dot{\boldsymbol{\gamma}}) + \ddot{\gamma}_j L_{v_j}(\boldsymbol{\gamma}, \dot{\boldsymbol{\gamma}}).$$

Finally, we can recognize the right hand side as a total derivative, by the product rule: it is nothing but

$$\frac{d}{dt} \sum_j L_{v_j}(\boldsymbol{\gamma}, \dot{\boldsymbol{\gamma}})\dot{\gamma}_j;$$

hence rearranging our equation (35) yields

$$\frac{d}{dt} \left( -L + \sum_j L_{v_j}(\boldsymbol{\gamma}, \dot{\boldsymbol{\gamma}})\dot{\gamma}_j \right) = 0.$$

Thus we obtain the following:

**Proposition 7.7.** *If  $L_t = 0$ , the quantity  $H(\boldsymbol{\gamma}, \dot{\boldsymbol{\gamma}})$  is conserved, where  $H$  is given by*

$$H(\mathbf{x}, \mathbf{v}) \equiv -L(\mathbf{x}, \mathbf{v}) + \sum_j L_{v_j}(\mathbf{x}, \mathbf{v})v_j.$$

In mechanics, this quantity is known as the *Hamiltonian* and is of central importance, since systems that are time-independent are extremely common: to a very good approximation, the laws of nature do not seem to change as time evolves. In our favorite case of potential motion with  $L = (1/2)|\dot{\mathbf{x}}|^2 - V(\mathbf{x})$  we obtain simply

$$H = \frac{1}{2}|\dot{\mathbf{x}}|^2 + V(\mathbf{x}),$$

in other words, the Hamiltonian comes out to be the kinetic plus potential energy, and *we have recovered the law of energy conservation* in the setting

of Lagrangian mechanics. Its validity depended essentially on the time-independence of the Lagrangian.

*Exercise 7.3.* Use the conservation of the Hamiltonian to simplify and solve the equations for the minimal surface of rotation from §7.4.2 as follows.

- (1) Use Proposition 7.7 to show that for the solution to the Euler-Lagrange equations (33), the quantity

$$(36) \quad H = f(1 + (f')^2)^{1/2} - \frac{f(f')^2}{(1 + (f')^2)^{1/2}}$$

satisfies  $dH/ds = 0$ .

- (2) Show that (36) simplifies to read

$$(37) \quad f^2 = H^2(1 + (f')^2).$$

- (3) Recalling that the hyperbolic trigonometric functions satisfy

$$\cosh' s = \sinh s, \quad \cosh^2 s - \sinh^2 s = 1,$$

show that the most general solution to (37) is

$$f(s) = H \cosh(H^{-1}s + s_0).$$

- (4) Show that the solution satisfying the boundary conditions  $f(\pm 1) = R$  is given by

$$f(s) = H \cosh(H^{-1}s)$$

where  $H$  satisfies the transcendental equation

$$\cosh H^{-1} = RH^{-1}.$$

Explain why such an  $H$  exists as long as  $R$  is sufficiently large, but show that if  $R$  is a sufficiently small positive number, no solutions exist. **HINT:** Try graphing the two functions  $\cosh u$  and  $Ru$  to get an idea of what is going on here.

- (5) The surface of rotation of  $H \cosh H^{-1}s$  is called a *catenoid*, and since this minimizes the area, it is the shape that a soap film spanning two rings will take on. Can you guess what the soap film will do for the values of  $R$  where no solutions of this form exist?

**7.6. Why Lagrangian mechanics?** At this point, we will pause to address the natural question: why should we care about the Lagrangian formulation of mechanics? In the case of a particle moving in a potential, anyway, all it seems to do is give us a recipe for producing Newton's equations of motions, which we knew in the first place. It turns out, though, that there are some powerful features of this new formulation of mechanics that will clarify what are otherwise some rather hard problems.

What follow are some of the reasons that Lagrangian mechanics will be of use. They split roughly into the two kinds: there are physical situations where Lagrangian mechanics makes it easier to *formulate* the equations of motion; and there are situations where it may not be hard to write down, say, the Newtonian equations of motion, but the Lagrangian formulation

may make it easier to *solve* the equations of motion, or at least to understand the qualitative behavior of solutions.

- (1) MOTION WITH CONSTRAINTS. This is the the first new application of Lagrangian mechanics that we will encounter. If a particle is constrained to slide on a curve or surface, it is much easier to find its equations of motion by passing via the Lagrangian formulation than by a simple appeal to Newton's laws. We discuss this in §7.8 below.
- (2) CHANGE OF COORDINATES. In changing to a new coordinate system (say, spherical coordinates) that may be best adapted to a physical problem, it is often easier and cleaner to write the Lagrangian in the new coordinates than to try to find how the second-order equations of motion transform. Since the Principle of Least Action was coordinate-invariant (i.e., had nothing to do with what coordinate system we chose to write our action integral in) we know that the Euler-Lagrange equations must hold for the transformed Lagrangian in the new coordinate system.
- (3) CONSERVED QUANTITIES AND SYMMETRIES. One of the most powerful features of the Lagrangian formulation of mechanics is the recipe that it offers, via Noether's Theorem, for generating conserved quantities of the motion from symmetries of the underlying Lagrangian. This can be a huge help in solving the equations of motion or of understanding how the solutions behave and will be discussed extensively in §7.9 below.
- (4) RELATIONSHIP WITH QUANTUM MECHANICS. Richard Feynman, following a suggestion of P.A.M. Dirac, famously developed a formalism for describing the behavior of a quantum particle in terms of the underlying classical action integral. A detailed exposition of Feynman's work is beyond the scope of the course, but we merely observe here that the essential quantity for computing quantum evolution (the "propagator") is given by

$$(38) \quad \int e^{iS[\gamma]/\hbar} \mathcal{D}(\gamma)$$

where

$$S[\gamma] = \int_0^T L(\gamma, \dot{\gamma}, t) dt$$

is the classical action,  $\hbar$  is a physical constant (the famous Planck constant divided by  $2\pi$ ) and the integral (38), with its funny notation  $\mathcal{D}(\gamma)$  denotes an integral taken over *all paths  $\gamma$  going from  $\mathbf{x}_0$  to  $\mathbf{x}_1$  in time  $T$* . The space of such paths is immense, and making mathematical sense of the integral (38) is correspondingly ticklish. But you can see that in principle this is a simple recipe that simply takes our classical action and yields a computation of quantum evolution.

**7.7. Change of coordinates.** The world has no preferred coordinate system: much interesting physics follows from this observation and the consequent need to express the laws of physics in a manner that makes this *coordinate invariance* manifest. The Principle of Least Action is certainly a coordinate invariant statement, and we can take considerable advantage of our freedom to choose advantageous coordinates adapted to various physical problems.

For instance, suppose we are given the potential in  $\mathbb{R}^2$   $V(\mathbf{x}) = x_1^2 + x_2^2 = |\mathbf{x}|^2$ , hence we are dealing with the Lagrangian

$$L(\mathbf{x}, \dot{\mathbf{x}}) = \frac{1}{2}|\dot{\mathbf{x}}|^2 - |\mathbf{x}|^2.$$

We could certainly write the Euler–Lagrange equations for this potential in rectangular coordinates, but we might not find them immensely illuminating, since they don’t take advantage of the *rotational symmetry* of the problem we’re studying: if we rotate our view of space,  $V$  doesn’t change. Taking advantage of symmetries will be one of the main themes we’ll explore, and we’ll see how to do it from another perspective later on, but for now let’s notice that if we change to *polar coordinates*  $(r, \varphi)$  then  $V$  depends only on  $r$  and not on  $\varphi$ , and this seems like a simplification. So let us try to change coordinates; recall that polar coordinates in the plane are given by

$$\mathbf{x} = \begin{pmatrix} r \cos \varphi \\ r \sin \varphi \end{pmatrix}.$$

We now want to think of the motion in the  $r, \varphi$  plane as the fundamental thing to compute, and write the Principle of Least Action in terms of  $(r(t), \varphi(t))$  instead of  $(x_1(t), x_2(t))$ .

The only tricky point here is how to deal with the velocity term. We have of course

$$\mathbf{v} = \dot{\mathbf{x}},$$

and to deal with this in polar coordinates, we need to use the *chain rule*:  $x_1$  and  $x_2$  depend on  $r$  and  $\varphi$  which in turn depend on  $t$ , so we compute:

$$(39) \quad \dot{\mathbf{x}} = \begin{pmatrix} \dot{r} \cos \varphi - r \dot{\varphi} \sin \varphi \\ \dot{r} \sin \varphi + r \dot{\varphi} \cos \varphi \end{pmatrix} = \dot{r} \begin{pmatrix} \cos \varphi \\ \sin \varphi \end{pmatrix} + r \dot{\varphi} \begin{pmatrix} -\sin \varphi \\ \cos \varphi \end{pmatrix}$$

Thus, we compute

$$(40) \quad |\dot{\mathbf{x}}|^2 = \dot{r}^2 + r^2 \dot{\varphi}^2.$$

*Exercise 7.4.* Verify (40). You can do this computation one of two ways: the lowbrow way is just to take the sum of squares of components in (39), and note some beautiful cancellations. Perhaps a nicer way is to notice that the two vectors into which we’ve split the right-hand-side in (39) in the final expression are orthogonal to one another, and moreover that the vectors

$$\begin{pmatrix} \cos \varphi \\ \sin \varphi \end{pmatrix}, \quad \begin{pmatrix} -\sin \varphi \\ \cos \varphi \end{pmatrix},$$

are both unit vectors. This makes it very easy to compute the norm of the sum.

Of course the whole point of polar coordinates was that we can write  $V$  very simply in polar coordinates as  $V = r^2$ . So we finally arrive at the expression for the Lagrangian in polar coordinates

$$L(r, \varphi, \dot{r}, \dot{\varphi}) = \frac{1}{2}(\dot{r}^2 + r^2\dot{\varphi}^2) - r^2.$$

The Principle of Least Action then tells us that the Euler–Lagrange equations must apply to these function of  $(r(t), \varphi(t))$  in order to make the action stationary at the physical path: we thus write out the equations

$$L_r = \frac{d}{dt}L_{\dot{r}}, \quad L_{\varphi} = \frac{d}{dt}L_{\dot{\varphi}},$$

to obtain the equations of motion

$$r\dot{\varphi}^2 - 2r = \ddot{r}, \quad 0 = \frac{d}{dt}(r^2\dot{\varphi}).$$

These equations, and the second of them in particular, pretty well represent victory in understanding the problem. Note that the second equation yields a *constant of motion*, a quantity that has zero time derivative: if we set

$$\ell \equiv r^2\dot{\varphi},$$

we find that  $\dot{\ell} = 0$ . What is great about this is that it enables us to reduce what started as a two-dimensional problem into a one-dimensional one: noting that  $\dot{\varphi} = \ell/r^2$  we may then transform the other of the Euler–Lagrange equations, for  $\ddot{r}$ , to read

$$(41) \quad \ddot{r} = \frac{\ell^2}{r^3} - 2r.$$

Since  $\ell$  is constant, this is just a single second-order ODE for  $r$ , which is, it turns out, much easier to deal with than the second-order system of two equations that we started with. For a start, this would be pretty easy and efficient to solve numerically on a computer via Euler’s method. But we can do much better in analyzing it qualitatively, in fact. Let us set

$$V_{\text{eff}}(r) = \frac{\ell^2}{2r^2} + r^2,$$

where we have simply taken minus the anti-derivative of the right side of (41). Now pretend we live in a one-dimensional world with coordinate  $r$  and we are interested in studying the motion in the potential  $V_{\text{eff}}(r)$ . Well, the Euler–Lagrange equation, a.k.a. Newton’s Second Law, tells us that once again,

$$\ddot{r} = -\frac{dV_{\text{eff}}}{dr} = \frac{\ell^2}{r^3} - 2r.$$

In other words, the equation we’ve found that  $r$  satisfies in our original problem is the same as it would satisfy if it were simply subject to one-dimensional motion in the “effective” potential  $V_{\text{eff}}$ . So we can draw phase

plane diagrams for  $r(t)$ , analyze its equilibria, closed orbits, periods, etc., using all the wonderful technology we've developed for one dimensional potential motion. We will discuss the physical interpretation of the conserved quantity  $\ell$  and its relationship to the symmetries of the original problem a bit later in the course.

*Exercise 7.5.* Choose a value for  $\ell$  and draw a phase plane diagram for the motion of  $r$  in the potential  $V_{\text{eff}}(r)$ . (Restrict attention just to the physically meaningful set where  $r > 0$ ; why can't the particle cross over to  $r < 0$ ?)

**7.8. Motion with constraints.** One of the most powerful features of the Lagrangian formulation of mechanics is that it allows us to deal easily with problems of *motions with constraints*, in which a particle is constrained, by forces whose analysis is not our primary interest, to lie on some curve or surface. For instance, the real pendulum should be thought of as a particle in  $\mathbb{R}^2$ , subject to the downward force of terrestrial gravity, constrained to lie on the circle  $x_1^2 + x_2^2 = 1$ . This constraint is enforced by the rigid pendulum pulling or pushing on the mass at the end to hold it in place, but we will do our best to ignore the exact mechanism which is holding it there.

We reiterate here that the role of the principle of least action in mechanics is for us two-fold: it helps both with finding the equations of motion for a given physical system; and it often also gives us help in solving that system. This section gives an example of its utility in the first sense: the variational principle will give us a very simple way of finding the equations of motion satisfied by a constrained system:<sup>17</sup>

*Definition 7.8.* Let  $X \subset \mathbb{R}^n$ , and let  $L(\mathbf{x}, \mathbf{v}, t)$  be a Lagrangian defined on all of  $\mathbb{R}^n$ , with a corresponding action integral

$$S(\gamma) = \int_0^T L(\gamma, \dot{\gamma}, t) dt.$$

We define the *restricted action*  $S_X$  to be the restriction of the action functional  $S$  to paths lying in  $X$ . We define the motion of a particle constrained to  $X$  to be the motion along paths that are stationary for the restricted action  $S_X$ .

We adopt the same definition when the constraint  $X$  is dependent on the time-variable  $t$ , as well.

In other words, in checking whether a path is stationary, we are no longer comparing to all nearby paths, but rather only to those nearby paths that lie in  $X$ . You may be used to solving analogous problems in multi-variable calculus, in which you need to find the extrema of a function when you

---

<sup>17</sup>We will take this to be an *axiom* but you might like to think about how you might derive it as a limit of motion of almost-constrained systems, in which we add a potential  $V_{\text{constraint}}(\mathbf{x})$  to an unconstrained system, where  $V_{\text{constraint}}(\mathbf{x})$  is, say, zero along the set  $X$  and very rapidly growing away from it, so that, by conservation of energy, we see that the motion of a particle is forced to stay very close to  $X$ .

restrict attention to a curve or surface in  $\mathbb{R}^2$  or  $\mathbb{R}^3$ . One method you may have learned to do this is the method of Lagrange multipliers. There is a perfectly analogous method of Lagrange multipliers that you can use to solve constrained problems in the calculus of variations. We will pursue a different method, however, which is in many ways even simpler.

To do this, we remember that a smooth curve or surface in  $\mathbb{R}^3$  (say) is one that can be *parametrized* by a coordinate  $s \in \mathbb{R}^1$  or  $u, v \in \mathbb{R}^2$  via a map

$$s \mapsto \mathbf{f}(s) = \begin{pmatrix} f_1(s) \\ f_2(s) \\ f_3(s) \end{pmatrix}$$

in the case of a curve, and

$$(u, v) \mapsto \mathbf{f}(u, v) = \begin{pmatrix} f_1(u, v) \\ f_2(u, v) \\ f_3(u, v) \end{pmatrix}$$

in the case of a surface. In order that the curve or surface really be smooth, we will require that the derivative

$$\dot{\mathbf{f}}(s)$$

in the case of a curve be non-vanishing, and that the matrix

$$(\mathbf{f}_u \ \mathbf{f}_v) = \begin{pmatrix} \frac{\partial f_1}{\partial u} & \frac{\partial f_1}{\partial v} \\ \frac{\partial f_2}{\partial u} & \frac{\partial f_2}{\partial v} \\ \frac{\partial f_3}{\partial u} & \frac{\partial f_3}{\partial v} \end{pmatrix}$$

have rank 2 (i.e., as large as possible) everywhere. We will not emphasize this technical assumption, and refer the student to a good textbook on multi-variable calculus, or on the differential geometry of curves and surfaces for further explanation.

For example, if  $X \subset \mathbb{R}^3$  is the unit sphere, then we may parametrize most of it by two variables, usually denoted  $\theta, \varphi$ , with the mappings

$$(\theta, \varphi) \mapsto \begin{pmatrix} \sin \theta \cos \varphi \\ \sin \theta \sin \varphi \\ \cos \theta \end{pmatrix}$$

This parametrization is of course nothing but spherical coordinates. Another example might be the paraboloid  $x_3 = x_1^2 + x_2^2$  in  $\mathbb{R}^3$ . This can easily be parametrized by

$$(u, v) \mapsto (u, v, u^2 + v^2).$$

More generally, if we have a surface given by a graph  $x_3 = f(x_1, x_2)$  then we have the parametrization

$$(u, v) \mapsto (u, v, g(u, v))$$

for free. This seems like a very special example, but is more general than it appears: if you think about it, a little patch of a surface always looks like such a graph, at least after you turn your head a little (i.e., rotate your

coordinates) in case you have been unlucky and ended up with a piece that looks exactly vertical (see Figure ??).

Now all we have to do to compute our Euler-Lagrange equations for a constrained system is use the parametrization of  $X$  to write down the most general path  $\gamma$  lying in  $X$ . We will illustrate with the example of the *spherical pendulum*, which is the three-dimensional version of the real pendulum: a particle in  $\mathbb{R}^3$  subjected to terrestrial gravity, hence with

$$L = \frac{1}{2}|\mathbf{v}|^2 - x_3$$

restricted to the unit sphere,  $S^2 \subset \mathbb{R}^3$ . Using our parametrization of the sphere above, we see that every point can be written

$$\mathbf{f}(\theta, \varphi) \equiv \begin{pmatrix} \sin \theta \cos \varphi \\ \sin \theta \sin \varphi \\ \cos \theta \end{pmatrix}.$$

Hence a path on the sphere can be described by the two functions  $\theta(t), \varphi(t)$ , i.e. any path can be written

$$(42) \quad \gamma(t) = \mathbf{f}(\theta(t), \varphi(t))$$

for some choices of  $\theta(t), \varphi(t)$ . You can think of  $(\theta(t), \varphi(t))$  as describing a curve in  $\mathbb{R}^2$ , and you will see that we are in the process of rewriting our whole problem as a two-dimensional problem, since the sphere is two-dimensional.

Now the only hard part is to compute  $\dot{\gamma}$  for  $\gamma$  given by (42). We have to use the chain rule to compute

$$\begin{aligned} \dot{\gamma}(t) &= \frac{d}{dt} \begin{pmatrix} \sin \theta(t) \cos \varphi(t) \\ \sin \theta(t) \sin \varphi(t) \\ \cos \theta(t) \end{pmatrix} \\ (43) \quad &= \begin{pmatrix} \cos \theta \dot{\theta} \cos \varphi - \sin \theta \sin \varphi \dot{\varphi} \\ \cos \theta \dot{\theta} \sin \varphi + \sin \theta \cos \varphi \dot{\varphi} \\ -\sin \theta \dot{\theta} \end{pmatrix} \\ &= \dot{\theta} \begin{pmatrix} \cos \theta \cos \varphi \\ \cos \theta \sin \varphi \\ -\sin \theta \end{pmatrix} + \dot{\varphi} \begin{pmatrix} -\sin \theta \sin \varphi \\ \sin \theta \cos \varphi \\ 0 \end{pmatrix} \end{aligned}$$

We are interested in

$$L(\gamma, \dot{\gamma}, t) = \frac{1}{2}|\dot{\gamma}|^2 - \gamma_3$$

so we now compute

$$|\dot{\gamma}|^2 = \dot{\theta}^2 + \sin^2 \theta \dot{\varphi}^2.$$

Thus, we can finally rewrite

$$L(\gamma, \dot{\gamma}, t) = \frac{1}{2}(\dot{\theta}^2 + \sin^2 \theta \dot{\varphi}^2) - \cos \theta$$

completely in terms of the functions  $\theta(t), \varphi(t)$ . This is our restricted Lagrangian: it gives us the Lagrangian  $L$  in terms of an arbitrary path  $\mathbf{f}(\theta(t), \varphi(t))$

lying on  $S^2$ . As in our computations in polar coordinates in Section 7.7, stationarity of  $S_{\text{restricted}}$  among paths of this form means simply that *the Euler-Lagrange equations for  $\theta(t), \varphi(t)$  should be satisfied*. So it remains only to write down the Euler-Lagrange equations in these variables. The equation in  $\theta$  gives

$$(44) \quad \ddot{\theta} = \sin \theta + \sin \theta \cos \theta \dot{\varphi}^2,$$

while the equation in  $\varphi$  gives

$$(45) \quad \frac{d}{dt}(\sin^2 \theta \dot{\varphi}) = 0.$$

These equations really are *coupled*: the variable  $\theta$  occurs in the expression for  $\ddot{\varphi}$  and  $\varphi$  occurs in the expression for  $\ddot{\theta}$ . Note that since  $\sin^2 \theta \dot{\varphi}$  has zero derivative,  $\theta^2 \dot{\varphi} \equiv \ell$  is a *constant*. Thus we may substitute this knowledge into (44) and obtain

$$\ddot{\theta} = \sin \theta + \frac{\cos \theta}{\sin^3 \theta} \ell^2.$$

This one is tough to deal with, but has the virtue of being a single equation for  $\theta$ , now uncoupled from  $\varphi$ . If we let  $p = \dot{\theta}$  then we can split this second order equation into a pair of first order ones

$$(46) \quad \begin{aligned} \dot{\theta} &= p \\ \dot{p} &= \sin \theta + \frac{\cos \theta}{\sin^3 \theta} \ell^2, \end{aligned}$$

and we can do our usual phase plane analysis for this system.

*Exercise 7.6.* Draw a phase plane picture for the solutions of the system (46). Notice that the motion is bounded, following closed curves.

In fact, time independence of the Lagrangian of course gives *second* conserved quantity in addition to  $\ell$ , which will help us analyze this system. We will pursue the consequences of this shortly.

More generally, let us analyze the following problem: a particle moving on a parametrized surface with, with *no potential at all*. In free space, the motion of a particle with  $V = 0$  is not very interesting: solutions to the equations of motion are simply the straight lines  $\mathbf{x}(t) = \mathbf{x}_0 + \mathbf{v}_0 t$ . But if we constrain the particle to a surface, things are not so simple.

Assume the parametrization is by

$$u, v \mapsto \mathbf{f}(s) = \begin{pmatrix} f_1(u, v) \\ f_2(u, v) \\ f_3(u, v) \end{pmatrix},$$

the most general set-up. A general path on the surface can then be represented

$$\gamma(t) = \mathbf{f}(u(t), v(t))$$

where  $u(t), v(t)$  is a parametrized curve in  $\mathbb{R}^2$ . Then since the Lagrangian is just the quantity

$$L(\mathbf{x}, \mathbf{v}, t) = \frac{1}{2} |\dot{\mathbf{x}}|^2 = \frac{1}{2} (\dot{x}_1^2 + \dot{x}_2^2 + \dot{x}_3^2)$$

we have by the chain rule

$$L(\gamma, \dot{\gamma}, t) = \frac{1}{2} (((f_1)_u \dot{u} + (f_1)_v \dot{v})^2 + ((f_2)_u \dot{u} + (f_2)_v \dot{v})^2 + ((f_3)_u \dot{u} + (f_3)_v \dot{v})^2).$$

(with subscripts on the  $f_j$  indicating partial derivatives). This Lagrangian is thus of the form

$$(47) \quad L = E(u, v) \dot{u}^2 + 2F(u, v) \dot{u} \dot{v} + G(u, v) \dot{v}^2,$$

where  $E, F, G$  are the functions of  $u, v$  defined by

$$(48) \quad \begin{aligned} E(u, v) &= \frac{1}{2} ((f_1)_u^2 + (f_2)_u^2 + (f_3)_u^2), \\ F(u, v) &= \frac{1}{2} ((f_1)_u (f_1)_v + (f_2)_u (f_2)_v + (f_3)_u (f_3)_v), \\ G(u, v) &= \frac{1}{2} ((f_1)_v^2 + (f_2)_v^2 + (f_3)_v^2). \end{aligned}$$

*Exercise 7.7.* Compute these functions  $E, F, G$  in the case of the sphere  $S^2$ , parametrized by  $\theta, \varphi$ . Check that the answer you get agrees with the kinetic energy piece of the Lagrangian that we wrote down above for the spherical pendulum.

*Exercise 7.8.* Compute these functions  $E, F, G$  in the special case where the parametrized surface is the graph of a function:  $x_3 = f(x_1, x_2)$ .

The Euler-Lagrange equations for the Lagrangian (47) are easy enough to write down but are a little messy in this level of generality. They are extremely important equations in geometry, known as the *geodesic* equations. They describe the behavior of a free particle moving around on a surface.<sup>18</sup> We will however discuss them in some special cases. One especially good one is a generalization of our work on the sphere: we will study a *surface of revolution*. We choose two functions  $f(s), g(s)$  specifying a curve in  $\mathbb{R}^2$ . Define a parametrized surface by rotating this curve through  $\mathbb{R}^3$ , hence by

$$\mathbf{f}(s, \theta) = (f(s) \cos \theta, f(s) \sin \theta, g(s)).$$

*Exercise 7.9.* Compute  $E, F, G$  for the surface of revolution, and write down the geodesic equations.

*Exercise 7.10.* Write down the Euler-Lagrange equations for the free particle on a sphere  $S^2$ . Show that  $(\theta(t) = \pi/2, \varphi(t) = t)$  is a solution. What kind of curve is this? Can you use your knowledge of this one solution to figure out what all the other solutions of the equations look like, i.e. what curves they trace out?

<sup>18</sup>Of course, if this problem proves too simple, we can always add a potential to the problem to make life more interesting.

*Exercise 7.11.* Analyze the motion of a free particle moving on a paraboloid

$$x_3 = x_1^2 + x_2^2.$$

*Exercise 7.12.* What is the conserved Hamiltonian  $H$  resulting from the time-independence of the Lagrangian for geodesic motion?

A further interesting feature of the geodesic equations is that while we have derived them here from specifying a surface in  $\mathbb{R}^3$  and writing down the “free” Lagrangian (i.e., the one with zero potential), we could just as well choose three functions  $E, F$ , and  $G$  of two variables  $u, v$ , and use the expression

$$L = E(u, v)\dot{u}^2 + 2F(u, v)\dot{u}\dot{v} + G(u, v)\dot{v}^2$$

as the Lagrangian for the motion of a particle in the  $u, v$ -plane. It is usual to choose  $E, F, G$  in such a way that the symmetric matrix

$$\begin{pmatrix} E & F \\ F & G \end{pmatrix}$$

is everywhere positive definite, so that the Lagrangian

$$L = \begin{pmatrix} \dot{u} & \dot{v} \end{pmatrix} \begin{pmatrix} E & F \\ F & G \end{pmatrix} \begin{pmatrix} \dot{u} \\ \dot{v} \end{pmatrix}$$

is always positive, as a kinetic energy should probably be. This is the modern, “abstract” point of view of Riemannian geometry, and we can study the resulting geodesic equations just as happily as if they came from a real surface<sup>19</sup> in  $\mathbb{R}^3$ . A very important example is if we take just the half-plane consisting of  $(u, v)$  with  $v > 0$ , and use the very simple functions

$$E = \frac{1}{v^2}, \quad F = 0, \quad G = \frac{1}{v^2}.$$

This turns the upper half-plane into a geometric object known as the “hyperbolic plane.”

*Exercise 7.13.* (1) Compute the geodesic equations for the hyperbolic plane. Show that vertical lines  $u = \text{constant}$  solve these equations for some appropriate parametrization  $v = v(t)$ .

(2) Write down the conserved quantity corresponding to the fact that the Lagrangian is independent of  $u$ .

We will return to the equations in the hyperbolic plane later on, when we study Hamiltonian dynamics.

*Exercise 7.14.* A particle, acting on by gravitational force  $\mathbf{F} = (0, 0, -1)$ , is confined to lie on a spinning hoop, which consists of a circle of radius 1 spinning at one revolution per  $2\pi/\omega$  seconds about the  $x_3$  axis, which is one of its diameters. Thus if  $\theta$  measures the location of the particle on the hoop,

<sup>19</sup>It is a fascinating open problem in geometry to know, given three functions  $E, F, G$  in  $\mathbb{R}^2$ , if there is at least a little patch of a parametrized surface in  $\mathbb{R}^3$  for which these arise in the process described above. This is known as the “isometric imbedding problem”

with  $\theta = 0$  corresponding to straight down, the particle's position at time  $t$  is given by

$$\mathbf{x}(t) = \begin{pmatrix} \sin \theta \sin \omega t \\ \sin \theta \cos \omega t \\ -\cos \theta \end{pmatrix}.$$

- (1) Draw a picture to illustrate this assertion about where the particle is.
- (2) Write down the Lagrangian for the motion (in terms of  $\theta, \dot{\theta}$ ) by plugging  $\mathbf{x}, \dot{\mathbf{x}}$  into the standard Lagrangian for a particle in a gravitational potential.
- (3) Write down the Euler-Lagrange equation describing  $\theta(t)$ .
- (4) Determine all equilibrium points of the motion, and whether they are stable or unstable. For stable equilibria, find the approximate frequency of oscillation about the equilibrium. (Hint: you want points where  $\ddot{\theta} = 0$ ; to find the small oscillations nearby, linearly approximate the effective force  $\ddot{\theta}$ . Or else write the motion in terms of an effective potential and use the usual methods. Note also that you will probably need to split into different cases depending on the value of the angular velocity parameter  $\omega$ .)

**7.9. Conservation Laws II.** Recall that for a Lagrangian  $L$  that is independent of a variable  $x_j$  we obtained a conservation law, with a conserved quantity

$$L_{v_j}(\gamma, \dot{\gamma}, t).$$

Additionally, for Lagrangians independent of  $t$ , we obtained another conservation law. Another way to express the notion of being independent of one variable ( $x_j$  or  $t$ ) is that the Lagrangian enjoys a *symmetry* under translation in that variable. We let

$$T_a^j(x_1, \dots, x_n, v_1, \dots, v_n) = (x_1, \dots, x_j + a, \dots, x_n, v_1, \dots, v_n)$$

denote the operation of shifting the  $j$ 'th coordinate by  $a$ . Then the statement that  $L$  is independent of  $x_j$  is simply that

$$L(\mathbf{x}, \mathbf{v}, t) = L(T_a^j(\mathbf{x}, \mathbf{v}), t).$$

This is an example of a *continuous symmetry*, i.e. a family of transformations that leave something (in this case,  $L$ ), unchanged. That  $L$  enjoys this continuous symmetry and that it consequently enjoys a conservation law is our first and simplest instance of the following general, and profound, principle:

A continuous symmetry gives rise to a conservation law.

We remark that our notion of continuous symmetry will not include what might be more familiar notions of symmetry, which are often discrete symmetries. For instance, the bilateral symmetry of the human body, corresponding to invariance under reflection in a mirror, is a discrete symmetry: there is just a single transformation under which the body is apparently

invariant—reflection in a plane bisecting the body vertically—not a continuous family of them.

In addition to the translation symmetry denote  $T_a^j$  above, another useful example to keep in mind is that of *rotation*. This is, say, the family of symmetries obtained by, say, rotation  $\mathbb{R}^3$  about the  $z$ -axis through angle  $\theta$ . The angle  $\theta$  can take on any real value, hence this is indeed a continuous family of symmetries.

To understand symmetries other than translation of  $\mathbb{R}^n$ , we have to understand how they act simultaneously on position and on velocity. In general, if

$$F = (f_1(\mathbf{x}), \dots, f_n(\mathbf{x})) : \mathbb{R}^n \rightarrow \mathbb{R}^n$$

is a smooth mapping, we can use  $F$  to transform a path  $\gamma(t)$  into the new path

$$F(\gamma(t)) = (f_1(\gamma(t)), \dots, f_n(\gamma(t))).$$

Now we compute the tangent to this new path, and obtain

$$\frac{d}{dt}(F(\gamma(t))) = \sum_{j=1}^n \left( \frac{\partial f_1}{\partial x_j} \dot{\gamma}_j(t), \dots, \frac{\partial f_n}{\partial x_j} \dot{\gamma}_j(t) \right),$$

which it is most convenient and elegant to write as<sup>20</sup>

$$dF \cdot \dot{\gamma}$$

where  $dF$  denotes the matrix of partial derivatives

$$dF = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_n} \end{pmatrix}.$$

Thus, under the transformation  $F$ , if  $\mathbf{v}$  denotes the velocity of a path, then  $\mathbf{v}$  gets transformed to

$$dF \cdot \mathbf{v}.$$

Thus, we associate to the transformation  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$  a map, denoted  $F_*$ , of  $\mathbb{R}_{\mathbf{x}}^n \times \mathbb{R}_{\mathbf{v}}^n \rightarrow \mathbb{R}_{\mathbf{x}}^n \times \mathbb{R}_{\mathbf{v}}^n$ , given by

$$F_*(\mathbf{x}, \mathbf{v}) = (F(\mathbf{x}), dF(\mathbf{x}) \cdot \mathbf{v}).$$

This denotes the action of  $F$  both on points and on tangent vectors simultaneously.

Now let  $F^a$  be a family of transformations of  $\mathbb{R}^n$  with  $F^0(\mathbf{x}) = \mathbf{x}$  and let  $F_*^a$  denote the corresponding transformations of  $\mathbb{R}_{\mathbf{x}}^n \times \mathbb{R}_{\mathbf{v}}^n$ .

The following key result explaining our dogma above that symmetries lead to conservation laws is known as “Noether’s theorem” after Emmy Noether, who proved it in much greater generality than we consider here, in 1915.

---

<sup>20</sup>Indeed, this matrix formulation is the best way to think of the multi-variable chain rule.

**Theorem 7.9.** *Suppose that  $L(F_*^a(\mathbf{x}, \mathbf{v}), t) = L(\mathbf{x}, \mathbf{v}, t)$  for all  $t$  in an interval containing 0. Then the quantity*

$$p = \sum_{j=1}^n L_{v_j} W_j$$

*is conserved under the evolution of the mechanical system, where  $\mathbf{W}$  is the vector field*

$$\mathbf{W}(\mathbf{x}) = \left. \frac{d}{da} \right|_{a=0} F^a(\mathbf{x}).$$

You should think of the vector field  $\mathbf{W}$  as the initial direction in which the family of maps  $F^a$  pushes the point  $\mathbf{x}$ . More precisely, we have the first order approximation

$$(49) \quad F^a(\mathbf{x}) = \mathbf{x} + a\mathbf{W}(\mathbf{x}) + O(a^2)$$

where  $O(a^2)$  means a term that is bounded by a multiple of  $a^2$  as  $a \rightarrow 0$ .

*Proof.* Our only actual use of the symmetry

$$(50) \quad L(\mathbf{x}, \mathbf{v}, t) = L(F_*^a(\mathbf{x}, \mathbf{v}), t) \quad \text{for all } a$$

will be to take its derivative at  $a = 0$ , where the map  $F^a$  is simply the identity. (Consequently, we really just need what you might call an “infinitesimal symmetry.”)

We have already decided to call  $dF^a/da = \mathbf{W}$ , so the components of the  $a$ -derivative of  $L(F_*^a(\mathbf{x}, \mathbf{v}), t)$  involving  $L_{x_i}$ ’s are easy to compute. To deal with the  $\mathbf{v}$  components of the partial derivative, we note that

$$\left. \frac{d}{da} \right|_{a=0} dF^a(\mathbf{x}) = \begin{pmatrix} \left. \frac{\partial^2}{\partial a \partial x_1} \right|_{a=0} f_1^a & \cdots & \left. \frac{\partial^2}{\partial a \partial x_n} \right|_{a=0} f_1^a \\ \vdots & \ddots & \vdots \\ \left. \frac{\partial^2}{\partial a \partial x_1} \right|_{a=0} f_n^a & \cdots & \left. \frac{\partial^2}{\partial a \partial x_n} \right|_{a=0} f_n^a \end{pmatrix}.$$

On the other hand, by definition of  $\mathbf{W}$  we have  $(\partial f_i^a / \partial a)|_{a=0} = W_i$ , hence by commutativity of mixed partial derivatives,

$$\left. \frac{d}{da} \right|_{a=0} dF^a(\mathbf{x}) = \begin{pmatrix} \frac{\partial W_1}{\partial x_1} & \cdots & \frac{\partial W_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial W_n}{\partial x_1} & \cdots & \frac{\partial W_n}{\partial x_n} \end{pmatrix}.$$

Thus, by the chain rule, we can now compute the  $a$ -derivative of (50):

$$\begin{aligned} 0 &= \left. \frac{d}{da} \right|_{a=0} L(F_*^a(\mathbf{x}, \mathbf{v}), t) \\ &= \sum_{i=1}^n L_{x_i}(\mathbf{x}, \mathbf{v}, t) W_i(\mathbf{x}) + \sum_{i,j=1}^n L_{v_i}(\mathbf{x}, \mathbf{v}, t) (\partial W_i(\mathbf{x}) / \partial x_j) v_j. \end{aligned}$$

In particular, if we evaluate this expression at  $\mathbf{x} = \boldsymbol{\gamma}(t)$ ,  $\mathbf{v} = \dot{\boldsymbol{\gamma}}(t)$  we obtain

$$(51) \quad -\sum_{i=1}^n L_{x_i}(\boldsymbol{\gamma}(t), \dot{\boldsymbol{\gamma}}(t), t) W_i(\boldsymbol{\gamma}(t)) = \sum_{i,j=1}^n L_{v_i}(\boldsymbol{\gamma}(t), \dot{\boldsymbol{\gamma}}(t), t) (\partial W_i(\boldsymbol{\gamma}(t)) / \partial x_j) \dot{\gamma}_j(t)$$

Recognizing that by the chain rule,  $(d/dt)W_i(\boldsymbol{\gamma}(t)) = \sum_{j=1}^n (\partial W_i(\boldsymbol{\gamma}(t)) / \partial x_j) \dot{\gamma}_j(t)$ , we can rewrite this as

$$(52) \quad -\sum_{i=1}^n L_{x_i}(\boldsymbol{\gamma}(t), \dot{\boldsymbol{\gamma}}(t), t) W_i(\boldsymbol{\gamma}(t)) = \sum_{i=1}^n L_{v_i}(\boldsymbol{\gamma}(t), \dot{\boldsymbol{\gamma}}(t), t) \frac{d}{dt} W_i(\boldsymbol{\gamma}(t)).$$

Finally, we now return to the Euler-Lagrange equations. We have for each  $i = 1, \dots, n$

$$-L_{x_i}(\boldsymbol{\gamma}(t), \dot{\boldsymbol{\gamma}}(t), t) + \frac{d}{dt} L_{v_i}(\boldsymbol{\gamma}(t), \dot{\boldsymbol{\gamma}}(t), t) = 0.$$

multiplying this by  $W_i$  and summing over  $i$  yields

$$-\sum_{i=1}^n L_{x_i}(\boldsymbol{\gamma}(t), \dot{\boldsymbol{\gamma}}(t), t) W_i(\boldsymbol{\gamma}(t)) + \sum_{i=1}^n W_i(\boldsymbol{\gamma}(t)) \frac{d}{dt} L_{v_i}(\boldsymbol{\gamma}(t), \dot{\boldsymbol{\gamma}}(t), t) = 0$$

and we can recognize the first term as the left side of (52). Thus we now have

$$\sum_{i,j=1}^n \frac{d}{dt} (W_i(\boldsymbol{\gamma}(t))) L_{v_i}(\boldsymbol{\gamma}(t), \dot{\boldsymbol{\gamma}}(t), t) + \sum_{i=1}^n W_i(\boldsymbol{\gamma}(t)) \frac{d}{dt} L_{v_i}(\boldsymbol{\gamma}(t), \dot{\boldsymbol{\gamma}}(t), t) = 0$$

By the product rule, this just says that

$$\frac{d}{dt} (W_i(\boldsymbol{\gamma}(t)) L_{v_i}(\boldsymbol{\gamma}(t), \dot{\boldsymbol{\gamma}}(t), t)) = 0,$$

which is simply the assertion of the theorem.  $\square$

We now turn to applications of this theorem, which are many and fascinating. First, we revisit conservation of momentum. For this, the symmetry was

$$F^a(x_1, \dots, x_n) = (x_1, \dots, x_j + a, \dots, x_n)$$

and  $dF^a = \text{Id}$ . We easily compute in this case that  $(d/da)F^a = (0, \dots, 1, \dots, 0) = \mathbf{e}_j$ . Hence the conserved quantity in Noether's theorem is just

$$L_{v_j}(\boldsymbol{\gamma}, \dot{\boldsymbol{\gamma}}, t),$$

the momentum in the  $j$ 'th coordinate.

A more interesting example is *rotational symmetry*. Let us suppose we have a physical system in  $\mathbb{R}^3$  that is rotationally symmetric about the  $x_3$ -axis. This means that our family of symmetries is given by rotation by angle  $a$  about this axis, i.e.

$$F^a(\mathbf{x}) = \begin{pmatrix} \cos a & -\sin a & 0 \\ \sin a & \cos a & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \mathbf{x} \equiv R_a \cdot \mathbf{x}$$

Then we also compute<sup>21</sup>

$$dF = R_a$$

as well. Hence the rotational symmetry we are requiring is

$$L(R_a \cdot \mathbf{x}, R_a \cdot \mathbf{v}, t) = L(\mathbf{x}, \mathbf{v}, t).$$

For instance in the 3-dimensional harmonic oscillator we have

$$L(\mathbf{x}, \mathbf{v}, t) = \frac{1}{2}|\mathbf{v}|^2 + k_1x_1^2 + k_2x_1^2 + k_3x_1^2.$$

As long as  $k_1 = k_2$ , this symmetry holds, as the whole system is symmetric about the  $x_3$ -axis.

*Exercise 7.15.* Check that

$$L(\mathbf{x}, \mathbf{v}, t) = \frac{1}{2}|\mathbf{v}|^2 + k_1x_1^2 + k_2x_1^2 + k_3x_1^2$$

enjoys rotational symmetry about the  $x_3$ -axis. Note that  $|\mathbf{v}|^2$  certainly has this symmetry (indeed, about any axis), as

$$|\mathbf{v}|^2 = \langle \mathbf{v}, \mathbf{v} \rangle = \langle R \cdot \mathbf{v}, R \cdot \mathbf{v} \rangle$$

for any matrix  $R$  with the property  $R^t R = \text{Id}$ .

Now we try to find the conserved quantity. We have

$$\begin{aligned} \mathbf{W} &= \left. \frac{dF^a(\mathbf{x})}{da} \right|_{a=0} \\ &= \left. \frac{d}{da} \right|_{a=0} R_a \cdot \mathbf{x} \\ &= \begin{pmatrix} -\sin 0 & -\cos 0 & 0 \\ \cos 0 & -\sin 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \cdot \mathbf{x} \\ &= \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \cdot \mathbf{x} \\ &= \begin{pmatrix} -x_2 \\ x_1 \\ 0 \end{pmatrix}. \end{aligned}$$

Thus, the conserved quantity in question is

$$\sum_{i=1}^3 L_{v_i} W_i = -x_2 L_{v_1} + x_1 L_{v_2}.$$

In the case of our harmonic oscillator—or, indeed, of any system with a potential symmetric about the  $x_3$ -axis—this comes out to be

$$-x_2 v_1 + x_1 v_2 = -x_2 \dot{x}_1 + x_1 \dot{x}_2,$$

---

<sup>21</sup>We are taking the derivative of a linear transformation here, so it should not be surprise that the derivative of the map and the map itself agree.

which is usually called the *angular momentum* about the  $x_3$ -axis.

*Exercise 7.16.* Check, by actually differentiating, that the angular momentum

$$-x_2\dot{x}_1 + x_1\dot{x}_2$$

is conserved along the motion of the 3-dimensional oscillator provided  $k_1 = k_2$ .

*Exercise 7.17.* Show that for a particle in  $\mathbb{R}^3$  with Lagrangian of the form

$$L(\mathbf{x}, \mathbf{v}) = \frac{1}{2}|\mathbf{v}|^2 - V$$

with the potential of the special form

$$V = V(|\mathbf{x}|)$$

(i.e., *spherically symmetric*), the *total angular momentum*

$$\mathbf{L} = \mathbf{x} \times \dot{\mathbf{x}}$$

is conserved.

Another excellent example is that of a multi-particle system. Suppose for a start that we have two particles, at positions  $\mathbf{x}^1$  and  $\mathbf{x}^2$  (each in  $\mathbb{R}^3$ ) with masses  $m_1$  and  $m_2$ , interacting via Newtonian gravity. Then the gravitational potential associated to this system is

$$V(\mathbf{x}^1, \mathbf{x}^2) = -\frac{Gm_1m_2}{|\mathbf{x}^1 - \mathbf{x}^2|^2},$$

and the Lagrangian is the sum of the kinetic energies of the particles minus this potential energy:

$$L(\mathbf{x}^1, \mathbf{x}^2, \mathbf{v}^1, \mathbf{v}^2) = \frac{1}{2}m_1|\mathbf{v}^1|^2 + \frac{1}{2}m_2|\mathbf{v}^2|^2 + \frac{Gm_1m_2}{|\mathbf{x}^1 - \mathbf{x}^2|}.$$

The essential feature of this system is that the potential term only depends on the *difference* between the particle positions and not on the positions themselves. To put it in a more portentous way, this means that there is no preferred origin of our coordinate system: the law of gravitation by which these particles interact doesn't depend whether they are in our solar system or far across the Milky Way. Now the mathematical expression of this invariance is simply that the system has symmetries as follows: for each  $j = 1, 2, 3$ , we can consider the map

$$F^a(\mathbf{x}^1, \mathbf{x}^2) = (\mathbf{x}^1 + a\mathbf{e}_j, \mathbf{x}^2 + a\mathbf{e}_j)$$

that translates each position vector by  $a$  units along the  $x_j$ -axis, in other words, just shifts the whole system in that direction. It is easy to see that this does not change the velocities, so  $dF^a = \text{Id}$ . So the Lagrangian in question enjoys the symmetry  $F^a$ . Now we find the conserved quantity

guaranteed by Noether. The infinitesimal symmetry is the vector field in  $\mathbb{R}^6 = \mathbb{R}^3 \times \mathbb{R}^3$  given by

$$\begin{aligned}\mathbf{W} &\equiv (\mathbf{W}^1, \mathbf{W}^2) \\ &= \frac{d}{da} F^a \Big|_{a=0}(\mathbf{x}^1, \mathbf{x}^2) \\ &= (\mathbf{e}_j, \mathbf{e}_j)\end{aligned}$$

i.e., consisting of the unit vector  $\mathbf{e}_j$  in each factor of  $\mathbb{R}^3$ . Thus the conserved quantity is

$$P \equiv \sum_{i=1}^3 L_{v_i^1} W_i^1 + L_{v_i^2} W_i^2 = m_1 v_j^1 + m_2 v_j^2$$

which is the *total momentum* in the  $x_j$ -direction. Thus, while individual momenta may not be conserved in this setting, the momentum of the whole system remains constant.

More generally, suppose we have  $N$  particles, each moving in  $\mathbb{R}^n$ . We will label the different particles by Greek letters, to avoid confusion with spatial dimensions, and let  $\mathbf{x}^\alpha$  denote the position of particle  $\alpha$  and  $m_\alpha$  its mass, where  $\alpha = 1, \dots, N$ . Thus,  $x_i^\alpha$  is the  $i$ 'th coordinate of the position of the  $\alpha$ 'th particle. Suppose now that these particles interact, and the Lagrangian for their interaction is given by a potential *that only depends on the differences in the positions of the particles*. Again, this means that there is no preferred origin of our coordinate system. Thus

$$L = \sum_{\alpha=1}^N \frac{1}{2} m_\alpha |\mathbf{v}^\alpha|^2 + V$$

where  $V = V(\mathbf{x}^\alpha - \mathbf{x}^\beta)$  for  $\alpha, \beta = 1, \dots, N$ . Thus, for each  $j = 1, \dots, n$  we have a symmetry

$$F^a(\mathbf{x}^1, \dots, \mathbf{x}^N) = (\mathbf{x}^1 + a\mathbf{e}_j, \dots, \mathbf{x}^N + a\mathbf{e}_j).$$

The infinitesimal symmetry is the vector field

$$\frac{d}{da} F^a \Big|_{a=0}(\mathbf{x}^1, \dots, \mathbf{x}^N) = (\mathbf{e}_j, \dots, \mathbf{e}_j)$$

in  $(\mathbb{R}^n)^N$  consisting of the unit vector  $\mathbf{e}_j$  in each slot. The conserved quantity is again total momentum:

$$P = \sum_{\alpha,j} L_{v_j^\alpha} W_{\alpha,j} = \sum_{\alpha=1}^N m_\alpha v_j^\alpha.$$

*Exercise 7.18.* Show that in a two particle system in  $\mathbb{R}^3$  with

$$L(\mathbf{x}^1, \mathbf{x}^2, \mathbf{v}^1, \mathbf{v}^2) = \frac{1}{2} m_1 |\mathbf{v}^1|^2 + \frac{1}{2} m_2 |\mathbf{v}^2|^2 - V$$

with a potential of the form

$$V = V(|\mathbf{x}^1 - \mathbf{x}^2|),$$

the total angular momentum of the system

$$m_1 \mathbf{x}^1 \times \dot{\mathbf{x}}^1 + m_2 \mathbf{x}^2 \times \dot{\mathbf{x}}^2$$

is conserved. (Cf. Exercise 7.17.)

## 8. TWO PARTICLE INTERACTIONS

A particularly useful invariant we found in Noether's theorem was the following: suppose that  $N$  particles are moving in  $\mathbb{R}^n$ , and interacting via a symmetry that only depends on the differences of their position vectors, with Lagrangian

$$L = \frac{1}{2} \sum m_\alpha |\mathbf{v}^\alpha|^2 - V, \quad V = V(\mathbf{x}^\alpha - \mathbf{x}^\beta).$$

Then all  $n$  components of the *total momentum*

$$\mathbf{P} = \sum m_\alpha \mathbf{v}_\alpha$$

are conserved. In particular, if we have two particles with

$$(53) \quad L = \frac{1}{2} m_1 |\mathbf{v}^1|^2 + \frac{1}{2} m_2 |\mathbf{v}^2|^2 + V(\mathbf{x}^1 - \mathbf{x}^2)$$

then  $\mathbf{P} = m_1 \mathbf{v}^1 + m_2 \mathbf{v}^2$  is a constant, in other words

$$\frac{d^2}{dt^2} (m_1 \mathbf{x}^1 + m_2 \mathbf{x}^2) = 0.$$

Thus, if we consider a fictitious particle with location  $m_1 \mathbf{x}^1 + m_2 \mathbf{x}^2$ , this particle moves freely, as if there were no potential influencing it. Note that if, for the sake of elegance, we rescale the location of this particle by the total mass and set

$$\mathbf{X} = \frac{1}{m_1 + m_2} (m_1 \mathbf{x}^1 + m_2 \mathbf{x}^2)$$

then  $\mathbf{X}$  is the average position of the two particles making up our system, weighted by their masses. Thus it has a special name: we call  $\mathbf{X}$  the *center of mass* of the system. Conservation of total momentum means that it moves at a fixed rate along a straight line.

Uniform motion of the center of mass is convenient, but no motion would be even more convenient! Motivated by this, we make a change of coordinates in order to separate out the (simple and predictable) motion of the center of mass and the (potentially complicated) motions of the particles relative to the center of mass. In doing this, we begin by noting that while  $\mathbf{x}^1, \mathbf{x}^2$  were very natural coordinates on  $\mathbb{R}^n \times \mathbb{R}^n$  representing the positions of our two particles, and giving the Lagrangian the nice form (53), there is no law that says that we cannot make a change of coordinates on  $\mathbb{R}^n \times \mathbb{R}^n$  that mixes up the two factors, and write the Lagrangian in these coordinates. Now  $\mathbf{X}$  turned out to be conserved under the motion, so it is very natural to take  $\mathbf{X}$  to be part of our new coordinate system on  $\mathbb{R}^n \times \mathbb{R}^n$ . This leaves  $n$  more coordinates to be specified, and many different choices would work

for the rest of these: we could for instance use  $\mathbf{x}^1$  or  $\mathbf{x}^2$ , since specifying  $\mathbf{x}^1$  (say) and  $\mathbf{X}$  would allow us to solve for  $\mathbf{x}^2$  and give an invertible change of coordinates. But this would break up the symmetry, such as it was, between the two particles, and seems a little un-natural. Why not instead use

$$\mathbf{y} \equiv \mathbf{x}^1 - \mathbf{x}^2,$$

since that is what occurs in our potential  $V$ ?

Thus, we use the change of coordinates

$$(54) \quad \begin{aligned} \mathbf{y} &= \mathbf{x}^1 - \mathbf{x}^2, \\ \mathbf{X} &= \frac{1}{m_1 + m_2} (m_1 \mathbf{x}_1 + m_2 \mathbf{x}_2) \end{aligned}$$

and we easily compute that we can then solve for  $\mathbf{x}^i$  in terms of these new coordinates by

$$(55) \quad \begin{aligned} \mathbf{x}^1 &= \mathbf{X} + \frac{m_2 \mathbf{y}}{M} \\ \mathbf{x}^2 &= \mathbf{X} - \frac{m_1 \mathbf{y}}{M}, \end{aligned}$$

where we have set

$$M = m_1 + m_2$$

to denote the total mass.

The virtue and clarity of the Lagrangian approach is that in order to compute the evolution of  $\mathbf{y}$  and  $\mathbf{X}$ , we need only rewrite the Lagrangian in these coordinates, and then compute the Euler-Lagrange equations. Equation (55) gives

$$\begin{aligned} L &= \frac{1}{2} m_1 |\dot{\mathbf{x}}^1|^2 + \frac{1}{2} m_2 |\dot{\mathbf{x}}^2|^2 + V(\mathbf{x}^1 - \mathbf{x}^2) \\ &= \frac{1}{2} m_1 (\dot{\mathbf{X}} + (m_2/M) \dot{\mathbf{y}})^2 + \frac{1}{2} (\dot{\mathbf{X}} - (m_1/M) \dot{\mathbf{y}})^2 - V(\mathbf{y}) \\ &= \frac{m_1 m_2}{2M} |\dot{\mathbf{y}}|^2 + \frac{1}{2} M \dot{\mathbf{X}}^2 - V(\mathbf{y}). \end{aligned}$$

As the Lagrangian has no explicit dependence on  $\mathbf{X}$ , we easily recover the fact that

$$\frac{\partial L}{\partial \dot{\mathbf{X}}} = M \dot{\mathbf{X}}$$

is conserved: this is just the conservation of total momentum that we observed above. Meanwhile, the motion of  $\mathbf{y}$  now decouples completely from the linear motion of  $\mathbf{X}$ : it satisfies the Euler-Lagrange equations

$$\frac{m_1 m_2}{2M} \ddot{\mathbf{y}} = -\nabla V(\mathbf{y}).$$

In other words, *the variable  $\mathbf{y}$  evolves as if it represented the motion of a single particle in  $\mathbb{R}^n$ , with potential  $V(\mathbf{y})$  and mass<sup>22</sup>  $m_1 m_2 / M$* . Consequently, we have reduced the problem of two-particle motion with a potential of

---

<sup>22</sup>This important quantity is often known as *reduced mass*.

the form  $V(\mathbf{x}^1 - \mathbf{x}^2)$  to the problem of one-particle motion in the potential  $V(\mathbf{y})$ : the difference vector between the particle positions evolves according to motion in this potential, while the center of mass moves linearly. If you want to return to the original particle positions, it's moreover easy to do so: by (55),

$$\begin{aligned}\mathbf{x}^1 - \mathbf{X} &= \frac{m_2 \mathbf{y}}{M} \\ \mathbf{x}^2 - \mathbf{X} &= -\frac{m_1 \mathbf{y}}{M},\end{aligned}$$

so in fact if we subtract off the center of mass (i.e., use it as our origin of coordinates), the individual particle motions become scaled versions of the motion of  $\mathbf{y}$ , hence scaled versions of one particle motion in the potential  $V$ .

We began with more than two particles, so you might wonder how this generalizes. The good news is that, as you will show in the following exercise, if we have, say, three particles with a potential depending on differences of positions, we can always eliminate one particle from the problem by introducing center of mass coordinates.

## 9. RADIAL POTENTIALS

We now consider motion in a potential of the form  $V(|\mathbf{x}|)$  depending only on the *norm* of a vector. The central example to keep in mind is that of Newtonian gravity, with

$$V = -\frac{Gm_1m_2}{|\mathbf{x}|},$$

but we will postpone the analysis of this special case until after analyzing the features of the more general class of radial potentials. Owing to the discussion of the previous section, our analysis will be applicable to two different situations:

- (1) A single particle at position  $\mathbf{x}$ , affected by the potential  $V(|\mathbf{x}|)$ . For instance, we might consider the earth orbiting around the sun: to a very good approximation, the sun's location is not affected by the gravitational pull of the earth, so we may consider the sun to be fixed at the origin of coordinates and the earth to move in the potential

$$V = -\frac{Gm_{\text{sun}}m_{\text{earth}}}{|\mathbf{x}|},$$

- (2) Two particles at positions  $\mathbf{x}_1$  and  $\mathbf{x}_2$ , with masses  $m_1, m_2$ , interacting via a potential

$$V(|\mathbf{x}_1 - \mathbf{x}_2|).$$

Then by the results of §8, this problem is equivalent to the study of the motion of a single particle of mass

$$\frac{m_1m_2}{m_1 + m_2}$$

at position  $\mathbf{x} = \mathbf{x}_1 - \mathbf{x}_2$ . This case would thus include a more honest account of the earth-sun system, in which the earth does affect the sun's motion. In the case of the earth and sun, however, we have

$$m_{\text{earth}} = 6.0 \times 10^{24} \text{ kg}, \quad m_{\text{sun}} = 2.0 \times 10^{30} \text{ kg}.$$

Hence the effective mass

$$\frac{m_{\text{earth}}m_{\text{sun}}}{m_{\text{earth}} + m_{\text{sun}}}$$

is extremely close to  $m_{\text{earth}}$ , hence the Lagrangian for  $\mathbf{x} \equiv \mathbf{x}_1 - \mathbf{x}_2$  is very well approximated by the one-body Lagrangian

$$L = \frac{1}{2}m_{\text{earth}}|\dot{\mathbf{x}}|^2 + \frac{Gm_{\text{earth}}m_{\text{sun}}}{|\mathbf{x}|}.$$

Bearing in mind, then, that our analysis will deal with both one-particle motion in a potential and two-particle interactions, we will specialize to three dimensions, and will analyze the motion of a particle in  $\mathbb{R}^3$  with Lagrangian

$$L = \frac{1}{2}|\mathbf{v}|^2 - V(|\mathbf{x}|)$$

where  $V(r)$  is a function on  $(0, \infty)$ . Our first observations regard conserved quantities for the motion. As the Lagrangian is time-independent, we have seen that energy is conserved, hence

$$H = \frac{1}{2}|\mathbf{v}|^2 + V(|\mathbf{x}|)$$

is conserved. Furthermore, the rotational symmetry, about any axis we choose, yields conservation of the total angular momentum

$$\mathbf{L} = \mathbf{x} \times \mathbf{v}.$$

Now for any given initial conditions  $\mathbf{x}_0, \mathbf{v}_0$  for position and velocity, we have

$$\mathbf{L} = \mathbf{x} \times \mathbf{v} = \mathbf{x}_0 \times \mathbf{v}_0$$

for all time. Hence in particular the vector  $\mathbf{L}$  remains perpendicular to the vector  $\mathbf{x}$  for all time: this means that *the motion is entirely in the plane  $\mathbf{L}^\perp$  of vectors perpendicular to  $\mathbf{L}$* . (If  $\mathbf{L}$  happens to be the zero vector, there is a bit of a special case: the motion must in fact be along a line through the origin.)

By rotating our coordinate system, we may take  $\mathbf{L} = x_3$ , so that the motion is in fact in the  $x_1 - x_2$  plane, with  $x_3(t) = 0$  for all  $t$ . Thus, we have reduced the problem of radial potentials in  $\mathbb{R}^3$  to that in  $\mathbb{R}^2$ : it suffices to study the Lagrangian

$$L = \frac{1}{2}|\mathbf{v}|^2 - V(|\mathbf{x}|)$$

in  $\mathbb{R}^2$ . The conserved angular momentum is now the quantity

$$\ell = x_1v_2 - x_2v_1$$

and the conserved energy is

$$H = \frac{1}{2}(v_1^2 + v_2^2) + V(|(x_1, x_2)|)$$

At this point it is instructive to switch to polar coordinates—recall that the ease with which we are able to do this is one of the attractive features of the Lagrangian formalism. If we have

$$(x_1(t), x_2(t)) = (r(t) \cos \varphi(t), r(t) \sin \varphi(t))$$

then differentiating both sides yields

$$\begin{aligned} (\dot{x}_1(t), \dot{x}_2(t)) &= (\dot{r}(t) \cos \varphi(t) - r(t) \sin \varphi(t) \dot{\varphi}(t), \dot{r}(t) \sin \varphi(t) + r(t) \cos \varphi(t) \dot{\varphi}(t)) \\ &\quad \dot{r}(t)(\cos \varphi(t), \sin \varphi(t)) + r(t) \dot{\varphi}(t)(-\sin \varphi(t), \cos \varphi(t)) \end{aligned}$$

Thus if  $\gamma(t) = (x_1(t), x_2(t))$  is any path, we found (Exercise 7.4) that

$$|\dot{\gamma}(t)|^2 = \dot{r}(t)^2 + r(t)^2 \dot{\varphi}(t)^2.$$

Hence we may write our Lagrangian in polar coordinates as

$$(56) \quad L(\gamma(t), \dot{\gamma}(t)) = \frac{1}{2}(\dot{r}(t)^2 + r(t)^2 \dot{\varphi}(t)^2) - V(r(t)).$$

It is now easy to find the Euler-Lagrange equations in  $r, \varphi$ . Since  $\varphi(t)$  does not even appear in the equation, the latter is exceptionally easy: we have

$$\frac{d}{dt}(r(t)^2 \dot{\varphi}(t)) = 0.$$

In other words, this yields the conserved quantity

$$r(t)^2 \dot{\varphi}(t).$$

Is this news? Not really: the absence of  $\varphi$  from the Lagrangian just expressed the rotational symmetry of the problem, which we had already exploited via Noether's theorem: this quantity is just the angular momentum yet again, i.e.,

$$(57) \quad \ell = r(t)^2 \dot{\varphi}(t).$$

*Exercise 9.1.* Check via explicit computation that

$$r(t)^2 \dot{\varphi}(t) = x_1 \dot{x}_2 - x_2 \dot{x}_1.$$

Now, we could write down the second Euler-Lagrange equation, the one in  $r$ . But we are not going to bother! Instead, we note that there is a second conserved quantity guaranteed us, owing to the time-invariance of the Lagrangian, and that is that Hamiltonian. This comes out to be

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

This equation can now be drastically simplified by using our other conservation law (57) to substitute  $\dot{\varphi} = -\ell/r^2$  and eliminate  $\varphi$  from the equation altogether. This procedure yields

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

As in Section 7.7, an illuminating feature of this Hamiltonian is that it is once again exactly the same that we would have got from considering the one-dimensional motion of a particle in a potential that not quite given by  $V(r)$  but by the modified “effective” potential

$$V_{\text{eff}}(r) \equiv V(r) + \frac{\ell^2}{2r^2}.$$

As in Section 7.7, we see that with this choice of effective potential, our original Euler–Lagrange equation for  $r$  is equivalent to the one for the one-dimensional problem with Lagrangian

$$L = \frac{1}{2}\dot{r}^2 - V_{\text{eff}}(r),$$

and hence the equation of motion is

$$(58) \quad \ddot{r}(t) = -V'_{\text{eff}}(r).$$

and the conserved energy is

$$(59) \quad H = \frac{1}{2}\dot{r}^2 + V_{\text{eff}}(r),$$

*Exercise 9.2.* Write down the Euler–Lagrange equation in  $r$  arising from (56) and check that it does agree with the motion (58).

The analysis of our problem of central force motion is now considerably simplified. Having reduced to the problem of motion in the plane, we may compute the conserved quantities  $H, \ell$  from the initial position and velocities of a particle. Then we can solve (58) as usual by solving (59) for  $\dot{r}$  and integrating

$$\frac{dt}{dr} = \frac{1}{\sqrt{2H - 2V_{\text{eff}}(r)}},$$

to yield

$$t(r) = \int \frac{1}{\sqrt{2H - 2V_{\text{eff}}(r)}} dt.$$

In principle we can then invert this equation to obtain  $r(t)$ . If we are lucky we may be able to do this integral exactly, but more generally it is quite easy to compute numerically, so we declare victory at this point: the problem is reduced to simply computing an integral.<sup>23</sup>

<sup>23</sup>That this constitutes success may take a little getting used to if you like exact formulae, but contrast this with the problem we would have been faced with in solving for the motion of a particle in some less symmetric potential in  $\mathbb{R}^3$ : we have to solve three completely coupled second order differential equations of the form (9). One can try to do

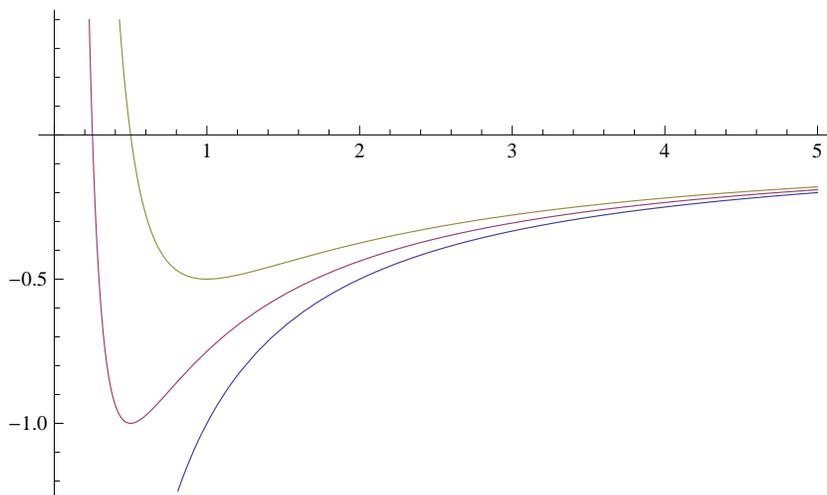


FIGURE 6. Effective potential  $-1/r + \frac{\ell^2}{2r^2}$  for  $\ell^2 = 0, 1/2, 1$ .

Even without solving explicitly, we note that we can read excellent qualitative information off from our effective potential. For instance, Figure 6 shows the effective potentials in the case of Newtonian gravity,  $V(r) = 1/r$ , with various values of  $\ell$ . When  $\ell = 0$  we easily see that any particle that starts moving in the direction of decreasing  $r$  will fall to  $r = 0$ , while a particle that starts moving to the right will escape to infinity if and only if  $H > 0$ —otherwise, it turns around and falls into the potential well. This case is extremely special, as any non-zero value of angular momentum results in  $1/r^2$  growth of the potential at  $r = 0$ , hence no particle with nonzero angular momentum can fall to the origin. When  $\ell \neq 0$  the effective potential does have a minimum, which we easily compute to be at  $r_0 = \ell^2$ . This corresponds to a stable equilibrium solution of the equation for  $r(t)$  at  $r = r_0$ . While  $r = r_0$  is fixed,  $\theta$  is of course varying along this orbit with  $r_0^2 \dot{\theta} = \ell$ . Hence  $\theta = \theta_0 + t/\ell$  increases uniformly in time, and we easily see that this is a stable circular orbit.

*Exercise 9.3.* Let  $V(r) = -1/r^2$ . Sketch the effective potential in this case for various values of  $\ell$ . When, if ever, do there exist stable circular orbits? Can orbits with nonzero angular momentum fall into the potential singularity?

Let us now suppose for the moment that we are interested in the motion of a particle that is bounded, i.e. for which the sub-level set  $\{r : V_{\text{eff}}(r) \leq H\}$  is a bounded set. Thus, the particle is in orbit around the central potential,

---

this on a computer, but even there it is an immensely more difficult problem, prone to numerical instabilities if not handled with considerable care. Finding qualitative features of the motion, as we will do below for specific choices of radial potential, is in general out of the question.

with  $r$  oscillating back and forth between maximum and minimum values. We will now pursue the question of the shape of this orbit.

Note that in general, for a non-circular orbit, it is not so easy to solve for  $\varphi$ : we have

$$\dot{\varphi} = \ell/r(t)^2,$$

so we can only hope to integrate this equation and solve for  $\varphi$  once we have found  $r(t)$ . But we remark that it is always the case that  $\dot{\varphi} > 0$ , so it makes sense to try and eliminate  $t$  from consideration and use  $\varphi$  as a parameter along the orbit instead, hence finding (by abuse of notation) the function  $r(\varphi)$  giving the shape of the orbit. This is easily done in principle at least: we have

$$\begin{aligned} \frac{dr}{d\varphi} &= \frac{dr}{dt} \frac{dt}{d\varphi} \\ &= \sqrt{2H - 2V_{\text{eff}}(r)} \frac{r^2}{\ell} \end{aligned}$$

In other words, separating variables and integrating yields

$$\int \frac{\ell}{r^2 \sqrt{2H - 2V_{\text{eff}}(r)}} dt = \varphi - \varphi_0,$$

hence we can in principle now invert to find  $r(\varphi)$ .

What, then, does an orbit look like? Let us consider the time it takes for  $r$  to pass from its maximum value  $r_+$  along its trajectory down to its minimum  $r_0$ , and back up to its maximum  $r_+$ . Let  $\tau$  denote this time. After time  $\tau$ , the  $\varphi$  variable has been incremented by some amount,

$$\begin{aligned} \Delta\varphi &= \int_0^\tau \dot{\varphi} dt \\ &= \int_0^\tau \ell r(t)^{-2} dt. \end{aligned}$$

If it happens to be the case that  $\Delta\varphi = 2\pi$  then the trajectory is periodic, as  $r(\tau) = r(0)$ ,  $\varphi(\tau) = \varphi(0)$  have returned to their initial values, while  $\dot{r}(\tau) = 0$  (since  $r$  is at its maximum) and  $\dot{\varphi} = \ell/r_+^2$  at both  $t = 0$  and  $t = \tau$ . Thus, the whole system evolves periodically, with  $r$  and  $\varphi$  returning to their initial values after every increment of time  $\tau$ . If, as might seem more likely,  $\Delta\varphi \neq 2\pi$ , then this is not in general the case: the orbits will not close up.

Figure 7 shows three particle trajectories, each with initial conditions

$$\mathbf{x}(0) = (1, 0), \quad \dot{\mathbf{x}}(0) = (0, 1/2),$$

and for  $t \in [0, 30]$ , but in three different central potentials:  $V(\mathbf{x}) = |\mathbf{x}|^p$  with  $p = 0.9, 1, 1.1$ . Note the remarkable fact that the orbit with  $p = 1$ —the case of Newtonian gravity—does appear to be periodic. We will demonstrate this periodicity in the next section.

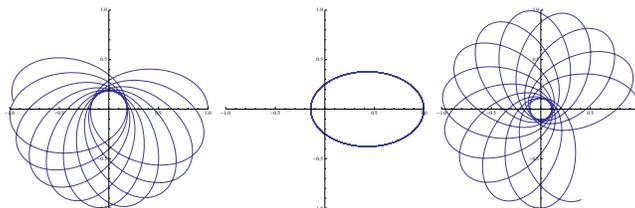


FIGURE 7. Three pictures of motion with the same initial conditions,

$$\mathbf{x}(0) = (1, 0), \quad \dot{\mathbf{x}}(0) = (0, 1/2)$$

in the potentials  $V(\mathbf{x}) = |\mathbf{x}|^p$  with  $p = 0.9, 1, 1.1$  respectively, with  $t \in [0, 30]$ .

*Exercise 9.4.* In each of the three cases shown in Figure 7, use the given initial conditions to find the values of the parameters  $H$  and  $\ell$  occurring in the solution of the motion.

## 10. THE KEPLER PROBLEM

In this section, we discuss the special case of the “Kepler problem,” which is to say, the problem of the motion of a particle in a bound orbit under the influence of Newtonian gravity:  $V = 1/r$ . Newton’s solution of this problem, deriving Kepler’s laws of planetary motion from the inverse square law of force (first published in his *Principia* in 1687) surely constitutes one of the great triumphs of human reasoning.

In the case of Newtonian gravity, we have

$$(60) \quad V_{\text{eff}}(r) = -\frac{1}{r} + \frac{\ell^2}{2r^2},$$

hence we are faced with the integral<sup>24</sup>

$$\begin{aligned} \varphi &= \int \frac{\ell}{r^2 \sqrt{2H - 2V_{\text{eff}}(r)}} dr \\ &= \ell \int \frac{dr}{r^2 \sqrt{2H + 2/r - \ell^2/r^2}}. \end{aligned}$$

To deal with this integral, do not panic: it is much improved by the substitution  $u = 1/r$ . It then becomes

$$-\ell \int \frac{du}{\sqrt{2H + 2u - \ell^2 u^2}}.$$

<sup>24</sup>We will take  $\varphi_0 = 0$  by rotating our coordinate system as needed.

Completing the square in the denominator, we may write it as

$$\begin{aligned} -\ell \int \frac{du}{\sqrt{2H + 2u - \ell^2 u^2}} &= - \int \frac{du}{\sqrt{2H/\ell^2 + 2u/\ell^2 - u^2}} \\ &= - \int \frac{du}{\sqrt{1/\ell^4 + 2H/\ell^2 - (u - 1/\ell^2)^2}}. \end{aligned}$$

The integral

$$\int \frac{du}{\sqrt{A^2 - (u - B)^2}}$$

is easily evaluated by the substitution  $u - B = A \sin \rho$ : we come out with

$$\arcsin((u - B)/A).$$

Thus, recalling that  $u = r^{-1}$ , we have (retaining the notation  $A = (1/\ell^4 + 2H/\ell^2)^{1/2}$ )

$$\varphi = - \arcsin((r^{-1} - \ell^{-2})/A)$$

i.e.,

$$r^{-1} = \left(\frac{1}{\ell^2} - A \sin \varphi\right).$$

This in principle solves our problem of analyzing the shape of the orbit: we have found the orbit as the polar coordinate graph of the function  $r(\varphi) = (\frac{1}{\ell^2} - A \sin \varphi)^{-1}$ . To see what kind of curve this is, though, we return to rectangular coordinates: we have

$$\begin{aligned} 1 &= \frac{r}{\ell^2} - Ar \sin \varphi \\ &= \frac{(x_1^2 + x_2^2)^{1/2}}{\ell^2} - Ax_2, \end{aligned}$$

i.e.,

$$\frac{(x_1^2 + x_2^2)^{1/2}}{\ell^2} = Ax_2 + 1.$$

Squaring and rearranging gives

$$\ell^{-4}x_1^2 + (\ell^{-4} - A^2)x_2^2 - 2Ax_2 = 1.$$

Recalling that  $A^2 = \ell^{-4} + 2H/\ell^2$  we can rewrite it as

$$(61) \quad \ell^{-4}x_1^2 - 2H\ell^{-2}x_2^2 - 2Ax_2 = 1.$$

As this is quadratic in  $x_1, x_2$  we recognize it as some kind of conic section, i.e., an ellipse, a parabola, or a hyperbola. Now recall that for bound orbits, we have  $H < 0$ , since these are the orbits for which the energy is less than the limit at infinity of the effective potential. In this case, the coefficients of  $x_1^2$  and  $x_2^2$  are both positive, and we recognize the curve as an ellipse. For the unbound situation when  $H > 0$ , we likewise have a hyperbola, and in the special case when  $H = 0$ , then the equation simply reads

$$2Ax_2 = -1 + \ell^{-4}x_1^2,$$

which is manifestly the equation of a parabola.

The elliptical character of bound orbits is Kepler's first law. Kepler's second law says that the orbits sweep out equal areas in equal times. By writing the area of a graph in polar coordinates as

$$\frac{1}{2} \int r(\varphi)^2 d\varphi = \frac{1}{2} \int r(t)^2 \dot{\varphi}(t) dt,$$

we see that the area swept out in time  $t$  is simply  $\ell t$  and the assertion is just that of conservation of angular momentum. We note that this law, unlike the first, has nothing to do with the specific potential  $V(r) = -1/r$  and would in fact hold for any central potential  $V(r)$ , since conservation of angular momentum applies in that generality. Finally, Kepler's third law says that the square of the orbit period is proportional to the cube of the semi-major axis of the ellipse.

*Exercise 10.1.* In this problem, we prove Kepler's third law.

Recall that we have found that orbits are given by

$$(\star) \quad \ell^{-4} x_1^2 - 2H\ell^{-2} x_2^2 - 2Ax_2 = 1.$$

Here

$$\ell = r^2 \dot{\varphi}$$

is the conserved angular momentum,

$$(\dagger) \quad H = \frac{1}{2} \dot{r}^2 + \frac{\ell^2}{2r^2} - \frac{1}{r}$$

is the conserved Hamiltonian, and we write

$$A = \left( \frac{1}{\ell^4} + \frac{2H}{\ell^2} \right)^{1/2}.$$

- (1) Show that in order for the equation  $(\dagger)$  to have solutions for some pairs  $(r, \dot{r})$  we must have

$$H \geq -\frac{1}{2\ell^2}.$$

(Hint: show that  $H$  must be as least as big as the minimum of the effective potential; also remember here and below that  $H < 0$  on bound orbits.)

- (2) Recall that the area of a figure in the plane given in polar coordinates by  $0 \leq r \leq r(\varphi)$ ,  $\varphi \in [0, 2\pi]$  is equal to

$$\int_0^{2\pi} \frac{1}{2} r(\varphi)^2 d\varphi;$$

use this to show that the period of the motion is given by

$$\tau = \frac{2E}{\ell},$$

where  $E$  is the area of the ellipse in  $(\star)$ .

- (3) Show that you can rewrite the equation ( $\star$ ) as

$$\ell^{-4}x_1^2 - 2H\ell^{-2}(x_2 - B)^2 = -\frac{1}{2H\ell^2}$$

for some value  $B$ . (Hint: complete the square in  $x_2$ , and then substitute in the definition of  $A$  after rearranging.)

- (4) Show that the two axes of the ellipse (longest and shortest diameters) are  $2a$  and  $2b$  with

$$a = \frac{\ell}{\sqrt{-2H}}, \quad b = \frac{1}{-2H}.$$

- (5) Use the result of part 1 to show that the larger of these two diameters is in fact  $2b$ , hence *semimajor axis*, that is, half the larger diameter, is  $b$ .
- (6) Recalling<sup>25</sup> that the area of an ellipse with semiaxes  $a, b$  is  $E = \pi ab$ , show that the ratio of the square of the period and the cube of the semimajor axis is proportional to  $a^2/\ell^2 b$  with  $a, b$  as above.
- (7) Compute this quantity in terms of  $H, \ell$  to show that in fact the ratio of the square of the period and the cube of the semimajor axis is a constant, *independent of  $H, \ell$* ; in other words, this ratio is the same for every orbit. This is Kepler's Third Law.

*Exercise 10.2.* In this extended exercise, we explore a different theory of the motion of a particle under the influence of a large gravitating body. This is the solution afforded by the theory of General Relativity (“GR”).

In GR, the Lagrangian for a body attracted by a star of mass  $M$  is given in spherical coordinates by the Lagrangian<sup>26</sup>

$$(62) \quad \frac{1}{2}((1 - 2M/r)^{-1}\dot{r}^2 + r^2(\dot{\theta}^2 + \sin^2\theta\dot{\varphi}^2) - (1 - 2M/r)^{-1}).$$

Note that the term  $r^2(\dot{\theta}^2 + \sin^2\theta\dot{\varphi}^2)$  is just the angular part of what we would get from expressing the usual free particle Lagrangian  $(1/2)|\dot{\mathbf{x}}|^2$  in spherical coordinates, but the radial part is quite different. Note also that this expression differs from its counterpart in Newtonian gravity not just by a change of potential  $(1 - 2M/r)^{-1}$  but also by a change in the kinetic energy term, just as happened when we studied motion constrained to a surface in  $\mathbb{R}^3$ . This corresponds to the famous “warping of space” which GR tells us is the mechanism by which the star affects the orbit of other particles. The Lagrangian (62) is easily seen to blow up rather disagreeably at  $r = 2M$ ;

<sup>25</sup>You can get this from the area of the circle using the change of variables formula for integration in  $\mathbb{R}^2$ , by using a coordinate transformation to stretch the circle to the ellipse.

<sup>26</sup>We are suppressing here the crucial complication that the interpretation of the “time” variable in GR is interesting and subtle. But we will focus on the shape of the orbit and not on its time-parametrization, so this defect in our treatment is inessential. In a fuller treatment, there would indeed be no potential term in this expression, but one involving  $\dot{t}$ , the derivative of the “time” coordinate with respect to a parameter along the geodesic (“proper time”).

in one interpretation this is supposed to be well inside the star, in a region where different physics is occurring owing to the stuff the star is made of, and hence not relevant to planetary motion.<sup>27</sup> Or, alternatively, you can think of this Lagrangian as describing the exterior of a black hole;  $r = 2M$  is the “event horizon.”

- (1) Show that angular momentum is conserved, just as in Euclidean space, so that just as in the case of central force motion we may take the motion of the particle to be in the plane  $\theta = \pi/2$ , with the quantity

$$\ell = r^2 \dot{\varphi}$$

conserved.

- (2) Since the Lagrangian is time independent, show that the quantity

$$H = \frac{1}{2} \left( (1 - 2M/r)^{-1} \dot{r}^2 + \frac{\ell^2}{r^2} + (1 - 2M/r)^{-1} \right)$$

is conserved.

- (3) Write this conservation law in the form

$$\frac{1}{2} \dot{r}^2 + V_{\text{eff}}(r) = E$$

with  $E$  a constant and

$$V_{\text{eff}}(r) = \frac{2MH}{r} + \frac{\ell^2}{2r^2} - \frac{\ell^2 M}{r^3}$$

an effective potential.

- (4) Graph  $V_{\text{eff}}(r)$  for various values of  $\ell$  and  $H$ . Compare this effective potential to the one in the Newtonian case (60). Show that there are many orbits with  $\ell \neq 0$  that tend toward the origin  $r = 0$ .
- (5) Describe all circular orbits of the particle by finding critical points of the effective potential. Are these orbits stable or unstable?

## 11. HAMILTONIAN MECHANICS

**11.1. Hamilton’s equations of motion.** We have seen the for a time independent Lagrangian

$$L = L(\mathbf{x}, \mathbf{v}),$$

there is automatically a conserved quantity, the *Hamiltonian*

$$H = H(\mathbf{x}, \mathbf{v}) = \sum_j L_{v_j}(\mathbf{x}, \mathbf{v}) v_j - L(\mathbf{x}, \mathbf{v}).$$

In the case

$$L = \frac{1}{2} |\mathbf{v}|^2 - V(\mathbf{x})$$

---

<sup>27</sup>This was Schwarzschild’s original notion.

this yielded the usual law of the conservation of energy, with

$$H = \frac{1}{2}|\mathbf{v}|^2 + V(\mathbf{x}).$$

When there were other conserved quantities, e.g. those we found by applying Noether's theorem, we found it more convenient to write down and analyze the equations of conservation, which were first order equations, than to deal with the second-order equations given by the Euler-Lagrange equations associated to  $L$  (a.k.a. Newton's law). Recall that in the special case when  $x_i$  was a cyclic variable, i.e., when  $\partial L/\partial x_i = 0$  then the quantity

$$p_i \equiv \frac{\partial L}{\partial v_i}$$

was conserved under the time-evolution of the system. In a simple example like a particle in  $\mathbb{R}^n$  in a potential, written in rectangular coordinates, we have

$$L = \frac{1}{2}m|\mathbf{v}|^2 - V(\mathbf{x})$$

and of course

$$p_i = mv_i,$$

so these are not so different from the velocities: the constant factor  $m$  is not so exciting. But in a slightly more exotic example, for instance as simple as a free particle in  $\mathbb{R}^2$  but with the Lagrangian expressed in polar coordinates, we have

$$L = \frac{1}{2}m(v_r^2 + r^2v_\varphi^2)$$

(Here we have written  $v_r$  and  $v_\varphi$  as the names of the variables into which  $\dot{r}$  and  $\dot{\varphi}$  will be inserted, for consistency of notation with the previous example.) Then the quantities

$$p_r = \frac{\partial L}{\partial v_r} = mv_r$$

$$p_\varphi = \frac{\partial L}{\partial v_\varphi} = mr^2v_\varphi$$

are a little more interesting: the first is not conserved under the evolution, even though it looks much like a momentum; by contrast the second one looks less like linear momentum but *is conserved* (it is of course angular momentum).

Recall that the expression  $p_i = L_{v_i}$  in fact occurs throughout Lagrangian mechanics: the Euler-Lagrange equations are nothing but

$$\dot{p}_i = L_{x_i}$$

of course, and the conserved quantities in Noether's theorem are likewise constructed as

$$\sum p_i W_i.$$

Thus, these are in many ways more useful functions than  $v_i$  to think about in general, so we now propose to change coordinates in our study of the

equations of motion, from  $x_i, v_i$  to  $x_i, p_i$ . Since it is confusing to reuse the same name for a coordinate when it is part of two different coordinate systems, we will rename our  $x$  variables as  $q$  when they are part of a coordinate system with the  $p$ 's. Thus, our change of coordinates is

$$(x_1, \dots, x_n, v_1, \dots, v_n) \rightarrow (q_1 = x_1, \dots, q_n = x_n, p_1 = L_{v_1}(\mathbf{x}, \mathbf{v}, t), \dots, p_n = L_{v_n}(\mathbf{x}, \mathbf{v}, t))$$

The  $p_i$ 's are important enough to have a special name: they are called “generalized momenta.” (Of course if  $L$  is kinetic plus potential energy in rectangular coordinates, they are just ordinary momenta.) Note that the inverse coordinate transformation is a map taking

$$(\mathbf{q}, \mathbf{p}) \mapsto (\mathbf{x} = \mathbf{q}, \mathbf{v} = \mathbf{v}(\mathbf{q}, \mathbf{p})),$$

i.e. it's of course trivial to compute the space components, but the velocities are subtler.

As long as we are replacing variables with ones that might arise as conserved quantities, it pays to consider  $H$ , the Hamiltonian, which is the conserved quantity that we get if  $L$  is independent of  $t$ : we have

$$H = \sum_{j=1}^n L_{v_j} v_j - L = \sum p_j v_j - L.$$

The latter expression for  $H$  is a little suspect, as it's simultaneously using coordinates from the  $(x, v)$  and  $(q, p)$  coordinate systems, so it might be clearer to write it just in  $(q, p)$  coordinates as

$$H = \sum p_j v_j(\mathbf{q}, \mathbf{p}) - L(\mathbf{x}(\mathbf{q}, \mathbf{p}), \mathbf{v}(\mathbf{q}, \mathbf{p}), t).$$

By the chain rule we can compute its partial derivatives in our new coordinate system. For brevity of notation, however, we will suppress the  $\gamma$  notation. We find that

$$\begin{aligned} \frac{\partial H}{\partial p_i} &= v_i + \sum_j p_j \frac{\partial v_j}{\partial p_i} - \sum_j L_{v_j} \frac{\partial v_j}{\partial p_i} - \sum L_{x_j} \frac{\partial x_j}{\partial p_i} \\ &= v_i + \sum_j p_j \frac{\partial v_j}{\partial p_i} - \sum_j L_{v_j} \frac{\partial v_j}{\partial p_i} \\ &= v_i \end{aligned}$$

where we have used the fact that  $\partial x_j / \partial p_i$  is zero since  $x_j = q_j$  is independent of the  $p$  variables. We also compute

$$\begin{aligned} \frac{\partial H}{\partial q_i} &= \sum p_j \frac{\partial v_j}{\partial q_i} - \sum_j L_{v_j} \frac{\partial v_j}{\partial q_i} - \sum L_{x_j} \frac{\partial x_j}{\partial q_i} \\ &= \sum p_j \frac{\partial v_j}{\partial q_i} - \sum_j p_j \frac{\partial v_j}{\partial q_i} - L_{x_i} \\ &= -L_{x_i} \end{aligned}$$

Now, finally, we specialize this computation to the situation where  $\mathbf{x} = \mathbf{x}(t)$  is the trajectory of a physical particle and  $\mathbf{v} = \dot{\mathbf{x}}$  is its time derivative, so that  $\mathbf{x}, \mathbf{v}$  satisfy the equations

$$(63) \quad \dot{x}_i = v_i,$$

$$(64) \quad L_{x_i}(\mathbf{x}, \mathbf{v}, t) = \frac{d}{dt} L_{v_i}(\mathbf{x}, \mathbf{v}, t),$$

with the second set being of course the Euler-Lagrange equations. We can now write  $\mathbf{q}(t) = \mathbf{x}(t)$  and  $\mathbf{p}(t) = \mathbf{p}(\mathbf{x}(t), \mathbf{v}(t))$  for the corresponding values of  $\mathbf{q}$  and  $\mathbf{p}$  along the trajectory. This improves the appearance of the right-hand-side of (64) which becomes  $\dot{p}_i$ ; similarly, we can recognize  $v_i$  as  $\dot{q}_i$  in our new coordinates.

Rewriting our equations for the partial derivatives of  $H$  then yields:

$$\begin{aligned} \frac{\partial H}{\partial p_i} &= \dot{q}_i \\ \frac{\partial H}{\partial q_i} &= -\dot{p}_i \end{aligned}$$

Turning these equations around we may regard them as equations

$$(65) \quad \begin{aligned} \dot{q}_i &= \frac{\partial H}{\partial p_i}, \\ \dot{p}_i &= -\frac{\partial H}{\partial q_i}. \end{aligned}$$

specifying the derivatives of  $\mathbf{q}, \mathbf{p}$  in terms of  $\mathbf{q}, \mathbf{p}$ , i.e. as a system of first-order ODEs for these  $2n$  variables. These equations, known as *Hamilton's equations of motion* thus express the time evolution of position and generalized momentum in a very efficient and pleasingly symmetrical way. As soon as we know  $H$ , we can write down this first order system of equations and use *it* as the equations of motion rather than the second-order Euler-Lagrange equations.

Let us now compute these equations of motion in some simple examples. We will begin with the motion of a particle of mass  $m$  in a one-dimensional potential. Then

$$L = \frac{1}{2}mv^2 - V(x)$$

and

$$p = L_v = mv.$$

The conserved energy is of course

$$H = \frac{1}{2}mv^2 + V(x),$$

and rewriting it in terms of the new variables ( $q = x, p = mv$ ) gives

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

Then Hamilton's equations of motion are

$$\dot{q} = m^{-1}p, \quad \dot{p} = -V'(q).$$

If we take  $V(q) = (1/2)kq^2$ , the harmonic oscillator potential, then the equations read

$$\dot{q} = m^{-1}p, \quad \dot{p} = -kq.$$

It is especially nice notation to bundle together  $(q, p)$  into a single vector and write

$$\frac{d}{dt} \begin{pmatrix} q \\ p \end{pmatrix} = \begin{pmatrix} m^{-1}p \\ -kq \end{pmatrix}$$

If  $k = m = 1$  we recognize the right-hand side  $(p, -q)$  as a *rotation vector field* in the  $(q, p)$ -plane.

A more sophisticated example is to take the Kepler problem in polar coordinates, as above. Then as above we have

$$L = \frac{1}{2}(v_r^2 + r^2v_\varphi^2) + \frac{1}{r}$$

hence as we saw before,

$$p_r = v_r, \quad p_\varphi = r^2v_\varphi$$

and

$$H = \frac{1}{2}(v_r^2 + r^2v_\varphi^2) - \frac{1}{r} = \frac{p_r^2}{2} + \frac{p_\varphi^2}{2q_r^2} - \frac{1}{q_r}.$$

Here to be completely pedantic, we have renamed the  $r$  variable  $q_r$  and will also rename  $\varphi$  as  $q_\varphi$  below, to emphasize that we have changed coordinate systems from  $r, \varphi, v_r, v_\varphi$  to  $q_r = r, q_\varphi = \varphi, p_r = v_r, p_\varphi = r^2v_\varphi$ . (In future we'll stop doing this silly renaming of space variables, at least most of the time.) Note that  $p_\varphi$  is just what we called  $\ell$ , the angular momentum, in our earlier treatment of the Kepler problem. Now Hamilton's equations read:

$$\begin{aligned} \dot{q}_r &= p_r, \\ \dot{p}_r &= \frac{p_\varphi^2}{q_r^3} - \frac{1}{q_r^2}, \\ \dot{q}_\varphi &= \frac{p_\varphi}{q_r^2}, \\ \dot{p}_\varphi &= 0 \end{aligned}$$

Note that since the last equation tells us that  $p_\varphi$  is constant, the first two equations then stand on their own as a system of two coupled equations, reflecting that fact that  $r$  (a.k.a.  $q_r$ ) is moving in a one-dimensional effective potential.

Note that it is not so amazing that the equations of motion can be written as a first order system of equations in  $2n$  variables: indeed, we could always have set

$$v_i = \dot{x}_i$$

and regarded the equations

$$\begin{aligned}\dot{x}_i &= v_i, \\ \frac{d}{dt}L_{v_i} &= L_{x_i}\end{aligned}$$

as a system of equations for the  $2n$  variables  $x, v$  that is first order. Indeed, this is pretty much what we did in doing phase plane analyses in one dimension, where we tracked the values of  $x, v$  simultaneously. It just turns out that  $(\mathbf{q}, \mathbf{p})$  are *better* coordinates to use in the phase plane, for a variety of reasons.

**11.2. Vector fields and flows.** We will now start to take  $H(\mathbf{q}, \mathbf{p}, t)$  as the fundamental quantity to study rather than  $L(\mathbf{x}, \mathbf{v}, t)$ . We can determine  $H$  from  $L$  as discussed above, and it turns out that we can go back the other way, as well, although we will not pursue this point.

For a start, let us compute the derivative of  $H$  itself along the flow. We know that if  $L$  is independent of  $t$  then  $H$  ought to be conserved, but let us try and see this directly. We have

$$\frac{d}{dt}H(\mathbf{q}, \mathbf{p}, t) = \sum H_{q_i} \dot{q}_i + H_{p_i} \dot{p}_i + H_t$$

(with subscripts denoting partial derivatives) hence by Hamilton's equations of motion

$$\frac{d}{dt}H(\mathbf{q}, \mathbf{p}, t) = \sum (-p_i) \dot{q}_i + p_i \dot{p}_i + H_t = H_t.$$

We record this as a proposition:

**Proposition 11.1.**

$$\frac{d}{dt}H(\mathbf{q}, \mathbf{p}, t) = H_t(\mathbf{q}, \mathbf{p}, t).$$

*In particular, if  $H$  is independent of  $t$ , then the value of  $H$  is conserved along the flow.*

This is not earth-shattering news, but is at least quite reassuring.

Henceforth we will assume that  $H$  is independent of  $t$ . Thus the Hamiltonian is conserved under the time evolution. If we think of  $(\mathbf{q}, \mathbf{p})$  evolving in  $\mathbb{R}^n \times \mathbb{R}^n$  according to (65), then we know that it stays on a fixed level set of  $H$  (i.e., the set where  $H$  has a fixed, constant, value). Typically this is a *hypersurface*, i.e., a higher dimension generalization of a curve or surface having dimension  $2n - 1$  (rather than a curve's 1 dimension or a surface's 2).

In order to think some more about the motion in the  $(\mathbf{q}, \mathbf{p})$  phase space, we will change our perspective a little bit. First let us discuss some background on vector fields and ODEs, valid in greater generality.

Consider a vector field  $\mathbf{v}(\mathbf{x})$  in  $\mathbb{R}^N$ . Given such a vector field, we may consider the system of ODEs

$$\dot{\mathbf{x}} = \mathbf{v}(\mathbf{x})$$

which is of course just shorthand for the system of  $N$  equations

$$\dot{x}_i = v_i(x_1, \dots, x_N), \quad i = 1, \dots, N.$$

Conversely, if we have any system of ODEs

$$\dot{x}_i = f_i(\mathbf{x})$$

in which the RHS is independent of  $t$  (these are called *autonomous* ODEs) then we can make the right hand side into a vector field  $\mathbf{v} = (f_1, \dots, f_N)$  and the ODE is of the form we have considered above.

Now let us reinterpret our ODE system geometrically. Consider a solution curve  $\mathbf{x}(t)$  to the autonomous system of ODEs

$$(66) \quad \dot{\mathbf{x}} = \mathbf{v}(\mathbf{x}).$$

Then we as usual can compute for, any  $t$  and for  $\epsilon$  small,

$$\mathbf{x}(t + \epsilon) \sim \mathbf{x}(t) + \epsilon \dot{\mathbf{x}}(t) = \mathbf{x}(t) + \epsilon \mathbf{v}(\mathbf{x}(t)).$$

In other words,  $\mathbf{v}(t)$  is just the tangent to the parametrized curve  $\mathbf{x}(t)$ . Thus, solving the system of ODEs (66) with a given initial condition  $\mathbf{x}(0) = \mathbf{x}_0$  is equivalent to finding a curve in  $\mathbb{R}^N$  through this point that is everywhere tangent to the specified vector field  $\mathbf{v}$ . The solutions curves are called *integral curves* of the given vector field  $\mathbf{v}$ .

Now we return to mechanics. Here we are concerned specifically with  $\mathbb{R}^{2n} = \mathbb{R}^n \times \mathbb{R}^n$  with coordinates  $(\mathbf{q}, \mathbf{p})$ , and with the vector field

$$\mathbf{v}_H = \begin{pmatrix} \frac{\partial H}{\partial \mathbf{p}} \\ -\frac{\partial H}{\partial \mathbf{q}} \end{pmatrix}$$

which we write as shorthand for

$$\begin{pmatrix} \frac{\partial H}{\partial p_1} \\ \vdots \\ \frac{\partial H}{\partial p_n} \\ -\frac{\partial H}{\partial q_1} \\ \vdots \\ -\frac{\partial H}{\partial q_n} \end{pmatrix}$$

Then, continuing with our shorter notation, the vector equation

$$\frac{d}{dt} \begin{pmatrix} \mathbf{q} \\ \mathbf{p} \end{pmatrix} = \mathbf{v}_H$$

is just a fancier way of writing Hamilton's equations of motion (65).

As the vector field  $\mathbf{v}_H$  is of central importance in this business, it has a special name:

*Definition 11.2.* Let  $f(\mathbf{q}, \mathbf{p})$  be a function on  $\mathbb{R}^n \times \mathbb{R}^n$ . The *Hamilton vector field*  $\mathbf{v}_f$  associated to  $f$  is the vector field

$$\mathbf{v}_f = \begin{pmatrix} \frac{\partial f}{\partial \mathbf{p}} \\ -\frac{\partial f}{\partial \mathbf{q}} \end{pmatrix}.$$

*Example 11.3.* Consider

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

This is the Hamiltonian associated to the Lagrangian  $(1/2)v^2 - (1/2)x^2$ , and describes the harmonic oscillator. Then

$$\mathbf{v}_H = \begin{pmatrix} p \\ -q \end{pmatrix}.$$

This is a *rotation vector field* in the plane. We know that its integral curves must lie along sets where  $H$  is constant, and of course these are just circles centered at the origin.

So Hamiltonian mechanics is governed by the integral curves of the Hamilton vector field  $\mathbf{v}_H$ . These turns out to have various special properties. To understand them, we will make one further step in interpreting our ODE. Let us consider a point

$$(\mathbf{q}_0, \mathbf{p}_0) \in \mathbb{R}^n \times \mathbb{R}^n.$$

Then given  $t \in \mathbb{R}$ , there is a unique point  $(\mathbf{q}(t), \mathbf{p}(t))$  given by solving Hamilton's equations of motion with initial data  $(\mathbf{q}(0), \mathbf{p}(0)) = (\mathbf{q}_0, \mathbf{p}_0)$ . Let us call this point

$$\Phi_t(\mathbf{q}_0, \mathbf{p}_0).$$

In other words, we define a family of maps

$$\Phi_t : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^n \times \mathbb{R}^n$$

given by evolving a point in phase space for time  $t$  under the equations of motion.<sup>28</sup> We can in fact do this quite generally by starting with any vector field—it does not have to be a Hamilton vector field.

*Definition 11.4.* Given a vector field  $\mathbf{v}$ , we call the family of maps  $\Phi_t$  constructed by the procedure above the *flow* generated by the vector field. If  $\mathbf{v} = \mathbf{v}_H$  is a Hamilton vector field, we call it the *Hamilton flow* of the function  $H$ .

**Lemma 11.5.** *For all  $s, t \in \mathbb{R}$  we have*

$$\Phi_{t+s} = \Phi_t \circ \Phi_s.$$

*In particular,  $\Phi_0 = \text{Id}$  and  $\Phi_{-t} = \Phi_t^{-1}$  (so that these maps are necessarily invertible).*

---

<sup>28</sup>Note that this construction is not special to Hamilton vector fields: we can (and will) discuss the same map associated to solving any system of autonomous ODEs.

*Proof.* Fix any  $s \in \mathbb{R}$  and any point  $\mathbf{w}$ . Consider the trajectories

$$\gamma_1(t) = \Phi_t(\Phi_s(\mathbf{w})), \quad \gamma_2(t) = \Phi_{t+s}(\mathbf{w}).$$

We have  $\gamma_1(0) = \gamma_2(0) = \Phi_s(\mathbf{w})$ . Both paths also satisfy the ODE. Therefore they must be equal, by uniqueness of solutions to an initial value problem.  $\square$

In the particular example  $H = (1/2)p^2 + (1/2)q^2$  that we considered above, the map  $\Phi_t$  is just clockwise rotation of the plane by angle  $t$ . To see this, note that  $\dot{q} = p$ ,  $\dot{p} = -q$  implies that  $\ddot{q} = -q$  so that  $q = A \cos t + B \sin t$ , as usual; then  $\dot{p} = -q$  means that  $p = -A \sin t + B \cos t$ . Evaluating for the initial conditions  $(q_0, p_0)$  gives  $A = q_0$ ,  $B = p_0$ , so that in vector form we may write

$$\begin{pmatrix} q(t) \\ p(t) \end{pmatrix} = \begin{pmatrix} q_0 \cos t + p_0 \sin t \\ -q_0 \sin t + p_0 \cos t \end{pmatrix} = \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix} \begin{pmatrix} q_0 \\ p_0 \end{pmatrix}.$$

So we have for any  $t$ ,

$$\Phi_t(q, p) = \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix} \begin{pmatrix} q \\ p \end{pmatrix}.$$

The fact that  $\Phi_t$  ends up being a family of *linear* maps from  $\mathbb{R}^2$  to itself, i.e., is given by multiplication by a matrix, is very special to this example and should by no means be regarded as the typical behavior of a mechanical system.

The following is a useful property of the flow along a Hamilton vector field. It is known as *Liouville's theorem*

**Theorem 11.6.** *For any set  $\Omega \subset \mathbb{R}^n \times \mathbb{R}^n$  and any  $t \in \mathbb{R}$ ,*

$$\text{Vol}(\Phi_t(\Omega)) = \text{Vol}(\Omega).$$

We thus say that the flow  $\Phi_t$  is *volume-preserving*.

Note that if we instead decided that our phase space should have coordinates  $(\mathbf{x}, \mathbf{v})$  as before, then the phase space flow would not preserve the volume.

Liouville's theorem is the consequence of a more general theorem about when the flow along a vector field in  $\mathbb{R}^N$  is volume preserving: the criterion is that the vector field have vanishing divergence.

**Proposition 11.7.** *Let  $\Phi_t$  be the flow generated by the vector field  $\mathbf{v}$  in  $\mathbb{R}^N$ . If  $\nabla \cdot \mathbf{v} = 0$  then  $\Phi_t$  is volume-preserving.*

To see that this proposition proves Liouville's theorem, we need only compute the divergence of a Hamilton vector field. We have

$$\begin{aligned}\nabla \cdot \mathbf{v}_H &= \nabla \cdot \begin{pmatrix} \frac{\partial H}{\partial \mathbf{p}} \\ -\frac{\partial H}{\partial \mathbf{q}} \end{pmatrix} \\ &= \sum_{j=1}^n \frac{\partial}{\partial q_j} \left( \frac{\partial H}{\partial p_j} \right) + \sum_{j=1}^n \frac{\partial}{\partial p_j} \left( -\frac{\partial H}{\partial q_j} \right) \\ &= 0\end{aligned}$$

by the equality of mixed partial derivatives.

Thus, it remains to prove Proposition 11.7

*Proof.* It will of course suffice to show that for every  $t \in \mathbb{R}^N$  we have

$$\frac{d}{dt} \text{Vol } \Phi_t(\Omega) = 0 \text{ for all } \Omega \subset \mathbb{R}^N.$$

To begin with, we claim that it suffices to verify this claim *only for*  $t = 0$ . Indeed, if we know this for  $t = 0$ , then for any  $T \in \mathbb{R}$ , if we let  $\Omega' = \Phi_T(\Omega)$  we compute

$$\begin{aligned}\frac{d}{dt} \text{Vol } \Phi_t(\Omega) \Big|_{t=T} &= \lim_{h \rightarrow 0} \frac{\text{Vol } \Phi_{T+h}(\Omega) - \text{Vol } \Phi_T(\Omega)}{h} \\ &= \lim_{h \rightarrow 0} \frac{\text{Vol } \Phi_h(\Phi_T(\Omega)) - \text{Vol } \Phi_T(\Omega)}{h} \\ &= \lim_{h \rightarrow 0} \frac{\text{Vol } \Phi_h(\Omega') - \text{Vol } \Phi_0(\Omega')}{h} \\ &= \frac{d}{dt} \text{Vol } \Phi_t(\Omega') \Big|_{t=0}.\end{aligned}$$

Thus, the claim that the derivative vanishes for the set  $\Omega$  at time  $T$  follows from the claim that it vanishes for  $\Omega'$  at time 0.

So we have to compute, for arbitrary  $\Omega \subset \mathbb{R}^N$ ,

$$\frac{d}{dt} \text{Vol } \Phi_t(\Omega) \Big|_{t=0}$$

We may write the volume of a set as the integral of the function 1 over that set:

$$\text{Vol } \Phi_t(\Omega) = \int_{\Phi_t(\Omega)} 1 \, dq \, dp.$$

This suggests that we perform a change of variables, setting  $(\mathbf{q}, \mathbf{p}) = \Phi_t(\mathbf{q}', \mathbf{p}')$ ; thus,  $(\mathbf{q}, \mathbf{p}) \in \Phi_t(\Omega)$  exactly when  $(\mathbf{q}', \mathbf{p}') \in \Omega$ , and we may rewrite

$$\text{Vol } \Phi_t(\Omega) = \int_{\Omega} 1 \det d\Phi_t(\mathbf{q}', \mathbf{p}') \, dq' \, dp'.$$

Here the determinant  $\det d\Phi_t(\mathbf{q}', \mathbf{p}')$  is, crucially, the Jacobian determinant resulting from the change of variables. Technically we ought to have the absolute value of the determinant here, but we remark that for  $t = 0$   $\Phi_t$  is the identity map, hence  $d\Phi_0$  has determinant 1. Since the determinant of

$d\Phi_t$  changes continuously as  $t$  evolves, then certainly it remains positive for small<sup>29</sup>  $t$ .

Thus, we are now interesting in showing that the quantity

$$\frac{d}{dt} \text{Vol } \Phi_t(\Omega)|_{t=0} = \int_{\Omega} 1 \frac{d}{dt} \det d\Phi_t(\mathbf{q}', \mathbf{p}')|_{t=0} dq' dp'$$

vanishes. It thus suffices to show that

$$\frac{d}{dt} \det d\Phi_t(\mathbf{q}', \mathbf{p}')|_{t=0} = 0.$$

To do this, we appeal to a lemma about the time-derivative of a family of matrices: note that if

$$A(t) = \begin{pmatrix} \lambda_1(t) & 0 & \dots & 0 \\ 0 & \lambda_2(t) & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & \lambda_N(t) \end{pmatrix}$$

is a family of matrices depending on  $t$  that happens to be diagonal, then we have

$$\log \det A(t) = \sum \log \lambda_j(t)$$

so that

$$\frac{d}{dt} \log \det A(t) = \sum \frac{\dot{\lambda}_j(t)}{\lambda_j(t)} = \text{Tr } \dot{A}(t)A^{-1}(t).$$

This turns out to be true more generally, whenever  $A(t)$  is a family of invertible matrices, depending on a parameter  $t$  (Lemma 11.8). Consequently, we have

$$\frac{(d/dt) \det A(t)}{\det A(t)} = \text{Tr } \dot{A}(t)A^{-1}(t),$$

i.e.,

$$(d/dt) \log \det A(t) = \text{Tr } \dot{A}(t)A^{-1}(t).$$

Now specializing to  $A(t) = d\Phi_t$  we have

$$\frac{d}{dt} \det d\Phi_t(\mathbf{q}', \mathbf{p}')|_{t=0} = \det(d\Phi_0) \text{Tr}(d/dt)(d\Phi_t)|_{t=0}(d\Phi_0)^{-1}.$$

Since  $\Phi_0 = \text{Id}$ , we certainly have  $d\Phi_0 = \text{Id}$ , hence we simply obtain

$$\frac{d}{dt} \det d\Phi_t(\mathbf{q}', \mathbf{p}')|_{t=0} = \text{Tr}(d/dt)(d\Phi_t)|_{t=0}$$

---

<sup>29</sup>In fact it must always remain positive, as  $\Phi_t$  is always invertible, so that the determinant of its derivative can never pass through the value 0.

We can commute time and space derivatives to rewrite

$$\begin{aligned}
 \frac{d}{dt}(d\Phi_t)|_{t=0} &= d((d/dt)\Phi_t|_{t=0}) \\
 &= d \begin{pmatrix} v_1(\mathbf{x}) \\ \vdots \\ v_N(\mathbf{x}) \end{pmatrix} \\
 (67) \qquad &= \begin{pmatrix} \frac{\partial v_1}{\partial x_1} & \cdots & \frac{\partial v_1}{\partial x_N} \\ \vdots & \ddots & \vdots \\ \frac{\partial v_N}{\partial x_1} & \cdots & \frac{\partial v_N}{\partial x_N} \end{pmatrix}
 \end{aligned}$$

Hence

$$\begin{aligned}
 \text{Tr} \frac{d}{dt}(d\Phi_t)|_{t=0} &= \sum \frac{\partial v_i}{\partial x_i} \\
 &= \nabla \cdot \mathbf{v} \\
 &= 0
 \end{aligned}$$

by hypothesis. □

**Lemma 11.8.** *If*

$$A(t) = \begin{pmatrix} a_{11}(t) & \cdots & a_{1n}(t) \\ \vdots & \ddots & \vdots \\ a_{n1}(t) & \cdots & a_{nn}(t) \end{pmatrix}$$

*is a family of invertible square matrices with time-varying coefficients, then*

$$\frac{d}{dt} \log \det A(t) = \text{Tr} A'(t)A(t)^{-1}.$$

We leave the proof as an exercise, sketched as follows:

*Exercise 11.1.* Prove Lemma 11.8 by completing the following outline:

- (1) Let  $P$  be an invertible square matrix and  $Q$  another matrix. Let  $h \in \mathbb{R}$ . Show that

$$\det(P + hQ) = \det P \det(I + hQP^{-1}).$$

- (2) If  $B$  is any square matrix, show that

$$\det(I + hB) = 1 + h \text{Tr} B + O(h^2)$$

where  $O(h^2)$  is shorthand for terms with powers  $h^2$  and higher.<sup>30</sup>

HINT: You can proceed one of three ways. Option 1: try to write the determinant out as a sum of products of entries and collect all the terms in the determinant that end up just having the power  $h^0$  or  $h^1$  (rather than a higher power of  $h$ ). Option 2: if  $B$  is diagonalizable, you can write it as  $C\Lambda C^{-1}$  where  $\Lambda$  is diagonal. This approach can be made rigorous by noting that diagonalizable matrices are dense

---

<sup>30</sup>More generally, this means “a term  $f(h)$  such that  $|f(h)| \leq \text{const} \cdot h^2$  for  $h$  small.”

in all matrices, but we'll leave that step for another class; you may assume  $B$  is diagonalizable. Option 3 (probably the best one): write the thing you're trying to compute in terms of the *characteristic polynomial* of  $B$  and remember what you learned in linear algebra.

(3) Show that

$$\det(P + hQ) = \det P(1 + hQP^{-1}).$$

(4) Using the fact that  $A(t + h) = A(t) + hA'(t) + O(h^2)$ , complete the proof.

### 11.3. Poincaré recurrence.

11.4. **Poisson brackets.** Let  $a(\mathbf{q}, \mathbf{p})$  be a function on phase space  $\mathbb{R}^n \times \mathbb{R}^n$ . We sometimes call such a function an *observable*, as it is something that we may observe by doing a physical experiment on the state of a particle. We may consider the time-derivative of  $a$  along the time-evolution of a system governed by the Hamiltonian function  $H(\mathbf{q}, \mathbf{p})$ . We then have (using Hamilton's equations of motion for the second equality below):

$$\begin{aligned} \frac{da}{dt} &= \sum_{i=1}^n \frac{\partial a}{\partial q_i} \dot{q}_i + \frac{\partial a}{\partial p_i} \dot{p}_i \\ (68) \qquad &= \sum_{i=1}^n \frac{\partial a}{\partial q_i} \frac{\partial H}{\partial p_i} - \frac{\partial a}{\partial p_i} \frac{\partial H}{\partial q_i}. \end{aligned}$$

The form of this interesting expression leads to the following more general definition.

*Definition 11.9.* For  $f, g \in C^\infty(\mathbb{R}^n \times \mathbb{R}^n)$  we define the *Poisson bracket* of  $f$  and  $g$  as the new function

$$\{f, g\} = \sum_{i=1}^n \frac{\partial f}{\partial q_i} \frac{\partial g}{\partial p_i} - \frac{\partial f}{\partial p_i} \frac{\partial g}{\partial q_i}.$$

Then (68) is telling us that

$$\frac{da}{dt} = \{a, H\}.$$

Another, pretty much equivalent, way to think of the Poisson bracket is in terms of the Hamilton vector field:

**Proposition 11.10.**  $\{f, g\} = (\mathbf{v}_g \cdot \nabla)f$ .

This simply follows directly from the definition, or, if you prefer, from the fact that the time derivative along the flow generated by the Hamiltonian function  $g$  is simply the directional derivative along the vector field  $\mathbf{v}_g$ .

**Proposition 11.11.** *The Poisson bracket is bilinear and anti-symmetric, i.e. for  $c_i \in \mathbb{R}$  we have*

$$\{c_1 f_1 + c_2 f_2, g\} = c_1 \{f_1, g\} + c_2 \{f_2, g\},$$

$$\{f, c_1g_1 + c_2g_2\} = c_1\{f, g_1\} + c_2\{f, g_2\},$$

and

$$\{f, g\} = -\{g, f\}.$$

The anti-symmetry of the Poisson bracket is perhaps rather surprising: it tells us that if we reverse the roles of the two functions  $a$  and  $H$  considering the evolution of the observable  $H$  along the flow generated by the Hamiltonian function  $a$ , we get minus the time derivative of  $a$  along the flow generated by  $H$ !

We remark now that  $a$  is a *conserved quantity* for the system with Hamiltonian  $H$  if and only if  $\dot{a} = 0$ , i.e., by our remarks above, exactly when

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

Thus, having zero Poisson bracket (“Poisson commuting”) with  $H$  is a key property. Somewhat surprisingly, perhaps, this property is preserved by Poisson bracket! To see this, we begin by proving an important result known as the *Jacobi identity*:

**Theorem 11.12.** For  $f, g, h \in C^\infty(\mathbb{R}^n \times \mathbb{R}^n)$ ,

$$(69) \quad \{\{f, g\}, h\} + \{\{g, h\}, f\} + \{\{h, f\}, g\} = 0.$$

The way to remember this is that the expression is simply triple Poisson bracket of  $f, g, h$  plus all its cyclic permutations in which we cycle through  $f, g, h$  preserving the order, with the “cyclicity” convention that  $f$  comes after  $h$ .

*Proof.* Examining the form of the general term in the LHS of (69) we note that it is of the form

$$f''g'h', \quad f'g''h', \quad \text{or} \quad f'g'h'',$$

i.e., involves exactly two derivatives falling on one of the functions and one derivative on each of the others. As the expression is symmetric in  $f, g, h$ , it clearly suffices to show that the sum of terms of the form  $f'g'h''$  vanish. This simplifies matters considerably as the first term on the LHS of (69) then does not contribute at all, and we need to collect the  $h''$  terms in

$$\left\{ \sum_i g_{q_i} h_{p_i} - g_{p_i} h_{q_i}, f \right\} + \left\{ \sum_i h_{q_i} f_{p_i} - h_{p_i} f_{q_i}, g \right\};$$

this gives

$$\begin{aligned}
& \sum_{i,j} g_{q_i} h_{p_i q_j} f_{p_j} - g_{p_i} h_{q_i q_j} f_{p_j} - g_{q_i} h_{p_i p_j} f_{q_j} + g_{p_i} h_{q_i p_j} f_{q_j} \\
& + \sum_{i,j} h_{q_i q_j} f_{p_i} g_{p_j} - h_{p_i q_j} f_{q_i} g_{p_j} - h_{q_i p_j} f_{p_i} g_{q_j} + h_{p_i p_j} f_{q_i} g_{q_j} \\
& = \sum_{i,j} h_{q_i q_j} (-g_{p_i} f_{p_j} + f_{p_i} g_{p_j}) + h_{p_i p_j} (-g_{q_i} f_{q_j} + f_{q_i} g_{q_j}) \\
& + \sum_{i,j} h_{q_i p_j} (g_{p_i} f_{q_j} - f_{p_i} g_{q_j}) + h_{q_j p_i} (g_{q_i} f_{p_j} - f_{q_i} g_{p_j}) \\
& = 0,
\end{aligned}$$

where we have of course used the equality of mixed partial derivatives.  $\square$

**Corollary 11.13.** *If  $\{f, H\} = 0$  and  $\{g, H\} = 0$  then  $\{\{f, g\}, H\} = 0$ . Hence the Poisson bracket of two conserved quantities is again conserved.*

The following exercise gives the nicest application of this result:

*Exercise 11.2.* For a system in  $\mathbb{R}^3 \times \mathbb{R}^3$ , let

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

denote the angular momentum. Show that if the first two components  $L_1, L_2$  are conserved quantities, then so is  $L_3$ .

**11.5. Symplectic structure.** Our description of Hamilton vector fields can be streamlined if we introduce a little more notation. We let  $J$  denote the  $2n \times 2n$  matrix

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

with  $I$  denoting the  $n \times n$  identity matrix and 0 denoting an  $n \times n$  block of zeroes. Then we note that given a function  $f(\mathbf{q}, \mathbf{p})$  on phase space we may write

$$\nabla f = \begin{pmatrix} \frac{\partial f}{\partial q_1} \\ \vdots \\ \frac{\partial f}{\partial q_n} \\ \frac{\partial f}{\partial p_1} \\ \vdots \\ \frac{\partial f}{\partial p_n} \end{pmatrix}$$

The following is then easy to compute:

**Lemma 11.14.** *The Hamilton vector field  $\mathbf{v}_f$  is given by  $J(\nabla f)$ .*

In other words, the Hamilton vector field of  $f$  is simply the gradient of  $f$  multiplied by the matrix  $J$ .

It is sometimes convenient to be able to change coordinates on phase space. If we do this, however, we may need to dramatically change our procedure for computing the Hamilton vector field. For instance, in one dimension, if we set  $(q', p') = (q^2, p)$  then the Hamilton vector field of the function  $q^2$  is  $(0, -2q)$  in the  $(q, p)$  coordinates, while if we compute in “primed” coordinates, we are just computing the Hamilton vector field of the function  $q'$ , which is  $(0, -1)$ . These vector fields do not agree, even if we now try to write them both in the same coordinate system.

Under what circumstances would this computation have worked out? First we have to remember how vector fields transform under differentiable maps: remember from our discussion of tangent vectors to curves in §7.9 that  $\Phi$  takes the vector  $\mathbf{v}$  to the vector

$$d\Phi \mathbf{v}$$

where  $d\Phi$  is the matrix of partial derivatives of the map  $\Phi$ . Now we want computing the Hamilton vector field and then transforming under  $\Phi$  to be the same as transforming under  $\Phi$  and then computing Hamilton vector field, i.e., we want

$$\begin{aligned} (d\Phi)(J(\nabla f)) &= J\nabla(f \circ \Phi^{-1}) \\ &= J((d\Phi)^{-1})^t(\nabla f). \end{aligned}$$

If this equation is to hold for *every*  $f$  it must in fact be the case that

$$(d\Phi)J = J((d\Phi)^{-1})^t$$

i.e., that for each  $(\mathbf{q}, \mathbf{p})$

$$(d\Phi(\mathbf{q}, \mathbf{p}))J(d\Phi(\Phi^{-1}(\mathbf{q}, \mathbf{p})))^t = J$$

*Definition 11.15.* A smooth transformation  $\Phi : \mathbb{R}^{2n} \rightarrow \mathbb{R}^{2n}$  is said to be a *symplectomorphism* if it has a smooth inverse, and if

$$(d\Phi)J(d\Phi)^t = J$$

The symplectomorphisms form a *group*:

**Proposition 11.16.** *The composition of two symplectomorphisms is again a symplectomorphism, and the inverse of a symplectomorphism is a symplectomorphism.*

*Exercise 11.3.* Prove the proposition.

Rather remarkably, the transformation of phase space effected by classical dynamics (i.e. generated by the flow along a Hamilton vector field) is *itself* a symplectomorphism:

**Theorem 11.17.** *Let  $\Phi_t$  be the flow generated by the Hamilton vector field  $\mathbf{v}_f$ . Then for all  $t$ ,  $\Phi_t$  is a symplectomorphism.*

## APPENDIX A. SEPARABLE FIRST ORDER DIFFERENTIAL EQUATIONS

A first order differential equation for a function  $x(t)$  that can be written in the form

$$\frac{dx}{dt} = f(x),$$

for some function  $f$ , is known as a separable equation. There is a simple procedure for “solving” these equations, which in practice reduces the problem to doing an indefinite integral and then taking the inverse function of a given function. This is often kind of messy in practice, but it can usually be done efficiently on a computer in practical situations where the computations become truly bad.

Here is how the procedure works in brief: we rewrite the equation as

$$\frac{dx}{f(x)} = dt$$

and integrate both sides to get  $t$  as a function of  $x$ . Then we solve the resulting equation for  $x$  in terms of  $t$ .

If this argument alarms you (and it probably should), a more honest but less memorable way of thinking of this process is to write the equation as

$$\frac{1}{f(x)} \frac{dx}{dt} = 1$$

and integrate both sides of *that*, making the substitution  $u = x(t)$ .

Here is a simple example: to solve

$$\frac{dx}{dt} = x^2$$

we rewrite the problem as

$$\frac{dx}{x^2} = dt,$$

hence integration of both sides yields

$$-\frac{1}{x} = t - C$$

(we have combined the constant of integration from both sides into a single constant which we have chosen to write as  $-C$ ). Hence

$$x = \frac{1}{C - t}$$

for some constant  $C$ .

## APPENDIX B. CONSTANT COEFFICIENT SECOND ORDER EQUATIONS

In this section we consider *constant coefficient linear homogeneous second order ODE*. What do these words mean? *Linear* means equations of the form

$$a(t)\ddot{x}(t) + b(t)\dot{x}(t) + c(t)x(t) = y(t),$$

so that the LHS depends *linearly* on the function  $x$ . *Constant coefficient* means that furthermore,  $a$ ,  $b$ , and  $c$  are independent of  $t$ , i.e., constant. And *homogeneous* means that  $y$  is 0. So we're left with

$$(70) \quad a\ddot{x}(t) + b\dot{x}(t) + cx(t) = 0$$

with  $a, b, c \in \mathbb{R}$ .

To solve, the best strategy is to make a good guess. In the special case of  $\ddot{x} - x = 0$  we see that  $e^{\pm t}$  are solutions. This may motivate trying a solution more generally of the form  $e^{\lambda t}$ . Plugging into (70) and factoring out  $e^{\lambda t}$  gives the equation

$$e^{\lambda t}(a\lambda^2 + b\lambda + c) = 0.$$

Since  $e^{\lambda t}$  is never 0 this must entail

$$a\lambda^2 + b\lambda + c = 0;$$

this last is called the *characteristic equation* and we can solve by the quadratic formula to get the (usually) two solutions

$$\lambda_{\pm} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

This then gives two different possible solutions to (70) of the form

$$e^{\lambda_+ t}, e^{\lambda_- t}.$$

Taking linear combinations of those two gives us the desired two-parameter family,

$$Ae^{\lambda_+ t} + Be^{\lambda_- t}$$

which, by Picard-Lindelöf, should be all the solutions there are.

There are two interesting wrinkles to this. The first is that if  $b^2 - 4ac < 0$  these are *complex* rather than real numbers: if  $\lambda = \alpha + i\beta$  with  $\alpha, \beta \in \mathbb{R}$  we get solutions of the form

$$e^{(\alpha+i\beta)t} = e^{\alpha t} e^{i\beta t}.$$

To understand these solutions better, recall that

$$e^{i\beta t} = \cos \beta t + i \sin \beta t,$$

so this factor is in fact an oscillating one. Also note that since  $a, b, c$  were taken to be real numbers, we can take the real or imaginary part of a solution of (70) to get another solution, so if we prefer we can write two different solutions

$$e^{\alpha t} \cos \beta t, e^{\alpha t} \sin \beta t.$$

*Exercise B.1.* Since the solutions to the characteristic equation  $a\lambda^2 + b\lambda + c$  should come in complex conjugate pairs, it looks like we should now get *four* possible solutions, two from  $\lambda_+$  and two from  $\lambda_-$ . Why is this not the case?

The second wrinkle, though, comes in the case where the equation

$$a\lambda^2 + b\lambda + c = 0$$

has a double-root. This happens when  $b^2 = 4ac$ , and then we only get *one* solution to (70) by this method, the solution

$$e^{\lambda t}$$

with  $\lambda = -b/2a$ . Now the surprising fact is that there is still an easy way to produce a second solution, but it's not obvious. The second solution is now given by  $te^{\lambda t}$ , and to see an explanation of that it's helpful to take a serious linear algebra course in which Jordan form and matrix exponentials are discussed.

*Exercise B.2.* Check that  $te^{\lambda t}$  is in fact a second solution of (70) in the case of a double root.

*Email address:* `jwunsch@math.northwestern.edu`