

Classical Mechanics

An Advanced Course

Nikhil Kalyanapuram

E-mail: contact@nkalyanapuram.com

Contents

1	Introduction	1
2	Newton's Laws and Equations of Motion	3
2.1	Frames of Reference and Galilean Symmetries	4
2.1.1	Mechanical Equivalence and Inertial Frames	5
2.1.2	Generalized Coordinates	6
2.2	The Euler-Lagrange Equations	6
2.2.1	The Free Particle	9
2.2.2	Interacting Particles	11
2.3	Noether's Theorem	13
3	Harmonic Oscillators	17
3.1	The Free Oscillator	17
3.1.1	The Harmonic Approximation	19
3.1.2	Normal Modes	20
3.2	The Damped Oscillator	21
3.3	Forced Oscillators	23
3.3.1	Green's Functions and General Forces	24
3.4	Anharmonic Oscillators and Perturbation Theory	27
4	Theory of Orbits	30
4.1	The Two-Body Problem	30
4.2	The Kepler Problem	32
4.2.1	Closed Orbits and Kepler's Laws	33
4.2.2	Open Orbits	36
4.3	The Eccentricity Vector	37
5	Rigid Body Dynamics	40
5.1	The Inertia Tensor	40
5.2	The Equations of Motion	43
5.2.1	The Euler Equations	43
5.2.2	Free Symmetric Tops	45
5.3	Euler Angles	46
5.3.1	Angular Velocity in the Body Frame	48
5.3.2	The Heavy Symmetric Top	49
5.4	The Free Asymmetric Top	52
5.5	Non-Inertial Reference Frames	57
5.5.1	Fictitious Forces	57

5.5.2	Rotating Frames	59
6	The Phase Space Formalism	61
6.1	The Hamiltonian	61
6.2	Poisson Brackets	63
6.2.1	Flows in Phase Space	63
6.2.2	Coordinate and Momentum Brackets	65
6.3	Canonical Transformations	66
6.3.1	The Symplectic Structure	66
6.3.2	Infinitesimal Canonical Transformations	68
6.4	Liouville's Theorem	70
A	Calculus of Variations	71
A.1	Minimizing Arc-Lengths	71
A.2	The Brachistochrone Curve	72
B	Fourier Analysis	74
B.1	The Fourier Transform	74
B.2	The Dirac Delta Function	76
B.3	Properties of the Fourier Transform	77
B.4	Analytic Aspects	78
C	Elliptic Integrals and Elliptic Functions	80
C.1	Elliptic Integrals	80
C.2	Applications of Elliptic Integrals	82
C.3	Elliptic Functions	83
C.4	Theta Functions	86
	Bibliographical Notes	89

1 Introduction

Classical mechanics is, as the name suggests, a mechanical theory to study physical phenomena. It involves the study of particles, their motions, and how the interactions of the particles affect these motions. The word mechanics should naturally invoke images of rigid rods, oscillating springs and the orbits of planets, but beyond that, it provides a unifying framework that casts all of these phenomena as different avatars of the same underlying laws.

The other half of the heading is even more important. Classical mechanics is a *classical* theory. It is meant to be applied to the classic problems of the natural world, and fails rather spectacularly for regimes that we normally do not *see*. The behavior of large and energetic astrophysical phenomena, the behavior of electrons and collisions of elementary particles naturally fall out of the scope of classical mechanics. Nevertheless, classical mechanics is an extremely rich subject, both mathematically and physically, and is one of the most consistent and solid frameworks modern physics has to offer.

In this monograph, we will be focused on studying classical mechanics as we have understood it due to the innovations introduced by Isaac Newton, especially calculus. Prior to Newton, there was a very rigorous and technical approach to mechanics characterized by detailed and baroque geometric arguments, that today would basically be a foreign language relative to how we do physics today. Personally, I find it interesting to occasionally revisit these old texts, but it would be cruel and unusual to inflict them on any modern student. References will be given for those interested.

Further, even after Newton, there have been developments to classical mechanics that are related to the Newtonian approach in nontrivial ways, largely driven by Joseph-Louis Lagrange and William Rowan Hamilton. These frameworks, now known as Lagrangian and Hamiltonian mechanics not only place Newtonian laws on firmer mathematical footing, but also provide a bridge between classical physics, and later advances in field theory and quantum mechanics. Accordingly, it *is* worth spending a fair bit of time on these.

The modern theory of gravity is due to Albert Einstein, known as the general theory of relativity, and is considerably beyond the scope of what we will study in the present text. However, a classical theory of gravity was developed by Newton, and can be studied within the framework of classical mechanics. Most familiar consequences of gravity, including planetary orbits, their precessions, the tides and projectile trajectories are described so well in the classical framework that it took over two centuries to discover that there was more behind the scenes.

Now for the things we will not be covering. In this volume we will not be studying the behavior of fluids and continua of particles. While these are described not only by classical equations of motion, but were actually studied by Newton himself, a detailed

study of the phenomenology involved would drag us down routes that will ultimately detract from the more universal phenomena we seek to study here. Additionally, we will not study large ensembles of particles in this book either. In fact, this actually happens to be outside of the scope of classical mechanics, requiring the more subtle concepts of statistical physics and kinetic theory to understand.

2 Newton’s Laws and Equations of Motion

At the heart of classical mechanics are three laws of motion due to Isaac Newton, which laid out for the first time truly universal principles describing the motion of bodies. Unfortunately, Newton himself presented these laws as, well, laws (or axioms of motion), and did not bother explaining why and how he arrived at them. With the benefit of hindsight and a few hundred years, we can do somewhat better. The goal of the present chapter then will be more convincing; and in doing so we will discuss Newton’s laws, as well as the broader concepts of dynamics needed to justify them well.

Prelude: Aristotelian Physics

Physics before Newton was rather different. Perhaps the greatest pre-Newtonian point of distinction was the fact that people were concerned less by *dynamics*, and more by *statics*. Where dynamics is the study of changes and causes of motion, statics is really the study of steady, stable systems, and the constraints that need to be satisfied to guarantee this stability. As a result of this, a predominant method of doing physics in those days was ordinary Euclidean geometry. So long as a complicated physical system at rest could be recast as a geometric diagram to scale, the ratios and relationships between the lengths of the diagram gave you those constraints.

In this light, Aristotelian physics occupies an uncomfortable departure, in that it attempts to make qualitative statements about dynamical aspects of mechanics, when the right tools to do so simply did not exist. As a result, the “laws of motion” had to be understood entirely empirically, and now appear rather strange. Put simply, there are two laws of motion due to Aristotle. The first of these is that objects move at a rate proportional to their weights, or,

$$v \propto W. \tag{2.1}$$

The second says that objects move at a rate inversely proportional to the density of the medium,

$$v \propto \frac{1}{\rho}. \tag{2.2}$$

That’s basically it. We now know that these are not quite right, but amusingly it’s actually the case that these laws *can* be derived, taking a very specific limit of Newton’s laws. In many ways, Newtonian mechanics is really understanding the implicit assumptions made in the Aristotelian world, and loosening these assumptions one by one to arrive at more fundamental rules. This highlights a key aspect of the progression of physics more generally; laws aren’t necessarily right or wrong so much as right or wrong within certain domains.

Two of the assumptions underlying Aristotelian mechanics are obvious; the first is that he assumes a natural *direction of motion*. The weight of bodies is naturally associated with their falling towards the earth, implying an assumed direction of motion towards the ground. Further, the second law assumes the existence of a medium in which the bodies fall. Newtonian mechanics begins where these assumptions end; henceforth, neither a preferred direction nor a preferred medium will be assumed.

2.1 Frames of Reference and Galilean Symmetries

Let's start with definitions. A particle is a body of dimension zero. At any given point in time, a particle occupies a single point in space. A choice of specifying the location of a particle is known as a coordinate system, or frame of reference. We often shorten this to simply *frame*. The purpose of equations of motion will be to specify the position of a particle (or a system of particles) in a frame of reference as a function of time, which will be denoted by the position vector $x_i(t)$. The time derivative of this quantity will be the *velocity* $v_i(t)$.

Two different frames of reference can be related by coordinate transformations. For instance, if a particle is at $(0, 1, 0)$ in frame A , it would be at $(0, 0, 0)$ in frame B , obtained by translating frame A by 1 unit along the y -axis. Given that we can transform our frames of reference, the first step is to understand which of those transformations are physically meaningful and which are not.

The average adult can throw a baseball at around 75 miles per hour, or 33 meters per second. Ignoring gravity for the time being, suppose I were to hurl a baseball in an empty field. Now say I threw the ball in a different direction. Changing your direction of throwing the ball is simply performing a rotation of its velocity, namely

$$v_i \longrightarrow S_{ij}v_j \tag{2.3}$$

where the matrix S satisfies

$$SS^T = S^TS = 1. \tag{2.4}$$

Physics really isn't concerned in which direction I choose to do this, and the trajectory of the ball is going to look the same regardless. The technical word for this is *isotropy*, which means, roughly, the same in all directions. The same is true if I simply move in some direction, and throw the ball again. Such a *translation* simply changes my initial location x_i to some new location according to

$$x_i \longrightarrow x_i + a_i \tag{2.5}$$

for constant a_i .

Now suppose I were to throw the ball again, but this time on a ship, moving steady on the water. It turns out that, so long as the velocity of the ship does not

change, either in terms of its direction or its magnitude, the laws of mechanics do not change. In technical terms, this means that the transformations

$$x_i \longrightarrow x_i + v_i t \tag{2.6}$$

for constant v_i do not change the laws of mechanics.

It's not hard to see that these transformations: rotations, translations and boosts (the technical term for tacking on a constant velocity) form a *group*. In other words, composing any of these (or any combination) will yield something that is once again a rotation, translation, boost, or some combination of the three. This group is known as the Galilean group. The study of classical mechanics is simply the study of the laws of physics that do not change under transformations due to the Galilean group.

Of the three Galilean transformations, it was the third that challenges pre-Newtonian intuition the most. As we pointed out earlier, Aristotle implicitly assumed that there was both a preferred direction of velocity, as well as a preferred *frame* relative to which the velocity is being measured. The first of these is easier to give up than the second. The idea that nature doesn't prefer any specific direction seems almost natural, but the idea that nature doesn't prefer any velocity is less so.

What this means is that a statement regarding the velocity of a particle without specifying what the velocity is relative to is meaningless. On the other hand, statements about acceleration are meaningful regardless of the choice of inertial frame. Seeing this isn't hard. Say I were to claim a particle is moving with a velocity u_i without specifying the frame. A boost into another frame can be used to transform this into whatever velocity I like. On the other hand, suppose I were to say that the body is not accelerating in a frame (let's call it A), amounting to the statement

$$\ddot{x}_i = 0. \tag{2.7}$$

We now make a Galilean boost into another frame A' , effecting the transform $x_i \longrightarrow x_i + v_i t = x'_i$. Rather trivially, we still find that

$$\ddot{x}'_i = 0. \tag{2.8}$$

This hints at what the laws of motion ought to talk about. Laws of motion that make statements about just the velocity tell us nothing physically interesting; they need to make statements about the acceleration of particles to be physically meaningful.

2.1.1 Mechanical Equivalence and Inertial Frames

The principle of relativity under Galilean transformations is the statement that physics ought to remain the same when such a transformation is performed. But keep in mind that Galilean transformations really effect a shift from one coordinate system (or frame) to another. As a result, whether or not this statement of equivalence, which is known as mechanical equivalence, holds is dependent on whether or

not any such frame exists in the first place. If it turns out that it is never possible to pick a frame in which the laws of physics do not depend on location for instance, then all this would not be relevant.

As it turns out, such frames do in fact exist. In particular, there are indeed frames in which space is homogeneous, in that translations do not affect mechanical laws, and time is homogeneous as well. Such frames are called *inertial reference frames*. Any two inertial reference frames will be related by some combination of translations in space and time, rotations and boosts.

2.1.2 Generalized Coordinates

From time to time, we're going to be interested in studying not the motions of individual particles themselves, but in some other feature of the dynamics¹. A simple example of this is two-body systems, where we're not really interested in the trajectories of the individual bodies, but rather in the motion of the particles *relative to each other*. A simple choice of coordinates for such a system would be

$$\begin{aligned}\vec{r}_{\text{com}} &= \frac{m_1 \vec{r}_1 + m_2 \vec{r}_2}{m_1 + m_2} \\ \vec{r}_{\text{rel}} &= \vec{r}_1 - \vec{r}_2\end{aligned}\tag{2.9}$$

where we have switched to vector notation in the interest of clarity. In this specific case, it's clear that new choice of coordinates also transforms under the Galilean group in the same way. The lesson here is that, oftentimes, we may choose to characterize the trajectories of mechanical systems using *generalized coordinates*, rather than explicit coordinates of the individual particles. We will denote generalized coordinates and their velocities by q_i and $\dot{q}_i$ respectively. More quantitatively, if the positional coordinates of the particles are given by x_i , where $i = 1, 2, \dots, 3N$ (each of the N particles has an x, y, z position) for each, generalized coordinates may be any functions

$$q_i := q_i(x_1, \dots, x_{3N}, t).\tag{2.10}$$

In practice however, generalized functions tend to be simple, not very complicated, functions of the actual coordinates.

Accordingly, we will often phrase our laws of motion describing the trajectories $q_i(t)$, keeping in mind of course invariance under Galilean transformations.

2.2 The Euler-Lagrange Equations

We mentioned earlier that even though Newton characterized his laws of motion as axioms, we would derive them from first-principles. This wasn't entirely honest;

¹This is especially true in constrained systems, in which the motions of the particles may be subject to complicated conditions, but combinations of their positions may be more easily studied.

instead, what we will actually do is replace the three axioms with one axiom, albeit one that is less intuitive. This axiom is known as the *principle of least action*, of *Hamilton's principle*.

The principle of least action requires that for every dynamical system, there exists a quantity known as the *Lagrangian*. We denote it by L . The Lagrangian is a function of the generalized coordinates, generalized velocities and time. Supposing now that the system evolves between two sets $(q_i(t_1), \dot{q}_i(t_1))$ and $(q_i(t_2), \dot{q}_i(t_2))$, the trajectory is determined by extremizing the integral

$$S[q_i, \dot{q}_i, t] = \int_{t_1}^{t_2} L(q_i, \dot{q}_i, t) dt. \quad (2.11)$$

Originally, the principle of least action was inspired by the principle of least time due to Fermat, who argued that light always travels in a path that minimizes the time of travel. In a sense, the principle of least action is similar. Any trajectory, given two fixed end points, comes with a natural notion of “resistance” along that trajectory, given by the action. The trajectory picked by the system is one that is least resisted, thereby minimizing the action. This point of view will show up later when encountering quantum mechanics.

This optimization problem is then one that selects, among all possible trajectories respecting the boundary conditions, that trajectory which minimizes the action. A good place to start is to ask what happens to S when a given trajectory is perturbed, just a little bit. Since exploring the entire space of paths is hard, we’re going to have to be content with seeing how S varies slowly instead. Specifically, we’d like to look at the difference

$$\delta S = S[q_i + \delta q_i, \dot{q}_i + \delta \dot{q}_i, t] - S[q_i, \dot{q}_i, t]. \quad (2.12)$$

Since we fix our boundary conditions, we must have

$$\delta q_i(t_1) = \delta q_i(t_2) = 0. \quad (2.13)$$

Expanding to first order, we find

$$\delta S = \int_{t_1}^{t_2} \left(\frac{\partial L}{\partial q_i} \delta q_i + \frac{\partial L}{\partial \dot{q}_i} \delta \dot{q}_i \right) dt. \quad (2.14)$$

The second term is readily integrated by parts to yield

$$\delta S = \frac{\partial L}{\partial \dot{q}_i} \delta q_i \Big|_{t_1}^{t_2} + \int_{t_1}^{t_2} \left(\frac{\partial L}{\partial q_i} + \frac{d}{dt} \frac{\partial L}{\partial \dot{q}_i} \right) \delta q_i dt. \quad (2.15)$$

The first term identically vanishes due to the enforced boundary conditions. Regarding the second, the condition for the existence of an extremizing trajectory for S would be the vanishing of the variation δS around the extremizing trajectory. Since

this would hold for any small variation, the desired trajectory exists if and only if we have

$$\frac{\partial L}{\partial q_i} + \frac{d}{dt} \frac{\partial L}{\partial \dot{q}_i} = 0 \quad (2.16)$$

for each i . These equations, known as the *Euler-Lagrange* equations are the cornerstone of modern analytical mechanics. It turns out that any problem, including those with constraints, symmetries and so on, can be phrased in this language. The problem devolves upon finding the right Lagrangian, and solving the Euler-Lagrange equations will yield the dynamics of the system.

What are some elementary properties that the Lagrangian would have to satisfy? For one thing, if there is any internal symmetry of the problem, the Lagrangian would need to reflect this. In addition to this, we don't want the action to be unbounded from below. This sets limitations on the overall sign of the Lagrangian, which we will find useful in deriving the Lagrangian of certain simple systems.

Lagrangians aren't unique, but they are unique in the sense of belonging to an equivalence class. The equivalence class is defined by being related by a total derivative; two Lagrangians L and L' are equivalent if they are related according to

$$L' - L = \frac{df}{dt}. \quad (2.17)$$

It's easy to see why this is true; any such total derivative in time would simply change the action by a constant:

$$S \longrightarrow S + f(t_2) - f(t_1) \quad (2.18)$$

which does not affect the minimization problem.

We note now that there actually isn't a good axiomatic reason for why the principle of least action needs to hold. It could very well have been that the right time evolution of classical systems would be described by paths that did not admit of a Lagrangian description, or did not satisfy the principle of least action even if they did. And in fact, it turns out that the non-extremizing paths are physically relevant, just not classically². The ultimate justification for the principle of least action comes from understanding how classical physics arises out of a specific limit of quantum mechanics, and will have to wait.

Calculus of Variations

The right mathematical framework for understanding problems like the one discussed above is the calculus of variations. A full presentation of this could easily form a monograph in and of itself. More practically, the method begins with the notion of a

²It also turns out that there are bizarre theories of physics that do not admit of Lagrangian descriptions, but these are well out of the scope of any elementary exposition.

functional, which takes a function (it may be defined over some domain $D \subset \mathbb{R}$), and returns a real number. Generally, given a domain D , a functional would be defined as a map

$$F : C^2(D) \longrightarrow \mathbb{R}. \quad (2.19)$$

The problem of the calculus of variations is to find conditions on functions $f \in C^2(D)$ such that a given $F(f)$ is minimized. The Euler-Lagrange equations are one avatar of such a set of conditions.

A staggering number of problems—abstract and very concrete—can be phrased in the language of minimizing a functional. As we’ve already seen, this takes the form of minimizing an action problem in classical dynamics, but the use of variational calculus extends considerably beyond that. Historically, variational methods were used extensively to study minimum energy configurations of complex systems, such as molecules, for instance.

There are conditions under which a given variational problem may not be solvable. Rather, it’s more accurate to say that there are properties that have to be satisfied by the space of functions to ensure that the variational problem makes sense to begin with. For readers interested in formal problems of this kind, we have provided details in the references.

2.2.1 The Free Particle

An elementary illustration of how the Lagrangian is often determined entirely in terms of simple symmetry considerations is provided by the case of the free particle. Consider a single free particle (which by definition cannot be “interacting” with anything), under conditions of Galilean invariance. Spatial isotropy and invariance under translations tell us that the Lagrangian cannot depend on the position x_i of the particle, as any such function would introduce spatial anisotropy into the Lagrangian. Accordingly we have

$$L := L(\dot{x}_i). \quad (2.20)$$

Further, the condition of invariance under rotations tells us that the Lagrangian can only depend on the squared velocity. Note that the squared velocity is defined by

$$\dot{x}^2 = \dot{x}_i \delta_{ij} \dot{x}_j \quad (2.21)$$

in which form it’s easy to see invariance under rotations, which are orthogonal matrices. Finally, consider what happens when a small boost is performed; we do this by boosting the velocity according to

$$\dot{x}'_i = \dot{x}_i + \epsilon_i \quad (2.22)$$

where ϵ_i is an infinitesimal boost vector. The Lagrangian in the new reference frame becomes

$$L(\dot{x}'_i) = L(\dot{x}_i) + \epsilon_i \frac{\partial L}{\partial \dot{x}_i}. \quad (2.23)$$

For the two of these to be mechanically equivalent, we would require the second term to be a total time derivative. The simplest such choice is

$$\frac{\partial L}{\partial \dot{x}_i} = C_1 \dot{x}_i \quad (2.24)$$

which when integrated becomes

$$L(\dot{x}_i) = \frac{1}{2} C_1 \dot{x}^2 + C_2. \quad (2.25)$$

The second of these constants may be dropped without any loss of generality; it has no effect on the dynamical equations of motion. The constant C_1 , largely due to historical reasons, has been normalized differently, and used to define a quantity called the *mass*,

$$C_1 = m. \quad (2.26)$$

That the mass should be positive is required to ensure that the Lagrangian is not unbounded below. To understand what the mass measures, we need to understand how particles react in the presence of interactions, to be considered in the next section.

In the meantime, we'll compute the Euler-Lagrange equations for the free particle Lagrangian. Note that since the Lagrangian does not depend on x_i , we have

$$\ddot{x}_i = 0. \quad (2.27)$$

Recall that the Lagrangian was determined by making reference to the Galilean symmetries, under which the system was assumed to be invariant. Thus, a free particle cannot accelerate, under conditions of Galilean invariance. This leads us to an alternate definition of inertial frames, which if you recall, were frames in which space and time were isotropic and homogeneous. An inertial frame is simply a frame in which free particles do not accelerate, or, one in which Newton's first law holds.

It's worth looking at what the free particle Lagrangian looks like in different coordinate systems. In cartesian coordinates, we have

$$L = \frac{1}{2} m \dot{x}^2 + \frac{1}{2} m \dot{y}^2 + \frac{1}{2} m \dot{z}^2 \quad (2.28)$$

and in spherical coordinates we find

$$L = \frac{1}{2} m \dot{r}^2 + \frac{1}{2} m r^2 \dot{\theta}^2 + \frac{1}{2} m r^2 \sin^2 \theta \dot{\phi}^2. \quad (2.29)$$

The right choice of coordinate system often makes calculations considerably simpler. We will return to that particular section when considering various particular physical systems.

2.2.2 Interacting Particles

Interactions are only possible when we deal with several particles. In mechanical systems, we need multiple particles to introduce dynamics, since single particles are naturally defined in terms of the elementary kinetics described in the preceding system. A non-interacting system of many particles will naturally incorporate the same notions of spatial homogeneity, since any notion of proximity between the particles won't really impact the mechanics. In this case, the Lagrangian is a simple sum

$$L = \sum_a \frac{1}{2} m_a \dot{x}_a^2 \quad (2.30)$$

where the index a denotes the particles (each x_a may itself be a vector). Physically, interactions are characterized by the relative positions of the particles; particles far apart will interact less, for instance. This naturally positional character of interactions may be encapsulated by a function $U(\vec{x}_1, \vec{x}_2, \dots)$, modifying the Lagrangian according to

$$L = \sum_a \frac{1}{2} m_a \dot{x}_a^2 - U(x_a). \quad (2.31)$$

This is about as general as Lagrangians tend to get; we'll see that it leads to a surprisingly rich variety of physical problems.

The Euler-Lagrange equations for this Lagrangian now start to look familiar:

$$m_a \ddot{x}_a = - \frac{\partial U}{\partial x_a} \quad (2.32)$$

where we have one set of (potentially vector) equations for each choice of a . For instance, in 3 dimensions, we will have $3N$ equations for a system with N particles. By defining the *force* on the a -th particle as

$$F_a = - \frac{\partial U}{\partial x_a} \quad (2.33)$$

the Euler-Lagrange equation is known as Newton's second law of motion. Acceleration, in other words, is proportional to the applied force on a particle. The force itself is a derived quantity, obtained as the negative gradient of the potential function $U(x_a)$.

Most systems are translationally invariant, meaning that the physics tends to be invariant under general translations $\vec{x}_a \rightarrow \vec{x}_a + \vec{a}$ (where the constant translation

$\vec{a}$ is not to be confused with the index a). This implies that the potential does not depend on the positions themselves, but on the differences

$$x_{ab} = x_a - x_b. \quad (2.34)$$

An upshot of this is that we will always have

$$\frac{\partial U_{ab}}{\partial x_a} = -\frac{\partial U_{ab}}{\partial x_b} \quad (2.35)$$

for all $a \neq b$. Here, we have used U_{ab} to denote that part of the interaction that controls the interaction between particles a and b . The preceding relationship encodes what is known as Newton's third law of motion. In mechanics, when any two particles are interacting, the forces exerted by them on each other are equal in magnitude, and opposite in direction.

Constant and Linear Forces

An elementary illustration of the interaction term is to identify what it looks like when all the forces are constants. In other words we have

$$m\ddot{\vec{x}}_a = \vec{F}_a \quad (2.36)$$

implying that

$$\vec{F}_a = -\frac{\partial U}{\partial \vec{x}_a} \quad (2.37)$$

which can be immediately integrated to yield

$$U = -\sum_a \vec{F}_a \cdot \vec{x}_a. \quad (2.38)$$

When we have a single particle instead of a system of particles, this reduces down to the Lagrangian of a single particle in an external field

$$L = \frac{1}{2}m\dot{\vec{x}}^2 + \vec{F} \cdot \vec{x}. \quad (2.39)$$

It's easy to see that this actually breaks spatial isotropy; there is now a preferred direction, along the vector prescribed by the force acting on the particle.

In the discussions so far, we have naturally assumed that the mass term m is a constant. It turns out however to be meaningful to relax this assumption. Say our mass term is a function of time $m(t)$. In that case, we have

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{x}_a} = \frac{d(m(t)\dot{x}_a)}{dt} = \dot{m}(t)\dot{x}_a + m\ddot{x}_a. \quad (2.40)$$

This has the effect of modifying the force, but via a term that is linear in the velocity. As it happens, terms of this kind show up in dissipative systems, that aren't closed,

in the sense of being isolated from their backgrounds. Systems like this lose energy (a term we will define soon), and the friction term accounts for the rate at which this energy is lost. We will study circumstances such as this when analyzing oscillators.

2.3 Noether's Theorem

The Lagrangian formalism is useful in another way. Since the Lagrangian by definition respects the symmetries of the mechanical system, it reveals mathematical upshots that are obscured in the more traditional manner of doing mechanics. Among the more surprising of these upshots is a mathematical identity known as Noether's theorem.

Noether's theorem is a statement that relates global symmetries, as in symmetries that render the Lagrangian invariant, and quantities that are conserved along the motion of the system. Quantitatively, it states that associated with each global symmetry is a conserved quantity (call it Q), that satisfies

$$\frac{dQ}{dt} = 0 \quad (2.41)$$

along the trajectory specified by the Euler-Lagrange equations. The most salient examples of such conserved quantities arise out of the Galilean group; as it happens there are conserved quantities associated with invariance under translations along time, translations along space, as well as rotations. For the time being, we will focus on the first two of these, pushing the discussion of the conserved quantities due to rotations when we study the dynamics of rigid bodies.

Let's start with invariance along time. We assume that the Lagrangian stays fixed when we perturb time by a small parameter, as in

$$t \longrightarrow t + \delta t, \quad (2.42)$$

or in other words that we have

$$\frac{\partial L}{\partial t} = 0. \quad (2.43)$$

Given this, note that the *total derivative* of the Lagrangian reduces to

$$\frac{dL}{dt} = \frac{\partial L}{\partial q_a} \dot{q}_a + \frac{\partial L}{\partial \dot{q}_a} \ddot{q}_a \quad (2.44)$$

where we can drop the partial derivative term with respect to t . Noting now that on the support of the Euler-Lagrange equations we have

$$\frac{\partial L}{\partial q_a} \dot{q}_a = \frac{d}{dt} \frac{\partial L}{\partial \dot{q}_a} q_a \quad (2.45)$$

we find

$$\frac{dL}{dt} = \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}_a} \dot{q}_a \right) \quad (2.46)$$

or

$$\frac{dH}{dt} = 0 \quad (2.47)$$

where

$$H = \frac{\partial L}{\partial \dot{q}_a} \dot{q}_a - L \quad (2.48)$$

is known as the *hamiltonian* function of the motion. Time translation invariance tells us that $H = E_0$ for some constant E_0 during the motion. It actually turns out that classical mechanics can be equivalently formulated using H as the fundamental function instead of L , which will form the subject of study in chapter 6.

In the case of spatial translations, it's even easier to find the conserved quantities. Consider an infinitesimal translation

$$q_a \longrightarrow q_a + \epsilon_a. \quad (2.49)$$

We know that under this, we expect the Lagrangian to transform as

$$\delta_\epsilon L = \frac{\partial L}{\partial q_a} \cdot \epsilon_a. \quad (2.50)$$

The invariance of the Lagrangian under such a translation requires this variation to vanish; using the Euler-Lagrange equation and the fact that the ϵ_a are arbitrary, we find

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}_a} = 0. \quad (2.51)$$

The quantities

$$p_a = \frac{\partial L}{\partial \dot{q}_a} \quad (2.52)$$

are called generalized momenta, and are the constants of motion corresponding to invariance under translations of the generalized coordinates³. The realizations of energy and momenta as conserved quantities due to Galilean invariance are a genuine prediction of modern analytical dynamics, which cannot be understood in the plain Newtonian framework alone.

³It should be pointed out that it's not always the case that these match the general idea of momentum that's more familiar to students. That said, when the generalized coordinates essentially represent actual spatial coordinates, the two do match.

Noether’s theorem is quite a deep result: it tells us that conservation laws are a reflection of underlying symmetries that enforce mechanical similarity between systems that may look different, but actually have the same physical dynamics. The full scope of Noether’s theorem is best understood in the context of field theory, rather than ordinary mechanics, but the simple instances of its applications here hopefully serve as a hint to just how deep the general lessons of the theorem are.

Solving Problems by Quadrature

In the broader scheme of history, computers haven’t really been around all that long. Among other things, this meant that historically, physics was much harder, and exactly solvable problems in mechanics were valued much more than they are today.

One of the general names such problems went by was “*problems soluble by quadrature*”. What this really meant is that the equations of motion could be recast into basically one integral expression, that could be computed—usually by hand—numerically if needed. While the myriad methods of doing these integrals are somewhat *passé* at this point, the overall algorithm is instructive.

The gold standard of solving by quadrature is found in single-particle problems, where the Lagrangian can be written as

$$L = \frac{1}{2}m\dot{q}^2 - V(q) \quad (2.53)$$

where q is a one dimensional generalized coordinate. This Lagrangian is obviously invariant under translations of time, which means that the hamiltonian

$$H = \frac{1}{2}m\dot{q}^2 + V(q) \quad (2.54)$$

is a constant of the motion. Say this constant is E_0 . This tells us that⁴

$$\frac{dq}{dt} = \sqrt{\frac{2}{m}(E_0 - V(q))}. \quad (2.55)$$

This can be directly integrated by rearranging; we have

$$t - t_0 = \int_{q(t_0)}^{q(t)} \frac{dq}{\sqrt{\frac{2}{m}(E_0 - V(q))}} \quad (2.56)$$

This is a full solution to the equations of motion. In general, how this works in practice is that we find $q(t_0)$ by solving for a certain choice of E_0 , by setting

$$E_0 = V(q_0) \quad (2.57)$$

⁴We have ignored the subtlety of the plus or minus sign on the square root; when this sign changes, the direction of the motion will adjust accordingly.

which amounts to setting the initial velocity to 0. Then the integral can be directly evaluated, numerically if necessary. The upshot of all of this is that so long as a mechanical problem can be reduced to a one dimensional problem of this kind, we've basically solved the equations of motion. Indeed, much of classical mechanics really boils down to this: convert a high dimensional problem into a one dimensional problem, and then solve by quadrature.

3 Harmonic Oscillators

In many ways, the harmonic oscillator is kind of like a theoretical laboratory. It's complicated enough that the dynamics differ quite substantially from that of free particle motion, but is also simple enough that it can be studied under increasingly complicated assumptions, which can probe various notions found elsewhere, such as damping, resonance, perturbation theory and so on. Sidney Coleman once remarked that a young physicist's career really consists of being increasingly sophisticated in his handling of the harmonic oscillator. Characteristic of a lot of Coleman's remarks, it has a reductive element to it, but it's true enough; the basic techniques learned in analyzing the harmonic oscillator in the classical context often carry over more or less unchanged to virtually every other domain of physics.

3.1 The Free Oscillator

The harmonic oscillator is the physical system obtained by applying a force proportional to the displacement on a free particle

$$\vec{F} = -k\vec{x}, \quad (3.1)$$

where k is a positive constant. Implicit in this is the fact that we no longer have invariance under spatial translations. The force itself depends on how much the particle has been displaced with respect to an origin, which we may define as the point at which the particle does not experience a force. Further, the fact that the force is opposing the displacement from the origin at all times implies that it's *restorative*. The more the displacement, the more the force seeks to return the particle to a point where the force is zero.

In terms of the Lagrangian formalism, it's easy to check by direct computation that this force can be derived from the following potential function

$$U(x) = \frac{1}{2}k\vec{x}^2, \quad (3.2)$$

which results in the following Lagrangian

$$L = \frac{1}{2}m\dot{\vec{x}}^2 - \frac{1}{2}k\vec{x}^2. \quad (3.3)$$

In 3 dimensions, it's easy to see that this is really the sum of three Lagrangians, each of which defines an independent harmonic oscillator⁵. Accordingly, we lose no generality in simply studying the one dimensional case:

$$L = \frac{1}{2}m\dot{x}^2 - \frac{1}{2}kx^2. \quad (3.4)$$

⁵The easiest way to see this is that there are no cross terms in either the kinetic or potential terms; terms such as $\sim xy$ are absent.

Coming now to the Euler-Lagrange equation, we find the well-known form of

$$m\ddot{x} = -kx, \quad (3.5)$$

which can be written as

$$Dx = 0 \quad (3.6)$$

for

$$D = \frac{d^2}{dt^2} + \frac{k}{m}. \quad (3.7)$$

The mathematically precise formulation of this is to state that the solution of the Euler-Lagrange equation lives in the *kernel* $\ker(D)$ of the differential operator D . The question now comes down to actually finding this kernel. Before doing so, it's worth making some elementary remarks on what the kernel is.

Quantitatively speaking, the kernel is a subspace of the set $C^2(\mathbb{R})$ of second-differentiable functions on the real line, since our displacement $x(t)$ is a function from $\mathbb{R}$ to $\mathbb{R}$. One particularly useful basis for this space happens to be the family of exponential functions $e^{i\omega t}$ where ω is a complex number. What this means is that any function $f \in C^2(\mathbb{R})$ can be expanded as

$$f(t) \sim \sum_{\omega} a_{\omega} e^{i\omega t} \quad (3.8)$$

where the sum runs over some family of ω . Ultimately, finding the kernel boils down to figuring out what this family of ω ought to look like.

Phrased like this, the problem is simple enough. We see that

$$\ker(D) = \text{span}\{e^{\pm i\omega_0 t}\} \quad (3.9)$$

where

$$\omega_0 = \sqrt{\frac{k}{m}}. \quad (3.10)$$

In other words we have

$$x(t) = a_+ e^{i\omega t} + a_- e^{-i\omega t}. \quad (3.11)$$

It should be noted that this manner of analyzing linear differential equations, known as *Fourier analysis* is not how the harmonic oscillator was historically studied (and happens to also not be the way it is first studied in most formal courses today). That being said, the versatility of this way of studying linear differential equations will reveal itself when studying not only various perturbations to the harmonic oscillator,

but also when encountering problems in more advanced areas of physics. As with any powerful tool, the sooner it is understood well, the better.

The coefficients $a_{\pm}$ may be fixed by a suitable choice of boundary conditions. The simplest way to do this is to once again make use of the fact that due to the invariance of the Lagrangian along translations of time, we know that the energy is a constant. Assuming now that

$$x(0) = 0 \tag{3.12}$$

we can define

$$\dot{x}(0) := v_0 = \sqrt{\frac{2E}{m}}. \tag{3.13}$$

This gives us the following set of algebraic equations

$$\begin{aligned} a_+ + a_- &= 0 \\ i\omega(a_+ - a_-) &= v_0. \end{aligned} \tag{3.14}$$

which upon solving yields

$$a_+ = -a_- = -i\sqrt{\frac{E}{2k}} \tag{3.15}$$

producing the solution

$$x(t) = \sqrt{\frac{2E}{k}} \sin(\omega_0 t). \tag{3.16}$$

Needless to say, this is the same solution that would have been obtained had we solved the Euler-Lagrange equation using standard methods of differential equations.

3.1.1 The Harmonic Approximation

Coleman's flippant remark really comes from a rather elementary observation; to a pretty good approximation, any potential can be written as a harmonic oscillator, plus additional stuff. How this works is as follows. Suppose we have a single particle Lagrangian that takes the standard form

$$L = \frac{1}{2}m\dot{x}^2 - U(x) \tag{3.17}$$

in one dimension. $U(x)$ in this case is some generic function of the position x . Consider now a point of minima, which would correspond to a point x_0 where

$$\frac{\partial U}{\partial x}(x_0) = 0 \tag{3.18}$$

and

$$\frac{\partial^2 U}{\partial x^2}(x_0) > 0. \quad (3.19)$$

Expanding the Lagrangian around this point gives us the expansion

$$L = \frac{1}{2}m\dot{x}^2 - U(x_0) - \frac{1}{2}\left(\frac{\partial^2 U}{\partial x^2}(x_0)\right)(x - x_0)^2 + \cdots \quad (3.20)$$

By translating $x \rightarrow x - x_0$, this just becomes the Lagrangian for a harmonic oscillator with constant

$$k = \frac{\partial^2 U}{\partial x^2}(x_0). \quad (3.21)$$

In other words, any potential (as long as it has a minimum), can be resolved into a harmonic oscillator, plus corrections, around that minimum. Treating the corrections requires perturbation theory, which we will discuss in a later section.

3.1.2 Normal Modes

In certain physical applications that are known to have harmonic oscillator behavior, it may not be possible to straightforwardly cast the Lagrangian in the form of N decoupled harmonic oscillators. The most general such case would amount to a simple quadratic Lagrangian that takes the form

$$L = \frac{1}{2}m_{ij}\dot{q}_i\dot{q}_j - \frac{1}{2}k_{ij}q_iq_j \quad (3.22)$$

where the Lagrangian has been expressed in generalized coordinates. Here, the matrices m_{ij} and k_{ij} are naturally symmetric, due to the form of the Lagrangian. Although this looks complicated, the same algorithm, applied more or less in the same manner turns out to work. The first step is to write down the Euler-Lagrange equations, which become

$$m_{ij}\ddot{q}_i + k_{ij}q_j = 0. \quad (3.23)$$

In Fourier space, this amounts to finding the kernel of $D_{ij}(\omega)$, which is defined by

$$D_{ij}(\omega) = -\omega^2 m_{ij} + k_{ij}. \quad (3.24)$$

Another way of phrasing this is to find the nullspace of D_{ij} , treated as a matrix now. This is a standard problem of linear algebra; the kernel is simply spanned by exponentials $e^{i\omega t}$ where ω satisfies the algebraic equation

$$\det(-\omega^2 m_{ij} + k_{ij}) = 0. \quad (3.25)$$

3.2 The Damped Oscillator

The free oscillator is characterized by an invariance, specifically that under translations of time. This has the effect of making the energy a conserved quantity, which aided in the isolation of the solution of the Euler-Lagrange equation. In the real-world, truly free oscillators are rare, and in most practical cases of interest, the oscillators experience damping forces, which need to be incorporated in the Euler-Lagrange equations to fully solve the problem.

Generally, to a good approximation, such forces (which tend to be empirically reconstructed) can be characterized as conventional friction,

$$F_d = -\gamma \dot{x} \quad (3.26)$$

where γ is a constant. This modifies the Euler-Lagrange equations into

$$m\ddot{x} = -\gamma\dot{x} - kx. \quad (3.27)$$

The corresponding differential operator, the kernel of which will span the solutions of the Euler-Lagrange equations, will now be given by

$$D = \frac{d^2}{dt^2} + \frac{\gamma}{m} \frac{d}{dt} + \frac{k}{m}. \quad (3.28)$$

In terms of the Fourier modes, this reduces to the problem of solving the following algebraic equation

$$\omega^2 - i\omega b - \omega_0^2 = 0 \quad (3.29)$$

where we have redefined $b = \frac{\gamma}{m}$. Since this is a quadratic equation, we can just write down the two solutions

$$\omega_{\pm} = \frac{ib \pm \sqrt{4\omega_0^2 - b^2}}{2} \quad (3.30)$$

which means that the general solution to the differential equation can be expanded according to (writing $\delta^2 = \omega_0^2 - \frac{b^2}{4}$)

$$x(t) = e^{-\frac{b}{2}t} (a_+ e^{i\delta t} + a_- e^{-i\delta t}). \quad (3.31)$$

The damped oscillator then differs from the free oscillator in two ways. First of all, the amplitude itself “damps”, in that it declines in magnitude in a characteristic fashion, given by the exponential factor $e^{-\frac{b}{2}t}$. All other things being equal, this means that the motion comes to a halt in the asymptotic future $t \rightarrow \infty$. Further, we have the exponentials that depend on δ , which actually change depending on whether or not $\delta^2 > 0$. In the former case, when $\delta^2 > 0$, δ is real and we have the portion in the brackets as oscillating functions. In this case, using the same arguments as in the

free case, but keeping in mind now that the velocity at $t = 0$ is defined in terms of the *starting energy* E_0 , we have

$$x(t) = e^{-\frac{b}{2}t} \sqrt{\frac{2E_0}{k}} \cos(\delta t). \quad (3.32)$$

An upshot of this is that by writing this instead as

$$x(t) = \sqrt{\frac{2E(t)}{k}} \cos(\delta t) \quad (3.33)$$

we find an approximate formula for energy decay

$$E(t) = E_0 e^{-bt}. \quad (3.34)$$

The general decay of energy would have to found by noting that the Noether charge is not conserved.

The case $\delta^2 < 0$ is more subtle. In this case, we have two imaginary solutions for the Fourier modes, given by

$$\omega_{\pm} = i \left(\frac{b}{2} \pm \delta \right). \quad (3.35)$$

A general solution of the Euler-Lagrange equation in this instance will be written down as

$$x(t) = a_+ e^{-(\frac{b}{2} + \delta)t} + a_- e^{-(\frac{b}{2} - \delta)t}. \quad (3.36)$$

The degenerate case of $\delta = 0$ causes the solution space to shrink; a standard theorem of differential equations can be used to show that the most general solution in this case becomes

$$x(t) = (a_1 + a_2 t) e^{-\frac{b}{2}t}. \quad (3.37)$$

As mentioned before, there isn't a clean way to express the equations of motion for the damped oscillator in terms of a specific Lagrangian, as doing so would violate Galilean invariance. This is due to the natural violation of energy conservation that is a characteristic of the damped oscillator. In general, this isn't much of an issue, because the damping term is usually phenomenological anyway, to the extent that it is used to describe real physical systems. In such cases, it ends up being more convenient to simply start with the equation of motion as a given, and start from there.

3.3 Forced Oscillators

In the same way that an oscillator can be damped on account of frictional forces, it can also be driven by an external force. External forces in actual physical systems take the form of external fields, and a microscopic description of them usually requires some knowledge of the elementary underlying dynamics. However, the solution for the motion of the oscillator in the presence of a generic force is a completely solvable problem, and illustrates some of the essential analytic features of the oscillator that make it such a commonly encountered physical system.

Since the study of an oscillator subject to a generic driving force exhibits a lot of interesting analytic properties, we will go over it in some detail, making the relevant mathematical points as we go along.

An oscillator is said to be driven when it is subject to an external force, that can be defined by a generic function $F(t)$ according to

$$F_d := F(t). \quad (3.38)$$

This has the effect of modifying the equation of motion into

$$m\ddot{x} + kx = F(t), \quad (3.39)$$

in the absence of friction. On the level of the Lagrangian, the same equation can be recovered as the standard Euler-Lagrange equation for the modified Lagrangian

$$L = \frac{1}{2}m\dot{x}^2 - \frac{1}{2}kx^2 + F(t)x. \quad (3.40)$$

Modifications of this form show up in a number of phenomena, and will become especially relevant when studying the interactions between fields and particles, as well as that between various fields. The linear coupling between the position x and the forcing function $F(t)$ is generally known as a *source* term, as terms of this sort have the interpretation of being sources when dealing with fields.

To keep our discussion as general as possible, we will include friction terms, giving rise to the general equation of motion for a damped oscillator subject to a drive

$$m\ddot{x} + \gamma\dot{x} + kx = F(t). \quad (3.41)$$

Let's start with the simplest example of a driving force, namely a periodic driving force at a fixed frequency ω ,

$$F(t) = F_0 e^{i\omega t}. \quad (3.42)$$

We have chosen to keep the force in terms of a constant exponential, keeping in mind that the real or imaginary part can always be taken to specialize to a real-valued

sine or cosine wave. In this case, we would expect for the displacement to follow an approximately similar periodic dependence

$$x(t) \sim Ae^{i\omega t}. \quad (3.43)$$

Direct computation shows us that we must have

$$A = \frac{F_0}{m} \frac{1}{\omega_0^2 - \omega^2 + ib\omega}. \quad (3.44)$$

The quantity

$$G(\omega) = \frac{1}{\omega_0^2 - \omega^2 + ib\omega} \quad (3.45)$$

is known as the frequency-space Green's function, and shows up in virtually every domain of physics in which oscillatory phenomena are important. Splitting this into real and imaginary parts, we find

$$G(\omega) = |G(\omega)|e^{i\delta(\omega)} \quad (3.46)$$

where

$$\tan \delta(\omega) = \frac{b\omega}{\omega^2 - \omega_0^2}. \quad (3.47)$$

$\delta(\omega)$ is popularly known as the *phase-shift*, and as the name suggests, physically it corresponds to a delay between the action of the force, and when the motion of the particle “sees” the force. When there is zero friction, this delay vanishes, and the force and the particle motion are perfectly in phase away from resonance. On the other hand, as friction begins to dominate, the phase offset grows, and for very high amounts of friction, the applied force and the motion of the particle are 90 degrees out of phase.

3.3.1 Green's Functions and General Forces

We start again with the problem of solving the equation of motion

$$m\ddot{x} + \gamma\dot{x} + kx = F(t) \quad (3.48)$$

assuming this time that $F(t)$ can be any generic function. As we will see, even this claim can't be made quite so loose; there are conditions that $F(t)$ will have to satisfy, but we will proceed with this for the time being.

Our first claim is that there exists a function $G(t - t')$ that satisfies

$$(\partial_t^2 + b\partial_t + \omega_0^2) G(t - t') = \delta(t - t') \quad (3.49)$$

where $\delta(t)$ is the Dirac delta function. The second claim is that the solution of the Euler-Lagrange equation can be written down as

$$x(t) = \frac{1}{m} \int_{-\infty}^t G(t-t') F(t') dt'. \quad (3.50)$$

The notation here is meant to be suggestive; the function $G(t-t')$ is also known as the Green's function. In fact, it's just the Fourier transformed version of the Green's function $G(\omega)$ we saw in the preceding section. To see this, we recall the spectral expansion of the delta function

$$\delta(t-t') = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega(t-t')} d\omega. \quad (3.51)$$

If we represent $G(t-t')$ in a spectral expansion according to

$$G(t-t') = \frac{1}{2\pi} \int_{-\infty}^{\infty} G(\omega) e^{i\omega(t-t')} d\omega \quad (3.52)$$

we find by direct evaluation that we must have

$$(-\omega^2 + ib\omega + \omega_0^2)G(\omega) = 1 \quad (3.53)$$

for all ω , which matches (3.45). The second claim is then even easier to verify; applying the operator $(\partial_t^2 + b\partial_t + \omega_0^2)$ to the claimed solution for $x(t)$ we get

$$(\partial_t^2 + b\partial_t + \omega_0^2) x(t) = \frac{1}{m} \int_{-\infty}^t (\partial_t^2 + b\partial_t + \omega_0^2) G(t-t') F(t') dt \quad (3.54)$$

which simplifies to

$$(\partial_t^2 + b\partial_t + \omega_0^2) x(t) = \frac{1}{m} \int_{-\infty}^t \delta(t-t') F(t') dt = \frac{1}{m} F(t) \quad (3.55)$$

obviously satisfying the Euler-Lagrange equation. Ultimately, this means that finding the solution of the Euler-Lagrange equation is equivalent to finding the Green's function, and then performing a convolution integral with the forcing function.

Now the definition of the Green's function as the integral expression

$$G(t-t') = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{i\omega(t-t')}}{\omega_0^2 - \omega^2 + ib\omega} d\omega \quad (3.56)$$

tells us that direct calculation of the Green's function is difficult, not even getting to the point of integrating this against the forcing function. However, quite a bit can be extracted by working in frequency space instead, and we can identify features of the Green's function directly from the integral expression. First, observe that if we

require a *causal* Green's function, meaning that for $t < t'$ $G(t - t') = 0$, we require that

$$\text{Im}(\omega) > 0. \quad (3.57)$$

When this is the case, both of the poles of the Green's function $G(\omega)$ lie in the upper half plane. When $t > t'$, the integral can be closed in the upper half plane, enclosing the two poles and giving us a nonzero $G(t - t')$, while for $t < t'$, the integral is well-defined when the contour is closed in the lower half plane, and identically vanishes.

Coming now to the solution of the Euler-Lagrange equation in frequency space, we can do this by utilizing the two expansions

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} x(\omega) e^{i\omega t} d\omega \quad (3.58)$$

and

$$F(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{i\omega t} d\omega. \quad (3.59)$$

For this integral to exist, we require that the force satisfy

$$\int_{-\infty}^{\infty} |F(\omega)| d\omega < \infty. \quad (3.60)$$

Functions that satisfy this property are known as absolutely integrable functions. When dealing with Fourier transforms rigorously, it turns out to be best to deal with functions that vanish outside a finite region (or rapidly decaying functions, known as Schwartz functions). Getting into the weeds with the analysis can take us out of the main point of this text, but it's worth keeping in mind that we need to deal with relatively well behaved forcing functions for all of this to be well-defined. Readers interested in more details may consult the appendix for a short course on Fourier transforms.

Coming back to the Euler-Lagrange equation, the spectral expansion of $x(t)$ and $F(t)$ results in the equivalent equation

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} m(-\omega^2 + ib\omega + \omega_0^2) x(\omega) e^{i\omega t} d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{i\omega t} d\omega. \quad (3.61)$$

By noting that each $e^{i\omega t}$ is linearly independent, the Euler-Lagrange equation reduces to an (infinite) collection of equations for each ω

$$mG(\omega)^{-1} x(\omega) = F(\omega) \quad (3.62)$$

which yields

$$x(\omega) = \frac{1}{m}G(\omega)F(\omega). \quad (3.63)$$

This is the solution of the Euler-Lagrange equation for a generic forcing function. It's a straightforward evaluation to verify that this is equivalent to the more baroque version presented in (3.50).

3.4 Anharmonic Oscillators and Perturbation Theory

Let's go back to our understanding that a general potential $U(x)$ is, to a first approximation around a minimum, a harmonic oscillator. But what about higher order terms. Indeed, a general polynomial potential is going to look something like this

$$U(x) = U(x_0) + (x - x_0)^2 U''(x_0) + (x - x_0)^3 U'''(x_0) + \dots \quad (3.64)$$

Accordingly, we can consider the so-called *anharmonic* oscillator, in which we do not neglect these higher order terms. In the following, we will consider the anharmonic oscillator with potential energy

$$U(x) = \frac{1}{2}kx^2 + \frac{1}{3}\mu x^3 + \frac{1}{4}\lambda x^4. \quad (3.65)$$

Generally, in such potentials we want $\lambda > 0$, since this ensures that for large values of x , the potential does not go to $-\infty$. This is to ensure that the local minima correspond to true minima, rather than local artifacts. It's for the same reason that we don't deal with the pure μx^3 term, and only consider it in combination with the quartic term.

As it turns out, the Euler-Lagrange equation for such a potential in the absence of a friction term can be solved exactly in terms of elliptic functions. We go over this example in the appendix, but it's not especially illuminating for the discussion here, where we want to study the problem for *small* values of μ and λ . The systematic way of doing this is known as perturbation theory, and is really the only way to do calculations for any practical problem that shows up in the physical world. As tends to be the case for the harmonic oscillator, the case of the anharmonic oscillator is a very convenient toy example to understand the mechanics of perturbation theory.

The basic idea is as follows. Since the anharmonic parameters are small, we can think of the full solution to the problem as simply “harmonic oscillator plus corrections”. The way we do this is by using a bookkeeping parameter ϵ by rescaling

$$(\mu, \lambda) \longrightarrow (\epsilon\mu, \epsilon\lambda) \quad (3.66)$$

which will be set to $\epsilon = 1$ at the end. The purpose of the parameter ϵ is to control the “degree” of smallness; by organizing the calculation in orders of ϵ , we can isolate out corrections to the free harmonic oscillator in a systematic way. In this vein, we assume that the full solution to the anharmonic oscillator is given by the expansion

$$x(t) = x^{(0)}(t) + \epsilon x^{(1)}(t) + \epsilon^2 x^{(2)}(t) + \dots \quad (3.67)$$

and require that each coefficient of ϵ^n vanish independently. At ϵ^0 we find the bare Euler-Lagrange equations

$$\ddot{x}^{(0)}(t) + b\dot{x}^{(0)}(t) + \omega_0^2 x^{(0)}(t) = 0. \quad (3.68)$$

At ϵ^1 we find the more complicated equation

$$\ddot{x}^{(1)}(t) + b\dot{x}^{(1)}(t) + \omega_0^2 x^{(1)}(t) = \mu x^{(0)}(t)^2 + \lambda x^{(0)}(t)^3 \quad (3.69)$$

and more complicated terms at higher order. Even at first sight, we can see what's going on here. The lower order solutions basically act as forcing functions for the higher order contributions, and the exact character of the force depends on the form of the anharmonic potential. To identify how this works in practice, let's write down the general solution at leading order

$$x^{(0)}(t) = a_+ e^{i\omega_+ t} + a_- e^{i\omega_- t} \quad (3.70)$$

where $\omega_{\pm}$ are the solutions of

$$\omega^2 - ib\omega - \omega_0^2 = 0. \quad (3.71)$$

We assume $\delta^2 > 0$, and rewrite the solution in terms of its real part

$$x^{(0)}(t) = \text{Re}(a_+ e^{i\omega_+ t} + a_- e^{i\omega_- t}) = A e^{-\frac{b}{2}t} \cos(\delta t + \phi) \quad (3.72)$$

where A and ϕ are the amplitude and phase to be determined by the boundary conditions. Coming now to the ϵ^1 equations, we compute

$$\begin{aligned} \mu x^{(0)}(t)^2 + \lambda x^{(0)}(t)^3 &= \frac{\mu A^2 e^{-bt}}{2} + \frac{\mu A^2 e^{-bt}}{2} \cos(2\delta t + 2\phi) \\ &+ \frac{3\lambda A^3 e^{-\frac{3}{2}bt}}{4} \cos(\delta t + \phi) + \frac{\lambda A^3 e^{-\frac{3}{2}bt}}{4} \cos(3\delta t + \phi). \end{aligned} \quad (3.73)$$

where we have simplified the square and cubed terms using standard trigonometric identities.

The best way to solve the Euler-Lagrange equations at this point is to rewrite these as real parts of exponentials, and then use the Green's functions. Since we have a constant, and three harmonics at $\omega = \delta, 2\delta, 3\delta$, this will involve four different terms with different $G(\omega)$.

The physical phenomena at play here show up in considerably more complex physical systems. The source at ϵ^1 has four terms; the constant term has $\omega = 0$, and

is known as the DC⁶ term. The others are known as harmonics; 2δ corresponds to the second harmonic, for instance. This particular anharmonic oscillator is a model of non-linear processes in crystals for example, and the higher harmonics actually correspond to the combination of different pulses of light, to give a new one at a higher frequency. Furthermore—and unsurprisingly—the higher harmonics decay more rapidly in time. This accords with a physically meaningful picture that far into the future, systems tend to equilibrate away from deviations from harmonicity.

⁶Standing for direct current.

4 Theory of Orbits

A good operative definition of orbits is as trajectories due to forces that only depend on the linear distance between particles. This means that between particle i and particle j , the force mediated between them will be only a function of the radial separation

$$F_{ij} := F_{ij}(|\vec{x}_i - \vec{x}_j|). \quad (4.1)$$

The most canonical of orbital trajectories is the orbit due to the inverse square law, which provides the classical description of gravity. In this chapter, we will study orbits generally at first, followed by a more careful treatment of gravitational orbits in particular. The study of orbits is one of the more concrete applications of classical mechanics, and orbital perturbation theory in particular is a significant test of Newtonian mechanics, exemplified by the computation of the precession of perihelions. We will study this particular phenomenon in the last section.

4.1 The Two-Body Problem

The two-body problem is the question of determining the orbit of two particles which interact via a force that only depends on their mutual separation. Specifically, the Lagrangian can be expanded as

$$L = \frac{1}{2}m_1\dot{\vec{x}}_1^2 + \frac{1}{2}m_2\dot{\vec{x}}_2^2 - U(|\vec{x}_1 - \vec{x}_2|). \quad (4.2)$$

We are assuming that the particles propagate in three dimensions, meaning that we have

$$\begin{aligned} \vec{x}_1 &= (x_1, y_1, z_1) \\ \vec{x}_2 &= (x_2, y_2, z_2) \end{aligned} \quad (4.3)$$

As presented, the Lagrangian actually hides the fact that the two-body problem is really a misnomer; the Lagrangian actually turns out to describe a one-body problem, due to a cute trick. Consider the following coordinate transformation; we have the separation vector

$$\vec{r} = \vec{x}_1 - \vec{x}_2 \quad (4.4)$$

and the center of mass

$$\vec{r}_{\text{COM}} = \frac{m_1\vec{x}_1 + m_2\vec{x}_2}{m_1 + m_2}. \quad (4.5)$$

It's a simple algebraic exercise to show that we have

$$\frac{1}{2}m_1\dot{\vec{x}}_1^2 + \frac{1}{2}m_2\dot{\vec{x}}_2^2 = \frac{1}{2}\mu\dot{\vec{r}}^2 + \frac{1}{2}(m_1 + m_2)\dot{\vec{r}}_{\text{COM}}^2 \quad (4.6)$$

where the reduced mass μ is defined by

$$\mu = \frac{m_1 m_2}{m_1 + m_2}. \quad (4.7)$$

But now the Lagrangian becomes very simple; in the new coordinates we find

$$L = \frac{1}{2} \mu \dot{\vec{r}}^2 - U(r) + L_{\text{COM}} \quad (4.8)$$

where the center of mass Lagrangian is just that of a free particle, and has trivial dynamics. All of the nontrivial physics is contained in just the part of the Lagrangian that depends on the separation vector, which can without any loss of generality be treated as the generalized coordinate of a new particle of mass μ . In this manner, the two-body problem is really just a one-body problem in disguise⁷. Henceforth, we will drop the center of mass part, and deal with only the separation.

An immediate consequence of the new Lagrangian is the conservation of angular momentum; observe that we have the following Euler-Lagrange equation

$$\mu \ddot{\vec{r}} = -U'(r) \hat{r} \quad (4.9)$$

which is easy enough to verify. This means that we have

$$\vec{r} \times \ddot{\vec{r}} = 0 \quad (4.10)$$

or

$$\frac{d}{dt}(\vec{r} \times \dot{\vec{r}}) = 0. \quad (4.11)$$

Essentially, there exists a vector $\vec{L} \propto \vec{r} \times \dot{\vec{r}}$ that is conserved (both in magnitude as well as in direction) during the motion. But a constant vector specifies a plane perpendicular to it, meaning that the entirety of the motion takes place on a fixed plane. The three-dimensional two-body problem is really a two-dimensional one body problem. In two dimensions, the Lagrangian is even simpler:

$$L = \frac{1}{2} \mu \dot{r}^2 + \frac{1}{2} \mu r^2 \dot{\phi}^2 - U(r) \quad (4.12)$$

where r is the radial direction and ϕ is the angular direction. Just inspecting the Euler-Lagrange equations tells us that there is one further simplification; the Lagrangian does not depend on ϕ . This means that

$$\frac{\partial L}{\partial \phi} = \mu r^2 \dot{\phi} \quad (4.13)$$

is also a constant. Calling this constant ℓ , we have the final form of the Lagrangian

⁷Unlike the two-body problem of finding jobs, which is a genuine two-body problem.

$$L = \frac{1}{2}\mu\dot{r}^2 + \frac{\ell^2}{2\mu r^2} - U(r) \quad (4.14)$$

which is a one-body problem in one dimension, easily solved by quadrature. At this point, we can just write the solution down. The energy is conserved due to time translation invariance; we have

$$E = \frac{1}{2}\mu\dot{r}^2 + \frac{\ell^2}{2\mu r^2} + U(r) \quad (4.15)$$

which can be rearranged to get

$$dt = \frac{dr}{\sqrt{\frac{2}{\mu}(E - U(r)) - \frac{\ell^2}{\mu^2 r^2}}}. \quad (4.16)$$

An occasionally more easily evaluated form is obtained by rewriting this in terms of $d\phi$ instead, using the constancy of $\ell = \mu r^2 \dot{\phi}$,

$$d\phi = \frac{\ell dr}{\mu r^2} \frac{1}{\sqrt{\frac{2}{\mu}(E - U(r)) - \frac{\ell^2}{\mu^2 r^2}}} \quad (4.17)$$

This integral can be hard, or easy to evaluate, based on exactly what kind of potential $U(r)$ we are dealing with. Fortunately, the potential most relevant to real-world orbits makes the previous integral elementary (in the sense of elementary functions, not necessarily trivial). More complicated choices will result in more complicated functions, such as elliptic, and hyperelliptic functions⁸ for high enough degrees.

4.2 The Kepler Problem

The Kepler problem is the paradigmatic classical problem solved by Newton. In modern language, it is the study of orbits for the choice of potential

$$U(r) = \frac{\alpha}{r} \quad (4.18)$$

where α is a constant⁹. The force corresponding to this is the well-known inverse square law

$$-\nabla U(r) = \frac{\alpha}{r^2}. \quad (4.19)$$

For $\alpha < 0$, this is the law of gravitational attraction discovered by Newton. Values of α greater than zero are also interesting, but do not result in closed orbits. The

⁸Superelliptic functions do not, unfortunately, show up all that often in physics.

⁹For the simple two-body case we're looking at, this constant is inferred from Newton's law of gravitation. In other words, it's just $\alpha = Gm_1m_2$.

Kepler problem itself can be solved for any choice of α , and specialized cases can be considered afterward.

Given this choice of potential, the solution by quadrature is given by the integral

$$d\phi = \frac{\ell dr}{\mu r^2} \frac{1}{\sqrt{\frac{2}{\mu} \left(E - \frac{\alpha}{r}\right) - \frac{\ell^2}{\mu^2 r^2}}}. \quad (4.20)$$

Now by making a change of variables $x = \frac{1}{r}$, the integral simplifies into the elementary problem

$$d\phi = -\frac{\ell}{\mu} \frac{dx}{\sqrt{\frac{2}{\mu} (E - \alpha x) - \frac{\ell^2 x^2}{\mu^2}}}. \quad (4.21)$$

Evaluating the integral, we find

$$\phi = \cos^{-1} \left(\frac{\ell x + \frac{\mu \alpha}{\ell}}{\sqrt{2\mu E + \frac{\alpha^2 \mu^2}{\ell^2}}} \right) + \text{const.} \quad (4.22)$$

A little rearrangement tidies it up considerably; we can rewrite this as

$$e \cos(\theta - \theta_0) = \frac{p}{r} - 1 \quad (4.23)$$

where

$$e = \sqrt{1 + \frac{2E\ell^2}{\alpha^2 \mu}} \quad (4.24)$$

and

$$p = \frac{\ell^2}{\mu \alpha}. \quad (4.25)$$

Equation (4.23) is the parametric form of the trajectory defined by a conic section. It isn't obvious looking at it, but treating e as the eccentricity, this turns out to not be so hard to prove. Unsurprisingly, due to the dependence of e on E (which in turn depends on α), we will see that choosing different signs for α will result in different conic sections.

4.2.1 Closed Orbits and Kepler's Laws

We'll work backwards, since it's easier that way. We claim that for $e < 1$, we have closed orbits. This implies that we must have

$$0 > E > -\frac{\alpha^2 \mu^2}{2\ell^2}. \quad (4.26)$$

It's not hard to see that this is only possible when $\alpha < 0$. Looking back at the expression for the energy

$$E = \frac{1}{2}\mu\dot{r}^2 + \frac{\ell^2}{2\mu r^2} + \frac{\alpha}{r} \quad (4.27)$$

it's clear that the energy can take non-positive values only if $\alpha < 0$. In fact, this can be seen directly by plotting the “effective potential”

$$U_{\text{eff}}(r) = \frac{\ell^2}{2\mu r^2} + \frac{\alpha}{r}. \quad (4.28)$$

Coming back now to the form of the orbit, we have

$$\frac{p}{r} = 1 + e \cos(\theta - \theta_0) \quad (4.29)$$

where $e < 1$. Since θ_0 is simple an initial phase, by choosing $x(0) = 0$, we may define

$$\begin{aligned} x(\theta) &= r \cos(\theta - \theta_0) \\ y(\theta) &= r \sin(\theta - \theta_0). \end{aligned} \quad (4.30)$$

Expanding the orbit solution using this we have

$$x^2 + y^2 = (p - ex)^2 \quad (4.31)$$

which after some algebra is the standard equation for an ellipse

$$\frac{(x - \xi)^2}{a^2} + \frac{y^2}{b^2} = 1 \quad (4.32)$$

where $a = \frac{p}{1-e^2}$, $b = \frac{p}{\sqrt{1-e^2}}$ and $\xi = \frac{pe}{1-e^2}$. In the limit $e = 0$, this reduces down to the equation of a circle with radius p . The constants a and b are the semi-major and semi-minor axes respectively, denoting the length and breadth of the ellipse. The observation that planets orbit the sun in elliptical orbits is the content of *Kepler's first law*. We can see from the original expression for the orbit that the maximum and minimum values of r are given by $\frac{p}{1-e}$ and $\frac{p}{1+e}$ respectively. For general orbits, these are known as the apoapsis and periapsis respectively, and for our solar system, they are called the aphelion and perihelion.

As it happens, we've already seen the second law, in the form of a conserved quantity. Recall that we had

$$\mu r^2 \dot{\phi} = \ell \quad (4.33)$$

Rearranging and noting that $r \times r d\phi = r ds = 2dA$, where ds is an infinitesimal arc length along the orbit and dA is the infinitesimal area swept by the particle in the orbit

$$dA = \frac{\ell}{2} dt. \quad (4.34)$$

In particular, for a given interval of time Δt , the area ΔA swept will be the same, regardless of the exact starting and ending location of the particle. This is known as *Kepler's second law*.

Unfortunately, things aren't quite as slick when it comes to the third law, which is somewhat tougher to prove. We actually need the fact that the force is given by the inverse square law to prove it. The main content of the law is to establish a relationship between the characteristic length scales of the orbit, and the period of the planet moving in the orbit. We start with the formula for the area of an ellipse, which is

$$A = \pi ab. \quad (4.35)$$

Now we can use a trick; since the second law tells us that the same area is swept up in the same time, we can just integrate the second law to give us

$$T = \frac{2A}{\ell} \quad (4.36)$$

which we now rewrite in terms of p , using the formulas for a , b and (4.25),

$$T = \frac{2p^2}{p^{1/2} \sqrt{\mu\alpha} (1 - e^2)^{3/2}} \quad (4.37)$$

or

$$T^2 \propto a^3 \quad (4.38)$$

which is *Kepler's third law*.

Elliptical orbits are of fundamental importance to the study of our solar system. All of the planets, as well as their respective satellites, travel on elliptical orbits. The planets themselves describe elliptical orbits around the sun, which the satellites orbit both the sun and the earth. Incidentally, this is the reason why the Ptolemaic model involves so-called epicycles. An epicycle is simply the trajectory described by a body which moves in a small circle, such that the circle itself has a circular orbit around a fixed center. In fact, by setting

$$z(t) = A_1 e^{i\omega_1 t} + A_2 e^{i\omega_2 t} \quad (4.39)$$

such that $A_1 \gg A_2$, the real and imaginary parts would determine the $x(t)$ and $y(t)$ positions of such an object, with respect to the origin. Fixing the earth as the center and ignoring ellipticity, this is a pretty good model of all observed movements of planets in the night sky. With hindsight, we can see that this was an early attempt at a Fourier series, taking only the first few terms into account.

Table 1. Eccentricities of Planets in the Solar System

Planet	Eccentricity
Mercury	0.2056
Venus	0.0068
Earth	0.0167
Mars	0.0934
Jupiter	0.0484
Saturn	0.0541
Uranus	0.0472
Neptune	0.0086
Pluto	0.249

The eccentricity itself varies widely among the different planets and other orbiters of the sun. The earth for instance has a rather small eccentricity of 0.0167, which means that to an excellent approximation, the orbit is basically circular. Halley's comet on the other hand has an eccentricity of 0.967, and has a highly exaggerated and narrow orbit around the sun.

4.2.2 Open Orbits

Open orbits occur when the energy $E > 0$. Recalling the formula for the eccentricity,

$$e = \sqrt{1 + \frac{2E\ell^2}{\alpha^2\mu}} \quad (4.40)$$

such an orbit must always have $e > 1$. Once again, it's just an algebraic exercise to show that the orbital equation can be rewritten as

$$\frac{(x - \xi)^2}{a^2} - \frac{y^2}{b^2} = 1 \quad (4.41)$$

where this time we have

$$\begin{aligned} a &= \frac{p}{(e^2 - 1)} \\ b &= \frac{p}{\sqrt{e^2 - 1}} \end{aligned} \quad (4.42)$$

and $\xi = \frac{pe}{e^2 - 1}$. This is the standard form of the equation for a hyperbola. The orbit is unbounded in the sense that there is a solution for θ for any value of $r > 0$; for $r \rightarrow \infty$, the solution is

$$\theta - \theta_0 = \cos^{-1} \left(-\frac{1}{e} \right). \quad (4.43)$$

The degenerate limit of $e = 1$ is much easier; the algebra simplifies quite a bit more to give us

$$y^2 = p^2 - 2pex \quad (4.44)$$

which is the standard expression for a parabola. In this case, $E = 0$.

We don't see hyperbolic orbits as often as elliptic ones; comets that fly by the earth never to return are hyperbolic by definition. If they weren't, we'd see them come back.

4.3 The Eccentricity Vector

There is a distinguished vector that introduces an additional, but not essentially new, conserved quantity that can simplify calculations *post facto*. While not particularly useful from the viewpoint of the calculations done in the present chapter, this vector turns out to be important when understanding what happens when we perturb orbits. Although we will not be considering perturbed orbits in the present text, understanding this vector will give us some idea of what needs to be studied when dealing with more complex and realistic cases.

That said, the vector in question is the so-called Runge-Lenz vector, or eccentricity vector more simply. It's given by the strange looking expression

$$\vec{A} = \vec{p} \times \vec{\ell} - \mu\alpha\hat{r} \quad (4.45)$$

where as discussed earlier $\vec{\ell} = \mu\vec{r} \times \vec{v}$ is the conserved angular momentum of the orbit. It should be emphasized here that this expression is with respect to the reduced one-body problem; we have for instance

$$\vec{p} = \mu\dot{\vec{r}}. \quad (4.46)$$

As it happens, not only is the eccentricity vector a conserved quantity along a Kepler orbit, but it also has some nice properties. Proving that's its conserved is a straightforward, but tedious, exercise in algebra. See that on account of $\dot{\vec{L}} = 0$ we obtain

$$\dot{\vec{A}} = \dot{\vec{p}} \times \vec{\ell} - \mu\alpha\dot{\hat{r}}. \quad (4.47)$$

Now use

$$\dot{\vec{p}} = -\frac{\alpha}{r^2}\vec{r} \quad (4.48)$$

and

$$\dot{\hat{r}} = -\frac{\vec{r}\dot{r}}{r^2} + \frac{\dot{\vec{r}}}{r} \quad (4.49)$$

with the identity $\dot{r} = \hat{r} \cdot \vec{v}$. Using the triple cross product, it's easy to see that we have

$$\dot{\vec{A}} = 0. \quad (4.50)$$

The interesting thing about this vector is that it's related to something we've already seen. To find out what, take the dot product with $\vec{r}$. Denoting the angle between $\vec{A}$ and $\vec{r}$ as φ , this becomes

$$Ar \cos \varphi = \ell^2 - \alpha \mu r \quad (4.51)$$

or

$$\frac{p}{r} = 1 + \frac{A}{\alpha \mu} \cos \varphi \quad (4.52)$$

which is just the equation of the orbit, so long as we make the identifications $\varphi = \theta - \theta_0$ and $A = \mu \alpha e$, where e is the eccentricity. Now that we have the magnitude of the eccentricity vector (why we chose the name should be obvious at this point), what remains is to determine the magnitude. We'll use a trick this time. Since it's a conserved vector, we only need to find its direction at one moment in the orbit, and this will always be valid. The quick way of figuring this out is to solve for $\cos \varphi = 0$, which happens when $\vec{A}$ is parallel to $\vec{r}$. But when this is the case, we see that

$$r = \frac{p}{1 + e} \quad (4.53)$$

which is the location of the periapsis (or perihelion for the solar system). Thus, $\vec{A}$ always points towards the periapsis of the orbit.

Why exactly is this useful? For one thing, clean conservation laws are quite rare, and it's good to stock up on them when we can. Less flippant however is the fact that this provides one additional quantity that is likely to not remain conserved when we perturb the simple Kepler problem. For instance, a famous historical problem that can only really be treated with perturbation theory is the precession of the perihelion of mercury. A full discussion of this would likely require a detailed analysis considerably out of the scope of the present text, but the basic idea is as follows. It turns out that for real planets, $\vec{A}$ is not conserved, but actually precesses. Essentially, it's possible to show that

$$\dot{\vec{A}} = \vec{\Omega}_A \times \vec{A} \quad (4.54)$$

where $\vec{\Omega}_A$ is rather small. Specializing to the case of Mercury, this would be due to the influence of the other planets, which have the effect of making the perihelion move just a little bit every year. As far as observations are concerned, most of this is due to the fact that the Earth itself rotates. But adjusting for that, it was found

that the perihelion of Mercury precesses by an amount equal to 575 seconds¹⁰ every century. Of this, approximately 532 is due to the perturbations of Mercury's orbit due to the other planets. The remaining—or anomalous—bit of 43 seconds a century remained a thorny issue in celestial mechanics for a long time, until it was discovered that general relativity actually corrects Newton's law. This small correction supplies the remaining 43 seconds perfectly.

¹⁰Recall that 3600 seconds is one degree.

5 Rigid Body Dynamics

The idea of rigid bodies is meant to invoke images of objects whose shape and general structure does not change during the dynamics. Among other things, this means that the relative positions of the particles composing the body do not change with time. What do change however is the *orientation* of the body itself, and the natural framework to understand this is in the study of rotations. So far, we have been mostly concerned with trajectories in real space. In other words, we wanted to understand $\vec{r}(t)$, the location of a particle as a function of time. The study of rigid bodies on the other hand will involve the study of the orientations, or angles, of the body with respect to a fixed coordinate system as a function of time. Needless to say, this requires some preliminary formalism, as well as a novel set of generalized coordinates, one that captures how this orientation can be understood dynamically.

5.1 The Inertia Tensor

Let's begin with some notation, and set up our system of coordinates. In everything we will study in the present chapter, we keep in mind a fixed, inertial frame. We'll denote the axes of this frame using standard (x, y, z) notation. Given an object O , which generally speaking will be some kind of extended object (think a top, or even a planet), we will also occasionally make reference to a frame of reference fixed to the object. In other words, as the object moves with respect to the fixed inertial frame, the body frame will move accordingly. We will denote the body frame coordinates using (X, Y, Z) .

To describe the rotational motion of the object, the first step is to clarify what a rotation does to any given vector (scalars of course are unaffected). Suppose we have a vector $\vec{a}$, and we rotate the frame by an angle θ . Now without any loss of generality, we can assume that we rotate about the z axis, by simply defining the axis of rotation to be z in the first place¹¹. The rotated vector will be given by

$$a'_i = R_{ij}(\theta)a_j \quad (5.1)$$

where we have now switched to index notation for clarity. The rotation matrix takes the well known form

$$R(\theta) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (5.2)$$

Since we're interested in dynamics, it's important to understand this action infinitesimally. Expanding the rotation matrix for small $\delta\theta$, we find

¹¹Sneaky tricks like this are commonplace, it's best to get used to them.

$$a'_i = a_i + (-\theta a_y, \theta a_x, 0) \quad (5.3)$$

or

$$\delta \vec{a} = \delta \vec{\theta} \times \vec{a} \quad (5.4)$$

where $\delta \vec{\theta} = \delta \theta \hat{z}$. Since the choice of z as the axis of rotation was arbitrary, this formula holds more generally, with the only change being noting that $\delta \vec{\theta}$ is along the axis of rotation, with magnitude being given by the small angular change.

Since this is an infinitesimal equation, and holds for any vector, we may divide throughout by δt to obtain an equation for the time derivative, where now $\vec{\omega} = \frac{d\vec{\theta}}{dt}$ is now the angular frequency, the rate at which the object in question is being rotated. By setting $\vec{a} = \vec{r}$ we find

$$\vec{v} = \vec{\omega} \times \vec{r}. \quad (5.5)$$

It's important to keep in mind that this is the change in position, with respect to the fixed coordinate frame, due to rotations. The object has a whole (the center of mass, really) may be moving with some velocity as well, so the general formula for it's velocity *vis a vis* the fixed frame would be

$$\vec{v}_{\text{tot}} = \vec{v}_{\text{com}} + \vec{\omega} \times \vec{r} \quad (5.6)$$

where $\vec{v}_{\text{com}}$ in this case would be the movement of the center of mass of the extended object, while $\vec{r}$ would be the location of some point on it. For this reason any discussion of rigid bodies by its construction is only in the sense of describing the various points on the extended body independently, assuming that their relative positions do not change. This of course is no longer possible to assume when the objects are deformable, a situation we will not consider in the present text.

Keeping this in mind, we can derive the Lagrangian of a rigid body, specifically the kinetic piece. Since we assume that the particles composing the rigid body do not change their relative positions, we can approximate the extended body as a system of discrete particles, and simply add up their respective kinetic energies in the fixed coordinate frame,

$$L = \sum_m \frac{1}{2} m \vec{v}_{\text{tot}}^2 \quad (5.7)$$

where the sum over m is meant to invoke the idea that we are actually summing over all discrete particles that compose the rigid body. We'll make this more precise in what follows. Starting by expanding this out, we get

$$L = \sum_m \frac{1}{2} m (\vec{v}_{\text{com}}^2 + (\vec{\omega} \times \vec{r})^2) + \sum_m m \vec{v}_{\text{com}} \cdot (\vec{\omega} \times \vec{r}). \quad (5.8)$$

The last term can be simplified

$$\sum_m m \vec{v}_{\text{com}} \cdot (\vec{\omega} \times \vec{r}) = (\vec{v}_{\text{com}} \times \omega) \cdot \sum_m m \vec{r} \quad (5.9)$$

where we have brought out the $\vec{\omega}$ and $\vec{v}_{\text{com}}$, as they are the same for all of the particles composing the extended body. Noting that $\sum_m m \vec{r}$ is proportional to the center of mass of the motion, we lose no generality by assuming an inertial frame of reference in which the center of mass is at rest at the origin. This is just because we're only interested in the effects of rotation, and the full range of motion may be recovered by simply going to the moving frame. This lets us drop the last term (we also drop the first, which is just the motion of the center of mass), giving us

$$L = \frac{1}{2} I_{ij} \omega_i \omega_j \quad (5.10)$$

where I_{ij} (which we have found by expanding the cross product of the second term) is given by

$$I_{ij} = \sum_m m (r^2 \delta_{ij} - r_i r_j) \quad (5.11)$$

and is known as the inertia tensor. Now to make the sum itself precise, this is easy enough. Since we've approximated the extended body as a sum over discrete particles, we may simply drop this and rewrite the sum as an integral instead. Each m then becomes a unit mass $\rho(\vec{r})dV$, where $\rho(\vec{r})$ is the density of the object. We then have

$$I_{ij} = \int_O \rho(\vec{r}) (r^2 \delta_{ij} - r_i r_j) dV. \quad (5.12)$$

A common tradition historically was to compute I_{ij} for very strange looking objects, which is entertaining, but largely unnecessary practically. Linear algebra is more useful here. Notice that since I_{ij} is a symmetric tensor, it can be diagonalized. In a frame of reference in which the coordinate system is chosen such that I_{ij} is diagonal, we may always write

$$I_{ij} = \begin{pmatrix} I_x & 0 & 0 \\ 0 & I_y & 0 \\ 0 & 0 & I_z \end{pmatrix} \quad (5.13)$$

and

$$L = \frac{1}{2} I_x \omega_x^2 + \frac{1}{2} I_y \omega_y^2 + \frac{1}{2} I_z \omega_z^2 \quad (5.14)$$

which is the Lagrangian for a free object that may be rotating, in the chosen coordinate system. It should be remarked that in general, when we're dealing with a

rotating body, we're not really dealing with a system that is invariant under Galilean symmetry in the strictest sense anymore. This is because even in free rotation, the body may exhibit asymmetries (the height may be different from the width, say), which would manifestly break the invariance under rotations. This means that the frame of reference that diagonalizes the inertia tensor would itself be dynamical, and change during the motion. For this reason, it turns out to actually be easier to work in the rotating frame, the consequences of which we will encounter soon.

But to come back, upshot of the simple form of the Lagrangian is that the Euler-Lagrange equations are equivalently clarifying; since we have no potential term the derivative of L with respect to ω_i is a constant:

$$\frac{d\ell_i}{dt} = 0 \quad (5.15)$$

where

$$\begin{aligned} \ell_i &= \frac{\partial L}{\partial \omega_i} \\ &= I_{ij}\omega_j \end{aligned} \quad (5.16)$$

is the angular momentum. The name isn't an accident; it's identical to the angular momentum of planets we discussed earlier. In that case the fixed frame was the one in which the sun was at rest, while the rotation was about the center of mass. It's a simple enough exercise; we have

$$\vec{\ell} = \vec{r} \times \vec{p}. \quad (5.17)$$

Setting $\vec{p} = m\vec{v} = m\vec{\omega} \times \vec{r}$, we can evaluate the triple cross product

$$\vec{\ell} = m\vec{\omega}r^2 - m\vec{r}(\vec{r} \cdot \vec{\omega}) \quad (5.18)$$

or in index notation

$$\ell_i = m(r^2\delta_{ij} - r_i r_j)\omega_j. \quad (5.19)$$

Since this is true for any m , this is equivalent to the formula in (5.16). This means that formulas can be derived for the definition of angular momentum for discrete masses, after which we can just translate into the continuum form.

5.2 The Equations of Motion

5.2.1 The Euler Equations

The Euler-Lagrange equations derived in the previous section reveal a canonical feature of rotations; nontrivial rotations are ones where the angular momentum is not conserved. Recall that the free Euler-Lagrange equations are just

$$\frac{d\ell_i}{dt} = 0. \quad (5.20)$$

Naturally, we would expect this to change when there is a force that's being applied on the bodies. Fortunately, we can derive this without having to make direct reference to a Lagrangian. To start with, let's do it for discrete particles, and then move to the continuum description as mentioned earlier. Writing this out,

$$\dot{\vec{\ell}} = \dot{\vec{r}} \times \vec{p} + \vec{r} \times \dot{\vec{p}} \quad (5.21)$$

which noting that $\dot{\vec{r}} = \vec{v}$ (and therefore parallel to $\vec{p}$) and $\dot{\vec{p}} = \vec{F}$ we have

$$\dot{\vec{\ell}} = \vec{r} \times \vec{F}. \quad (5.22)$$

The quantity $\vec{\tau} = \vec{r} \times \vec{F}$ is known as the torque. The Euler-Lagrange equation is pretty simple now

$$\frac{d\ell_i}{dt} = \tau_i. \quad (5.23)$$

Keep in mind that this has been derived in the fixed frame of reference.

In principle these are all the equations of motion that we need to get a handle on the exact dynamics of any rigid body. Unfortunately, things are complicated by the fact that in the fixed frame of reference, the inertia tensor and the ω_i are in general not constants of the motion, which makes the dynamical equations quite a bit more difficult to deal with. A clever way to bypass this is to move to the rotating frame instead. Recall that the rotating frame is related to the fixed frame by an angular velocity, given by ω_i . In particular, we can isolate out the change in the angular momentum that is due purely to the rotation; in the fixed system, we can split the change in ℓ_i into a change that is purely due to the rotation, as well as an intrinsic change within the rotating frame. We do this by setting

$$\left(\frac{d\vec{\ell}}{dt} \right)_{\text{rotation}} = \vec{\omega} \times \vec{\ell}. \quad (5.24)$$

By then decomposing the time derivative of the angular momentum, into a fixed part (corresponding to its change in the moving frame) and the rotational part, we find *Euler's equations*¹²

$$\left(\frac{d\vec{\ell}}{dt} \right) + \vec{\omega} \times \vec{\ell} = \vec{\tau}. \quad (5.25)$$

¹²Do not confuse these with the Euler-Lagrange equations. Euler discovered a number of things, his name shows up a lot.

So long as we keep in mind that the components of $\vec{\ell}$ are to be treated in the moving frame, we can drop the brackets in favor of clarity. Further, assume that within the moving frame, we always pick a coordinate system such that I_{ij} is diagonal. This makes the Euler equations particularly simple

$$\begin{aligned} I_1 \frac{d\omega_1}{dt} + (I_3 - I_2)\omega_2\omega_3 &= \tau_1 \\ I_2 \frac{d\omega_2}{dt} + (I_1 - I_3)\omega_3\omega_1 &= \tau_2 \\ I_3 \frac{d\omega_3}{dt} + (I_2 - I_1)\omega_1\omega_2 &= \tau_3. \end{aligned} \tag{5.26}$$

The limit of *free rotation* is found by setting $\tau_i = 0$. We'll see that there are cases where free rotation is fairly nontrivial.

5.2.2 Free Symmetric Tops

The case of free rotations is especially interesting because it happens to be amenable to solution for all possible choices of I_i . When all of them, or two out of three, are equal, we call such rotating bodies under free rotation *free symmetric tops*. In the completely asymmetric case, where all three moments of inertia are different, the solution turns out to be far harder to obtain, and as such we postpone a discussion of that to a later section. But the cases of the free symmetric tops are quite a bit simpler.

The fully symmetric case $I_i = I$ is completely trivial, since the second terms disappear from the Euler equations; each of the equations reduces down to

$$I\dot{\omega}_i = 0 \tag{5.27}$$

or that $\omega_i = c_i$ for constants c_i . An equivalent form of this is to see that

$$\ell_i = I\omega_i \tag{5.28}$$

which tells us that the angular velocities are just proportional to the angular momenta, which are themselves conserved. Fortunately, the axisymmetric configuration, where two out of three I_i are equal, is a bit more interesting. Without loss of generality, we can assume that $I_1 = I_2 = I$ and $I_3 = I'$. The Euler equations are

$$\begin{aligned} I\dot{\omega}_1 + \Delta I\omega_2\omega_3 &= 0 \\ I\dot{\omega}_2 - \Delta I\omega_3\omega_1 &= 0 \\ I'\dot{\omega}_3 &= 0 \end{aligned} \tag{5.29}$$

where $\Delta I = I' - I$ has been substituted for clarity. The last of these is solved readily; $\omega_3 = \Omega$ is just a constant. Now we see that the first two can be written more cleverly as a matrix equation

$$I \frac{d}{dt} \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix} = \begin{pmatrix} 0 & -\Delta I \Omega \\ \Delta I \Omega & 0 \end{pmatrix} \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix} \quad (5.30)$$

An amusing way of solving this is to see that the matrix

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad (5.31)$$

is the two-dimensional representation of the imaginary unit $-i$. As a result, the matrix equation is really equivalent to

$$\frac{d\vec{\omega}}{dt} = -i \frac{\Delta I \Omega}{I} \vec{\omega} \quad (5.32)$$

where we have abused notation to write $\vec{\omega}$ is the row vector defined previously. Resolving the particular solution $\exp\left(\frac{\Delta I \Omega}{I} t\right)$ into the real and imaginary parts using this matrix dictionary, we find

$$\begin{aligned} \omega_1 &= \omega_0 \sin\left(\frac{\Delta I \Omega}{I} t\right) \\ \omega_2 &= -\omega_0 \cos\left(\frac{\Delta I \Omega}{I} t\right). \end{aligned} \quad (5.33)$$

The angular velocity in the body frame is not a constant; we see that it is periodic, with period given by $\frac{2\pi I}{\Delta I \Omega}$. The technical term for it is *precession*.

As it happens, this turns out to be applicable to the free rotation of the Earth. The Earth is an oblate spheroid, and can be approximated as an axisymmetric top. This results in an overall precession that can be observed in the movement of stars on the celestial sphere over time. Per century, the rate of precession of the Earth's rotation is about 5029". In fact, this is the largest component of the *measured* precession rate of the perihelion of mercury. It's only after adjusting for this precession that the true precession due to perturbations of the other planets and general relativity can be isolated.

5.3 Euler Angles

As we've seen, it's convenient to switch between the body frame and the fixed space frame when dealing with the motion of rigid bodies; depending on exactly what we're trying to answer, working in one or the other coordinate system may be necessary. In order to avoid the messiness of keeping track of two different frames of reference, it's useful to have a dynamical dictionary that interpolates between them. This dictionary is commonly known as *Euler angles*.

Put simply, Euler angles are a set of three angles that allow us to move between the body frame and the space frame. Since these angles translate between a fixed frame and one that's in motion, the Euler angles themselves will be dynamical.

Although it's not always practical to solve the Euler equations directly in terms of Euler angles, they may be computed after the fact, telling us exactly how the motion is expressed by these angles.

There are several conventions for Euler angles, some used more than others. It's also the case that different fields use different conventions, and this can cause a fair bit of confusion. As a result, we'll first illustrate the idea of Euler angles intuitively, pick a convention (the physics one), and stick to it.

The main idea of Euler angles is that we only need three angles at any given point to determine the transformation from one coordinate system to another. A full proof of this isn't especially illuminating, but we can visualize this relatively simply in our heads. As before, denote the axes in the body frame by (x, y, z) , and the fixed coordinate system by (X, Y, Z) . Since $\hat{z}$ will just be some unit vector that (generally speaking) will not be in the YZ plane, we can simply rotate by some angle α around the Z axis to bring z into the YZ plane. Let's call the transformation of (x, y, z) under this transformation (x', y', z') . In other words

$$R_z(\alpha) : (x, y, z) \longrightarrow (x', y', z') \quad (5.34)$$

where z' is in the YZ plane. Now to align z' along Z , we can perform a rotation around x' (recall that we need to keep z' on its plane, so any of the two perpendicular directions will do; we just pick x'). If this angle is β , we now have the image

$$R_x(\beta) : (x', y', z') \longrightarrow (x', y'', Z). \quad (5.35)$$

The remaining coordinates (x', y'') will simply be related to (X, Y) by some rotation γ around Z ; we have then

$$R_z(\gamma) : (x', y'', Z) \longrightarrow (X, Y, Z). \quad (5.36)$$

This means that the combined rotation

$$R(\alpha, \beta, \gamma) = R_z(\gamma)R_x(\beta)R_z(\alpha) \quad (5.37)$$

can interpolate between the body and fixed frames. When the body is in motion, the angles (α, β, γ) will themselves be dynamical quantities. It should be remarked that this particular zxz convention is the standard one picked in physics. Engineers use different conventions, but they can all be shown to ultimately work.

It isn't easy to visualize the Euler angles, but calculating with them isn't nearly as hard. Generally, keep in mind that the change in the last Euler angle α corresponds to intrinsic rotation of the body about its z axis, β corresponds to a kind of wobbling motion, known as nutation, while γ corresponds to extrinsic rotation, known as precession.

The full rotation matrix is useful when written down, but it's too cumbersome to fit into the page. But it's easy enough to calculate when needed, using the following rotation matrices

$$R_x(\theta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & -\sin(\theta) \\ 0 & \sin(\theta) & \cos(\theta) \end{pmatrix} \quad (5.38)$$

and

$$R_z(\theta) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (5.39)$$

One particular case worth calculating is when $R(\alpha, \beta, \gamma)$ is applied to a vector along the z axis. We find in this case

$$R(\alpha, \beta, \gamma) \begin{pmatrix} 0 \\ 0 \\ M \end{pmatrix} = \begin{pmatrix} M \sin \beta \sin \gamma \\ -M \sin \beta \cos \gamma \\ M \cos \beta \end{pmatrix}. \quad (5.40)$$

5.3.1 Angular Velocity in the Body Frame

As we've already discussed, all of this is to make it possible for us to work in the body frame, since we at least have the benefit that the moments of inertia are constant in that frame. As a result, we only need to keep track of the angular changes of the body frame with respect to the fixed frame, and together with the information about the moments of inertia in the body frame, we have all we need to completely characterize the dynamics.

The first step in this will be to write down the angular velocity in Euler angles. The easiest way to do this is to first write them in the intermediate coordinate systems, and then move back into the body frame. Keep in mind that we have the following: $\dot{\alpha}$ corresponds to the rotation around z , $\dot{\beta}$ corresponds to the rotation around x' , and $\dot{\gamma}$ corresponds to the rotation around Z . So we have

$$\vec{\omega} = \dot{\alpha}\hat{z} + \dot{\beta}\hat{x}' + \dot{\gamma}\hat{Z}. \quad (5.41)$$

We now use the following relationships (abusing notation somewhat)

$$\begin{aligned} (x', y', z') &= R_z(\alpha)(x, y, z) \\ (x', y'', Z) &= R_x(\beta)R_z(\alpha)(x, y, z) \end{aligned} \quad (5.42)$$

to infer that

$$\hat{x}' = \cos(\alpha)\hat{x} - \sin(\alpha)\hat{y} \quad (5.43)$$

and

$$\hat{Z} = \sin(\alpha) \sin(\beta) \hat{x} + \sin(\beta) \cos(\alpha) \hat{y} + \cos(\beta) \hat{z}. \quad (5.44)$$

We can then write down the angular velocities in terms of the Euler angles directly in the body frame

$$\begin{aligned} \omega_x &= \dot{\beta} \cos(\alpha) + \dot{\gamma} \sin(\alpha) \sin(\beta) \\ \omega_y &= -\dot{\beta} \sin(\alpha) + \dot{\gamma} \cos(\alpha) \sin(\beta) \\ \omega_z &= \dot{\alpha} + \dot{\gamma} \cos(\beta) \end{aligned} \quad (5.45)$$

5.3.2 The Heavy Symmetric Top

The use of Euler angles opens up to solution the more complicated case of a top, which although axisymmetric, is not free. Specifically, it's heavy, and as a consequence is subject to gravity. We characterize gravity as a constant force along the z -axis, and the potential function according to

$$U = -m\vec{g} \cdot \vec{h} \quad (5.46)$$

where the vector $\vec{h}$ points towards the center of mass and m is the mass of the top. That this is how gravity acts on extended objects is a theorem that can be proven using analytical statics. We won't do so here, since the proof isn't particularly illuminating. But this will do for our purposes. We assume that for the top $I_x = I_y = I_1$ and $I_z = I_2$, which means that the center of mass is along the z axis. In the inertial frame, this would be on Z , which tells us that

$$\vec{h} = \begin{pmatrix} 0 \\ 0 \\ h \end{pmatrix} \quad (5.47)$$

where h is the height of the center of mass. In the rotating frame, this is

$$\vec{h} = \begin{pmatrix} h \sin \beta \sin \gamma \\ -h \sin \beta \cos \gamma \\ h \cos \beta \end{pmatrix} \quad (5.48)$$

and given that $\vec{g}$ is pointed along $-Z$, the potential function is just

$$U = mgh \cos(\beta). \quad (5.49)$$

The kinetic part of the Lagrangian can be computed directly using the Euler angles; we have

$$\frac{1}{2}I_1(\omega_1^2 + \omega_2^2) + \frac{1}{2}I_2\omega_3^2 = \frac{1}{2}I_1(\dot{\beta}^2 + \dot{\gamma}^2 \sin^2(\beta)) + \frac{1}{2}I_2(\dot{\alpha} + \dot{\gamma} \cos(\beta))^2. \quad (5.50)$$

The Lagrangian is now

$$L = \frac{1}{2}I_1 \left(\dot{\beta}^2 + \dot{\gamma}^2 \sin^2(\beta) \right) + \frac{1}{2}I_2 (\dot{\alpha} + \dot{\gamma} \cos(\beta))^2 - mgh \cos(\beta). \quad (5.51)$$

AS in the Kepler problem, the first step is to search for conserved quantities. The lack of any explicit time dependence naturally results in the energy

$$E = \frac{1}{2}I_1 \left(\dot{\beta}^2 + \dot{\gamma}^2 \sin^2(\beta) \right) + \frac{1}{2}I_2 (\dot{\alpha} + \dot{\gamma} \cos(\beta))^2 + mgh \cos(\beta) \quad (5.52)$$

being one of them. The others are obtained through the usual use of Noether's theorem. Since there is no α or γ dependence, we obtain two more conserved quantities

$$\frac{\partial L}{\partial \dot{\alpha}} = I_2 (\dot{\alpha} + \dot{\gamma} \cos(\beta)) \quad (5.53)$$

which is just the angular momentum ℓ_α and

$$\frac{\partial L}{\partial \dot{\gamma}} = I_1 \dot{\gamma} \sin^2(\beta) + I_2 \cos(\beta) (\dot{\alpha} + \dot{\gamma} \cos(\beta)). \quad (5.54)$$

The trick now is to observe that we can express $\dot{\alpha}$ and $\dot{\gamma}$ in terms of only β , and then show that β itself can be reduced to a problem of quadrature. Defining the two constants

$$\begin{aligned} \mu &= \frac{1}{I_1} \ell_\alpha \\ \lambda &= \frac{1}{I_1} \frac{\partial L}{\partial \dot{\gamma}} \end{aligned} \quad (5.55)$$

we can write

$$\dot{\gamma} = \frac{\lambda - \mu \cos(\beta)}{\sin^2(\beta)} \quad (5.56)$$

and

$$\dot{\alpha} = \frac{I_1}{I_2} \mu - \frac{(\lambda - \mu \cos(\beta)) \cos(\beta)}{\sin^2(\beta)}. \quad (5.57)$$

The problem then reduces to one that's only dependent on finding a solution for β as a function of time. This is done by using the conserved energy. Rewriting it in terms of the two angular conserved quantities, we have

$$E = \frac{1}{2}I_1 \dot{\beta}^2 + V_{\text{eff}}(\beta) \quad (5.58)$$

where the effective potential is

$$V_{\text{eff}}(\beta) = \frac{1}{2}I_1 \frac{(\lambda - \mu \cos(\beta))^2}{\sin^2(\beta)} + \frac{\ell_\alpha^2}{2I_2} + mgh \cos(\beta). \quad (5.59)$$

To reduce this to quadrature, the following standard substitution is used

$$\begin{aligned} u &= \cos(\beta) \\ \dot{u} &= -\dot{\beta} \sin(\beta) \end{aligned} \quad (5.60)$$

which allows us to rearrange the energy to reveal

$$\dot{u}^2 = f(u) \quad (5.61)$$

where

$$f(u) = \left(\frac{2E}{I_1} - \frac{\ell_\alpha^2}{I_1 I_2} \right) (1 - u^2) - (\lambda - \mu u)^2 - \frac{2mgh}{I_2} u(1 - u^2). \quad (5.62)$$

We can simplify $f(u)$ further, but we don't really need to. The most important thing is to observe that it's a cubic function, which means that in terms of its roots $r_{1,2,3}$ and an overall scaling factor c ¹³, it can be written as

$$f(u) = c(u - r_1)(u - r_2)(u - r_3). \quad (5.63)$$

In other words, the solution to the problem of quadrature is the elliptic integral

$$\frac{du}{\sqrt{(u - r_1)(u - r_2)(u - r_3)}} = \sqrt{c} \times dt. \quad (5.64)$$

Ordering the roots $r_1 < r_2 < r_3$, the substitution

$$u = r_1 + (r_2 - r_1)y \quad (5.65)$$

with

$$k^2 = \frac{r_2 - r_1}{r_3 - r_1} \quad (5.66)$$

puts the differential equation into the form

$$\frac{dy}{\sqrt{4y(y - 1)(k^2 y - 1)}} = \sqrt{\frac{c(r_3 - r_1)}{4}} \times dt \quad (5.67)$$

which seems like a handful. Fortunately, the left side turns out to be the differential form of the sn^2 function, with modulus k argument $\frac{t}{\tau}$ where

¹³It's easy enough to see that $c = \frac{2mgh}{I_2}$, which has units of $[\text{Time}]^{-2}$.

$$\tau^2 = \frac{4}{c(r_3 - r_1)}. \quad (5.68)$$

In other words

$$\cos(\beta(t)) = r_1 + (r_2 - r_1) \operatorname{sn}^2\left(\frac{t}{\tau}; k\right). \quad (5.69)$$

The period of this is just $4\tau K(k)$, where $K(k)$ is the complete elliptic integral of the second kind. This means that the period motion corresponding to wobbling is periodic, but more complicated than a simple sinusoid.

It isn't really worth further integrating these expressions to get α and γ as functions of time (integrals of elliptic functions are pretty messy, and they all integrate to various elliptic integrals with the same modulus anyway). But some qualitative features can be extracted from the quadrature form.

The time dependence in the quadrature form of the solutions comes from the fact that u is a function of time, and $f(u)$ is a cubic function. For a given motion, the time evolution will generically be between two of the roots, say u_1 and u_2 . The precise shape of the motion will then depend on whether or not the sign of α changes during the motion. The three possibilities are:

1. $\dot{\alpha} > 0$ at both roots, in which case the top simply traces a non-intersecting loop,
2. $\dot{\alpha}$ changes signs, in which case the top traces out a loop that intersects itself and
3. $\dot{\alpha} = 0$ at one of the roots, which results in a singular cusp, before it traces back its path along the sphere.

Virtually all standard books have pretty pictures describing these motions. We will not be reproducing them.

5.4 The Free Asymmetric Top

We come now to the free asymmetric top, where all three moments of inertia are different and the top itself is not acted upon by any external force. This includes gravity; we assume the top is massless.

The free rotation of an asymmetric top turns out to have extremely complex dynamics, and involves a fairly intricate interplay of elliptic functions. Since it's an abnormally complex problem that actually admits of an exact solution, it's worth pushing through the full analysis. This section may be skipped by those generally uninterested in elliptic functions and related mathematical techniques, but it's worth doing as a matter of technical exercise.

First, the Lagrangian. It's very simple, since there is no potential term; we have

$$L = \frac{1}{2}I_1\omega_1^2 + \frac{1}{2}I_2\omega_2^2 + \frac{1}{2}I_3\omega_3^2. \quad (5.70)$$

We also note that due to manifest time independence, we obtain $E = L$, a conserved quantity.

The fact that the rotation is free means that the angular momentum is a conserved quantity in the inertial frame. We make the simplifying assumption that the angular momentum is aligned with the Z axis. Useful later will be the expression of this in the rotating frame, which we obtain by using the earlier expression for $\hat{Z}$ in the body frame,

$$\vec{\ell} = \ell \sin(\alpha) \sin(\beta) \hat{x} + \ell \cos(\alpha) \sin(\beta) \hat{y} + \ell \cos(\beta) \hat{z}. \quad (5.71)$$

This doesn't cost us any generality, but will turn out to be useful later. Since the angular momentum is conserved, we note that its magnitude

$$\ell^2 = I_1^2 \omega_1^2 + I_2^2 \omega_2^2 + I_3^2 \omega_3^2 \quad (5.72)$$

is also conserved. Finally, we choose the ordering $I_1 < I_2 < I_3$.

The reduction to quadratures is done by eliminating two out of three of the ω_i , and then solving the Euler equation of the third. The convention here has been to solve the Euler equation for ω_2 , although this isn't a hard and fast rule. Eliminating ω_3 and ω_1 in turn, we find the following relations

$$\begin{aligned} \omega_1^2 &= \frac{(2EI_3 - \ell^2) - I_2(I_3 - I_2)\omega_2^2}{I_1(I_3 - I_1)} \\ \omega_3^2 &= \frac{(\ell^2 - 2EI_1) - I_2(I_2 - I_1)\omega_2^2}{I_3(I_3 - I_1)} \end{aligned} \quad (5.73)$$

which we substitute into

$$\dot{\omega}_2 = \frac{I_3 - I_1}{I_2} \omega_1 \omega_3. \quad (5.74)$$

The algebraic simplifications are straightforward, but tedious enough to render the steps worth skipping. Getting some common factors out, we get

$$\dot{\omega}_2 = \frac{1}{I_2 \sqrt{I_2 I_3}} \sqrt{(2EI_3 - \ell^2)(\ell^2 - 2EI_1)} \sqrt{1 - \left(\frac{\omega_2}{\Omega_2}\right)^2} \sqrt{1 - \left(\frac{\omega_2}{\Omega_3}\right)^2} \quad (5.75)$$

where

$$\begin{aligned} \Omega_1^2 &= \frac{2EI_3 - \ell^2}{I_1(I_3 - I_1)} \\ \Omega_2^2 &= \frac{2EI_3 - \ell^2}{I_2(I_3 - I_2)} \\ \Omega_3^2 &= \frac{\ell^2 - 2EI_1}{I_3(I_3 - I_1)}. \end{aligned} \quad (5.76)$$

We have defined Ω_1^2 in anticipation of its role later. These three auxiliary constants will help in simplifying some of the later expressions for the angular momenta. We now rescale by defining

$$\begin{aligned} u &= \frac{\omega_2}{\Omega_2} \\ k &= \frac{\Omega_2}{\Omega_3} \\ \frac{1}{\tau} &= \sqrt{\frac{(I_3 - I_2)(\ell^2 - 2EI_1)}{I_1 I_2 I_3}} \end{aligned} \tag{5.77}$$

the Euler equation simplifies considerably; we have

$$\tau \dot{u} = \sqrt{(1 - u^2)(1 - k^2 u^2)}. \tag{5.78}$$

Now we can recognize this as the differential equation for the elliptic function sn with modulus k , but with the time variable rescaled according to $t \rightarrow t/\tau = s$. Thus,

$$u = \text{sn}(s; k). \tag{5.79}$$

We can see, basically by inspection at this point that we have

$$\vec{\omega} = (\Omega_1 \text{cn}(s; k), \Omega_2 \text{sn}(s; k), \Omega_3 \text{dn}(s; k)). \tag{5.80}$$

We see that in the rotating frame, $\vec{\omega}$ exhibits the periodicity characteristic of elliptic functions; the replacement

$$s \rightarrow s + 4K(k) \tag{5.81}$$

where $K(k)$ is the complete elliptic integral of the first kind leaves the components invariant. We thus see that there is period of $4K(k)\tau$. This completes the solution of the angular momenta in the moving frame. But we would also like to know how this changes with respect to the fixed inertial frame, for which we have to find the corresponding Euler angles as a function of time.

The analysis to follow is famously complicated, and the only complete version I am aware of is due to Whittaker. It involves a rather deft manipulation of theta functions, much of which isn't particularly clarifying. As a result, in the following I will provide a more direct calculation, albeit at the price of making the final result a little less manifest.

Two of the Euler angles turn out to be relatively easy to recover from the $\vec{\omega}$, and the last is essentially the source of all the difficulty. Specifically, recall that our convention was that $\vec{\ell} = \ell \hat{Z}$. Expanding this in the rotating frame, we have

$$\vec{\ell} = \ell \sin(\alpha) \sin(\beta) \hat{x} + \ell \cos(\alpha) \sin(\beta) \hat{y} + \ell \cos(\beta) \hat{z}. \tag{5.82}$$

Relating this to the fact that each component of the angular momentum satisfies $\ell_i = I_i \omega_i$, we can immediately identify

$$\begin{aligned}\tan(\alpha) &= \frac{I_1 \Omega_1 \operatorname{cn}(s; k)}{I_2 \Omega_2 \operatorname{sn}(s; k)} \\ \cos(\beta) &= \frac{I_3 \Omega_3}{\ell} \operatorname{dn}(s; k)\end{aligned}\tag{5.83}$$

leaving γ as the last Euler angle to be determined. Unfortunately, this requires going to the expansion of $\vec{\omega}$ in Euler angles. We use the following

$$\begin{aligned}\omega_1 &= \dot{\beta} \cos(\alpha) + \dot{\gamma} \sin(\alpha) \sin(\beta) \\ \omega_2 &= -\dot{\beta} \sin(\alpha) + \dot{\gamma} \cos(\alpha) \sin(\beta).\end{aligned}\tag{5.84}$$

Eliminating $\dot{\beta}$,

$$\dot{\gamma} = \frac{\omega_1 \sin(\alpha) + \omega_2 \cos(\alpha)}{\sin(\beta)}.\tag{5.85}$$

Finally, we use

$$\begin{aligned}\ell \sin(\alpha) \sin(\beta) &= I_1 \omega_1 \\ \ell \cos(\alpha) \sin(\beta) &= I_2 \omega_2 \\ \frac{I_1^2 \omega_1^2 + I_2^2 \omega_2^2}{\ell^2} &= \sin^2(\beta)\end{aligned}\tag{5.86}$$

to eliminate α and β ; we find

$$\dot{\gamma} = \ell \frac{I_1 \omega_1^2 + I_2 \omega_2^2}{I_1^2 \omega_1^2 + I_2^2 \omega_2^2}.\tag{5.87}$$

Given that ω_1 and ω_2 are elliptic functions, this promises to be a difficult integral. The complexity here isn't really reducible very much, it's a feature of the problem itself. Inserting the elliptic function solutions to ω_1 and ω_2 and using the identity

$$\operatorname{cn}^2(s; k) + \operatorname{sn}^2(s; k) = 1\tag{5.88}$$

we find the expression

$$\dot{\gamma} = \ell \frac{I_1 \Omega_1^2 + (I_2 \Omega_2^2 - I_1 \Omega_1^2) \operatorname{sn}^2(s; k)}{I_1^2 \Omega_1^2 + (I_2^2 \Omega_2^2 - I_1^2 \Omega_1^2) \operatorname{sn}^2(s; k)}.\tag{5.89}$$

Integrating this is achieved by partial fractioning out the sn^2 from the numerator; it's an painful but elementary exercise from which we get

$$\dot{\gamma} = \ell \frac{I_2 \Omega_2^2 - I_1 \Omega_1^2}{I_2^2 \Omega_2^2 - I_1^2 \Omega_1^2} + \ell \left(\frac{(I_2^2 - I_1 I_2) \Omega_2^2}{I_1 (I_2^2 \Omega_2^2 - I_1^2 \Omega_1^2)} \right) \frac{1}{1 + n \operatorname{sn}^2(s; k)}\tag{5.90}$$

where

$$n = \frac{I_2^2 \Omega_2^2 - I_1^2 \Omega_1^2}{I_1^2 \Omega_1^2}. \quad (5.91)$$

This is actually integrable; the incomplete elliptic integral of the third kind is defined by

$$\Pi(n; \operatorname{am}(x; k) | k) = \int_0^x \frac{dx}{1 + n \operatorname{sn}^2(x; k)} \quad (5.92)$$

where $\operatorname{am}(x; k)$ is the Jacobi period function. We do need to keep in mind that s is related to t by a scaling of $1/\tau$, but the final result is

$$\gamma(s) - \gamma_0 = s \left(\tau \ell \frac{I_2 \Omega_2^2 - I_1 \Omega_1^2}{I_2^2 \Omega_2^2 - I_1^2 \Omega_1^2} \right) + \tau \ell \left(\frac{(I_2^2 - I_1 I_2) \Omega_2^2}{I_1 (I_2^2 \Omega_2^2 - I_1^2 \Omega_1^2)} \right) \Pi(n; \operatorname{am}(s; k) | k) \quad (5.93)$$

where we have expressed everything in dimensionless variables to keep things consistent. From a formal viewpoint, this is a rather clean expression, but it obscures some of the periodicity. Recall that γ is a phase. This means that we need to check periods up to 2π . Essentially, we need to find transformations of s that leave $e^{i\gamma(s)}$ invariant.

The first term in $\gamma(s)$ is simply proportional to s , so we leave that aside for the moment. Π is periodic, but only up to an additive constant. In particular

$$\Pi(n; \operatorname{am}(s + 2K(k); k) | k) = \Pi(n; \operatorname{am}(s; k) | k) + 2\Pi(n; k). \quad (5.94)$$

This quasiperiodicity may be eliminated by adding and subtracting $\frac{s}{2K(k)} \Pi(n; k)$; more precisely we decompose

$$\gamma(s) - \gamma_0 = \gamma_1(s) + \gamma_2(s) \quad (5.95)$$

where

$$\gamma_1 = s \tau \ell \left(\frac{I_2 \Omega_2^2 - I_1 \Omega_1^2}{I_2^2 \Omega_2^2 - I_1^2 \Omega_1^2} + \frac{1}{K(k)} \frac{(I_2^2 - I_1 I_2) \Omega_2^2}{I_1 (I_2^2 \Omega_2^2 - I_1^2 \Omega_1^2)} \Pi(n; k) \right) \quad (5.96)$$

and

$$\gamma_2(s) = \tau \ell \left(\frac{(I_2^2 - I_1 I_2) \Omega_2^2}{I_1 (I_2^2 \Omega_2^2 - I_1^2 \Omega_1^2)} \right) \left(\Pi(n; \operatorname{am}(s; k) | k) - \frac{s}{K(k)} \Pi(n; k) \right). \quad (5.97)$$

γ_2 now exhibits invariance under the shift $s \longrightarrow s + 2mK(k)$ where m is any integer. This is equivalent to saying that it is invariant under $t \longrightarrow t + 2K(k)\tau$. This is half

the period we found for the ω_i . That said, note that $e^{i\gamma_1}$ remains the same when $t \longrightarrow t + T'$ ¹⁴ where

$$\frac{2\pi}{T'} = \ell \left(\frac{I_2\Omega_2^2 - I_1\Omega_1^2}{I_2^2\Omega_2^2 - I_1^2\Omega_1^2} + \frac{1}{2K(k)} \frac{(I_2^2 - I_1I_2)\Omega_2^2}{I_1(I_2^2\Omega_2^2 - I_1^2\Omega_1^2)} \Pi(n; k) \right). \quad (5.98)$$

What this means is that not only do γ_1 and γ_2 have different periods, but the ratio of the periods is a generically transcendental function. Further, T' is incommensurate with $4K(k)\tau$ as well, the periods of $\alpha(t)$ and $\beta(t)$. The practical consequence of this is the fact that the motion of the top never crosses the same point twice.

5.5 Non-Inertial Reference Frames

Any two inertial reference frames are related by a Galilean transformation, and do not as such encode any dynamics. In a sense, the study of accelerating frames is by its very nature the study of non-inertial reference frames, since the coordinate system that moves with an accelerating body naturally provides an instance of a fundamentally non-inertial frame. That said, there is something to be said about understanding what happens to the laws of motion in the non-inertial frames themselves.

We already saw an avatar of this when we studied rotations. The Euler equations, although difficult to solve in the rest frame, became much more tractable in the rotating frame. Unfortunately, in connecting those results to the lab frame we had to use Euler angles, which reintroduced the same complexity. In this section, we will generalize this translation by looking at the generic relationship between a given inertial reference frame and another that is related to the first by a non-inertial transformation.

5.5.1 Fictitious Forces

Say we have a free particle in an inertial frame of reference. It's naturally described by the free Euler-Lagrange equations (assuming for the sake of simplicity a one-dimensional motion)

$$m\ddot{x} = 0. \quad (5.99)$$

Now say we observe this particle from a frame of reference that is related to the first by a non-inertial transformation. If we express this in terms of a velocity boost, we can write down the relationship between the velocity in the new frame and the velocity in the old frame more quantitatively as

¹⁴It's worth remarking that a serious attempt has been made to prove the equivalence of this to the more famous result due to Whittaker. Despite my best efforts however, this particular proof has eluded me, although there is little doubt that it is the case. The full proof will likely involve an expansion of the complete and incomplete elliptic integrals of the third kind in theta functions, and a conspiracy between the resulting expressions to produce the classical result.

$$\dot{x}'(t) = \dot{x}(t) - \xi(t) \quad (5.100)$$

where $\xi(t)$ is no longer a constant. The Euler-Lagrange equations in the new frame now become

$$m\ddot{x}' = -m\dot{\xi}. \quad (5.101)$$

We see that the non-inertial relationship characterized by $\xi(t)$ is experienced in the new frame in terms of a “force”, but not one due to a potential or any interaction. The force is fictitious, an artifact due to the new reference frame not being an inertial one. Further, since the introduction of the non-inertial transformation manifestly involves time, energy will no longer be a conserved quantity in the non-inertial frame.

Fictitious forces show up every so often when we would like to encapsulate what may be a complicated set of forces acting on a system in terms of a coordinate transformation. It may seem unnecessary to do so, but consider the reverse scenario. A body that is subject to a constant force field (to keep things concrete) satisfies the Euler-Lagrange equations

$$m\ddot{x} = f \quad (5.102)$$

where f is a constant. This means that the acceleration is inside a non-inertial frame where we can observe the body falling under the force. Now consider a new frame where

$$x' = x - \frac{f}{2}t^2. \quad (5.103)$$

In this frame, the Euler-Lagrange equation is

$$\ddot{x}' = 0. \quad (5.104)$$

As it turns out, the ability to make such transformations—ones that render a non-inertial frame inertial—is purely local. In the presence of real forces, there can be no global transformation of the frame of reference to eliminate the forces and bring the Euler-Lagrange equations into a free particle version. This insight lies at the heart of general relativity, which exploits this fact to recast gravity as a fundamentally local phenomenon, one that we observe due to local changes in the spacetime geometry.

To come back to our earlier discussion, the effect of a non-inertial transformation in the presence of a potential can also be studied. Say we have the following Lagrangian

$$L = \frac{1}{2}m\dot{x}^2 - U(x) \quad (5.105)$$

in an inertial frame. Moving to the non-inertial frame we find

$$L = \frac{1}{2}m\dot{x}'^2 + m\dot{x}'\xi(t) + \frac{1}{2}\dot{\xi}^2 - U(x'). \quad (5.106)$$

Fortunately, the third term in the Lagrangian is trivially the time derivative of some function, and drops out of the Euler-Lagrange equations. Direct computation now gives us the modified Euler-Lagrange equations

$$m\ddot{x}' = -\frac{\partial U}{\partial x'} - m\ddot{\xi}. \quad (5.107)$$

Parenthetically, this can be absorbed into the potential by defining

$$U_{\text{eff}} = U(x') - m\dot{x}'\ddot{\xi}. \quad (5.108)$$

Once again, via an appropriate choice of $\xi(t)$, a particle under the influence of a force may “appear” to move freely, by making U_{eff} (it’s derivative, more precisely) vanish.

Needless to say, by introducing indices and dot products this entire analysis can be repeated in 3 dimensions as well.

5.5.2 Rotating Frames

The equations of motion in a non-inertial reference frame, when specified to rotations, end up being especially simple and illuminating. Simply put, we look at a frame that is rotating with some angular velocity $\vec{\Omega}$ with respect to an inertial frame; the velocity in the new frame will be given by

$$\vec{v}' = \vec{v} + \vec{r} \times \vec{\Omega} \quad (5.109)$$

in line with the relationship between proper and angular velocity we have already derived. The Lagrangian of a free particle is then transformed to

$$L = \frac{1}{2}m\vec{v}'^2 - m\vec{v} \cdot (\vec{r} \times \vec{\Omega}) + \frac{1}{2}m(\vec{r} \times \vec{\Omega})^2. \quad (5.110)$$

To calculate the Euler-Lagrange equations, we first note that

$$\begin{aligned} \frac{\partial L}{\partial \vec{r}} &= m\vec{v}' \times \vec{\Omega} + \vec{\Omega} \times (\vec{r} \times \vec{\Omega}) \\ \frac{\partial L}{\partial \vec{v}'} &= m\vec{v}' - m\vec{r} \times \vec{\Omega} \end{aligned} \quad (5.111)$$

whence we find

$$m\dot{\vec{v}}' = 2m\vec{v}' \times \vec{\Omega} + \vec{\Omega} \times (\vec{r} \times \vec{\Omega}) + m\vec{r} \times \dot{\vec{\Omega}}. \quad (5.112)$$

In writing the Euler-Lagrange equations we have used the fact that $\dot{\vec{r}} = \vec{v}'$, since the rotational motion does not affect the radial vector.

We see that the introduction of the non-inertial frame induces three fictitious forces. The first of these, given by $2m\vec{v}' \times \vec{\Omega}$ is known as the *Coriolis force*. The

second, due to $\vec{\Omega} \times (\vec{r} \times \vec{\Omega})$ is called the *centrifugal force*. The third given by $m\vec{r} \times \dot{\vec{\Omega}}$ is a result of non-uniform rotations, and vanishes when $\vec{\Omega}$ is a constant. We'll assume that hereafter, to keep the analysis simple. It doesn't affect the dynamics quantitatively anyway.

A classic application of the Euler-Lagrange equations in a rotating frame of reference is to the small harmonic motions of a suspended object, as affected by the rotations of the Earth. Align the axis of rotation of the Earth with the z axis, and neglect small motions along it. Now the equations of motion in the xy plane (for any potential really) for small $\vec{\Omega} = \Omega \hat{z}$ (thereby neglecting the centrifugal force) will become

$$\begin{aligned}\ddot{x} + \omega_0^2 x &= 2\dot{y}\Omega \\ \ddot{y} + \omega_0^2 y &= -2\dot{x}\Omega.\end{aligned}\tag{5.113}$$

The normal modes are the solutions to

$$\det \begin{pmatrix} -\omega^2 + \omega_0^2 & 2i\omega\Omega \\ -2i\omega\Omega & -\omega^2 + \omega_0^2 \end{pmatrix} = 0\tag{5.114}$$

as an equation in ω . For small Ω , we find

$$\omega_{\pm} = \pm\omega_0 + \Omega.\tag{5.115}$$

This means that a simple pendulum will record the effect of the rotation as a small change in its period. This device is known as the *Foucault pendulum*¹⁵. The additional Ω acts as a precession, and defines a second period over which the pendulum must move to return to its starting point.

Now it turns out that in the case of the Earth, Ω is actually latitude dependent, given by $\Omega(\varphi) = \Omega_0 \sin(\varphi)$, where φ runs from 0 to 90 degrees on each hemisphere. Ω_0 itself is roughly 7.3×10^{-5} radians per second. For instance, at the Tropic of Cancer, $\varphi = 23.5^\circ$, and we have $\Omega = 2.9 \times 10^{-5}$ radians per second. This means that in

$$T = \frac{2\pi}{\Omega(23.5^\circ)} \times \frac{\text{hr}}{3600} = 59.95 \text{ hours}\tag{5.116}$$

the pendulum undergoes one full precession due to the Coriolis force. For comparison, at the North Pole the same calculation would yield a period of 23.9 hours. The length of the day, in other words.

¹⁵Not to be confused with the book; also not to be confused with the French philosopher.

6 The Phase Space Formalism

The use of Lagrangians took us from simply discussing dynamics in terms of laws of motion to expressing mechanics in terms of the interactions between various degrees of freedom of the system. Quantitatively, we describe the degrees of freedom in terms of the generalized coordinates and generalized velocities, of which the Lagrangian is an invariant function. Occasionally however, it ends up being more convenient to use a dual system of generalized variables, which generally speaking is referred to as the Hamiltonian formalism.

At the same time, the use of the dual set of variables characterizes a geometric space known as phase space. Further, a number of dynamical laws can be rephrased as “trajectories” on this generalized phase space. Although a lot of this ends up being a little unnecessary and abstract when dealing with classical systems, their importance becomes more pronounced in the context of quantum mechanical systems. In this chapter, we discuss this formalism in some depth.

6.1 The Hamiltonian

We’ve actually seen the basic transformation to phase space already. Recall that we had defined generalized momenta, in addition to generalized velocities, in terms of the derivative

$$p_i = \frac{\partial L}{\partial \dot{q}_i}. \quad (6.1)$$

In moving from ordinary $(q_i, \dot{q}_i)$ coordinates to phase space, the essential idea is that we use p_i instead of $\dot{q}_i$. This has the effect of “democratizing” the degrees of freedom somewhat; instead of having one set of variables defined as derivatives of the other, we have two *a priori* independent sets of degrees of freedom, constrained by the equations of motion. The algorithm for this is as follows. Consider first, assuming that the Lagrangian has no time dependence, the total derivative

$$dL = \frac{\partial L}{\partial q_i} dq_i + \frac{\partial L}{\partial \dot{q}_i} d\dot{q}_i \quad (6.2)$$

which in terms of the generalized momenta becomes

$$dL = \dot{p}_i dq_i + p_i d\dot{q}_i. \quad (6.3)$$

Note that in this we have made use of the Euler-Lagrange equations. Critically, the transformations to follow only apply when we assume that the equations of motion are satisfied. Since we would like to remove the dependence on $\dot{q}_i$, we focus on the second term; rewrite it as a total derivative of the form

$$p_i d\dot{q}_i = d(p_i \dot{q}_i) - \dot{q}_i dp_i \quad (6.4)$$

and note that we now have

$$dH = -\dot{p}_i dq_i + \dot{q}_i dp_i \quad (6.5)$$

where

$$H = p_i \dot{q}_i - L. \quad (6.6)$$

The function H is known as the Hamiltonian. Recalling Noether's theorem, the Hamiltonian is an expression of the total energy of the system. In defining it, we need to ensure that the $\dot{q}_i$ are appropriately expressed in terms of p_i ; in general, the Hamiltonian is a function of the form $H = H(q_i, p_i)$.

From (6.6), we note the following relations

$$\begin{aligned} \frac{\partial H}{\partial q_i} &= -\dot{p}_i, \\ \frac{\partial H}{\partial p_i} &= \dot{q}_i \end{aligned} \quad (6.7)$$

known as Hamilton's equations. Note first that unlike in the case of Lagrangian mechanics, for N particles we have $6N$ equations here, in contrast with the $3N$ Euler-Lagrange equations.

Consider the most general kind of Lagrangian, given by a free and interaction part

$$L = \sum_i \frac{1}{2} m \dot{q}_i^2 - U(q_i) \quad (6.8)$$

where we assume for the sake of simplicity that all the masses are identical. The generalized momenta then become

$$\frac{\partial L}{\partial \dot{q}_i} = m \dot{q}_i. \quad (6.9)$$

Now computing

$$\sum_i p_i \dot{q}_i - L = \sum_i \frac{1}{2} m \dot{q}_i^2 + U(q_i). \quad (6.10)$$

Recalling that we need to rewrite this in terms of q_i and p_i , we note

$$\dot{q}_i = \frac{p_i}{m} \quad (6.11)$$

giving us

$$H = \sum_i \frac{1}{2m} p_i^2 + U(q_i). \quad (6.12)$$

To see the equivalence with Lagrangian mechanics, note that the first set of Hamilton's equations give us

$$\frac{\partial U}{\partial q_i} = -\dot{p}_i \quad (6.13)$$

which is simply the Euler-Lagrange equations. The second set of Hamilton's equations just act as a restatement of the definition of the generalized momenta.

6.2 Poisson Brackets

6.2.1 Flows in Phase Space

In general, the use of Hamiltonians is more a matter of seeing symmetries and conservation laws, rather than actually simplifying calculations. The reason for this is that rather than determining the equations of motion *per se*, the Hamiltonian formalism recasts the equations of motion as a kind of flow in phase space. In the following, we'll see how this works.

Consider a function that is naturally defined on phase space, and may or may not be time-dependent. For N particles, such a function will naturally depend on the $6N$ variables given by the generalized coordinates q_i and generalized momenta p_i , as well as t . For such a function, we naturally have the identity

$$df(t, q_i, p_i) = \frac{\partial f}{\partial t} dt + \sum_i \left(\frac{\partial f}{\partial q_i} dq_i + \frac{\partial f}{\partial p_i} dp_i \right). \quad (6.14)$$

Dividing throughout by dt , we obtain

$$df(t, q_i, p_i) = \frac{\partial f}{\partial t} dt + \sum_i \left(\frac{\partial f}{\partial q_i} \dot{q}_i + \frac{\partial f}{\partial p_i} \dot{p}_i \right). \quad (6.15)$$

Substituting Hamilton's equations for the time derivative, we find

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \sum_i \left(\frac{\partial f}{\partial q_i} \frac{\partial H}{\partial p_i} - \frac{\partial f}{\partial p_i} \frac{\partial H}{\partial q_i} \right). \quad (6.16)$$

The quantity inside the large parenthesis is known as the *Poisson bracket*; more specifically for two functions f and g on phase space we define it as

$$\{f, g\} = \sum_i \left(\frac{\partial f}{\partial q_i} \frac{\partial g}{\partial p_i} - \frac{\partial f}{\partial p_i} \frac{\partial g}{\partial q_i} \right). \quad (6.17)$$

The Poisson bracket has some formal properties that make it especially easy to prove theorems on evolution in phase space. For one thing, it acts as a derivative on the space of functions; note that we have

$$\{f, gh\} = \{f, g\}h + g\{f, h\} \quad (6.18)$$

as a trivial consequence of the product rule. An especially interesting (and slightly annoying to prove) is the cyclic rule, also known as the Jacobi identity¹⁶

$$\{f, \{g, h\}\} + \{g, \{h, f\}\} + \{h, \{f, g\}\} = 0. \quad (6.19)$$

More trivially, the Poisson bracket is bilinear and antisymmetric as well. Given the Poisson bracket, a distinguished operator emerges, that controls time evolution in phase space, denoted by

$$\hat{L} = -\{H, \cdot\} \quad (6.20)$$

whose action on any function in phase space becomes

$$\hat{L}f = -\{H, f\}, \quad (6.21)$$

known as the *Liouville* operator. This results now in the formal equivalence

$$\frac{d}{dt} = \partial_t + \hat{L}. \quad (6.22)$$

A nice physical picture revealed by this is as follows: time evolution on phase space breaks down naturally into two distinct pieces. On the one hand we have a purely temporal piece, which governs changes of functions at every point separately. The second piece is more intrinsic to the structure of phase space itself, and sensitive to the specifics of the Hamiltonian defined on it. In other words, for functions independent of time, the Hamiltonian characterizes a natural “flow” on phase space; we find that in this case

$$\dot{f} = -\{H, f\} \quad (6.23)$$

or equivalently that for conserved quantities, their Poisson brackets with H identically vanish. The realization of this fact as a flow can be made more concrete by noting that the Hamilton equations can themselves be written as

$$\begin{aligned} \dot{q}_i &= \{q_i, H\} \\ \dot{p}_i &= \{p_i, H\}. \end{aligned} \quad (6.24)$$

Once again, this encodes the basic idea that the Hamiltonian traces out flows in phase space; if the position and momenta are constants along the motion, they naturally have vanishing Poisson brackets with the Hamiltonian. This arises in the case of free particle motion for instance; there we have a Hamiltonian that only depends on p_i , resulting in a vanishing Poisson bracket with all of the p_i , aligning with our intuition that a free particle always has constant momentum.

¹⁶Recall that the vector cross product also satisfies the Jacobi identity.

The Liouville operator in a sense is really a map, that takes a point in phase space and returns a pair of vectors, which characterize the flow along the physical trajectories specified by the Hamiltonian.

6.2.2 Coordinate and Momentum Brackets

A particular class of Poisson brackets, namely those between simple functions of the generalized coordinates and momenta reveal a simple algebraic structure, which although relatively quaint in classical physics, translates largely verbatim to quantum mechanics. To begin with, notice that we have

$$\{q_i, p_j\} = \delta_{ij} \quad (6.25)$$

which can be verified by direct calculation. Notice that defining the operator (in the same manner as we defined the Liouville operator)

$$P_i = -\{p_i, \cdot\} \quad (6.26)$$

we obtain the relationship

$$\delta q P_i(q_i) = \delta q. \quad (6.27)$$

Although this looks trivial, what it means is that the operator P_i generates small variations of q_i ; by acting $1 + \delta q P_i$ on q_i , we get $q_i + \delta q$, which corresponds to a small variation. More generally, by direct calculation we have¹⁷

$$P_i f(q_i) = \frac{\partial f}{\partial q_i} \quad (6.28)$$

resulting in the formal equivalence

$$P_i \longleftrightarrow \frac{\partial}{\partial q_i}. \quad (6.29)$$

Entirely analogously, a similar relationship can be established between

$$Q_i = \{q_i, \cdot\} \quad (6.30)$$

and p_i .

Whenever there is a bracket between the generalized coordinates and momenta in this fashion, they are said to be *canonically conjugate variables*. This is a fancy way of saying that we have a natural structure imposed by the dynamics—the Poisson bracket structure in this case—that results in the coordinates and momenta acting as each others “duals”. The coordinates generate small variations of the momenta, and vice versa. It turns out that there is a more or less verbatim translation of these

¹⁷It’s an amusing Taylor series exercise to verify that $\exp(\delta q P_i)f(q_i) = f(q_i + \delta q)$.

relations to quantum mechanics, where the Poisson bracket structure becomes the commutator between position and momenta operators treated as matrices. In that context, the fact that the commutator is non-zero is the origin of the uncertainty principle, as it renders the two operators mutually incommensurate.

6.3 Canonical Transformations

Since the Hamiltonian formalism is naturally defined on trajectories obeying the equations of motion, there are implicit constraints imposed on transformations of the generalized coordinates and generalized momenta. Only those transformations that respect these conditions are permitted, since other transformations will naturally move us outside the allowed trajectories. We call such transformations canonical. In the present section, we will study canonical transformations in some detail, and see that not only do they unify the kinematic and dynamical aspects of the formalism itself, but also reveal rather elegant underlying mathematics.

6.3.1 The Symplectic Structure

The most elementary definition of a canonical transformation is a transformation of the form

$$(q_i, p_i) \longrightarrow (\tilde{q}_i(q_i, p_i), \tilde{p}_i(q_i, p_i)) \quad (6.31)$$

which are guaranteed to preserve the form of Hamilton's equations. Recall that the particular structure of Hamilton's equations were derived assuming that we stayed on the solutions of the equations of motion; canonical transformations as a result may alter the specific appearance of the Hamiltonian, but the equations themselves cannot change.

That said, there are two ways to understand canonical transformations. One is to work out the precise conditions they have to follow by hand, which is somewhat tedious, and the second is to exploit a little additional mathematical abstraction—beyond what we've already established—to see more conceptually what is going on. The approach we follow here is to write the Hamilton equations in a manner that clarifies what the canonical transformations are doing. We start by collecting the generalized coordinates into a single vector with $6N$ components as follows

$$\xi_I = (q_i, p_i) \quad (6.32)$$

where I runs from 1 to $6N$. Now defining the $6N \times 6N$ matrix

$$J = \begin{pmatrix} 0 & 1_{3N \times 3N} \\ -1_{3N \times 3N} & 0 \end{pmatrix} \quad (6.33)$$

Hamilton's equations can be written in a more unified way as

$$\dot{\xi}_I = J_{IJ} \frac{\partial H}{\partial \xi_I}. \quad (6.34)$$

Phase space endowed with a matrix of the form taken by J is said to be endowed with a *symplectic structure*. We'll see now that canonical transformations are those that preserve the symplectic structure.

Consider a canonical transformation $\xi_I \longrightarrow \xi'_I$. We now require that Hamilton's equations stay invariant under it. This means that we must have

$$\dot{\xi}'_I = J_{IJ} \frac{\partial H}{\partial \xi'_J} \quad (6.35)$$

which upon effecting the change of variables says

$$\frac{\partial \xi'_I}{\partial \xi_J} \dot{\xi}_J = J_{IJ} \frac{\partial H}{\partial \xi_K} \frac{\partial \xi_K}{\partial \xi'_J}. \quad (6.36)$$

To simplify notation, we define the matrices

$$\frac{\partial \xi'_I}{\partial \xi_J} = M_{IJ}^{-1} \quad (6.37)$$

so that we have, in matrix notation

$$M^{-1} \dot{\xi} = J M^T \frac{\partial H}{\partial \xi} \quad (6.38)$$

or equivalently

$$\dot{\xi} = (M J M^T) \frac{\partial H}{\partial \xi}. \quad (6.39)$$

For this to be equivalent to Hamilton's equations, we must have

$$J = M J M^T. \quad (6.40)$$

Although this looks abstract, consider the fact that we've already seen a version of this before. Recall what happens when we perform an ordinary rotation in 3 dimensions; it results in the coordinates being transformed according to

$$x_i \longrightarrow O_{ij} x_j \quad (6.41)$$

where O is an orthogonal matrix. But orthogonal simply means that it preserves the identity matrix, which is the “metric structure” on flat space. Rotations are transformations that preserve the metric structure, and canonical transformations are those that preserve the symplectic structure. The theme of endowing spaces with a structure and then looking at transformations that preserve it is a recurring theme in physics.

In addition to being easier to define, this way of characterizing canonical transformations has the benefit of making more or less manifest other features of canonical transformations. For instance, canonical transformations preserve the Poisson bracket. To see this, observe that the Poisson bracket in the transformed coordinates can be rewritten as

$$\{f, g\} = \sum_{I,J} \frac{\partial f}{\partial \xi'_I} J_{IJ} \frac{\partial g}{\partial \xi'_J}. \quad (6.42)$$

To show that this is equivalent to the original Poisson brackets, note that

$$\sum_{I,J,K,L} \frac{\partial f}{\partial \xi'_K} M_{KI} J_{IJ} M_{JL}^T \frac{\partial g}{\partial \xi'_L} = \sum_{K,L} \frac{\partial f}{\partial \xi'_K} J_{KL} \frac{\partial g}{\partial \xi'_L} \quad (6.43)$$

which is simply the Poisson bracket in the original phase space variables. An alternative, and admittedly simpler way of seeing this is to observe that

$$\{\xi_I, \xi_J\} = J_{IJ} \quad (6.44)$$

making the preservation of the Poisson bracket under canonical transformations manifest.

6.3.2 Infinitesimal Canonical Transformations

Since a canonical transformation is simply a map between two different coordinate systems, we can study what happens when we perform an infinitesimal transformation, allowing us to discover the local mechanism of preserving the symplectic structure. Consider then the following infinitesimal map

$$\begin{aligned} q_i &\longrightarrow q_i + \epsilon u_i(q_i, p_i) \\ p_i &\longrightarrow p_i + \epsilon v_i(q_i, p_i) \end{aligned} \quad (6.45)$$

where ϵ is an infinitesimal constant. The Jacobian matrix M^{18} is then given by

$$M^{-1} = \begin{pmatrix} 1 + \epsilon \frac{\partial u_i}{\partial q_j} & \epsilon \frac{\partial u_i}{\partial p_j} \\ \epsilon \frac{\partial v_i}{\partial q_j} & 1 - \epsilon \frac{\partial v_i}{\partial p_j} \end{pmatrix}. \quad (6.46)$$

By enforcing the invariance of the symplectic structure under this transformation (and dropping anything that's order ϵ^2), we find the following condition

$$\frac{\partial u_i}{\partial q_j} = -\frac{\partial v_j}{\partial p_i}. \quad (6.47)$$

This actually happens to be the condition for the vector (u_i, v_i) to be a total derivative; in other words

¹⁸Keep in mind that $MJM^T = J$ also guarantees that $M^{-1}JM^{-1T} = J$ applies as well.

$$\begin{aligned} u_i &= \frac{\partial G}{\partial p_i} \\ v_i &= -\frac{\partial G}{\partial q_i}. \end{aligned} \tag{6.48}$$

The technical word for such a function G is the generator of the infinitesimal transformations. An interesting upshot of this is the following. Consider what happens under an infinitesimal time evolution of a point $(q_i(t), p_i(t))$,

$$(q_i(t), p_i(t)) \longrightarrow (q_i(t), p_i(t)) + \delta t (\dot{q}_i(t), \dot{p}_i(t)) \tag{6.49}$$

But recalling now the relations

$$\begin{aligned} \dot{q}_i &= \frac{\partial H}{\partial p_i} \\ \dot{p}_i &= -\frac{\partial H}{\partial q_i} \end{aligned} \tag{6.50}$$

we see that the Hamiltonian is the generator of the canonical transformation corresponding to time evolution. An alternative way of thinking about this is that time evolution is simply a transformation of phase space that preserves the symplectic structure, one that happens to be generated by the Hamiltonian.

Finally, let's see how the Hamiltonian responds to canonical transformations. Consider the variation of the Hamiltonian under a canonical transformation

$$\delta H = \sum_i \left(\frac{\partial H}{\partial q_i} \delta q_i + \frac{\partial H}{\partial p_i} \delta p_i \right). \tag{6.51}$$

If this is generated by a generator G , we can write

$$\delta H = \epsilon \left(\frac{\partial H}{\partial q_i} \frac{\partial G}{\partial p_i} - \frac{\partial H}{\partial p_i} \frac{\partial G}{\partial q_i} \right) \tag{6.52}$$

which equates to

$$\delta H = \epsilon \{H, G\}. \tag{6.53}$$

When the Hamiltonian is invariant under this canonical transformation, we see then that the Poisson bracket between H and G identically vanishes. This also tells us that G is an invariant of the motion. The physical meaning of this is that invariants of the motion, also known as conserved quantities, generate canonical transformations in phase space. In fact, they generate a particular class of canonical transformations that actually leave the Hamiltonian as well as the symplectic structure invariant.

6.4 Liouville's Theorem

An ancillary consequence of characterizing canonical transformations in terms of generators is an elementary proof of Liouville's theorem, which simply states that the volume element of phase space stays the same along physical trajectories. In fact, it also stays the same under any canonical transformation, as we will see. Recall the volume element in phase space

$$d\Gamma = \prod_i dq_i dp_i. \quad (6.54)$$

Now let's perform an infinitesimal canonical transformation generated by G , according to

$$\begin{aligned} q_i &\longrightarrow q_i + \epsilon \frac{\partial G}{\partial p_i} \\ p_i &\longrightarrow p_i - \epsilon \frac{\partial G}{\partial q_i} \end{aligned} \quad (6.55)$$

We then have for the transformations of dq_i and dp_i the following

$$\begin{aligned} dq_i &\longrightarrow dq_i + \epsilon \frac{\partial^2 G}{\partial p_i \partial q_i} dq_i \\ dp_i &\longrightarrow dp_i - \epsilon \frac{\partial^2 G}{\partial q_i \partial p_i} dp_i \end{aligned} \quad (6.56)$$

Now it's elementary to see that to leading order in ϵ , this preserves the phase space factor; note that

$$dq_i dp_i \longrightarrow dq_i dp_i \left(1 + \epsilon \frac{\partial^2 G}{\partial p_i \partial q_i} - \epsilon \frac{\partial^2 G}{\partial p_i \partial q_i} + \mathcal{O}(\epsilon^2) \right) = dq_i dp_i. \quad (6.57)$$

Incidentally, the statement of Liouville's theorem is immediate as a result of this; since physical trajectories simply correspond to canonical transformations where $G = H$ and $\epsilon = \delta t$, the volume of phase space is always conserved during the time evolution of the physical system.

Crucially, an important implication of Liouville's theorem is that it's really just the volume of a given region of phase space that has to stay the same with time evolution. In fact, the time evolution can deform the shape of the region beyond recognition; the integrated volume however always remains what it was at the beginning of the evolution.

A Calculus of Variations

The core result of the calculus of variations is the Euler-Lagrange equations, and generally using variations of functions to minimize functionals. Since these have been largely covered in the main text, in this appendix we will focus on a couple of classic problems that motivated the study of the calculus of variations.

A.1 Minimizing Arc-Lengths

The calculus of variations is a systematic approach to the problem of how to minimize a number that's dependent on a function. The elementary example that clarifies the salience of this question is the arc-length for a general curve $f(x)$

$$\text{Length} = \int_a^b \sqrt{1 + f'(x)^2} \, dx. \quad (\text{A.1})$$

Given this arc-length, the question is to find the $f(x)$ for which this length is minimized. It appears *a priori* that the problem is rather difficult; we need to consider *all possible functions* $f(x)$, and pick out the one that minimizes the arc length. While the functional analysis involved in formulating this rigorously can get rather tedious, the fortunate fact is that for most spaces of functions that aren't especially pathological, solutions to problems of this kind aren't too hard to get.

In this particular case the manner of getting at this is as follows. The function that minimizes the arc length is a function $f(x)$ such that performing a small variation $f(x) \rightarrow f(x) + \delta f(x)$ which obeys

$$\delta f(a) = \delta f(b) = 0 \quad (\text{A.2})$$

and renders the difference

$$\text{Length}[f + \delta f] - \text{Length}[f] = 0. \quad (\text{A.3})$$

In other words, the *variational derivative* of the arc length vanishes when evaluated at the function that minimizes it. If what we're doing is correct, the result obtained from this should match our intuition, that the shortest distance between two points is a straight line.

To verify this, let's do the variational calculation. To leading order in δf , we get

$$\delta \text{Length} = \int_a^b \frac{f'(x) \delta f'(x)}{\sqrt{1 + f'(x)^2}} dx. \quad (\text{A.4})$$

Since the primary variation was in terms of f , not of f' , we need to integrate this by parts; doing this we get

$$\delta \text{Length} = \left. \frac{f'(x) \delta f(x)}{\sqrt{1 + f'(x)^2}} \right|_a^b - \int_a^b \frac{d}{dx} \left(\frac{f'(x)}{\sqrt{1 + f'(x)^2}} \right) \delta f(x) dx. \quad (\text{A.5})$$

Fortunately, the boundary conditions guarantee that the term outside the integral vanishes, since δf is 0 at the boundaries. Further, since the variation itself was arbitrary, to ensure that the variation of the arc-length vanishes we must have

$$\frac{d}{dx} \left(\frac{f'(x)}{\sqrt{1 + f'(x)^2}} \right) = 0. \quad (\text{A.6})$$

Solving this, we find

$$\text{const} \times f'(x)^2 = 1 \quad (\text{A.7})$$

which is just the statement that $f'(x)$ is a constant, or that $f(x)$ is a linear function of x . In other words, the function that minimizes the length between two points is in fact a straight line, as our intuition tells us.

A.2 The Brachistochrone Curve

The brachistochrone curve is a cycloid; a curve formed by a point placed on a wheel as the wheel moves (without slipping). It turns out that this is the solution to a classic problem in the calculus of variations. Specifically: what is the curve between two points that minimizes the time of travel as a body travels down it under gravity? Solving this is an instructive example of how to treat variational problems, and even offers an avatar of the conserved quantities we found that helped in integrated the equations of motion in mechanical systems.

We want to find a curve that minimizes the time of descent between two given points (x_1, y_1) and (x_0, y_0) . We characterize the trajectory by the locus (x, y) , where x and y are the distances from x_0 and y_0 respectively. The first step is to define the time of descent. To do this, note that the time differential is given by

$$dt = \frac{ds}{v(x, y)} \quad (\text{A.8})$$

where $v(x, y)$ is the velocity of the particle at the point (x, y) . This can actually be inferred, assuming the particle starts at velocity $v = 0$ at $y = 0$ by conserving energy

$$mgy(x) = \frac{1}{2}mv^2 \quad (\text{A.9})$$

or that

$$v = \sqrt{2gy}. \quad (\text{A.10})$$

Recalling now that we have

$$ds = \sqrt{1 + y'(x)^2} dx \quad (\text{A.11})$$

we find for the functional to minimize the following time integral

$$T(a, b) = \int_a^b \sqrt{\frac{1 + y'^2}{2gy}} dx. \quad (\text{A.12})$$

The most straightforward solution would be obtained by solving the Euler-Lagrange equations. But we've actually seen an easier approach to this already. Note that in the present problem, x really plays the role of 'time', and y, y' play the role of the generalized coordinate and velocity. This means, that y virtue of the manifest independence of the Lagrangian on time, we must have that the "Hamiltonian"

$$y' \frac{\partial L}{\partial y'} - L = \text{const} \quad (\text{A.13})$$

where

$$L = \sqrt{\frac{1 + y'^2}{2gy}}. \quad (\text{A.14})$$

In the most general formulation of variational problems, this is known as Beltrami's identity, and holds whenever there is a manifest independence of the 'time' variable. The specific avatar of this in the context of classical particles happens to be the conservation of energy.

After some algebra, the conservation of the Hamiltonian results in the following differential equation:

$$y' \sqrt{\frac{y}{A - y}} = 1. \quad (\text{A.15})$$

This can either be solved directly, or the substitution $y = A \cos^2 \theta$ can be used, where θ is a parameter. Simplifying, we get

$$-2A \cos^2 \theta d\theta = dx. \quad (\text{A.16})$$

Collecting everything, we find

$$\begin{aligned} x(\theta) &= -\frac{A}{2}(2\theta - \sin 2\theta) \\ y(\theta) &= \frac{A}{2}(1 + \cos 2\theta) \end{aligned} \quad (\text{A.17})$$

The precise range of θ and value of A will be determined by solving

$$\begin{aligned} x(\theta_f) &= x_2 - x_1 \\ y(\theta_f) &= y_2 - y_1. \end{aligned} \quad (\text{A.18})$$

B Fourier Analysis

B.1 The Fourier Transform

Fourier transforms arose out of the study of periodic functions. Say we have a function defined on the interval $[-T, T]$, with periodic boundary conditions. This simply means that

$$f(t + 2T) = f(t). \quad (\text{B.1})$$

The two most elementary functions that satisfy this are the standard sine and cos waves,

$$\begin{aligned} s_n(t) &= \sin\left(\frac{n\pi}{T}t\right) \\ c_n(t) &= \cos\left(\frac{n\pi}{T}t\right). \end{aligned} \quad (\text{B.2})$$

Each of these functions “wraps” around the circle with circumference $2T$ a total of n times as t goes from 0 to $2T$. Now it turns out that any function (up to a constant) on $[-T, T]$, so long as it isn’t particularly pathological can be expanded in a linear basis given by these functions; basically we have

$$f(t) = a_0 + \sum_{n=1}^{\infty} (a_n s_n(t) + b_n c_n(t)) \quad (\text{B.3})$$

where (a_n, b_n) are constants. The constants themselves are found by supplying the right boundary conditions of f . For instance, if $f(T) = f(-T) = 0$, we would require all of the b_n to vanish.

Since it can be annoying to keep track of sines and cosines separately, what is often done is to absorb them into a complex exponential, but rewriting the function as

$$f(t) = \sum_{n=-\infty}^{\infty} f_n \exp\left(\frac{in\pi}{T}t\right) \quad (\text{B.4})$$

where now the f_n are complex constants, and simply taking the real limit at the end. This has the added advantage of eliminating the asymmetry of adding the constant part by hand.

Fourier analysis is the study of what happens when we take $T \rightarrow \infty$. Essentially, we’re now studying all possible functions on the real line, with certain conditions on the function that prohibit it from being especially singular. While the rigorous study of this limit happens to be very interesting mathematically, most salient for our purposes is the fact that this also turns out to be very relevant to physical systems, especially those which have fundamentally oscillatory behavior.

The Fourier transform is formally defined as follows. Note that from (B.4), we must have

$$f_n = \frac{1}{2T} \int_{-T}^T f(t') \exp\left(-\frac{in\pi}{T}t'\right) dt'. \quad (\text{B.5})$$

Replacing this into the formula (B.4) we find

$$f(t) = \int_{-T}^T f(t') \Delta_T(t - t') dt' \quad (\text{B.6})$$

where

$$\Delta_T(t - t') = \frac{1}{2T} \sum_{n=-\infty}^{\infty} \exp\left(\frac{in\pi}{T}(t - t')\right). \quad (\text{B.7})$$

Now in the limit of $T \rightarrow \infty$, let's define

$$\frac{2\pi}{2T} = d\omega. \quad (\text{B.8})$$

The rationale for this is that if we define the “frequencies” in this interval by the series $\omega_n = \frac{n\pi}{T}$, in the limit of very large T , the difference between two consecutive frequencies approaches an infinitesimal. The sum over n then really becomes an integral over ω , and we have

$$\Delta_T(t - t') \rightarrow \delta(t - t') = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega(t-t')} d\omega. \quad (\text{B.9})$$

This quantity is known as the Dirac delta function. It's not really a function, but it's called one. We'll get back its properties later.

The limiting case of the Fourier series can now be obtained as well. Note that (B.4) becomes

$$f(t) = \frac{1}{2\pi} \int \hat{f}(\omega) e^{i\omega t} d\omega \quad (\text{B.10})$$

where

$$\hat{f}(\omega) = \int_{-\infty}^{\infty} e^{-i\omega t} f(t) dt. \quad (\text{B.11})$$

We have fiddled around with the constants to arrive at these. The reader should reproduce these formulas himself. Formula (B.11) is known as the *Fourier transform* and (B.10) is known as the inverse Fourier transform. Different authors use different conventions for the factors of 2π and the signs in the exponentials (in fact, we have used the opposite convention for the signs in the main text). None of these really affect the final results if proper care is taken to keep track of them.

B.2 The Dirac Delta Function

As mentioned earlier, the quantity defined as

$$\delta(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega t} d\omega \quad (\text{B.12})$$

is known as the Dirac delta function. Furthermore, as pointed out, it's not really a function.

The Dirac delta function is what is known as a distribution. It's a dual of a function; it takes a function and returns a number. In other words, under an integral sign, the Dirac delta function does much the same thing a functional does. We've already seen one of these properties in the limiting case of (B.6),

$$f(t) = \int_{-\infty}^{\infty} f(t') \delta(t - t') dt. \quad (\text{B.13})$$

This property is known as the sifting property of the Dirac delta function. It shifts the value of the function by an amount set in the argument. Another property of the Dirac delta function is that the integral of it over the line is 1. To see this, note that

$$\int_{-T}^T \Delta_T(t) dt = 1 \quad (\text{B.14})$$

since only the constant term survives; all other exponentials drop out when integrated over a period. Now take the limit of $T \rightarrow \infty$ to get

$$\int_{-\infty}^{\infty} \delta(t) dt = 1. \quad (\text{B.15})$$

In fact, we can see from the definition of the Dirac delta function $\delta(t)$ that it actually vanishes everywhere other than at $t = 0$, where it's formally infinite. It should be clear at this point that no classical function can satisfy this, although it does happen to be the case that a number of classical functions can smoothly approximate it. One such case is the function

$$\delta_R(t) = \frac{1}{2R} \quad (\text{B.16})$$

for $t \in [-R, R]$ and zero otherwise. This function has the shape of a tall step, where the step runs to infinity as we approach $R \rightarrow 0$. In fact, there is a rigorous sense in which

$$\lim_{R \rightarrow \infty} \delta_R(t) = \delta(t). \quad (\text{B.17})$$

How this works is as follows. Define the functional F_g as follows

$$F_g[f] = \int_{-\infty}^{\infty} f(x)g(x)dx. \quad (\text{B.18})$$

Now, for this to be convergent, we obviously require certain conditions to hold on g and f . Normally, such functions are either compactly supported, or rapidly decreasing. In what follows this will be assumed. Ultimately, this defines a linear functional on the space of all functions like f ; it takes a function and returns a number. Now define

$$\langle \delta_R - \delta, f \rangle = \int_{-\infty}^{\infty} (\delta_R(x) - \delta(x))f(x)dx. \quad (\text{B.19})$$

For f either compactly supported or rapidly decreasing, the rigorous sense in which δ_R converges to the Dirac delta function is that we have

$$\lim_{R \rightarrow \infty} \langle \delta_R - \delta, f \rangle = 0. \quad (\text{B.20})$$

Incidentally, derivatives of the Dirac delta function can be defined entirely analogously, using the representation of the Dirac delta function as a distribution. For instance, any distribution will satisfy

$$F_{g'}[f] = -F_g[f'] \quad (\text{B.21})$$

or, applied to the case of the Dirac delta function

$$\langle \delta', f \rangle = - \int_{-\infty}^{\infty} \delta(x)f'(x)dx = -f(0). \quad (\text{B.22})$$

B.3 Properties of the Fourier Transform

The Fourier transform exhibits and reveals a number of analytic properties of functions, and places conditions on how the functions that admit of a well-defined Fourier transform ought to behave. We'll go over some of the properties of the Fourier transforms themselves first, assuming for the time being that they are well-defined, and later make remarks on the conditions that need to be satisfied to guarantee that the manipulations were justified.

Say we have a product of two functions f and g , and would like to compute the Fourier transform of the product. We have

$$\int_{-\infty}^{\infty} f(t)g(t)e^{-i\omega t}dt = \iiint_{-\infty}^{\infty} \hat{f}(\omega_1)\hat{g}(\omega_2)e^{i\omega_1 t + i\omega_2 t - i\omega t}d\omega_1 d\omega_2 dt. \quad (\text{B.23})$$

In the triple integral, the time portion is now trivialized by writing

$$\delta(\omega_2 - (\omega - \omega_1)) = \int_{-\infty}^{\infty} e^{i\omega_1 t + i\omega_2 t - i\omega t}dt. \quad (\text{B.24})$$

We then have

$$\iint_{-\infty}^{\infty} \hat{f}(\omega_1) \hat{g}(\omega_2) \delta(\omega_2 - (\omega - \omega_1)) d\omega_1 d\omega_2 = \int_{-\infty}^{\infty} \hat{f}(\omega_1) \hat{g}(\omega - \omega_1) d\omega_1. \quad (\text{B.25})$$

The quantity

$$(\hat{f} \star \hat{g})(\omega) = \int_{-\infty}^{\infty} \hat{f}(\omega_1) \hat{g}(\omega - \omega_1) d\omega_1 \quad (\text{B.26})$$

is known as the convolution of $\hat{f}$ and $\hat{g}$. The product is resolved as a convolution in Fourier space, and a product in Fourier space is a convolution in ordinary space. A trivial instance of the latter is that the product of $\hat{f}(\omega)$ and 1 in Fourier space is equal to the convolution of f and δ in ordinary space (δ is the Fourier transform of 1).

Now consider a derivative of a function, and the computation of its Fourier transform. Notice that

$$\int_{-\infty}^{\infty} f'(t) e^{-i\omega t} dt = \int_{-\infty}^{\infty} f(t) \times (-i\omega) e^{-i\omega t} dt \quad (\text{B.27})$$

where upon integration by parts, the boundary term can be dropped for compactly supported or rapidly decaying f . Thus we have

$$\widehat{\partial_t f} = -i\omega \hat{f} \quad (\text{B.28})$$

or, that derivatives are trivialized as multiplication by the Fourier variable. This has the upshot of allowing us to define fractional derivatives as the inverse Fourier transform of multiplying by $(-i\omega)^a$ for any value of a .

It is for this reason that linear differential equations were so easily trivialized and inverted in the chapter on harmonic oscillators using Fourier transforms. Being naturally oscillatory, they admit of Fourier transformations, and the solutions to the linear oscillators are immediate once we observe that the Euler-Lagrange equations can be written as simply a product of a polynomial in $i\omega$ —corresponding to the differential operator—and the position $x(\omega)$.

B.4 Analytic Aspects

We've alluded to conditions a function needs to have to make the Fourier transform meaningful. Since a detailed discussion of this will lead us not only down a path far outside the scope of the present text, but also one involving a fairly heavy bit of abstract functional analysis, we will simply provide some of the heuristics, asking the reader to keep these properties in mind. As it turns out, some of these convergence and existence criteria become more important when handling waves and scattering in advanced aspects of field theory, so it is worth noting them early on.

Formally speaking, the Fourier transform is a mapping between two classes of spaces of functions. In particular, it is an automorphism between two function spaces, one defined on the original variable, and the other on the Fourier variable. A function space is simply a set of functions on a domain that behaves like a vector space. The sum of two functions must belong to it, there is a zero function, and if f is a function in the space, so is $-f$. In the following, we will assume the function is defined on $\mathbb{R}$, but the generalization to general $\mathbb{R}^n$ is just a linguistic one.

The first class of relevant functions is the space of functions of compact support. These are defined by

$$\mathcal{C}(\mathbb{R}) = \{f : \mathbb{R} \longrightarrow \mathbb{R} \mid \text{supp}(f) = \Omega\} \quad (\text{B.29})$$

where Ω is a compact subset of $\mathbb{R}$. Another class of relevant functions is the space of Schwartz functions, which are defined by

$$\mathcal{S}(\mathbb{R}) = \{f : \mathbb{R} \longrightarrow \mathbb{R} \mid \sup |x^\alpha \partial_x^\beta f(x)| < \infty\} \quad (\text{B.30})$$

which is a fancy way of saying that the function falls off more rapidly at infinity than any reciprocal power of x , as do any of its derivatives.

The Fourier transform takes a function defined on one of these spaces, and returns another one possessing the same properties. However, the Fourier transform “warps” the space in a nontrivial way. It’s for this reason that definitions of things like the Dirac delta function only really make sense as limits of functions which fall into one of these categories. For instance, δ_R is a compactly supported function, vanishing outside of $[-R, R]$.

In terms of physics, what this really means is that any function that is obtained by solving the Euler-Lagrange equations via performing Fourier transforms must ultimately enjoy these properties. For instance, in the case of the forced oscillator, the solution is only defined rigorously when the forcing function is either a Schwartz function or a rapidly decreasing one. To a degree this makes physical sense; a function that does not taper off far away in time corresponds to a force with infinite energy. On the other hand, if we want to use a forcing function that doesn’t obey this, the rigorous approach would be to “regulate” it, using some kind of rapidly decaying prefactor like e^{-at^2} , and then finding the limit of $a \longrightarrow 0$ at the end. Once again, most of these issues can be ignored in the classical context, but become important when studying more advanced topics in field theory.

C Elliptic Integrals and Elliptic Functions

In the same way as trigonometric functions arise out of computing arc lengths on a sphere, elliptic integrals have their origins in the calculation of elliptic arc lengths. However, it turns out that this one generalization results in a considerably richer analytic structure, and makes the study of elliptic integrals and functions quite a bit harder than their circular counterparts¹⁹

Despite their relative obscurity today²⁰, the subject is so mathematically intricate that in the author’s opinion, one would be remiss to neglect them. As with other appendices in the present work, a detailed discussion of elliptic integrals and functions would warrant one—or many—dedicated monographs. Accordingly, the present appendix will highlight some of their essential features, and give a flavor of some of the depth involved in their analysis.

C.1 Elliptic Integrals

Elliptic integrals are a class of explicit functions that arise out of the study of two categories of problems. The first, as already remarked, is the computation of arc lengths on ellipses. The second is in the general study of problems solvable by quadrature. A considerable number of problems in classical mechanics can be reduced to a finite set of these elliptic integrals, and it’s accordingly easiest to simply define the integrals as such.

The three canonical elliptic integrals are as follows. The elliptic integral of the first kind is defined by

$$K(x; k) = \int_0^x \frac{dt}{\sqrt{(1-t^2)(1-k^2t^2)}}. \quad (\text{C.1})$$

The constant k is known as the modulus, and is ultimately responsible for the distinction between elliptic and circular integrals. We’ll see later that it corresponds to the eccentricity, or the extent of deviation from circularity. The elliptic integral of the second kind is defined by

$$E(x; k) = \int_0^x \sqrt{\frac{1-k^2t^2}{1-t^2}} dt. \quad (\text{C.2})$$

The elliptic integral of the second kind is most closely related to the computation of arc lengths. Finally, the elliptic integral of the third kind is defined by

¹⁹Jacob Bourjaily once remarked to me that the transition from trigonometric functions to elliptic functions feels similar to the transition from driving automatic to stick. Challenging at first, and largely natural once it becomes a habit.

²⁰There was a time—back when Britain was an empire—during which the study of elliptic functions was a standard fixture of a course in analysis and classical physics. As these tragically receded into history, so did the empire as well.

$$\Pi(n; x|k) = \int_0^x \frac{dt}{(1 - nt^2)\sqrt{(1 - t^2)(1 - k^2t^2)}}. \quad (\text{C.3})$$

Any integral of a rational function $f(t, \sqrt{Q(t)})$, where Q is either a cubic or quartic in t can be reduced to a linear combination of these three integrals. The most direct proof of this involves the computation of the integral cohomology, and isn't altogether illuminating. We'll take this as fact for the purposes of brevity here, but references will be provided in the Bibliographical Essay.

A substitution that's often used is to set $\sin \phi = t$, which results in

$$dt = \cos \phi d\phi = \sqrt{1 - t^2} d\phi. \quad (\text{C.4})$$

In this case, we have for the three elliptic integrals

$$\begin{aligned} K(\phi; k) &= \int_0^\phi \frac{d\phi}{\sqrt{1 - k^2 \sin^2 \phi}} \\ E(\phi; k) &= \int_0^\phi \sqrt{1 - k^2 \sin^2 \phi} d\phi \\ \Pi(n; \phi|k) &= \int_0^\phi \frac{d\phi}{(1 - n \sin^2 \phi) \sqrt{1 - k^2 \sin^2 \phi}} \end{aligned} \quad (\text{C.5})$$

When defined in terms of integrals over t , these are known as the Jacobi form, and when defined in terms of the angular integrals over ϕ , they are known as the Legendre forms.

In all of these cases, where there is a dependence of these integrals on x (or equivalently ϕ), the corresponding elliptic integrals are said to be *incomplete*. The complete elliptic functions are those for which we have $\phi = \frac{\pi}{2}$. This choice of angle is largely conventional; since it's the standard definition we will use it in our treatment. The complete elliptic integrals are simply denoted by $K(k)$, $E(k)$ and $\Pi(n|k)$, dropping the t or ϕ . Two occasionally useful series expansions of the first two complete elliptic integrals are

$$K(k) = \frac{\pi}{2} \sum_{n=0}^{\infty} \left(\frac{1}{2^{2n}} \binom{2n}{n} \right)^2 k^{2n} \quad (\text{C.6})$$

and

$$E(k) = \frac{\pi}{2} \sum_{n=0}^{\infty} \left(\frac{1}{2^{2n}} \binom{2n}{n} \right)^2 \frac{k^{2n}}{1 - 2n}. \quad (\text{C.7})$$

Unfortunately, a substantially similar expression for the complete elliptic integral of the third kind is harder to write down.

C.2 Applications of Elliptic Integrals

Generally speaking, the practical use of elliptic integrals involves reducing a relevant problem down to one of the three elliptic integrals, which can then be evaluated relatively easily numerically (they happen to be quite stable even when implemented somewhat naïvely). In the event that we need to use the complete avatars of the elliptic integrals, the series expansions can be employed as well.

For instance, say we have an ellipse given by

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \quad (\text{C.8})$$

whose arc length we wish to compute. By parametrizing the curve using

$$\begin{aligned} x &= a \cos t \\ y &= b \sin t \end{aligned} \quad (\text{C.9})$$

we may write

$$ds = \sqrt{x'(t)^2 + y'(t)^2} dt = b \sqrt{1 - e^2 \sin^2 t} dt \quad (\text{C.10})$$

where $e^2 = \frac{b^2 - a^2}{b^2}$ is the eccentricity of the ellipse. The arc length from $t = 0$ to $t = \theta$ will be

$$\text{Length}(\theta) = b \int_0^\theta \sqrt{1 - e^2 \sin^2 t} dt. \quad (\text{C.11})$$

This is simply an incomplete elliptic integral of the second kind. The total perimeter of the ellipse is obtained when $\theta = 2\pi$, which is just $4bE(e)$.

Note how this reduces to the appropriate formula for the perimeter of the circle when $e \rightarrow 0$. We find using the series expansion

$$\text{Length} = \frac{4b\pi}{2} (1 - e^2 + \mathcal{O}(e^4)) \rightarrow 2b\pi \quad (\text{C.12})$$

which is the circumference of a circle of radius b . Due to the fact that the leading correction is of order e^2 , we see that in physical cases it will be exceedingly small. For instance, in the orbit of Mercury, this contribution will change the circumference by roughly 4%. In the case of Earth's orbit on the other hand, the correction will be around 0.02%.

The elliptic integral of the first kind shows up on a very elementary mechanics problem, the motion of the simple pendulum. Assuming the length of the string does not change, the Lagrangian is a simple combination of a kinetic and potential term according to

$$L = \frac{1}{2} m \ell^2 \dot{\theta}^2 + m g \ell \cos(\theta) \quad (\text{C.13})$$

where θ is the angle from the vertical. Since it's manifestly independent of time, we can directly solve by quadrature using the conservation of energy; in particular,

$$E = \frac{1}{2}m\ell^2\dot{\theta}^2 - mg\ell \cos(\theta). \quad (\text{C.14})$$

If the starting angle of the pendulum is θ_0 , we may write $E = -mg\ell \cos \theta_0$. Rearranging now we find

$$4\frac{g}{\ell} \left(\sin^2 \frac{\theta_0}{2} - \sin^2 \frac{\theta}{2} \right) = \dot{\theta}^2 \quad (\text{C.15})$$

whence

$$dt = \frac{1}{2} \sqrt{\frac{\ell}{g}} \frac{d\theta}{\sqrt{\sin^2 \frac{\theta_0}{2} - \sin^2 \frac{\theta}{2}}}. \quad (\text{C.16})$$

A full period corresponds to 4 times the time taken between 0 and θ_0 ; evaluating this integral after making the replacement

$$\sin \phi = \frac{\sin \frac{\theta}{2}}{\sin \frac{\theta_0}{2}} \quad (\text{C.17})$$

we get

$$T = 4 \sqrt{\frac{\ell}{g}} \int_1^{\frac{\pi}{2}} \frac{d\phi}{\sqrt{1 - k^2 \sin^2 \phi}} \quad (\text{C.18})$$

where $k = \sin \frac{\theta_0}{2}$. In other words,

$$T = 4 \sqrt{\frac{\ell}{g}} K \left(\sin \frac{\theta_0}{2} \right) \quad (\text{C.19})$$

which recovers the period $2\pi\sqrt{\frac{\ell}{g}}$ in the $\theta_0 \rightarrow 0$ limit corresponding to small oscillations.

C.3 Elliptic Functions

Consider the integral expression

$$u = \int_0^x \frac{dt}{\sqrt{1 - t^2}}. \quad (\text{C.20})$$

As this happens to be an elementary integral, with antiderivative $\sin^{-1}$, we know that

$$u = \sin^{-1} x \quad (\text{C.21})$$

or

$$x = \sin(u). \quad (\text{C.22})$$

Elliptic functions are the analogue of sine and cosine, but for elliptic integrals instead of the integral in (C.20); they make it possible to recover the limit of integration x , given an incomplete elliptic integral.

Recall the incomplete elliptic integral of the first kind

$$K(x; k) = \int_0^x \frac{d\theta}{\sqrt{1 - k^2 \sin^2 \theta}}. \quad (\text{C.23})$$

Setting $u = K(x, k)$, we define the Jacobi amplitude function by

$$\text{am}(u; k) = x. \quad (\text{C.24})$$

It is, in other words, essentially the inverse of the elliptic integral of the first kind, when expressed in Legendre form. We trivially have

$$\begin{aligned} (K \circ \text{am})(u) &= u \\ (\text{am} \circ K)(x) &= x. \end{aligned} \quad (\text{C.25})$$

Differentiating (C.23) under the integral sign, we find the identity

$$\frac{du}{dx} = \frac{1}{\sqrt{1 - \sin^2 x}} \quad (\text{C.26})$$

where since we have $x = \text{am}(u; k)$, we now have

$$\frac{d}{du} \text{am}(u; k) = \sqrt{1 - k^2 \sin^2(\text{am}(u; k))}. \quad (\text{C.27})$$

This motivates the following definitions

$$\begin{aligned} \text{sn}(u; k) &= \sin(\text{am}(u; k)) \\ \text{cn}(u; k) &= \cos(\text{am}(u; k)) \\ \text{dn}(u; k) &= \sqrt{1 - k^2 \sin^2(\text{am}(u; k))} \end{aligned} \quad (\text{C.28})$$

These three functions are the canonical *Jacobi elliptic functions*. They satisfy identities and can be manipulated in much the same way that trigonometric functions can be, just with more effort and care. It's not especially useful to treat this with excessive detail, but we'll go over the properties most salient to the calculations in the main text.

The most important feature of these functions is their periodicity. Note that since sn and cn are trigonometric functions of the Jacobi amplitude, they are periodic under the replacements

$$\text{am}(u; k) \longrightarrow \text{am}(u; k) + 2\pi. \quad (\text{C.29})$$

At the same time, recall that for $\text{am}(u; k) = \frac{\pi}{2}$, we have $u = K(k)$. In other words, as u is incremented by $4K(k)$, the Jacobi amplitude is shifted by 2π . In other words we have

$$\begin{aligned} \text{sn}(u + 4K(k); k) &= \text{sn}(u; k) \\ \text{cn}(u + 4K(k); k) &= \text{cn}(u; k). \end{aligned} \quad (\text{C.30})$$

Since $\text{dn}(u; k)$ is an even function of $\text{sn}(u; k)$, it has a period of $2K(k)$, similar to the tangent function in ordinary trigonometry.

Expressed as functions of the Jacobi amplitude, each of the elliptic integrals can be seen to be quasiperiodic when the argument u is shifted to $u + 2K(k)$. To establish this, note that

$$K(\phi + \pi; k) = 2K(k) + \int_{\pi}^{\phi + \pi} \frac{d\theta}{\sqrt{1 - k^2 \sin^2 \theta}} \quad (\text{C.31})$$

Performing the replacement $\theta = \theta' + \pi$, the integral expression becomes $K(\phi; k)$. In other words

$$K(\phi + n\pi; k) = 2nK(k) + K(\phi; k) \quad (\text{C.32})$$

with equivalent identities holding when K is replaced by E and Π . In terms of the Jacobi amplitude am , this amounts to the shift

$$u \longrightarrow u + 2K(k). \quad (\text{C.33})$$

This was the key identity used in the resolution of the Euler angle for the asymmetric top, into two separately periodic pieces.

Jacobi elliptic functions satisfy derivative and addition formulas similar to—but more complex than—their trigonometric counterparts. For instance,

$$\frac{d}{du} \text{sn}(u; k) = \text{cn}(u; k) \frac{d}{du} \text{am}(u; k) \quad (\text{C.34})$$

which equals

$$\frac{d}{du} \text{sn}(u; k) = \text{cn}(u; k) \text{dn}(u; k). \quad (\text{C.35})$$

Similarly, we have

$$\frac{d}{du} \text{cn}(u; k) = -\text{sn}(u; k) \text{dn}(u; k) \quad (\text{C.36})$$

and

$$\frac{d}{du} \operatorname{dn}(u; k) = -k^2 \operatorname{sn}(u; k) \operatorname{cn}(u; k) \quad (\text{C.37})$$

Parenthetically, the derivative formula can be used to show that sn is the inverse function of the elliptic integral of the second kind, but this time expressed in Jacobi form. In a sense, the map between the functions am and sn furnish the translation between the Legendre forms and Jacobi forms of elliptic integrals. We also have, simply as a matter of definition

$$\begin{aligned} \operatorname{sn}^2(u; k) + \operatorname{cn}^2(u; k) &= 1 \\ \operatorname{dn}^2(u; k) + k^2 \operatorname{sn}^2(u; k) &= 1. \end{aligned} \quad (\text{C.38})$$

Occasionally useful are the addition formulas, which are complicated, both in terms of their forms and derivations. We simply state them below

$$\begin{aligned} \operatorname{sn}(u + v; k) &= \frac{\operatorname{sn}(u; k) \operatorname{cn}(v; k) \operatorname{dn}(v; k) + \operatorname{sn}(v; k) \operatorname{cn}(u; k) \operatorname{dn}(u; k)}{1 - k^2 \operatorname{sn}^2(u; k) \operatorname{sn}^2(v; k)} \\ \operatorname{cn}(u + v; k) &= \frac{\operatorname{cn}(u; k) \operatorname{cn}(v; k) - \operatorname{sn}(u; k) \operatorname{sn}(v; k) \operatorname{dn}(u; k) \operatorname{dn}(v; k)}{1 - k^2 \operatorname{sn}^2(u; k) \operatorname{sn}^2(v; k)} \\ \operatorname{dn}(u + v; k) &= \frac{\operatorname{dn}(u; k) \operatorname{dn}(v; k) - k^2 \operatorname{sn}(u; k) \operatorname{sn}(v; k) \operatorname{cn}(u; k) \operatorname{cn}(v; k)}{1 - k^2 \operatorname{sn}^2(u; k) \operatorname{sn}^2(v; k)} \end{aligned} \quad (\text{C.39})$$

C.4 Theta Functions

Consider the solutions of the following algebraic equation in $\mathbb{P}^1$

$$y^2 = Q(x) \quad (\text{C.40})$$

where Q is either a cubic or a quartic. Curves of this kind are known as elliptic curves, integrals of which we have just studied. At the same time, a natural question to pose is: what would a coordinate system on such a curve look like?

For a circle, we simply have the usual trigonometric coordinate system, but this turns out to be a rather complicated question for elliptic curves. A class of functions known as theta functions provide the correct description for this coordinate system in the case of elliptic curves. As it happens, these functions also enjoy a number of convenient relationships with elliptic functions.

The full theory of theta functions is extensive, and a full treatment of them cannot be achieved here. Accordingly, we'll record a few results worth remembering for the typical physicist, with a more detailed guide to be found in the Bibliographical Essay. We'll start with definitions; the four theta functions are defined by the following series expansions

$$\begin{aligned}
\theta_1(z|\tau) &= \sum_{n=-\infty}^{\infty} (-1)^{n-\frac{1}{2}} q^{\left(n+\frac{1}{2}\right)^2} e^{i(2n+1)z} \\
\theta_2(z|\tau) &= \sum_{n=-\infty}^{\infty} q^{\left(n+\frac{1}{2}\right)^2} e^{i(2n+1)z} \\
\theta_3(z|\tau) &= \sum_{n=-\infty}^{\infty} q^{n^2} e^{i2nz} \\
\theta_4(z|\tau) &= \sum_{n=-\infty}^{\infty} (-1)^n q^{n^2} e^{i2nz}
\end{aligned} \tag{C.41}$$

where the nome q is defined as

$$q = e^{i\pi\tau}. \tag{C.42}$$

We won't prove the following, but it turns out that if we set

$$z(u) = \frac{\pi u}{2K(k)} \tag{C.43}$$

and

$$\tau(k) = i \frac{K(k)}{K(\sqrt{1-k^2})} \tag{C.44}$$

the following identities hold

$$\begin{aligned}
\operatorname{sn}(u; k) &= \frac{\theta_3(0|\tau(k))\theta_1(z(u)|\tau(k))}{\theta_2(0|\tau(k))\theta_4(z(u)|\tau(k))} \\
\operatorname{cn}(u; k) &= \frac{\theta_4(0|\tau(k))\theta_2(z(u)|\tau(k))}{\theta_2(0|\tau(k))\theta_4(z(u)|\tau(k))} \\
\operatorname{dn}(u; k) &= \frac{\theta_4(0|\tau(k))\theta_3(z(u)|\tau(k))}{\theta_3(0|\tau(k))\theta_4(z(u)|\tau(k))}.
\end{aligned} \tag{C.45}$$

Conventionally, the u, k arguments, and the 0's are suppressed, to make things more readable. For instance, $\theta_3(0|\tau(k)) = \theta_3(\tau)$.

Another theta function, which shows up somewhat often when dealing with problems of heat transfer, is the Jacobi theta function, defined by

$$\vartheta(z|\tau) = 1 + 2 \sum_{n=1}^{\infty} q^{n^2} \cos(2n\pi z). \tag{C.46}$$

The complex analysis associated with theta functions is rather rich, but can also be mathematically more involved. They happen to be the part of the theory of elliptic functions that survived their decline; even now they are used extensively in the study of Abelian forms, and the study of Green's functions on Riemann surfaces. Both of

these subjects—needless to say—are considerably outside the scope of a study of classical mechanics.

Bibliographical Notes

We start with general references. The most canonical is the beautiful monograph due to Landau and Lifshitz:

1. L. D. Landau and E. M. Lifshitz, *Mechanics*, vol. 1 of *Course of Theoretical Physics*. Pergamon Press, Oxford, 3 ed., 1976

The book is written in classic Russian style. The prose is terse, and very little—if anything—is said that doesn’t need to be said. In particular, Chapter V, dealing with oscillations, is especially thorough.

Standard graduate treatments which more than suffice for a “working” knowledge of classical mechanics are the two textbooks

2. L. N. Hand and J. D. Finch, *Analytical Mechanics*. Cambridge University Press, Cambridge, 1998
3. H. Goldstein, C. P. Poole, and J. L. Safko, *Classical Mechanics*. Addison-Wesley, San Francisco, 3 ed., 2002

Hand and Finch is simpler than Goldstein (et al), and has the benefit of being better written as well. It has an elegant derivation of the precession of the equinox due to the oblateness of the Earth, the largest contributing factor to the precession of Mercury’s orbit.

Goldstein is a favorite for graduate courses on mechanics, a cultural phenomenon that is beyond my ken. It suffers from poor notation, but makes up for it by being verbose and relatively thorough. There is no discussion of advanced aspects of the top, both the symmetric and asymmetric versions.

The most extensive treatment of classical mechanics is due to ET Whittaker

4. E. T. Whittaker, *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies: With an Introduction to the Problem of Three Bodies*. Cambridge University Press, Cambridge, 4 ed., 1937

Legendary is perhaps a good word to describe this book, as is notorious. In approximately 600 dense pages, virtually any aspect of mechanics can be found treated in relative depth in this work. The asymmetric top is solved in gory detail, using theta functions. An extensive discussion of the three body problem can be found here as well.

For those interested in the development of mechanics from a historical lens, there is a retelling of the *Principia* due to Chandrasekhar

5. S. Chandrasekhar, *Newton’s Principia for the Common Reader*. Clarendon Press, Oxford, 1995;

with two standard translations being due to Motte and Cohen and Whitman (the latter of which was used by Chandra)

6. A. Motte, *Sir Isaac Newton's Mathematical Principles of Natural Philosophy and His System of the World*. University of California Press, Berkeley, 1934
7. I. B. Cohen and A. Whitman, *The Principia: Mathematical Principles of Natural Philosophy*. University of California Press, Berkeley, 1999

An interesting side note: Chandra treats the infamous Proposition CXVI (and its 22 corollaries) of the *Principia*, in which Newton studies the three body problem. From what I can tell, it's the only time Newton's geometric arguments in this particular proposition were translated into modern calculus. On that note, Chandrasekhar suggests—rather convincingly—that the *Principia* may have been the last of the great “geometric” works of mechanics, paving the way to giving up this old manner of physics. The extremely baroque arguments Newton had to use to do “calculus” likely convinced people that there had to be a better way of doing things.

For some modern books, three excellent textbooks are due to Tong and Fitzpatrick respectively.

8. D. Tong, *Classical Mechanics*, vol. 1 of *Lectures on Theoretical Physics*. Cambridge University Press, Cambridge, 2025
9. R. Fitzpatrick, *Newtonian Dynamics: An Introduction*. CRC Press, Boca Raton, FL, 1 ed., 2022
10. R. Fitzpatrick, *An Introduction to Celestial Mechanics*. Cambridge University Press, Cambridge, 2012

Tong's textbook is based on two lecture notes that are freely available, which I cannot recommend highly enough

11. D. Tong, “Classical dynamics,” Lecture notes, University of Cambridge, 2005
12. D. Tong, “Dynamics and relativity,” Lecture notes, University of Cambridge, 2013

Fitzpatrick's *An Introduction to Celestial Mechanics* can also be accessed online as lecture notes

13. R. Fitzpatrick, “Classical mechanics: An introductory course,” Online course text, University of Texas at Austin, 2011

Perhaps the clearest treatment I am aware of discussing orbital perturbation theory can be found in these notes. The computation of the precession of Mercury's perihelion due to the surrounding planets is especially clarifying.

For those interested in a mathematical treatment of classical mechanics, and symplectic theory in particular, will find Arnold's book quite valuable.

14. V. Arnold, *Mathematical Methods of Classical Mechanics*, vol. 60 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2 ed., 1989

Finally, the importance of studying function theory and special functions cannot be emphasized enough. Perhaps the worst advice I ever received was to not study special functions—which are special for a reason—with any seriousness. A classic, but somewhat dated reference is

15. E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis*. Cambridge University Press, Cambridge, 4 ed., 1927

A much more accessible reference is

16. P. Denner and A. Krzywicki, *Mathematics for Physicists*. Dover Publications, Mineola, NY, dover edition ed., 1996

There do not seem to be very many elementary books on elliptic integrals and elliptic functions. A book that compensates for this with extreme thoroughness is

17. H. Hancock, *Lectures on the Theory of Elliptic Functions*. Dover Phoenix Editions. Dover Publications, Mineola, NY, 2004

Finally, we note the following historical references

18. Ptolemy, *Ptolemy's Almagest*. Princeton Paperbacks. Princeton University Press, Princeton, NJ, english translation by g. j. toomer ed., 1998
19. R. Fitzpatrick, *A Modern Almagest: An Updated Version of Ptolemy's Model of the Solar System*. Lulu, 2011
20. C. Rovelli, “Aristotle’s physics: A physicist’s look,” *Journal of the American Philosophical Association* **1** no. 1, (2015) 23–40

The first of these is the canonical translation of the monumental *Almagest* due to Ptolemy. Despite how complicated it may appear at first, readers will likely find the geometric arguments simple when they go over them more carefully. The second of these is a modern translation of the *Almagest*, which ends up being pared down in size a fair bit when converted into modern notation. Finally, the last of these (which can also be found as an arXiv preprint at 1312.4057) is a fun article due to Rovelli, which ought to convince the skeptic that Aristotelian physics wasn’t all that crazy after all. It was just an approximation of its more accurate Newtonian cousin, in the same way as Newtonian mechanics approximates quantum mechanics or relativistic theory.

