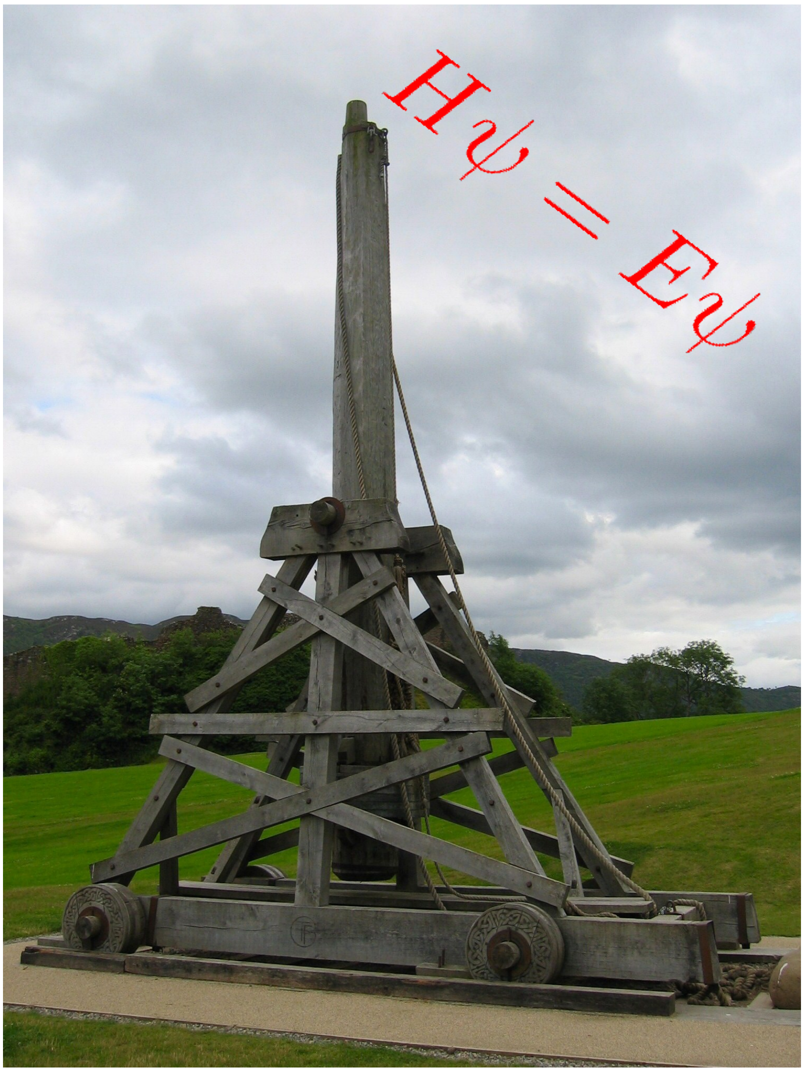
Classical Mechanics

with Modern and Medieval Applications



Jay D. Tasson
Department of Physics and Astronomy, Carleton College

About these notes

These notes have been assembled by Jay Tasson based on the Phys 231 class at Carleton College. Though assembled by Jay, much of the material is based on organization, problems, and ideas from my colleagues, most notably Bill Titus from whom I inherited much, including the trebuchet project. Enjoy reading and using these notes, but avoid sharing them beyond Carleton (e.g. posting online) without talking to Jay.

Contents

Chapter 1

Preliminaries

1.1 About the title

Why *Classical Mechanics with Modern and Medieval Applications*? First, it's a bit of joke based on the potentially confusing use of the term "Classical" on our liberal arts campus. The term may refer to the classical period, being the era of ancient Greek and Roman civilizations extending from the 7^{th} or 8^{th} century BC through the 5^{th} century AD. It may also refer to the classical period in Western music, 1750-1820. In physics, classical is typically used to refer to anything not involving quantum mechanics, which was developed in the 1920s. Thus "classical" in the present context refers to physics not involving quantum mechanics. We'll also restrict attention primarily to nonrelativistic physics, which means systems for which the effects of special and general relativity are negligible.

If we're sticking to classical physics, why then does the Schrödinger equation (which you may or may not have met yet) appear on the cover? That brings me to the second part of the title, "modern applications". Classical mechanics may be old, but it is certainly not a dry or irrelevant subject. It has vast applications ranging from engineering to the most modern of physics. As a case in point, the H appearing in the Schrödinger equation is the Hamiltonian, a concept well-understood in classical mechanics long before quantum physics was imagined. As the course proceeds we'll make contact with these modern connections, though you need not worry if you don't know any modern physics. I'll explain any relevant modern physics. Note that you might be used to seeing the Schrödinger equation with the replacement

$$H = -\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + U. \tag{1.1}$$

Finally, why "Medieval Applications"? Primarily, because trebuchets are a really cool piece of Medieval weaponry of historical significance that are based on classical mechanics, and they make a fun system for us to model using the tools we'll develop in the course.

1.2 Preclass Questions (PCQ)

Throughout these notes I've embedded Preclass Questions (PCQ). They are intended to encourage you to think about some things as you read and to get you ready to learn what's coming next. They usually fall into one of two classes: those that ask you to develop your own answer to a question that doesn't really have a unique answer and those that ask you to answer a conceptual question or fill in some steps in an example or derivation, in which case there is a correct answer. For questions of the first type, my own personal answer to the question often appears later in the notes. *Please do not search ahead for my answer.* Doing so will only poison the well of you own thoughts. That would be unfortunate, since you might generate something better than what I have. I'm really interested in getting you to think and hearing what you have to say.

Your answers will be turned in at the beginning of class, and you will get participation credit for turning in thoughtful, complete responses. You should take action such that you have access to your PCQ answers during class time.

Often a set of questions about the next chapter will appear at the end of the previous. Since in the case of Chapter 1 there is no previous chapter, I'll put these questions here. Write answers to these questions before reading on.

PCQ 1.1

Produce a definition of "physics" in your own words using a few sentences.

PCQ 1.2

In a sentence or two, compare and contrast physics and biology.

1.3 Forest Goals

In the course of studying an area of physics, many people find that they spend much time focused deeply on what one might call "the trees" of the subject – trying to solve an equation, understand a definition, etc. Clearly these things are an important part of your education, but it's equally if not more important to consider understanding on a much bigger scale, that of the forest. Efforts of the "forest" scale might include why are we doing this, how does it relate to what we learned in this other cool class, how do all of these laws or equations fit together, why does the subject take its present shape, etc. Throughout the course, we'll try to keep these "forest" type questions in mind as well. For a visual, see Fig. 1.1.

Figure 1.1: Forest and trees.

While we'll perhaps address many forest goals in the course, I think I can identify several key ones right at the start. They include: gaining a picture of classical mechanics as a useful field with vast application, brain training, understanding the development of physical theory, experiencing theoretical physics, experiencing computational methods. The second goal is most easily summed up by Einstein, who said "the true purpose of education is to train the mind to think." I find that the more hard questions I think about, the better I get at tackling new ones, even if they are completely unrelated. As we struggle with hard questions in this course, your brain will get stronger, and you will get better at tackling questions in whatever field you choose to study. The final two goals we should pause and explore further.

1.3.1 Physical Theory

We are about to devote a term to studying something we call "physics". Some of you may devote a large part of your life to doing "physics". If this is worth doing, we should be able to define it. Moreover, in this course, we'll want to understand what it means to build a physical theory. We'll study things like Newton's Laws as examples of physical theories, and we'll even find that Newton had some choice in how to develop his laws. In order to have these types of discussions, we need to try to decide what physics is, how to build a physical theory, and what approach physicists typically take to problems. If you didn't answer the PCQs at the beginning of the chapter, make sure you do so now before reading further. Once you've got your ideas in hand, you can read on to hear what I (and in a few case some other people) think. Before we get started, I should make a few disclaimers. First, you should feel free to disagree with anything I say here. There is no right answer and my positions may even change over time. Second, I make some comparisons between physics and other sciences at some points in the following comments. My comments are in no way meant to imply that I think physics is better than what any of these other folks do. I have great respect for what people in other sciences do. I'm glad there are people studying the expression of genes and function of the human brain.

Physics

Dictionary.com defines physics as follows: "the science that deals with matter, energy, motion, and force." To me, that feels both too narrow and too broad. It feels too narrow in that physicists also study waves, angular momentum, spacetime curvature, etc., which don't seem to be included here. It also feels too broad. Don't chemists and biologists also study matter and energy?

I like, "the interplay of mathematical modeling and experiment usually applied to fundamental natural problems." It seems to me that the heart of what physicists do is trying to find some math that predicts an observation and then trying to find an observation that checks a further prediction of the math. "Fundamental natural problems" is a little harder to get a feel for. To understand what I mean, we should first step back and look at history.

There was a point where the Dictionary.com definition might have seemed more suitable. This was back in a time when physics, chemistry, geology, biology, psychology, etc. were believed to be completely separate, and physics was the branch that studied motion, matter, etc. The conclusions found about these things were usually thought to have an origin quite separate from the results found in other sciences. We now see that when you ask why a variety of complex systems behave as they do, the answers seem to point toward a common origin. That is, all of the sciences seem to be studying different consequences of the same natural principles. For example, we now see that it's really the principles of quantum mechanics that are responsible for chemistry, the chemistry within living organisms that is responsible for biology, the biology of the brain that is behind the discoveries of psychology, etc.

Physics initially asked what one might call the simplest of questions. Why do objects move? What is temperature? What is matter? Hence it seems that the more complex questions asked by chemistry, for example, have been linked to the "fundamental" answers achieved by physics. This does not mean that chemistry or psychology are not worth while. Although we see that quantum mechanics explains chemistry, it's not practical to actually calculate all of the conclusions of chemistry explicitly from quantum mechanics. In some cases it's just not possible in practice. Getting the behavior of the human brain using quantum mechanics is something that seems very likely possible in principle, but we really really have no clue how to do it. For more comments on the fundamental nature of physics, see for example, Refs. [1] and [2].

So one of the aims of physics is to continue to seek simpler mathematical descriptions that will explain more known results in terms of a smaller set of ideas. This process is usually referred to as a reductionist approach. A particularly pointed example from history is Newton's law of universal gravitation. In generating it, Newton "explained" Kepler's laws of planetary motion as well as why apples fall to Earth. Thus the number of "separate" phenomena is reduced.

The present situation in physics in one in which things have been essentially reduced to the following equation:

$$\begin{aligned} S \;=\; & \int d^4x \Big[\tfrac{1}{2} iee^{\mu}{}_{a} \overline{L}_A \gamma^a \overset{\leftrightarrow}{D}_{\mu} L_A + \tfrac{1}{2} iee^{\mu}{}_{a} \overline{R}_A \gamma^a \overset{\leftrightarrow}{D}_{\mu} R_A \tfrac{1}{2} iee^{\mu}{}_{a} \overline{Q}_A \gamma^a \overset{\leftrightarrow}{D}_{\mu} Q_A + \tfrac{1}{2} iee^{\mu}{}_{a} \overline{U}_A \gamma^a \overset{\leftrightarrow}{D}_{\mu} U_A \\ & + \tfrac{1}{2} iee^{\mu}{}_{a} \overline{D}_A \gamma^a \overset{\leftrightarrow}{D}_{\mu} D_A - \big[(G_L)_{AB} e \overline{L}_A \phi R_B + (G_U)_{AB} e \overline{Q}_A \phi^c U_B + (G_D)_{AB} e \overline{Q}_A \phi D_B \big] + \text{h.c.} \\ & - e (D_{\mu}\phi)^{\dagger} D^{\mu}\phi + \mu^2 e \phi^{\dagger}\phi - \frac{\lambda}{3!} e (\phi^{\dagger}\phi)^2, - \tfrac{1}{2} e \mathrm{Tr}(G_{\mu\nu} G^{\mu\nu}) - \tfrac{1}{2} e \mathrm{Tr}(W_{\mu\nu} W^{\mu\nu}) - \tfrac{1}{4} e B_{\mu\nu} B^{\mu\nu} \\ & + \tfrac{1}{16\pi G} R - 2e\Lambda \Big]. \end{aligned} \tag{1.2}$$

This is the action for the standard model of particle physics along with the action of general relativity. I provide it here for entertainment purposes only, but it's fun to think that this one equation in principle could explain everything from why my pencil falls to why I like to do physics. The equation is an example of an action, which is a concept we'll meet later in the course, although in much simpler terms. This equation is so fundamental that two of my former colleagues who got married to each other saw fit to put it on their wedding cake. See Fig. 1.2.

Figure 1.2: Standard model Lagrangian on a cake.

Another angle that I think the Dictionary.com definition misses is related more specifically to the way physicists approach problems. Physicists generally work from a very small number of simple principles that apply very widely. When presented with a system that's too complex to be solved exactly using this small number of simple principles, physicists often proceed by finding a series of well-understood simplifying assumptions or approximations until they get a system that's similar to the original that they can solve exactly using first principles. This approach is no-doubt familiar as you've likely studied motion under the assumption of negligible aerodynamic drag. Thus in addition to trying to find a ever smaller set of more fundamental principles, physicists also work to apply the known principles to solve new problems of higher complexity. We even here of fields like "bio-physics" which is the application of physics-style problem solving to topics traditionally thought of as biology.

A different approach to problems is sometimes found in other sciences. Rather than solving a simpler system exactly, they often look for rules and patterns in the full system, with all of its complexity. Clearly both methods have complementary strengths and weaknesses. It's also the case that Biologists sometimes approach problems from first principles and physicists sometimes look for patterns in a complex system, but the two fields tend to favor different problem-solving approaches. The difference can be a source of frustration for students taking classes in both areas. Physics students in biology classes often crave "deeper" answers based on foundational principles. The approximations used in physics can feel fake to biology students, and derivations from first principles feel foreign. The spherical cow joke exemplifies this difference. If you're not familiar with the cow joke, a variation goes like this: "A farmer has a problem with his cows. A biologist comes in, examines the cows and begins running tests on them. A chemist comes in and begins running tests on the food and water in the cows' enclosure. A physicist looks at the situation and says, lets begin by approximating the cows as spheres." For a number of problems concerning cows, approximating them as spheres may be a very reasonable model to start with, and for this reason, a bio-physicist friend of mine said that you're a physicist if you don't really think the spherical cow joke is funny.

Theory

The word theory is used by many people and it often has rather different meaning. Non-scientists usually mean "a plausible explanation". The American Association for the Advancement of Science says: "A scientific theory is a well-substantiated explanation of some aspect of the natural world, based on a body of facts that have been repeatedly confirmed through observation and experiment. Such fact-supported theories are not 'guesses' but reliable accounts of the real world. The theory of biological evolution is more than 'just a theory.' It is as factual an explanation of the universe as the atomic theory of matter or the germ theory of disease. Our understanding of gravity is still a work in progress. But the phenomenon of gravity, like evolution, is an accepted fact." This seems to match what I've heard from biologists.

The use of the term theory in physics seems to be slightly different still. Someone on Wikipedia [3] said it better than I ever could. They said "In physics the term theory is generally used for a mathematical framework – derived from a small set of basic postulates – which is capable of producing experimental predictions for a given category of physical systems." This seems to differ from the use in some other sciences in at least two important ways. First, everything is usually based on a small set of postulates, and every aspect of the predictions come, usually in a mathematical way, from these postulates. Second, physics does not seem to impose the requirement that the predictions match experiment as is typically required in other areas of science. That is, the name "theory" can be applied in physics to a logical structure that makes reasonable predictions whether or not they are known to match experiment. This further highlights a difference between physics and the other sciences. Physics often has a bottom-up approach as I describe in the construction of a theory above. Other sciences frequently take a more top-down approach in which they seek patterns in a set of observations, form a hypothesis, then test the hypothesis further.

The properties that a theory of physics must have are usually stated as follows: Logically Consistent – no contradictory assumptions, Quantitative – a complete physical theory should not just say that there is an increase in some quantity, a number should be specified, Predictive – a new theory should make predictions about results not yet observed, and it should be possible to falsify the theory based on the results of these observations. In particle physics, the term theory is often used synonymously with writing down an action, since an action basically does all of the above. That's another reason learning about the action will be a fun thing to do later in the course.

To make all of the trouble with the word theory even worse, there are a whole bunch of other related terms. Here's my take on a few of them. Hypothesis – an educated guess that guides further investigation. It seems that other sciences talk about hypotheses much more than physicists, presumably because physicist develop what they call theories in the more formal way discussed above. Model – this is a little like a theory, but it need not be as general. It also may be more of a calculational or conceptual trick rather than how things "really" work, as in the wave model of light, for example. Law – like a theory, but usually very simple and widely applied. The burden of experimental verification is usually required here by physicists too before using the word law. Principle – often a key assumption built into a theory that is known to be valid at some level. The Principle of Relativity is one example. Conjecture – an educated guess, like a hypothesis, but usually focused on some mathematical point that could be proven from the mathematics that is already in place. For example, Newton may have conjectured that elliptical orbits are a consequence of his theory before working it out.

PCQ 1.3

How well did you read this? I'll expect you to know the properties a physical theory must have. Did you absorb this well enough to explain each point and remember them in 3 weeks?

PCQ 1.4

Reflect on your definition of physics, my definition of physics, what theories, laws, etc. are and the other material in this section. Develop a question or comment about this material. Warning: I'm going to ask you to share this answer in class.

1.3.2 Branches of Physics

Historically physics has been divided into two classifications – theoretical physics and experimental physics. If I was to flash you a picture of a stereotypical theoretical physicist at work, you might see an office with a blackboard full of equations, a large shelf of books, and a table filled with papers and pencils. The theorist sitting at the table would be trying to develop mathematical models that describe nature in the simplest way, at the most fundamental level.

If I was to do the same for an experimental physicist, the focal point of the picture would be a complicated apparatus with many wires, dials, hoses, and knobs. The rest of the picture would be filled with shelves of tools and equipment. The experimentalist would be recording observations in a notebook amid the search for a more clever measurement technique, or measurement to make, that would shine light on a deep natural question.

As a side note, as you were reading these descriptions, there was almost certainly a person in the mental picture that I painted. What did that person look like in your mental picture? The answer might provide an interesting chance to explore your own implicit bias.

The pictures that I provided here of experimentalists and theorists are stereotypes. As such, they contain threads of truth along with many holes and oversimplifications, and they certainly don't apply to every physicist of either type. I provide them only to give a sense of the traditional divide. A natural question at this stage is, "why the divide?" The day-to-day activities of each group (in the stereotypical picture) are quite different, and each type of work requires a large skill set of its own. Thus physicists tend to self-sort based on their interests and the skills that come most naturally to them. That said, there are many folks who sit very close to the line. Although they'll identify themselves as members of one camp or the other, they may cross the line from time to time. It's also the case that nearly all theorists need to have some idea of what it means to do experiments, and experimentalists need to have some understanding of the theories they test.

A tool which was strangely absent from my stereotypical pictures was the computer. Today a computer is prevalent in both pictures. Moreover, an additional classification has appeared and been named "computational physics". In the stereotypical picture, these people have access to a huge computer. They study the development or use of computer algorithms to obtain solutions to problems. These are typically problems for which a fundamental mathematical model exists, but there are interesting problems for which exact "analytical" solutions can't be found. If you took Atomic and Nuclear Physics, you likely have some familiarity with the use of computers for doing data acquisition and analysis, "experimental-type" computer uses as well as some experience with simulation. In Sec. 1.3.4, I'll say a few words about how theoretical and computational physicists use computers.

Another word that you sometimes here in the classification of physics and physicists is phenomenology/phenomenologist. This is a subclass of theoretical physics that is aimed at working out the implications of new theoretical ideas for specific experiments or systems. Jay does a fair amount of this – e.g. working out the speed of a gravitational wave in a theory that's not general relativity.

Finally there are two other groups that deserve mention in this context. Observational astronomers play a role similar to experimental physicists, but they can't change the systems they study. Thus they "observe" rather than "experiment". This is both a practical and a fundamental difference in approach. It requires a set of related but slightly different skills. Engineers, also in an oversimplified view, seek to apply "understood" physical theory to "manipulate nature" for very specific goals. For example, engineers use the laws of mechanics and laws governing material structure to manipulate nature into the from of bridges.

PCQ 1.5

I've provided a very brief window into various "branches" of physics above (which provides a delightful pun in our present discussion of forests and trees). Write one question about this classification scheme and be prepared to discuss in class.

In this course, we'll first think like theorists as we discuss the theories of classical mechanics at the most fundamental level, and we'll seek exact solutions using, for example the technology of integral calculus. Hence the name "analytical mechanics". We'll then work as computational physicists by studying and applying computer algorithms to solve problems that don't have analytical solutions. Finally, we'll work a little like engineers as we apply the theory to our trebuchets. Hopefully all of this will provide a window into how theoretical and computational physicists think.

1.3.3 Theoretical Physics

As we kick off our study of theoretical physics, proceed to read the following short article written by Steven Weinberg, Nobel Prize winning physicist, about the style of thought involved in doing theory. It's a story that has great resonance with my own experience. The version that appears here is as reprinted in the book *Facing Up: Science and its cultural adversaries* [1]. The book is a collection of essays written by Weinberg on a variety of topics relating to how science works and interacts with other areas of human thought and activity. I've placed the book on our course reserve shelf if you'd like to take a look at the other things Weinberg has to say. You should be aware however that many of the essays are controversial in nature, and the views expressed are not necessarily endorsed by Jay.

16

The Red Camaro

The magazine *George* used to have an "Eyewitness" feature, consisting of a brief account of some important moment that had occurred just thirty years earlier. Somehow its editors found out in 1997 that my first paper on the eletroweak theory was written in 1967, so they asked me to contribute a short article describing just how I came to do this work. *George* was a glossy magazine filled with fashion advertisements, the sort of magazine to which I would never have thought of contributing, but they offered a fee per word higher than any I had ever received, so I suppressed my puritan instincts and wrote the piece below. The editorial staff at *George* was very helpful, and we wound up with the fullest account I have given of the circumstances surrounding my work on the electroweak theory.

As it turned out, the article appeared in the magazine on a page opposite an advertisement for Liz Claiborne perfume. The ad was printed in scented ink, so anyone who read the original version of this article carried away a memory of a nice musky fragrance, an effect that Harvard University Press has wisely decided not to try to reproduce.

On October 15, 1764, Edward Gibbon conceived the idea of writing the history of the decline and fall of the Roman Empire while he was listening to barefoot monks singing vespers in the ruins of the Roman Capitol. I wish I could say I worked in settings that glamorous. I got the idea for my best-known work while I was driving my red Camaro in Cambridge, Massachusetts, on the way to my office in the physics department at the Massachusetts Institute of Technology.

I was feeling strung out. I had taken a leave of absence from my regular professorship at Berkeley a year earlier so that my wife could study at Harvard Law School. We had just gone through the trauma of moving from one rented house in Cambridge to another, and I had taken over the responsibility of getting our daughter to nursery school, playgrounds, and all that. More to the point, I was also stuck in my work as a theoretical physicist.

Like other theorists, I work with just pencil and paper, trying to make simple explanations of complicated phenomena. We leave it to the experimental physicists to decide whether our theories actually describe the real world. It was this opportunity to explain something about nature by noodling around with mathematical ideas that drew me into theoretical physics in the first place. For the previous two years, I had made progress in understanding what physicists call the strong interactions—the forces that hold particles together inside atomic nuclei. Some of my calculations had even been confirmed by experiments. But now these ideas seemed to be leading to nonsense. The new theories of the strong interactions I had been playing with that autumn implied that one of the particles of high energy nuclear physics should have no mass at all, but this particle was known to be actually quite heavy. Making predictions that are already known to be wrong is no way to get ahead in the physics game.

Often, when you're faced with a contradiction like this, it does no good to sit at your desk doing calculations—you just go round and round in circles. What does sometimes help is to let the problem cook on your brain's back burner while you sit on a park bench and watch your daughter play in a sandbox.

After this problem had been cooking in my mind for a few weeks, suddenly on my way to MIT (on October 2, 1967, as near as I can remember), I realized there was nothing wrong with the sort of theory on which I had been working. I had the right answer, but I had been working on the wrong problem. The mathematics I had been playing with had nothing to do with the strong interactions, but it gave a beautiful description of a different kind of force, known as the weak interaction. This is the force that is responsible, among other things, for the first step in the chain of nuclear reactions that produces the heat of the sun. There were inconsistencies in all previous theories of this force, but suddenly I saw how they could be solved. And I realized the massless particle in this theory that had given me so much trouble had nothing to do with the heavy particles that feel the strong interaction; it was the photon, the particle of which light is composed, that is responsible for electric and magnetic forces and that indeed has zero mass. I realized that what I had cooked up was an approach not just to understanding the weak interactions but to unifying the theories of the weak and electromagnetic forces into what has since come to be called the electroweak theory. This is just the sort of thing physicists love—to see several things that appear different as various aspects of one underlying phenomenon. Unifying the weak and electromagnetic forces might not have applications in medicine or technology, but if successful, it would be one more step in a centuries-old process of showing that nature is governed by simple, rational laws.

Somehow, I got safely to my office and started to work out the details of the theory. Where before I had been going around in circles, now everything was easy. Two weeks later, I mailed a short article on the electroweak theory to *Physical Review Letters*, a journal widely read by physicists.

The theory was proved to be consistent in 1971. Some new effects predicted by the theory were detected experimentally in 1973. By 1978, it was clear that measurements of these effects agreed precisely with the theory. And in 1979, I received the Nobel Prize in physics, along with Sheldon Glashow and Abdus Salam, who had done independent work on the electroweak theory. I have since learned that the paper I wrote in October 1967 has become the most cited article in the history of elementary particle physics.

I kept my red Camaro until it was totaled by one too many Massachusetts winters, but it never again took me so far.

1.3.4 Computation

Since Weinberg said a little about how traditional theoretical physics is done, I'll now say a few words about computation. Fundamentally, electronic computers are useful to theoretical physics because the can do many repetitive calculations quickly and accurately, not because they do anything magical that a human can't. Many of the ideas now used in computational physics predate electronic computers, when "computers" were people. In some cases "computer algorithms" were implemented by passing papers around a room of people, each of whom did a step of the calculation assembly-line style. Electronic computers now replace what would otherwise be enormous numbers of human computers making computational work much more practical. The whole thing has some similarities to robots replacing humans on auto assembly lines.

Computers can sometimes be used to build a sort of virtual system or "computer model". You'll do this later in the course for the trebuchets, and if you took atomic and nuclear physics, you did it for the decay of silver lab. With such a virtual system, you can see how any change in the set up will effect the outcome, sometimes at the touch of a button. For example, set friction to anything you want, or turn it off, change relative dimensions of the system, etc. You can also make a large number of similar systems and see how they interact.

Computers can also "solve" integrals that have no exact solution. Remember an integral is a sum, and it can be viewed as an area under a curve. Rather than using the methods of calculus, one could divide the area under the curve into many tiny rectangles, find their area, and add them up as illustrated in Fig. 1.3. Since

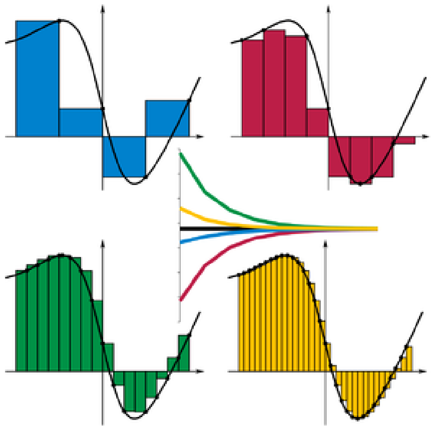


Figure 1.3: Riemann sum.

electronic computers are fast and accurate, they can divide the area into very many very small rectangles and get a very good approximation to a hard integral. Similar strategies apply to other mathematical problems. These strategies are called "numerical solutions". Computers can be used to generate numerical solutions in isolation or as a part of a computational model.

When solving a problem on a computer, most of the real thinking is done by humans. The humans must make a plan for how the computer will solve the problem, break the problem into steps the computer can do, decide what order the steps should be done in, etc. This process is not really different than it would be using the old human computer strategy. The problem needs to be prepared and planned for the assembly line. You then need some way of giving the computer your list of instructions, which will be interpreted by more computer code to eventually arrange circuits in the computer. There are 3 basic classes of options here: coding languages, numerical manipulation packages, and computer algebra systems. Each of these has strengths and weaknesses.

Coding languages are in some sense the most "primitive". There is less built-in functionality and, among the options mentioned, your instructions are "closest to the circuits", meaning that comparatively less additional interpretation of your commands takes place behind the scenes. This implies that the user often has to do more work, but has more flexibility and control. All things being equal, it also typically means the computer will get the job done faster once you get things set up. As far as I can tell, Python, C (and its variations like C++), and Fortran are the most common coding languages used in computational physics, though in principle almost any language could be used. Basically everything in this category can be set up for free on almost any computer.

Numerical manipulation packages are computer programs that allow the user to do a huge variety of numerical tasks: matrix manipulations, numerical integration, plotting, etc. The most common is Matlab, which is not free. To provide a general idea of the costs involved, as of January 2017, a basic home license cost \$150, and a commercial license cost over \$2,000. At Carleton, some research groups have licenses and students can install Matlab on their own computers for educational purposes at no cost `https://wiki.carleton.edu/x/sIN8Ag`. There is also an open-source competitor, which is free and largely compatible with Matlab, called Octave. These packages typically have more built in functionality over a primitive coding language. For example, in a numerical manipulation package you would likely find a one-line command to create a 6×6 identity matrix. In Fortran, you'd need several lines of more basic code to do such a thing.

The final class of tools are computer algebra systems. These were originally developed to do all of the symbolic algebra typically done in high school algebra and college calculus. They know that $a + a = 2a$, $\int \sin x dx = -\cos x + C$, etc., and a lot more higher-level math. The most common commercial products are Mathematica and Maple, which come with price tags similar to Matlab. However, Carleton has a full site license for Mathematica and it's one of the tools we'll use extensively in this course. It's on computers all over campus and you can install it on your own computer. See the resources tab on the Physics and Astronomy web page. There are also open-source computer algebra systems such as Maxima and SageMath.

For each class of products, the functionality I've emphasized is that associated with its origins, and the functionality for which it continues to excel. However, over time there has been considerable expansion of functionality in each class, such that these classes now overlap considerably. For example, one can now write what looks and behaves like primitive computer code in computer algebra systems and numerical packages. Numerical packages have a lot of computer algebra capability and the SymPy package for python makes computer algebra possible there. The NumPy package provides may numerical functions in python, and the computer algebra systems now have extensive built-in numerical capability. All of this is to say that at the level of the work seen in an undergraduate physics major, almost any of the above choices could get the job done.

The tool chosen for use by a company, research group, instructor, etc. is often determined by how the strengths of a given tool match the needs of the project(s) along with the culture of the subfield and the personal preferences of the individuals involved. Each tool has its own syntax (rules for typing ideas). Something as simple as a^b will be coded differently in a number of the options above. Most people start by learning one tool, then take a class with a different instructor, join a different research group, get a different job, etc., and find that their new colleagues are using a different tool that they must now learn. The good news is that learning the logic behind what computers can do and how to set up tasks for them is largely independent of the tool chosen. It's the syntax along with various built in packages that differ. In this class, we'll mostly use Mathematica, but we'll do a few adventures with other tools to give you a feel for the variation and some broader exposure. The golden rule of computation is as follows: if you know what you want to do, but need the syntax for some tool, Google it!

PCQ 1.6

Google "powers in Fortran" and "powers in Mathematica". How would you type a^b in each tool?

1.4 Tree Goals

The part that many people find easy about a course is focusing on the trees – the specific things to learn. This might include specific equations or techniques. Not that all of these things are easy, but it's easier to focus on them than the big picture forest goals. In classical mechanics, I find these coming in two basic forms. One form is the results and equations of classical mechanics, which might include things like Newton's Second Law or The Principle of Least Action. The other form are what we might call tools, which might include things like Taylor series expansion or separation of variables. These are little tricks that it's nice to have at your disposal by knowing what they are and when they're helpful, i.e. to have them in your "toolbox." Richard Feynman, another Nobel-Prize winning theorist tells an entertaining story about his toolbox in his biography *Surely You're Joking Mr. Feynman* [4], which is included below in slightly abridged form. Although he doesn't use the word "toolbox" until the end, there is another good "toolbox" thing he discusses in getting there. You might also be struck by the divide between theoretical physics and math. The entire book contains lots of little bits about how theoretical physicists think, but unfortunately it also contains some examples of bad behavior.

At the Princeton graduate school, the physics department and the math department shared a common lounge, and every day at four o'clock we would have tea. It was a way of relaxing in the afternoon, in addition to imitating an English college. People would sit around playing Go, or discussing theorems...

I still remember a guy sitting on the couch, thinking very hard, and another guy standing in front of him, saying, "And therefore such-and-such is true."

"Why is that?" the guy on the couch asks.

"Its trivial! Its trivial!" the standing guy says, and he rapidly reels off a series of logical steps ... which goes on at high speed for about fifteen minutes!

Finally the standing guy comes out the other end, and the guy on the couch says, "Yeah, yeah. Its trivial."

We physicists were laughing, trying to figure them out. We decided that "trivial" means "proved." So we joked with the mathematicians: "We have a new theoremthat mathematicians can prove only trivial theorems, because every theorem that's proved is trivial."

The mathematicians didn't like that theorem, and I teased them about it...

Topology was not at all obvious to the mathematicians. There were all kinds of weird possibilities that were "counterintuitive." Then I got an idea. I challenged them: "I bet there isn't a single theorem that you can tell mewhat the assumptions are and what the theorem is in terms I can understandwhere I can't tell you right away whether it's true or false."

It often went like this: They would explain to me, "You've got an orange, OK? Now you cut the orange into a finite number of pieces, put it back together, and it's as big as the sun. True or false?"

"No holes?"

"No holes."

"Impossible! There ain't no such a thing."

"Ha! We got him! Everybody gather around! It's So-and-so's theorem of immeasurable measure!"

Just when they think they've got me, I remind them, "But you said an orange! You can't cut the orange peel any thinner than the atoms."

"But we have the condition of continuity: We can keep on cutting!"

"No, you said an orange, so I assumed that you meant a real orange."

So I always won. If I guessed it right, great. If I guessed it wrong, there was always something I could find in their simplification that they left out.

Actually, there was a certain amount of genuine quality to my guesses. I had a scheme, which I still use today when somebody is explaining something that Im trying to understand: I keep making up examples. For instance, the mathematicians would come in with a terrific theorem, and they're all excited. As they're telling me the conditions of the theorem, I construct something which fits all the conditions. You know, you have a set (one ball)disjoint (two halls). Then the balls turn colors, grow hairs, or whatever, in my head as they put more conditions on. Finally they state the theorem, which is some dumb thing about the ball which isn't true for my hairy green ball thing, so I say, "False!"

If it's true, they get all excited, and I let them go on for a while. Then I point out my counterexample.

"Oh. We forgot to tell you that it's Class 2 Hausdorff homomorphic."

"Well, then," I say, "It's trivial! It's trivial!" By that time I know which way it goes, even though I dont know what Hausdorff homomorphic means...

Although I gave the mathematicians a lot of trouble, they were always very kind to me...

Paul Olum and I shared a bathroom. We got to be good friends, and he tried to teach me mathematics. He got me up to homotopy groups, and at that point I gave up. But the things below that I understood fairly well.

One thing I never did learn was contour integration. I had learned to do integrals by various methods shown in a book that my high school physics teacher Mr. Bader had given me... It had Fourier series, Bessel functions, determinants, elliptic functionsall kinds of wonderful stuff that I didnt know anything about.

h

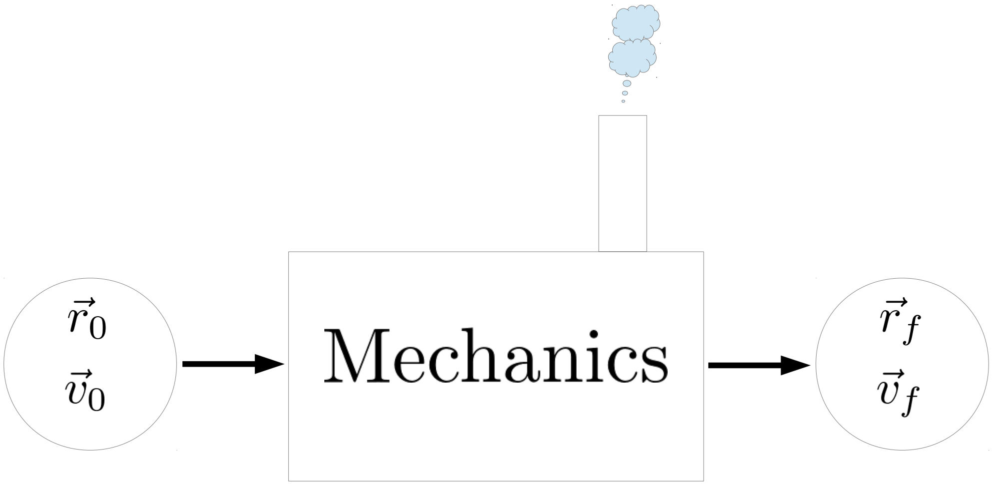


Figure 1.4: The mechanics machine.

> That book also showed how to differentiate parameters under the integral sign – it's a certain operation. It turns out that's not taught very much in the universities; they don't emphasize it. But I caught on how to use that method... So because I was self-taught using that book, I had peculiar methods of doing integrals.
>
> The result was, when [others]... had trouble doing a certain integral, it was because they couldn't do it with the standard methods they had learned in school. If it was contour integration, they would have found it; if it was a simple series expansion, they would have found it. Then I come along and try differentiating under the integral sign, and often it worked. So I got a great reputation for doing integrals, only because my box of tools was different from everybody else's, and they had tried all their tools on it before giving the problem to me.

No pair of excerpts could capture every aspect of theoretical physics and there are a few potentially misleading omissions here that I'd like to address. First, I'm not stuck on quoting Nobel Laureates. It's just that winning the Nobel tends to expand peoples' opportunities to write things like this. Next, I fear that one could get the impression that theoretical physics is only done by super geniuses that skipped third grade and choose to solve integrals over going to parties. This is certainly not the case. In fact, I know a very successful theorist who failed a course equivalent to the one you're taking now the first time around. That said, he worked very very hard to overcome this, which is part of a common theme. To steal a quote from one of colleagues, "physics is often better suited to the stubborn than the brilliant." Collaboration is really important in doing science. However for most theorists, it's at times little hidden. Theorists often work in small groups and work on the problem alone until they are either stuck on something or think they have solved a key piece, then they talk. They also build heavily on the work of others that they may or may not know through reading papers. If you look up Wienbergs 1967 paper, you'll find that it builds very directly on a lot of other work, and it acknowledges a useful conversation with someone at the end. Finally, opportunities to highlight demographic diversity are limited in the context of two selections. Though the population that has done theoretical physics up to now falls far short of matching the diversity of the human population at large, major contributions have been made by a diverse group, and we'll see some of that diversity as the course continues. Feynman's example of an "alternate tool box" actually highlights one of the benefits of diversity when a group of people are working on a problem.

1.5 Mechanics

Perhaps the final term we should define before leaving this "preliminaries" section is "Mechanics". In the present context it refers to a scheme or schemes that model the motion of a particle or particles as a function of time. In pictures, it's a machine that takes in initial conditions, the position $\vec{r}_0$ and velocity $\vec{v}_0$ of some particle or particles at some initial time and produces the position $\vec{r}_f$ and velocity $\vec{v}_f$ of the particles at some later time as shown in figure 1.4. Really, it gives position and velocity of the particles as a function of time, $\vec{r}[t]$ and $\vec{v}[t]$, where $\vec{r}_0$ is the value of the function at some initial time $t = t_0$ and $\vec{r}_f$ is the value of the function at some later time $t = t_f$.

A few notes on my notation. Note that I use square brackets [ ] for function notation as a matter of consistency with Mathematica. Note also that when I say $\vec{r}$, I mean the position of a particle in 3 dimensional space. Ultimately this vector could then be expressed in any coordinate system. When I say x, I mean the position of the particle along the x axis. When I define a vector like $\vec{r}$ and then use it with out the vector sign, I mean the magnitude of the vector. That is

$$r \equiv |\vec{r}|. \tag{1.3}$$

A triple line equal sign is standard physics notation for a definition. In this case, I'm introducing the new symbol r and defining it via Eq. (1.3). Mathematicians usually use := for the same purpose.

What can play the role of the machine in the figure? It should be a physical theory as discussed in Sec. 1.3.1, and what physical theories work will be a significant part of the course. It turns out that there is essentially more than one equivalent choice. One choice is Newton's laws along with the definitions of acceleration and velocity, as you probably already know from an intro physics course. We'll consider this version of mechanics, Newtonian Mechanics, in perhaps a little more detail than you've seen before in the next chapter. Then we'll soon meet other options.

PCQ 1.7

Based on your existing knowledge of Newton's Laws, do you consider them "obvious"? As in, do you say to yourself something like, "Clearly force equals mass times acceleration. Why does that Newton guy get so much credit anyway?" Answer yes or no, and use a few sentences to explain your position.

PCQ 1.8

A 2 kg ball is initially at rest at $\vec{x} = 0$. It experiences a constant 5 N force in the horizontal direction. Find the position of the ball after 3 s using your existing knowledge of Newtonian Mechanics. Use a few sentences to comment on how Newtonian Mechanics is doing the job of the machine drawn in Fig. 1.4.

Bibliography

[1] S. Weinberg, *Facing Up: Science and its cultural adversaries*, Harvard University Press, Cambridge, Massachusetts, 2001.

[2] T.S. Kuhn, *The Structure of Scientific Revolutions*, University of Chicago Press, Chicago, 1962.

[3] http://en.wikipedia.org/wiki/Scientific_theory, accessed 12/2012.

[4] R.P. Feynman, *Surely You're Joking Mr. Feynman*, Norton, New York, 1997.

Chapter 2

Anatomy of a theory: Newton's Laws

You have all seen and used Newton's laws. An alternative title for this section could be "basics mechanics from an advanced perspective". My goals in starting here is as follows: to provide a reminder of Newton's laws for those that have not thought about them for a while, to get you to think more deeply about them than you might have before, to treat them as an example of a theory to pave the way for constructing less familiar theories later, and to put in place a familiar environment for leaning some new tools.

You already know about how to use them, at least for some cases of varying complexity, but surprisingly, picking apart what they mean can be harder than knowing how to apply them to simple cases. In the next few sections, I give one interpretation of what they mean. Note however, that my interpretation is not unique, and there are others out there (see, for example [1]). I also must confess that after more than 2 decades of exposure to Newton's Laws, my own understanding is still evolving.

2.1 Newton *à la* Newton

So here they are, as stated by Newton himself (actually a translation) [2] in 1686.

AXIOMS, OR LAWS OF MOTION[1]

LAW I

Every body continues in its state of rest, or of uniform motion in a right line, unless it is compelled to change that state by forces impressed upon it.

PROJECTILES continue in their motions, so far as they are not retarded by the resistance of the air, or impelled downwards by the force of gravity. A top, whose parts by their cohesion are continually drawn aside from rectilinear motions, does not cease its rotation, otherwise than as it is retarded by the air. The greater bodies of the planets and comets, meeting with less resistance in freer spaces, preserve their motions both progressive and circular for a much longer time.

LAW II[2]

The change of motion is proportional to the motive force impressed; and is made in the direction of the right line in which that force is impressed.

If any force generates a motion, a double force will generate double the motion, a triple force triple the motion, whether that force be impressed altogether and at once, or gradually and successively. And this motion (being always directed the same way with the generating force), if the body moved before, is added to or subtracted from the former motion, according as they directly conspire with or are directly contrary to each other; or obliquely joined, when they are oblique, so as to produce a new motion compounded from the determination of both.

LAW III

To every action there is always opposed an equal reaction: or, the mutual actions of two bodies upon each other are always equal, and directed to contrary parts.

Whatever draws or presses another is as much drawn or pressed by that other. If you press a stone with your finger, the finger is also pressed by the

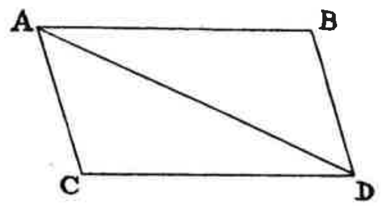
[[1] Appendix, Note 14.] [[2] Appendix, Note 15.]

stone. If a horse draws a stone tied to a rope, the horse (if I may so say) will be equally drawn back towards the stone; for the distended rope, by the same endeavor to relax or unbend itself, will draw the horse as much towards the stone as it does the stone towards the horse, and will obstruct the progress of the one as much as it advances that of the other. If a body impinge upon another, and by its force change the motion of the other, that body also (because of the equality of the mutual pressure) will undergo an equal change, in its own motion, towards the contrary part. The changes made by these actions are equal, not in the velocities but in the motions of bodies; that is to say, if the bodies are not hindered by any other impediments. For, because the motions are equally changed, the changes of the velocities made towards contrary parts are inversely proportional to the bodies. This law takes place also in attractions, as will be proved in the next Scholium.

COROLLARY I

A body, acted on by two forces simultaneously, will describe the diagonal of a parallelogram in the same time as it would describe the sides by those forces separately.

If a body in a given time, by the force M impressed apart in the place A, should with an uniform motion be carried from A to B, and by the force N impressed apart in the same place, should be carried from A to C, let the parallelogram ABCD be completed, and, by both forces acting together, it will in the same time be carried in the diagonal from A to D. For since the force N acts in the direction ofthe line AC, parallel to BD, this force (by the second Law) will not at all alter the velocity generated by the other force M, by which the body is carried towards the line BD. The body therefore will arrive at the line BD in the same time, whether the force N be impressed or not; and therefore at the end of that time it will be found somewhere in the line BD. By the same argument, at the end of the same time it will be found somewhere in the line CD. Therefore it will be found in the point D, where both lines meet. But it will move in a right line from A to D, by Law I.



2.2 Basics

Newton's laws amount to a set of assumptions about how objects move. These assumptions have been found to apply very broadly and aspects of them have been found to be valid to a high (but not infinite!) degree of sensitivity in experiments. Before we pick these laws apart one at a time in the following sections, there are a few general things to point out.

Based on our discussion of a physical theory in Sec. 1.3.1, we should construct Newtonian mechanics, as a logical structure based on definitions, assumptions, etc. In a logical structure, new concepts are defined in terms of existing ones. This is great once you get going, but it's impossible to get started, since in the beginning you have no existing concepts from which to begin defining new ones. Thus a logical structure usually starts with what are known as primitive notions: basic concepts left undefined. You might remember from high-school geometry that a point is an undefined term, or primitive notion in geometry. In mechanics, one choice of primitive notions is space $\vec{r}$, time t, and particle.

Next, we can define additional terms based on our primitive notions. For example velocity can be defined in terms of time and space:

$$\vec{v} \equiv \lim_{\Delta t \to 0} \frac{\Delta \vec{x}}{\Delta t} = \frac{d\vec{x}}{dt}. \tag{2.1}$$

We can then define more things based on existing definitions, for example acceleration can be defined in terms of velocity and time:

$$\vec{a} \equiv \lim_{\Delta t \to 0} \frac{\Delta \vec{v}}{\Delta t} = \frac{d\vec{v}}{dt}. \tag{2.2}$$

Newton's laws continue making some definitions. They also make assumptions about how some of the relevant quantities relate that are logically consistent, quantitative, and predictive. Thus to my mind, I think they make a good example of a physical theory, and hopefully, you've already convinced yourself that they do the job of the box labeled mechanics in Fig. 1.4.

2.3 Newton's First Law

Here Newton is extending to 3 dimensions and formalizing somewhat Galileo's law of inertia. The law seems quite innocent to us, but was quite a shock in Galileo's time as the prevailing perspective was that the natural state of an object was at rest. If we try to identify the logical structure in Newton's laws, we also find that it does a lot.

(1) Newton's first law begins to define a force. He says "compelled to change that state by a force impressed on it." This seems to define force as an external thing that causes acceleration, having origin outside of the particle.

(2) He is assuming that constant velocity does not need an explanation and focuses our attention on explaining changes in velocity. This is the key departure from the previous ideas about the natural state of objects being at rest. Previous thinkers had tried to explain why things might move at constant velocity rather than being at rest.

(3) He assumes that objects for which there are no external influences travel at constant velocity. This is not always true. Consider yourself sitting in a seat, wearing a seat belt, in a rocket ship drifting through deep space. Suppose a marble is floating in front of your nose inside the ship. Now suppose the rocket booster is fired. The marble formerly at rest in front of your nose will accelerate to the back of the ship. There was no external influence on the marble, yet it accelerated from your perspective. Thus Newton's first law restricts our attention to using perspectives, known as inertial frames of reference, in which objects do not appear to accelerate when no forces act. Once you find one inertial frame, there are a whole family of them that consists of all frames moving at constant velocity relative to the one you found, as you might recall from your special relativity studies. It actually turns out that one can still use Newton's laws in noninertial frames, by introducing what's known as fictitious forces. We talk about these in Advanced Classical Mechanics.

2.4 Newton's Second Law (N2)

Here Newton makes the relation between force and acceleration quantitative. Translating his words into symbols, we can write

$$\vec{F} = m\vec{a}. \tag{2.3}$$

Note that Newton never uses the word mass here. He just says "proportional", so the m here is just a proportionality constant, which we can give the name "mass", that we don't know anything about yet. It is also the case that this need not be the same m that appears in Newton's law of gravitation, as we'll discuss in the next chapter.

Many people feel that when Newton says "motion" he means momentum, which we can define at this stage to be $\vec{p} = m\vec{v}$. Thus one could state newton's second law as

$$\vec{F} = \frac{d\vec{p}}{dt}. \tag{2.4}$$

For constant m, this is just a repackaging of symbols. For now, let's not allow m's that are a function of time.

We also usually write Eq. (2.3) with a subscript "net" on the $\vec{F}$:

$$\vec{F}_{\text{net}} = m\vec{a}. \tag{2.5}$$

This says that when we have multiple forces they should be added as vectors. This follows from the single force version by virtue of the fact that $\vec{a}$ is a vector and a particle can only have one acceleration. Newton states this bit as a corollary at the end of the excerpt I've included above. He goes on from here to state a number of corollaries. Some people also refer to this as Newton's forth law.

Remember (at least from the current perspective) Newton's second law is an assumption used in constructing the theory. As such, it, along with the results that flow from it, should make quantitative predictions about nature that are falsifyable by experiment. Newton's second law is one place where it's easy to see how Newton had some choice, and I think this freedom implies his laws are not obvious. We'll explore some options in Sec. 2.7.

2.5 Newton's Third Law

First note that when Newton says action, he means force. This is particularly tricky for us, since action is defined to be something entirely different later in the course.

So Newton is assuming here that when a particle 1 exerts a force on particle 2 (pushes it for example), particle 2 exerts and equal and opposite force on 1. In symbols

$$\vec{F}_{1\,\text{on}\,2} = -\vec{F}_{2\,\text{on}\,1}. \tag{2.6}$$

As a result, Newton's third law gives us a way to compare masses. Suppose we have the two particles above exerting forces on each other as described by Eq. (2.6). By inserting Newton's second law, we have

$$m_2\vec{a}_2 = -m_1\vec{a}_1. \tag{2.7}$$

If we take the magnitude of both sides and do a little algebra, we have

$$\frac{m_1}{m_2} = \frac{a_2}{a_1}. \tag{2.8}$$

So by measuring the acceleration of each particle as they exert a force on each other, we can measure the ratio of their masses. Evidently the best we can do is compare two masses. Until November of 2018, there was a block in Paris that was defined to be 1 kg (there is now a definition in terms of fundamental constants). When you said your object was 2 kg, you mean that when it exerts a force on the block in Paris, the block in Paris accelerates twice as much as your object. That's really all we can say about mass in the context of Newton's Laws.

It turns out that for some kinds of action at a distance forces, Newton's third law is not true. We'll talk about this in Sec. 2.6

PCQ 2.1

So I've now given you my perspective on Newtons 3 laws. Come up with at least one question about the material in chapter 2 up to Sec. 2.5

2.6 Why Newton's Laws Are Useful Concepts

In the reductionist spirit of physics, I feel compelled to draw a connection between Newton's Laws and the law of conservation of momentum. Conservation of momentum says that there is a quantity called momentum $\vec{p}$ that nature feels it can't loose or gain. For classical nonrelativistic point particles, we have $\vec{p} = m\vec{v}$ as already stated. Therefore if two particles interact and particle 1's momentum changes by an amount $d\vec{p}_1$, particle 2's momentum must change by an amount $d\vec{p}_2 = -d\vec{p}_1$. Dividing both sides by a chuck of time dt over which this occurs gives

$$\frac{d\vec{p}_2}{dt} = -\frac{d\vec{p}_1}{dt}. \tag{2.9}$$

We can then use Newton's second law, written in terms of momentum to recover Newton's third law as a special case of conservation of momentum.

From this perspective, Newton's third law tells us that when two particles interact, the momentum lost by one is gained by the other. Force is a measure of how much the particles interact, and Newton's second law tells us how the size of the interaction controls how much momentum the particles exchange. If momentum could be just gained or lost in nature, tracking its flow would probably not be so helpful.

So far, I haven't really reduced the number of independent concepts here, but this still feels reductionist to me, because I know that momentum conservation applies across all of physics, not just Newton's laws. We'll also see later in the course that momentum is conserved as a result of the fact that the laws of nature don't depend on position $\vec{r}$.

As a final note here, I said earlier that there are some cases in which Newton's third law fails. This issue occurs when there are other places for momentum to live besides as mass times velocity of the particles. The other place it can be is in the fields, like the gravitational or magnetic fields, but our understanding of fields was far too limited at the time of the Principia for Newton to say this.

2.7 Alternate Versions

When people say Newton's Laws are obvious, I think it denies some of the creativity of science. In this section, we'll consider some other things that Newton could have written down instead of his second law. One of the main goals of this section is to experience "thinking like a theorist." The style of thinking required here is exactly like doing theoretical research. After writing down an alternative version of N2, we'll then need to check several things to see if Newton really could have chosen such a form. We can classify each version of N2 as follows: (a) not logically consistent with N1 or N3, (b) not complete in the sense that it does not give a relation between $\vec{F}$ and $\vec{a}$, (c) logically consistent, but equivalent to the usual form (d) logically consistent, different from the usual form, but ruled out experimentally, (e) logically consistent, different from the usual, and experimentally viable. Note that experimentally viable in this context means that for at least some values of the parameters in the proposed alternate N2, it has not been ruled out by existing experimental data. Note that we have not yet explicitly considered any of the separate theories that govenern the forces. While there is sometimes some flexability as to what goes into the force laws and what goes into N2, we consider the Fs as coming from the same usual force laws in all cases.

PCQ 2.2

Classify the following alternate versions of N2 based on the above scheme:
(If you don't feel confident that you can get this right, it's ok, just give it your best shot. Remember the PCQs are just intended to get you thinking, and participation credit requires only a thoughtful answer, not the right answer.)
1. $\vec{F} = b^2\vec{a}$, where b is some constant that may differ from particle to particle.
2. $\vec{F} = m\vec{a} + \vec{b}$, where $\vec{b}$ has constant nonzero components that may differ from particle to particle.

3.
$$\begin{pmatrix} F_x \\ F_y \\ F_z \end{pmatrix} = \begin{pmatrix} m_x & 0 & 0 \\ 0 & m_y & 0 \\ 0 & 0 & m_z \end{pmatrix} \begin{pmatrix} a_x \\ a_y \\ a_z \end{pmatrix},$$
where m_x, m_y, and m_z are particle dependent constants, that are not necessarily all equal.
4. $\vec{F} = \frac{m}{(1-v^2/c^2)^{3/2}}\vec{a}$, where c is the speed of light.

Some implications of 3 have been worked out by some of us at Carleton [3]

2.8 Application

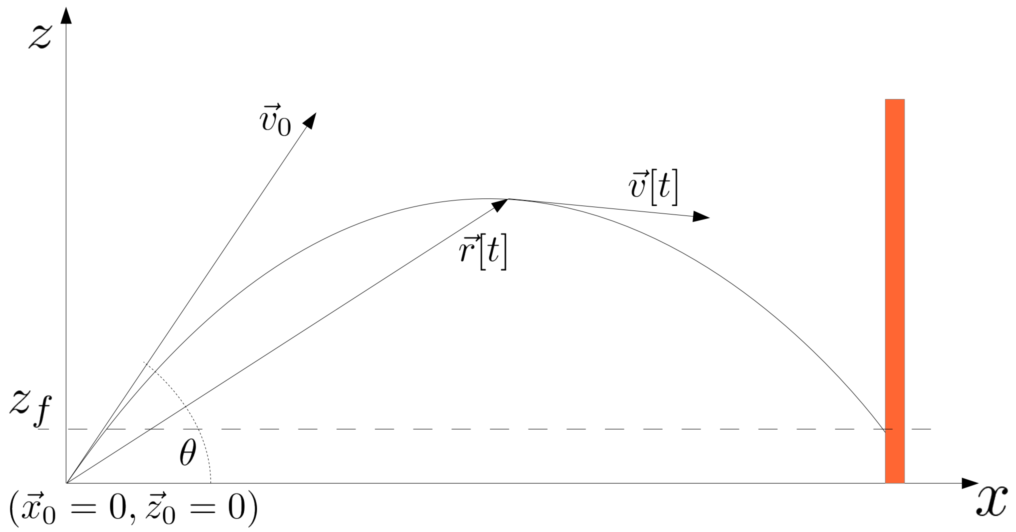
In this section we'll first consider a simple example of the application of Newton's Laws. It will be of the variety you've no doubt solved before. I've chosen it for the following reasons: (a) review, (b) illustrate the problem-solving strategy, (c) give you a chance to think about *how* Newton's Laws are working without really having to struggle to solve the problem, (d) point out a number of tools that will be useful to us later as the problems become more challenging, and (e) we'll want to add a drag force to this problem later so it will be good to have this simple drag-free case to refer back to later. After the example, we'll highlight what it is that we're doing in applying this "Newtonian Game".

2.8.1 Simple Example

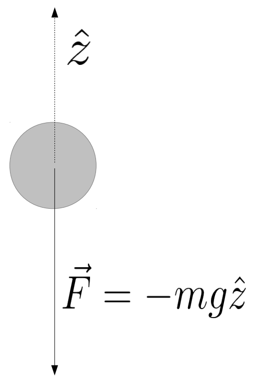
Bill the medieval engineer has been asked to lay siege to the castle St. Olaf. He wishes to get an initial estimate of whether or not he can position his trebuchet out of rang of the arrows the Ole's will fire from the top of the castle wall, while still having the wall in range of his trebuchet. For this initial estimate, he ignores the effects of aerodynamic drag. The ground is flat and horizontal. In working out his solution, Bill chooses an origin centered on his trebuchet, picks an x axis along the horizontal, and a z axis that points vertically up. To break the wall, he must hit it near the base, perhaps a small distance z_f above it. For this initial example, I'll take $z_f = 0$. To be out of range of the Ole's arrows, the wall must be at least a distance x_f from the trebuchet. The trebuchet is set up to fire at an angle θ above the horizontal, and it has a firing velocity of $\vec{v}_o$. Can Bill make a successful shot from a safe distance?

Being a successful and obedient medieval engineer (like all of you will be!), Bill produces the following solution following the problem-solving strategy exactly. Try to follow Bill's steps. Avoid any temptation to take a short cut. We're trying to illustrate methods here, not just jump to an answer.

Note also that although we have not actually met any forces yet in this course, we'll do that in Chapter 3, I'll rely on your existing knowledge that near the surface of the Earth the gravitational force is $\vec{F} = mg$ down. Sometimes people make a sign error here. Note that g is defined as a positive quantity and you should put an appropriate sign in front of it to match your coordinates. Here I'm using $\hat{z}$ up, so I should write $\vec{F} = -mg\hat{z}$.
Sketch:



Free Body Diagram (FBD):



Note that in free body diagrams all force vectors should have their tails on the center of the particle.
Target: We have some choice here. We just need to get the necessary equations to check if the numbers work. One approach is to take x_f as the target given $\vec{v}_0$ and see if it works.
Physical Principle(s): Gravitational force law as shown in the FBD, Newton's second law

$$\vec{F}_{\text{net}} = m\vec{a}, \tag{2.10}$$

along with the definitions of velocity and acceleration.
Solve for the variable of interest: With Eq. (2.10) and the gravitational force from the FBD we have

$$m\vec{a} = -mg\hat{z}. \tag{2.11}$$

Canceling the m and noting that acceleration is the second time derivative (dot over something indicates a time derivative) of position we have

$$\ddot{\vec{r}}[t] = -g\hat{z}, \tag{2.12}$$

which is an example of something called an Ordinary Differential Equation (ODE). We'll talk quite a bit about ODEs as we proceed. This can also be written

$$\dot{\vec{v}}[t] = -g\hat{z}, \quad \dot{\vec{r}}[t] = \vec{v}[t], \tag{2.13}$$

which, along with the initial conditions

$$\vec{v}[0] = \vec{v}_0, \quad \vec{r}[0] = \vec{0}, \tag{2.14}$$

form the initial value problem (more on this in the next section). The solution in this case is the familiar kinematic equation

$$z[t] = v_{0z}t - \tfrac{1}{2}gt^2, \quad x[t] = v_{0x}t. \tag{2.15}$$

(We'll discuss more formally how to get this in the next section). Here we're taking $z[t_f] = 0$, so the $z[t]$ equation can be solved for time yielding

$$
\begin{aligned}
t_f &= \frac{2v_{0z}}{g} &\quad (2.16)\\
&= \frac{2v_0}{g}\sin\theta. &\quad (2.17)
\end{aligned}
$$

Plugging this into the $x[t]$ equation yields the target

$$
\begin{aligned}
x[t_f] &= \frac{2v_{0x}v_{0z}}{g} &\quad (2.18)\\
&= \frac{2v_0^2}{g}\sin\theta\cos\theta &\quad (2.19)\\
&= \frac{v_0^2}{g}\sin[2\theta] &\quad (2.20)
\end{aligned}
$$

Plug in numbers: (none to put in)
Check: (1) $x[t_f] = 0$ if $\theta = 0$ ✓, (2) $x[t_f] = \max$ if $\theta = 45^o$ ✓, (3) $x[t_f] = 0$ if $v_0 = 0$ ✓.

PCQ 2.3
Check by differentiating that Eqs. (2.15) along with an appropriate y equation are the solution to ODE (2.12).

Note that in Eq. (2.20), Bill has used the identity

$$\sin\theta\cos\theta = \tfrac{1}{2}\sin[2\theta] \quad (2.21)$$

This is a good trick that should be tucked away in your toolbox. Note also the kinds of things that Bill does to check his result.

So now Bill can get a rough sense of whether his mission is doable. He knows the firing velocity v_0 and the firing angle θ. The solution $x[t_f]$ is the furthest away he can be and still hit the wall. He can then compare this to the archer's range. He should use some care in interpreting his results in the context of his assumptions.

2.8.2 Newtonian Game

We can identify the following recipe in what we just did, that could be referred to as the "Newtonian Game".

1. Verify that you have an inertial frame.
2. Identify the forces acting on the particle at its present location.
3. Add the forces as vectors to get $\vec{F}_{\text{net}}$.
4. Plug $\vec{F}_{\text{net}}$ into N2 and identify the initial conditions. This sets up the initial value problem, which consists of the differential equation(s) from N2 along with the initial conditions.
5. Use the initial value problem to find the position and velocity of the particle at a later time. This makes it clear that we've filled the mechanics machine in Chapter 1.

There are two basic ways to carryout step 5. One can sometimes use various techniques to be discussed in the next section to find a familiar function that solves the initial value problem. This is known as an analytical solution. Sometimes (or even much of the time) this is hard or not possible. An alternative approach is to use computational methods as we'll discuss later in the course.

2.9 Summary and Additional Reading/Practice

Based on this chapter, you should be able to:

- identify which of Newton's laws apply in a given situation

- apply Newton's laws to find forces, masses, or accelerations that are unknown in a given system

- evaluate a physical theory to assess its logical consistency and experimental viability

If you want to (optionally) read another advanced treatment of Newton's laws by another author, see Taylor Chapter 1 or Marion and Thornton Chapter 2. Each of these contain discussion of the meaning of Newton's laws, discussions of how to use them, and provides some advanced examples. There are problems in these chapters for additional practice. Your intro text, those on the course shelf, or the OpenStax text `https://openstax.org/details/books/university-s̄physics-s̄volume-s̄1` also have less advanced discussion, good examples, and practice problems. If you'd like to test your knowledge further after doing homework 1, answering some of the questions and working some of the problems in Giancoli Chapter 4 starting on page 103 or OpenStax Chapters 5 and 6 might be good. You can also challenge yourself by asking a friend to come up with an alternative version of Newton's 2nd law for you to evaluate.

Bibliography

[1] S.T. Thornton and J.B. Marion, *Classical Dynamics of Particles and Systems*, Brooks/Cole – Thomson, Belmont, California, 2004.

[2] I. Newton, *Newton's Principia. The mathematical principles of natural philosophy*, A. Motte, tr., D. Adee, New York, 1848.

[3] T.H. Bertschinger, N.A. Flowers, and J.D. Tasson, arXiv:1308.6572 `http://arxiv.org/abs/arXiv:1308.6572`; T.H. Bertschinger *et al.*, Symmetry, **11**, 22 (2019) `https://doi.org/10.3390/sym11010022`.

Chapter 3

Tricks of the Trade

Applying Newton's laws to more complex situations and solving the resulting equations to answer the relevant questions requires some mathy tools and tricks. The same tools are useful later when we get to Lagrangian and Hamiltonian mechanics, and many of them are usefull all over physics. So while we're thinking about deep *ideas* related to the meaning of physics, etc., this section will give us some more calculational bits to focus on.

3.1 Taylor Expansion

You may have learned about Taylor expansion in a calculus class, or you may have used some of its special cases without even knowing it. The small angle approximation

$$\sin\theta \approx \theta, \tag{3.1}$$

if $\theta << 1$ that you may have met in your introductory lab or the binomial approximation

$$(1+x)^a \approx 1 + ax, \tag{3.2}$$

if $x << 1$ that you may have met in special relativity are special cases of truncated Taylor expansions. This section will provide the basic math behind generating the Taylor expansion of a function. We'll talk more about what it means and why it's so awesome for doing physics in class.

The key behind Taylor expansion is that any suitably well-behaved function can be approximated in some neighborhood by a polynomial. To convince yourself that this so, consider the function $f[x] = \cos x$ in the neighborhood of $x = 0$. Figure 3.1 demonstrates that this function is well approximated near $x = 0$ by the polynomial $p[x] = 1 - \frac{1}{2}x^2$. Of course $f[x]$ and $p[x]$ will continue to diverge as we move away from $x = 0$ since $p[x]$ is not equal to $f[x]$. It is only an approximation for $f[x]$ near $x = 0$.

So, given a suitably well-behaved function $f[x]$, how do you find a polynomial approximation for it near a specific value x_0? The answer is the Taylor series defined as follows

$$f[x] = f[x_0] + \frac{f'[x_0]}{1!}(x - x_0) + \frac{f''[x_0]}{2!}(x - x_0)^2 + \frac{f'''[x_0]}{3!}(x - x_0)^3 + \dots \tag{3.3}$$

To be specific, this equation is called "the Taylor series of f centered at x_0". Here primes denote differentiation of a function with respect to its argument e.g. $f''[x_0]$ means differentiate the function $f[x]$ two times with respect to x, then plug in the number x_0. For some other common language, people often say that the polynomial is generated by "expanding f about x_0" or the "expansion is centered at x_0." Notice that there is an equals sign between the function and the polynomial. This is exact (rather than an approximation) if the infinite series of terms is included. The point x_0 is known as an expansion point. For points near the expansion point, the terms in the polynomial typically get smaller with increasing powers of $(x - x_0)$. Hence a polynomial *approximation* for $f[x]$ near x_0 with a finite number of terms can be generated by truncating the series in Eq. 3.3. Where should you truncate? The more terms you keep, the better your approximation will be.

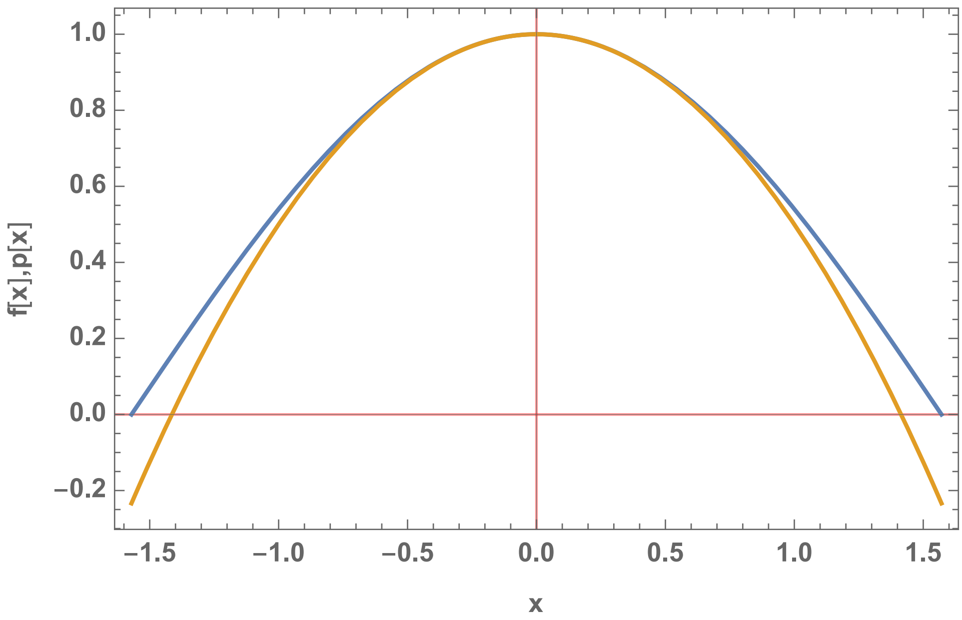


Figure 3.1: The function $f[x] = \cos x$ (upper blue curve) and the polynomial $p[x] = 1 - \frac{1}{2}x^2$ (lower yellow curve).

PCQ 3.1

In our example in Fig. 3.1, $p[x]$ is generated from the first few terms of the Taylor series of $\cos x$ centered at 0. See if you can obtain $p[x]$ using Eq. 3.3 with $f[x] = \cos x$ and $x_0 = 0$.

Some random additional information . . . Note that to truncate the series and get something meaningful in our example involving $\cos x$, we need $x << 1$. This tells us that the higher powers in the series are getting less and less significant. Note also that I can't say $x << 1$ unless x has no units, since clearly 1 has no units. Here I'm ok, since it's an angle in radians. When a Taylor series of f is centered at 0, this special case is called the Maclaurin series of f, but I tend not to use this word and stick with Taylor instead. When I say "suitably well-behaved", I mean the function f and all of its derivatives need to exist on the interval of interest, but I rarely think about these technicalities. When Taylor expansion jumps out as a usefull tool in physics, they are almost always satisfied. Taylor series sometimes converge for suitable values of x.

The "Series[ ]" function in Mathematica will produce a Taylor series. Do "?Series" in Mathematica to get instructions on how to use it.

If you want some practice, try showing the following by expanding the function of x about $x = 0$

$$\sqrt{1+x} = 1 + \tfrac{1}{2}x - \tfrac{1}{8}x^2 + \ldots \tag{3.4}$$

$$\cos[a+x] = \cos a - x \sin a + \ldots \tag{3.5}$$

For additional mathy discussion, as well as examples and exercises, see section 8.8 starting on page 602 of *Calculus* by Bradley and Smith on the course reserve shelf, or any other calculus text. For applications in physics and more conceptual background, see our in-class discussion and/or videos on Moodle.

PCQ 3.2

Computational Corner 1: A number of our readings will direct you to a tutorial to build some skills with some computational tool. As today's Computational corner, download and open the file "00Mathematica.nb" on a computer that has Mathematica installed. It's the first of a series that will introduce you to some of Mathematica's features. Work through the tutorial. Even if you have some experience with Mathematica, I suspect you'll learn something new as you do it. As with any part of the reading, be ready to ask questions about anything you got stuck on. You don't need to submit your work on the tutorial. As your submission for this PCQ, answer the following question: Why should symbols that you define in Mathematica start with a lower-case letter?

3.2 Coordinates and Vectors

3.2.1 Coordinates

Note that in our chapter 2 example, Bill chose Cartesian coordinates (x, y, z), and he chose both the origin and directions of the axis to be useful for the given problem. This is an important skill. In classical mechanics we work with 3 dimensional (3D) flat spaces and Cartesian coordinates are often a useful choice. You should note however that they are not the only choice and there are a great many alternatives.

We'll meet a few alternatives to Cartesian coordinates in this course. I'll introduce them as the come up, but I'll provide one now just to give you a specific example of an alternative to have in mind. Lets stick to the 2D plane so that I can draw pictures. Thus the Cartesian choice would be (x, y). An alternative known as plane polar coordinates uses (r, ϕ). Here r is a distance from the origin and ϕ is an angle around the origin from the x axis as shown in Fig. 3.2. Note that every point (except the origin) has a unique pair of (r, ϕ) coordinates just like it has a unique set of (x, y) coordinates. So for a 2D plane, plane polar coordinates is a method of labeling points that does the same job as Cartesian coordinates.

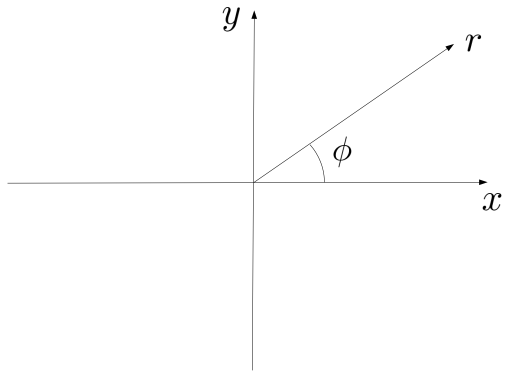


Figure 3.2: plane polar coordinates.

As one might expect, there is a relationship between the two systems. That is, x and y can be expressed in terms of r and ϕ, and such relationships are useful in changing to a new choice of coordinates. You'll explore this relationship in the exercise below:

PCQ 3.3

Here is x in terms of r and ϕ:

$$x = r\cos\phi. \tag{3.6}$$

Find y in terms of r and ϕ.

Note that in keeping with our notation, $\vec{r}$ is the position vector, while r is its magnitude, which also happens to be the radial coordinate of a particle sitting at a particular point in plane polar coordinates.

Why would anyone want to use such a system over Cartesian coordinates? The answer is usually that a particular choice of coordinates sometimes simplifies the problem. The idea is similar to the way in which choosing a particular origin can simplify a problem. If you imagine, for example, a case in which a particle moves in a circle around the origin, you might expect plane polar coordinates to be particularly helpful.

3.2.2 Vectors

Newton's laws make use of vectors. You've all met vectors at some point in your past, and most of you probably think of them as little straight line segments with arrow heads on one end sitting in three dimensional space (like the ones in the FBD of our example). The length of the line segment gives its magnitude, and the orientation of the arrow head gives its direction. You might also recall that vectors can be moved around in 3D space without changing them. Thus the location of the vector is a piece of information not contained within the vector itself that must be specified separately.

We've already been using a lot of vector terminology. Since most of us have dealt with this stuff already, I'll not review all of the details here, but **you should know the following**: meaning of magnitude and component, how to find the components given the magnitude and an angle, how to find the magnitude give the components, how to add and subtract vectors, how to find the dot product, and how to scalar multiply a vector. These things are a part of what we might call vector algebra. If you feel rusty, consult an intro text (there are some on the course reserve shelf) and if necessary Jay, your classmates, the problem-solving facilitators, etc.

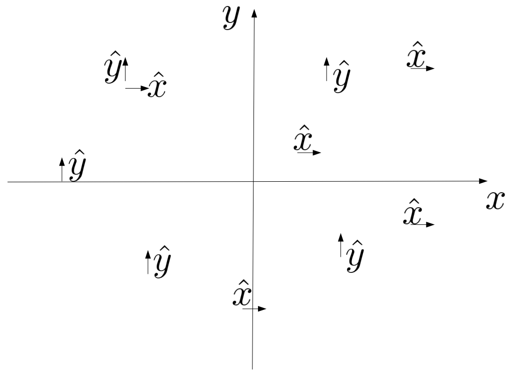


Figure 3.3: sample Cartesian coordinate basis vectors.

Unit vectors and basis vectors are two tricks that will be useful to us, that you might not have thought a lot about in the intro course. Unit vectors are simply vectors of magnitude 1 (not even any units). I'll denote them with a "hat" as opposed to a vector sign. The easiest way to make a unit vector is to simply divide a vector by its magnitude as in

$$\hat{r} = \frac{\vec{r}}{r}. \tag{3.7}$$

One place we'll find unit vectors helpful is in creating what's known as a unit basis. A set of basis vectors are a collection of vectors that one can express any other vector in terms of. If all of the members of this set happen to also be 1 unit long, it can be particularly neat and the basis becomes a unit basis. Basis vectors go along with a set of coordinates and are tangent to the coordinate curves. In the example above, I was using $\hat{x}$ as a unit vector along the x direction (tangent to "x" curves), $\hat{y}$ as a unit vector along the y direction, and $\hat{z}$ as a unit vector along the z direction. These unit basis vectors can be thought of as existing everywhere in space as shown in Fig. 3.3, hence we might call them a vector field. Thus any vector can be expressed in terms of these basis vectors. In our example we had

$$\vec{a} = 0\hat{x} + 0\hat{y} - g\hat{z} = -g\hat{z}. \tag{3.8}$$

Note that the names of the basis vectors are not universal. Another popular choice for the Cartesian unit basis vectors is $\hat{\imath}, \hat{\jmath}, \hat{k}$.

PCQ 3.4

In plane polar coordinates, $\hat{r}$ and $\hat{\phi}$ form a unit basis. Draw the x, y plane. Draw 4 sample $\hat{r}$ vectors and 4 sample $\hat{\phi}$ basis vectors. Try to choose your samples to illustrate "interesting" behavior.

Relations between the Cartesian unit bases and other basis vectors can also be found. For example in plane polar coordinates

$$\hat{r} = \hat{x}\cos\phi + \hat{y}\sin\phi. \tag{3.9}$$

PCQ 3.5

Show that $\dot{\hat{r}} = \dot{\phi}\hat{\phi}$. Does this make sense given your answer to the last PCQ?

This highlights a trap that can catch you when using non-Cartesian basis: if the velocity is $\vec{v} = k\hat{\phi}$, where k is a constant, the velocity vector is not constant and the acceleration is not zero! Another feature of non-Cartesian systems you might want to be aware of is that they often have a few points where they are not well defined. Here $\hat{\phi}$ is not well defined at the origin.

Some people like a notation popular in calc 3 texts in which components of vectors are expressed as $\vec{a} =< 0, 0, -g >$. This can lead to trouble in the context of alternative coordinate systems since the choice of coordinates and basis vectors is hidden in the notation. For example, we could have a vector in plane polar coordinates $\vec{a} = -g\sin\phi\hat{r} - g\cos\phi\hat{\phi}$. The calc 3 type notation does not make clear whether our components are $\hat{x}, \hat{y}$ components or $\hat{r}, \hat{\phi}$ components. When we do calculus with vectors, we find that hiding the basis vectors can be particularly dangerous since $\frac{d\hat{\phi}}{dx} \neq 0$.

It's perhaps interesting to note that there is much more to vectors and coordinates than we're considering here. For example, the idea of a "straight line segment" does not make sense beyond flat space, so for curved spaces a modified definition of a vector is needed. Here we're considering orthogonal basis vectors; however, one need not do so. One can have vectors in spacetime rather than just space. If you have a chance to take a General Relativity course or a Differential Geometry course, both offer cool opportunities to learn more about vectors and coordinates.

For additional discussion of polar coordinates, see Taylor section 1.7.

3.3 ODEs

Newton's second law always results in at least one differential equation to be solved. Our other formulations of mechanics will too. The equations can be written in many forms, but for now consider the following form of the equation we found in the chapter 2 example in Sec. 2.8.1 as a specific example:

$$\ddot{z}[t] = -g. \tag{3.10}$$

So we know what two time derivatives of the function $z[t]$ is, but we'd like to know the function explicitly. Thus, in the context of the present example, we want a function $z[t]$ that, when differentiated twice, is a constant $-g$. So a differential equation then is a relation involving derivatives of a function that the function must satisfy, and solving the differential equation means finding the function. The adjective "ordinary" refers to the fact that ordinary, as opposed to partial, derivatives are involved.

There is a whole branch of mathematics that studies differential equations. The investigation of when solutions to various types of equations exist and techniques for finding those solutions are just a few examples of the things one can study. If you're lucky, you might be doing the differential equations course now and learning all of this fun stuff in parallel with the applications we see in this course. If you're not doing the ODE course now, it's ok. I'll tell you all of the techniques you need to know.

ODEs are classified based on a number of features. The ODE (3.10) is an example of a second-order equation, because the highest derivative of the function we seek is a second derivative. The equation

$$\dddot{z}[t] + \ddot{z}[t]^2 = 0 \tag{3.11}$$

would be an example of a third order ODE. ODE (3.10) is a linear equation, since the highest power of the function or its derivatives is 1. ODE (3.11) is a nonlinear equation, since a derivative of the function is raised to a power. Finally ODE (3.11) is homogeneous, since there are no terms that don't involve the function we seek, while ODE (3.10) is inhomogeneous. Putting it all together, ODE (3.10) is a linear, second order, inhomogeneous ODE, while ODE (3.11) is a nonlinear, third order, homogeneous ODE.

In mechanics, the second order ODEs can be rewritten as a pair first order ODEs. Here ODE (3.10) can be rewritten

$$\dot{v}_z[t] = -g \quad \dot{z}[t] = v_z[t]. \tag{3.12}$$

We can see the answer to the first equation via the fundamental theorem of calculus

$$v_z[t] = C - \int g dt = C - gt. \tag{3.13}$$

Since $v_z[0] = C$, we can interpret the constant as the z component of the velocity at $t = 0$. If we don't like $t = 0$ as the initial time, we can split up the constant as follows:

$$v_z[t] = v_{z0} - g(t - t_0), \tag{3.14}$$

where the original constant C has been renamed $v_{z0} + gt_0$. We can then apply the same procedure to the second equation in (3.12) and we find (taking $t_0 = 0$)

$$z[t] = z_0 + v_{z0}t - \tfrac{1}{2}gt^2. \tag{3.15}$$

Note that we need two "initial conditions" to fully specify the solution. A differential equation along with enough initial conditions to fully specify the solution is known as an initial value problem (IVP).

Above I used the fundamental theorem of calculus to write down a solution. Another way of viewing this process that has some advantages in more complicated cases is separation of variables. That is, in the context of the first equation in (3.12), put the v's on one side and the t's on the other. Then integrate both sides as follows:

$$\int_{v_{z0}}^{v_z} dv_z' = \int_{t_0}^{t} -g dt'. \tag{3.16}$$

This will yield the answer in the form (3.14) directly. Note that it's advisable to use definite integrals here and the limits must be "corresponding" in the sense that v_{z0} is the velocity at time t_0. We'll introduce a number of additional techniques for solving ODEs as they come up.

Note that Eq. (2.12) is a vector differential equation. The z component is the most interesting case, so I've pulled out that part as a scalar differential equation in Eq. (3.10). The other two components are

$$\ddot{x} = 0, \quad \ddot{y} = 0. \tag{3.17}$$

They are easily solved by the equations

$$x[t] = v_{x0}t + x_0, \quad y[t] = v_{y0}t + y_0. \tag{3.18}$$

Make sure you see how to get these solutions and why they make sense.

3.4 Characteristic Quantities and Unitless Variables

This section introduces two tricks that appear all over physics: unitless variables, and characteristic quantities. You'll want them in your toolbox.

3.4.1 Characteristic Quantities

In many physical problems there are constants that set the basic scale for variables that have particular units. They are given the name "characteristic (name of relevant quantity)". For example, a characteristic acceleration is a name given to some constant in the problem that has units of acceleration. A characteristic speed is a name given to some constant in the problem that has units of speed. Sometimes they appear directly in the problem.

PCQ 3.6

In our example in Sec. 2.8.1, a characteristic acceleration a_c and a characteristic speed v_c can be identified with constants that appear explicitly in the problem. Find them. (Hint: this shouldn't be too hard. You're looking for the two constants in the example that have the required units.)

Sometimes, you'd like to find a characteristic quantity, but no constant of the correct units appears directly in the problem. For example, you might want a characteristic time in our example in Sec. 2.8.1, but there is no constant in the problem with units of time. Often these can be constructed from other characteristic quantities that you can find directly in the problem.

PCQ 3.7

Using various combinations of the characteristic acceleration and speed you found in PCQ 3.6, find the characteristic time for the problem t_c and the characteristic length l_c. (Ok fine, I'll do one for you: $t_c = v_0/g$. So it's a constant in the problem that has units of time, hence a characteristic time.)

So what good are characteristic quantities? There might be lots of answers, but I can immediately identify 2. In this course, they'll be really helpful in constructing unitless variables, as we'll discuss in the next subsection. They are also helpful in getting a sense of what the basic scale of a problem is likely to be.

An example of the "basic scale" type application is found in one of the key questions facing theoretical physics today. Physics has proceeded with its reductionist approach to a point at which a description of nature based on 2 fundamental theories has emerged: the Standard Model of particle physics (basically a description of all forces except gravity using quantum mechanics), and General Relativity (a description of gravity based on spacetime curvature that does not include quantum mechanics). I already wrote down these two theories in Eq. (1.2). The combination of these two theories provides an accurate account of all experimental results achieved to date. That is, there is no experimental problem with either of them or using them both together. However, at energy scales very far above even the reach of the LHC (Large Hadron Collider, our largest particle accelerator at the moment), it appears likely that there are problems with the completeness or logically consistency of applying these two theories together. Thus most physicists expect that there should be some quantum theory of gravity, to match the other forces of nature. (For a more extensive, though I think still digestible, discussion of the need for quantum gravity, see the introduction of Ref. [2].)

What does all of this have to do with us right now? Well, we can use our new technology of characteristic variables to get a sense of where we could expect to find problems. There are lots of ways to approach the question. One way is to ask, how small must a section of space be in order for quantum effects on its curvature to be relevant? The constant that controls the scale of quantum effects is $\hbar = 1.06 \times 10^{-34}$ Js, while the constants that control curvature in general relativity are Newton's constant (the one from his universal law of gravitation) $G = 6.67 \times 10^{-11}$ Nm^2/kg^2 and the speed of light $c = 3.00 \times 10^8$ m/s. These can be packaged into a characteristic length, known as the Planck length l_p, as follows:

$$l_p = \sqrt{\frac{\hbar G}{c^3}} = 1.6 \times 10^{-35}\text{m}. \tag{3.19}$$

Our present experiments are very far indeed from probing sizes this small (distance across an atomic nucleus is of order 10^{-15} m). Our only hope for seeing these effects in the foreseeable future appears to be the hope that nature may have left some clue in physics accessible with current technology, or that which might be built in the near future.

Before moving on, let's see how the "basic scale" argument works in the context of our familiar example from Sec. 2.8.1. If the t_c we chose above really is a characteristic time for the problem and there is no "fine tuning" (see below), then we'd expect to see interesting things happen at times that are within an order of magnitude or so of t_c. Indeed that's the case. One finds that the ball reaches the maximum height along its trajectory at a time $t_{\max} = t_c \sin\theta$. This will be within an order of magnitude or so of t_c unless θ is set to a value that is very very close to zero using a sensitive instrument, i.e. "fine tuned". One also finds that the ball arrives at the base of the wall in a time $t_f = 2t_c \sin\theta$, which is also within an order of magnitude or so of t_c provided we don't fine tune the angle. If we find that it takes $1000t_c$ or $0.0001t_c$ for something interesting to happen in the problem, then we might suspect that either we made a mistake or something special is going on. Either way, we should investigate the problem further.

PCQ 3.8

After hearing a wonderful presentation on characteristic quantities from gentle, charming, intelligent, medieval engineer Tasson, sassy student, Major Payne-Diaz of the royal army, raises the following thoughtful question (raised politely in the past by many real and thoughtful mechanics students). "What if the base of the castle is in a pit 1000 m (or more) deep? Huh Tasson? Then the time to hit the wall will be much longer than your t_c. So I think this characteristic quantity stuff is pointless." What should Tasson say to the Major (if he doesn't want to loose his head) to address her concern?

3.4.2 Unitless Variables

Now that we have characteristic quantities, unitless or dimensionless variables are straight forward to define. For example, a unitless acceleration A can be defined as $A \equiv a/a_c$. By construction, this new quantity has no units since it's the ratio of two things with the same units. It's also an acceleration, since it's the acceleration divided by a constant. Hence it would typically be referred to as "unitless acceleration" or "dimensionless acceleration". This is sort of like choosing cleaver units for the problem. If our characteristic acceleration were $a_c = g = 9.81\text{m/s}^2$, then when $A = 1$, $a = g = 9.81\text{m/s}^2$, when $A = 2$, $a = 2g = 19.62\text{m/s}^2$, etc. Note, since A is a unitless variable, it is not appropriate to list any units for the value it takes. (I know, since kindergarten people have been drilling you to include units. Now I'm telling you not to.) Proceeding then, dimensionless time is $T = t/t_c$, dimensionless speed is $V = v/v_c$, etc. Note that this notation using capitals for dimensionless variables is not at all standard. Many people use the corresponding Greek letter. That's great for t becomes τ, but harder for l, c, etc.

You've likely already seen some use of dimensionless variables, but you may not have called them that. Any exponent must be dimensionless. For example, $(10\text{s})^{5\text{m}}$ doesn't make sense. After all, what units would the answer have? Similarly most functions must eat a unitless thing. The expression, $\sin(t)$, where $t = 5$ s also does not make sense. You've probably seen things like $\sin(\omega t)$, where ω is angular frequency with units of rad/s. But rad is not a "real" unit, it's just a place holder to remind the reader that the quantity is a unitless angle. An angle in radians is a ratio of arc length to radius, hence it's unitless. So when you see $\sin(\omega t)$, $T = \omega t$ is a dimensionless time, and $t_c = 1/\omega$ is a characteristic time.

So we can now ask the question, what are dimensionless variables good for? We've already seen one use of them above, and I know of at least 3 more.

The most important use of dimensionless variables in this course will be in plotting results. Suppose you want to use *Mathematica* to plot something like x_f as a function of launch angle θ in our example in Sec. 2.8.1 where we found

$$x_f = \tfrac{v_0^2}{g}\sin[2\theta]. \tag{3.20}$$

You can't do that without picking some values for v_0 and g. It turns out that the fraction on the right hand side is the characteristic length. So in terms of dimensionless variables, the equation can be written

$$X_f = \sin[2\theta]. \tag{3.21}$$

Thus dimensionless range can be plotted vs launch angle without the need to pick any values.

A second use is in generating cleaner expressions. If you're going to do a bunch more algebra, Eq. (3.21) is neater than Eq. (3.20). This advantage grows as problems become more complex.

Finally, one can often extract the functional behavior of an integral without doing the integral. Consider the question, "how does F depend on α, where

$$F = \int_0^\infty e^{-\alpha t} dt \tag{3.22}$$

is the definition of F?" By converting to the dimensionless variable $T = \alpha t$, the integral becomes

$$F = \frac{1}{\alpha}\int_0^\infty e^{-T} dT. \tag{3.23}$$

The upper limit is still infinity, since infinity times α is still infinity and the lower limit is still 0 since 0 times α is still 0. Now T is the only variable left in the integral and it is an integration variable. So once the integral is done, it will just be a number and the only occurrence of α is on the outside.

Thus F goes like $1/\alpha$. Incidentally, how did I know that the dimensions of α were 1/time?

PCQ 3.9

Computational Corner 2: Do tutorial "01Mathematica.nb". As your response to this PCQ, create a notebook with a nice plot of $\sin[x]$ from 0 to 2π. Give the plot the title "Your Name's Sin Function", where you should replace "Your Name" with your name.

Bibliography

[1] S.T. Thornton and J.B. Marion, *Classical Dynamics of Particles and Systems*, Brooks/Cole – Thomson, Belmont, California, 2004.

[2] C. Callender and N. Huggett, *Physics Meets Philosophy at the Planck Scale*, Cambridge University Press, Cambridge, 2001.

Chapter 4

Force and Energy

In our interpretation, Newton's Laws give us a method of predicting the future position and velocity of a particle once we know the force on it. Thus to make further progress we need additional theories that predict properties of the forces that seem to be present in nature. These forces can be constructed based on definitions, assumptions, etc. following the rules highlighted in the previous chapters, however, logical consistency requires that we don't write down things that are inconsistent with Newton's laws.

A complete treatment of all of the details of the theories of various forces can fill entire courses. Fortunately, we can make a lot of progress on the cases of interest to us with the rather minimal treatment considered here. Although we won't dig so deep at the fundamental level here, we'll still see some cool consequences of some interesting forces. First I'll remind you of some familiar forces in Sec. 4.1, then we'll consider aerodynamic drag in Sec. 4.2, which will be very important for our siege project. We'll address the connection with energy in Sec. 4.3. We'll also find that some of these forces, or combinations of them, when plugged into Newton's second law will yield ODEs that we can't solve using the analytical methods we've discussed. Section 5.1 talks about how to solve these equations using some numerical techniques.

4.1 Familiar Forces

You are probably already familiar with a number of forces from introductory physics. This section will review those forces and probably point out some aspects you might not have considered before. For more information about these forces and opportunities to practice with them, see an introductory physics text such as Giancoli or OpenStax on the course reserve shelf.

4.1.1 Electricity and Magnetism

Most of the forces we deal with in classical mechanics are really manifestations of Electricity and Magnetism (E&M), which you'll study later in much more detail. In this class, we'll deal primarily with manifestations of E&M that are most easily treated without reference to the theory at the fundamental level. In fact, as I'll point out in the sections to follow, most of the forces I write here could really set under this subsection heading, though in practice they are well-disguised. Many were well known before they were recognized as consequences of E&M. This is another example of the reductionist nature of physics we discussed in chapter 1.

One case in which we can make use of a rather fundamental prediction of the theory of E&M is the force experienced by a point particle with electric charge q_2 due to another point particle with charge q_1. The relevant force law is known as Coulomb's law and can be written

$$\vec{F}_{1\,\mathrm{on}\,2} = \frac{kq_1q_2}{r_{1,2}^2}\hat{r}_{1,2}. \tag{4.1}$$

Here $k = 9.0 \times 10^9$ N$\cdot$m/C^2 is a constant that can be informally viewed as "controlling" the size of the electric force, $r_{1,2}$ is the magnitude of a vector $\vec{r}_{1,2}$ pointing from particle 1 to particle 2, and $\hat{r}_{1,2}$ is the corresponding unit vector.

PCQ 4.1

When Coulomb's law is written as in Eq. (4.1), it's possible to seen from the form of the equation that Newton's third law is satisfied by the electric force. Describe how to see this.

Though the direction information is perfectly actuate as usually written in Eq. 4.1, be warned that it is often easier in practice when solving problems in Cartesian coordinates to just get the magnitude of the force from Coulomb's law and construct the Cartesian components with the correct signs by examining the picture.

4.1.2 Gravity

Gravity is perhaps the only force we'll consider in the course whose roots can't be traced to E&M, and just like E&M, one could spend years studying the theory and it's applications. Some of you might spend a term studying the fundamental theory as we know it, General Relativity, and you may spend much time in astronomy and geophysics studying its implications. Here we'll treat just the gravitational force on a point masses m_2 due to a point mass m_1, a result known as Newton's law of universal gravitation, which can be written

$$\vec{F}_{1\,\mathrm{on}\,2} = -\frac{Gm_1m_2}{r_{1,2}^2}\hat{r}_{1,2}. \tag{4.2}$$

Note the striking resemblance to Coulomb's law. We've just replaced k with Newton's constant G, charges with masses, and added a minus sign. Everything else is the same.

PCQ 4.2

Take a guess at why Coulomb's law and Newton's law of universal gravitation have the same form.

It turns out that Newton's law of universal gravitation works for any spherically symmetric masses, not just point masses. After you take E&M in the spring, you should be able to show this. With this result and the wonderful tool of Taylor expansion one can show that the gravitational force on a particle of mass m near the surface of the Earth is (approximately) the familiar

$$\vec{F} = -mg\hat{r}, \tag{4.3}$$

where $\hat{r}$ is a unit vector pointing away from the center of the Earth. The familiar g here is an appropriate packaging of constants from Newton's law of gravitation as applied to Earth,

$$g = \frac{GM_E}{R_E}, \tag{4.4}$$

where M_E and R_E are the mass and radius of Earth respectively. It should be referred to as the gravitational field of the Earth near the surface, but is often referred to as the acceleration of gravity near Earth's surface.

The m's appearing in the law of gravitation (which we could call gravitational mass) and the m's appearing in Newton's second law (which we could call inertial mass) are the same physical quantity to a high (but not infinite) degree of experimental sensitivity. However, if they are not the same, perfectly reasonable theories can be constructed. The fact that they are the same, is a key to General Relativity as we know it, and is sometimes referred to as The Weak Equivalence Principle [1]. Since it is of fundamental importance, there is a whole industry of experiments testing Weak Equivalence, including a recent space mission [2], the results of which several of us at Carleton used for further study [3]. The equality of inertial and gravitational mass means that a particle experiencing only the gravitational force has acceleration equal to the gravitational field justifying the nomenclature "acceleration of gravity" for g.

PCQ 4.3

When you drop a tennis ball, you find that the acceleration is somewhat different from g given by Eq. (4.4). What other forces are involved?

For additional advanced discussion of Newtonian gravity, see Marion and Thornton chapter 5.

4.1.3 Spring

You're probably seen Hooke's law for the force exerted by a stretched or compressed spring:

$$\vec{F} = -k\vec{x}, \tag{4.5}$$

where $\vec{x}$ is the displacement of the end of the spring from equilibrium and k is known as the spring constant, which effectively characterizes the stiffness of the spring.

Hooke's law works for all kinds of stretchy or spring situations not just things you usually think of as springs. It is also a really important case, as we'll see later in the course, since may other types of forces can be modeled as a spring force near an equilibrium (place where the force is zero). This is the first of the many cases we'll see where what we really have is a manifestation of electromagnetic forces at the atomic and molecular scale inside the material.

4.1.4 Normal Force

Here the use of the word "normal" means perpendicular. Hence a normal force is a force perpendicular to a surface. These forces can be understood as E&M at the atomic level if we consider the structure of the material. Here we'll consider hard surfaces of objects where the normal force provides whatever force is needed to prevent another object from penetrating the surface.

Often this will come up in the context of a plane like a wall or a table that remains at rest. When a particle pushes against the plane, the plane pushes back with a force equal and opposite to the perpendicular component of the force on the plane by the particle in accord with N3. I always find it cool that normal force adjusts to provide what ever is needed. When I stand on the floor, the normal force on me is equal in magnitude to my gravitational force. When I hold my coffee cup, the normal force on me increases.

PCQ 4.4

Does a bathroom scale measure gravitational force or normal force?

Normal force often acts as a "constraint force". Constraint forces place constraints on the position of a particle. For example, the Normal force provided by the plane of my office floor, which I'll put at $z = 0$ constrains the location of my chair to the region $z > 0$. When we do Lagrangian mechanics, one of the key advantages will be that we don't have to consider these forces unless we're interested in them.

4.1.5 Tension

Like normal force, tension is a constraint force. It provides whatever force is needed to keep the objects attached to the ends of a string, chain, or other object from moving apart. Normal force is a way to model objects that can't be penetrated by other objects, while tension models objects that can't be stretched or torn apart. At the atomic level, it's E&M too.

PCQ 4.5

A massless rope has objects attached to each end. The forces provided by these objects are the only forces acting on the rope. Use Newton's laws to argue that the forces the rope exerts on each object must be equal independent of whether or not the system accelerates.

4.1.6 Buoyancy

Buoyant forces are those that cause things to float. They are responsible for ships floating on water and for helium balloons floating in air. It turns out that the buoyant force on an object is equal in magnitude to the gravitational force that the fluid it pushed out of the way would have experienced.

There is a very cleaver argument that shows this. Consider some fluid, like a bucket of water at rest on a table. Enclose a region of the water with an imaginary membrane (draw it now). The water in the region you've divided off does not accelerate. It just sits there at rest in the bucket. Thus the net force on it must be zero. Hence the buoyant force provided by the neighboring water pushing upward on the water inside of your membrane must exactly cancel the downward gravitational force on the water inside of your membrane (add free-body diagram arrows to your bucket drawing). Now imagine replacing the water inside of your membrane with a rock of the same shape. The neighboring water still pushes in just the same way. Hence the buoyant force on your rock must be the same as the buoyant force that acted on the water it pushed out of the way. Hence the buoyant force experienced by the rock is equal to the weight of the water displaced. (Did you follow that? If not, you might try reading it again, thinking carefully about each step.)

PCQ 4.6

A rocket ship in deep space is filled with air similar to Earth's atmosphere (so the people aboard can breath comfortably). They are celebrating a birthday party and have a helium balloon.
(a) The ship is initially drifting with no acceleration. In this configuration, what is the balloon doing inside the ship? (Rising to the ceiling? etc.)
(b) The ship's captain then fires the rockets so that the ship accelerates with acceleration g and the ship's inhabitants can comfortably dance on the floor as a part of their party. When the rockets are fired, does the helium balloon go slamming to the floor, rise to the ceiling, or is it unaffected? Try to support your answers with an argument along the same line the one above.

The argument above the PCQ provides the answer for what buoyant force does, but it doesn't really explain how the buoyant force works. For this, we need the idea that the pressure (force per area) exerted by the water increases with depth. Thus no matter the shape, the water pushes upward harder on the lower parts of a submerged object than it pushes downward on the upper parts. This results in a net upward force. We'll play with this idea more in class.

For more on buoyant force, see any intro text. Giancoli on our course shelf has nice discussions.

4.1.7 Friction

Friction is an interaction between surfaces parallel to the surfaces (as apposed to normal force which is perpendicular). It too is E&M at the fundamental level owing to E&M interactions between the atoms and molecules of the surfaces. Friction is a complex phenomenon not completely understood. In fact, a faculty member at St. Olaf has a research program centered on studying friction at the microscopic level.

A model that provides a first approximation to friction is

$$F_{k,s} = \mu_{k,s} F_N. \tag{4.6}$$

Here the notation k, s means there's really 2 equations; one if you chose k and another if you choose s. The parameter μ_k is the coefficient of kinetic friction. It characterizes how "sticky" the surfaces are for cases where the surfaces are moving with respect to one another. Hence F_k is the force of kinetic friction. The parameter μ_s is the coefficient of static friction, which is typically slightly greater than μ_k. It characterizes how "sticky" the surfaces are for cases where the surfaces are at rest with respect to one another. Hence F_s is the maximum possible force of static friction. Note that I must say "maximum possible" since this is actually the maximum force friction can provide before the surfaces begin to slip. Note also that Eq. (4.6) is a relation among scalars and not a vector equation.

This model of friction is just that – a model. We don't really have a fundamental understanding of why it's this way. It just seems to describe observations in a way that's about right but not exactly. You should be able to convince yourself that this is a rather different animal than Newton's law of universal gravitation. The Feynman lectures [4] provide a nice discussion of the problems with friction that I've included below.

That the formula $F = \mu N$ is approximately correct can be demonstrated by a simple experiment. We set up a plane, inclined at a small angle θ, and place a block of weight W on the plane. We then tilt the plane at a steeper angle, until the block just begins to slide from its own weight. The component of the weight downward along the plane is $W \sin \theta$, and this must equal the frictional force F when the block is sliding uniformly. The component of the weight normal to the plane is $W \cos \theta$, and this is the normal force N. With these values, the formula becomes $W \sin \theta = \mu W \cos \theta$, from which we get $\mu = \sin \theta / \cos \theta = \tan \theta$. If this law were exactly true, an object would start to slide at some definite inclination. If the same block is loaded by putting extra weight on it, then, although W is increased, all the forces in the formula are increased in the same proportion, and W cancels out. If μ stays constant, the loaded block will slide again at the same slope. When the angle θ is determined by trial with the original weight, it is found that with the greater weight the block will slide at about the same angle. This will be true even when one weight is many times as great as the other, and so we conclude that the coefficient of friction is independent of the weight.

In performing this experiment it is noticeable that when the plane is tilted at about the correct angle θ, the block does not slide steadily but in a halting fashion. At one place it may stop, at another it may move with acceleration. This behavior indicates that the coefficient of friction is only roughly a constant, and varies from place to place along the plane. The same erratic behavior is observed whether the block is loaded or not. Such variations are caused by different degrees of smoothness or hardness of the plane, and perhaps dirt, oxides, or other foreign matter. The tables that list purported values of μ for "steel on steel," "copper on copper," and the like, are all false, because they ignore the factors mentioned above, which really determine μ. The friction is never due to "copper on copper," etc., but to the impurities clinging to the copper.

In experiments of the type described above, the friction is nearly independent of the velocity. Many people believe that the friction to be overcome to get something started (static friction) exceeds the force required to keep it sliding (sliding friction), but with dry metals it is very hard to show any difference. The opinion probably arises from experiences where small bits of oil or lubricant are present, or where blocks, for example, are supported by springs or other flexible supports so that they appear to bind.

It is quite difficult to do accurate quantitative experiments in friction, and the laws of friction are still not analyzed very well, in spite of the enormous engineering value of an accurate analysis. Although the law $F = \mu N$ is fairly accurate once the surfaces are standardized, the reason for this form of the law is not really understood. To show that the coefficient μ is nearly independent of velocity requires some delicate experimentation, because the apparent friction is much reduced if the lower surface vibrates very fast. When the experiment is done at very high speed, care must be taken that the objects do not vibrate relative to one another, since apparent decreases of the friction at high speed are often due to vibrations. At any rate, this friction law is another of those semiempirical laws that are not thoroughly understood, and in view of all the work that has been done it is surprising that more understanding of this phenomenon has not come about. At the present time, in fact, it is impossible even to estimate the coefficient of friction between two substances.

It was pointed out above that attempts to measure μ by sliding pure substances such as copper on copper will lead to spurious results, because the surfaces in contact are not pure copper, but are mixtures of oxides and other impurities. If we try to get absolutely pure copper, if we clean and polish the surfaces, outgas the materials in a vacuum, and take every conceivable precaution, we still do not get μ. For if we tilt the apparatus even to a vertical position, the slider will not fall off—the two pieces of copper stick together! The coefficient μ, which is ordinarily less than unity for reasonably hard surfaces, becomes several times unity! The reason for this unexpected behavior is that when the atoms in contact are all of the same kind, there is no way for the atoms to "know" that they are in different pieces of copper. When there are other atoms, in the oxides and greases and more complicated thin surface layers of contaminants in between, the atoms "know" when they are not on the same part. When we consider that it is forces between atoms that hold the copper together as a solid, it should become clear that it is impossible to get the right coefficient of friction for pure metals.

The same phenomenon can be observed in a simple home-made experiment with a flat glass plate and a glass tumbler. If the tumbler is placed on the plate and pulled along with a loop of string, it slides fairly well and one can feel the coefficient of friction; it is a little irregular, but it is a coefficient. If we now wet the glass plate and the bottom of the tumbler and pull again, we find that it binds, and if we look closely we shall find scratches, because the water is able to lift the grease and the other contaminants off the surface, and then we really have a glass-to-glass contact; this contact is so good that it holds tight and resists separation so much that the glass is torn apart; that is, it makes scratches.

PCQ 4.7

Computational Corner Do Mathematica tutorial "02Mathematica.nb". Respond with one of the following: a question about the tutorial, something new that you learned, or a connection with something you already knew.

4.2 Drag

In your intro physics course, you probably solved many projectile problems and always ignored aerodynamic drag. Sometimes this is justified, sometimes not. Here we'll see how to account for drag. When people say they are talking about drag, they usually mean a resistive force due to motion through a fluid (liquid or gas) that is approximately opposite the direction of motion. The fluid many exert such forces in other directions as in the lift on an airplane wing, but we won't deal with that here. The issue of drag is quite relevant for our siege, as we'll want to include its effects on the balls we fling from our trebuchet toward the castle wall. From a more modern standpoint, reducing drag on cars is a really big deal. The teardrop shape of the Prius is one of the reasons it gets such good mileage.

Our goal will be to find an expression for the drag force that we can stick into Newton's laws, get an equation of motion, and get a more realistic solution for projectile motion. We'll consider objects moving slow compared to the speed of sound. Under these conditions, it turns out that the drag force takes the form

$$\vec{F}_D = (av + bv^2)(-\hat{v}). \tag{4.7}$$

We might have been able to guess this form for two reasons. First, we know from experience that the drag force increases with velocity (stick your hand out the window of a car and you'll feel this). Second, we are seeking the answer for speeds small compared to the speed of sound, and this looks like a Taylor expansion at small v.

4.2.1 Viscosity

The first term in Eq. (4.7) arises physically from viscosity. Viscosity can be thought of as an internal friction in fluids. Among other things, it can make them pour slowly. Honey has a higher viscosity than water. In the present context fluid is dragged along with the object as it moves, and layers of fluid near the object slide over one another creating the force. The viscosity of a given fluid is characterized by a coefficient of viscosity η having units kg/m·s.

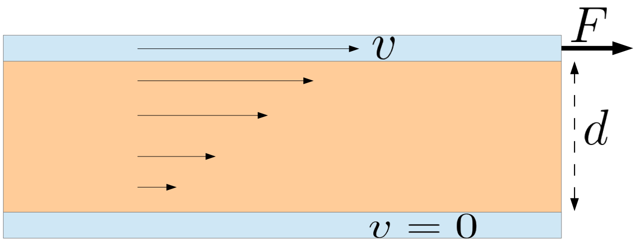


Figure 4.1: simple viscosity example.

Though the following example is not directly relevant for our interest in drag on projectiles, it helps motivate the units for η and gives some intuition for the concept of viscosity. Imagine a pair of horizontal flat plates with a fluid between them such that the upper plate floats on the fluid as shown in fig. 4.1. Now imagine pulling the top plate off to the right at constant speed by applying a force F as shown. As you pull the plate, it's the viscous force you're countering. The force is given by the following equation

$$F = \eta \frac{vA}{d}, \tag{4.8}$$

where d is the distance between the plates, v is the speed at which you pull the plate, and A is the area of the plates. This result should make some sense. If the plates had a bigger area, there would be more fluid rubbing together and the force needed to move the plates would be greater. If the plates were further apart (larger d) there would be a greater "buffer" between the plates and it would be easier to slide the top plate. If you slide slower, their is less stress on the fluid since the difference in the velocity between the top plate and the bottom plate is not so great, hence the velocity dependence. You might say, but how do you know it goes like v and not v^3? One perspective is to think of Eq. (4.8) as a hypothesis validated by experiment. Another approach would be to used our knowledge of intermolecular forces to try and explain Eq. (4.8) from the fundamental level.

4.2.2 Pressure Force

While the first term in Eq. (4.7) is do to viscous forces, the second is due to what is often called a pressure force. This force arises do to pressure differences between the front and back of a projectile as it boars its way through the fluid. This force depends on the size scale of the disturbance in the fluid d. For example, if you have a bigger crossectional area perpendicular to the direction of motion, you create a bigger disturbance in the fluid and experience more drag. That's why sky divers fall more slowly when they lay horizontally than when they make their bodies vertical, and fall slower still once they open the parachute. It also depends on the density ρ of the stuff being disturbed. It takes a bigger force to disturb more massive stuff. This force also goes like v^2, hence identification with the second term in Eq. (4.7). Real objects feel both the pressure force and the viscous force as they move through a fluid; however, in many cases one of the forces is much more relevant that the other, as we'll see.

4.2.3 A Neater Way

We've now seen that the drag force depends on 4 things: the viscosity of the fluid η, the size scale of the disturbance d, the velocity of the object v, and the density of the fluid ρ. It also depends on the shape of the object. Here we'll consider a sphere (the one we'll shoot at the castle), in which case d can be identified with the diameter.

Our situation with drag force in the form of Eq. (4.7) has a very annoying feature. The constants that appear there are some function of η, ρ, and d. To find the functional form of those constants, we would need to do a number of experiments. Perhaps first holding v, η, and ρ constant while varying d and measuring F, then holding v, η, and d constant while varying ρ etc. Fortunately our tool of dimensionless variables gives us a better way forward. This takes a few steps to set up, so bear with me as we put them in place. Try to focus on understanding each piece as you read through the first time. Then read through it again to make sure you understand the whole picture.

Stating our situation in equations, we could say that the drag force is some function of η, ρ, d and v. That is

$$\vec{F}_D = f[\eta, \rho, d, v](-\hat{v}). \tag{4.9}$$

There is a theorem, Buckingham's Pi Theorem [5], which states that if you have n physical quantities constructed from r basic units there will be exactly $n - r$ dimensionless combinations of the n quantities. Here we have $n = 4$ physical quantities appearing in our function f in Eq. (4.9). They are constructed from $r = 3$ basic units: length, time, and mass. Thus there is exactly one dimensionless combination of them which is given the name Reynolds Number:

$$R_e = \frac{\rho d v}{\eta}. \tag{4.10}$$

One can show that a characteristic force can be found and takes the form

$$F_c = \rho v^2 d^2. \tag{4.11}$$

We could then rewrite Eq. (4.9) in terms of the dimensionless force

$$\frac{\vec{F}_D}{\rho v^2 d^2} = g[\eta, \rho, d, v](-\hat{v}). \tag{4.12}$$

(May the dimensionless force be with you!) One can view Eq. (4.12) as arising by dividing both sides of Eq. (4.9) by F_c. That means the right hand side will also change so I give the function a new name g. We now have something dimensionless on the left hand side of Eq. (4.12). Thus the right hand side must also be dimensionless. This means that η, ρ, d and v can only appear on the right-hand side in the combination R_e. So we many rewrite Eq. (4.12) in the form

$$\frac{\vec{F}_D}{\rho v^2 d^2} = g\left[\frac{\rho d v}{\eta}\right](-\hat{v}) = g[R_e](-\hat{v}). \tag{4.13}$$

Solving for the original variable of interest, the drag force, we have

$$\vec{F}_D = g[R_e]\rho v^2 d^2(-\hat{v}), \tag{4.14}$$

where our function g is known as the drag coefficient. While use of Eq. (4.14) as written would be perfectly fine, many people choose to reshuffle things as follows by pulling a $\pi/8$ out of the function g. This means that the function changes slightly, hence I give it a new name C_D:

$$\vec{F}_D = \frac{\pi}{8} C_D[R_e]\rho v^2 d^2(-\hat{v}). \tag{4.15}$$

This can then be rearranged to the form

$$\vec{F}_D = \tfrac{1}{2} C_D[R_e]\rho v^2 A(-\hat{v}), \tag{4.16}$$

where A is the crossectional area of the sphere. A lot of people like to write it this way since the $\frac{1}{2}\rho v^2$ is kinetic energy per volume. Equation (4.16) contains the same information as Eq. (4.7), but the form has several advantages as we'll now demonstrate.

The advantages of (4.16) (or equivalently (4.14)) all stem from the fact that we now have *one unknown function* C_D *of one variable* R_e. Thus we only need to do one experiment to determine C_D: vary R_e and measure F_D. We can do this by varying any of the original 4 physical quantities η, ρ, d and v while holding the others constant. Once the experiment is done, we'll have a completely known function for the drag force on a sphere of any size. The results of the experiment are shown below, presented as C_D vs. R_e. (If, for example, you measure F_D and vary d, then you can just do a little algebra to preset it as C_D vs. R_e instead.)

Note that log/log graph paper is used in Fig. 4.2, and the function looks linear for small Reynolds number on this paper. This region can be fit using

$$\log C_D = \log\alpha - \log R_e, \quad R_e < 1, \tag{4.17}$$

where one finds $\alpha = 24$. Manipulation of this result yields

$$C_D = \frac{24}{R_e}, \quad R_e < 1. \tag{4.18}$$

Thus for small R_e we find that the drag force takes the form

$$\vec{F}_D \approx 3\pi\eta d v(-\hat{v}), \quad R_e < 1, \tag{4.19}$$

by plugging Eq. (4.18) into Eq. (4.16). This is known as Stokes's Law, and it is the viscous force which dominates at low R_e. Note that this is the piece that corresponds to the linear term in Eq. (4.7).

PCQ 4.8

Find Eq. (4.19) from Eq. (4.18) and Eq. (4.16).

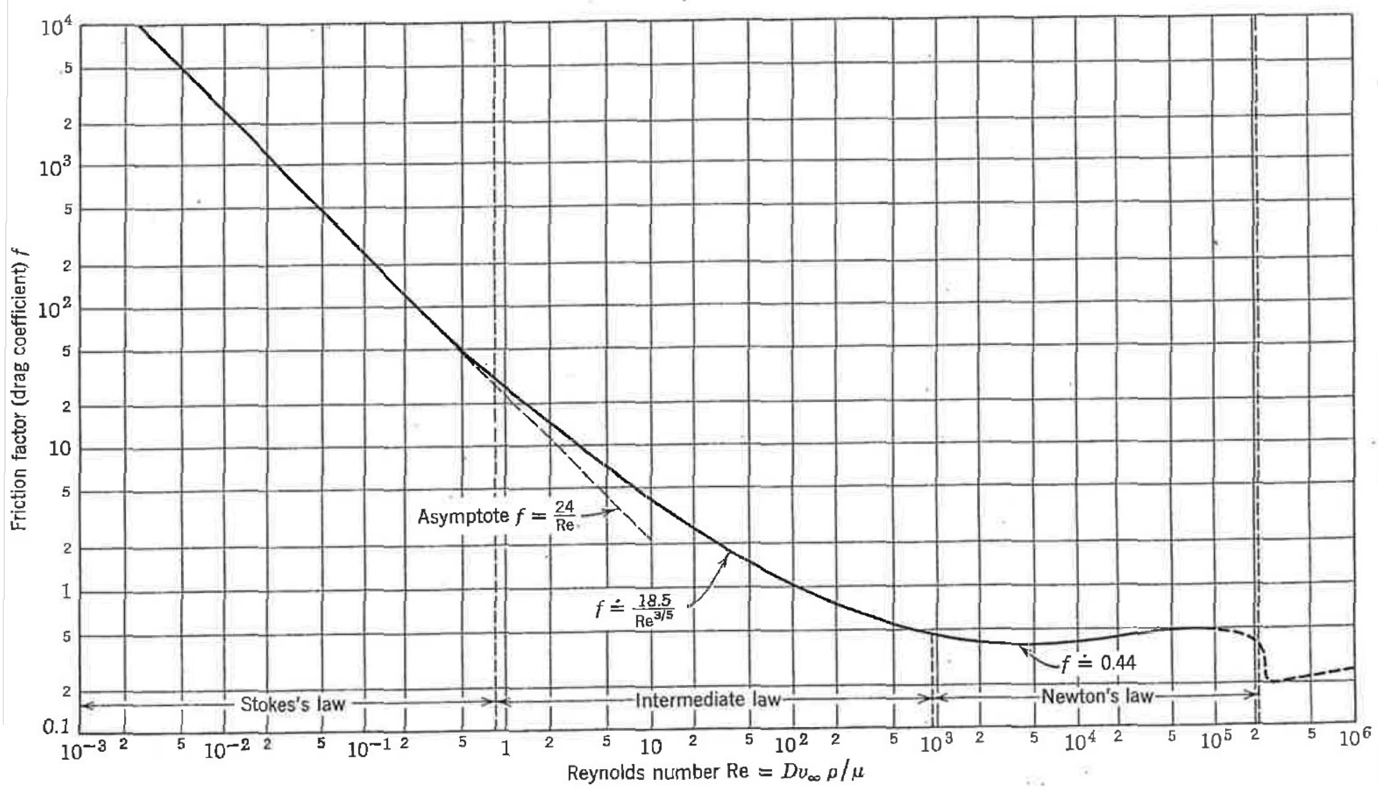


Fig. 6.3–1. Friction factor (or drag coefficient) for spheres moving relative to a fluid with a velocity v_∞. See definition of f in Eq. 6.1–5. [Curve taken from C. E. Lapple, "Dust and Mist Collection," in *Chemical Engineers' Handbook* (ed. by J. H. Perry), McGraw-Hill, New York (1950), Third Edition, p. 1018.]

Figure 4.2: C_D vs. R_e for a sphere. Note that that $f = C_D$ in the notation of this figure [6].

PCQ 4.9

Describe how to see that Eq. (4.17) is a line on a log plot.

At Reynolds numbers in the range $10^3 < R_e < 2 \times 10^5$, the drag coefficient is nearly constant:

$$C_D = 0.44, \quad 10^3 < R_e < 2 \times 10^5. \tag{4.20}$$

Plugging Eq. (4.20) into Eq. (4.16) yields

$$\vec{F}_D \approx 0.22 \rho A v^2 (-\hat{v}), \quad 10^3 < R_e < 2 \times 10^5. \tag{4.21}$$

This is known as Newton's law. Here the pressure force dominates and the result corresponds to the second term in Eq. (4.7). A popular fit to the whole region $0 < R_e < 2 \times 10^5$ is

$$C_D = \frac{24}{R_e} + \frac{6.0}{1 + \sqrt{R_e}} + 0.40. \tag{4.22}$$

So I've spent some time developing the theory of drag. One could ask many more questions here, but answering many of them would take us too far into fluid dynamics than our current priorities allow. We'll mostly want to proceed to investigate the behavior of projectiles and other bodies subject to these forces since that's the mechanics part. Still, the physics here is really fascinating, so I'll add just a few more conceptual comments. First, a part of the power of drag formulated in terms of Reynolds number is that it allows experiments on models since the same Reynolds number can be achieved by making d small while compensating with the other parameters η, ρ, and v. Thus you may be able to get good idea of the aerodynamic forces on a 787 dreamliner by doing experiments on a small model. In fact, all fluid flows at a given Reynolds number are geometrically similar. You also might ask why the curve in Fig. 4.2 has this strange shape. The answer has to do with the strange way in which fluid flows around an object at various Reynolds numbers. Figure 4.3 provides another plot, this time for a cylinder. It also provides some cartoons of the fluid flow responsible. Note that there is a sharp drop in C_D as a function of Reynolds number on both plots beyond the region of constant C_D. The flow shown in the final cartoon of Fig. 4.3 is responsible for this behavior. This is known as "the drag crisis" and it is unusual because the drag force may actually reduce here as speed increases. You may have noticed that golf balls have little dimples on them. Counter intuitively, dimpled balls actually fly further than smooth ones. While the dimples increase the viscous force, the pressure force is dominant at Reynolds numbers relevant for golf balls. The disturbance created by the dimples induces the drag crisis and reduces the overall drag force notably.

Our primary interest will be in modeling motion of particles in the presence of drag forces. We'll work out some examples in class, and a number of interesting cases are worked out in Taylor Chapter 2 [7].

PCQ 4.10

No doubt, you have all heard of energy. Try to define energy in your own words in preparation for the next section.

4.3 Energy

4.3.1 Foundations of Energy

One could imagine getting into energy in the following way. One could imagine that over time folks would have noticed that there was a set of quantities in nature having the same units whose sum never seemed to change. This set of quantities is collectively called energy. To me, a sum of really weird things that nature feels it must preserve would cry out for an explanation. Emmy Noether found such an explanation in 1915 in showing that conserved quantities are always due to symmetries in nature. She achieved this impressive result in spite of great persecution for her gender. In the case of energy conservation, the relevant symmetry is known as time-translation invariance – the fact that the laws of nature don't change with time. Hence the best definition of energy that I can find is "energy is the quantity that is conserved when there is time translation symmetry in the system."

This statement is most easily unpacked in the context of an example. Consider once again a particle dropped from rest under gravity in the laboratory such that the system can be modeled using the following equation of motion

$$m\ddot{z} = -mg. \tag{4.23}$$

If we change the time coordinate in the following way

$$t \to t + b, \tag{4.24}$$

where b is some constant (that is, do a time translation), then there will be no change to the equation of motion and hence no change to our physical description of the system. Intuitively, our description of this system is the same no matter when we perform the experiment (now, next week, etc.). So the system is said to have "time translation symmetry".

Under these conditions, the equation of motion can be integrated once to find the conserved quantity. Here, we'll do this as follows. Multiply both sides of Eq. (4.23) by $\dot{z}$, such that it reads

$$m\dot{z}\ddot{z} = -mg\dot{z}z. \tag{4.25}$$

This just makes the integration easier. Now we can be rewritten as

$$0 = \tfrac{d}{dt}(\tfrac{1}{2}m\dot{z}^2 + mgz). \tag{4.26}$$

This is probably not too obvious, but if you act with the derivative on the thing in parenthesis, you'll get back the thing in Eq. (4.25).

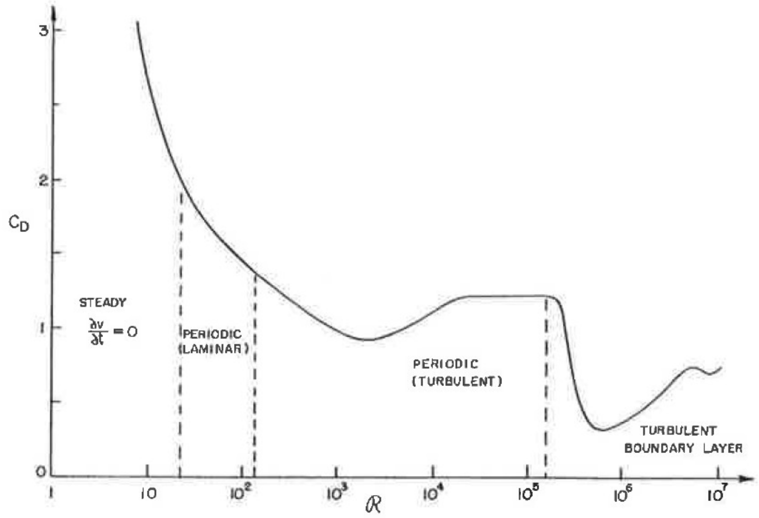


Fig. 41–4. The drag coefficient C_D of a circular cylinder as a function of the Reynolds number.

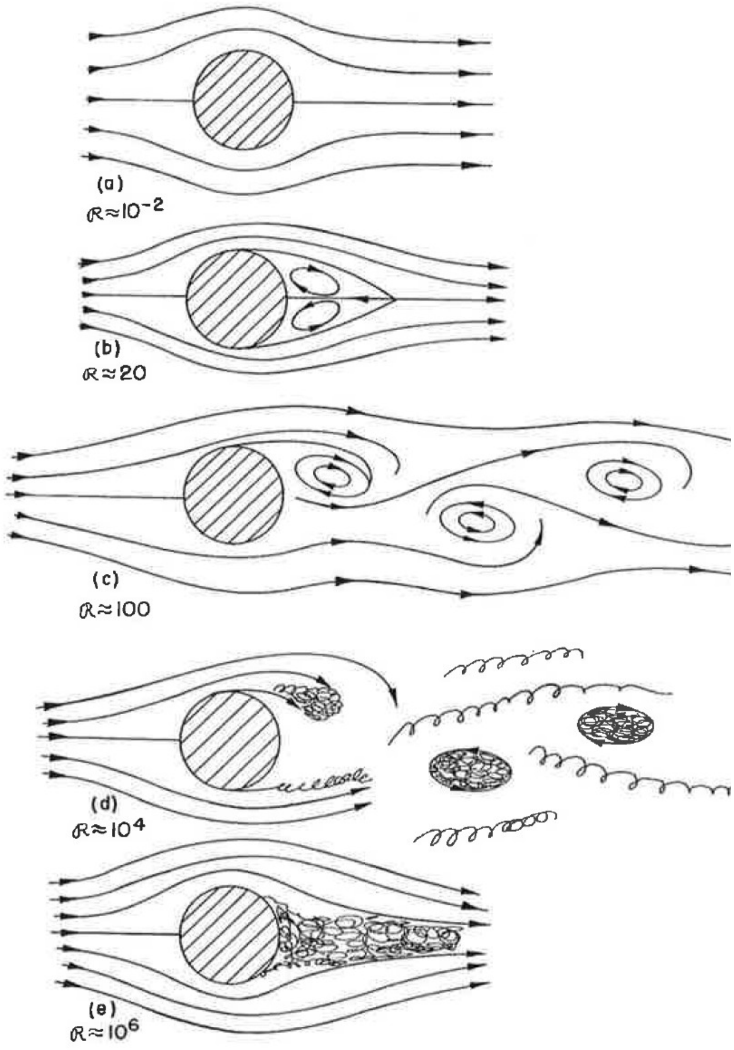


Fig. 41–6. Flow past a cylinder for various Reynolds numbers.

Figure 4.3: C_D vs. R_e for a cylinder, along with some flow patterns. Image from Ref. [4]

PCQ 4.11
Check that Eq. (4.26) is equivalent to Eq. (4.25).

Now, if the time derivative of the stuff in parenthesis is zero, then that material must be equal to a constant. I will name that constant E so that I can write

$$E = \tfrac{1}{2}m\dot{z}^2 + mgz. \tag{4.27}$$

Note that we now have the sum of two things equal to a constant. In other words, the sum on the right is conserved. So this is an example of a system in which time translation invariance exists and there is a conserved quantity as a result that we call energy. You'll no-doubt recognize the two parts as the things you would have called kinetic and gravitational potential energy in intro physics.

In some sense, the simplest kind of energy is kinetic energy defined (nonrelativistically) as

$$T = \tfrac{1}{2}mv^2. \tag{4.28}$$

Note that I'm using T here for kinetic energy. This, unfortunately, is very standard notation and I want to use it even though it conflicts with our convention of using capital letters for dimensionless quantities. If in a give case confusion could arise between kinetic energy and dimensionless time, we'll define a new symbol for dimensionless time, like maybe τ. In what follows, we'll want to consider 2 other broad classes of energy: potential energy and thermal energy. Kinetic energy along with potential energy together comprise something known as mechanical energy. Later in this section, I'll have much more to say about each type.

But first, let's consider a case where there is no such quantity as a conserved energy. If Newton's constant G were a function of time $G[t]$ rather than being a constant, then there would not be time translation invariance. The outcome of the experiment would depend on when we did it (because G and hence forth g are changing). Things would be different if we do the experiment now vs. next week. Mathematically, the equation of motion would change if we did our time translation in Eq. (4.24). If we tried to integrate the equation of motion once as we did in going from Eq. (4.26) to Eq. (4.27), the time dependence in g would prevent us from finding a constant. So without time translation invariance, there would be no conserved energy. The connection between symmetries and conserved quantities is something we'll explore later in the course after we develop Lagrangian mechanics.

4.3.2 The general case

In the example above, I considered gravity, but one could do a similar calculation for a generic force $\vec{F}$. This time, dot a $\dot{\vec{r}}$ into both sides as follows

$$m\dot{\vec{r}}\cdot\ddot{\vec{r}} = \dot{\vec{r}}\cdot\vec{F}. \tag{4.29}$$

Rearrange as follows

$$\tfrac{d}{dt}(\tfrac{1}{2}mv^2) = \vec{F}\cdot\tfrac{d\vec{r}}{dt}. \tag{4.30}$$

PCQ 4.12

Start with Eq. (4.30) and see if you can show that it is the same as Eq. (4.29) as I claim.

Notice that the we have the kinetic energy appearing in parenthesis. Now integrate both sides of Eq. (4.30) over time. On the left I'll get $T[t] - T[t_0]$ by the fundamental theorem of calculus. So I can write

$$\Delta T = \int \vec{F}\cdot\tfrac{d\vec{r}}{dt}dt, \tag{4.31}$$

or using the un-chain-rule on the right,

$$\Delta T = \int \vec{F}\cdot d\vec{r}. \tag{4.32}$$

You'll recognize the right as the integral form of work. Hence this is the work-energy theorem. If the integral on the right has special properties, which we'll discuss below, then it can be turned into a potential energy as we did with gravity above, and we'll have a conserved energy.

As an aside, note that work is the transfer of energy via a force. There is another way to transfer energy known as heat that you'll meet in thermodynamics.

4.3.3 Conservative Forces and Potential Energy

In this section, we'll define conservative forces and identify many types of potential energy. Conservative forces are defined as forces for which the following two conditions hold: (1) The force depends at most on position $\vec{r}$ and not on any other variables like time, velocity, etc. (2) The work done by the force on a particle as the particle travels between two points a and b does not depend on the choice of path. When conservative forces act, energy does not enter or leave the system as described by classical mechanics, hence the name conservative. Note however that all forces seem to obey conservation of energy, but those that are not conservative may transform energy to forms that are not addressed by classical mechanics of macroscopic objects as we'll see below.

For conservative forces, one can define a potential energy function as follows:

$$U = -w = -\int \vec{F} \cdot d\vec{r}, \tag{4.33}$$

where the integral is carried out along a path from the arbitrarily defined zero of potential energy to the location of interest. Note that this potential energy is well defined for conservative forces, but may not be for nonconservative forces. For example, if the work done depends on the path taken, there is no unique potential energy function corresponding to the given force. Similarly, if the force is a function of time, velocity, etc. it may depend on when or how the particle traveled along the path. Note also that the minus signs make sense in the context of the work energy theorem. When a force does work on a particle, potential energy is lost, while kinetic energy is gained.

Newtonian gravity is a nice example of a conservative force. From it's definition, it depends only on position. Also, in the absence of friction, one can take a slide of any shape down a hill and be going just as fast at the bottom. Hence the work done by gravity is path independent. Friction, on the other hand, is a good example of a nonconservative force. The work done by friction is path dependent. You'll expend much less energy if you push a box straight across a room from point a to point b over taking some longer path.

Of the forces we met at the beginning of the chapter, electric, gravitational, and spring are well known as being conservative forces with potential energies usually written as follows:

Type	Equation
Gravitational	$U_g = \frac{-Gm_1m_2}{r}$
Gravitational near surface	$U_g \approx mgz$
Electric	$U_e = \frac{kq_1q_2}{r}$
Spring	$U_s = \frac{1}{2}kx^2$

Note, however, that these definitions can be altered by the addition of a constant. Friction and aerodynamic drag are good examples of nonconservative forces.

PCQ 4.13

Is buoyancy a conservative force?

It's often the case that one wants to get the force from the potential. In other words, undo Eq. (4.33). To see how to do this, consider an infinitesimal change in potential energy

$$dU[\vec{r}] = \frac{\partial U}{\partial x}dx + \frac{\partial U}{\partial y}dy + \frac{\partial U}{\partial z}dz. \tag{4.34}$$

The right hand side is the chain rule. If U were a function of x only, then $dU = \frac{dU}{dx}dx$ as an application of the ordinary old chain rule of calc 1. (If that still bothers you, consider playing around with differentiating a specific function like $U = cx$.) In the present context, the function U in general depends on x, y, z. Thus an arbitrary change in the function could come from a change in any of x, y, z. For those that have not had multi variable, the symbol $\partial/\partial x$ is a partial derivative. For our present purposes, it means differentiate with respect to x treating the other variables as constants.

Before I arrive at my grand result, I need one more piece of notation. The gradient operator is defined as follows by its action on some function. So here f is just a generic function name:

$$\vec{\nabla} f = \frac{\partial f}{\partial x}\hat{x} + \frac{\partial f}{\partial y}\hat{y} + \frac{\partial f}{\partial z}\hat{z}. \tag{4.35}$$

In other words, when the gradient operator acts on a scalar function f it generates a vector whose components are the derivative of the function f along each direction. The operation is called "the gradient". The symbol ∇ is called "the Del operator" or "Nabla" (`\nabla` in LaTeX). Thus I can rewrite Eq. (4.34) as

$$dU[\vec{r}] = \vec{\nabla}U \cdot d\vec{r}. \tag{4.36}$$

Using the first equality in Eq. (4.33) and the definition of work, we can also write

$$dU = -\vec{F} \cdot d\vec{r}. \tag{4.37}$$

Comparing Eqs. (4.36) and (4.37) one can identify

$$\vec{F} = -\vec{\nabla}U. \tag{4.38}$$

PCQ 4.14

Check that Eq. (4.38) works for the gravitational force and the gravitational potential energy in the "near surface" approximation.

As a final note for those that know vector calculus, it can be shown that requiring the curl of the force to be zero

$$\vec{\nabla} \times \vec{F} = 0 \tag{4.39}$$

is equivalent to the requirement of path independence in identifying conservative forces.

4.3.4 Thermal Energy

So when forces like friction act, mechanical energy is not conserved. However, total energy is always conserved (as far as we know). So where does it go? The answer is usually thermal energy. Thermal energy is really just kinetic and potential energy on the microscopic scale. That is, when the thermal energy is higher, atoms and molecules are bouncing around more in a material and it feels warmer. Hence when it's -20^o F, you rub your hands together to warm them since the mechanical energy lost by your hands through the frictional force becomes thermal energy of your hands.

There is a fact about thermal energy that is strange at first. Once energy becomes thermal energy, it's hard to get it back as mechanical energy. When a box slides down a ramp, it is never observed to cool itself and use the energy to go back up, even though that's allowed by conservation of energy. One can convert *some* thermal energy to mechanical energy, that's how a steam engine works. You build a fire and use the thermal energy to give a train kinetic energy, but it turns out that you can never convert all of the thermal energy generated by the fire to kinetic energy of the train. It is not just that we're not good at building the perfect engine. One can prove theoretically that is not possible, no matter how good the engineers get at building engines. This fact is known as the Second Law of Thermodynamics.

You'll learn a lot about the Second Law if you take the Thermodynamics course, but in very simple terms it works as follows. When a box slides down a ramp, 10^{23} particles are all doing something together in a very coordinated way – sliding down the ramp together. When that energy gets converted to thermal energy, it becomes energy shared among the 10^{23} particles making up the box and the ramp in a very random way as they all bounce around in the material. While it is statistically possible that at any given moment, the 10^{23} randomly bouncing particles might happen to all bounce in such a way that they propel the box back up the ramp, this is very very ... very very = very^{30} unlikely. It's so unlikely that it is never observed to happen. So it's the spreading of energy into a large number of random configurations that makes thermal energy hard to get back as mechanical energy. The measure of this distribution of energy across configurations is known as entropy, and as mechanical energy converts to thermal energy, entropy increases substantially. Thus when politicians talk about "running out of energy" they should really be talking about running out of energy in low-entropy forms. And perhaps we should have a United States Department of Entropy instead of Energy.

As a final technical note, the words heat and thermal energy are often confused. Heat is an energy transfer method analogous to work. You would never say that a speeding bullet has a lot of work. You would say it has a lot of kinetic energy. Similarly, you should not say that a red hot piece of metal has a lot of heat, you should say that it has a lot of thermal energy.

For alternative discussions of energy and more practice problems see Taylor chapter 4 or an introductory physics text.

Bibliography

[1] For some discussion of the Weak Equivalence Principle, see, A. Nobili *et al.*, Am. J. Phys. **83**, 39 (2015). `https://doi.org/10.1119/1.4798583`

[2] For initial results of the MICROSCOPE mission, see, P. Touboul *et al.*, Phys. Rev. Lett. **119**, 231101 (2017). `https://doi.org/10.1103/PhysRevLett.119.231101`

[3] H. Pihan-le Bars . . . G. Mo, J. Tasson *et al.*, Phys. Rev. Lett. **123**, 231102 (2019). `https://doi.org/10.1103/PhysRevLett.123.231102` or for the arXiv version `http://arxiv.org/abs/arXiv:1912.03030`

[4] R. Feynman, R. Leighton, and M. Sands, *The Feynman Lectures on Physics*, (Addison-Wesley, 1963).

[5] P.K. Kundu and I.M. Cohen, *Fluid Mechanics* Fourth Ed., (Academic Press, 2010).

[6] R. Bird, W. Steward, E. Lightfoot, *Transport Phenomena*, (Wiley, 1960).

[7] J.R. Taylor, *Classical Mechanics*, (University Science Books, 2005).

Chapter 5

Numerical Methods

5.1 Numerical Methods I

Suppose one faces an initial value problem of the form

$$\begin{aligned} \frac{dv[t]}{dt} &= \tfrac{1}{m}F[x[t],v[t],t] \\ \frac{dx[t]}{dt} &= v[t] \\ v[t_0] &= v_0 \\ x[t_0] &= x_0. \end{aligned} \tag{5.1}$$

In all of the problems we've faced so far, we've found functions $x[t], v[t]$, constructed from combinations of familiar functions like polynomials, trig functions, etc., which, when plugged into the ODE yield a true statement. See, for example, Chapter 2 for the case where $F =$ constant, and the solution is the standard kinematic equations for constant acceleration from intro physics. Note that in Eq. (5.1) I've been extra liberal with showing explicit functional dependence. People typically add this in and remove it to suit the clarity of the situation. For example, it would be equivalent to simply write

$$\begin{aligned} \frac{dv}{dt} &= \tfrac{1}{m}F \\ \frac{dx}{dt} &= v \\ v[t_0] &= v_0 \\ x[t_0] &= x_0, \end{aligned} \tag{5.2}$$

where here the functional dependence is implied in the first 2 lines.

It is not always possible to find such functions that solve the differential equations. Here, we'll explore the most basic approach to solving the IVP numerically. Lets focus on the first ODE in Eq. (5.1). In it's current form, it refers to all of the functions at a single time t, and the derivatives, by definition, refer to the rate of change in the function at that time. To prepare for numerical work, we can instead approximate the derivative as a finite difference:

$$\frac{dv[t]}{dt} \approx \frac{v[t+\Delta t] - v[t]}{\Delta t}, \tag{5.3}$$

where Δt is some (probably small) increment of time. We can then write an approximate version of the 1st ODE as follows:

$$v[t+\Delta t] \approx v[t] + \tfrac{1}{m}F[x[t],v[t],t]\Delta t. \tag{5.4}$$

This is sometimes called a velocity update equation. If you know the velocity and position of a particle at some time t, it gives an approximate value for the velocity a time Δt later. We could do the same procedure with the second ODE.

PCQ 5.1

Write an equation analogous to (5.4) for the second ODE in Eq. (5.1) by filling in the blank below

$$x[t + \Delta t] \approx x[t] + \underline{\hspace{3em}}\,. \tag{5.5}$$

One can now use Eqs. (5.4) and (5.5) to do the job of our mechanics machine. Given an initial velocity and position, one can pick a small time step Δt, plug this information into the right hand side of Eqs. (5.4) and (5.5) and have the position and velocity a time Δt later. One can then plug these answers back in on the right and finding the position and velocity at time $t + 2\Delta t$. Continuing to "iterate" the process can yield an approximate value for the position and velocity at an arbitrary time.

As an example, consider the following IVP:

$$\begin{aligned} \dot{v} &= \tfrac{1}{m}F = -\tfrac{b}{m}v \\ \dot{z} &= v \\ v[0] &= v_0 = 2.0\ \tfrac{\text{m}}{\text{s}} \\ Z[0] &= z_0 = 10\ \text{m} \end{aligned} \tag{5.6}$$

The exact answer is

$$\begin{aligned} v[t] &= v_0 e^{-\frac{b}{m}t} \\ z[t] &= z_0 + \frac{v_0}{\frac{b}{m}}(1 - e^{-\frac{b}{m}t}). \end{aligned} \tag{5.7}$$

PCQ 5.2

Check that Eq. (5.7) is a solution to (5.6).

Since this forms a problem where we know the exact answer and can check approximate work against it to see how the approximation is working it forms a nice place to start. For a case where $\frac{b}{m} = 0.10\text{s}^{-1}$, and taking $\Delta t = 3.0$ s, we'll find the position and velocity at time $t = 9$ s by hand by applying Eqs. (5.4) and (5.5) to our example. I find the following:

t (s)	v (m/s)	z (m)
0.0	2.0	10
3.0	1.4	16
6.0	1.0	20
9.0	-	-

Here the first line of the table is obvious, it's just the initial conditions. The second line is obtained via

$$\begin{aligned} v_0 - \tfrac{b}{m}v_0\Delta t &= 1.4\ \text{s} \\ z_0 + v_0\Delta t &= 16\ \text{m}. \end{aligned} \tag{5.8}$$

PCQ 5.3

Fill in the last 2 lines of the table.

Figures 5.1 and 5.2 show the exact solution (smooth curve) and the points from our approximation, which have been connected by lines for velocity vs. time and position vs. time respectively.

PCQ 5.4
What do you notice about the slope of the segment connecting the first and second point in each plot? Is this expected based on the algorithm we set up?

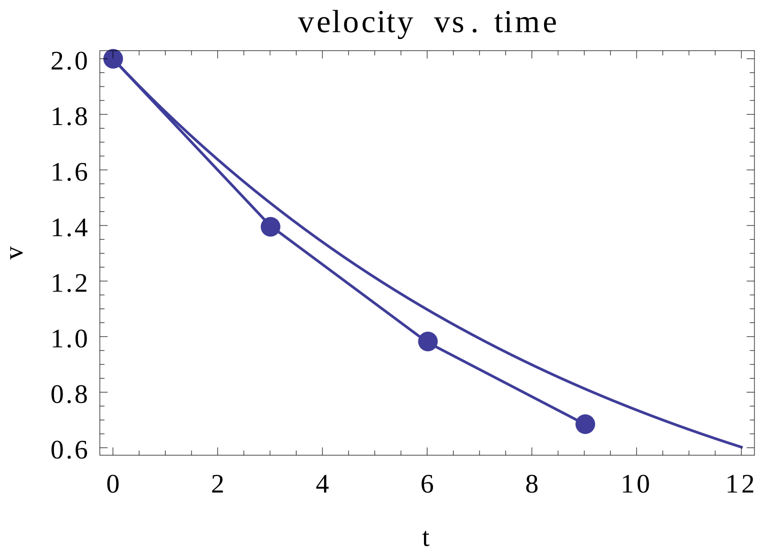


Figure 5.1: Numerical example: velocity vs. time.

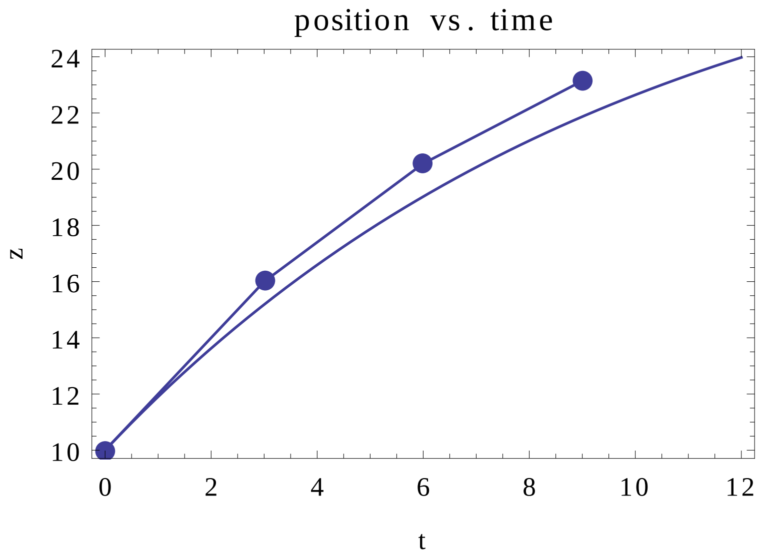


Figure 5.2: Numerical example: position vs. time.

You'll notice that there is some error here (the points don't lie exactly on the exact solution). There are 2 basic ways to improve. One is to use a smaller Δt, but as you can see, you'll get tired fast if you make Δt small. So we'll want to get a computer to do this for us. It turns out that computers are really good at repeating the same instruction over and over fast. The construct is called a loop and you'll be introduced to it in a Mathematica tutorial. Another way to improve the outcome is to improve the algorithm, or the approximation procedure we used at the start. Doing this will yield better results with the same Δt, and hence the same number of calculations. We'll do this later.

This iterative approach highlights the "local" nature of Newton's laws. In working with the Newtonian perspective, we find the force at a point, that tells us the acceleration at that point, which tells us where the particle will go next and what its new velocity will be. We then use its new location and velocity to find the new force at the new point, and the process repeats.

PCQ 5.5

Computational Corner Do Mathematica tutorial "03Mathematica.nb". Respond with one of the following: a question about the tutorial, something new that you learned, or a connection with something you already knew.

5.2 Numerical Methods II

Will there be a Numerical Methods II? We'll see.

Chapter 6

Rotation

Up to now in the course we've considered the motion of point particles, or objects that we treated as points. In this Chapter, we'll develop the technology to treat some of the motion of extended rigid bodies, like hockey sticks, iphones, and trebuchet arms. Really, there is nothing new here since these objects are really made up of a bunch of particles that we can treat as points in classical mechanics, but we don't want to treat all 10^{23} particles individually. Hence the need for the methods we'll develop here. Note however that the idea of a rigid body is really an approximation. All of the things we'll consider are flexible and stretchable at some level. A lot of what I'll have to say will be about rotation, since that's the key new part that we want to develop over point particle motion; however, I'll say a few words about center of mass first.

6.1 Center of Mass

Center of mass is a weighted average position of particles in a system, where the weights are the masses. In equations, it's defined as

$$\vec{r}_{\rm cm} = \frac{1}{m_t}\sum_n m_n \vec{r}_n, \tag{6.1}$$

where the sum over n runs over all particles in the system, $\vec{r}_n$ is the position of the $\mathrm{n^{th}}$ particle in the system, m_n is the mass of the $\mathrm{n^{th}}$ particle, and m_t is the total mass of the system.

PCQ 6.1

Show that

$$\frac{d\vec{r}_{\rm cm}}{dt} = \text{constant} \tag{6.2}$$

when the net force on the system is zero.

For continuous materials, one can write

$$\vec{r}_{\rm cm} = \frac{1}{m_t}\int \rho \vec{r} d^3 r, \tag{6.3}$$

where ρ is the density of the material and $d^3 r$ denotes a volume integral.

However, even if the net force on a group of particles is zero, its motion can still be interesting. In fact, the collection of particles can be moving and interacting in all sorts of interesting ways, it's just that the weighted average of their position above does not change. Imagine an isolated galaxy in deep space. The net force on the system may be very small, but there are still many many interesting motions within the system. The motion of any system can always be viewed as the motion of the center of mass plus the motion of the individual particles making up the system about the center of mass.

When we consider rigid bodies, we usually mean that the particles making up the body don't move with respect to each other to a good approximation. This is the case when we want to describe things like the motion of a baseball flying through the air. In this approximation, we really have only two motions. The motion of the center of mass through space and rotation about the center of mass. Breaking the motion into these two pieces will often be very useful for us.

PCQ 6.2

When you throw a shoe, it's the center of mass that follows the usual parabolic trajectory, while the rest of the shoe may rotate about that point. Explain this behavior with a sentence or two in light of the above description.

6.2 Angular Momentum

6.2.1 Point Particles

Angular momentum,

$$\vec{l} = \vec{r} \times \vec{p}, \tag{6.4}$$

is a conserved quantity in nature. As stated above, this is the angular momentum of a point particle, and indeed, angular momentum is a useful concept in certain types of motion where the point particle approximation is good. However, our immediate interest will be in describing rigid bodies and we'll work this equation into a form that is more useful for describing rigid bodies in the following sections.

Note that in spite of the fact that there is a $\vec{p}$ in here, angular momentum is a separate conserved quantity, and its conservation does not follow directly from momentum conservation. That is, one can find theories (which have not been confirmed by experiment, at least not yet) that have momentum conservation and lack angular momentum conservation.

PCQ 6.3

We've now met Noether, who found that a conservation law always comes from a symmetry in nature. See if you can guess what symmetry generates angular momentum conservation.

You might recall that the directions of vectors are weird in the description or rotation. Most of them point in directions where nothing is really happening. We can already see this in the case of angular momentum. From the right hand rule, $\vec{l}$ points perpendicular to the plane of the particle's motion.

The motion of a particle moving in a circle of fixed radius (like a ball whirled on a string) is conveniently described with angles. In a convenient coordinate system with the plane of the circle as the $x-y$ plane and the origin at the center, the location of the particle can be described with an angle ϕ around the origin increasing from the $+x$ axis. The change in angle is then related to a change in the position of the particle by

$$d\vec{r} = d\vec{\phi} \times \vec{r}. \tag{6.5}$$

Note that with this definition, the direction associated with $\vec{\phi}$ is perpendicular to the plane of motion, just like $\vec{l}$. We're sort of forced to have this weird direction for $\vec{l}$ since that's the conserved quantity, and the rest of the directions in rotational motion theory are chosen to work nicely with it. Note also that when magnitudes are considered here, one recovers the definition of an angle in radians. Radians is the unitless ratio of arc length over radius (radians is just a label to remind us that it's unitless).

If we consider the time rate of change of $\vec{r}$ and $\vec{\phi}$ we find

$$\vec{v} = \frac{d\vec{\phi}}{dt} \times \vec{r} = \vec{\omega} \times \vec{r}, \tag{6.6}$$

where $\vec{\omega}$ is the angular velocity, and $\vec{v}$ is a velocity tangent to the circle. Note that in this example, I don't really need to say this since there can be no velocity component along $\vec{r}$ for circular motion at fixed radius. Taking another time derivative yields

$$\vec{a} = \frac{d\vec{\omega}}{dt} \times \vec{r} + \vec{\omega} \times \vec{v}. \tag{6.7}$$

Note that the first term is tangent to the circle

$$\vec{a}_t = \vec{\alpha} \times \vec{r}, \tag{6.8}$$

where $\vec{\alpha}$ is known as angular acceleration. This is associated with an increase in the speed of the particle on its circular path. The second term points toward the center and has magnitude v^2/r, i.e. the centripetal acceleration. The centripetal acceleration arises due to the change in direction of $\vec{v}$ with time, while the tangential acceleration arises due to changes in the magnitude of $\vec{v}$. I suppose one could continue differentiating here and define an angular jerk, angular snap, angular pop, and angular crackle, but I'll stop here.

Using Eq. (6.6) one can rewrite the angular momentum of our particle as

$$\vec{l} = m\vec{r} \times (\vec{\omega} \times \vec{r}). \tag{6.9}$$

Note that if $\vec{r}$ and $\vec{\omega}$ are perpendicular, as they are with our simple point particle in the chosen coordinates, this reduces to the perhaps familiar

$$\vec{l} = mr^2\vec{\omega}. \tag{6.10}$$

6.2.2 Multiple Particles

You are perhaps familiar with the fact that when you have more than one particle in a system, the total momentum of the system is the vector sum of the momenta of the particles. The same is true of angular momentum. For a collection of N particles engaged in arbitrary motions the total angular momentum of the collection can be written

$$\vec{l} = \sum_n^N \vec{r}_n \times \vec{p}_n, \tag{6.11}$$

where n is a summation index labeling the particles of the system. So we're adding up the angular momentum of particle 1 plus particle 2 etc.

If we consider a rigid body rotating about some axis with angular speed ω, all of the particles in the body must have the same angular speed, but they clearly don't all have the same position. Thus for a rigid body, we can express the angular momentum of the body in terms of ω as

$$\vec{l} = \sum_n^N m_n\vec{r}_n \times (\vec{\omega} \times \vec{r}_n). \tag{6.12}$$

There is a vector identity known as the "bac-cab" rule for obvious reasons, which reads

$$\vec{A} \times (\vec{B} \times \vec{C}) = \vec{B}(\vec{A} \cdot \vec{C}) - \vec{C}(\vec{A} \cdot \vec{B}). \tag{6.13}$$

Application of the bac-cab rule for the case of Eq. (6.12) yields

$$\vec{l} = \sum_n^N m_n \left(r_n^2\vec{\omega} - (\vec{\omega} \cdot \vec{r}_n)\vec{r}_n\right). \tag{6.14}$$

Although it's perhaps not completely obvious, this can be repackaged as follows

$$\vec{l} = \widetilde{I}\vec{\omega}, \tag{6.15}$$

where $\widetilde{I}$ is a 3×3 matrix known as the moment of inertia tensor, which takes the form

$$\widetilde{I} = \sum_n^N m_n \begin{pmatrix} r_n^2 - r_{nx}r_{nx} & -r_{nx}r_{ny} & -r_{nx}r_{nz} \\ -r_{nx}r_{ny} & r_n^2 - r_{ny}r_{ny} & -r_{ny}r_{nz} \\ -r_{nx}r_{nz} & -r_{ny}r_{nz} & r_n^2 - r_{nz}r_{nz} \end{pmatrix}. \tag{6.16}$$

My use of the tilde over the I here is intended to indicate that this is a tensor and distinguish it from vectors and scalars. The notation r_n here is the magnitude of the position of the n^{th} particle in the body. The notation r_{nx} is the x component of that vector. Note that this implies $\vec{l}$ and $\vec{\omega}$ are not in the same direction for every situation. Most books at this level would not tell you the whole truth about moment of inertia. They think you can't handle the truth, but I don't agree so I did this out in all of its glory. I'll make some comments about how one gets the thing that introductory books usually do below, and we won't actually solve any problems that require the full result above.

As an aside, which is rather beyond our present scope, you might wonder what a tensor is. Like vectors, there are varying levels of formality with which one can attempt a definition. You might remember some formal discussion of vectors in linear algebra, and a similar discussion can be applied to tensors. One way to get a basic sense of tensors is to note that a scalar is a single number (a 1×1 matrix), a vector (in 3D) is a column matrix of 3 numbers. The thing I introduced above and called a tensor, is actually a 2-tensor, which in 3D can be represented with a 3×3 matrix. A 3-tensor can be represented with a $3 \times 3 \times 3$ cube of numbers, and a 1-tensor is actually a vector. One complicating factor here is that not every set of 3 numbers that you might choose to write as a column matrix is a vector. Only such objects that transform in a particular way under transformations, (in the case of 3D Euclidean space it's rotations), are vectors. Under rotations, the components of vectors get mixed, without changing the magnitude. Higher-rank tensors have similar requirements on their transformation properties.

PCQ 6.4

Do a little of the matrix multiplication in Eq. (6.15). Do just enough to argue that Eq. (6.15) is equivalent to Eq. (6.14).

Many times one wishes to consider an object that can be modeled as a continuous object. In a way similar to the case of center of mass, this can be done by converting the discrete sum into an integral. One can do this separately for each entry in the matrix. For example, the upper right entry, which we could call I_{xx} becomes

$$I_{xx} = \int \rho(r^2 - r_x r_x) d^3 r. \tag{6.17}$$

Perhaps this seems a little daunting at this stage, but don't despair. There is a trick that makes it simpler for many cases you'll encounter. It is a fact of matrix theory that one can find coordinates in which $\widetilde{I}$ is diagonal. These coordinates are called the principle axies. Moreover, these coordinates can easily be guessed in many cases. For objects with a high degree of symmetry, the coordinates that make $\widetilde{I}$ diagonal are those that lie along symmetry axies of the object. In this class, we'll use objects and situations, where you can't help but deal with a principle axis, and need only one component of the moment of inertia tensor.

As a simple example, consider a point mass m doing circular motion with radius r about the origin in the $x - y$ plane. Working in Cartesian coordinates, its position vector $\vec{r}$ will have intermittently nonzero x and y components, but never a nonzero z component. Its angular velocity $\vec{\omega}$ will point only in the z direction. If we put this information into Eq. (6.15), we'll find the following equation:

$$\begin{pmatrix} l_x \\ l_y \\ l_z \end{pmatrix} = m \begin{pmatrix} - & - & 0 \\ - & - & 0 \\ - & - & r^2 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ \omega_z \end{pmatrix}, \tag{6.18}$$

where an entry of "$-$" in the matrix indicates a quantity that I didn't bother with, since it's not relevant for the answer. Note that the sum is irrelevant, since we have only one particle. Hence we find

$$l_z = m r^2 \omega_z, \tag{6.19}$$

consistent with Eq. (6.9). Hence people usually say, "the moment of inertia of a point mass is $I_{\text{point}} = mr^2$," in reference to the only relevant piece of the moment of inertia tensor in this simple case. This language and notation is typical when only one on-diagonal component of the moment of inertia tensor is needed, as will be the case for all problems we'll deal with in this class. Under such circumstances people typically write

$$\vec{l} = I \vec{\omega}, \tag{6.20}$$

where I is the relevant component of the moment of inertial tensor typically referred to as, "the moment of inertial of a (insert name of object), about (insert description of the axis)," as in, "the moment of inertial of uniform thin rod of length a and mass M about an axis perpendicular to the rod through its center of mass is $I = \frac{1}{12} M a^2$."

The following table [1] provides some common moments of inertia. Each of these can be found by doing an integral of the form (6.17).

Schaum's Outline: Mathematical Handbook of formulas + tables; Spiegel + Liu

11 SPECIAL MOMENTS OF INERTIA

The table below shows the moments of inertia of various rigid bodies of mass M. In all cases it is assumed the body has uniform [i.e. constant] density.

TYPE OF RIGID BODY	MOMENT OF INERTIA
11.1 Thin rod of length a	
(a) about axis perpendicular to the rod through the center of mass,	$\frac{1}{12}Ma^2$
(b) about axis perpendicular to the rod through one end.	$\frac{1}{3}Ma^2$
11.2 Rectangular parallelepiped with sides a, b, c	
(a) about axis parallel to c and through center of face ab,	$\frac{1}{12}M(a^2+b^2)$
(b) about axis through center of face bc and parallel to c.	$\frac{1}{12}M(4a^2+b^2)$
11.3 Thin rectangular plate with sides a, b	
(a) about axis perpendicular to the plate through center,	$\frac{1}{12}M(a^2+b^2)$
(b) about axis parallel to side b through center.	$\frac{1}{12}Ma^2$
11.4 Circular cylinder of radius a and height h	
(a) about axis of cylinder,	$\frac{1}{2}Ma^2$
(b) about axis through center of mass and perpendicular to cylindrical axis,	$\frac{1}{12}M(h^2+3a^2)$
(c) about axis coinciding with diameter at one end.	$\frac{1}{12}M(4h^2+3a^2)$
11.5 Hollow circular cylinder of outer radius a, inner radius b and height h	
(a) about axis of cylinder,	$\frac{1}{2}M(a^2+b^2)$
(b) about axis through center of mass and perpendicular to cylindrical axis,	$\frac{1}{12}M(3a^2+3b^2+h^2)$
(c) about axis coinciding with diameter at one end.	$\frac{1}{12}M(3a^2+3b^2+4h^2)$
11.6 Circular plate of radius a	
(a) about axis perpendicular to plate through center,	$\frac{1}{2}Ma^2$
(b) about axis coinciding with a diameter.	$\frac{1}{4}Ma^2$

11.7 Hollow circular plate or ring with outer radius a and inner radius b	
(a) about axis perpendicular to plane of plate through center,	$\frac{1}{2}M(a^2+b^2)$
(b) about axis coinciding with a diameter.	$\frac{1}{4}M(a^2+b^2)$
11.8 Thin circular ring of radius a	
(a) about axis perpendicular to plane of ring through center,	Ma^2
(b) about axis coinciding with diameter.	$\frac{1}{2}Ma^2$
11.9 Sphere of radius a	
(a) about axis coinciding with a diameter,	$\frac{2}{5}Ma^2$
(b) about axis tangent to the surface.	$\frac{7}{5}Ma^2$
11.10 Hollow sphere of outer radius a and inner radius b	
(a) about axis coinciding with a diameter,	$\frac{2}{5}M(a^5-b^5)/(a^3-b^3)$
(b) about axis tangent to the surface.	$\frac{2}{5}M(a^5-b^5)/(a^3-b^3)+Ma^2$
11.11 Hollow spherical shell of radius a	
(a) about axis coinciding with a diameter,	$\frac{2}{3}Ma^2$
(b) about axis tangent to the surface.	$\frac{5}{3}Ma^2$
11.12 Ellipsoid with semi-axes a, b, c	
(a) about axis coinciding with semi-axis c,	$\frac{1}{5}M(a^2+b^2)$
(b) about axis tangent to surface, parallel to semi-axis c and at distance a from center.	$\frac{1}{5}M(6a^2+b^2)$
11.13 Circular cone of radius a and height h	
(a) about axis of cone,	$\frac{3}{10}Ma^2$
(b) about axis through vertex and perpendicular to axis,	$\frac{3}{20}M(a^2+4h^2)$
(c) about axis through center of mass and perpendicular to axis.	$\frac{3}{80}M(4a^2+h^2)$
11.14 Torus with outer radius a and inner radius b	
(a) about axis through center of mass and perpendicular to plane of torus,	$\frac{1}{4}M(7a^2-6ab+3b^2)$
(b) about axis through center of mass and in the plane of the torus.	$\frac{1}{4}M(9a^2-10ab+5b^2)$

Let's try a simple example – the thin rod of mass M and length a about its center. What we mean by 'thin' is that we can treat it as one dimensional, a line of mass. Let the x axis stab through the center of the rod perpendicular to it. Let the y axis point down the rod, and the z complete the set. Since x is a symmetry axis of the system, it will be a principle axis and we can get away without the full tensor technology. For simple cases, particularly with objects of less than 3 dimensions, a conceptual approach can be used. I'll first take this conceptual route, then use the general results here to see how they work.

A conceptual approach: The moment of inertia of a point mass is $I_{\text{point}} = mr^2$, where m is the mass of the point and r is the distance to the rotation axis. But any extended object can be thought of as being made up of an infinite number of infinitesimal point masses having mass dm. So to get the moment of inertia of the extended object, we can just add up the moments of inertia of the individual points by writing,

$$I_{\text{rod}} = \int r^2 dm, \tag{6.21}$$

where r is the distance from the axis to the little chunk dm. To turn it into an integral we can solve, we need to express dm in terms of position. We can do so for a uniform 1D object by defining a mass per length $\lambda = M/a$, and writing $dm = \lambda dy$. That is, mass per length, times a little piece of length is a little piece of mass. So our integral is now

$$I_{\text{rod}} = \lambda \int_{-a/2}^{a/2} y^2 dy. \tag{6.22}$$

Note that I've replaced r with y since y is the variable that describes the distance from the rotation axis to the mass. Now I can do the integral:

$$I_{\text{rod}} = \frac{1}{12} M a^2. \tag{6.23}$$

Just as shown on the sheet.

The formal approach: It's I_{xx} we're after. Since the x component of the position vector r_x is just x, I'll write it that way here

$$I_{xx} = \int \rho(r^2 - r_x r_x) d^3r = \int \rho(x^2 + y^2 + z^2 - x^2) dxdydz. \tag{6.24}$$

The density ρ is a mass per volume, but if we treat the rod as one dimensional, then it doesn't really have a volume. The mass exists only at $x = 0$ and at $z = 0$. So how can we divide its mass by it's volume to get a density? There is a tool for this situation called the Dirac delta "function" $\delta[x]$. It is a "function" that has the following properties:
(1) infinite height at the point where its argument is zero,
(2) is zero everywhere else,
(3) the area under the curve is one,
(4) it's units are the inverse units of the argument,
(5) if you integrate it times some other function, it works as follows:

$$\int \delta[x - x_0] x^2 dx = x_0^2, \tag{6.25}$$

That is, the result is the other function (here x^2) evaluated at the place where the arguement of the delta is zero. For a visual, see Fig. 6.1. Property (5) might make some sense if you think about it as follows. Since the delta is zero everywhere except x_0, there is no contribution to the sum except at x_0. So we just get the value of the other function at this point. Function is in quotes because it does not satisfy the mathematical definition of a function. It ought to be called a distribution instead. With this technology, we can define our mass per volume as follows:

$$\rho = \lambda \delta[x]\delta[z]. \tag{6.26}$$

Now we can do the x and z integrals easily yielding

$$I_{xx} = \int_{-a/2}^{a/2} \lambda y^2 dy \tag{6.27}$$

as we had via our conceptual approach. While this more formal approach doesn't look very helpful here, for 3D objects it's nice, as we'll see in class.

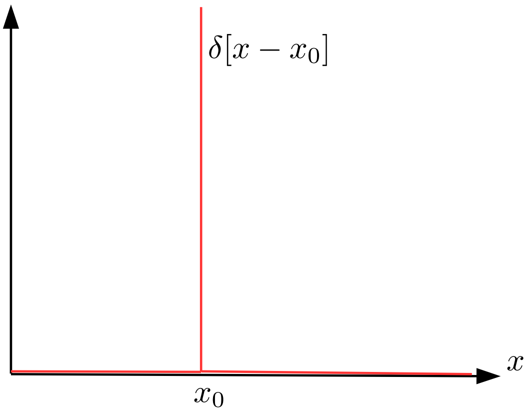


Figure 6.1: A sample Dirac delta.

A final piece of useful technology associated with moment of inertia is called the parallel axis theorem. It allows one to find the moment of inertia about a particular axis through the center of mass when it is known about another parallel axis or visa verse. A version of this can be found for any of the parts of the moment of inertia tensor, but I'll just do it for the case when only one on-diagonal part of the tensor is needed. Here the theorem takes the form

$$I_{||} = I_{\mathrm{CM}} + ma^2, \tag{6.28}$$

where I_{CM} is the moment of inertia about the center of mass, $I_{||}$ is the moment of inertia about a parallel axis not through the center of mass, a is the distance from the center of mass to the parallel axis, and m is the mass of the body. This should make some sense. The moment of inertia is smallest about the center of mass. About some parallel axis, it's what it was about the center of mass, plus the whole body acting as a point mass at some distance a from the new axis.

PCQ 6.5

Show that the moments of inertia for a thin rod in Eq. 11.1(a) and 11.1(b) of the moment of inertia table are consistent with the parallel axis theorem.

6.3 Torque and Energy

Now that we've paid our dues and worked out the moment of inertia tensor and the form of the angular momentum of a body, it's not too hard to guess the rest of the mechanics applying to rotation. Just as Newton's 2nd Law provided the concept of force for the smooth transfer of momentum, we can define a rotational version of Newton's 2nd Law that provides a concept of torque for the smooth transfer of angular momentum between objects. Thus we define the torque τ as

$$\vec{\tau} = \frac{d\vec{l}}{dt} = \widetilde{I}\vec{\alpha}. \tag{6.29}$$

For cases dealing with principle axis requiring only one component of the moment of inertial tensor, we can write

$$\vec{\tau} = I\vec{\alpha}. \tag{6.30}$$

One can also show easily from the defining equation of angular momentum that

$$\vec{\tau} = \vec{r} \times \vec{F}, \tag{6.31}$$

where F is a force applied at a position r.

It is also the case that a body can have kinetic energy associated with both the linear motion of its center of mass as well as the rotational motion about its center of mass. The form of this rotational kinetic energy is

$$T_r = \tfrac{1}{2}\vec{\omega}\widetilde{I}\vec{\omega}, \tag{6.32}$$

or for our simple cases

$$T_r = \tfrac{1}{2} I\omega^2. \tag{6.33}$$

This is perhaps guessable as the only form that matches the $m \to I$, $\vec{v} \to \vec{\omega}$ procedure that takes one from linear quantities to rotational quantities. It can be shown in a number of ways. One is to follow a chain of steps starting from the linear kinetic energy of each particle making up the body and converting velocities into angular velocities. This method parallels the derivation of $\vec{l}$ in terms of $\widetilde{I}$.

PCQ 6.6

Computational Corner Do Mathematica tutorial "04Mathematica.nb". Then answer the following question: at what value of x is the expression $8 + 4x + 0.5x^3$ equal to zero? Figure this out in 2 ways. First with Mathematica's FindRoot function, then with Mathematica's Solve function. Respond to this PCQ by either writing down or printing your code for these 2 steps and comment on any differences between results generated by the 2 approaches.

Chapter 7

Lagrangian Mechanics

Lagrangian Mechanics, published by Lagrange in 1788, some 100 years after the Principia, is an alternative formulation of mechanics entirely equivalent to Newtonian Mechanics. The method "feels simple" and is appealing to the reductionist style of physics. It tends to make solving complex problems easier, coming with a very definite prescription for how to proceed. The key equation associated with the method, the Lagrangian, has become perhaps the most fundamental equation in physics, with all cutting edge work on fundamental interactions in nature formulated through it. As a key tool in quantum mechanics and general relativity alike, the Lagrangian goes far beyond a cute trick for classical mechanics. Here we'll first consider how to apply the method, then we'll have some discussion about how/why it works.

7.1 Basics

The Lagrangian in classical mechanics is defined as the kinetic energy minus the potential energy:

$$L[q_1, q_2, \ldots, q_n, \dot{q}_1, \dot{q}_2, \ldots, \dot{q}_n, t] = T - U. \tag{7.1}$$

As I've displayed explicitly, the Lagrangian will in general be a function of n generalized coordinates, which I've given the generic labels $q_1 - q_n$, the corresponding generalized velocities $\dot{q}_1 - \dot{q}_n$, and possibly time. In this one paragraph, I know I've introduced a lot. Be patient as I pull apart the pieces in the next few paragraphs. You might hold one of the example problems that we solved in class next to you as you read. Try to match the general discussion here with the things that happened in the example.

The number n here is the number of degrees of freedom of the system. This corresponds to the number of independent ways the system can move. This is also the number of variables that are needed to completely describe its motion. As some examples, a point particle in 3 dimensions has 3 degrees of freedom – one for each of the 3 independent directions it can move. An rigid body in 3 dimensions has 6 degrees of freedom – the 3 directions in which it can translate and the 3 independent ways in which it can rotate. So if a system consists of N rigid bodies, the number of degrees of freedom in the system could be up to $6N$. The number of degrees of freedom may be limited by constraints. For example, in the case of a particle sliding down a wedge under gravity, we usually require the particle to remain in the 2D plane of the wedge, and to remain in contact with the wedge. These constraints reduce the number of degrees of freedom for the particle from 3 to 1. So for the wedge problem, we only needed 1 generalized coordinate, the position of the particle along the surface of the wedge.

You might recall back in our discussion of forces that I called things like normal force "constraint forces". These are forces that enforce a constraint. It's the normal force in the wedge example that constrains the particle to stay on the wedge (as opposed to passing through it). You also know very well that constraint forces are a big part of Newtonian mechanics, while you might have noticed that they seem to be absent from our solutions using Lagrangian mechanics. In fact, the value of the Normal force was never a part of our Lagrangian solution of the particle on the wedge. The fact that we don't have to deal with these constraint forces is a key advantage of Lagrangian mechanics. We'll see later that if we want information about the Normal force (or any other constraint force) as a part of our solution, we can ask for it using an extension of the Lagrangian formulation known as the method of Lagrange multipliers.

Note that generalized coordinates are not necessarily coordinates. They are simply some variables that describe the degrees of freedom of the system. They don't have to be a part of a coordinate system and they don't need to have units of length. The generalized velocities are just a time derivative of the generalized coordinates.

Once the Lagrangian has been established for the system the equations of motion are obtained by applying the Euler-Lagrange equations:

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}_j} - \frac{\partial L}{\partial q_j} = 0. \tag{7.2}$$

Note that by q_j I mean a particular one of the $q_1 - q_n$ appearing in the Lagrangian. There will be n "copies" of the Euler-Lagrange equations, one for each of the n degrees of freedom each yielding an equation of motion for the corresponding generalized coordinate. For example, for an unconstrained particle in 3D, three Euler-Lagrange equations will be needed, and they will roughly correspond to the 3 equations of motion one would get from Newton's second law in the Newtonian formulation of this problem.

Note that the generalized coordinates, and generalized velocities are treated as independent in the Euler-Lagrange equations. That is,

$$\frac{\partial \dot{q}}{\partial q} = \frac{\partial q}{\partial \dot{q}} = 0. \tag{7.3}$$

This often seems odd at first. Folks often say, "but when a particle is in free fall, my solution to the problem clearly allows me to express the speed of the particle as a function of its position." While that's true, the key phrase is "my solution to the problem clearly allows me ...". Velocity can only be a known function of position after the problem is solved. In general these quantities are independent. For example, a marble can in general have any speed at height 7 m. It's only after I tell you that it fell to that height under gravity from an initial height of 10 m that you know what its speed will be.

7.2 Tasson's 6-Step Program

These are a set of steps I use when solving Lagrangian problems. I believe that if you follow them carefully, it's almost impossible to get the problem wrong.

Step 1: Define Cartesian coordinates for the body/bodies of interest in an inertial frame of reference. It is also helpful to pick a convenient origin. In the case of a simple pendulum of constant length l hanging from the ceiling, I pick x horizontally along the ceiling, y vertically down, and the origin at the pivot.

Step 2: Choose generalized coordinates that build in the constraints. In the case of our pendulum, the angle θ the pendulum makes with the vertical is a good choice. It clearly provides a complete description of the motion. You should check that your choice completely describes the motion. If you have multiple generalized coordinates, you should also verify that they are independent of one another.

Step 3: Write T and U in Cartesian coordinates. For our pendulum problem, that's

$$T = \tfrac{1}{2}m(\dot{x}^2 + \dot{y}^2), \quad U = -mgy. \tag{7.4}$$

Step 4: Find expressions for the Cartesian coordinates in terms of the generalized coordinates. Differentiate to find the Cartesian velocities in terms of the generalized velocities. For our pendulum problem that's

$$x = l\sin\theta \quad y = l\cos\theta \tag{7.5}$$

and

$$\dot{x} = l\dot{\theta}\cos\theta \quad \dot{y} = -l\dot{\theta}\sin\theta. \tag{7.6}$$

Step 5: Write the Lagrangian in terms of generalized coordinates and generalized velocities by plugging your transformation equations from step 4 into your energies from step 3. Sometimes one might feel that they can cheat steps 3 and 4 and write down the Lagrangian directly. I do this once in a while for cases I know by heart, but if I feel at all unsure, I follow the steps. In the current example, I find

$$L = \tfrac{1}{2}ml^2\dot{\theta}^2 + mgl\cos\theta. \tag{7.7}$$

Step 6: Apply the Euler-Lagrange equations. Here I find

$$ml^2\ddot{\theta} = -mgl\sin\theta. \tag{7.8}$$

Now we just need to solve the differential equation of motion, but that's all the same as Newtonian mechanics. Whoever's mechanics you use, it results in the same (or physically equivalent) differential equations to solve.

7.3 Least Action

This section is perhaps one way to get at the question of why Lagrangian Mechanics works. I'm providing this because you probably should at this stage have some idea of the meaning of Least Action, but I'm not going to work out all of the details. That's a job for PHYS 355.

Lagrangian Mechanics as I've formulated it above can be viewed as a result of the Principle of Least Action. (Some people call this Hamilton's Principle, others draw some distinctions between the Principle of Least Action and Hamilton's Principle. For our present purposes, the ideas are about the same.)

The action is defined as the integral of the Lagrangian along a path from a point A to a point B

$$S = \int L dt. \tag{7.9}$$

The Principle of Least Action states that the particle will take the path from A to B for which the value of the action turns out to be a minimum as compared with all of the other paths the particle might consider. So the idea is that the particle stands at A, sniffs out the path of least action to B, and takes that one. Evidently for the case of a particle in free fall under gravity, the path of least action is a vertical line along which the particle accelerates with magnitude g.

There is a branch of mathematics called The Calculus of Variations, which we will not explore here, that addresses extremizing (maximizing or minimizing) objects that have a form such as the action. It turns out that the condition which must be satisfied such that S is an extremum is that the thing in the integral, in this case L, satisfies the Euler-Lagrange equations. So the Lagrangian formulation can be viewed as coming from the Principle of Least Action.

It's also worth noting that the action has rather deep significance beyond classical mechanics in addition to being the route through which new theories are usually constructed. There is a formulation of quantum mechanics known as the path integral formulation in which a particle traveling from a point A to a point B takes all possible paths at once, not just the one of least action as it does in classical mechanics. As one approaches the classical limit, the other paths besides the classical one become less and less significant. In this interpretation, it turns out that the phase of the wave function is proportional to the action. It also turns out that for free particles in special or general relativity, the action is proportional to the proper time. If you have not yet met any quantum mechanics or relativity, my comments here might not make so much sense, but you should at least get the picture that the action is a really big deal.

So why does it work? Lagrangian Mechanics is a physical theory, built up from undefined terms, definitions, and assumptions, most notably the Principle of Least Action. The assumptions are a part of a logically consistent framework that makes quantitative predictions. It just so happens that those predictions turn out to be correct. Since Newtonian Mechanics also yields the same correct results, they are evidently two equivalent theories that do the same job.

PCQ 7.1

Consider the solution to the problem of a particle in free fall under gravity in the Lagrangian formulation. Which part of the Euler-Lagrange equation gives you the $m\vec{a}$ part of Newton's 2nd Law? Which part gives the $\vec{F}_g$ part of Newton's 2nd Law? What sense do you get about the match between our 2 formulations of mechanics from these results?

PCQ 7.2

What advantages and disadvantages do you see for Lagrangian Mechanics over Newtonian Mechanics?

7.4 Generalized Momentum and Generalized Force

In our development of Lagrangian mechanics thus far, we've met generalized coordinates and generalized velocities. We've seen that in simple cases they correspond directly to what we might normally think of as Cartesian coordinates and velocities, but in more complicated cases they appear as some other variables that seem to be doing the same job. Here we'll take a little closer look at what's going on. Our path will provide perhaps another answer to the question, "Why does Lagrangian mechanics work?"

Consider the simple Lagrangian of a point particle falling under gravity

$$L = \tfrac{1}{2}m\dot{z}^2 - mgz, \tag{7.10}$$

where I've chosen the Cartesian z axis pointing up as my generalized coordinate. So this is perhaps the simplest case where the generalized coordinate is just a coordinate.

As we work out the Euler-Lagrange equation, we do

$$\frac{\partial L}{\partial \dot{z}} = m\dot{z}. \tag{7.11}$$

Note that the thing on the right is the momentum of the particle in the z direction, p_z, which is the only relevant direction in this problem. If our generalized coordinate was something more general we would not get the usual momentum from Newtonian physics, but we would get something that plays a similar role. Thus a quantity called generalized momentum is defined as

$$p_q = \frac{\partial L}{\partial \dot{q}}. \tag{7.12}$$

Just like generalized coordinates are sometimes just Cartesian coordinates and sometimes not, this will sometimes be Cartesian momentum and sometimes not.

If we return to our simple example, perhaps the next thing we would do is take the total time derivative of what we've now called the generalized momentum:

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{z}} = m\ddot{z}. \tag{7.13}$$

From Newton's laws, we know that the total time derivative of momentum is equal to the force. Here we have the total time derivative of generalized momentum, which by purely psyching out this naming scheme, we might expect to be equal to a generalized force, and this is the case. If we write the Euler-Lagrange equations in the following leading form,

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}} = \frac{\partial L}{\partial q}, \tag{7.14}$$

we can identify the part on the right as a generalized force. In the context of our present simple example we again find that the generalized force is the usual Newtonian force

$$\frac{\partial L}{\partial q} = -mg. \tag{7.15}$$

There are two key points about this result. First, when people ask, "Why does Lagrangian mechanics work?" sometimes they really mean, "How can I see that it's the same as Newton's laws?" I think the previous example answers that. Second, generalized anything, means the usual thing re-expressed using convenient variables, or transformed to a set of convenient variables. Depending on the variables chosen, one may find that identifying the generalized quantities with a counterpart that seems more "Newtonian" could be hard. In the homework, we'll see that when an angle is chosen as a generalized coordinate, the corresponding generalized velocity is angular velocity, the generalized momentum is angular momentum, and the generalized force is torque.

PCQ 7.3

Inspect the Lagrangian and the definition of generalized momentum. What must be true about L in order for p to be conserved in a given problem?

7.5 Nonconservative Forces

Thus far, we've seen how to handle conservative forces through the Lagrangian and the Euler Lagrange equations. We've seen that the Lagrangian formulation can be viewed as coming from the principle of least action, a nifty assumption in deed. But nonconservative forces present a problem for the Lagrangian as we've defined it. I'll first offer an argument that not incorporating nonconservative forces is not really a limitation of the Lagrangian approach at the fundamental level. I'll then offer a work-around that allows us to bring nonconservative forces into the discussion in a way that carries many of the advantages of the Lagrangian approach.

7.5.1 Fundamental Physics

If we consider the nonconservative forces we've met, surface friction and aerodynamic drag, both are really approximate models for handling the effect of 10^{23} little particles on the few massive particles that we're attempting to model with classical mechanics. They account for energy leaving the system of interest and going to a place that we don't really care about in classical mechanics. As we stated when we met these forces, they are really associated with kinetic and potential energy on the microscopic level. Thus if we choose to build a theory at the microscopic level, we don't have to deal with them. Their effects will be contained in the kinetic terms and interaction terms associated with the fundamental particles we have in the Lagrangian. Thus for building theories at the fundamental level, the principle of least action appears to be a great way to proceed, and at this level, there are no nonconservative forces messing us up. The problem we face is associated solely with the kinds of approximations we need to make in doing classical mechanics. Thus I view the beauty of the Lagrangian and the principle of least action as unspoiled by this issue, but you can decide for yourself if you agree.

Of course we'd still like to use our approximate models of friction and drag in doing classical mechanics. For that, you need the work-around that I present in the next subsection.

7.5.2 Generalized Nonconservative Forces

The way to handle nonconservative forces is to just add them in at the level of the Euler-Lagrange equations:

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}} = \frac{\partial L}{\partial q} + Q_q. \tag{7.16}$$

Here Q_q is the generalized nonconservaitve force associated with the q coordinate. In the simplest of cases, it will correspond to the component of the nonconservative force in q direction.

In the context of a given problem, one should use the following recipe to get Q_q. Step one: identify the nonconservative force $\vec{F}$ in Cartesian coordinates as one would do in Newtonian mechanics. Step two: develop the transformation equations from Cartesian coordinates to generalized coordinates. Step three: if $\vec{F}$ is a function of the generalized coordinates or generalized velocities, use the transformation equations to eliminate the coordinates and velocities in favor of the generalized coordinates and velocities. Step four: carry out the following transformation of the force vector to generalized coordinates:

$$Q_q = F_x\frac{\partial x}{\partial q} + F_y\frac{\partial y}{\partial q} + F_z\frac{\partial z}{\partial q}. \tag{7.17}$$

This turns out to be the general way of converting a vector from one coordinate system to another. For example, if one has a vector $\vec{A}$ in 2D Cartesian coordinates x, y and one wants to know the $\hat{\phi}$ component of the vector in plane polar coordinates, one would do

$$A_\phi = A_x\frac{\partial x}{\partial \phi} + A_y\frac{\partial y}{\partial \phi}. \tag{7.18}$$

PCQ 7.4

You developed expressions for the coordinate transformation from Cartesian coordinates to plane polar coordinates in PCQ 2.3. Use them to find an expression for A_ϕ in terms of A_x, A_y, r, ϕ.

Evidently, handling nonconservative forces using Eq. (7.16) offers advantages in some cases associated with the fact that generalized coordinates often provide a simpler more systematic way of solving the problem over proceeding directly with Newton's laws. It also provides the opportunity to get a better sense of how the Lagrangian formulation and Newtonian formulation match up. Personally, I have not found occasion to apply this formulation involving generalized nonconservative forces beyond simple examples that illustrate how it's working, but that might mean I'm just not proficient enough with the tool to recognize where it would really be helpful. Chapter 12 of Arya's text [1] offers much more on this type of formulation in terms of generalized forces.

PCQ 7.5

Computational Corner Work through Mathematica tutorial 05Mathematica.nb. As your response to this PCQ, draw a picture associated with the Riemann sum program that I present in the tutorial. Comment on why the approximation improves as dx gets smaller.

7.6 Lagrange Multipliers

At the start, two potential issues with Lagrangian mechanics as compared with Newtonian mechanics that one might have identified were nonconservative forces and constraint forces. We addressed nonconservative forces in the last section, here we'll address constraint forces. Happily I find the resolution in the case of constraint forces much more fulfilling than the case of nonconservative forces.

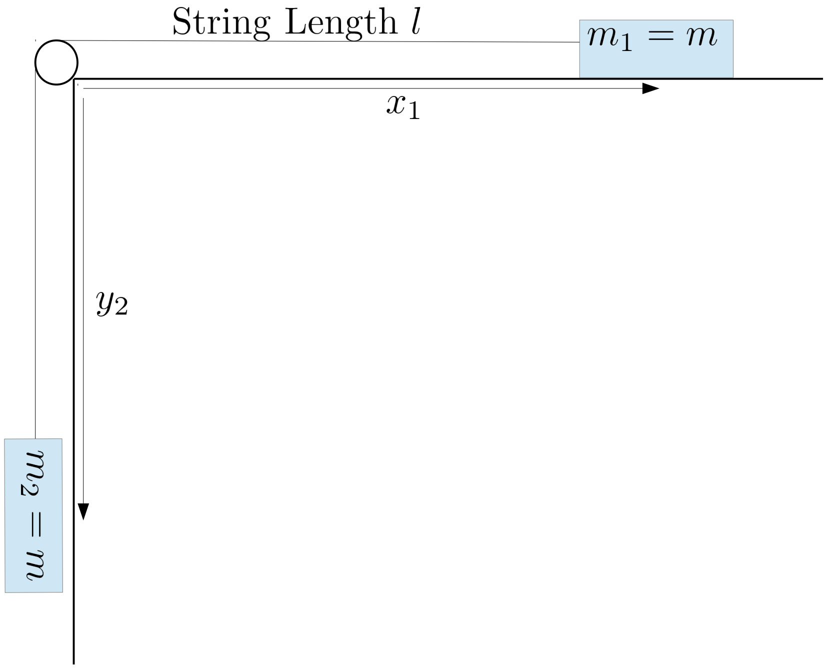
Recall that constraint forces are forces that enforce a constraint. Since that sentence is not very informative, I'll appeal again to an example. When we solved the simple pendulum, the tension in the string supporting the bob was a constraint force. It provided as much force as was needed to keep the bob a fixed distance, l, from the suspension point. Hence it enforced the constraint on the position of the particle $\sqrt{x^2 + y^2} = l$.

PCQ 7.6

Come up with another example of a constraint force in the context of a simple example problem.

When one solves a problem using Newtonian Mechanics, one must take detailed account of the constraint forces in solving the problem just like any other force. One of the advantages of Lagrangian mechanics is that one can get away without consideration of the constraint forces. However, if the constraint force is the target in a give problem, this is not an advantage. This might be the case, for example, if you want to know how strong the string must be in constructing our pendulum.

The way in which the constraint forces have been eliminated in our treatment of Lagrangian mechanics so far has been by using the constraints up front to reduce the number of degrees of freedom. For example, when we considered the problem below, we used $l = x_1 + y_2$ to reduce the number of independent variables in choosing generalized coordinates.



If we want the force that holds this constraint to show up in our theory, we should leave an extra degree of freedom in our description at the start. Then the force that holds the constraint will appear. In the context of the present example, that means using 2 generalized coordinates x_1 and y_2 even though there is only one degree of freedom. Then tension in the cord will then appear as the constraint force.

The way to do this is using the method of Lagrange multipliers, which works via the following steps. Step 1: define a function of the generalized coordinates, which is equal to zero that imposes the constraint. In this case, it could be chosen as

$$G[x_1, y_2] = x_1 + y_2 - l = 0. \tag{7.19}$$

Step 2: add an arbitrary multiplier (the Lagrange multiplier) times your function to the Lagrangian. In this case it's

$$L = L_0 + \lambda G[x_1, y_2] = \tfrac{1}{2}m(\dot{x}_1^2 + \dot{y}_2^2) + mgy_2 + \lambda G[x_1, y_2], \tag{7.20}$$

where L_0 is the Lagrangian without the additional term from the Lagrange Multiplier. Note that since I can add zero to anything without changing it, this is perfectly ok. Now apply the Euler-Lagrange equations. There will be two of them, one for each generalized coordinate, which take the form

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{x}_1} = \frac{\partial L}{\partial x_1} \tag{7.21}$$

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{y}_2} = \frac{\partial L}{\partial y_2}. \tag{7.22}$$

We could also write them more explicitly as follows:

$$\frac{d}{dt}\frac{\partial L_0}{\partial \dot{x}_1} = \frac{\partial L_0}{\partial x_1} + \lambda\frac{\partial G}{\partial x_1} \tag{7.23}$$

$$\frac{d}{dt}\frac{\partial L_0}{\partial \dot{y}_2} = \frac{\partial L_0}{\partial y_2} + \lambda\frac{\partial G}{\partial y_2}. \tag{7.24}$$

In fact, many people don't bother to add the bit to the Lagrangian and rather just add the pieces shown above to the Euler-Lagrange equations. Note that the additional pieces look like a generalized force. In the case of our present example with Cartesian generalized coordinates, $\lambda\frac{\partial G}{\partial x_1}$ is the tension force on m_1 and $\lambda\frac{\partial G}{\partial y_2}$ is the tension force on m_2. However in the case of more general generalized coordinates, these will not quite correspond to force. We now have 3 unknowns $x_1[t], y_2[t]$, and λ, and we have 3 equations: the two Euler-Lagrange equations and the constraint equation. Solving this system yields the position as a function of time of both bodies as well as the constraint force.

Working out the Euler-Lagrange equations explicitly in the present example yields the system of equations

$$m\ddot{x}_1 = \lambda \tag{7.25}$$
$$m\ddot{y}_2 = mg + \lambda \tag{7.26}$$
$$x_1 + y_2 - l = 0. \tag{7.27}$$

Manipulating the system yields

$$\ddot{x}_1 = -\frac{g}{2} \tag{7.28}$$
$$\ddot{y}_2 = \frac{g}{2} \tag{7.29}$$
$$\lambda = -\frac{mg}{2}. \tag{7.30}$$

This should match what you get from Newton's laws in this problem, and $\lambda \frac{\partial G}{\partial x_1} = \lambda$ should make sense as the tension acting on particle 1.

It's perhaps also worth noting that this method can be extended to any number of constraints. One just needs to introduce an additional Lagrange multiplier for each constraint. For example if we have two constraint equations $G[q_1, q_2, \ldots, q_k] = 0$ and $F[q_1, q_2, \ldots, q_k] = 0$ the Lagrangian becomes

$$L = L_0 + \lambda G + \beta F, \tag{7.31}$$

where we now have 2 Lagrange multipliers λ and β.

Note also that in general the constraints we've been using here are known as holonomic constraints, which are those of the form

$$G[q_1, q_2, \ldots, q_k, t] = \text{constant}. \tag{7.32}$$

One could imagine other things such as

$$G[q_1, q_2, \ldots, q_k, t] < \text{constant} \tag{7.33}$$

that are known as nonholonomic, which we have not addressed here.

I think the presentation of Lagrange multipliers here provides an intuitive sense of what's happening, and enough information to use the method in practice. One could also ask some deeper questions about the mathematics of exactly how this works. If you're interested, you could certainly read on about this idea in a number of places [2, 3], but I don't propose to present more here as I believe it would take us too far a field.

PCQ 7.7

Complete the following statements:
(1) Lagrangian mechanics is cool because ...
(2) The thing I dislike most about Lagrangian mechanics is ...
(3) One thing I still don't feel like I understand about Lagrangian mechanics is ...
Think about these carefully. You'll be asked to work with a group in class shortly to synthesize a set of group responses, which we'll discuss as a class.

Bibliography

[1] A.P. Arya, *Introduction to Classical Mechanics* 2nd Edition, Prentice-Hall, Upper Saddle River, New Jersey, 1998.

[2] S.T. Thornton and J.B. Marion, *Classical Dynamics of Particles and Systems*, Brooks/Cole – Thomson, Belmont, California, 2004.

[3] L.N. Hand and J.D. Finch, *Analytical Mechanics*, Cambridge University Press, New York, New York, 1998.

Chapter 8

Central Forces and Orbits

An early triumph for Newtonian mechanics was a match to Kepler's description of orbits. In the first section here, you'll read an elementary treatment of the derivation of some of Kepler's empirical conclusions from Newton's laws. This is from the introductory text, *Physics for Scientists and Engineers* by D.C. Giancoli. In the second section, we'll take a more comprehensive and general Lagrangian approach. The excerpt here is from *Classical Mechanics* by J.R. Taylor. These sections really amount to a very detailed, and very widely applied example of Newtonian mechanics and Lagrangian mechanics, respectively.

8.1 An 'Elementary' Treatment

6–5 Kepler's Laws and Newton's Synthesis

More than a half century before Newton proposed his three laws of motion and his law of universal gravitation, the German astronomer Johannes Kepler (1571–1630) had written a number of astronomical works in which we can find a detailed description of the motion of the planets about the Sun. Kepler's work resulted in part from the many years he spent examining data collected by Tycho Brahe (1546–1601) on the positions of the planets in their motion through the heavens. Among Kepler's writings were three empirical findings that we now refer to as **Kepler's laws of planetary motion**. These are summarized as follows, with additional explanation in Figs. 6–15 and 6–16.

Kepler's first law: The path of each planet about the Sun is an ellipse with the Sun at one focus (Fig. 6–15).

Kepler's second law: Each planet moves so that an imaginary line drawn from the Sun to the planet sweeps out equal areas in equal periods of time (Fig. 6–16).

Kepler's third law: The ratio of the squares of the periods of any two planets revolving about the Sun is equal to the ratio of the cubes of their semimajor axes. [The semimajor axis is half the long (major) axis of the orbit, as shown in Fig. 6–15, and represents the planet's average distance from the Sun.[†]] That is, if T_1 and T_2 represent the periods (the time needed for one revolution about the Sun) for any two planets, and s_1 and s_2 represent their semimajor axes, then

$$\left(\frac{T_1}{T_2}\right)^2 = \left(\frac{s_1}{s_2}\right)^3.$$

We can rewrite this as

$$\frac{s_1^3}{T_1^2} = \frac{s_2^3}{T_2^2},$$

meaning that s^3/T^2 should be the same for each planet. Present-day data are given in Table 6–2; see the last column.

[†]The semimajor axis is equal to the planet's average distance from the Sun in the sense that it equals the average of the planet's nearest and farthest distances from the Sun (points Q and R in Fig. 6–15). Most planetary orbits are close to circles, and for a circle the semimajor axis is the radius of the circle.

TABLE 6–2 Planetary Data Applied to Kepler's Third Law

Planet	Average Distance from Sun, s (10^6 km)	Period, T (Earth years)	s^3/T^2 (10^{24} km^3/y^2)
Mercury	57.9	0.241	3.34
Venus	108.2	0.615	3.35
Earth	149.6	1.0	3.35
Mars	227.9	1.88	3.35
Jupiter	778.3	11.86	3.35
Saturn	1427	29.5	3.34
Uranus	2870	84.0	3.35
Neptune	4497	165	3.34
Pluto	5900	248	3.33

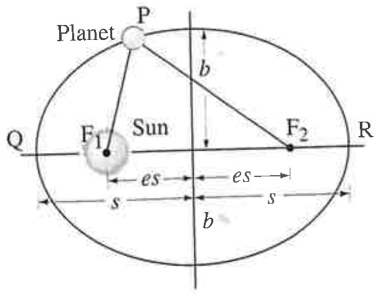


FIGURE 6–15 (a) *Kepler's first law.* An ellipse is a closed curve such that the sum of the distances from any point P on the curve to two fixed points (called the foci, F_1 and F_2) remains constant. That is, the sum of the distances, $F_1P + F_2P$, is the same for all points on the curve. A circle is a special case of an ellipse in which the two foci coincide, at the center of the circle. The semimajor axis is s (that is, the long axis is $2s$) and the semiminor axis is b, as shown. The *eccentricity*, e, is defined so that es is the distance from the center to either focus. The Earth and most of the other planets have nearly circular orbits. For Earth $e = 0.017$.

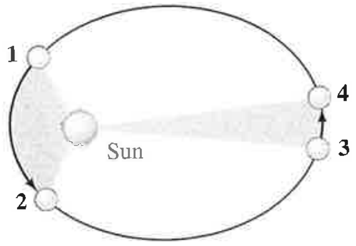


FIGURE 6–16 *Kepler's second law.* The two shaded regions have equal areas. The planet moves from point 1 to point 2 in the same time as it takes to move from point 3 to point 4. Planets move fastest in that part of their orbit where they are closest to the Sun. Exaggerated scale.

Kepler arrived at his laws through careful analysis of experimental data. Fifty years later, Newton was able to show that Kepler's laws could be derived mathematically from the law of universal gravitation and the laws of motion. He also showed that for any reasonable form for the gravitational force law, only one that depends on the inverse square of the distance is fully consistent with Kepler's laws. He thus used Kepler's laws as evidence in favor of his law of universal gravitation. Eq. 6–1.

Derivation of Kepler's third law

We will derive Kepler's second law in Chapter 11 when we study angular momentum. Here we derive Kepler's third law, and we do it for the special case of a circular orbit, in which case the semimajor axis is the radius r of the circle. (Note that most planetary orbits are close to a circle.) First, we write down Newton's second law of motion $\Sigma F = ma$. Then for F we substitute the law of universal gravitation, Eq. 6–1, and for a the centripetal acceleration, v^2/r:

$$\Sigma F = ma$$

$$G\frac{m_1 M_S}{r_1^2} = m_1 \frac{v_1^2}{r_1}.$$

Here m_1 is the mass of a particular planet, r_1 its distance from the Sun, and v_1 its average speed in orbit; M_S is the mass of the Sun, since it is the gravitational attraction of the Sun that keeps each planet in its orbit. The period T_1 of the planet is the time required for one complete orbit, a distance equal to $2\pi r_1$, the circumference of a circle, so

$$v_1 = \frac{2\pi r_1}{T_1}.$$

We substitute this formula for v_1 into the equation above:

$$G\frac{m_1 M_S}{r_1^2} = m_1 \frac{4\pi^2 r_1}{T_1^2}.$$

We rearrange this to get

$$\frac{T_1^2}{r_1^3} = \frac{4\pi^2}{GM_S}. \qquad \textbf{(6–6)}$$

We derived this for planet 1 (say, Mars). The same derivation would apply for a second planet (say, Saturn):

$$\frac{T_2^2}{r_2^3} = \frac{4\pi^2}{GM_S},$$

where T_2 and r_2 are the period and orbit radius, respectively, for the second planet. Since the right sides of the two previous equations are equal, we have $T_1^2/r_1^3 = T_2^2/r_2^3$ or, rearranging,

Kepler's third law

$$\left(\frac{T_1}{T_2}\right)^2 = \left(\frac{r_1}{r_2}\right)^3, \qquad \textbf{(6–7)}$$

which is Kepler's third law. Equations 6–6 and 6–7 are valid also for elliptical orbits if we replace r with the semimajor axis s.

The derivations of Eqs. 6–6 and 6–7 (Kepler's third law) are general enough to be applied to other systems. For example, we could determine the mass of the Earth from Eq. 6–6 using the period of the Moon about the Earth and the Moon's distance from the Earth, or the mass of Jupiter from the period and distance of one of its moons (this is indeed how masses are determined; see the Problems). We can also use Eqs. 6–6 and 6–7 to compare objects that orbit other attracting centers, such as the Moon and a weather satellite orbiting Earth. But be careful not to use Eq. 6–7 to compare, say, the Moon's orbit around the Earth to the orbit of Mars around the Sun because they depend on different attracting centers.

In the following examples, we assume the orbits are circles, although it is not quite true in general.

EXAMPLE 6–8 **Where is Mars?** Mars' period (its "year") was first noted by Kepler to be about 687 days (Earth-days), which is $(687\,\text{d}/365\,\text{d}) = 1.88\,\text{yr}$. Determine the distance of Mars from the Sun using the Earth as a reference.

SOLUTION The period of the Earth is $T_E = 1\,\text{yr}$, and the distance of Earth from the Sun is $r_{ES} = 1.50 \times 10^{11}\,\text{m}$. From Kepler's third law (Eq. 6–7):

$$\frac{r_{MS}}{r_{ES}} = \left(\frac{T_M}{T_E}\right)^{\frac{2}{3}} = \left(\frac{1.88\,\text{yr}}{1\,\text{yr}}\right)^{\frac{2}{3}} = 1.52.$$

So Mars is 1.52 times the Earth's distance from the Sun, or $2.28 \times 10^{11}\,\text{m}$.

EXAMPLE 6–9 **The Sun's mass determined.** Determine the mass of the Sun given the Earth's distance from the Sun as $r_{ES} = 1.5 \times 10^{11}\,\text{m}$.

PHYSICS APPLIED

Determining the Sun's mass

SOLUTION We can use Eq. 6–6 and solve for M_S:

$$M_S = \frac{4\pi^2 r_{ES}^3}{GT_E^2} = \frac{4\pi^2(1.5 \times 10^{11}\,\text{m})^3}{(6.67 \times 10^{-11}\,\text{N}\cdot\text{m}^2/\text{kg}^2)(3.16 \times 10^7\,\text{s})^2} = 2.0 \times 10^{30}\,\text{kg}$$

where we used the fact that

$$T_E = 1\,\text{yr} = (365\tfrac{1}{4}\,\text{d})(24\,\text{h/d})(3600\,\text{s/h}) = 3.16 \times 10^7\,\text{s}.$$

EXAMPLE 6–10 **ESTIMATE** **Geosynchronous satellite, simplified.** A *geosynchronous* satellite of the Earth (as mentioned in Example 6–6) is one that stays above the same point on the equator of the Earth. Estimate the height above the Earth's surface needed for a geosynchronous weather satellite. (This is to be a "lunchtime" calculation done on a napkin, without calculator, as compared to our earlier calculation in Example 6–6.)

SOLUTION To use Kepler's third law we must compare the satellite to some other object that orbits Earth. The simplest choice is the Moon because we know its period and distance. The Moon's period is about $T_M \approx 27\,\text{d}$ and its distance from the Earth about $r_{ME} \approx 380{,}000\,\text{km}$. The period of the weather satellite needs to be $T_{Sat} = 1\,\text{d}$ so that it stays above the same place on the Earth. Hence,

$$r_{Sat} = r_{ME}\left(\frac{T_{Sat}}{T_M}\right)^{\frac{2}{3}} = r_{ME}\left(\frac{1\,\text{d}}{27\,\text{d}}\right)^{\frac{2}{3}} = r_{ME}\left(\frac{1}{3}\right)^2 = \frac{r_{ME}}{9}.$$

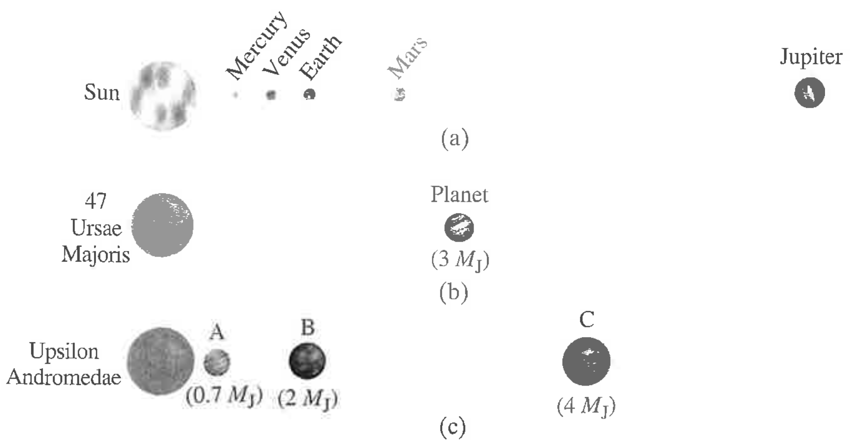
(How nice that the Moon's approximate period turns out to be a perfect cube.) A geosynchronous satellite must be $\frac{1}{9}$ the distance to the Moon, which is 42,000 km from the center of the Earth or 36,000 km above the Earth's surface. This is about 6 Earth radii high.

Accurate measurements on the orbits of the planets indicated that they did not precisely follow Kepler's laws. For example, slight deviations from perfectly elliptical orbits were observed. Newton was aware that this was to be expected from the law of universal gravitation ("every body in the universe attracts every other body…") because each planet exerts a gravitational force on the other planets. Since the mass of the Sun is much greater than that of any planet, the force on one planet due to any other planet will be small in comparison to the force on it due to the Sun. (The derivation of perfectly elliptical orbits ignores the forces due to other planets.) But because of this small force, each planetary orbit should depart from a perfect ellipse, especially when a second planet is fairly close to it. Such deviations, or **perturbations**, as they are called, from perfect ellipses are indeed observed. In fact, Newton's recognition of perturbations in the orbit of Saturn was a hint that helped him formulate the law of universal gravitation, that all

PHYSICS APPLIED

Perturbations and discovery of planets

FIGURE 6–17 (a) Our solar system, compared to recently discovered planets orbiting (b) the star 47 Ursae Majoris, and (c) the star Upsilon Andromedae with at least three planets. M_J is the mass of Jupiter. (Sizes not to scale.)



bodies attract gravitationally. Observation of other perturbations later led to the discovery of Neptune and Pluto. Deviations in the orbit of Uranus, for example, could not all be accounted for by perturbations due to the other known planets. Careful calculation in the nineteenth century indicated that these deviations could be accounted for if there were another planet farther out in the solar system. The position of this planet was predicted from the deviations in the orbit of Uranus, and telescopes focused on that region of the sky quickly found it; the new planet was called Neptune. Similar but much smaller perturbations of Neptune's orbit led to the discovery of Pluto in 1930.

PHYSICS APPLIED

Planets around other stars

More recently, in 1996, planets revolving about distant stars (Fig. 6–17) were inferred from the regular "wobble" of each star due to the gravitational attraction of the revolving planet.

Newton's synthesis

The development by Newton of the law of universal gravitation and the three laws of motion was a major intellectual achievement. For with these laws, Newton was able to describe the motion of objects on Earth and in the heavens. The motions of heavenly bodies and bodies on Earth were seen to follow the same laws (something not previously recognized generally, although Galileo and Descartes had argued in its favor). For this reason, and also because Newton integrated the results of earlier workers into his system, we sometimes speak of Newton's "synthesis."

EXAMPLE 11–6 **Derivation of Kepler's second law.** *Kepler's second law* states that each planet moves so that a line from the Sun to the planet sweeps out equal areas in equal periods of time (see Section 6–5). Use the law of conservation of angular momentum to show this.

SOLUTION The planet moves in an ellipse as shown in Fig. 11–14. In a time dt, the planet moves a distance $v\,dt$ and sweeps out an area dA equal to the area of a triangle of base r and height $v\,dt\sin\theta$ (shown exaggerated in Fig. 11–14). Hence

$$dA = \tfrac{1}{2}(r)(v\,dt\sin\theta)$$

and

$$\frac{dA}{dt} = \tfrac{1}{2}rv\sin\theta.$$

The magnitude of the angular momentum $\mathbf{L}$ about the Sun is

$$L = |\mathbf{r}\times m\mathbf{v}| = mrv\sin\theta,$$

so

$$\frac{dA}{dt} = \frac{1}{2m}L.$$

But L = constant, since the gravitational force $\mathbf{F}$ is directed toward the Sun so the torque it produces is zero (we ignore the pull of the other planets). Hence dA/dt = constant, which is what we set out to prove.

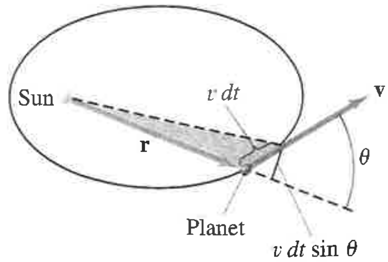


FIGURE 11–14 Kepler's second law of planetary motion (Example 11–6).

8.2 A Lagrangian Approach

PCQ 8.1

This reading is a chance for you to get some practice reading a "real" physics text. As you do so, fill in some missing steps in between the equations and ask some questions about what you're reading. Submit you calculations and questions as the pcq associated with each day's reading from the material to follow.

CHAPTER 8

Two-Body Central-Force Problems

In this chapter, I shall discuss the motion of two bodies each of which exerts a conservative, central force on the other but which are subject to no other, "external," forces. There are many examples of this problem: the two stars of a binary star system, a planet orbiting the sun, the moon orbiting the earth, the electron and proton in a hydrogen atom, the two atoms of a diatomic molecule. In most cases the true situation is more complicated. For example, even if we are interested in just one planet orbiting the sun, we cannot completely neglect the effects of all the other planets; likewise, the moon–earth system is subject to the external force of the sun. Nevertheless, in all cases, it is an excellent starting approximation to treat the two bodies of interest as being isolated from all outside influences.

You may also object that the examples of the hydrogen atom and the diatomic molecule do not belong in classical mechanics, since all such atomic-scale systems must really be treated by quantum mechanics. However, many of the ideas I shall develop in this chapter (the important idea of reduced mass, for instance) play a crucial role in the quantum mechanical two-body problem, and it is probably fair to say that the material covered here is an essential prerequisite for the corresponding quantum material.

8.1 The Problem

Let us consider two objects, with masses m_1 and m_2. For the purposes of this chapter, I shall assume the objects are small enough to be considered as point particles, whose positions (relative to the origin O of some inertial reference frame) I shall denote by $\mathbf{r}_1$ and $\mathbf{r}_2$. The only forces are the forces $\mathbf{F}_{12}$ and $\mathbf{F}_{21}$ of their mutual interaction, which I shall assume is conservative and central. Thus the forces can be derived from a potential energy $U(\mathbf{r}_1, \mathbf{r}_2)$. In the case of two astronomical bodies (the earth and the sun, for instance) the force is the gravitational force $Gm_1m_2/|\mathbf{r}_1 - \mathbf{r}_2|^2$, with the corresponding potential energy (as we saw in Chapter 4)

$$U(\mathbf{r}_1, \mathbf{r}_2) = -\frac{Gm_1m_2}{|\mathbf{r}_1 - \mathbf{r}_2|}. \tag{8.1}$$

For the electron and proton in a hydrogen atom, the potential energy is the Coulomb PE of the two charges (e for the proton and $-e$ for the electron),

$$U(\mathbf{r}_1, \mathbf{r}_2) = -\frac{ke^2}{|\mathbf{r}_1 - \mathbf{r}_2|}, \tag{8.2}$$

where k denotes the Coulomb force constant, $k = 1/4\pi\epsilon_o$.

In both of these examples, U depends only on the difference $(\mathbf{r}_1 - \mathbf{r}_2)$, not on $\mathbf{r}_1$ and $\mathbf{r}_2$ separately. As we saw in Section 4.9, this is no accident: Any isolated system is translationally invariant, and if $U(\mathbf{r}_1, \mathbf{r}_2)$ is translationally invariant it can only depend on $(\mathbf{r}_1 - \mathbf{r}_2)$. In the present case there is a further simplification: As we saw in Section 4.8, if a conservative force is central, then U is independent of the direction of $(\mathbf{r}_1 - \mathbf{r}_2)$. That is, it only depends on the *magnitude* $|\mathbf{r}_1 - \mathbf{r}_2|$, and we can write

$$U(\mathbf{r}_1, \mathbf{r}_2) = U(|\mathbf{r}_1 - \mathbf{r}_2|) \tag{8.3}$$

as is the case in the examples (8.1) and (8.2).

To take advantage of the property (8.3), it is convenient to introduce the new variable

$$\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2. \tag{8.4}$$

As shown in Figure 8.1, this is just the position of body 1 relative to body 2, and I shall refer to $\mathbf{r}$ as the **relative position**. The result of the previous paragraph can be rephrased to say that the potential energy U depends only on the magnitude r of the relative position $\mathbf{r}$,

$$U = U(r). \tag{8.5}$$

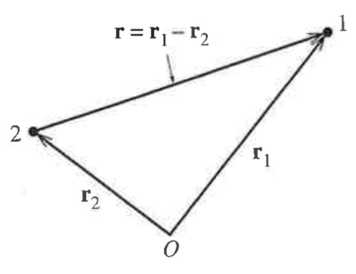


Figure 8.1 The relative position $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$ is the position of body 1 relative to body 2.

We can now state the mathematical problem that we have to solve: We want to find the possible motions of two bodies (the moon and the earth, or an electron and a proton), whose Lagrangian is

$$\mathcal{L} = \tfrac{1}{2}m_1\dot{\mathbf{r}}_1^2 + \tfrac{1}{2}m_2\dot{\mathbf{r}}_2^2 - U(r). \tag{8.6}$$

Of course, I could equally have stated the problem in Newtonian terms, and I shall in fact feel free to move back and forth between the Lagrangian and Newtonian formalisms according to which seems the more convenient. For the present, the Lagrangian formalism is the more transparent.

8.2 CM and Relative Coordinates; Reduced Mass

Our first task is to decide what generalized coordinates to use to solve our problem. There is already a strong suggestion that we should use the relative position $\mathbf{r}$ as one of them (or as three of them, depending on how you count coordinates), because the potential energy $U(r)$ takes such a simple form in terms of $\mathbf{r}$. The question is then, what to choose for the other (vector) variable. The best choice turns out to be the familiar *center of mass* (or CM) position, $\mathbf{R}$, of the two bodies, defined as in Chapter 3 to be

$$\mathbf{R} = \frac{m_1\mathbf{r}_1 + m_2\mathbf{r}_2}{m_1 + m_2} = \frac{m_1\mathbf{r}_1 + m_2\mathbf{r}_2}{M}, \tag{8.7}$$

where as before M denotes the total mass of the two bodies:

$$M = m_1 + m_2.$$

As we saw in Chapter 3, the CM of two particles lies on the line joining them, as shown in Figure 8.2. The distances of the center of mass from the two masses m_2 and m_1 are in the ratio m_1/m_2. In particular, if m_2 is much greater than m_1, then the CM is very close to body 2. (In Figure 8.2, the ratio m_1/m_2 is about 1/3, so the CM is a quarter of the way from m_2 to m_1.)

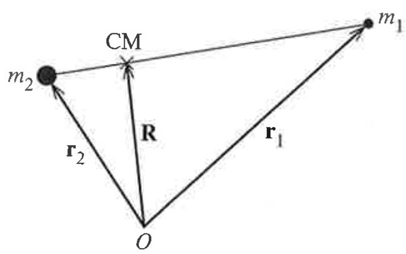


Figure 8.2 The center of mass of the two bodies lies at the position $\mathbf{R} = (m_1\mathbf{r}_1 + m_2\mathbf{r}_2)/M$ on the line joining the two bodies.

We saw in Section 3.3 that the total momentum of the two bodies is the same as if the total mass $M = m_1 + m_2$ were concentrated at the CM and were following the CM as it moves:

$$\mathbf{P} = M\dot{\mathbf{R}}. \tag{8.8}$$

This result has important simplifying consequences: We know, of course, that the total momentum is constant. Therefore, according to (8.8), $\dot{\mathbf{R}}$ is constant; and this means we can choose an inertial reference frame in which the CM is at rest. This **CM frame** is an especially convenient frame in which to analyze the motion, as we shall see.

I am going to use the CM position $\mathbf{R}$ and the relative position $\mathbf{r}$ as generalized coordinates for our discussion of the motion of our two bodies. In terms of these coordinates, we already know that the potential energy takes the simple form $U = U(r)$. To express the kinetic energy in these terms, we need to write the old variables $\mathbf{r}_1$ and $\mathbf{r}_2$ in terms of the new $\mathbf{R}$ and $\mathbf{r}$. It is a straightforward exercise to show that (see Figure 8.2)

$$\mathbf{r}_1 = \mathbf{R} + \frac{m_2}{M}\mathbf{r} \qquad \text{and} \qquad \mathbf{r}_2 = \mathbf{R} - \frac{m_1}{M}\mathbf{r}. \tag{8.9}$$

Thus the kinetic energy is

$$\begin{aligned} T &= \tfrac{1}{2}\left(m_1\dot{\mathbf{r}}_1^2 + m_2\dot{\mathbf{r}}_2^2\right) \\ &= \tfrac{1}{2}\left(m_1\left[\dot{\mathbf{R}} + \frac{m_2}{M}\dot{\mathbf{r}}\right]^2 + m_2\left[\dot{\mathbf{R}} - \frac{m_1}{M}\dot{\mathbf{r}}\right]^2\right) \\ &= \tfrac{1}{2}\left(M\dot{\mathbf{R}}^2 + \frac{m_1 m_2}{M}\dot{\mathbf{r}}^2\right). \end{aligned} \tag{8.10}$$

The result (8.10) simplifies further if we introduce the parameter

$$\mu = \frac{m_1 m_2}{M} \equiv \frac{m_1 m_2}{m_1 + m_2} \qquad \text{[reduced mass]} \tag{8.11}$$

which has the dimensions of mass and is called the **reduced mass**. You can easily check that μ is always less than both m_1 and m_2 (hence the name). If $m_1 \ll m_2$, then μ is very close to m_1. Thus the reduced mass for the earth–sun system is almost exactly the mass of the earth; the reduced mass of the electron and proton in hydrogen is almost exactly the mass of the electron. On the other hand, if $m_1 = m_2$, then obviously $\mu = \frac{1}{2}m_1$.

Returning to (8.10), we can rewrite the kinetic energy in terms of μ as

$$T = \tfrac{1}{2}M\dot{\mathbf{R}}^2 + \tfrac{1}{2}\mu\dot{\mathbf{r}}^2. \tag{8.12}$$

This remarkable result shows that the kinetic energy is the same as that of two "fictitious" particles, one of mass M moving with the speed of the CM, and the other of mass μ (the reduced mass) moving with the speed of the relative position $\mathbf{r}$. Even more significant is the corresponding result for the Lagrangian:

$$\begin{aligned} \mathcal{L} = T - U &= \tfrac{1}{2}M\dot{\mathbf{R}}^2 + \left(\tfrac{1}{2}\mu\dot{\mathbf{r}}^2 - U(r)\right) \\ &= \mathcal{L}_{\text{cm}} + \mathcal{L}_{\text{rel}}. \end{aligned} \tag{8.13}$$

We see that by using the CM and relative positions as our generalized coordinates, we have split the Lagrangian into two separate pieces, one of which involves only the CM coordinate $\mathbf{R}$ and the other only the relative coordinate $\mathbf{r}$. This will mean that we can solve for the motions of $\mathbf{R}$ and $\mathbf{r}$ as two separate problems, which will greatly simplify matters.

8.3 The Equations of Motion

With the Lagrangian (8.13), we can write down the equations of motion of our two-body system. Because $\mathcal{L}$ is independent of $\mathbf{R}$, the $\mathbf{R}$ equation (really three equations, one each for X, Y, and Z) is especially simple,

$$M\ddot{\mathbf{R}} = 0 \qquad \text{or} \qquad \dot{\mathbf{R}} = \text{const.} \tag{8.14}$$

We can explain this result in several ways: First (as we already knew), it is a direct consequence of conservation of total momentum. Alternatively, we can view it as reflecting that $\mathcal{L}$ is independent of $\mathbf{R}$, or, in the terminology introduced in Section 7.6, the CM coordinate $\mathbf{R}$ is "ignorable." More specifically, $\mathcal{L}_{\text{cm}} = \frac{1}{2}M\dot{\mathbf{R}}^2$ (which is the only part of $\mathcal{L}$ that involves $\mathbf{R}$) has the form of the Lagrangian of a *free* particle of mass M and position $\mathbf{R}$. Naturally, therefore (Newton's first law), $\mathbf{R}$ moves with constant velocity.

The Lagrange equation for the relative coordinate $\mathbf{r}$ is a little less simple but equally beautiful: $\mathcal{L}_{\text{rel}}$, the only part of $\mathcal{L}$ that involves $\mathbf{r}$, is mathematically indistinguishable from the Lagrangian for a single particle of mass μ and position $\mathbf{r}$, with potential energy $U(r)$. Thus the Lagrange equation corresponding to $\mathbf{r}$ is just (check it and see!)

$$\mu\ddot{\mathbf{r}} = -\nabla U(\mathbf{r}). \tag{8.15}$$

To solve for the relative motion, we have only to solve Newton's second law for a single particle of mass equal to the reduced mass μ and position $\mathbf{r}$, with potential energy $U(r)$.

The CM Reference Frame

Our problem becomes even easier to think about if we make a clever choice of reference frame. Specifically, because $\dot{\mathbf{R}} = \text{const}$, we can choose an inertial reference frame, the so-called **CM frame**, in which the CM is at rest and the total momentum

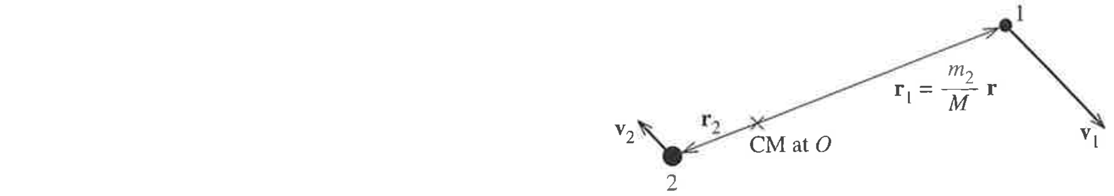


Figure 8.3 In the CM frame the center of mass is stationary at the origin. The relative position $\mathbf{r}$ is the position of particle 1 relative to particle 2; therefore, the position of particle 1 relative to the origin is $\mathbf{r}_1 = (m_2/M)\mathbf{r}$.

is zero. In this frame, $\dot{\mathbf{R}} = 0$ and the CM part of the Lagrangian is zero ($\mathcal{L}_{\text{cm}} = 0$). Thus in the CM frame

$$\mathcal{L} = \mathcal{L}_{\text{rel}} = \tfrac{1}{2}\mu\dot{\mathbf{r}}^2 - U(r) \tag{8.16}$$

and the problem really is reduced to a one-body problem. This dramatic simplification illustrates the curious terminology of the "ignorable coordinate." Recall that a coordinate q_i is said to be ignorable if $\partial\mathcal{L}/\partial q_i = 0$. We see that, in the present case at least, the motion associated with the ignorable coordinate $\mathbf{R}$ really is something that we can ignore.

It is worth taking a moment to consider what the motion looks like in the CM frame, as shown in Figure 8.3. The CM is stationary, and we naturally take it to be the origin. Both particles are moving, but with equal and opposite momenta. If m_2 is much greater than m_1 (as is often the case), the CM is close to m_2 and particle 2 has a small velocity. (In the figure, $m_2 = 3m_1$ and hence $v_2 = \frac{1}{3}v_1$.) It is important to note that the relative position $\mathbf{r}$ is the position of particle 1 relative to particle 2, and is not the actual position of either particle. As shown in the picture, the position of particle 1 is actually $\mathbf{r}_1 = (m_2/M)\mathbf{r}$. However, if $m_2 \gg m_1$, then the CM is very close to particle 2, which is almost stationary, and $\mathbf{r}_1 \approx \mathbf{r}$; that is, $\mathbf{r}$ is very nearly the same thing as $\mathbf{r}_1$.

The equation of motion in the CM frame is derived from the Lagrangian $\mathcal{L}_{\text{rel}}$ of (8.16) and is just Equation (8.15). This is precisely the same as the equation for a single particle of mass equal to the reduced mass μ, in the fixed central force field of the potential energy $U(r)$. In the equations of this chapter, the repeated appearance of the mass μ will serve to remind you that the equations apply to the relative motion of two bodies. However, you may find it easier to *visualize* a single body (of mass μ) orbiting about a fixed force center. In particular, if $m_2 \gg m_1$, these two problems are for practical purposes exactly the same. Moreover, if your interest actually is in a single body, of mass m say, orbiting a fixed force center, then you can use all of the same equations, simply replacing μ with m. In any event, any solution for the relative coordinate $\mathbf{r}(t)$ always gives us the motion of particle 1 relative to particle 2. Equivalently, using the relations of Figure 8.3, knowledge of $\mathbf{r}(t)$ tells us the motion of particle 1 (or particle 2) relative to the CM.

Conservation of Angular Momentum

We already know that the total angular momentum of our two particles is conserved. Like so many other things, this condition takes an especially simple form in the CM frame. In any frame, the total angular momentum is

$$\begin{aligned} \mathbf{L} &= \mathbf{r}_1 \times \mathbf{p}_1 + \mathbf{r}_2 \times \mathbf{p}_2 \\ &= m_1 \mathbf{r}_1 \times \dot{\mathbf{r}}_1 + m_2 \mathbf{r}_2 \times \dot{\mathbf{r}}_2. \end{aligned} \tag{8.17}$$

In the CM frame, we see from (8.9) (with $\mathbf{R} = 0$) that

$$\mathbf{r}_1 = \frac{m_2}{M}\mathbf{r} \qquad \text{and} \qquad \mathbf{r}_2 = -\frac{m_1}{M}\mathbf{r}. \tag{8.18}$$

Substituting into (8.17), we see that the angular momentum in the CM frame is

$$\begin{aligned} \mathbf{L} &= \frac{m_1 m_2}{M^2}(m_2 \mathbf{r} \times \dot{\mathbf{r}} + m_1 \mathbf{r} \times \dot{\mathbf{r}}) \\ &= \mathbf{r} \times \mu \dot{\mathbf{r}} \end{aligned} \tag{8.19}$$

where I have replaced $m_1 m_2 / M$ by the reduced mass μ.

The most remarkable thing about this result is that the total angular momentum in the CM frame is exactly the same as the angular momentum of a single particle with mass μ and position $\mathbf{r}$. For our present purposes the important point is that, because angular momentum is conserved, we see that the vector $\mathbf{r} \times \dot{\mathbf{r}}$ is constant. In particular, the *direction* of $\mathbf{r} \times \dot{\mathbf{r}}$ is constant, which implies that the two vectors $\mathbf{r}$ and $\dot{\mathbf{r}}$ remain in a fixed plane. That is, in the CM frame, the whole motion remains in a fixed plane, which we can take to be the xy plane. In other words, in the CM frame, the two-body problem with central conservative forces is reduced to a two-dimensional problem.

The Two Equations of Motion

To set up the equations of motion for the remaining two-dimensional problem, we need to choose coordinates in the plane of the motion. The obvious choice is to use the polar coordinates r and ϕ, in terms of which the Lagrangian (8.16) is

$$\mathcal{L} = \tfrac{1}{2}\mu(\dot{r}^2 + r^2\dot{\phi}^2) - U(r). \tag{8.20}$$

Since this Lagrangian is independent of ϕ, the coordinate ϕ is ignorable, and the Lagrange equation corresponding to ϕ is just

$$\frac{\partial \mathcal{L}}{\partial \dot{\phi}} = \mu r^2 \dot{\phi} = \text{const} = \ell \qquad [\phi \text{ equation}]. \tag{8.21}$$

Since $\mu r^2 \dot{\phi}$ is the angular momentum ℓ (strictly, the z component ℓ_z), the ϕ equation is just a statement of conservation of angular momentum.

The Lagrange equation corresponding to r (often called the **radial equation**) is

$$\frac{\partial \mathcal{L}}{\partial r} = \frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{r}},$$

or

$$\mu r\dot{\phi}^2 - \frac{dU}{dr} = \mu\ddot{r} \qquad [r \text{ equation}]. \tag{8.22}$$

As we already saw in Example 7.2 [Equations (7.19) and (7.20)], if we move the centripetal term $\mu r\dot{\phi}^2$ over to the right, this is just the radial component of $\mathbf{F} = m\mathbf{a}$ (or rather, $\mathbf{F} = \mu\mathbf{a}$, since μ has replaced m).

8.4 The Equivalent One-Dimensional Problem

The two equations of motion that we have to solve are the ϕ equation (8.21) and radial equation (8.22). The constant ℓ (the angular momentum) in the ϕ equation is determined by the initial conditions, and our main use for the ϕ equation is to solve it for $\dot{\phi}$,

$$\dot{\phi} = \frac{\ell}{\mu r^2}, \tag{8.23}$$

which will let us eliminate $\dot{\phi}$ from the radial equation in favor of the constant ℓ. The radial equation can be rewritten as

$$\mu\ddot{r} = -\frac{dU}{dr} + \mu r\dot{\phi}^2 = -\frac{dU}{dr} + F_{\text{cf}} \tag{8.24}$$

which has the form of Newton's second law for a particle in *one* dimension with mass μ and position r, subject to the actual force $-dU/dr$ plus a "fictitious" outward centrifugal force[1]

$$F_{\text{cf}} = \mu r\dot{\phi}^2. \tag{8.25}$$

In other words, the particle's radial motion is exactly the same as if the particle were moving in one dimension, subject to the actual force $-dU/dr$ *plus* the centrifugal force F_{cf}.

We have now reduced the problem of the relative motion of two bodies to a single one-dimensional problem, as expressed by (8.24). Before we discuss what the solutions are going to look like, it is helpful to rewrite the centrifugal force, using the ϕ equation (8.23) to eliminate $\dot{\phi}$ in favor of the constant ℓ,

$$F_{\text{cf}} = \frac{\ell^2}{\mu r^3}. \tag{8.26}$$

Even better, we can now express the centrifugal force in terms of a centrifugal potential energy,

$$F_{\text{cf}} = -\frac{d}{dr}\left(\frac{\ell^2}{2\mu r^2}\right) = -\frac{dU_{\text{cf}}}{dr}, \tag{8.27}$$

[1] This centrifugal force may be a little more familiar if I write it in terms of the azimuthal velocity $v_\phi = r\dot{\phi}$ as $F_{\text{cf}} = \mu {v_\phi}^2/r$.

where the centrifugal potential energy U_{cf} is defined as

$$U_{cf}(r) = \frac{\ell^2}{2\mu r^2}. \tag{8.28}$$

Returning to (8.24), we can now rewrite the radial equation in terms of U_{cf} as

$$\mu\ddot{r} = -\frac{d}{dr}[U(r) + U_{cf}(r)] = -\frac{d}{dr}U_{eff}(r), \tag{8.29}$$

where the **effective potential energy** $U_{eff}(r)$ is the sum of the actual potential energy $U(r)$ and the centrifugal $U_{cf}(r)$:

$$U_{eff}(r) = U(r) + U_{cf}(r) = U(r) + \frac{\ell^2}{2\mu r^2}. \tag{8.30}$$

According to (8.29), the radial motion of the particle is exactly the same as if the particle were moving in one dimension with an effective potential energy $U_{eff} = U + U_{cf}$.

EXAMPLE 8.1 Effective Potential Energy for a Comet

Write down the actual and effective potential energies for a comet (or planet) moving in the gravitational field of the sun. Sketch the three potential energies involved and use the graph of $U_{eff}(r)$ to describe the motion of r. Since planetary motion was first described mathematically by the German astronomer Johannes Kepler, 1571–1630, this problem of the motion of a planet or comet around the sun (or any two bodies interacting via an inverse-square force) is often called the *Kepler problem.*

The actual gravitational potential energy of the comet is given by the well-known formula

$$U(r) = -\frac{Gm_1m_2}{r} \tag{8.31}$$

where G is the universal gravitational constant, and m_1 and m_2 are the masses of the comet and the sun. The centrifugal potential energy is given by (8.28), so the total effective potential energy is

$$U_{eff}(r) = -\frac{Gm_1m_2}{r} + \frac{\ell^2}{2\mu r^2}. \tag{8.32}$$

The general behavior of this effective potential energy is easily seen (Figure 8.4). When r is large, the centrifugal term $\ell^2/2\mu r^2$ is negligible compared to the gravitational term $-Gm_1m_2/r$, and the effective PE, $U_{eff}(r)$, is negative and

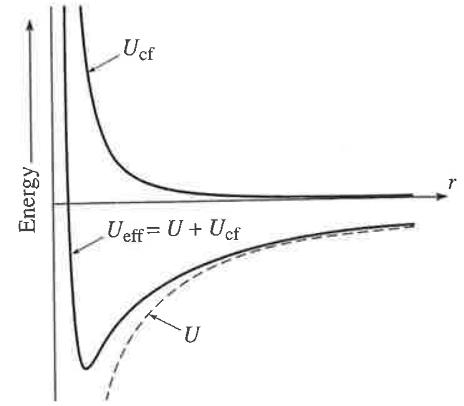


Figure 8.4 The effective potential energy $U_{eff}(r)$ that governs the radial motion of a comet is the sum of the actual gravitational potential energy $U(r) = -Gm_1m_2/r$ and the centrifugal term $U_{cf} = \ell^2/2\mu r^2$. For large r, the dominant effect is the attractive gravitational force; for small r, it is the repulsive centrifugal force.

sloping up as r increases. According to (8.29), the acceleration of r is down this slope. [The roller coaster car accelerates down the track defined by $U_{eff}(r)$.] Thus when a comet is far from the sun, $\ddot{r}$ is always inward.

When r is small, the centrifugal term $\ell^2/2\mu r^2$ dominates the gravitational term $-Gm_1m_2/r$ (unless $\ell = 0$), and near $r = 0$, $U_{eff}(r)$ is positive and slopes downward. Thus, as a comet gets closer to the sun, $\ddot{r}$ eventually becomes outward, and the comet starts to move away from the sun again. The one exception to this statement is when the angular momentum is exactly zero, $\ell = 0$, in which case (8.23) implies that $\dot{\phi} = 0$; that is, the comet is moving exactly radially, along a line of constant ϕ, and must at some time hit the sun.

Conservation of Energy

To find the details of the orbit we must look more closely at the radial equation (8.29). If we multiply both sides of that equation by $\dot{r}$, we find that

$$\frac{d}{dt}\left(\tfrac{1}{2}\mu\dot{r}^2\right) = -\frac{d}{dt}U_{eff}(r). \tag{8.33}$$

In other words,

$$\tfrac{1}{2}\mu\dot{r}^2 + U_{eff}(r) = \text{const.} \tag{8.34}$$

This result is, in fact, just conservation of energy: If we write out U_{eff} as $U + \ell^2/2\mu r^2$ and replace ℓ by $\mu r^2\dot{\phi}$, we see that

$$\begin{aligned}\tfrac{1}{2}\mu\dot{r}^2 + U_{eff}(r) &= \tfrac{1}{2}\mu\dot{r}^2 + \tfrac{1}{2}\mu r^2\dot{\phi}^2 + U(r)\\ &= E.\end{aligned} \tag{8.35}$$

This completes the rewriting of the two-dimensional problem of the relative motion as an equivalent one-dimensional problem involving just the radial motion. We see that the total energy (which we knew all along is constant) can be thought of as the one-dimensional kinetic energy of the radial motion, plus the effective one-dimensional potential energy U_{eff}, since the latter includes the actual potential energy U *and* the kinetic energy $\frac{1}{2}\mu r^2\dot{\phi}^2$ of the angular motion. This means that all of our experience with one-dimensional problems, both in terms of forces and in terms of energy, can be immediately transferred to the two-body central-force problem.

EXAMPLE 8.2 Energy Considerations for a Comet or Planet

Examine again the comet (or planet) of Example 8.1 and, by considering its total energy E, find the equation that determines the maximum and minimum distances of the comet from the sun, if $E > 0$ and, again, if $E < 0$.

In the energy equation (8.35) the term $\frac{1}{2}\mu\dot{r}^2$ on the left is always greater than or equal to zero. Therefore, the comet's motion is confined to those regions where $E \geq U_{\text{eff}}$. To see what this implies, I have redrawn in Figure 8.5 the graph of U_{eff} from Figure 8.4. Let us consider first the case that the comet's energy is greater than zero. In the figure I have drawn a dashed horizontal line at height E, labeled $E > 0$. A comet with this energy can move anywhere that this line is above the curve of $U_{\text{eff}}(r)$, but nowhere that the line is below the curve. This means simply that the comet cannot move anywhere inside the turning point labeled r_{min}, determined by the condition

$$U_{\text{eff}}(r_{\text{min}}) = E. \tag{8.36}$$

If the comet is initially moving in, toward the sun, then it will continue to do so until it reaches r_{min}, where $\dot{r} = 0$ instantaneously. It then moves outward, and, since there are no other points at which $\dot{r}$ can vanish, it eventually moves off to infinity, and the orbit is **unbounded**.

If instead $E < 0$, then the line drawn at height E (labeled $E < 0$) meets the curve of $U_{\text{eff}}(r)$ at the two turning points labeled r_{min} and r_{max}, and a comet with

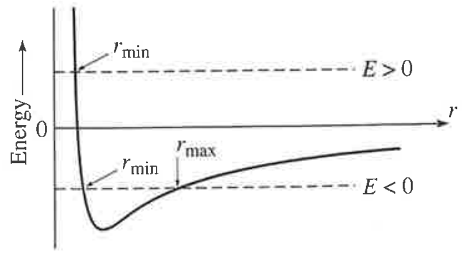


Figure 8.5 Plot of the effective potential energy $U_{\text{eff}}(r)$ against r for a comet. For a given energy E, the comet can only go where $E \geq U_{\text{eff}}(r)$. For $E > 0$ this means it cannot go inside the turning point at r_{min} where $U_{\text{eff}} = E$. For $E < 0$ it is confined between the two turning points labeled r_{min} and r_{max}.

$E < 0$ is trapped between these two values of r. If it is moving away from the sun ($\dot{r} > 0$) it continues to do so until it reaches $r_{\max}$, where $\dot{r}$ vanishes and reverses sign. The comet then moves inward until it reaches $r_{\min}$, where $\dot{r}$ reverses again. Therefore, the comet oscillates in and out between $r_{\min}$ and $r_{\max}$. For obvious reasons, this type of orbit is called a **bounded orbit**.[2]

Finally, if E is equal to the minimum value of $U_{\text{eff}}(r)$ (for a given value of the angular momentum ℓ), the two turning points $r_{\min}$ and $r_{\max}$ coalesce, and the comet is trapped at a fixed radius and moves in a circular orbit.

In this example, I considered just the case of an inverse-square force, but many two-body problems have the same qualitative features. For example, the motion of the two atoms in a diatomic molecule is governed by an effective potential that was sketched in Figure 4.12 and looks very like the gravitational curve of Figure 8.5. Thus all of our qualitative conclusions apply to the diatomic molecule and many other two-body problems.

In thinking about the radial motion of the two-body problem, you must not entirely forget the angular motion. According to (8.23), $\dot{\phi} = \ell/\mu r^2$, and ϕ is always changing, always with the same sign (continually increasing or continually decreasing). For example, as a comet with positive energy approaches the sun, the angle ϕ changes, at a rate that increases as r gets smaller; as the comet moves away, ϕ continues to change in the same direction, but at a rate that decreases as r gets larger. Thus the actual orbit of a positive-energy comet looks something like Figure 8.6. For the case of an inverse-square force (like gravity), the orbit of Figure 8.6 is actually a hyperbola, as we shall prove shortly, but the unbounded orbits (that is, orbits with $E > 0$) are qualitatively similar for many different force laws.

For the bounded orbits ($E < 0$), we have seen that r oscillates between the two extreme values $r_{\min}$ and $r_{\max}$, while ϕ continually increases (or decreases, but let's suppose the comet is orbiting counter-clockwise, so that ϕ is increasing). In the case of the inverse-square force, we shall see that the period of the radial oscillations happens to equal the time for ϕ to make exactly one complete revolution. Therefore, the motion repeats itself exactly once per revolution, as in Figure 8.7(a). (We shall also see that, for any inverse-square force, the bounded orbits are actually ellipses.) For most other force laws, the period of the radial motion is different from the time to make one revolution, and in most cases the orbit is not even closed (that is, it never returns to its initial conditions).[3] Figure 8.7(b) shows an orbit for which r goes from $r_{\min}$ to $r_{\max}$ and back to $r_{\min}$ in the time that the angle ϕ advances by about 330°, and the orbit certainly does not close on itself after one revolution.

[2] If we consider just one comet in orbit around the sun, then energy conservation implies that a bounded orbit ($E < 0$) can never change into an unbounded orbit ($E > 0$), nor vice versa. In reality a comet can occasionally come close enough to another comet or planet to change E, and the orbit can then change from bounded to unbounded or the other way.

[3] Besides the inverse square force, the only important exception is the isotropic harmonic oscillator, for which the orbits are also ellipses, as discussed in Section 5.3.

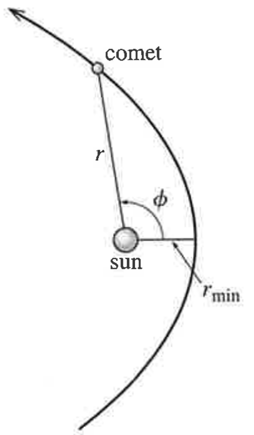


Figure 8.6 Typical unbounded orbit for a positive-energy comet. Initially r decreases from infinity to r_{min} and then goes back out to infinity. Meanwhile the angle ϕ is continually increasing.

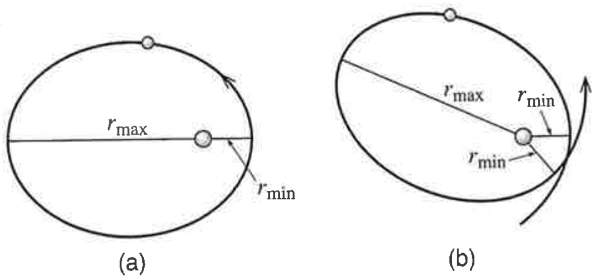


Figure 8.7 **(a)** The bounded orbits for any inverse-square force have the unusual property that r goes from r_{min} to r_{max} and back to r_{min} in exactly the time that ϕ goes from 0 to 360°. Therefore the orbit repeats itself every revolution. **(b)** For most other force laws, the period of oscillation of r is different from the time in which ϕ advances by 360°, and the orbit does not close on itself after one revolution. In this example, r completes one cycle from r_{min} to r_{max} and back to r_{min} while ϕ advances by about 330°.

8.5 The Equation of the Orbit

The radial equation (8.29) determines r as a function of t, but for many purposes we would like to know r as a function of ϕ. For example, the function $r = r(\phi)$ will tell us the shape of the orbit more directly. Thus we would like to rewrite the radial equation as a differential equation for r in terms of ϕ. There are two tricks for doing this, but let me first write the radial equation in terms of forces:

$$\mu\ddot{r} = F(r) + \frac{\ell^2}{\mu r^3} \tag{8.37}$$

where $F(r)$ is the actual central force, $F = -dU/dr$, and the second term is the centrifugal force.

The first trick to rewriting this equation in terms of ϕ is to make the substitution

$$u = \frac{1}{r} \quad \text{or} \quad r = \frac{1}{u} \tag{8.38}$$

and the second is to rewrite the differential operator d/dt in terms of $d/d\phi$ using the chain rule:

$$\frac{d}{dt} = \frac{d\phi}{dt}\frac{d}{d\phi} = \dot{\phi}\frac{d}{d\phi} = \frac{\ell}{\mu r^2}\frac{d}{d\phi} = \frac{\ell u^2}{\mu}\frac{d}{d\phi}. \tag{8.39}$$

(The third equality follows because $\ell = \mu r^2\dot{\phi}$, and the last results from the change of variables $u = 1/r$.)

Using the identity (8.39) we can rewrite $\ddot{r}$ on the left of the radial equation. First

$$\dot{r} = \frac{d}{dt}(r) = \frac{\ell u^2}{\mu}\frac{d}{d\phi}\left(\frac{1}{u}\right) = -\frac{\ell}{\mu}\frac{du}{d\phi}$$

and hence

$$\ddot{r} = \frac{d}{dt}(\dot{r}) = \frac{\ell u^2}{\mu}\frac{d}{d\phi}\left(-\frac{\ell}{\mu}\frac{du}{d\phi}\right) = -\frac{\ell^2 u^2}{\mu^2}\frac{d^2u}{d\phi^2}. \tag{8.40}$$

Substituting back into the radial equation (8.37) we find

$$-\frac{\ell^2 u^2}{\mu}\frac{d^2u}{d\phi^2} = F + \frac{\ell^2 u^3}{\mu}$$

or

$$u''(\phi) = -u(\phi) - \frac{\mu}{\ell^2 u(\phi)^2}F. \tag{8.41}$$

For any given central force F, this transformed radial equation is a differential equation for the new variable $u(\phi)$. If we can solve it, then we can immediately write down r as $r = 1/u$. In the next section, we shall solve it for the case of an inverse-square force and show that the resulting orbits are conic sections, that is, ellipses, parabolas, or hyperbolas. First, here is a simpler example.

EXAMPLE 8.3 The Radial Equation for a Free Particle

Solve the transformed radial equation (8.41) for a *free* particle (that is, a particle subject to no forces) and confirm that the resulting orbit is the expected straight line.

This example is probably one of the hardest ways of showing that a free particle moves along a straight line. Nevertheless, it is a nice check that the transformed radial equation makes sense. In the absence of forces, (8.41) is just

$$u''(\phi) = -u(\phi)$$

whose general solution we know to be

$$u(\phi) = A\cos(\phi - \delta), \tag{8.42}$$

where A and δ are arbitrary constants. Therefore, (renaming the constant $A = 1/r_o$)

$$r(\phi) = \frac{1}{u(\phi)} = \frac{r_o}{\cos(\phi - \delta)}. \tag{8.43}$$

This unpromising-looking equation is in fact the equation of a straight line in polar coordinates, as you can see from Figure 8.8. In that picture Q is a fixed point with polar coordinates (r_o, δ), and the line in question is the line through Q perpendicular to OQ. It is easy to see that the point P with polar coordinates (r, ϕ) lies on this line if and only if $r\cos(\phi - \delta) = r_o$. In other words, Equation (8.43) is the equation of this straight line.

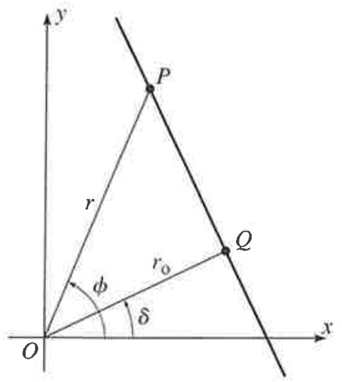


Figure 8.8 The fixed point Q has polar coordinates (r_o, δ) relative to the origin O. The point P with polar coordinates (r, ϕ) lies on the line through Q perpendicular to OQ if and only if $r\cos(\phi - \delta) = r_o$. That is, the equation of this line is (8.43).

In the next section, I shall use the same transformed radial equation (8.41) to solve a much less trivial problem, finding the path of a comet or any other body held in orbit by an inverse-square force.

8.6 The Kepler Orbits

Let us now return to the Kepler problem, the problem of finding the possible orbits of a comet or any other object subject to an inverse-square force. The two important examples of this problem are the motion of comets or planets around the sun (or earth satellites around the earth), in which case the force is the gravitational force $-Gm_1m_2/r^2$, and the orbital motion of two opposite charges q_1 and q_2, in which case the force is the Coulomb force kq_1q_2/r^2. To include both cases and to simplify the equations, I shall write the force as (remember that $u = 1/r$)

$$F(r) = -\frac{\gamma}{r^2} = -\gamma u^2, \tag{8.44}$$

where γ is the "force constant," equal to Gm_1m_2 in the gravitational case.[4]

Thanks to our elaborate preparations, we can now solve the main problem very easily. Inserting the force (8.44) into the transformed radial equation (8.41), we find that $u(\phi)$ must satisfy

$$u''(\phi) = -u(\phi) + \gamma\mu/\ell^2. \tag{8.45}$$

Notice that it is a unique feature of the inverse-square force that the last term in this equation is a constant, since only in this case does the u^2 of the force cancel the $1/u^2$ in (8.41). Because this last term is constant, we can solve (8.45) very easily: If we substitute

$$w(\phi) = u(\phi) - \gamma\mu/\ell^2,$$

the equation becomes

$$w''(\phi) = -w(\phi),$$

which has the general solution

$$w(\phi) = A\cos(\phi - \delta), \tag{8.46}$$

where A is a positive constant and δ is a constant that we can take to be zero by a suitable choice of the direction $\phi = 0$. Thus the general solution for $u(\phi)$ can be written as

$$u(\phi) = \frac{\gamma\mu}{\ell^2} + A\cos\phi = \frac{\gamma\mu}{\ell^2}(1 + \epsilon\cos\phi) \tag{8.47}$$

[4] The constant γ is positive for the gravitational force and for the force between two opposite charges. As discussed in Problem 8.31, for two charges of the same sign, γ is negative. For now, we'll assume it is positive.

where ϵ is just a new name for the dimensionless positive constant $A\ell^2/\gamma\mu$. Since $u = 1/r$, the constant $\gamma\mu/\ell^2$ on the right has the dimensions [1/length], and I shall introduce the length

$$c = \frac{\ell^2}{\gamma\mu} \tag{8.48}$$

in terms of which our solution becomes

$$\frac{1}{r(\phi)} = \frac{1}{c}(1 + \epsilon\cos\phi)$$

or

$$r(\phi) = \frac{c}{1 + \epsilon\cos\phi}. \tag{8.49}$$

This is our solution for r as a function of ϕ, in terms of the undetermined positive constant ϵ and the length $c = \ell^2/\gamma\mu$ (which is $\ell^2/Gm_1m_2\mu$ in the gravitational problem). I shall now explore its properties, first for the bounded orbits and then for the unbounded.

The Bounded Orbits

The behavior of the orbit $r(\phi)$ in (8.49) is controlled by the as-yet undetermined positive constant ϵ. A glance at (8.49) shows this behavior is very different according as $\epsilon < 1$ or $\epsilon \geq 1$. If $\epsilon < 1$, the denominator of (8.49) never vanishes, and $r(\phi)$ remains bounded for all ϕ. If $\epsilon \geq 1$ the denominator vanishes at some angle, and $r(\phi)$ approaches infinity as ϕ approaches that angle. Evidently the value $\epsilon = 1$ is the boundary between the bounded and unbounded orbits. I shall show shortly that this boundary corresponds exactly to the boundary between $E < 0$ and $E \geq 0$ discussed before. Meanwhile, let us start with the case that the constant ϵ is less than 1. With $\epsilon < 1$, the denominator of $r(\phi)$ in (8.49) oscillates as shown in Figure 8.9 between the values $1 \pm \epsilon$. Therefore, $r(\phi)$ oscillates between

$$r_{\text{min}} = \frac{c}{1+\epsilon} \quad \text{and} \quad r_{\text{max}} = \frac{c}{1-\epsilon} \tag{8.50}$$

with $r = r_{\text{min}}$ at the so-called **perihelion** when $\phi = 0$, and $r = r_{\text{max}}$ at the **aphelion** when $\phi = \pi$. Since $r(\phi)$ is obviously periodic in ϕ with period 2π, it follows that $r(2\pi) = r(0)$ and the orbit closes on itself after just one revolution. Thus the general appearance of the orbit is as in Figure 8.10.

While the orbit shown in Figure 8.10 certainly *looks* like an ellipse, I have not yet proved that it really is. However, it is a reasonably easy exercise (see Problem 8.16) to rewrite (8.49) in Cartesian coordinates and cast it in the form

$$\frac{(x+d)^2}{a^2} + \frac{y^2}{b^2} = 1 \tag{8.51}$$

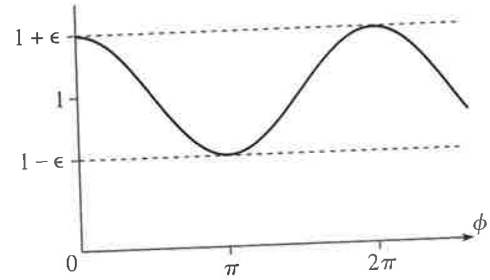


Figure 8.9 The denominator $1 + \epsilon \cos \phi$ in Equation (8.49) for $r(\phi)$ oscillates between $1+\epsilon$ and $1-\epsilon$, and is periodic with period 2π.

where (as you can easily check)

$$a = \frac{c}{1-\epsilon^2}, \qquad b = \frac{c}{\sqrt{1-\epsilon^2}}, \qquad \text{and} \qquad d = a\epsilon. \tag{8.52}$$

Equation (8.51) is the standard equation of an ellipse with semimajor and semiminor axes a and b, except that where we expect to see x we have $x + d$. This difference reflects that our origin, the sun, is not at the center of the ellipse, but at a distance d from it, as shown in Figure 8.10.

We can now identify the constant ϵ, which started life as an undetermined constant of integration in (8.47). According to (8.52) the ratio of the major to minor axes is

$$\frac{b}{a} = \sqrt{1-\epsilon^2}. \tag{8.53}$$

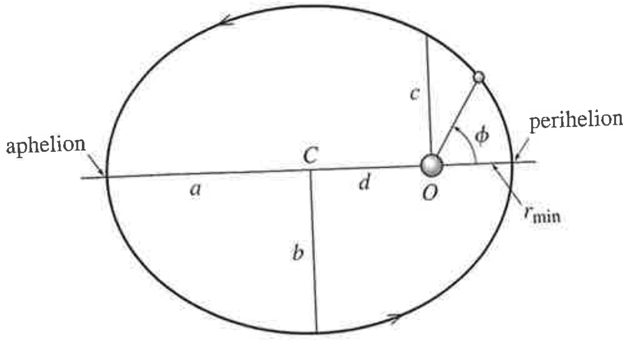


Figure 8.10 The bounded orbits of a comet or planet as given by Equation (8.49) are ellipses. The sun is at the origin O, which is one focus of the ellipse (*not* the center C). The distances a and b are called the semimajor and semiminor axes. The parameter $c = \ell^2/\gamma\mu$ introduced in (8.48) is the value of r when $\phi = 90°$. The points where the comet is closest and farthest from the sun are called the perihelion and aphelion.

Although you almost certainly don't remember it, this equation is the definition (or one possible definition) of the eccentricity of the ellipse. That is, this equation tells us that the constant ϵ is the eccentricity. Notice that if $\epsilon = 0$, then $b = a$ and the ellipse is a circle; if $\epsilon \to 1$, then $b/a \to 0$ and the ellipse becomes very thin and elongated.

Having identified the constant ϵ as the eccentricity, we can now identify the position of the sun in relation to the ellipse. According to (8.52) the distance from the center C to the sun at O is $d = a\epsilon$, and (though again you may not remember it) $a\epsilon$ is the distance from the center to either focus of the ellipse. Thus the position of the sun is actually one of the ellipse's two focuses, and we have now proved **Kepler's first law**, that the planets (and comets whose orbits are bounded) follow orbits that are ellipses with the sun at one focus.

EXAMPLE 8.4 Halley's Comet

Halley's comet, named for the English astronomer Edmund Halley (1656–1742), follows a very eccentric orbit with $\epsilon = 0.967$. At closest approach (the perihelion) the comet is 0.59 AU from the sun, fairly close to the orbit of Mercury. (The AU or astronomical unit is the mean distance of the earth from the sun, about 1.5×10^8 km.) What is the comet's greatest distance from the sun, that is, the distance of the aphelion?

The given distance is $r_{\text{min}} = 0.59$ AU, and, according to (8.50), $r_{\text{max}}/r_{\text{min}} = (1+\epsilon)/(1-\epsilon)$. Therefore

$$r_{\text{max}} = \frac{1+\epsilon}{1-\epsilon} r_{\text{min}} = \frac{1.967}{0.033} r_{\text{min}} = 60\, r_{\text{min}} = 35 \text{ AU}.$$

This means that at its greatest distance Halley's comet is outside the orbit of Neptune.

The Orbital Period; Kepler's Third Law

We can now find the period of the elliptical orbits of the comets and planets. According to Kepler's second law (Section 3.4), the rate at which a line from the sun to a comet or planet sweeps out area is

$$\frac{dA}{dt} = \frac{\ell}{2\mu}.$$

Since the total area of an ellipse is $A = \pi ab$, the period is

$$\tau = \frac{A}{dA/dt} = \frac{2\pi ab\mu}{\ell}.$$

If we square both sides and use (8.53) to replace b^2 by $a^2(1-\epsilon^2)$, this becomes

$$\tau^2 = 4\pi^2 \frac{a^4(1-\epsilon^2)\mu^2}{\ell^2} = 4\pi^2 \frac{a^3 c\mu^2}{\ell^2},$$

where in the last equality I used (8.52) to replace $a(1-\epsilon^2)$ by c. Since the length c was defined in (8.48) as $\ell^2/\gamma\mu$, this implies that

$$\tau^2 = 4\pi^2 \frac{a^3\mu}{\gamma}. \tag{8.54}$$

Finally, γ is the constant in the inverse-square force law $F = -\gamma/r^2$, and, for the gravitational force, $\gamma = Gm_1m_2 = G\mu M$ where M is the total mass, $M = m_1 + m_2$. (Notice the handy identity that $m_1m_2 = \mu M$.) In our case $m_2 = M_s$, the mass of the sun, which is very much greater than m_1, the mass of the comet or planet. Thus, to an excellent approximation, $M \approx M_s$, and

$$\gamma = Gm_1m_2 \approx G\mu M_s.$$

Therefore, the factor of μ in (8.54) cancels, and we find that

$$\tau^2 = \frac{4\pi^2}{GM_s} a^3. \tag{8.55}$$

This is **Kepler's third law**: Because the mass of the comet (or planet) has canceled out, the law says that for all bodies orbiting the sun, the square of the period is proportional to the cube of the semimajor axis. (For circular orbits, we can replace a^3 by r^3.) The law applies equally to the satellites of any large body. For example, all satellites of the earth, including the moon, obey the same law [with M_s replaced by the earth's mass M_e in (8.55)], and the same applies to all the moons of Jupiter.

EXAMPLE 8.5 Period of a Low-Orbit Earth Satellite

Use Kepler's third law to estimate the period of a satellite in a circular orbit a few tens of miles above the earth's surface.

The period is given by (8.55) with M_s replaced by M_e. Since the orbit is circular, we can replace a by r, and since the orbit is close to the earth's surface, $r \approx R_e$, the radius of the earth. Therefore

$$\tau^2 = \frac{4\pi^2}{GM_e} R_e{}^3.$$

This simplifies if we recall that $GM_e/R_e{}^2 = g$, the acceleration of gravity on the earth's surface, and we find that

$$\tau = 2\pi\sqrt{\frac{R_e}{g}} = 2\pi\sqrt{\frac{6.38 \times 10^6 \text{ m}}{9.8 \text{ m/s}^2}} = 5070 \text{ s} \approx 85 \text{ min}, \tag{8.56}$$

in agreement with the well-known observation that low-orbit satellites circle the earth in about one and a half hours.

Relation between Energy and Eccentricity

Finally, we can relate the eccentricity ϵ of the orbit to the energy E of the comet or other orbiting body. The simplest way to do this is to remember that, at its distance of closest approach r_{min}, the comet's energy is equal to the effective potential energy U_{eff} [Equation (8.36)],

$$\begin{aligned} E = U_{\text{eff}}(r_{\text{min}}) &= -\frac{\gamma}{r_{\text{min}}} + \frac{\ell^2}{2\mu r_{\text{min}}{}^2} \\ &= \frac{1}{2r_{\text{min}}}\left(\frac{\ell^2}{\mu r_{\text{min}}} - 2\gamma\right). \end{aligned} \tag{8.57}$$

Now we know from (8.50) that $r_{\text{min}} = c/(1+\epsilon)$, and from its definition (8.48) that $c = \ell^2/\gamma\mu$. Therefore

$$r_{\text{min}} = \frac{\ell^2}{\gamma\mu(1+\epsilon)}$$

and, substituting into (8.57),

$$\begin{aligned} E &= \frac{\gamma\mu(1+\epsilon)}{2\ell^2}[\gamma(1+\epsilon) - 2\gamma] \\ &= \frac{\gamma^2\mu}{2\ell^2}(\epsilon^2 - 1). \end{aligned} \tag{8.58}$$

The calculations leading to (8.58) are equally valid for bounded and unbounded orbits, and they imply the following expected correlations: Negative energies ($E < 0$) correspond to eccentricities $\epsilon < 1$, which in turn correspond to bounded orbits. Positive energies ($E > 0$) correspond to eccentricities $\epsilon > 1$, which in turn correspond to unbounded orbits. Equation (8.58) is a useful relation between the mechanical properties E and ℓ and the geometrical property ϵ. It implies some interesting connections. For example, for a given value of the angular momentum ℓ, the orbit of lowest possible energy is the circular orbit with $\epsilon = 0$ (a connection which has an important counterpart in quantum mechanics).

8.7 The Unbounded Kepler Orbits

In the previous section, we found the general Kepler orbit, as given by (8.49),

$$r(\phi) = \frac{c}{1 + \epsilon\cos\phi}, \tag{8.59}$$

and examined in detail the bounded orbits — those for which $\epsilon < 1$ or, equivalently, as we have seen, $E < 0$. In this section, I shall sketch the corresponding analysis of the unbounded orbits, with $\epsilon \geq 1$ and $E \geq 0$.

The boundary between the bounded and unbounded orbits comes when $\epsilon = 1$ or $E = 0$. With $\epsilon = 1$, the denominator of (8.59) vanishes when $\phi = \pm\pi$. Therefore,

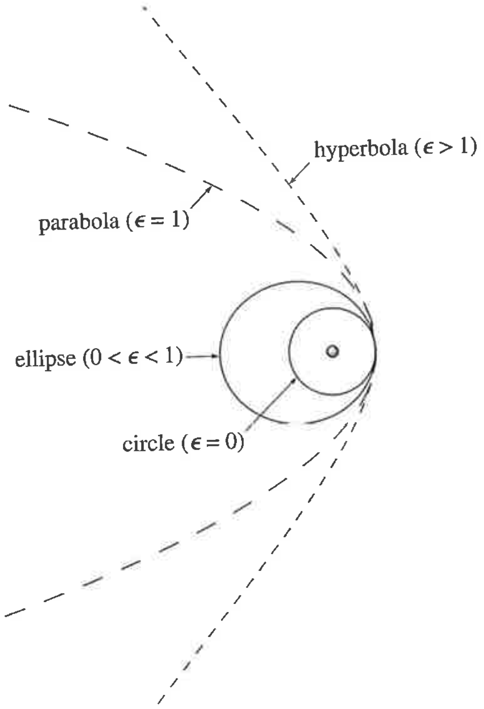


Figure 8.11 Four different Kepler orbits for a comet: a circle, an ellipse, a parabola, and a hyperbola. For clarity, the four orbits were chosen with the same values for $r_{\rm min}$ and with the closest approaches all in the same direction.

$r(\phi) \to \infty$ as $\phi \to \pm\pi$. That is, if $\epsilon = 1$, the orbit is unbounded and goes off to infinity as the comet approaches $\phi = \pm\pi$. Some elementary algebra, parallel to what led to (8.51), shows that with $\epsilon = 1$ the Cartesian version of (8.59) is

$$y^2 = c^2 - 2cx \tag{8.60}$$

which is the equation of a parabola. This orbit is shown (with the long dashes) in Figure 8.11.

If $\epsilon > 1$ (or $E > 0$), the denominator of (8.59) vanishes at a value $\phi_{\rm max}$ determined by the condition

$$\epsilon \cos(\phi_{\rm max}) = -1.$$

Thus $r(\phi) \to \infty$ when $\phi \to \pm\phi_{\rm max}$ and the orbit is confined to the range of angles $-\phi_{\rm max} < \phi < \phi_{\rm max}$. This gives the orbit the general appearance sketched in Figure 8.6. With $\epsilon > 1$ the Cartesian form of (8.59) is (Problem 8.30)

$$\frac{(x-\delta)^2}{\alpha^2} - \frac{y^2}{\beta^2} = 1, \tag{8.61}$$

where you can easily identify the constants α, β, and δ (Problem 8.30). This is the equation of a hyperbola, and we have proved that, as anticipated, the positive energy Kepler orbits are hyperbolas. One such orbit is shown (with the smaller dashes) in Figure 8.11.

Summary of Kepler Orbits

Our results for the Kepler orbits can be summarized as follows: All of the possible orbits are given by Equation (8.59),

$$r(\phi) = \frac{c}{1 + \epsilon \cos \phi}, \tag{8.62}$$

and are characterized by the two constants of integration[5] ϵ and c. The dimensionless constant ϵ is related to the comet's energy by (8.58),

$$E = \frac{\gamma^2 \mu}{2\ell^2}(\epsilon^2 - 1). \tag{8.63}$$

It is, as we have seen, the eccentricity of the orbit that determines the orbit's shape as follows:

eccentricity	energy	orbit
$\epsilon = 0$	$E < 0$	circle
$0 < \epsilon < 1$	$E < 0$	ellipse
$\epsilon = 1$	$E = 0$	parabola
$\epsilon > 1$	$E > 0$	hyperbola

You can see from (8.62) that the constant c is a scale factor that determines the size of the orbit. It has the dimensions of length and is the distance from sun to comet when $\phi = \pi/2$. It is equal to $\ell^2/\gamma\mu$ or, since γ is the force constant Gm_1m_2,

$$c = \frac{\ell^2}{Gm_1m_2\mu}, \tag{8.64}$$

where m_1 is the mass of the comet, m_2 that of the sun, and μ is the reduced mass $\mu = m_1m_2/(m_1 + m_2)$, which is exceedingly close to m_1 since m_2 is so large.

8.8 Changes of Orbit

In this final section, I shall discuss how a satellite can change from one orbit to another. For example, a spacecraft wishing to visit Venus may want to transfer from a circular

[5] Since Newton's second law is a second-order differential equation and the motion is in two dimensions, there are actually four constants of integration in all. The third is the constant δ in (8.46) which we chose to be zero, forcing the axis of the orbit to be the x axis. The fourth is the comet's position on the orbit at time $t = 0$.

Chapter 9

Conserved quantities

Earlier in the course when we encountered conserved quantities, I advertised that they were always due to a continuous symmetry in the system – a result due to the great Emmy Noether and known as Noether's theorem. Now that we have Lagrangian mechanics, we can see how it works explicitly.

Before doing anything by way of proof, recall PCQ 7.3. Go back and look at it again. Consider a one degree of freedom problem with a generalized coordinate z. Hopefully you concluded that when the Lagrangian is independent of z, then p_z is conserved. This can be seen from the Euler-Lagrange equations:

$$\frac{\partial L}{\partial z} = \frac{d}{dt}\frac{\partial L}{\partial \dot{z}}. \tag{9.1}$$

The quantity $\frac{\partial L}{\partial z}$ is zero when L is independent of z. We then have $\frac{d}{dt}\frac{\partial L}{\partial \dot{z}} = 0$, which means $\frac{\partial L}{\partial \dot{z}}$ must be a constant, something independent of time. This thing is the generalized momentum, which in simple cases is the ordinary momentum. So we see that a generalized momentum is conserved when the Lagrangian is independent of the corresponding generalized coordinate.

When the Lagrangian is independent of z, the system is said to have a "symmetry". You can shift the whole system by some amount ϵ and nothing changes. Physically, this means pick up the system, move it a distance ϵ and repeat whatever experiment you did to see if anything changes. Mathematically, it means replace z with $z + \epsilon$ and note that you still have exactly the same Lagrangian when z does not appear in it. Note that $\dot{z}$ may appear in the Lagrangian, but it won't change under the transformation we're considering here. So

$$L = \tfrac{1}{2}m\dot{z}^2 \tag{9.2}$$

is invariant under translations in z, and $m\dot{z}$ will be a conserved quantity, while

$$L = \tfrac{1}{2}m\dot{z}^2 - \tfrac{1}{2}kz^2 \tag{9.3}$$

is not invariant and $m\dot{z}$ is not conserved.

PCQ 9.1

Why physically is $m\dot{z}$ not conserved in the second case above?

Let me now prove a version of Noether's theorem. Noether's theorem can be proven in a variety of degrees of generality. I'll do it for a case of minimal generality that I think will serve best as an illustration of the point. I'll then comment on some generalizations. Before we begin, recall from the definition of a derivative we have

$$\frac{f[x+\epsilon] - f[x]}{\epsilon} \approx \frac{df[x]}{dx}, \tag{9.4}$$

where approximately would become exact as ϵ goes to zero. Similarly, if we have a function of multiple variables, we have

$$\frac{f[x+\epsilon, y] - f[x, y]}{\epsilon} \approx \frac{\partial f[x]}{\partial x}. \tag{9.5}$$

Ok, on with the proof. Consider a general Lagrangian involving some set of generalized coordinates where we call an arbitrary generalized coordinate q_n. Now consider a transformation on the system of the form

$$q_n \to q_n + \epsilon f_n[t], \tag{9.6}$$

which we'll assume is a symmetry transformation, one that does not change the Lagrangian. Here ϵ is a small quantity and $f_n[t]$ is a function of time. Note that our simple example above would fit in here with q_n being z and $f_n[t]$ being 1. This implies that the generalized velocities transform as follows:

$$\dot{q}_n \to \dot{q}_n + \epsilon \dot{f}_n[t]. \tag{9.7}$$

Now consider the difference between the new Lagrangian and the old one

$$\Delta L = L[q_n + \epsilon f_n, \dot{q}_n + \epsilon \dot{f}_n, t] - L[q_n, \dot{q}_n, t] = \sum_n \left(\frac{\partial L}{\partial q_n} \epsilon f_n + \frac{\partial L}{\partial \dot{q}_n} \epsilon \dot{f}_n] \right) = 0. \tag{9.8}$$

This equals zero, since we've assumed it's a symmetry. The second to last equality makes use of the definition of a derivative with infinitesimal ϵ. Using the equation of motion, we can replace $\frac{\partial L}{\partial q_n}$ with $\frac{d}{dt}\frac{\partial L}{\partial \dot{q}_n}$, yielding

$$\sum_n \left(\frac{d}{dt}\frac{\partial L}{\partial \dot{q}_n} \epsilon f_n + \frac{\partial L}{\partial \dot{q}_n} \epsilon \dot{f}_n] \right) = 0. \tag{9.9}$$

Now pull out a time derivative and an ϵ to get

$$\epsilon \frac{d}{dt} \left(\sum_n \frac{\partial L}{\partial \dot{q}_n} f_n \right) = 0. \tag{9.10}$$

If the left hand side is really equal to zero, then we need either $\epsilon = 0$ or

$$\sum_n \frac{\partial L}{\partial \dot{q}_n} f_n = \text{constant}. \tag{9.11}$$

Clearly we want the latter since if $\epsilon = 0$, we didn't make a transformation, which would be in conflict with the problem we originally set up. So we've concluded that when we have a symmetry, that is $\Delta L = 0$, some stuff is constant, in other words, conserved.

Example 1: consider a Lagrangian for N interacting particles confined to 1 dimension, where the particles interact via a potential energy function U that depends only on the differences in particle positions:

$$L = \sum_n^N \tfrac{1}{2} m_n \dot{q}_n - U[q_1 - q_2, \ldots, q_1 - q_n, q_2 - q_3, \ldots]. \tag{9.12}$$

This Lagrangian will be unchanged under the transformation

$$q_n \to q_n + \epsilon \tag{9.13}$$

and we would conclude from Noether's theorem that $\sum_n \frac{\partial L}{\partial \dot{q}_n}$ is conserved, that is the total momentum of this system is conserved as expected since it does not interact with the outside world. We have only internal forces here.

Example 2: consider a free particle and the transformation

$$\vec{r} \to \vec{r} + \epsilon \hat{n} \times \vec{r}. \tag{9.14}$$

This amounts to a rotation by an angle ϵ about an axis $\hat{n}$. The Lagrangian will be invariant for infinitesimal ϵ and the conserved quantity will be $\sum_{i=x,y,z} \frac{\partial L}{\partial \dot{r}_i} (\hat{n} \times \vec{r})_i$, which is $m\dot{\vec{r}} \cdot (\hat{n} \times \vec{r})$ or after the bac-cab rule $\hat{n} \cdot (\vec{r} \times m\dot{\vec{r}})$. This is the angular momentum about the $\hat{n}$ axis.

So I've proved and used a version of Noether's theorem, but there are more general versions. I assumed a particular type of symmetry transformation. I could generalize this. I also took a stronger than needed definition of a symmetry – no change in L when we make the transformation. It turns out that can weaken this and still have symmetry, an easy example would be if the new L differed from the old by a constant. Since the Euler-Lagrange equations involve only derivatives of L, then the system would still be unchanged. So I could also have proved more general versions of Noether's theorem by considering these more general notions of invariance. Finally, everything we're doing in this class with point particles can be done with fields. For example, Maxwell's Equations are equations of motion that can be derived from a Lagrangian involving electric and magnetic fields. There is a generalized version of Noether's theorem that works for fields too.

Chapter 10

Hamiltonian Mechanics

Hamiltonian mechanics provides a third formulation of mechanics. It feels similar in spirit to the the Lagrangian formulation, but makes use of different variables. The method has a variety of advantages in more advanced treatments of mechanics, and provides a nice connection to the Schrödinger/Heisenberg formulation of quantum mechanics, which is how you'll likely first meet the subject of quantum mechanics. (As we noted before, there is a path-integral formulation of quantum mechanics that makes use of the Lagrangian, but this formulation is less commonly used.) Although many of the advantages of the Hamiltonian formulation don't really shine in the types of problems we consider in the course, it's important to understand the method here, where you'll at least appreciate its use as another approach along side Newtonian and Lagrangian methods. I'll also describe some of the advantages, and use it as another chance to explore conserved quantities.

10.1 Basics

The Hamiltonian is defined from the Lagrangian via:

$$H[q_1, q_2, \ldots, q_n, p_1, p_2, \ldots, p_n, t] = \sum_j \dot{q}_j p_j - L[q_1, q_2, \ldots, q_n, \dot{q}_1, \dot{q}_2, \ldots, \dot{q}_n, t], \tag{10.1}$$

where p_j are the generalized momenta defined as before:

$$p_k \equiv \frac{\partial L}{\partial \dot{q}_k}. \tag{10.2}$$

As I've displayed explicitly, the Hamiltonian will in general be a function of n generalized coordinates, which I've given the generic labels $q_1 \ldots q_n$, the corresponding generalized momenta $p_1 \ldots p_n$, and possibly time. So the Hamiltonian formulation could be viewed as a fancy change of variables: the function L is exchanged for H and the generalized velocities are exchanged for generalized momenta. This type of transformation is an example of something called a Legendre transformation, which pops up other places in physics such as thermodynamics, for example. If you're interested in exploring the general idea of a Legendre transformation further, see, for example, Goldstein's Hamiltonian chapter. The easiest way to obtain the Hamiltonian is to follow steps 1-5 of the Lagrangian procedure to get the Lagrangian, then apply Eq. (10.2) to get the generalized momenta, plug the Lagrangian and the generalized momenta into Eq. (10.1), and use the results of Eq. (10.2) to remove the generalized velocities in favor of the generalized momenta.

Once the Hamiltonian has been obtained, Hamilton's equations of motion can be used to get differential equations that can be solved for the motion of the system. Hamilton's equations take the form

$$\dot{q}_k = \frac{\partial H}{\partial p_k} \tag{10.3}$$

$$\dot{p}_k = -\frac{\partial H}{\partial q_k}. \tag{10.4}$$

These equations can be solved directly to obtain the motion of the system. It is also typically possible to verify in a given problem that they yield information equivalent to the Newtonian or Lagrangian formulation by time differentiating the first equation and plugging it into the second.

Lets look at a really simple example that illustrates all of the points made above. Consider what is perhaps the simplest example that we've considered many times: a particle of mass m falls from rest under gravity near Earth from an initial height z_0. We'll use one generalized coordinate z, which measures distance up from the surface of Earth. Under these conditions we have

$$L = \tfrac{1}{2}m\dot{z}^2 - mgz. \tag{10.5}$$

We can then apply Eq. (10.2) to get the generalized momentum. As we've seen before in this simple case, it's just the z component of the ordinary old momentum:

$$p = m\dot{z}. \tag{10.6}$$

Now we can plug L and p into Eq. (10.1), where we'll find

$$H = \tfrac{1}{2}m\dot{z}^2 + mgz. \tag{10.7}$$

PCQ 10.1

Check that plugging p and L into Eq. (10.1) indeed yields Eq. (10.7).

Note that the Hamiltonian we find here is just the total energy $T + U$.

Warning #1: the fact that we found $H = T + U$ here is a special feature of certain systems, so don't assume it will be true. Work out H via the procedure above unless you've verified in advance that you have a system where $H = T + U$. We'll explore the requirements for this later in the chapter.

Warning #2: We are not yet done "finding the Hamiltonian". To be done, we need to express H in terms of the generalized momenta. The Hamiltonian is a function of generalized coordinates and generalized momenta. They are the basic variables. So far, we have an equation for H that contains generalized velocities. Before we really have the Hamiltonian in all its glory, we need to eliminate $\dot{z}$ in favor of p. We can do this by solving Eq. (10.6) for $\dot{z}$ and plugging it into Eq. (10.7). Doing so yields the final form for H:

$$H = \frac{p^2}{2m} + mgz. \tag{10.8}$$

We can now apply Hamilton's equations of motion. With just one degree of freedom, we can suppress the subscript k in Eq. (10.4), note that our generalized coordinate is z, and plug in our H. We find:

$$\dot{z} = \frac{\partial H}{\partial p} = \frac{p}{m} \tag{10.9}$$

$$\dot{p} = -\frac{\partial H}{\partial z} = -mg. \tag{10.10}$$

As usual, the job of our Mechanics Machine is to find later positions given initial positions, so we'll solve these equations for $z[t]$. In this case, we can start by integrating the p equation, which yields:

$$p = p_0 - mgt = -mgt, \tag{10.11}$$

where the second equality follows from the initial condition $p_0 = 0$. We can now plug this into the first equation yielding

$$\dot{z} = \frac{-mgt}{m}. \tag{10.12}$$

Integrating, we find:

$$z[t] = z_0 - \tfrac{1}{2}gt^2, \tag{10.13}$$

i.e. the usual constant acceleration result for position as a function of time.

For more examples of slowly increasing difficulty, you might want to look at examples 13.1-13.4 in Taylor's text and the examples appearing the discussions to follow.

10.2 Origin of Hamiltonian Mechanics

Now that we've seen how Hamiltonian Mechanics works in a simple example, let's look at why/how it's working. We saw that the Euler-Lagrange equations of the Lagrangian formulation followed from the principle of least action. Hamilton's equations follow from the principle of least action too. The easiest way to see it is to use the Euler-Lagrange equations to get the Hamilton equations. For ease of notation, we'll do this here for the case of one generalized coordinate, where Eq. (10.1) reduces to

$$H = \dot{q}p - L, \tag{10.14}$$

and I'll also suppress the functional dependence.

To get the equations, the basic idea will be to write the differential of H in 2 ways and compare them. First, I'll write

$$dH = \frac{\partial H}{\partial q}dq + \frac{\partial H}{\partial p}dp + \frac{\partial H}{\partial t}dt. \tag{10.15}$$

This is just the chain rule for all of the independent variables in H. Now I'll write the differential of H by considering Eq. (10.14):

$$dH = \dot{q}dp + pd\dot{q} - dL, \tag{10.16}$$

but

$$dL = \frac{\partial L}{\partial q}dq + \frac{\partial L}{\partial \dot{q}}d\dot{q} + \frac{\partial L}{\partial t}dt. \tag{10.17}$$

PCQ 10.2

Using the Euler-Lagrange equations, show that Eq. (10.17) can be written

$$dL = \dot{p}dq + pd\dot{q} + \frac{\partial L}{\partial t}dt. \tag{10.18}$$

Substituting into (10.16), we get

$$dH = \dot{q}dp - \dot{p}dq - \frac{\partial L}{\partial t}dt. \tag{10.19}$$

Upon comparing (10.15) and (10.19), we see that the stuff in front of the dp must be equal, and that equality is the first Hamilton equation. Similarly, the stuff in front of the dq must be equal, and that is the second Hamilton equation. To aid your eye, I've repeated those equations below with the terms reordered to match up:

$$\begin{array}{rcl} dH & = & \dfrac{\partial H}{\partial q}dq + \dfrac{\partial H}{\partial p}dp + \dfrac{\partial H}{\partial t}dt, \\ & & \ \ \updownarrow \qquad\quad \updownarrow \\ dH & = & -\dot{p}\ dq + \ \ \dot{q}\ \ dp - \dfrac{\partial L}{\partial t}dt. \end{array}$$

10.3 H as a Conserved Quantity

Recall that

$$\frac{\partial L}{\partial q_k} = 0 \tag{10.20}$$

implied that p_k was conserved. Given our knowledge from special relativity that q is to t as p is to the energy E, we might expect that

$$\frac{\partial L}{\partial t} = 0 \tag{10.21}$$

implies that energy is conserved, and we would be right. Well, it's actually H that's conserved, but H is the total energy in a closed system.

Let's prove it. Again, I'll do so for the case of one generalized coordinate. I want to see that when (10.21) holds,

$$\frac{dH}{dt} = 0, \tag{10.22}$$

which is what is meant by conserved. I'll start with the chain rule:

$$\frac{dL}{dt} = \frac{\partial L}{\partial q}\frac{dq}{dt} + \frac{\partial L}{\partial \dot{q}}\frac{d\dot{q}}{dt} + \frac{\partial L}{\partial t}. \tag{10.23}$$

That is, the first term handles the time dependence that enters L from q, the second term handles the time dependence that enters L from $\dot{q}$, and the last term handles the time dependence which could occur in L explicitly. Reverting to dot notation, this is

$$\frac{dL}{dt} = \frac{\partial L}{\partial q}\dot{q} + \frac{\partial L}{\partial \dot{q}}\ddot{q} + \frac{\partial L}{\partial t}. \tag{10.24}$$

By using the Euler-Lagrange equations as in the PCQ, we can write this as

$$\begin{aligned} \frac{dL}{dt} &= \dot{p}\dot{q} + p\ddot{q} + \frac{\partial L}{\partial t} && (10.25) \\ &= \frac{d}{dt}(p\dot{q}) + \frac{\partial L}{\partial t}. && (10.26) \end{aligned}$$

Rearranging, we have

$$\begin{aligned} 0 &= \frac{d}{dt}(p\dot{q} - L) + \frac{\partial L}{\partial t} && (10.27) \\ &= \frac{dH}{dt} + \frac{\partial L}{\partial t}. && (10.28) \end{aligned}$$

Thus H is conserved when there is no explicit time dependence in L. That is, when the physics does not depend on time, just as momentum was conserved when the physics did not depend on position.

PCQ 10.3

For each of the following Lagrangians, state whether or not $\frac{\partial L}{\partial t} = 0$:

$$\begin{aligned} L &= \tfrac{1}{2}m\dot{z}^2 + mgz && (10.29) \\ L &= \tfrac{1}{2}m\dot{z}^2 + mgz + \alpha t z && (10.30) \\ L &= \tfrac{1}{2}m\dot{z}^2 + mgz + f[t]z, && (10.31) \end{aligned}$$

where α is a positive constant and the function $f[t]$ appearing in the last Lagrangian is $f[t] = \alpha t$.

Perhaps the most typical situation is to have a conserved H that is equal to the total energy. However, I have now told you that 2 scenarios can come up: 1) H might not be conserved, and 2) H might not be the total energy. These are 2 separate issues in the sense that issue 1 may arise with or without issue 2 and visa versa. Lets look at a few examples.

10.3.1 $H = E$, but it's not conserved

Suppose we have a spring and mass system, for which the Lagrangian is

$$L = \tfrac{1}{2}m\dot{x}^2 - \tfrac{1}{2}kx^2, \tag{10.32}$$

but suppose that the spring constant is not a constant but some sort of function of time. That is, $k = k[t]$. The Lagrangian now has explicit time dependence, $\frac{\partial L}{\partial t} = -\frac{1}{2}\frac{\partial k}{\partial t}x^2$, so we expect that H is not conserved. While we're looking at the Lagrangian, lets get Lagrange's equation of motion for comparison with our Hamiltonian results:

$$\ddot{x} = -\frac{k}{m}x. \tag{10.33}$$

Now transform, to get the Hamiltonian:

$$H = \tfrac{1}{2}m\dot{x}^2 + \tfrac{1}{2}k[t]x^2 = \frac{p_x^2}{2m} + \tfrac{1}{2}k[t]x^2. \tag{10.34}$$

Note that this is $T + U$. Hamilton's equations will be

$$\begin{aligned} \dot{x} &= \frac{\partial H}{\partial p_x} = \frac{p_x}{m} \\ \dot{p}_x &= -\frac{\partial H}{\partial x} = -k[t]x. \end{aligned} \tag{10.35}$$

Before I could consider integrating this, I would need to specify $k[t]$, but we can show explicitly in this example (as we did generally in the prior section) that Hamilton's equations contain the same information as Lagrange's equations by differentiating the first equation, yielding

$$\ddot{x} = \frac{\dot{p}_x}{m} \tag{10.36}$$

and plugging in the second for $\dot{p}_x$ on the right to yield Eq. (10.33).

We can show explicitly that H is not conserved. Consider

$$\frac{dH}{dt} = \frac{p_x \dot{p}_x}{m} + k[t]x\dot{x} + \tfrac{1}{2}\dot{k}[t]x^2. \tag{10.37}$$

Plug in Hamilton's equations (10.35) to get

$$\frac{dH}{dt} = \frac{p_x(-k[t]x)}{m} + k[t]x(p_x/m) + \tfrac{1}{2}\dot{k}[t]x^2 = \tfrac{1}{2}\dot{k}[t]x^2. \tag{10.38}$$

So in the usual spring and mass problem with constant k, the we would have only the first 2 terms above, which cancel and H would be conserved. But here, we have the extra remaining term due to the explicit time dependence. Note that the issue is coming up because when we have $k[t]$, the spring potential is not a true potential energy. When we first defined potential energy, we said that it should depend only on coordinates and be path independent. None the less, we can (and sometimes will want to) consider such "psudopotentials" in Lagrangian and Hamiltonian mechanics. Earlier in the course, we looked at the *Physical Review Letter* that considered Newton's constant not being constant. That would also fall in this category.

10.3.2 $H \neq E$ but it is conserved

Consider a wire with one end at the origin that points out radially from the origin in the $x - y$ plane. That is, the wire lies along the r direction in plane polar coordinates. Let a bead of mass m be mounted on the wire such that it's free to slide in and out (toward smaller r or larger r). Let the wire rotate about the origin at constant angular speed ω. There is just one degree of freedom in this problem. The only choice the bead has is to go in or go out, so r is a good choice for the only generalized coordinate. Note also, that there is no potential energy in this problem. The rotating wire is an example of a time-dependent constraint. That is, the transformation equations have time appearing in them explicitly:

$$x = r\cos\omega t \quad y = r\sin\omega t. \tag{10.39}$$

The wire constrains the bead, but this constraint is 'moving' resulting in a time dependence. With this you can show that the Lagrangian takes the form

$$L = \tfrac{1}{2}m(\dot{r}^2 + \omega^2 r^2). \tag{10.40}$$

Note that the Lagrangian is in fact the total energy here, since there is no U to subtract from the T. In other words, $L = T = E$. The Hamiltonian is

$$H = \tfrac{1}{2}m(\dot{r}^2 - \omega^2 r^2) = \frac{p_r^2}{2m} - \tfrac{1}{2}m\omega^2 r^2. \tag{10.41}$$

PCQ 10.4
Check that this is the Hamiltonian.

Note that this is not the total energy. It is one part of the kinetic energy minus another. The issue here has arisen because the time-dependent constraint has altered the form of the kinetic term. When there are no time-dependent constraints, each term in the kinetic energy has a product of 2 generalized velocities. We say that the kinetic energy is "a homogeneous quadratic function of velocities", where homogeneous here means it involves only quadratic terms and no linear or constant terms.

Note that although $H \neq E$, it is still conserved. We know this since $\frac{\partial L}{\partial t} = 0$, and we can show it explicitly. Consider

$$\frac{dH}{dt} = \frac{p_r \dot{p}_r}{m} - m\omega^2 r\dot{r}. \tag{10.42}$$

Hamilton's equations of motion take the form

$$\begin{aligned} \dot{r} &= \frac{\partial H}{\partial p_r} = \frac{p_r}{m} \\ \dot{p}_r &= -\frac{\partial H}{\partial r} = m\omega^2 r. \end{aligned} \tag{10.43}$$

Plugging them into Eq. (10.42) demonstrates that H is constant. We can also show that although H is conserved, the total energy is not. This can be demonstrated mathematically by plugging Hamilton's equations into the derivative of the total energy:

$$\frac{dE}{dt} = \frac{p_r \dot{p}_r}{m} + m\omega^2 r\dot{r} = 2\omega^2 r p_r, \tag{10.44}$$

which is not zero. Physically, it's because whatever motor keeps the wire rotating at ω can do work on the bead.

Although in these 2 examples, a psudopotential led to an H that's not conserved and a time-dependent constraint led to $H \neq E$, either of these conditions could lead to either outcome or both. If there are only true potentials, and there are no time-dependent constraints, we'll find $H = E$ and H is conserved.

To round out this second example, we could also solve the equations of motion for a bead initially at rest at a position r_0. To begin, we can manipulate the left side of the second Hamilton equation with the vdv/dx trick:

$$\frac{dp_r}{dt} = \frac{dp_r}{dr}\frac{dr}{dt} = \frac{p_r}{m}\frac{dp_r}{dr}, \tag{10.45}$$

where for the last equality, I've plugged in the first Hamilton equation for $\dot{r}$. With this, the second Hamilton equation can be written:

$$p_r \frac{dp_r}{dr} = m^2\omega^2 r, \tag{10.46}$$

which can be integrated to yield

$$p_r^2 = m^2\omega^2(r^2 - r_0^2). \tag{10.47}$$

This can now be inserted into the right side of the first Hamilton equation, and we can again separate the variables:

$$\frac{dr}{\omega\sqrt{r^2 - r_0^2}} = dt. \tag{10.48}$$

Integrating, we get

$$\omega t = \ln\left[\frac{r + \sqrt{r^2 - r_0^2}}{r_0}\right], \tag{10.49}$$

which can be solved for r to get

$$r[t] = \tfrac{1}{2} r_0 \left(e^{\omega t} + e^{-\omega t}\right). \tag{10.50}$$

10.4 Hamiltonian Advantages

Here I'll briefly introduce some additional technology related to the Hamiltonian and comment on some its advantages.

10.4.1 phase space

The Hamiltonian is a function of n coordinates and n generalized momenta where n is the number of degrees of freedom. Specifying each of these $2n$ quantities uniquely specifies the state of the system. One can visualize the possible configurations of the system as points in a $2n$ dimensional phase space. As the system evolves from the initial conditions, the momenta and coordinates will change, tracing out a path in phase space.

PCQ 10.5
I work out an example phase space in the notebook phase_space located in the computational resources folder. Work through it and answer the questions contained.

Here I'll briefly mention 2 areas of physics where discussions of phase space come up quite a lot. Earlier in the course, we mentioned the idea of chaotic systems. It turns out that for some systems, tiny differences in initial conditions rapidly turn into large differences in the state of the system. Two otherwise identical such systems started with almost the same initial conditions quickly look very different. The study of these types of systems is a very active research area in physics and math. Phase space plots turn out to be a key tool in exploring the properties of these systems. If you take advanced classical mechanics, you'll likely explore these issues in some detail. Accelerator physics usually involves a lot of particles distributed over neighboring points in phase space. Certain beam properties are often desirable such as specific ranges of momenta and position. Visualizing this information as well as developing results for how the system will evolve is significantly aided by phase space descriptions.

10.4.2 quantum connections

Our most fundamental descriptions of nature are in terms of quantum mechanical theories. This has many implications, but among them is the realization that it's actually only probabilities of events that can be calculated in these theories. The macroscopic objects of classical mechanics are typically made up of many microscopic objects, the correct description of which is quantum, but the quantum properties 'disappear' or become irrelevant as one approaches the macroscopic scale. This is why classical mechanics is still so useful and provides a good match to experiment.

If you took Phys 228, you would have developed a quantum mechanical description of certain phenomena using an object called the wave function $\Psi[x, t]$. If you haven't taken 228, it's ok, just try to get a basic sense of the importance of the Hamiltonian from the general ideas of this section. In your description, the wave function contained the probabilistic information about the state of the system. The time evolution of the system came from the Schrödinger Equation:

$$i\hbar\frac{d\Psi[x,t]}{dt} = \left(-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V[x]\right)\Psi[x,t]. \tag{10.51}$$

The most likely value (technically the expectation value) of some quantity of interest, like say the position of the particle, came from sandwiching an operator between 2 of these wave functions in the following way:

$$\langle x\rangle = \int_{-\infty}^{\infty} \Psi^* x \Psi dx, \tag{10.52}$$

where * is complex conjugate. The momentum operator, the thing you sandwich to get the expectation value of the momentum of the system you learned was $\hat{p} = -i\hbar\frac{\partial}{\partial x}$.

One might expect that there are some connections between classical theories and quantum ones, and the Hamiltonian provide the opportunity to highlight some of them. First, we can identify the thing in the parenthesis on the right hand side of the Schrödinger Equation as the Hamiltonian. If we start with the Hamiltonian for a particle with some potential energy $V[x]$, we would have

$$H = \frac{p^2}{2m} + V[x] \tag{10.53}$$

as we've seen throughout the chapter. If you insert the momentum operator, you should see that you get the thing in parenthesis. So we've already made a connection, half of the Schrödinger Equation is roughly the same

Hamiltonian from classical mechanics. So we could write the Schrödinger Equation as

$$i\hbar \frac{d\Psi[x,t]}{dt} = \hat{H}\Psi[x,t], \tag{10.54}$$

where $\hat{H}$ is the Hamiltonian as a quantum operator. That is, the momentum has been replaced with the momentum operator.

This connection is deeper, but I need to introduce a little more notation to make the point. Suppose you have some function of the dynamical variables $f[q,p]$ in classical mechanics. I don't intend any specific meaning for f, I just mean any old function of the generalized coordinates and velocities that you might want to talk about. Consider the time derivative of that function

$$\frac{df}{dt} = \frac{\partial f}{\partial q}\frac{dq}{dt} + \frac{\partial f}{\partial p}\frac{dp}{dt}. \tag{10.55}$$

Here the time dependence is contained in the dynamical variables only, so we've applied the chain rule. If we insert Hamilton's equations of motion for the $\dot{q}$ and $\dot{p}$, we get

$$\frac{df}{dt} = \frac{\partial f}{\partial q}\frac{\partial H}{\partial p} - \frac{\partial H}{\partial q}\frac{\partial f}{\partial p}. \tag{10.56}$$

Notice the antisymmetric role of H and f on the right side of this equation. There is an operation called a Poisson bracket that is defined by the structure on the right. We could define this operation acting on 2 generic objects, A and B as follows

$$\{A,B\} = \frac{\partial A}{\partial q}\frac{\partial B}{\partial p} - \frac{\partial B}{\partial q}\frac{\partial A}{\partial p}. \tag{10.57}$$

Thus we could rewrite Eq. (10.56) as

$$\frac{df}{dt} = \{f,H\}. \tag{10.58}$$

What we can see from this is that the change in our function of the dynamical variables, f, over time is related deeply to the Hamiltonian. The technical statement is that 'H is the generator of time translations'. That is, plopping H in the Poisson bracket with the quantity of interest tells you how the quantity of interest evolves in time. This might not be surprising if you consider the fact that H is the dynamical object that is conserved as a result of time translation invariance. You can already get a hint of how this is playing out in quantum. We see that the Schrödinger Equation reads, 'the action of H on the wave function is proportional to the time rate of change in the wave function.'

You might complain that I haven't really shown the connection yet between the Poisson bracket of H and the Schrödinger Equation, and you'd be right. To do so, I need a little more advanced perspective on Quantum Mechanics. Rather than describe the probabilistic properties of the system with the wave function, I'll describe them with a vector in a complex space (that is a vector with complex components) called $|\psi\rangle$ and it's complex conjugate transpose $\langle\psi|$. You don't need to worry about why it's complex or what the space looks like, just think of it as a vector with complex components that somehow represents the state of the system. The operators can now be thought of as matrices that act on the state vectors. So the Schrödinger Equation could be written in pretty much the same form:

$$i\hbar \frac{d|\psi\rangle}{dt} = \hat{H}|\psi\rangle. \tag{10.59}$$

You could also write its complex conjugate transpose:

$$-i\hbar \frac{d\langle\psi|}{dt} = \langle\psi|\hat{H}. \tag{10.60}$$

Note that complex conjugation turns i into $-i$ as usual, and transpose changes the order of the matrix multiplication, a fact you might or might not know from linear algebra.

One perspective on how to build a quantum theory from foundational assumptions is to take the classical theory, turn observable quantities, like position and momentum into operators, and replace the Poisson brackets with $1/i\hbar$ times something called a commutator. A commutator of 2 quantum operators is just the difference of the matrix operators in the opposite order:

$$[\hat{A},\hat{B}] = \hat{A}\hat{B} - \hat{B}\hat{A}. \tag{10.61}$$

So in this scheme, we would replace something like $\{f, H\}$ with $\frac{1}{i\hbar}[\hat{f}, \hat{H}]$.

If we look back on Eq. (10.56), the quantum version of this would be to replace the function of the dynamical variables, f, with the expectation value of the corresponding quantum operator $\langle\psi|\hat{f}|\psi\rangle$. One could then use the product rule to expand a time derivative on this object

$$\frac{d}{dt}\langle\psi|\hat{f}|\psi\rangle = \left(\frac{d}{dt}\langle\psi|\right)\hat{f}|\psi\rangle + \langle\psi|\hat{f}\left(\frac{d}{dt}|\psi\rangle\right). \tag{10.62}$$

We can now insert the Schrödinger Equation on each side and turn it into a commutator:

$$\begin{aligned}\frac{d}{dt}\langle\psi|\hat{f}|\psi\rangle &= \left(\frac{d}{dt}\langle\psi|\right)\hat{f}|\psi\rangle + \langle\psi|\hat{f}\left(\frac{d}{dt}|\psi\rangle\right)\\ &= -\frac{1}{i\hbar}\langle\psi|\hat{H}\hat{f}|\psi\rangle + \frac{1}{i\hbar}\langle\psi|\hat{f}\hat{H}|\psi\rangle\\ &= \frac{1}{i\hbar}\langle\psi|(\hat{f}\hat{H} - \hat{H}\hat{f})|\psi\rangle\\ &= \frac{1}{i\hbar}\langle\psi|[\hat{f}, \hat{H}]|\psi\rangle.\end{aligned} \tag{10.63}$$

Note that this has the same structure as Eq. (10.58) except that we're considering the expectation value of $\hat{f}$ instead of the classical function, and the Poisson bracket has been replaced with $1/i\hbar$ times the commutator. These are the 2 foundational assumptions I said I was going to make in constructing the quantum theory from the classical one. So we see that the role played by the Hamiltonian in the Schrödinger equation is the same as the role it plays in classical mechanics, that of the generator of time translations.

10.4.3 canonical transformations

As we've seen many times, a change of coordinates can make a problem much easier. In Lagrangian mechanics, one can change to different generalized coordinates such that the new coordinates are mixtures of the old. Hamiltonian mechanics admits an even more flexible set of transformations. The Hamiltonian is invariant under certain transformations that mix the generalized coordinates and generalized momenta called canonical transformations. These are transformations in the $2n$ dimensional phase space. An analogous thing is not possible in Lagrangian mechanics because the generalized velocities are related directly to the coordinates via a time derivative. Though application of this method would take us beyond our current scope, I mention it as further motivation for consideration of the Hamiltonian approach along side Newton and Lagrange.

Chapter 11

Harmonic Oscillator

The harmonic oscillator is a system in which a restoring force pulls a particle back to a stable equilibrium point, and that force is proportional to the displacement from equilibrium. The spring and mass system is the classic example:

$$\vec{F} = -k\vec{x}. \tag{11.1}$$

The system is called the simple harmonic oscillator when this is the only force on the particle. In this chapter, we'll take a look at solutions to the simple harmonic oscillator and systems that can be approximated as the simple harmonic oscillator. We'll then consider the addition of drag in the harmonic oscillator, called a damped harmonic oscillator, and the further addition of a driving force that adds energy to the system, which results in the damped, driven harmonic oscillator. All of these cases are of great importance because large numbers of systems from the Tacoma narrows bridge to atomic bonds to electric circuits can be modeled in this way.

11.1 Simple Harmonic Oscillator

Consider a spring and mass system governed as usual by Newton's second law:

$$m\ddot{x} = -kx, \tag{11.2}$$

or rearranged as

$$\ddot{x} = -\omega^2 x, \tag{11.3}$$

where $\omega^2 = k/m$. If you ask yourself what the solution to this equation is, you'll conclude that you need a function, which when differentiated twice gives itself back. Trig functions do this. It is a second order differential equation, so it will involve 2 undetermined integration constants that can be interpreted as being associated with the 2 initial conditions as usual. You might also note that either sin or cos will work. So the general solution can be written

$$x[t] = A\sin\omega t + B\cos\omega t. \tag{11.4}$$

PCQ 11.1

Check that Eq. (11.4) is indeed a solution to ODE (11.3) by plugging it in on both sides and getting a true statement.

If we consider the position at $t = 0$, we find

$$x[0] = B. \tag{11.5}$$

That is, we find that our arbitrary constant B is actually the initial position of the particle at the end of the spring. So B could be renamed x_0. If we take the derivative of Eq. (11.4), we get the velocity of the particle at the end of the spring

$$v[t] = A\omega\cos\omega t - B\omega\sin\omega t. \tag{11.6}$$

If we consider the velocity at time $t = 0$, we find the initial velocity is

$$v[0] = A\omega. \tag{11.7}$$

So with this we could replace the arbitrary constant A with v_0/ω. Hence the whole solution could be written

$$x[t] = \frac{v_0}{\omega}\sin\omega t + x_0\cos\omega t. \tag{11.8}$$

This is not the only way to write the solution to ODE (11.3). It turns out that the following forms also yield a true statement when plugged into (11.3) on both sides.

$$x[t] = Ce^{i\omega t} + De^{-i\omega t} \tag{11.9}$$
$$x[t] = E\cos[\omega t + \delta] \tag{11.10}$$
$$x[t] = F\sin[\omega t + \phi] \tag{11.11}$$

Each of these forms is equivalent to Eq. (11.4). The objects C, D, E, F, δ, ϕ are integration constants which can be identified with the initial conditions just as we did with A and B.

PCQ 11.2

Pick at least one of the above forms and verify that it is also a solution to Eq. (11.3) by plugging it in.

As another exercise, lets see how Eq. (11.10) is related to Eq. (11.4). We'll begin by applying a trig identity to Eq. (11.10).

$$\begin{aligned} x[t] &= E\cos[\omega t + \delta] \\ &= E\cos\delta\cos\omega t - E\sin\delta\sin\omega t \\ &= A\cos\omega t + B\sin\omega t \end{aligned} \tag{11.12}$$

In the last line, we just note that $E\cos\delta$ is one undetermined constant and $-E\sin\delta$ is another, so we might as well just rename them A and B and we've recovered Eq. (11.4). Each of the forms of the solution describes a sinusoidal solution with some combination of the integration constants providing both the amplitude and the phase. That's all we get to pick with our initial conditions.

One of the reasons this is such an important problem in physics is that so many systems can be approximated as the simple harmonic oscillator. Consider the pendulum. As we've seen many times, the equation of motion looks like

$$\ddot{\theta} = -\frac{g}{l}\sin\theta. \tag{11.13}$$

If we take just the first term from the Taylor expansion of $\sin\theta$ for small θ, we get

$$\ddot{\theta} \approx -\frac{g}{l}\theta, \tag{11.14}$$

which is the simple harmonic oscillator equation with the replacements $x \to \theta, k/m \to g/l$. So we can see right away that for small angles, the angular frequency of the pendulum is approximately $\sqrt{g/l}$. We can make a rather general point if we look at a potential energy plot such as Fig. 11.1. The figure shows the (dimensionless) potential $U = -\cos\theta$ overlayed with the (dimensionless) potential $U = -1+\frac{1}{2}\theta^2$, which is same as the harmonic oscillator potential (the -1 is irrelevant since it's a constant that will go away when we take the derivatives to get the equation of motion). What we see is that near any stable equilibrium, the spring potential is a good approximation to the the real potential.

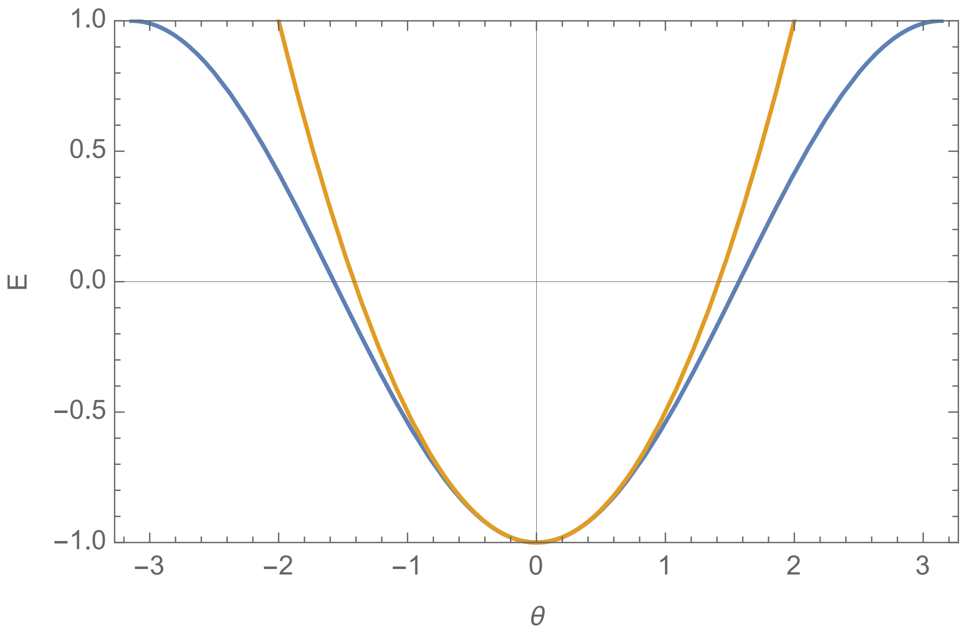


Figure 11.1: Parabolic potential in gold, $-\cos\theta$ potential in blue.

11.2 Damped Harmonic Oscillator

If we add a damping force of the form $F_D = -bv$, the equation of motion would become

$$m\ddot{x} = -kx - b\dot{x}, \tag{11.15}$$

or

$$\ddot{x} = -\frac{k}{m}x - \frac{b}{m}\dot{x}. \tag{11.16}$$

For later convenience, I'll introduce some new symbols

$$\omega_0^2 = \frac{k}{m} \quad 2\Gamma = \frac{b}{m}. \tag{11.17}$$

So the equation of motion can be written

$$\ddot{x} = -\omega_0^2 x - 2\Gamma\dot{x}. \tag{11.18}$$

Here ω_0^2 is called the natural frequency and Γ is the damping constant or decay constant. Note that when the damping is zero, the natural frequency is just the frequency of oscillations. We'll see later that with damping, this conclusion will change.

The equation (11.18) is a linear, homogeneous, ODE. It turns out that the solutions to such equations can always be written as combinations of exponentials. If you take the ODE course, you might prove this. Here, we'll just state the result and use it. So to solve Eq. (11.18), we'll try plugging in $x = e^{pt}$ and find values of p that work. After plugging into Eq. (11.18) we find

$$p^2 + 2\Gamma p + \omega_0^2 = 0 \tag{11.19}$$

PCQ 11.3
Check that you get Eq. (11.19).

We can then apply the quadratic equation and get the values of p that work

$$p = -\Gamma \pm \sqrt{\Gamma^2 - \omega_0^2} = -\Gamma \pm \beta, \tag{11.20}$$

where I've defined $\beta = \sqrt{\Gamma^2 - \omega_0^2}$ So the general solution will be a linear combination of e^{pt} with each value of p. That is,

$$x[t] = e^{-\Gamma t}(Ae^{\beta t} + Be^{-\beta t}), \tag{11.21}$$

where A and B are again our 2 integration constants. We can also write this as

$$x[t] = e^{-\Gamma t}(C \cosh \beta t + D \sinh \beta t), \tag{11.22}$$

where we've packaged the integration constants differently in analogy to the simple harmonic oscillator case. In this second form, the arbitrary constants can be found in terms of the initial conditions as follows

$$C = x_0 \quad D = \frac{v_0}{\beta} + \frac{\Gamma x_0}{\beta}. \tag{11.23}$$

It turns out that there are 3 interesting possibilities for β. The conditions and names are as follows:

β imaginary	$(\Gamma < \omega_0)$	under damped
β real	$(\Gamma > \omega_0)$	over damped
$\beta = 0$	$(\Gamma = \omega_0)$	critically damped

We shall consider each case in turn.

11.2.1 under damped

If β is imaginary, we can define a real quantity ω as follows:

$$\omega = -i\beta = \sqrt{\omega_0^2 - \Gamma^2} \tag{11.24}$$

We can then write the solution

$$\begin{aligned} x[t] &= e^{-\Gamma t}(Ae^{i\omega t} + Be^{-i\omega t}) && (11.25) \\ &= e^{-\Gamma t}(C \cos \omega t + D \sin \omega t). && (11.26) \end{aligned}$$

So this is sinusoidal oscillations at frequency ω with an exponentially decaying amplitude. Note that $\omega < \omega_0$. That is, in the damped system, the frequency of oscillations is lower than in the zero damping case. Evidently, the damping "slows down" the oscillations. We can plot a particular case as in Fig. 11.2

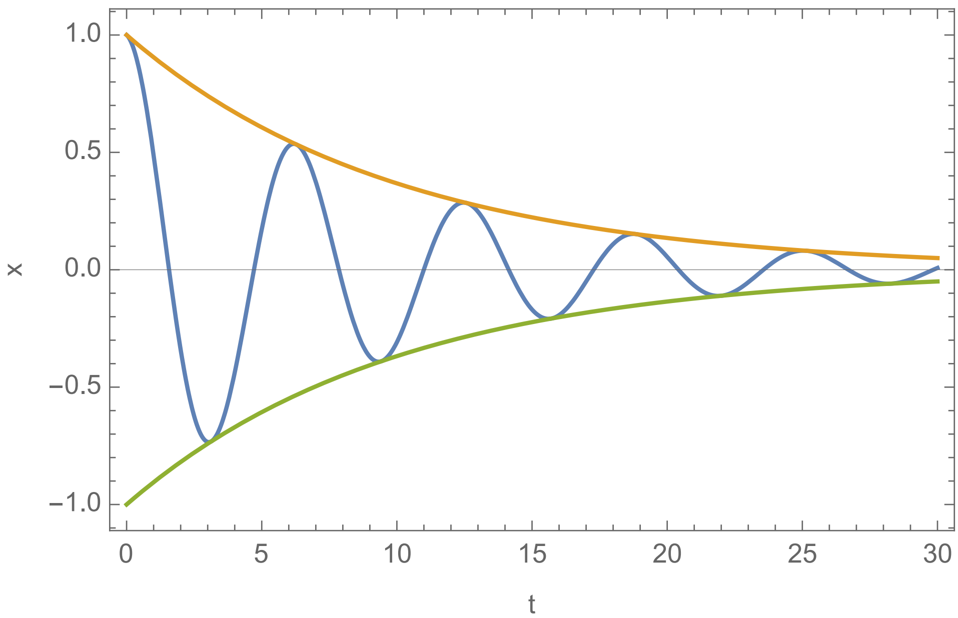


Figure 11.2: Under damped oscillator: $x[t]$ in blue, damping envelope in green and gold.

Here the solution is in blue in the middle. I've also plotted the decaying amplitude or"decay envelop" separately. Basically every mass on a spring you've seen in the lab is really doing this since aerodynamic drag is acting and the oscillations die out.

11.2.2 over damped

Lets just stare at the general solution again.

$$x[t] = e^{-\Gamma t}(Ae^{\beta t} + Be^{-\beta t}), \tag{11.27}$$

With real β here in the over damped case, it sort of makes sense to just write the Γ and β exponentials together as follows:

$$x[t] = Ae^{-(\Gamma-\beta)t} + Be^{-(\Gamma+\beta)t}. \tag{11.28}$$

These are both decaying exponentials, so there is no oscillations here. For example, if we start the system by pulling it away from equilibrium and letting it go, it will just decay to the equilibrium. Imagine you have a spring and mass system in cold maple syrup. If you stretched the mass on the spring and let it go, it would just slowly move back to equilibrium without overshooting. The motion looks like the blue curve in Fig. 11.3. As the damping is increased, the time it takes for the oscillator to return to equilibrium increases.

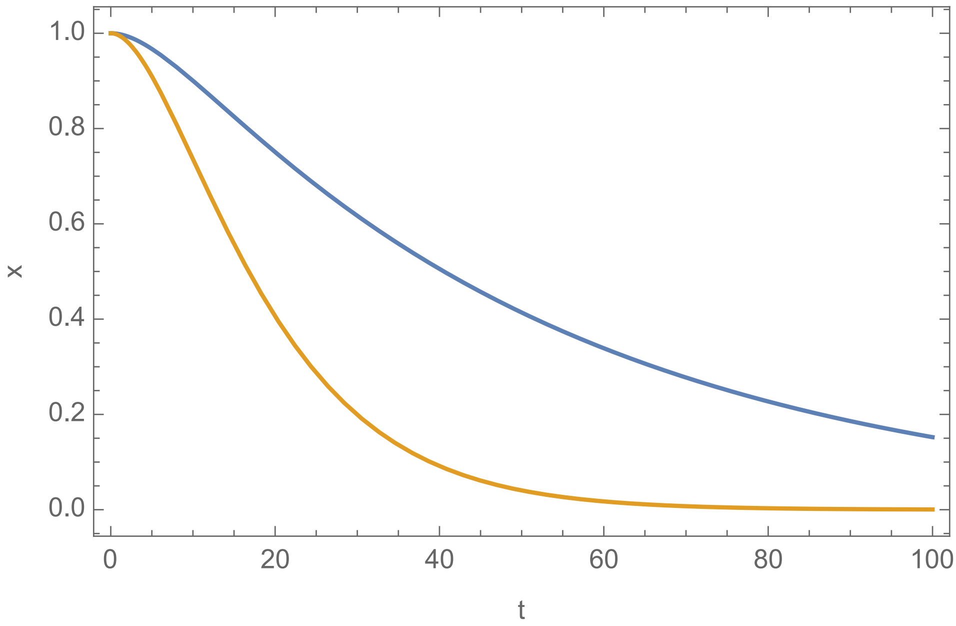


Figure 11.3: Over damped oscillator $x[t]$ in blue, critically damping oscillator in gold.

11.2.3 critically damped

Here it's easiest to start with the hyperbolic trig form of the general solution:

$$x[t] = e^{-\Gamma t}(x_0 \cosh \beta t + \tfrac{1}{\beta}(v_0 + \Gamma x_0) \sinh \beta t). \tag{11.29}$$

We can then take the limit as β goes to zero. The second term is indeterminant, so we'll either need l'Hopital's rule, or we'll need to consider the Taylor expansion of sinh. Either way, the result is

$$x[t] = e^{-\Gamma t}(x_0 + (v_0 + \Gamma x_0)t). \tag{11.30}$$

This result is plotted as the gold curve in Fig. 11.3. Critical damping is the fastest way to get the system to rest at the equilibrium. This has lots of engineering applications, including door closing springs and automotive suspensions.

11.3 Damped Driven Oscillator

The simplest example of a damped driven oscillator is a kid on a swing being pushed by an adult. Aerodynamic drag provides the damping, removing energy from the system, and the adult provides the driving, adding energy to the system. As you know, if the adult is to add energy, they need to push in the right direction at the right time. If they push the wrong way at the wrong time, the system will slow down. This fact is a consequence of "resonance", which will be a key idea in this section.

To keep things simple at the start, I'll consider a simpler (but closely related) system. I'll focus on a mass on a spring with damping as before, but I'll now add a sinusoidal driving force of amplitude F_0. The particular form of this force is just a choice. In principle, I could consider any force here. Think of this additional driving force as provided by some motor that periodically pushes the mass. So the equation of motion will become

$$\ddot{x} + 2\Gamma\dot{x} + \omega_0^2 x = \frac{F_0}{m}\cos[\Omega t + \phi_0]. \tag{11.31}$$

Note that the left hand side as I've written it is just the damped harmonic oscillator. The right side is the driving force, where Ω is the frequency of the driving force and ϕ_0 sets the initial value of the force. So we're applying this external driving force that varies from $+F_0$ to $-F_0$ and back periodically at frequency Ω to our mass.

PCQ 11.4

If $\phi_0 = -\pi/2$, what is the driving force at time $t = 0$? How about if $\phi_0 = 0$?

You might wonder why I'm choosing a driving force of this particular form. There are probably 3 (somewhat interrelated) answers to this: (a) It's a case that I know how to solve that has interesting features. (b) It's a good model for a lot of systems of interest. (c) Fourier series tells us that any (well behaved) force function could be represented as a sum of sin functions and cos functions of various frequencies, which implies that it's really more general that it might seem.

PCQ 11.5

In class you picked up a number of plots. The plots show the solution to Eq. (11.31), $x[t]$, and the driving force as a function of time for different values of Ω. Come up with one observation about these plots.

Note that Eq. (11.31) is again an inhomogeneous, 1st order, linear ode. We could think of the solution as constructed from a homogeneous solution and a particular solution. The homogeneous solution is just the one we got for the damped (undriven) case in the previous section. Since we happened to find solutions that decay away with time, we might notice that if we wait long enough (a time of order $1/\Gamma$) this homogeneous solution will die out and we'll be left with the particular solution. You might also have noticed in the plots of driven oscillator responses that the solutions after long time, which are the particular solutions, have frequency Ω. While these are important insights, the particular solution is challenging to guess, so I'll now show you another solution method.

To find the particular solution, I'm going to use a method I call make the problem more *complex* to make it easier. Let $z[t] = x[t] + iy[t]$ be a complex function, where $x[t]$ and $y[t]$ are real functions. Consider the complex differential equation

$$\ddot{z} + 2\Gamma\dot{z} + \omega_0^2 z = \frac{F_0}{m}e^{-i(\Omega t+\phi_0)}. \tag{11.32}$$

Note that the real part of this equation is (11.31). If I find a function $z[t]$ that satisfies this complex equation then its real part will satisfy Eq. (11.31). Since I know from observation that the particular solution will be oscillatory with frequency Ω, I'm going to guess

$$z_p[t] = ce^{-i(\Omega t+\phi_0)}, \tag{11.33}$$

where c is a complex constant. I'll now plug this into (11.32) to see if I can find a c that will work. I claim that I'll find

$$c = \frac{F_0/m}{(\omega_0^2 - \Omega^2) - 2i\Omega\Gamma}. \tag{11.34}$$

This is a complex number and I can write it in the usual maner of a complex number (as a real and imaginary part) by multiplying the top and the bottom by the complex conjugate of the bottom

$$
\begin{aligned}
c &= \frac{F_0/m}{(\omega_0^2-\Omega^2)-2i\Omega\Gamma} \times \frac{(\omega_0^2-\Omega^2)+2i\Omega\Gamma}{(\omega_0^2-\Omega^2)+2i\Omega\Gamma} && (11.35)\\
&= \frac{F_0}{m}\frac{\omega_0^2-\Omega^2}{(\omega_0^2-\Omega^2)^2+4\Omega^2\Gamma^2} + 2i\frac{F_0}{m}\frac{\Omega\Gamma}{(\omega_0^2-\Omega^2)^2+4\Omega^2\Gamma^2} && (11.36)\\
&= \mathcal{A} + i\mathcal{B}. && (11.37)
\end{aligned}
$$

Since the real and imaginary parts of c are big and clunky, I'm defining $\mathcal{A}$ and $\mathcal{B}$ as the real and imaginary parts in the last line here.

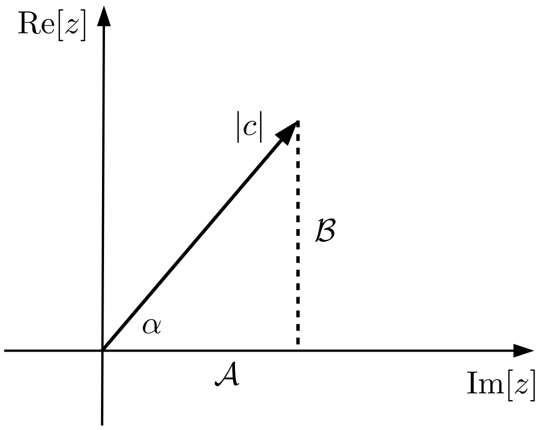


Figure 11.4: The constant c on the complex plane

We can view the complex constant c on the complex plane as in Fig. 11.4. Note that $\mathcal{A}$ is the real "component" and $\mathcal{B}$ is the imaginary "component". The angle α is

$$
\alpha = \arctan\left[\frac{\mathcal{B}}{\mathcal{A}}\right] = \arctan\left[\frac{2\Omega\Gamma}{\omega_0^2-\Omega^2}\right] \tag{11.38}
$$

The magnitude of c is found by multiplying c by its complex conjugate and taking the square root yielding

$$
|c| = \frac{F_0/m}{\sqrt{(\omega_0^2-\Omega^2)^2+2\Omega^2\Gamma^2}}. \tag{11.39}
$$

We could now write z_p as

$$
\begin{aligned}
z_p &= |c|e^{i\alpha}e^{-i(\Omega t+\phi_0)}\\
&= |c|e^{-i(\Omega t+\phi_0-\alpha)} && (11.40)
\end{aligned}
$$

Taking the real part and writing things out explicitly, we have

$$
x_p = \frac{F_0\cos[\Omega t+\phi_0-\alpha]}{m\sqrt{(\omega_0^2-\Omega^2)^2+2\Omega^2\Gamma^2}}. \tag{11.41}
$$

We can now notice that the amplitude of the oscillation is the biggest when $\Omega = \omega_0$. That is, when we drive the system at its natural frequency. This is known as driving on resonance. We also see that x_p leads the drive signal by a phase α. On resonance, we see that $\alpha = \pi/2$, or 90^o, consistent with what you'll see in the plots. Note also that the ultimate amplitude of the oscillations that results when driving on resonance is controlled by the damping constant. Less damping implies greater amplitude at resonance.

Another popular way of quantifying the damping in a system is quality factor. High quality factor means there is little damping in the system. One way to explore the resonant frequency and the quality factor of a system experimentally is to run the driving force at a variety of frequencies and plot the resulting squared amplitudes of the steady state oscillations. Figure 11.5 shows what one might find. Here the (dimensionless) damping factor is 0.5 and the resonant frequency is 20. Note that the width of the resulting peak at half the max is 2Γ centered on ω_0. One definition for the quality factor is

$$Q = \frac{\text{position}}{\text{width}} = \frac{\omega_0}{2\Gamma}. \tag{11.42}$$

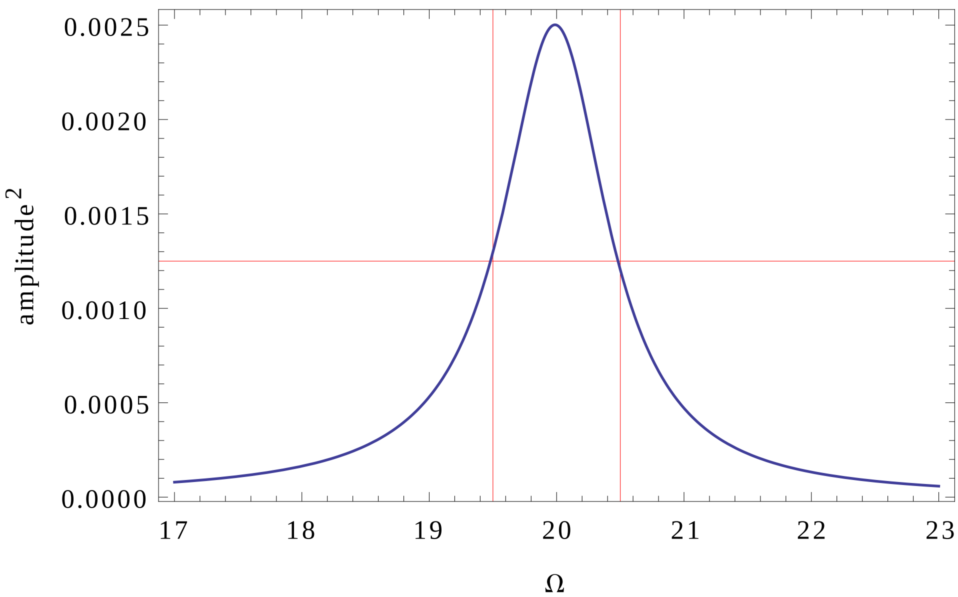


Figure 11.5: Squared amplitude vs. driving frequency near resonance.

The quality factor also turns out to be inversely proportional to the rate of energy loss from the system divided by the total energy in the system averaged over a period. We can find the rate of energy loss from the work done by the drag force:

$$\frac{dE}{dt} = -2m\Gamma\dot{x}\cdot\dot{x}. \tag{11.43}$$

The total energy in the system is

$$E = \tfrac{1}{2}m\dot{x}^2 + \tfrac{1}{2}m\omega_0^2 x^2. \tag{11.44}$$

If we consider weak damping

$$\begin{aligned} x &\approx A\cos\omega_0 t \\ \dot{x} &\approx -A\omega_0 \sin\omega_0 t. \end{aligned} \tag{11.45}$$

Plugging this into the energy expressions yields

$$\frac{1}{E}\frac{dE}{dt} = -4\Gamma\sin^2\omega_0 t. \tag{11.46}$$

We can now consider the following calculation:

$$\begin{aligned} Q^{-1} &= \frac{1}{2\pi}\int_0^T \left|\frac{1}{E}\frac{dE}{dt}\right| dt \\ &= \frac{2\Gamma}{\pi}\int_0^T \sin^2\omega_0 t dt \\ &= \frac{2\Gamma}{\omega_0\pi}\int_0^{2\pi} \sin^2\theta d\theta \\ &= \frac{2\Gamma}{\omega_0\pi}\pi \\ &= \frac{2\Gamma}{\omega_0}. \end{aligned} \tag{11.47}$$

Note that this definition yields the same result as the width at half the max definition.