Chapter 13

Appendices

Problems and Solutions in Introductory Mechanics (Draft version, August 2014)
David Morin, morin@physics.harvard.edu

13.1 Appendix A: Vectors

13.1.1 Basics

Figure 13.1

Although it is possible to define a *vector* in a more precise way, the definition that will suffice for our purposes is that a vector is a mathematical object that has both a *magnitude* (that is, a length) and a *direction*. For example, Fig. 13.1 shows a vector that points upward and rightward with a length of about 3 cm. In general, a vector can point in an arbitrary direction in 3-D space, but we'll deal mainly with 2-D vectors in the plane of the page. Vectors also exist in 1-D, but in contrast with 2-D and 3-D where a vector can point in an infinite number of directions, a vector in 1-D can point in only two directions – to the right or to the left, if we're dealing with a horizontal line. So the direction of a 1-D vector can be specified by simply writing a positive or negative sign.

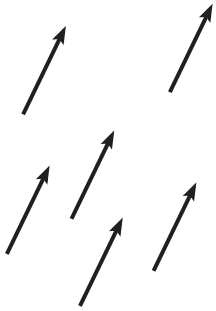

Figure 13.2

A vector in any dimension is completely determined by its magnitude and direction. This means that all of the vectors in Fig. 13.2 are actually the same vector. It doesn't matter where the tail and the tip of a given vector are, as long as the segment joining them has a given length and points in a given direction.

We'll denote vectors with boldface letters, such as $\mathbf{r}$ or $\mathbf{v}$, etc. And we'll denote the magnitude (length) with an italic latter, such as r or v, etc. The length is also sometimes denoted with absolute value bars around the vector, so $|\mathbf{r}|$ means the same thing as r. The length is a *scalar*, that is, just a number (with units).

It is important to distinguish between a vector and its magnitude. The magnitude is only one piece of the complete information contained in the entire vector (with the other piece being the direction). The classic example of this distinction is the *velocity* $\mathbf{v}$ of an object versus the *speed* v. By "velocity" we mean the complete vector, and by "speed" we mean just the magnitude. A statement such as "My velocity is 10 m/s" uses incorrect terminology, whereas statements such as "My speed is 10 m/s" and "My velocity is 10 m/s in the northwest direction" use correct terminology.

Examples of vectors that come up in this book are: position $\mathbf{r}$, velocity $\mathbf{v}$, acceleration $\mathbf{a}$, momentum $\mathbf{p}$, force $\mathbf{F}$, angular momentum $\mathbf{L}$, and angular velocity $\boldsymbol{\omega}$. Examples of scalars are: length ℓ, speed v, energy E, mass m, and angular speed ω.

13.1.2 Cartesian coordinates

If you want to do anything quantitative with vectors, it is usually necessary to use a *coordinate system* to describe them. There are many different kinds of coordinate systems you can choose from, the most common being Cartesian, polar, cylindrical, and spherical. Additionally, you have the freedom to choose the orientation of your coordinate axes.

The simplest and most common type of coordinate system is a Cartesian one, which in 3-D space involves the three mutually perpendicular x, y, and z axes. The Cartesian *components* of

a vector **a** are written as (a_x, a_y, a_z). To understand what this notation means, we can write **a** in terms of the *basis* vectors $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, $\hat{\mathbf{z}}$. These basis vectors are unit vectors (that is, they have length 1) pointing along the three coordinate axes. In terms of the basis vectors, we have

$$\mathbf{a} = (a_x, a_y, a_z) = a_x\hat{\mathbf{x}} + a_y\hat{\mathbf{y}} + a_z\hat{\mathbf{z}}. \tag{13.1}$$

So (a_x, a_y, a_z) is just shorthand for $a_x\hat{\mathbf{x}} + a_y\hat{\mathbf{y}} + a_z\hat{\mathbf{z}}$. We'll discuss the addition of vectors in Section 13.1.4 below, but in short, the $\mathbf{a} = a_x\hat{\mathbf{x}} + a_y\hat{\mathbf{y}} + a_z\hat{\mathbf{z}}$ relation says that if you march a distance a_x in the x direction, then a_y in the y direction, and then a_z in the z direction, you will end up at the tip of the **a** vector.

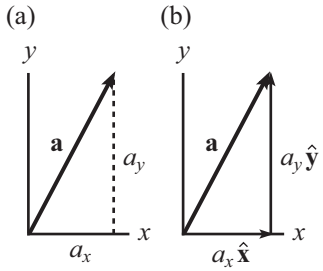

Figure 13.3

The word "component" can be used in two slightly different ways. If you ask someone for the x component of a vector **a**, the answer might be a_x (which is a number; see Fig. 13.3(a)), or the answer might be $a_x\hat{\mathbf{x}}$ (which is a vector; see Fig. 13.3(b)). The difference is just semantics. Either answer is acceptable, provided that it is used in the appropriate way, depending on whether it is a number or a vector. For example, if a person says, "A vector is the sum of its components," then she is clearly thinking of the components as vectors and making the correct statement that $\mathbf{a} = a_x\hat{\mathbf{x}} + a_y\hat{\mathbf{y}} + a_z\hat{\mathbf{z}}$, as opposed to the incorrect statement that $\mathbf{a} = a_x + a_y + a_z$. This latter equation certainly can't be correct, because it equates a vector with a nonvector. Additionally, the magnitude isn't even correct.

The length of a vector **a** whose Cartesian coordinates are (a_x, a_y, a_z) is

$$a \equiv |\mathbf{a}| = \sqrt{a_x^2 + a_y^2 + a_z^2}. \tag{13.2}$$

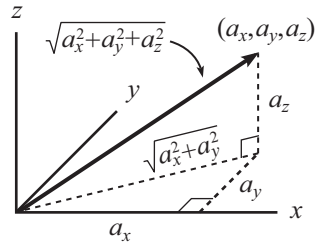

Figure 13.4

This follows from two applications of the Pythagorean theorem in Fig. 13.4, first with the right triangle in the x-y plane, and then with the vertical right triangle extending upward from the x-y plane to the point (a_x, a_y, a_z).

It should be emphasized that a vector has a meaning independent of any coordinate system (type, choice of origin, orientation of axes). Components are used to describe a vector, but a vector is what it is, independent of how we choose to describe it. The vector **a** we drew in Fig. 13.1 has length 3 cm and points up to the right. But we don't even need to give these bits of information to describe it. All we need to do is draw it on the page, as we did in Fig. 13.1. That arrow is the vector **a**, period.

However, the fact of the matter is that using components makes it much easier to deal with vectors. It would be a pain to have to keep drawing vectors whenever you needed to do something with them, such as adding them (see below). And if your drawing skills are shaky, then you might not trust the answer you obtain, anyway. Components provide a reliable and rigorous way of describing vectors. If the task of a given problem is to find a certain vector, then in practice the most common ways of stating the answer are to give the Cartesian coordinates or to give the magnitude and direction.

REMARK: Here is a somewhat subtle point, concerning the distinction between *points* and *vectors*. If someone gives you the triplet of numbers (a_x, a_y, a_z), is this a point in space, or is it a vector? Well, it could be either, so you would need to be told which it is. The main difference between a point (a_x, a_y, a_z) and a vector (a_x, a_y, a_z) is that whereas a vector doesn't depend on the choice of origin, a point very much does. The coordinates of the point (a_x, a_y, a_z) are measured with respect to a given origin, so if you don't know where that origin is, then you don't know where the point (a_x, a_y, a_z) is. In contrast, the components of the vector (a_x, a_y, a_z) give the position of the head with respect to the tail, so it doesn't matter where the origin is. In Fig. 13.4 we actually used (a_x, a_y, a_z) to mean both a point and a vector.

Given two points in space, the position of one with respect to the other is a vector, because the relative position doesn't depend on the choice of origin. If you are given the coordinates of a particular house, then you also need to be given the origin that these coordinates are measured with respect to. But if you walk from one house to another (in which case your displacement will be described by a vector), then you're going to take the same walk independent of the location of the point that someone arbitrarily decides to call the origin.

13.1.3 Polar coordinates

Consider a *point* in 2-D space (and yes, we're talking about a point here, not a vector). We can use Cartesian coordinates to describe the location of this point relative to a given origin, as we did above. But we can also use polar coordinates. The polar coordinates of a point are the distance r from the origin and the angle θ relative to the horizontal axis, as shown in Fig. 13.5. By looking at the right triangle in the figure, the Cartesian coordinates (x, y) of the point can be obtained from the polar coordinates (r, θ) by

$$x = r\cos\theta \qquad \text{and} \qquad y = r\sin\theta. \tag{13.3}$$

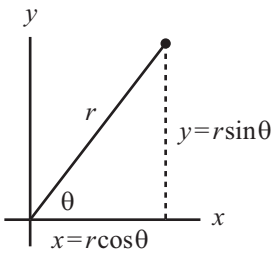

Figure 13.5

Equivalently, the polar coordinates (r, θ) can be obtained from the Cartesian coordinates (x, y) by

$$r = \sqrt{x^2 + y^2} \qquad \text{and} \qquad \theta = \arctan(y/x). \tag{13.4}$$

The second expression here doesn't quite determine θ, due to the $\tan(\theta + \pi) = \tan\theta$ ambiguity. But the correct angle can be determined by the sign of x (or y); this narrows down the quadrant the angle is in. In 3-D, the generalization of polar coordinates is *spherical coordinates*, which involve an additional angle associated with circling around a sphere (the longitudinal angle on the earth).

Let's now look at what form a *vector* (as opposed to a point) takes when written in polar coordinates.[1] Things aren't as simple as they are in Cartesian coordinates, because while the Cartesian basis vectors $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, $\hat{\mathbf{z}}$ don't depend on position, the polar basis vectors $\hat{\mathbf{r}}$ and $\hat{\boldsymbol{\theta}}$ do. At a given point, the unit vector $\hat{\mathbf{r}}$ is defined to point in the radial direction, and the unit vector $\hat{\boldsymbol{\theta}}$ is defined to point in the counterclockwise tangential direction. These directions depend on where the point is located in the plane, as shown at points P_1 and P_2 in Fig. 13.6. The polar basis vectors $\hat{\mathbf{r}}$ and $\hat{\boldsymbol{\theta}}$ are undefined at the origin, because the radial and tangential directions aren't uniquely defined there.

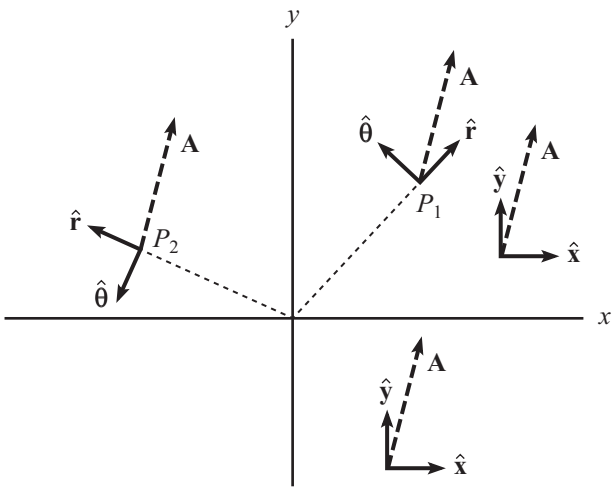

Figure 13.6

To be clear, the $\hat{\mathbf{r}}$ vector pointing up and to the right at point P_1 in Fig. 13.6 is the same as a similar vector pointing up and to the right, no matter where we place it in the plane (as in Fig. 13.2). However, this $\hat{\mathbf{r}}$ vector has nothing to do with what we define to be the $\hat{\mathbf{r}}$ vector at, say, point P_2, where $\hat{\mathbf{r}}$ points up and to the left. Although we use the same *symbols* $\hat{\mathbf{r}}$ and $\hat{\boldsymbol{\theta}}$ for the

[1]The rest of this subsection on polar coordinates isn't important for this book; the $\hat{\mathbf{r}}$ and $\hat{\boldsymbol{\theta}}$ basis vectors don't appear anywhere except in this appendix. But it is natural to include the following discussion, given that we've introduced polar coordinates. In more advanced topics in physics, this discussion is highly relevant.

basis vectors at any point, this doesn't mean that the basis vectors at different points are actually the same *vectors*.

So what does a vector look like in when written in polar coordinates? Consider the vector **A** shown above in Fig. 13.6. This vector is the same vector, no matter where we place it in the plane (as in Fig. 13.2). In Cartesian coordinates, let's say **A** is given by $(0.7)\hat{\mathbf{x}} + (2.3)\hat{\mathbf{y}}$ (just a rough estimate, by comparing **A** with the unit basis vectors). This expression describes **A** in Cartesian coordinates, no matter where we draw **A**. But in polar coordinates, the description depends on what basis vectors we use. That is, it depends on where we put **A** (more precisely, where we put the tail of **A**). At point P_1, **A** happens to be $(2.1)\hat{\mathbf{r}} + (1.2)\hat{\boldsymbol{\theta}}$, whereas at point P_2, **A** happens to be $(0.4)\hat{\mathbf{r}} - (2.4)\hat{\boldsymbol{\theta}}$. These are two different expressions for the same vector. Note that all of the above expressions for **A** have the same length (at least up to rounding errors), as they must. The length happens to be about 2.4.

Polar (or spherical) coordinates are very useful when describing vectors in setups that possess circular (or spherical) symmetry. For example, the full vector form of the gravitational force on a mass m due to a mass M located at the origin is

$$\mathbf{F} = -\frac{GMm}{r^2}\hat{\mathbf{r}}, \tag{13.5}$$

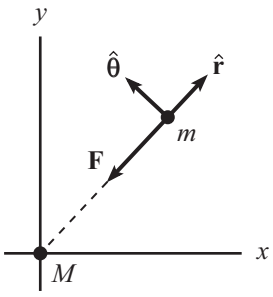

Figure 13.7

as shown in Fig. 13.7. Here we are using the $\hat{\mathbf{r}}$ basis vector at the most reasonable point to pick, which is the location of the mass m. (The only other special point in the plane is the origin, where M is. But as mentioned above, the polar basis vectors are undefined there.) The above polar expression for **F** is much cleaner than the Cartesian expression,

$$\mathbf{F} = -\frac{GMm}{x^2+y^2}\left(\frac{x}{\sqrt{x^2+y^2}}\hat{\mathbf{x}} + \frac{y}{\sqrt{x^2+y^2}}\hat{\mathbf{y}}\right). \tag{13.6}$$

This is indeed the same force as in Eq. (13.5), because when $\hat{\mathbf{r}}$ is written in terms of Cartesian coordinates, it equals the vector in parentheses in Eq. (13.6). This is true because this vector has length 1, and it is proportional to the vector (x, y) which points in the $\hat{\mathbf{r}}$ direction by definition.

The gravitational force **F** is an example of a *vector field*. In a vector field, every point in space has a vector associated with it. Another example of a vector field is wind velocity; every point is space has associated with it the local velocity of the wind. (In contrast, in a *scalar field*, every point has a number (with units) associated it. An example is the temperature.) When describing a vector field, it is understood that when writing a vector at a given point in terms of basis vectors, you are using the basis vectors at that point, although this specification isn't necessary in Cartesian coordinates because the basis vectors are the same everywhere.

Remark: If someone working with polar coordinates gives you the pair of numbers (a_r, a_θ), what does it represent? As with Cartesian coordinates, the expression (a_r, a_θ) is ambiguous. It could mean a point in the plane, as in Fig. 13.5, with $a_r = r$ and $a_\theta = \theta$ (in which case you would need to be told where the origin is). Or it could mean the components of the vector $a_r\hat{\mathbf{r}} + a_\theta\hat{\boldsymbol{\theta}}$ (in which case you would need to be told what basis vectors $\hat{\mathbf{r}}$ and $\hat{\boldsymbol{\theta}}$ to use). In practice, most people would assume that (a_r, a_θ) represents a point in the plane. Note that technically the ambiguity is removed if you pay attention to units/dimensions. In the case where (a_r, a_θ) represents a point in the plane, a_r has dimensions of length and a_θ is a dimensionless angle. But in the case where (a_r, a_θ) represents a vector, a_r and a_θ must have the same dimensions. This is true because the dimensions of $\hat{\mathbf{r}}$ and $\hat{\boldsymbol{\theta}}$ are the same; they are both dimensionless unit vectors, just as $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ are. In an expression such as $x\hat{\mathbf{x}}$, the units of meters are in the x, not the $\hat{\mathbf{x}}$.

13.1.4 Adding and subtracting vectors

What is the sum of two vectors, for example, the two vectors **a** and **b** shown in Fig. 13.8(a)? To add two vectors, we put them "tail-to-head," as shown in Fig. 13.8(b). It doesn't matter which vector you start with; you can add them in either order and you will obtain the same result, as shown with the parallelogram in the figure. In other words, vector addition is *commutative*. In

the top triangle, the vector from the tail of **a** to the head of **b** is the desired sum; you can think of this as **a** + **b**. In the bottom triangle, the vector from the tail of **b** to the head of **a** is also the desired sum; you can think of this as **b** + **a**.

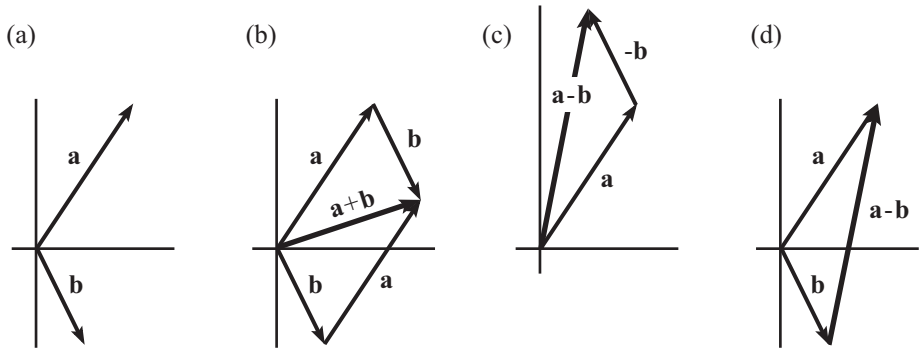

Figure 13.8

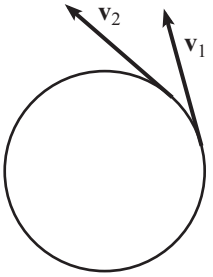

Figure 13.9

The tail-to-head procedure makes intuitive sense: If you go for a walk and end up with a displacement of **a**, and from there you go for another walk and end up with a displacement of **b** (relative to where you began your second walk), then your total displacement is given by the **a** + **b** vector in Fig. 13.8(b).

If you want to subtract two vectors, this operation is the same as adding on the negative of one of the vectors. That is, $\mathbf{a} - \mathbf{b} = \mathbf{a} + (-\mathbf{b})$. (The negative of a vector is the vector that has the same magnitude but points in the opposite direction.) If we consider the same vectors **a** and **b** as in Fig. 13.8(a), the difference **a** − **b** is shown in Fig. 13.8(c); we have added −**b** to **a**. Equivalently, **a** − **b** is the vector from the tip of **b** to the tip of **a** (with the tails of the vectors coinciding), because **a** − **b** is the vector that you need to add to **b** to obtain **a**. This equality $\mathbf{b} + (\mathbf{a} - \mathbf{b}) = \mathbf{a}$ is shown in Fig. 13.8(d).

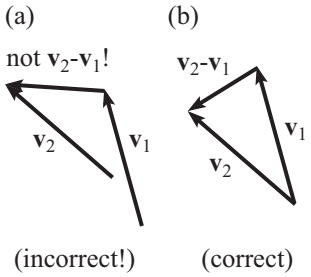

Figure 13.10

If you want to use the method in Fig. 13.8(d) to calculate the difference between two vectors, you must remember to put the tails of the two vectors at the same point. A common error is to draw the (incorrect) difference vector as the vector that connects the tips of two given vectors whose tails don't coincide. Consider, for example, a particle traveling around a circle at constant speed. The velocity vectors at two different times are shown in Fig. 13.9; they have the same length but different directions. The difference $\mathbf{v}_2 - \mathbf{v}_1$ between these vectors is *not* the vector shown in Fig. 13.10(a), obtained by simply connecting the tips of the vectors as they appear on the page in Fig. 13.9. The correct difference, obtained by having the tails of $\mathbf{v}_1$ and $\mathbf{v}_2$ coincide, is shown in Fig. 13.10(b). Of course, you can also obtain this difference by adding $-\mathbf{v}_1$ to $\mathbf{v}_2$, as you can check.

If you want to multiply a vector by a scalar n (a number), you just need to multiply the length by n and keep the direction the same. In the special case where n is an integer, this is the same as lining up n copies of the vector. For example, since $3\mathbf{b} = \mathbf{b} + \mathbf{b} + \mathbf{b}$, the vector 3**b** is shown in Fig. 13.11. If you want to multiply a vector by another vector, there are two possible ways to do this, involving the *dot product* and the *cross product*. These are discussed below in Sections 13.1.6 and 13.1.7.

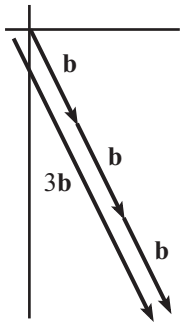

Figure 13.11

As mentioned earlier, although it is possible in theory to manipulate vectors by doing nothing other than drawing them (which has been our strategy thus far in this subsection), it is usually necessary to make use of coordinates if we want to be quantitative. If we are adding two vectors, we simply need to add the corresponding Cartesian components. For simplicity let's work in 2-D, where a vector is described by two Cartesian components (a_x, a_y). In the addition example in Fig. 13.8(b), let's say that the two vectors are $\mathbf{a} = (2,3)$ and $\mathbf{b} = (1,-2)$. Then the sum is

$$\mathbf{a} + \mathbf{b} = (2,3) + (1,-2) = (2+1,\, 3-2) = (3,1), \tag{13.7}$$

as shown in Fig. 13.12 (we've scaled up the size of the figure, relative to Fig. 13.8). It is clear

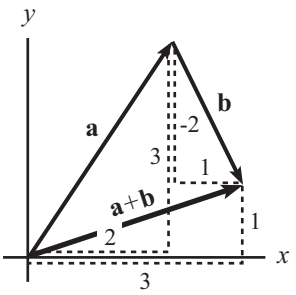

Figure 13.12

from this figure that the x component of the sum equals the sum of the x components of the two individual vectors. Likewise for the y components. A common mistake when adding vectors is to forget that they are vectors and simply add the magnitudes. Remember that vectors aren't just numbers, and that you always need to first break vectors into their Cartesian components, and then you can add the corresponding components. To subtract two vectors, you simply need to subtract the components.

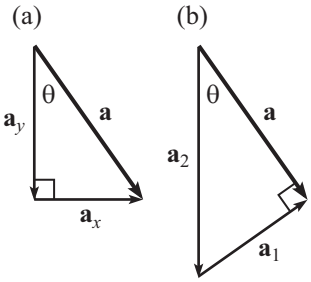

Figure 13.13

13.1.5 Using components

What is the vertical component of the vector $\mathbf{a}$ shown in Fig. 13.13? Is it the vertical vector $\mathbf{a}_y$ in Fig. 13.13(a) with length $a\cos\theta$, or is it the vertical vector $\mathbf{a}_2$ in Fig. 13.13(b) with length $a/\cos\theta$? The correct choice is the first one, $\mathbf{a}_y$. When breaking a given vector into components, the vector is always the hypotenuse of the triangle representing the components. The axes of our coordinate systems are always orthogonal, so the components must always be orthogonal, which isn't the case in Fig. 13.13(b). A vector is the sum of its components, and while it is certainly true that $\mathbf{a} = \mathbf{a}_2 + \mathbf{a}_1$ in Fig. 13.13(b), just as it is true that $\mathbf{a} = \mathbf{a}_y + \mathbf{a}_x$ in Fig. 13.13(a), the former of these decompositions isn't helpful in general (because $\mathbf{a}_1$ has a vertical component, whereas $\mathbf{a}_x$ does not).

When solving a problem, choosing your coordinate system wisely can make the problem much easier to solve. Cartesian and polar coordinates are the most common options. But additionally, if you are using Cartesian coordinates, you might find that one orientation of your axes makes things easier than another. For example, in setups involving objects on inclined planes, you can pick your axes to be horizontal and vertical, or you can pick them to be parallel and perpendicular to the plane. It often isn't obvious which is the better set of axes, so you might need to try both. When using the latter set, you will need to break the gravitational acceleration $\mathbf{g}$ (or equivalently, the gravitational force $m\mathbf{g}$) into components parallel and perpendicular to the plane. These components have lengths $g\sin\theta$ and $g\cos\theta$, as shown in Fig. 13.14. (You can do the geometry to determine which angles in the figure are equal to θ, or you can just use the limiting-case reasoning in Multiple-Choice Question 1.12.) The rectangle formed by the components is shown in Fig. 13.15. It's personal preference which of the two right triangles you draw to find the components.

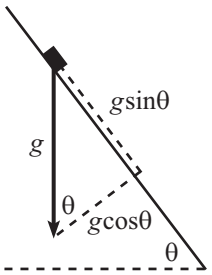

Figure 13.14

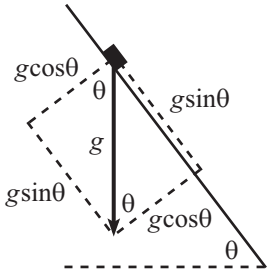

Figure 13.15

The downward $m\mathbf{g}$ force is a vector, and a vector is the sum of its components. So the sum of the $mg\sin\theta$ force along the plane and the $mg\cos\theta$ force perpendicular to the plane equals the original $m\mathbf{g}$ force. (There is also a normal force from the plane acting on the object, but that won't concern us here.) In other words, the object on the plane behaves in exactly the same way whether it is acted on by the single gravitational force or by two people pulling along the plane and perpendicular to the plane with forces of $mg\sin\theta$ and $mg\cos\theta$. The sum of the *magnitudes* of these two forces doesn't equal the magnitude of the original $m\mathbf{g}$ force (the sum is larger), but that is irrelevant. All that matters is that the sum of the two component *vectors* equals the original $m\mathbf{g}$ vector. When solving problems, it is often helpful to forget all about the original force vector and instead consider it to be two separate forces generated by two people pulling/pushing along the directions of the components.

Many equations in physics are vector equations. An example is $\mathbf{F} = m\mathbf{a}$. Another example is a conservation-of-momentum equation of the form $\mathbf{p}_{\text{initial}} = \mathbf{p}_{\text{final}}$. Since these are vector equations, they don't say simply that the magnitudes are the same on both sides; they say that the complete vectors (including the directions) are the same. Equivalently, they say that the corresponding components on each side are equal. In other words, a vector equation is really three equations, one for each component. So $\mathbf{F} = m\mathbf{a}$ says the same thing as the combination of the three $F_x = ma_x$, $F_y = ma_y$, and $F_z = ma_z$ equations. (If you're dealing with a 2-D system in the x-y plane, then you really have only two equations, because $F_z = ma_z$ is the trivial statement that $0 = 0$.) Therefore, when dealing with vectors, you will invariably need to break them up into components and deal with these components separately.

13.1.6 Dot product

As mentioned in Section 13.1.4, there are two possible ways to multiply two vectors together: the *dot product* and the *cross product*. These products don't come up too often in this book, but since they do appear, we'll provide a brief review of them. The dot product shows up in the definition of work, and the cross product shows up in the definitions of torque, angular momentum, and fictitious forces.

Consider two vectors whose Cartesian components are given by

$$\mathbf{a} = (a_x, a_y, a_z) \qquad \text{and} \qquad \mathbf{b} = (b_x, b_y, b_z). \tag{13.8}$$

The *dot product*, or *scalar product*, between these vectors is defined as

$$\mathbf{a} \cdot \mathbf{b} \equiv a_x b_x + a_y b_y + a_z b_z. \tag{13.9}$$

The dot product takes two vectors and produces a scalar, which is just a number (with units). You can quickly use Eq. (13.9) to show that the dot product is commutative (that is, $\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a}$) and distributive (that is, $(\mathbf{a} + \mathbf{b}) \cdot \mathbf{c} = \mathbf{a} \cdot \mathbf{c} + \mathbf{b} \cdot \mathbf{c}$). Note that the dot product of a vector with itself is $\mathbf{a} \cdot \mathbf{a} = a_x^2 + a_y^2 + a_z^2$, which is just its length squared. So $\mathbf{a} \cdot \mathbf{a} = |\mathbf{a}|^2 \equiv a^2$.

Taking the sum of the products of the corresponding components of two vectors, as we did in Eq. (13.9), might seem like a silly and arbitrary thing to do. Why don't we instead look at, say, the sum of the cubes of the products of the corresponding components? The reason is that the dot product as we've defined it has many nice properties, the most useful of which is that it can be written as

$$\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}||\mathbf{b}| \cos\theta \equiv ab\cos\theta, \tag{13.10}$$

where θ is the angle between the two vectors. We can demonstrate this as follows. Consider the dot product of the vector $\mathbf{c} \equiv \mathbf{a} + \mathbf{b}$ with itself, which is simply the square of the length of $\mathbf{c}$. Using the distributive and commutative properties, we have

$$\begin{aligned} c^2 \;=\; (\mathbf{a}+\mathbf{b}) \cdot (\mathbf{a}+\mathbf{b}) \;&=\; \mathbf{a}\cdot\mathbf{a} + 2\mathbf{a}\cdot\mathbf{b} + \mathbf{b}\cdot\mathbf{b} \\ &=\; a^2 + 2\mathbf{a}\cdot\mathbf{b} + b^2. \end{aligned} \tag{13.11}$$

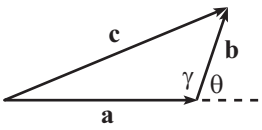

Figure 13.16

But from the law of cosines applied to the triangle in Fig. 13.16, we have

$$c^2 = a^2 + b^2 - 2ab\cos\gamma = a^2 + b^2 + 2ab\cos\theta, \tag{13.12}$$

because $\gamma = \pi - \theta$. Equating the two results for c^2 in Eqs. (13.11) and (13.12) yields $\mathbf{a} \cdot \mathbf{b} = ab\cos\theta$, as desired. The angle between two vectors is therefore given by

$$\cos\theta = \frac{\mathbf{a}\cdot\mathbf{b}}{|\mathbf{a}||\mathbf{b}|}. \tag{13.13}$$

A nice corollary of this result is that if the dot product of two vectors is zero, then $\cos\theta = 0$, which means that the vectors are perpendicular. If someone gives you the vectors $(1, -2, 3)$ and $(4, 5, 2)$, it is by no means obvious visually that they are perpendicular. But we know from Eq. (13.13) that they indeed are.

The dot product $\mathbf{a} \cdot \mathbf{b}$ can be written as either $a(b\cos\theta)$ or $b(a\cos\theta)$. So $\mathbf{a} \cdot \mathbf{b}$ equals the length of $\mathbf{a}$ times the component of $\mathbf{b}$ along $\mathbf{a}$. Or vice versa, depending on which length you want to group with the $\cos\theta$ factor. If we rotate our coordinate system, the dot product of two vectors remains the same, because from Eq. (13.10) it depends only on their lengths and the angle between them; and these are unaffected by the rotation. In other words, the dot product is a scalar. (The technical definition of a scalar is something that is invariant under a rotation of the coordinate system.) This certainly isn't obvious from looking at the original definition in Eq. (13.9), because the coordinates get all messed up during the rotation.

13.1.7 Cross product

The *cross product*, or *vector product*, between two vectors is defined via a determinant as

$$\begin{aligned}\mathbf{a}\times\mathbf{b} &\equiv \begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ a_x & a_y & a_z \\ b_x & b_y & b_z \end{vmatrix} \\ &= \hat{\mathbf{x}}(a_y b_z - a_z b_y) + \hat{\mathbf{y}}(a_z b_x - a_x b_z) + \hat{\mathbf{z}}(a_x b_y - a_y b_x). \end{aligned} \tag{13.14}$$

The cross product takes two vectors and produces another vector. As with the dot product, you can show that the cross product is distributive. However, it is *anti*-commutative (that is, $\mathbf{a}\times\mathbf{b} = -\mathbf{b}\times\mathbf{a}$), which is evident from Eq. (13.14); interchanging two rows of a determinant negates its value. This implies that the cross product of any vector with itself is zero.

As with the dot product, the reason why we study this particular combination of components is that it has many nice properties, the most useful of which are that the direction of $\mathbf{a}\times\mathbf{b}$ is perpendicular to both $\mathbf{a}$ and $\mathbf{b}$ (in the orientation determined by the right-hand rule; see below) and that the magnitude is

$$|\mathbf{a}\times\mathbf{b}| = |\mathbf{a}||\mathbf{b}|\sin\theta \equiv ab\sin\theta. \tag{13.15}$$

We'll derive this relation below, but let's first show that $\mathbf{a}\times\mathbf{b}$ is indeed perpendicular to both $\mathbf{a}$ and $\mathbf{b}$. We'll do this by making use of the fact that if the *dot* product of two vectors is zero, then the vectors are perpendicular. We have

$$\mathbf{a}\cdot(\mathbf{a}\times\mathbf{b}) = a_x(a_y b_z - a_z b_y) + a_y(a_z b_x - a_x b_z) + a_z(a_x b_y - a_y b_x) = 0, \tag{13.16}$$

as desired; you can check that all of the terms cancel in pairs. And likewise for $\mathbf{b}\cdot(\mathbf{a}\times\mathbf{b})$. So $\mathbf{a}\times\mathbf{b}$ is perpendicular to both $\mathbf{a}$ and $\mathbf{b}$.

However, there is still an ambiguity in the direction of $\mathbf{a}\times\mathbf{b}$, because although we know that $\mathbf{a}\times\mathbf{b}$ points along the direction perpendicular to the plane spanned by $\mathbf{a}$ and $\mathbf{b}$, there are two possible directions along this line. Assuming that our coordinate system has been chosen to be "right-handed" (which means that if you point the fingers of your right hand in the direction of $\hat{\mathbf{x}}$ and then swing them to $\hat{\mathbf{y}}$, your thumb points along $\hat{\mathbf{z}}$), then the direction of $\mathbf{a}\times\mathbf{b}$ is determined by the right-hand rule. That is, if you point the fingers of your right hand in the direction of $\mathbf{a}$ and then swing them to $\mathbf{b}$ (through the angle that is less than 180°), then your thumb points along $\mathbf{a}\times\mathbf{b}$. This is consistent with the fact that Eq. (13.14) gives $(1,0,0)\times(0,1,0) = (0,0,1)$, which is the statement that $\hat{\mathbf{x}}\times\hat{\mathbf{y}} = \hat{\mathbf{z}}$.

Let's now demonstrate the result in Eq. (13.15), which is equivalent to $|\mathbf{a}\times\mathbf{b}|^2 = a^2b^2(1-\cos^2\theta)$, which in turn is equivalent to $|\mathbf{a}\times\mathbf{b}|^2 = a^2b^2 - (\mathbf{a}\cdot\mathbf{b})^2$. Written in terms of the components, this last equation is

$$\begin{aligned}(a_y b_z - a_z b_y)^2 + (a_z b_x - a_x b_z)^2 + (a_x b_y - a_y b_x)^2 &= (a_x^2 + a_y^2 + a_z^2)(b_x^2 + b_y^2 + b_z^2) \\ &\quad - (a_x b_x + a_y b_y + a_z b_z)^2. \end{aligned} \tag{13.17}$$

If you stare at this long enough, you'll see that it's true. The three different types of terms agree on both sides. For example, both sides have an $a_y^2 b_z^2$ term, a $-2a_y b_y a_z b_z$ term, and no $a_x^2 b_x^2$ term.

13.2 Appendix B: Taylor series

13.2.1 Basics

Taylor series are extremely useful for checking limiting cases, in particular in situations where a given parameter is small. A Taylor series expresses a given function of x as a series expansion in powers of x. The general form of a Taylor series is (the primes here denote differentiation)

$$f(x_0 + x) = f(x_0) + f'(x_0)x + \frac{f''(x_0)}{2!}\,x^2 + \frac{f'''(x_0)}{3!}\,x^3 + \cdots . \tag{13.18}$$

This equality can be verified by taking successive derivatives of both sides of the equation and then setting $x = 0$. For example, taking the first derivative and then setting $x = 0$ gives $f'(x_0)$ on the left. And this operation also gives $f'(x_0)$ on the right, because the first term is a constant and gives zero when differentiated, the second term gives $f'(x_0)$, and all of the rest of the terms give zero once we set $x = 0$ because they all contain at least one power of x. Likewise, if we take the second derivative of each side and then set $x = 0$, we obtain $f''(x_0)$ on both sides. And so on for all derivatives. Therefore, since the two functions on each side of Eq. (13.18) are equal at $x = 0$ and also have their nth derivatives equal at $x = 0$ for all n, they must in fact be the same function (assuming that they're nicely behaved functions, as we generally assume in physics).

Some specific Taylor series that often come up are listed below. They are all expanded around $x = 0$; that is, $x_0 = 0$ in Eq. (13.18). We use these series countless times throughout this book when checking how expressions behave in the limit of some small quantity. The series are all derivable via Eq. (13.18), but sometimes there are quicker ways of obtaining them. For example, Eq. (13.20) is most easily obtained by taking the derivative of Eq. (13.19), which itself is simply the sum of a geometric series.

$$\frac{1}{1+x} = 1 - x + x^2 - x^3 + \cdots \tag{13.19}$$

$$\frac{1}{(1+x)^2} = 1 - 2x + 3x^2 - 4x^3 + \cdots \tag{13.20}$$

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots \tag{13.21}$$

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots \tag{13.22}$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots \tag{13.23}$$

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots \tag{13.24}$$

$$\sqrt{1+x} = 1 + \frac{x}{2} - \frac{x^2}{8} + \cdots \tag{13.25}$$

$$\frac{1}{\sqrt{1+x}} = 1 - \frac{x}{2} + \frac{3x^2}{8} + \cdots \tag{13.26}$$

$$(1+x)^n = 1 + nx + \binom{n}{2}x^2 + \binom{n}{3}x^3 + \cdots \tag{13.27}$$

These series might look a little scary, but in most situations there is no need to include terms beyond the first-order term in x. For example, $\sqrt{1+x} \approx 1 + x/2$ is usually a good enough approximation. The smaller x is, the better the approximation is, because any term in the expansion is smaller than the preceding term by a factor of order x. Note that you can quickly verify that the $\sqrt{1+x} \approx 1 + x/2$ expression is valid to first order in x, by squaring both sides to obtain $1 + x \approx 1 + x + x^2/4$. Similar reasoning at second order shows that $-x^2/8$ is correctly the next term in the expansion.

As mentioned in Footnote 4 in Chapter 1, we won't worry about taking derivatives to rigorously derive all of the above Taylor series. We'll just take them as given, which means that if

you haven't studied calculus yet, that's no excuse for not using Taylor series! Instead of deriving them, let's just check that they're believable. This can easily be done with a calculator. For example, consider what e^x looks like if x is a very small number, say, $x = 0.0001$. Your calculator (or a computer, if your want more digits) will tell you that

$$e^{0.0001} = 1.0001000050001666\ldots \tag{13.28}$$

This can be written more informatively as

$$\begin{aligned} e^{0.0001} &= 1.0 \\ &+ 0.0001 \\ &+ 0.000000005 \\ &+ 0.0000000000001666\ldots \\ &= 1 + (0.0001) + \frac{(0.0001)^2}{2!} + \frac{(0.0001)^3}{3!} + \cdots . \end{aligned} \tag{13.29}$$

This last line agrees with the form of the Taylor series for e^x in Eq. (13.22). If you made x smaller (say, 0.000001), then the same pattern would form, but just with more zeros between the numbers than in Eq. (13.28). If you kept more digits in Eq. (13.28), you could verify the $x^4/4!$ and $x^5/5!$, etc., terms in the e^x Taylor series. But things aren't quite as obvious for these terms, because we don't have all the nice zeros as we do in the first 12 digits of Eq. (13.28).

Note that the left-hand sides of all of the Taylor series listed above involve only 1's and x's. So how do we make an approximation to an expression of the form, say, $\sqrt{N+x}$? We could of course use the general Taylor-series expression in Eq. (13.18) and generate the series from scratch by taking derivatives. But we can save ourselves some time by making use of the similar-looking series in Eq. (13.25). We can turn the N into a 1 by factoring out an N from the square root, which gives $\sqrt{N}\sqrt{1+x/N}$. Having generated a 1, we can now apply Eq. (13.25), with the only modification being that the small quantity x that appears in that equation is replaced by the small quantity x/N. This gives (to first order in x)

$$\sqrt{N+x} = \sqrt{N}\sqrt{1+\frac{x}{N}} \approx \sqrt{N}\left(1+\frac{1}{2}\frac{x}{N}\right) = \sqrt{N} + \frac{x}{2\sqrt{N}} . \tag{13.30}$$

Again, you can quickly verify that this expression is valid to first order in x by squaring both sides. If $N = 100$ and $x = 1$, then this approximation gives $\sqrt{101} \approx 10 + 1/20 = 10.05$, which is very close to the actual value of $\sqrt{101} = 10.0499\ldots$.

13.2.2 How many terms to keep?

When making a Taylor-series approximation, how do you know how many terms in the series to keep? For example, if the exact answer to a given problem takes the form of $e^x - 1$, then the Taylor series $e^x \approx 1 + x$ tells us that our answer is approximately equal to x. You can check this by picking a small value for x (say, 0.01) and plugging it in your calculator. This approximate form makes the dependence on x (for small x) much more transparent than the original expression $e^x - 1$ does.

But what if our exact answer had instead been $e^x - 1 - x$? The Taylor series $e^x \approx 1 + x$ would then yield an approximate answer of zero. And indeed, the answer *is* approximately zero. However, when making approximations, it is generally understood that we are looking for the *leading-order* term in the answer (that is, the smallest power of x with a nonzero coefficient). If our approximate answer comes out to be zero, then that means we need to go (at least) one term further in the Taylor series, which means $e^x \approx 1 + x + x^2/2$ in the present case. Our approximate answer is then $x^2/2$. (You should check this by letting $x = 0.01$.) Similarly, if the exact answer had instead been $e^x - 1 - x - x^2/2$, then we would need to go out to the term of order x^3 in the Taylor series for e^x.

Be sure to be consistent in the powers of x that you deal with. If the exact answer is, say, $e^x - 1 - x - x^2/3$, and if you use the Taylor series $e^x \approx 1 + x$, then you will obtain an approximate

answer of $-x^2/3$. This is incorrect, because it is inconsistent to pay attention to the $-x^2/3$ term in the exact answer while ignoring the corresponding $x^2/2$ term in the Taylor series for e^x. Including both terms gives the correct approximate answer as $x^2/6$.

So what is the answer to the above question: How do you know how many terms in the series to keep? Well, the answer is that before you do the calculation, there's really no way of knowing how many terms to keep. The optimal strategy is probably to just hope for the best and start by keeping only the term of order x. This will often be sufficient. But if you end up with a result of zero, then you can go to order x^2, and so on. Of course, you could play it safe and always keep terms up to, say, fourth order. But that is invariably a poor strategy, because you will probably never need to go out that far in a series.

13.2.3 Dimensionless quantities

Note that whenever you use a Taylor series from the above list to make an approximation in a physics problem, the parameter x must be *dimensionless*. If it weren't dimensionless, then the terms with the various powers of x in the series would all have different units, and it makes no sense to add terms with different units.

As an example of an expansion involving a properly dimensionless quantity, consider the approximation made in going from Eq. (1.3) to Eq. (1.4) in the beach-ball example in Chapter 1. In this setup, the small dimensionless quantity x is the bt/m term that appears in the exponent in Eq. (1.3). This quantity is indeed dimensionless, because from the original expression for the drag force, $F_d = -bv$, we see that b has units of N/(m/s), or equivalently kg/s. Hence bt/m is dimensionless.

We can restate the above dimensionless requirement in a more physical way. Consider the question, "What is the velocity $v(t)$ in Eq. (1.3), in the limit of small t?" This question is meaningless, because t has dimensions. Is a year a large or small time? How about a hundredth of a second? There is no way to answer this without knowing what situation we're dealing with. A year is short on the time scale of galactic evolution, but a hundredth of a second is long on the time scale of a nuclear process. It makes sense only to look at the limit of a large or small *dimensionless* quantity. And by "large or small," we mean compared with the number 1.

Equivalently, in the beach-ball example the quantity m/b has dimensions of time, so the value of m/b is a time that is inherent to the system. It therefore *does* make sense to look at the limit where $t \ll m/b$ (that is, $bt/m \ll 1$), because we are comparing two things, namely t and m/b, that have the *same* dimensions. We will sometimes be sloppy and say things like, "In the limit of small t." But you know that we really mean, "In the limit of a small dimensionless quantity that has a t in the numerator," or, "In the limit where t is much smaller than a certain quantity that has dimensions of time."

After you make an approximation, how do you know if it is a "good" one? Well, just as it makes no sense to ask if a dimensionful quantity is large or small without comparing it to another quantity with the same dimensions, it makes no sense to ask if an approximation is "good" or "bad" without stating what accuracy you want. In the beach-ball example, let's say that we're looking at a value of t for which $bt/m = 1/100$. In Eq. (1.4) we kept the bt/m term in the Taylor series for $e^{-bt/m}$, and this directly led to our answer of $-gt$. We ignored the $(bt/m)^2/2$ term in the Taylor series. This is smaller than the bt/m term that we kept, by a factor of $(bt/m)/2 = 1/200$. So the error is roughly half a percent. (The corrections from the higher-order terms will be even smaller.) If this is enough accuracy for whatever purpose you have in mind, then the approximation is a good one. If not, then it's a bad one, and you need to add more terms in the series until you get your desired accuracy.

13.3 Appendix C: Limiting cases, scientific method

In Section 1.1.3 we discussed the strategy of checking limiting (or special) cases after solving a problem. It turns out that checking limiting cases is directly analogous to the scientific method (the procedure of testing hypotheses against experiments, and then modifying the hypotheses if needed). To see how this analogy comes about, recall that the point of checking limiting cases is that although it is often difficult to determine how a system behaves in general, your intuition usually gives you a very good idea of how a system behaves in certain limiting cases. Of course, it is quite possible that your intuition can lead you astray. But for the purposes of the present discussion, we'll assume that your intuition is always correct.

Once you've solved a problem and obtained an answer, checking a limiting case leads to two possible results:

1. Your answer (which you are hoping is correct) doesn't agree with what your intuition says (which we are assuming is correct). In this case you conclude that your answer must be *incorrect.*

2. Your answer *does* agree with what your intuition says. In this case all you can conclude is that your answer *might* be correct. Note well that checking a limiting case can never tell you that your answer is *definitely* correct.

The above two cases are exactly analogous to what happens in real life with the scientific method. In the real world, everything comes down to experiment. If you have a theory or hypothesis that you think is correct, then you need to check that its predictions are consistent with experiments. The specific experiments you do are the analog of the special cases you check after solving a problem. That is, when checking special cases, you check your answer to a problem (what you hope is true) against your intuition about a special case (what you know is true). And when using the scientific method, you check your theory (what you hope is true) against the result of an experiment (what you know is true). This is summarized in the following table.

	Physics class	*Science*
What you *hope* is true	Answer to problem	Theory
What you *know* is true	Intuition about special cases	Result of experiment

In the real world, if your theory *isn't* consistent with the experiments, then you need to go back and fix it, just as you would need to go back and fix your answer to a physics problem if it weren't consistent with the special cases. Your theory might need minor tweaks, or it might be total garbage and need a complete overhaul. If, on the other hand, your theory *is* consistent with the experiments, then although this is nice, the only thing it actually tells you is that your theory *might* be correct. And considering the way things usually turn out, the odds are that it actually isn't correct, but rather the limiting case of a more correct theory.

For example, Newtonian physics (which is what this book is about) is consistent with any "everyday" type of experiment that you might do. But that doesn't mean that Newtonian physics is correct. In fact, Newtonian physics is certainly *not* correct. It's a perfectly good theory for everyday scenarios, but it breaks down when things get very small (when quantum mechanics takes over) or very fast (when relativity takes over). Newtonian physics is simply a limiting theory of these *more* correct theories. And we use the term "more correct" here, because these theories (quantum mechanics and relativity) aren't correct either. They're both just limiting theories of a more correct theory (quantum field theory), which in turn is a limiting theory of something else. And so on and so forth. Turtles, turtles, all the way down – or at least much farther than we can presently see.

Every time you do an experiment, you don't actually prove anything. Instead, what you do is narrow down the possibilities of what *might* be true, by ruling out incorrect theories. And that's *all* you can do. You can solve a physics problem for your class, of course, but no one has solved the problem of the universe yet. So the only thing we can do is perform experiments and narrow

things down. Imagine a multiple-choice question that has a million possible answers to choose from. And imagine that there is no possibility of directly deducing which answer is correct. Then all we can do is start chipping away and eliminating answers by looking at special cases. It's a slow and daunting process, but it's the only process we have.

The collective endeavor of science is therefore to squeeze down the set of possible theories until there's only one left. We have a long way to go, of course! At any stage in the process, the theories that are still on the table are the ones we haven't been able to disprove. So that's how science works. You can't actually *prove* anything, so you learn to settle for the things you can't *disprove.*

Consider, when seeking gestalts,
The theories that science exalts.
It's not that they're known
To be written in stone.
It's just that we can't say they're false.

13.4 Appendix D: Problems requiring calculus

The problems listed below require calculus. "1 M" means "Chapter 1 Multiple Choice," and "1 P" means "Chapter 1 Problems," etc. This list represents less than a sixth of the total number of about 400 questions/problems. So if you haven't studied calculus yet, this book can still be very useful. The chapter introductions also occasionally use calculus, but this won't prevent you from using the given results. We aren't counting Taylor series as a calculus topic, because as mentioned in Footnote 4 in Chapter 1, the application of Taylor series in this book involves only algebra.

1	M:	none
1	P:	none (3 appears to require calculus, but it doesn't actually)
2	M:	3
2	P:	none
3	M:	none
3	P:	2, 3, 8, 9, 10, 12, 22
4	M:	none
4	P:	none
5	M:	none
5	P:	1, 2, 3, 6, 14, 16, 19, 21, 23
6	M:	18, 20
6	P:	7, 13, 14, 22, 23, 24, 25 (also, 5 can be solved with sums or integrals)
7	M:	none
7	P:	6, 10, 11, 12, 14, 22, 33 (1, 2, 3, 4, 5 can be solved with sums or integrals)
8	M:	none
8	P:	2, 4, 11 (1 can be solved with sums or integrals)
9	M:	none
9	P:	none
10	M:	1, 2
10	P:	3, 4, 6, 17, 18, 19
11	M:	none
11	P:	1, 5, 7, 8, 11, 12, 13, 14, 18, 20
12	M:	none
12	P:	4, 6, 8