We break the course up into roughly three pieces, though not sequentially, which we then review in brief, without providing examples. These are roughly as follows:

Integration:

- Integration has two aspects: finding areas under curves and finding antiderivatives.
- These two aspects are related by the two fundamental theorems of calculus.
- The first fundamental theorem of calculus allows us to find areas, and other quantities represented by definite integrals, by finding antiderivatives.
- Substitution and integration by parts are techniques which can be used to find antiderivatives.
- Definite integrals (which are limits of Riemann sums) can be applied to many problems beyond just that of finding the area under a curve: among others, finding velocities and positions, finding volumes, and arc length.
- We can extend the definition of the definite integral to unbounded regions, either vertically (integrating functions which go to infinity) or horizontally (integrating functions on infinite regions).

Differential equations:

- Differential equations allow us to apply integrals to more complicated situations, where rates of change depend on the quantity changing.
- Some differential equations can be solved explicitly by separation of variables. In cases where we cannot solve explicitly we can use Euler's method.
- Plotting slope fields is a way of getting qualitative information about the behaviour of solutions to differential equations.
- The SIR model is a particular system of differential equations which can be used to model disease outbreaks.

Sequences, series, and Taylor series:

- Sequences and series can be used to find approximations to numbers and functions, and also to model limiting behaviours.
- A sequence gives rise to a series through an auxiliary sequence of partial sums.
- Convergence of a series can often be determined by one of the divergence test, the integral test, and the ratio test, which only require us to know the sequence of terms, not the sequence of partial sums.
- Taylor polynomials give approximations to functions at a point, and for most of the functions we deal with, Taylor series give the functions exactly.

Integration

General introduction

Consider a function $f(x)$ defined on an interval $[a, b]$. Suppose for the moment that $f(x) \geq 0$ on $[a, b]$, and that we wish to calculate the area between the graph of $y = f(x)$ and the x-axis over the interval $[a, b]$. For a general function f there is no obvious way to do this using only the geometric formulæfor area with which we are familiar. We may however find an *approximation* to the area by subdividing the interval $[a, b]$ into subintervals $[x_k, x_{k+1}]$, approximating f on each subinterval by its value at some particular point in each subinterval, and finding the total area of the resulting rectangles. In particular, if we use subintervals of equal lengths, and require the point in each subinterval at which we evaluate the function f to be either the left or right endpoint of the interval, then we obtain respectively the *left sum*

$$\sum_{k=0}^{n-1} f(x_k)\Delta x$$

and the *right sum*

$$\sum_{k=1}^{n} f(x_k)\Delta x,$$

where in both cases $\Delta x = \frac{b-a}{n}$ and $x_k = a + k\Delta x$. Here n is the number of subintervals we are using; note that each sum contains exactly n terms.

If a function f is always decreasing on $[a, b]$, then the left sum will be an overestimate to the area under the graph of $y = f(x)$ over that interval, while the right sum will be an underestimate; and if f is always increasing than the opposite is true (the left sum will be an underestimate while the right sum will be an overestimate). For a general function f, however, either or both may be an overestimate or an underestimate.

As mentioned, the sums above give only *approximations* to the area under the graph of f. It seems reasonable, though, that if we were able to somehow take a limit as $n \to \infty$,[1] meaning that we take more and more subintervals (which are therefore smaller and smaller), the sums will approach the *exact* value of the area. For functions which are continuous on $[a, b]$, such limits exist for both the left and right sums, and are moreover equal. We denote this common limit by

$$\int_a^b f(x)\,dx.$$

(Here the quantity dx cannot be left out. One can think of it as representing intuitively an infinitesimal change in x, in the which case $f(x)\,dx$ represents the area of an infinitesimal rectangle of height $f(x)$ and width dx, and the integral represents the sum of the areas of all of these infinitesimal rectangles – of which there are, of course, infinitely many.)

Now this limit is almost impossible to compute for any specific function f, which makes it seem that perhaps we have just replaced one impossible problem (finding an area) with another (finding a limit). This would be the case were it not for the *fundamental theorem of calculus*, which can be stated as follows:

If f is continuous on $[a, b]$, and F is a function on $[a, b]$ which satisfies $F' = f$ there, then

$$\int_a^b f(x)\,dx = F(b) - F(a).$$

We can make sense of this result by consulting the foregoing definition of the definite integral. Let us consider a left sum approximation to $\int_a^b f(x)\,dx$ under the hypotheses in the above theorem. This will be of the form

$$\sum_{k=0}^{n-1} f(x_k)\Delta x = \sum_{k=0}^{n-1} F'(x_k)\Delta x.$$

But now $F'(x_k)$ is the rate of change of f at $x_k = a + k\frac{b-a}{n}$, while $\Delta x = \frac{b-a}{n}$, so that, at least for n large, we have approximately

$$F'(x_k)\Delta x \approx f(x_k + \Delta x) - f(x_k) = F(x_{k+1}) - F(x_k);$$

this is just the tangent-line approximation to F at x_k. If we assume that this approximation is good enough that we can substitute it into the sum above, then we have

$$\begin{aligned}\sum_{k=0}^{n-1} F'(x_k)\Delta x &\approx \sum_{k=0}^{n-1} [F(x_{k+1}) - F(x_k)] \\ &= [F(x_1) - F(x_0)] + [F(x_2) - F(x_1)] + \cdots + [F(x_n) - F(x_{n-1})] \\ &= F(x_n) - F(x_0) = F(b) - F(a).\end{aligned}$$

In other words, we have

$$\sum_{k=0}^{n-1} f(x_k)\Delta x \approx F(b) - F(a),$$

and this approximation will become exact in the limit as $n \to \infty$.[2]

[1] In order to dispel possible misconceptions, we note that the above limit is *not* an example of an infinite series since the individual terms in the sum depend themselves on n (in an infinite sum, for example, a_0 is always a_0 no matter which partial sum we are looking at, but for the right sum above the first term will be $f(a + \frac{b-a}{n})$, which depends on n).

[2] In other words, while the *definition* of $\int_a^b f(x)\,dx$ by itself more or less does just replace one hard problem by another one, it does lead to another much easier method of solution through the fundamental theorem of calculus.

In other words, we have now turned the problem of finding areas into the problem of finding antiderivatives.

Now since the derivative of a constant is always 0[3], a function does not have a unique antiderivative. Any two antiderivatives of a given function will however only differ by a constant. We use the notation $\int f(x)\,dx$ for the set of all antiderivatives of the function f; in other words, if $F' = f$, then we write

$$\int f(x)\,dx = F(x) + C,$$

where C represents an arbitrary constant. In other words, we may think of an expression like $\int f(x)\,dx$ is that it represents the *antiderivative* of f with respect to x in the same way that $\frac{df}{dx}$ represents the *derivative* of f with respect to x.

This allows us to compute areas under the graphs of many simple functions just by inverting what we know about derivatives. For example, we get the following antidifferentiation rules, which we phrase in terms of indefinite integrals:

$$\int x^n\,dx = \frac{1}{n+1}x^{n+1} + C, \quad n \neq -1$$

$$\int \frac{1}{x}\,dx = \ln|x| + C$$

$$\int \cos x\,dx = \sin x + C$$

$$\int \sin x\,dx = -\cos x + C$$

$$\int e^x\,dx = e^x + C$$

$$\int a^x\,dx = \frac{1}{\ln a}a^x + C, \quad a > 0,\ a \neq 1$$

We can obtain antiderivatives of more general functions by inverting our differentiation rules. Thus, for example,

$$\int f(x) + g(x)\,dx = \int f(x)\,dx + \int g(x)\,dx$$

$$\int cf(x)\,dx = c\int f(x), \quad c \text{ constant}$$

(Note carefully that this last rule is most definitely NOT true when c is not a constant!) We may also invert the chain and product rules: this gives rise to two integration techniques which we discuss separately since they are more complicated than the foregoing. These are *substitution* and *integration by parts*.

Substitution

Recall the *chain rule*:

$$\frac{d}{dx}F(g(x)) = F'(g(x))g'(x).$$

This gives the corresponding integration rule

$$\int F'(g(x))g'(x)\,dx = F(g(x)) + C.$$

Now it often happens that one needs to do more than simply apply this rule once; perhaps one has to apply it twice, or apply it and then another rule. Thus we typically rephrase it as a simplification rule, as follows:

$$\int f(g(x))g'(x)\,dx = \int f(u)\,du$$

$$\begin{cases} u = g(x) \\ du = g'(x)\,dx \end{cases}$$

[3]This is something eminently worthy of committing to memory and never forgetting!

What does this mean?[4] By what we said above, $\int f(u)\,du$ represents the (set consisting of every) antiderivative of f *with respect to* u; in other words, if we differentiate $\int f(u)\,du$ with respect to u, we shall get $f(u)$. Now if we differentiate it instead with respect to x, we obtain by the chain rule

$$\begin{aligned}\frac{d}{dx}\int f(u)\,du &= \frac{d}{du}\int f(u)\,du \cdot \frac{du}{dx}\\ &= f(u)\frac{du}{dx} = f(g(x))g'(x),\end{aligned}$$

exactly the function whose antiderivative we sought. This is the sense in which the above equation is true. The equation is true since it allows us to replace one integral by a manifestly simpler one.

In any practical problem the key to applying substitution is to find a function g such that the original integrand can be written in the form $f(g(x))g'(x)$ for some f. Usually what trips us up is the necessity to include the factor $g'(x)$.

Integration by parts

The formula for integration by parts is fortunately rather easier to grasp. Recall the *product rule*:

$$\frac{d}{dx}\left(f(x)g(x)\right) = f'(x)g(x) + g'(x)f(x).$$

If we invert this, we obtain the integration rule

$$\int f'(x)g(x) + g'(x)f(x)\,dx = f(x)g(x) + C.$$

Now most of the time our integrand does not come to us nicely written out in the form $f'(x)g(x)+g'(x)f(x)$, so the above formula cannot be used exactly as written. The method of integration by parts consists in moving one part of the integral on the left-hand side above to the right-hand side, namely

$$\int f'(x)g(x)\,dx = f(x)g(x) - \int g'(x)f(x)\,dx;$$

we may drop the C since it is included in the indefinite integral on the right-hand side.

This method is useful precisely because it often happens that $g'(x)f(x)$ is more amenable to integration than was $f'(x)g(x)$. In applying this method, then, we seek to write our original integrand as a product of two functions, one of which (corresponding to $f'(x)$) is something whose antiderivative we know (or can find), and which is such that $g'(x)f(x)$ is more easily integrated than $f'(x)g(x)$.

Applications of definite integrals

Perhaps the simplest situation where definite integrals can be applied to other than the problem of finding areas is in finding quantities given their rates of change. Thus, consider an object moving along a line, and suppose that we are given the velocity $v(t)$ as a function of time and want to know the position as a function of time. If we let $x(t)$ denote the position as a function of time, then we have by the fundamental theorem of calculus, since $x'(t) = v(t)$ by definition,

$$\int_0^t v(s)\,ds = x(t) - x(0);$$

thus if we are given $x(0)$ we can compute $x(t)$ for all t simply by integrating v. (Note that, as in the second fundamental theorem of calculus, we must use a variable other than t inside the integral; here we use s.)

[4]It is the opinion of the current author that the 'deep meaning' of the above formula is quite difficult to write out in any explicit way. Integration being an inverse process, and substitution being a method of integration, there is something about the process of substitution which is necessarily rather backwards. (One is reminded of crab fugues at this point.) Facility with the method, as with so much else in calculus (and in life), comes with practice. What follows probably cannot hope to be any better than a rather awkward mixture of proof and explanation.

Similarly, we can obtain $v(t)$ if we are given the acceleration $a(t) = v'(t)$ and the initial velocity $v(0)$. (If we are given instead $v(t_0)$ for some t_0 we can proceed in exactly the same fashion, but take the lower limit of our integral to be t_0 instead of 0.)

More interesting applications come from geometry. Specifically, we may use integrals to find volumes of three-dimensional regions and arclengths of curves.

Let us consider volumes first. Suppose that we are given a three-dimensional solid extending along some axis, say from a to b along the x-axis (though it could be the y-axis), and that we are told that, for every x in $[a, b]$, the cross-section of the solid in a plane perpendicular to the x-axis has area $A(x)$. Then the volume of the solid is equal to

$$\int_a^b A(x)\,dx.$$

We can see why this makes sense as follows. Remember that we started out approximating the area under the graph of f by approximating f by constant values on subintervals. We can do something similar here: let us split $[a, b]$ up into n subintervals of equal width $[x_k, x_{k+1}]$, $k = 0, \ldots, n-1$, just as we did for areas, and then approximate the cross-section of the solid on each of these subintervals by the cross-section at x_k (say). Then the volume of the piece corresponding to $[x_k, x_{k+1}]$ will be $A(x_k)\Delta x$, and we obtain the following as an approximation to the total volume:

$$\sum_{k=0}^{n-1} A(x_k)\Delta x.$$

In the limit as $n \to \infty$, it is reasonable that the above approximation should become exact, and the limit of the above sum is by definition the definite integral above.

There is a specific case of the foregoing worth noting. Suppose that we are told that the region arises by taking a two-dimensional region in the xy plane and rotating it around the x-axis. Let us consider two cases: one where the region is that between the graph of $y = f(x)$ and the x-axis, and one where the region is that between the graphs of $y = f(x)$ and $y = g(x)$, $f(x) \geq g(x)$. In the first case, the cross-section at a point x will be a circle of radius $f(x)$ and hence of area $A(x) = \pi\left[f(x)\right]^2$, and the volume will be

$$\int_a^b \pi\left[f(x)\right]^2\,dx.$$

In the second case, the cross-section will be an annulus with outer radius $f(x)$ and inner radius $g(x)$, which has area $\pi\left(\left[f(x)\right]^2 - \left[g(x)\right]^2\right)$, and the volume will be

$$\int_a^b \pi\left(\left[f(x)\right]^2 - \left[g(x)\right]^2\right)\,dx.$$

(Note that we square *before* taking the difference; this is very important!)

There are other similar applications of integrals to problems involving various notions of density. Suppose, for example, that we are told that a certain quantity exists in amount $\delta(x)$ per unit length. Then the total amount of that quantity in an interval $[a, b]$ will be

$$\int_a^b \delta(x)\,dx;$$

this can be shewn by considering breaking the interval $[a, b]$ up into subintervals as we did for volumes just now. A more interesting case is where we have a density in two dimensions, in other words, when we are told that a certain quantity exists in amounts δ per unit area. In general, finding the amount of the quantity in a region will require doing double integrals, which are beyond the scope of this course. In cases where the function has some symmetry though we can reduce the situation to one we can handle. Examples of this were done in class.

One final application of definite integrals is in finding the length of curves. The length of the curve $y = f(x)$ from $x = a$ to $x = b$ can be shewn to be equal to the integral

$$\int_a^b \sqrt{1 + [f'(x)]^2}\, dx.$$

Improper integrals

In our definition of the integral and our statement of the fundamental theorem of calculus, we assumed that we were working with a continuous function on a bounded interval. When one or both of these conditions does not hold, we can sometimes still determine the value of the integral by replacing evaluation at one of the bounds by a limit. We distinguish two cases: one where the integrand $f(x)$ is discontinuous at an endpoint, and one where one (or both) of the endpoints is $\pm\infty$. In the first case, suppose that $f(x)$ is continuous on $[a, b]$ except at $x = a$. Then if F is any antiderivative of f on $(a, b]$, we have for any $c \in (a, b]$ that

$$\int_c^b f(x)\, dx = F(b) - F(c).$$

If $\lim_{c\to a^+} F(c)$ exists, then it makes sense to *define*

$$\int_a^b f(x)\, dx = F(b) - \lim_{c\to a^+} F(c).$$

A similar definition can be given if f is discontinuous at b:

$$\int_a^b f(x)\, dx = \lim_{c\to b^-} F(c) - F(a).$$

Note that in both cases the direction of the limit is chosen so that the quantity c is always inside the interval (a, b).

Now suppose that f is continuous everywhere but $b = +\infty$. If $c > a$ then we have

$$\int_a^c f(x)\, dx = F(c) - F(a);$$

if $\lim_{c\to+\infty} F(c)$ exists, then again it makes sense to *define*

$$\int_a^{+\infty} f(x)\, dx = \lim_{c\to+\infty} F(c) - F(a).$$

A similar definition can be given if $a = -\infty$.

More complicated situations can be handled by breaking the original integral up into a sum of integrals. Thus, for example, if f is discontinuous at both endpoints, we may write

$$\int_a^b f(x)\, dx = \int_a^c f(x)\, dx + \int_c^b f(x)\, dx,$$

where $c \in (a, b)$ is arbitrary, and evaluate each of these integrals as indicated previously. (If f is discontinuous at an interior point in the interval, say at $c \in (a, b)$, then we may also break up the integral this way.) We may do something similar if $a = -\infty$ and $b = +\infty$.

Integrals defined by limits as above are called *improper integrals*. When the appropriate limits exist we say that the integral *converges*.

Sometimes it is useful to know whether a particular integral converges even in cases where we cannot compute its value exactly. We may use the following comparison test:

> Let $f(x)$ and $g(x)$ be two functions satisfying $0 \leq f(x) \leq g(x)$. If an improper integral for g converges, then the same improper integral for f also converges; while if an improper integral for f diverges, then the same improper integral for g also diverges.

In other words, smaller than convergent is convergent, and greater than divergent is divergent.

Differential equations

It often happens in modelling problems that the rate of change of a particular quantity depends on that quantity itself; for example, the rate of growth of bacteria in a culture depends on the amount of bacteria, how fast a hot object cools depends on how much hotter it is than the surrounding air, and so forth. Situations like this can be modelled by *differential equations*.

The most general type of differential equation we usually consider is an equation of the form

$$y' = f(x, y). \tag{1}$$

A *solution* to such an equation is a function $g(x)$ such that $y = g(x)$ satisfies the above equation in the sense that

$$y' = g'(x) = f(x, y) = f(x, g(x));$$

in other words, for every point x (in some interval), the slope of the curve $y = g(x)$ at the point $(x, g(x))$, which is what is represented by the left-hand side of the equation (1), is exactly equal to the right-hand side, namely $f(x, y) = f(x, g(x))$.

In the event that the right-hand side of (1) can be written as a product of a function of x and a function of y, say $h(x)k(y)$, we can sometimes find a solution analytically as follows:

$$\begin{aligned} y' = \frac{dy}{dx} &= h(x)k(y) \\ \frac{dy}{k(y)} &= \frac{dx}{h(x)} \\ \int \frac{dy}{k(y)} &= \int \frac{dx}{h(x)}. \end{aligned}$$

Provided that both of these integrals can be evaluated, we will then arrive with a relation giving y in terms of x implicitly.

Note that both of the integrals in the above solution are indefinite integrals; this means that there will be an undetermined constant of integration in the general solution. This is a general property of differential equations: in order to get a unique solution, we must also specify a point through which the solution curve passes. In other words, we do not just solve the equation $y' = f(x, y)$ but rather the *initial value problem*

$$y' = f(x, y), \qquad y(x_0) = y_0,$$

for some point (x_0, y_0). This means that in addition to the solution's satisfying the differential equation in the sense explained above, the graph of the solution curve also passes through the point (x_0, y_0).

As we have seen at many points in this course, many integrals which we can write down cannot be evaluated explicitly in terms of elementary functions. Similarly, most differential equations which we can write down cannot be solved explicitly in terms of elementary functions. For these differential equations, as with integrals, it is often helpful to be able to find approximate solutions using numerical techniques. One of the simplest of these is *Euler's method*; it is very similar to our method of approximating an integral by a left or right Riemann sum. In detail, suppose that we seek the solution to the above initial value problem:

$$y' = f(x, y), \qquad y(x_0) = y_0.$$

Then we know that the solution curve passes through the point (x_0, y_0). Moreover, we also know that the solution curve has slope $y'|_{x_0} = f(x_0, y_0)$ there, since the curve is the graph of a solution to the differential equation. Thus we know that the tangent line to the solution curve at the point (x_0, y_0) has the equation

$$y - y_0 = f(x_0, y_0)(x - x_0). \tag{2}$$

Now we may use this to approximate the solution at a point near x_0. In particular, let $\Delta x > 0$ be small, and let $x_0 = x_1 + \Delta x$. Then the tangent line (2) tells us that the solution at x_1 can be approximated by

$$\begin{aligned} y_1 &= y_0 + f(x_0, y_0)(x_1 - x_0) \\ &= y_0 + f(x_0, y_0)\Delta x. \end{aligned}$$

In other words, the solution curve will pass near the point $(x_1, y_1) = (x_0 + \Delta x, y_0 + f(x_0, y_0)\Delta x)$. If we now start from *this*[5] point, and apply a further tangent-line approximation, we will find that at $x = x_2 = x_1 + \Delta x = x_0 + 2\Delta x$ the solution can be approximated by

$$\begin{aligned} y_2 &= y_1 + f(x_1, y_1)(x_2 - x_1) \\ &= y_1 + f(x_1, y_1)\Delta x. \end{aligned}$$

In general, then, we find that we can approximate the solution at the points $x_n = x_0 + n\Delta x$ by values y_n which satisfy the recurrence relation

$$y_{n+1} = y_n + f(x_n, y_n)\Delta x.$$

This is called *Euler's method*.

A pictorial method of visualising qualitative features of solutions to differential equations which is related to Euler's method is using so-called *slope fields*. The differential equation

$$y' = f(x, y)$$

means that a solution passing through the point (x, y) must have slope $f(x, y)$ there. Thus suppose we make a grid of points on the (x, y) plane, and at each of these points we plot a small line segment with slope equal to the value of f there; i.e., at the point (x, y) we plot a small line segment with slope $f(x, y)$. This plot of small line segments is called the *slope field* of the differential equation. By looking at the slope field we can deduce information about the solutions to the differential equation.

An analogy may be helpful at this point. The information a slope field gives us about a differential equation is roughly similar to the information plotting points $(x, f(x))$ gives us about the graph of a function $y = f(x)$; in other words, we only get discrete information at a select number of points. When we are sketching the graph of a function, we can get a better picture by connecting the points we have plotted with a smooth curve. While we do not do quite the same thing with slope fields, we can get a better idea of parts of the solution by plotting specific solution curves, either determined analytically – in the rare cases where that is possible – or numerically (say by using Euler's method).

Sometimes we are more interested in qualitative information about the behaviour of solutions than in exact quantitative information. Oftentimes the slope field can give us this qualitative information. For example, if we plot the slope field for the equation

$$y' = 1 - y,$$

we will immediately see that all solutions tend towards the line $y = 1$ as $x \to \infty$. We will also see that the line $y = 1$ appears to be a solution to the equation, which can be verified directly since $y = 1$ implies $y' = 0$, $1 - y = 0$, so $y' = 1 - y$. More complicated qualitative information can also be deduced, such as threshold values.

Systems of equations; SIR model

[5]Note that since the exact solution curve does not, in general, pass exactly through this point, we are now taking the tangent line approximation to a solution curve which is only an approximation to our original curve; in other words, we are taking an approximation to an approximation. A detailed study of the error resulting from approximating a solution to a differential equation using Euler's method is an interesting problem in numerical analysis which is unfortunately beyond the scope of the present course. Here we simply note that the error can be estimated.

Often we run into cases where we wish to model more than one value at the same time, and where the rates of change depend on multiple quantities. Examples are the robin-worm and SIR models discussed in section 11.8 of the textbook. Here let we look at a simpler example. An object of mass 1 connected to a spring with spring constant 1 (both in whatever system of units we happen to be using) which oscillates along a line has a coordinate x and velocity v which satisfy the system of equations

$$\begin{aligned}\frac{dx}{dt} &= v \\ \frac{dv}{dt} &= -x.\end{aligned}$$

Now it can be shown (and is not hard to check) that if we give also the initial conditions $x(0) = 1$, $v(0) = 0$, then the solution to this system is $x = \cos t$, $v = -\sin t$. But suppose that we do not know how to solve this equation to get x and v as functions of t. Is there anything we can do? Well, perhaps we are not just interested in how x and v depend on t, but also in how they depend on each other. Then we can see whether we might be able to find a differential equation relating just x and v.[6] Now by the chain rule we have

$$\frac{dx}{dv} = \frac{dx}{dt} \cdot \frac{dt}{dv}.$$

Similarly, by the chain rule we may write[7]

$$\frac{dt}{dt} = 1 = \frac{dt}{dv}\frac{dv}{dt},$$

so

$$\frac{dt}{dv} = \frac{1}{\frac{dv}{dt}},$$

and we have

$$\begin{aligned}\frac{dx}{dv} &= \frac{\frac{dx}{dt}}{\frac{dv}{dt}} \\ &= \frac{v}{-x} = -\frac{v}{x}.\end{aligned}$$

This is clearly separable, and applying separation of variables gives

$$\begin{aligned}-x\,dx &= v\,dv \\ -\frac{1}{2}x^2 &= \frac{1}{2}v^2 + C \\ v^2 + x^2 &= -2C,\end{aligned}$$

which is, for $C \leq 0$, the equation of a circle with centre $(0, 0)$ and radius $\sqrt{-2C}$. In our example above, we have $x(0) = 1$, $v(0) = 0$, which gives $-2C = 1$, so that the point (x, v) will be on the unit circle. In other words, while it may not have been a simple matter to determine where on the circle x and v are at a certain time t, we can say for certain that having started out on the unit circle, the point (x, v) will stay on the unit circle for all time.

We note one other feature which the above manipulation reveals. What happens if $C = 0$? A circle of radius 0 is just a point! Now for C to be zero we must have $x(t) = v(t) = 0$ for some t (for example, $t = 0$), and so what this says is that if at any point in time we have $x(t) = v(t) = 0$, then we must have $x(t) = v(t) = 0$ for all other times. A solution with this property is called a *stationary solution* or an *equilibrium solution*, since it does not depend on time.

[6] We hasten to point out that for general systems this is not possible. The robin-worm and SIR models, as well as the mass-on-a-spring system we consider here, are special cases in this sense.

[7] We are assuming that the function $v = v(t)$ can be inverted to obtain $t = t(v)$, at least in principle

The SIR model discussed in the textbook has another feature. This model is given by (see p. 617 of the textbook)

$$\frac{dS}{dt} = -aSI$$
$$\frac{dI}{dt} = aSI - bI.$$

Here S represents the number of people susceptible to a certain disease in a population, I represents the number of people infected, a represents the rate at which infected people infect susceptible people, and b represents the rate at which infected people recover. Now for this system it is not a simple matter to find an exact expression for S and I as functions of time. (In the textbook example curves are shown which were found numerically.) We can however attempt to find a curve of I versus S the way we did for the example system above. If we divide $\frac{dI}{dt}$ by $\frac{dS}{dt}$ as we did above, we obtain

$$\frac{\frac{dI}{dt}}{\frac{dS}{dt}} = \frac{dI}{dS} = \frac{aSI - bI}{-aSI}$$
$$= -1 + \frac{b}{aS}$$
.

As before, this equation may be solved by separation of variables:

$$dI = \left(-1 + \frac{a}{bS}\right) dS$$
$$I = -S + \frac{a}{b} \ln S + C.$$

We may then plot I versus S and examine the resulting curves.

One particular feature of the above system can be determined directly from the equations. Since a, S, and I are always positive, we see that $\frac{dS}{dt}$ is always negative; in other words, S always decreases with time. Now suppose that initially S is such that $-1 + \frac{b}{aS} > 0$, i.e., that $S < \frac{b}{a}$. Since S decreases with time, this means that $S < \frac{b}{a}$ for all time, so that $-1 + \frac{b}{aS} = \frac{dI}{dS} > 0$ for all time. This means that I will always move in the same direction as S: i.e., if S increases, so does I; if S decreases, so does I. Now since S always decreases with time, I must also always decrease with time – in other words, no matter how many infections we start out with, the number of infections will always decrease with time, which means that the disease cannot spread out of control. This point $S = \frac{b}{a}$ is called a *threshold value*: the number of infections can only increase with time if the initial number of susceptible people is greater than this value. See the textbook for additional discussion.

Sequences, series, and Taylor series

[Since fairly extensive notes for much of this section of the course have already been posted, I will keep this section brief.]

There are many situations in this course where we have talked about approximating a certain quantity to a higher and higher degree of accuracy. Thus, for example, integrals are approximated by Riemann sums, and the accuracy gets better if we take more and more subintervals; we have just discussed using Euler's method to approximate solutions to differential equations, and there too the approximation gets better as the number of steps gets larger (equivalently, as the step size Δx gets smaller).[8] Another example is given by finding square roots, as we discussed in class (we can find a sequence of rational numbers getting closer and closer to the square root of 2, for example).

[8]We are being a bit loose here. In particular it may not be the case that the accuracy increases monotonically with the number of subintervals: there are some cases where a 4-subinterval estimate, for example, may be worse than a 2-subinterval estimate. Our statement is to be taken in the limiting sense that we can make the accuracy as good as we want by taking the number of subintervals sufficiently large.

In general, a *sequence* is an ordered list $\{s_n\}$ of numbers $s_0, s_1, \ldots$,[9] and we say that it has a *limit* S, or that it *converges to* S, if s_n is arbitrarily close to S as long as n is sufficiently large. (Note that this means that s_n must remain arbitrarily close to S as n gets large.) Here are a few simple examples.
EXAMPLES. (a) The sequence given by $s_n = \frac{1}{n}$ converges to 0, since if n is very large, $\frac{1}{n}$ will be a very small positive number.
(b) The sequence given by $s_n = \sin 2\pi n$ converges since it is always equal to the number 0. (This is called a *constant sequence.*)
(c) The sequence given by $s_n = (-1)^n$ does *not* converge since it is close to (in fact, equal to!) 1 some of the time and -1 the rest of the time, and this happens for arbitrarily large values of n.

A sequence which does not converge is said to *diverge*. Note that convergence, divergence, and the value of the limit (when it exists) does not depend on the first terms of the sequence, but only on the terms s_n for sufficiently large n.

Our interest in sequences is mostly for their application to *series*. Suppose that we have a sequence $\{a_n\}_{n=0}^{\infty}$. Then the *infinite series*

$$\sum_{k=0}^{\infty} a_k,$$

when it converges, is defined as follows. For every $n \geq 0$ define the nth *partial sum*

$$S_n = \sum_{k=0}^{n} a_k,$$

i.e., the sum of the first $n+1$ (since we are starting with $k = 0$!) terms of the sequence. This gives us a new sequence $\{S_n\}_{n=0}^{\infty}$, called the sequence of partial sums. If this sequence has a limit, say S, then we call this limit the sum of the infinite series $\sum_{k=0}^{\infty} a_k$, and say that the series *converges* to this value.[10]

As with integrals, it is generally not possible, even for very simple sequences $\{a_n\}$, to find an explicit form for the partial sums S_n. Fortunately, it is often possible to determine whether a series converges by working directly with the terms $\{a_n\}$ instead of the partial sums. Specifically, we have the following three tests:

1. The DIVERGENCE TEST: If $\lim_{n\to+\infty} a_n$ does not exist or exists but does not equal 0, then $\sum a_n$ diverges.
2. The RATIO TEST: If $\lim_{n\to+\infty} \left|\frac{a_{n+1}}{a_n}\right| = L$, then $\sum a_n$ converges if $L < 1$ and diverges if $L > 1$, while if $L = 1$ it may either converge or diverge.
3. The INTEGRAL TEST: If $a_n = f(n)$, where $f(x)$ is a function which is positive and decreasing, then $\sum a_n$ converges if $\int_b^{\infty} f(x)\,dx$ converges, and $\sum a_n$ diverges if $\int_b^{\infty} f(x)\,dx$ diverges.

The most important series for us are the *power series*, which are the *Taylor series* of their limits where these exist. More precisely, suppose that we have a function $f(x)$ on an interval $(a-r, a+r)$ for some $r > 0$, and that all the derivatives of f exist at the point a. Then the *Taylor series of f around the point a* is the series

$$\sum_{k=0}^{\infty} \frac{1}{k!} f^{(k)}(a)(x-a)^k.$$

Note that, assuming it is permissible to differentiate the series term-by-term, this is exactly the series whose kth derivative at $x = a$ is equal to that of f. Similarly, if we have a power series

$$\sum_{k=0}^{\infty} a_k (x-a)^k \tag{3}$$

[9]It makes no fundamental difference to anything we do in this paragraph whether the index n starts at 0, or 1, or 235, or 2,957,120, or any other number. If we wish to explicitly specify ths starting point, we may write, for example, $\{s_n\}_{n=0}^{\infty}$.
[10]Technically, the series $\sum_{k=0}^{\infty} a_k$ only exists if the limit of the sequence of partial sums converges, and hence it is meaningless to write it out until we know whether this sequence converges. We uniformly ignore this technicality and gleefully write $\sum_{k=0}^{\infty} a_k$ whenever convenient regardless of whether it actually exists or not.

which converges to the function f for x near a, then it must be the Taylor series for f, i.e., we must have

$$a_k = \frac{1}{k!} f^{(k)}(a).$$

The convergence of a general power series may be determined using the ratio test. This gives rise to the following. Consider the power series in (3). Consider the limit $\lim_{k\to\infty} \left|\frac{a_{k+1}}{a_k}\right|$. If this limit exists and is equal to some positive number, say $\frac{1}{R}$, then the power series will converge on the interval $(a-R, a+R)$ and we call the number R the *radius of convergence*. If the limit exists and equals zero, then the power series converges for all x; in this case we say that the radius of convergence is infinite. If the limit does not exist then we must do something else to determine convergence of the series.

In the first case above, the series may converge at either, both, or neither of the endpoints of the interval $(a-R, a+R)$. The set of all points on which the series converges is called the *interval of convergence*; in the case under consideration, it is always one of the four intervals

$$(a-R, a+R), \qquad [a-R, a+R), \qquad (a-R, a+R], \qquad [a-R, a+R].$$

Convergence at the endpoints must be determined by a test other than the ratio test as the ratio test always fails at the endpoints by definition of R.[11]

The following series expansions should be memorised:

$$\begin{aligned}
\frac{1}{1+x} &= \sum_{k=0}^{\infty} x^k, \qquad x \in (-1, 1) \\
e^x &= \sum_{k=0}^{\infty} \frac{x^k}{k!}, \qquad x \in (-\infty, +\infty) \\
\cos x &= \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)!} x^{2k}, \qquad x \in (-\infty, +\infty) \\
\sin x &= \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!} x^{2k+1}, \qquad x \in (-\infty, +\infty).
\end{aligned}$$

(It is worth noting here that when we speak of the nth order term of a Taylor series, we mean the term containing x^n, *not* the nth nonzero term, which is generally something different. Thus, for example, the second order term in the Taylor series for $\sin x$ around $x = 0$ is 0 since the series above does not have a quadratic term. It is *not* $-\frac{1}{6}x^3$, which is the second nonzero term.)

Finding Taylor series does not always require finding derivatives as new Taylor series can be obtained by algebraic manipulations, composition, and differentiation and integration. In particular, if the function f has a Taylor series

$$\sum_{k=0}^{\infty} a_k (x-a)^k$$

(where of course we must have $a_k = \frac{1}{k!} f^{(k)}(a)$ as always) which converges on $(a-r, a+r)$, then on this interval the function f' is represented by the Taylor series

$$\sum_{k=1}^{\infty} k a_k (x-a)^{k-1},$$

[11] Note that convergence of a Taylor series does not imply that the series actually converges to the function. However, the functions we generally deal with do have Taylor series which converge to the original function, when they converge at all.

which is exactly the term-by-term derivative of the series for f. Similarly, series may also be integrated term-by-term, and we may apply normal algebraic and functional composition to them. An example of this last procedure is the series for e^{-x^2}:

$$e^{-x^2} = \sum_{k=0}^{\infty} \frac{1}{k!} \left(-x^2\right)^k = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} x^{2k}.$$

Finally, Taylor series can also be used to find solutions to differential equations, either exact solutions or approximations. Suppose, for example, that we have the equation

$$y' = cy + d.$$

This can be easily solved by separation of variables. To solve it using Taylor series, we suppose that y has a series expansion

$$y = \sum_{k=0}^{\infty} a_k (x-a)^k.$$

(If we are given $y|_{x=a}$, then we take a_0 to be this value.) If we assume that this series converges to y on some interval, then we know that we have

$$y' = \sum_{k=1}^{\infty} k a_k (x-a)^{k-1} = \sum_{k=0}^{\infty} (k+1) a_{k+1} (x-a)^k,$$

while

$$cy + d = \sum_{k=0}^{\infty} (ca_k + d)(x-a)^k;$$

for y to satisfy the differential equation, these expressions must be equal, and hence we must have

$$(k+1)a_{k+1} = ca_k + d$$

for all k, so the coefficients a_k must satisfy the recurrence relation

$$a_{k+1} = \frac{ca_k + d}{k+1}.$$

Given a_0, we are thus able to find all other coefficients. In the general case it is not clear how to find a simple expression for a_k, but in the case $d = 0$ it is not hard to see that

$$a_k = a_0 \frac{c^k}{k!},$$

so that the solution is

$$a_0 \sum_{k=0}^{\infty} \frac{c^k}{k!} (x-a)^k = a_0 \sum_{k=0}^{\infty} \frac{1}{k!} \left(c[x-a]\right)^k = a_0 e^{c(x-a)},$$

which is what we would have found with separation of variables.

In the event that we are not able to find a simple expression for a_k, or that having found one we are not able to find a simple expression for the sum of the resulting series, we may still find an approximate solution by keeping only the first several terms of the series.