MATH 1432 (CALCULUS II) LECTURE NOTES
INVITATION-ONLY SECTION

VAUGHN CLIMENHAGA

Contents

Date: July 15, 2020.

Part I. Integration

Lecture 1 Review of integration and the substitution rule

Stewart §5.5, Spivak Ch. 19

1.1. Definite and indefinite integrals

Last semester, we motivated the introduction of integrals by considering the question of how to determine areas. This led us to two definitions:

(1) the definite integral $\int_a^b f(x)\,dx$ is a *number* obtained as a limit of Riemann sums, which depends on the interval $[a,b]$ and can be interpreted as an area;
(2) the indefinite integral $\int f(x)\,dx$ is a *function* whose derivative is $f(x)$.

The two are related by the Fundamental Theorem of Calculus, which has two halves.

The first half says that definite integrals can be used to find indefinite integrals (antiderivatives), since $\frac{d}{dx}\int_a^x f(t)\,dt = f(x)$.

The second half goes in the opposite direction, and says that indefinite integrals can be used to find definite integrals: if $F(x) = \int f(x)\,dx$ is an indefinite integral of f, so that $F'(x) = f(x)$ at every x, then $\int_a^b f(x)\,dx = F(b) - F(a)$.

Although the first half guarantees that every continuous function has an indefinite integral, it does not give a general procedure for writing down an elementary formula for $\int f(x)\,dx$. Our emphasis for the next little while will be on this process, which is essential if we are to use the second half of the FTC effectively.

By "elementary formula", we mean a formula that can be written down in terms of constants, polynomials, rational functions, exponentials, trigonometric functions, and logarithms using addition, subtraction, multiplication, and division. For example, $F(x) = \tan^{-1}(x)$ is an elementary formula, but $F(x) = \int_0^x \frac{1}{1+t^2}\,dt$ is not elementary because it involves an integral, even though it represents the same function.

Given an integral $\int f(x)\,dx$, then, our goal will be to find an elementary formula for it. Bear the following warning in mind, though: not every integral admits an elementary formula. For example, it is possible to show[1] that $\int \sin(x^2)\,dx$ does not have an elementary formula, and in fact there is a sense in which *most* indefinite integrals do not have elementary formulas. Nevertheless, a great many of them do, including some of the most important ones, and so we will turn our attention now to finding them.

1.2. Substitution rule

The first method of integration is by direct inspection: we have a list of functions $F(x)$ whose derivatives $f(x) = F'(x)$ are known, and if f happens to appear on the corresponding list of derivatives, then we can simply read off the indefinite integral $\int f(x)\,dx = F(x) + C$.

[1] The proof involves tools that go beyond the scope of this course, and we will not discuss it.

The second method, which we encountered briefly last semester, is the substitution rule. This is a consequence of the chain rule for differentiation, which says that if F, g are differentiable functions, then $F \circ g$ is differentiable and has $(F \circ g)'(x) = F'(g(x))g'(x)$. In particular, if $F'(x) = f(x)$ so that F gives the indefinite integral of f, then we have $(F \circ g)' = (f \circ g) \cdot (g')$; this can be written in the form

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

It is usually easier to remember and apply this rule if we introduce a new variable $u = g(x)$, and observe that $\frac{d}{du}F(u) = f(u)$, so that the above formula becomes

$$\int f(g(x))g'(x)\,dx = \int f(u)\,du. \tag{1.1}$$

It is common to rewrite the formula $g'(x) = \frac{du}{dx}$ as $du = g'(x)\,dx$, in which case (1.1) appears to become almost trivial:

$$\int f(\underbrace{g(x)}_{u})\,\underbrace{g'(x)\,dx}_{du} = \int f(u)\,du.$$

We emphasize, though, that the formula $du = g'(x)\,dx$ is purely a bookkeeping device rather than a valid part of a proof, because we have not yet given du and dx any independent meaning of their own. We will continue to use it because it simplifies the appearance of various computation, but please remember the logical order of things: (1.1) justifies this formula, rather than the other way round.

Example 1.1. We can compute $\int x\sqrt{1+x^2}\,dx$ by putting $u = 1+x^2$ so that $du = 2x\,dx$, and we obtain

$$\int x\sqrt{1+x^2}\,dx = \int \underbrace{\sqrt{1+x^2}}_{\sqrt{u}} \cdot \underbrace{x\,dx}_{\frac{1}{2}du} = \int \frac{1}{2}u^{1/2}\,du = \frac{1}{2}\cdot\frac{2}{3}u^{3/2} + C = \frac{1}{3}(1+x^2)^{3/2} + C.$$

Example 1.2. To find $\int \tan x\,dx$, we can write $\tan x = \frac{\sin x}{\cos x}$ and notice that the derivative of $\cos x$ appears in the numerator (up to a negative sign), so putting $u = \cos x$ gives $du = -\sin x\,dx$ and

$$\begin{aligned}\int \tan x\,dx = \int \frac{\sin x}{\cos x}\,dx = \int \frac{-du}{u} &= -\ln|u| + C = -\ln|\cos x| + C = \ln|1/\cos x| + C \\ &= \ln|\sec x| + C.\end{aligned}$$

There is no universal procedure telling us how to make the change of variables $u = g(x)$, but these examples illustrate some guidelines that are helpful to keep in mind: it is reasonable to try setting u as the input of some function in the integrand (the square root function in Example 1.1), or as an expression whose derivative also appears in the integrand (the cosine function in Example 1.2). Sometimes it even works to let u be the entire integrand: for example, in $\int \sqrt{2x+1}\,dx$ we can take $u = \sqrt{2x+1}$ so that $u^2 = 2x+1$ and $2u\,du = 2\,dx$, and we get

$$\int \underbrace{\sqrt{2x+1}}_{u}\,\underbrace{dx}_{u\,du} = \int u \cdot u\,du = \frac{1}{3}u^3 + C = \frac{1}{3}(2x+1)^{3/2} + C.$$

Note that the substitution $u = 2x + 1$ would also work here; there is often more than one route to the correct answer!

When computing an indefinite integral via the substitution rule, it is important to remember that the final answer must always be written in terms of the *original* variable, not the substituted one. Thus the last step in each of the above examples was to convert an expression involving u into an expression involving x.

The substitution rule can also be used for definite integrals, either by first computing the indefinite integral and then applying the FTC, or by applying the change of variables $u = g(x)$ to the limits of integration as well.

Example 1.3. To compute $\int_1^2 (1-2x)^{-2}\,dx$, we can write $u = 1-2x$ so that $du = -2\,dx$ and the new integral goes from $u = -1$ to $u = -3$:

$$\int_1^2 \frac{dx}{(1-2x)^2} = -\frac{1}{2}\int_{-1}^{-3} u^{-2}\,du = \frac{1}{2u}\Big|_{-1}^{-3} = \frac{1}{2(-3)} - \frac{1}{2(-1)} = -\frac{1}{6} + \frac{1}{2} = \frac{1}{3}.$$

Lecture 2 Integration by parts

Stewart §7.1, Spivak Ch. 19

2.1. A consequence of the product rule

We found the substitution rule for integrals by looking at the chain rule for derivatives, and exploring its consequences for integrals. We can also do this with the product rule, which says that if f, g are differentiable functions, then

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

This can be rewritten as

$$f(x)g'(x) = \frac{d}{dx}\big(f(x)g(x)\big) - g(x)f'(x) = \frac{d}{dx}\Big(f(x)g(x) - \int g(x)f'(x)\,dx\Big),$$

and we conclude that

$$\int f(x)g'(x)\,dx = f(x)g(x) - \int g(x)f'(x)\,dx. \tag{2.1}$$

We do not write a constant of integration because the right-hand side still contains an indefinite integral. The relationship (2.1) is called *integration by parts* and is a powerful tool for evaluating many integrals, especially when f, g can be chosen so that gf' is easier to integrate than fg'.

Example 2.1. Suppose we want to evaluate $\int x\cos x\,dx$. Then we might try $f(x) = x$ and $g'(x) = \cos x$; to get this, we should put $g(x) = \sin x$, and then (2.1) gives

$$\int x\cos x\,dx = x\sin x - \int \underbrace{(\sin x)}_{g(x)} \cdot \underbrace{1}_{f'(x)}\,dx = x\sin x - (-\cos x) + C = x\sin x + \cos x + C.$$

And indeed, we can verify this by differentiating and using the product rule:

$$\frac{d}{dx}(x\sin x + \cos x) = (\sin x + x\cos x) - \sin x = x\cos x.$$

Remark 2.2. Since antiderivatives are only determined up to a constant, the fact that $g'(x) = \cos x$ actually only tells us that $g(x) = \sin x + C$ for some C. You can check that using this $g(x)$ still gives us the same answer. Because we can choose $g(x)$ to be *any* antiderivative of $g'(x)$, we may as well choose it to be the antiderivative that is the simplest to write down, which usually happens when we put $C = 0$.

Remark 2.3. As with the substitution rule, not all choices are helpful! For example, if we put $f(x) = \cos x$ and $g'(x) = x$ in the example above, we would get $g(x) = \frac{1}{2}x^2$ and $f'(x) = -\sin x$, so

$$\int x\cos x\,dx = \frac{1}{2}x^2\cos x - \int \frac{1}{2}x^2(-\sin x)\,dx = \frac{1}{2}\Big(x^2\cos x + \int x^2\sin x\,dx\Big).$$

We have done nothing wrong – the equation we derived is true – but we have not done anything helpful, either, since we do not know how to evaluate $\int x^2\sin x\,dx$.

We pause a moment to recall that we can write the chain rule in an alternate form by writing $u = g(x)$ and $y = f(u) = f(g(x))$, so that we have the following diagram:

$$\begin{tikzcd}
x \arrow[r, mapsto, "g"'] \arrow[rr, bend left, "f\circ g"] & u = g(x) \arrow[r, mapsto, "f"'] & y = f(u) = f(g(x))
\end{tikzcd}$$

Then the chain rule becomes the very sensible-looking equation $\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}$.

Remark 2.4. It looks like we are simply cancelling the two appearances of the term du, but this is not quite right; we have not given any independent meaning to the symbols dy, du, and dx outside of a derivative like $\frac{dy}{dx}$, or an integral like $\int f(x)\,dx$. Thus this should be regarded as a bookkeeping tool more than anything else; however, it is in some ways easier to remember, and the fact that it conforms to our expectation of how fractions should behave suggests that the notation $\frac{dy}{dx}$ is appropriate to use.

A similar bookkeeping tool is useful for integration by parts. Using the notation $u = f(x)$ and $v = g(x)$, we write $du = f'(x)\,dx$ and $dv = g'(x)\,dx$ (despite the fact that dx, du, and dv have no independent meaning in their own right!) and rewrite (2.1) in the following form, which is easier to remember:

$$\int u\,dv = uv - \int v\,du. \tag{2.2}$$

In Example 2.1 we would put $u = x$, $du = dx$, $dv = \cos x\,dx$, and $v = \sin x$, obtaining the same result as before.

Example 2.5. To evaluate $\int \ln x\,dx$, put $u = \ln x$, $dv = dx$, $du = \frac{1}{x}\,dx$, and $v = x$:

$$\int \underbrace{\ln x}_{u}\underbrace{dx}_{dv} = \underbrace{x}_{v}\underbrace{\ln x}_{u} - \int \underbrace{x}_{v}\underbrace{\frac{1}{x}\,dx}_{du} = x\ln x - \int 1\,dx = x\ln x - x + C.$$

2.2. Iterated integration by parts

Example 2.6. To evaluate $\int t^2 e^t\,dt$, we can put $u = t^2$ and $dv = e^t\,dt$, so $du = 2t\,dt$ and $v = e^t$, giving

$$\int t^2 e^t\,dt = \underbrace{t^2 e^t}_{uv} - \underbrace{\int 2te^t\,dt}_{\int v\,du}. \tag{2.3}$$

To evaluate the last integral we use integration by parts a second time; bring out the factor of 2 and compute $\int te^t\,dt$ by putting $u = t$, $dv = e^t\,dt$, $du = dt$, $v = e^t$, giving

$$\int te^t\,dt = te^t - \int e^t\,dt = te^t - e^t.$$

Using this in (2.3) gives

$$\int t^2 e^t\,dt = t^2 e^t - 2\int te^t\,dt = t^2 e^t - 2(te^t - e^t) + C = t^2 e^t - 2te^t + 2e^t + C.$$

Exercise 2.7. Follow this same approach to show that if $f(t)$ is a polynomial of degree n, then using integration by parts n times gives

$$\int f(t)e^t\,dt = \big(f(t) - f'(t) + f''(t) - \cdots + (-1)^n f^{(n)}(t)\big)e^t + C.$$

Sometimes by using integration by parts multiple times, we end up with an expression that does not yield the integral directly, but which gives an equation that can be solved for it. This is best illustrated with an example.

Example 2.8. To evaluate $\int e^x \sin x\,dx$, we integrate by parts twice:

$$\begin{aligned}\int e^x \sin x\,dx &= -e^x \cos x + \int e^x \cos x\,dx && (u = e^x \text{ and } dv = \sin x\,dx)\\ &= -e^x \cos x + \Big(e^x \sin x - \int e^x \sin x\,dx\Big) && (u = e^x \text{ and } dv = \cos x\,dx).\end{aligned}$$

Since this last expression contains the original integral, one might at first think that we have gotten nowhere. But in fact, we are nearly done! Adding $\int e^x \sin x\,dx$ to both sides of the equation gives

$$2\int e^x \sin x\,dx = e^x(\sin x - \cos x) \qquad \Rightarrow \qquad \int e^x \sin x\,dx = \frac{1}{2}e^x(\sin x - \cos x) + C,$$

where we add a constant of integration to get the most general antiderivative.

2.3. Definite integrals

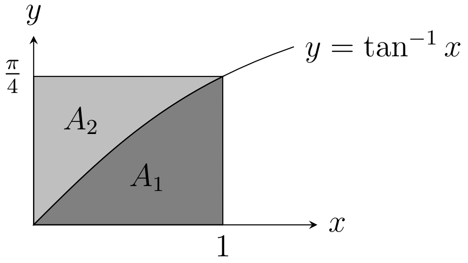
By the FTC, (2.1) has a counterpart for definite integrals:

$$\begin{aligned}\int_a^b f(x)g'(x)\,dx &= \big[f(x)g(x)\big]_a^b - \int_a^b g(x)f'(x)\,dx\\ &= f(b)g(b) - f(a)g(a) - \int_a^b g(x)f'(x)\,dx.\end{aligned} \tag{2.4}$$

Example 2.9. To evaluate $\int_0^1 \tan^{-1} x\,dx$, we put $f(x) = \tan^{-1} x$ and $g'(x) = 1$, so $g(x) = x$ and $f'(x) = \frac{1}{1+x^2}$, giving

$$\begin{aligned}\int_0^1 \tan^{-1} x\,dx &= \big[x\tan^{-1} x\big]_0^1 - \int_0^1 \frac{x}{1+x^2}\,dx && \text{(then substitute } u = 1+x^2\text{)}\\ &= \frac{\pi}{4} - \frac{1}{2}\int_1^2 \frac{1}{u}\,du && \text{(using } du = 2x\,dx\text{)}\\ &= \frac{\pi}{4} - \frac{\ln 2}{2}.\end{aligned}$$

Remark 2.10. The integral $\int_0^1 \tan^{-1} x\,dx$ represents the area A_1 in the diagram below. The area A_2 can be computed by observing that $A_1 + A_2 = \frac{\pi}{4}$, the area of the rectangle, or via the integral $A_2 = \int_0^{\pi/4} \tan y\,dy$, since it is the area to the left of the curve $x = \tan y$. Thus we conclude that $\int_0^{\pi/4} \tan y\,dy = \frac{\pi}{4} - A_1 = \frac{\ln 2}{2}$. This is consistent with the fact that (as we computed last semester using the substitution rule) $\int \tan y\,dy = \ln|\sec y|$ and thus $\int_0^{\pi/4} \tan y\,dy = \ln\sqrt{2}$.



Lecture 3 Trigonometric integrals

Stewart §7.2, Spivak Ch. 19

3.1. Powers of sine

As Remark 2.10 showed, there may be more than one way to correctly calculate a given integral. For another example of this, consider $\int \sin^2 x\,dx$. One approach is to use the identity

$$\cos(2x) = \cos^2 x - \sin^2 x = 1 - 2\sin^2 x \quad\Rightarrow\quad \sin^2 x = \frac{1}{2}(1-\cos(2x)) \tag{3.1}$$

together with the substitution $u = 2x$, $du = 2\,dx$ to get

$$\begin{aligned}\int \sin^2 x\,dx = \frac{1}{2}\int(1-\cos(2x))\,dx &= \frac{x}{2} - \frac{1}{4}\int \cos u\,du\\ &= \frac{x}{2} - \frac{1}{4}\sin u + C = \frac{x}{2} - \frac{1}{4}\sin(2x) + C.\end{aligned} \tag{3.2}$$

A second, equally good, approach is to use integration by parts to get

$$\begin{aligned}\int \sin^2 x\,dx &= \int \underbrace{(\sin x)}_{u}\underbrace{(\sin x)\,dx}_{dv} = \underbrace{(\sin x)}_{u}\underbrace{(-\cos x)}_{v} - \int \underbrace{(-\cos x)}_{v}\underbrace{(\cos x)\,dx}_{du}\\ &= -\sin x\cos x + \int \cos^2 x\,dx = -\sin x\cos x + \int (1-\sin^2 x)\,dx\\ &= -\sin x\cos x + x - \int \sin^2\ dx;\end{aligned}$$

then we can add $\int \sin^2 x\,dx$ to both sides and divide by 2, obtaining

$$\int \sin^2 x\,dx = \frac{1}{2}(x - \sin x\cos x) + C. \tag{3.3}$$

This agrees with (3.2) because $\frac{1}{4}\sin(2x) = \frac{1}{4}\cdot 2\sin x\cos x = \frac{1}{2}\sin x\cos x$.

So why bother with two different approaches? One reason is that they generalize to solve different classes of problems, as we will soon see: for some integrals, the first approach via trigonometric identities and substitution is better, while for others, the second approach via iterated integration by parts has advantages.

3.2. Products of sines and cosines

Let us return to the first way of computing $\int \sin^2 x\,dx$, where we used trigonometric identities and substitutions. Can we use this to compute $\int \sin^n x\,dx$ for other values of n, or more generally $\int \sin^m x\cos^n x\,dx$?

For $n = 3$ we quickly see that the half-angle formula (3.1) does not seem to help:

$$\int \sin^3 x\,dx = \int \sin x\cdot\frac{1}{2}(1-\cos 2x)\,dx = \dots?$$

For $n = 4$, on the other hand, we have

$$\begin{aligned}\int \sin^4 x\,dx &= \int \frac{1}{4}(1-\cos 2x)^2\,dx = \frac{1}{4}\int (1 - 2\cos 2x + \cos^2 2x)\,dx\\ &= \frac{1}{4}x - \frac{1}{4}\sin 2x + \frac{1}{4}\int \cos^2 2x\,dx;\end{aligned}$$

to compute this last integral, observe that (3.1) gives $\cos^2 y = \frac{1}{2}(1+\cos 2y)$, so

$$\int \cos^2 2x\,dx = \frac{1}{2}\int (1+\cos 4x)\,dx = \frac{1}{2}x + \frac{1}{8}\sin 4x + C,$$

and we conclude that

$$\int \sin^4 x\,dx = \frac{1}{4}x - \frac{1}{4}\sin 2x + \frac{1}{4}\Big(\frac{1}{2}x + \frac{1}{8}\sin 4x\Big) + C = \frac{3}{8}x - \frac{1}{4}\sin 2x + \frac{1}{32}\sin 4x + C.$$

A similar approach works for any even power of $\sin x$, but not for odd powers, as the case $n = 3$ illustrates. For odd powers, though, we can use the substitution rule without using a half-angle identity: writing $u = -\cos x$, so that $du = \sin x\,dx$, we get

$$\int \sin^3 x\,dx = \int (\sin^2 x)(\sin x)\,dx = \int (1-\cos^2 x)\sin x\,dx = \int (1-u^2)\,du$$

$$= u - \frac{1}{3}u^3 + C = -\cos x + \frac{1}{3}\cos^3 x + C.$$

The same substitution will work for any odd power, although the computation will become longer. We could similarly compute $\int \cos^n x\, dx$ whenever n is odd by using $u = \sin x$, $du = \cos x\, dx$.

Now suppose that we want to compute $\int \sin^2 x \cos^5 x\, dx$. By making the substitution $u = \sin x$, $du = \cos x\, dx$, we get

$$\begin{aligned}\int \sin^2 x \cos^5 x\, dx &= \int (\sin^2 x)(\cos^2 x)^2 \cos x\, dx = \int u^2(1-u^2)^2\, du \\ &= \int u^2(1-2u^2+u^4)\, du = \int u^2 - 2u^4 + u^6\, du \\ &= \frac{1}{3}u^3 - \frac{2}{5}u^5 + \frac{1}{7}u^7 + C = \frac{1}{3}\sin^3 x - \frac{2}{5}\sin^5 x + \frac{1}{7}\sin^7 x + C.\end{aligned}$$

Indeed, this substitution works to compute $\int \sin^m x \cos^n x\, dx$ whenever $m, n \geq 0$ and n is odd: if $n = 2k+1$, then $u = \sin x$, $du = \cos x\, dx$ gives

$$\int \sin^m x \cos^n x\, dx = \int \sin^m x(1-\sin^2 x)^k \cos x\, dx = \int u^m(1-u^2)^k\, du. \tag{3.4}$$

The last integral can be computed by expanding $(1-u^2)^k$ and using $\int u^\ell\, du = \frac{u^{\ell+1}}{\ell+1}$. In the case when m is odd, the substitution $u = \cos x$, $du = -\sin x\, dx$ lets us do a similar computation. Now we can summarize the overall strategy.

Technique 3.1. To compute $\int \sin^m x \cos^n x\, dx$ when $m, n \geq 0$, do the following:

(1) if n is odd, use the substitution $u = \sin x$, $du = \cos x$ as in (3.4);
(2) if m is odd, use the substitution $u = \cos x$, $du = -\sin x\, dx$;
(3) if n, m are both even, use the trigonometric identities $\sin^2 x = \frac{1}{2}(1-\cos 2x)$ and $\cos^2 x = \frac{1}{2}(1+\cos 2x)$ to rewrite the integral.

Example 3.2. With $m = 4$ and $n = 2$, we use the half-angle formulas to get

$$\begin{aligned}I = \int \sin^4 x \cos^2 x\, dx &= \int \frac{1}{4}(1-\cos 2x)^2 \frac{1}{2}(1+\cos 2x)\, dx \\ &= \frac{1}{8}\int (1 - 2\cos 2x + \cos^2 2x)(1+\cos 2x)\, dx \\ &= \frac{1}{8}\int (1 - \cos 2x - \cos^2 2x + \cos^3 2x)\, dx.\end{aligned}$$

The first two terms are easy to integrate. For the third we use the half-angle formula again to get

$$\int \cos^2 2x\, dx = \int \frac{1}{2}(1+\cos 4x)\, dx = \frac{1}{2}x + \frac{1}{8}\sin 4x + C.$$

For the fourth, we use $u = \sin 2x$ and $du = 2\cos 2x\, dx$ to get

$$\begin{aligned}\int \cos^3 2x\, dx &= \int (1-\sin^2 2x)\cos 2x\, dx = \int (1-u^2)\frac{du}{2} \\ &= \frac{1}{2}u - \frac{1}{6}u^3 + C = \frac{1}{2}\sin 2x - \frac{1}{6}\sin^3 2x + C.\end{aligned}$$

Putting it all together gives

$$\begin{aligned} I &= \frac{1}{8}x - \frac{1}{16}\sin 2x - \frac{1}{8}\Big(\frac{1}{2}x + \frac{1}{8}\sin 4x\Big) + \frac{1}{8}\Big(\frac{1}{2}\sin 2x - \frac{1}{6}\sin^3 2x\Big) + C \\ &= \frac{1}{16}x - \frac{1}{64}\sin 4x - \frac{1}{48}\sin^3 2x + C. \end{aligned}$$

3.3. Products of tangents and secants

The technique above works well enough when $m, n \geq 0$. But what if one or both of them is negative? For the moment we consider the case when cos appears in the denominator, and see that converting the expression to tangents and secants is useful. (When sin is in the denominator, one should use cot and csc instead, and the story is similar. When both sin and cos are in the denominator, things become more difficult, and we will not consider this case.)

Example 3.3. $\displaystyle\int \frac{\sin x}{\cos^2 x}\,dx = \int \tan x \sec x\,dx = \sec x + C.$

More generally, whenever $n \geq m \geq 0$ we can write

$$\int \frac{\sin^m x}{\cos^n x}\,dx = \int \tan^m x \sec^{n-m} x\,dx,$$

so now we will study integrals of the form $\int \tan^m x \sec^k x\,dx$.

Exercise 3.4. If $m > n \geq 0$, show that we can always use the identity $\sin^2 x = 1 - \cos^2 x$ to write $\int \frac{\sin^m x}{\cos^n x}\,dx$ in terms of integrals of products of tangents and secants, as in the following:

$$\int \frac{\sin^3 x}{\cos^2 x} = \int \Big(\frac{\sin x}{\cos^2 x} - \frac{\sin x \cos^2 x}{\cos^2 x}\Big)\,dx = \int (\sec x \tan x - \sin x)\,dx = \sec x + \cos x + C.$$

Since the substitutions $u = \sin x$ and $u = \cos x$ worked in the previous section, it is natural to try the substitutions $u = \tan x$ and $u = \sec x$ to evaluate $\int \tan^m x \sec^k x\,dx$.

- $u = \tan x$ gives $du = \sec^2 x\,dx$, so for this to be effective we need to peel off a factor of $\sec^2 x$ and then be able to use the identity $\sec^2 x = \tan^2 x + 1$:

  $$\int \tan^m x \sec^k x\,dx = \int (\tan^m x \sec^{k-2} x)(\sec^2 x)\,dx = \int u^m (1+u^2)^{\frac{k-2}{2}}\,du.$$

  As long as k is even, this will lead to a polynomial that we can integrate.
- $u = \sec x$ gives $du = \sec x \tan x\,dx$, which helps if we can to remove a factor of $\sec x \tan x$ and then use the identity $\tan^2 x = \sec^2 x - 1 = u^2 - 1$:

  $$\int \tan^m x \sec^k x\,dx = \int (\tan^{m-1} x \sec^{k-1} x)(\sec x \tan x)\,dx = \int (u^2-1)^{\frac{m-1}{2}} u^{k-1}\,du$$

  This leads to a polynomial if m is odd.

Thus for $\int \tan^m x \sec^k x\,dx$, we have the following analogue of Technique 3.1: use the substitution $u = \tan x$ if $k \geq 2$ is even, and $u = \sec x$ if $m \geq 1$ is odd (as long as $k \geq 1$). If k is even and m is odd, then either substitution can be used.

Example 3.5. When $m = 2$ and $k = 4$ we use $u = \tan x$, $du = \sec^2 x\,dx$ to get

$$\begin{aligned}\int \tan^2 x \sec^4 x\,dx &= \int \tan^2 x(1+\tan^2 x)\sec^2 x\,dx = \int u^2(1+u^2)\,du \\ &= \int (u^2+u^4)\,du = \frac{1}{3}u^3 + \frac{1}{5}u^5 + C = \frac{1}{3}\tan^3 x + \frac{1}{5}\tan^5 x + C.\end{aligned}$$

Example 3.6. When $m = k = 5$ we use $u = \sec x$, $du = \sec x \tan x\,dx$ together with $\tan^2 x = \sec^2 x - 1 = u^2 - 1$ to get

$$\begin{aligned}\int \tan^5 x \sec^5 x\,dx &= \int (\tan^2 x)^2 \sec^4 x(\sec x \tan x)\,dx = \int (u^2-1)^2 u^4\,du \\ &= \int (u^8 - 2u^6 + u^4)\,du = \frac{1}{9}\sec^9 x - \frac{2}{7}\sec^7 x + \frac{1}{5}\sec^5 x + C.\end{aligned}$$

So far we have seen that $\int \tan^m x \sec^k x\,dx$ can be computed by the substitution $u = \tan x$ if $k \geq 2$ is even, and by $u = \sec x$ if $m \geq 1$ is odd (unless $k = 0$). The remaining cases not covered by this approach are the following:

(1) $k = 0$, so there are no powers of $\sec x$ to remove.
(2) The power on $\tan x$ is even, and the power on $\sec x$ is odd.

In the first case, we have $\int \tan^m x\,dx$. When $m = 1$ we recall that this can be computed by the substitution $u = \cos x$:

$$\int \tan x\,dx = \int \frac{\sin x}{\cos x}\,dx = -\int \frac{1}{u}\,du = -\ln|\cos x| + C = \ln|\sec x| + C.$$

For $m = 2$ we can use the identity $\tan^2 x = \sec^2 x - 1$ to get

$$\int \tan^2 x\,dx = \int (\sec^2 x - 1)\,dx = \tan x - x + C.$$

For $m \geq 3$, we can use the same identity and the substitution $u = \tan x$ to write

$$\begin{aligned}\int \tan^m x\,dx &= \int \tan^{m-2} x \sec^2 x\,dx - \int \tan^{m-2} x\,dx = \int u^{m-2}\,du - \int \tan^{m-2} x\,dx \\ &= \frac{1}{m-1}\tan^{m-1} x - \int \tan^{m-2} x\,dx.\end{aligned}$$

Iterating this, we eventually reach either $\int \tan x\,dx$ or $\int \tan^2 x\,dx$.

Example 3.7. $\displaystyle\int \tan^3 x\,dx = \frac{1}{2}\tan^2 x - \int \tan x\,dx = \frac{1}{2}\tan^2 x - \ln|\sec x| + C.$

What about the second case above, $\int \tan^{2m} x \sec^{2k+1} x\,dx$? In this case we can still use the identity $\tan^{2m} x = (\tan^2 x)^m = (\sec^2 x + 1)^m$ to write the integral in terms of integrals of the form $\int \sec^{2\ell+1} x\,dx$. But how do we evaluate such integrals?

Lecture 4 More trigonometric integrals

Stewart §7.2, Spivak Ch. 19

4.1. The integral of secant

The last lecture left open the problem of how to evaluate $\int \sec^{2\ell+1} x\, dx$, where $\ell \geq 0$. Let us focus on the case $\ell = 0$ and compute $\int \sec x\, dx$. It turns out that the substitution rule is enough, but we need to be quite clever about how we use it. We have

$$\int \sec x\, dx = \int \frac{1}{\cos x}\, dx = \int \frac{\cos x}{\cos^2 x}\, dx = \int \frac{\cos x}{1 - \sin^2 x}\, dx,$$

which looks like it is moving in the wrong direction (getting more complicated). However, upon putting $u = \sin x$ we get $du = \cos x\, dx$ and thus

$$\int \sec x\, dx = \int \frac{1}{1-u^2}\, du = \int \frac{1}{(1+u)(1-u)}\, du = \frac{1}{2} \int \Big(\frac{1}{1+u} - \frac{1}{1-u} \Big)\, du,$$

where the second equality is natural to do, and the third can be easily checked but is probably not the first thing that would have popped into your head.[2] Once we are at this point, however, we are nearly done! Indeed, since $\int \frac{1}{1+u}\, du = \ln|1+u|$ and $\int \frac{1}{1-u}\, du = -\ln|1-u|$, we have

$$\int \sec x\, dx = \frac{1}{2}\big(\ln|1+u| - \ln|1-u| \big) = \frac{1}{2} \ln\Big| \frac{1+u}{1-u} \Big|,$$

omitting the constant of integration for the time being. Recalling that $u = \sin x$, we multiply top and bottom by $(1+u)$ to obtain $1 - u^2 = 1 - \sin^2 x = \cos^2 x$ in the denominator, and get

$$\int \sec x\, dx = \frac{1}{2} \ln\Big| \frac{(1+u)^2}{1-u^2} \Big| = \frac{1}{2} \ln\Big| \frac{(1+\sin x)^2}{\cos^2 x} \Big| = \ln\Big| \frac{1+\sin x}{\cos x} \Big| = \ln|\sec x + \tan x|.$$

In order to get the most general antiderivative we add the constant of integration:

$$\int \sec x\, dx = \ln|\sec x + \tan x| + C. \tag{4.1}$$

Although this argument does not use any rules that you have not learned yet, it is certainly not one that I would expect you to come up with on your own! It does, however, illustrate a little bit of the nature of computing indefinite integrals; there are many different steps that one might take next at any given stage, and it is a little bit of an art form to decide which one is most likely to be useful. Certain tricks appear over and over again – factor a difference of squares, add and subtract the same thing, multiply and divide by the same thing, look for any useful trigonometric identities that may be relevant – but in the end there is no substitute for just working through lots of problems and gaining practice and experience in integrating.

4.2. More trigonometric identities

One more set of trigonometric identities is worth mentioning at this point.

Exercise 4.1. Use the formulas for $\cos(A \pm B)$ and $\sin(A \pm B)$ to prove that

$$\sin A \cos B = \frac{1}{2}\big[\sin(A+B) + \sin(A-B) \big],$$

[2] It will seem more natural once we have discussed *partial fractions*.

$$\sin A \sin B = \frac{1}{2}\big[\cos(A-B) - \cos(A+B)\big],$$
$$\cos A \cos B = \frac{1}{2}\big[\cos(A-B) + \cos(A+B)\big].$$

These identities can be used to evaluate integrals involving products of $\sin(mx)$ and $\cos(nx)$.

Example 4.2. The first identity above gives

$$\begin{aligned}\int \sin 2x \cos 7x\,dx &= \frac{1}{2}\int \big[\sin(2x+7x) + \sin(2x-7x)\big]\,dx \\ &= \frac{1}{2}\int \sin 9x\,dx - \frac{1}{2}\int \sin 5x\,dx = -\frac{1}{18}\cos 9x + \frac{1}{10}\cos 5x + C.\end{aligned}$$

4.3. *A reduction formula for powers of sine

Now let us return to the second approach given in §3.1 to compute $\int \sin^2 x\,dx$, using integration by parts, and see what happens if we try to use this approach to compute $\int \sin^n x\,dx$. Mimicking the integration by parts from that section, we can write

$$\begin{aligned}\int \sin^n x\,dx &= \int \underbrace{(\sin x)^{n-1}}_{u}\underbrace{(\sin x)\,dx}_{dv} \\ &= \underbrace{(\sin x)^{n-1}}_{u}\underbrace{(-\cos x)}_{v} - \int \underbrace{(-\cos x)}_{v}\underbrace{(n-1)(\sin x)^{n-2}(\cos x)\,dx}_{du} \\ &= -\cos x \sin^{n-1} x + (n-1)\int \sin^{n-2} x \cos^2 x\,dx.\end{aligned}$$

Then using $\sin^{n-2} x \cos^2 x = \sin^{n-2} x(1-\sin^2 x) = \sin^{n-2} x - \sin^n x$, we get

$$\int \sin^n x\,dx = -\cos x \sin^{n-1} x + (n-1)\int \sin^{n-2} x\,dx - (n-1)\int \sin^n x\,dx.$$

Adding $(n-1)\int \sin^n x\,dx$ to both sides gives

$$n\int \sin^n x\,dx = -\cos x \sin^{n-1} x + (n-1)\int \sin^{n-2} x\,dx,$$

and dividing by n we obtain

$$\int \sin^n x\,dx = -\frac{1}{n}\cos x \sin^{n-1} x + \frac{n-1}{n}\int \sin^{n-2} x\,dx. \tag{4.2}$$

This is not a complete answer in and of itself, but it lets us reduce the problem to a similar question for a smaller value of n, and by iterating the procedure we will eventually reach $\int \sin x\,dx$ or $\int \sin^2 x\,dx$, both of which we know how to compute.

Example 4.3. Using $n = 3$ in (4.2) gives

$$\int \sin^3 x\,dx = -\frac{1}{3}\cos x \sin^2 x + \frac{2}{3}\int \sin x\,dx = -\frac{1}{3}\cos x \sin^2 x - \frac{2}{3}\cos x + C.$$

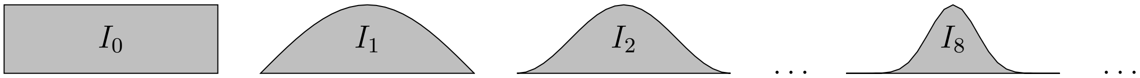
Exercise 4.4. Use integration by parts to prove a similar reduction formula for $\int \sec^2 x\,dx$; together with §4.1 this lets you compute $\int \sec^n x\,dx$ for all $n \in \mathbb{N}$.

4.4. *The Wallis product

Now suppose we compute the definite integrals associated to these examples over the interval $[0,\pi]$. That is, we consider for each $n = 0, 1, 2, 3, \dots$ the real number

$$I_n = \int_0^\pi \sin^n x \, dx.$$

Before doing any computations, observe that the sequence I_n represents the areas of the regions shown here.



Remark 4.5. It appears that these regions are getting smaller and smaller, so that $\lim_{n\to\infty} I_n = 0$. This turns out to be true, but it takes a little bit of work to prove, and we will not do so here.

The first two terms are easy to compute:

$$\begin{aligned} I_0 &= \int_0^\pi 1 \, dx = \pi, \\ I_1 &= \int_0^\pi \sin x \, dx = \big[-\cos x \big]_0^\pi = -\cos\pi + \cos 0 = 2. \end{aligned}$$

For larger values of n, we use the reduction formula (4.2):

$$I_n = \int_0^\pi \sin^n x \, dx = \left[-\frac{1}{n} \cos x \sin^{n-1} x \right]_0^\pi + \frac{n-1}{n} \int_0^\pi \sin^{n-2} x \, dx.$$

Since $\sin 0 = \sin\pi = 0$, the first term on the RHS vanishes, and the last integral is just I_{n-2}, so we get

$$I_n = \frac{n-1}{n} I_{n-2}. \tag{4.3}$$

Thus the next few terms in the sequence are

$$\begin{aligned} I_2 &= \frac{1}{2} I_0 = \pi \cdot \frac{1}{2}, & I_3 &= \frac{2}{3} I_1 = 2 \cdot \frac{2}{3}, \\ I_4 &= \frac{3}{4} I_2 = \pi \cdot \frac{1}{2} \cdot \frac{3}{4}, & I_5 &= \frac{4}{5} I_3 = 2 \cdot \frac{2}{3} \cdot \frac{4}{5}, \end{aligned}$$

and so on. The general formula is

$$I_{2n} = \pi \cdot \frac{1}{2} \cdot \frac{3}{4} \cdot \dots \cdot \frac{2n-1}{2n}, \qquad I_{2n+1} = 2 \cdot \frac{2}{3} \cdot \frac{4}{5} \cdot \dots \cdot \frac{2n}{2n+1}.$$

So far this is kind of cute, but now something surprising happens. We see that there is one rule for the even terms in the sequence I_n, and another rule for the odd terms. What happens if we compare two consecutive terms, one even and one odd? Are they close together, or far apart? Note that since $0 \le \sin x \le 1$ for all $x \in [0,\pi]$, we have $\sin^{n+1} x \le \sin^n x$ for all n, and thus

$$I_{n+1} = \int_0^\pi \sin^{n+1} x \, dx \le \int_0^\pi \sin^n x \, dx = I_n.$$

In particular, this gives $I_{2n} \geq I_{2n+1} \geq I_{2n+2}$, and dividing through by I_{2n} gives

$$1 = \frac{I_{2n}}{I_{2n}} \geq \frac{I_{2n+1}}{I_{2n}} \geq \frac{I_{2n+2}}{I_{2n}} = \frac{2n+1}{2n+2} \quad \text{using (4.3).}$$

Since $\lim_{n\to\infty} \frac{2n+1}{2n+2} = \lim_{n\to\infty} 1 = 1$, the Squeeze Theorem implies that

$$\lim_{n\to\infty} \frac{I_{2n+1}}{I_{2n}} = 1. \tag{4.4}$$

From the formulas for I_{2n} and I_{2n+1}, we see that

$$\frac{I_{2n+1}}{I_{2n}} = \frac{2 \cdot \frac{2}{3} \cdot \frac{4}{5} \cdot \cdots \cdot \frac{2n}{2n+1}}{\pi \cdot \frac{1}{2} \cdot \frac{3}{4} \cdot \cdots \cdot \frac{2n-1}{2n}} = \frac{2}{\pi} \cdot \frac{2}{1} \cdot \frac{2}{3} \cdot \frac{4}{3} \cdot \frac{4}{5} \cdots \frac{2n}{2n-1} \frac{2n}{2n+1}.$$

Together with (4.4), this implies that

$$1 = \lim_{n\to\infty} \left(\frac{2}{\pi} \cdot \frac{2}{1} \cdot \frac{2}{3} \cdot \frac{4}{3} \cdot \frac{4}{5} \cdots \frac{2n}{2n-1} \frac{2n}{2n+1} \right),$$

or if you prefer, after multiplying both sides by $\frac{\pi}{2}$,

$$\frac{\pi}{2} = \lim_{n\to\infty} \left(\frac{2}{1} \cdot \frac{2}{3} \cdot \frac{4}{3} \cdot \frac{4}{5} \cdots \frac{2n}{2n-1} \frac{2n}{2n+1} \right). \tag{4.5}$$

This is the *Wallis product*, a formula for π that was discovered in 1655 by the English mathematician John Wallis. It is often written as an *infinite product*:

$$\frac{\pi}{2} = \frac{2}{1} \cdot \frac{2}{3} \cdot \frac{4}{3} \cdot \frac{4}{5} \cdot \frac{6}{5} \cdot \frac{6}{7} \cdot \frac{8}{7} \cdot \frac{8}{9} \cdots.$$

We will have more to say about expressions like this when we study sequences and series at the end of the course; for the time being I merely urge extreme caution. Because this expression is infinite and not finite, it does not always behave in the way we might expect. For example, one might be tempted to say that because it does not matter in which order we multiply and divide things, we could just as well write the final expression as

$$\frac{2}{3} \cdot \frac{2}{3} \cdot \frac{4}{5} \cdot \frac{4}{5} \cdot \frac{6}{7} \cdot \frac{6}{7} \cdots \tag{4.6}$$

by moving all the denominators one spot to the left. But this turns out to be quite wrong! Indeed, if we write $x_n = \frac{2}{1} \cdot \frac{2}{3} \cdot \frac{4}{3} \cdot \frac{4}{5} \cdots \frac{2n}{2n-1} \frac{2n}{2n+1}$ and $y_n = \frac{2}{3} \cdot \frac{2}{3} \cdot \frac{4}{5} \cdot \frac{4}{5} \cdots \frac{2n}{2n+1} \frac{2n}{2n+1}$, then (4.5) says that $\lim_{n\to\infty} x_n = \frac{\pi}{2}$, and we see clearly that $y_n = \frac{x_n}{2n+1}$, so the product in (4.6) should be interpreted as

$$\lim_{n\to\infty} y_n = \lim_{n\to\infty} \frac{x_n}{2n+1} = 0.$$

But all this is really a discussion for another time, mentioned here merely to illustrate why we should exercise some care when treating infinite expressions.

Lecture 5 Trigonometric substitutions

Stewart §7.3, Spivak Ch. 19

5.1. Reversing the substitution rule

We know that the area of the unit circle is π, so the area under the curve $y = \sqrt{1-x^2}$ from 0 to 1 is $\frac{\pi}{4}$: in other words,

$$\int_0^1 \sqrt{1-x^2}\,dx = \frac{\pi}{4}.$$

Can we compute this integral using the fundamental theorem of calculus, by finding an antiderivative? The function $\sqrt{1-x^2}$ is not on our list of known derivatives, and integration by parts will not get us anywhere. Neither will the first substitutions we might try: $u = x^2$, $u = 1-x^2$, $u = \sqrt{1-x^2}$. On the other hand, there is a substitution we can make that helps, but it looks rather different: instead of replacing x with a new variable that is a function of x, we write x as a function of a new variable t by putting $x = \sin t$. Then we have $dx = \cos t\,dt$, and the usual trigonometric identities give

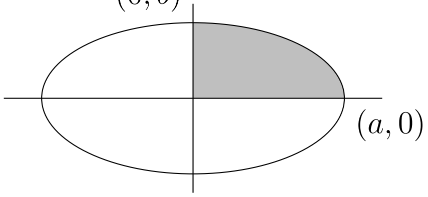
$$\begin{aligned}\int \sqrt{1-x^2}\,dx &= \int \left(\sqrt{1-\sin^2 t}\right)\cos t\,dt = \int \cos^2 t\,dt = \frac{1}{2}\int (1+\cos 2t)\,dt \\ &= \frac{1}{2}t + \frac{1}{4}\sin 2t + C = \frac{1}{2}t + \frac{1}{2}\sin t\cos t + C \\ &= \frac{1}{2}\sin^{-1} x + \frac{1}{2}x\sqrt{1-x^2} + C.\end{aligned}$$

In previous lectures we have made some effort to point out that dx and dt do not have an independent meaning in their own right, so the proper way to justify the above substitution is to define t by $t = \sin^{-1} x$, and then observe that writing $g(t) = \sin t$, the usual substitution rule gives

$$\int f(g(t))g'(t)\,dt = \int f(x)\,dx. \tag{5.1}$$

In this example, though we are going the reverse of the usual direction. In previous applications of the substitution rule, we wanted to find functions f and g so that the integral we were given could be rewritten as the LHS of (5.1), and then transformed into the RHS; in the present example, we start with the integral on the RHS and look for a function g such that transforming it into the LHS is productive.

Remark 5.1. Since sin is not 1-1 on its entire domain, any use of the inverse function $\sin^{-1}$ must always come with a choice of which branch we use. It is standard to choose $\sin^{-1} x \in [-\frac{\pi}{2}, \frac{\pi}{2}]$, which is why we chose $\cos t = \sqrt{1-\sin^2 t}$ instead of $\cos t = -\sqrt{1-\sin^2 t}$ in the above computation; the latter choice would correspond to a different branch of the inverse function, although it would still lead to a valid antiderivative.



Example 5.2. To find the area enclosed by the ellipse $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$, where $a, b > 0$ are the lengths of the two semi-axes of the ellipse, we can solve for y in the first quadrant and get

$$\frac{y^2}{b^2} = 1 - \frac{x^2}{a^2} = \frac{a^2 - x^2}{a^2} \qquad \Rightarrow \qquad y = \frac{b}{a}\sqrt{a^2 - x^2},$$

so that the total area of the ellipse is $A = 4\int_0^a \frac{b}{a}\sqrt{a^2 - x^2}\, dx$. Then we use the substitution $x = a\sin\theta$, $dx = a\cos\theta\, d\theta$, to get

$$\begin{aligned}\int_0^a \sqrt{a^2 - x^2}\, dx &= \int_0^{\pi/2} \sqrt{a^2 - a^2\sin^2\theta} \cdot a\cos\theta\, d\theta = a^2 \int_0^{\pi/2} \cos^2\theta\, d\theta \\ &= \frac{a^2}{2}\int_0^{\pi/2} (1 + \cos 2\theta)\, d\theta = \frac{a^2}{2}\left[\theta + \frac{1}{2}\sin 2\theta\right]_0^{\pi/2} = \frac{\pi a^2}{4},\end{aligned}$$

and we conclude that the area of the ellipse is

$$A = \frac{4b}{a}\int_0^a \sqrt{a^2 - x^2}\, dx = \frac{4b}{a} \cdot \frac{\pi a^2}{4} = \pi ab.$$

Notice that in the case $a = b$ this reduces to the familiar formula for the area of a circle.

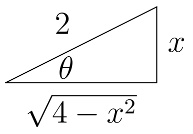
Technique 5.3. Trigonometric substitutions such as the one above are useful for simplifying integrals involving quadratic polynomials inside square roots.

- If you see $\sqrt{a^2 - x^2}$, use $x = a\sin\theta$, $dx = a\cos\theta\, d\theta$, and $1 - \sin^2\theta = \cos^2\theta$.
- For $\sqrt{a^2 + x^2}$, use $x = a\tan\theta$, $dx = a\sec^2\theta\, d\theta$, and $1 + \tan^2\theta = \sec^2\theta$.
- For $\sqrt{x^2 - a^2}$, use $x = a\sec\theta$, $dx = a\sec\theta\tan\theta\, d\theta$, and $\sec^2\theta - 1 = \tan^2\theta$.

5.2. More examples

Example 5.4. To compute $\int \frac{\sqrt{4-x^2}}{x^2}\, dx$, put $x = 2\sin\theta$, $dx = 2\cos\theta\, d\theta$, and get

$$\begin{aligned}\int \frac{\sqrt{4 - x^2}}{x^2}\, dx &= \int \frac{\sqrt{4 - 4\sin^2\theta}}{4\sin^2\theta} 2\cos\theta\, d\theta = \int \frac{\cos^2\theta}{\sin^2\theta}\, d\theta = \int \cot^2\theta\, d\theta \\ &= \int (\csc^2\theta - 1)\, d\theta = -\cot\theta - \theta + C.\end{aligned}$$



To complete the solution we must write this in terms of x. It is useful to draw the triangle shown at right, where the edges are determined by the condition that $\sin\theta = x/2$ together with the Pythagorean theorem. We see that $\cot\theta = \frac{\sqrt{4-x^2}}{x}$, and obtain

$$\int \frac{\sqrt{4 - x^2}}{x^2}\, dx = -\frac{\sqrt{4 - x^2}}{x^2} - \sin^{-1}\frac{x}{2} + C.$$

Remark 5.5. We could also evaluate this integral using integration by parts, with $u = \sqrt{4 - x^2}$ and $dv = x^{-2}\, dx$, so $v = -1/x$, and we would obtain the same result. But the approach using trigonometric substitution is a little more routine in that we do not have to guess at a choice of u and v that work.

Example 5.6. To compute $\int \frac{1}{x\sqrt{x^2+1}}\,dx$ we put $x = \tan\theta$, $dx = \sec^2\theta\,d\theta$, and get

$$\int \frac{1}{x\sqrt{x^2+1}}\,dx = \int \frac{\sec^2\theta}{\tan\theta\sqrt{\tan^2\theta+1}}\,d\theta = \int \frac{\sec^2\theta}{\tan\theta|\sec\theta|}\,dx.$$

To eliminate the absolute value signs we choose $\theta \in (-\frac{\pi}{2}, \frac{\pi}{2})$ so that $\sec\theta > 0$, and we get

$$\int \frac{1}{x\sqrt{x^2+1}}\,dx = \int \frac{\sec\theta}{\tan\theta}\,d\theta = \int \frac{1/\cos\theta}{\sin\theta/\cos\theta}\,d\theta = \int \csc\theta\,d\theta.$$

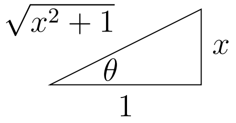
We have not evaluated this before, but we did compute $\int \sec\theta\,d\theta = \ln|\tan\theta + \sec\theta|$. Thus it is natural to expect that the integral of $\csc\theta$ is related to $\ln|\cot\theta + \csc\theta|$, and differentiating this expression gives

$$\frac{d}{d\theta}\ln|\cot\theta + \csc\theta| = \frac{-\csc^2\theta - \csc\theta\cot\theta}{\cot\theta + \csc\theta} = -\csc\theta.$$

We conclude that

$$\int \frac{1}{x\sqrt{x^2+1}}\,dx = \int \csc\theta\,d\theta = -\ln|\cot\theta + \csc\theta|.$$

To write this in terms of x, we use the triangle shown to get $\cot\theta = \frac{1}{x}$ and $\csc\theta = \frac{\sqrt{x^2+1}}{x}$, so that



$$\int \frac{1}{x\sqrt{x^2+1}}\,dx = -\ln\left|\frac{1}{x} + \frac{\sqrt{x^2+1}}{x}\right| = \ln\left|\frac{x}{1+\sqrt{x^2+1}}\right| = \ln|x| - \ln(1+\sqrt{x^2+1}) + C.$$

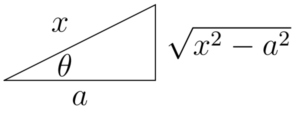
Example 5.7. For $\int \frac{x}{\sqrt{x^2+1}}\,dx$, we use $x = \tan\theta$ and $dx = \sec^2\theta\,d\theta$ with $\theta \in (-\frac{\pi}{2}, \frac{\pi}{2})$ to get

$$\int \frac{x}{\sqrt{x^2+1}}\,dx = \int \frac{\tan\theta\sec^2\theta}{\sqrt{\tan^2\theta+1}}\,d\theta = \int \tan\theta\sec\theta\,d\theta = \sec\theta + C = \sqrt{x^2+1} + C.$$

Observe that we could also have used the substitution $u = x^2 + 1$.

Example 5.8. To compute $\int \frac{1}{\sqrt{x^2-a^2}}\,dx$, where $a > 0$, we put $x = a\sec\theta$ and $dx = a\sec\theta\tan\theta\,d\theta$, using the range $\theta \in (0, \frac{\pi}{2}) \cup (\pi, \frac{3\pi}{2})$ so that $\tan\theta > 0$, and get

$$\begin{aligned}\int \frac{1}{\sqrt{x^2-a^2}}\,dx &= \int \frac{a\sec\theta\tan\theta}{\sqrt{a^2\sec^2\theta - a^2}}\,d\theta = \int \frac{\sec\theta\tan\theta}{\tan\theta}\,d\theta \\ &= \int \sec\theta\,d\theta = \ln|\sec\theta + \tan\theta| + C \\ &= \ln\left|\frac{x}{a} + \frac{\sqrt{x^2-a^2}}{a}\right| + C.\end{aligned}$$



To obtain a marginally simpler expression we can absorb $-\ln a$ into the constant of integration and write $\int \frac{1}{\sqrt{x^2-a^2}}\,dx = \ln|x + \sqrt{x^2-a^2}| + C$.

Example 5.9. The previous example could also be computed by using the *hyperbolic* substitution $x = a\cosh t$, $dx = a\sinh t\,dt$, since then we have

$$\int \frac{1}{\sqrt{x^2-a^2}}\,dx = \int \frac{a\sinh t}{\sqrt{a^2\cosh^2 t - a^2}}\,dt = \int \frac{\sinh t}{\sinh t}\,dt = t + C = \cosh^{-1}\Big(\frac{x}{a}\Big) + C.$$

Recalling the definition $\cosh t = \frac{e^t+e^{-t}}{2}$, one can use the quadratic formula to verify that the two solutions agree with each other.

Lecture 6 Complicated quadratics

Stewart §7.3, Spivak Ch. 19

6.1. Trigonometric substitutions for complicated quadratics

If an integral contains the square root of a quadratic polynomial in one of the three simple forms from Technique 5.3, then the corresponding trigonometric substitution is clear. For more complicated quadratic polynomials, a preliminary substitution can be used to bring the expression to one of the forms there.

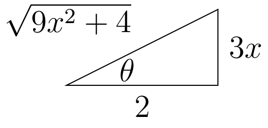
Example 6.1. If we are confronted with the integral $\int x^2(9x^2+4)^{-3/2}\,dx$, then we can first make the substitution $u = 3x$ to write $9x^2+4 = u^2+4$, and then make the trigonometric substitution $u = 2\tan\theta$. In terms of x, this is $3x = 2\tan\theta$, so $x = \frac{2}{3}\tan\theta$, $dx = \frac{2}{3}\sec^2\theta\,d\theta$, and we get

$$\begin{aligned}\int x^2(9x^2+4)^{-3/2}\,dx &= \int \frac{4}{9}\tan^2\theta(4\tan^2\theta+4)^{-3/2}\cdot\frac{2}{3}\sec^2\theta\,d\theta \\ &= \frac{8}{27}\int 4^{-3/2}\tan^2\theta(\sec^2\theta)^{-3/2}\sec^2\theta\,d\theta = \frac{1}{27}\int\frac{\tan^2\theta}{\sec\theta}\,d\theta,\end{aligned}$$

where we choose $\theta \in (-\frac{\pi}{2}, \frac{\pi}{2})$ to guarantee that $\sec\theta > 0$. Writing everything in terms of sine and cosine gives

$$\begin{aligned}\int x^2(9x^2+4)^{-3/2}\,dx &= \frac{1}{27}\int\frac{\sin^2\theta}{\cos\theta}\,d\theta = \frac{1}{27}\int\Big(\frac{1}{\cos\theta} - \cos\theta\Big)\,d\theta \\ &= \frac{1}{27}\big(\ln|\sec\theta+\tan\theta| - \sin\theta\big) + C.\end{aligned}$$

Using the triangle at right, we have $\sec\theta = \frac{\sqrt{9x^2+4}}{2}$, $\tan\theta = \frac{3x}{2}$, and $\sin\theta = \frac{3x}{\sqrt{9x^2+4}}$, so



$$\int x^2(9x^2+4)^{-3/2}\,dx = \frac{1}{27}\Big(\ln|\sqrt{9x^2+4}+3x| - \frac{3x}{\sqrt{9x^2+4}}\Big) + C,$$

where as before we absorb a factor of $\frac{1}{27}\ln 2$ into the constant of integration.

If the quadratic contains a linear term, then we need to complete the square first.

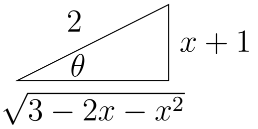
Example 6.2. To compute $I = \displaystyle\int \frac{x^2}{\sqrt{3-2x-x^2}}\,dx$, we complete the square as

$$3 - 2x - x^2 = -(x+1)^2 + 4,$$

and so the preliminary substitution to make is $u = x+1$. Then the quantity inside the square root is $4 - u^2$, so we make the substitution $u = 2\sin\theta$. Doing both substitutions successively gives

$$\begin{aligned}\int \frac{x^2}{\sqrt{3-2x-x^2}}\,dx &= \int \frac{(u-1)^2}{\sqrt{4-u^2}}\,du = \int \frac{(2\sin\theta - 1)^2}{\sqrt{4 - 4\sin^2\theta}} \cdot 2\cos\theta\,d\theta \\ &= \int (4\sin^2\theta - 4\sin\theta + 1)\,d\theta = \int (2(1-\cos 2\theta) - 4\sin\theta + 1)\,d\theta \\ &= 2\theta - \sin 2\theta + 4\cos\theta + \theta + C.\end{aligned}$$

To transition back to x we use the triangle to get $\sin\theta = \frac{x+1}{2}$, $\cos\theta = \frac{1}{2}\sqrt{3-2x-x^2}$, and thus



$$\begin{aligned} I &= 3\sin^{-1}\left(\frac{x+1}{2}\right) - 2\cdot\frac{x+1}{2}\cdot\frac{\sqrt{3-2x-x^2}}{2} + 4\cdot\frac{\sqrt{3-2x-x^2}}{2} + C \\ &= 3\sin^{-1}\left(\frac{x+1}{2}\right) + (3-x)\frac{\sqrt{3-2x-x^2}}{2} + C.\end{aligned}$$

Lecture 7 Rational functions

Stewart §7.4, Spivak Ch. 19

7.1. Polynomial long division

We know how to integrate polynomials using linearity and the power rule. But what about rational functions? The following example is instructive.

Example 7.1.

$$\int \frac{2x+3}{x+1}\,dx = \int \frac{2(x+1)+1}{x+1}\,dx = \int \left(2 + \frac{1}{x+1}\right)dx = 2x + \ln|x+1| + C.$$

Observe that while it was not clear how to integrate the original rational function directly, we were able to transform it into the sum of a polynomial (in this case the constant function 2) and a new rational function ($\frac{1}{x+1}$), where the new rational function has a simpler numerator and is easier to integrate.

We can carry this out more generally. Suppose we want to integrate a rational function $\frac{P(x)}{Q(x)}$, where P, Q are polynomials. The first step is to use *polynomial long division* to write

$$P(x) = S(x)Q(x) + R(x),$$

where S, R are polynomials and $\deg R < \deg Q$. Then we have

$$\int \frac{P(x)}{Q(x)}\,dx = \int S(x)\,dx + \int \frac{R(x)}{Q(x)}\,dx.$$

The first integral on the RHS can be computed directly. For the second, we can compute it directly if Q is linear (so that $R(x)$ is constant), while for more general Q we will need the method of *partial fractions* that we introduce in the next section.

Example 7.2. Suppose we want to compute $\int \frac{x^3-x}{x+2}\,dx$. Then $P(x) = x^3 - x$ and $Q(x) = x + 2$, so polynomial long division gives $x^3 - x = (x^2 - 2x + 3)(x + 2) - 6$, via the following computation.

$$\begin{array}{rr}
 & x^2 - 2x + 3 \\ \cline{2-2}
x+2\big) & x^3 \qquad\qquad - x \\
 & \underline{-x^3 - 2x^2 \qquad\qquad} \\
 & -2x^2 \quad - x \\
 & \underline{2x^2 + 4x} \\
 & 3x \\
 & \underline{-3x - 6} \\
 & -6
\end{array}$$

Recall how the algorithm goes: since $\deg P = 3$ and $\deg Q = 1$, we must have $\deg S = 3 - 1 = 2$, so we start by determining the quadratic coefficient of S. This must be 1 in order for the leading terms of $P(x)$ and $S(x)Q(x)$ to agree, so we write x^2 in the top line and then subtract $x^2Q(x)$ from $S(x)$, obtaining $-2x^2 - x$. This polynomial now plays the role of P, and we repeat the process until the remaining polynomial has degree smaller than $\deg Q = 1$.

Using the result of the long division, we get

$$\begin{aligned}
\int \frac{x^3 - x}{x + 2}\,dx = \int \frac{(x^2 - 2x + 3)(x + 2) - 6}{x + 2} &= \int \Big(x^2 - 2x + 3 - \frac{6}{x + 2}\Big)\,dx \\
&= \frac{1}{3}x^3 - x^2 + 3x - 6\ln|x + 2| + C.
\end{aligned}$$

7.2. Partial fraction decompositions

The above procedure is not always enough; it may still not be immediately clear how to integrate the resulting rational function. For example, when we derived the formula for $\int \sec\theta\,d\theta$, an important step was to compute $\int \frac{1}{1-x^2}\,dx$, which is not as easy as integrating $\frac{1}{ax+b}$. The trick was to notice that

$$\frac{1}{1-x} + \frac{1}{1+x} = \frac{(1+x) + (1-x)}{(1-x)(1+x)} = \frac{2}{1-x^2},$$

which let us write

$$\int \frac{1}{1-x^2}\,dx = \frac{1}{2}\int \frac{1}{1-x} + \frac{1}{1+x}\,dx = \frac{1}{2}(\ln|1+x| - \ln|1-x|) + C.$$

This is an example of a *partial fraction decomposition.* But how did we come up with it? And how can we use a similar trick to help us compute other integrals?

Example 7.3. To compute $\int \frac{x+5}{x^2-2x-3}\,dx$, we can factor the denominator as $x^2-2x-3 = (x-3)(x+1)$ and conjecture that

$$\frac{x+5}{x^2-2x-3} = \frac{A}{x-3} + \frac{B}{x+1} \text{ for some choice of } A, B \in \mathbb{R}. \tag{7.1}$$

Observe that the RHS can be rewritten as

$$\frac{A(x+1)+B(x-3)}{(x-3)(x+1)} = \frac{(A+B)x+(A-3B)}{x^2-2x-3},$$

and so we want to choose A, B such that

$$(A+B)x+(A-3B) = x+5 \text{ for every } x \in \mathbb{R}.$$

This happens if and only if $A+B=1$ and $A-3B=5$; this is a system of two equations in two variables, which we can easily solve to get $A=2$, $B=-1$, so that

$$\int \frac{x+5}{x^2-2x-3}\,dx = \int \frac{2}{x-3} - \frac{1}{x+1}\,dx = 2\ln|x-3| - \ln|x+1| + C.$$

Technique 7.4. A similar method works anytime the denominator can be factored into distinct linear polynomials: if $\deg P < \deg Q = n$ and $Q(x) = (x-r_1)\cdots(x-r_n)$, so that $r_1, \dots, r_n$ are the roots of Q, then our goal is to find $A_1, \dots, A_n \in \mathbb{R}$ such that

$$\frac{P(x)}{Q(x)} = \frac{A_1}{x-r_1} + \cdots + \frac{A_n}{x-r_n} \text{ for every } x \in \mathbb{R} \setminus \{r_1, \dots, r_n\}. \tag{7.2}$$

Putting the RHS over a common denominator equal to $Q(x)$, one sees that (7.2) is true if and only if $A_1, \dots, A_n$ satisfy a certain system of n linear equations in n variables, obtained by comparing the coefficients of the polynomial $P(x)$ (up to degree $n-1$) to the coefficients of the numerator on the RHS. Once the values of $A_1, \dots, A_n$ are found, the RHS can easily be integrated using $\int \frac{A}{x-r}\,dx = A\ln|x-r|$.

The procedure of rewriting the rational function $\frac{P(x)}{Q(x)}$ as the RHS of (7.2) is called a *partial fraction decomposition.*

Remark 7.5. We stress that (7.2) is *not* an equation to be solved for x; rather, it is a condition that is supposed to hold for *every* x, and this then determines the values of the numbers $A_1, \dots, A_n$ using comparison of coefficients.

Technique 7.6. An alternate technique for finding the coefficients in (7.2) is to put the RHS over a common denominator and then instead of comparing coefficients, evaluate both $P(x)$ and the numerator of the RHS at n specific points. It makes sense to choose points where the RHS takes a simple form, and one can often achieve this by evaluating it at the points $r_1, \dots, r_n$.

Example 7.7. In Example 7.3, we could rewrite (7.1) as

$$\frac{x+5}{x^2-2x-3} = \frac{A(x+1)+B(x-3)}{(x-3)(x+1)},$$

just as we did before, so that we want to find A, B such that

$$x+5 = A(x+1)+B(x-3) \text{ for all } x;$$

then instead of comparing coefficients, we could evaluate this equation at $x = 3$ and $x = -1$, where it gives

$$3 + 5 = A(3 + 1) + B \cdot 0 \text{ and } -1 + 5 = A \cdot 0 + B(-1 - 3),$$

which are easily solved to give $A = 2$ and $B = -1$, just as before.

It is important to observe that both of these methods *fail* as currently formulated if the factors of $Q(x)$ are not distinct.

Example 7.8. If $P(x) = x$ and $Q(x) = (x + 1)^2$, then (7.2) becomes

$$\frac{x}{(x+1)^2} = \frac{A_1}{(x+1)^2} + \frac{A_2}{(x+1)^2} = \frac{A_1 + A_2}{(x+1)^2};$$

this can only be satisfied if $A_1 + A_2 = x$ for every x, which is impossible. (The corresponding system of linear equations is $A_1 + A_2 = 0$, $0 = 1$.)

Now there are three questions that need to be addressed.

(1) Does the system of equations coming from (7.2) always have a solution if the linear factors are all distinct?
(2) What do we do if the factors are not distinct; how do we deal with repeated roots of $Q(x)$?
(3) What do we do if $Q(x)$ does not factor into linear polynomials? For example, what if $Q(x) = x^2 + 1$?

We will address the second and third questions in the following sections. For the first question, we start by thinking about the case $n = 2$. Suppose that $P(x) = ax + b$ and $Q(x) = (x - r_1)(x - r_2)$. Then (7.2) becomes

$$\frac{ax + b}{(x - r_1)(x - r_2)} = \frac{A_1}{x - r_1} + \frac{A_2}{x - r_2} = \frac{A_1(x - r_2) + A_2(x - r_1)}{(x - r_1)(x - r_2)}$$

and thus A_1, A_2 must satisfy

$$ax + b = A_1(x - r_2) + A_2(x - r_1) \text{ for every } x.$$

We could expand the RHS and compare coefficients, but it is easier to evaluate the above equation at $x = r_1$ and $x = r_2$, when it gives

$$ar_1 + b = A_1(r_1 - r_2) \quad \text{and} \quad ar_2 + b = A_2(r_2 - r_1).$$

If $r_1 \neq r_2$, then we can immediately solve for A_1 and A_2, and see that they are uniquely determined by a, b, r_1, r_2. On the other hand, we observe that $ar_i + b \neq 0$ for both $i = 1$ and $i = 2$ (since otherwise we could have simplified the expression $\frac{ax+b}{(x-r_1)(x-r_2)}$ by dividing top and bottom by $x - r_i$), and thus if $r_1 = r_2$ then there can be no solution A_1, A_2, since the right-hand sides vanish.

Proposition 7.9. *Suppose that $Q(x)$ factors as $Q(x) = (x - r_1)(x - r_2) \cdots (x - r_n)$, and $P(r_i) \neq 0$ for all i.*[3] *Then there are real numbers $A_1, \dots, A_n$ satisfying* (7.2) *if and only if the roots $r_1, \dots, r_n$ are all distinct. Morever, in this case there is exactly one solution: the values of $A_1, \dots, A_n$ are uniquely determined by P, Q.*

[3] Again, the assumption on P is reasonable because otherwise $P(x)$ would have $(x - r_i)$ as a factor, and we could cancel this term from both P and Q.

Proof. Collecting the terms in the RHS of (7.2) over a common denominator, we get

$$P(x) = A_1(x-r_2)(x-r_3)\cdots(x-r_n) + A_2(x-r_1)(x-r_3)\cdots(x-r_n) + \cdots + A_n(x-r_1)\cdots(x-r_{n-1}),$$

where in each term we multiply A_j by the product of the factors $x - r_i$ taken over all $i \neq j$. In particular, the only term on the RHS that does *not* include a factor of $(x - r_1)$ is the first, and thus evaluating the above equation at $x = r_1$ gives

$$P(r_1) = A_1(r_1 - r_2)(r_1 - r_3)\cdots(r_1 - r_n).$$

Similarly, evaluating at $x = r_2$ gives

$$P(r_2) = A_2(r_2 - r_1)(r_2 - r_3)\cdots(r_2 - r_n).$$

Continuing in this way we get n equations, one for each A_i. If all the roots r_i are distinct, then each A_i is multiplied by a nonzero number to get $P(r_i)$, and we can solve for A_i to get the unique solution. On the other hand, if $r_i = r_j$ for some $i \neq j$, then we get $P(r_j) = A_j \cdot 0 = 0$, contradicting the assumption that $P(r_j) \neq 0$. Thus when the roots are not distinct, there is no solution. □

Lecture 8 General partial fraction decompositions

Stewart §7.4, Spivak Ch. 19

8.1. Repeated factors

Now we address the second question, about what to do when $Q(x)$ has a repeated root, so that $\frac{P(x)}{Q(x)}$ does not admit a partial fraction decomposition of the form (7.2), where all the numerators are constants and all the denominators are linear.

Start by considering Example 7.8, and observe that we can still perform the following computation:

$$\int \frac{x}{(x+1)^2}\,dx = \int \frac{(x+1)-1}{(x+1)^2}\,dx = \int \frac{1}{x+1} - \frac{1}{(x+1)^2}\,dx = \ln|x+1| + \frac{1}{x+1} + C.$$

This suggests a general way of dealing with repeated factors.

Technique 8.1. To integrate $\frac{P(x)}{(x-r)^n}$, where P is a polynomial of degree $< n$, find real numbers $A_1, \dots, A_n$ such that

$$P(x) = A_1(x-r)^{n-1} + A_2(x-r)^{n-2} + \cdots + A_{n-1}(x-r) + A_n.$$

Then we obtain the following result:

$$\begin{aligned}\int \frac{P(x)}{(x-r)^n} &= \int \frac{A_1}{x-r} + \frac{A_2}{(x-r)^2} + \cdots + \frac{A_n}{(x-r)^n}\,dx \\ &= A_1 \ln|x-r| - \frac{A_2}{x-r} - \frac{1}{2}\cdot\frac{A_3}{(x-r)^2} - \cdots - \frac{1}{n-1}\cdot\frac{A_n}{(x-r)^{n-1}} + C.\end{aligned}$$

Example 8.2. To integrate $\frac{1}{x(x+1)^3}$, we combine the ideas from (7.2) and Technique 8.1: we want to find real numbers A, B, C, D such that

$$\frac{1}{x(x+1)^3} = \frac{A}{x} + \frac{B}{x+1} + \frac{C}{(x+1)^2} + \frac{D}{(x+1)^3}.$$

Putting everything over a common denominator, we see that our goal is

$$\frac{1}{x(x+1)^3} = \frac{A(x+1)^3 + Bx(x+1)^2 + Cx(x+1) + Dx}{x(x+1)^3}.$$

This holds if and only if the numerators agree for all x, that is, if

$$1 = A(x+1)^3 + Bx(x+1)^2 + Cx(x+1) + Dx. \tag{8.1}$$

To find A, B, C, D, we can use either Technique 7.4 and compare coefficients, or Technique 7.6 and evaluate both sides at appropriate values of x.

Comparing coefficients: Expanding the RHS of (8.1) and comparing coefficients between the two sides, we obtain a system of four linear equations in four variables:

$$\begin{aligned} 1 &= A(x^3 + 3x^2 + 3x + 1) + Bx(x^2 + 2x + 1) + C(x^2 + x) + Dx \\ &= (A + B)x^3 + (3A + 2B + C)x^2 + (3A + B + C + D)x + A, \end{aligned}$$

which yields

$$\begin{aligned} 0 &= A + B && \text{cubic coefficients} \\ 0 &= 3A + 2B + C && \text{quadratic coefficients} \\ 0 &= 3A + B + C + D && \text{linear coefficients} \\ 1 &= A && \text{constant coefficients.} \end{aligned}$$

The fourth equation gives $A = 1$, then the first gives $B = -A = -1$, then the second gives $C = -3A - 2B = -3 + 2 = -1$, then the third gives $D = -3A - B - C = -3 + 1 + 1 = -1$.

Evaluating at specific values: When $x = 0$, the RHS of (8.1) is equal to A, so we conclude that $A = 1$. Similarly, when $x = -1$, the RHS is equal to $-D$, and we conclude that $D = -1$. Because 0 and -1 are the only roots of $Q(x)$, it is not clear what two other values of x to use in order to find B and C. Choosing two values essentially at random would lead to a system of two equations in two variables, which could then be solved. Alternately, we can take one of the following two approaches.

Simplify and divide: Using $A = 1$ and $D = -1$, (8.1) can be rewritten as

$$1 = (x+1)^3 + Bx(x+1)^2 + Cx(x+1) - x; \tag{8.2}$$

adding $x - (x+1)^3$ to both sides gives

$$\begin{aligned} Bx(x+1)^2 + Cx(x+1) &= 1 + x - (x+1)^3 = (x+1)(1 - (x+1)^2) \\ &= (x+1)(1 - x^2 - 2x - 1) = -x(x+1)(x+2). \end{aligned}$$

Divide both sides by $x(x+1)$ to get

$$B(x+1) + C = -x - 2.$$

Again, recall that this is an equation that we want to be true for all x. Evaluating both sides at $x = -1$ gives $C = -(-1) - 2 = 1 - 2 = -1$, and thus we are left with

$$B(x+1) = -x - 2 - (-1) = -x - 1 = -(x+1),$$

so $B = -1$.

Differentiate: Another way to find B and C is to observe that the LHS and RHS of (8.2) are two different ways of writing the same function, so their derivatives must also be equal:

$$0 = 3(x+1)^2 + B(x+1)^2 + 2Bx(x+1) + C(x+1) + Cx - 1.$$

Evaluating this at $x = -1$ gives $C = -1$, and so the equation becomes

$$0 = 3(x+1)^2 + B(x+1)^2 + 2Bx(x+1) - 2(x+1).$$

Differentiating again gives

$$0 = 6(x+1) + 2B(x+1) + 2B(x+1) + 2Bx - 2,$$

and putting $x = -1$ gives $B = -1$.

Whichever of the above approaches we use, we conclude that $A = 1$ and $B = C = D = -1$, so

$$\begin{aligned}\int \frac{1}{x(x+1)^3}\,dx &= \int \frac{1}{x} - \frac{1}{x+1} - \frac{1}{(x+1)^2} - \frac{1}{(x+1)^3}\,dx \\ &= \ln|x| - \ln|x+1| + \frac{1}{x+1} + \frac{1}{2(x+1)^2} + C,\end{aligned}$$

where this last C is a constant of integration (not the coefficient from the earlier computations).

At some level it is a matter of taste which approach we use to determine the coefficients in any given problem. However, it may be the case that one of the methods is easier than the others, and thus it is useful to be familiar with all of them. And indeed, there are other ways to proceed as well.

Example 8.3. Revisiting $\int \frac{1}{x(x+1)^3}\,dx$, suppose we start by only going partway in the partial fraction decomposition (here A, B, C, D are not the same as in the previous computations):

$$\frac{1}{x(x+1)^3} = \frac{A}{x} + \frac{Bx^2 + Cx + D}{(x+1)^3}$$

Collecting the RHS over a common denominator we get

$$1 = A(x+1)^3 + Bx^3 + Cx^2 + Dx = (A+B)x^3 + (3A+C)x^2 + (3A+D)x + A,$$

so $A = 1$, $B = -1$, $C = -3$, and $D = -3$. Then we make the substitution $u = x + 1$ and get

$$\begin{aligned}\int \frac{1}{x(x+1)^3}\,dx = \int \frac{1}{x} - \frac{x^2+3x+3}{(x+1)^3}\,dx &= \ln|x| - \int \frac{(u-1)^2 + 3(u-1) + 3}{u^3}\,du \\ = \ln|x| - \int \frac{u^2+u+1}{u^3}\,du &= \ln|x| - \int (u^{-1} + u^{-2} + u^{-3})\,du\end{aligned}$$

$$
\begin{aligned}
&= \ln|x| - \ln|u| - u^{-1} - \frac{1}{2}u^{-2} + C \\
&= \ln|x| - \ln|x+1| - \frac{1}{x+1} - \frac{1}{2(x+1)^2} + C.
\end{aligned}
$$

8.2. Quadratic factors

Now we have enough tools to integrate $\frac{P(x)}{Q(x)}$ whenever $Q(x)$ factors into linear terms. But what if it does not? The prototypical example of a polynomial that does not factor into linear terms is $Q(x) = x^2 + 1$, and we recall from our work on differentiating trigonometric functions that

$$\int \frac{1}{1+x^2}\,dx = \tan^{-1} x + C. \tag{8.3}$$

More generally, given $a > 0$ we can use the trigonometric substitution $x = a\tan\theta$, $dx = a\sec^2\theta\,d\theta$ to obtain

$$\int \frac{1}{x^2+a^2}\,dx = \int \frac{a\sec^2\theta}{a^2\tan^2\theta + a^2}\,d\theta = \int \frac{1}{a}\,d\theta = \frac{\theta}{a} + C = \frac{1}{a}\tan^{-1}\frac{x}{a} + C. \tag{8.4}$$

Technique 8.4. Any integral of the form $\int \frac{Ax+B}{x^2+bx+c}\,dx$, where x^2+bx+c is irreducible, can be evaluated by completing the square as $x^2+bx+c = (x-\frac{b}{2})^2 + (c - \frac{b^2}{4})$, making the substitution $u = x - \frac{b}{2}$ so that the denominator becomes $u^2 + a^2$ for $a = \sqrt{c - \frac{b^2}{4}}$, and then using (8.4).

For integrals of the form $\int \frac{Ax+B}{(x^2+bx+c)^k}\,dx$, we can do the same substitution to obtain an integrand with denominator $(u^2+a^2)^k$, and then use the substitution $u = a\tan\theta$ as above to obtain an integrand in terms of tangents and secants.

This turns out to be the final piece of the puzzle.

Technique 8.5 (Integrating rational functions by partial fractions)**.** Given *any* rational function $\frac{P(x)}{Q(x)}$, we can use the method of partial fractions to integrate it by going through the following steps.

(1) Use polynomial long division to reduce to the case when $\deg P < \deg Q$.
(2) Factor $Q(x)$ as a product of linear and quadratic terms, where the quadratic terms have no real roots.[4]
(3) Use any of the techniques described earlier to find coefficients that let us write $\frac{P(x)}{Q(x)}$ as a sum of expressions of one of the following forms:

$$\frac{A}{x-r}, \quad \frac{A}{(x-r)^k}, \quad \frac{Ax+B}{x^2+bx+c}, \quad \frac{Ax+B}{(x^2+bx+c)^k}.$$

(4) Integrate each of these individually:
  (a) $\displaystyle\int \frac{A}{x-r}\,dx = A\ln|x-r|;$

[4] The fact that this is always possible is called the *Fundamental Theorem of Algebra*, and its proof is beyond the scope of this course.

(b) $\displaystyle\int \frac{A}{(x-r)^k} = \frac{A}{k-1}(x-r)^{-(k-1)}$ when $k \geq 2$;

(c) For the last two types, use Technique 8.4: complete the square, make a u-substitution, and then either use (8.4) or make a further trigonometric substitution.

Example 8.6. To compute $\int \frac{1}{1+x^3}\,dx$, we factor the denominator as

$$1 + x^3 = (1+x)(1-x+x^2),$$

but we cannot go any further because $1-x+x^2$ has no real roots (this follows from the quadratic formula since $(-1)^2 - 4(1)(1) = -3 < 0$). As in Example 8.3, though, we can write

$$\frac{1}{1+x^3} = \frac{A}{1+x} + \frac{Bx+C}{1-x+x^2},$$

and taking a common denominator we see that A, B, C must satisfy

$$1 = A(1-x+x^2) + (Bx+C)(1+x) \text{ for all } x.$$

As before, there are several ways to solve this. Putting $x = -1$ immediately gives $1 = A(1-(-1)+1) = 3A$, so $A = \frac{1}{3}$. Although $Q(x) = 1+x^3$ has no other real roots, we can observe that the expressions obtained for $x = 0$ and $x = 1$ are not so complicated:

$$x = 0 \;\Rightarrow\; 1 = \frac{1}{3} + C \quad\text{and}\quad x = 1 \;\Rightarrow\; 1 = \frac{1}{3} + 2(B+C).$$

Thus $C = \frac{2}{3}$ and $B + C = \frac{1}{3}$, so $B = -\frac{1}{3}$, and we have

$$\int \frac{1}{1+x^3}\,dx = \frac{1}{3}\int \frac{1}{1+x} + \frac{-x+2}{1-x+x^2}\,dx.$$

The first part integrates as $\frac{1}{3}\int \frac{1}{1+x}\,dx = \frac{1}{3}\ln|x+1|$, so it remains to integrate the last part. Completing the square gives $x^2 - x + 1 = (x-\frac{1}{2})^2 + \frac{3}{4} = u^2 + a^2$ for $u = x - \frac{1}{2}$ and $a = \sqrt{3}/2$, so we get

$$\begin{aligned}\int \frac{-x+2}{1-x+x^2}\,dx &= \int \frac{-(u+\frac{1}{2})+2}{u^2+a^2}\,du = -\frac{1}{2}\int \frac{2u}{u^2+a^2}\,du + \frac{3}{2}\int \frac{1}{u^2+a^2}\,du \\ &= -\frac{1}{2}\ln(u^2+a^2) + \frac{3}{2}\frac{1}{a}\tan^{-1}\frac{u}{a} \\ &= -\frac{1}{2}\ln(x^2-x+1) + \sqrt{3}\tan^{-1}\frac{x-\frac{1}{2}}{\sqrt{3}/2} + C.\end{aligned}$$

Putting it all together gives

$$\int \frac{1}{1+x^3}\,dx = \frac{1}{3}\ln|x+1| - \frac{1}{6}\ln(x^2-x+1) + \frac{\sqrt{3}}{3}\tan^{-1}\frac{2x-1}{\sqrt{3}} + C.$$

We give one more example to illustrate the procedure in the presence of a repeated quadratic factor.

Example 8.7. To compute $\int \frac{1-x}{x(x^2+1)^2}\,dx$, we first find $A, B, C, D, E \in \mathbb{R}$ such that

$$\frac{1-x}{x(x^2+1)^2} = \frac{A}{x} + \frac{Bx+C}{x^2+1} + \frac{Dx+E}{(x^2+1)^2}$$

for all x, which upon putting things over a common denominator is equivalent to

$$1 - x = A(x^2+1)^2 + (Bx+C)(x^2+1)x + (Dx+E)x.$$

Evaluating at $x = 0$ gives $A = 1$, so we must find B, C, D, E satisfying

$$(Bx+C)(x^2+1)x + (Dx+E)x = 1 - x - (1 + 2x^4 + x^4) = -x - 2x^2 - x^4.$$

Dividing both sides by x, we want

$$\begin{aligned} -1 - 2x - x^3 &= (Bx+C)(x^2+1) + (Dx+E) \\ &= Bx^3 + Cx^2 + (B+D)x + (C+E), \end{aligned}$$

and comparing coefficients gives

$$B = -1, \quad C = 0, \quad B + D = -2, \quad C + E = -1 \quad \Rightarrow \quad D = -1, \quad E = -1.$$

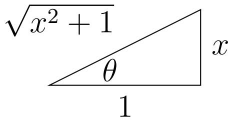
We conclude that

$$\begin{aligned} \int \frac{1-x}{x(x^2+1)^2}\,dx &= \int \frac{1}{x} - \frac{x}{x^2+1} - \frac{x+1}{(x^2+1)^2}\,dx \\ &= \ln|x| - \frac{1}{2}\ln|x^2+1| - \frac{1}{2}(x^2+1)^{-1} - \int \frac{1}{(x^2+1)^2}\,dx. \end{aligned}$$

To evaluate the final integral we use the substitution $x = \tan\theta$, $dx = \sec^2\theta\,d\theta$ and obtain

$$\begin{aligned} \int \frac{1}{(x^2+1)^2}\,dx = \int \frac{\sec^2\theta}{(\tan^2\theta+1)^2}\,d\theta = \int \frac{\sec^2\theta}{\sec^4\theta}\,d\theta = \int \cos^2\theta\,d\theta \\ = \frac{1}{2}\int (1 + \cos 2\theta)\,d\theta = \frac{1}{2}\theta + \frac{1}{4}\sin 2\theta + C. \end{aligned}$$

The triangle at right gives $\sin\theta = \frac{x}{\sqrt{x^2+1}}$ and $\cos\theta = \frac{1}{\sqrt{x^2+1}}$, so $\sin 2\theta = 2\sin\theta\cos\theta = \frac{2x}{x^2+1}$, and we get



$$\begin{aligned} \int \frac{1-x}{x(x^2+1)^2}\,dx &= \ln|x| - \frac{1}{2}\ln(x^2+1) - \frac{1}{2}(x^2+1)^{-1} - \frac{1}{2}\tan^{-1}x + \frac{x}{2(x^2+1)} + C \\ &= \ln|x| - \frac{1}{2}\ln(x^2+1) + \frac{x-1}{2(x^2+1)} - \frac{1}{2}\tan^{-1}x + C. \end{aligned}$$

At this point we now have the ability to integrate *any* rational function, although the computations involved may be quite intimidating. For example, the partial fraction decomposition

$$\frac{x^3+x^2+1}{x(x-1)(x^2+x+1)(x^2+1)^2} = \frac{A}{x} + \frac{B}{x-1} + \frac{Cx+D}{x^2+x+1} + \frac{Ex+F}{x^2+1} + \frac{Gx+H}{(x^2+1)^2}$$

leads to a system of 8 linear equations in 8 variables, which would be tedious to solve by hand (though a computer would do it very quickly), and then we would be left with 8 separate integrals to compute, some immediate, others requiring appropriate substitutions.

Lecture 9 Numerical integration

Stewart §7.7, Spivak Ch. 19

9.1. Endpoint and midpoint rules

When we are tasked with evaluating a definite integral $\int_a^b f(x)\,dx$, our usual approach is to find an antiderivative $F(x) = \int f(x)\,dx$ and then apply the FTC to get $\int_a^b f(x)\,dx = F(b) - F(a)$. However, as the discussion above illustrates, we may not be able to find a formula for an antiderivative, even if we know the formula for f. And it may be the case that we do not even know the formula for f, for example if the function is known only experimentally. In such cases we turn to a different approach to computing definite integrals, which goes back to their original definition via Riemann sums.

Recall that given $n \in \mathbb{N}$, we can partition the interval $[a, b]$ into n subintervals $[x_{i-1}, x_i]$ for $i = 1, \dots, n$, where $x_i = a + i\Delta x$, and $\Delta x = \frac{b-a}{n}$ is the width of each subinterval. Then upon selecting a point x_i^* inside each subinterval $[x_{i-1}, x_i]$, the corresponding Riemann sum is

$$\sum_{i=1}^n f(x_i^*)\Delta x.$$

There are three natural choices to make for x_i^*: we might choose the left endpoint, the right endpoint, or the midpoint of $[x_{i-1}, x_i]$. Using left endpoints gives the *left endpoint approximation*

$$L_n = \sum_{i=1}^n f(x_{i-1})\Delta x,$$

and similarly, the *right endpoint approximation* is

$$R_n = \sum_{i=1}^n f(x_i)\Delta x.$$

Choosing $x_i^* = \bar{x}_i := \frac{1}{2}(x_{i-1} + x_i)$ gives the *midpoint approximation*

$$M_n = \sum_{i=1}^n f(\bar{x}_i)\Delta x.$$

It follows from the general theory of integration that all three approximations converge to $\int_a^b f(x)\,dx$ as $n \to \infty$; however, we are also interested in the *speed* of approximation. Indeed, if you have an application that requires a numerical answer precise to within 10^{-4}, then it is necessary to know how large n must be in order to guarantee this degree of precision. We state the following theorem without proof.

Theorem 9.1. *If $f\colon [a, b] \to \mathbb{R}$ is twice differentiable and $K \in \mathbb{R}$ has the property that $|f''(x)| \le K$ for all $x \in [a, b]$, then the error term in the midpoint approximation can be bounded as follows:*

$$\left| M_n - \int_a^b f(x)\,dx \right| \le \frac{K(b-a)^3}{24n^2}.$$

For simplicity's sake, suppose that a, b, K have the property that $K(b-a)^3/24 = 1$. Then the error bound is n^{-2}, and so to guarantee precision of 10^{-4} using the midpoint rule, we would need to take $n = 10^2 = 100$. It turns out that the corresponding error estimate for the left and right endpoint approximations has a factor of n, not n^2, in the denominator, and to get $n^{-1} = 10^{-4}$ requires $n = 10^4$; this illustrates that to get an estimate with a very small error bound, it is useful to use the more efficient midpoint approximation.

9.2. Trapezoid rule

There are two more methods that are worth mentioning here; both involve replacing the rectangles used in Riemann sums with more general shapes.

For the *trapezoid rule*, instead of using a rectangle with height $f(x_i^*)$ for some $x_i^* \in [x_{i-1}, x_i]$, we use a trapezoid with two vertices on the x-axis, at $(x_{i-1}, 0)$ and $(x_i, 0)$, and the other two vertices on the graph of the function, at $(x_{i-1}, f(x_{i-1}))$ and $(x_i, f(x_i))$. The area of this trapezoid is

$$\text{average height} \times \text{base} = \frac{f(x_{i-1}) + f(x_i)}{2} \cdot \Delta x,$$

and adding up the areas of the n trapezoids over the intervals $[x_{i-1}, x_i]$ for $i = 1, \dots, n$, we get the following approximation for $\int_a^b f(x)\,dx$:

$$T_n := \sum_{i=1}^n \frac{f(x_{i-1}) + f(x_i)}{2} \cdot \Delta x, \qquad \Delta x := \frac{b-a}{n}, \quad x_i := a + i\Delta x. \tag{9.1}$$

We can rewrite this as

$$T_n = \frac{\Delta x}{2}\big(f(x_0) + 2f(x_1) + 2f(x_2) + \dots + 2f(x_{n-1}) + f(x_n)\big).$$

Theorem 9.1 has an analogue for the trapezoid rule, except that the 24 in the denominator is replaced by 12; the trapezoid rule is actually not quite as good as the midpoint rule in general.

9.3. Simpson's rule

Another way to interpret the trapezoid rule is that on each interval $[x_{i-1}, x_i]$, we replaced the function f with a linear function $g \approx f$ that agrees with f at the endpoints, and then integrated g instead of f. (Of course, we use a different function g on each small interval $[x_{i-1}, x_i]$.) To get a better approximation, we might try using a quadratic function instead. Recall that a quadratic function is determined by its values at three points, so now we should ask for g to agree with f at the endpoints *and* at the midpoint. Notationally, it will be easier to assume that n is even and then approximate by quadratics on $[x_0, x_2]$, $[x_2, x_4]$, and so on. The key computation is contained in the following lemma.

Lemma 9.2. *Suppose we are given $h > 0$, three points $x_0 < x_1 < x_2$ related by $x_1 = x_0 + h$ and $x_2 = x_1 + h$, and three values $y_0, y_1, y_2 \in \mathbb{R}$. Let $g(x)$ be the unique quadratic polynomial $g(x)$ such that $g(x_i) = y_i$ for $i = 0, 1, 2$. Then*

$$\int_{x_0}^{x_2} g(x)\,dx = \frac{h}{3}(y_0 + 4y_1 + y_2). \tag{9.2}$$

Proof. Without loss of generality we can assume that $x_0 = -h$, $x_1 = 0$, and $x_2 = h$, since translating the graph horizontally does not change the area underneath it. Since g is a quadratic polynomial, we must have $A, B, C \in \mathbb{R}$ such that $g(x) = Ax^2 + Bx + C$. Evaluating this at $x = -h, 0, h$ and using the fact that $g(x_i) = y_i$ for $i = 0, 1, 2$, we get

$$\begin{aligned} y_0 &= Ah^2 - Bh + C, \\ y_1 &= C, \\ y_2 &= Ah^2 + Bh + C. \end{aligned}$$

Adding the first and the third equations gives $y_0 + y_2 = 2Ah^2 + 2C$, and the second equation gives $C = y_1$, so we can evaluate the integral as

$$\begin{aligned} \int_{-h}^{h} (Ax^2 + Bx + C)\,dx &= \Big[\frac{A}{3}x^3 + \frac{B}{2}x^2 + Cx\Big]_{-h}^{h} = \frac{2Ah^3}{3} + 2Ch = \frac{h}{3}(2Ah^2 + 6C) \\ &= \frac{h}{3}(y_0 + y_2 + 4C) = \frac{h}{3}(y_0 + y_2 + 4y_1), \end{aligned}$$

which proves the lemma. □

Now return to the question of approximating $\int_a^b f(x)\,dx$. Given $n \in \mathbb{N}$ even and $x_i = a + i\Delta x$, where $\Delta x = (b-a)/n$, let $y_i = f(x_i)$. When we add up the areas under the parabolas over $[x_0, x_2]$, $[x_2, x_4]$, and so on, we obtain the following approximation for $\int_a^b f(x)\,dx$, known as *Simpson's rule*:

$$\begin{aligned} S_n &= \frac{\Delta x}{3}(y_0 + 4y_1 + y_2) + \frac{\Delta x}{3}(y_2 + 4y_3 + y_4) + \cdots + \frac{\Delta x}{3}(y_{n-2} + 4y_{n-1} + y_n) \\ &= \frac{\Delta x}{3}(y_0 + 4y_1 + 2y_2 + 4y_3 + 2y_4 + \cdots + 2y_{n-2} + 4y_{n-1} + y_n). \end{aligned} \tag{9.3}$$

For Simpson's rule we have the following improvement on Theorem 9.1, which again we do not prove here.

Theorem 9.3. *If f is four times differentiable on $[a,b]$ and $K \in \mathbb{R}$ is such that $|f^{(4)}(x)| \leq K$ for all $x \in [a,b]$, then the error term in Simpson's rule can be bounded as follows:*

$$\Big| S_n - \int_a^b f(x)\,dx \Big| \leq \frac{K(b-a)^5}{180n^4}.$$

The factor of n^4 in the denominator means that we can obtain a very precise approximation with a relatively small value of n, which makes this approximation very useful and explains why we may consider it an improvement over the approximations described earlier.

Lecture 10 Improper integrals

Stewart §7.8, Spivak exercises 14.25–30

10.1. Infinite width

We know that if $f\colon [a,b] \to (0,\infty)$ is a positive function, then $\int_a^b f(x)\,dx$ represents the area underneath the graph of $y = f(x)$ over the bounded interval $[a,b]$. But what if we consider the area under the graph over an *unbounded* interval? Can we still make sense of this notion?

Start with a concrete example: consider the region beneath the graph of $y = \frac{1}{x^2}$, above the x-axis, and to the right of the line $x = 1$. If we truncate this region by cutting off everything to the right of the line $x = t$ for some fixed $t > 1$, then the truncated region has area

$$A(t) = \int_1^t \frac{1}{x^2}\,dx = -\frac{1}{x}\Big|_1^t = 1 - \frac{1}{t}.$$

Viewing the truncated region as an approximation to the region we are interested in, we see that the approximation gets better the larger t gets, and that $\lim_{t\to\infty} A(t) = 1$, so it seems reasonable to say that the region originally described has area 1, and to write $\int_1^\infty \frac{1}{x^2}\,dx = 1$. This serves as a template for a more general definition.

Definition 10.1. Let $f\colon [a,\infty) \to \mathbb{R}$ be such that

(1) f is integrable on $[a,t]$ for every $t \geq a$, and
(2) $\lim_{t\to\infty} \int_a^t f(x)\,dx$ exists and is finite.

Then we write $\int_a^\infty f(x)\,dx := \lim_{t\to\infty} \int_a^t f(x)\,dx$, and call this an *improper integral (of type 1)*; in this case we say that the improper integral is *convergent*. If the limit does not exist, we say that it is *divergent*.

The improper integral $\int_{-\infty}^b f(x)\,dx$ is defined similarly when $f\colon (-\infty,b] \to \mathbb{R}$ is integrable on every $[t,b]$, provided the limit $\lim_{t\to-\infty} \int_t^b f(x)\,dx$ exists and is finite.

Finally, if both $\int_{-\infty}^a f(x)\,dx$ and $\int_a^\infty f(x)\,dx$ are convergent, then we write $\int_{-\infty}^\infty f(x)\,dx = \int_{-\infty}^a f(x)\,dx + \int_a^\infty f(x)\,dx$.

Exercise 10.2. Prove that in the last part of Definition 10.1, it does not matter what value of a we choose: if the two improper integrals are convergent for some value of a, then they are convergent for any other value of a, and their sum has the same value.

In the case when $f \geq 0$, we can interpret an improper integral as an area, just as with more familiar definite integrals.

Example 10.3. $\displaystyle\lim_{t\to\infty} \int_1^t \frac{1}{x}\,dx = \lim_{t\to\infty} \big[\ln x\big]_1^t = \lim_{t\to\infty} \ln t = \infty$, so the improper integral $\int_1^\infty \frac{1}{x}\,dx$ is divergent.

Example 10.4.

$$\begin{aligned}\int_{-\infty}^0 xe^x\,dx &= \lim_{t\to-\infty} \int_t^0 \underbrace{x}_{u}\,\underbrace{e^x\,dx}_{dv} = \lim_{t\to-\infty} xe^x\Big|_t^0 - \int_t^0 e^x\,dx = \lim_{t\to-\infty} \Big[xe^x - e^x\Big]_t^0 \\ &= \lim_{t\to-\infty} 0e^0 - e^0 - te^t + e^t = -1,\end{aligned}$$

so the improper integral is convergent.

Example 10.5. To evaluate $\int_{-\infty}^{\infty} \frac{1}{1+x^2}\,dx$, we first compute

$$\int_0^\infty \frac{1}{1+x^2}\,dx = \lim_{t\to\infty} \int_0^t \frac{1}{1+x^2}\,dx = \lim_{t\to\infty} \Big[\tan^{-1} x\Big]_0^t = \lim_{t\to\infty} \tan^{-1} t = \frac{\pi}{2},$$

and a similar computation gives $\int_{-\infty}^0 \frac{1}{1+x^2}\,dx = \frac{\pi}{2}$, so

$$\int_{-\infty}^\infty \frac{1}{1+x^2}\,dx = \int_{-\infty}^0 \frac{1}{1+x^2}\,dx + \int_0^\infty \frac{1}{1+x^2}\,dx = \frac{\pi}{2} + \frac{\pi}{2} = \pi.$$

Example 10.6. Suppose we fix a positive real number $p > 0$ and consider the improper integral $\int_1^\infty \frac{1}{x^p}\,dx$. For which values of p is this integral convergent? Note that Example 10.3 showed that it is divergent when $p = 1$. For $p \neq 1$, we have

$$\lim_{t\to\infty} \int_1^t \frac{1}{x^p}\,dx = \lim_{t\to\infty} \left[\frac{x^{1-p}}{1-p}\right]_1^t = \lim_{t\to\infty} \frac{t^{1-p}-1}{1-p} = \begin{cases} \infty & \text{if } p < 1, \\ \frac{1}{p-1} & \text{if } p > 1. \end{cases}$$

Thus $\int_1^\infty \frac{1}{x^p}\,dx$ is convergent if $p > 1$, and divergent if $p \leq 1$.

10.2. Infinite height

Another type of improper integral arises from vertical asymptotes. Recall that our original definition of the definite integral $\int_a^b f(x)\,dx$ required the function f to be bounded on $[a, b]$ (as well as some other requirements). If f has a vertical asymptote at one of the endpoints, then we can define $\int_a^b f(x)\,dx$ as an improper integral by using a limiting procedure similar to the one in the previous section.

Definition 10.7. Let $f\colon [a, b) \to \mathbb{R}$ be continuous and suppose that

(1) $\lim_{x\to b^-} f(x)$ does not exist, and
(2) $\lim_{t\to b^-} \int_a^t f(x)\,dx$ exists and is finite.

Then we write $\int_a^b f(x)\,dx := \lim_{t\to b^-} \int_a^t f(x)\,dx$ for the corresponding *improper integral (of type 2)*, which we call *convergent*. If the limit does not exist, we say that the improper integral is *divergent*.

Similarly, if f is continuous everywhere on $[a, b]$ except for the left endpoint a, then we write $\int_a^b f(x)\,dx = \lim_{t\to a^-} \int_t^b f(x)\,dx$ provided the limit exists.

Finally, if f is continuous everywhere on $[a, b]$ except for some point $c \in (a, b)$, then we write $\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx$ *provided both improper integrals converge.*

Exercise 10.8. Prove that if f is continuous on all of $[a, b]$, then all of the definitions above agree with the usual definition of $\int_a^b f(x)\,dx$.

Example 10.9. $f(x) = \frac{1}{\sqrt{x-2}}$ has a vertical asymptote at $x = 2$, so

$$\int_2^5 \frac{1}{\sqrt{x-2}}\,dx = \lim_{t\to 2^+} \int_t^5 \frac{dx}{\sqrt{x-2}} = \lim_{t\to 2^+} \Big[2\sqrt{x-2}\Big]_t^5 = \lim_{t\to 2^+} 2\sqrt{3} - 2\sqrt{t-2} = 2\sqrt{3},$$

and the improper integral converges.

Example 10.10. $\sec x$ has a vertical asymptote at $x = \pi/2$, so

$$\int_0^{\frac{\pi}{2}} \sec x\,dx = \lim_{t\to\frac{\pi}{2}^-}\int_0^t \sec x\,dx = \lim_{t\to\frac{\pi}{2}^-}\Big[\ln|\sec x+\tan x|\Big]_0^t = \lim_{t\to\frac{\pi}{2}^-}\ln|\sec t+\tan t| = \infty,$$

and the improper integral is divergent.

Example 10.11. To evaluate $\int_0^3 \frac{dx}{x-1}$, observe that $\frac{1}{x-1}$ has a vertical asymptote at $x = 1$, so we need to independently evaluate $\int_0^1 \frac{dx}{x-1}$ and $\int_1^3 \frac{dx}{x-1}$. The first of these is

$$\int_0^1 \frac{dx}{x-1} = \lim_{t\to 1^-}\int_0^t \frac{dx}{x-1} = \lim_{t\to 1^-}\Big[\ln|x-1|\Big]_0^t = \lim_{t\to 1^-}\ln|t-1| = -\infty,$$

and thus $\int_0^3 \frac{dx}{x-1}$ is divergent.

Remark 10.12. The previous example shows the need for caution when applying the FTC. It would be all too easy to unthinkingly push symbols around and write $\int_0^3 \frac{dx}{x-1} = [\ln|x-1|]_0^3 = \ln 2$, but this is wrong. Observe that the FTC does not apply here, because it requires the function to be integrable (and in particular, bounded) on the entire interval.

Example 10.13.

$$\int_0^1 \ln x\,dx = \lim_{t\to 0^+}[x\ln x - x]_t^1 = \lim_{t\to 0^+}(-1 - t\ln t + t) = -1,$$

where we use the fact that $\lim_{t\to 0^+} t\ln t = 0$. Thus the improper integral is convergent.

10.3. Comparison theorems

Sometimes determining whether or not an improper integral is convergent is significantly easier than establishing its numerical value.

Theorem 10.14. *Suppose $f, g\colon [a,\infty) \to \mathbb{R}$ are continuous and satisfy $f(x) \geq g(x) \geq 0$ for all $x \geq a$. Then the following are true.*

(1) If $\int_a^\infty f(x)\,dx$ is convergent, then $\int_a^\infty g(x)\,dx$ is convergent.
(2) If $\int_a^\infty g(x)\,dx$ is divergent, then $\int_a^\infty f(x)\,dx$ is divergent.

Proof. We start by proving the first claim. Suppose that $\int_a^\infty f(x)\,dx$ is convergent, and let $G(t) := \int_a^t g(x)\,dx$. For every $t > a$, we have

$$G(t) = \int_a^t g(x)\,dx \leq \int_a^t f(x)\,dx \leq \int_a^\infty f(x)\,dx,$$

where the first inequality uses $g \leq f$ and properties of integrals, and the second inequality uses the fact that $f \geq 0$. Thus the function G is bounded above. Moreover, for every $t_1 \leq t_2 > a$ we have

$$G(t_2) = \int_a^{t_2} g(x)\,dx = \int_a^{t_1} g(x)\,dx + \int_{t_1}^{t_2} g(x)\,dx = G(t_1) + \int_{t_1}^{t_2} g(x)\,dx \geq G(t_1),$$

where the last inequality uses the fact that $g \geq 0$. Thus G is a nondecreasing function. By the monotone convergence theorem, $\lim_{t\to\infty} G(t)$ exists (and is equal to $\sup\{G(t) : t > a\}$). This means that $\int_a^\infty g(x)\,dx$ is convergent.

The second claim in the theorem is equivalent to the first one, so this proves the theorem. □

Example 10.15. To determine convergence of $\int_0^\infty e^{-x^2}\,dx$, we observe that $x^2 \geq x$ for all $x \geq 1$, and thus $e^{-x^2} \leq e^{-x}$ for all $x \geq 1$. Since $\int_1^\infty e^{-x}\,dx$ is convergent, Theorem 10.14 implies that $\int_1^\infty e^{-x^2}\,dx$ is convergent as well. This in turn implies that $\int_0^\infty e^{-x^2}\,dx$ is convergent, because

$$\begin{aligned}\int_0^\infty e^{-x^2}\,dx = \lim_{t\to\infty}\int_0^t e^{-x^2}\,dx &= \lim_{t\to\infty}\int_0^1 e^{-x^2}\,dx + \int_1^t e^{-x^2}\,dx \\ &= \int_0^1 e^{-x^2}\,dx + \lim_{t\to\infty}\int_1^t e^{-x^2}\,dx = \int_0^1 e^{-x^2}\,dx + \int_1^\infty e^{-x^2}\,dx.\end{aligned}$$

Remark 10.16. In fact, using more sophisticated techniques it is possible to show that $\int_0^\infty e^{-x^2}\,dx = \sqrt{\pi}/2$, but this requires tools that we have not yet developed.

10.4. *Cauchy Principal Value integral

Let us return to the world of Example 10.11 for a moment.[5] We declared the integral $\int_0^3 \frac{1}{x-1}\,dx$ divergent because $\int_0^1 \frac{1}{x-1}\,dx = -\infty$. But someone looking at the picture might argue that the negative part of the graph here should exactly cancel with the positive part of the graph from $x = 1$ to $x = 2$, leaving us with a finite integral on the interval $[0, 3]$.

Consider the similar example $\int_{-1}^1 \frac{1}{x}\,dx$; the graph is symmetric around the origin, and so one may argue that the integral should be 0, despite the fact that $\int_{-1}^0 \frac{1}{x}\,dx$ and $\int_0^1 \frac{1}{x}\,dx$ both diverge. One way of making this precise is to use something called the *Cauchy Principal Value integral*, which says that if $f\colon [a,b] \to \mathbb{R}$ is continuous everywhere except for one point $c \in (a,b)$, then we put

$$\mathrm{PV}\!\int_a^b f(x)\,dx := \lim_{t\to 0^+}\Big(\int_a^{c-t} f(x)\,dx + \int_{c+t}^b f(x)\,dx\Big).$$

The notation is meant to remind us that this stands for something different than the usual integral.

Exercise 10.17. Show that if $\int_a^b f(x)\,dx$ is convergent in the sense of Definition 10.7, then $\mathrm{PV}\!\int_a^b f(x)\,dx = \int_a^b f(x)\,dx$.

Observe that with this definition, we have

$$\mathrm{PV}\!\int_{-1}^1 \frac{1}{x}\,dx = \lim_{t\to 0^+}\Big(\int_{-1}^{-t}\frac{1}{x}\,dx + \int_t^1 \frac{1}{x}\,dx\Big) = \lim_{t\to 0^+} 0 = 0,$$

consistent with our earlier intuition. However, there is a problem, as the following two examples illustrate.

[5] This section will not appear on any tests, and largely follows an explanation I read on the website of Dave Rusin (University of Texas).

Example 10.18.

$$\begin{aligned}
\mathrm{PV}\!\int_{-1}^{1} \frac{2x+4}{(x^2+4x)^3}\,dx &= \lim_{t\to 0^+} \Big(\int_{-1}^{-t} \frac{2x+4}{(x^2+4x)^3}\,dx + \int_{t}^{1} \frac{2x+4}{(x^2+4x)^3}\,dx \Big) \\
&= \lim_{t\to 0^+} \Big(\Big[-(x^2+4x)^{-2} \Big]_{-1}^{-t} + \Big[-(x^2+4x)^{-2} \Big]_{t}^{1} \Big) \\
&= \lim_{t\to 0^+} \Big(\frac{1}{9} - \frac{1}{(t^2-4t)^2} + \frac{1}{(t^2+4t)^2} - \frac{1}{25} \Big) \\
&= \frac{1}{9} - \frac{1}{25} + \lim_{t\to 0^+} \frac{1}{t^2} \Big(\frac{1}{(t+4)^2} - \frac{1}{(t-4)^2} \Big) \\
&= \frac{16}{225} + \lim_{t\to 0^+} \frac{(t-4)^2-(t+4)^2}{t^2(t^2-16)^2} = \frac{16}{225} + \lim_{t\to 0^+} \frac{-16}{t(t^2-16)^2} = -\infty.
\end{aligned}$$

Example 10.19. Evaluating the same integral using the substitution $u = x^2 + 4x$ gives

$$\mathrm{PV}\!\int_{-1}^{1} \frac{2x+4}{(x^2+4x)^3}\,dx = \mathrm{PV}\!\int_{-3}^{5} \frac{1}{u^3}\,du = \int_{3}^{5} \frac{1}{u^3}\,du = \frac{1}{9} - \frac{1}{25} = \frac{16}{225}.$$

So we see that if we allow ourselves to evaluate integrals around a broader class of vertical asymptotes using the Cauchy principal value integral, then we need to come to terms with the fact that the substitution rule no longer works! One may reasonably conclude (as we do in this course) that this is too high a price to pay, and thus we will refrain from assigning finite values to any integrals that are divergent in the sense of Definition 10.7.

Remark 10.20. A similar phenomenon occurs for improper integrals of the form $\int_{-\infty}^{\infty} f(x)\,dx$. If the integral is convergent, then we have

$$\begin{aligned}
\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{0} f(x)\,dx + \int_{0}^{\infty} f(x)\,dx &= \lim_{t\to\infty} \int_{-t}^{0} f(x)\,dx + \lim_{t\to\infty} \int_{0}^{t} f(x)\,dx \\
= \lim_{t\to\infty} \left(\int_{-t}^{0} f(x)\,dx + \int_{0}^{t} f(x)\,dx \right) &= \lim_{t\to\infty} \int_{-t}^{t} f(x)\,dx.
\end{aligned}$$

In light of this, one might be tempted to define $\int_{-\infty}^{\infty} f(x)\,dx$ as $\lim_{t\to\infty} \int_{-t}^{t} f(x)\,dx$, provided the latter limit exists. However, **this turns out to be a bad idea**. Imagine that we adopt this definition. Then since $f(x) = x$ has $\int_{-t}^{t} x\,dx = \frac{1}{2}x^2|_{-t}^{t} = 0$ for all t, the limit exists and is equal to 0, so our new (bad!) definition would give $\int_{-\infty}^{\infty} x\,dx = 0$. But if we recall how integrals are supposed to behave, then we expect the following two properties to be true:

1. shifting the graph to the left or right does not change the integral;
2. shifting the graph up or down does change the integral.

Observe that shifting the graph of $f(x) = x$ one unit to the left has the same effect as shifting it one unit up. So according to the first rule, this shift should not change the value of the integral, but according to the second rule, it should change it! This contradiction can only be avoided by declaring that $\int_{-\infty}^{\infty} x\,dx$ is undefined (divergent), and indeed, using the true definition we observe that this improper integral is divergent because $\int_{0}^{\infty} x\,dx$ is divergent.

Review of integration strategies

Stewart §7.5, Spivak Ch. 19

This review is not included in a numbered lecture, but will be/was done during the hour preceding the first class test.

Now that we have learned several different tools for integration, it is worth reviewing them and describing an overall strategy.

Step 1: Check list of basic examples

The following list of integrals should be committed to memory, so that once we see one of these integrals appear, we know how to complete the solution. (To avoid cluttering up the display, we omit the constants of integration.)

$$\int x^n \, dx = \frac{x^{n+1}}{n+1} \qquad \int \frac{1}{x} \, dx = \ln|x|$$

$$\int e^x \, dx = e^x \qquad \int b^x \, dx = \frac{b^x}{\ln b}, \quad b > 0$$

$$\int \sin^x \; dx = -\cos x \qquad \int \cos x \, dx = \sin x$$

$$\int \sec^2 x \, dx = \tan x \qquad \int \csc^2 x \, dx = -\cot x$$

$$\int \sec x \tan x \, dx = \sec x \qquad \int \csc x \cot x \, dx = -\csc x$$

$$\int \sec x \, dx = \ln|\sec x + \tan x| \qquad \int \csc x \, dx = -\ln|\csc x + \cot x|$$

$$\int \tan x \, dx = \ln|\sec x| \qquad \int \cot x \, dx = \ln|\sin x|$$

$$\int \sinh x \, dx = \cosh x \qquad \int \cosh x \, dx = \sinh x$$

$$\int \frac{dx}{x^2 + a^2} = \frac{1}{a} \tan^{-1} \frac{x}{a} \qquad \int \frac{dx}{\sqrt{a^2 - x^2}} = \sin^{-1} \frac{x}{a}, \quad a > 0$$

$$\int \frac{dx}{x^2 - a^2} = \frac{1}{2a} \ln \left| \frac{x-a}{x+a} \right| \qquad \int \frac{dx}{\sqrt{x^2 \pm a^2}} = \ln|x + \sqrt{x^2 \pm a^2}|.$$

Remark 10.21. Some of these have alternate forms; for example Stewart's book lists the integral of $\csc x$ as $\ln|\csc x - \cot x|$. A short computation using properties of logarithms and the identity $\csc^2 x - \cot^2 x = 1$ shows that this agrees with the form here.

Step 2: Simplify if possible

If the integrand can be simplified using standard algebraic manipulations or trigonometric identities, this is the next thing to do.

Example 10.22.

$$\int (\sin x + \cos x)^2 \, dx = \int (\sin^2 x + 2\sin x \cos x + \cos^2 x) \, dx = \int (1 + \sin 2x) \, dx.$$

Step 3: Make an obvious substitution, if there is one

If there is a clear choice for u such that du naturally appears in the integrand, then it is worth trying this substitution.

Example 10.23. $\displaystyle\int \frac{x}{x^2-1} \, dx$ has the derivative of the denominator in the numerator (up to a constant), so $u = x^2 - 1$ is natural and transforms the integral into $\displaystyle\frac{1}{2}\int \frac{1}{u} \, du$.

Step 4: Classify the integral as a type that we know how to deal with

There are four general classes of integrals that we have developed a procedure for dealing with by now.

(1) Trigonometric integrals such as $\int \sin^4 x \cos^3 x \, dx$, which can be handled using various substitutions and identities.
(2) Rational functions, which can be handled using partial fractions.
(3) Integrals of the form $\int f(x)g(x) \, dx$, where $g(x)$ is something we can integrate and $f(x)$ gets simpler after differentiating; the most important case is when f is a polynomial, but this also includes things like $f(x) = \ln x$ or $f(x) = \tan^{-1} x$. For integrals like this, integration by parts with $u = f(x)$ and $dv = g(x) \, dx$ is likely to be useful.
(4) Integrals involving quadratic polynomials inside square roots, for which an appropriate trigonometric substitution is often helpful.

Step 5: Get creative

Sometimes with a little more creativity we can find an algebraic manipulation or a substitution that helps, even if one was not obvious upon initial inspection.

Example 10.24.

$$\begin{aligned}\int \frac{1}{1-\sin x} \, dx = \int \frac{1+\sin x}{1-\sin^2 x} \, dx = \int \frac{1+\sin x}{\cos^2 x} \, dx \\ = \int (\sec^2 x + \sec x \tan x) \, dx = \tan x + \sec x + C.\end{aligned}$$

Example 10.25. The substitution $u = \sqrt{x}$ has $x = u^2$, $dx = 2u \, du$, so

$$\int e^{\sqrt{x}} \, dx = \int 2ue^u \, du,$$

which can be integrated by parts.

Similarly, some functions that do not immediately look like rational functions can be transformed into rational functions via an appropriate substitution, and then integrated using partial fractions.

Example 10.26. Using the substitution $u = \sqrt{x+4}$, $x = u^2 - 4$, $dx = 2u\,du$, we have

$$\begin{aligned}\int \frac{\sqrt{x+4}}{x}\,dx &= \int \frac{u}{u^2-4} 2u\,du = 2\int \frac{u^2}{u^2-4}\,du = 2\int 1 + \frac{4}{u^2-4}\,du \\ &= 2u + 2\int \frac{1}{u-2} - \frac{1}{u+2}\,du = 2u + 2\ln|u-2| - 2\ln|u+2| + C \\ &= 2\sqrt{x+4} - \ln\frac{x+8-4\sqrt{x+4}}{x+8+4\sqrt{x+4}} + C.\end{aligned}$$

where we use the computation

$$(u \pm 2)^2 = (\sqrt{x+4} \pm 2)^2 = x + 4 \pm 4\sqrt{x+4} + 4 = x + 8 \pm 4\sqrt{x+4},$$

and have omitted the computations to determine the partial fraction decomposition. Just to continue the fun, we point out that the last term on the first line can also be integrated with the substitution $u = 2\sec\theta$, so

$$\int \frac{4}{u^2-4}\,du = \int \frac{4}{4\tan^2\theta} 2\sec\theta\tan\theta\,d\theta = 2\int \frac{\sec\theta}{\tan\theta}\,d\theta = 2\int \csc\theta\,d\theta,$$

which is a known integral. Turning θ back into u, and then into x, gives the same result as above.

Remark 10.27. It turns out that we can also integrate any rational function of trigonometric functions, such as $f(x) = \frac{\cos^2 x - \sin^3 x}{\tan x + \sec x}$, by using the *Weierstrass substitution* $t = \tan\frac{x}{2}$, which (after some work) yields $\cos x = \frac{1-t^2}{1+t^2}$, $\sin x = \frac{2t}{1+t^2}$, and $dx = \frac{2}{1+t^2}\,dt$, so that $\int f(x)\,dx = \int g(t)\,dt$ for some rational function g.

By now it should be clear that the task of finding formulas for antiderivatives – *symbolic integration* – is rather harder than the task of finding formulas for derivatives – *symbolic differentiation.* The latter task is fairly routine thanks to various results like the product rule, the chain rule, etc., which let us write down a formula for the derivative of any function that is written in terms of polynomials, radicals, exponentials, logarithms, and trigonometric functions. As we have seen, symbolic integration is an entirely different matter, and it turns out that there are some integrals that *cannot* be evaluated in terms of the 'elementary' functions we are used to dealing with; this include relatively innocuous-looking expressions such as $\int e^{-x^2}\,dx$ and $\int \frac{1}{\ln x}\,dx$.[6]

[6] To make this claim of impossibility rigorous, one needs to formulate clearly the class of functions that we consider, and then provide a proof of impossibility; this is beyond the scope of this course.

Part II. Applications of integration

Lecture 11 Arc length and the catenary

Stewart §8.1, Spivak exercise 13.25

11.1. A formula for arc length

We know how to compute the length of a straight line segment: it is simply the distance between the two endpoints (x_1, y_1) and (x_2, y_2) given by Pythagoras' formula

$$\text{distance} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}.$$

If we consider a "piecewise linear" curve that is a sequence of straight line segments connecting the points $P_0, P_1, \dots, P_n$, then we can similarly compute the total length of the curve as $\sum_{i=1}^n \text{distance}(P_{i-1}, P_i)$.

Given a more general curve in the plane, it is reasonable to approximate it by a piecewise linear curve, compute the length of the approximation, and then take a limit as the endpoints of the approximating line segments get closer and closer together. To make this more precise, suppose we consider the graph of $y = f(x)$ between $x = a$ and $x = b$. Then we might fix a large integer $n \in \mathbb{N}$; choose points in the interval $[a, b]$ by $x_i = a + i\Delta x$ for $0 \le i \le n$, where $\Delta x = \frac{b-a}{n}$; denote the point $(x_i, f(x_i))$ by P_i; and then declare the length of the curve to be

$$\text{length} = \lim_{n\to\infty} \sum_{i=1}^n \text{distance}(P_{i-1}, P_i). \tag{11.1}$$

This looks suspiciously similar to a limit of Riemann sums. Using Pythagoras' formula we get

$$\text{distance}(P_{i-1}, P_i) = \sqrt{(x_i - x_{i-1})^2 + (f(x_i) - f(x_{i-1}))^2}. \tag{11.2}$$

If f is continuous on $[a, b]$ and differentiable on (a, b), then the Mean Value Theorem says that for each $1 \le i \le n$ there is $x_i^* \in [x_{i-1}, x_i]$ such that

$$f(x_i) - f(x_{i-1}) = f'(x_i^*)(x_i - x_{i-1}).$$

Using this in (11.2) and recalling that $x_i - x_{i-1} = \Delta x$, we get

$$\text{distance}(P_{i-1}, P_i) = \sqrt{(\Delta x)^2 + f'(x_i^*)^2(\Delta x)^2} = \sqrt{1 + f'(x_i^*)^2} \cdot \Delta x.$$

Taking a sum over i from 1 to n, and then a limit as $n \to \infty$, we see that

$$\lim_{n\to\infty} \sum_{i=1}^n \text{distance}(P_{i-1}, P_i) = \lim_{n\to\infty} \sum_{i=1}^n \sqrt{1 + f'(x_i^*)^2} \cdot \Delta x = \int_a^b \sqrt{1 + (f'(x))^2}\, dx$$

provided f' is continuous on (a,b). Thus we make the following definition: if f is continuously differentiable,[7] the *arc length* L of the curve $y=f(x)$ from $x=a$ to $x=b$ is

$$L = \text{length} = \int_a^b \sqrt{1+(f'(x))^2}\,dx. \tag{11.3}$$

It is also often useful to define the following *arc length function*: given a continuously differentiable function $f\colon [a,b]\to\mathbb{R}$ and a value $x\in(a,b)$, the arc length of the section of curve from $(a,f(a))$ to $(x,f(x))$ is given by

$$s(x) = \int_a^x \sqrt{1+(f'(t))^2}\,dt. \tag{11.4}$$

By the FTC, we have

$$s'(x) = \sqrt{1+(f'(x))^2}. \tag{11.5}$$

Writing $y=f(x)$ and using Leibniz notation $f'(x)=\frac{dy}{dx}$, we can write (11.3) as

$$L = \int_a^b \sqrt{1+\Big(\frac{dy}{dx}\Big)^2}\,dx,$$

and (11.5) can be rewritten as

$$\frac{ds}{dx} = \sqrt{1+\Big(\frac{dy}{dx}\Big)^2}. \tag{11.6}$$

In an abuse of notation (since s, dx, and dy have no independent meaning), this is sometimes written as

$$(ds)^2 = (dx)^2 + (dy)^2,$$

which is an infinitesimal version of Pythagoras' formula (11.2).

11.2. Examples of arc length

Example 11.1. To find the arc length L of the parabola $y=x^2$ from $(0,0)$ to $(1,1)$, we use (11.3) with $a=0$, $b=1$, $f'(x)=2x$ to write

$$\begin{aligned} L &= \int_0^1 \sqrt{1+(2x)^2}\,dx && \Big(u=2x,\ dx=\frac{1}{2}du\Big)\\ &= \frac{1}{2}\int_0^2 \sqrt{1+u^2}\,du && \Big(u=\tan\theta,\ du=\sec^2\theta\,d\theta\Big)\\ &= \frac{1}{2}\int_0^\alpha \sec^3\theta\,d\theta && (\tan\alpha=2,\ \sec^2\alpha=1+\tan^2\alpha=5). \end{aligned}$$

Thus

$$\begin{aligned} 2L &= \int_0^\alpha \underbrace{(\sec\theta)}_{u}\underbrace{(\sec^2\theta)\,d\theta}_{dv} = [\underbrace{\sec\theta}_{u}\underbrace{\tan\theta}_{v}]_0^\alpha - \int_0^\alpha \sec\theta\tan^2\theta\,d\theta\\ &= \sec\alpha\tan\alpha - \sec 0\tan 0 - \int_0^\alpha \sec\theta(\sec^2\theta-1)\,d\theta \end{aligned}$$

[7] This means that f is differentiable and that f' is continuous.

$$= 2\sqrt{5} - \int_0^\alpha \sec^3\theta\,d\theta + \int_0^\alpha \sec\theta\,d\theta = 2\sqrt{5} - 2L + \big[\ln|\sec\theta + \tan\theta|\big]_0^\alpha,$$

and solving for L gives

$$\begin{aligned} L &= \frac{1}{4}\Big(2\sqrt{5} + \ln|\sec\alpha + \tan\alpha| - \overbrace{\ln|\sec 0 + \tan 0|}^{=\ln|1+0|=0}\Big) \\ &= \frac{\sqrt{5}}{2} + \frac{\ln(\sqrt{5}+2)}{4}. \end{aligned}$$

We get a similar formula for arc length if x is written as a function of y; if $x = g(y)$ and the curve runs from $y = a$ to $y = b$, then the arc length is $\int_a^b \sqrt{1 + (g'(y))^2}\,dy$.

Example 11.2. Consider the curve $x^2 = y^3$ running from $(1,1)$ to $(8,4)$. If we write y as a function of x, then we get $y = x^{2/3}$ and the formula for arc length gives

$$L = \int_1^8 \sqrt{1 + \Big(\frac{2}{3}x^{-1/3}\Big)^2}\,dx = \int_1^8 \sqrt{1 + \frac{4}{9}x^{-2/3}}\,dx.$$

It is not at all clear how to evaluate this. On the other hand, if we write x as a function of y then we have $x = y^{3/2}$ (note that the curve is in the first quadrant so $x > 0$) and the arc length is

$$\begin{aligned} L &= \int_1^4 \sqrt{1 + \Big(\frac{3}{2}y^{1/2}\Big)^2}\,dy = \int_1^4 \sqrt{1 + \frac{9}{4}y}\,dy \qquad (u = 1 + \tfrac{9}{4}y,\ du = \tfrac{9}{4}\,dy) \\ &= \frac{4}{9}\int_{13/4}^{10} \sqrt{u}\,du = \frac{4}{9}\Big[\frac{2}{3}u^{3/2}\Big]_{13/4}^{10} = \frac{8}{27}\Big(10^{3/2} - \Big(\frac{13}{4}\Big)^{3/2}\Big) \\ &= \frac{1}{27}(80\sqrt{10} - 13\sqrt{13}). \end{aligned}$$

As is already apparent from the previous examples, the presence of the square root in the arc length formula often leads to a nasty integral.

Example 11.3. Consider the hyperbola $x^2 - y^2 = 1$. Let L denote the arc length from $(1,0)$ to $(2,\sqrt{3})$. Writing $y = \sqrt{x^2-1}$ gives $\frac{dy}{dx} = \frac{x}{\sqrt{x^2-1}}$, so

$$L = \int_1^2 \sqrt{1 + \frac{x^2}{x^2-1}}\,dx = \int_1^2 \sqrt{\frac{2x^2-1}{x^2-1}}\,dx.$$

The presence of a quadratic inside a square root suggests a trigonometric substitution; but there are two quadratics in play here! Writing $x = \sec\theta$ gives $\sqrt{x^2-1} = \tan\theta$ and $dx = \sec\theta\tan\theta$, so

$$\int \sqrt{\frac{2x^2-1}{x^2-1}}\,dx = \int \frac{\sqrt{2\sec^2\theta - 1}}{\tan\theta}\cdot\sec\theta\tan\theta\,d\theta,$$

and it is not at all clear where to go from here, since none of the usual trigonometric identities help us simplify $\sqrt{2\sec^2\theta - 1}$. In fact it turns out that this integral cannot be evaluated in elementary terms using the functions that we have introduced so far, and so we cannot compute L exactly. Given this, our best bet would be to turn to numerical integration and use something like Simpson's rule to obtain an approximate value for the integral.

In fact, there is one more point to be made here. Note that the integrand $\sqrt{\frac{2x^2-1}{x^2-1}}$ has a vertical asymptote at $x = 1$, which was one of the limits of integration. Thus this is actually an *improper* integral! Since our discussion of numerical integration did not include any tools for dealing with improper integrals (and we would need to start by using a comparison theorem to check that this improper integral actually converges), we are better off writing $x = \sqrt{y^2 + 1}$ and working with the integral

$$L = \int_0^{\sqrt{3}} \sqrt{1 + \left(\frac{dx}{dy}\right)^2}\, dy = \int_0^{\sqrt{3}} \sqrt{1 + \frac{y^2}{y^2+1}}\, dy.$$

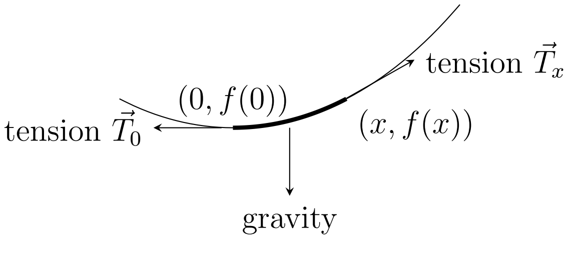
We still cannot evaluate this explicitly, but at least the integrand here is a bounded continuous function, and so we do not need to resort to improper integrals.

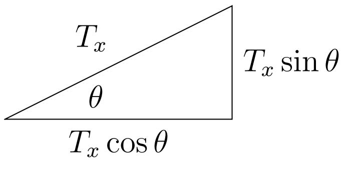
The appearance of the improper integral in the first expression comes because the curve has a vertical tangent line at $(1, 0)$, so the slope $\frac{dy}{dx}$ that appears in the arc length formula is infinite at this point. This is a phenomenon you should watch out for when computing arc lengths.

11.3. *The catenary

Consider a cable that is suspended at its endpoints and hangs freely in between them, such as a power line or telephone wire between two poles. What shape will it make?

Assume that the cable is relatively thin, so that it can be well-approximated by a curve $y = f(x)$, and that it is flexible, so that the tension at any point in the cable is in a direction tangent to the curve. Choose some point on the curve as the origin for x, and consider the part of the cable that lies between 0 and x. As shown in the picture, there are three forces acting on this segment of cable: tension pulling it to the left at the point $(0, f(0))$ (labeled $\vec{T_0}$), tension pulling it to the right at the point $(x, f(x))$ (labeled $\vec{T_x}$), and gravity pulling it downward. Let T_0 and T_x denote the magnitude of the tension forces; then since the tension at x points in the direction of the tangent line, which has slope $\tan\theta = f'(x)$, we see that the horizontal and vertical components of $\vec{T_0}$ are as shown in the picture at right.





Because the cable is not moving, the forces must all balance out, so $T_0 = T_x \cos\theta$, and $T_x \sin\theta$ equals the magnitude of the force due to gravity. This force is $m(x)g$, where g is the gravitational constant and $m(x)$ is the mass of the segment of cable between 0 and x. Assume that the cable has uniform density ρ (mass per unit length), so that $m(x) = \rho s(x)$, where $s(x)$ is the length of the section of cable from 0 to x. Then we have

$$T_x \cos\theta = T_0 \quad \text{and} \quad T_x \sin\theta = m(x)g = \rho g s(x).$$

Dividing these two equations gives

$$f'(x) = \tan\theta = \frac{T_x \sin\theta}{T_x \cos\theta} = \frac{\rho g}{T_0} s(x). \tag{11.7}$$

Our goal is to find a formula for the function f that allows us to write y as a function of x via $y = f(x)$. We will do this by first writing both x and y as functions of arc length s and as functions of a new variable t; this procedure of obtaining *parametrizations* for the curve is one that we will later return to and study in greater detail.

Using (11.6) and writing $a = \frac{T_0}{\rho g} > 0$ for a parameter that depends on the physical characteristics of the situation, we see that the function $x \mapsto s(x)$ satisfies

$$\frac{ds}{dx} = \sqrt{1 + \left(\frac{dy}{dx}\right)^2} = \sqrt{1 + \left(\frac{s}{a}\right)^2} = \frac{\sqrt{a^2 + s^2}}{a}, \tag{11.8}$$

where the second equality uses the fact that $\frac{dy}{dx} = f'(x) = \frac{s(x)}{a}$. Since the derivative of the inverse function $s \mapsto x(s)$ is the reciprocal of the derivative $s'(x)$, we conclude that

$$\frac{dx}{ds} = \frac{a}{\sqrt{a^2 + s^2}} \quad \Rightarrow \quad x(s) = \int \frac{a}{\sqrt{a^2 + s^2}}\, ds.$$

To evaluate this integral, we could use the trigonometric substitution $s = a\tan\theta$ and the identity $1 + \tan^2\theta = \sec^2\theta$, but it turns out to be simpler to use the substitution $s = a\sinh t$ and the identity $1 + \sinh^2 t = \cosh^2 t$:

$$x(s) = \int \frac{a}{\sqrt{a^2 + a^2\sinh^2 t}} \cdot a\cosh t\, dt = \int a\, dt = at + C.$$

Note that when $s = 0$ we have $x = 0$ and $t = \sinh^{-1}\frac{s}{a} = 0$, so the constant of integration is $C = 0$, and we have $t = x/a$.

To determine $y(s)$ we first use the chain rule to write

$$\frac{dy}{ds} = \frac{dy}{dx}\frac{dx}{ds} = \frac{s}{a} \cdot \frac{a}{\sqrt{a^2 + s^2}} \quad \Rightarrow \quad y(s) = \int \frac{s}{\sqrt{a^2 + s^2}}\, ds = \sqrt{a^2 + s^2} + b,$$

where b is a constant of integration. Since $\sqrt{a^2 + s^2} = \sqrt{a^2 + a^2\sinh^2 t} = a\cosh t$, we conclude that

$$y = f(x) = \sqrt{a^2 + s^2} + b = a\cosh(t) + b = a\cosh\left(\frac{x}{a}\right) + b.$$

Thus the curve formed by a hanging cable – called a *catenary* – is described by the hyperbolic cosine function. Note that here a, b are parameters determined by the physical characteristics of the situation, including the strength of gravity, the density of the cable, and the location of the two points at which it is suspended.

Lecture 12 Surface area

Stewart §8.2, Spivak appendix to Ch. 19

12.1. Surfaces of revolution

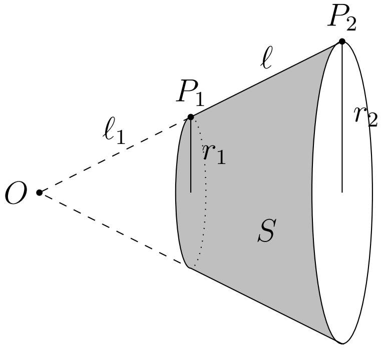
Suppose we are given a function $f\colon [a,b] \to \mathbb{R}$ whose graph describes a curve $\{(x, f(x)) : a \leq x \leq b\}$. The *surface of revolution* associated to this curve is the surface $S \subset \mathbb{R}^3$ that is obtained by rotating the curve around the x-axis. Observe that if (x, y, z) is a point on S, then the distance from (x, y, z) to $(x, 0, 0)$ must be equal to $f(x)$, and thus a precise description of S can be given by

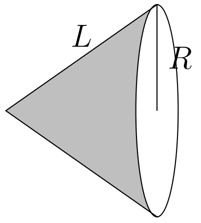
$$S = \{(x, y, z) : \sqrt{y^2 + z^2} = f(x)\},$$

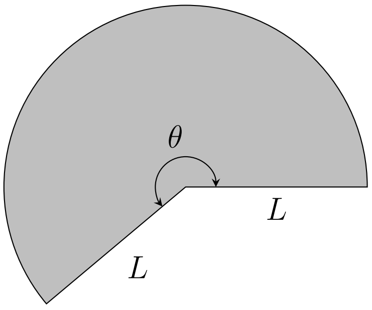
though we will not use this in what follows. Our goal in this section is to find a formula for the surface area of S.

As usual, the simplest case occurs when f is linear; $f(x) = mx + c$. If $m = 0$ so that f is constant, then the corresponding surface of revolution is a cylinder with radius c and depth $b - a$; this cylinder can be unrolled into a rectangle with the same surface area, whose dimensions are $(b - a) \times 2\pi c$, so the surface area of S is given by $2\pi c(b - a)$. We will find it convenient to write this as $2\pi r\ell$, where $r = c$ is the radius, and $\ell = b - a$ is the distance between $(a, f(a))$ and $(b, f(b))$ (since $f(a) = f(b)$).

If $m \neq 0$, then S is a truncated cone. To compute the surface area of S, write $r_1 = f(a)$ and $r_2 = f(b)$ for the radii of the circles that form the ends of the truncated cone. For concreteness, assume that $m > 0$ so that $r_1 < r_2$ (the case $m < 0$ is similar). Let $P_1 = (a, f(a))$ and $P_2 = (b, f(b))$ be the two endpoints of the line segment, and let ℓ be the distance between them. Let O be the point where the line $y = mx + c$ intersects the x-axis, and let ℓ_1 be the distance from O to P_1, as shown in the picture at left.







To find the surface area of S, we need to find the formula for the surface area of a cone. Consider the cone shown in the second picture, where the base has radius R and the diagonal side has length L. If we cut this cone along a line from the base to the tip and then unroll it, we obtain a shape such as the one shown in the third picture, which has the same area as the cone. The arc that forms the outer boundary has length θL by the formula for length of a circular arc; it also has length $2\pi R$ since this boundary is obtained by unrolling the circle at the cone's base. Thus we have $\theta L = 2\pi R$, and moreover, the area of this region is given by

$$\text{area} = \frac{\theta}{2\pi} \cdot \pi L^2 = \frac{1}{2}\theta L^2 = \frac{1}{2} \cdot \frac{2\pi R}{L} \cdot L^2 = \pi R L.$$

Returning to the surface area of S, observe that S is obtained by taking a cone with $L = \ell + \ell_1$ and $R = r_2$, and then removing from it a cone with $L = \ell_1$ and $R = r_1$. Thus the surface area of S is

$$\text{area}(S) = \pi r_2(\ell + \ell_1) - \pi r_1 \ell_1 = \pi\big(\ell_1(r_2 - r_1) + \ell r_2\big).$$

Observe that $\frac{\ell_1}{r_1} = \frac{\ell_1+\ell}{r_2}$, so $\ell_1 r_2 = r_1\ell_1 + r_1\ell$, and thus $\ell_1(r_2 - r_1) = \ell r_1$, which gives

$$\text{area}(S) = \pi(\ell r_1 + \ell r_2) = \pi(r_1 + r_2)\ell.$$

Note that this agrees with the formula from above for the surface area of a cylinder when f is constant, since in this case we have $r_1 = r_2 = r$. In order to write everything directly in terms of the function f, we observe that

$$r_1 + r_2 = f(a) + f(b) = 2f(x^*), \quad \text{where } x^* = \frac{a+b}{2},$$

since $f(x) = mx + c$ is linear, and moreover

$$\ell = \sqrt{(b-a)^2 + (f(b) - f(a))^2} = \sqrt{(b-a)^2 + (m(b-a))^2} = (b-a)\sqrt{1+m^2}.$$

Thus we have proved the following.

Proposition 12.1. *If $f\colon [a,b] \to [0,\infty)$ is linear (in other words, its graph is a line segment), then the surface area of the corresponding surface of revolution is*

$$\text{area} = 2\pi f(x^*)\sqrt{1 + (f'(x^*))^2} \cdot (b-a),$$

where $x^* = \frac{a+b}{2}$ is the midpoint of $[a,b]$.

In light of Proposition 12.1, it is reasonable to define the surface area of a surface of revolution S for an *arbitrary* continuously differentiable function $f\colon [a,b] \to [0,\infty)$ by approximating the graph of f using a piecewise linear curve with n pieces, whose corresponding surface of revolution has an area that can be computed using the proposition, and then taking a limit as $n \to \infty$. Thus we define the surface area of S to be

$$\text{area}(S) = \lim_{n\to\infty} \sum_{i=1}^n 2\pi f(x_i^*)\sqrt{1 + f'(x_i^*)^2}\Delta x = \int_a^b 2\pi f(x)\sqrt{1 + f'(x)^2}\,dx, \tag{12.1}$$

where $\Delta x = \frac{b-a}{n}$, $x_i = a + i\Delta x$, and $x_i^* = \frac{1}{2}(x_{i-1} + x_i)$.

Using Leibniz notation, (12.1) can be written as $\int_a^b 2\pi y\sqrt{1 + (\frac{dy}{dx})^2}\,dx$, or even more compactly as $\int_a^b 2\pi y\,ds$, where ds is shorthand for $\sqrt{1 + (\frac{dy}{dx})^2}\,dx$, which is integrated to get arc length; this is a useful way to remember the formula for surface area.

Remark 12.2. The astute reader may notice that in Proposition 12.1, $f(x^*)$ and $f'(x^*)$ referred to the *linear* function f, while in (12.1) $f(x_i^*)$ and $f'(x_i^*)$ refer to the *original (nonlinear)* function f, rather than to its linear approximation. One can proceed as in the argument for arc length and use the MVT to produce x_i^* for which $f'(x_i^*)$ takes the value we expect, but even after doing this $f(x_i^*)$ need not be exactly equal to $\frac{1}{2}(f(x_{i-1}) + f(x_i))$. Thus to give a complete proof, one can use continuity of f to estimate the difference between these two quantities, and then prove that this difference goes to 0 as $n \to \infty$. We omit the details, but suggest this as a worthwhile exercise for the reader who wishes to see everything completely justified.

12.2. Examples of surface area

Example 12.3. Consider the surface of revolution obtained by rotating the curve $y = \sqrt{2-x}$ for $0 \le x \le 1$ around the x-axis. Then $\frac{dy}{dx} = -\frac{1}{2\sqrt{2-x}}$, so the surface area is

$$\begin{aligned} S &= \int_0^1 2\pi y \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx = \int_0^1 2\pi \sqrt{2-x} \cdot \sqrt{1 + \frac{1}{4(2-x)}}\, dx \\ &= 2\pi \int_0^1 \sqrt{2-x} \cdot \sqrt{\frac{9-4x}{4(2-x)}}\, dx = \pi \int_0^1 \sqrt{9-4x}\, dx. \end{aligned}$$

Making the substitution $u = 9 - 4x$, $du = -4\, dx$ gives

$$S = -\frac{\pi}{4} \int_9^5 \sqrt{u}\, du = \frac{\pi}{4} \int_5^9 \sqrt{u}\, du = \frac{\pi}{4} \left[\frac{2}{3} u^{3/2}\right]_5^9 = \frac{\pi}{6}\left(9^{3/2} - 5^{3/2}\right).$$

Because the function $x \mapsto y = \sqrt{2-x}$ is 1-1 on $[0,1]$, we could also study this surface treating x as a function of y: solving for x gives $x = 2 - y^2$, and y ranges over the interval $[1, \sqrt{2}]$. To make this change of variables in the integral, we replace dx with $\frac{dx}{dy}\, dy$ (here we are using the substitution rule) and write

$$S = \int_1^{\sqrt{2}} 2\pi y \sqrt{1 + \left(\frac{dy}{dx}\right)^2} \cdot \frac{dx}{dy}\, dy = \int_1^{\sqrt{2}} 2\pi y \sqrt{\left(\frac{dx}{dy}\right)^2 + \left(\frac{dy}{dx} \cdot \frac{dx}{dy}\right)^2}\, dy.$$

By the rule for derivatives of inverse functions, we have $\frac{dy}{dx} \cdot \frac{dx}{dy} = 1$, and so this formula can be rewritten as

$$S = \int_1^{\sqrt{2}} 2\pi y \sqrt{1 + \left(\frac{dx}{dy}\right)^2}\, dy \tag{12.2}$$

Recall from our discussion of arc length that the symbol ds can be interpreted either as $\sqrt{1 + (\frac{dy}{dx})^2}\, dx$ or as $\sqrt{1 + (\frac{dx}{dy})^2}\, ds$; then the mnemonic formula $\int 2\pi y\, ds$ for surface area can reasonably be interpreted as standing for either

$$\int_a^b 2\pi y \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx \qquad \text{or} \qquad \int_c^d 2\pi y \sqrt{1 + \left(\frac{dx}{dy}\right)^2}\, dy, \tag{12.3}$$

where $[a,b]$ is the integral over which x ranges and $[c,d]$ is the integral over which y ranges.

Remark 12.4. Be careful to note that in both versions of the surface area formula (12.3), the first part of the integrand is $2\pi y$, regardless of whether we are integrating with respect to x or y. This part of the integrand represents the fact that the surface is constructed by rotation around the x-axis, so that y represents the distance from the axis and $2\pi y$ represents the circumference of the circle obtained by cutting a cross-section of the surface at a given value of x. The part of the integral that depends on which variable we integrate with respect to is the derivative appearing inside the square root.

Returning to Example 12.3, we see that the surface area can also be computed using (12.2) and the formula $\frac{dx}{dy} = -2y$:

$$S = \int_1^{\sqrt{2}} 2\pi y\sqrt{1+(-2y)^2}\,dy = 2\pi \int_1^{\sqrt{2}} y\sqrt{1+4y^2}\,dy \qquad (u = 1+4y^2,\ du = 8y\,dy)$$
$$= 2\pi \int_5^9 \frac{1}{8}\sqrt{u}\,du = \frac{\pi}{4}\int_5^9 \sqrt{u}\,du,$$

which is exactly the same integral we obtained the first time around (after making the substitution $u = 9 - 4x$).

Remark 12.5. One can also consider surfaces of revolution around the y-axis, and in this case the roles of x and y are reversed, so the general formula is $\int 2\pi x\,ds$.

Example 12.6. To find the surface area of a sphere of radius R, we can treat it as the surface of revolution around the y-axis of the curve $x = \sqrt{R^2 - y^2}$ for $-R \le y \le R$, which has $\frac{dx}{dy} = -y/\sqrt{R^2-y^2}$, and we get

$$S = \int_{-R}^{R} 2\pi x\sqrt{1+\Big(\frac{dx}{dy}\Big)^2}\,dy = 2\pi \int_{-R}^{R} \sqrt{R^2-y^2}\cdot\sqrt{1+\frac{y^2}{R^2-y^2}}\,dy$$
$$= 2\pi \int_{-R}^{R} \sqrt{R^2-y^2}\cdot\sqrt{\frac{R^2}{R^2-y^2}}\,dy = 2\pi\int_{-R}^{R} R\,dy = 2\pi\Big[Ry\Big]_{-R}^{R} = 2\pi R(2R) = 4\pi R^2.$$

Example 12.7. Consider the curve $\{(x, \frac{1}{x}) : x \in [1,\infty)\}$, which has infinite length. The corresponding surface of revolution is called *Gabriel's horn.* Its surface area is given by the improper integral

$$S = \int_1^\infty 2\pi\cdot\frac{1}{x}\sqrt{1+\Big(-\frac{1}{x^2}\Big)^2}\,dx = \int_1^\infty \frac{2\pi}{x}\sqrt{1+\frac{1}{x^4}}\,dx. \tag{12.4}$$

Observe that for every $x \ge 1$, we have

$$\frac{2\pi}{x}\sqrt{1+\frac{1}{x^4}} \ge \frac{2\pi}{x} \ge \frac{1}{x},$$

and that we showed earlier that the improper integral $\int_1^\infty \frac{1}{x}\,dx$ is divergent. By the Comparison Theorem, the integral in (12.4) is divergent, which we interpret as meaning that Gabriel's horn has infinite surface area.

On the other hand, the *volume* of the region enclosed by Gabriel's horn is given by

$$V = \int_1^\infty \pi y^2\,dx = \int_1^\infty \frac{\pi}{x^2}\,dx = \lim_{t\to\infty}\Big[\frac{-\pi}{x}\Big]_1^t = \pi,$$

so this improper integral is convergent and the volume is finite.[8] This is a somewhat counter-intuitive state of affairs; can you explain it?

[8] You may also observe that if $P \subset \mathbb{R}^3$ is a plane containing the x-axis, then the corresponding cross-section of the enclosed region (its intersection with P) has infinite area, while this area is finite for any plane not containing the x-axis.

Lecture 13 Physical applications

Stewart §8.3

13.1. *Hydrostatic force and pressure

Here is an application from engineering. Suppose we have a dam that is holding back water, and we want to compute the total force that the water exerts on the dam; this is the *hydrostatic force*. The force per unit area at a given point is the *hydrostatic pressure*, and varies from point to point; near the surface of the water the pressure is relatively small, while deep down it is greater. Thus the force is obtained by integrating the pressure.

Imagine a small cube of water at depth d. If the water is motionless (we are at equilibrium) then all 6 faces of the cube experience the same force from the surrounding water; if it were not so, then the cube would move or deform. The top face experiences a downward force due to the column of water above it, which has mass ρAd, where ρ is the density of the fluid and A is the surface area of the top face of the cube. Thus the total force on the top face is $g\rho Ad$, where g is the gravitational constant, and thus the pressure in any given direction is force/area$=g\rho d$.

Example 13.1. Suppose that we consider a dam shaped like a trapezoid whose bottom and top edges are horizontal, with lengths 10 m and 18 m, respectively; suppose the total height of the dam is 16 m; and suppose that the water is 3/4 of the way to the top of the dam, so it is 12 m deep.

Let $w(x)$ denote the width of the dam at a depth x below the surface of the water. Then if we divide the interval $[0,12]$ into n pieces $[x_{i-1},x_i]$ of equal length $\Delta x = 12/n$, the strip of the dam between depths x_{i-1} and x_i is roughly a rectangle with width $w(x_i)$ and height Δx, so its area is $w(x_i)\Delta x$ and it experiences a hydrostatic force of pressure × area$=\rho gx\cdot w(x_i)\Delta x$. Summing up over all n strips and taking a limit as $n\to\infty$ gives a total force of

$$F = \lim_{n\to\infty}\sum_{i=1}^n \rho gxw(x_i)\Delta x = \int_0^{12} \rho gxw(x)\,dx. \tag{13.1}$$

In this case we see that $w(x)=ax+b$ for some $a,b\in\mathbb{R}$, which can be determined by using the fact that $w(12)=10$ (at the deepest point) and $w(-4)=18$ (at the top of the dam), so we have

$$12a+b=10 \quad\text{and}\quad -4a+b=18.$$

Subtracting the two equations gives $16a=-8$, so $a=-1/2$, and thus $b=18+4a=18-2=16$, which gives $w(x)=16-x/2$, and the total hydrostatic force on the dam is

$$F=\int_0^{12}\rho g\Big(16x-\frac{x^2}{2}\Big)\,dx=\rho g\Big[8x^2-\frac{x^3}{6}\Big]_0^{12}=\rho g\Big(8\cdot(12)^2-\frac{(12)^3}{6}\Big)$$

$$=\rho g\cdot 144\cdot\Big(8-\frac{12}{6}\Big)=\rho g\cdot 144\cdot 6=864\rho g.$$

Note that the number 864 represents $\int_0^{12} xw(x)\,dx$ and thus has units m^3. Using the values $g = 9.8\,\mathrm{m/s}^2$ and $\rho = 1000\,\mathrm{kg/m}^3$, we get

$$F \approx 8.47 \times 10^6\,\mathrm{N},$$

where the units of force are Newtons, $1\,\mathrm{N} = 1\,\mathrm{kg\,m/s}^2$.

Example 13.2. An undersea laboratory is built on the ocean floor where the water is 100 m deep. The end of the lab has the shape of a sine function, with width 10 m and height 5 m. How much hydrostatic force does the end of the lab experience?

Let y be the height above the ocean floor; then the depth of the water at any given point is $100 - y$, and the same arguments that lead to (13.1) show that the total force is

$$F = \int_0^5 \rho g(100 - y)w(y)\,dy,$$

where $w(y)$ is the width of the lab at height y. Taking the x-axis to be centred at the centre of the lab, the height of the lab at position x is given by $y(x) = 5\cos(\frac{\pi x}{10})$ (draw the graph of this function and observe that it has the height and width specified), and so if y is given, we have $x = \pm\frac{10}{\pi}\cos^{-1}(\frac{y}{5})$. The distance between these two x-coordinates is $\frac{20}{\pi}\cos^{-1}(\frac{y}{5})$, and this is our value for $w(y)$. Thus the total force is

$$\begin{aligned} F &= \int_0^5 \frac{20}{\pi}\rho g(100 - y)\cos^{-1}\left(\frac{y}{5}\right)dy \qquad\qquad (z = \tfrac{y}{5},\ dy = 5\,dz) \\ &= \frac{100}{\pi}\rho g\int_0^1 (100 - 5z)\cos^{-1}(z)\,dz, \end{aligned}$$

and we can integrate this using parts with $u = \cos^{-1} z$, $dv = (20 - z)\,dz$ to get

$$\begin{aligned} F &= \frac{500}{\pi}\rho g\int_0^1 \underbrace{\cos^{-1}(z)}_{u}\,\underbrace{(20 - z)\,dz}_{dv} \\ &= \frac{500}{\pi}\rho g\Big[\cos^{-1}(z)(20z - \tfrac{1}{2}z^2)\Big]_0^1 - \frac{500}{\pi}\rho g\int_0^1 \frac{20z - \frac{1}{2}z^2}{-\sqrt{1 - z^2}}\,dz. \end{aligned}$$

Observe that $\cos^{-1}(1) = 0$, so the first term vanishes at both $z = 0$ and $z = 1$, giving

$$F = \frac{500}{\pi}\rho g\int_0^1 \frac{20z - \frac{1}{2}z^2}{\sqrt{1 - z^2}}\,dz.$$

To evaluate the first part of the integral we observe that

$$\int_0^1 \frac{z}{\sqrt{1 - z^2}}\,dz = \Big[-\sqrt{1 - z^2}\Big]_0^1 = 0 - (-1) = 1.$$

For the second part, we put $z = \sin\theta$, $dz = \cos\theta\,d\theta$, $\sqrt{1 - z^2} = \cos\theta$, and get

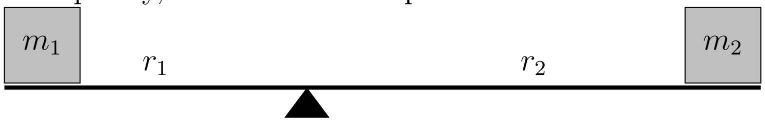
$$\begin{aligned} \int_0^1 \frac{z^2}{\sqrt{1 - z^2}}\,dz &= \int_0^{\pi/2} \frac{\sin^2\theta}{\cos\theta}\cos\theta\,d\theta = \int_0^{\pi/2} \sin^2\theta\,d\theta \\ &= \int_0^{\pi/2} \frac{1 - \cos 2\theta}{2}\,d\theta = \frac{\pi}{4} - \left[\frac{1}{4}\sin 2\theta\right]_0^{\pi/2} = \frac{\pi}{4}. \end{aligned}$$

Putting it all together gives

$$F = \frac{500}{\pi}\rho g\Big(20 \cdot 1 - \frac{1}{2} \cdot \frac{\pi}{4}\Big) = \rho g\Big(\frac{1000}{\pi} - \frac{125}{2}\Big) \approx 3.07 \times 10^8 \,\mathrm{N}.$$

13.2. Center of mass

Suppose we have a rigid plank supported on a fulcrum, with two masses m_1 and m_2 placed on opposite sides of the fulcrum, at distances r_1 and r_2, as shown in the picture. For simplicity, assume that the plank is massless.



We want to determine conditions on m_1, m_2, r_1, r_2 such that the system balances; that is, if the masses are initially at rest, then they remain at rest.[9] Use a coordinate system in which the initial height of the masses is 0; then their initial potential energy is 0, and so is their initial kinetic energy. Let $\theta(t)$ be the angle made by the plank with the horizontal at time t, and let $v_1(t), v_2(t)$ be the velocities of the two masses at time t. Then the total energy at time t is

$$E(t) = \underbrace{\frac{1}{2}m_1v_1^2 + \frac{1}{2}m_2v_2^2}_{\text{kinetic energy}} + \underbrace{m_1g(-r_1\sin\theta) + m_2g(r_2\sin\theta)}_{\text{(gravitational) potential energy}}$$

By conservation of energy, we must have $E(t) = 0$ for all t. Observe that if $m_1r_1 = m_2r_2$, then the potential energy is 0 no matter what θ is. Thus the kinetic energy must also be 0, which means that $m_1v_1^2 + m_2v_2^2 = 0$, but this is only possible if $v_1 = v_2 = 0$ for all t. Thus we have proven the following.

Proposition 13.3 (Law of the lever)**.** *If $m_1r_1 = m_2r_2$ in the situation above, then the system is balanced and remains in equilibrium, motionless.*

Exercise 13.4. Prove that if $m_1r_1 > m_2r_2$, then the plank will rotate counterclockwise – m_1 will sink and m_2 will rise – and vice versa if $m_1r_1 < m_2r_2$.

Now suppose we have a finite set of masses $m_1, m_2, \dots, m_n$ at locations $x_1, x_2, \dots, x_n$ along the plank, and that the fulcrum is located at position $\bar{x}$. Note that these values can be either positive or negative, since we are not specifying which side of the fulcrum each mass lies on, and we do not require the fulcrum to lie at 0. In particular, the location of the mass m_i *relative to the fulcrum* is given not by x_i, but by $x_i - \bar{x}$, with a negative value indicating that the mass is to the left of the fulcrum, and a positive value indicating that it is to the right.

Repeating the same reasoning as before, we see that the system will be in equilibrium if and only if the values of $m_i, x_i, \bar{x}$ have the property that the change in potential energy is 0 no matter what value θ takes. If the plank is at angle θ, then mass m_i is at height $(x_i - \bar{x})\sin\theta$, and thus the total change in potential energy is

$$\sum_{i=1}^{n} m_i(x_i - \bar{x})\sin\theta.$$

[9] I learned the argument given here from a short write-up by Peter McLoughlin.

We need this to vanish for all θ; equivalently, we require that

$$0 = \sum_{i=1}^{n} m_i(x_i - \bar{x}) = \Big(\sum_{i=1}^{n} m_i x_i\Big) - \Big(\sum_{i=1}^{n} m_i\Big)\bar{x}.$$

Thus we have proved the following.

Proposition 13.5. *The plank with masses $m_1, \dots, m_n$ placed at positions $x_1, \dots, x_n$ is in equilibrium if and only if the fulcrum is placed at position*

$$\bar{x} = \frac{\sum_{i=1}^{n} m_i x_i}{\sum_{i=1}^{n} m_i}. \tag{13.2}$$

The point $\bar{x}$ where the fulcrum must be placed to ensure equilibrium is called the *center of mass* of the system; it is also called the *center of gravity* or the *centroid*. The numerator in (13.2) is called the *moment*[10] *of the system about the origin*, and represents the tendency that the system would have to rotate clockwise (if the moment is positive) or counterclockwise (if the moment is negative) **if** we were to place the fulcrum at the origin. The denominator in (13.2) is of course the total mass of the system.

Lecture 14 Two- and three-dimensional objects

Stewart §8.3

14.1. Center of mass in two dimensions

Now suppose we have a system of masses $m_1, \dots, m_n$ located in the plane $\mathbb{R}^2$, at positions $(x_1, y_1), \dots, (x_n, y_n)$. We would like to find the point $(\bar{x}, \bar{y})$ with the property that if our masses are placed on a flat (massless) surface, which is then placed on a fulcrum located at $(\bar{x}, \bar{y})$, then the system would balance in equilibrium. This point $(\bar{x}, \bar{y})$ will again be called the center of mass, or centroid, of the system.

First imagine that we support the surface not on a fulcrum placed at a single point, but on a rod that is oriented parallel to the y-axis, so that rotation is only possible around this axis. Then the y-coordinates of the masses are irrelevant, and all that matters is their x-coordinates. As in the previous section, we see that

$$\bar{x} = \frac{M_y}{m} \qquad \text{where } M_y = \sum_{i=1}^{n} m_i x_i \text{ and } m = \sum_{i=1}^{n} m_i. \tag{14.1}$$

We call M_y the *moment around the y-axis.* A similar argument reveals that

$$\bar{y} = \frac{M_x}{m} \qquad \text{where } M_x = \sum_{i=1}^{n} m_i y_i \text{ and } m = \sum_{i=1}^{n} m_i, \tag{14.2}$$

where M_x is the *moment around the x-axis.*

[10] If we multiply the moment by the gravitational constant g, we get the *moment of force*, or *torque*.

Example 14.1. Suppose we place three small objects with masses 1, 2, and 4 at positions $(0,1)$, $(1,1)$, and $(2,3)$, respectively. Then the two moments are

$$M_y = 1 \cdot 0 + 2 \cdot 1 + 4 \cdot 2 = 10 \quad \text{and} \quad M_x = 1 \cdot 1 + 2 \cdot 1 + 4 \cdot 3 = 15.$$

Since $m = 1 + 2 + 4 = 7$, we see that the center of mass is at $(\frac{10}{7}, \frac{15}{7})$.

Remark 14.2. If we have a set of masses with moments M_y and M_x, then (14.1) and (14.2) can be rewritten as

$$M_y = m\bar{x} \quad \text{and} \quad M_x = m\bar{y};$$

the moments of the entire set of masses are the same as the moments of a single point mass located at the centroid $(\bar{x}, \bar{y})$ with mass m. In other words, the moments are unchanged if we move all of the masses to the centroid.

Remark 14.3. Observe that if we have two sets S_1 and S_2 of masses, and compute their moments $M_y(S_1)$ and $M_y(S_2)$ independently, then the moment of the overall system comprising all the masses is given by $M_y(S_1 \cup S_2) = M_y(S_1) + M_y(S_2)$. A similar result holds for the the moments around the x-axis.

14.2. Continuous objects

Now we consider the continuous case – a plate with uniform density ρ. Let $R \subset \mathbb{R}^2$ be the region describing the shape of the plate, and let $C(R) \in \mathbb{R}^2$ denote the centroid of R; that is, the point at which a fulcrum must be placed in order for the plate to balance. As before, we have $C(R) = (M_y(R)/m, M_x(R)/m)$, where m is the total mass of the plate and $M_y(R)$, $M_x(R)$ are the moments of R around the y- and x-axes, respectively. The difference is that this time we do not have a formula for M_y and M_x; we must derive one. To do this, we assume that the centroid and moments obey the following principles.

(1) *Symmetry:* If R is symmetric around a line ℓ, then $C(R)$ lies on ℓ.
(2) *Replacement:* If all of the mass of R is moved to a single point located at $C(R)$, then the moments M_y and M_x are unchanged.
(3) *Additivity:* If R_1 and R_2 are disjoint regions, then $M_y(R_1 \cup R_2) = M_y(R_1) + M_y(R_2)$, and similarly for M_x.

Observe that the second and third principles are analogues of Remarks 14.2 and 14.3, respectively.

For simplicity we first assume that the plate is described by the set

$$R = \{(x,y) : a \leq x \leq b, 0 \leq y \leq f(x)\} = \bigcup_{x \in [a,b]} \{x\} \times [0, f(x)] \subset \mathbb{R}^2$$

for some function $f \colon [a,b] \to [0,\infty)$. As usual we approximate R by taking $n \in \mathbb{N}$ large, dividing $[a,b]$ into n intervals of length $\Delta x = (b-a)/n$ with endpoints $x_i = a + i\Delta x$, and considering the union of rectangles $R_i := [x_{i-1}, x_i] \times [0, f(\bar{x}_i)]$, where $\bar{x}_i = \frac{1}{2}(x_{i-1}, x_i)$. As long as f is continuous, it is reasonable to expect that

$$M_y(R) = \lim_{n\to\infty} M_y\Big(\bigcup_{i=1}^{n} R_i\Big) \quad \text{and} \quad M_x(R) = \lim_{n\to\infty} M_x\Big(\bigcup_{i=1}^{n} R_i\Big). \tag{14.3}$$

By the third principle above (additivity), we have

$$M_y\Big(\bigcup_{i=1}^n R_i\Big) = \sum_{i=1}^n M_y(R_i) \quad \text{and} \quad M_x\Big(\bigcup_{i=1}^n R_i\Big) = \sum_{i=1}^n M_x(R_i). \tag{14.4}$$

The centroid of R_i lies at $(\bar{x}_i, \frac{1}{2}f(\bar{x}_i))$ by the first principle above (symmetry), and the mass of R_i is $\rho f(\bar{x}_i)\Delta x$. Thus the second principle above (replacement) gives

$$M_y(R_i) = \rho \bar{x}_i f(\bar{x}_i)\Delta x \quad \text{and} \quad M_x(R_i) = \rho \cdot \frac{1}{2} f(\bar{x}_i)^2 \Delta x. \tag{14.5}$$

Combining (14.3)–(14.5) gives

$$\begin{aligned} M_y(R) &= \lim_{n\to\infty} \sum_{i=1}^n \rho \bar{x}_i f(\bar{x}_i)\Delta x = \rho \int_a^b x f(x)\,dx, \\ M_x(R) &= \lim_{n\to\infty} \sum_{i=1}^n \rho \cdot \frac{1}{2} f(\bar{x}_i)^2 \Delta x = \rho \int_a^b \frac{1}{2}\big(f(x)\big)^2\,dx. \end{aligned}$$

Since $m = \rho \int_a^b f(x)\,dx$, we conclude that the centroid of R has coordinates given by

$$\bar{x} = \frac{\int_a^b x f(x)\,dx}{\int_a^b f(x)\,dx} \quad \text{and} \quad \bar{y} = \frac{\int_a^b \frac{1}{2} f(x)^2\,dx}{\int_a^b f(x)\,dx}. \tag{14.6}$$

Example 14.4. We find the centroid of a semicircular region R with radius r. For concreteness take the upper half of the circle centered at the origin. Since R is symmetric around the y-axis we immediately have $\bar{x} = 0$. For $\bar{y}$, we describe the region via $f(x) = \sqrt{r^2 - x^2}$ on $[-r, r]$ and observe that $\int_{-r}^r f(x)\,dx = \frac{1}{2}\pi r^2$ by the formula for circle area, so that (14.6) gives

$$\begin{aligned} \bar{y} &= \frac{\int_{-r}^r \frac{1}{2} f(x)^2\,dx}{\int_{-r}^r f(x)\,dx} = \frac{1}{\pi r^2}\int_{-r}^r (r^2 - x^2)\,dx = \frac{2}{\pi r^2}\int_0^r (r^2 - x^2)\,dx \\ &= \frac{2}{\pi r^2}\Big[r^2 x - \frac{1}{3}x^3\Big]_0^r = \frac{2}{\pi r^2}\Big(r^3 - \frac{1}{3}r^3\Big) = \frac{2}{\pi r^2}\cdot\frac{2}{3}r^3 = \frac{4r}{3\pi}. \end{aligned}$$

Thus the centroid of the region is located at $(0, \frac{4r}{3\pi})$.

If we consider a more general region described as

$$R = \{(x,y) : x \in [a,b], y \in [g(x), f(x)]\}, \tag{14.7}$$

where $g, f\colon [a,b] \to \mathbb{R}$ are continuous functions with $g \le f$, then similar arguments give

$$\bar{x} = \frac{1}{A}\int_a^b x(f(x) - g(x))\,dx \quad \text{and} \quad \bar{y} = \frac{1}{A}\int_a^b \frac{1}{2}\big(f(x)^2 - g(x)^2\big)\,dx, \tag{14.8}$$

where $A = \int_a^b (f(x) - g(x))\,dx$ is the area of R.

14.3. Pappus's theorem

Theorem 14.5 (Pappus's theorem)**.** *Let R be a region in the plane that lies entirely to one side of some line ℓ, and let V be the volume of the solid of revolution formed by rotating R around the line ℓ. Let A be the area of R and let d be the distance traveled by the centroid of R as it revolves around ℓ. Then $V = Ad$.*

Proof. Without loss of generality, take ℓ to be the y-axis, and let R be given in terms of functions g, f as in (14.7).[11] Recall how we find volume by cylindrical shells:

(1) the area of the annulus with inner radius p and outer radius q is $\pi q^2 - \pi p^2 = \pi(q^2 - p^2) = 2\pi m(q-p)$, where $m = \frac{p+q}{2}$;
(2) thus the volume of the cylindrical shell formed by rotating the rectangle $[x_{i-1}, x_i] \times [g(\bar{x}_i), f(\bar{x}_i)]$ around the y-axis is $2\pi\bar{x}_i(f(\bar{x}_i) - g(\bar{x}_i))$, where $\bar{x}_i$ is the midpoint of $[x_{i-1}, x_i]$;
(3) the volume of R is

$$V = \lim_{n\to\infty} \sum_{i=1}^{n} 2\pi\bar{x}_i(f(\bar{x}_i) - g(\bar{x}_i))\Delta x = \int_a^b 2\pi x(f(x) - g(x))\,dx,$$

where $\Delta x = (b-a)/n$ and $x_i = a + i\Delta x$.

Recalling the first half of (14.8), we have

$$V = 2\pi \int_a^b x(f(x) - g(x))\,dx = 2\pi A\bar{x},$$

and since $2\pi\bar{x}$ is the distance d traveled by the centroid as it rotates, this proves the theorem. □

Example 14.6. Consider a disc with center $(R, 0)$ and radius r, where $0 < r < R$. The centroid of the disc is its center (by the symmetry principle), and the corresponding solid of revolution is a torus, whose volume is

$$V = Ad = (\pi r^2)(2\pi R) = 2\pi^2 r^2 R.$$

Lecture 15 *Probability

Stewart §8.5

A *random variable* is a quantity that depends on some random factors. For example, any of the following could be described by a random variable:

- W = the sum of the numbers on a pair of dice after they are rolled;
- X = the number of students who come to class on a randomly selected day;
- Y = the height of a randomly selected person;
- Z = the amount of rainfall during a randomly selected week.

[11] If R cannot be written in this form, you first need to decompose it as a finite union of such regions.

The first two examples above, W and X, are *discrete* random variables, meaning that we can make a list of all the possible values they can take, and then assign a probability to each individual value. The last two examples, Y and Z, are *continuous* random variables, meaning that they can take a continuum of values; instead of listing all possible values, we allow the value to be any real number. (Of course, some parts of the real line may have zero probability: both Y and Z will have probability 1 of being ≥ 0.)

The *probability distribution* of a random variable tells us the probabilities associated to the different values it can take. For a discrete random variable, we can describe the distribution by simply listing the probabilities associated to each of the possible values: for example, if W is the sum of the numbers on a pair of dice, then $\mathbb{P}(W = 2) = \frac{1}{36}$ because the $6 \times 6 = 36$ equally likely outcomes include exactly one that produces a sum of 2, and we can similarly list $\mathbb{P}(W = 3)$, $\mathbb{P}(W = 4)$, and so on.

For a continuous random variable X, we must do something else, since we cannot list all the possible values. Rather, we describe the distribution by a *probability density function*; this is a function $f \colon \mathbb{R} \to [0, \infty)$ with the property that

$$\underbrace{\mathbb{P}(a \leq X \leq b)}_{\text{probability that } a \leq X \leq b} = \int_a^b f(x)\,dx \text{ for every } a < b \in \mathbb{R}.$$

The interpretation of this is that if we make n independent observations of the random variable X, then the proportion of observations for which $a \leq X \leq b$ will converge to $\int_a^b f(x)\,dx$ as $n \to \infty$ (this is called the *law of large numbers*).

Probability density functions are required to have $f(x) \geq 0$ for all x, and to satisfy $\int_{-\infty}^{\infty} f(x)\,dx = 1$. The first condition guarantees that probabilities are always ≥ 0, and the second condition guarantees that the probability that *something* happens is equal to 1.

Example 15.1. An *exponentially distributed random variable* takes only positive values and has a probability density function (PDF) that decays exponentially as $x \to \infty$; that is, $f(x) = 0$ for $x < 0$, and there are $c, \lambda > 0$ such that $f(x) = ce^{-\lambda x}$ for $x \geq 0$. Random variables like this are often used to model *waiting time* phenomena in which X represents the amount of time until the next occurrence of a particular event, such as my dog barking at a car that drives past my house.

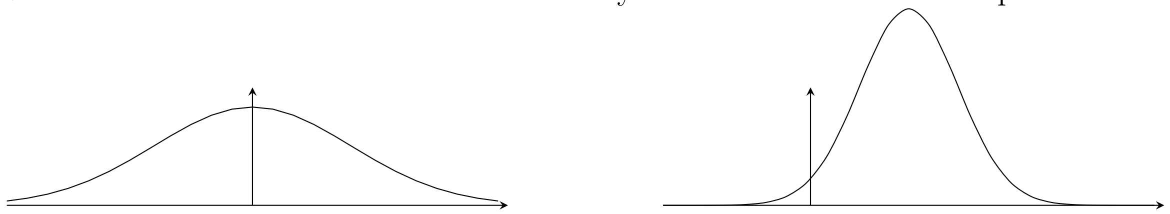
The value of λ reflects the rate at which the PDF decays; smaller λ means that X is more likely to take larger values, while larger λ means that it is more likely to take smaller values. We need to determine the value of c to guarantee that f is normalized: $\int_{-\infty}^{\infty} f(x)\,dx = 1$. From the definition of f we get

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_0^{\infty} f(x)\,dx = \lim_{t\to\infty} \int_0^t ce^{-\lambda x}\,dx = \lim_{t\to\infty} \Big[-\frac{c}{\lambda} e^{-\lambda x} \Big]_0^t = \frac{c}{\lambda}.$$

Thus we must put $c = \lambda$, obtaining a PDF of $f(x) = \lambda e^{-\lambda x}$. Then the probability that X lies in an interval $[a, b]$ for $a \geq 0$ is given by

$$\mathbb{P}(a \leq X \leq b) = \int_a^b \lambda e^{-\lambda x}\,dx = \Big[-e^{-\lambda x} \Big]_a^b = e^{-\lambda b} - e^{-\lambda a}.$$

Example 15.2. A *normally distributed random variable* can take both positive and negative values and has a PDF given by $f(x) = \frac{1}{A}e^{-(x-\mu)^2/(2\sigma^2)}$, where μ is the *mean* of the distribution, σ is the *standard deviation*, and $A = \int_{-\infty}^{\infty} e^{-(x-\mu)^2/(2\sigma^2)}\,dx$ is the normalizing constant that guarantees the property $\int_{-\infty}^{\infty} f(x)\,dx = 1$. This is also called a *Gaussian distribution* or sometimes informally a *bell curve* due to its shape.



$\mu = 0$, $\sigma = 1$ $\qquad$ $\mu = 1$, $\sigma = 0.5$

It is possible to prove that $A = \sqrt{2\pi\sigma^2}$, but this requires tools that we have not yet developed (recall that we cannot find $\int e^{-x^2}\,dx$ explicitly). Note that varying μ has the effect of sliding the graph of f to the left or right; the graph is symmetric around the line $x = \mu$. Varying σ has the effect of squeezing or stretching it horizontally, so that when σ is small more of the area under the graph is concentrated closer to the line $x = \mu$, and when σ is large more area is located further away from this line. Thus σ quantifies how likely the value of X is to be close to the mean μ.

In the example of the normal distribution, the symmetry of the PDF makes it reasonable to interpret μ as an average, or mean, since for every range of values greater than μ, there is a range of values on the opposite side of μ that are achieved with equal probability. But how do we find the mean of an arbitrary random variable?

First recall that if we measure a random variable N times and record the results of the measurements as $X_1, \dots, X_N$, then the *observed* average value is

$$\bar{X} = \frac{1}{N}\sum_{j=1}^{N} X_j.$$

Suppose for a moment that we have a discrete random variable, which only takes values from a finite set $\{x_1, \dots, x_n\}$. Then for each i we can write k_i for the number of times that we see the value x_i appear in the list $(X_1, \dots, X_N)$, and obtain

$$\bar{X} = \frac{1}{N}\sum_{j=1}^{N} X_j = \frac{1}{N}\sum_{i=1}^{n} k_i x_i. \tag{15.1}$$

Now return to the case of a continuous random variable. Suppose we fix a large $t > 0$, a large $n \in \mathbb{N}$, and split the interval $[-t, t]$ into n intervals of length $\Delta x = 2t/n$ by putting $x_i = -t + i\Delta x$. If we measure the random variable N different times, we expect $\approx N\int_{x_{i-1}}^{x_i} f(x)\,dx \approx Nf(x_i)\Delta x$ of these measurements to lie in the interval $[x_{i-1}, x_i]$. Thus (15.1) gives

$$\text{average of } X \approx \frac{1}{N}\sum_{i=1}^{n} Nf(x_i)\Delta x \cdot x_i = \sum_{i=1}^{n} x_i f(x_i)\Delta x.$$

Once again we recognize this as a Riemann sum, whose limit as $n \to \infty$ is $\int_{-t}^{t} xf(x)\,dx$. Taking a limit as $t \to \infty$, we see that the average value (mean) of the random variable X with probability density function f is given by

$$\mu = \int_{-\infty}^{\infty} xf(x)\,dx. \tag{15.2}$$

Exercise 15.3. Use the symmetry of the normal distribution to show that this agrees with the use of the notation μ there.

Remark 15.4. In light of Remark 10.20, you should be mildly uneasy (at least) with our casual use of the relationship $\int_{-\infty}^{\infty} xf(x)\,dx = \lim_{t\to\infty} \int_{-t}^{t} f(x)\,dx$. This works fine provided the improper integral $\int_{-\infty}^{\infty} xf(x)\,dx$ is convergent; however, if the improper integral is divergent then (15.2) is invalid, and in fact we must say that in this case the mean does not exist!

Exercise 15.5. Find $c > 0$ such that $f(x) = \frac{c}{1+x^2}$ is a probability density function, and show that in this case the improper integral in (15.2) is divergent.

Remark 15.6. The mean μ is sometimes called the *first moment*. Observe that it is given by the same integral that we used to compute the moment around the y-axis of a region in $\mathbb{R}^2$. Since $\int_{-\infty}^{\infty} f(x)\,dx = 1$ for a PDF, this means that the centroid of the region under the graph of f lies on the line $x = \mu$.

In probability theory one also needs to study *higher moments* such as $\int_{-\infty}^{\infty} x^2 f(x)\,dx$, $\int_{-\infty}^{\infty} x^3 f(x)\,dx$, and so on. As with the mean, these integrals may or may not be convergent, depending on which probability density function we consider.

Example 15.7. For the exponential distribution given by $f(x) = \lambda e^{-\lambda x}$, the mean is

$$\begin{aligned}\mu &= \int_0^{\infty} \lambda x e^{-\lambda x}\,dx = \lim_{t\to\infty} \Big[\lambda x \cdot (-\lambda^{-1} e^{-\lambda x})\Big]_0^t - \int_0^t \lambda(-\lambda^{-1} e^{-\lambda x})\,dx \\ &= \lim_{t\to\infty} -te^{-\lambda t} + \int_0^t e^{-\lambda x}\,dx = \lim_{t\to\infty} \Big[-\frac{1}{\lambda} e^{-\lambda x}\Big]_0^t = \frac{1}{\lambda}.\end{aligned}$$

Part III. Differential equations

Lecture 16 Ideas and examples

Stewart §9.1 and §9.2

16.1. Real-world problems modeled by DEs

When we write down a model describing some kind of real-world situation in which our goal is to determine a particular function f, we often end up with a *differential equation* (DE) that contains both f and some of its derivatives. For example, this occurred when we considered the hanging cable problem and discovered that the equation $y = f(x)$ describing the catenary could be determined by first finding the arc length function $s(x)$, which satisfies the equation (11.8): $\frac{ds}{dx} = \frac{1}{a}\sqrt{a^2 + s^2}$. We were able to solve this by rewriting it as $\frac{dx}{ds} = a/\sqrt{a^2 + s^2}$ and then integrating with respect to s; note that this represents the simplest sort of differential equation, where the function to be determined (in this case $x(s)$) appears only on the LHS in terms of its derivative, and thus can be found by computing a single integral. Most of the DEs we encounter from now on will be more involved than this.

One instructive example arose last semester when we studied population growth. If $P(t)$ represents the size of a particular population at time t, then the simplest model describing how P evolves in time simply accounts for the change due to reproduction and death:. Write $k_r > 0$ for the rate at which reproduction happens, so that the population increase due to reproduction in a short time interval Δt is $k_r P(t)\Delta t$, and $k_d > 0$ for the rate at which members of the population die, so that the decrease due to death in time Δt is $-k_d P(t)\Delta t$. Thus

$$\frac{dP}{dt} = \lim_{\Delta t \to 0} \frac{\Delta P(t)}{\Delta t} = \lim_{\Delta t \to 0} \frac{k_r P(t)\Delta t - k_d P(t)\Delta t}{\Delta t} = (k_r - k_d)P(t).$$

Writing $k := k_r - k_d$, we see that the population function satisfies the differential equation

$$\frac{dP}{dt} = kP. \tag{16.1}$$

If $k_r > k_d$ then $k > 0$ and the population grows; if $k_r < k_d$ then $k < 0$ and the population shrinks. We saw last semester that (16.1) can be solved by dividing both sides by P and using logarithmic derivatives:

$$\frac{d}{dt}\ln P = \frac{P'}{P} = \frac{kP}{P} = k \quad \Rightarrow \quad \ln P(t) = kt + C \quad \Rightarrow \quad P(t) = C_0 e^{kt},$$

where $C_0 = e^C$ is a constant of integration which can take any value in $(0, \infty)$. (Here we assume that the population is positive; if $P(t) = 0$ then there is nothing to model.) To determine C_0 we need to know the value of the population at some point in time: if we know the population at some time t_0, then we have

$$P(t_0) = C_0 e^{kt_0} \quad \Rightarrow \quad C_0 = P(t_0)e^{-kt_0} \quad \Rightarrow \quad P(t) = P(t_0)e^{-kt_0}e^{kt} = P(t_0)e^{k(t-t_0)}.$$

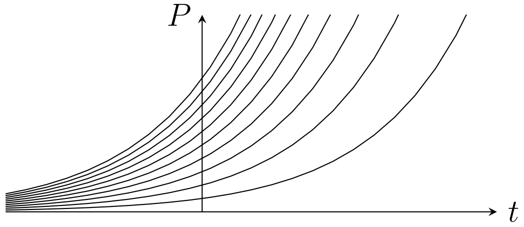
In particular, if we know the population at time 0 then we have

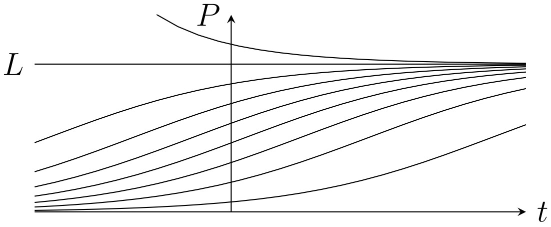
$$P(t) = P(0)e^{kt}.$$

The problem of finding $P(t)$ given that

(1) P satisfies (16.1) and
(2) $P(t_0)$ is known

is called an *initial value problem.* In an initial value problem, we expect to get a single function as the solution. If all we have is a differential equation but are not given the initial value, then we expect to get a whole family of solutions, such as $P(t) = C_0 e^{kt}$; here the constant C_0 can be thought of as a parameter telling us which member of the family we are looking at. The picture at left shows some of the members of this family for the DE in (16.1) when $k > 0$.





Of course this model is not entirely realistic, because sooner or later the population will start to run out of resources and growth will slow. A more realistic model incorporates the *carrying capacity* of the environment in which the population lives, and has solutions with the shape shown in the right-hand figure. To describe it quantitatively, let L be the largest population that the environment can sustainably support; then we would like to have $P' \approx kP$ when P is small ($P \ll L$), while P'/P decreases for larger values of P, with P' becoming *negative* when $P > L$. This last requirement represents the idea that if the population is too large, then it will shrink towards the carrying capacity L. One DE that meets these requirements is the following *logistic DE* introduced by Verhulst in the 1840s:

$$\frac{dP}{dt} = kP\Big(1 - \frac{P}{L}\Big). \tag{16.2}$$

This is not quite so easy to solve as (16.1) was: dividing both sides by P does not help, because the RHS still contains P and so a straightforward integration does not solve the problem. We will see how to solve DEs like this in a few days. In the meantime, we can make some qualitative observations.

(1) $\frac{dP}{dt} = 0$ if and only if $P = 0$ or $P = L$. In particular, $P(t) = 0$ and $P(t) = L$ are both solutions of (16.2). Solutions such as these, where the function in question is constant, are called *equilibrium solutions.*
(2) $\frac{dP}{dt} > 0$ when $P \in (0, L)$, and $\frac{dP}{dt} < 0$ when $P \in (L, \infty)$. The picture suggests (and we will later prove) that $\lim_{t\to\infty} P(t) = L$ as long as the initial condition is positive.

Example 16.1. Consider a mass m attached to a spring, moving horizontally on a frictionless surface. Let $x(t)$ denote the displacement of the mass from its equilibrium position at time t. Then *Hooke's law* says that the spring exerts a force $F = -kx$ on the mass, where $k > 0$ is a constant depending on the stiffness of the spring. Since $F = ma = m\ddot{x}$, the position function x satisfies the DE

$$\frac{d^2x}{dt^2} = -\frac{k}{m}x. \tag{16.3}$$

Exercise 16.2. Show that writing $\omega = \sqrt{\frac{k}{m}}$, the functions $x(t) = \sin(\omega t)$ and $x(t) = \cos(\omega t)$ are both solutions of (16.3). Can you think of any others?

Definition 16.3. The *order* of a differential equation is the order of the highest derivative that appears in the equation.

The population DEs (16.1) and (16.2) are first-order differential equations, while the spring equation (16.3) is second-order.

16.2. Explicit solutions using logarithms

The problem of finding the indefinite integral of a function f can be viewed as a differential equation $F' = f$, where the indefinite integral F is the solution of the DE. As we saw already, it is not always possible to find an elementary formula for the indefinite integral (such as when $f(x) = e^{-x^2}$) and thus one should not expect to always be able to write down an elementary formula for a differential equation. Indeed, in general the problem of solving differential equations is substantially more difficult than the problem of finding indefinite integrals, and there is no single technique that we can rely on to always lead us to the answer.

Remark 16.4. A significant item in the theory of differential equations is to determine whether or not a given DE even has a solution (*existence*), and if so, whether it is possible to have multiple solutions with the same initial conditions, or whether there is only one (*uniqueness*). Such existence and uniqueness results are not part of this course, however.

With that said, there are many classes of DEs for which it is possible to find a solution by reducing the problem to that of finding an indefinite integral. The population DE (16.1) illustrated this, and the technique used there works for any DE of the form

$$\frac{dy}{dx} = f(x)y, \tag{16.4}$$

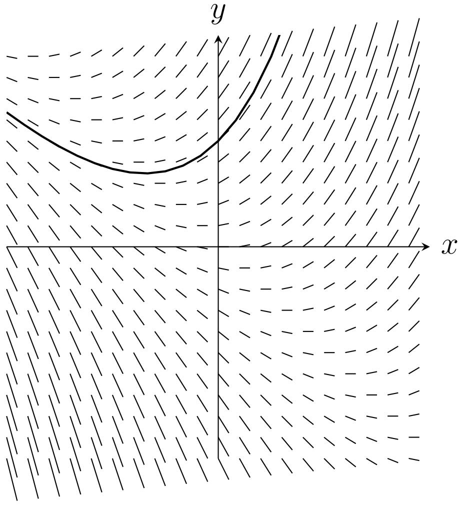
where $f(x)$ is any given integrable function. Dividing both sides by y gives

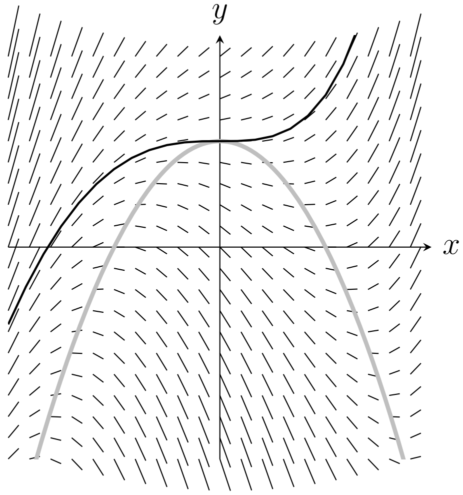
$$\frac{d}{dx}\log y = \frac{y'}{y} = f(x) \quad \Rightarrow \quad \log y(x) = \int f(x)\,dx \quad \Rightarrow \quad y(x) = e^{\int f(x)\,dx}.$$

Example 16.5. To solve $y' = xy$ with $y(1) = 1$, we write it as $(\log y)' = x$, so $\log y = \frac{1}{2}x^2 + C$, and $\log y(1) = \log 1 = 0$ together with $\log y(1) = \frac{1}{2} + C$ gives $C = -\frac{1}{2}$, so the solution of the initial value problem is $y = e^{-1/2}e^{x^2/2}$.

16.3. *Qualitative solutions using direction fields

Suppose we are confronted with the differential equation $y' = x + y$. This does not immediately reduce to a simple integration like the examples we solved so far. But we can still at least sketch the general shape of the solutions by using a *direction field* (also called a *slope field*), where at each point $(x, y) \in \mathbb{R}^2$ we put a short line segment with slope $x + y$, as shown in the first picture below. Then every solution of the DE will be tangent to these lines at all the points it passes through; the picture shows the specific solution with initial condition $y(0) = 1$.





This procedure works for every first-order DE of the form $y' = F(x, y)$; at each point (x, y) we put a short line segment with slope $F(x, y)$.

Example 16.6. The DE $y' = x^2 + y - 1$ has a direction field as in the second picture above. Observe that the points at which the direction field is horizontal can be found by solving $0 = y' = x^2 + y - 1$ to get $y = 1 - x^2$; this is the parabola in the picture. Below this parabola, solutions of the DE are decreasing; above it, solutions are increasing. The other curve in the picture is the solution with $y(0) = 1$.

Definition 16.7. A DE of the form $y' = F(x, y)$ is *autonomous* if the function F only depends on y, and not on x, so that it can actually be written as $y' = F(y)$. In the case when the independent variable is time, this can be thought of as "time-independence" of the system; the rule governing how y' is related to y does not change depending on t, but is the same for all time.

The two DEs above are not autonomous. The logistic DE $P' = kP(1 - \frac{P}{L})$ is autonomous. This has the consequence that its direction field looks the same if we shift it horizontally, and thus any solution curve remains a solution curve if we shift it left or right.

16.4. *Euler's method

As was the case when we computed definite integrals, there are situations in which it is better to take a numerical approach and try to find an approximate solution to an initial value problem (IVP). Such methods can become very sophisticated, but in this course we only consider the simplest one, called *Euler's method.*

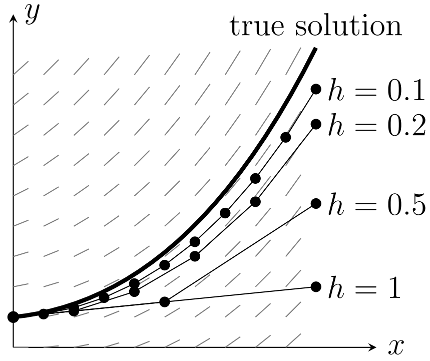
Roughly speaking, the idea of Euler's method is to move along the direction field in small steps of size h, where at each step we look at the direction field to see which way to move, then move that predetermined distance, and then look again at the direction field to get our instructions for the next step.

A little more precisely, the algorithm is this. Suppose we are given the IVP whose DE is $y' = F(x, y)$ and whose initial condition is $y(x_0) = y_0$. Fixing a step size h, we define (x_n, y_n) iteratively by

$$x_{n+1} = x_n + h, \qquad y_{n+1} = y_n + hF(x_n, y_n).$$

Thus the x-coordinate always increments by the step size, and the y-coordinate increments by the amount that it would change if F were constant and took the value that it does at (x_n, y_n).

The picture at right shows several applications of Euler's method to the initial value problem $y' = x + y$, $y(0) = 0.1$, with varying values of h. Observe that as h decreases it appears that the approximate solutions given by Euler's method are converging to the true solution. Whether this in fact occurs as $h \to 0^+$ is an important question in numerical analysis.



Lecture 17 *Separable differential equations

Stewart §9.3

17.1. Separable differential equations

Consider the first-order DE

$$\frac{dy}{dx} = \frac{x}{y}. \tag{17.1}$$

Based on our experience with the 'logarithm trick' for solving the DE $\frac{dy}{dx} = xy$ in Example 16.5, we might expect to get somewhere by multiplying both sides by y and writing $yy' = x$. In the previous example, the next step was to recognize that $\frac{y'}{y} = (\log y)'$. To proceed here, we need to replace $\log y$ with something that gives yy' upon differentiation by x.

After a little thought, you might realize that $\frac{d}{dx}(\frac{1}{2}y^2) = y\frac{dy}{dx}$, so (17.1) becomes $\frac{d}{dx}(\frac{1}{2}y^2)' = x$, or equivalently $\frac{d}{dx}(y^2) = 2x$, and integrating with respect to x gives $y^2 = x^2 + C$, so every solution of (17.1) has the form $y = \sqrt{x^2 + C}$ for some C. (Here we consider positive solutions; one could also consider negative solutions, but note that $y = 0$ is forbidden since y appears in the denominator of (17.1).)

To make this procedure into a more general strategy, let us replace the words "After a little thought" in the previous paragraph with the following more helpful argument: after writing (17.1) as $y\frac{dy}{dx} = x$, integrate both sides with respect to x to obtain

$$\int y\frac{dy}{dx}\,dx = \int x\,dx.$$

By the substitution rule, the integral on the left-hand side can be rewritten as $\int y\,dy$, and thus we get

$$\int y\,dy = \int x\,dx,$$

which upon evaluation gives the same solution as before.

The general strategy, then, is this: given a first-order DE $\frac{dy}{dx} = F(x, y)$, we say that the equation is *separable* if the RHS can be written as $F(x, y) = g(x)f(y)$, where g depends only on x and f depends only on y. Then we have

$$\frac{dy}{dx} = g(x)f(y) \quad\Rightarrow\quad \frac{1}{f(y)}\frac{dy}{dx} = g(x) \quad\Rightarrow\quad \int\frac{dy}{f(y)} = \int\frac{1}{f(y)}\frac{dy}{dx}\,dx = \int g(x)\,dx,$$

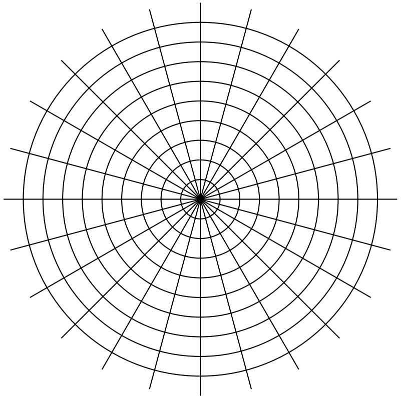
where the penultimate equality once again uses the substitution rule. Observe that our solutions of the DEs $y' = xy$ and $y' = x/y$ both used this strategy.

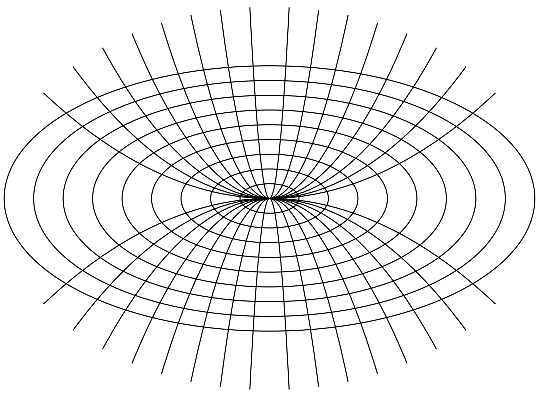
In general, evaluating the integrals is not quite the final step of the solution, because we still need to solve the resulting equation to find y in terms of x.

17.2. Orthogonal trajectories

Suppose we are given a family of curves, such as the set of lines through the origin in $\mathbb{R}^2$. It is occasionally of interest to find curves with the property that they intersect every curve in our original family at a right angle. For example, in an electrostatic field, the lines (curves) of force are always perpendicular to the lines (curves) of constant potential.

In the case of the set of lines through the origin, it is easy to see that the curves intersecting every such line orthogonally are just the circles centered at the origin, as shown in the picture at left. But what if we start with the family of parabolas whose vertex is at the origin, and which open up or down?





The family of parabolas just described comprises all the curves $y = kx^2$ where $k \in \mathbb{R}$. The picture at right suggests that the orthogonal trajectories for this family are ellipses. To confirm this, we first observe that if the point (x, y) lies on the parabola $y = kx^2$, then the slope of the parabola at this point is $2kx$. We can eliminate k by observing that $y = kx^2$ implies $k = y/x^2$, so the slope at this point is $2y/x$.

Now recall that two lines are perpendicular if and only if the product of their slopes is -1. Thus the slope of an orthogonal trajectory through (x, y) must be $-\frac{x}{2y}$ at this point. We conclude that a curve $x \mapsto y(x)$ describes an orthogonal trajectory if and only if it has the property that

$$\frac{dy}{dx} = -\frac{x}{2y}$$

everywhere. But this is a separable DE! So we can solve it by writing

$$2y\frac{dy}{dx} = -x \quad \Rightarrow \quad \int 2y\,dy = -\int x\,dx \quad \Rightarrow \quad y^2 = -\frac{1}{2}x^2 + C.$$

Thus the orthogonal trajectories to the family of parabolas are indeed the ellipses with equations $\frac{1}{2}x^2 + y^2 = C$.

17.3. Mixing problems

The following example gives another situation where a separable DE arises. Suppose mercury is leaking into a certain lake at a rate of γ g/min, and that water is flowing into the lake (from upstream) and out of the lake (downstream) at a rate of R L/min. (Since these rates are equal, the total volume of water in the lake remains constant.) Suppose also that at time $t = 0$, the lake is clean; there is no mercury in it. How much mercury is in the lake at time t?

Let V be the volume of the lake, which is constant. Let $y(t)$ be the mass of the mercury in the lake at time t, and let $\rho(t) = y(t)/V$ be the concentration. We make the simplifying assumption that mixing happens instantaneously, so that the concentration of mercury is the same throughout the lake. Then the rate at which mercury flows out of the lake is $\rho R = yR/V$ g/min, and since it flows in with rate γ g/min, we conclude that

$$\frac{dy}{dt} = \gamma - y\frac{R}{V}. \tag{17.2}$$

This is autonomous, and hence separable, so we can divide both sides by $\gamma - yR/V$ and then integrate, obtaining

$$\int \frac{dy}{\gamma - y\frac{R}{V}}\,dy = \int dt = t + C.$$

The integral on the LHS can be computed as follows:

$$\int \frac{dy}{\gamma - y\frac{R}{V}}\,dy = \frac{V}{R}\int \frac{dy}{\frac{\gamma V}{R} - y}\,dy = -\frac{V}{R}\ln\Big(\frac{\gamma V}{R} - y\Big),$$

and we conclude that

$$-\ln\Big(\frac{\gamma V}{R} - y\Big) = \frac{R}{V}t + C_1 \quad \Rightarrow \quad \frac{\gamma V}{R} - y = Ae^{-Rt/V},$$

so that the total amount of mercury in the lake at time t is given by

$$y(t) = \frac{\gamma V}{R} - Ae^{-Rt/V},$$

where A is a constant. To determine A we observe that at time $t = 0$ we have $0 = y = \frac{\gamma V}{R} - A$, so $A = \frac{\gamma V}{R}$, and we obtain

$$y(t) = \frac{\gamma V}{R}\left(1 - e^{-Rt/V}\right).$$

17.4. Solving the logistic model

The logistic DE $\frac{dP}{dt} = kP(1 - \frac{P}{L})$ from (16.2) is separable because the right-hand side does not depend on t, so we can divide both sides by $P(1 - \frac{P}{L})$ and then integrate:

$$\int \frac{dP}{P(1 - \frac{P}{L})} = \int k\,dt. \tag{17.3}$$

The integral on the RHS is easy. For the one on the left we use partial fractions to write

$$\int \frac{dP}{P(1 - \frac{P}{L})} = \int \frac{L}{P(L-P)}\,dP = \int \left(\frac{1}{P} + \frac{1}{L-P}\right) dP$$
$$= \ln P - \ln|L-P| = \ln\frac{P}{|L-P|},$$

and thus (17.3) gives

$$\ln\frac{P}{|L-P|} = kt + C.$$

Taking the exponential of both sides gives

$$\frac{P}{|L-P|} = e^C e^{kt}.$$

Let $Q = e^C$ if $L > P$ and $Q = -e^C$ if $L < P$; then $\frac{P}{L-P} = Qe^{kt}$, and we can solve for P:

$$P = LQe^{kt} - PQe^{kt} \quad\Rightarrow\quad P(1 + Qe^{kt}) = LQe^{kt} \quad\Rightarrow\quad P = \frac{LQe^{kt}}{1 + Qe^{kt}} = \frac{L}{1 + Q^{-1}e^{-kt}}.$$

This gives the general solution of the logistic DE. To find a particular solution given an initial population P_0 at time 0, we observe that $P_0 = P(0) = L/(1 + Q^{-1})$, so $1 + Q^{-1} = L/P_0$, and thus $Q^{-1} = \frac{L}{P_0} - 1$. Thus it is convenient to write the solution of the IVP as

$$P(t) = \frac{L}{1 + Ae^{-kt}} \quad \text{where } A = \frac{L}{P_0} - 1 = \frac{L - P_0}{P_0}.$$

Remark 17.1. Recall that the logistic DE is autonomous; the RHS does not depend on the independent variable. The example above illustrates the general principle that *every* autonomous DE is separable, because it can be written as $\frac{dy}{dx} = f(y)$, and thus can in principle be solved by writing $\frac{1}{f(y)}\frac{dy}{dx} = 1$ and integrating to get $\int \frac{1}{f(y)}\,dy = x + C$. There are then two obstacles to turning this into a complete solution:

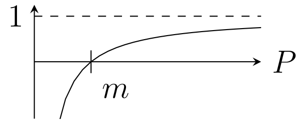
(1) the integral may be difficult or impossible to evaluate explicitly;
(2) it may be difficult or impossible to solve the resulting equation explicitly for y and write down a formula giving y in terms of x.

Lecture 18 *Other population models

Stewart §9.4

Beyond the logistic DE, there are other population models that are worth considering in certain situations. For example, suppose we expect that our population needs to be above a certain minimum size m to maintain itself, and that a population below this critical value will eventually die out. Then we might add another factor to the logistic DE that forces $\frac{dP}{dt}$ to be negative whenever $P < m$; we would like this factor to have the property that

- it is negative when $P < m$;
- it is positive when $P > m$;
- it is close to 1 for large values of P (when the population is well above the critical threshold, the original logistic DE should still be nearly accurate).



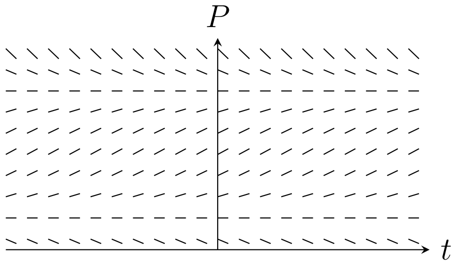
These suggest that its graph should have the general shape shown in the picture. An example of such a function is $(1 - \frac{m}{P})$, so we might multiply the RHS of the logistic DE by this factor and consider the DE

$$\frac{dP}{dt} = kP\Big(1 - \frac{P}{L}\Big)\Big(1 - \frac{m}{P}\Big). \tag{18.1}$$

We can rewrite the RHS as

$$\frac{dP}{dt} = \frac{k}{L}(L - P)(P - m).$$

To understand the behavior of this DE's slope field, we can draw its slope field.



This looks an awful lot like the slope field for the logistic DE:

(1) there are two equilibrium solutions, at $P = m$ and $P = L$;
(2) for $P \in (m, L)$, the population grows over time and appears to approach L;
(3) for $P > L$, the population decreases over time and appears to approach L.

The extra feature here is that there are positive values of P that are *below* the smaller equilibrium solution, and if the initial value of P lies in this range, then P decreases and eventually becomes 0, so the population goes extinct.

One could find an explicit solution of (18.1) by the same method as we used for the logistic DE, but we omit the details of this. Instead, we make the following observation: suppose that we write $y = P - m$ for the amount by which the population exceeds the critical threshold m. Then we have $P = y + m$ and can write

$$\frac{dy}{dt} = \frac{dP}{dt} = \frac{k}{L}(L - (y + m))y = \frac{k}{L}y(L - m - y) = \frac{k(L - m)}{L}y\Big(1 - \frac{y}{L - m}\Big).$$

But this means that y satisfies the original logistic DE! Granted, we need to change the parameters – the growth rate for y is $k(L-m)/L$ (instead of k) and the "carrying capacity" is $L-m$ (instead of L) – but this observation means that we can write any solution of (18.1) in terms of a solution for the logistic DE, and vice versa, so that in this sense the two problems are equivalent.

Remark 18.1. In fact, a similar change of variables (or substitution, if you prefer), can be used to turn any DE of the form $y' = ay^2 + by + c$ into the logistic DE, provided $b^2 - 4ac > 0$ so the DE has two equilibrium solutions.

So far, our population DEs have depended on parameters that affect the *quantitative* values of the solutions, but do not affect their *qualitative* form; that is, changing the parameters resulted in a new system that had the same number of equilibrium solutions, same overall description of types of solutions, etc. The next example is different.

Suppose $P(t)$ represents a population of fish that follows logistic growth but is also harvested at a constant rate c. Then the DE that it should satisfy is

$$\frac{dP}{dt} = kP\Big(1 - \frac{P}{L}\Big) - c. \tag{18.2}$$

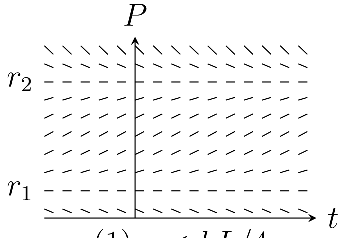
Again, we could solve this explicitly by dividing both sides by the quadratic on the RHS and then integrating, but instead of plunging blindly ahead with symbol manipulation, it is more instructive to take a moment and think about the overall picture. In particular, we want to understand for which values of P the RHS is positive, negative, and 0. Rewrite the DE as

$$\frac{dP}{dt} = -\frac{k}{L}P^2 + kP - c.$$

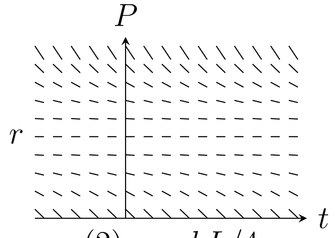
Observe that the RHS is a quadratic with discriminant[12] given by

$$k^2 - 4(-k/L)c = k^2 + 4ck/L = k(k + 4c/L).$$

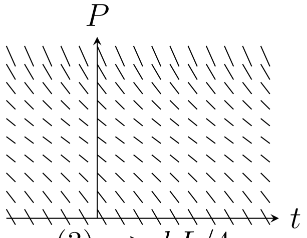
Since $k > 0$, we see that the sign of the discriminant is the same as the sign of $k + 4c/L$. This is determined by how the harvesting rate c compares to $kL/4$. There are three cases, whose slope fields are shown in the pictures.



(1) $c < kL/4$



(2) $c = kL/4$



(3) $c > kL/4$

We describe these one by one.

(1) When $c < kL/4$, the discriminant $k(k + 4c/L)$ is positive and thus the quadratic $-\frac{k}{L}P^2 + kP - c$ has two real roots r_1 and r_2, which correspond to equilibrium solutions of (18.2). When P lies between these roots, the quadratic is positive so

[12] Recall that the *discriminant* of the quadratic polynomial $ax^2 + bx + c$ is $b^2 - 4ac$, which is the expression that appears under the square root in the quadratic formula for the roots of the polynomial. The polynomial has two real roots if the discriminant is positive, one if it is 0, and none if it is negative.

$\frac{dP}{dt} > 0$ and the population grows, converging to r_2. When $P > r_2$, the quadratic is negative and the population shrinks, again converging to r_2. When $P < r_1$, the population shrinks and eventually goes extinct.[13]

(2) When $c = kL/4$, the discriminant is 0 and thus the quadratic has exactly one real root r, so (18.2) has exactly one equilibrium solution $P = r$. When $P > r$ the quadratic is negative so the population shrinks, converging to r. When $P < r$ then quadratic is again negative and the population shrinks, then goes extinct.

(3) When $c > kL/4$, the discriminant is negative and the quadratic has no real roots. Thus no matter what value P takes, the quadratic is negative and the population shrinks, eventually going extinct.

The first case corresponds to a harvesting rate that is sustainable provided the initial population is between r_1 and r_2. The final case corresponds to an unsustainable harvesting rate that eventually wipes out the population. The second case is borderline and unstable; although $P = r$ is an equilibrium solution, any fluctuation below this population (due perhaps to some effects not included in the model) will eventually lead to extinction.

Remark 18.2. The phenomenon seen here, wherein the solutions of a DE change dramatically and exhibit qualitatively different behavior as a parameter (or family of parameters) is varied, is called a *bifurcation*, and we say that $kL/4$ is a *bifurcation value* for the parameter c. Such parameter values are extremely important in the study of differential equations and other models of real-world systems.

Lecture 19 *Linear differential equations

Stewart §9.5

19.1. Linear first-order DEs

Example 19.1. Consider the DE

$$xy' + y = 2x. \tag{19.1}$$

This is a first-order DE, but it is not written in a form where we can immediately determine if it is separable. To determine this, we need to solve for y' and get $y' = 2 - \frac{y}{x}$; since we cannot find functions $g(x)$ and $f(y)$ such that $g(x)f(y) = 2 - \frac{y}{x}$, this DE is not separable. So what do we do?

In the end, there is only one thing we know how to do: integrate. If we could integrate both sides of (19.1) with respect to x, then we might hope to once again end up with an equation that could be solved to determine y. To integrate the LHS, we can first use integration by parts with $u = x$, $v = y$ to write

$$\int \underbrace{x}_{u} \underbrace{y'\,dx}_{dv} = xy - \int y\,dx,$$

[13] Note that this looks just like the picture we gave for (18.1) above, and indeed, as suggested in Remark 18.1, the two DEs can be related by a change of variables.

and then obtain

$$\int (xy' + y)\,dx = \int xy'\,dx + \int y\,dx = xy, \tag{19.2}$$

so that

$$xy = \int 2x\,dx = x^2 + C,$$

and the solution of (19.1) is

$$y = x + \frac{C}{x}.$$

In retrospect it should not be surprising that the antiderivative in (19.2) has the form that it does; the LHS of (19.1) has two terms, one of which includes y' and the other of which includes y, so it is reasonable to expect that its antiderivative would have the form $R(x)y$ for some function R. Indeed, the product rule gives

$$(R(x)y)' = R(x)y' + R'(x)y,$$

and we see that (19.2) works because $R(x) = x$ has $R'(x) = 1$. Thus it would be reasonable to use this approach anytime we have a DE where

- the LHS has the form $R(x)y' + R'(x)y$ for some function $R(x)$, and
- the RHS depends only on x (not on y).

Now that we have a hammer, let's go looking for some nails; are there many DEs like this?

Definition 19.2. A *linear first-order differential equation* is a DE that can be written in the form

$$f(x)\frac{dy}{dx} + g(x)y = h(x) \tag{19.3}$$

for some functions f, g, h.

Taking $f(x) = x$, $g(x) = 1$, and $h(x) = 2x$ gives (19.1).

Remark 19.3. A linear first-order DE can always be rewritten in the form

$$\frac{dy}{dx} + P(x)y = Q(x) \tag{19.4}$$

by dividing both sides of (19.3) by $f(x)$ and writing $P(x) = g(x)/f(x)$ and $Q(x) = h(x)/f(x)$.

Example 19.4. The DE

$$x^2y' + 2xy = 1$$

is a linear first-order DE, for which we can use the approach described above: we want a function $R(x)$ for which the LHS is $(R(x)y)'$, and since $R(x) = x^2$ has $R'(x) = 2x$, we see that indeed we can rewrite the DE as

$$\frac{d}{dx}(x^2y) = 1,$$

and integrating gives

$$x^2y = x + C,$$

so that the solution is $y = \frac{1}{x} + \frac{C}{x^2}$.

In both of the examples we have done so far, the solution was to let $R(x)$ be the function in front of y'; however, this only worked because we got lucky (and because the examples were engineered to work out nicely). Indeed, in order for the linear first-order DE

$$f(x)y' + g(x)y = h(x)$$

to have a LHS that can be written as $(R(x)y)'$, we must have $R(x) = f(x)$ and $R'(x) = g(x)$; in other words, we must have $f'(x) = g(x)$. If the DE we are given does not have this property, then we need to do a little more work.

19.2. General solution to first-order linear DEs

Example 19.5. The DE

$$y' = x + y \tag{19.5}$$

appeared in §16.3, when we introduced direction fields to sketch the general shape of its solutions because we did not yet have the tools to solve it exactly. It is a linear first-order differential equation since we can rewrite it as

$$y' - y = x. \tag{19.6}$$

However, the LHS of this last equation cannot be written as $(R(x)y)'$, because we have $f(x) = 1$ and $g(x) = -1$, so $f'(x) \neq g(x)$. So what are we to do?

The solution is to observe that we can multiply the entire DE (19.6) by an *integrating factor* $I(x)$, which if we choose it correctly, will make the previous trick work out. So we rewrite (19.6) as

$$I(x)y' - I(x)y = I(x)x. \tag{19.7}$$

This is again a linear first-order DE with $f(x) = I(x)$, $g(x) = -I(x)$, and $h(x) = xI(x)$. We want to choose $I(x)$ so that $f'(x) = g(x)$; in other words, we need $I'(x) = -I(x)$. This is again a DE, but it is one we know how to solve! We can put $I(x) = e^{-x}$, and then (19.7) becomes

$$e^{-x}y' - e^{-x}y = xe^{-x}.$$

The LHS has antiderivative $e^{-x}y$, so we can integrate both sides with respect to x and get

$$e^{-x}y = \int xe^{-x}\,dx = -xe^{-x} + \int e^{-x}\,dx = -(x+1)e^{-x} + C.$$

Multiplying both sides by e^x gives the general solution

$$y = Ce^x - (x+1).$$

This technique works for any linear first-order DE as in (19.3). It is easiest if we first divide through by $f(x)$ to write the DE in the form (19.4), and then multiply through by a (not yet determined) integrating factor to obtain

$$I(x)y' + P(x)I(x)y = Q(x)I(x). \tag{19.8}$$

We want the LHS to be equal to $(I(x)y)'$, which is true if and only if I satisfies the DE

$$I'(x) = P(x)I(x).$$

We can solve this DE by dividing by $I(x)$ and then using logarithms:

$$\frac{I'}{I} = P \quad \Rightarrow \quad \ln I(x) = \int P(x)\,dx \quad \Rightarrow \quad I(x) = e^{\int P(x)\,dx}.$$

Then (19.8) gives

$$(Iy)' = Iy' + PIy = QI \quad \Rightarrow \quad Iy = \int QI\,dx \quad \Rightarrow \quad y = \frac{1}{I}\int QI\,dx.$$

This is a general procedure for solving linear first-order DEs. Observe that the process involves two indefinite integrals: one to find $\ln I$, and a second to find y. In the first of these, we can take the constant of integration to be any value we like; it is enough to take $\ln I$ to be *any* antiderivative of P. In the second integral, on the other hand, we need to include the constant of integration, because it is ultimately determined by the initial condition of the DE.

Example 19.6. Consider the DE

$$y' + 3x^2y = 6x^2.$$

Multiplying through by an unknown integrating factor I gives

$$Iy' + 3x^2Iy = 6x^2I.$$

We want to choose I such that $I' = 3x^2I$, so

$$\log I = \int 3x^2\,dx = x^3 \quad \Rightarrow \quad I = e^{x^3}.$$

Thus the second form of the DE gives

$$(e^{x^3}y)' = e^{x^3}y' + 3x^2e^{x^3}y = 6x^2e^{x^3},$$

and we conclude that

$$e^{x^3}y = \int 6x^2e^{x^3}\,dx = 2e^{x^3} + C.$$

Thus the solution of the DE is

$$y = e^{-x^3}(2e^{x^3} + C) = 2 + Ce^{-x^3}.$$

19.3. Another solution of the logistic DE

We already solved the logistic DE $P' = kP(1-P/L)$ in §17.4, but just for fun let's do it again, via a different approach. Let $P(t)$ be a solution of the logistic DE, and define a new function $y(t)$ by $y = 1/P$. Then we have

$$y' = -\frac{P'}{P^2} = -\frac{kP - \frac{k}{L}P^2}{P^2} = -\frac{k}{P} + \frac{k}{L} = -ky + \frac{k}{L},$$

so $y(t)$ is a solution of the first-order linear DE

$$y' + ky = \frac{k}{L}.$$

This can be solved by the method introduced in this lecture; multiplying by an integrating factor I gives

$$Iy' + Iky = \frac{k}{L}I, \tag{19.9}$$

and we want I to satisfy $I' = Ik$, so we choose $I(t) = e^{kt}$. Then the left-hand side of (19.9) is $\frac{d}{dt}(e^{kt}y)$, and integrating both sides of (19.9) gives

$$e^{kt}y = \int \frac{k}{L}e^{kt}\,dt = \frac{1}{L}e^{kt} + C.$$

Multiplying through by e^{-kt} gives

$$y = \frac{1}{L} + Ce^{-kt},$$

and since $y = 1/P$ we see that the solution to the logistic DE is given by

$$P(t) = \frac{1}{y(t)} = \frac{1}{\frac{1}{L} + Ce^{-kt}} = \frac{L}{1 + CLe^{-kt}},$$

which agrees with the solution in §17.4 (by putting $Q = (CL)^{-1}$).

Lecture 20 Coupled differential equations

Stewart §9.6

20.1. Predator-prey models

Before we leave our discussion of differential equations, we consider two more examples, starting with a population model. This time instead of considering a single population, we consider two populations that interact with each other as predator and prey.

For concreteness, let $R(t)$ represent the population of rabbits in a given area, and $W(t)$ the population of wolves. We suppose that if there were no wolves, then the rabbits would reproduce according to the simple population growth DE $\frac{dR}{dt} = kR$, where $k > 0$. On the other hand, if there were no rabbits, then the wolves would have no food source and their population would decay following the DE $\frac{dW}{dt} = -rW$, where again $r > 0$.

Each of these DEs is easy to solve on its own. Things get interesting (and harder!) when we consider the interaction between the two populations. If $R > 0$ and $W > 0$, then some of the rabbits will be eaten by wolves, which decreases $\frac{dR}{dt}$ and increases $\frac{dW}{dt}$. A reasonable assumption is that the contribution to the derivatives due to predation is proportional to RW, since this number represents the number of possible rabbit-wolf pairs. Thus we arrive at the *Lotka–Volterra equations*

$$\frac{dR}{dt} = kR - aRW, \qquad \frac{dW}{dt} = -rW + bRW, \tag{20.1}$$

where $a, b, k, r > 0$ are parameters determined by the physical characteristics of the populations, their environment, and their interactions.

Remark 20.1. The DE (20.1) is not a single DE, but rather two DEs coupled together. This kind of situation arises very often in real-world models, and has the potential to increase the complexity of the situation tremendously. In particular, we should not expect to be able to write down explicit formulas for the solutions to such systems.

It turns out that autonomous systems of two DEs can be more or less completely understood at a qualitative level, similarly to our qualitative analysis of the various population models in §18, and the range of possible behaviors are very limited. However, with three or more DEs, there is the possibility of *chaotic* behavior, which has the appearance (in a way that can be made precise) of being nearly entirely random over long time scales, despite the fact that it is governed by deterministic equations.

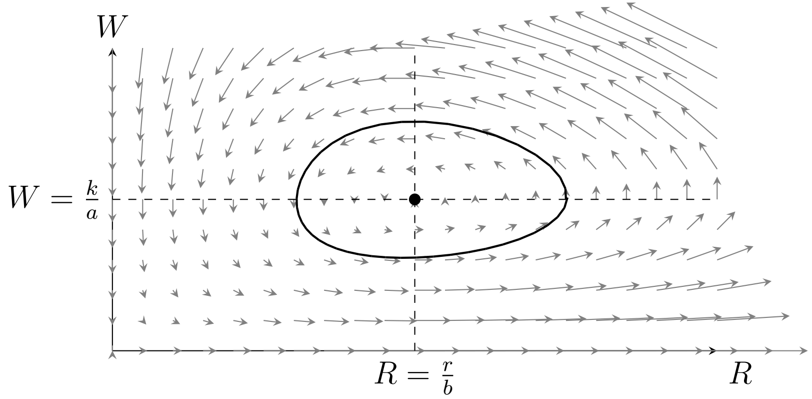
To understand the qualitative behavior of the Lotka–Volterra model, it is useful to find equilibrium solution(s), and more generally to find in which regions R and W are decreasing and increasing; we should also consider any special cases where the situation simplifies.

To find any equilibria, we see that

$$\frac{dR}{dt} = 0 \quad \Leftrightarrow \quad kR = aRW \quad \Leftrightarrow \quad R = 0 \text{ or } W = \frac{k}{a},$$
$$\frac{dW}{dt} = 0 \quad \Leftrightarrow \quad rW = bRW \quad \Leftrightarrow \quad W = 0 \text{ or } R = \frac{r}{b}.$$

Thus there are exactly two equilibrium solutions: the trivial solution where $R = W = 0$ (no rabbits, no wolves), and a nontrivial solution $W = k/a$, $R = r/b$.

To proceed further we draw an analogue of the direction field. The difference is that this time the line segment we draw at the point (R, W) has direction given by $(\frac{dR}{dt}, \frac{dW}{dt})$, and since this can be pointed in any direction (not just into the first or fourth quadrants, as was the case for our earlier direction fields) we will put arrowheads at the end of each of the line segments. The picture we obtain is called a *vector field*.



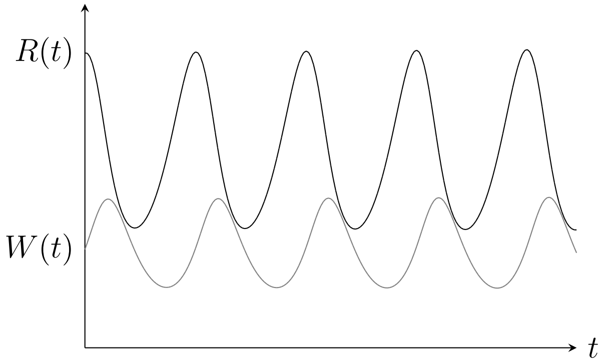
We refer to the region $\{(R, W) : R \geq 0, W \geq 0\}$ as the *phase space* of the system; each pair (R, W) represents a state that the system can be in. The horizontal line $W = k/a$ and the vertical line $R = r/b$ partition phase space into four regions. Observe that

- if $W < \frac{k}{a}$ then $aW < k$, so $\frac{dR}{dt} = R(k - aW) > 0$ (rabbit population increases when wolf population is small);
- if $W > \frac{k}{a}$ then $\frac{dR}{dt} < 0$ (rabbit population decreases when wolf population is large);
- if $R < \frac{r}{b}$ then $bR < r$, so $\frac{dW}{dt} = W(bR - r) < 0$ (wolf population decreases when rabbit population is small);
- if $R > \frac{r}{b}$ then $\frac{dW}{dt} > 0$ (wolf population increases when rabbit population is large).

Thus in the lower left region, where W and R are both below the thresholds, we see that W is decreasing and R is increasing, and all the arrows point down and to the right. In the lower right region, they point up and to the right. In the upper right region, they point up and to the left, and in the upper left region they point down and to the left.

As time progresses, the point $(R(t), W(t))$ moves in a counterclockwise direction around the equilibrium solution $(\frac{r}{b}, \frac{k}{a})$. The picture shows a typical solution curve, computed numerically. The numerical computations suggest that the curve returns to its starting point and then repeats periodically. Is this actually what happens? After all, *a priori* it would be equally reasonable for the curve to spiral in towards the equilibrium point, or to spiral away from it. In fact, the curve is indeed closed, as the numerical evidence suggests, but we will set this question aside and move to other things.[14]

We make one final observation: if we graph the rabbit and wolf populations and superimpose the pictures, then the oscillatory behavior shown above leads to a picture reminiscent of sine and cosine: two oscillating functions whose phases are offset, with the peaks of one lagging behind the peaks of the other.



20.2. Systems with more than two variables

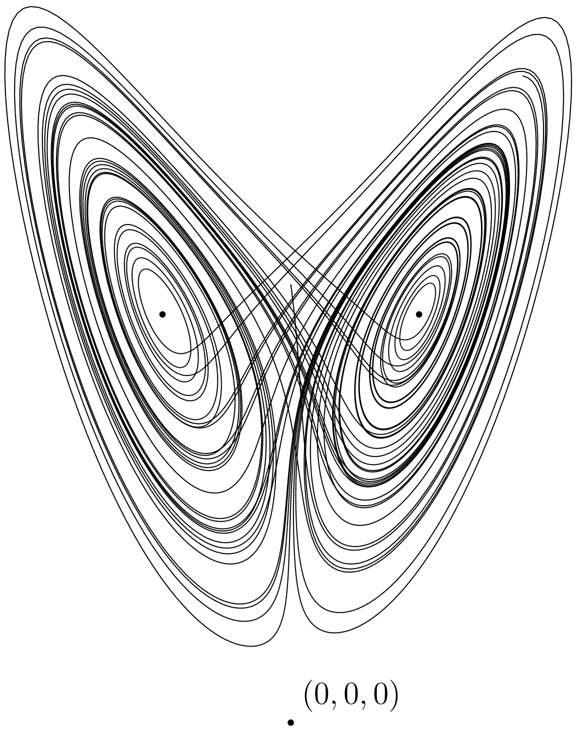
It turns out that for a system of two coupled autonomous differential equations, the only possible behaviors (from a qualitative point of view) are the ones we have encountered already; solutions can approach a fixed equilibrium solution as in the logistic DE, or diverge to infinity, or approach a periodic solution that oscillates endlessly and repeats itself exactly as in the Lotka–Volterra model. The precise theorem that describes all the possible behaviors is called the *Poincaré–Bendixson theorem*, and its details lie beyond the scope of this course; the basic idea is that solution curves cannot "get past" each other because everything lies in a two-dimensional space.

When we have more than two coupled DEs, on the other hand, life changes dramatically. In 1963, the meteorologist Edward Lorenz studied the following set of three coupled DEs as part of a simplified model of atmospheric convection:

$$\begin{aligned} \frac{dx}{dt} &= \sigma(y - x), \\ \frac{dy}{dt} &= x(\rho - z) - y, \\ \frac{dz}{dt} &= xy - \beta z. \end{aligned} \tag{20.2}$$

[14] The idea is to find a function $H(R, W)$ that depends on the size of both populations and that has the property that it does not change over time, so that the solution curve lies on a *level set* $\{(R, W) : H(R, W) = H_0\}$ for some value of H_0. It turns out that $H(R, W) = aW + bR - k \ln W - r \ln R$ does the job.

Here x, y, z are three functions of t whose physical interpretations we omit, and σ, ρ, β are three real-valued parameters reflecting certain physical properties of the system being studied; Lorenz used the values $\sigma = 10$, $\beta = 8/3$, and $\rho = 28$.



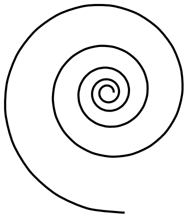
It is not so difficult to find the equilibrium solutions here: if $\frac{dx}{dt} = \frac{dy}{dt} = \frac{dz}{dt} = 0$, then the first equation in (20.2) gives $y = x$, and the second becomes $x(\rho - z - 1) = 0$, so either $x = y = 0$ or $z = \rho - 1$. If $x = y = 0$ then the third equation gives $z = 0$, so one equilibrium solution is $x = y = z = 0$. If $z = \rho - 1$ then the third equation gives $x = y = \pm\sqrt{\beta z} = \pm\sqrt{\beta(\rho - 1)}$, so there are two other equilibrium solutions. These three equilibria are shown in the picture at right, which also draws a single (numerically computed) solution of the system for some randomly chosen non-equilibrium initial condition. Observe that this solution does not appear to have any of the long-term behaviors described above: it does not approach an equilibrium solution, nor does it escape off to infinity, nor does it approach a periodic solution. Rather, it seems to spiral around one of the two nonzero equilibria for some time, then switches to spiral around the other, and so on in some manner that does not follow any readily discernible pattern.

The butterfly-like object shown in the picture is sometimes called a *strange attractor* and is emblematic of the field of *chaos theory*; the Lorenz equations display the phenomenon of *sensitive dependence on initial conditions*, which is a mechanism by which systems that follow deterministic rules can still exhibit behavior that appears random. If you search online for animations of the Lorenz attractor, you should have no trouble finding videos showing how solution curves that start very close to each other can follow each other for a while and then very quickly diverge so that their behavior is quite different. This means that if you only know the *approximate* state of a system to start off with (which reflects the reality that any measurement we make includes some error), then as time progresses you lose information about what state the system is in, which can be interpreted as a 'growth of randomness'. This can be made more precise by studying entropy, decay of correlations, and other topics in the field of *dynamical systems*, but these lie well beyond the scope of this course.

Part IV. Parametric curves and polar coordinates

Lecture 21 Parametric curves

Stewart §10.1, Spivak Chapter 12 appendix



Suppose we want to write an equation that describes the curve shown at right. Our usual approach to describing a curve by an equation is to write y as a function of x, or in some cases, x as a function of y. However, neither of these is an option here, since the curve fails both the vertical line test (so it cannot be written as the graph of $y = f(x)$) and the horizontal line test (so it cannot be written as the graph of $x = g(y)$).

In such situations, we can describe the curve by writing both x and y as functions of a new independent variable, instead of writing one as a function of the other. Thus we introduce a new variable t, called a *parameter*, and write $x = g(t)$, $y = f(t)$.

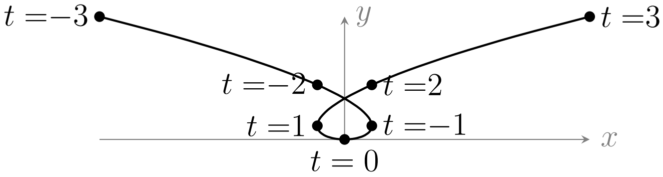
We have actually seen this situation several times already. It first appeared when we solved the catenary problem; although we ended up writing y as a function of x, an important intermediate step was to write both x and y as a function of arclength s, and also as a function of another parameter t. We also saw parametric curves appear in the last lecture on predator-prey models, where a solution of the system of differential equations was given by a curve written in terms of the parameter t, which represented time.

Example 21.1. The description of points on the unit circle as $x = \cos\theta$, $y = \sin\theta$ describes the circle as a parametric curve, where θ is the parameter.

As when graphing curves of the form $y = f(x)$, a useful approach to graphing a parametric curve is to make a table of values of t together with the corresponding values of x and y. For example, the curve shown below has the parametrization

$$x = t^3 - 3t, \qquad y = t^2, \tag{21.1}$$

and the table at right shows the values of x and y for integer values of t between -3 and 3; the corresponding points are marked on the curve.



t	-3	-2	-1	0	1	2	3
x	-18	-2	2	0	-2	2	18
y	9	4	1	0	1	4	9

Observe that one needs to be careful to connect the dots in the right order; based on the positions one might be tempted to connect the dot for $t = -2$ to the dot for $t = 1$, but this would give a very different shape to the curve.

The part of the curve shown in the picture above corresponds to parameter values lying in the interval $[-3, 3]$. When we restrict a curve to parameters $a \leq t \leq b$, the point $(x(a), y(a))$ is called the *initial point* of the curve, and $(x(b), y(b))$ is called the *terminal point.*

Remark 21.2. One should be careful to distinguish between a *curve*, which is a subset of $\mathbb{R}^2$, and a *parametric curve*, which is a curve equipped with a particular parametrization. The same curve can be equipped with many different parametrizations. For example, $x = t^2$, $y = t$, $-1 \leq t \leq 1$ describes an arc of a parabola opening to the right with vertex at the origin. This same curve is also described by $y = \cos t$, $x = \cos^2 t$. The difference between two parametrizations is analogous to the difference between two cars following the same road but with different (and varying) speeds.

As this example illustrates, any curve that can be described as the graph of a function (y in terms of x, or x in terms of y) can also be given as a parametric curve. The graph of $y = f(x)$ admits a parametrization $x = t$, $y = f(t)$, and the graph of $x = g(y)$ admits a parametrization $x = g(t)$, $y = t$. Thus our new technique is a more general one.

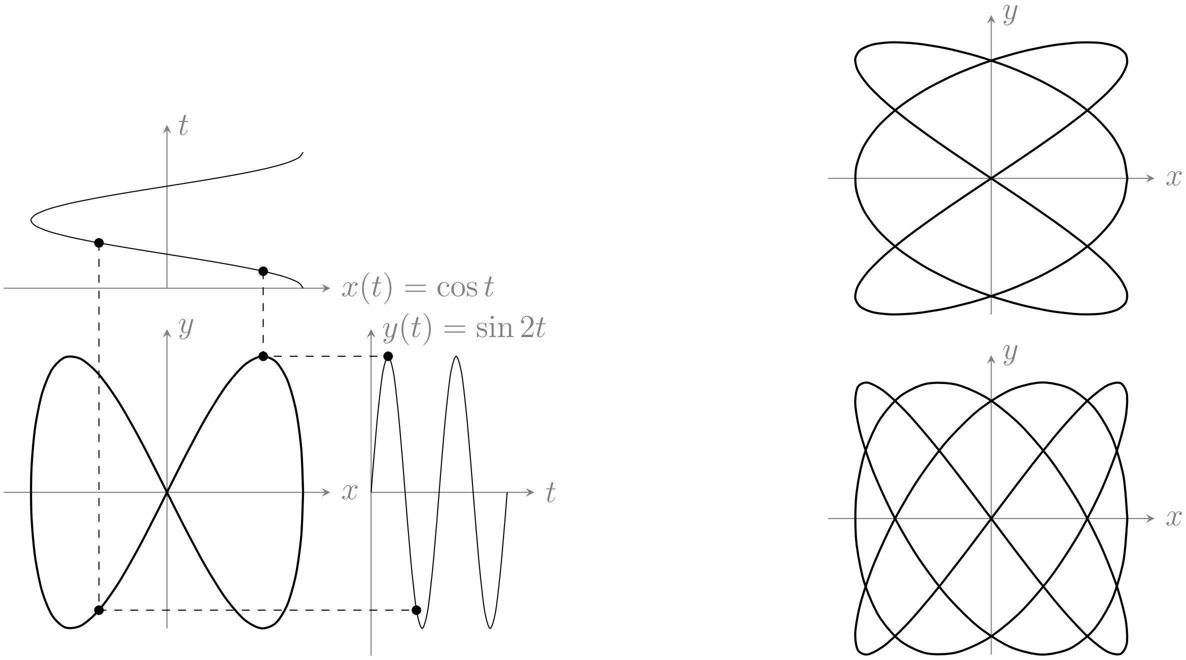
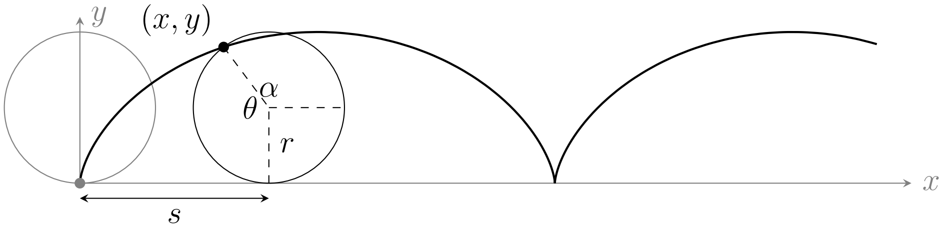


FIGURE 1. Lissajous figures

Example 21.3. The parametric curve $x = \cos t$, $y = \sin 2t$ has the appearance shown in the 'figure-eight' picture at left in Figure 1. Above and to the right of this curve are drawn the graphs of x and y with respect to the parameter t for $0 \leq t \leq 2\pi$, which covers the entire curve by periodicity. (Actually the t-axis in both graphs is compressed to save space.) Note that in the graph of $x(t)$, we plot t along the vertical axis and x along the horizontal axis.

This is an example of a *Lissajous figure*, a family of curves given by parametrizations $x = \cos(at)$, $y = \sin(bt)$ where $a, b \in \mathbb{N}$. (One can also add a phase shift by replacing at with $at + c$.) The pictures at right in Figure 1 show Lissajous figures for $a = 3$, $b = 2$ (top) and $a = 3$, $b = 4$ (bottom).

Example 21.4. Here is a physical example that is easier to describe via a parametric curve. Consider a circle rolling along the ground; the curve traced out by a point on the circle is called a *cycloid*.



For simplicity, assume that when we start, the point we are interested in is on the ground, and take this as the origin. Now start rolling the circle to the right, and let (x, y) be the location of the point we marked. To write a parametric equation for the cycloid, we let r be the radius of the circle, and write θ for the angle through which it has rotated so far. Then the total horizontal distance s that the circle has rolled is equal to the arc length from the bottom of the circle to (x, y), so we have $s = r\theta$, and we see that the center of the circle is currently at $(r\theta, r)$. The displacement of (x, y) from the center can be given in terms of θ:

(1) if α denotes the angle from the positive horizontal to (x, y) as we move around the circle (see the picture), then $x = r\theta + r\cos\alpha$ and $y = r + r\sin\alpha$;
(2) $\alpha + \theta = \frac{3\pi}{2}$, so $\cos\alpha = \cos(\frac{3\pi}{2} - \theta) = \cos\frac{3\pi}{2}\cos\theta + \sin\frac{3\pi}{2}\sin\theta = -\sin\theta$, and $\sin\alpha = \sin(\frac{3\pi}{2} - \theta) = \sin\frac{3\pi}{2}\cos\theta - \cos\frac{3\pi}{2}\sin\theta = -\cos\theta$.

We obtain the following parametrization for the cycloid in terms of θ:

$$x = r\theta - r\sin\theta = r(\theta - \sin\theta), \qquad y = r - r\cos\theta = r(1 - \cos\theta). \tag{21.2}$$

Remark 21.5. It turns out that the cycloid arises in the solution of two questions of historical interest. One of these is the *brachistochrone* problem, which asks to find the curve connecting two points A and B in the plane along which an object will slide from A to B the fastest under the influence of gravity, without friction. It turns out that the answer is not a straight line, as one might initially expect; rather, it is an (upside-down) cycloid.

The *tautochrone* problem asks for a curve with the property that the time it takes an object to slide down to the lowest point of the curve is independent of the initial height. It turns out that once again, the solution is an inverted cycloid. The proofs of these facts, however, require tools from the *calculus of variations*, which is well beyond the scope of this course.

Lecture 22 Calculus with parametrizations

Stewart §10.2, Spivak Chapter 12 appendix

22.1. Slopes

Suppose we want to find the slope of a parametric curve $(x, y) = (f(t), g(t))$ at a given point. We can only do this if near this point, there is a differentiable function F such that $y = F(x)$ describes the curve. When is this possible? The following diagram is useful.

$$\begin{tikzcd}
& t \arrow[dl, mapsto, "f"] \arrow[dr, mapsto, "g"] & \\
x \arrow[ur, mapsto, bend left, "f^{-1}"] \arrow[rr, mapsto, "F=g\circ f^{-1}"'] & & y
\end{tikzcd}$$

The parametrization lets us write y as a function of t, so we can write y as a function of x whenever we can write t as a function of x. This happens whenever the function f is invertible near the value of t that we are interested in; moreover, the chain rule gives the slope $F'(x)$ as $(f^{-1})'(x)g'(t)$.

Recall from last semester that the inverse function f^{-1} is defined and differentiable near $x = f(t)$ as long as $f'(t) \neq 0$, and that in this case we have $(f^{-1})'(f(t)) = 1/f'(t)$. Thus we have proved the following.

Proposition 22.1. *If $(x, y) = (f(t), g(t))$ is a parametric curve, where f, g are differentiable, and $t_0 \in \mathbb{R}$ is such that $f'(t_0) \neq 0$, then near t_0 we can write $y = F(x)$ for some differentiable function F, and the slope of the curve at $(f(t_0), g(t_0))$ is given by*

$$\frac{dy}{dx}\Big|_{x=f(t_0)} = F'(f(t_0)) = \frac{g'(t_0)}{f'(t_0)} = \frac{dy/dt|_{t=t_0}}{dx/dt|_{t=t_0}}. \tag{22.1}$$

If f and g are differentiable and $g'(t_0) = 0$ while $f'(t_0) \neq 0$, then at this point the curve has a horizontal tangent line; in other words, a horizontal tangent line occurs when $\frac{dy}{dt} = 0 \neq \frac{dx}{dt}$. A vertical tangent line occurs when $f'(t_0) = 0$ and $g'(t_0) \neq 0$; equivalently, when $\frac{dx}{dt} = 0 \neq \frac{dy}{dt}$.

When testing for horizontal or vertical tangent lines, the condition that the other derivative *not* vanish at t_0 is very important, as the following example shows.

Example 22.2. Let $x = t^3$ and $y = t^3$; then both $\frac{dx}{dt}$ and $\frac{dy}{dt}$ vanish when $t = 0$, but the tangent line is neither horizontal nor vertical here since the curve is just the graph of $y = x$.

22.2. Convexity

What about the second derivative? If we want to determine whether the curve is convex or concave, it is useful to compute $F''(x)$. To do this we use (22.1) to write $F'(x) = \frac{f'(t)}{g'(t)}$, and differentiating gives

$$F''(x) = \frac{d}{dx}F'(x) = \frac{d}{dx}\frac{g'(t)}{f'(t)} = \frac{dt}{dx}\frac{d}{dt}\frac{g'(t)}{f'(t)} = \frac{1}{dx/dt}\frac{d}{dt}\frac{g'(t)}{f'(t)} = \frac{1}{f'(t)}\frac{d}{dt}\frac{g'(t)}{f'(t)}, \tag{22.2}$$

where the third equality uses the chain rule, and the fourth uses the rule for derivatives of inverse functions. We can use the quotient rule to expand this as

$$F''(x) = \frac{f'(t)g''(t) - g'(t)f''(t)}{f'(t)^3},$$

but it is often easier to just work with the formula in (22.2), which can also be rewritten as

$$\frac{d^2y}{dx^2} = \frac{d}{dx}\frac{dy}{dx} = \frac{\frac{d}{dt}\frac{dy}{dx}}{\frac{dx}{dt}}. \tag{22.3}$$

Remark 22.3. Naive analogy with (22.1) might lead us to expect that the second derivative is given by $\frac{d^2y/dt^2}{d^2x/dt^2}$, since we may feel like we could "cancel the two appearances of dt^2", but we see from the above that this is not the case. This illustrates the dangers of treating higher derivatives as if they are fractions.

Example 22.4. Recall the parametric curve $x = t^3 - 3t$, $y = t^2$ from (21.1). This has vertical tangent lines when $0 = \frac{dx}{dt} = 3t^2 - 3$, so $t = \pm 1$; this corresponds to the points $(\mp 2, 1)$. Everywhere else we have $\frac{dx}{dt} \neq 0$ so we can use (22.1) to write

$$\frac{dy}{dx} = \frac{2t}{3t^2 - 3}.$$

We see that the only point with a horizontal tangent line occurs when $t = 0$, when the curve passes through the origin.

Note that the curve intersects itself where it crosses the y-axis; writing $x = 0$ gives $t = 0$ (at the origin) or $t^2 - 3 = 0$, so $t = \pm\sqrt{3}$, and both parameter values correspond to the point $(0, 3)$. Using $t = \sqrt{3}$, the slope is $2\sqrt{3}/6 = \sqrt{3}/3$; using $t = -\sqrt{3}$, the slope is $-\sqrt{3}/3$. These correspond to the tangent lines to the two different 'branches' of the curve passing through this point.

To determine concavity and convexity, we use (22.3) to write

$$\frac{d^2y}{dx^2} = \frac{\frac{d}{dt}\left(\frac{2t}{3t^2-3}\right)}{\frac{d}{dt}(t^3 - 3t)} = \frac{1}{3t^2 - 3} \cdot \frac{(3t^2 - 3) \cdot 2 - 2t(6t)}{(3t^2 - 3)^2} = \frac{-6t^2 - 6}{(3t^2 - 3)^3} = -\frac{2}{9}\left(\frac{t^2 + 1}{(t^2 - 1)^3}\right).$$

This never vanishes, so the graph has no inflection points. The second derivative is undefined at $t = \pm 1$, which makes sense because the first derivative is also undefined there. For $|t| < 1$ we have $\frac{d^2y}{dx^2} > 0$ and the graph is convex; for $|t| > 1$ we have $\frac{d^2y}{dx^2} < 0$ and the graph is concave.

Example 22.5. Consider the cycloid given by the parametrization $x = r(\theta - \sin\theta)$, $y = r(1 - \cos\theta)$, where $r > 0$ is the radius of the circle, and $\theta \in \mathbb{R}$ is the parameter. Then at a point (x, y) on the cycloid, the slope of the tangent line is given by

$$\frac{dy}{dx} = \frac{dy/d\theta}{dx/d\theta} = \frac{r\sin\theta}{r(1 - \cos\theta)} = \frac{\sin\theta}{1 - \cos\theta}.$$

It is tempting to immediately say "the tangent line is horizontal if and only if $\sin\theta = 0$". However, the full picture is a little more subtle, because when $\theta = 2n\pi$ for some $n \in \mathbb{Z}$, we have $\sin\theta = 0 = 1 - \cos\theta$, so both numerator and denominator vanish. This corresponds to the 'cusp' at the bottom of the cycloid, where the curve is not differentiable, although we can observe that $\lim_{\theta\to 2n\pi^\pm} \frac{\sin\theta}{1-\cos\theta} = \pm\infty$, which reflects the fact that the tangent line approaches vertical as (x, y) approaches a cusp.

The remaining values of θ for which $\sin\theta = 0$ are $\theta = (2n+1)\pi$ for some $n \in \mathbb{Z}$, and in this case we have $1 - \cos\theta = 2$, so the slope is indeed horizontal; this corresponds to the highest point on each loop of the cycloid.

The only values of θ for which the denominator vanishes are $\theta = 2n\pi$, when $\cos\theta = 1$, and as we saw above we have $\frac{dy}{d\theta} = \frac{dx}{d\theta} = 0$ at these points.

Lecture 23 Geometry of parametric curves

Stewart §10.2, Spivak Chapter 12 appendix

23.1. Area

Consider the parametric curve $(x, y) = (f(t), g(t))$, where $\alpha \le t \le \beta$ and f, g are differentiable. Suppose that $g \ge 0$ everywhere and that f is increasing, so that writing $a = f(\alpha)$ and $b = f(\beta)$, the function $f\colon [\alpha, \beta] \to [a, b]$ is invertible. Then $F = g \circ f^{-1}$ gives y as a function of x:

$$y = g(t) = g(f^{-1}(x)) = (g \circ f^{-1})(x) = F(x).$$

We know that the area under the curve $y = F(x)$, where $a \le x \le b$, is given by $A = \int_a^b F(x)\,dx$. Using the substitution rule to write this integral in terms of t, which is related to x by $x = f(t)$, we have

$$A = \int_a^b F(x)\,dx = \int_a^b y\,dx = \int_\alpha^\beta y\frac{dx}{dt}\,dt = \int_\alpha^\beta g(t)f'(t)\,dt. \tag{23.1}$$

Example 23.1. The area under one loop of the cycloid is

$$\begin{aligned} A &= \int_0^{2\pi r} y\,dx = \int_0^{2\pi} r(1-\cos\theta)\big(r(1-\cos\theta)\big)\,d\theta = r^2\int_0^{2\pi}(1 - 2\cos\theta + \cos^2\theta)\,d\theta \\ &= r^2\int_0^{2\pi}\Big(1 - 2\cos\theta + \frac{1}{2}(1+\cos 2\theta)\Big)\,d\theta = r^2\Big[\frac{3}{2}\theta - 2\sin\theta + \frac{1}{4}\sin 2\theta\Big]_0^{2\pi} \\ &= r^2\cdot\frac{3}{2}\cdot 2\pi = 3\pi r^2. \end{aligned}$$

23.2. Arc length

As above, consider the parametric curve $(x, y) = (f(t), g(t))$ on the interval $t \in [\alpha, \beta]$, where f, g are differentiable. If $f' > 0$ everywhere, then we can find the arc length of this curve by following the procedure above and writing it as $y = F(x)$ where $F = g \circ f^{-1}$; then the substitution rule gives the arc length as

$$L = \int_a^b \sqrt{1 + F'(x)^2}\,dx = \int_a^b \sqrt{1 + \Big(\frac{dy}{dx}\Big)^2}\,dx = \int_\alpha^\beta \sqrt{1 + \Big(\frac{dy/dt}{dx/dt}\Big)^2}\cdot\frac{dx}{dt}\,dt,$$

and simplifying gives

$$L = \int_\alpha^\beta \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\, dt. \tag{23.2}$$

The integrand in this final expression can be viewed as an infinitesimal version of the Pythagorean formula.

What if f' is not always positive? What if the curve fails to satisfy the vertical line test and cannot be written as $y = F(x)$? In this case we can still follow the procedure from Lecture 11.1 and consider polygonal approximations to the curve. Partitioning the interval $[\alpha, \beta]$ into n subintervals $[t_{i-1}, t_i]$ of equal length $\Delta t = (\beta - \alpha)/n$, where $t_i = \alpha + i\Delta t$ for $0 \leq i \leq n$, and writing $P_i = (f(t_i), g(t_i))$, we can once again declare the length of the curve to be given by (11.1), so that

$$L = \lim_{n\to\infty} \sum_{i=1}^n \text{distance}(P_{i-1}, P_i) = \lim_{n\to\infty} \sum_{i=1}^n \sqrt{(f(t_i) - f(t_{i-1}))^2 + (g(t_i) - g(t_{i-1}))^2}$$

Applying the mean value theorem to f and g on $[t_{i-1}, t_i]$ gives t_i^* and t_i^{**} in this subinterval such that

$$f(t_i) - f(t_{i-1}) = f'(t_i^*)(t_i - t_{i-1}) = f'(t_i^*)\Delta t \quad \text{and} \quad g(t_i) - g(t_{i-1}) = g'(t_i^{**})\Delta t.$$

Thus we can compute the arc length as

$$\begin{aligned} L &= \lim_{n\to\infty} \sum_{i=1}^n \sqrt{(f'(t_i^*)\Delta t)^2 + (g'(t_i^{**})\Delta t)^2} = \lim_{n\to\infty} \sum_{i=1}^n \sqrt{(f'(t_i^*))^2 + (g'(t_i^{**}))^2} \cdot \Delta t \\ &= \int_\alpha^\beta \sqrt{f'(t)^2 + g'(t)^2}\, dt, \end{aligned}$$

where as in Remark 12.2 we observe that the expression on the first line is not quite a Riemann sum because f' and g' are evaluated at different values of t, but the sum nevertheless converges to the integral. Observe that the formula we obtained here is the same as the formula in (23.2). Thus we have the following.

Definition 23.2. If a curve C admits a parametrization $(x, y) = (f(t), g(t))$, $\alpha \leq t \leq \beta$, where f, g are differentiable and the curve C is traversed exactly once as t ranges from α to β, then the arc length of C is given by

$$L = \int_\alpha^\beta \sqrt{f'(t)^2 + g'(t)^2}\, dt = \int_\alpha^\beta \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\, dt. \tag{23.3}$$

Exercise 23.3. Show that the arc length does not depend on the choice of parametrization; that is, show that if $(f_1(t), g_1(t))$ and $(f_2(t), g_2(t))$ are two parametrizations of the same curve, then they give the same value of L in (23.3).

Example 23.4. For the circle $x = \cos t$, $y = \sin t$, $t \in [0, 2\pi]$, we have $\frac{dx}{dt} = -\sin t$ and $\frac{dy}{dt} = \cos t$, so (23.3) gives the arc length

$$\int_0^{2\pi} \sqrt{\sin^2 t + \cos^2 t}\, dt = \int_0^{2\pi} 1\, dt = 2\pi,$$

as expected. If we reparametrize the circle as $x = \cos(t^2)$, $y = \sin(t^2)$, then $\frac{dx}{dt} = -2t\sin(t^2)$ and $\frac{dy}{dt} = 2t\cos(t^2)$, and we return to the starting point $(1, 0)$ when $t = \sqrt{2\pi}$, so (23.3) gives the arc length

$$\int_0^{\sqrt{2\pi}} \sqrt{4t^2 \sin^2 t^2 + 4t^2 \cos^2 t^2}\, dt = \int_0^{\sqrt{2\pi}} 2t\, dt = \Big[t^2\Big]_0^{\sqrt{2\pi}} = 2\pi,$$

which agrees with the earlier answer.

Example 23.5. The arc length of one loop of the cycloid, which has $\frac{dx}{d\theta} = r(1 - \cos\theta)$ and $\frac{dy}{d\theta} = r\sin\theta$, is given by

$$\begin{aligned} L &= \int_0^{2\pi} \sqrt{r^2(1-\cos\theta)^2 + r^2\sin^2\theta}\, d\theta = \int_0^{2\pi} r\sqrt{1 - 2\cos\theta + \cos^2\theta + \sin^2\theta}\, d\theta \\ &= r\int_0^{2\pi} \sqrt{2 - 2\cos\theta}\, d\theta = r\int_0^{2\pi} \sqrt{4\sin^2\frac{\theta}{2}}\, d\theta = r\int_0^{2\pi} 2\sin\frac{\theta}{2}\, d\theta \\ &= r\Big[-4\cos\frac{\theta}{2}\Big]_0^{2\pi} = r\big(-4(-1) - (-4)(1)\big) = 8r. \end{aligned}$$

23.3. Surface area

Consider a parametric curve $(x, y) = (f(t), g(t))$ on the interval $t \in [\alpha, \beta]$, where f, g are differentiable and $g > 0$; let S be the surface area of the surface of revolution obtained by rotating this curve around the x-axis. Then similar arguments to those in Lecture 12.1 show that

$$S = \int_\alpha^\beta 2\pi y\, ds = \int_\alpha^\beta 2\pi y\sqrt{\Big(\frac{dx}{dt}\Big)^2 + \Big(\frac{dy}{dt}\Big)^2}\, dt.$$

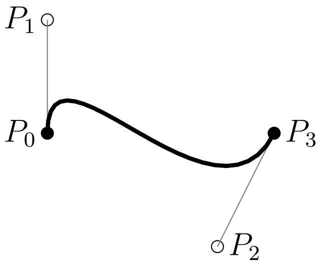
Example 23.6. The sphere of radius r is the surface of revolution for $x = r\cos t$, $y = r\sin t$, $t \in [0, \pi]$, so its surface area is

$$S = \int_0^\pi 2\pi r\sin t\sqrt{r^2\sin^2 t + r^2\cos^2 t}\, dt = \int_0^\pi 2\pi r^2 \sin t\, dt = \Big[-2\pi r^2\cos t\Big]_0^\pi = 4\pi r^2.$$

This is a little simpler than the computation we did in Example 12.6.

23.4. Bezier curves

One useful application of parametric curves is given by *Bézier curves*, which are widely used in graphics, design, animation, and other related fields. A *cubic Bézier curve* is given by four *control points* P_0, P_1, P_2, P_3, as shown in the picture. Intuitively, the curve starts at P_0 and ends at P_3, with P_1 used to determine the tangent direction at P_0, and P_2 to determine the tangent direction at P_3.



A nice animation illustrating how to construct these curves can be found online at `https://www.jasondavies.com/animated-bezier/`, and an interactive applet that lets you see how the curve responds to changes in the locations of the four control points can be found at `https://www.desmos.com/calculator/cahqdxeshd`.

Lecture 24 Polar coordinates

Stewart §10.3, Spivak Chapter 4 appendix 3

24.1. Rectangular and polar coordinates

We are accustomed to using a rectangular coordinate system[15] to describe points in the plane: the two real numbers x and y uniquely determine a point P in the plan as the point that you reach by starting at the origin, moving x units to the right, and moving y units up. Now we describe a new coordinate system, called *polar coordinates*.

Start by fixing a reference point, called the *pole* – usually we choose the origin. Fix an infinite ray starting at this point, called the *polar axis* – usually we choose the positive x-axis. Given real numbers r and θ, the *polar coordinates* (r, θ) describe a point P in the plane as follows:

(1) standing at the pole, face in the direction of the polar axis and then rotate θ radians counterclockwise;
(2) move a distance r in the direction you are now facing.

The point P is the point that you reach at the end of this procedure. To put it another way, the polar coordinates of P are the real numbers r and θ such that $r = |OP|$ is the distance from the origin to P, and θ is the angle from the positive x-axis to the line segment OP.

Remark 24.1. The numbers r and θ uniquely determine P, but (in sharp constrast to the situation with rectangular coordinates) the other direction requires some choice.

- When $r = 0$, *any* value of θ puts P at the origin.
- For $r > 0$, the angles θ and $\theta + 2\pi$ give the same point P. Thus θ is only determined up to multiples of 2π. We will often choose $\theta \in (-\pi, \pi]$, but one could just as easily choose $\theta \in [0, 2\pi)$, or any other half-open interval with length 2π.
- The second procedure described above, for obtaining r and θ from P, always returns a nonnegative value of r. However, the first procedure, for obtaining P from r and θ, makes sense even when r is negative, provided we interpret the second step for a negative value of r as meaning "move backwards by a distance of $|r|$". Then we see that the polar coordinates $(-r, \theta)$ and $(r, \theta + \pi)$ both correspond to the same point.

Exercise 24.2. Plot the points with polar coordinates $(1, \frac{\pi}{4})$, $(2, \frac{\pi}{2})$, $(3, -\frac{3\pi}{4})$, and $(4, \pi)$.

To compare rectangular and polar coordinates, observe that after rotating by an angle θ, we are standing at the origin and facing in the direction of the point on the unit circle with rectangular coordinates $(\cos\theta, \sin\theta)$. (Indeed, this is one definition of cos and sin.) Walking a distance r in this direction moves us to the point $(r\cos\theta, r\sin\theta)$. In other words, rectangular and polar coordinates are related by the equations

$$x = r\cos\theta, \qquad y = r\sin\theta. \tag{24.1}$$

[15] Also called *Cartesian* coordinates, after René Descartes.

These describe the first procedure above; converting polar coordinates to rectangular coordinates.

Example 24.3. The point with polar coordinates $(2, \frac{\pi}{3})$ has $r = 2$ and $\theta = \frac{\pi}{3}$, so its rectangular coordinates are

$$x = 2\cos\frac{\pi}{3} = 2 \cdot \frac{1}{2} = 1, \quad y = 2\sin\frac{\pi}{3} = 2 \cdot \frac{\sqrt{3}}{2} = \sqrt{3}.$$

Remark 24.4. The relationship (24.1) between polar coordinates and rectangular coordinates can be written in a single equation involving complex numbers. Recall that given a real number θ, the complex exponential function is $e^{i\theta} = \cos\theta + i\sin\theta$, and thus given $r \geq 0$ we have

$$re^{i\theta} = r\cos\theta + ir\sin\theta = x + iy,$$

where x, y are the real and imaginary parts, respectively, of the complex number $re^{i\theta}$. If $z = re^{i\theta}$, then the number r is called the *modulus* of z, and θ is called the *argument*. Observe that θ is only defined up to a multiple of 2π.

What about the other direction? If a point has rectangular coordinates (x, y), then squaring the two equations in (24.1) and adding them together gives

$$x^2 + y^2 = r^2\cos^2\theta + r^2\sin^2\theta = r^2.$$

If $x^2 + y^2 = 0$ then we must have $x = y = 0$, so the point is the origin and can be represented as $r = 0$, $\theta =$ any real number. If $x^2 + y^2 \neq 0$, then we can choose r to be the positive square root $r = \sqrt{x^2 + y^2}$ and convert (24.1) to

$$\cos\theta = \frac{x}{r}, \qquad \sin\theta = \frac{y}{r}.$$

Together these uniquely determine θ in the interval $(-\pi, \pi]$, or in any half-open interval of length 2π. Note that this restriction reflects the ambiguity mentioned in Remark 24.1 above: (r, θ) and $(r, \theta + 2\pi)$ represent the same point, because $\cos(\theta + 2\pi) = \cos\theta$ and $\sin(\theta + 2\pi) = \sin\theta$.

Dividing the two halves of (24.1) gives another useful formula,

$$\tan\theta = \frac{r\sin\theta}{r\cos\theta} = \frac{y}{x}.$$

Then θ is determined by any two of the three values $\cos\theta, \sin\theta, \tan\theta$.

It gets a little messy if we try to explicitly write down a formula for θ in terms of x, y, r using inverse trigonometric functions. One is tempted to simply write

$$\theta = \cos^{-1}\frac{x}{r};$$

however, in order to invert the cosine function, we must restrict it to an interval on which it is 1-1. The usual choice is $[0, \pi]$, but then we would always have $\sin\theta \geq 0$, and so we would need to choose $r < 0$ to represent points with $y < 0$. Thus we should actually follow a two-step procedure: first look at the sign of y to determine which branch of $\cos^{-1}$ to use, and then apply $\cos^{-1}$ to find θ. More precisely:

(1) if $y < 0$, restrict cos to $(-\pi, 0)$ and then invert, so that $\theta = \cos^{-1}\frac{x}{r} \in (-\pi, 0)$;
(2) if $y \geq 0$, restrict cos to $[0, \pi]$ and then invert, so that $\theta = \cos^{-1}\frac{x}{r} \in [0, \pi]$.

Another way of describing this is to take $\theta = \cos^{-1}\frac{x}{r} \in [0,\pi]$ and then check the sign of y: if $y \geq 0$, then leave θ as it is, and if $\theta < 0$, then replace θ by $-\theta$.

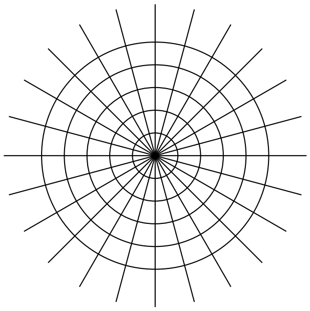
Exercise 24.5. Describe similar procedures for finding θ using $\sin^{-1}\frac{y}{r}$ and $\tan^{-1} yx$; in both cases the inverse trigonometric function yields a value in $[-\frac{\pi}{2},\frac{\pi}{2}]$, and then we must look at the sign of x to determine whether θ is given by this value or by a related one.

Example 24.6. If P has rectangular coordinates $(1,-1)$, then $x = 1$ and $y = -1$, so $r = \sqrt{x^2+y^2} = \sqrt{2}$, and thus

$$\cos^{-1}\frac{x}{r} = \cos^{-1}\frac{1}{\sqrt{2}} = \frac{\pi}{4}.$$

Since $y < 0$, the point P lies below the x-axis and we have $\theta = -\cos^{-1}\frac{x}{r} = -\frac{\pi}{4}$.

24.2. Curves in polar coordinates



We know three ways to describe a curve in rectangular coordinates:

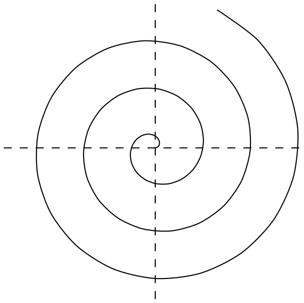
(1) *explicitly* as the graph of $y = f(x)$ or $x = g(y)$;
(2) *implicitly* as the solution set of $F(x,y) = 0$;
(3) *parametrically* as $x = f(t)$, $y = g(t)$.

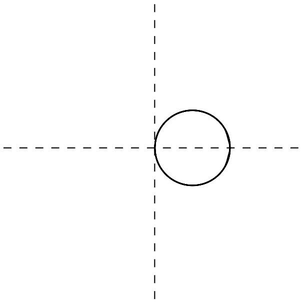
We can use each of these methods in polar coordinates as well. First consider the curves $r = c$ and $\theta = t$, where c, t are constants. These curves are shown in the picture at right. Given $c > 0$, the equation $r = c$ describes the circle centered at the origin with radius c. Indeed, since $r = \sqrt{x^2+y^2}$ this formula can be rewritten in rectangular coordinates as the (implicit) formula $x^2+y^2 = r^2 = c^2$. Given $t \in \mathbb{R}$, we either have $\cos t = 0$ (if t is an odd multiple of $\frac{\pi}{2}$), in which case $\sin t = \pm 1$ and the curve is the y-axis ($x = 0$, $y = \pm r$), or $\cos t \neq 0$ in which case $\frac{y}{x} = \tan t$, so the curve is the line $y = (\tan t)x$. This shows that the curves of constant r are concentric circles around the origin, while curves of constant θ are lines through the origin.

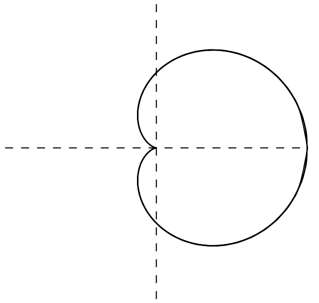
More generally, a curve of the form $r = f(\theta)$ can be written parametrically as

$$x = f(\theta)\cos\theta, \quad y = f(\theta)\sin\theta. \tag{24.2}$$

The three pictures below illustrate the curves $r = \theta$, $r = \cos\theta$, and $r = 1+\cos\theta$, which we discuss next.







Example 24.7. $r = \theta$ gives a spiral curve as shown in the left-hand picture; observe that increasing θ corresponds to moving around the origin in a counterclockwise direction,

and that each time we cross the next axis (having increased θ by $\frac{\pi}{2}$) the value of r has increased and we are further from the origin.

Example 24.8. With the curve $r = \cos\theta$, we see that r decreases from 1 to 0 as θ goes from 0 to $\frac{\pi}{2}$. Then when θ goes over the interval $[\frac{\pi}{2}, \pi]$, where we might expect the curve to lie in the second quadrant, we have $\cos\theta \leq 0$, so in fact the curve lies in the fourth quadrant. Moreover, when $\theta = \pi$ we have $r = -1$ and thus $x = 1$, $y = 0$, which is where the curve starts at $\theta = 0$; thus the entire curve is covered by the parameter range $\theta \in [0, \pi]$. In fact, the curve is the circle with center at $(\frac{1}{2}, 0)$ (in rectangular coordinates) and radius $\frac{1}{2}$; to see this, observe that $x = r\cos\theta = r^2 = x^2 + y^2$, so this is the curve defined in rectangular coordinates by the implicit equation

$$0 = x^2 - x + y^2 = \left(x - \frac{1}{2}\right)^2 + y^2 - \frac{1}{4}.$$

Example 24.9. With $r = 1 + \cos\theta$, we have $r \geq 0$ for all θ, so the curve passes through all four quadrants. The value of r decreases from 2 to 0 as θ ranges from 0 to π; this is the top half of the curve shown. The bottom half of the curve corresponds to $\theta \in [\pi, 2\pi]$, when r increases from 0 back to 2. This curve is called the *cardioid* because of its heart-like shape.

Lecture 25 Calculus with polar coordinates

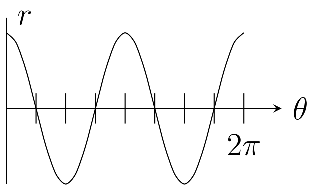
Stewart §10.4.

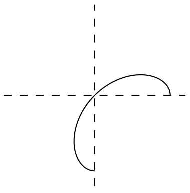
25.1. Slopes of tangent lines

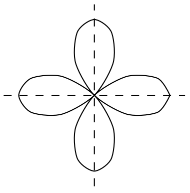
Consider a curve given in polar coordinates by the formula $r = f(\theta)$, where f is differentiable. Using the parametric representation of the curve in (24.2), we can compute the slope of the tangent line at any point by using Proposition 22.1 to write

$$\frac{dy}{dx} = \frac{dy/d\theta}{dx/d\theta} = \frac{f'(\theta)\sin\theta + f(\theta)\cos\theta}{f'(\theta)\cos\theta - f(\theta)\sin\theta} = \frac{\frac{dr}{d\theta}\sin\theta + r\cos\theta}{\frac{dr}{d\theta}\cos\theta - r\sin\theta}. \tag{25.1}$$

Example 25.1. Consider the curve with polar formula $r = \cos 2\theta$. The first picture below shows the graph of r as a function of θ where these are taken as *rectangular* coordinates; this is helpful in order to visualize how r decreases and increases as θ varies, which in turn lets us picture the graph. The second picture shows the graph of the curve on the interval $\theta \in [0, \frac{\pi}{2}]$. Observe how r decreases from 1 to 0 on $[0, \frac{\pi}{4}]$, and then to -1 on $[\frac{\pi}{4}, \frac{\pi}{2}]$, so that on this second interval the curve actually lies in the third quadrant. The third picture shows the complete curve, which consists of four copies of this first piece, each rotated by $\frac{\pi}{2}$ from the previous one.







Using (25.1), we see that the slope of the tangent line to the curve $r = \cos 2\theta$ is

$$\begin{aligned}\frac{dy}{dx} &= \frac{-2\sin 2\theta \sin\theta + \cos 2\theta \cos\theta}{-2\sin 2\theta\cos\theta - \cos 2\theta \sin\theta} = \frac{-4\sin^2\theta\cos\theta + (1-2\sin^2\theta)\cos\theta}{-4\sin\theta\cos^2\theta - (2\cos^2\theta - 1)\sin\theta} \\ &= \frac{\cos\theta(1-6\sin^2\theta)}{\sin\theta(1-6\cos^2\theta)}.\end{aligned}$$

Considering $\theta \in [0, 2\pi)$ to get one full circuit around the curve, we see that the numerator vanishes if and only if $\cos\theta = 0$ or $\sin^2\theta = \frac{1}{6}$. The first possibility occurs at the values $\theta = \frac{\pi}{2}$ and $\theta = \frac{3\pi}{2}$, while the second occurs for one value of θ in each quadrant. Writing $\theta_0 = \sin^{-1}\frac{1}{\sqrt{6}} \in (0, \frac{\pi}{2})$ for the value of θ in this interval at which $\sin^2\theta = \frac{1}{6}$, we see that the four values at which this occurs are $\theta = \theta_0, \pi - \theta_0, \pi + \theta_0, 2\pi - \theta_0$. This gives six points at which the numerator vanishes.

Similarly, the denominator vanishes if and only if $\sin\theta = 0$ or $\cos^2\theta = \frac{1}{6}$. The first possibility occurs at $\theta = 0$ and $\theta = \pi$, while the second occurs at one point in each quadrant. Writing $\theta_1 = \cos^{-1}\frac{1}{\sqrt{6}} \in (0, \frac{\pi}{2})$ for the value of θ in this interval with $\cos^2\theta = \frac{1}{6}$, we see that the four values at which this occurs are $\theta_1, \pi - \theta_1, \pi + \theta_1, 2\pi - \theta_1$. This gives six points at which the denominator vanishes; observe that this does not include any of the points at which the numerator vanishes.

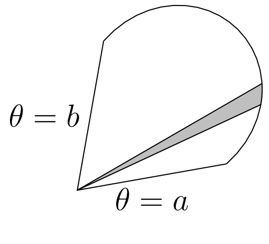
Since the denominator is nonzero everywhere that the numerator vanishes, we see that the tangent line is horizontal when $\theta = \frac{\pi}{2}, \frac{3\pi}{2}$, which corresponds to the points $(0, \pm 1)$, and when $\theta \in \{\theta_0, \pi - \theta_0, \pi + \theta_0, 2\pi - \theta_0\}$. At $\theta = \theta_0$ we have

$$\begin{aligned} x &= \cos 2\theta \cos\theta = (1 - 2\sin^2\theta)\cos\theta = \frac{2}{3}\sqrt{1 - \sin^2\theta_0} = \frac{2}{3}\sqrt{\frac{5}{6}}, \\ y &= \cos 2\theta \sin\theta = (1 - 2\sin^2\theta)\sin\theta = \frac{2}{3}\sqrt{\frac{1}{6}}.\end{aligned}$$

Thus the four points with horizontal tangent lines are $(\pm\frac{2}{3}\sqrt{\frac{5}{6}}, \pm\frac{2}{3}\sqrt{\frac{1}{6}})$.

Similarly, since the numerator is nonzero everywhere that the denominator vanishes, the tangent line is vertical when $\theta \in \{0, \pi, \theta_1, \pi - \theta_1, \pi + \theta_1, 2\pi - \theta_1\}$, and these six points have coordinates $(\pm 1, 0)$ and $(\pm\frac{2}{3}\sqrt{\frac{1}{6}}, \pm\frac{2}{3}\sqrt{\frac{5}{6}})$.

25.2. Area in polar coordinates



We know that in rectangular coordinates, the region bounded by the curves $x = a$, $x = b$, $y = 0$, and $y = f(x)$ has area $\int_a^b f(x)\,dx$. What about polar coordinates? What is the area of the region bounded by the curves $\theta = a$, $\theta = b$, and $r = f(\theta)$?

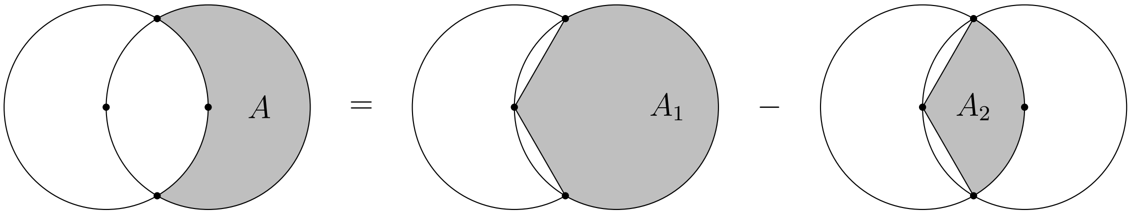
As usual, we fix a large $n \in \mathbb{N}$ and divide the parameter interval $[a, b]$ into n subintervals of equal length $\Delta\theta = (b - a)/n$, with endpoints $\theta_i = a + i\Delta\theta$. Then the ith interval $[\theta_{i-1}, \theta_i]$ determines a 'wedge' such as the one shown in the picture, whose area is $\approx \frac{1}{2}f(\theta_i^*)^2\Delta\theta$, where $\theta_i^* \in [\theta_{i-1}, \theta_i]$ and we remember that a sector of a circle with angle

θ and radius r has area $\frac{1}{2}\theta r^2$. Adding these areas together gives a Riemann sum, and taking a limit as $n \to \infty$ we see that the area of the region is given by

$$A = \lim_{n\to\infty} \sum_{i=1}^{n} \frac{1}{2} f(\theta_i^*)\Delta\theta = \int_a^b \frac{1}{2} f(\theta)^2 \, d\theta. \tag{25.2}$$

Example 25.2. The right-most leaf of the "clover" shape from Example 25.1 corresponds to the parameter interval $\theta \in [-\frac{\pi}{4}, \frac{\pi}{4}]$, so its area is

$$\begin{aligned} A &= \int_{-\pi/4}^{\pi/4} \frac{1}{2} r^2 \, d\theta = \int_{-\pi/4}^{\pi/4} \frac{1}{2}\cos^2 2\theta \, d\theta = \int_0^{\pi/4} \cos^2 2\theta \, d\theta \\ &= \int_0^{\pi/4} \frac{1}{2}(1 + \cos 4\theta)\, d\theta = \frac{1}{2}\Big[\theta + \frac{1}{4}\sin 4\theta\Big]_0^{\pi/4} = \frac{\pi}{8}. \end{aligned}$$



Example 25.3. Consider two circles with radius 1 whose centers are a distance 1 apart. What is the area of the region that lies outside one circle and inside the other?

Choose polar coordinates in which the first circle is centered at the origin, so its polar equation is $r = 1$. Recall that $r = \cos\theta$ gives a circle centered at $(\frac{1}{2}, 0)$ with radius $\frac{1}{2}$, so the second circle has polar equation $r = 2\cos\theta$. Observe that these circles intersect when $1 = r = 2\cos\theta$, so $\cos\theta = \frac{1}{2}$, which occurs when $\theta = \pm\frac{\pi}{3}$. As shown in the picture, the region in which we are interested in is given by the inequalities $-\frac{\pi}{3} \le \theta \le \frac{\pi}{3}$ and $1 \le r \le 2\cos\theta$. Its area is $A = A_1 - A_2$, where A_1 is the area of the region inside $r = 2\cos\theta$ and A_2 is the area of the region inside $r = 1$. Our area formula gives $A_1 = \int_{-\pi/3}^{\pi/3} \frac{1}{2}(2\cos\theta)^2 \, d\theta$ and $A_2 = \int_{-\pi/3}^{\pi/3} \frac{1}{2} \cdot 1 \, d\theta$, so we get

$$\begin{aligned} A &= \int_{-\pi/3}^{\pi/3} \frac{1}{2}(4\cos^2\theta - 1)\, d\theta = \int_0^{\pi/3} (4\cos^2\theta - 1)\, d\theta \\ &= \int_0^{\pi/3} (2\cos(2\theta) + 1)\, d\theta = \Big[\sin(2\theta) + \theta\Big]_0^{\pi/3} = \sin\frac{2\pi}{3} + \frac{\pi}{3} = \frac{\sqrt{3}}{2} + \frac{\pi}{3}. \end{aligned}$$

Remark 25.4. In the above example we found the intersection points of two curves $r = f(\theta)$ and $r = g(\theta)$ by finding the values of θ for which $f(\theta) = g(\theta)$. There is one caveat that comes with this process. Suppose we look for intersection points of the four-leaf clover $r = \cos 2\theta$ with the circle $r = \frac{1}{2}$. Solving $\cos 2\theta = \frac{1}{2}$ produces 4 solutions in $[0, 2\pi)$, but it is clear from the picture following Example 25.1 that the circle intersects the clover in 8 places. The other 4 intersections come from points where r is negative; in other words, they correspond to solutions of $f(\theta + \pi) = g(\theta)$. When we are dealing with curves for which r may take negative values, we must be on the alert for this phenomenon.

25.3. Arc length in polar coordinates

To find the arc length of a curve $r = f(\theta)$ given in polar coordinates, we can once again proceed by writing it as a parametric curve

$$x = f(\theta)\cos\theta, \quad y = f(\theta)\sin\theta,$$

so that

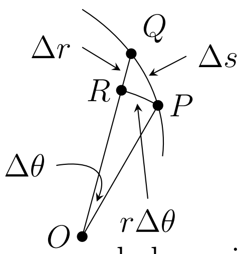
$$\frac{dx}{d\theta} = \frac{dr}{d\theta}\cos\theta - r\sin\theta, \quad \frac{dy}{d\theta} = \frac{dr}{d\theta}\sin\theta + r\cos\theta,$$

and the derivative of the arc length function s has square given by

$$\begin{aligned}\left(\frac{ds}{d\theta}\right)^2 = \left(\frac{dx}{d\theta}\right)^2 + \left(\frac{dy}{d\theta}\right)^2 = \left(\frac{dr}{d\theta}\right)^2 &\cos^2\theta - 2r\frac{dr}{d\theta}\cos\theta\sin\theta + r^2\sin^2\theta \\ + \left(\frac{dr}{d\theta}\right)^2 &\sin^2\theta + 2r\frac{dr}{d\theta}\cos\theta\sin\theta + r^2\cos^2\theta,\end{aligned}$$

which simplifies to

$$\left(\frac{ds}{d\theta}\right)^2 = \left(\frac{dr}{d\theta}\right)^2 + r^2. \tag{25.3}$$



As a mnemonic aid to remembering (25.3), we can multiply through by $(d\theta)^2$ to get

$$(ds)^2 = (dr)^2 + r^2(d\theta)^2, \tag{25.4}$$

where once again we have the caveat that we have not given these symbols an independent meaning. The formula (25.4) can be remembered by considering the diagram shown, in which P has polar coordinates (r,θ), Q has polar coordinates $(r+\Delta r, \theta+\Delta\theta)$, and R has polar coordinates $(r, \theta+\Delta\theta)$. Then the circular arc from P to R has length $r\Delta\theta$ and the line segment RQ has length Δr. The piece of curve from P to Q is not quite the hypotenuse of a right triangle with legs $r\Delta\theta$ and Δr, but it is very close to being this, and thus a good approximation to its length is given by

$$(\Delta s)^2 = (r\Delta\theta)^2 + (\Delta r)^2.$$

As P and Q get closer together, this approximation becomes better, and the meaning of (25.4) is that in the limit it gives exactly the integrand we need to compute arc length. In particular, using (25.3) we conclude that the arc length over the interval $a \le \theta \le b$ is

$$L = \int_a^b \sqrt{r^2 + \left(\frac{dr}{d\theta}\right)^2}\, d\theta. \tag{25.5}$$

Example 25.5. The curve $r = 2\cos\theta$ for $0 \le \theta \le \pi$ has arc length

$$L = \int_0^\pi \sqrt{(2\cos\theta)^2 + (-2\sin\theta)^2}\, d\theta = \int_0^\pi 2\, d\theta = 2\pi,$$

which is reassuring since this is a circle with radius 1.

Example 25.6. The arc length of the cardioid $r = 1 + \cos\theta$ $(0 \le r \le 2\pi)$ is

$$L = \int_0^{2\pi} \sqrt{(1+\cos\theta)^2 + (-\sin\theta)^2}\, d\theta = \int_0^{2\pi} \sqrt{1 + 2\cos\theta + \cos^2\theta + \sin^2\theta}\, d\theta$$

$$= \int_0^{2\pi} \sqrt{2 + 2\cos\theta}\, d\theta = \sqrt{2} \int_0^{2\pi} \sqrt{1 + \cos\theta}\, d\theta.$$

Multiplying top and bottom by $\sqrt{1 - \cos\theta}$ gives

$$\int_0^{2\pi} \sqrt{1+\cos\theta}\, d\theta = \int_0^{2\pi} \frac{\sqrt{(1+\cos\theta)(1-\cos\theta)}}{\sqrt{1-\cos\theta}}\, d\theta = \int_0^{2\pi} \frac{\sqrt{1-\cos^2\theta}}{\sqrt{1-\cos\theta}}\, d\theta$$

$$= \int_0^{2\pi} \frac{|\sin\theta|}{\sqrt{1-\cos\theta}}\, d\theta = \int_0^{\pi} \frac{\sin\theta}{\sqrt{1-\cos\theta}}\, d\theta + \int_\pi^{2\pi} \frac{-\sin\theta}{\sqrt{1-\cos\theta}}\, d\theta.$$

The substitution $u = 1 - \cos\theta$ has $du = \sin\theta$ and thus

$$\int \frac{\sin\theta}{\sqrt{1-\cos\theta}}\, d\theta = \int u^{-1/2}\, du = 2\sqrt{u} + C = 2\sqrt{1-\cos\theta} + C.$$

Using this we can evaluate the above integrals and conclude that

$$\int_0^{2\pi} \sqrt{1+\cos\theta}\, d\theta = \Big[2\sqrt{1-\cos\theta}\Big]_0^{\pi} - \Big[2\sqrt{1-\cos\theta}\Big]_{\pi}^{2\pi} = 2\sqrt{2} - 0 - (0 - 2\sqrt{2}) = 4\sqrt{2}.$$

Thus the arc length of the cardioid is $L = \sqrt{2} \cdot 4\sqrt{2} = 8$.

An alternate method for evaluating $\int_0^{2\pi} \sqrt{1+\cos\theta}\, d\theta$ (instead of the algebraic trick we used) is to use the identity $\cos\theta = 2\cos^2\frac{\theta}{2} - 1$ to write

$$(25.6) \qquad \int_0^{2\pi} \sqrt{1+\cos\theta}\, d\theta = \int_0^{2\pi} \sqrt{2\cos^2\frac{\theta}{2}}\, d\theta = \sqrt{2} \int_0^{2\pi} \Big|\cos\frac{\theta}{2}\Big|\, d\theta.$$

Since $\cos\frac{2\pi-\theta}{2} = \cos(\pi - \frac{\theta}{2}) = -\cos\frac{\theta}{2}$, we see that the function $\theta \mapsto |\cos\frac{\theta}{2}|$ is symmetric around the line $\theta = \pi$, and thus $\int_0^\pi |\cos\frac{\theta}{2}|\, d\theta = \int_\pi^{2\pi} |\cos\frac{\theta}{2}|\, d\theta$, so we conclude that

$$\int_0^{2\pi} \Big|\cos\frac{\theta}{2}\Big|\, d\theta = 2\int_0^{\pi} \Big|\cos\frac{\theta}{2}\Big|\, d\theta = 2\int_0^{\pi} \cos\frac{\theta}{2}\, d\theta = 2\Big[2\sin\frac{\theta}{2}\Big]_0^{\pi} = 4,$$

where the second equality uses the fact that $\cos\frac{\theta}{2} \geq 0$ for all $\theta \in [0, \pi]$. Together with (25.6) this once again gives $\int_0^{2\pi} \sqrt{1+\cos\theta}\, d\theta = 4\sqrt{2}$, thus $L = 8$.

Part V. Sequences and series

Lecture 26 Sequences

Stewart §11.1, Spivak Ch. 22

26.1. Sequences and limits

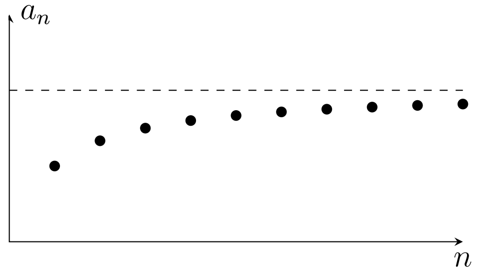
A *sequence* is a list of numbers $a_1, a_2, a_3, \dots$ in a given order (so $1, 2, 3, 4, \dots$ is a different sequence from $1, 3, 2, 4, \dots$); equivalently, a sequence is a function from $\mathbb{N}$ to $\mathbb{R}$. We refer to a_n as the *nth term* of the sequence. We will often write $\{a_n\}$, $\{a_n\}_{n=1}^{\infty}$, (a_n), or $(a_n)_{n=1}^{\infty}$ to refer to the sequence as a whole.

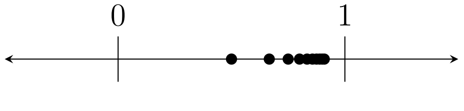
A sequence may or may not be given in terms of a nice formula: for example, $a_n = \frac{n}{n+1}$ has a nice explicit formula for each term, while the sequence

$$b_1 = 0, \quad b_{n+1} = 1 + \sqrt{b_n}$$

is defined *recursively*, and there is no simple formula for its nth term. Or we might consider the sequence whose nth term c_n is the nth digit of the decimal expansion of π, and then there is neither an explicit nor recursive formula that is readily available. A similar thing occurs with the sequence $2, 3, 5, 7, 11, 13, 17, \dots$, where the nth term p_n is the nth prime number.

We can plot a sequence $\{a_n\}$ by drawing a dot at each of the points (n, a_n); this is the graph of the function $\mathbb{N} \to \mathbb{R}$ defined by $n \mapsto a_n$. The first picture shows the result for $a_n = \frac{n}{n+1}$ (with the horizontal axis compressed to save space).





Another option is to draw a number line and put a dot at a_n for each value of n, as shown in the second picture. This second method has the advantage of providing a more compact representation, but the disadvantage that it loses all information about the *order* in which the terms of the sequence appear, since permuting these terms would result in the same picture; moreover, if several terms of the sequence are close together then it becomes difficult to distinguish them. In the end, we tend not to rely on graphical representations of sequences nearly as much as we do for functions $\mathbb{R} \to \mathbb{R}$, and so we will not use either of these methods that often.

Many of our basic definitions and theorems about limits for functions have analogues for sequences.

Definition 26.1. A sequence $\{a_n\}$ has a *limit* $L \in \mathbb{R}$ if for every $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that for every $n \geq N$, we have $|a_n - L| < \epsilon$. In this case we write $\lim_{n\to\infty} a_n = L$, or sometimes "$a_n \to L$ as $n \to \infty$". If the sequence a_n has a limit, we say that *the sequence converges.* If it does not have a limit, we say that *the sequence diverges.*

The following simple fact will occasionally be useful.

Exercise 26.2*.* Prove that $\lim_{n\to\infty} a_{n+1} = \lim_{n\to\infty} a_n$ whenever the sequence converges.

Definition 26.3. One type of diverging sequence is worth particular mention. We write $\lim_{n\to\infty} a_n = \infty$ (or sometimes "$a_n \to \infty$ as $n \to \infty$") if for every $M > 0$ there exists $N \in \mathbb{N}$ such that for every $n \geq N$, we have $a_n \geq M$, and $\lim_{n\to\infty} a_n = -\infty$ (or $a_n \to -\infty$) if for every $M > 0$ there exists $N \in \mathbb{N}$ such that for every $n \geq N$ we have $a_n \leq -M$.

Exercise 26.4*.* Show that the sequence $x_n = (-1)^n$ is divergent.

All of the limit laws still work, just as they did for functions. Thus we have

$$\lim_{n\to\infty} \frac{n}{n+1} = \lim_{n\to\infty} \frac{1}{1+\frac{1}{n}} = \frac{1}{\lim_{n\to\infty}(1+\frac{1}{n})} = \frac{1}{1+\lim_{n\to\infty}\frac{1}{n}} = \frac{1}{1+0} = 1.$$

Similarly, the squeeze theorem still holds.

Theorem 26.5 (Squeeze theorem)**.** *Given three sequences satisfying $a_n \leq b_n \leq c_n$ for all n, if we have $\lim_{n\to\infty} a_n = \lim_{n\to\infty} c_n = L$, then $\lim_{n\to\infty} b_n = L$ as well.*

Proof. Exercise: recall the proof of the squeeze theorem for functions, and adapt it. Observe that as part of the proof, you must show that the sequence b_n converges. □

Proposition 26.6. *A sequence a_n converges to 0 if and only if $|a_n|$ also converges to 0.*

Proof. We have $-|a_n| \leq a_n \leq |a_n|$ for all n, so if $|a_n| \to 0$ then $-|a_n| \to 0$ by the limit laws, and the squeeze theorem implies that $a_n \to 0$ as well. The other direction is a short exercise using the definition. □

Theorem 26.7. *If a sequence $\{a_n\}$ and a function $f\colon \mathbb{R} \to \mathbb{R}$ are related by $a_n = f(n)$, and if moreover we have $\lim_{x\to\infty} f(x) = L$, then $\lim_{n\to\infty} a_n = L$.*

Proof. Exercise. □

Example 26.8. If $b_n = \frac{\ln n}{n}$, then we have $b_n = f(n)$ where $f(x) = \frac{\ln x}{x}$. Since $\ln x \to \infty$ as $x \to \infty$, we see that $\lim_{x\to\infty} f(x)$ has indeterminate form, and so l'Hospital's rule together with Theorem 26.7 gives

$$\lim_{n\to\infty} b_n = \lim_{x\to\infty} \frac{\ln x}{x} = \lim_{x\to\infty} \frac{1/x}{1} = 0.$$

Example 26.9. The sequence $(-1)^n$ whose terms are $-1, 1, -1, 1, -1, 1, \ldots$ diverges, but the sequence $\frac{(-1)^n}{n}$ whose terms are $-\frac{1}{n}, \frac{2}{n}, -\frac{3}{n}, \frac{4}{n}, \ldots$ converges to 0 by Proposition 26.6, since $\frac{1}{n} \to 0$.

Theorem 26.10. *If f is a function that is continuous at L, and a_n is a sequence in the domain of f such that $\lim_{n\to\infty} a_n = L$, then $\lim_{n\to\infty} f(a_n) = f(L)$.*

Proof. Exercise (use the definition of continuity). □

Example 26.11. Since $\frac{\pi}{n} \to 0$ as $n \to \infty$ and since $\theta \mapsto \sin\theta$ is continuous at 0, we have

$$\lim_{n\to\infty} \sin\frac{\pi}{n} = \sin\left(\lim_{n\to\infty}\frac{\pi}{n}\right) = \sin 0 = 0.$$

Example 26.12. Consider the sequence $a_n = \frac{n!}{n^n}$. The numerator and denominator both diverge to ∞, so this has indeterminate form, but we cannot use l'Hospital's rule without first finding some differentiable function $f(x)$ such that $f(n) = n!$. Since we do not know any such function,[16] we use a different argument, and observe that for every n we have

$$0 \le a_n = \frac{1}{n}\cdot\left(\frac{2}{n}\cdot\frac{3}{n}\cdots\frac{n}{n}\right) \le \frac{1}{n}.$$

Since $\frac{1}{n} \to 0$, the squeeze theorem implies that $\frac{n!}{n^n} \to 0$ as $n \to \infty$.

Example 26.13. Recall from our study of exponential functions that

$$\lim_{x\to\infty} a^x = \begin{cases} 0 & \text{if } 0 \le a < 1, \\ 1 & \text{if } a = 1, \\ \infty & \text{if } a > 1. \end{cases}$$

Using Theorem 26.7, this implies that given $r \ge 0$, the sequence r^n satisfies

$$\lim_{n\to\infty} r^n = \begin{cases} 0 & \text{if } 0 \le r < 1, \\ 1 & \text{if } r = 1, \\ \infty & \text{if } r > 1. \end{cases}$$

In particular, given any $r \in (-1, 1)$, we have

$$|r^n| = |r|^n \to 0 \quad \text{since } |r| \in [0, 1).$$

By Proposition 26.6, this implies that $r^n \to 0$. We conclude that $r^n \to 0$ for every $|r| < 1$, and $r^n \to 1$ when $r = 1$. For all other values of r, the sequence r^n diverges.

26.2. Monotonic sequences

A sequence a_n is called *increasing* if $a_{n+1} > a_n$ for every n, and *decreasing* if $a_{n+1} < a_n$ for every n. If one of these conditions holds, then the sequence is called *monotonic.*

Remark 26.14. If we weaken the condition to $a_{n+1} \ge a_n$ for all n, then we say that the sequence is *nondecreasing.* Similarly if $a_{n+1} \le a_n$ for all n, then the sequence is *nonincreasing.* You should be warned that some authors use "increasing" to mean "nondecreasing", and say "strictly increasing" when they mean $a_{n+1} > a_n$; similarly for "decreasing" and "strictly decreasing. Thus if you encounter the words "increasing" or "decreasing" when you read a piece of mathematics, it is worth checking to see in which sense the author is using them.

Example 26.15.

(1) The sequence 3, 3.1, 3.14, 3.141, 3.1415, 3.14159, . . . is increasing.[17]

[16] In fact there is such a function, called the *gamma function*, but we have not studied this yet.

[17] Actually to make this completely true, we need to add two digits whenever we encounter a 0 in the decimal expansion of π; as given, the sequence is merely nondecreasing.

(2) The sequence $a_n = \frac{1}{n}$ is decreasing, since $n+1 > n$ implies $\frac{1}{n+1} < \frac{1}{n}$.
(3) The sequence $b_n = n$ is increasing.
(4) The sequence $c_n = (-1)^n$ is neither increasing nor decreasing.
(5) The sequence $d_n = \frac{n}{n^2+1}$ is decreasing. To prove this we can observe that $d_n = f(n)$ where $f(x) = \frac{x}{x^2+1}$ has derivative

$$f'(x) = \frac{(x^2+1)\cdot 1 - x\cdot 2x}{(x^2+1)^2} = \frac{1-x^2}{(x^2+1)^2} < 0 \quad \text{for all } x > 1$$

and thus is decreasing on $(1,\infty)$. Alternately we can use the direct computation

$$\begin{aligned} d_{n+1} - d_n &= \frac{n+1}{(n+1)^2+1} - \frac{n}{n^2+1} = \frac{(n+1)(n^2+1) - n(n^2+2n+2)}{(n^2+2n+2)(n^2+1)} \\ &= \frac{(n^3+n^2+n+1) - (n^3+2n^2+2n)}{(n^2+2n+2)(n^2+1)} = \frac{1-n-n^2}{(n^2+2n+2)(n^2+1)} < 0. \end{aligned}$$

Definition 26.16. A sequence $\{a_n\}$ is *bounded above* if there exists $M \in \mathbb{R}$ such that $a_n \le M$ for all $n \in \mathbb{N}$; in this case M is called an *upper bound* for the sequence.

Similarly, the sequence is *bounded below* if there exists $m \in \mathbb{R}$ such that $a_n \ge m$ for all $n \in \mathbb{N}$; in this case m is a *lower bound* for the sequence.

We say that $\{a_n\}$ is *bounded* if it is bounded above and bounded below.

Exercise 26.17. Show that $\{a_n\}$ is bounded if and only if $\{|a_n|\}$ is bounded above.

Example 26.18.

(1) The sequence 3, 3.1, 3.14, 3.141, 3.1415, ... is bounded; 3 is a lower bound, and π is an upper bound.
(2) The sequence $a_n = \frac{1}{n}$ is bounded; 0 is a lower bound, and 1 is an upper bound.
(3) The sequence $b_n = n$ is bounded below by 1, but is not bounded above.
(4) The sequence $c_n = (-1)^n$ is bounded; -1 is a lower bound, and 1 is an upper bound.
(5) The sequence $d_n = \frac{n}{n^2+1}$ is bounded; 0 is a lower bound, and $d_1 = \frac{1}{2}$ is an upper bound because the sequence is decreasing.

Observe that in each of these cases, the lower and upper bounds that are quoted are in fact optimal. For example, -1 is also a lower bound for the sequence $a_n = \frac{1}{n}$, but it seems better to use the (larger) lower bound 0, since this carries more information. Similarly, 2 is an upper bound for this sequence, but the upper bound 1 is in some sense better. This line of thinking motivates the following definition.

Definition 26.19. A real number M is the *least upper bound* for a sequence $\{a_n\}$ if

- M is an upper bound ($a_n \le M$ for all $n \in \mathbb{N}$), and
- no number smaller than M is an upper bound (for every $L < M$, there is $n \in \mathbb{N}$ such that $a_n > L$).

In this case we also call M the *supremum* of the sequence, and write $M = \sup_n a_n$.

Similarly, m is the *greatest lower bound* for $\{a_n\}$ if

- m is a lower bound ($a_n \ge m$ for all $n \in \mathbb{N}$), and
- no number larger than m is an upper bound (for every $\ell > m$, there is $n \in \mathbb{N}$ such that $a_n < \ell$).

In this case we also call m the *infimum* of the sequence, and write $m = \inf_n a_n$.

It is easy to identify the supremum or infimum when it occurs as a term in the sequence; in the example above, this was the case for the infimums of the increasing sequences 3, 3.1, 3.141, … and $b_n = n$, and for the supremums of the decreasing sequences $a_n = \frac{1}{n}$ and $d_n = \frac{n}{n^2+1}$. It was also the case for $c_n = (-1)^n$, where every term is either ± 1.

When the supremum or infimum does not occur as a term in the sequence, we rely on the following fundamental property of the real numbers.[18]

Least Upper Bound Property. *If $\{a_n\}$ is a sequence of real numbers that is bounded above, then it has a least upper bound in the real numbers. Similarly, if $\{a_n\}$ is a sequence of real numbers that is bounded below, then it has a greatest lower bound in the real numbers.*

Remark 26.20. The Least Upper Bound Property is not a theorem that we are going to prove; rather, it is a fundamental property of the real numbers, which we assume as an axiom. Later in your mathematical career, you will learn how to *construct* the real numbers in such a way that this property is satisfied. For now we content ourselves with the observation that this property fails dramatically if we work with the rational numbers instead of the real numbers. Indeed, the first sequence in Example 26.18 is a sequence of rational numbers that admits a rational upper bound (4 will work) but does *not* have a least upper bound in the rational numbers (because π is irrational).

Theorem 26.21 (Monotone Convergence Theorem)**.** *If a_n is a nondecreasing sequence that is bounded above, then it converges to its supremum. Similarly, if b_n is a nonincreasing sequence that is bounded below, then it converges to its infimum.*

Proof. Let M be the least upper bound of the sequence a_n. Then for every $\epsilon > 0$, the numbers $M - \epsilon$ is not an upper bound (by the definition of least upper bound), so there is some $N \in \mathbb{N}$ such that $a_N > M - \epsilon$. But since the sequence is nondecreasing, this implies that for every $n \geq N$ we have $M - \epsilon < a_N \leq a_n \leq M$, which verifies the definition of a limit and proves the first half of the theorem. The second half follows by observing that $a_n = -b_n$ is nondecreasing and is bounded above. □

Observe that the first, second, and last sequences in Example 26.18 illustrate the theorem; the first sequence converges to its supremum π, while the second and last sequences converge to their infimum 0.

Example 26.22. Define a sequence a_n recursively by $a_1 = 1$, $a_{n+1} = \frac{1}{2}(a_n + 2)$. We claim that $a_n \leq 2$ for all n, and that a_n is nondecreasing. Observe that if $a_n \leq 2$, then $a_{n+1} \leq 2$ as well, so the first claim follows by induction since $a_1 = 1 < 2$. Moreover, if $a_n \leq 2$, then $a_{n+1} = \frac{1}{2}(a_n + 2) \geq \frac{1}{2}(a_n + a_n) = a_n$, so a_n is nondecreasing. By the Monotone Convergence Theorem, $L = \lim_{n\to\infty} a_n$ exists. Thus we have

$$L = \lim_{n\to\infty} a_{n+1} = \lim_{n\to\infty} \frac{1}{2}(a_n + 2) = \frac{1}{2}\left(\left(\lim_{n\to\infty} a_n\right) + 2\right) = \frac{1}{2}(L + 2),$$

and solving for L gives $L = 2$. Thus $a_n \to 2$ as $n \to \infty$.

[18] We state the property for sequences, but in fact it, and the definitions of infimum and supremum above, are valid for any subset of $\mathbb{R}$.

Lecture 27 Summing an infinite series

Stewart §11.2, Spivak Ch. 23

27.1. Convergence and divergence

Suppose we want to add together all of the terms of a sequence $a_1, a_2, a_3, \dots$. We refer to this as an *infinite series* (often just *series*) and write

$$a_1 + a_2 + a_3 + \cdots + a_n + \cdots = \sum_{n=1}^{\infty} a_n = \sum a_n,$$

where the last notation is a shorthand that we will usually avoid, preferring to write the bounds of summation explicitly to avoid confusion.

Does this notion make sense? What does it mean to add infinitely many numbers together? Certainly we feel as though we run into trouble if we try to compute $1+2+3+4+\cdots$. On the other hand, if we are confronted with the sum $\sum_{n=1}^{\infty} \frac{1}{2^n} = \frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\cdots$, then we may reasonably observe that the first n terms in the sum admit the explicit formula

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots + \frac{1}{2^n} = 1 - \frac{1}{2^n},$$

which can easily be proved by induction. The RHS converges to 1 as $n \to \infty$, so it is reasonable to say that the infinite sum $\sum_{n=1}^{\infty} \frac{1}{2^n}$ also converges to 1.

Definition 27.1. Given a sequence $\{a_n\}$, the corresponding series is $\sum_{n=1}^{\infty} a_n$. The *partial sums* of the series are the numbers $S_n = \sum_{k=1}^{n} a_k$. If the sequence of partial sums converges to a real number S, then we say that the series $\sum a_n$ is *convergent*, and write $\sum_{n=1}^{\infty} a_n = S$; we call S the *sum* of the series. If the sequence of partial sums does not converge, we say that the series is *divergent*.

A good way of remembering this is by the notation

$$\sum_{n=1}^{\infty} a_n = \lim_{N\to\infty} \sum_{n=1}^{N} a_n,$$

which is clearly analogous to the way we dealt with improper integrals:

$$\int_1^{\infty} f(x)\,dx = \lim_{t\to\infty} \int_1^{t} f(x)\,dx.$$

We will develop the relationship between infinite series and improper integrals further in a little while.

27.2. Geometric series

Example 27.2. A *geometric series* is a series of the form

$$a + ar + ar^2 + ar^3 + \cdots = \sum_{n=1}^{\infty} ar^{n-1},$$

where $a, r \in \mathbb{R}$. If $r = 1$ then clearly this series diverges since the nth partial sum is $S_n = an$. When $r \neq 1$, we can write the nth partial sum explicitly by observing that

$$\begin{aligned} S_n &= a + ar + ar^2 + \cdots + ar^{n-1}, \\ rS_n &= ar + ar^2 + ar^3 + \cdots + ar^n. \end{aligned}$$

Subtracting these two gives

$$S_n - rS_n = a - ar^n \quad \Rightarrow \quad S_n = \Big(\frac{1 - r^n}{1 - r}\Big)a.$$

This diverges if $|r| \geq 1$, while if $|r| < 1$ then we have

$$\lim_{n\to\infty} S_n = \Big(\frac{1 - \lim_{n\to\infty} r^n}{1 - r}\Big)a = \frac{a}{1 - r}.$$

The result of this example is important enough to be worth stating as a theorem.

Theorem 27.3. *The geometric series $\sum_{n=1}^\infty ar^{n-1}$ is convergent if and only if $|r| < 1$, and in this case the sum is $\frac{a}{1-r}$.*

Example 27.4. $\sum_{n=1}^\infty 2^{2n}3^{1-n} = \sum_{n=1}^\infty \frac{4^n}{3^n} \cdot 3$ diverges because it is a geometric series with $r = \frac{4}{3}$.

Example 27.5. The repeating decimal $3.2\overline{41} = 3.2414141414141\ldots$ can be written using a geometric series:

$$\begin{aligned} 3.2\overline{41} &= 3.2 + \frac{41}{10^3} + \frac{41}{10^5} + \frac{41}{10^7} + \cdots = 3.2 + \frac{41}{10^3}\sum_{n=1}^\infty (10^{-2})^{n-1} \\ &= \frac{32}{10} + \frac{41}{10^3} \cdot \frac{1}{1 - \frac{1}{100}} = \frac{32}{10} + \frac{41}{10 \cdot 99} = \frac{3209}{990}. \end{aligned}$$

It is also worth highlighting the case of a geometric series with $a = 1$: given any $|x| < 1$, we have

$$\sum_{n=0}^\infty x^n = \sum_{n=1}^\infty x^{n-1} = \frac{1}{1 - x}.$$

This is our first example of a *power series* representation of a function, which we will spend more time on later.

27.3. Other examples

Example 27.6. Consider the series

$$\sum_{n=1}^\infty \frac{1}{n(n+1)} = \frac{1}{1\cdot 2} + \frac{1}{2\cdot 3} + \frac{1}{3\cdot 4} + \cdots .$$

To compute the partial sums and determine convergence or divergence, we can use the observation that

$$\frac{1}{n(n+1)} = \frac{1}{n} - \frac{1}{n+1},$$

and thus

$$\begin{aligned} S_n &= \sum_{k=1}^{n} \frac{1}{k(k+1)} = \sum_{k=1}^{n} \Big(\frac{1}{k} - \frac{1}{k+1}\Big) \\ &= \Big(1 - \frac{1}{2}\Big) + \Big(\frac{1}{2} - \frac{1}{3}\Big) + \Big(\frac{1}{3} - \frac{1}{4}\Big) + \cdots + \Big(\frac{1}{n} - \frac{1}{n+1}\Big) = 1 - \frac{1}{n+1}. \end{aligned}$$

The sum $\sum_{k=1}^{n}(\frac{1}{k} - \frac{1}{k+1})$ is called a *telescoping sum* because it collapses into the short easy-to-handle expression $1 - \frac{1}{n+1}$. We now see that $S_n \to 1$ as $n \to \infty$, so the series is convergent and the infinite sum is 1.

Example 27.7. The series $\sum_{n=1}^{\infty} \frac{1}{n}$ is called the *harmonic series*. We claim that it is divergent. To prove this, observe that

$$S_1 = 1, \quad S_2 = 1 + \frac{1}{2}, \quad S_4 = 1 + \frac{1}{2} + \underbrace{\frac{1}{3} + \frac{1}{4}}_{>2\cdot\frac{1}{4}=\frac{1}{2}} > 1 + 2 \cdot \frac{1}{2},$$

and that similar estimates are available for S_8, S_{16}, etc.:

$$S_8 = S_4 + \underbrace{\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}}_{>4\cdot\frac{1}{8}=\frac{1}{2}} > S_4 + \frac{1}{2} > 1 + 3 \cdot \frac{1}{2},$$

$$S_{16} = S_8 + \sum_{n=9}^{16} \frac{1}{n} > S_8 + \sum_{n=9}^{16} \frac{1}{16} = S_8 + \frac{1}{2} > 1 + 4 \cdot \frac{1}{2}.$$

In general, we have

$$S_{2^{n+1}} = S_{2^n} + \sum_{k=2^n+1}^{2^{n+1}} \frac{1}{k} > S_{2^n} + \sum_{k=2^n+1}^{2^{n+1}} \frac{1}{2^{n+1}} = S_{2^n} + 2^n \cdot \frac{1}{2^{n+1}} = S_{2^n} + \frac{1}{2},$$

and it follows by induction that for every n we have $S_{2^n} > 1 + \frac{n}{2}$. Since the RHS goes to ∞ as $n \to \infty$, we conclude that the partial sums diverge, hence the harmonic series diverges.

Remark 27.8. Writing $N = 2^n$, the lower bound above gives $\sum_{k=1}^{N} \frac{1}{k} > \frac{n}{2} = \frac{1}{2}\log_2 N$. In fact a better estimate is $\sum_{k=1}^{N} \frac{1}{k} \approx \ln N$, but this takes a little more work.

27.4. Basic theorems

Theorem 27.9. *If the series $\sum_{n=1}^{\infty} a_n$ is convergent, then the sequence of terms a_n converges to* 0.

Proof. As usual, let $S_n = \sum_{k=1}^{n} a_k$, and observe that $a_n = S_n - S_{n-1}$. If the series is convergent then $L = \lim_{n\to\infty} S_n$ exists, and by Exercise 26.2 and the limit laws we have

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} (S_n - S_{n-1}) = \lim_{n\to\infty} S_n - \lim_{n\to\infty} S_{n-1} = L - L = 0. \qquad \square$$

Remark 27.10. It is worth reiterating that every series has two sequences associated to it: the sequence of terms, which we often denote a_n, and the sequence of partial sums which we often denote S_n. Theorem 27.9 says that if S_n converges, then a_n converges to 0.

Remark 27.11. The converse of this theorem is not true; $a_n \to 0$ does not guarantee that the series converges, as the example of the harmonic series illustrates.

Corollary 27.12. *If a_n is a divergent sequence, or a convergent sequence whose limit is not equal to 0, then the corresponding series $\sum_{n=1}^{\infty} a_n$ is divergent.*

Example 27.13. $\sum_{n=1}^{\infty} \frac{2n^2}{n^2+1}$ diverges because $\frac{2n^2}{n^2+1} \to 2$.

Theorem 27.14. *If the series $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$ both converge, then so do the series $\sum_{n=1}^{\infty}(a_n + b_n)$, $\sum_{n=1}^{\infty}(a_n - b_n)$, and $\sum_{n=1}^{\infty} ca_n$, where $c \in \mathbb{R}$. Moreover, we have*

$$\sum_{n=1}^{\infty}(a_n \pm b_n) = \Big(\sum_{n=1}^{\infty} a_n\Big) \pm \Big(\sum_{n=1}^{\infty} b_n\Big), \qquad \sum_{n=1}^{\infty}(ca_n) = c\Big(\sum_{n=1}^{\infty} a_n\Big).$$

Proof. We prove the result for addition and leave the others as exercises. Observe that the partial sums $S_n = \sum_{k=1}^{n}(a_k + b_k)$ satisfy

$$S_n = \Big(\sum_{k=1}^{n} a_k\Big) + \Big(\sum_{k=1}^{n} b_k\Big),$$

and the two sequences of partial sums on the RHS converge by assumption, so the limit law for addition gives

$$\lim_{n\to\infty} S_n = \lim_{n\to\infty}\Big(\sum_{k=1}^{n} a_k\Big) + \lim_{n\to\infty}\Big(\sum_{k=1}^{n} b_k\Big) = \sum_{n=1}^{\infty} a_n + \sum_{n=1}^{\infty} b_n.$$

The other two results are similar, using the corresponding limit laws. □

Example 27.15.

$$\sum_{n=1}^{\infty}\Big(\frac{2}{n(n+1)} + \frac{3}{2^n}\Big) = 2\sum_{n=1}^{\infty}\frac{1}{n(n+1)} + \frac{3}{2}\sum_{n=1}^{\infty}\Big(\frac{1}{2}\Big)^{n-1} = 2\cdot 1 + \frac{3/2}{1-\frac{1}{2}} = 2 + 3 = 5.$$

We conclude with one more general observation: convergence only depends on the ‘tail’ of the series, and is not affected if we change finitely many terms. The following exercise makes this precise.

Exercise 27.16. Let $\sum a_n$ and $\sum b_n$ be two series with the property that there exists $N \in \mathbb{N}$ such that $a_n = b_n$ for all $n \geq N$; in other words, we can obtain (b_n) from (a_n) by changing finitely many terms. Show that $\sum a_n$ converges if and only if $\sum b_n$ converges.

For example, if $a_n = \frac{1}{n}$ for $n < 1000$ and $a_n = 2^{-n}$ for $n \geq 1000$, then $\sum a_n$ converges even though the first part (the first 1000 terms) looks like the (divergent) harmonic series, because we can obtain a_n from the (convergent) geometric series $\sum 2^{-n}$ by changing finitely many terms.

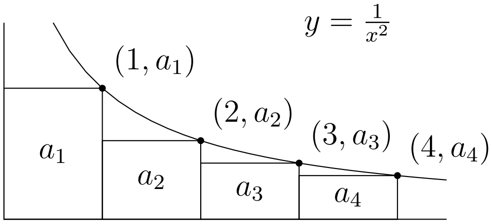
Lecture 28 The integral test

Stewart §11.3, Spivak Ch. 23

28.1. Some examples and a theorem

In Example 27.6 we showed that the series $\sum \frac{1}{n(n+1)} = \sum \frac{1}{n^2+n}$ converges. What about the series $\sum_{n=1}^{\infty} \frac{1}{n^2}$? In this case we have no nice formula for $S_n = 1 + \frac{1}{2^2} + \frac{1}{3^2} + \cdots + \frac{1}{n^2}$, so it is not clear how to check whether the partial sums converge.

One approach is to observe that $a_n = \frac{1}{n^2} = f(n)$, where $f(x) = \frac{1}{x^2}$, and the integral of $f(x)$ is easy to compute; then we need to compare $\sum_{k=1}^{n} \frac{1}{k^2}$ and $\int_1^n \frac{1}{x^2}\,dx$. The picture at right shows how to do this. The rectangles shown have areas $a_1, a_2, \dots$, and they all lie underneath the graph of $\frac{1}{x^2}$. In particular, we see that for any $n \geq 2$, the region covered by the rectangles with areas $a_2, a_3, \dots, a_n$ lies inside the region underneath the graph between $x = 1$ and $x = n$, so we have



$$a_2 + a_3 + \cdots + a_n \leq \int_1^n \frac{1}{x^2}\,dx = \Big[-\frac{1}{x}\Big]_1^n = 1 - \frac{1}{n}.$$

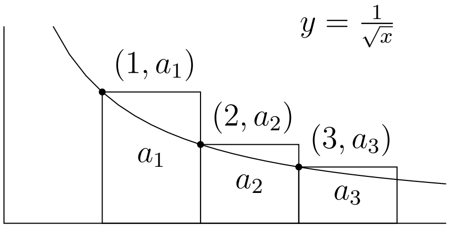
We conclude that the partial sums of the series satisfy

$$S_n = a_1 + a_2 + a_3 + \cdots + a_n \leq 2 - \frac{1}{n} \leq 2$$

for all n, so the sequence of partial sums is bounded above. Since all the terms a_n are positive, the sequence S_n is also increasing, and thus by the monotone convergence theorem it converges.

Remark 28.1. Note that the above argument gives us no information about the actual value of the infinite sum $\sum_{n=1}^{\infty} \frac{1}{n^2}$, merely that it converges. In fact one can prove that the value is $\frac{\pi^2}{6}$, but this takes significantly more work than we will enter into here.

Now consider another example, the series $\sum_{n=1}^{\infty} \frac{1}{\sqrt{n}}$. In this case the above argument does not yield an upper bound for the partial sums, because $\int_1^n \frac{1}{\sqrt{x}}\,dx = [2\sqrt{x}]_1^n = 2(\sqrt{n} - 1) \to \infty$. This suggests that perhaps we should try to prove that this series is divergent. And indeed, by modifying the above picture slightly, sliding each rectangle one unit to the right so that their tops lie *above* the graph of the function, we obtain the bound



$$S_n = a_1 + a_2 + a_3 + \cdots + a_n \geq \int_1^{n+1} \frac{1}{\sqrt{x}}\,dx = 2(\sqrt{n+1} - 1).$$

The RHS diverges to ∞ as $n \to \infty$, so we conclude that the series $\sum_{n=1}^{\infty} \frac{1}{\sqrt{n}}$ diverges.

The arguments used in these two examples lead to the following result.

Theorem 28.2 (Integral test for series)**.** *Consider the series $\sum a_n$. Suppose that $f\colon [1,\infty) \to [0,\infty)$ is a continuous, nonnegative, nonincreasing function such that $f(n) = a_n$ for all $n \in \mathbb{N}$. Then $\sum a_n$ converges if and only if the improper integral $\int_1^{\infty} f(x)\,dx$ converges.*

We will prove this theorem, together with some more detailed estimates, in Proposition 28.6 below. First we point out a couple applications.

Example 28.3. Given $p \in \mathbb{R}$, the *p-series* $\sum_{n=1}^{\infty} \frac{1}{n^p}$ converges if and only if $p > 1$. To see this, observe that for all $p \leq 0$, the terms do not converge to 0, so the series diverges by Corollary 27.12. For $p > 0$, the function $f(x) = x^{-p}$ is continuous, positive, and decreasing on $[1, \infty)$, so by the integral test the series converges if and only if $\int_1^\infty x^{-p}\,dx$ converges, which occurs if and only if $p > 1$.

Example 28.4. To check convergence of $\sum \frac{\ln n}{n^2}$, we attempt to use the integral test with $f(x) = \frac{\ln x}{x^2}$. Differentiating to check whether the function is decreasing, we see that

$$f'(x) = \frac{x^2 \cdot \frac{1}{x} - (\ln x)2x}{x^4} = \frac{1 - 2\ln x}{x^3},$$

which is < 0 for all $x > \sqrt{e}$. Thus the function is decreasing on $(\sqrt{e}, \infty)$; this is not quite what Theorem 28.2 asked for, but as we will see below, it turns out to be enough, and the integral test still works. We can compute the integral by parts:

$$\int_1^t \frac{\ln x}{x^2}\,dx = \Big[-\frac{\ln x}{x}\Big]_1^t + \int_1^t \frac{1}{x^2}\,dx = \Big[-\frac{\ln x}{x} - \frac{1}{x}\Big]_1^t = -\frac{1+\ln t}{t} - (-1) = 1 - \frac{1+\ln t}{t}.$$

This converges to 1 as $t \to \infty$, so the series is convergent as well.

The hypothesis that f is nonincreasing is vital, as the following exercise shows.

Exercise 28.5. Define a function f by setting $f(n) = 0$ and $f(n + \frac{1}{2}) = 1$ for all integers n, and then connecting these points on the graph with straight lines, so that $f(n+t) = f(n-t) = 2t$ for $t \in [0, \frac{1}{2}]$. Sketch the graph of f and show that $\int_1^\infty f(x)\,dx$ diverges but $\sum_{n=1}^\infty f(n)$ converges.

28.2. Estimating the remainder

It is often important to understand how quickly a series converges to its limit S, by estimating the *remainder*

$$R_n := S - S_n = \sum_{k=1}^{\infty} a_k - \sum_{k=1}^{n} a_k = \lim_{N\to\infty} \sum_{k=1}^{N} a_k - \sum_{k=1}^{n} a_k = \lim_{N\to\infty} \sum_{k=n+1}^{N} a_k = \sum_{k=n+1}^{\infty} a_k.$$

Proposition 28.6. *Given a series $\sum a_n$ and a natural number $n \in \mathbb{N}$, suppose that $f\colon [n,\infty) \to [0,\infty)$ is a continuous, nonnegative, nonincreasing function such that $f(k) = a_k$ for all $k \in \mathbb{N}$. Then for every $N > n$, we have*

$$\int_{n+1}^{N+1} f(x)\,dx \leq \sum_{k=n+1}^{N} a_k \leq \int_n^N f(x)\,dx. \tag{28.1}$$

In particular, $\sum_{k=1}^\infty a_k$ converges if and only if the improper integral $\int_n^\infty f(x)\,dx$ converges, and in this case we have

$$\int_{n+1}^{\infty} f(x)\,dx \leq \sum_{k=n+1}^{\infty} a_k \leq \int_n^\infty f(x)\,dx. \tag{28.2}$$

Proof. For the lower bound in both cases, we let $g(x) = f(\lfloor x \rfloor)$, so that $g(x) = a_k$ for all $x \in [k, k+1)$. Then $g(x) \geq f(x)$ for all x because x is nondecreasing, and thus

$$\sum_{k=n+1}^{N} a_k = \int_{n+1}^{N+1} g(x)\,dx \geq \int_{n+1}^{N+1} f(x)\,dx.$$

Since $f \geq 0$, the only way for the improper integral to diverge is if it goes to ∞, and thus a divergent improper integral leads to a divergent sequence of partial sums, which proves one half of the claim following (28.1). For the other inequality, and the other half of the claim, let $g(x) = f(\lceil x \rceil)$, so that $g(x) = a_k$ for all $x \in (k-1, k]$, and observe that $g(x) \leq f(x)$ for all x because x is nondecreasing. Thus

$$\sum_{k=n+1}^{N} a_k = \int_{n}^{N} g(x)\,dx \leq \int_{n}^{N} f(x)\,dx.$$

If the improper integral converges, then since $f \geq 0$ we have $\int_n^N f(x)\,dx < \int_n^\infty f(x)\,dx$ for all N, and thus the partial sums $\sum_{k=n+1}^N a_k$ form a nondecreasing sequence that is bounded above, which implies convergence. In this case the estimates in (28.2) follow from (28.1) by taking a limit as $N \to \infty$. □

Observe that the integral test as formulated in Theorem 28.2 is a specific case of this proposition.

Example 28.7. Consider the series $\sum \frac{1}{n^2}$, and suppose we wish to find how many terms it takes for the partial sum to get within $\frac{1}{100}$ of the limit. Using (28.2) we see that

$$R_n \leq \int_n^\infty \frac{1}{x^2}\,dx = \lim_{t\to\infty} \Big[-\frac{1}{x} \Big]_n^t = \lim_{t\to\infty} \Big(\frac{1}{n} - \frac{1}{t} \Big) = \frac{1}{n}.$$

Thus we get the desired error estimate when $\frac{1}{n} \leq \frac{1}{100}$, so we need to take $n = 100$ terms.

Note that we could get a better approximation to the limit in Example 28.7 by adding one of the integrals from (28.2) to the partial sum. Indeed, under the assumptions of Proposition 28.6, the quantity $\big(\sum_{k=1}^n a_k\big) + \int_n^\infty f(x)\,dx$ is generally a better approximation to $\sum_{k=1}^\infty a_k$ than the partial sum is on its own, and we can bound the error between the approximation and the true value as follows:

$$\begin{aligned}\Big| \sum_{k=1}^\infty a_k - \Big(\sum_{k=1}^n a_k + \int_n^\infty f(x)\,dx \Big) \Big| &= \Big| \sum_{k=n+1}^\infty a_k - \int_n^\infty f(x)\,dx \Big| \\ &\leq \Big| \int_{n+1}^\infty f(x)\,dx - \int_n^\infty f(x)\,dx \Big| = \int_n^{n+1} f(x)\,dx.\end{aligned}$$

In Example 28.7, we see that this error bound is equal to

$$\int_n^{n+1} f(x)\,dx = \int_n^{n+1} \frac{1}{x^2}\,dx = -\frac{1}{x}\Big|_n^{n+1} = \frac{1}{n} - \frac{1}{n+1} = \frac{1}{n(n+1)} < \frac{1}{n^2},$$

and so to get within $\frac{1}{100}$ of the limit we could use $n = 10$ and then add the improper integral (which we can calculate explicitly in this case).

Lecture 29 Comparison tests and alternating series

Stewart §11.4 and §11.5, Spivak Ch. 23

29.1. Comparison tests

We know that the geometric series $\sum_{n=1}^{\infty} \frac{1}{2^n}$ converges. What about $\sum_{n=1}^{\infty} \frac{1}{2^n+1}$? This is not a geometric series but looks similar enough that we might expect similar convergence behavior. And indeed, if we compare the partial sums $S_n = \sum_{k=1}^{n} \frac{1}{2^k+1}$ to the partial sums $T_n = \sum_{k=1}^{n} \frac{1}{2^k}$, we can observe that the inequality $\frac{1}{2^k+1} \leq \frac{1}{2^k}$ immediately implies that

$$S_n = \sum_{k=1}^{n} \frac{1}{2^k+1} \leq \sum_{k=1}^{n} \frac{1}{2^k} = T_n \leq T := \lim_{n\to\infty} T_n,$$

where the inequality $T_n \leq T$ uses the fact that the terms $\frac{1}{2^n}$ are nonnegative so the sequence of partial sums T_n is nondecreasing. The partial sums S_n are nondecreasing for the same reason, and they are bounded above by T, so the monotone convergence theorem implies that $\sum \frac{1}{2^n+1}$ is convergent.

This argument is worth codifying. Note the analogy between the following result and Theorem 10.14 for improper integrals.

Theorem 29.1 (Comparison test)**.** *Consider two series $\sum a_n$ and $\sum b_n$ whose terms are all nonnegative: $a_n, b_n \geq 0$.*

(1) If $\sum b_n$ is convergent and $a_n \leq b_n$ for all n, then $\sum a_n$ is convergent.
(2) If $\sum b_n$ is divergent and $a_n \geq b_n$ for all n, then $\sum a_n$ is divergent.

Proof. Consider the partial sums $S_n = \sum_{k=1}^{n} a_k$ and $T_n = \sum_{k=1}^{n} b_k$. For the first claim, we have $T = \lim_{n\to\infty} T_n$, so $S_n \leq T_n \leq T$ for all n, and thus S_n is a bounded monotonic sequence, which therefore converges. The second half of the theorem follows from the first half by taking a contrapositive and reversing the roles of a_n and b_n. □

Exercise 29.2. Prove that the theorem remains true if the inequalities are only assumed to hold for all *sufficiently large* n. That is, in part (1) we can replace the assumption that $a_n \leq b_n$ for all n with the assumption that there exists $N \in \mathbb{N}$ such that $a_n \leq b_n$ for all $n \geq N$, and similarly in part (2).

Example 29.3. The series $\sum \frac{\ln n}{n}$ has $\frac{\ln n}{n} \geq \frac{1}{n}$ for all $n \geq 3$; since the harmonic series $\sum \frac{1}{n}$ diverges, the series $\sum \frac{\ln n}{n}$ diverges as well.

Example 29.4. Consider the series $\sum \frac{\ln n}{n^2}$. We saw in Example 28.4 that this converges by the integral test. We can also prove this using the comparison test. Recall that $\lim_{n\to\infty} \frac{\ln n}{\sqrt{n}} = 0$, and thus $\ln n \leq \sqrt{n}$ for all sufficiently large n. For all such n we have

$$\frac{\ln n}{n^2} \leq \frac{\sqrt{n}}{n^2} = n^{\frac{1}{2}-2} = n^{-3/2} = \frac{1}{n^{3/2}}.$$

The p-series $\sum 1/n^{3/2}$ is convergent (since $\frac{3}{2} > 1$), so the comparison test shows that $\sum \frac{\ln n}{n^2}$ is convergent as well.

As these examples show, it is often the case that a series can be compared to either a p-series or a geometric series, and so these are usually the first candidates that you should consider.

Remark 29.5. As in Proposition 28.6 and (28.2), one can use the comparison test to estimate the remainder in a convergent sum: if $a_n \leq b_n$ for all $n \geq N$, then the remainder term for $\sum a_n$ is bounded above the the remainder term for $\sum b_n$.

Sometimes it is easier to compare two sequences asymptotically than it is to go term-by-term. The following result shows that this is enough to study convergence.

Theorem 29.6 (Limit comparison test)**.** *Consider two series $\sum a_n$ and $\sum b_n$ whose terms are all positive: $a_n, b_n > 0$. Suppose that there is a real number $c > 0$ such that $\lim_{n\to\infty} \frac{a_n}{b_n} = c$. Then $\sum a_n$ is convergent if and only if $\sum b_n$ is convergent.*

Proof. By the assumption that $\frac{a_n}{b_n} \to c > 0$, there exists $N \in \mathbb{N}$ such that for all $n \geq N$ we have $\frac{c}{2} < \frac{a_n}{b_n} < 2c$, or equivalently, $\frac{c}{2} b_n < a_n < 2cb_n$. By Theorem 27.14, $\sum \frac{c}{2} b_n$ and $\sum 2cb_n$ converge if and only if $\sum b_n$ converges. Thus convergence of $\sum b_n$ implies convergence of $\sum 2cb_n$, and hence convergence of $\sum a_n$ by the comparison test. Similarly, convergence of $\sum a_n$ implies convergence of $\sum \frac{c}{2} b_n$ by the comparison test, and hence convergence of $\sum b_n$. □

Example 29.7. The series $\sum \frac{1}{2^n - 1}$ converges by applying the limit comparison test with the reference series $\sum \frac{1}{2^n}$, which is a convergent geometric series: observe that

$$\lim_{n\to\infty} \frac{1/(2^n - 1)}{1/2^n} = \lim_{n\to\infty} \frac{1}{1 - 2^{-n}} = 1.$$

Example 29.8. Consider the series

$$\sum_{n=1}^{\infty} \frac{2n^2 + 3n}{\sqrt{5 + n^5}}.$$

To determine what series to compare this to, observe that for large n we have

$$\frac{2n^2 + 3n}{\sqrt{5 + n^5}} \approx \frac{2n^2}{n^{5/2}} = \frac{2}{n^{1/2}}.$$

Since the p-series $\sum n^{-1/2}$ is divergent, we can prove that $\sum \frac{2n^2+3n}{\sqrt{5+n^5}}$ is divergent by observing that

$$\lim_{n\to\infty} \frac{(2n^2 + 3n)/(\sqrt{5 + n^5})}{n^{-1/2}} = \lim_{n\to\infty} \frac{2n^2 + 3n}{n^{-1/2}\sqrt{5 + n^5}} \cdot \frac{n^{-2}}{n^{-2}} = \lim_{n\to\infty} \frac{2 + 3n^{-1}}{\sqrt{5n^{-5} + 1}} = 2,$$

and then applying the limit comparison test.

29.2. Alternating series

The integral test and comparison tests only apply to series with nonnegative entries. It is also sometimes important to understand series with both positive and negative entries. The simplest class of series like this is the following.

Definition 29.9. A series $\sum a_n$ is *alternating* if its terms alternate between positive and negative, so that writing $b_n = |a_n|$, we have $a_n = (-1)^n b_n$ for every n. We also call the series alternating if we have $a_n = (-1)^{n-1} b_n$ for every n.

The following theorem says that for alternating series, the condition in Theorem 27.9 is actually sufficient for convergence of the series, in sharp distinction to what happens for more general series.

Theorem 29.10. *If $\sum a_n$ is an alternating series for which $b_n = |a_n|$ is a nonincreasing sequence ($b_{n+1} \le b_n$ for all n) that converges to 0 ($\lim_{n\to\infty} b_n = 0$), then $\sum a_n$ converges.*

Proof. Suppose that $a_n = (-1)^{n-1} b_n$ (the case with $(-1)^n$ is similar), so that the series is

$$b_1 - b_2 + b_3 - b_4 + b_5 - b_6 + \cdots .$$

Then the even partial sums satisfy

$$S_{2n+2} = S_{2n} + a_{2n+1} + a_{2n+2} = S_{2n} + b_{2n+1} - b_{2n+2} \ge S_{2n},$$

where the last inequality uses the fact that $b_{2n+2} \le b_{2n}$. This shows that the sequence of even partial sums is nondecreasing. Moreover, for every n we have

$$\begin{aligned} S_{2n} &= b_1 - b_2 + b_3 - b_4 + b_5 - \cdots + b_{2n-1} - b_{2n} \\ &= b_1 - (b_2 - b_3) - (b_4 - b_5) - \cdots - (b_{2n-2} - b_{2n-1}) - b_{2n} \le b_1, \end{aligned} \tag{29.1}$$

where the last inequality uses the fact that each term in brackets is nonnegative (since b_k is nonincreasing). By the monotone convergence theorem, the sequence of even partial sums S_{2n} converges to some limit S. Since $b_k \to 0$, we also have

$$\lim_{n\to\infty} S_{2n+1} = \lim_{n\to\infty} (S_{2n} + b_{2n+1}) = \lim_{n\to\infty} S_{2n} + \lim_{n\to\infty} b_{2n+1} = S + 0 = S.$$

We leave it as an exercise to show that the two results $\lim_{n\to\infty} S_{2n} = S$ and $\lim_{n\to\infty} S_{2n+1} = S$ together imply $\lim_{n\to\infty} S_n = S$, which completes the proof. □

Example 29.11. Although the harmonic series $\sum \frac{1}{n}$ diverges, the *alternating* harmonic series $\sum \frac{(-1)^{n-1}}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \cdots$ converges.

Theorem 29.12. *Under the conditions of Theorem 29.10, if $S = \sum_{n=1}^{\infty} a_n$, then the remainder term $R_n = S - S_n$ satisfies $|R_n| \le |a_{n+1}|$ for all n.*

Proof. Observe that the odd partial sums and the even partial sums converge to S from different sides, so S is always between S_n and S_{n+1}. In particular,

$$|S - S_n| \le |S_{n+1} - S_n| = |a_{n+1}|.$$

□

Lecture 30 Absolute convergence, ratio and root tests

Stewart §11.6, Spivak Ch. 23

30.1. Absolute convergence

Now that we are discussing series with both positive and negative terms, the following definition becomes important.

Definition 30.1. A series $\sum a_n$ is *absolutely convergent* if $\sum |a_n|$ is convergent.

Theorem 30.2. *If $\sum a_n$ is absolutely convergent, then it is convergent.*

Proof. For every n, we have $0 \le a_n + |a_n| \le 2|a_n|$, so $\sum(a_n + |a_n|)$ is convergent by the comparison test. Then $\sum a_n = \sum((a_n + |a_n|) - |a_n|)$ is convergent by Theorem 27.14 as the difference of two convergent series. □

Definition 30.3. A series is *conditionally convergent* if it is convergent, but not absolutely convergent.

Example 30.4. The alternating series $\sum \frac{(-1)^{n-1}}{n^2}$ is absolutely convergent, while $\sum \frac{(-1)^{n-1}}{n}$ is only conditionally convergent.

Example 30.5. $\sum \frac{\cos n}{n^2}$ is absolutely convergent by the comparison test, since $|\frac{\cos n}{n^2}| \le \frac{1}{n^2}$ and the series $\sum \frac{1}{n^2}$ is convergent.

The crucial difference between absolute and conditional convergence is the way in which rearrangements of a series behave. We are used to the idea that rearranging the terms in a sum does not change its value. The following two exercises ask you to prove that this continues to be true for an absolutely convergent infinite series.

Exercise 30.6. Let $\sum a_n$ be a series whose terms are all nonnegative ($a_n \ge 0$). Prove that the value of the infinite sum is the supremum of all the partial sums, and use this fact to deduce that $\sum a_n$ remains the same no matter what order the terms of the series are written in.

Exercise 30.7. Show that given any series $\sum a_n$, there are sequences $b_n, c_n \ge 0$ such that the $a_n = b_n - c_n$ for all n. (Hint: one of b_n, c_n should be $|a_n|$, and the other should be 0.) Then show that $\sum a_n$ is absolutely convergent if and only if $\sum b_n$ and $\sum c_n$ are both convergent, and use the previous exercise to deduce that for an absolutely convergent series, the value of the infinite sum is unchanged by rearranging the terms of the series.

When a series is only conditionally convergent, the story changes dramatically, as the following example illustrates: let S be the sum of the alternating harmonic series, so

$$S = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \frac{1}{7} - \frac{1}{8} + \frac{1}{9} - \frac{1}{10} + \cdots .$$

Multiplying every term by $\frac{1}{2}$, Theorem 27.14 gives

$$\frac{1}{2}S = 0 + \frac{1}{2} + 0 - \frac{1}{4} + 0 + \frac{1}{6} + 0 - \frac{1}{8} + 0 + \frac{1}{10} + \cdots .$$

Adding these two sequences and using Theorem 27.14 again gives

$$\frac{3}{2}S = 1 + 0 + \frac{1}{3} - \frac{1}{2} + \frac{1}{5} + 0 + \frac{1}{7} - \frac{1}{4} + \frac{1}{9} + 0 + \cdots .$$

Observe that every odd-numbered term in this last series is the same as the corresponding term in the first series, so all of the terms $\frac{1}{2n+1}$ appear exactly once. Every term $-\frac{1}{2n}$ appears exactly once in the last series also, but they are 'stretched out' further by placing 0's in between. Thus this last series is a rearrangement of the first one, and both series are convergent, but their sums are different!

In fact, the story is even stranger; it is possible to show that if $\sum a_n$ is a conditionally convergent series, then for *any* $r \in \mathbb{R}$ there is a conditionally convergent series $\sum b_n$ with sum r that is a rearrangement of $\sum a_n$. Thus conditionally convergent series must be treated with a certain amount of caution.

30.2. Ratio test

In light of the previous discussion, it is useful to be able to determine when a series is absolutely convergent.

Theorem 30.8 (Ratio test)**.** *Consider a series $\sum a_n$ with nonzero terms, and let $L = \lim_{n\to\infty} |\frac{a_{n+1}}{a_n}|$ if the limit exists.*

(1) If $L < 1$, then $\sum a_n$ is absolutely convergent.
(2) If $L > 1$, or if the limit is ∞, then $\sum a_n$ is divergent.
(3) If $L = 1$, or if the limit does not exist, then the ratio test is inconclusive and gives no information.

Proof. For the first part, choose $r \in (L, 1)$; then there is $N \in \mathbb{N}$ such that $|\frac{a_{n+1}}{a_n}| < r$ for all $n \geq N$. This gives $|a_{n+1}| < r|a_n|$, and iterating gives $|a_{N+k}| < |a_N|r^k$ for all $k \geq 1$. Since $\sum_{k=1}^{\infty} |a_N|r^k$ is convergent (a geometric series with $|r| < 1$), the comparison test implies that $\sum |a_n|$ is convergent as well.

For the second part, $L > 1$ implies that there is N such that $|a_{n+1}| > |a_n|$ for all $n \geq N$, so $\lim a_n \neq 0$, and by Corollary 27.12 the series is divergent. □

Example 30.9. Consider the series $\sum(-1)^n \frac{n^3}{2^n}$. The limit in the ratio test is

$$L = \lim_{n\to\infty} \frac{(n+1)^3/2^{n+1}}{n^3/2^n} = \lim_{n\to\infty} \left(1 + \frac{1}{n}\right)^3 \cdot \frac{1}{2} = \frac{1}{2} < 1,$$

so the series is absolutely convergent.

Example 30.10. The series $\sum \frac{2^n}{n!}$ is absolutely convergent because

$$\lim_{n\to\infty} \frac{2^{n+1}/(n+1)!}{2^n/n!} = \lim_{n\to\infty} \frac{2}{n+1} = 0.$$

Observe that the ratio test does not catch all convergent series. Indeed, although $\sum \frac{1}{n^2}$ is convergent, the ratio test gives

$$L = \lim_{n\to\infty} \frac{1/(n+1)^2}{1/n^2} = \lim_{n\to\infty} \frac{n^2}{(n+1)^2} = \lim_{n\to\infty} \frac{1}{(1+\frac{1}{n})^2} = 1,$$

and thus is inconclusive.

30.3. Root test

Theorem 30.11 (Root test)**.** *Consider a series $\sum a_n$, and let $L = \lim_{n\to\infty} \sqrt[n]{|a_n|}$ if the limit exists.*

1. *If $L < 1$, then $\sum a_n$ is absolutely convergent.*
2. *If $L > 1$, or if the limit is ∞, then $\sum a_n$ is divergent.*
3. *If $L = 1$, or if the limit does not exist, then the ratio test is inconclusive and gives no information.*

Proof. In the first case, once again choose $r \in (L, 1)$, so there is $N \in \mathbb{N}$ such that for all $n \geq N$, we have $\sqrt[n]{|a_n|} = |a_n|^{1/n} < r$. Raising both sides to the nth power gives $|a_n| < r^n$, and since $\sum r^n$ converges, the comparison test implies that $\sum |a_n|$ converges as well. In the second case, a similar argument shows that $|a_n| \to \infty$, so $\sum a_n$ diverges. □

Example 30.12. The series $\sum(\frac{n}{2n+1})^n$ is absolutely convergent, because

$$\lim_{n\to\infty} \sqrt[n]{\left(\frac{n}{2n+1}\right)^n} = \lim_{n\to\infty} \frac{n}{2n+1} = \frac{1}{2} < 1.$$

Although the ratio test would work in this example, it would be rather messier to carry out the computations. (Try it!) In some other cases, the root test may work where the ratio test fails.

Exercise 30.13. Let $a_n = \frac{1}{3^n}$ when n is odd, and $a_n = \frac{2}{3^n}$ when n is even. Use the root test to prove that $\sum a_n$ converges. Show that the limit in the ratio test does not exist.

On the other hand, if the ratio test fails because the limit is equal to 1, then the root test will not work either.

Exercise 30.14. Prove that if the limit in the ratio test exists, then the limit in the root test exists as well, and the two limits are the same.

This last exercise implies that whenever the ratio test works, the root test would also work, although the computations might be harder (they could also be easier). If $L = 1$ in either the ratio test or the root test, then neither of the tests will determine convergence.[19]

Lecture 31 Power series

Stewart §11.8, Spivak Ch. 24

Definition 31.1. A *power series* is a series of the form $\sum_{n=0}^{\infty} c_n x^n$, where $c_n \in \mathbb{R}$ are constants, called the *coefficients* of the power series, and $x \in \mathbb{R}$ is a variable. For a given value of x, a power series becomes a series in the sense we have been studying so far,

[19] If you read the preceding passage carefully, though, you will see that it is possible to have $L = 1$ in the root test while the limit in the ratio test does not exist.

and can be either convergent or divergent. The *domain* of the power series is the set of all x for which the power series converges. When x lies in this domain, we write

$$f(x) = \sum_{n=0}^{\infty} c_n x^n = c_0 + c_1 x + c_2 x^2 + c_3 x^3 + \cdots$$

for the function determined by the power series.

Example 31.2. We already saw from the formula for the sum of a geometric series that the function $\frac{1}{1-x}$ can be represented by the power series

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n = 1 + x + x^2 + x^3 + \cdots ,$$

where the coefficients are $c_n = 1$, and the series converges iff $|x| < 1$.

It is possible for the domain to be all of $\mathbb{R}$.

Example 31.3. Let $f(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}$. Then for all $x \in \mathbb{R}$, the terms $a_n = \frac{x^n}{n!}$ satisfy

$$\left|\frac{a_{n+1}}{a_n}\right| = \left|\frac{x^{n+1}/(n+1)!}{x^n/n!}\right| = \left|\frac{x}{n+1}\right| \to 0 \text{ as } n \to \infty$$

and thus the series is convergent by the ratio test.[20]

It is also possible for the domain to be a single point.

Example 31.4. Let $f(x) = \sum_{n=0}^{\infty} n! x^n$. Then the series converges for $x = 0$ because all terms are 0, but for $x \neq 0$ the terms $a_n = n! x^n$ satisfy

$$\left|\frac{a_{n+1}}{a_n}\right| = \left|\frac{(n+1)! x^{n+1}}{n! x^n}\right| = |(n+1)x| \to \infty \text{ as } n \to \infty$$

and thus the series is divergent by the ratio test.

One can also center the series at a point $a \neq 0$ by replacing x^n with $(x-a)^n$.

Example 31.5. Consider the series

$$\sum_{n=1}^{\infty} \frac{(x-2)^n}{n}.$$

For a given $x \in \mathbb{R}$, the ratio of successive terms (in absolute value) is

$$\left|\frac{(x-2)^{n+1}/(n+1)}{(x-2)^n/n}\right| = \left|\frac{(x-2)n}{n+1}\right| \to |x-2|.$$

Thus by the root test the series converges when $|x-2| < 1$ (that is, when $1 < x < 3$), and diverges when $|x-2| > 1$. At $x = 1$ it converges (alternating harmonic series) and at $x = 3$ it diverges (harmonic series).

Theorem 31.6. *Given a power series $\sum_{n=0}^{\infty} c_n (x-a)^n$, one of the following three things happens.*

(1) The series converges when $x = a$ and diverges for all $x \neq a$.

[20] Compare this to Example 30.10, which did this same calculation in the case $x = 2$.

(2) *The series converges absolutely for all $x \in \mathbb{R}$.*
(3) *There exists $R > 0$ such that the series converges absolutely when $|x - a| < R$ and diverges when $|x - a| > R$.*

Definition 31.7. The number R from Theorem 31.6 is called the *radius of convergence* of the power series.

Before proving Theorem 31.6, we observe that while this result gives absolute convergence in the interior of the interval, it is silent on what happens at the endpoints, where we can have either convergence or divergence.

Example 31.8. Fixing any $p \in \mathbb{R}$, we see that the power series $\sum_{n=0}^{\infty} n^{-p} x^n$ has the property that

$$\left| \frac{(n+1)^{-p} x^{n+1}}{n^{-p} x^n} \right| = \left| \left(1 + \frac{1}{n}\right)^{-p} x \right| \to |x| \text{ as } n \to \infty,$$

and thus by the ratio test its radius of convergence is $R = 1$. The behavior at the endpoints $x = \pm 1$ depends on p.

(1) For $p = 0$, $\sum x^n$ diverges at both endpoints.
(2) For $p = 1$, $\sum \frac{x^n}{n}$ converges conditionally when $x = -1$, and diverges when $x = 1$.
(3) For $p = 2$, $\sum \frac{x^n}{n^2}$ converges absolutely at both endpoints.

The following exercises illustrate the remaining possible behaviors at the endpoints.

Exercise 31.9. Prove that $\sum(-2x)^n/n$ has radius of convergence $1/2$, converges conditionally at $x = 1/2$, and diverges at $x = -1/2$.

Exercise 31.10. Prove that the power series

$$x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \frac{x^9}{9} - \cdots$$

has radius of convergence 1 and converges conditionally at both endpoints.

Exercise 31.11. Prove that if a power series converges absolutely at one endpoint of its interval of convergence, then it converges absolutely at the other endpoint as well.

Now we prove Theorem 31.6, starting with a lemma.

Lemma 31.12. *Let c_n be a sequence of coefficients and $r \in \mathbb{R}$ a real number such that $\sum c_n r^n$ converges. Then given any $a, x \in \mathbb{R}$ with $|x - a| < |r|$, the series $\sum c_n (x - a)^n$ converges absolutely.*

Proof. Convergence of $\sum c_n r^n$ implies that $c_n r^n \to 0$, so there is $N \in \mathbb{N}$ such that $|c_n r^n| < 1$ for all $n \geq N$. For such n we then have

$$|c_n (x-a)^n| = |c_n r^n| \cdot \left| \frac{(x-a)^n}{r^n} \right| < \left| \frac{x-a}{r} \right|^n,$$

and thus $\sum |c_n (x-a)^n|$ converges by the comparison test, using the geometric series $\sum |\frac{x-a}{r}|^n$, which is convergent because $|x - a| < |r|$. □

Proof of Theorem 31.6. Consider the set $A = \{r \geq 0 : \sum c_n r^n \text{ converges}\}$. This is nonempty because $0 \in A$. If it is unbounded then for every $x \in \mathbb{R}$ there exists $r \in A$ such that $r > |x-a|$, and thus by Lemma 31.12, the series $\sum c_n(x-a)^n$ converges at x. This puts us in case (2).

Now suppose that A is bounded, and let $R = \sup A$ be its least upper bound. Then for every $x \in \mathbb{R}$ with $|x-a| > R$, we see that $\sum c_n(x-a)^n$ diverges, otherwise Lemma 31.12 would imply that $\sum c_n r^n$ converges for some $r \in (R, |x-a|)$, contradicting the claim that R is an upper bound for A. If $R = 0$, then we are in case (1); the series diverges for all $x \neq a$. If $R > 0$, then for every x with $|x-a| < R$ we can choose $r \in A$ with $|x-a| < r$ (since R is the *least* upper bound) and use Lemma 31.12 to deduce that $\sum c_n(x-a)^n$ converges absolutely. This puts us in case (3). □

Theorem 31.6 tells us that every power series converges on an interval (which could be a single point, or all of $\mathbb{R}$), and diverges on its complement. The radius of convergence can often – but not always – be determined by using either the ratio test or the root text.

Theorem 31.13 (Ratio test for radius of convergence)**.** *Suppose $\sum c_n(x-a)^n$ is a power series for which the limit $L = \lim_{n\to\infty} |c_{n+1}/c_n|$ exists. Then the radius of convergence is $R = 1/L$. If $L = 0$ then the radius of convergence is ∞; if $L = \infty$ then the radius of convergence is* 0.

Proof. We apply the ratio test. Given $x \in \mathbb{R}$, we have

$$\lim_{n\to\infty} \left| \frac{c_{n+1}(x-a)^{n+1}}{c_n(x-a)^n} \right| = L|x-a|.$$

This is < 1 when $|x-a| < 1/L$, implying absolute convergence, and > 1 when $|x-a| > 1/L$, implying divergence. If $L = 0$ then the limit is always 0, giving absolute convergence, and if $L = \infty$ then the limit is ∞ for all $x \neq a$, giving divergence. □

Theorem 31.14 (Root test for radius of convergence)**.** *Suppose $\sum c_n(x-a)^n$ is a power series for which the limit $L = \lim_{n\to\infty} |c_n|^{1/n}$ exists. Then the radius of convergence is $R = 1/L$. If $L = 0$ then the radius of convergence is ∞; if $L = \infty$ then the radius of convergence is* 0.

Proof. This is exactly the same as the previous proof except we use the following computation:

$$\lim_{n\to\infty} |c_n(x-a)^n|^{1/n} = |x-a| \lim_{n\to\infty} |c_n|^{1/n} = L|x-a|. \qquad \square$$

Lecture 32 Calculus with power series

Stewart §11.9, Spivak Ch. 24

Of the various elementary functions that we have encountered so far, polynomials are among the easiest to work with; they can be added, subtracted, and multiplied relatively easily, and differentiation and integration are also straightforward using the rules

$$\frac{d}{dx}x^n = nx^{n-1} \quad \text{and} \quad \int x^n\,dx = \frac{x^{n+1}}{n+1} + C.$$

Since polynomials are finite sums of expressions like these, they can be manipulated without incident. For power series, which are *infinite* sums, more care is needed, as indicated by the results about rearrangements of conditionally convergent series. The following theorem says that differentation and integration work as expected.

Theorem 32.1. *Let $\sum_{n=0}^\infty c_n(x-a)^n$ be a power series with radius of convergence $R > 0$, and let $f\colon (a-R, a+R) \to \mathbb{R}$ be the function defined by $f(x) = \sum_{n=0}^\infty c_n(x-a)^n$. Then f is continuously differentiable on this interval, and therefore also integrable; moreover, we can represent f' and $\int f$ by the following power series:*

$$f'(x) = \sum_{n=1}^\infty nc_n(x-a)^{n-1},$$
$$\int f(x)\,dx = C + \sum_{n=0}^\infty c_n \frac{(x-a)^{n+1}}{n+1}.$$

Both of these power series also have radius of convergence R.

Proof. The proof that these power series have the same radius of convergence R can be given by a mild modification of Lemma 31.12, which we leave as an exercise.

The proof that they actually give the derivative and integral of f is more difficult and was omitted in the lecture.[21] For simplicity we prove the result for the derivative, with $a = 0$. The result for the integral is a corollary, and the proof for other values of a is the same; one simply needs to replace x with $x - a$ everywhere that it appears. Thus we need to show that if we define two functions $f, g\colon (-R, R) \to \mathbb{R}$ by the power series

$$f(x) = \sum_{n=0}^\infty c_n x^n \quad \text{and} \quad g(x) = \sum_{n=1}^\infty nc_n x^{n-1},$$

then $f'(x) = g(x)$ for all $x \in (-R, R)$. By definition of the derivative, we have

$$f'(x) = \lim_{y\to x} \frac{1}{y-x} \sum_{n=1}^\infty c_n(y^n - x^n) = \lim_{y\to x} \sum_{n=1}^\infty c_n(y^{n-1} + xy^{n-2} + x^2y^{n-3} + \cdots + x^{n-2}y + x^{n-1}),$$

where we used the factorization $y^n - x^n = (y-x)(y^{n-1} + xy^{n-2} + \cdots + x^{n-1})$. The expression in brackets can be written as $\sum_{j=0}^{n-1} y^j x^{n-1-j}$, and we can write nx^{n-1} as $\sum_{j=0}^{n-1} x^{n-1}$, so we conclude that

$$f'(x) - g(x) = \lim_{y\to x} \sum_{n=1}^\infty c_n \Big(\sum_{j=0}^{n-1} (y^j x^{n-1-j} - x^{n-1}) \Big). \tag{32.1}$$

We must show that this quantity vanishes. It is tempting to move the limit inside the sums and observe that $\lim_{y\to x} y^j x^{n-1-j} - x^{n-1} = 0$; however, while this would be allowed

[21] I learned the proof here from a blog post by Tim Gowers at `https://gowers.wordpress.com/2014/02/22/differentiating-power-series/` – the "more common" proof of this theorem uses the concept of *uniform convergence* of a sequence of functions, which is beyond the scope of this course.

if the sums were both finite, it is *not* always allowed for infinite sums. Indeed, the infinite sum is itself a limit, and we could more properly write the above equation as

$$f'(x) - g(x) = \lim_{y\to x} \lim_{N\to\infty} \sum_{n=1}^{N} c_n \Big(\sum_{j=0}^{n-1} (y^j x^{n-1-j} - x^{n-1}) \Big).$$

Thus before we can pass $\lim_{y\to x}$ inside the sums, we would need to interchange the order of the limits. This is an issue that has not arisen for us so far, and that we will not treat in any detail, save by issuing this warning: be very, very careful if anyone tries to sell you a computation in which the order of two limits are interchanged. Sometimes it is valid, and sometimes it is not; in this course we have not developed the tools to tell the difference.

Instead, we will estimate the magnitude of the inner sum as follows:

$$\Big| \sum_{j=0}^{n-1} (y^j x^{n-1-j} - x^{n-1}) \Big| \le \sum_{j=0}^{n-1} |x|^{n-1-j} |y^j - x^j| = \sum_{j=0}^{n-1} |x|^{n-1-j} |y - x| \Big| \sum_{i=0}^{j-1} y^i x^{j-1-i} \Big|.$$

Here the last equality once again uses the factorization for $y^j - x^j$ that we used before. Recalling that R is the radius of convergence of the power series and $|x| < R$, fix r such that $|x| < r < R$, and choose y close enough to x that $|y| < r$. Then we have $|\sum_{i=0}^{j-1} y^i x^{j-1-i}| \le \sum_{i=0}^{j-1} r^{j-1} = jr^{j-1}$, and the above estimate gives

$$\Big| \sum_{j=0}^{n-1} (y^j x^{n-1-j} - x^{n-1}) \Big| \le \sum_{j=0}^{n-1} r^{n-1-j} |y-x| \cdot j r^{j-1} = \sum_{j=0}^{n-1} j r^{n-2} |y-x| = \frac{n(n-1)}{2} r^{n-2} |y-x|.$$

Returning to the expressions in (32.1), we see that

$$\Big| \sum_{n=1}^{\infty} c_n \Big(\sum_{j=0}^{n-1} (y^j x^{n-1-j} - x^{n-1}) \Big) \Big| \le \sum_{n=1}^{\infty} c_n \cdot \frac{n(n-1)}{2} r^{n-2} |y - x|$$

and thus

$$|f'(x) - g(x)| \le \lim_{y\to x} |y - x| \sum_{n=1}^{\infty} \frac{n(n-1)}{2} c_n r^{n-2} = 0,$$

provided the last sum converges. The fact that it converges for $r \in (0, R)$ is a consequence of the exercise at the beginning of this proof, since this is the power series that “should” represent $f''(r)$, and you were asked to prove in that exercise that formal term-by-term differentiation does not change the radius of convergence. □

Example 32.2. We have seen that on $(-1, 1)$, the function $f(x) = \frac{1}{1-x}$ is represented by the power series

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n = 1 + x + x^2 + x^3 + \cdots . \tag{32.2}$$

Differentiating and integrating term-by-term, as in Theorem 32.1, we see that

$$f'(x) = \sum_{n=1}^{\infty} n x^{n-1} \quad \text{and} \quad \int f(x)\, dx = C + \sum_{n=0}^{\infty} \frac{x^{n+1}}{n+1}.$$

Since $f'(x) = \frac{1}{(1-x)^2}$, we obtain the new power series representation

$$\frac{1}{(1-x)^2} = \sum_{n=1}^{\infty} nx^{n-1} = 1 + 2x + 3x^2 + 4x^3 + 5x^4 + \cdots . \tag{32.3}$$

Similarly, since $\int f(x)\,dx = -\ln(1-x) + C$, we obtain

$$\ln(1-x) = C - \sum_{n=0}^{\infty} \frac{x^{n+1}}{n+1}.$$

When $x = 0$ the LHS vanishes, so the constant of integration is $C = 0$, and we have the power series representation

$$\ln(1-x) = -\sum_{n=1}^{\infty} \frac{x^n}{n} = -x - \frac{x^2}{2} - \frac{x^3}{3} - \frac{x^4}{4} - \cdots . \tag{32.4}$$

Remark 32.3. It is possible to show that (32.4) remains valid not just on the interval $(-1, 1)$, but also at the endpoint $x = -1$; see the extra credit problems on the homework for an outline of the proof of *Abel's theorem*, which establishes this fact[22] Observe that then this series gives

$$\ln 2 = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \cdots ,$$

so that $\ln 2$ is the value of the sum of the alternating harmonic series.

Example 32.4. Replacing x in (32.2) by $(-x^2)$, we obtain the power series representation

$$\frac{1}{1+x^2} = \frac{1}{1-(-x^2)} = \sum_{n=0}^{\infty}(-x^2)^n = \sum_{n=0}^{\infty}(-1)^n x^{2n} = 1 - x^2 + x^4 - x^6 + x^8 - \cdots ,$$

which is valid on the interval $(-1, 1)$. Integrating gives

$$\tan^{-1}(x) = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \cdots ,$$

where the constant of integration is 0 because $\tan^{-1}(0) = 0$. This is valid on the interval $(-1, 1)$ (though we observe that the function $\tan^{-1}(x)$ is defined on all of $\mathbb{R}$). Once again, Abel's theorem can be used to extend its validity to include $x = 1$, and we obtain the following formula:

$$\frac{\pi}{4} = \tan^{-1}(1) = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \cdots .$$

Example 32.5. The *Bessel function of order* 0 is given by the power series

$$J_0(x) = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{4^n (n!)^2}.$$

[22] At $x = 1$, the series diverges, which is consistent with the fact that $\ln 0$ is undefined.

Using the root test one can check that it converges for all $x \in \mathbb{R}$. Using Theorem 32.1 one can compute its derivative:

$$J_0'(x) = \sum_{n=1}^{\infty} \frac{(-1)^n 2n x^{2n-1}}{4^n (n!)^2}$$

Similarly one can compute J_0'', and verify that J_0 is a solution of the differential equation

$$x^2 f''(x) + x f'(x) + x^2 f(x) = 0, \tag{32.5}$$

which arises (among other places) when one studies the shape of a vibrating drumhead.

The DE in (32.5) does not admit a closed form solution in terms of functions we have studied earlier, so this last example illustrates the utility of power series as a tool. Even in situations where a problem can be solved using other methods, the power series is often easier to compute: for example, we could compute $\int \frac{1}{1+x^4}\,dx$ using partial fractions, but it is fairly long and tedious to do so, while using power series we quickly get

$$\begin{aligned}\int \frac{1}{1+x^4}\,dx &= \int (1 - x^4 + x^8 - x^{12} + x^{16} - \cdots)\,dx \\ &= C + x - \frac{x^5}{5} + \frac{x^9}{9} - \frac{x^{13}}{13} + \cdots .\end{aligned}$$

Of course it may then be difficult or impossible to translate this power series back into a closed form for the integral, but in many cases the power series is just as useful, especially if what we are after is a numerical approximation.

Lecture 33 Taylor and Maclaurin series

Stewart §11.10, Spivak Ch. 24

33.1. Obtaining coefficients from higher derivatives

It is natural to ask whether a given function can be represented by a power series, and if so, how the coefficients of that series can be found. In light of Theorem 32.1, we see that any function represented by a power series needs to be at least differentiable on the interior of the interval of convergence; thus we cannot expect to represent $f(x) = |x|$ by a power series around 0.

Moreover, since a power series representation for f gives a power series for f' with the same radius of convergence, we see that f' must be differentiable as well. Continuing this line of reasoning, every derivative $f^{(n)}$ must exist. Is this enough? It turns out that the answer is no.

Exercise 33.1. Prove that the function

$$f(x) = \begin{cases} e^{-1/x^2} & x > 0, \\ 0 & x \le 0 \end{cases}$$

has derivatives of all orders at 0, but is not given by a power series in any open interval containing 0.

A function that has derivatives of all orders is called *smooth*. A function that is given by a convergent power series is called *analytic*. Analytic functions are smooth, but not every smooth function is analytic. For the moment, we address the question of how to find the coefficients of the power series, *assuming that f is indeed represented by a power series*. To this end, suppose that near $a \in \mathbb{R}$, a function f is given by a power series

$$f(x) = c_0 + c_1(x-a) + c_2(x-a)^2 + c_3(x-a)^3 + (\text{terms containing } (x-a)^4).$$

Then its derivative is given by

$$f'(x) = c_1 + 2c_2(x-a) + 3c_3(x-a)^2 + (\text{terms containing } (x-a)^3),$$

its second derivative is given by

$$f''(x) = 2c_2 + 2 \cdot 3 \cdot c_3(x-a) + (\text{terms containing } (x-a)^2),$$

and in general, its nth derivative is given by

$$f^{(n)}(x) = n!c_n + (\text{terms containing } (x-a)).$$

Since any term containing $(x-a)$ vanishes when we put $x = a$, we conclude that

$$f(a) = c_0, \quad f'(a) = c_1, \quad f''(a) = 2c_2, \quad \ldots \quad f^{(n)}(a) = n!c_n.$$

Thus we can recover the coefficients c_n from the values of the higher derivatives of f at a. We have proved the following theorem.

Theorem 33.2. *If f has a power series representation $\sum_{n=0}^{\infty} c_n(x-a)^n$, with radius of convergence $R > 0$, then we have*

$$c_n = \frac{f^{(n)}(a)}{n!} \text{ for all } n = 0, 1, 2, \ldots.$$

Thus for every $x \in (a-R, a+R)$, we have

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n. \tag{33.1}$$

The power series in (33.1) is called the *Taylor series* for f. In the case when $a = 0$, it is also called the *Maclaurin series*:

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!}x^n.$$

Example 33.3. For the exponential function $f(x) = e^x$, we have $f^{(n)}(x) = e^x$ for every $n = 0, 1, 2, \ldots$, and thus $c_n = \frac{f^{(n)}(0)}{n!} = \frac{1}{n!}$. Thus the Maclaurin series for e^x (the Taylor series around 0) is

$$\sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots. \tag{33.2}$$

Observe that for every $x \in \mathbb{R}$, we have

$$\frac{x^{n+1}/(n+1)!}{x^n/(n!)} = \frac{x}{n+1} \to 0 \text{ as } n \to \infty,$$

so the series converges absolutely by the ratio test. Thus the radius of convergence is $R = \infty$.

What the above example does *not* immediately tell us is whether or not the power series in (33.2) actually converges to e^x. Could it converge to something else instead? We will investigate this question in the next section.

33.2. Approximation by polynomials

Definition 33.4. Given $n \in \mathbb{N}$, the nth *Taylor polynomial* for $f(x)$ around a is the nth partial sum of the Taylor series:

$$T_n(x) := \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k. \tag{33.3}$$

Thus the Taylor series converges to $f(x)$ at x if and only if $T_n(x) \to f(x)$ as $n \to \infty$. Equivalently, writing $R_n(x) = f(x) - T_n(x)$ for the nth *remainder* term at x, we have convergence if and only if $R_n(x) \to 0$. Exercise 33.1 gives an example where this does not occur. Can we give conditions under which it does occur?

Theorem 33.5 (Taylor's inequality)**.** *Let $f\colon (a-d, a+d)$ be $n+1$ times differentiable, and suppose that $|f^{(n+1)}(x)| \leq M$ for all $x \in (a-d, a+d)$. Then for every x in this interval, the nth remainder term $R_n(x) = f(x) - T_n(x)$ satisfies*

$$|R_n(x)| \leq \frac{M}{(n+1)!}|x-a|^{n+1}.$$

Proof. We prove this by induction in n when $x \in (a, a+d)$; the proof for $x \in (a-d, a)$ is similar. First consider the case $n = 0$. In this case we have $T_0(x) = f(a)$ for all x, so $R_0(x) = f(x) - f(a)$, and by assumption $|f'(x)| \leq M$ for all $x \in (a, a+d)$, so

$$|R_0(x)| = |f(x) - f(a)| = \Big| \int_a^x f'(t)\,dt \Big| \leq \int_a^x |f'(t)|\,dt \leq \int_a^x M\,dt = M(x-a).$$

Now suppose that $n \geq 1$ and that the result holds for $n-1$. Then observe that since $R_n(x) = f(x) - T_n(x)$ by definition, we have $R_n'(x) = f'(x) - T_n'(x)$. We claim that T_n' is the degree $(n-1)$ Taylor polynomial for f': indeed, differentiating (9.1) gives

$$T_n'(x) = \sum_{k=1}^{n} \frac{f^{(k)}(a)}{(k-1)!}(x-a)^{k-1} = \sum_{j=0}^{n-1} \frac{f^{(j+1)}(a)}{j!}(x-a)^j,$$

and since $f^{(j+1)}(a) = (f')^{(j)}(a)$, this proves the claim. Moreover, we have $|(f')^{(n)}(x)| = |f^{(n+1)}(x)| \leq M$ for all $x \in (a, a+d)$, so by the inductive hypothesis we obtain

$$|R_n'(x)| \leq \frac{M}{n!}|x-a|^n.$$

Integrating this gives

$$|R_n(x)| = \Big| \int_a^x R_n'(t)\,dt \Big| \leq \int_a^x |R_n'(t)|\,dt \leq \int_a^x \frac{M}{n!}(t-a)^n\,dt = \Big[\frac{M(t-a)^{n+1}}{(n+1)!}\Big]_a^x$$

This last expression is equal to $M(x-a)^{n+1}/(n+1)!$, which proves the theorem. □

Remark 33.6. In fact, one can prove the following more explicit formulas for the remainder term:

$$R_n(x) = \frac{1}{n!}\int_a^x (x-t)^n f^{(n+1)}(t)\,dt,$$
$$R_n(x) = \frac{f^{(n+1)}(t)}{(n+1)!}(x-a)^{n+1} \text{ for some } t \text{ between } x \text{ and } a.$$

Observe that each of these implies Taylor's inequality. Note also that the second of these reduces to the Mean Value Theorem in the case $n = 0$.

Returning to the case of $f(x) = e^x$, we see from Taylor's inequality that for every $|x| \leq d$ and every $n \in \mathbb{N}$, we have

$$|e^x - T_n(x)| \leq \frac{e^d}{(n+1)!}|x|^{n+1} \to 0 \text{ as } n \to \infty,$$

where the convergence to 0 follows because $\sum \frac{x^n}{n!}$ converges. This proves that the Maclaurin series $\sum \frac{x^n}{n!}$ does indeed converge to e^x, so that we can write

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} \text{ for all } x \in \mathbb{R}. \tag{33.4}$$

In particular, putting $x = 1$ gives the following infinite series for e:

$$e = \sum_{n=0}^{\infty} \frac{1}{n!} = 1 + 1 + \frac{1}{2} + \frac{1}{3!} + \frac{1}{4!} + \cdots.$$

Observe also that if we differentiate (33.4) term-by-term, we get the same power series, consistent with the fact that $\frac{d}{dx}e^x = e^x$.

Example 33.7. We could also take the Taylor series of e^x around another point; for example, with $a = 1$ we see that $f^{(n)}(a) = e^a = e^1 = e$ for all n, and thus

$$e^x = \sum_{n=0}^{\infty} \frac{e}{n!}(x-1)^n,$$

where the argument that the remainder terms go to 0 is similar to the one given above.

Example 33.8. For $f(x) = \sin x$, we have

$$f'(x) = \cos x, \quad f''(x) = -\sin x, \quad f'''(x) = -\cos x,$$

and in general,

$$f^{(n)}(x) = \begin{cases} \sin x & \text{if } n \equiv 0 \pmod 4, \\ \cos x & \text{if } n \equiv 1 \pmod 4, \\ -\sin x & \text{if } n \equiv 2 \pmod 4, \\ -\cos x & \text{if } n \equiv 3 \pmod 4. \end{cases}$$

Thus the nth derivatives at 0 are $0, 1, 0, -1, 0, 1, 0, -1, \ldots$, and the Maclaurin series is

$$x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!}.$$

Since all derivatives have absolute value ≤ 1 for every x, we can take $M = 1$ in Taylor's inequality and obtain

$$|R_n(x)| \leq \frac{|x|^{n+1}}{(n+1)!} \to 0,$$

which proves that the Maclaurin series converges to $\sin x$ for every $x \in \mathbb{R}$, and thus

$$\sin x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!} = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots . \tag{33.5}$$

By making a similar argument for $f(x) = \cos x$, or by differentiating (33.5) term-by-term and applying Theorem 32.1, we get

$$\cos x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!} = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots . \tag{33.6}$$

Remark 33.9. Using (33.4), (33.5), and (33.6), we see that writing i for a (complex) square root of -1, we have

$$\begin{aligned} e^{ix} &= 1 + (ix) + \frac{(ix)^2}{2!} + \frac{(ix)^3}{3!} + \frac{(ix)^4}{4!} + \frac{(ix)^5}{5!} + \frac{(ix)^6}{6!} + \frac{(ix)^7}{7!} + \frac{(ix)^8}{8!} + \cdots \\ &= 1 + ix - \frac{x^2}{2!} - i\frac{x^3}{3!} + \frac{x^4}{4!} + i\frac{x^5}{5!} - \frac{x^6}{6!} - i\frac{x^7}{7!} + \frac{x^8}{8!} + \cdots \\ &= \Big(1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!} + \cdots\Big) + i\Big(x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots\Big) \\ &= \cos x + i \sin x, \end{aligned}$$

where the equality in the third line uses the fact that Taylor series converge *absolutely* on the interior of the interval of convergence, and thus we can rearrange the series without changing the sum.

33.3. Binomial series

Recall from the Binomial Theorem that given a positive integer n and any real number x, we can write

$$(1+x)^n = 1 + nx + \binom{n}{2} x^2 + \cdots + \binom{n}{n-2} x^{n-2} + nx^{n-1} + x^n = \sum_{k=0}^{n} \binom{n}{k} x^k. \tag{33.7}$$

Writing $f(x) = (1+x)^n$, we see that for $0 \leq k \leq n$, we have

$$f^{(k)}(x) = \frac{d^k}{dx^k}(1+x)^n = n(n-1)\cdots(n-k+1)(1+x)^{n-k},$$

and so $f^{(k)}(0) = n(n-1)\cdots(n-k+1) = n!/(n-k)!$. When $k > n$ we have $f^{(k)}(0) = 0$, and so the Maclaurin series for $(1+x)^n$ is

$$\sum_{k=0}^{\infty} \frac{1}{k!} f^{(k)}(0) x^k = \sum_{k=0}^{n} \frac{1}{k!} \frac{n!}{(n-k)!} x^k = \sum_{k=0}^{n} \binom{n}{k} x^k,$$

which recovers the formula in (33.7). But we can compute the Maclaurin series even if n is *not* an integer; consider the function $f(x) = (1+x)^\alpha$ for an arbitrary real number α. Then for any $k \geq 0$, we have

$$f^{(k)}(x) = \frac{d^k}{dx^k}(1+x)^\alpha = \alpha(\alpha-1)\cdots(\alpha-k)(1+x)^{\alpha-k}.$$

Note that if α happens to be a positive integer, then this expression vanishes whenever $k \geq \alpha$. Evaluating at $x = 0$ gives

$$f^{(k)}(0) = \alpha(\alpha-1)\cdots(\alpha-k),$$

and so the Maclaurin series for $(1+x)^\alpha$ is

$$\sum_{k=0}^{\infty} \frac{\alpha(\alpha-1)\cdots(\alpha-k+1)}{k!}x^k. \tag{33.8}$$

Extending the notation from the integer case, we write

$$\binom{\alpha}{k} := \frac{\alpha(\alpha-1)\cdots(\alpha-k+1)}{k!} \tag{33.9}$$

and refer to these numbers as *binomial coefficients*. Once again we see that if α is a positive integer, then $\binom{\alpha}{k} = 0$ for all $k > \alpha$, while for $0 \leq k \leq \alpha$ the formula in (33.9) reduces to our usual definition of binomial coefficients $\frac{\alpha!}{k!(\alpha-k)!}$.

When α is *not* a positive integer, we can determine the radius of convergence of the power series in (33.8) by applying the ratio test (Theorem 31.13):

$$\left|\frac{\binom{\alpha}{k+1}}{\binom{\alpha}{k}}\right| = \left|\frac{\alpha(\alpha-1)\cdots(\alpha-k)}{(k+1)!}\frac{k!}{\alpha(\alpha-1)\cdots(\alpha-k+1)}\right| = \left|\frac{\alpha-k}{k+1}\right| \to 1$$

as $k \to \infty$, and thus the radius of convergence is 1. (Can you determine when it does and does not converge at the endpoints ± 1?)

Theorem 33.10. *For every $\alpha \in \mathbb{R}$ and $|x| < 1$, we have $(1+x)^\alpha = \sum_{k=0}^\infty \binom{\alpha}{k}x^k$.*

Proof. (This proof was omitted in the lecture.) Let $g(x) = \sum_{k=0}^\infty \binom{\alpha}{k}x^k$. Our goal is to write a differential equation involving g' and g that we can solve to find a closed-form expression for g, which will turn out to be $(1+x)^\alpha$.

Differentiating g term-by-term and using Theorem 32.1, we see that

$$\begin{aligned} g'(x) &= \sum_{k=1}^\infty \binom{\alpha}{k} kx^{k-1} = \sum_{k=1}^\infty \frac{\alpha(\alpha-1)\cdots(\alpha-k+1)}{k!}kx^{k-1} \\ &= \sum_{k=1}^\infty \frac{\alpha(\alpha-1)\cdots(\alpha-k+1)}{(k-1)!}x^{k-1} = \sum_{j=0}^\infty \frac{\alpha(\alpha-1)\cdots(\alpha-j)}{j!}x^j, \end{aligned}$$

where in the last step we reindexed the sum by putting $j = k-1$. Multiplying the second-to-last series by x gives

$$xg'(x) = \sum_{k=1}^\infty \frac{\alpha(\alpha-1)\cdots(\alpha-k+1)}{(k-1)!}x^k,$$

and adding these two formulas (renaming both indices to i for consistency) gives

$$\begin{aligned} g'(x)+xg'(x) &= \alpha + \sum_{i=1}^{\infty}\Big(\frac{\alpha(\alpha-1)\cdots(\alpha-i)}{i!}+\frac{\alpha(\alpha-1)\cdots(\alpha-i+1)}{(i-1)!}\Big)x^i \\ &= \alpha + \sum_{i=1}^{\infty}\frac{\alpha(\alpha-1)\cdots(\alpha-i+1)}{(i-1)!}\Big(\frac{\alpha-i}{i}+1\Big)x^i \\ &= \alpha + \alpha\sum_{i=1}^{\infty}\frac{\alpha(\alpha-1)\cdots(\alpha-i+1)}{i!}x^i = \alpha g(x). \end{aligned}$$

Thus we have proved that

$$g'(x)=\frac{\alpha g(x)}{1+x},$$

and we conclude that

$$\frac{d}{dx}\ln g(x)=\frac{g'(x)}{g(x)}=\frac{\alpha}{1+x}.$$

Using the fact that $g(0)=1$, we obtain

$$\ln g(x)=\ln g(0)+\int_0^x\frac{\alpha}{1+t}\,dt=\alpha\ln(1+t)\Big|_0^x=\alpha\ln(1+x).$$

Taking exponentials gives $g(x)=(1+x)^\alpha$ and completes the proof. □

33.4. Power series arithmetic

If we want to find the Maclaurin series for $f(x)=e^x\sin x$, we could proceed by computing all of its derivatives and using Theorem 33.2. However, if we want to avoid using the product rule over and over again, there is another way. We already know the power series representations of e^x and $\sin x$, and it turns out (though we will not prove it) that multiplying these series as though they were polynomials gives the power series representation of their product:

$$\begin{aligned} e^x\sin x &= \Big(1+x+\frac{1}{2}x^2+\frac{1}{6}x^3+\cdots\Big)\Big(x-\frac{1}{6}x^3+\cdots\Big) \\ &= \Big(x+x^2+\frac{1}{2}x^3+\frac{1}{6}x^4+\cdots\Big)+\Big(-\frac{1}{6}x^3-\frac{1}{6}x^4-\frac{1}{12}x^5-\frac{1}{36}x^6-\cdots\Big)+\cdots \\ &= x+x^2+\frac{1}{3}x^3+\cdots, \end{aligned}$$

where each instance of $\cdots$ indicates terms of degree 4 or higher. A similar computation can be done when we divide power series, using a long-division-type procedure analogous to polynomial long division; this can be used, for example, to find the Maclaurin series for $\tan x = \frac{\sin x}{\cos x}$.

Review of convergence for series

Stewart §11.7, Spivak Ch. 23

This review is not included in a numbered lecture, but will appear in a separate online video.

When determining convergence or divergence of a given series, the first fundamental fact to keep in mind is Corollary 27.12: if the sequence of terms a_n does not converge to 0, then the series $\sum a_n$ diverges.

If the sequence of terms does go to 0 – that is, if $\lim_{n\to\infty} a_n = 0$ – then the series $\sum a_n$ might converge and might diverge. If all the terms are ≥ 0, then it is reasonable to think of this convergence/divergence as being determined by whether the terms a_n go to 0 “quickly enough”.

Keep in mind the harmonic series $\sum \frac{1}{n}$ as a reminder that convergence to 0 of the *sequence* of terms does not imply convergence of the *series* (which requires convergence of the sequence of partial sums); this is an example where the terms go to 0 slowly enough that the series diverges.

Two classes of series are especially important to keep in mind.

- Given a real number p, the corresponding *p-series* is $\sum \frac{1}{n^p}$. The series converges if $p > 1$ and diverges if $p \leq 1$. (Note that the terms go to 0 for every $p > 0$.)
- Given real numbers a, r, the corresponding *geometric series* is $\sum ar^{n-1}$. Assuming $a \neq 0$, the series converges if $|r| < 1$ and diverges if $|r| \geq 1$. (Note that the terms go to 0 if and only if $|r| < 1$.)

For a more general series with terms $a_n \geq 0$ that go to 0, the question of “do the terms go to 0 quickly enough for the series to converge” can often be answered by comparing to one of these two kinds of series and using the Comparison Test or the Limit Comparison Test. Intuitively, one can think of ar^{n-1} as going to 0 *exponentially quickly*, and $\frac{1}{n^p}$ as going to 0 *polynomially quickly with degree p*. Then one way to verify that the terms a_n go to 0 “quickly enough for the series to converge” is to relate them to a sequence that goes to 0 either exponentially quickly, or polynomially quickly with degree > 1.[23]

For series with some negative terms, the situation is a little more subtle and there are two ways that convergence can happen.

- *Absolute convergence:* The terms a_n go to 0 quickly enough that $\sum |a_n|$ converges.
- *Conditional convergence:* The terms go to 0 slowly enough that $\sum |a_n|$ diverges, but there is enough cancellation between positive and negative terms that the series $\sum a_n$ still converges.

For examples of conditional convergence, we can look to alternating series of the form $\sum (-1)^n b_n$, where $b_n \geq 0$. The Alternating Series Test says that such a series converges if and only if $b_n \to 0$ (so the naive divergence test is actually a necessary and sufficient condition in this case), and so if we choose $b_n \to 0$ with $\sum b_n$ divergent, then $\sum (-1)^n b_n$ is conditionally convergent. This includes the alternating harmonic series $\sum (-1)^n \frac{1}{n}$.

Three other convergence tests are worth keeping in mind.

(1) If the terms a_n can be written as $a_n = f(n)$ where f is a continuous nonnegative nonincreasing function and we can determine the convergence or divergence of the improper integral $\int_1^\infty f(x)\,dx$, then the integral test can be applied.
(2) If the terms a_n contain factorials or other products, it is often useful to use the ratio test.
(3) If the terms a_n contain an nth power, it is often useful to use the root test.

[23] This is not an exhaustive list of the different rates with which a sequence can go to 0, but exponential and polynomial rates are the most important.

Example 33.11. $\sum \frac{n-1}{2n+1}$ diverges by Corollary 27.12 because

$$\lim_{n\to\infty} \frac{n-1}{2n+1} = \lim_{n\to\infty} \frac{1-\frac{1}{n}}{2+\frac{1}{n}} = \frac{1}{2} \neq 0.$$

Example 33.12. $\sum \frac{\sqrt{n^3+1}}{3n^3+4n^2+2}$ has terms on the same order of magnitude as $n^{3/2}/n^3 = n^{-3/2}$, so we use the limit comparison test and observe that

$$\lim_{n\to\infty} \frac{\sqrt{n^3+1}}{3n^3+4n^2+2} \div n^{-3/2} = \lim_{n\to\infty} \frac{n^{-3/2}\sqrt{n^3+1}}{n^{-3}(3n^3+4n^2+2)} = \lim_{n\to\infty} \frac{\sqrt{1+n^{-3}}}{3+4n^{-1}+2n^{-3}} = \frac{1}{3}.$$

Since $\sum n^{-3/2}$ is a convergent p-series, the limit comparison test implies that the original series converges as well.

Example 33.13. Consider $\sum ne^{-n^2}$. The function xe^{-x^2} can be integrated by the substitution $u = -x^2$ and is decreasing as soon as $\frac{d}{dx}(xe^{-x^2}) = e^{-x^2} - 2x^2e^{-x^2} < 0$, which is true for all $x > 1$; since $\int_1^t xe^{-x^2}\,dx = [-\frac{1}{2}e^{-x^2}]_1^t = \frac{1}{2}(e^{-1} - e^{-t}) \to \frac{1}{2e}$ as $t \to \infty$, the integral test tells us that $\sum ne^{-n^2}$ is convergent. This fact can also be proved using the ratio test, the root test, or the comparison test (using a geometric series as reference); try it!

For power series of the form $\sum_{n=0}^\infty c_n(x-a)^n$, there is always a *radius of convergence* R such that the series converges absolutely on $(a-R, a+R)$ and diverges when $|x-a| > R$. It is possible to have $R = 0$ or $R = \infty$; the power series $\sum \frac{x^n}{n!}$ is an important example with $R = \infty$.

When $x \in (a-R, a+R)$ so that $|x-a| < R$, the proof of absolute convergence goes by comparing the series to a geometric series (exponential behavior). At the endpoints $x = a \pm R$, we typically have to compare to a p-series (polynomial behavior) and there are multiple possibilities (each of the following examples has $R = 1$):

1. absolute convergence at both endpoints ($\sum \frac{1}{n^2}x^n$);
2. conditional convergence at both endpoints ($\sum \frac{(-1)^n}{2n+1}x^{2n+1}$);
3. conditional convergence at one endpoint and divergence at the other ($\sum \frac{1}{n}x^n$);
4. divergence at both endpoints ($\sum x^n$).

The radius of convergence can often be determined using a version of the Ratio or Root Tests. On the interval $(a-R, a+R)$, the power series defines a function f whose derivative and integral can be written as power series (with the same radius of convergence) using analogues of the familiar formulas for polynomials. Writing down the power series representations of the higher-order derivatives $f^{(n)}(x)$ and evaluating them at a reveals that when f admits a power series representation, it must be given by its *Taylor series* $\sum_{n=0}^\infty \frac{f^{(n)}(a)}{n!}(x-a)^n$. The partial sums of this series are the *Taylor polynomials*, and the difference between a function and its Taylor polynomial can be controlled by *Taylor's inequality* provided we have a good upper bound on $|f^{(n+1)}(x)|$.

We saw power series representations of e^x, $\sin x$, $\cos x$, $\frac{1}{1-x}$, $\ln(1-x)$, $\frac{1}{1+x^2}$, $\tan^{-1} x$, and $(1+x)^\alpha$ for $\alpha \in \mathbb{R}$; these were obtained with Taylor series, with the geometric series formula, and using differentiation and integration from known formulas.

Part VI. Conic sections, planetary motion

Lecture 34 Parabolas

Stewart §10.5, Spivak Chapter 4 appendix 2

34.1. Different descriptions

You may have encountered various ways to describe parabolas, or various properties that these curves have. Let us start our discussion of conic sections by recalling some (five!) of these descriptions; see Figure 2 below.

(1) *Analytic geometry – equation.* A parabola is the graph of the function $y = x^2$, or more generally $y = ax^2 + bx + c$, where $a, b, c \in \mathbb{R}$ are arbitrary parameters with $a \neq 0$. This gives a parabola that opens up (if $a > 0$) or down (if $a < 0$). For parabolas opening left and write we write $x = ay^2 + by + c$.

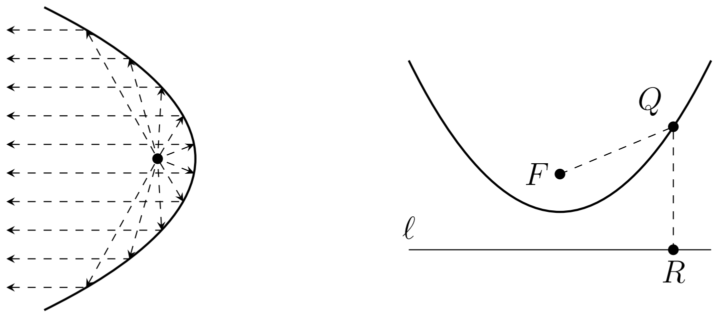
(2) *Physics – dynamics.* A projectile moving without air resistance in a uniform gravitational field flies along a parabola. Thus if we throw a ball, a parabola describes its flight path.

(3) *Physics – optics, acoustics.* A parabola has a distinguished point called the *focus* with the property that if a light bulb is placed at the focus and emits beams of light in all directions, then when these beams are reflected off of the parabola, they all become parallel to each other; see the first picture below. This is used in building headlights for cars. If the direction of the arrows is reversed this principle means that parallel incoming lines are all reflected to a single point (the focus), which is useful in building satellite dishes, radio telescopes, and parabolic microphones.

(4) *Two-dimensional geometry – focus and directrix.* A parabola has a point F, called the *focus*, and a line ℓ, called the *directrix*, with the property that given any point Q on the parabola, if R is the closest point to Q on the directrix ℓ, then $|QF| = |QR|$; see the second picture below. Notice that the point lying halfway between F and ℓ is on the parabola; this point is called the *vertex.*

(5) *Three-dimensional geometry – cross-section of cone.* Given a plane $\mathbf{P}$ and a cone C in three-dimensional space, if $\mathbf{P}$ is parallel to one of the lines containing the vertex of C, then the cross-section $\mathbf{P} \cap C$ is a parabola, as shown in the third picture below.

At first glance, it is not at all clear why the five different descriptions in the list above should all determine the same curve. Why should they be equivalent?



picture to be added

Figure 2. Different representations of a parabola.

34.2. Analytic geometry and projectile dynamics

We have already seen one equivalence: the first two descriptions are equivalent because if a projectile has constant horizontal velocity $v \neq 0$, initial vertical velocity w, and is subject to constant downward acceleration g, then its position (x, y) as a function of time t satisfies

$$\dot{x} = v, \quad \dot{y}(0) = w, \quad \ddot{y} = -g.$$

Integrating gives

$$\dot{y}(t) = \dot{y}(0) + \int_0^t \ddot{y}(\tau)\, d\tau = w - gt.$$

If the initial position is (x_0, y_0), then we have

$$x(t) = x_0 + vt, \quad y(t) = y_0 + \int_0^t \dot{y}(\tau)\, d\tau = y_0 + \int_0^t (w - g\tau)\, d\tau = y_0 + wt - \frac{g}{2}t^2.$$

This gives the trajectory as a parametric curve. To write y as a function of x we solve the first equation and get $t = (x - x_0)/v$ (recall that we assumed $v \neq 0$, so the projectile is not simply moving straight up and down), and deduce that

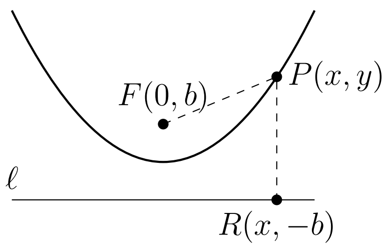
$$y = y_0 + w \cdot \frac{x - x_0}{v} - \frac{g}{2} \cdot \frac{(x - x_0)^2}{v^2}.$$

Thus y is a quadratic function of x, so the projectile follows a parabola.

34.3. Analytic geometry and plane Euclidean geometry

Now we show that the first and fourth descriptions from the list above are equivalent; that is, a curve with the focus-directrix property described there is in fact given as the graph of a quadratic polynomial.

Suppose C is a curve in the plane that has the focus-directrix property; that is, there is a point F and a line ℓ (not containing F) such that a point P in the plane lies on the curve C if and only if the distance $|PF|$ is equal to the distance from P to ℓ. We choose a coordinate system in which F lies on the positive y-axis and the origin is halfway between F and ℓ; thus $F = (0, b)$ and ℓ is given by the equation $y = -b$. Consider a point P with



coordinates (x, y). Then the closest point on ℓ to P is the point R that lies directly beneath P, which has coordinates $(x, -b)$. The focus-directrix property says that P lies on C if and only if $|PF| = |PR|$, or equivalently, $|PF|^2 = |PR|^2$. Observe that

$$\begin{aligned} |PF|^2 &= (x-0)^2 + (y-b)^2 = x^2 + y^2 - 2by + b^2, \\ |PR|^2 &= (y-(-b))^2 = (y+b)^2 = y^2 + 2by + b^2. \end{aligned}$$

Thus P lies on C if and only if

$$x^2 + y^2 - 2by + b^2 = y^2 + 2by + b^2 \quad \Leftrightarrow \quad x^2 = 4by \quad \Leftrightarrow \quad y = \frac{1}{4b}x^2.$$

In other words, C is the graph of the quadratic $y = ax^2$, where $a = \frac{1}{4b}$.

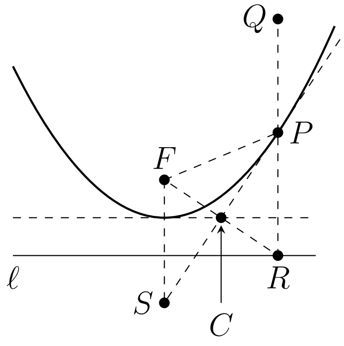
More generally, if C is a parabola with focus $F = (p, q)$ and directrix $y = r$ for some $p, q, r \in \mathbb{R}$ with $r \neq q$, then writing $k = (q+r)/2$, we can do a horizontal translation by p and a vertical translation by k to move F to $(0, b)$ and ℓ to $y = -b$, where $b = (q-r)/2$. The argument above gives the formula for the translated curve, so the original curve is the graph of

$$y - k = \frac{1}{4b}(x-p)^2 \quad \Leftrightarrow \quad y = \frac{q+r}{2} + \frac{1}{2(q-r)}(x-p)^2.$$

Remark 34.1. If we consider a parabola with a vertical directrix, then we interchange the roles of x and y in the above computations. If we consider a parabola with a directrix that is neither horizontal nor vertical, then the coordinates become more complicated, at least as long as we use a rectangular coordinate system, and we will not pursue this any further here.

34.4. Focus-directrix property and reflection property

Now we prove that the focus-directrix property implies the property that lines emanating from the focus are reflected to parallel lines, or equivalently, that an incoming line perpendicular to the directrix is reflected to the focus.



Consider such a line QP, and imagine a beam of light traveling along this line. When it reaches the point P on the parabola, what does it do? The law of reflection says that its outgoing angle is equal to its incoming angle. But angle with what? Whenever we discuss the angle that a line makes with a curve (or that two curves make with each other), what we mean is the angle that is made with the *tangent line* to the curve. In other words, if TS is the tangent line to the parabola at P, then the incoming beam is reflected towards the focus if and only if $\measuredangle FPS = \measuredangle QPT$. Since $\measuredangle QPT = \measuredangle RPS$, we conclude that

> *in order to prove that the incoming beam along QP is reflected towards the point F, it suffices to prove that the line bisecting the angle $\angle FPR$ is the tangent line to the parabola at P.*

At this point one might expect that we should introduce some coordinates and use the description of the parabola in terms of the graph of a quadratic polynomial, since describing a tangent line involves computing a derivative. But in fact, we can get a little more mileage out of a purely geometric approach.

Let ℓ' be the line bisecting $\angle FPR$, and let S be the point where ℓ' intersects the vertical line through F (in the picture, T also lies on ℓ'). Then $\measuredangle RPC = \measuredangle FPC$ and $|PF| = |PR|$ by the focus-directrix property, so the triangles FPC and RPC are congruent. In particular, C is the midpoint of FR, and ℓ' and FR are perpendicular.

Recall a fundamental property of perpendicular bisectors: ℓ' is the set of points in the plane that are the same distance from both F and R. If a point is on the same side of ℓ' as F is, then it is closer to F than it is to R, and vice versa. In particular, if X is *any* point on the parabola, then we have

$$|XF| = \text{distance from } X \text{ to } \ell \leq |XR|.$$

Moreover, the latter inequality is strict unless X lies directly above R; that is, unless $X = P$. This means that P is the only point where the parabola intersects ℓ', and that every other point on the parabola lies above ℓ'. Then the proof that ℓ' is the tangent line to the parabola at P, and thus that the focus-directrix property implies the reflection property, is completed by the following exercise.

Exercise 34.2. Let $I \subset \mathbb{R}$ be an open interval and suppose that $f\colon I \to \mathbb{R}$ is differentiable at a point $a \in I$. Suppose moreover that $y = mx + b$ is a line in the plane with the property that $f(a) = ma + b$, and $f(x) > mx + b$ for all $x \neq a$. Prove that $f'(a) = m$, so that in particular this line is the tangent line to the graph of f at a.

Remark 34.3. Without the assumption that f is differentiable at a, the conclusion of the exercise could fail; consider the absolute value function $f(x) = |x|$ and $a = 0$.

34.5. Dandelin spheres

The only remaining property to consider is the one that gives *conic sections* their names: a parabola is the cross-section obtained by intersecting a cone with a plane that is parallel to one of the lines that makes up the edge of the cone. For this we use a beautiful and elegant geometric argument discovered by the 19th century Belgian mathematician Germinal Dandelin.

In the following it is useful to keep Figure 3 in mind; we orient the cone so that it opens straight up, and consider the curve formed by intersecting the cone with a plane **P**. Now imagine that we drop a tiny sphere – like a small scoop of ice cream – into the cone. When it comes to rest near the bottom of the sphere, it will be tangent to it along a horizontal circle. If we increase the size of the sphere – changing our analogy, we may imagine that the sphere is a balloon that we inflate – then the sphere, and its circle of tangency, will rise higher on the cone. When the sphere is very small, it will lie entirely beneath the plane **P**. When it is sufficiently large, some points of it will lie above **P**. By the Intermediate Value Theorem, for some size of the sphere, it intersects the plane **P** in exactly one point.[24] Then one can deduce from Exercise 34.2 that **P** is tangent to

[24] It is a good exercise to make this statement a little more formal by writing down a continuous function that vanishes precisely when there is exactly one point of intersection.

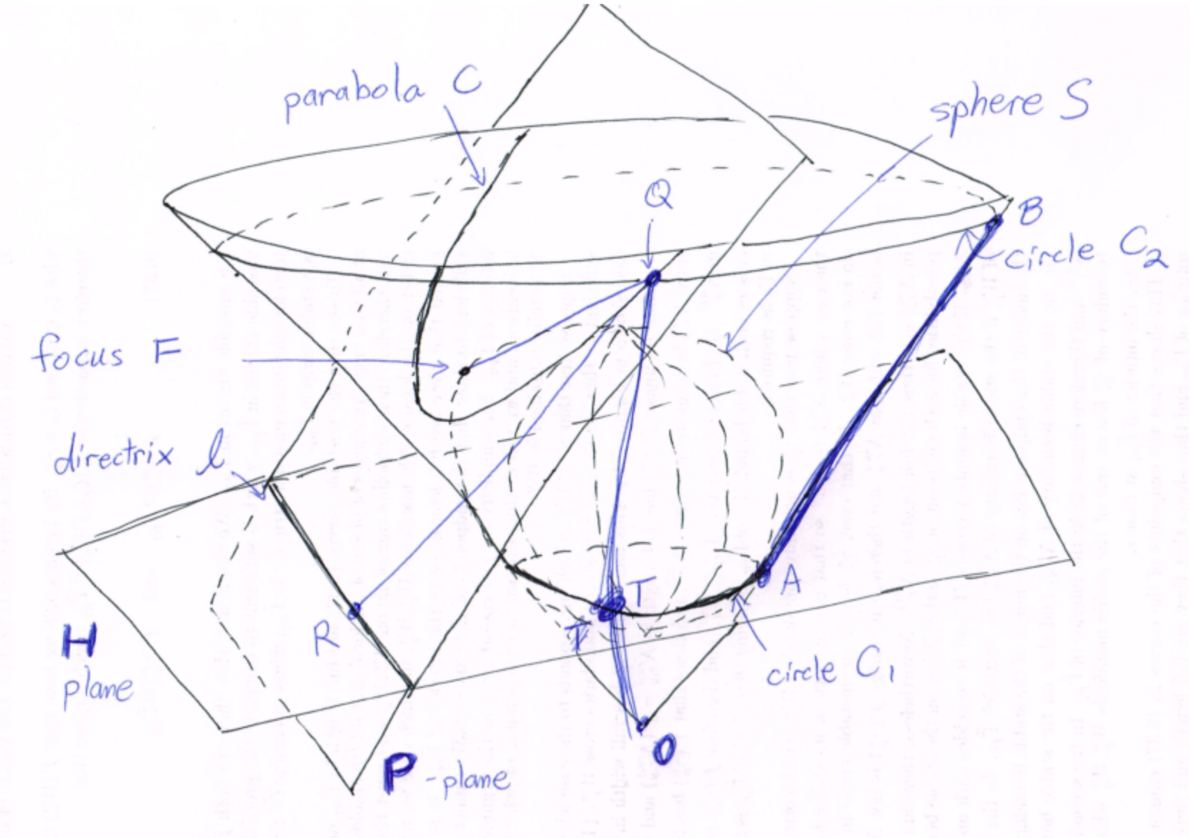


FIGURE 3. The Dandelin sphere for a parabola.

the sphere. The sphere obtained by this process is the *Dandelin sphere* associated to the cone and to the plane **P**; we denote it by S.

Let $F = S \cap \mathbf{P}$; this point will be the focus. Let C_1 be the circle of points where S intersects (and is tangent to) the cone. Since the picture is symmetric under rotation around a vertical axis, this lies in a horizontal plane **H**. Let $\ell = \mathbf{P} \cap \mathbf{H}$; this line will be the directrix. Given a point Q on the intersection of **P** and the cone, we must show that $|QF| = |Q\ell|$.

Start by drawing the line QO, where we recall that O is the vertex of the cone, and let T be the point where this line intersects **H**. This line is tangent to S, as is the line QF; thus if we write C for the center of S (not pictured), we see that $\angle QFC$ and $\angle QTC$ are both right angles, since a tangent line to a sphere is perpendicular to the radius at that point. Now Pythagoras gives

$$|QF|^2 = |QC|^2 - |CF|^2 = |QC|^2 - |CT|^2 = |QT|^2, \tag{34.1}$$

where the second equality uses the fact that CF and CT are radii of the sphere. In fact the computation in (34.1) proves the following general lemma, which is useful to record here for future reference.

Lemma 34.4. *Given a sphere S and a point X outside of the sphere, if XY and XZ are tangent to the sphere at Y and Z, respectively, then $|XY| = |XZ|$.*

So far we have not actually assumed that **P** is parallel to a line generating the sphere; this will come in useful in the next lecture when we consider more general conic sections. Now we add this assumption, and consider the line generating the sphere that is parallel to **P**; this is OA in the picture, where A is chosen to lie on C_1. Let C_2 be the horizontal circle containing Q, and let B denote the point where C_2 intersects the line OA. Then the line segment AB is obtained from QT by rotating around the vertical axis, so

$$|QT| = |AB| = |QR| = |Q\ell|$$

where R is the point on ℓ that is closest to Q, and the second equality follows because QR and BA are parallel line segments running between the same two horizontal planes (the planes containing C_1 and C_2). Combining this with (34.1) shows that $|QF| = |QT| = |Q\ell|$, and thus the point F and the line ℓ satisfy the focus-directrix property for the intersection of **P** with the cone.

Lecture 35 Ellipses (and hyperbolas)

Stewart §10.6

35.1. Focus-directrix description of conics, and polar coordinates

Now suppose we take a cross-section of a cone with an *arbitrary* plane **P**, which is not assumed to be parallel to any of the sides of the cone; see Figure 4. As before, take the axis of revolution of the cone to be vertical, and consider the Dandelin sphere S that is tangent to both the cone and the plane **P**, and lies below the plane. Let F be the point where S intersects **P**. As long as **P** is not horizontal (and in this case **P** intersects the cone in a circle, which we understand), **P** intersects **H** in a line ℓ.

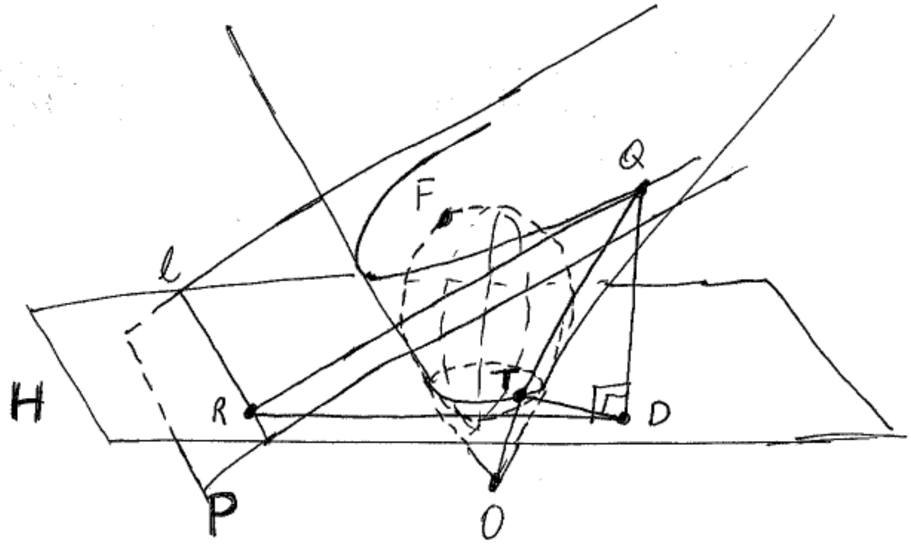


Figure 4. A Dandelin sphere for a general conic section.

We want to describe the curve in which $\mathbf{P}$ intersects the cone. Consider an arbitrary point Q on this curve. As before, Lemma 34.4 gives $|QF| = |QT|$, where T is the point in which the line QO intersects $\mathbf{H}$. And once again, we can choose a point R on the line ℓ such that $|Q\ell| = |QR|$. However, since $\mathbf{P}$ is not assumed to be parallel to any of the sides of the cone, we can no longer deduce that $|QT|$ and $|QR|$ are the same. Instead, we can compare both of these lengths to $|QD|$, where D is the point in $\mathbf{H}$ that lies directly below Q, so that in particular QD is vertical and $\angle QDR$, $\angle QDT$ are right angles. Then elementary trigonometry gives

$$\sin\measuredangle QTD = \frac{|QD|}{|QT|} \quad\text{and}\quad \sin\measuredangle QRD = \frac{|QD|}{|QR|}.$$

Observe that $\alpha = \measuredangle QTD$ is the angle that measures how wide or narrow the cone is, and does not depend on the specific choice of Q. Similarly, $\beta = \measuredangle QRD$ is the angle in which the planes $\mathbf{P}$ and $\mathbf{H}$ intersect, and once again is independent of Q. Thus we have

$$\frac{|QF|}{|Q\ell|} = \frac{|QT|}{|QR|} = \frac{|QD|/\sin\alpha}{|QD|/\sin\beta} = \frac{\sin\beta}{\sin\alpha}.$$

We have proved the following result.

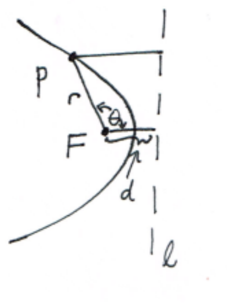
Theorem 35.1. *Consider a cone obtained as follows: take a line through the origin that makes an angle α with the horizontal plane, and rotate it around the vertical axis. Let C be a curve obtained by intersecting this cone with a plane $\mathbf{P}$ that makes a nonzero angle β with the horizontal. Let S be the Dandelin sphere for this cone and plane, and let $\mathbf{H}$ be the horizontal plane through the circle in which S intersects the cone. Let $\ell = \mathbf{H} \cap \mathbf{P}$ and $F = S \cap \mathbf{P}$. Then the curve C can be described via the following focus-directrix property: a point $Q \in \mathbf{P}$ lies on the curve C if and only if*

$$|QF| = e|Q\ell|, \quad \textit{where } e = \frac{\sin\beta}{\sin\alpha}. \tag{35.1}$$

The number $e > 0$ is called the *eccentricity* of the conic section C. When $\beta < \alpha$, we have $0 < e < 1$ and the curve C is an *ellipse*. When $\beta = \alpha$, we have $e = 1$ and the curve C is a *parabola*. When $\beta > \alpha$, we have $e > 1$ and the curve C is a *hyperbola*.[25]

We can use (35.1) to write a formula for C in polar coordinates. Put the focus F at the origin and let the directrix ℓ be the line $x = d$. Then given a point P at polar coordinates (r, θ), we have $|PF| = r$, while the x-coordinate of P is $r\cos\theta$, so $|P\ell| = d - r\cos\theta$. Thus P satisfies $|PF| = e|P\ell|$ if and only if $r = ed - er\cos\theta$. Solving for r, we see that in polar coordinates, this conic section is the graph of

$$r = \frac{ed}{1 + e\cos\theta}. \tag{35.2}$$



Remark 35.2. Choosing a directrix $x = -d$ gives the related equation $r = ed/(1 - e\cos\theta)$, and choosing a horizontal directrix $y = \pm d$ has the effect of replacing cos with sin.

[25] In fact, in this case C is one branch of a hyperbola; one typically considers also the reflection of the cone below the origin, so that the hyperbola has a corresponding branch in the lower half-space.

Another way of writing (35.2) is $r = R/(1 + e\cos\theta)$, where we no longer specify the distance to the directrix explicitly. Then putting $e = 0$ gives $r = R$, which is the polar equation of a circle, so we see that a circle is a conic section with eccentricity 0. (Note that this has no focus-directrix characterization, since the directrix would need to be "at infinity".)

35.2. Focus-focus description of ellipses, and rectangular coordinates

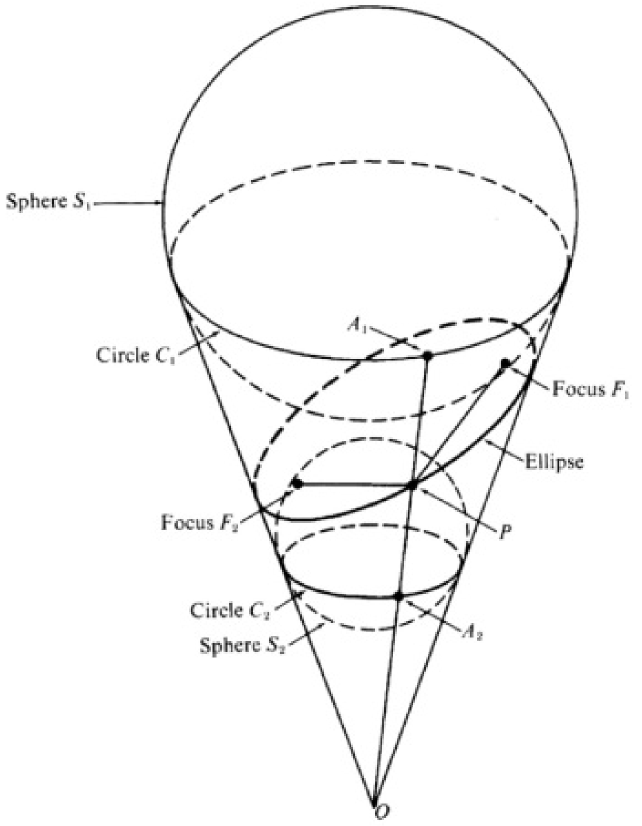


FIGURE 2.11 The ice-cream-cone proof.

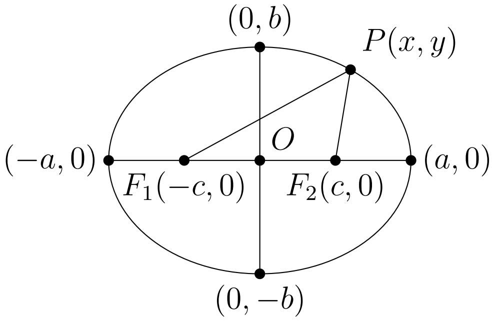
The ellipse also has a description not in terms of a focus and directrix, but in terms of *two* foci (plural of focus). This is illustrated in the picture at right, which for the moment is borrowed from Apostol's textbook (until I manage to produce one of my own). When the plane is not parallel to the generator of the cone, it actually has *two* Dandelin spheres, one below and one above. Writing F_1 and F_2 for the two points in which these spheres intersect the plane, we see that a point P on the ellipse has the property (by Lemma 34.4) that $|PF_1| = |PA_1|$ and $|PF_2| = |PA_2|$, where A_1 and A_2 are the points in which the line PO intersects the horizontal circles corresponding to the two spheres. But then $|PF_1|+|PF_2| = |PA_1|+|PA_2| = |A_1A_2|$, and this last quantity does not depend on P (by another application of Lemma 34.4, as in the proof for the parabola). Summarizing the result of this argument, we have proved the following.

Theorem 35.3. *If C is an ellipse (a conic section with eccentricity $e < 1$), then there are two points F_1, F_2, called* foci*, and a real number $r > 0$, such that a point P lies on C if and only if $|PF_1| + |PF_2| = r$.*

Note that if $F_1 = F_2$, so that the two foci coincide, then this condition reduces to the statement that the distance from P to the (single) focus is constant, which gives a circle.

Theorem 35.3 can be used to write an equation for an ellipse with a given eccentricity in rectangular coordinates. Instead of putting the origin at a focus, as we did with polar coordinates, we put the two foci on the x-axis with the origin at their midpoint, so that the foci F_1 and F_2 have coordinates $(\pm c, 0)$. Let $(a, 0)$ be the right-most point of the ellipse; by symmetry the left-most point is $(-a, 0)$. These two points are called the *vertices* of the ellipse. The line segment between the two vertices is called the *major* *axis*, and the line segment from the origin to one of the vertices is the *semimajor axis*. Similarly, the line between the top and bottom points $(0, \pm b)$ is the *minor axis*.

Observe that the vertex $Q = (a, 0)$ has $|QF_1| = a + c$ and $|QF_2| = a - c$, so the sum of the distances to the foci is $|QF_1| + |QF_2| = 2a$. Since the sum of the distances to the foci is constant for all points on the ellipse, we see that the ellipse is the set of point P such that $|PF_1| + |PF_2| = 2a$; in other words, the sum of the distances to the foci must always be equal to the length of the major axis. We also observe that when $P = (0, b)$, we have $|PF_1| = |PF_2| = \sqrt{b^2 + c^2}$, so each of these is equal to a, and in particular, a, b, c are related by



(35.3) $$a^2 = b^2 + c^2.$$

Now suppose P has coordinates (x, y). Using the Pythagorean formula to write $|PF_1| = \sqrt{(x + c)^2 + y^2}$ and $|PF_2| = \sqrt{(x - c)^2 + y^2}$, we see that the ellipse is the set of points (x, y) such that

$$\sqrt{(x + c)^2 + y^2} + \sqrt{(x - c)^2 + y^2} = 2a.$$

Isolating the first square root and then squaring both sides, this is equivalent to

$$\begin{aligned}(x + c)^2 + y^2 &= (\sqrt{(x - c)^2 + y^2} + 2a)^2 \\ &= (x - c)^2 + y^2 + 4a\sqrt{(x - c)^2 + y^2} + 4a^2.\end{aligned}$$

Expanding both sides gives

$$x^2 + 2cx + c^2 + y^2 = x^2 - 2cx + c^2 + y^2 + 4a\sqrt{(x - c)^2 + y^2} + 4a^2,$$

and after simplifying and isolating the square root we obtain

$$4cx - 4a^2 = 4a\sqrt{(x - c)^2 + y^2}.$$

Dividing by 4 and squaring both sides gives

$$\begin{aligned}(cx - a^2)^2 &= a^2\big((x - c)^2 + y^2\big), \\ c^2x^2 - 2a^2cx + a^4 &= a^2(x^2 - 2cx + c^2 + y^2) \\ &= a^2x^2 - 2a^2cx + a^2c^2 + a^2y^2, \\ a^4 - a^2c^2 &= (a^2 - c^2)x^2 + a^2y^2.\end{aligned}$$

Recalling from (35.3) that $a^2 - c^2 = b^2$, this is equivalent to

$$a^2b^2 = b^2x^2 + a^2y^2,$$

and dividing through by a^2b^2 gives the equation for the ellipse as

(35.4) $$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1.$$

Observe that when $a = b = r$ this becomes the familiar equation $x^2 + y^2 = r^2$ for a circle with radius r.

As with the parabola, we can shift this equation to describe an ellipse at other locations in the plane. If the ellipse has foci which lie on the same horizontal or vertical line, with midpoint (h,k), and if the lengths of the horizontal and vertical axes of the ellipse are $2a$ and $2b$, respectively, then the equation of the ellipse is

$$\frac{(x-h)^2}{a^2}+\frac{(y-k)^2}{b^2}=1. \tag{35.5}$$

35.3. Hyperbolas

The results for ellipses in the previous section have analogues for hyperbolas. We omit the details here, and merely mention the conclusions: the hyperbola corresponding to two foci F_1 and F_2 and a real number $r>0$ is the set of all points P in the plane such that

$$\big||PF_1|-|PF_2|\big|=r. \tag{35.6}$$

If $r\geq|F_1F_2|$, then the only way to satisfy (35.6) is if P is on the line ℓ through F_1 and F_2 but does not lie between them. This is a degenerate case that we ignore, so we assume that $0<r<|F_1F_2|$. In this case the hyperbola contains two points on ℓ, which lie between F_1 and F_2.

If we work in rectangular coordinates where the foci are at $(\pm c,0)$, and the points $(\pm a,0)$ are on the hyperbola (as before, we call these the *vertices*), then $a<c$ by the previous paragraph. Writing $b^2=c^2-a^2$ for convenience, a similar computation to the one in the previous section gives the equation for the hyperbola as

$$\frac{x^2}{a^2}-\frac{y^2}{b^2}=1. \tag{35.7}$$

If the foci lie on the y-axis then the roles of x,y are reversed. As in (35.5), this can be shifted to put the hyperbola elsewhere in the plane.

One feature specific to hyperbolas is worth mentioning. As x^2 gets large, y^2 must also get large, and the right-hand side of (35.7) becomes insignificant in comparison. Without this RHS, the equation would be $y=\pm\frac{b}{a}x$. These lines are the *asymptotes* of the hyperbola.

Exercise 35.4. Prove that as $x\to\infty$, the corresponding positive value of y (such that (x,y) lies on the hyperbola) has the property that the distance from (x,y) to the line $y=\frac{b}{a}x$ approaches 0.

35.4. List of characterizations

Our discussion of conic sections can be summarized by the following list, which gives equivalent ways of characterizing these curves.

(1) *Cross-section of a cone and a plane.* If α and β are the angles that the cone and plane, respectively, make with the horizontal, then $e=\sin\beta/\sin\alpha$ is the *eccentricity* of the resulting conic. $e=0$ gives a circle, $0<e<1$ gives an ellipse, $e=1$ gives a parabola, and $e>1$ gives a hyperbola.

(2) *Focus-directrix.* If the plane is not horizontal (the conic is not a circle), then the conic is described by a focus (point) F and a directrix (line) ℓ as the set of points

Q in the plane such that $|QF| = e|Q\ell|$. The focus and directrix can be found using a Dandelin sphere.

(3) *Polar coordinates.* If we choose a polar coordinate system with origin at the focus and such that the directrix is vertical, then the curve is given in polar coordinates as the graph of $r = R/(1 + e\cos\theta)$, where $R > 0$ is a constant and e is the eccentricity. When $e > 0$ we have $R = ed$, where d is the distance from the focus to the directrix.

(4) *Focus-focus.* If the curve is an ellipse then there are two foci F_1 and F_2 such that the conic is the set of points Q in the plane such that $|QF_1| + |QF_2|$ is equal to the length of the major axis. These two foci can be found using two Dandelin spheres. A similar characterization is available for hyperbolas (replacing sum with difference), but not for parabolas.

(5) *Rectangular coordinates.* Choosing a rectangular coordinate system with origin at the midpoint of the foci (for ellipses and hyperbolas) or at the midpoint of the focus and the directrix (for parabolas), the curve takes the familiar form $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ (ellipse), $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$ (hyperbola), or $y = ax^2$ (parabola), or possibly with x and y reversed depending on which orientation we choose.

(6) *Reflection property.* For a parabola, the lines emanating from the focus in all directions are reflected off of the parabola into a family of parallel lines. We proved this, and we leave the following laws for ellipses and hyperbolas as exercises.
   - For an ellipse, lines from one focus are reflected towards the other focus.
   - For a hyperbola, lines directed *towards* one focus, but with the hyperbola in the way, are reflected towards the other focus.

(7) *Motion in a gravitational field.* We proved that in a gravitational field that points uniformly downwards, an object moving without air resistance follows a parabola. Next we will turn our attention to movement in a gravitation field directed towards a single fixed point (the sun) that obeys an inverse square law, and show that the resulting trajectories are always conic sections.

Lecture 36 Kepler and Newton

Spivak Ch. 17

In the early 1600's, the German astronomer Johannes Kepler formulated the following three laws of planetary motion.

(1) *Elliptical motion*: Planets move in ellipses, with the sun at one focus.

(2) *Equal areas in equal times*: For a given planet, the area swept out by the line from the planet to the sun depends only on the amount of time elapsed, and not on when we start recording.

(3) *Harmonic law*: The ratio $(\text{major axis})^3/(\text{period})^2$ is the same for all planets.

Kepler's work was based on extensive observations and computations, and did not offer an explanation for *why* these laws should be true. An explanation of the mechanism behind the laws would have to wait for the work of Isaac Newton, who began developing the ideas of calculus in the 1660's, both at Cambridge and during a period of isolation in 1665-1666 when the university was closed due to an epidemic of the bubonic plague. Eventually Newton developed a theory of physics that he used to derive Kepler's laws in his *Principia*, published in 1687. The two crucial laws are the following.

- *Newton's second law*: The force F acting on an object, and its resulting acceleration a, are related by $F = ma$, where m is the mass of the object.
- *Law of universal gravitation*: Given two objects with masses M and m, each object attracts the other with force GMm/r^2, where r is the distance between the objects and G is a universal constant.

In the setting we are interested in, M denotes the mass of the sun, and m denotes the mass of the planet that we study. Although the planet moves in 3-dimensional space, its orbit is contained in a single 2-dimensional plane, so we will describe its position using both polar coordinates (r, θ) and rectangular coordinates (x, y). We will write $c(t)$ for the its position at time t. We will also write $\dot{c}(t) = \frac{d}{dt}c(t)$ for its velocity at time t and $\ddot{c}(t) = \frac{d^2}{dt^2}c(t)$ for its acceleration. Observe that because the object moves in a 2-dimensional plane, its velocity and acceleration are given by not just a magnitude, but also a direction; that is, they are *vectors* in this plane, as is the force F. You will study these further in a later calculus course. For the time being we merely observe that we can use rectangular coordinates to write

$$\dot{c} = (\dot{x}, \dot{y}) \quad \text{and} \quad \ddot{c} = (\ddot{x}, \ddot{y}),$$

where x and y are the coordinate functions describing the position c; both x and y are functions of time t, and the notation above represents their first and second derivatives with respect to t. We will similarly write $\dot{r}, \dot{\theta}, \ddot{r}, \ddot{\theta}$ for the first and second derivatives of the polar coordinates of $c(t)$ with respect to t.

With our notation established, let us begin our analysis. For simplicity we assume that $m \ll M$ and ignore the motion of the sun, assuming instead that the location of the sun is fixed; we will use this as the origin of our coordinate system.[26] We also ignore the effect of any other planets, asteroids, comets, etc., that may be lurking in the vicinity.[27] Under these assumptions, Newton's laws imply that a planet at position (r, θ) (in polar coordinates) has acceleration $a = GM/r^2$, directed along the line from the planet towards the sun. We prove the following three theorems, which demonstrate that this implies Kepler's laws. All three theorems use Newton's second law $F = ma$, but the first two theorems do not require the full strength of the law of universal gravitation.

Theorem 36.1. *Suppose that an object moves according to Newton's second law $F = ma$, and that F depends only on the object's current position (r, θ). (We do not yet*

[26] For a completely precise treatment, this assumption should be removed, and we should put the origin at the center of mass of the sun-planet system.

[27] This seems reasonable since the gravity exerted by these objects is extremely small relative to the gravity exerted by the sun. However, the cumulative effect of these perturbations over long (long!) periods of time can be substantial, and the question of asymptotic stability of the solar system remains extremely difficult; this was one of the questions that led to the development of the part of the theory of dynamical systems that is popularly known as *chaos theory*.

assume the law of universal gravitation.) Then the object's orbit satisfies Kepler's second law (equal areas in equal times) if and only if $F(r,\theta)$ always points along the line connecting the object to the origin. In this case, there is a constant K such that $r^2\dot{\theta} = K$ for all times t, and writing $a(t) := \ddot{r} - r(\dot{\theta})^2 = \ddot{r} - K^2/r^3$, we have

$$\ddot{c} = (a(t)\cos\theta, a(t)\sin\theta). \tag{36.1}$$

Informally, this says that an object moving in a force field has the property of "equal areas in equal times" if and only if the force field is *central* (always points along the line to the origin). Thus Newton's laws imply Kepler's second law.

Theorem 36.2. *Suppose an object moves in a central force field following an inverse square law, meaning that $\ddot{c}$ has magnitude Q/r^2 for some constant Q, and always points towards the origin. Then the object moves along a conic section. In particular, if the object's orbit is periodic, then it moves along an ellipse (or a circle).*

In fact, Theorem 36.2 is also an 'if and only if' – if every object moving in a central force field moves along a conic section, then the force satisfies an inverse square law. We will not prove this direction, however, and will content ourselves with the direction stated, which shows that Newton's laws imply Kepler's first law.

Theorem 36.3. *Under the conditions of Theorem 36.2, Kepler's third law is satisfied if and only if the constant Q is the same for all planets.*

This theorem shows that Kepler's third law holds if and only if the gravitational constant G is truly universal.

Proof of Theorem 36.1. First we determine how to write Kepler's second law in terms of r and θ. By (25.2), the area swept out by the curve c from time t_1 to t_2 is $\int_{\theta(t_1)}^{\theta(t_2)} \frac{1}{2}r^2\,d\theta$, where in the integral we consider r as a function of θ. Using the substitution rule to write the integral in terms of t, we see that the area is

$$\int_{t_1}^{t_2} \frac{1}{2}r(t)^2\dot{\theta}(t)\,dt. \tag{36.2}$$

Thus Kepler's second law – equal areas in equal times – is true if and only if $r^2\dot{\theta}$ is constant.

Now we look at the acceleration $\ddot{c}$, since this points in the same direction as the force. Writing $c = (x,y) = (r\cos\theta, r\sin\theta)$ and differentiating coordinate-wise gives

$$\begin{aligned} \dot{x} &= \dot{r}\cos\theta - r\dot{\theta}\sin\theta, \\ \dot{y} &= \dot{r}\sin\theta + r\dot{\theta}\cos\theta. \end{aligned} \tag{36.3}$$

Differentiating a second time gives

$$\begin{aligned} \ddot{x} &= \ddot{r}\cos\theta - 2\dot{r}\dot{\theta}\sin\theta - r\ddot{\theta}\sin\theta - r(\dot{\theta})^2\cos\theta = (\ddot{r} - r(\dot{\theta})^2)\cos\theta - (2\dot{r}\dot{\theta} + r\ddot{\theta})\sin\theta, \\ \ddot{y} &= \ddot{r}\sin\theta + 2\dot{r}\dot{\theta}\cos\theta + r\ddot{\theta}\cos\theta - r(\dot{\theta})^2\sin\theta = (\ddot{r} - r(\dot{\theta})^2)\sin\theta + (2\dot{r}\dot{\theta} + r\ddot{\theta})\cos\theta. \end{aligned} \tag{36.4}$$

Thus if we plot $\ddot{c} = (\ddot{x}, \ddot{y})$ in the plane, we can reach it by first moving a distance $(\ddot{r} - r(\dot{\theta})^2)$ in the direction of $(\cos\theta, \sin\theta)$, which is the direction of the line between the origin and the object, and then moving a distance of $(2\dot{r}\dot{\theta} + r\ddot{\theta})$ in the direction of

$(-\sin\theta, \cos\theta)$. Observe that this second motion is at right angles to the direction of the first motion, and thus $\ddot{c}$ points along the line to the origin if and only if $2\dot{r}\dot{\theta} + r\ddot{\theta} = 0$ (that is, if and only if our second motion had no distance).

Compare this to the criterion for Kepler's second law, that $r^2\dot{\theta}$ is constant. Differentiating $r^2\dot{\theta}$ w.r.t. t gives

$$\frac{d}{dt}(r^2\dot{\theta}) = 2r\dot{r}\dot{\theta} + r^2\ddot{\theta} = r(2\dot{r}\dot{\theta} + r\ddot{\theta}).$$

This shows that $r^2\dot{\theta}$ is constant if and only if $2\dot{r}\dot{\theta} + r\ddot{\theta} = 0$ at all times when the object is not at the origin, which shows that Kepler's second law holds if and only if the force always points along the line connecting the object to the origin. We saw already that $r^2\dot{\theta}$ is constant in this case, and then (36.1) follows immediately from (36.4). □

Proof of Theorem 36.2. If acceleration always has magnitude Q/r^2 and points towards the origin, then from (36.1) we have

$$\ddot{r} - \frac{K^2}{r^3} = -\frac{Q}{r^2} \quad \Rightarrow \quad \frac{d^2r}{dt^2} = \frac{K^2}{r^3} - \frac{Q}{r^2}. \tag{36.5}$$

A first this looks like a separable equation – the RHS depends only on r, which is what we want to find – so we might try dividing both sides by the RHS and then integrating. But the LHS is a *second* derivative, not a first derivative! So this is not actually a first-order separable DE like the ones we encountered earlier, and our techniques from before do not work.

Instead we need to reformulate things a little bit. Instead of writing r as a function of t, we consider r as a function of θ, and write (36.5) to obtain a DE in terms of $\frac{d}{d\theta}$, not $\frac{d}{dt}$. We can do this using the chain rule, but we need to be careful because a second derivative is involved. Recall from Theorem 36.1 that $\frac{d\theta}{dt} = \dot{\theta} = \frac{K}{r^2}$, and thus

$$\frac{d^2r}{dt^2} = \frac{d}{dt}\Big(\frac{dr}{dt}\Big) = \frac{d\theta}{dt}\frac{d}{d\theta}\Big(\frac{dr}{d\theta}\frac{d\theta}{dt}\Big) = \frac{K}{r^2}\frac{d}{d\theta}\Big(\frac{K}{r^2}\frac{dr}{d\theta}\Big),$$

where the first equality is the definition of second derivative, the second equality is two applications of the chain rule, and the third equality uses the formula for $\dot{\theta}$. Comparing this to (36.5) gives

$$\frac{K^2}{r^2}\frac{d}{d\theta}\Big(\frac{1}{r^2}\frac{dr}{d\theta}\Big) = \frac{K^2}{r^3} - \frac{Q}{r^2} \quad \Rightarrow \quad \frac{d}{d\theta}\Big(\frac{1}{r^2}\frac{dr}{d\theta}\Big) = \frac{1}{r} - \frac{Q}{K^2}.$$

We could expand the left-hand side using the product rule, but the resulting DE would not fit into any of the categories that we have a good procedure for solving at this point. Instead, the way forward is to make the observation that

$$\frac{d}{d\theta}\frac{1}{r} = -\frac{1}{r^2}\frac{dr}{d\theta},$$

and so the DE can be rewritten as

$$\frac{d}{d\theta}\Big(-\frac{d}{d\theta}\frac{1}{r}\Big) = \frac{1}{r} - \frac{Q}{K^2} \quad \Rightarrow \quad \frac{d^2}{d\theta^2}\frac{1}{r} = -\frac{1}{r} + \frac{Q}{K^2}.$$

Let $f(\theta) = \frac{1}{r} - \frac{Q}{K^2}$, and observe that $\frac{d^2}{d\theta^2} f(\theta) = \frac{d^2}{d\theta^2}\frac{1}{r}$; thus

$$\frac{d^2}{d\theta^2} f(\theta) = -f(\theta).$$

This is a DE that we can solve; the general solution is

$$f(\theta) = B\cos(\theta + \alpha),$$

where B, α are constants of integration determined by the initial values of f and $\frac{df}{d\theta}$. Thus

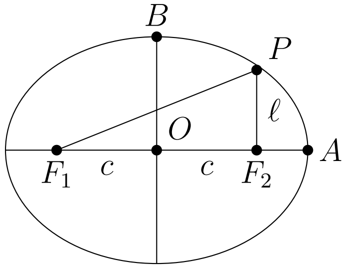
$$\frac{1}{r} - \frac{Q}{K^2} = B\cos(\theta + \alpha),$$

and solving for r gives

$$r = \frac{1}{\frac{Q}{K^2} + B\cos(\theta+\alpha)} = \frac{K^2/Q}{1 + \frac{BK^2}{Q}\cos(\theta+\alpha)}. \tag{36.6}$$

This is the polar equation for a conic section with eccentricity $e = BK^2/Q$, which proves the theorem. Observe that Q represents the strength of the central force divided by the mass of the object, while the parameters B, K, α are determined by the initial position and velocity. If these are such that the eccentricity is < 1, then the orbit is periodic and thus is an ellipse.

Observe that when the orbit is an ellipse, we can compare (36.6) to (35.2) and deduce that the eccentricity e and focus-directrix distance d satisfy $de = K^2/Q$. □



Before proving Theorem 36.3 we deduce a little more information about the shape of the ellipse in the case when $e < 1$. Referring to the picture at right, let ℓ be the length of the line segment F_2P, which goes through one of the foci and is perpendicular to the major axis (hence parallel to the minor axis). Observe that the distance from P to the directrix is the same as the distance d from F_2 to the directrix, and thus $\ell = de$ by definition. (Compare this with (35.2) and observe that P corresponds to $\theta = \frac{\pi}{2}$.) Moreover, recalling that the lengths a, b of the semi-major and semi-minor axes satisfy $b^2 + c^2 = a^2$, and that $|PF_1| + |PF_2| = 2a$, we have

$$2a = \ell + \sqrt{(2c)^2 + \ell^2} \quad \Rightarrow \quad (2c)^2 + \ell^2 = (2a - \ell)^2 = 4a^2 - 4a\ell + \ell^2.$$

Subtracting ℓ^2 from both sides gives $4c^2 = 4a^2 - 4a\ell$, and simplifying gives

$$a\ell = a^2 - c^2 = b^2.$$

Thus we have proved the following.

Lemma 36.4. *Let a, b be the lengths of the semimajor and semiminor axes of an ellipse with eccentricity e and focus-directrix distance d. Then $b^2 = dea = \ell a$, where ℓ is the length of the line segment from one focus to the edge of the ellipse, running perpendicular to the major axis.*

Proof of Theorem 36.3. Given a planet in an elliptical orbit as in Theorem 36.2, we need to relate the period of the orbit to the length of the major axis and to the constant Q. First observe that by (36.2) and the conclusion of Theorem 36.1, if we write T for the amount of time it takes the planet to complete one revolution (its period), then the area of the ellipse is

$$A = \int_0^T \frac{1}{2} r^2 \dot{\theta}\, dt = \int_0^T \frac{1}{2} K\, dt = \frac{KT}{2}, \tag{36.7}$$

where K is a constant (that may be different for different planets). On the other hand, we have $A = \pi ab$, where a, b are the lengths of the semimajor and semiminor axes, so

$$KT = 2\pi ab. \tag{36.8}$$

Using Lemma 36.4, we observe that $b^2 = \ell a = \frac{K^2}{Q} a$, and now we can complete the proof of Theorem 36.3 by squaring (36.8) and writing

$$K^2 T^2 = 4\pi^2 a^2 b^2 = 4\pi^2 a^2 \frac{K^2}{Q} a.$$

Solving for Q gives

$$Q = 4\pi^2 \frac{a^3}{T^2}, \tag{36.9}$$

which proves the theorem and establishes Kepler's third law as a consequence of the law of universal gravitation. □