

CHAPTER

4

Calculus Done Properly

Integration is where analysis, geometry and number theory meet, and where some of the most beautiful formulas in mathematics live.

IN THIS CHAPTER

- | | | |
|-----|--|---------------|
| 4.1 | What Is an Integral? | STEP III 1989 |
| 4.2 | Building Trigonometry from an Integral | STEP III 2017 |
| 4.3 | $22/7$ Is Bigger Than π | STEP I 1997 |
| 4.4 | Wallis, Beta and the Factorial of a Half | STEP II 1995 |
| 4.5 | The Sophomore's Dream | STEP III 2009 |
| 4.6 | Integrals With No Formula | STEP III 2016 |
| 4.7 | The Calculus of Variations | STEP III 2014 |
| 4.8 | Slicing a Spherical Loaf | STEP I 2001 |
| 4.9 | Holditch's Theorem | STEP III 2011 |

More Than Antiderivatives

I turn away with fright and horror from the lamentable evil of functions which do not have derivatives.

— Charles Hermite, in a letter of 1893

Ask a school student what integration is, and you will probably hear “the opposite of differentiation”. That is a wonderful fact, the *fundamental theorem of calculus*, and it makes most of the integrals you meet at school routine. But it is a theorem about integrals, not a definition of them, and it hides most of what makes integration interesting.

Integration is older than differentiation by about two thousand years. Archimedes found the area of a parabolic segment and the surface area of a sphere by summing thin slices, long before anyone thought of derivatives. And integration reaches places differentiation cannot. Every reasonable function can be differentiated by mechanical rules, but most functions *cannot* be integrated in a formula, which raises a natural question: when can they, and what do we do when they can’t? Integrals define new functions, prove inequalities between numbers, compute areas and volumes that have no other description, and give exact values of infinite sums and products. This chapter is about all of these.

What an integral really is

The chapter begins at the beginning, with the definition (Section 4.1). An integral is a limit of sums of thin rectangles; computing even $\int_0^a x^5 dx$ from that definition, without antiderivatives, is a substantial piece of work, and it makes you appreciate the fundamental theorem. With a precise definition comes a precise question: which functions are integrable? The answer, and the problems it raised, led in the early twentieth century to Lebesgue’s theory of integration, the foundation of modern probability.

Integrals that prove things

Some of the most surprising results in this chapter use integrals to prove facts that seem to have nothing to do with areas. A single integral of a positive function equals $\frac{22}{7} - \pi$, which proves $\pi < \frac{22}{7}$ in one line (Section 4.3). Powers of $\sin t$ lead to Wallis’s infinite product for π and to the value $\left(\frac{1}{2}\right)! = \frac{\sqrt{\pi}}{2}$ (Section 4.4). The integral of x^x turns into the sum $\sum (-1)^{n+1} n^{-n}$ (Section 4.5). And a single integral, $\int_0^x \frac{du}{1+u^2}$, can be used to build trigonometry from scratch (Section 4.2).

What you will meet at university

The first-year *Analysis* course proves the fundamental theorem of calculus properly, from the definition of the integral. Later courses build in several directions. *Measure theory* and the *Lebesgue integral* make integration work smoothly with limits and underpin probability. *Vector calculus* integrates over curves, surfaces and volumes, and Green’s and Stokes’s theorems relate them (Section 4.9 is a taste). *Complex analysis* evaluates real integrals by detours through the complex plane. The *calculus of variations* asks which function makes an integral as small as possible (Section 4.7), and leads to the principle of least action in physics. And *differential algebra* explains which integrals have formulas at all (Section 4.6).

The questions that run through this chapter

Three big questions

- 1. **What is an integral, and which functions have one?** How does the definition as a limit of sums connect with antiderivatives?
- 2. **What can integrals prove?** How can an area tell us about π , about factorials, about infinite sums?
- 3. **What can we do when there is no formula?** How do we know an integral has no formula, and what can we compute anyway?

Ideas in this chapter

§	Topic	Idea
4.1	The definition of the integral	Lower and upper sums; telescoping
4.2	Trigonometry from $\int \frac{du}{1+u^2}$	Substitutions that preserve an integral
4.3	$\frac{22}{7} > \pi$	An inequality as an integral of something positive
4.4	Wallis and $\left(\frac{1}{2}\right)!$	Reduction formulae, an invariant, a squeeze
4.5	$\int_0^1 x^x \, dx$	Series expansion and term-by-term integration
4.6	Integrals with no formula	Impossibility via polynomial algebra
4.7	The calculus of variations	Completing the square in an integral
4.8	Archimedes’ hat-box theorem	Simplifying an integrand reveals geometry
4.9	Holditch’s theorem	Area from a boundary integral; elimination

How to read this chapter

Several sections involve heavy but routine calculation: polynomial division, integration by parts, expanding products. Do these yourself, with pencil and paper, before reading the text. Then look for the idea that made the calculation work: an invariant, a symmetry, a cancellation. The calculation is what an examiner checks; the idea is what you will remember and use again.

4.1 What Is an Integral?

The question

(i) Prove that

$$\sum_{r=1}^n r(r+1)(r+2)(r+3)(r+4) = \frac{1}{6}n(n+1)(n+2)(n+3)(n+4)(n+5)$$

and deduce that

$$\sum_{r=1}^n r^5 < \frac{1}{6}n(n+1)(n+2)(n+3)(n+4)(n+5).$$

(ii) Prove that, if $n > 1$,

$$\sum_{r=0}^{n-1} r^5 > \frac{1}{6}(n-5)(n-4)(n-3)(n-2)(n-1)n.$$

(iii) Let f be an increasing function. If the limits

$$\lim_{n \rightarrow \infty} \sum_{r=0}^{n-1} \frac{a}{n} f\left(\frac{ra}{n}\right) \quad \text{and} \quad \lim_{n \rightarrow \infty} \sum_{r=1}^n \frac{a}{n} f\left(\frac{ra}{n}\right)$$

both exist and are equal, the definite integral $\int_0^a f(x) dx$ is defined to be their common value. Using this definition, prove that

$$\int_0^a x^5 dx = \frac{1}{6}a^6.$$

The hook

You have been integrating for years. You know that $\int_0^a x^5 dx = \frac{1}{6}a^6$ because “integration is the reverse of differentiation”. But that is a theorem, not a definition. What *is* an integral?

The honest answer is: an area, defined as a limit of sums of thin rectangles. This question asks you to compute an integral from that definition alone, with no antiderivatives, no fundamental theorem, nothing but adding up rectangles and taking a limit. It is harder than you might expect, which is exactly why the fundamental theorem of calculus,

which makes it easy, is such a remarkable result.

Before you start

Ask yourself The two sums in part (iii) are sums of areas of rectangles of width $\frac{a}{n}$. For an increasing function f , which sum is an underestimate of the area under the curve, and which an overestimate?

The first sum uses the height at the *left* end of each strip, $x = \frac{ra}{n}$ for $r = 0, \dots, n-1$. For an increasing function that is the lowest point on the strip, so the rectangles fit under the curve: a *lower sum*. The second uses the right ends, $r = 1, \dots, n$, giving rectangles that poke above the curve: an *upper sum* (Figure 4.1.1). The definition says the integral exists if both sums approach the same limit as the strips get thinner.

For $f(x) = x^5$ both sums involve $\sum r^5$. Parts (i) and (ii) bound that sum above and below.

Working it through

STAGE 1 A sum that telescopes

Ask yourself Write $P(r) = r(r+1)(r+2)(r+3)(r+4)(r+5)$, a product of six consecutive integers. What is $P(r) - P(r-1)$?

$P(r)$ and $P(r-1) = (r-1)r(r+1)(r+2)(r+3)(r+4)$ share the five factors $r(r+1)(r+2)(r+3)(r+4)$, so

$$\begin{aligned} P(r) - P(r-1) &= r(r+1)(r+2)(r+3)(r+4)[(r+5) - (r-1)] \\ &= 6r(r+1)(r+2)(r+3)(r+4). \end{aligned}$$

Summing from $r = 1$ to n , the right-hand sides telescope to $P(n) - P(0) = P(n)$. So

$$\sum_{r=1}^n r(r+1)(r+2)(r+3)(r+4) = \frac{1}{6}P(n) = \frac{1}{6}n(n+1)(n+2)(n+3)(n+4)(n+5).$$

Since $r^5 < r(r+1)(r+2)(r+3)(r+4)$ for every $r \geq 1$, the inequality follows. ■

STAGE 2 A lower bound

Pause and try Adapt Stage 1 to a product of five consecutive integers going down: $r(r-1)(r-2)(r-3)(r-4)$. Find its sum from $r = 0$ to $n-1$, and compare it with r^5 . ■

Let $Q(r) = (r-5)(r-4)(r-3)(r-2)(r-1)r$. Then $Q(r+1) - Q(r) = (r-4)(r-3)(r-2)(r-1)r[(r+1) - (r-5)] = 6(r-4)(r-3)(r-2)(r-1)r$. Summing from $r = 0$ to $n-1$ telescopes to $Q(n) - Q(0) = Q(n)$, so

$$\sum_{r=0}^{n-1} (r-4)(r-3)(r-2)(r-1)r = \frac{1}{6}(n-5)(n-4)(n-3)(n-2)(n-1)n.$$

Now compare term by term. For $r = 0$ both sides are 0. For $1 \leq r \leq 4$ the product contains a zero factor, while $r^5 > 0$. For $r \geq 5$ every factor is positive and at most r , and some are strictly smaller, so the product is less than r^5 . Since $n > 1$, the sum includes $r = 1$, where the inequality is strict, so

$$\sum_{r=0}^{n-1} r^5 > \frac{1}{6}(n-5)(n-4)(n-3)(n-2)(n-1)n. \quad \blacksquare$$

STAGE 3 The integral



Ask yourself Write the lower and upper sums for $f(x) = x^5$ in terms of $\sum r^5$. What power of n appears in the denominator?

With $f(x) = x^5$,

$$L_n = \sum_{r=0}^{n-1} \frac{a}{n} \left(\frac{ra}{n} \right)^5 = \frac{a^6}{n^6} \sum_{r=0}^{n-1} r^5, \quad U_n = \frac{a^6}{n^6} \sum_{r=1}^n r^5.$$

By parts (i) and (ii), and since $L_n < U_n$,

$$\frac{a^6}{6} \cdot \frac{(n-5)(n-4) \cdots n}{n^6} < L_n < U_n < \frac{a^6}{6} \cdot \frac{n(n+1) \cdots (n+5)}{n^6}.$$

Each of the two fractions is a product of six factors like $\frac{n-3}{n} = 1 - \frac{3}{n}$ or $\frac{n+3}{n} = 1 + \frac{3}{n}$, each tending to 1. So both outer bounds tend to $\frac{a^6}{6}$, and by the squeeze, so do L_n and U_n . Both limits exist and are equal, so by the definition

$$\int_0^a x^5 dx = \frac{1}{6}a^6. \quad \blacksquare$$

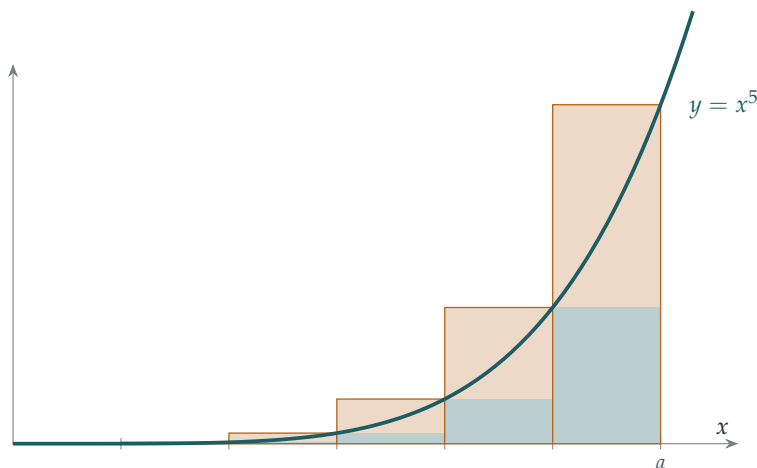


Figure 4.1.1 Lower and upper sums for x^5 on $[0, a]$ with $n = 6$ strips. The lower sum is the teal area (rectangles under the curve); the upper sum adds the amber pieces. The difference between them is exactly the area of the last, tallest rectangle, $\frac{a}{n} f(a)$, which tends to 0 as n grows.

Looking back

- **Telescoping again.** The sums of products of consecutive integers collapsed, exactly like the sums and products in Chapter 2.
- **Bound, then squeeze.** We never found $\sum r^5$ exactly (it is $\frac{1}{12}n^2(n+1)^2(2n^2+2n-1)$); bounds on both sides were enough.
- **The definition does the work.** No antiderivative appeared anywhere. Everything followed from the definition of the integral as a limit of sums.

What's really going on

Riemann's definition

The idea of finding areas by summing thin strips goes back to Archimedes, who found the area under a parabola by a version of exactly this argument around 250 BC. Newton and Leibniz turned integration into the reverse of differentiation, and for a century and a half that was how most people thought about it. In 1854 Bernhard Riemann gave the definition that university courses still begin with: divide $[a, b]$ into small pieces, form sums of (width) $\times$ (height of f somewhere in the piece), and say f is integrable if all such sums approach one limit as the pieces shrink. The version in this question, with equal strips and left and right endpoints, is a special case that works for increasing functions; Jean-Gaston Darboux's "upper and lower sums" make the general case clean.

Not every function has an integral

With a precise definition, you can ask which functions are integrable. Every continuous function on a closed interval is, and every increasing function. But consider Dirichlet's function: $f(x) = 1$ if x is rational and 0 if irrational. Every strip, however thin, contains both rationals and irrationals, so every upper sum is 1 and every lower sum is 0. They never agree, and $\int_0^1 f$ does not exist in Riemann's sense. In 1902 Henri Lebesgue invented a more powerful integral that handles this function (its Lebesgue integral is 0) and behaves much better under limits. The Lebesgue integral is the foundation of modern probability theory, and you will meet it in your second or third year.

Discrete calculus

The telescoping in Stage 1 is calculus in disguise. The “rising factorial” $r(r+1) \cdots (r+k-1)$ behaves under summation exactly as x^k behaves under integration: summing it raises the number of factors by one and divides by the new number of factors, just as $\int x^k dx = \frac{x^{k+1}}{k+1}$. This parallel between sums and integrals, differences and derivatives, is called the *calculus of finite differences*, and it is a powerful tool for evaluating sums.

The thinking skill

Work from the definition

When asked to prove something “using the definition”, use nothing else: no theorems that depend on what you are proving, no familiar shortcuts.

It is surprisingly hard, and it shows you exactly what the shortcuts are worth. At university, the first weeks of every course work this way: definitions first, then theorems that let you stop working from definitions.

Try this next

1 A limit as an integral. (From STEP III, 1999.) Justify by a sketch that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=1}^n f\left(1 + \frac{m}{n}\right) = \int_1^2 f(x) dx.$$

Deduce that

$$\lim_{n \rightarrow \infty} \left(\frac{1}{n+1} + \frac{1}{n+2} + \cdots + \frac{1}{2n} \right) = \ln 2,$$

and evaluate $\lim_{n \rightarrow \infty} \left(\frac{n}{n^2+1} + \frac{n}{n^2+4} + \cdots + \frac{n}{n^2+n^2} \right)$.

2 Archimedes' parabola. Using $\sum_{r=1}^n r^2 = \frac{1}{6}n(n+1)(2n+1)$ and the definition in part (iii), prove that $\int_0^a x^2 dx = \frac{1}{3}a^3$.

- 3 Rising factorials.** Show that $\sum_{r=1}^n r(r+1)\cdots(r+k-1) = \frac{n(n+1)\cdots(n+k)}{k+1}$ for every positive integer k .
- 4 The exact sum.** Use the identity $r(r+1)(r+2)(r+3)(r+4) = r^5 + 10r^4 + 35r^3 + 50r^2 + 24r$ and the known formulas for $\sum r^k$ with $k \leq 4$ to find $\sum_{r=1}^n r^5$ exactly.

Further reading

- **David Bressoud, *A Radical Approach to Lebesgue's Theory of Integration*** (Cambridge University Press). The story from Riemann's integral to Lebesgue's, told through the problems that forced each step.
- **Michael Spivak, *Calculus*** (Cambridge University Press). Chapter 13 defines the integral by upper and lower sums, exactly as here.
- **Graham, Knuth & Patashnik, *Concrete Mathematics*** (Addison-Wesley). Chapter 2 develops the calculus of finite differences, with rising and falling factorials.

4.2 Building Trigonometry from an Integral

The question

In this question, you are not permitted to use any properties of trigonometric functions or inverse trigonometric functions.

The function T is defined for $x > 0$ by

$$T(x) = \int_0^x \frac{1}{1+u^2} du,$$

and $T_\infty = \int_0^\infty \frac{1}{1+u^2} du$ (which has a finite value).

(i) By making an appropriate substitution in the integral for $T(x)$, show that

$$T(x) = T_\infty - T(x^{-1}).$$

(ii) Let $v = \frac{u+a}{1-au}$, where a is a constant. Verify that, for $u \neq a^{-1}$,

$$\frac{dv}{du} = \frac{1+v^2}{1+u^2}.$$

Hence show that, for $a > 0$ and $x < \frac{1}{a}$,

$$T(x) = T\left(\frac{x+a}{1-ax}\right) - T(a).$$

Deduce that

$$T(x^{-1}) = 2T_\infty - T\left(\frac{x+a}{1-ax}\right) - T(a^{-1})$$

and hence that, for $b > 0$ and $y > \frac{1}{b}$,

$$T(y) = 2T_\infty - T\left(\frac{y+b}{by-1}\right) - T(b).$$

(iii) Use the above results to show that $T(\sqrt{3}) = \frac{2}{3}T_\infty$ and $T(\sqrt{2}-1) = \frac{1}{4}T_\infty$.

The hook

What is π ? At school it is the ratio of a circle's circumference to its diameter, and $\sin$ and $\cos$ are defined by triangles. That is fine for geometry, but it leaves awkward questions. What exactly is "length" along a curve? Why does $\sin x$ have the Maclaurin series it does? How do we know the addition formulas hold for all angles, not just ones we can draw?

This question shows a completely different way in. Forget triangles and circles. Start from a single integral, $T(x) = \int_0^x \frac{du}{1+u^2}$, which needs nothing but the definition of the integral from Section 4.1. Define π to be $2T_\infty$. Then, using only substitutions, prove that $T(\sqrt{3}) = \frac{\pi}{3}$ and $T(\sqrt{2}-1) = \frac{\pi}{8}$, facts you know as $\arctan \sqrt{3} = 60^\circ$ and $\tan 22\frac{1}{2}^\circ = \sqrt{2}-1$. Trigonometry rebuilt from nothing.

Before you start

You know, secretly, that $T(x) = \arctan x$ and $T_\infty = \frac{\pi}{2}$. The question forbids you to use this, and that is the point: everything must come from the integral. But your secret knowledge is a useful guide.



Ask yourself Which trigonometric identities do parts (i) and (ii) correspond to? (Think of $\arctan x + \arctan \frac{1}{x}$, and the formula for $\tan(A+B)$.)

Part (i) is $\arctan x + \arctan \frac{1}{x} = \frac{\pi}{2}$. Part (ii) is the addition formula $\tan(A+B) = \frac{\tan A + \tan B}{1 - \tan A \tan B}$ in disguise: with $u = \tan A$ and $a = \tan B$, $v = \tan(A+B)$. The question proves these identities from the integral, without any geometry.

Working it through

STAGE 1 The reflection $u \mapsto 1/u$

Substitute $u = 1/w$ in $T(x)$. Then $du = -dw/w^2$, and $\frac{1}{1+u^2} = \frac{w^2}{w^2+1}$, so

$$T(x) = \int_\infty^{1/x} \frac{w^2}{w^2+1} \cdot \left(-\frac{1}{w^2}\right) dw = \int_{1/x}^\infty \frac{dw}{1+w^2} = T_\infty - T(x^{-1}). \quad \blacksquare$$

In other words, the area under $\frac{1}{1+u^2}$ from 0 to x equals the area from $\frac{1}{x}$ to ∞ (Figure 4.2.1).

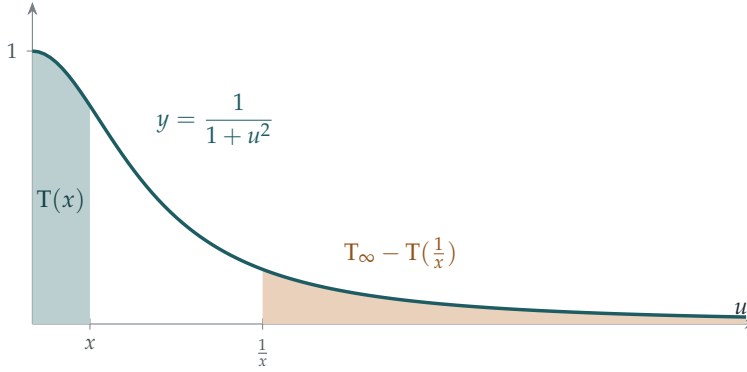


Figure 4.2.1 The substitution $u = 1/w$ shows that the two shaded areas under $\frac{1}{1+u^2}$ are equal: the area from 0 to x (teal) and the infinitely long area from $\frac{1}{x}$ onwards (amber). Here $x = \frac{1}{2}$.

STAGE 2 The addition substitution

? **Ask yourself** Differentiate $v = \frac{u+a}{1-au}$ by the quotient rule. Then compute $1+v^2$ as a single fraction and compare.

$$\begin{aligned}\frac{dv}{du} &= \frac{(1-au) + a(u+a)}{(1-au)^2} = \frac{1+a^2}{(1-au)^2}, \\ 1+v^2 &= \frac{(1-au)^2 + (u+a)^2}{(1-au)^2} = \frac{(1+a^2)(1+u^2)}{(1-au)^2}.\end{aligned}$$

Dividing, $\frac{1+v^2}{1+u^2} = \frac{1+a^2}{(1-au)^2} = \frac{dv}{du}$. ■

? **Ask yourself** This says $\frac{du}{1+u^2} = \frac{dv}{1+v^2}$. So the substitution $u \mapsto v$ doesn't change the integrand at all. What does it do to the limits of $T(x)$?

As u runs from 0 to x (with $x < \frac{1}{a}$, so $1-au > 0$ throughout and v is continuous and increasing), v runs from a to $\frac{x+a}{1-ax}$. So

$$T(x) = \int_0^x \frac{du}{1+u^2} = \int_a^{(x+a)/(1-ax)} \frac{dv}{1+v^2} = T\left(\frac{x+a}{1-ax}\right) - T(a). \quad \blacksquare$$

STAGE 3 Two more identities

Combining with part (i), twice:

$$T(x^{-1}) = T_\infty - T(x) = T_\infty - T\left(\frac{x+a}{1-ax}\right) + T(a) = 2T_\infty - T\left(\frac{x+a}{1-ax}\right) - T(a^{-1}). \quad \blacksquare$$

Now rename: put $y = 1/x$ and $b = 1/a$. The condition $x < 1/a$ becomes $1/y < b$, that is $y > 1/b$, and

$$\frac{x+a}{1-ax} = \frac{1/y + 1/b}{1 - 1/(by)} = \frac{b+y}{by-1}.$$

So $T(y) = 2T_\infty - T\left(\frac{y+b}{by-1}\right) - T(b).$ ■

STAGE 4 Special values

🕒 **Pause and try** Choose y and b in the last identity so that the argument $\frac{y+b}{by-1}$ equals y itself. What do you get?

Take $y = b = \sqrt{3}$: then $y > \frac{1}{b}$, and $\frac{\sqrt{3} + \sqrt{3}}{3 - 1} = \sqrt{3}$. The identity becomes $T(\sqrt{3}) = 2T_\infty - 2T(\sqrt{3})$, so

$$T(\sqrt{3}) = \frac{2}{3}T_\infty. \quad \blacksquare$$

❓ **Ask yourself** For $\sqrt{2} - 1$, try the addition identity with $x = a = \sqrt{2} - 1$. What is $\frac{2x}{1-x^2}$? And what is $T(1)$?

With $x = a = \sqrt{2} - 1$ (and $x < \frac{1}{a}$ since $x^2 = 3 - 2\sqrt{2} < 1$), the addition identity says $2T(x) = T\left(\frac{2x}{1-x^2}\right)$. Now $1 - x^2 = 2\sqrt{2} - 2 = 2(\sqrt{2} - 1) = 2x$, so $\frac{2x}{1-x^2} = 1$, and $2T(\sqrt{2} - 1) = T(1)$. Finally, part (i) with $x = 1$ gives $T(1) = T_\infty - T(1)$, so $T(1) = \frac{1}{2}T_\infty$. Therefore

$$T(\sqrt{2} - 1) = \frac{1}{4}T_\infty. \quad \blacksquare$$

Looking back

- **Substitutions that preserve the integrand.** Both key steps used substitutions under which $\frac{du}{1+u^2}$ is unchanged; only the limits moved. Such “symmetries” of an integral are powerful.
- **Fixed points.** The value $T(\sqrt{3})$ came from choosing numbers that made an expression equal to itself, so an identity became an equation.
- **Definitions can come from analysis.** We never used a triangle, yet we recovered $\frac{\pi}{3}$ and $\frac{\pi}{8}$ (with π defined as $2T_\infty$).

What's really going on

Building the elementary functions

In a rigorous analysis course, the functions $\exp$, $\sin$, $\cos$ and $\ln$ cannot simply be assumed; they have to be *constructed*, and their properties proved. There are three standard approaches. You can define them by power series ($\sin x = x - \frac{x^3}{3!} + \dots$). You can define them as solutions of differential equations, as in the 1986 specimen question in the exercises ($f' = f$, $f(0) = 1$ defines $\exp$). Or you can define inverse functions by integrals, as here: $\ln x = \int_1^x \frac{du}{u}$ and $\arctan x = \int_0^x \frac{du}{1+u^2}$, and then define $\pi = 2 \int_0^\infty \frac{du}{1+u^2}$. Each approach has its strengths, and proving that they all give the same functions is a satisfying exercise. Compare the functional-equation approach of Section 3.8.

Beyond the circle: elliptic integrals

The integral approach generalises in a way the triangle approach cannot. In 1718 Giulio Fagnano studied the arc length of the *lemniscate*, a figure-of-eight curve, which leads to $\int_0^x \frac{du}{\sqrt{1-u^4}}$, and found a substitution that doubles the arc, just as our substitution doubled T . Euler, reading Fagnano's work in 1751, found a full addition formula, and this led eventually, through Abel and Jacobi, to the theory of *elliptic functions*: inverse functions of integrals like this one, which are to the lemniscate what $\sin$ and $\cos$ are to the circle. We met the same integral in Gauss's work on the AGM (Section 2.6). Elliptic functions are now central to number theory; the elliptic curves of Section 1.9 are their geometric shadow.

Computing π

The addition identity is also the basis of the most famous formulas for computing π . Repeatedly applying it shows, for instance, that $T(1) = 4T(\frac{1}{5}) - T(\frac{1}{239})$, which is Machin's formula of 1706; combined with the series $T(x) = x - \frac{x^3}{3} + \frac{x^5}{5} - \dots$, it gave 100 digits of π and stayed the method of choice for two and a half centuries. That is the subject of Section 5.1.

The thinking skill

Symmetries of an integral

Look for substitutions that leave the integrand unchanged. They turn into identities for the function the integral defines.

Here $u \mapsto 1/u$ and $u \mapsto \frac{u+a}{1-au}$ preserve $\frac{du}{1+u^2}$, and they give the reflection and addition formulas. The same idea, applied to other integrals, underlies the theory of special functions.

Try this next

- 1 The exponential and the cosine.** (From the STEP II Specimen paper.) *Without using standard properties of exponential or trigonometric functions:* a function f satisfies $f' = f$ with $f(0) = 1$. Show that $\frac{d}{dx} [f(a-x)f(x)] = 0$ and deduce that $f(x)f(y) = f(x+y)$. Functions c and s satisfy $c' = -s$, $s' = c$, $s(0) = 0$, $c(0) = 1$. Show that $c(x+y) = c(x)c(y) - s(x)s(y)$. (*Hint:* fix y and consider $c(x+y) - c(x)c(y) + s(x)s(y)$ and a similar combination for s ; show the sum of their squares is constant.)
- 2 The inverse sine.** (From STEP III, 2007.) Let $S(x) = \int_0^x \frac{du}{\sqrt{1-u^2}}$. Using $y = \frac{u}{\sqrt{1+u^2}}$, show that $T(x) = S\left(\frac{x}{\sqrt{1+x^2}}\right)$ and deduce that $S\left(\frac{1}{\sqrt{2}}\right) = \frac{1}{2}T_\infty$.
- 3 An early Machin-type formula.** Use the addition identity to show that $T\left(\frac{1}{2}\right) + T\left(\frac{1}{3}\right) = T(1) = \frac{1}{2}T_\infty$.
- 4 T_∞ is finite.** Show that $T(x) \leq 1 + \int_1^x \frac{du}{u^2} < 2$ for all $x > 0$, and deduce from the monotone convergence theorem (Section 2.2) that T_∞ exists.

Further reading

- **Tom Körner, *Calculus for the Ambitious*** (Cambridge University Press). Builds the elementary functions rigorously, discussing the different approaches.
- **Walter Rudin, *Principles of Mathematical Analysis*** (McGraw-Hill). The famous “baby Rudin”; the prologue to Chapter 8 constructs $\exp$, $\sin$ and $\cos$ from power series in a few elegant pages. Advanced.
- **Victor Prasolov & Yuri Solov'yev, *Elliptic Functions and Elliptic Integrals*** (American Mathematical Society). For where the lemniscate leads. Advanced.

4.3 $22/7$ Is Bigger Than π

The question

Find constants $a_0, a_1, a_2, a_3, a_4, a_5, a_6$ and b such that

$$x^4(1-x)^4 = (a_6x^6 + a_5x^5 + a_4x^4 + a_3x^3 + a_2x^2 + a_1x + a_0)(x^2 + 1) + b.$$

Hence, or otherwise, prove that

$$\int_0^1 \frac{x^4(1-x)^4}{1+x^2} dx = \frac{22}{7} - \pi.$$

Evaluate $\int_0^1 x^4(1-x)^4 dx$ and deduce that

$$\frac{22}{7} > \pi > \frac{22}{7} - \frac{1}{630}.$$

The hook

Generations of students have been told that π is “about $\frac{22}{7}$ ”. It is a very good approximation: $\frac{22}{7} = 3.142857\dots$ and $\pi = 3.141592\dots$. But is $\frac{22}{7}$ too big or too small? And can you *prove* it, without simply looking up the digits of π , which someone else had to compute?

This question gives one of the most elegant proofs in all of mathematics. A single integral of a positive function, which therefore must be positive, turns out to equal exactly $\frac{22}{7} - \pi$. Positive integrand, positive integral, so $\frac{22}{7} > \pi$. The whole proof fits in a few lines, and it is so neat that it appears as a puzzle on mathematics T-shirts.

Before you start



Ask yourself Why does π come out of $\int_0^1 \frac{\text{polynomial}}{1+x^2} dx$? Which simple integral of this form gives π on its own?

$\int_0^1 \frac{dx}{1+x^2} = \arctan 1 = \frac{\pi}{4}$. When you divide a polynomial by $1+x^2$, you get a polynomial quotient (which integrates to a rational number) plus a constant remainder over $1+x^2$ (which integrates to a rational multiple of π). The first part of the question asks you to do exactly that division.

Working it through

STAGE 1 Polynomial division

First expand: $(1-x)^4 = 1 - 4x + 6x^2 - 4x^3 + x^4$, so

$$x^4(1-x)^4 = x^8 - 4x^7 + 6x^6 - 4x^5 + x^4.$$

? **Ask yourself** Divide by $x^2 + 1$, either by long division or by comparing coefficients. What is the remainder?

Comparing coefficients in $(a_6x^6 + \cdots + a_0)(x^2 + 1) + b$, from the top: $a_6 = 1$; $a_5 = -4$; $a_4 + a_6 = 6$, so $a_4 = 5$; $a_3 + a_5 = -4$, so $a_3 = 0$; $a_2 + a_4 = 1$, so $a_2 = -4$; $a_1 + a_3 = 0$, so $a_1 = 0$; $a_0 + a_2 = 0$, so $a_0 = 4$; and finally $a_1 = 0$ (the coefficient of x) and $a_0 + b = 0$, so $b = -4$. That is,

$$x^4(1-x)^4 = (x^6 - 4x^5 + 5x^4 - 4x^2 + 4)(x^2 + 1) - 4.$$

STAGE 2 The integral

Dividing by $1 + x^2$,

$$\frac{x^4(1-x)^4}{1+x^2} = x^6 - 4x^5 + 5x^4 - 4x^2 + 4 - \frac{4}{1+x^2},$$

so

$$\int_0^1 \frac{x^4(1-x)^4}{1+x^2} dx = \frac{1}{7} - \frac{4}{6} + \frac{5}{5} - \frac{4}{3} + 4 - 4 \cdot \frac{\pi}{4} = \frac{1}{7} - \frac{2}{3} + 1 - \frac{4}{3} + 4 - \pi.$$

Now $-\frac{2}{3} - \frac{4}{3} = -2$, so the rational part is $\frac{1}{7} + 3 = \frac{22}{7}$, and the integral is $\frac{22}{7} - \pi$. ■

STAGE 3 Bounding the integral

? **Ask yourself** The integrand is positive on $(0, 1)$, so $\frac{22}{7} > \pi$ at once. To get a lower bound for π we need an *upper* bound for the integral. How big can $\frac{1}{1+x^2}$ be on $[0, 1]$?

On $[0, 1]$, $1 \leq 1 + x^2 \leq 2$, so

$$\frac{1}{2}x^4(1-x)^4 \leq \frac{x^4(1-x)^4}{1+x^2} \leq x^4(1-x)^4,$$

with strict inequality except at the endpoints.

🔍 Pause and try Evaluate $\int_0^1 x^4(1-x)^4 dx$, and finish the proof.

Integrating the expansion term by term,

$$\int_0^1 x^4(1-x)^4 dx = \frac{1}{9} - \frac{4}{8} + \frac{6}{7} - \frac{4}{6} + \frac{1}{5} = \frac{1}{630}.$$

(Over the common denominator 2520: $280 - 1260 + 2160 - 1680 + 504 = 4$, and $\frac{4}{2520} = \frac{1}{630}$.) So

$$0 < \frac{22}{7} - \pi < \frac{1}{630},$$

which is $\frac{22}{7} > \pi > \frac{22}{7} - \frac{1}{630}$. ■

Numerically, $\frac{22}{7} - \frac{1}{630} = 3.14127\dots$, so we have proved $3.1412 < \pi < 3.1429$. Using the left-hand inequality as well, $\frac{1}{1260} < \frac{22}{7} - \pi$, which improves the upper bound to $\pi < 3.14207$.

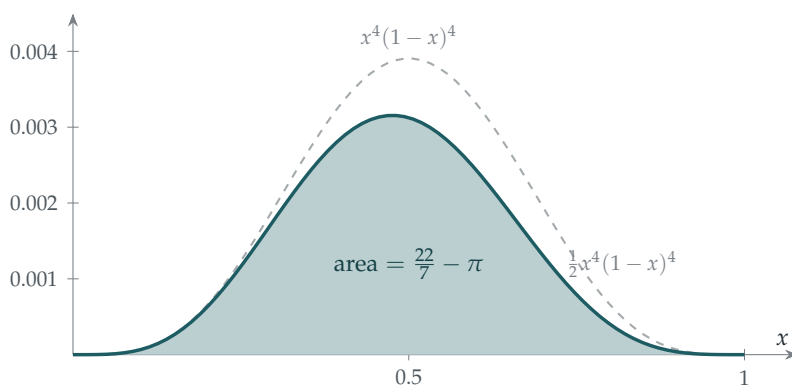


Figure 4.3.1 The integrand $\frac{x^4(1-x)^4}{1+x^2}$ (teal) is positive but tiny: its maximum is about 0.003. The shaded area is exactly $\frac{22}{7} - \pi \approx 0.00126$. The integrand lies between $\frac{1}{2}x^4(1-x)^4$ (dotted) and $x^4(1-x)^4$ (dashed), whose areas are $\frac{1}{1260}$ and $\frac{1}{630}$.

Looking back

- **Positivity proves inequalities.** The integral of a positive function is positive. Once the integral was identified as $\frac{22}{7} - \pi$, the inequality was immediate.
- **Small integrands, sharp bounds.** The factor $x^4(1-x)^4$ is tiny on $[0, 1]$ (at most $\frac{1}{256}$), which makes the integral small and the bound sharp.
- **Rational plus a multiple of π .** As in Section 2.9, an integral of the form “rational number + rational multiple of a constant” that is small gives a good rational approximation to the constant.

What's really going on

A proof with a history

Archimedes proved $\frac{223}{71} < \pi < \frac{22}{7}$ around 250 BC by computing the perimeters of 96-sided polygons inside and outside a circle (compare the polygons of Section 2.5). The integral proof is much more recent. It was published by D.P. Dalzell in 1944, and became famous when it was set as a problem in the 1968 Putnam competition for American undergraduates. Its charm is that it needs no numerical work at all: the inequality $\frac{22}{7} > \pi$ drops out of the fact that an integral of a positive function is positive.

Better approximations from the same idea

The method generalises. With $x^8(1-x)^8$ in place of $x^4(1-x)^4$, the same division gives

$$\int_0^1 \frac{x^8(1-x)^8}{1+x^2} dx = 4\pi - \frac{188684}{15015},$$

which proves $\pi > \frac{47171}{15015} = 3.1415917\dots$, correct to six significant figures. In 2009 Stephen Lucas found families of integrals of this kind with non-negative integrands that prove inequalities for other famous approximations, including $\frac{355}{113}$, which is correct to seven significant figures.

Approximations and irrationality

Why does π have such good fractional approximations as $\frac{22}{7}$ and $\frac{355}{113}$? Every irrational number has infinitely many fractions $\frac{p}{q}$ within $\frac{1}{q^2}$ of it, and π happens to have some exceptionally good ones early on, because its continued fraction $[3; 7, 15, 1, 292, \dots]$ contains the large entry 292 (see Section 2.2). Integrals like the one in this question, made smaller and smaller, are the starting point for proofs that numbers *cannot* be approximated too well by fractions, and so are irrational: Ivan Niven's famous one-page proof that π is irrational (1947) uses integrals of $\frac{x^n(1-x)^n}{n!} \sin(\pi x)$, the trigonometric cousins of those in Section 2.9.

The thinking skill

Encode the inequality as an integral of something positive

To prove $A > B$, find a positive function whose integral is $A - B$.

This turns an inequality into an identity plus an obvious fact. It is one of the most useful tricks in analysis, and the same idea, with sums of squares instead of integrals, proved AM–GM in Section 3.2.

Try this next

1 Check the numbers. Compute $\frac{22}{7} - \pi$ to four significant figures and confirm that it lies between $\frac{1}{1260}$ and $\frac{1}{630}$.

2 A Beta integral. Use the formula $\int_0^1 x^n(1-x)^n dx = \frac{(n!)^2}{(2n+1)!}$ (proved in Section 4.4) to check that $\int_0^1 x^4(1-x)^4 dx = \frac{1}{630}$.

3 A lower approximation. Show that

$$\int_0^1 \frac{x(1-x)^2}{1+x^2} dx = \frac{\pi}{2} - \frac{3}{2} \quad \text{and} \quad \int_0^1 \frac{x^2(1-x)^4}{1+x^2} dx = \pi - \frac{47}{15},$$

and deduce that $\pi > 3$ and $\pi > \frac{47}{15}$. Why are these less accurate than $\frac{22}{7}$?

4 Going further. **HARDER** Verify the identity for $\int_0^1 \frac{x^8(1-x)^8}{1+x^2} dx$ given above. (A computer algebra system will do the division; the point is to see that the remainder is 16, giving 4π .)

Further reading

- **Petr Beckmann, *A History of π*** (St Martin's Press). Archimedes' polygon method in detail, and much else.
- **Stephen Lucas, "Approximations to π derived from integrals with nonnegative integrands", *American Mathematical Monthly* 116 (2009).** Short and readable.
- **Ivan Niven, *Irrational Numbers*** (Mathematical Association of America). Includes Niven's proof that π is irrational.

4.4 Wallis, Beta and the Factorial of a Half

The question

Let

$$u_n = \int_0^{\frac{1}{2}\pi} \sin^n t \, dt$$

for each integer $n \geq 0$. By integrating

$$\int_0^{\frac{1}{2}\pi} \sin t \sin^{n-1} t \, dt$$

by parts, or otherwise, obtain a formula connecting u_n and u_{n-2} when $n \geq 2$ and deduce that

$$nu_n u_{n-1} = (n-1)u_{n-1}u_{n-2}$$

for all $n \geq 2$. Deduce that

$$nu_n u_{n-1} = \frac{1}{2}\pi.$$

Sketch graphs of $\sin^n t$ and $\sin^{n-1} t$, for $0 \leq t \leq \frac{1}{2}\pi$, on the same diagram and explain why $0 < u_n < u_{n-1}$. By using the result of the previous paragraph show that

$$nu_n^2 < \frac{1}{2}\pi < nu_{n-1}^2$$

for all $n \geq 1$. Hence show that

$$\left(\frac{n}{n+1}\right) \frac{1}{2}\pi < nu_n^2 < \frac{1}{2}\pi$$

and deduce that $nu_n^2 \rightarrow \frac{1}{2}\pi$ as $n \rightarrow \infty$.

The hook

What is $\left(\frac{1}{2}\right)!$? The question sounds like nonsense: $n!$ means $1 \times 2 \times \cdots \times n$, and you cannot multiply “half a number” of terms. Yet there is a natural answer, and it is

$$\left(\frac{1}{2}\right)! = \frac{\sqrt{\pi}}{2}.$$

Where does π come from? From integrals of powers of $\sin t$, which this question studies. Along the way it proves one of the oldest and strangest formulas for π , John Wallis’s infinite product of 1655,

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

and it finds the hidden constant in Stirling's formula for $n!$, which every probabilist and statistician uses.

Before you start



Ask yourself What are u_0 and u_1 ? And roughly what does the graph of $\sin^n t$ look like on $[0, \frac{\pi}{2}]$ when n is large?

$u_0 = \int_0^{\pi/2} 1 dt = \frac{\pi}{2}$ and $u_1 = \int_0^{\pi/2} \sin t dt = 1$. For large n , $\sin^n t$ is tiny except near $t = \frac{\pi}{2}$, where $\sin t$ is close to 1: the graph becomes a narrow spike against the right-hand edge (Figure 4.4.1). So u_n shrinks as n grows. The question finds exactly how fast: like $\sqrt{\pi/(2n)}$.

Working it through

STAGE 1 A reduction formula



Ask yourself Integrate $\int \sin t \cdot \sin^{n-1} t dt$ by parts, integrating $\sin t$ and differentiating $\sin^{n-1} t$. The new integrand contains $\cos^2 t$; how can you get back to powers of $\sin t$?

With $dv = \sin t dt$ (so $v = -\cos t$) and $w = \sin^{n-1} t$ (so $w' = (n-1) \sin^{n-2} t \cos t$),

$$u_n = \left[-\cos t \sin^{n-1} t \right]_0^{\pi/2} + (n-1) \int_0^{\pi/2} \sin^{n-2} t \cos^2 t dt.$$

The boundary term is zero for $n \geq 2$. Writing $\cos^2 t = 1 - \sin^2 t$, the integral is $u_{n-2} - u_n$, so $u_n = (n-1)(u_{n-2} - u_n)$, that is,

$$nu_n = (n-1)u_{n-2}. \quad \blacksquare$$

Multiplying both sides by u_{n-1} gives $nu_n u_{n-1} = (n-1)u_{n-1} u_{n-2}$. $\blacksquare$

STAGE 2 An invariant



Ask yourself The relation says the quantity $nu_n u_{n-1}$ does not change when n goes down by one. What is it equal to at $n = 1$?

By repeated use, $nu_n u_{n-1} = (n-1)u_{n-1} u_{n-2} = \cdots = 1 \cdot u_1 u_0 = 1 \cdot 1 \cdot \frac{\pi}{2}$. So $nu_n u_{n-1} = \frac{1}{2}\pi$ for all $n \geq 1$. $\blacksquare$

STAGE 3 A squeeze

For $0 < t < \frac{\pi}{2}$ we have $0 < \sin t < 1$, so $\sin^n t < \sin^{n-1} t$: the graph of $\sin^n t$ lies below that of $\sin^{n-1} t$ except at the endpoints. Integrating, $0 < u_n < u_{n-1}$.

🔍 **Pause and try** Use $u_n < u_{n-1}$ and $nu_n u_{n-1} = \frac{1}{2}\pi$ to trap nu_n^2 .

Replacing u_{n-1} by the smaller u_n gives $nu_n^2 < nu_n u_{n-1} = \frac{1}{2}\pi$. Replacing u_n by the larger u_{n-1} gives $\frac{1}{2}\pi = nu_n u_{n-1} < nu_{n-1}^2$. So $nu_n^2 < \frac{1}{2}\pi < nu_{n-1}^2$. ■

Applying the right-hand inequality with $n+1$ in place of n : $\frac{1}{2}\pi < (n+1)u_n^2$, so $nu_n^2 > \frac{n}{n+1} \cdot \frac{1}{2}\pi$. Together,

$$\left(\frac{n}{n+1}\right) \frac{1}{2}\pi < nu_n^2 < \frac{1}{2}\pi. \quad \blacksquare$$

Since $\frac{n}{n+1} \rightarrow 1$, the squeeze gives $nu_n^2 \rightarrow \frac{1}{2}\pi$. ■

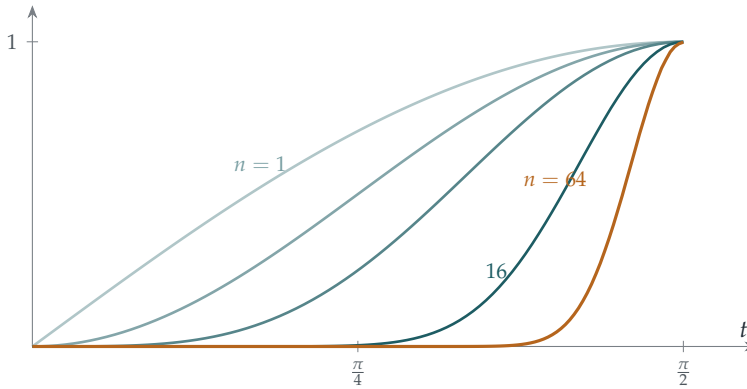


Figure 4.4.1 The graphs of $\sin^n t$ for $n = 1, 2, 4, 16, 64$. As n grows the area under the curve shrinks, concentrating near $t = \frac{\pi}{2}$. The question shows that $u_n \approx \sqrt{\pi/(2n)}$; for $n = 64$ that is about 0.157.

Looking back

- **Reduction formulae.** Integration by parts turned u_n into u_{n-2} , so all the u_n are determined by u_0 and u_1 .
- **Look for an invariant.** The product $nu_n u_{n-1}$ stayed fixed, just as the integral did under the AGM step in Section 2.6.
- **Monotonicity plus an invariant gives a squeeze.** Knowing $u_n < u_{n-1}$ and the exact value of $nu_n u_{n-1}$ trapped nu_n^2 between two quantities with the same limit.

What's really going on

Wallis's product

Running the reduction formula down to u_0 or u_1 gives explicit values:

$$u_{2m} = \frac{1 \cdot 3 \cdot 5 \cdots (2m-1)}{2 \cdot 4 \cdot 6 \cdots 2m} \cdot \frac{\pi}{2}, \quad u_{2m+1} = \frac{2 \cdot 4 \cdot 6 \cdots 2m}{3 \cdot 5 \cdot 7 \cdots (2m+1)}.$$

The squeeze shows $u_{2m}/u_{2m+1} \rightarrow 1$. Writing out that ratio and rearranging gives Wallis's product,

$$\frac{\pi}{2} = \lim_{m \rightarrow \infty} \frac{2 \cdot 2}{1 \cdot 3} \cdot \frac{4 \cdot 4}{3 \cdot 5} \cdots \frac{2m \cdot 2m}{(2m-1)(2m+1)}.$$

John Wallis found it in 1655 by a daring process of interpolation, guessing what $\int_0^1 (1-x^2)^{1/2} dx$ "should" be by fitting it into a table of integrals with whole-number exponents. It was one of the first infinite products in mathematics, after Viète's (Section 2.5), and the first to use only rational numbers.

The factorial of a half

Wallis's interpolation idea was made precise by Euler in 1729. He found a function, now called the *Gamma function*,

$$\Gamma(s) = \int_0^\infty t^{s-1} e^{-t} dt,$$

which satisfies $\Gamma(s+1) = s\Gamma(s)$ (integrate by parts) and $\Gamma(n+1) = n!$ for whole numbers n . It extends the factorial smoothly to all positive real numbers. The closely related *Beta function*

$$B(p, q) = \int_0^1 x^{p-1} (1-x)^{q-1} dx = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$$

includes the integrals of Section 4.3 and Section 2.9, and, after the substitution $x = \sin^2 t$, our u_n . In particular, $\int_0^1 x^{1/2} (1-x)^{1/2} dx = \frac{\pi}{8}$ (exercise 2), which gives $\Gamma(\frac{3}{2})^2/\Gamma(3) = \frac{\pi}{8}$, that is, $(\frac{1}{2})! = \Gamma(\frac{3}{2}) = \frac{\sqrt{\pi}}{2}$. Equivalently, $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, which is the same fact as the famous Gaussian integral $\int_{-\infty}^\infty e^{-x^2} dx = \sqrt{\pi}$, the reason π appears in the formula for the normal distribution.

Stirling's formula

Stirling's formula says $n! \approx \sqrt{2\pi n} (n/e)^n$. The $(n/e)^n$ part is relatively easy to obtain, but the constant $\sqrt{2\pi}$ is harder, and the standard way to find it is exactly this question: the explicit values of u_{2m} involve $(2m)!/(m!)^2$, and the limit $nu_n^2 \rightarrow \frac{\pi}{2}$ pins down the constant. A completely different route to the same constant, through probability, is in Section 12.23.

The thinking skill

Recurrences from integrals

Integration by parts often relates an integral with parameter n to the same integral with a smaller parameter. That recurrence can then be solved, or analysed for its behaviour as $n \rightarrow \infty$.

Watch for quantities that the recurrence leaves unchanged, and for monotonicity in n ; together they often give limits without any explicit formula.

Try this next

- 1 Wallis's product.** Using the explicit values of u_{2m} and u_{2m+1} , and $u_{2m+1} < u_{2m} < u_{2m-1}$, prove Wallis's product.
- 2 A Beta integral.** (From STEP II, 2013.) Let $I_n = \int_0^1 x^n (1-x)^n dx$. Show by integration by parts and symmetry that $I_n = \frac{n}{2(2n+1)} I_{n-1}$, and deduce that $I_n = \frac{(n!)^2}{(2n+1)!}$. Use $x = \sin^2 \theta$ to show that $I_{1/2} = \frac{\pi}{8}$. What does the formula $I_n = \frac{(n!)^2}{(2n+1)!}$ suggest $\left(\frac{1}{2}\right)!$ should be?
- 3 The Beta function.** (From STEP III, 2025.) For $p, q > 0$ let $b(p, q) = \int_0^1 x^{p-1} (1-x)^{q-1} dx$. Show that $b(p, q) = b(q, p)$, that $b(p+1, q) = b(p, q) - b(p, q+1)$, and, with the substitution $x = \frac{t}{1+t}$, that $b(p, q) = \int_0^\infty \frac{t^{p-1}}{(1+t)^{p+q}} dt$. Hence evaluate $\int_0^\infty \frac{t^{3/2}}{(1+t)^6} dt$. (The answer is $\frac{3\pi}{256}$.)
- 4 Central binomial coefficients.** **HARDER** Show that $u_{2m} = \binom{2m}{m} \frac{\pi}{2^{2m+1}}$, and use $nu_n^2 \rightarrow \frac{\pi}{2}$ to deduce that $\binom{2m}{m} \approx \frac{4^m}{\sqrt{\pi m}}$ for large m . Check this for $m = 10$.

Further reading

- **Julian Havil, *Gamma*** (Princeton University Press). Includes the Gamma function and its history.
- **Philip J. Davis, "Leonhard Euler's integral: a historical profile of the Gamma function"**, *American Mathematical Monthly* 66 (1959). A classic, prize-winning article, readable with this section behind you.
- **Emil Artin, *The Gamma Function*** (Dover). A short, beautiful monograph for when you have done some analysis.

4.5 The Sophomore's Dream

The question

Let m be a positive integer and let n be a non-negative integer.

- (i) Use the result $\lim_{t \rightarrow \infty} e^{-mt} t^n = 0$ to show that

$$\lim_{x \rightarrow 0} x^m (\ln x)^n = 0.$$

By writing x^x as $e^{x \ln x}$ show that

$$\lim_{x \rightarrow 0} x^x = 1.$$

- (ii) Let $I_n = \int_0^1 x^m (\ln x)^n dx$. Show that

$$I_{n+1} = -\frac{n+1}{m+1} I_n$$

and hence evaluate I_n .

- (iii) Show that

$$\int_0^1 x^x dx = 1 - \left(\frac{1}{2}\right)^2 + \left(\frac{1}{3}\right)^3 - \left(\frac{1}{4}\right)^4 + \cdots.$$

The hook

Every student learns, sooner or later, that $(a+b)^2 \neq a^2 + b^2$. The false version is jokingly called the “freshman’s dream”. This question proves something that looks just as much like a beginner’s mistake:

$$\int_0^1 x^x dx = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n^n} = 1 - \frac{1}{2^2} + \frac{1}{3^3} - \frac{1}{4^4} + \cdots,$$

and its companion $\int_0^1 x^{-x} dx = \sum_{n=1}^{\infty} \frac{1}{n^n}$. An integral of x^x turns into a sum of n^{-n} : it seems too good to be true, which is why it is called the *sophomore’s dream*. But it is true, and Johann Bernoulli found it in 1697.

The proof uses three ideas you will meet constantly at university: limits that compare exponentials with powers, reduction formulae, and expanding a function as a power series and integrating term by term.

Before you start

Ask yourself The function x^x is not defined at $x = 0$ by any formula you know (0^0 is ambiguous). How does it behave as x approaches 0 from above? Try $x = 0.1, 0.01, 0.001$ on a calculator.

$0.1^{0.1} \approx 0.794$, $0.01^{0.01} \approx 0.955$, $0.001^{0.001} \approx 0.993$. It seems to approach 1. Part (i) proves this, which tells us the integrand is well behaved near 0 and the integral makes sense. The graph of x^x on $(0, 1]$ starts near 1, dips to a minimum and returns to 1 at $x = 1$ (Figure 4.5.1).

Working it through

STAGE 1 Powers beat logarithms

Ask yourself As $x \rightarrow 0^+$, $\ln x \rightarrow -\infty$. What substitution turns $x \rightarrow 0^+$ into $t \rightarrow \infty$, and turns $x^m(\ln x)^n$ into something like $e^{-mt}t^n$?

Put $x = e^{-t}$. As $x \rightarrow 0^+$, $t \rightarrow \infty$, and $\ln x = -t$, so

$$x^m(\ln x)^n = e^{-mt}(-t)^n = (-1)^n e^{-mt} t^n \rightarrow 0.$$

In words: as $x \rightarrow 0$, any positive power of x goes to zero faster than any power of $\ln x$ blows up. ■

With $m = n = 1$, $x \ln x \rightarrow 0$, and since the exponential function is continuous, $x^x = e^{x \ln x} \rightarrow e^0 = 1$. ■

STAGE 2 A reduction formula

Ask yourself Integrate $I_{n+1} = \int_0^1 x^m (\ln x)^{n+1} dx$ by parts. Which factor should you integrate? What happens to the boundary term at $x = 0$?

Integrate x^m and differentiate $(\ln x)^{n+1}$:

$$I_{n+1} = \left[\frac{x^{m+1}}{m+1} (\ln x)^{n+1} \right]_0^1 - \int_0^1 \frac{x^{m+1}}{m+1} \cdot \frac{(n+1)(\ln x)^n}{x} dx.$$

At $x = 1$ the boundary term is 0 because $\ln 1 = 0$; as $x \rightarrow 0$ it tends to 0 by part (i) (with $m+1$ and $n+1$). The remaining integral is $\frac{n+1}{m+1} I_n$, so

$$I_{n+1} = -\frac{n+1}{m+1} I_n. \quad \blacksquare$$

Starting from $I_0 = \int_0^1 x^m dx = \frac{1}{m+1}$ and applying the formula n times,

$$I_n = (-1)^n \frac{n!}{(m+1)^n} \cdot \frac{1}{m+1} = \frac{(-1)^n n!}{(m+1)^{n+1}}. \quad \blacksquare$$

STAGE 3 The series

Ask yourself Expand $x^x = e^{x \ln x}$ using the series for e^y . What does each term look like? Which value of m makes the integrals in part (ii) appear?

Since $e^y = \sum_{n=0}^{\infty} \frac{y^n}{n!}$ for every y ,

$$x^x = e^{x \ln x} = \sum_{n=0}^{\infty} \frac{x^n (\ln x)^n}{n!}.$$

Pause and try Integrate term by term from 0 to 1, using part (ii) with $m = n$ (and treating $n = 0$ separately). What do you get?

The $n = 0$ term is 1, which integrates to 1. For $n \geq 1$, part (ii) with $m = n$ gives $\int_0^1 x^n (\ln x)^n dx = \frac{(-1)^n n!}{(n+1)^{n+1}}$, so the n th term contributes $\frac{(-1)^n}{(n+1)^{n+1}}$. Adding up,

$$\int_0^1 x^x dx = 1 - \frac{1}{2^2} + \frac{1}{3^3} - \frac{1}{4^4} + \cdots \quad \blacksquare$$

(The $n = 0$ term fits the same pattern, $\frac{1}{1^1}$.) The terms shrink extremely fast: seven terms give the value 0.783430... to six decimal places.

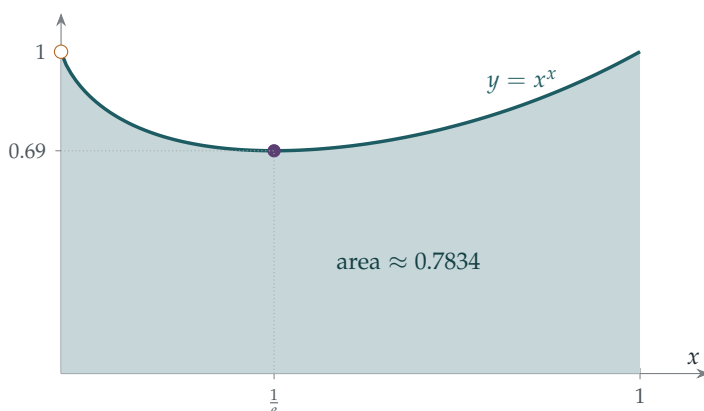


Figure 4.5.1 The graph of x^x on $(0, 1]$. It tends to 1 as $x \rightarrow 0$ (open circle), dips to a minimum of $e^{-1/e} \approx 0.692$ at $x = \frac{1}{e}$, and returns to 1 at $x = 1$. The shaded area is $\sum (-1)^{n+1} n^{-n} \approx 0.7834$.

Looking back

- **Substitute to compare growth rates.** Putting $x = e^{-t}$ turned a limit at 0 into a familiar limit at ∞ .
- **Boundary terms need checking.** The reduction formula depended on the boundary term vanishing at $x = 0$, which is where part (i) was used.

- **Series, then integrate.** Expanding the integrand as a power series and integrating each term turned an integral with no elementary antiderivative into a rapidly convergent sum.

What's really going on

Swapping sums and integrals

Stage 3 quietly swapped an infinite sum with an integral: $\int \Sigma = \Sigma \int$. That is not always allowed. At university you will meet examples where it fails, and theorems that say when it is safe. The simplest says that it is safe when the series converges *uniformly*: the tail of the series can be made small at every x simultaneously. Here, $|x \ln x| \leq \frac{1}{e}$ on $(0, 1]$, so the n th term is at most $\frac{e^{-n}}{n!}$ in size, whatever x is. The sum of these bounds converges, so the tails are uniformly small and the swap is justified. This argument is called the *Weierstrass M-test*, and it is one of the workhorses of analysis.

Why the “dream” works

The identity is special. It depends on the exact match between the powers in $\frac{(x \ln x)^n}{n!}$ and the integral $\int_0^1 x^n (\ln x)^n dx = \frac{(-1)^n n!}{(n+1)^{n+1}}$, where the $n!$ cancels perfectly. Underneath, this integral is a Gamma function in disguise: the substitution $x = e^{-t/(n+1)}$ turns it into $\frac{(-1)^n}{(n+1)^{n+1}} \int_0^\infty t^n e^{-t} dt = \frac{(-1)^n \Gamma(n+1)}{(n+1)^{n+1}}$, connecting with Section 4.4.

The name

The name “sophomore’s dream” was coined by Jonathan Borwein and his co-authors in their 2004 book *Experimentation in Mathematics*, as a companion to the “freshman’s dream” $(a + b)^n = a^n + b^n$ (which, curiously, is true in some number systems: in arithmetic modulo a prime p , $(a + b)^p = a^p + b^p$, because all the middle binomial coefficients are divisible by p). Borwein was a pioneer of *experimental mathematics*, in which computers are used to discover identities like this one, which are then proved by hand.

The thinking skill

Expand and integrate term by term

If an integrand has a power series whose terms you can integrate, integrate each term and add up the results.

It is a powerful way to evaluate integrals that have no elementary antiderivative. Always ask whether swapping the sum and the integral is justified; usually it is, but you should know why.

Try this next

- 1 **The companion.** Show that $\int_0^1 x^{-x} dx = \sum_{n=1}^{\infty} \frac{1}{n^n} \approx 1.2913$.
- 2 **The minimum.** Show that x^x has its minimum on $(0, \infty)$ at $x = \frac{1}{e}$, with value $e^{-1/e}$.
- 3 **How many terms?** For an alternating series with decreasing terms, the error after stopping is less than the first omitted term. How many terms of the series for $\int_0^1 x^x dx$ are needed for six decimal places?
- 4 **A cousin.** Use the same method to show that $\int_0^1 x^2 \ln x dx = -\frac{1}{9}$, and that $\int_0^1 \frac{\ln x}{1-x} dx = -\sum_{n=1}^{\infty} \frac{1}{n^2}$. (By Section 2.4, this is $-\frac{\pi^2}{6}$.)

Further reading

- **Jonathan Borwein, David Bailey & Roland Girgensohn, *Experimentation in Mathematics*** (A.K. Peters). The source of the name, and many more identities discovered by computer.
- **William Dunham, *The Calculus Gallery*** (Princeton University Press). Masterpieces of analysis from Newton to Lebesgue, including Bernoulli's work.
- **Mary Hart, *Guide to Analysis*** (Palgrave). Clear treatment of uniform convergence and when sums and integrals can be swapped.

4.6 Integrals With No Formula

The question

(i) Given that

$$\int \frac{x^3 - 2}{(x + 1)^2} e^x dx = \frac{P(x)}{Q(x)} e^x + \text{constant},$$

where $P(x)$ and $Q(x)$ are polynomials, show that $Q(x)$ has a factor of $x + 1$.

Show also that the degree of $P(x)$ is exactly one more than the degree of $Q(x)$, and find $P(x)$ in the case $Q(x) = x + 1$.

(ii) Show that there are no polynomials $P(x)$ and $Q(x)$ such that

$$\int \frac{1}{x + 1} e^x dx = \frac{P(x)}{Q(x)} e^x + \text{constant}.$$

You need consider only the case when $P(x)$ and $Q(x)$ have no common factors.

The hook

Differentiation is mechanical: there are rules, and any function built from the usual pieces can be differentiated by applying them. Integration is not. Every student has met integrals that resist every technique, and wondered whether they simply weren't clever enough. Sometimes the answer is: no one is clever enough, because *there is no formula*.

This question proves such a thing in a special case. The integral $\int \frac{e^x}{x+1} dx$ certainly exists, as an area. But it cannot be written in the form (rational function) $\times e^x$, which is the form you would expect from integration by parts. Proving that something is impossible, rather than finding something, is a different and very characteristic kind of mathematics.

Before you start



Ask yourself If $\int f(x)e^x dx = \frac{P}{Q}e^x + C$, what equation must P and Q satisfy? (Differentiate the right-hand side.)

Differentiating, $\frac{d}{dx} \left(\frac{P}{Q} e^x \right) = \frac{P'Q - PQ' + PQ}{Q^2} e^x$, and this must equal $f(x)e^x$. So the

whole question is about the polynomial identity

$$P'Q - PQ' + PQ = fQ^2.$$

Questions about integrals have become questions about polynomials: their factors and their degrees.

Working it through

STAGE 1 The factor $x + 1$

With $f(x) = \frac{x^3 - 2}{(x + 1)^2}$, multiply the identity through by $(x + 1)^2$:

$$(x + 1)^2(P'Q - PQ' + PQ) = (x^3 - 2)Q^2. \quad (+)$$

? **Ask yourself** Put $x = -1$ in (+). What must happen?

The left-hand side is 0 at $x = -1$. The right-hand side is $(-1 - 2)Q(-1)^2 = -3Q(-1)^2$. So $Q(-1) = 0$, and by the factor theorem, $x + 1$ is a factor of Q . ■

STAGE 2 Degrees

Let P have degree p and Q have degree q .

? **Ask yourself** In $P'Q - PQ' + PQ$, which term has the highest degree? Can it be cancelled by the others?

PQ has degree $p + q$, while $P'Q$ and PQ' have degree at most $p + q - 1$. So the left-hand side of (+) has degree exactly $2 + p + q$, and the right-hand side has degree $3 + 2q$. Equating, $p = q + 1$. ■

STAGE 3 Finding P

With $Q = x + 1$, we have $Q' = 1$ and P has degree 2. Dividing (+) by $(x + 1)^2$:

$$P'(x + 1) - P + P(x + 1) = x^3 - 2, \quad \text{that is,} \quad (x + 1)P' + xP = x^3 - 2.$$

Try $P = \alpha x^2 + \beta x + \gamma$. The left-hand side is

$$(x + 1)(2\alpha x + \beta) + \alpha x^3 + \beta x^2 + \gamma x = \alpha x^3 + (2\alpha + \beta)x^2 + (2\alpha + \beta + \gamma)x + \beta.$$

Matching with $x^3 - 2$: $\alpha = 1$, $2\alpha + \beta = 0$ so $\beta = -2$, and then $\gamma = 0$; the constant term $\beta = -2$ is consistent. So

$$P(x) = x^2 - 2x, \quad \int \frac{x^3 - 2}{(x + 1)^2} e^x dx = \frac{x^2 - 2x}{x + 1} e^x + C. \quad \blacksquare$$

(Check by differentiating, if you like; it is a good exercise in the quotient rule.)

STAGE 4 An impossibility

Now suppose $\int \frac{e^x}{x+1} dx = \frac{P}{Q} e^x + C$ with P and Q having no common factor. The identity becomes

$$(x+1)(P'Q - PQ' + PQ) = Q^2.$$

As in Stage 1, putting $x = -1$ shows $Q(-1) = 0$. Write $Q = (x+1)^k R$, where $k \geq 1$ and $R(-1) \neq 0$ (take out as many factors of $x+1$ as possible).

 **Pause and try** Substitute $Q = (x+1)^k R$, cancel as many factors of $(x+1)$ as you can, and then put $x = -1$. What do you learn about $P(-1)$?

Then $Q' = k(x+1)^{k-1}R + (x+1)^k R'$, and

$$P'Q - PQ' + PQ = (x+1)^{k-1} \left[(x+1)(P'R - PR' + PR) - kPR \right].$$

So the identity reads $(x+1)^k [(x+1)(P'R - PR' + PR) - kPR] = (x+1)^{2k} R^2$. Dividing by $(x+1)^k$:

$$(x+1)(P'R - PR' + PR) - kPR = (x+1)^k R^2.$$

At $x = -1$, since $k \geq 1$, this gives $-kP(-1)R(-1) = 0$. As $k \neq 0$ and $R(-1) \neq 0$, we must have $P(-1) = 0$. But then $x+1$ divides both P and Q , contradicting the assumption that they have no common factor. So no such P and Q exist. ■

Looking back

- **Translate into algebra.** “Can this be integrated in this form?” became “does this polynomial identity have a solution?”
- **Evaluate at a special point.** Putting $x = -1$ extracted information about factors, three times.
- **Proof by contradiction, using lowest terms.** As with $\sqrt{2}$ in Section 1.3, assuming the fraction is in lowest terms and finding a common factor gave the contradiction.

What's really going on

Poles and logarithms

There is a simple picture behind part (ii). The integrand $\frac{e^x}{x+1}$ has a *simple pole* at $x = -1$: near there it behaves like $\frac{e^{-1}}{x+1}$. Integrating $\frac{1}{x+1}$ gives $\ln|x+1|$, so the antiderivative should behave like $e^{-1} \ln|x+1|$ near -1 , growing to $-\infty$ logarithmically (Figure 4.6.1). But a rational function times e^x can only blow up like a power $(x+1)^{-k}$, and its

derivative then has a pole of order $k + 1 \geq 2$, never a simple pole. Logarithmic behaviour cannot be produced by rational functions; the algebra in Stage 4 is a rigorous version of this observation. In part (i), by contrast, the integrand has a *double* pole, which is exactly what the derivative of something with a simple pole looks like, so a rational answer was possible.

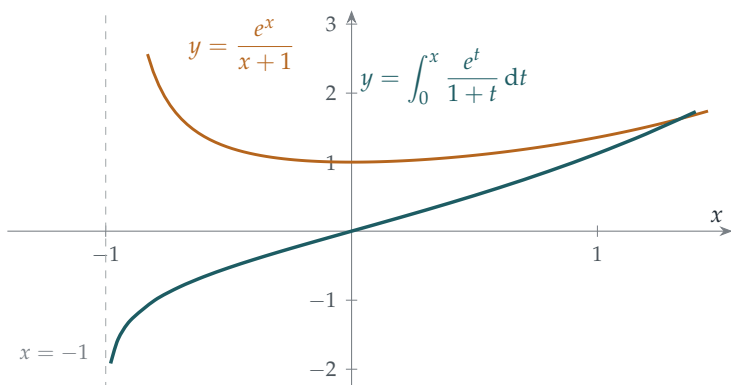


Figure 4.6.1 The integrand $\frac{e^x}{x+1}$ (amber) has a simple pole at $x = -1$. Its antiderivative (teal, computed numerically) exists perfectly well, but it tends to $-\infty$ *logarithmically* as $x \rightarrow -1^+$, like $e^{-1} \ln(x + 1)$. No rational function times e^x can behave like that.

Liouville and the Risch algorithm

In the 1830s Joseph Liouville developed a theory of which integrals can be expressed in terms of *elementary functions* (those built from polynomials, exponentials, logarithms, roots and trigonometric functions by the usual operations). One of his results says that $\int f(x)e^{g(x)} dx$, with f and g rational, is elementary only if it equals $R(x)e^{g(x)}$ for some rational R . So this question, which rules out the rational form, actually proves that $\int \frac{e^x}{x+1} dx$ is not elementary at all. The same theorem shows that $\int e^{-x^2} dx$ and $\int \frac{\sin x}{x} dx$ are not elementary. In 1969 Robert Risch turned Liouville's ideas into an algorithm that decides, for a large class of functions, whether an elementary integral exists and finds it if it does. Versions of it run inside computer algebra systems.

Special functions

When an integral has no elementary form, mathematicians give it a name and study it as a new function, just as we did with the Lambert W function in Section 3.9. The integral here is essentially the *exponential integral* $\text{Ei}(x) = \int_{-\infty}^x \frac{e^t}{t} dt$, which appears in heat flow and in the *logarithmic integral* $\text{li}(x) = \text{Ei}(\ln x)$, the best simple approximation to the number of primes up to x (Section 1.2).

The thinking skill

Proving impossibility

To show something cannot be done, assume it can, translate the assumption into precise algebra, and derive a contradiction.

The hard part is the translation: finding a precise statement of what “can be done” means. Here it meant “there are polynomials P, Q with no common factor satisfying this identity”. Many of the deepest results in mathematics are impossibility proofs of this kind.

Try this next

- 1 Two that work.** Show that $\int \frac{xe^x}{(x+1)^2} dx = \frac{e^x}{x+1} + C$ and $\int \frac{(x^2+1)e^x}{(x+1)^2} dx = \frac{x-1}{x+1} e^x + C$.
- 2 The Gaussian.** Show that there is no polynomial P with $\int e^{-x^2} dx = P(x)e^{-x^2} + C$. (*Hint: differentiate and compare degrees.*)
- 3 Another pole.** Adapt the argument of part (ii) to show that $\int \frac{e^x}{x} dx$ is not of the form $\frac{P(x)}{Q(x)} e^x + C$.
- 4 When is it possible?** Suppose f is a polynomial and $\int \frac{f(x)}{(x+1)^2} e^x dx = \frac{P(x)}{x+1} e^x + C$ for some polynomial P . Show that $(x+1)P' + xP = f$, and by evaluating this and its derivative at $x = -1$, show that $f(-1) + f'(-1) = 0$. Check that this condition holds for $f(x) = x^3 - 2$ and for the examples in exercise 1. **HARDER** Can you show that the condition is also sufficient?

Further reading

- **Brian Conrad, “Impossibility theorems for elementary integration”.** Freely available lecture notes explaining Liouville’s theorem with examples.
- **Elena Anne Marchisotto & Gholam-Ali Zakeri, “An invitation to integration in finite terms”,** *The College Mathematics Journal* 25 (1994). A gentle introduction to Liouville’s theory.
- **Manuel Bronstein, *Symbolic Integration I*** (Springer). How computer algebra systems actually integrate. Advanced.

4.7 The Calculus of Variations

The question

(i) Let

$$I = \int_0^1 ((y')^2 - y^2) dx \quad \text{and} \quad I_1 = \int_0^1 (y' + y \tan x)^2 dx,$$

where y is a given function of x satisfying $y = 0$ at $x = 1$. Show that $I - I_1 = 0$ and deduce that $I \geq 0$. Show further that $I = 0$ only if $y = 0$ for all x ($0 \leq x \leq 1$).

(ii) Let

$$J = \int_0^1 ((y')^2 - a^2 y^2) dx,$$

where a is a given positive constant and y is a given function of x , not identically zero, satisfying $y = 0$ at $x = 1$. By considering an integral of the form

$$\int_0^1 (y' + ay \tan bx)^2 dx,$$

where b is suitably chosen, show that $J \geq 0$. You should state the range of values of a , in the form $a < k$, for which your proof is valid.

In the case $a = k$, find a function y (not everywhere zero) such that $J = 0$.

The hook

Most optimisation problems at school ask for the best *number*: the value of x that maximises an area, say. The *calculus of variations* asks for the best *function*. Which curve between two points does a bead slide down fastest? What shape does a hanging chain take? Which path does light follow through a lens? In each case the quantity to be optimised is an integral, and the unknown is a whole function.

This question studies the simplest interesting example. The integral $J = \int_0^1 ((y')^2 - a^2 y^2) dx$ balances two competing terms: steepness pushes it up, size pushes it down. For small a , steepness always wins and $J \geq 0$ for every function. But there is a critical value of a at which the balance tips, and the question finds it exactly: $a = \frac{\pi}{2}$. Where π comes from will be the story of the section.

Before you start

? **Ask yourself** Why is it not obvious that $I \geq 0$? Try $y = 1 - x$, which satisfies $y(1) = 0$.

I is a difference of two positive quantities, so it could have either sign. For $y = 1 - x$: $\int_0^1 1 \, dx = 1$ and $\int_0^1 (1 - x)^2 \, dx = \frac{1}{3}$, so $I = \frac{2}{3} > 0$. But a cleverer function might make $\int y^2$ larger relative to $\int (y')^2$. The trick in the question is to write I as an integral of a *square*, which is obviously non-negative. This is “completing the square”, for integrals.

Working it through

STAGE 1 Completing the square

Expanding the square in I_1 :

$$I_1 = \int_0^1 ((y')^2 + 2yy' \tan x + y^2 \tan^2 x) \, dx,$$

so

$$I - I_1 = - \int_0^1 (y^2 + y^2 \tan^2 x + 2yy' \tan x) \, dx = - \int_0^1 (y^2 \sec^2 x + 2yy' \tan x) \, dx.$$

? **Ask yourself** The integrand $y^2 \sec^2 x + 2yy' \tan x$ is an exact derivative. Of what?

By the product rule, $\frac{d}{dx}(y^2 \tan x) = 2yy' \tan x + y^2 \sec^2 x$. So

$$I - I_1 = - \left[y^2 \tan x \right]_0^1 = -(y(1)^2 \tan 1 - y(0)^2 \tan 0) = 0,$$

because $y(1) = 0$ and $\tan 0 = 0$. So $I = I_1 \geq 0$, being the integral of a square. ■

STAGE 2 When is $I = 0$?

? **Ask yourself** If the integral of a continuous non-negative function is zero, what can you say about the function? So what equation must y satisfy if $I = 0$?

A continuous non-negative function with zero integral must be identically zero (if it were positive at one point, it would be positive on a small interval around it, contributing a positive amount). So $I = 0$ forces $y' + y \tan x = 0$ on $[0, 1]$. Multiplying by $\sec x$, this says $\frac{d}{dx} \left(\frac{y}{\cos x} \right) = \frac{y' \cos x + y \sin x}{\cos^2 x} = 0$, so $y = C \cos x$. Then $y(1) = C \cos 1 = 0$, and since $\cos 1 \neq 0$, $C = 0$. So $I = 0$ only if $y \equiv 0$. ■

STAGE 3 The general case

? **Ask yourself** Repeat Stage 1 with $\int_0^1 (y' + ay \tan bx)^2 dx$. What must b be for the leftover terms to form an exact derivative?

Call this integral K . Then

$$\begin{aligned} J - K &= - \int_0^1 (a^2 y^2 + a^2 y^2 \tan^2 bx + 2a y y' \tan bx) dx \\ &= - \int_0^1 (a^2 y^2 \sec^2 bx + 2a y y' \tan bx) dx. \end{aligned}$$

On the other hand $\frac{d}{dx}(ay^2 \tan bx) = 2a y y' \tan bx + ab y^2 \sec^2 bx$. These match exactly when $ab = a^2$, that is, $b = a$. Then

$$J - K = - \left[ay^2 \tan ax \right]_0^1 = -a y(1)^2 \tan a = 0,$$

so $J = K \geq 0$.

⌚ **Pause and try** Where can this argument break down? (Look at $\tan ax$ on the interval $0 \leq x \leq 1$.)

The argument needs $\tan ax$ to be defined and continuous for all x in $[0, 1]$; otherwise K might not exist and the fundamental theorem of calculus cannot be applied. As x runs over $[0, 1]$, ax runs over $[0, a]$, and $\tan$ is continuous there exactly when $a < \frac{\pi}{2}$. So the proof is valid for

$$a < k = \frac{\pi}{2}. \quad \blacksquare$$

STAGE 4 The critical case

? **Ask yourself** When $a = \frac{\pi}{2}$, the proof fails. In Stage 2, the function making $I = 0$ would have satisfied $y' + y \tan x = 0$. What is the analogous equation now, and what are its solutions?

The analogue is $y' + ay \tan ax = 0$, whose solutions are $y = C \cos ax$. With $a = \frac{\pi}{2}$, $y = \cos \frac{\pi x}{2}$ satisfies $y(1) = \cos \frac{\pi}{2} = 0$. Check directly:

$$\begin{aligned} J &= \frac{\pi^2}{4} \int_0^1 \left(\sin^2 \frac{\pi x}{2} - \cos^2 \frac{\pi x}{2} \right) dx = -\frac{\pi^2}{4} \int_0^1 \cos \pi x dx \\ &= -\frac{\pi^2}{4} \left[\frac{\sin \pi x}{\pi} \right]_0^1 = 0. \quad \blacksquare \end{aligned}$$

So at the critical value, a non-zero function makes J exactly zero.

Looking back

- **Completing the square in an integral.** Adding and subtracting an exact derivative turned J into the integral of a square plus a boundary term that vanished.

- **Check where the tools fail.** The method worked only while $\tan ax$ stayed finite on the interval. The point where it failed, $a = \frac{\pi}{2}$, was exactly the critical value.
- **The extreme case solves a differential equation.** The function giving $J = 0$ came from the first-order equation $y' + ay \tan ax = 0$, and it also satisfies $y'' + a^2y = 0$.

What's really going on

The brachistochrone and the birth of a subject

In 1696 Johann Bernoulli challenged the mathematicians of Europe: find the curve down which a bead slides from one point to another in the least time. Solutions came from Newton (anonymously, overnight; Bernoulli said he recognised “the lion by his claw”), Leibniz, l'Hôpital and Jakob Bernoulli. The answer is a cycloid. In the 1750s Euler and the young Joseph-Louis Lagrange turned these special tricks into a general method: if y makes $\int F(x, y, y') dx$ stationary, then y satisfies the *Euler–Lagrange equation*

$$\frac{d}{dx} \frac{\partial F}{\partial y'} = \frac{\partial F}{\partial y}.$$

For our J , $F = (y')^2 - a^2y^2$, and the Euler–Lagrange equation is $y'' + a^2y = 0$: simple harmonic motion.

Why $\pi/2$?

For $a < \frac{\pi}{2}$, J has a minimum value of 0, achieved only by $y \equiv 0$. At $a = \frac{\pi}{2}$ the minimum is still 0 but a second function achieves it. For $a > \frac{\pi}{2}$, the same function $y = \cos \frac{\pi x}{2}$ gives

$$J = \frac{\pi^2}{8} - \frac{a^2}{2} < 0,$$

and by scaling y up, J can be made as negative as you like: there is no minimum at all (Figure 4.7.1). The critical value is the first a for which $y'' + a^2y = 0$ has a non-zero solution satisfying the boundary conditions ($y(1) = 0$, and a “natural” condition $y'(0) = 0$ at the free end). That is an *eigenvalue problem*, and $a^2 = \frac{\pi^2}{4}$ is its smallest eigenvalue. Compare Section 3.7, where pinning the function at both ends of an interval of length π gave critical value 1; here the interval has length 1 and only one end is pinned, which is like half of an interval of length 2, so the critical a is $\frac{\pi}{2}$.

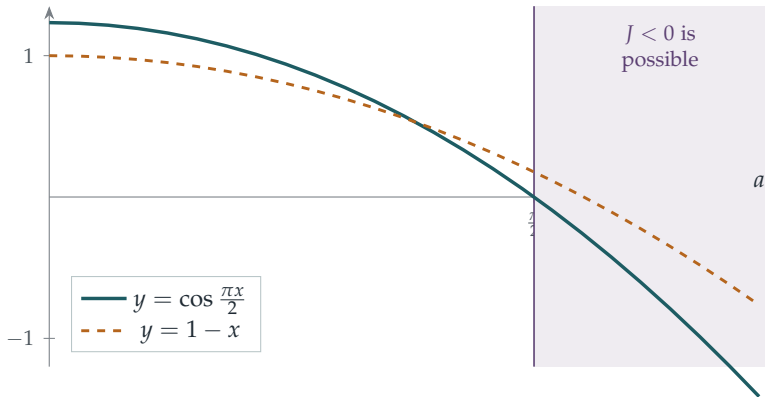


Figure 4.7.1 The value of J against a for two test functions with $y(1) = 0$. For $y = 1 - x$, $J = 1 - \frac{a^2}{3}$ becomes negative only when $a > \sqrt{3}$. For $y = \cos \frac{\pi x}{2}$, $J = \frac{\pi^2}{8} - \frac{a^2}{2}$ becomes negative as soon as $a > \frac{\pi}{2}$, and no function does better: for $a < \frac{\pi}{2}$, $J \geq 0$ for every y .

Physics and the principle of least action

The calculus of variations turned out to describe nature itself. In the eighteenth and nineteenth centuries it emerged that the laws of motion can be stated as: a mechanical system moves so as to make an integral called the *action* stationary. Light, similarly, travels along the path of least time (Fermat's principle), which explains refraction. In the twentieth century the same idea became the foundation of quantum mechanics, through Richard Feynman's "sum over paths". The quadratic integrals in this question are the simplest examples, and the eigenvalue problem they lead to reappears as the energy levels of a quantum oscillator in Section 8.6.

The thinking skill

Complete the square, then mind the boundary

To show a quadratic expression is non-negative, write it as a square plus something you can control. For integrals, the "something" is often an exact derivative, which integrates to boundary terms.

And always ask where the method stops working. The point of failure here was not an accident; it was the most important number in the problem.

Try this next

- Beyond the critical value.** Show that if $a > \frac{\pi}{2}$ then $y = \cos \frac{\pi x}{2}$ gives $J < 0$.

- 2 The Euler–Lagrange equation.** **HARDER** Suppose y minimises J among functions with $y(1) = 0$. For any function η with $\eta(1) = 0$ and any number ε , $J(y + \varepsilon\eta) \geq J(y)$. Expand $J(y + \varepsilon\eta)$ in powers of ε and deduce that $\int_0^1 (y'\eta' - a^2 y\eta) dx = 0$. Integrate by parts to show that $y'' + a^2 y = 0$ and $y'(0) = 0$ (assuming y is smooth enough).
- 3 A longer interval.** Show by a substitution that if $y(L) = 0$, then $\int_0^L ((y')^2 - a^2 y^2) dx \geq 0$ for all y exactly when $a < \frac{\pi}{2L}$.
- 4 The cycloid.** Look up the brachistochrone problem and find the parametric equations of the cycloid. Show that a bead released from rest anywhere on an upside-down cycloid reaches the bottom in the same time (the *tautochrone* property, found by Huygens in 1659).

Further reading

- **Paul Nahin, *When Least Is Best*** (Princeton University Press). The history and mathematics of optimisation, including the brachistochrone, written for students.
- **Richard Feynman, *The Feynman Lectures on Physics***, Volume II, Chapter 19 (“The principle of least action”). Free online.
- **Bruce van Brunt, *The Calculus of Variations*** (Springer). An undergraduate textbook; the first chapters need only A-level calculus.

4.8 Slicing a Spherical Loaf

The question

A spherical loaf of bread is cut into parallel slices of equal thickness. Show that, after any number of the slices have been eaten, the area of crust remaining is proportional to the number of slices remaining.

A European ruling decrees that a parallel-sliced spherical loaf can only be referred to as 'crusty' if the ratio of volume V (in cubic metres) of bread remaining to area A (in square metres) of crust remaining after any number of slices have been eaten satisfies $V/A < 1$. Show that the radius of a crusty parallel-sliced spherical loaf must be less than $2\frac{2}{3}$ metres.

[The area A and volume V formed by rotating a curve in the x - y plane round the x -axis from $x = -a$ to $x = -a + t$ are given by

$$A = 2\pi \int_{-a}^{-a+t} y \left(1 + \left(\frac{dy}{dx} \right)^2 \right)^{\frac{1}{2}} dx, \quad V = \pi \int_{-a}^{-a+t} y^2 dx.]$$

The hook

Slice a spherical loaf into slices of equal thickness. The slices near the middle are fat discs with a thin band of crust round the edge. The slices near the ends are small caps that seem to be almost all crust. Surely the end slices have more crust?

They don't. Every slice has exactly the same area of crust, however near the middle or the end it is cut. The thin central band is wide in circumference but nearly vertical; the end cap is small in circumference but slopes steeply, so its crust is stretched over a larger width. The two effects cancel exactly. This is a theorem of Archimedes, and it was so dear to him that he asked for a sphere inside a cylinder to be carved on his tomb. The second part of the question is a lighter-hearted piece of regulation, but it rests on the same calculation.

Before you start



Ask yourself The sphere of radius a is obtained by rotating which curve about the x -axis? What is $\frac{dy}{dx}$ for that curve?

Rotating the semicircle $y = \sqrt{a^2 - x^2}$, $-a \leq x \leq a$, gives the sphere. Then $\frac{dy}{dx} = \frac{-x}{\sqrt{a^2 - x^2}} = -\frac{x}{y}$. Keep that in mind: the slope is steep near the ends ($x \rightarrow \pm a$) and flat in the middle, which is exactly the effect described in the hook.

Working it through

STAGE 1 The crust formula simplifies



Ask yourself Compute $y\sqrt{1 + (y')^2}$ for the semicircle. Something remarkable happens.

With $y' = -x/y$,

$$1 + (y')^2 = 1 + \frac{x^2}{y^2} = \frac{y^2 + x^2}{y^2} = \frac{a^2}{y^2},$$

so $y\sqrt{1 + (y')^2} = y \cdot \frac{a}{y} = a$, a constant! The crust area between any two planes $x = x_1$ and $x = x_2$ is therefore

$$2\pi \int_{x_1}^{x_2} a \, dx = 2\pi a(x_2 - x_1).$$

It depends only on the distance between the planes, not on where they are.

STAGE 2 Crust is proportional to slices

If the loaf is cut into slices of thickness h , each slice has crust area $2\pi ah$, the same for all of them. Eat some slices, from one end, from both ends, or from anywhere, and the crust remaining is $2\pi ah$ times the number of slices remaining. ■

STAGE 3 The crusty condition

Suppose the slices are eaten from the end $x = a$, so what remains runs from $x = -a$ to $x = -a + t$ for some t between 0 and $2a$.



Ask yourself Compute V using the given formula. A substitution $s = x + a$ might make the algebra cleaner.

With $s = x + a$, $y^2 = a^2 - x^2 = a^2 - (s - a)^2 = 2as - s^2$, so

$$V = \pi \int_0^t (2as - s^2) \, ds = \pi \left(at^2 - \frac{t^3}{3} \right), \quad A = 2\pi at,$$

and

$$\frac{V}{A} = \frac{at^2 - t^3/3}{2at} = \frac{t}{2} - \frac{t^2}{6a}.$$



Pause and try The ruling requires $V/A < 1$ after any number of slices have been eaten, that is, for every t with $0 < t \leq 2a$. What is the largest value of $\frac{t}{2} - \frac{t^2}{6a}$ on this range?

Differentiating with respect to t : $\frac{1}{2} - \frac{t}{3a} = 0$ at $t = \frac{3a}{2}$, which is inside the range $(0, 2a]$, and this is a maximum (the expression is a downward parabola in t). The maximum value is

$$\frac{3a}{4} - \frac{9a^2/4}{6a} = \frac{3a}{4} - \frac{3a}{8} = \frac{3a}{8}.$$

So $V/A < 1$ for every t exactly when $\frac{3a}{8} < 1$, that is, $a < \frac{8}{3} = 2\frac{2}{3}$ metres. ■

Looking back

- **Simplify before integrating.** The integrand for the crust looked complicated but turned out to be constant. Always simplify an integrand before attacking it.
- **Read “for any” carefully.** The condition had to hold after *any* number of slices, so we needed the maximum of V/A , not its value for the whole loaf (which is $\frac{a}{3}$).
- **A surprising answer from a simple calculation.** The equal-crust property sounds implausible, but the calculation is two lines long.

What’s really going on

Archimedes’ hat-box theorem

Archimedes proved, in *On the Sphere and Cylinder*, that the area of the band of a sphere between two parallel planes equals the area of the band they cut from the cylinder that wraps around the sphere (Figure 4.8.1). Our calculation is this theorem: the cylinder of radius a has crust area $2\pi a$ per unit length, and so does the sphere. Taking the whole sphere, its surface area equals that of the cylinder’s curved surface, $2\pi a \times 2a = 4\pi a^2$. Archimedes regarded this, together with the fact that the sphere has two-thirds the volume of the cylinder, as his finest work. Cicero, visiting Syracuse in 75 BC, found the overgrown tomb by looking for the sphere and cylinder carved on it.

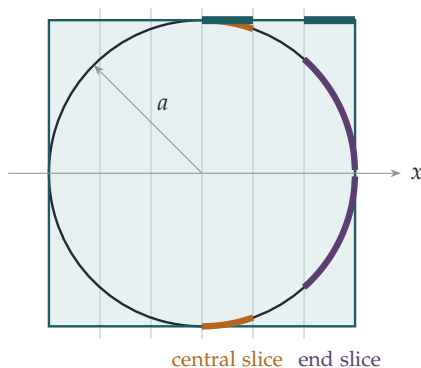


Figure 4.8.1 Cross-section of a spherical loaf of radius a inside its wrapping cylinder, cut into slices of equal thickness. The central slice (amber) has a short, nearly horizontal arc of crust far from the axis; the end slice (plum) has a long, steep arc close to the axis. When rotated about the x -axis, both give exactly the same crust area as the corresponding band of the cylinder (teal): $2\pi a$ times the thickness.

Maps of the Earth

The hat-box theorem gives a way to draw a map of the Earth that preserves area. Project each point of the globe horizontally outwards onto the wrapping cylinder, then unroll the cylinder. Bands of the sphere map to bands of the cylinder of the same area, and the same is true for any region. This is *Lambert's cylindrical equal-area projection* (1772). It distorts shapes badly near the poles, but areas are honest, unlike the familiar Mercator projection, which makes Greenland look as big as Africa (it is about fourteen times smaller). A famous theorem of Gauss, the *Theorema Egregium* (1827), shows that no map of a sphere can preserve both areas and shapes: every flat map of the Earth must distort something.

Random points on a sphere

Here is a practical consequence. To choose a point uniformly at random on a sphere, you might try choosing a latitude and longitude uniformly, but that crowds points near the poles. The hat-box theorem gives the right recipe: choose the height x uniformly between $-a$ and a , and the longitude uniformly between 0 and 2π . Because area is proportional to height, this is exactly uniform. Computer simulations in physics and graphics use this method. (Remarkably, it works only in three dimensions: the corresponding statement for spheres in four or more dimensions is false.)

The thinking skill

Look for hidden simplicity

When a calculation produces something unexpectedly simple, such as a constant, stop and ask why. There is usually a geometric reason, and finding it is where the real understanding lies.

Here the integrand $y\sqrt{1+(y')^2}$ was constant because steepness and distance from the axis exactly compensate each other on a sphere: that is Archimedes' theorem.

Try this next

- 1 **The whole loaf.** Show that for the whole uncut loaf, $V/A = \frac{a}{3}$. Why is the condition in the question harder to satisfy than $\frac{a}{3} < 1$?
- 2 **The tropics.** What fraction of the Earth's surface lies between latitudes 23.4° south and 23.4° north? (Treat the Earth as a sphere. Since area is proportional to height, you only need $\sin 23.4^\circ$.)
- 3 **Other shapes.** Show that for a loaf in the shape of a paraboloid (rotate $y = \sqrt{x}$ about the x -axis), the crust per slice is *not* constant. For which curves $y(x)$ is $y\sqrt{1+(y')^2}$ constant? (Solve the differential equation.)
- 4 **Volumes.** Use the volume formula to show that a sphere has two-thirds the volume of its wrapping cylinder, as Archimedes found.

Further reading

- **Sherman Stein, *Archimedes: What Did He Do Besides Cry Eureka?*** (Mathematical Association of America). Archimedes' mathematics explained for a modern reader.
- **Timothy Gowers, *Mathematics: A Very Short Introduction*** (Oxford University Press). The chapter on geometry introduces the geometry of curved surfaces.
- **John Snyder, *Flattening the Earth*** (University of Chicago Press). Two thousand years of map projections.

4.9 Holditch's Theorem

The question

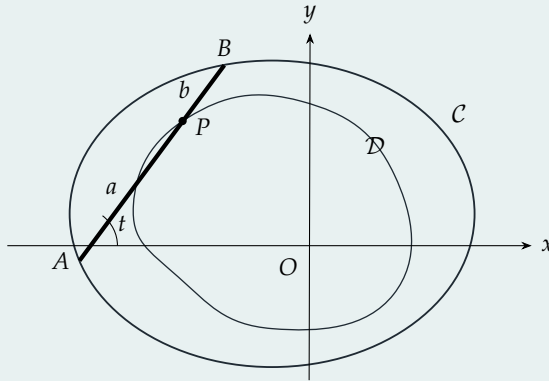
A movable point P has cartesian coordinates (x, y) , where x and y are functions of t . The polar coordinates of P with respect to the origin O are r and θ . Starting with the expression

$$\frac{1}{2} \int r^2 d\theta$$

for the area swept out by OP , obtain the equivalent expression

$$\frac{1}{2} \int \left(x \frac{dy}{dt} - y \frac{dx}{dt} \right) dt. \quad (*)$$

The ends of a thin straight rod AB lie on a closed convex curve $\mathcal{C}$. The point P on the rod is a fixed distance a from A and a fixed distance b from B . The angle between AB and the positive x direction is t . As A and B move anticlockwise round $\mathcal{C}$, the angle t increases from 0 to 2π and P traces a closed convex curve $\mathcal{D}$ inside $\mathcal{C}$, with the origin O lying inside $\mathcal{D}$, as shown in the diagram.



Let (x, y) be the coordinates of P . Write down the coordinates of A and B in terms of a, b, x, y and t .

The areas swept out by OA , OB and OP are denoted by $[A]$, $[B]$ and $[P]$, respectively. Show, using $(*)$, that

$$[A] = [P] + \pi a^2 - af$$

where

$$f = \frac{1}{2} \int_0^{2\pi} \left(\left(x + \frac{dy}{dt} \right) \cos t + \left(y - \frac{dx}{dt} \right) \sin t \right) dt.$$

Obtain a corresponding expression for $[B]$ involving b . Hence show that the area between the curves $\mathcal{C}$ and $\mathcal{D}$ is πab .

The hook

Take any oval, any shape at all as long as it is convex, and slide a rod of fixed length round inside it, keeping both ends on the curve. Paint a dot on the rod, dividing it into lengths a and b . As the rod goes once round, the dot traces out a smaller oval. What is the area of the ring between the two curves?

The astonishing answer is πab , *whatever the shape of the original curve*. A long thin oval, an egg, a rounded triangle: the area of the ring is always the same, and it depends only on where the dot is on the rod. The result was found in 1858 by the Reverend Hamnet Holditch, a fellow of Gonville and Caius College, Cambridge, and it is one of the most surprising theorems in elementary geometry. This question proves it with a single formula for area.

Before you start



Ask yourself Check the theorem in the simplest case. If $\mathcal{C}$ is a circle of radius R , what curve does P trace, and what is the area of the ring?

By symmetry, P traces a circle of some radius r with the same centre. The rod AB is a chord of the big circle through P , and the *intersecting chords theorem* (or “power of a point”) says that for any chord through P , $PA \times PB = R^2 - r^2$. So $ab = R^2 - r^2$, and the ring has area $\pi R^2 - \pi r^2 = \pi ab$. The theorem says this remains true for every convex curve, not just circles.

Working it through

STAGE 1 Area from coordinates



Ask yourself With $x = r \cos \theta$ and $y = r \sin \theta$, all functions of t , compute $x\dot{y} - y\dot{x}$, where dots denote $\frac{d}{dt}$.

$\dot{x} = \dot{r} \cos \theta - r \sin \theta \dot{\theta}$ and $\dot{y} = \dot{r} \sin \theta + r \cos \theta \dot{\theta}$. Then

$$x\dot{y} - y\dot{x} = r \cos \theta (\dot{r} \sin \theta + r \cos \theta \dot{\theta}) - r \sin \theta (\dot{r} \cos \theta - r \sin \theta \dot{\theta}) = r^2 \dot{\theta}.$$

$$\text{So } \frac{1}{2} \int r^2 d\theta = \frac{1}{2} \int r^2 \dot{\theta} dt = \frac{1}{2} \int (x\dot{y} - y\dot{x}) dt. \quad \blacksquare$$

This formula needs no polar coordinates, only the cartesian position as a function of a parameter. When a point goes once anticlockwise round a closed curve enclosing O , it

gives the area inside the curve.

STAGE 2 The ends of the rod

The rod points in the direction $(\cos t, \sin t)$ from A to B , with P a distance a from A and b from B . So

$$A = (x - a \cos t, y - a \sin t), \quad B = (x + b \cos t, y + b \sin t).$$

STAGE 3 The area swept by OA

Ask yourself Apply $(*)$ to A . Expand $x_A \dot{y}_A - y_A \dot{x}_A$ and sort the terms by powers of a .

With $x_A = x - a \cos t$, $y_A = y - a \sin t$, we have $\dot{x}_A = \dot{x} + a \sin t$ and $\dot{y}_A = \dot{y} - a \cos t$. Expanding,

$$\begin{aligned} x_A \dot{y}_A - y_A \dot{x}_A &= (x - a \cos t)(\dot{y} - a \cos t) - (y - a \sin t)(\dot{x} + a \sin t) \\ &= (x\dot{y} - y\dot{x}) + a^2(\cos^2 t + \sin^2 t) - a[(x + \dot{y}) \cos t + (y - \dot{x}) \sin t]. \end{aligned}$$

Integrating from 0 to 2π and halving: the first term gives $[P]$, the second gives $\frac{1}{2}a^2 \cdot 2\pi = \pi a^2$, and the third gives $-af$. So

$$[A] = [P] + \pi a^2 - af. \quad \blacksquare$$

Pause and try Now find $[B]$. (Hint: B is obtained from the formula for A by replacing a with what?)

$B = (x + b \cos t, y + b \sin t)$ is the formula for A with a replaced by $-b$. So

$$[B] = [P] + \pi b^2 + bf.$$

STAGE 4 Eliminating f

Ask yourself Both A and B go once round the same curve $\mathcal{C}$. What does that tell you about $[A]$ and $[B]$?

Each of A and B travels once anticlockwise round $\mathcal{C}$ as t goes from 0 to 2π , so $[A] = [B] = \text{area}(\mathcal{C})$. Equating the two expressions,

$$\pi a^2 - af = \pi b^2 + bf \implies f(a + b) = \pi(a^2 - b^2) \implies f = \pi(a - b).$$

Since P goes once round $\mathcal{D}$, $[P] = \text{area}(\mathcal{D})$, and

$$\text{area}(\mathcal{C}) - \text{area}(\mathcal{D}) = [A] - [P] = \pi a^2 - a\pi(a - b) = \pi ab. \quad \blacksquare$$

The mysterious quantity f , which depends on the shape of the curve, was never computed: it was eliminated.

Looking back

- **A formula suited to the problem.** The area formula (*) needed only coordinates as functions of a parameter, which is exactly how the rod's motion is described.
- **Two expressions, one unknown.** We could not compute f , but we had two equations involving it, so we eliminated it, like solving simultaneous equations.
- **Test the special case first.** The circle gave the answer πab before we started, and told us what to aim for.

What's really going on

Green's theorem and the planimeter

The formula (*) is a special case of *Green's theorem*, which relates an integral round a closed curve to an integral over the region inside it. In its area form, it says that the area enclosed by a curve is $\frac{1}{2} \oint (x dy - y dx)$, an integral that can be computed by travelling round the boundary alone. Engineers exploited this with the *planimeter*, a mechanical instrument invented in the nineteenth century (Jakob Amsler's 1854 design was the most successful): you trace the boundary of a shape on a map or diagram with a pointer, and a wheel records the enclosed area. Planimeters were used to measure land areas, to compute the work done in steam engines from indicator diagrams, and in many other applications, until digital methods replaced them. Holditch's theorem is closely related: the rod is like the arm of a planimeter.

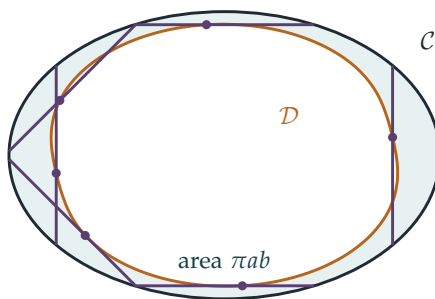


Figure 4.9.1 Holditch's theorem for an ellipse with semi-axes 3 and 2, and a rod of length 2.5 with $a = 1$, $b = 1.5$. Some positions of the rod are shown (plum), with the point P marked. The curve D traced by P (amber) was computed numerically, and the shaded ring between the curves has area $4.7124 \dots = 1.5\pi$, exactly as the theorem predicts.

The fine print

Holditch stated his theorem without much care for conditions, and it needs some. The rod must be able to go all the way round with both ends on the curve, and the traced

curve $\mathcal{D}$ must be a simple closed curve traced once; if the rod is too long compared with the width of $\mathcal{C}$, $\mathcal{D}$ can cross itself, and the “area” given by the formula counts some regions twice or with a negative sign. The question avoids these difficulties by assuming $\mathcal{D}$ is a convex curve containing O . Finding exactly when the theorem holds is a good example of how a beautiful idea must be made precise, a theme of Chapter 3.

The thinking skill

Eliminate what you cannot compute

If a quantity is too complicated to calculate, look for two different equations that contain it, and eliminate it.

Here f depended on every detail of the curve, but it appeared in the expressions for both $[A]$ and $[B]$, and the fact that $[A] = [B]$ removed it. The answer, πab , could only be independent of the curve because the curve-dependent part cancelled.

Try this next

- 1 **The area of an ellipse.** Use $(*)$ with $x = p \cos t$, $y = q \sin t$ to show that an ellipse with semi-axes p and q has area πpq .
- 2 **The circle case.** Prove the intersecting chords theorem: if two chords of a circle meet at P , the products of the lengths of their two segments are equal. Deduce Holditch's theorem for a circle.
- 3 **The midpoint.** Show that $f = \pi(a - b)$ means that when P is the midpoint of the rod, $f = 0$. What is the ring's area then, in terms of the rod's length L ?
- 4 **A triangle.** HARDER Use $(*)$ to find the area of the triangle with vertices $(0, 0)$, (x_1, y_1) , (x_2, y_2) , traversing its boundary, and show it equals $\frac{1}{2}|x_1 y_2 - x_2 y_1|$. This is the start of the “shoelace formula” for the area of any polygon.

Further reading

- **Mark Levi**, *The Mathematical Mechanic* (Princeton University Press). Uses physical reasoning, with rods, wheels and springs, to prove mathematical results, in the spirit of this section.
- **H. M. Schey**, *Div, Grad, Curl, and All That* (Norton). An informal introduction to vector calculus, including Green's theorem.
- **Hamnet Holditch**, “Geometrical theorem”, *Quarterly Journal of Pure and Applied Mathematics* 2 (1858). The original note, a single short paragraph with none of the fine

print.

What We Have Learned

We began by asking what an integral really is, what integrals can prove, and what to do when an integral has no formula. Here is how the chapter's nine questions answered them.

1. What is an integral, and which functions have one?

In Section 4.1 we defined the integral as the common limit of lower and upper sums of thin rectangles, and computed $\int_0^a x^5 dx$ from that definition alone. It took some work: telescoping sums of consecutive products, bounds on both sides, and a squeeze. That effort is the measure of how much the fundamental theorem of calculus gives you for free.

The definition also told us that not every function is integrable. Dirichlet's function, 1 on the rationals and 0 on the irrationals, has upper sums all equal to 1 and lower sums all equal to 0. Problems like this one led Lebesgue to a more powerful theory of integration, which you will meet later at university.

2. What can integrals prove?

Far more than areas.

- ▶ **Inequalities between numbers.** An integral of a positive function equals $\frac{22}{7} - \pi$, so $\frac{22}{7} > \pi$ (Section 4.3). The idea, “write $A - B$ as the integral of something positive”, is as useful as it is simple.
- ▶ **Infinite products and special values.** Reduction formulae for $\int \sin^n t dt$, an invariant and a squeeze gave Wallis's product, the constant in Stirling's formula and $\left(\frac{1}{2}\right)! = \frac{\sqrt{\pi}}{2}$ (Section 4.4).
- ▶ **Infinite sums.** Expanding x^x as a power series and integrating term by term gave the sophomore's dream (Section 4.5).
- ▶ **The foundations of trigonometry.** Defining $T(x) = \int_0^x \frac{du}{1+u^2}$ and $\pi = 2T(\infty)$, substitutions alone produced the addition formula and special values like $T(\sqrt{3}) = \frac{\pi}{3}$ (Section 4.2).
- ▶ **Geometry.** An integrand that simplified to a constant revealed Archimedes' hat-box theorem (Section 4.8), and an area formula written as an integral round the boundary proved Holditch's astonishing theorem (Section 4.9).
- ▶ **Optimal functions.** Completing the square inside an integral showed exactly when $\int ((y')^2 - a^2 y^2)$ must be non-negative, and the critical value $\frac{\pi}{2}$ turned out to be an

eigenvalue (Section 4.7).

3. What can we do when there is no formula?

In Section 4.6 we proved that $\int \frac{e^x}{x+1} dx$ cannot be written as a rational function times e^x , and Liouville’s theory shows it has no elementary formula at all. The reason, seen through the lens of poles, is that the antiderivative must behave like a logarithm near $x = -1$, which rational functions cannot do. Yet the integral exists perfectly well, and we can compute it: numerically, by series (as for $\int x^x$ in Section 4.5), or by giving it a name and studying it as a new function, as we did with the W function in Section 3.9.

4. A toolkit for integration

Method	Where you used it
Working from the definition	$\int_0^a x^5$ from rectangles (4.1)
Telescoping sums	Sums of consecutive products (4.1)
Substitutions that preserve an integral	Trigonometry from $\int \frac{du}{1+u^2}$ (4.2)
Polynomial division	$\frac{22}{7} - \pi$ (4.3)
Positivity and bounding the integrand	$\pi < \frac{22}{7}$ (4.3)
Reduction formulae	$\int \sin^n t$ (4.4); $\int x^m (\ln x)^n$ (4.5)
Invariants and squeezes	$nu_n u_{n-1} = \frac{\pi}{2}$ (4.4)
Term-by-term integration	The sophomore’s dream (4.5)
Factor and degree arguments	No rational antiderivative (4.6)
Completing the square	The calculus of variations (4.7)
Simplifying before integrating	The spherical loaf (4.8)
Area as a boundary integral	Holditch’s theorem (4.9)

Where next

- The Gamma and Beta functions of Section 4.4, and the constant $\sqrt{2\pi}$ in Stirling’s formula, reappear in probability in Section 12.23.
- The eigenvalue problem of Section 4.7 becomes the quantum harmonic oscillator in Section 8.6.
- Machin’s formula, built from the addition identity of Section 4.2, is the subject of Section 5.1.
- Chapter 5 turns to complex numbers, which explain why so many of this chapter’s integrals involve π .

If you want to go further before university

- Paul Nahin, *Inside Interesting Integrals* (Springer): a feast of clever integrals

and the tricks behind them.

- ▶ **William Dunham, *The Calculus Gallery***: masterpieces of analysis from Newton to Lebesgue.
- ▶ **Michael Spivak, *Calculus***: the definition of the integral and the fundamental theorem, done properly.
- ▶ **Sherman Stein, *Archimedes***: for the ancient roots of integration.

Hermite recoiled from functions without derivatives, but integration has never had that problem: it is generous, accepting functions that differentiation cannot handle, and it keeps producing surprises. If a single theme runs through this chapter, it is that integrals are not just a technique for finding areas. They are a language for expressing and proving facts about numbers, functions and shapes, and you will be speaking it for the rest of your mathematical life.