4

Integration

When you can measure what you are speaking, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind: it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of science.—William Thomson, Lord Kelvin

4.0 Introduction

FIGURE 4.0.1.
Lord Kelvin (William Thomson, 1824–1907)

An actuary deciding what premium to charge for a life insurance policy needs integrals. So does a banker deciding what to charge for stock options. Black and Scholes received a Nobel Prize for this work, which involves a very fancy stochastic integral.

Chapters 1 and 2 began with algebra, then moved on to calculus. Here, as in chapter 3, we dive right into calculus. We introduce the relevant linear algebra (determinants) later in the chapter, where we need it.

When students first meet integrals, integrals come in two very different flavors – Riemann sums (the idea) and antiderivatives (the recipe) – rather as derivatives arise as limits, and as something to be computed using Leibnitz's rule, the chain rule, etc.

Since integrals can be systematically computed (by hand) only as antiderivatives, students often take this to be the definition. This is misleading: the definition of an integral is given by a Riemann sum (or by "area under the graph"; Riemann sums are just a way of making the notion of "area" precise). Section 4.1 is devoted to generalizing Riemann sums to functions of several variables. Rather than slice up the domain of a function $f : \mathbb{R} \to \mathbb{R}$ into little intervals and computing the "area under the graph" corresponding to each interval, we will slice up the "n-dimensional domain" of a function $f : \mathbb{R}^n \to \mathbb{R}$ into little n-dimensional cubes.

Computing n-dimensional volume is an important application of multiple integrals. Another is probability theory; in fact, probability has become such an important part of integration that integration has almost become a part of probability. Even such a mundane problem as quantifying how heavy a child is for his or her height requires multiple integrals. Fancier yet are the uses of probability that arise when physicists study turbulent flows, or engineers try to improve the internal combustion engine. They cannot hope to deal with one molecule at a time; any picture they get of reality at a macroscopic level is necessarily based on a probabilistic picture of what is going on at a microscopic level. We give a brief introduction to this important field in section 4.2.

Section 4.3 discusses what functions are integrable; in section 4.4, we use the notion of *measure* to give a sharper criterion for integrability (a criterion that applies to more functions than the criteria of section 4.3).

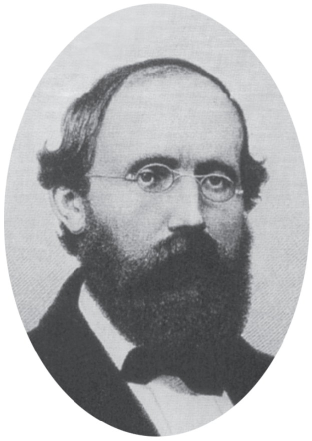

FIGURE 4.0.2.

Born in 1826, Bernhard Riemann died of tuberculosis in 1866. The Riemann integral, which he defined when working on trigonometric series, is far from his greatest accomplishment. He made contributions in many fields, including differential geometry and complex analysis. The *Riemann hypothesis* is probably the most famous unsolved problem in mathematics.

In section 4.5 we discuss Fubini's theorem, which reduces computing the integral of a function of n variables to computing n ordinary integrals. This is an important theoretical tool. Moreover, whenever an integral can be computed in elementary terms, Fubini's theorem is the key tool. Unfortunately, it is usually impossible to compute antiderivatives in elementary terms even for functions of one variable, and this tends to be truer yet of functions of several variables.

In practice, multiple integrals are most often computed using numerical methods, which we discuss in section 4.6. We will see that although the theory is much the same in $\mathbb{R}^2$ or $\mathbb{R}^{10^{24}}$, the computational issues are quite different. We will encounter some entertaining uses of Newton's method when looking for optimal points at which to evaluate a function, and some fairly deep probability in understanding why Monte Carlo methods work in higher dimensions.

Defining volume using dyadic pavings, as we do in section 4.1, makes most theorems easiest to prove, but such pavings are rigid; often we will want to have more "paving stones" where the function varies rapidly, and bigger ones elsewhere. Flexibility in choosing pavings is also important for the proof of the *change of variables formula*. Section 4.7 discusses more general pavings.

In section 4.8 we return to linear algebra to discuss higher-dimensional determinants. In section 4.9 we show that in all dimensions the determinant measures volumes; we use this fact in section 4.10, where we discuss the change of variables formula.

Many interesting integrals, such as those in Laplace and Fourier transforms, are not integrals of bounded functions over bounded domains. These use a different approach to integration, *Lebesgue integration*, discussed in section 4.11. Lebesgue integrals cannot be defined as Riemann sums, and require understanding the behavior of integrals under limits. The dominated convergence theorem is the key tool for this.

4.1 Defining the integral

Integration is a summation procedure; it answers the question, how much is there in all? In one dimension, $\rho(x)$ might be the density at point x of a bar parametrized by $[a, b]$; in that case

The Greek letter ρ, or "rho," is pronounced "row."

$$\int_a^b \rho(x)\, dx \tag{4.1.1}$$

is the total mass of the bar.

If instead we have a rectangular plate parametrized by $a \le x \le b$ and $c \le y \le d$, and with density $\rho\begin{pmatrix} x \\ y \end{pmatrix}$, then the total mass will be given by the *double integral*

$$\int\int_{[a,b]\times[c,d]} \rho\begin{pmatrix} x \\ y \end{pmatrix}\, dx\, dy, \tag{4.1.2}$$

where the domain of the entire double integral $\int\int$ is $[a,b]\times[c,d]$ (i.e., the plate). We will define such multiple integrals in this chapter. But you should always remember that the preceding example is too simple. We might want to understand the total rainfall in Britain, whose coastline is a very complicated boundary. (A celebrated article analyzes that coastline as a fractal, with infinite length.) Or we might want to understand the total potential energy stored in the surface tension of a foam; physics tells us that a foam assumes the shape that minimizes this energy.

We will see in section 4.5 that the double integral of equation 4.1.2 can be written

$$\int_c^d \left(\int_a^b \rho\begin{pmatrix} x \\ y\end{pmatrix} dx\right) dy.$$

We are not presupposing this equivalence in this section. One difference worth noting is that $\int_a^b$ specifies a direction: from a to b. (You will recall that direction makes a difference: $\int_a^b = -\int_b^a$.) Equation 4.1.2 specifies a domain but says nothing about direction.

Thus we want to define integration for rather bizarre domains and functions. Our approach will not work for truly bizarre functions, such as the function that equals 1 at all rational numbers and 0 at all irrational numbers; for that one needs Lebesgue integration (see section 4.11). But we still have to specify carefully what domains and functions we want to allow.

Our task will be somewhat easier if we keep the domain of integration simple, putting all the complication into the function to be integrated. If we wanted to sum rainfall over Britain, we would use $\mathbb{R}^2$, *not* Britain (with its fractal coastline!) as the domain of integration; we would then define our function to be rainfall over Britain, and 0 elsewhere.

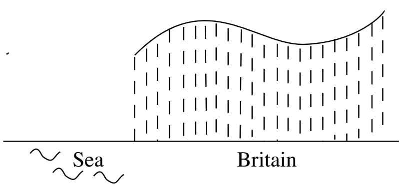

FIGURE 4.1.1.

The function that is rainfall over Britain and 0 elsewhere is discontinuous at the coast.

Thus, for a function $f:\mathbb{R}^n\to\mathbb{R}$, we will define the multiple integral

$$\int_{\mathbb{R}^n} f(\mathbf{x})\,|d^n\mathbf{x}|, \qquad 4.1.3$$

with $\mathbb{R}^n$ the domain of integration.

We emphatically do *not* want to assume that f is continuous, because most often it is not: if, for example, f is defined to be total rainfall for October over Britain, and 0 elsewhere, it will be discontinuous over most of the border of Britain, as shown in figure 4.1.1. What we actually have is a function g (e.g., rainfall) defined on some subset of $\mathbb{R}^n$ larger than Britain. We then consider that function only over Britain, by setting

$$f(\mathbf{x}) = \begin{cases} g(\mathbf{x}) & \text{if } \mathbf{x}\in \text{Britain} \\ 0 & \text{otherwise.}\end{cases} \qquad 4.1.4$$

We can express this another way, using the *indicator function* $\mathbf{1}$.

The indicator function is also known as the *characteristic function*.

Definition 4.1.1 (Indicator function). Let $A\subset\mathbb{R}^n$ be a bounded subset. The *indicator function* $\mathbf{1}_A$ is

$$\mathbf{1}_A(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x}\in A \\ 0 & \text{if } \mathbf{x}\notin A.\end{cases} \qquad 4.1.5$$

Equation 4.1.4 can then be rewritten

$$f(\mathbf{x}) = g(\mathbf{x})\mathbf{1}_{\text{Britain}}(\mathbf{x}). \qquad 4.1.6$$

This doesn't get rid of difficulties like the coastline of Britain – indeed, such a function f will usually have discontinuities on the coastline – but putting all the difficulties on the side of the function will make our definitions easier (or at least shorter).

We tried several notations before choosing $|d^n\mathbf{x}|$. First we used $dx_1 \dots dx_n$. That seemed clumsy, so we switched to dV. But it failed to distinguish between $|d^2\mathbf{x}|$ and $|d^3\mathbf{x}|$, and when changing variables we had to tack on subscripts to keep the variables straight.

But dV had the advantage of suggesting, correctly, that we are not concerned with direction (unlike integration in first year calculus, where $\int_a^b dx \neq \int_b^a dx$). We hesitated at first to convey the same message with absolute value signs, for fear the notation would seem forbidding, but decided that the distinction between oriented and unoriented domains is so important (it is a central theme of chapter 6) that our notation should reflect that distinction.

The notation Supp (support) should not be confused with sup (supremum). Recall that "supremum" and "least upper bound" are synonymous, as are "infimum" and "greatest lower bound" (definitions 1.6.5 and 1.6.7).

In section 4.11 we will discuss *Lebesgue integration*, which will allow us to define integrals of functions that are not bounded or do not have bounded support, or are locally extremely irregular.

You may have seen *improper integrals* of unbounded functions over unbounded domains. But this only works in dimension 1: improper integrals don't make sense in higher dimensions. In any case, improper integrals that are not absolutely convergent are very tricky, and those that are absolutely convergent also exist as Lebesgue integrals.

So while we really want to integrate g (i.e., rainfall) over Britain, we define that integral in terms of the integral of f over $\mathbb{R}^n$, setting

$$\int_{\text{Britain}} g(\mathbf{x})\,|d^n\mathbf{x}| = \int_{\mathbb{R}^n} f(\mathbf{x})\,|d^n\mathbf{x}|. \qquad 4.1.7$$

More generally, when integrating over a subset $A \subset \mathbb{R}^n$,

$$\int_A g(\mathbf{x})\,|d^n\mathbf{x}| = \int_{\mathbb{R}^n} g(\mathbf{x})\mathbf{1}_A(\mathbf{x})\,|d^n\mathbf{x}|. \qquad 4.1.8$$

Some preliminary definitions and notation

Before defining the Riemann integral, we need a few definitions.

Definition 4.1.2 (Support of a function: Supp(f)**).** The *support* of a function $f : \mathbb{R}^n \to \mathbb{R}$ is

$$\operatorname{Supp}(f) = \{\, \mathbf{x} \in \mathbb{R}^n \mid f(\mathbf{x}) \neq 0 \,\}. \qquad 4.1.9$$

Definition 4.1.3 ($M_A(f)$ and $m_A(f)$). If $A \subset \mathbb{R}^n$ is an arbitrary subset, we will denote by

$$\begin{aligned} M_A(f) &= \sup_{\mathbf{x}\in A} f(\mathbf{x}), \text{ the supremum of } f(\mathbf{x}) \text{ for } \mathbf{x} \in A \\ m_A(f) &= \inf_{\mathbf{x}\in A} f(\mathbf{x}), \text{ the infimum of } f(\mathbf{x}) \text{ for } \mathbf{x} \in A. \end{aligned} \qquad 4.1.10$$

Definition 4.1.4 (Oscillation). The *oscillation* of f over A, denoted $\operatorname{osc}_A(f)$, is the difference between its supremum and its infimum:

$$\operatorname{osc}_A(f) = M_A(f) - m_A(f). \qquad 4.1.11$$

Definition of the Riemann integral: dyadic pavings

In sections 4.1–4.10 we discuss integrals of functions $f : \mathbb{R}^n \to \mathbb{R}$ satisfying

1. $|f|$ is bounded, and
2. f has bounded support (i.e., there exists R such that $f(\mathbf{x}) = 0$ when $|\mathbf{x}| > R$).

With these restrictions on f, and for any subset $A \subset \mathbb{R}^n$, each quantity $M_A(f)$, $m_A(f)$, and $\operatorname{osc}_A(f)$, is a well-defined finite number. This is not true for a function like $f(x) = 1/x$, defined on the open interval $(0, 1)$. In that case $|f|$ is not bounded, and $\sup f(x) = \infty$.

There is quite a bit of choice as to how to define the integral; we will first use the most restrictive definition: *dyadic pavings* of $\mathbb{R}^n$.

To compute an integral in one dimension, we decompose the domain into little intervals, and construct on each the tallest rectangle fitting under the graph and the shortest rectangle containing it, as shown in figure 4.1.2.

The dyadic upper and lower sums correspond to decomposing the domain first at the integers, then the half-integers, then the quarter-integers, etc.