Caltech Division of the Humanities and Social Sciences

Notes on Integration by Parts

KC Border

Spring 1996
v. 2020.08.21::13.46

1 The Classical Fundamental Theorems of Calculus

We start with a review of the Fundamental Theorems of Calculus, as presented in Apostol [1]. The notion of integration employed is the Riemann integral.

1 Definition *An* ***indefinite integral*** *F of f over the interval I is any function F such that for some a in I,*

$$F(x) = \int_a^x f(s)\,ds \quad \text{for all } x \text{ in } I.$$

Different values of a give rise to different indefinite integrals of f.

An antiderivative is distinct from the concept of an indefinite integral.

2 Definition *A function P is a* ***primitive*** *or* ***antiderivative*** *of a function f on an interval I if*

$$P'(x) = f(x) \quad \text{for every } x \text{ in } I.$$

Leibniz' notation for this is $\int f(x)\,dx = P(x) + C$. Note that if P is an antiderivative of f, then so is $P + C$ for any constant function C.

Despite the similarity in notation, the statement that P is an antiderivative of f is a statement about the derivative of P, namely that $P'(x) = f(x)$ for all x in I; whereas the statement that F is an indefinite integral of f is a statement about the integral of f, namely that there exists some a in I with $\int_a^x f(s)\,ds = F(x)$ for all x in I. Nonetheless there is a close connection between the concepts, which justifies the similar notation. The connection is laid out in the two Fundamental Theorems of Calculus.

3 Theorem (First Fundamental Theorem of Calculus [1, Theorem 5.1, p. 202]) *Let f be integrable on $[a, x]$ for each x in $[a, b]$. Let $a \leqslant c \leqslant b$, and let F be the indefinite integral of f defined by*

$$F(x) = \int_c^x f(s)\,ds.$$

Then F is differentiable at every x in (a, b) where f is continuous, and at such points $F'(x) = f(x)$.

Therefore an indefinite integral of a continuous function f is also an antiderivative of f.

4 Theorem (Second Fundamental Theorem of Calculus [1, Theorem 5.3, p. 205]) *Let f be continuous on (a, b) and let P be any antiderivative of f on (a, b). The for each x and c in (a, b), we have*

$$P(x) = P(c) + \int_c^x f(s)\,ds.$$

That is, an antiderivative of a continuous function f is also an indefinite integral of f.

2 The Cantor ternary function

Given any number x with $0 \leqslant x \leqslant 1$ there is an infinite sequence $a_1, a_2, \ldots$, where each a_n belongs to $\{0, 1, 2\}$ such that $x = \sum_{n=1}^{\infty} \frac{a_n}{3^n}$. This sequence is called the **ternary representation** of x. If x is of the form $\frac{N}{3^m}$ (in lowest terms), then it has two ternary representations: $x = \sum_{n=1}^{\infty} \frac{a_n}{3^n}$, where $a_m > 0$ and $a_n = 0$ for $n > m$, and another representation of the form $x = \sum_{n=1}^{m-1} \frac{a_n}{3^n} + \frac{a_m - 1}{3^m} + \sum_{n=m+1}^{\infty} \frac{2}{3^n}$. But these are the only cases of a nonunique ternary representation, and there are only countably many such numbers. (See, e.g., Boyd [2, Theorem 1.23, p. 20].)

Given $x \in [0, 1]$, let $N(x)$ be the first n such that $a_n = 1$ in the ternary representation. If x has two ternary representations use the one that gives the larger value of $N(x)$. If x has a ternary representation with no $a_n = 1$, then $N(x) = \infty$. The **Cantor set** $\mathcal{C}$ consists of all numbers x in $[0, 1]$ for which $N(x) = \infty$. That is, those that have a ternary representation where no $a_n = 1$. That is, all numbers x of the form $x = \sum_{n=1}^{\infty} \frac{2b_n}{3^n}$, where each b_n belongs to $\{0, 1\}$. Each distinct sequence of 0s and 1s gives rise to a distinct element of $\mathcal{C}$. Indeed some authors identify the Cantor set with $\{0, 1\}^{\mathbb{N}}$ endowed with its product topology, since the mapping $(b_1, b_2, \ldots) \mapsto \sum_{n=1}^{\infty} \frac{2b_n}{3^n}$ is a homeomorphism. Also note that a sequence $(b_1, b_2, \ldots)$ of 0s and 1s also corresponds to a unique subset of $\mathbb{N}$, namely $\{n \in \mathbb{N} : b_n = 1\}$. Thus there are as many elements $\mathcal{C}$ as there are subset of $\mathbb{N}$, so the Cantor set is uncountable. (This follows from the Cantor diagonal procedure.) Yet the Cantor set includes no interval.

It is perhaps easier to visualize the complement of the Cantor set. Let

$$\mathcal{A}_n = \{x \in [0, 1] : N(x) = n\}.$$

The complement of the Cantor set is $\bigcup_{n=1}^{\infty} \mathcal{A}_n$. Define

$$\mathcal{C}_n = [0, 1] \setminus \bigcup_{k=1}^{n} \mathcal{A}_k,$$

so that $\mathcal{C} = \bigcap_{n=0}^{\infty} \mathcal{C}_n$. Now $\mathcal{A}_1$ consists of those x for which $a_1 = 1$ in its ternary expansion. This means that

$$\mathcal{A}_1 = \left(\tfrac{1}{3}, \tfrac{2}{3}\right) \qquad \text{and} \qquad \mathcal{C}_1 = \left[0, \tfrac{1}{3}\right] \cup \left[\tfrac{2}{3}, 1\right].$$

Note that $N(\frac{1}{3}) = \infty$ since $\frac{1}{3}$ can also be written as $\sum_{n=2}^{\infty} \frac{2}{3^n}$, so $a_1 = 0$, $a_n = 2$ for $n > 1$. Now $\mathcal{A}_2$ consists of those x for which $a_1 = 0$ or $a_1 = 2$ and $a_2 = 1$ in its ternary expansion. Thus

$$\mathcal{A}_2 = \left(\tfrac{1}{9}, \tfrac{2}{9}\right) \cup \left(\tfrac{7}{9}, \tfrac{8}{9}\right) \qquad \text{and} \qquad C_2 = \left[0, \tfrac{1}{9}\right] \cup \left[\tfrac{2}{9}, \tfrac{1}{3}\right] \cup \left[\tfrac{2}{3}, \tfrac{7}{9}\right] \cup \left[\tfrac{8}{9}, 1\right].$$

Each $\mathcal{C}_n$ is the union of 2^n closed intervals, each of length $\frac{1}{3^{n-1}}$, and $\mathcal{A}_{n+1}$ consists of the open middle third of each of the intervals in $\mathcal{C}_n$. The total length of the removed open segments is

$$\frac{1}{3} + 2 \cdot \frac{1}{9} + 4 \cdot \frac{1}{27} + \cdots = \sum_{n=0}^{\infty} \frac{2^n}{3^{n+1}} = \frac{1}{3} \sum_{n=0}^{\infty} \left(\frac{2}{3}\right)^n = \frac{1}{3} \cdot \frac{1}{1 - \frac{2}{3}} = 1.$$

Thus the total length of the Cantor set is $1 - 1 = 0$.

The Cantor ternary function f is defined as follows. On the open middle third $(\frac{1}{3}, \frac{2}{3})$ its value is $\frac{1}{2}$. On the open interval $(\frac{1}{9}, \frac{2}{9})$ its value is $\frac{1}{4}$ and on $(\frac{7}{9}, \frac{8}{9})$ its value is $\frac{3}{4}$. Continuing in this fashion, the function is defined on the complement of the Cantor set. The definition is extended to the entire interval by continuity. See Figure 1. A more precise but more opaque

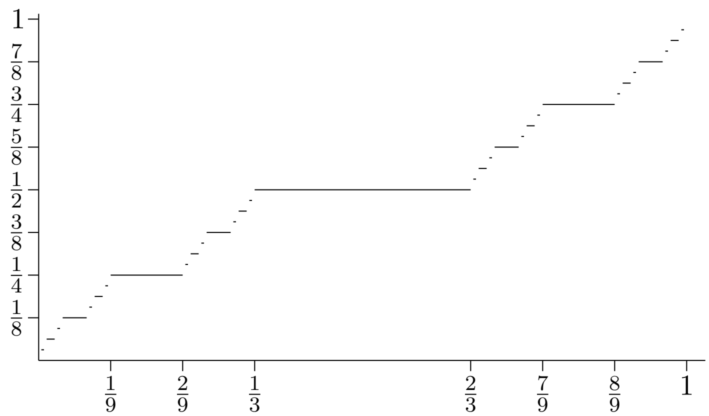

Figure 1. Partial graph of the Cantor ternary function.

definition is this:

$$f(x) = \begin{cases} \sum_{n=1}^{N(x)-1} \frac{\frac{1}{2}a_n}{2^n} + \frac{a_{N(x)}}{2^{N(x)}} & \text{if } N(x) < \infty, \\ \sum_{n=1}^{\infty} \frac{\frac{1}{2}a_n}{2^n} & \text{if } N(x) = \infty. \end{cases}$$

In any event notice that f is constant on each open interval in some $\mathcal{A}_n$, so it is differentiable there and $f' = 0$. Thus f is differentiable almost everywhere, and $f' = 0$ wherever it exists, but

$$f(1) - f(0) = 1 \quad \text{and} \quad \int_0^1 f'(x)\,dx = 0.$$

3 The Classic Integration by Parts Theorem

The Fundamental Theorems enable us to prove the following result.

5 Theorem (Integration by Parts) *Suppose f and g are continuously differentiable on the open interval I. Let $a < b$ belong to I. Then*

$$\int_a^b f(x)g'(x)\,dx + \int_a^b f'(x)g(x)\,dx = f(b)g(b) - f(a)g(a).$$

Proof based on Apostol [1, Section 5.9, pp. 217–218]: Define $h(x) = f(x)g(x)$. Then h is continuously differentiable on I and $h'(x) = f(x)g'(x) + f'(x)g(x)$. That is, h is an antiderivative of the continuous function $f(x)g'(x) + f'(x)g(x)$. So by the Second Fundamental Theorem of Calculus

$$h(b) - h(a) = \int_a^b f(x)g'(x) + f'(x)g(x)\,dx.$$

∎

This result is usually written less symmetrically as

$$\int_a^b f(x)g'(x)\,dx = f(b)g(b) - f(a)g(a) - \int_a^b f'(x)g(x)\,dx,$$

where the integral on the left is one that you already know how to evaluate. Sometimes one uses the language of change of variables: Letting $u = f(x)$ and $v = g(x)$, write $du = f'(x)\,dx$, $dv = g'(x)\,dx$, and

$$\int u\,dv = uv - \int v\,du.$$

4 A more general result

To apply the Second Fundamental Theorem of Calculus, we need $f'g + g'f$ to be a continuous function. The only reasonable sufficient condition for this is that f and g be continuously differentiable. However, Fubini's Theorem 10 on interchanging the order of integration allows us to prove the integration by parts formula under weaker conditions. All we need is that f and g be indefinite integrals. That is, we do not need f and g to be differentiable everywhere, only that they are indefinite integrals. This assumption is equivalent to what is called **absolute continuity**. In statistics, it is traditional to use upper case letters for distribution functions, and lower case letters for their densities, so I'll adopt that notation.

6 Theorem (Integration by Parts, Part II) *Suppose F and G satisfy*

$$F(x) = F(a) + \int_a^x f(s)\,ds$$

and

$$G(x) = G(a) + \int_a^x g(s)\,ds$$

for every x in $[a,b]$, where f and g are integrable over $[a,b]$ and fg is integrable over $[a,b]\times[a,b]$. Then

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

Proof based on Fubini's Theorem: Define the function $h\colon [a,b]\times[a,b] \to \boldsymbol{R}$ by

$$h(s,t) = f(s)g(t).$$

Fubini's Theorem 10 below tells us that we can compute the integral of h over the rectangle $[a,b] \times [a,b]$ as an iterated integral in either order:

$$\iint h(s,t)\,d(s,t) = \int_a^b \left(\int_a^b h(s,t)\,ds \right) dt = \int_a^b \left(\int_a^b h(s,t)\,dt \right) ds.$$

For our nice choice of h this becomes

$$\begin{aligned}
\iint h(s,t)\,d(s,t) &= \int_a^b \left(\int_a^b f(s)g(t)\,ds \right) dt \\
&= \int_a^b g(t) \left(\int_a^b f(s)\,ds \right) dt \\
&= \int_a^b g(t)\big(F(b) - F(a)\big)\,dt \\
&= \big[F(b) - F(a)\big] \int_a^b g(t)\,dt \\
&= \big[F(b) - F(a)\big]\big[G(b) - G(a)\big] \\
&= F(b)G(b) + F(a)G(a) - F(b)G(a) - F(a)G(b). \qquad (1)
\end{aligned}$$

But there is another way to compute the integral over the rectangle. Divide it into two triangles with the diagonal $s = t$. See Figure 2. This is equivalent to the following construction. Let $U = \{(s,t) : t > s\}$ and let $L = \{(s,t) : t > s\}$. Then for all (s,t),

$$f(s)g(t) = \mathbf{1}_U(s,t)f(s)g(t) + \mathbf{1}_L(s,t)f(s)g(t).$$

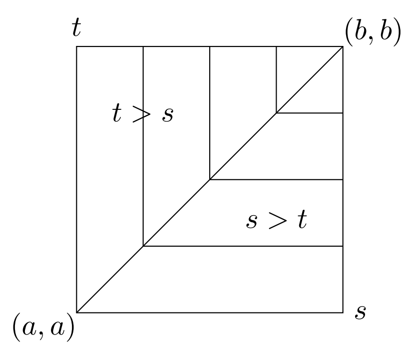

Figure 2. Splitting the rectangle.

Let $I_U = \iint \mathbf{1}_U(s,t)h(s,t)\,d(s,t)$ and $I_L = \iint \mathbf{1}_L(s,t)h(s,t)\,d(s,t)$. By Fubini's Theorem each of I_U and I_L can be calculated as an iterated integral in either order. So on the upper triangle U, for each s let's first integrate h with respect to t over the interval $(s,b]$ (shown as a vertical line segment in Figure 2) to get a function of s alone, and then integrate that result with respect to s as s runs over $[a, b]$.

$$\begin{aligned}
I_U &= \iint \mathbf{1}_U(s,t)h(s,t)\,d(s,t) \\
&= \int_a^b \left(\int_s^b f(s)g(t)\,dt \right) ds \\
&= \int_a^b f(s)\big(G(b) - G(s)\big)\,ds \\
&= G(b)\big[F(b) - F(a)\big] - \int_a^b f(s)G(s)\,ds.
\end{aligned}$$

In the lower rectangle L, we reverse the order of integration.

$$\begin{aligned}
I_L &= \iint \mathbf{1}_U(s,t)h(s,t)\,d(s,t) \\
&= \int_a^b \left(\int_t^b f(s)g(t)\,ds \right) dt \\
&= \int_a^b g(t)\big(F(b) - F(t)\big)\,dt \\
&= F(b)\big[G(b) - G(a)\big] - \int_a^b g(t)F(t)\,dt.
\end{aligned}$$

The integral rect over the rectangle is the sum of the two triangular integrals, so

$$\begin{aligned}
&F(b)G(b) + F(a)G(a) - F(b)G(a) - F(a)G(b) = \\
&\qquad G(b)\big[F(b) - F(a)\big] - \int_a^b f(s)G(s)\,ds + F(b)\big[G(b) - G(a)\big] - \int_a^b g(t)F(t)\,dt.
\end{aligned}$$

This simplifies to

$$\int_a^b f(s)G(s)\,ds + \int_a^b g(t)F(t)\,dt = F(b)G(b) - F(a)G(a).$$

Now we simply note that s and t are dummy variables and may be replaced by x to get the integration by part formula.

■

If you are astute and suspicious, you will not that I used the notation $\int_t^b$ to mean the integral over the closed interval $[t, b]$, but $\int_s^b$ to indicate the integral over the half-open interval $(s, b]$. Under our assumptions on F and G (absolute continuity), it doesn't make a difference, since points contribute zero to the value of the integral. But more generally, in statistics and probability we consider distributions where points may have strictly positive probability. For that case, the Riemann integral is not adequate, and the difference between closed and open intervals may matter. An integration by parts formula for that case is discussed in the next section.

5 Finite measures and nondecreasing functions

Let μ be a finite (nonnegative) measure on the Borel subsets of $\boldsymbol{R}$.[1] Define $F_\mu\colon \boldsymbol{R} \to \boldsymbol{R}_+$ by

$$F_\mu(x) = \mu(\{y \in \boldsymbol{R} : y \leqslant x\}).$$

The function F_μ is called the distribution function of μ, and has the following properties:

1. F_μ is nondecreasing.
2. F_μ is right continuous. That is, $F_\mu(x) = \lim_{y \downarrow x} F_\mu(y)$.
3. $\lim_{x\to-\infty} F_\mu(x) = 0$.
4. $\lim_{x\to\infty} F_\mu(x) = \mu(\boldsymbol{R})$.
5. $F(b) - F(a) = \mu\big((a,b]\big)$ for $a < b$.

Conversely, for any $F\colon \boldsymbol{R} \to \boldsymbol{R}_+$ satisfying

1. F is nondecreasing.
2. F is right continuous.
3. $\lim_{x\to-\infty} F(x) = 0$.
4. $\lim_{x\to\infty} F(x) < \infty$.

there is a unique nonnegative Borel measure μ_f satisfying $\mu_F\big((a,b]\big) = F(b) - F(a)$ for $a < b$. Given a distribution function $F\colon \boldsymbol{R} \to \boldsymbol{R}_+$ and a μ_F-integrable function g, the **Lebesgue–Stieltjes integral**

$$\int g\,dF = \int g\,d\mu_F$$

by definition. I learned this next theorem from Naresh Jain.

7 Integration by Parts for Distribution Functions *Let F and G be distribution functions on $\boldsymbol{R}$. Then*

$$\int_{(a,b]} F(x)\,dG(x) + \int_{(a,b]} G(x^-)\,dF(x) = F(b)G(b) - F(a)G(a), \tag{2}$$

where $G(x^-) = \lim_{y\uparrow x} G(y)$.

Proof: Define $A = \{(x,y) \in (a,b]^2 : x \leqslant y\}$. By Fubini's Theorem 10 on iterated integrals, we have

$$\iint \mathbf{1}_A\,d(\mu_G \times \mu_F) = \begin{cases} \displaystyle\int_{(a,b]} \left(\int_{(a,b]} \mathbf{1}_A\,d\mu_F\right) d\mu_G = \int_{(a,b]} \big(F(x) - F(a)\big)\,d\mu_G(x) \\ \text{or} \\ \displaystyle\int_{(a,b]} \left(\int_{(a,b]} \mathbf{1}_A\,d\mu_G\right) d\mu_F = \int_{(a,b]} \big(G(b) - G(y^-)\big)\,d\mu_F(y), \end{cases}$$

[1] See, e.g., Halmos [3] for more on Borel measures.

where $\mathbf{1}_A$ is the indicator function defined by $\mathbf{1}_A(x,y) = \begin{cases} 1 & (x,y) \in A \\ 0 & (x,y) \notin A \end{cases}$.

Rearrange to get

$$\int_{(a,b]} (F(x) - F(a))\, d\mu_G(x) = \int_{(a,b]} \big(G(b) - G(y^-)\big)\, d\mu_F(y)$$

or

$$\int_{(a,b]} F(x)\, dG(x) - F(a)\big(G(b) - G(a)\big) = G(b)\big(F(b) - F(a)\big) - \int_{(a,b]} G(y^-)\, d\mu_F(y),$$

from which the conclusion follows. ∎

8 Corollary *If either F or G is continuous, then*

$$\int_{[a,b]} F(x)\, dG(x) + \int_{[a,b]} G(x)\, dF(x) = F(b)G(b) - F(a)G(a). \tag{2'}$$

Proof: If G is continuous, then $G(x^-) = G(x)$, so IP-df implies IP-df'. Now simply note that the statement is symmetric in F and G. ∎

9 Corollary *Let F be a cumulative distribution function with $F(0) = 0$ and $\lim_{x\to\infty} F(x) = 1$, that is, the cumulative distribution function of a nonnegative random variable. Then for any $p > 0$,*

$$\int_{[0,\infty)} x^p\, dF(x) = p \int_0^\infty (1 - F(x)) x^{p-1}\, dx$$

Proof: Fix $b > 0$ and set

$$G_b(x) = \begin{cases} 0 & x \leqslant 0 \\ x^p & 0 \leqslant x \leqslant b \\ b^p & x \geqslant b \end{cases}$$

and note that G_b is a continuous distribution function. By Corollary 8,

$$\begin{aligned} \int_0^b x^p\, dF &= F(b)b^p - \int_0^b F(x)\, dG_b(x). \\ &= F(b)b^p - p\int_0^b F(x) x^{p-1}\, dx \\ &= p\int_0^b (F(b) - F(x)) x^{p-1}\, dx, \end{aligned}$$

since G_b has derivative px^{p-1} on $(0,b)$. Now let $b \to \infty$. ∎

6 Fubini's Theorem

There is a collection of related results that are all referred to as Fubini's theorem. This version is taken from Halmos [3, Theorem C, p. 148].

10 Fubini's Theorem *Let* $(X, \mathcal{S}, \mu)$ *and* $(Y, \mathcal{T}, \nu)$ *be* σ*-finite measure spaces. If* $f\colon X \times Y \to \boldsymbol{R}$ *is* $\mu \times \nu$*-integrable, then* $x \mapsto \int_Y f(x, y)\, d\nu(y)$ *and* $y \mapsto \int_X f(x, y)\, d\mu(x)$ *are* μ*-integrable and* ν*-integrable respectively, and*

$$\int_{X \times Y} f(x, y)\, d(\mu \times \nu)(x, y) = \int_X \left(\int_Y f(x, y)\, d\nu(y) \right) d\mu(x) = \int_Y \left(\int_X f(x, y)\, d\mu(x) \right) d\nu(y).$$

References

[1] Apostol, T. M. 1967. *Calculus*, 2d. ed., volume 1. Waltham, Massachusetts: Blaisdell.

[2] Boyd, D. 1972. Classical analysis, volume 1. Notes prepared for Ma 108 abc. Published occasionally since at least 1972 by the Department of Mathematics, California Institute of Technology, 253-37, Pasadena CA 91125.

[3] Halmos, P. R. 1974. *Measure theory*. Graduate Texts in Mathematics. New York: Springer–Verlag. Reprint of the edition published by Van Nostrand, 1950.