Improper Integrals

Louis A. Talman
Department of Mathematical & Computer Sciences
Metropolitan State College of Denver

November 8, 2006

1 Improper Integrals on Finite Intervals

The standard definitions for improper integrals often raise questions in students' minds. In order to treat the reasons why we make the definitions as we do, we begin with a naïve approach to the topic.

Let f be a function which is continuous at every point of the interval $(a, b]$. If we choose R so that $a < R \le b$, then the standard existence theorem for the definite integral assures us that $\int_R^b f(t)\,dt$ exists. But the standard theorem gives us no information regarding $\int_a^b f(t)\,dt$. Indeed, $\int_a^b f(t)\,dt$ need not exist in the same sense that the integrals $\int_R^b f(t)\,dt$ exist. The following example makes this statement precise. (We use the notation $\mathcal{S}(f, [0,1], \mathcal{P}, \mathcal{P}')$ for the Riemann sum $\sum_{k=1}^n f(\xi_k)(x_k - x_{k-1})$, where $\mathcal{P} = \{x_0, x_1, x_2, \ldots, x_n\}$ is a partition of the interval $[0,1]$ and $\mathcal{P}' = \{\xi_k : x_{k-1} \le \xi_k \le \xi_{k-1}$ and $k = 1, 2, \ldots, n\}$ is a network of points ξ_k, $k = 1, \ldots, n$, such that $x_{k-1} \le \xi_k \le x_k$ for every k.)

1.1 Example:

Let f be given by $f(x) = 1/\sqrt{x}$ on $(0,1]$. Then $\lim_{\|\mathcal{P}\|\to 0} \mathcal{S}(f, [0,1], \mathcal{P}, \mathcal{P}')$ does not exist.

Analysis: Consider the Riemann sums

$$\mathcal{S}(f, [0,1], \mathcal{P}_n, \mathcal{P}_n') = \sum_{k=1}^{n} f(\xi_k)(x_k - x_{k-1}), \tag{1}$$

where the points x_k of the partition $\mathcal{P}_n$ are chosen so that $x_1 = 1/n$, and the points ξ_k of the network $\mathcal{P}_n'$ are chosen so that $\xi_1 = 1/n^4$. (We are indifferent as to how the other points of the partition $\mathcal{P}_n$ and the network $\mathcal{P}_n'$ are chosen, as long as the standard requirements

$0 = x_0 < x_1 < x_2 < \cdots < x_{n-1} < x_n = 1$ and $x_{k-1} \le \xi_k \le x_k$ are all met.) For each $n \in \mathbb{N}$, we have

$$\mathcal{S}(f, [0,1], \mathcal{P}_n, \mathcal{P}'_n) = \sum_{k=1}^{n} f(\xi_k)(x_k - x_{k-1}) \tag{2}$$

$$= \frac{1}{\sqrt{1/n^4}} \cdot \frac{1}{n} + \sum_{k=2}^{n} f(\xi_k)(x_k - x_{k-1}) \tag{3}$$

$$\ge \frac{n^2}{1} \cdot \frac{1}{n} = n. \tag{4}$$

We can therefore find partitions of arbitrarily small norm for which the corresponding Riemann sums can be arbitrarily large. This means that $\lim_{\|\mathcal{P}\| \to 0} \mathcal{S}(f, [0,1], \mathcal{P}, \mathcal{P}')$ does not exist.●

The reader should verify that, with the same choices we have made above, but with $g(x) = 1/x^2$, a similar phenomenon occurs. In fact, if h is *any* function continuous on an interval $(a, b]$, but for which $\lim_{x \to a^+} h(x) = \infty$, a similar argument shows that $\lim_{\|\mathcal{P}\| \to 0} \mathcal{S}(h, [a,b], \mathcal{P}, \mathcal{P}')$ does not exist. The underlying idea is that, no matter how small the norm of our partition may be, the integrand is unbounded in the left-most subdivision. In consequence, we can find a point there where the value of the function is so large that its product with the length of the subdivision is as large as we wish.

There is, however, an important difference between the two examples f and g of the preceding paragraphs. If we choose $R \in (0, 1]$, then we have

$$\int_R^1 f(t)\,dt = \int_R^1 \frac{dt}{\sqrt{t}} \tag{5}$$

$$= 2\sqrt{t}\Big|_R^1 \tag{6}$$

$$= 2 - 2\sqrt{R}, \tag{7}$$

and $\lim_{R \to 0^+} \int_R^1 f(t)\,dt$ exists. On the other hand,

$$\int_R^1 g(t)\,dt = \int_R^1 \frac{dt}{t^2} \tag{8}$$

$$= -\frac{1}{t}\Big|_R^1 \tag{9}$$

$$= -1 + \frac{1}{R}, \tag{10}$$

and $\lim_{R\to 0+}\int_R^1 g(t)\,dt$ doesn't exist. This striking difference in behaviors is what prompts the standard definition of an *improper integral* $\int_a^b f(t)\,dt$ for a function given continuous on $(a,b]$.

1.2 Definition (Improper Integral Ia):

Let f be a function which is continuous on an interval $(a,b]$. If $\lim_{R\to a^+}\int_R^b f(t)\,dt$ exists, we say that the improper integral $\int_a^b f(t)\,dt$ converges, *and we assign the value of the limit to the integral. If the limit does not exist, we say that* the improper integral $\int_a^b f(t)\,dt$ diverges, *and we assign no value to the integral.*

And, of course, the situation is symmetric; *mutatis mutandis*, we find that the same kinds of things happen for functions continuous on intervals $[a,b)$. Thus, we have another definition.

1.3 Definition (Improper Integral Ib):

Let f be a function which is continuous on an interval $[a,b)$. If $\lim_{S\to b^-}\int_a^S f(t)\,dt$ exists, we say that the improper integral $\int_a^b f(t)\,dt$ converges, *and we assign the value of the limit to the integral. If the limit does not exist, we say that* the improper integral $\int_a^b f(t)\,dt$ diverges, *and we assign no value to the integral.*

In both instances, the term *improper* reflects the fact that, because f has a discontinuity at one end of the interval $[a,b]$, we don't expect $\int_a^b f(t)\,dt$ to have its usual meaning as a limit of Riemann sums. As we have seen, the latter limit may not exist; improper integrals really are something different from definite integrals.

We remark that some authors ([1], for example) make it explicit that an integral $\int_a^b f(t)\,dt$ is an improper integral of the kind in Definition 1.2 by writing $\int_{a^+}^b f(t)\,dt$ for such integrals. The corresponding notation for integrals of the Definition 1.3 flavor is $\int_a^{b^-} f(t)\,dt$. This notation is helpful, but not commonly encountered.

So far, so good, but we have dealt only with functions that have just one singularity—and that at an end-point of the interval of integration. Two questions now confront us.

1. What are we to do if the integrand has a singularity *interior* to the interval of integration?

2. What are we to do if the integrand has singularities at *both* ends of the interval of integration?

The key to answering these two questions lies in this: *We would like for the algebra of our new, extended, notion of an integral to be the same as the algebra of our original definite integral.* The show-stopper here turns out to be the requirement that when $a < c < b$ we want to have

$$\int_a^b f(t)\,dt = \int_a^c f(t)\,dt + \int_c^b f(t)\,dt. \tag{11}$$

This innocent-looking equation causes difficulty in some—but not all—instances when f has an unbounded discontinuity at c, and in some—but not all—instances when f has unbounded discontinuities at both a and b. So we must take care in the way we assign meaning to the symbol $\int_a^b f(t)\,dt$ in such cases.

1.4 Example:

Let g be the function given by $g(x) = 1/x$. There is no way, consistent with the algebra of definite integrals, to assign numerical meaning to the improper integral $\int_{-1}^1 g(t)\,dt$.

Analysis: An argument like that of Example 1.1 shows that $\lim_{\|\mathcal{P}\|\to 0} \mathcal{S}(g, [-1, 1], \mathcal{P}, \mathcal{P}')$ does not exist. We must be a little more finicky about our partitions than we were in Example 1.1, however. We can consider partitions symmetric about the origin that contain consecutive points $-1/n$, 0, and $1/n$ with associated networks that contain the points $-1/n \in [-1/n, 0]$ and $1/n^2 \in [0, 1/n]$ but are otherwise also symmetric about the origin. Every Riemann sum arising from such a choice of partition and associated network yields $n - 1$, so the integral is indeed improper in the sense that we can't interpret it as a limit of Riemann sums.

Because $f(-x) = -f(x)$, it is tempting to assign the value 0 to the integral $\int_{-1}^1 (1/t)\,dt$. This is certainly consistent with the idea that $\int_{-1}^1 (1/t)\,dt$ should be the limit, as $R \to 0^+$ of the sum $\int_{-1}^{-R} (1/t)\,dt + \int_R^1 (1/t)\,dt$. But asking that $\int_{-1}^{-1} (1/t)\,dt = \int_{-1}^0 (1/t)\,dt + \int_0^1 (1/t)\,dt$ leads to profound difficulties, because *neither of the improper integrals on the right side of the latter equation has meaning.* Notice that we cannot dismiss this difficulty lightly by writing $\int_{-1}^0 (1/t)\,dt = -\infty$, $\int_0^1 (1/t)\,dt = \infty$, and then passing off $-\infty + \infty$ as zero. Doing so gets us into immediate trouble with $\int_{-2}^1 (1/t)\,dt$, which then becomes $\int_{-2}^0 (1/t)\,dt + \int_0^1 (1/t)\,dt$. The latter must be $-\infty + \infty$ (which we have tried to call zero) on the one hand, but must also be $\int_{-2}^{-1} (1/t)\,dt + \int_{-1}^1 (1/t)\,dt$ (which consistency requires us to evaluate as $-\ln 2 + 0$) on the other hand.•

In fact, we must not assign a general meaning to the expression "$-\infty + \infty$". And the reason that this is so is precisely the existence of examples, like the one we are examining, where assigning a general meaning leads us to a contradiction. Recall that the symbolism $\int_0^1 (1/t)\,dt = \infty$ means that $\lim_{R\to 0^+} \int_R^1 (1/t)\,dt = \infty$, and the symbolism $\lim_{x\to a^+} F(x) = \infty$ is a short-hand for the statement that $\lim_{x\to a^+} F(x)$ does not exist on account of a specific kind of behavior indulged in by the function F. And we may never assert the existence, without further justification, of the limit of a sum when the two limits of the summands do not exist. In this case, there is no way to find further justification.

There is nothing inherently wrong with wanting equation (11) to work for improper integrals. If we interpret the equation carefully enough in the case where the integrand has a discontinuity at c, we can achieve a workable definition. The important insight that leads to a workable definition is that we must evaluate the two improper integrals on the right side of the equation *independently* of each other; that is, we must insist

1. that $\int_a^c f(t)\,dt$ converge and
2. that $\int_c^b f(t)\,dt$ converge

before we try to assign meaning to $\int_a^b f(t)\,dt$. Notice that this meaning is implicit in the symbols of equation (11), because we assign no meaning to the integral from a to c when it diverges, and we assign no meaning to the integral from c to b when it diverges. Thus, (11) is meaningless if *either* of the integrals on the right side diverges. When both the integral from a to c and the integral from c to b converge, we can replace each with a real number—and we can always add real numbers. Thus, the definition for an improper integral where the integrand has a singularity interior to the interval of integration has to go as follows.

1.5 Definition (Improper Integral Ic):

Let f be a function which is continuous on $[a, c) \cup (c, b]$. If $\lim_{S\to c^-} \int_a^S f(t)\,dt$ exists and $\lim_{R\to c^+} \int_R^b f(t)\,dt$ exists, we say that the improper integral $\int_a^b f(t)\,dt$ converges, *and we assign the sum of the limits as the value of $\int_a^b f(t)\,dt$. If either limit fails to exist (or if both fail to exist), we say that* the improper integral $\int_a^b f(t)\,dt$ diverges, *and we assign no value to the integral.*

A very similar analysis, also based upon (11), leads to a definition when f has singularities at both endpoints of $[a, b]$.

1.6 Definition (Improper Integral Ic):

Let f be a function which is continuous on (a, b). Choose c arbitarily in (a, b). If $\lim_{S\to a^+} \int_S^c f(t)\,dt$ exists and $\lim_{R\to b^-} \int_c^R f(t)\,dt$ exists, we say that the improper integral $\int_a^b f(t)\,dt$ converges, *and we assign the sum of the limits as the value of $\int_a^b f(t)\,dt$.*

If either limit fails to exist (or if both fail to exist), we say that the improper integral $\int_a^b f(t)\,dt$ diverges, *and we assign no value to the integral.*

Technically, we must show that this definition is meaningful, because different people might choose different numbers for c, and we want to be sure that those different people will arrive at common values for $\int_a^b f(t)\,dt$. However, if $a < c < c' < b$, then whatever $S \in (a, c)$ may be, $\int_S^{c'} f(t)\,dt = \int_S^c f(t)\,dt + \int_c^{c'} f(t)\,dt$, so that

$$\lim_{S\to a^+} \int_S^{c'} f(t)\,dt = \lim_{S\to a^+} \left[\int_S^c f(t)\,dt + \int_c^{c'} f(t)\,dt\right] \tag{12}$$

$$= \lim_{S\to a^+} \left[\int_S^c f(t)\,dt\right] + \int_c^{c'} f(t)\,dt, \tag{13}$$

while

$$\lim_{R\to b^-} \int_{c'}^{R} f(t)\,dt = \lim_{R\to b^-} \left[\int_{c'}^c f(t)\,dt + \int_c^{R} f(t)\,dt\right] \tag{14}$$

$$= \lim_{R\to b^-} \left[\int_c^R f(t)\,dt\right] - \int_c^{c'} f(t)\,dt, \tag{15}$$

These observations show that our fears that different people might get different numbers were without foundation because we may now write

$$\begin{aligned}\lim_{S\to a^+} \int_S^{c'} f(t)\,dt + \lim_{R\to b^-} \int_{c'}^{R} f(t)\,dt &= \lim_{S\to a^+} \left[\int_S^c f(t)\,dt\right] + \int_c^{c'} f(t)\,dt \\ &\quad + \lim_{R\to b^-} \left[\int_c^R f(t)\,dt\right] - \int_c^{c'} f(t)\,dt \qquad (16) \\ &= \lim_{S\to a^+} \left[\int_S^c f(t)\,dt\right] + \lim_{R\to b^-} \left[\int_c^R f(t)\,dt\right] \qquad (17)\end{aligned}$$

regardless of the choices made for c and c'.

2 Improper Integrals on Infinite Intervals

We now turn to integrals of the form $\int_a^\infty f(t)\,dt$, where we suppose that f is continuous on the interval $[a, \infty)$. Once again, we find that we cannot interpret the symbol $\int_a^\infty f(t)\,dt$ as a limit of Riemann sums, so the term *improper* is appropriate for these integrals. However, this time the reason why no such limit can enter the picture is even more compelling than the reason we found in our earlier analysis: No finite partition of the interval $[a, \infty)$ can possibly result in intervals $[x_{k-1}, x_k]$, $k = 0, 1, \ldots, n$, of finite length but whose union is the entire interval $[a, \infty)$, even if we fudge and allow $x_n = \infty$. Thus, it isn't even possible to *form* the Riemann sums $\mathcal{S}(f, [a, \infty), \mathcal{P}, \mathcal{P}')$ that we need before we can pass to a limit as $\|\mathcal{P}\| \to 0$.

But the fundamental existence theorem does guarantee us that if we choose any real number T, $a < T < \infty$, the integral $\int_a^T f(t)\,dt$ exists. Examples such as

$$\int_1^T \frac{dt}{t^2} = -\frac{1}{t}\Big|_1^T \tag{18}$$

$$= -\frac{1}{T} + 1 \to 1 \text{ as } T \to \infty \tag{19}$$

and

$$\int_1^T \frac{dt}{\sqrt{t}} = 2\sqrt{t}\Big|_1^T \tag{20}$$

$$= 2\sqrt{T} - 2 \to \infty \text{ as } T \to \infty \tag{21}$$

convince us that, in general, $\lim_{T\to\infty} \int_a^\infty f(t)\,dt$ may or may not exist. These observations lead us to the standard definition for this kind of improper integral.

2.1 Definition (Improper Integral IIa):

Let f be a function which is continuous on $[a, \infty)$. If $\lim_{T\to\infty} \int_a^T f(t)\,dt$ exists, we say that the improper integral $\int_a^\infty f(t)\,dt$ converges, *and we assign the value of the limit as the value of the integral. If the limit does not exist, we say that* the improper integral $\int_a^\infty f(t)\,dt$ diverges, *and we assign no value to the integral.*

It is now clear what we must do with integrals of the form $\int_\infty^a f(t)\,dt$.

2.2 Definition (Improper Integral IIb):

Let f be a function which is continuous on $(-\infty, a)$. If $\lim_{U\to-\infty} \int_U^a f(t)\,dt$ exists, we say that the improper integral $\int_{-\infty}^a f(t)\,dt$ converges, *and we assign the value of the limit as the value of the integral. If the limit does not exist, we say that* the improper integral $\int_{-\infty}^a f(t)\,dt$ diverges, *and we assign no value to the integral.*

It remains to deal with combinations of the five different kinds of improper integrals we have so far discussed. Because a naïve approach causes problems similar to those we encountered when we tried to interpret $\int_{-1}^1 (1/t)\,dt$ as zero, we must always break such combinations up as we did with this latter integral. Thus, if f is continuous on the whole real line we must intepret $\int_{-\infty}^{\infty} f(t)\,dt$ as the sum of the two improper integrals $\int_{-\infty}^{a} f(t)\,dt$ and $\int_{a}^{\infty} f(t)\,dt$. (We get to pick the number a at our convenience—see the discussion following Definition 1.6.) Both $\int_{-\infty}^{a} f(t)\,dt$ and $\int_{a}^{\infty} f(t)\,dt$ must converge if we are to allow that $\int_{-\infty}^{\infty} f(t)\,dt$ converges. If g is continuous on $(a,b) \cup (b,\infty)$, we must consider all of the improper integrals $\int_{a}^{c} f(t)\,dt$, $\int_{c}^{b} f(t)\,dt$, $\int_{b}^{d} f(t)\,dt$, and $\int_{d}^{\infty} f(t)\,dt$, where c is a point chosen from (a,b) and d is a point chosen from (b,∞). All four of these improper integrals must converge before we can conclude that $\int_{a}^{\infty} f(t)\,dt$ converges.

References

[1] Widder, D. V., *Advanced Calculus*, second edition, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1961