LIMITS AT INFINITY AND IMPROPER INTEGRALS

MATH 153, SECTION 59 (VIPUL NAIK)

Corresponding material in the book: Section 11.7.

What students should already know: The intuitive definition of limit, at least a dim memory of the ε-δ definition of limit, graphing a function, the concepts of vertical tangents and cusps, vertical asymptotes, and horizontal asymptotes. Basic rules for limits at infinity.

What students should definitely get: The few key ideas about taking limits at infinity involving logarithmic, exponential, trigonometric, and inverse trigonometric functions. The application of these to improper integrals.

Executive summary

Words ...

(1) The integral $\int_a^\infty f(x)\,dx$ is defined as the limit $\lim_{L\to\infty} \int_a^L f(x)\,dx$. If F is an antiderivative of f, this equals $\lim_{L\to\infty} F(L) - F(a)$.
(2) The integral $\int_{-\infty}^a f(x)\,dx$ is defined as the limit $\lim_{L\to-\infty} \int_L^a f(x)\,dx$. If F is an antiderivative of f, this equals $F(a) - \lim_{L\to-\infty} F(L)$.
(3) The integral $\int_{-\infty}^\infty f(x)\,dx$ is defined as a double limit, where the upper limit of integration is limited to infinity, while the lower limit of integration is limited to negative infinity. If F is an antiderivative of f, this equals $\lim_{L\to\infty} F(L) - \lim_{M\to-\infty} F(M)$.
(4) Another kind of improper integral occurs where the function is integrated over an interval and is not defined at an endpoint of the interval. Here, we take the limit over intervals of integration where the interval gradually tends towards the trouble points. For instance, if a function f is to be integrated over $[a, b]$ but b is a trouble point, we take $\lim_{c\to b} \int_a^c f(x)\,dx$. If F is an antiderivative of f, then if F extends continuously to b, this is just equal to $F(b) - F(a)$.
(5) In general, if there are multiple trouble points, we first partition the interval of integration so that all the trouble points are at the partition boundaries. We then use the limiting procedure on each piece and add up across the pieces.

Actions ...

(1) The most straightforward way of computing an indefinite integral is to compute the corresponding antiderivative and take the difference between the upper and lower limits.
(2) In some cases, this is either infeasible or terribly messy. In these cases, we may use the various other methods for computing definite integrals that bypass computing the antiderivative. These include the use of symmetry and a combination of u-substitution plus noticing that after the substitution, the upper and lower limits of integration become the same.
(3) In yet other cases, taking the limit of the antiderivative may be hard, and we may need to use all the techniques discussed in preceding subsections for computing this antiderivative.

1. Graphing: a bird's eye review

You have probably studied how to graph a function. I have taught graphing in 151/152 in previous years following the outline in the book (Section 4.6) with a bit of augmentation. The book breaks the problem down into seven easy steps, and I augmented some additional details to the steps. The overall idea was: *find the domain*, *find the intervals of interest*, *find the points and limiting behaviors of interest*, and *find the symmetries of interest*. Thus, we determined the intervals on which the function was increasing/decreasing or concave up/down, we found the critical points, inflection points, and local extreme values, as well as the points of discontinuity, the vertical tangents and cusps, and the vertical and horizontal asymptotes. We also checked for even functions (more generaly, mirror symmetry) and odd functions (more generally, half-turn symmetry). Then, we combined all this information to graph the function.

2. Limits at infinity

Let us now explore the question of how we in general determined the limit for a function at $\pm\infty$. Recall that, from the perspective of the function's graph, this is equivalent to determining the horizontal asymptotes. At the time we studied this problem, we were limited by the small class of functions we were dealing with. Now, we have expanded our collection of functions considerably, and must accordingly deal with the much larger number of limit questions.

(If these points do not all seem familiar to you, please review the discussion of limits at infinity, which was a subtopic the previous quarter, covered in the lecture on infinity, cusps, tangents, and asymptotes).

(1) There is a bunch of rules, such as $(\to\infty) \times (\to\infty) = \to\infty$.
(2) For a nonconstant polynomial, the limit at $+\infty$ depends on the sign of its leading coefficient, while the limit at $-\infty$ depends on the sign of its leading coefficient as well as the parity of its degree (i.e., whether the degree is even or odd). Specifically, a nonconstant polynomial of even degree and positive leading coefficient goes to $+\infty$ in both directions. A nonconstant polynomial of odd degree and positive leading coefficient goes to $+\infty$ in the positive direction and $-\infty$ in the negative direction. If the leading coefficient is negative, each limit becomes the negative of what it would have been if the leading coefficient were positive. The proof of this essentially follows from the limit statements for the function x^n.
(3) For a rational function, if the degree of the numerator is *less than* the degree of the denominator, the limit as $x \to \pm\infty$ is zero. This can be proved by factoring out the largest power of x from both numerator and denominator. If the degrees of the numerator and denominator are equal, the limit is the quotient of the leading coefficients. If the degree of the numerator is greater than the degree of the denominator, the limits are infinities, with the signs depending of the sign of the quotient of the leading coefficients and the parity of the difference of degrees. (That's a long-winded sentence, but you should already know what this means).

We now look at some of the newer functions that we have added to our discussion: trigonometric functions, exponential functions, logarithmic functions, and inverse trigonometric functions. Here are some general principles:

(1) Remember that $(\to\infty)$ times anything that is bounded below by a positive number is still $\to\infty$. Similarly, $\to 0$ times anything bounded from both above and below is still $\to 0$. Thus, for instance as $x \to \infty$, $(\sin x + \cos x)/x^2$ goes to zero because $\sin x + \cos x$ is bounded from both above and below, and $1/x^2 \to 0$. Similarly, $x^3(2 + \sin x)$ goes to $+\infty$ as $x \to \infty$ because $2 + \sin x \geq 1$ for all x. On the other hand, $x \sin x$ is wildly oscillating because as $x \to \infty$, $\sin x$ oscillates between $+1$ and -1, while $x \to \infty$, so the product oscillates wildly with every-increasing magnitude.
(2) As $x \to \infty$, $e^x \to \infty$ and as $x \to -\infty$, $e^x \to 0$. What is particularly important to remember (we will see a justification for this later in the course) is that e^x *grows faster* than any polynomial. In particular, if p is a polynomial with positive leading coefficient, then $\lim_{x\to\infty} e^x/p(x) = \infty$. Even though the $1/p(x)$ part goes to zero, the e^x part goes to ∞ much faster. *Exponential growth* is much faster than *polynomial growth*. For similar reasons, $\lim_{x\to-\infty} p(x)e^x = 0$ for all polynomials p. As $x \to -\infty$, $e^x \to 0$ so quickly that no polynomial is able to salvage it.
(3) As $x \to \infty$, $\ln(x) \to \infty$. What is notable is that $\ln(x)$ approaches ∞ slower than any polynomial or positive power function. Thus, $(\ln x)/x^r \to 0$ as $x \to \infty$ for any $r > 0$ and $(\ln x)/p(x) \to 0$ as $x \to \infty$ for any nonconstant polynomial p. Also, as $x \to 0$, $\ln x \to -\infty$, but it goes to $-\infty$ slower than x^{-r} does for any r. Thus, $x^r \ln x \to 0$ as $x \to 0$ for any $r > 0$ and $p(x) \ln x \to 0$ as $x \to 0$ for any polynomial p without constant term. Note that the existence of a constant term would send it to $\pm\infty$, where the sign of infinity depends on the sign of the constant term.

One way of thinking of these functions and their behavior at ∞ is in terms of a *hierarchy* where:

(1) $\ln x$ and polynomials in $\ln x$ are at the bottom.
(2) x and polynomials in x and power functions of x are in between.
(3) $\exp(x)$ is at the top.

As $x \to \infty$, all of these tend to ∞, but logarithmic functions are just no match for polynomial functions, which in turn are no match for exponential functions. That is why, in colloquial language, we use the term *exponential* for superfast, *polynomial* for ordinarily fast, and *logarithmic* for a snail's pace. Note that within each level, there is a considerably wide range of incomparable functions. For polynomials, for instance, the larger the degree, the faster it goes to $+\infty$. However, this rigidly hierarchical society nonetheless looks monolithically weak and primitive when compared to the infinitely superior exponentials, while together they trample on the snail-like logarithms.

Aside: logarithm as something like x^0. One way of thinking of the logarithm function is that it grows faster than constant functions but slower than x^r for any $r > 0$. Thus, in mathematically imprecise language, the logarithm function grows like x^ε where ε is an infinitesimal number – greater than 0 but smaller than any positive number.

Another way of justifying this is to note that x^r arises as the antiderivative (up to constants) of x^{r-1}, for $r \neq 0$. On the other hand, the antiderivative of x^{-1} is $\ln x$. Since $-1 + 1 = 0$, $\ln x$ is, in some sense, like x^0.

In a similar vein, e^x can be thought of as something like x^∞ in the sense that it beats out x^r for every finite r.

3. The logarithmic transformation

One of the more useful and less talked about transformations in the context of understanding functions is to study the logarithm of a function instead of the original function. Why do we do this? In some cases, we do this because the expression for $\ln \circ f$ might be nicer and more instructive than the expression for f. For instance, the Gompertz function is described as:

$$e^{\frac{a - ke^{-abt}}{b}}$$

The logarithm of this looks nicer:

$$\frac{a - ke^{-abt}}{b}$$

There is also a more conceptual reason for often looking at $\ln \circ f$. For some functions, it is the *multiplicative rate of growth* that is more relevant than the *additive rate of growth*. This is often true for population growths, which are sort-of exponential.

If we do *not* use the logarithmic transformation in these cases, the problem that arises is that a fractional change from 10 to 20 looks a lot bigger than a fraction change from 0.1 to 0.2, even though we want to accord the same visual weight to both fractional changes.

Here are some important features of the logarithmic transformation:

(1) The logarithmic transformation preserves where the function is increasing and where it is decreasing.
(2) As $f(x) \to \infty$, $\ln(f(x)) \to \infty$, and as $f(x) \to 0$, $\ln(f(x)) \to -\infty$. Thus, under the logarithmic transformation, the upper half of the y-axis gets stretched to the entire y-axis.
(3) The logarithmic transformation need *not* preserve the sense of concavity. We often say that a function f is *logarithmically* concave up/down if $\ln \circ f$ is concave up/down. It is an easy exercise to work out how the sense of concavity of $\ln \circ f$ depends on f and its derivatives.

4. Improper integrals

So far, we have looked at definite integration as the following kind of problem:

$$\int_a^b f(x)\,dx$$

where a continuous function f is integrated over a (finite) closed interval $[a, b]$. The fundamental theorem of calculus essentially states that if F if a function defined around $[a, b]$ such that $F' = f$, then:

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

We now look at some situations where we still want to integrate between limits, but something fails.

4.1. **Integration where the interval of integration goes to infinity.** Consider for instance the integral:

$$\int_1^{\infty} \frac{dx}{x^2+1}$$

First, what does this *mean*? It can be interpreted as:

$$\lim_{L \to \infty} \int_1^L \frac{dx}{x^2+1}$$

This can now be evaluated. Specifically, it is:

$$\lim_{L \to \infty} [\arctan L - \frac{\pi}{4}]$$

As $L \to \infty$, $\arctan L \to \pi/2$, so we obtain:

$$\frac{\pi}{2} - \frac{\pi}{4} = \frac{\pi}{4}$$

The same general procedure can be followed for other functions. We look at three cases:

(1) $\int_a^{\infty} f(x)\,dx := \lim_{L \to \infty} \int_a^L f(x)\,dx$. In particular, if F is an antiderivative for f, it is $\lim_{L \to \infty}[F(L) - F(a)]$. Note that since $F(a)$ is constant, this can be rewritten as $(\lim_{L \to \infty} F(L)) - F(a)$. The former limit is the horizontal asymptote value for F as $L \to \infty$.

(2) $\int_{-\infty}^{a} f(x)\,dx := \lim_{L \to -\infty} \int_L^a f(x)\,dx$. In particular, if F is an antiderivative for f, it is $\lim_{L \to -\infty}[F(a) - F(L)]$. Since $F(a)$ is constant, this is $F(a) - \lim_{L \to -\infty} F(L)$. The latter limit is the horizontal asymptote value for F as $L \to -\infty$.

(3) $\int_{-\infty}^{\infty} f(x)\,dx = \lim_{L \to -\infty} \lim_{M \to \infty} \int_L^M f(x)\,dx$. Really making proper sense of this requires us to know what a *double limit* means – and this is something that we will come to later. Intuitively, if F is an antiderivative, this is the difference between the limiting values of F at $+\infty$ and $-\infty$. Thus, for instance:

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

because the limiting value at $+\infty$ is $\pi/2$ and the limiting value at $-\infty$ is $-\pi/2$. Graphically, this is the width of the strip between the two horizontal asymptotes – that at $+\infty$ and that at $-\infty$.

4.2. **Integration where the integrand goes to infinity at an endpoint.** Consider, for instance, the integral:

$$\int_0^1 \frac{dx}{\sqrt{1-x^2}}$$

The integrand approaches $+\infty$ as $x \to 1$. To make sense of this integral, we defined it as:

$$\lim_{L \to 1} \int_0^L \frac{dx}{\sqrt{1-x^2}}$$

This simplifes to:

$$\lim_{L \to 1} [\arcsin L - \arcsin 0] = \frac{\pi}{2}$$

Note something important that is happening here. Although the integrand itself goes off to ∞ at one of the endpoints of integration, it has an antiderivative that has a finite limit at that endpoint (where it has a one-sided vertical tangent). That is the reason the definite integral is in fact well-defined and finite.

More generally:

(1) If we have $\int_a^b f(x)\,dx$ with $a < b$ where f is (possibly) undefined or not continuous at b but is defined and continuous everywhere else, the integral is $\lim_{L\to b^-} \int_a^L f(x)\,dx$. If F is an antiderivative for f, the integral is $\lim_{L\to b} F(L) - F(a)$. Note that, as the arcsin example shows, f may go to ∞ while F has a well-defined limit. In other cases, it happens that f is oscillatory but F has a well-defined limit.
(2) If we have $\int_a^b f(x)\,dx$ with $a < b$ where f is (possibly) undefined or not continuous at b but is defined and continuous everywhere else, the integral is $\lim_{L\to a^+} \int_L^b f(x)\,dx$.
(3) If we have problems at both endpoints, we need to do a double limit approaching both of them.
(4) We could have a mixed situation where one of the limits is an actual infinity and the other is a trouble endpoint. The same limiting approach works.

4.3. **If it's broken, partition first and then approach endpoints gently.** In addition to problems at the endpoints of the interval of integration, we could have problems in the *interior* of the interval of integration – points inside the interval of integration where the function is not continuous. The discontinuity may be a *jump* discontinuity, an *infinite* discontinuity, or some other kind of discontinuity. In such cases, we adopt a two-step procedure:

(1) We first partition the original interval of integration into smaller intervals such that the trouble points are all at the ends of intervals of integration.
(2) Next, for each of these smaller intervals, we write the integral as a limit, as described in the previous part.
(3) Finally, we add up the integrals on each of the intervals. This is the grand piecing together.

5. Some subtleties and aspects of improper integration

5.1. **What does it mean for the integral to infinity to be finite?** Suppose f is a continuous function defined on $[a,\infty)$ with the property that $\int_a^\infty f(x)\,dx$ is finite. What can we conclude about f.

Let F be an antiderivative for f on $[a,\infty)$ (with only a one-sided derivative match at a). Then the fact that $\int_a^\infty f(x)\,dx$ is finite is equivalent to the fact that $\lim_{x\to\infty} F(x)$ is finite. Since $f = F'$, this means that F has a horizontal asymptote. What does this imply about $\lim_{x\to\infty} f(x)$?

Think about it. If F has a horizontal asymptote, what does its derivative go to? You may be tempted to say that the derivative must go to zero. This is *sort of* true, but there are counterexamples. Specifically, what *is* true is the following: *if* $\lim_{x\to\infty} f(x)$ exists, it must be zero. However, it is possible that the limit does not exist at all. This can happen if f has very occasional spikes combined with longer lulls, i.e., if the motion of F is somewhat jerky.

5.2. **Integrating a function over all real numbers.** Suppose f is a continuous function defined on all of $\mathbb{R}$. What does it mean to say that $\int_{-\infty}^\infty f(x)\,dx$ is finite? We saw earlier that this involves a two-step limiting procedure: we use a limiting procedure for the upper end of integration going to $+\infty$, and another limiting procedure for the lower end of the integration going to $-\infty$.

The idea we are going to mention now is extremely important. The question is: can we define the integral as follows?

$$\int_{-\infty}^{\infty} f(x)\,dx \stackrel{?}{=} \lim_{a\to\infty} \int_{-a}^{a} f(x)\,dx$$

In other words, is the integral on the whole of the real line simply the limit of the integral values on intervals symmetric about the origin, i.e., intervals of the form $[-a,a]$?

First, let's understand how this definition ostensibly differs from the earlier definition offered. In the earlier definition, we had two separate variables, one going to $+\infty$, and the other going to $-\infty$, and the two variables bore no relation to another. We first took a limit involving one variable, and then, having got an answer that depends upon the other variable, we then took the limit in terms of the other variable.

In the new version, we have a *single* variable that controls both the approach to $+\infty$ and the approach to $-\infty$, and both approaches are happening in a coordinated fashion – rather than one variable going off and reaching infinity first before we start moving the other variable. The limit in the new version is thus a *much more specific limit* than the limit in the general version.

Here's what is true: *if the limit in the general version exists and is finite*, then it can be computed using the more specific version. Basically, the general definition says that the limit would exist however we did the approaches to the respective infinities. Our new definition gives one such way of making the approach.

However, it is possible that the limit does not exist as per the general definition and we still get an answer with the new definition. That answer is wrong and (for the most part) meaningless.

Let's make this concrete with a discussion of odd functions.

5.3. **Odd functions integrated over the real line.** Consider an odd function f defined over the entire real line $\mathbb{R}$. By *odd*, we mean that $f(-x) = -f(x)$ for all $x \in \mathbb{R}$. We know that integrating an odd function over any symmetric interval of the form $[-a, a]$ gives the value 0.

So, we get:

$$\lim_{a \to \infty} \int_{-a}^{a} f(x)\, dx = 0$$

So, can we conclude the following?

$$\int_{-\infty}^{\infty} f(x)\, dx \stackrel{?}{=} 0$$

No. What we can conclude is that *if* the integral on the left side is a well defined finite number, *then* it must be zero.

What's happening? Think again to *why* the integral of an odd function on a symmetric interval $[-a, a]$ is 0. The reason is that the integral on $[-a, 0]$ cancels the integral on $[0, a]$. So, we have these canceling numbers. As a becomes larger, the numbers keep canceling each other.

But here's the rub. What if the two numbers that are canceling each other off are also becoming larger and larger in magnitude, or perhaps staying roughly constant in magnitude? Then, it becomes very critical that we are moving the two markers (the ones oing to $+\infty$ and $-\infty$) in tandem – if we displace one of them even a bit, then the integral value will shift abruptly.

For instance, consider the function $f(x) := x$. This is an odd function. We know that $\int_0^a f(x)\, dx = a^2/2$. And $\int_{-a}^0 f(x)\, dx = -a^2/2$, so the total integral on $[-a, a]$ is 0. But the point is that the two integrals individually are becoming larger and larger in magnitude. The fact that they cancel out is a fortuitous consequence of the way we're moving the markers. If we moved the two markers at even slightly different rates, we would get a markedly different limit. For instance:

$$\lim_{a \to \infty} \int_{-a}^{a+1} f(x)\, dx = \lim_{a \to \infty} a + (1/2) = \infty$$

Also:

$$\lim_{a \to \infty} \int_{-a}^{a-1} f(x)\, dx = \lim_{a \to \infty} -a + (1/2) = -\infty$$

Similarly, if we consider the function $f(x) := \sin x$, integrating on symmetric intervals gives 0, but if we displaced things slightly, we would no longer get 0.

5.4. **What's a good guarantee that stuff actually can be integrated?** For *nonnegative* functions, the problems above do not arise. More generally, if $\lim_{a \to \infty} \int_{-a}^{a} |f(x)|\, dx$ is finite, then we do have:

$$\int_{-\infty}^{\infty} f(x)\, dx = \lim_{a \to \infty} \int_{-a}^{a} f(x)\, dx$$

We will explore these criteria, and explain the role of *absolute values*, in a fascinating later discussion about debts, borrowing, scams, and Ponzi-Madoff schemes.

An example of an odd function that we can immediately (not right now, but after covering some more stuff in sequences and series) see is integrable (and hence integrates to 0) is $f(x) := x/(x^2 + 1)^2$.

6. Integrability summary for power functions

6.1. Positive power functions. Positive power functions are integrable on all closed and bounded (finite) intervals, since they are continuous. They are not integrable as $x \to \infty$ or as $x \to -\infty$.

6.2. Negative power functions. Here, we break into cases. For simplicity, we consider only $(0, \infty)$, including improper integrals at the endpoints. In each case, we compute an antiderivative and study its limiting behavior at 0 and at ∞.

(1) For $r \in (-1, 0)$, the proper integral is defined and finite on any finite interval $[a, b]$ with $0 < a < b < \infty$. Also, the improper integral with lower limit 0 is defined and finite. However, an integral with upper limit $+\infty$ goes off to ∞. This is because the antiderivative starts with value 0 at 0 and goes to $+\infty$ at ∞.

(2) For $r = -1$, the proper integral is defined and finite on any finite interval $[a, b]$ with $0 < a < b < \infty$. The improper integral with lower limit 0 is $+\infty$, and the improper integral with upper limit $+\infty$ is also $+\infty$. This is because the antiderivative, which is ln, starts at $-\infty$ at 0 and goes to $+\infty$ at ∞.

(3) For $r < -1$, the proper integral is defined and finite on any finite interval $[a, b]$ with $0 < a < b < \infty$. The improper integral with lower limit 0 is $+\infty$. The improper integral with positive lower limit and upper limit $+\infty$ is finite. This is because the antiderivative starts off at $-\infty$ at 0 and goes to 0 as $x \to \infty$.

The ideas behind these come up repeatedly in other situations, as we shall see.