An Axiomatic Approach to the Integral

Leonard Gillman

American Mathematical Monthly, Volume 100, Issue 1 (Jan., 1993), 16-25.

Stable URL:
http://links.jstor.org/sici?sici=0002-9890%28199301%29100%3A1%3C16%3AAAATTI%3E2.0.CO%3B2-6

An Axiomatic Approach to the Integral

Leonard Gillman

1. THE RIEMANN INTEGRAL. Have you ever watched an engineer or physicist set up an integral? They don't mention Riemann sums nor pick an arbitrary point z_k in the kth segment; they don't even mention a partition. Instead, they draw FIGURE 1 (to use area as an example) and say: *Here's the strip at x of width dx (where dx is small). Then dA is equal to* [writing it down]

$$f(x)\,dx.$$

Then they prefix an integral sign.

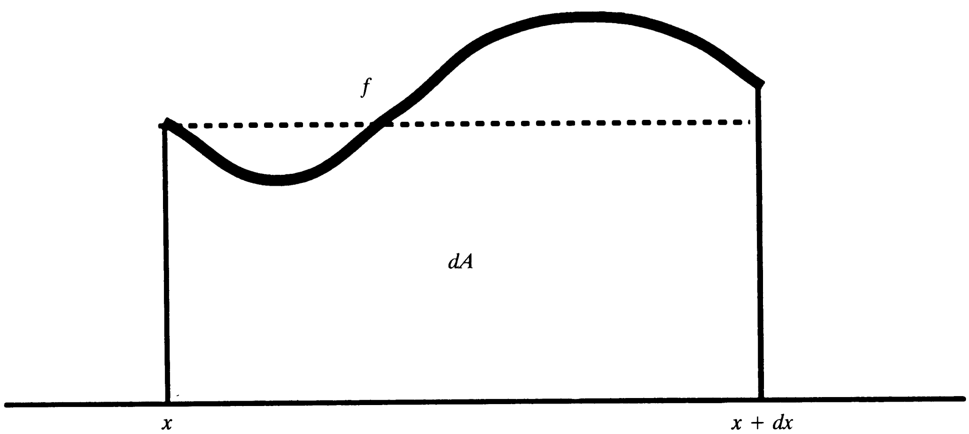

Figure 1

This invariably gets a good laugh at a lecture. Maybe instead it deserves applause for being so efficient and logical. Concentrating on a single segment permits circumventing the notational paraphernalia of all those subscripts. Calling the width dx permits skipping the summation sign and going directly to the integral sign. (Some concerned teachers will argue that it is better to remind the student of the sums on which the integral is based. Others will liken that to solving quadratics by completing the square.)

The usual way is to pick an arbitrary z_k in the kth segment of a partition, form the corresponding Riemann sum, and proclaim that since arbitrary Riemann sums approach the integral (as the norm goes to 0), so then do these arbitrary sums. But then so do particular sums. The practicing scientist picks $z_k = x_{k-1}$. (It is assumed in all this that f is continuous.)

Actually, there are applications in which one traditionally chooses a particular z_k—for example, in length of arc, where z_k is determined from the Mean-Value Theorem. (By that time the students are wondering whether it's legal.) Our scientist sticks to $z_k = x_{k-1}$ here as well. Presumably the resulting partial sums also approach "what should be" or "what the physicists intuition affirms is" the volume, or the work, etc.

2. THE DARBOUX INTEGRAL. Can this hope be replaced by a theorem? We suggest an approach based on the Darboux integral, defined as the unique number lying between all lower sums and all upper sums. Probably every calculus text bases at least one of its derivations on this condition. We will exploit it systematically. Our point of departure is the well-known properties of additivity and "betweenness":

$$\textbf{(A)} \qquad \int_a^{x+\Delta x} \varphi = \int_a^x \varphi + \int_x^{x+\Delta x} \varphi,$$

$$\textbf{(B)} \qquad \Big(\min_{[x,x+\Delta x]} \varphi\Big)\Delta x \le \int_x^{x+\Delta x} \varphi \le \Big(\max_{[x,x+\Delta x]} \varphi\Big)\Delta x$$

(where φ is continuous on an interval $[a, b]$, and $a \le x < x + \Delta x \le b$).

It should come as no surprise that the integral is the *only* function satisfying **(A)** and **(B)**; this is stated formally as Theorem 1 below. In each application, to show that the quantity of interest is an integral, we show that it has these two properties. As a result, our intuition is relieved of the responsibility of making predictions about infinite processes. (We don't even mention Riemann sums.) Instead our assumptions refer to concepts that are more real to the student, such as that the area of an enclosing region is greater than that of the enclosed region; and we put these assumptions up front. Integrals are thus derived as *theorems* rather than announced as definitions. As a byproduct, we know that the volume by either discs or shells will be the same, that the area by either rectangular or polar coordinates will be the same, and so on.

In the usual treatment of the integral, after all the elaborate preparation, there always comes that anticlimactic moment when you confront a nonRiemann sum and have to mumble that yes, that will work too but the proof is too hard for this course. (The rough-and-ready scientist skirts around this problem.) Our more general Theorem 2 covers these cases as well.

Most of all this has been done before in various degrees of thoroughness, but a review seems worth while. What I believe to be new are the improved version of the general theorem, a more natural axiom for surface area, and the observation that this axiom and those for arc length are equivalent (in the presence of additivity) to the corresponding formulas.

3. BACKGROUND

Theorem 1. *Let φ be continuous on an interval $[a, b]$, and let I_u^v be defined for $a \le u \le v \le b$. Suppose that*

$$\textbf{(A)} \qquad I_a^{x+\Delta x} = I_a^x + I_x^{x+\Delta x}$$

and

$$\textbf{(B)} \qquad \Big(\min_{[x,x+\Delta x]} \varphi\Big)\Delta x \le I_x^{x+\Delta x} \le \Big(\max_{[x,x+\Delta x]} \varphi\Big)\Delta x,$$

when $\Delta x > 0$. *Then*

$$I_a^b = \int_a^b \varphi(x)\, dx.$$

Note that **(A)** includes the condition $I_a^a = 0$ (the case $x = a$).

Proof: Consider an arbitrary partition of $[a, b]$. By hypothesis, **(B)** holds for every segment $[x, x + \Delta x]$. Hence by **(A)** (applied repeatedly), the lower and upper sums L and U satisfy $L \leq I_a^b \leq U$. Therefore I_a^b is the Darboux integral.

The earliest reference I know to the properties **(A)** and **(B)** as characteristic of the integral is Hahn and Rosenthal [**2**, 149–150], which in fact adopts them as the *definition*. I believe that Howard Levi [**4**, 60–70] is the first to recognize that they characterize the integral and then follow up by invoking them systematically in applications; this book, with its wealth of imaginative ideas, deserves to be better known. Serge Lang [**3**] also invokes this characterization. Gillman and McDowell [**1**] adopts **(A)** and **(B)** as the definition of the integral and applies them systematically to an extensive selection, including polar coordinates and multiple integrals. The two-variable version of Theorem 1 permits an effortless proof that the value of the double integral is given by each of the two iterated integrals.

This illustrates the advantage in using **(A)** and **(B)** rather than the Darboux integral itself, which is essentially the same thing: they focus attention on the underlying principles. (Recall that **(A)** and **(B)** constitute the two steps of the proof of the Fundamental Theorem of Calculus.)

We now illustrate some standard geometric applications. The letters A, V, L, S, with appropriate indices, represent area, volume, arc length, surface area.

4. APPLYING THEOREM 1. The bread-and-butter problems are (a) the area under the graph of a function f, and (b) the volume generated by revolving the graph of f about the x-axis. Most textbooks use a betweenness argument in these two problems, so for them we describe our method only in outline. In (a), we bound the area on $[x, x + \Delta x]$ by two rectangular regions on that same base, of heights min f and max f; then we use the fact that the area of a rectangle is the product of its dimensions, and the axiom about the areas of enclosed and enclosing regions. In (b), we bound the volume on $[x, x + \Delta x]$ by two right circular cylinders, of radii min f and max f; then we use the fact that the volume of a right circular cylinder is $\pi r^2 h$, and the axiom that if one of two solids encloses another, the enclosing one has the larger volume.

We proceed to examples where the bounds are less obvious.

Example 1. Length of arc. Let f and f' be continuous on $[a, b]$. To define the length of the graph of f, first note axiom **(A)**: $L_a^{x+\Delta x} = L_a^x + L_x^{x+\Delta x}$.

Now we look at FIGURE 2 and wonder what to do. The chord of the arc is a lower bound for the length; what is an upper bound? A cue is the observation that of two straight-line graphs on the same interval, the one with the larger absolute slope is the longer (FIGURE 3). We extend this principle to graphs with variable slopes. Imagine walking across a field, heading eastward but edging north as you go. Suppose I do the same, keeping due north of you at all times and, *at every instant*, heading *more* northward than you (FIGURE 4). Everyone agrees I walk farther than you. Also, south is as good as north. We summarize this in the axiom:

(L) *If* $|f_1'(t)| \leq |f_2'(t)|$ *for each t in* $[x, x + \Delta x]$, *then* $L_x^{x+\Delta x}(f_1) \leq L_x^{x+\Delta x}(f_2)$.

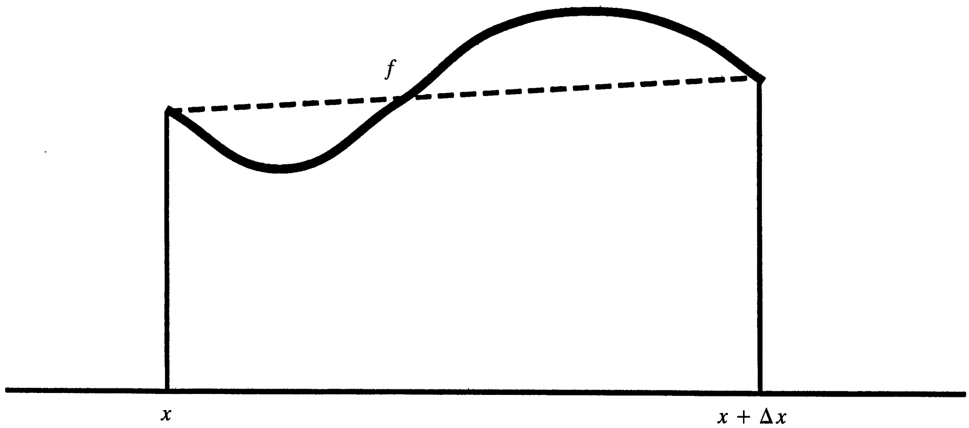

Figure 2

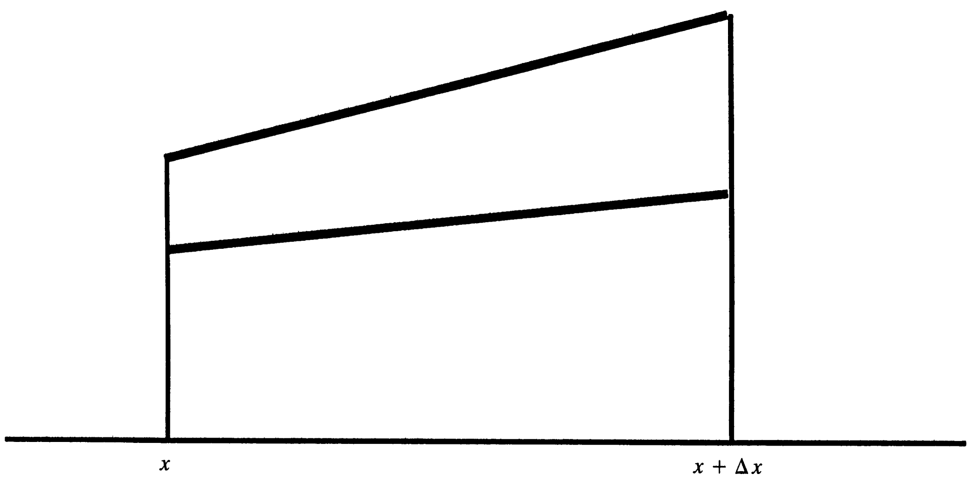

Figure 3

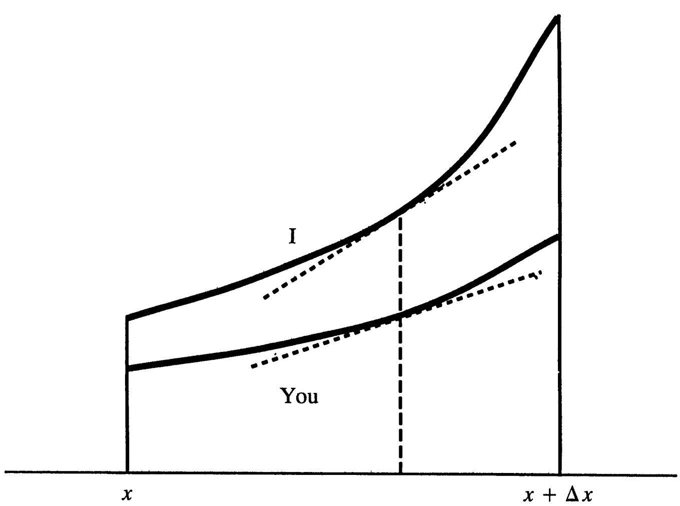

Figure 4

According to this axiom, the segment on $[x, x+\Delta x]$ with slope $\min|f'|$ is shorter than the graph, which in turn is shorter than the segment with slope $\max|f'|$. (Here and later we could state the axiom for the special case and then it would look less forbidding.) Noting that the length of a segment of slope m is $\sqrt{1+m^2}\,\Delta x$, and that $\sqrt{1+\min|f'|^2} = \min\sqrt{1+f'^2}$ and similarly for the max, we have

$$\min_{[x,x+\Delta x]} \sqrt{1+f'^2}\,\Delta x \le L_x^{x+\Delta x} \le \max_{[x,x+\Delta x]} \sqrt{1+f'^2}\,\Delta x. \tag{1}$$

This is property **(B)** with respect to the function $\varphi(x) = \sqrt{1+f'(x)^2}$. By Theorem 1,

$$L_a^b = \int_a^b \sqrt{1+f'(x)^2}\,dx. \tag{2}$$

Note that we never did use the original chordal lower bound. The weaker lower bound in (1) is adequate, and its value is expressible directly, without the Mean-Value Theorem.

Axiom **(L)** is stated and applied in **[4]** and again in **[1]**. It may be that the notion of approximating an arc by chords holds such a strong intuitive appeal that the foregoing derivation may be dispensed with. In any case, we will want these ideas in the discussion of surface area, where intuition tends to be weak.

5. SETTING UP INTEGRALS. Let us recapitulate. To set up an integral on $[a, b]$, consider a typical subinterval $[x, x+\Delta x]$. Denote by I_a^b the quantity in the application to be represented by the integral. First verify the properties **(A)** and **(B)** *within the field of application*—on the basis of knowledge of the field, or intuition, or advice from a physicist, etc. *From then on, the rest is mathematics*: Theorem 1 tell us that I_a^b is the integral.

To verify **(A)** and **(B)**: (i) Draw a picture. (ii) Verify **(A)**. This is an *axiom* in the application and turns out to be automatic in just about every case. (iii) Find a formula that holds when all the variables are constants. (iv) Use this formula, along with additional axioms for the application, to obtain the bounds to be used in **(B)**. Note that **(B)** is a *theorem* in the application.

6. THE GENERAL THEOREM. Theorem 1 proves to be inadequate in many applications, as we will shortly see. We need the following generalization.

Theorem 2. *Let φ be continuous on $[a, b]$ and let I_u^v be defined for $a \le u \le v \le b$. Suppose that for $\Delta x > 0$,*

(A) $$I_a^{x+\Delta x} = I_a^x + I_x^{x+\Delta x}$$

and

(B′) $$\alpha\,\Delta x \le I_x^{x+\Delta x} \le \beta\,\Delta x,$$

where α and β both approach $\varphi(x)$ as $\Delta x \to 0$. Then

$$I_a^b = \int_a^b \varphi(x)\,dx.$$

Proof: We show as in the Fundamental Theorem that the function

$$F(x) \equiv I_a^x$$

is an antiderivative of φ. (The notation will be for the case $\Delta x > 0$.) By **(A)**,

$$\frac{F(x+\Delta x) - F(x)}{\Delta x} = \frac{I_x^{x+\Delta x}}{\Delta x}.$$

By **(B)**,

$$\alpha \le \frac{I_x^{x+\Delta x}}{\Delta x} \le \beta.$$

Since α and β approach $\varphi(x)$ as $\Delta x \to 0$, so does the quantity between them; therefore $F'(x) = \varphi(x)$. Finally, by the other half of the Fundamental Theorem,

$$\int_a^b \varphi(x)\, dx = I_a^b - I_a^a = I_a^b.$$

A special case of this theorem appears in **[4]**. The full theorem is presented in **[1]** for both the single and double integral, but the statements are somewhat awkward and the proof for the double integral is ε's and δ's; I have since written up a proof for myself modeled after the one just given.

7. APPLYING THE GENERAL THEOREM. The procedure for setting up integrals is the same as that outlined in Section 5, except possibly in step (iv), where there may be more than one choice for α and β.

Example 2. Area between two graphs. Let f and g be continuous on $[a, b]$, with $f \ge g$. To define the area between their graphs, I would first add the constant $|\min_{[a,b]} g|$ to both functions (if necessary) to reduce to the case $g(x) \ge 0$ for all x (with the axiom that the area between the graphs is not affected by the rigid motion), then note (by an axiom of general additivity) that the area between f and g is the area under f minus the area under g. But the very simplicity of the problem, free of distractions, makes it a good one for illustrating our methods. Again, we first note axiom **(A)**: $A_a^{x+\Delta x} = A_a^x + A_x^{x+\Delta x}$.

Consider the strip on $[x, x + \Delta x]$ (Figure 5). Is it an axiom that the area between the graphs is greater than that of a rectangle on the same base with height $\min_{[x, x+\Delta x]}(f - g)$, and less than one of height $\max_{[x, x+\Delta x]}(f - g)$? Note that in this example, the region does not *contain* any rectangle on that base of height

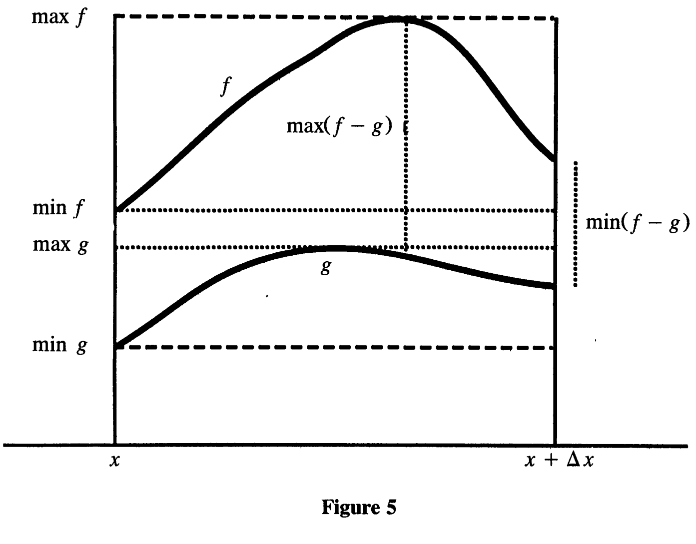

Figure 5

$\min_{[x,x+\Delta x]}(f-g)$, nor is it contained in any rectangle of height $\max_{[x,x+\Delta x]}(f-g)$. If you want that axiom nevertheless, then you obtain

$$\min_{[x,x+\Delta x]}(f-g)\,\Delta x \le A_x^{x+\Delta x} \le \max_{[x,x+\Delta x]}(f-g)\,\Delta x. \tag{3}$$

This is property (**B**), with $\varphi = f - g$, and you are finished.

However, the general theorem does not require your intuition to be that fine. You need only compare regions of which one is a subset of the other. To get an upper bound in the present example, pick the enclosing rectangle of height

$$\max_{[x,x+\Delta x]} f - \min_{[x,x+\Delta x]} g.$$

The lower bound comes with a little twist. If $\min_{[x,x+\Delta x]} f > \max_{[x,x+\Delta x]} g$, then the strip encloses a rectangle of height

$$\min_{[x,x+\Delta x]} f - \max_{[x,x+\Delta x]} g$$

(FIGURE 5); otherwise, $\min_{[x,x+\Delta x]} f - \max_{[x,x+\Delta x]} g$ is zero or negative. In either case,

$$\left(\min_{[x,x+\Delta x]} f - \max_{[x,x+\Delta x]} g\right)\Delta x \le A_x^{x+\Delta x} \le \left(\max_{[x,x+\Delta x]} f - \min_{[x,x+\Delta x]} g\right)\Delta x.$$

These inequalities are weaker than (3) and are not in the form (**B**). However, each of the two bounds (in parentheses) approaches $f(x) - g(x)$ as $\Delta x \to 0$. The inequalities are therefore of the form (**B′**), with $\varphi = f - g$. By the general theorem,

$$A_a^b = \int_a^b (f-g).$$

Example 3. Volume of a solid of revolution; shell method. Let f and g be continuous on $[a,b]$, where $a \ge 0$, and $f(x) \ge g(x)$ for all x, and revolve the region between the graphs about the *y-axis* to generate a solid. To define its volume by the shell method, we again first note axiom (**A**): $V_a^{x+\Delta x} = V_a^x + V_x^{x+\Delta x}$.

Consider the strip on $[x, x+\Delta x]$ (FIGURE 5). Its contribution to the total volume is less than that from the enclosing rectangle of height $\max_{[x,x+\Delta x]} f - \min_{[x,x+\Delta x]} g$, and is greater than that from the enclosed rectangle of height $\min_{[x,x+\Delta x]} f - \max_{[x,x+\Delta x]} g$ (perhaps a "negative" rectangle). The solid created by rotating a rectangle is the difference between two cylinders; the formula for the volume is $V = 2\pi\bar{r}h\,\Delta r$, where Δr is the difference of the two radii, $\bar{r}$ is their average, and h is the height of the rectangle. Again we invoke the axiom that the larger volume goes with the enclosing solid. Consequently,

$$2\pi\bar{x}\left(\min_{[x,x+\Delta x]} f - \max_{[x,x+\Delta x]} g\right)\Delta x \le V_x^{x+\Delta x} \le 2\pi\bar{x}\left(\max_{[x,x+\Delta x]} f - \min_{[x,x+\Delta x]} g\right)\Delta x,$$

where $\bar{x} = x + \frac{1}{2}\Delta x$, the average radius. These inequalities are in the form (**B′**), with $\varphi(x) = 2\pi x[f(x) - g(x)]$. By the general theorem,

$$V_a^b = 2\pi\int_a^b x[f(x) - g(x)]\,dx.$$

Example 4. Length of an arc defined parametrically. For an arc defined by parametric equations $x = f(t)$, $y = g(t)$, where f, f', g, and g' are continuous on an interval $a \le t \le b$, the idea is the same as in Example 1. Again we note axiom (**A**): $L_a^{t+\Delta t} = L_a^t + L_t^{t+\Delta t}$.

Consider the strip on $[t, t+\Delta t]$. Will you accept as an axiom that:

$$(\mathbf{LL}_0)\quad \begin{array}{l} \textit{If } f_1'(\tau)^2 + g_1'(\tau)^2 \le f_2'(\tau)^2 + g_2'(\tau)^2 \textit{ for each } \tau \textit{ in } [t, t+\Delta t], \\ \textit{then } L_t^{t+\Delta t}(f_1, g_1) \le L_t^{t+\Delta t}(f_2, g_2)? \end{array}$$

If so, then you obtain

$$\min_{[t, t+\Delta t]} \sqrt{f'^2 + g'^2}\, \Delta t \le L_t^{t+\Delta t} \le \max_{[t, t+\Delta t]} \sqrt{f'^2 + g'^2}\, \Delta t.$$

This is property **(B)**, with $\varphi(t) = \sqrt{f'(t)^2 + g'(t)^2}$, and you are finished.

The general theorem does not require you to be so imaginative. Instead, consider the following more pedestrian axiom:

$$(\mathbf{LL})\quad \begin{array}{l} \textit{If } |f_1'(\tau)| \le |f_2'(\tau)| \textit{ and } |g_1'(\tau)| \le |g_2'(\tau)| \textit{ for each } \tau \textit{ in } [t, t+\Delta t], \\ \textit{then } L_t^{t+\Delta t}(f_1, g_1) \le L_t^{t+\Delta t}(f_2, g_2). \end{array}$$

This leads to the inequalities

$$\sqrt{\min_{[t, t+\Delta t]} f'^2 + \min_{[t, t+\Delta t]} g'^2}\, \Delta t \le L_t^{t+\Delta t} \le \sqrt{\max_{[t, t+\Delta t]} f'^2 + \max_{[t, t+\Delta t]} g'^2}\, \Delta t,$$

which is property **(B′)**, with $\varphi(t) = \sqrt{f'(t)^2 + g'(t)^2}$. By the general theorem,

$$L_a^b = \int_a^b \sqrt{f'(t)^2 + g'(t)^2}\, dt. \tag{4}$$

Note that this agument carries over at once to the three-dimensional case $x = f(t)$, $y = g(t)$, $z = h(t)$.

Example 5. Area of a surface of revolution. Let f and f' be continuous on $[a, b]$, with f nonnegative. When the graph of f is revolved about the x-axis, it generates a surface. To define the area of such a surface, we first note axiom **(A)**: $S_a^{x+\Delta x} = S_a^x + S_x^{x+\Delta x}$.

Now we need information about some basic surface of revolution. The simplest is the one obtained by revolving a horizontal segment. If the length of the segment is h and the radius of revolution is r, then the segment sweeps out a right circular cylinder of radius r and "height" h; its area is $2\pi rh$ (found by the "slit-and-unroll" procedure).

Consider two such cylinders with parameters r_1, h_1 and r_2, h_2; obviously, if $r_1 h_1 > r_2 h_2$, then the area of the first is greater than the area of the second. We wish to find a similar comparison for revolving any two graphs. Perhaps you feel confident that:

$$(\mathbf{S}_0)\quad \begin{array}{l} \textit{If } f_1(t)\sqrt{1 + f_1'(t)^2} \le f_2(t)\sqrt{1 + f_2'(t)^2} \textit{ for each } t \textit{ in } [x, x+\Delta x], \\ \textit{then } S_x^{x+\Delta x}(f_1) \le S_x^{x+\Delta x}(f_2). \end{array}$$

If so then you obtain

$$2\pi \min_{[x, x+\Delta x]} \left(f \cdot \sqrt{1 + f'^2}\right) \Delta x \le S_x^{x+\Delta x} \le 2\pi \max_{[x, x+\Delta x]} \left(f \cdot \sqrt{1 + f'^2}\right) \Delta x.$$

This is property **(B)**, with $\varphi(t) = f(t)\sqrt{1 + f'(t)^2}$, and you are finished.

But again your intuition does not have to be that creative. Instead, consider the cruder assumption that if the graph of f_1 lies below the graph of f_2, and if

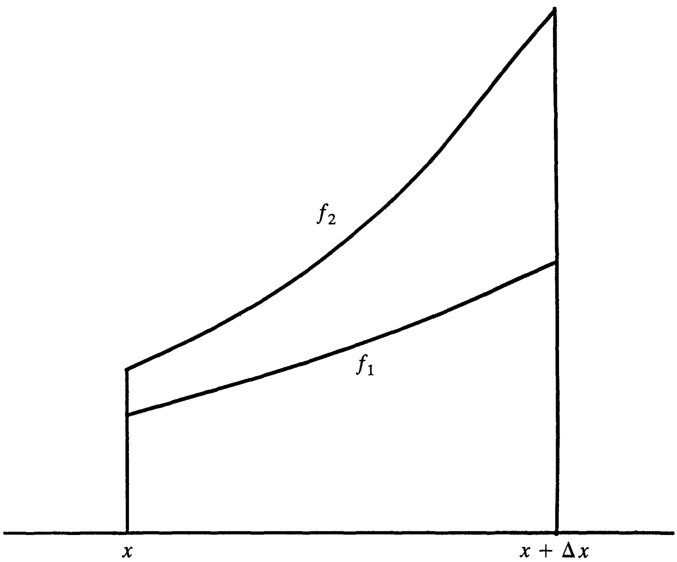

Figure 6

$|f_1'| \leq |f_2'|$ pointwise (FIGURE 6), then f_1 generates the smaller area:

(S) *If* $f_1(t) \leq f_2(t)$ *and* $|f_1'(t)| \leq |f_2'(t)|$ *for each* t *in* $[x, x+\Delta x]$, *then* $S_x^{x+\Delta x}(f_1) \leq S_x^{x+\Delta x}(f_2)$.

This axiom appears in **[1]** in a slightly different form. The present version follows along the lines suggested by the referee as being more intuitive.

We return to the given function f. Its graph is (say) as in FIGURE 7. The two accompanying line segments serve as lower and upper bounds: the lower has

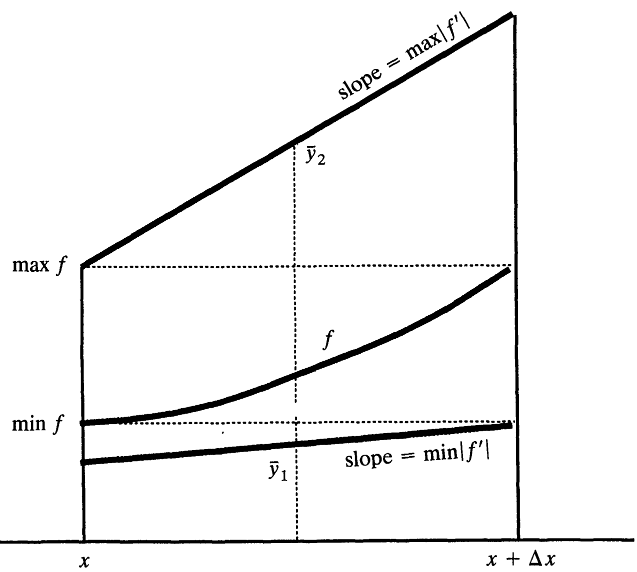

Figure 7

maximum value $\min_{[x, x+\Delta x)} f$ and slope $\min_{[x, x+\Delta x]}|f'|$; the upper has minimum value $\max_{[x, x+\Delta x]} f$ and slope $\max_{[x, x+\Delta x]}|f'|$. When revolved about the x-axis, each segment generates a frustum of a cone. According to Axiom (**S**), the areas of these frusta are lower and upper bounds (resp.) for the area generated by revolving the graph of f. Now, the area of a frustum is equal to $2\pi \bar{r} l$, where $\bar{r}$ is the average of the two extreme radii of revolution, and l is the slant height, the length of the segment being revolved. (Slit and unroll again.) Consequently,

$$2\pi \bar{y}_1\left(\min_{[x, x+\Delta x]} \sqrt{1+f'^2}\right)\Delta x \le S_x^{x+\Delta x} \le 2\pi \bar{y}_2\left(\max_{[x, x+\Delta x]} \sqrt{1+f'^2}\right)\Delta x, \quad (5)$$

where $\bar{y}_1$ and $\bar{y}_2$ are the average radii (FIGURE 6). Since $\min_{[x, x+\Delta x]} f$ and $\max_{[x, x+\Delta x]} f$ both approach $f(x)$ as $\Delta x \to 0$, $\bar{y}_1$ and $\bar{y}_2$ also approach $f(x)$. The inequalities (5) therefore assert (**B**′), with $\varphi(x) = 2\pi f(x)\sqrt{1+f'(x)^2}$. By the general theorem,

$$S_a^b = 2\pi \int_a^b f(x)\sqrt{1+f'(x)^2}\,dx. \quad (6)$$

8. EQUIVALENCE OF THE AXIOMS WITH THE FORMULAS. Are axioms (**L**), (**LL**), and (**S**) convincing? Although self-evidence cannot be legislated, it may nevertheless help to know that each one, in conjunction with (**A**), is *equivalent* to the corresponding formula, (2), (4), or (6). We have seen in each case that the axiom, together with (**A**), implies the formula. Conversely, each formula implies (**A**), by additivity of the integral; and, clearly, (2) ⇒ (**L**), (4) ⇒ (**LL**$_0$) ⇒ (**LL**), and (6) ⇒ (**S**$_0$) ⇒ (**S**).

I wish to thank the referee for an extremely careful reading of the manuscript and for several thoughtful suggestions in addition to the contribution cited in Example 5. I am also grateful to Dan Kalman for a number of helpful comments.

REFERENCES

1. Leonard Gillman and Robert H. McDowell, *Calculus*, W. W. Norton, New York, 1973.
2. Hans Hahn and Arthur Rosenthal, *Set Functions*, The University of New Mexico Press, Albuquerque, 1948.
3. Serge Lang, *A First Course in Calculus*, Addison Wesley, Reading, Mass., 1968.
4. Howard Levi, *Polynomials, Power Series, and Calculus*, Van Nostrand, Princeton, 1968.

Department of Mathematics
University of Texas
Austin, TX 78712