Integration, Average Behavior: Area of Regions in the Plane*

Lawrence Baggett

Abstract

This module contains an axiom of choice and also includes various theorems and exercises related to the usage of integrals in finding the area of regions in the plane.

It would be desirable to be able to assign to each subset S of the Cartesian plane R^2 a nonnegative real number $A(S)$ called its area. We would insist based on our intuition that (i) if S is a rectangle with sides of length L and W then the number $A(S)$ should be LW, so that this abstract notion of area would generalize our intuitively fundamental one. We would also insist that (ii) if S were the union of two disjoint parts, $S = S_1 \cup S_2$, then $A(S)$ should be $A(S_1) + A(S_2)$. (We were taught in high school plane geometry that the whole is the sum of its parts.) In fact, even if S were the union of an infinite number of disjoint parts, $S = \cup_{n=1}^{\infty} S_n$ with $S_i \cap S_j = \varnothing$ if $i \neq j$, we would insist that (iii) $A(S) = \sum_{n=1}^{\infty} A(S_n)$.

The search for such a definition of area for every subset of R^2 motivated much of modern mathematics. Whether or not such an assignment exists is intimately related to subtle questions in basic set theory, e.g., the *Axiom of Choice* and the *Continuum Hypothesis*. Most mathematical analysts assume that the Axiom of Choice holds, and as a result of that assumption, it has been shown that there can be no assignment $S \to A(S)$ satisfying the above three requirements. Conversely, if one does not assume that the Axiom of Choice holds, then it has also been shown that it is perfectly consistent to assume as a basic axiom that such an assignment $S \to A(S)$ does exist. We will not pursue these subtle points here, leaving them to a course in Set Theory or Measure Theory. However, Here's a statement of the Axiom of Choice, and we invite the reader to think about how reasonable it seems.

1:

AXIOM OF CHOICE Let $\mathcal{S}$ be a collection of sets. Then there exists a set A that contains exactly one element out of each of the sets S in $\mathcal{S}$.

The difficulty mathematicians encountered in trying to define area turned out to be involved with defining $A(S)$ for **every** subset $S \in R^2$. To avoid this difficulty, we will restrict our attention here to certain " reasonable" subsets S. Of course, we certainly want these sets to include the rectangles and all other common geometric sets.

Definition 1:

By a (open) *rectangle* we will mean a set $R = (a,b) \times (c,d)$ in R^2. That is, $R = \{(x,y) : a < x < b \text{ and } c < y < d\}$. The analogous definition of a *closed rectangle*$[a,b] \times [c,d]$ should be clear: $[a,b] \times [c,d] = \{(x,y) : a \leq x \leq b,\ c \leq y \leq d\}$.

*Version 1.2: Dec 9, 2010 4:59 pm -0600

By the *area* of a (open or closed) rectangle $R=(a,b)\times(c,d)$ or $[a,b]\times[c,d]$ we mean the number $A(R)=(b-a)(d-c)$. .

The fundamental notion behind our definition of the area of a set S is this. If an open rectangle $R=(a,b)\times(c,d)$ is a subset of S, then the area $A(S)$ surely should be greater than or equal to $A(R)=(b-a)(d-c)$. And, if S contains the disjoint union of several open rectangles, then the area of S should be greater than or equal to the sum of their areas.

We now specify precisely for which sets we will define the area. Let $[a,b]$ be a fixed closed bounded interval in R and let l and u be two continuous real-valued functions on $[a,b]$ for which $l(x)<u(x)$ for all $x\in(a,b)$.

Definition 2:

Given $[a,b], l$, and u as in the above, let S be the set of all pairs $(x,y)\in R^2$, for which $a<x<b$ and $l(x)<y<u(x)$. Then S is called an *open geometric set.* If we replace the $<$ signs with $\leq$ signs, i.e., if S is the set of all (x,y) such that $a\leq x\leq b$, and $l(x)\leq y\leq u(x)$, then S is called a *closed geometric set.* In either case, we say that S is bounded on the left and right by the vertical line segments $\{(a,y): l(a)\leq y\leq u(a)\}$ and $\{(b,y): l(b)\leq y\leq u(b)\}$, and it is bounded below by the graph of the function l and bounded above by the graph of the function u. We call the union of these four bounding curves the *boundary* of S, and denote it by C_S.

If the bounding functions u and l of a geometric set S are smooth or piecewise smooth functions, we will call S a *smooth* or *piecewise smooth* geometric set.

If S is a closed geometric set, we will indicate the corresponding open geometric set by the symbol S^0.

The symbol S^0 we have introduced for the open geometric set corresponding to a closed one is the same symbol that we have used previously for the interior of a set. Study the exercise that follows to see that the two uses of this notation agree.

Exercise 1

a. Show that rectangles, triangles, and circles are geometric sets. What in fact is the definition of a circle?
b. Find some examples of sets that are **not** geometric sets. Think about a horseshoe on its side, or a heart on its side.
c. Let f be a continuous, nonnegative function on $[a,b]$. Show that the "region" under the graph of f is a geometric set.
d. Show that the intersection of two geometric sets is a geometric set. Describe the left, right, upper, and lower boundaries of the intersection. Prove that the interior $(S_1\cap S_2)^0$ of the intersection of two geometric sets S_1 and S_2 coincides with the intersection $S_1^0\cap S_2^0$ of their two interiors.
e. Give an example to show that the union of two geometric sets need not be a geometric set.
f. Show that every closed geometric set is compact.
g. Let S be a closed geometric set. Show that the corresponding open geometric set S^0 coincides with the interior of S, i.e., the set of all points in the interior of S. HINT: Suppose $a<x<b$ and $l(x)<y<u(x)$. Begin by showing that, because both l and u are continuous, there must exist an $\varepsilon>0$ and a $\delta>0$ such that $a<x-\delta<x+\delta<b$ and $l(x)<y-\varepsilon<y+\varepsilon<u(x)$.

Now, given a geometric set S (either open or closed), that is determined by an interval $[a,b]$ and two bounding functions u and l, let $P=\{x_0<x_1<...<x_n\}$ be a partition of $[a,b]$. For each $1\leq i\leq n$, define numbers c_i and d_i as follows:

$$c_i=\underset{x_{i-1}<x<x_i}{sup} l(x),\ \text{and}\ d_i=\underset{x_{i-1}<x<x_i}{inf} u(x). \tag{1}$$

Because the functions l and u are continuous, they are necessarily bounded, so that the supremum and infimum above are real numbers. For each $1\leq i\leq n$ define R_i to be the open rectangle $(x_{i-1},x_i)\times(c_i,d_i)$.

Of course, d_i may be $< c_i$, in which case the rectangle R_i is the empty set. In any event, we see that the partition P determines a finite set of (possibly empty) rectangles $\{R_i\}$, and we denote the union of these rectangles by the symbol $\mathcal{C}_P. = \cup_{i=1}^n (x_{i-1}, x_i) \times (c_i, d_i)$.

The area of the rectangle R_i is $(x_i - x_{i-1})(d_i - c_i)$ if $c_i < d_i$ and 0 otherwise. We may write in general that $A(R_i) = (x_i - x_{i-1}) max((d_i - c_i), 0)$. Define the number A_P by

$$A_P = \sum_{i=1}^n (x_i - x_{i-1})(d_i - c_i). \tag{2}$$

Note that A_P is not exactly the sum of the areas of the rectangles determined by P because it may happen that $d_i < c_i$ for some i's, so that those terms in the sum would be negative. In any case, it is clear that A_P is less than or equal to the sum of the areas of the rectangles, and this notation simplifies matters later.

For any partition P, we have $S \supseteq \mathcal{C}_P$, so that, if $A(S)$ is to denote the area of S, we want to have

$$\begin{array}{rcl} A(S) & \geq & \sum_{i=1}^n A(R_i) \\ & = & \sum_{i=1}^n (x_i - x_{i-1}) max((d_i - c_i), 0) \\ & \geq & \sum_{i=1}^n (x_i - x_{i-1})(d_i - c_i) \\ & = & A_P. \end{array} \tag{3}$$

Definition 3:

Let S be a geometric set (either open or closed), bounded on the left by $x = a$, on the right by $x = b$, below by the graph of l, and above by the graph of u. Define the *area*$A(S)$ of S by

$$A(S) = \sup_P A_P = \sup_{P=\{x_0<x_1<...<x_n\}} \sum_{i=1}^n (x_i - x_{i-1})(d_i - c_i), \tag{4}$$

where the supremum is taken over all partitions P of $[a, b]$, and where the numbers c_i and d_i are as defined above.

Exercise 2

a. Using the notation of the preceding paragraphs, show that each rectangle R_i is a subset of the set S and that $R_i \cap R_j = \varnothing$ if $i \neq j$. It may help to draw a picture of the set S and the rectangles $\{R_i\}$. Can you draw one so that $d_i < c_i$?
b. Suppose S_1 is a geometric set and that S_2 is another geometric set that is contained in S_1. Prove that $A(S_2) \leq A(S_1)$. HINT: For each partition P, compare the two A_P's.

Exercise 3

Let T be the triangle in the plane with vertices at the three points $(0,0), (0,H)$, and $(B,0)$. Show that the area $A(T)$, as defined above, agrees with the formula $A = (1/2) BH$, where B is the base and H is the height.

The next theorem gives the connection between area (geometry) and integration (analysis). In fact, this theorem is what most calculus students think integration is all about.

Theorem 1:

Let S be a geometric set, i.e., a subset of R^2 that is determined in the above manner by a closed bounded interval $[a, b]$ and two bounding functions l and u. Then

$$A(S) = \int_a^b (u(x) - l(x))\, dx. \tag{5}$$

Proof:
Let $P = \{x_0 < x_1 < ... < x_n\}$ be a partition of $[a,b]$, and let c_i and d_i be defined as above. Let h be a step function that equals d_i on the open interval (x_{i-1}, x_i), and let k be a step function that equals c_i on the open interval (x_{i-1}, x_i). Then on each open interval (x_{i-1}, x_i) we have $h(x) \leq u(x)$ and $k(x) \geq l(x)$. Complete the definitions of h and k by defining them at the partition points so that $h(x_i) = k(x_i)$ for all i. Then we have that $h(x) - k(x) \leq u(x) - l(x)$ for all $x \in [a,b]$. Hence,

$$A_P = \sum_{i=1}^{n} (x_i - x_{i-1})(d_i - c_i) = \int_a^b (h-k) \leq \int_a^b (u-l). \quad (6)$$

Since this is true for every partition P of $[a,b]$, it follows by taking the supremum over all partitions P that

$$A(S) = \sup_P A_P \leq \int_a^b (u(x) - l(x))\, dx, \quad (7)$$

which proves half of the theorem; i.e., that $A(S) \leq \int_a^b u - l$.

To see the other inequality, let h be any step function on $[a,b]$ for which $h(x) \leq u(x)$ for all x, and let k be any step function for which $k(x) \geq l(x)$ for all x. Let $P = \{x_0 < x_1 < ... < x_n\}$ be a partition of $[a,b]$ for which both h and k are constant on the open subintervals (x_{i-1}, x_i) of P. Let $a_1, a_2, ..., a_n$ and $b_1, b_2, ..., b_n$ be the numbers such that $h(x) = a_i$ on (x_{i-1}, x_i) and $k(x) = b_i$ on (x_{i-1}, x_i). It follows, since $h(x) \leq u(x)$ for all x, that $a_i \leq d_i$. Also, it follows that $b_i \geq c_i$. Therefore,

$$\int_a^b (h-k) = \sum_{i=1}^{n} (a_i - b_i)(x_i - x_{i-1}) \leq \sum_{i=1}^{n} (x_i - x_{i-1})(d_i - c_i) = A_P \leq A(S). \quad (8)$$

Finally, let $\{h_m\}$ be a nondecreasing sequence of step functions that converges uniformly to u, and let $\{k_m\}$ be a nonincreasing sequence of step functions that converges uniformly to l. See part (d) of here[1]. Then

$$\int_a^b (u-l) = \lim_m \int_a^b (h_m - k_m) \leq A(S), \quad (9)$$

which proves the other half of the theorem.

OK! Trumpet fanfares, please!

Theorem 2:
($A = \pi r^2$.) If S is a circle in the plane having radius r, then the area $A(S)$ of S is πr^2.

Proof:
Suppose the center of the circle S is the point (h,k). This circle is a geometric set. In fact, we may describe the circle with center (h,k) and radius r as the subset S of R^2 determined by the closed bounded interval $[h-r, h+r]$ and the functions

$$u(x) = k + \sqrt{r^2 - (x-h)^2} \quad (10)$$

and

$$l(x) = k - \sqrt{r^2 - (x-h)^2}. \quad (11)$$

[1]"Integration, Average Behavior: Integrable Functions", Exercise 3 <http://cnx.org/content/m36209/latest/#fs-id1170766144622>

By the preceding theorem, we then have that

$$A(S) = \int_{h-r}^{h+r} 2\sqrt{r^2 - (x-h)^2}\, dx = \pi r^2. \tag{12}$$

We leave the verification of the last equality to the following exercise.

Exercise 4

Evaluate the integral in the above proof:

$$\int_{h-r}^{h+r} 2\sqrt{r^2 - (x-h)^2}\, dx. \tag{13}$$

Be careful to explain each step by referring to theorems and exercises in this book. It may seem like an elementary calculus exercise, but we are justifying each step here.

2:

REMARK There is another formula for the area of a geometric set that is sometimes very useful. This formula gives the area in terms of a "double integral." There is really nothing new to this formula; it simply makes use of the fact that the number (length) $u(x) - l(x)$ can be represented as the integral from $l(x)$ to $u(x)$ of the constant 1. Here's the formula:

$$A(S) = \int_a^b \left(\int_{l(x)}^{u(x)} 1\, dy \right) dx. \tag{14}$$

The next theorem is a result that justifies our definition of area by verifying that the whole is equal to the sum of its parts, something that any good definition of area should satisfy.

Theorem 3:

Let S be a closed geometric set, and suppose $S = \cup_{i=1}^n S_i$, where the sets $\{S_i\}$ are closed geometric sets for which $S_i^0 \cap S_j^0 = \varnothing$ if $i \neq j$. Then

$$A(S) = \sum_{i=1}^n A(S_i). \tag{15}$$

Proof:

Suppose S is determined by the interval $[a, b]$ and the two bounding functions l and u, and suppose S_i is determined by the interval $[a_i, b_i]$ and the two bounding functions l_i and u_i. Because $S_i \subseteq S$, it must be that the interval $[a_i, b_i]$ is contained in the interval $[a, b]$. Initially, the bounding functions l_i and u_i are defined and continuous on $[a_i, b_i]$, and we extend their domain to all of $[a, b]$ by defining $l_i(x) = u_i(x) = 0$ for all $x \in [a, b]$ that are not in $[a_i, b_i]$. The extended functions l_i and u_i may not be continuous on all of $[a, b]$, but they are still integrable on $[a, b]$. (Why?) Notice that we now have the formula

$$A(S_i) = \int_{a_i}^{b_i} (u_i(x) - l_i(x))\, dx = \int_a^b (u_i(x) - l_i(x))\, dx. \tag{16}$$

Next, fix an x in the open interval (a, b). We must have that the vertical intervals $(l_i(x), u_i(x))$ and $(l_j(x), u_j(x))$ are disjoint if $i \neq j$. Otherwise, there would exist a point y in both intervals, and this would mean that the point (x, y) would belong to both S_i^0 and S_j^0, which is impossible by hypothesis. Therefore, for each $x \in (a, b)$, the intervals $\{(l_i()x), u_i(x)\}$ are pairwise disjoint open intervals, and they are all contained in the interval $(l(x), u(x))$, because the S_i's are subsets of S. Hence, the sum of the lengths of the open intervals $\{(l_i(x), u_i(x))\}$ is less than or equal to the length of $(l(x), u(x))$. Also, for any point y in the closed interval $[l(x), u(x)]$, the point (x, y)

must belong to one of the S_i's, implying that y is in the closed interval $[l_i(x), u_i(x)]$ for some i. But this means that the sum of the lengths of the closed intervals $[l_i(x), u_i(x)]$ is greater than or equal to the length of the interval $[l(x), u(x)]$. Since open intervals and closed intervals have the same length, we then see that $(u(x) - l(x) = \sum_{i=1}^{n} (u_i(x) - l_i(x))$.

We now have the following calculation:

$$\begin{array}{rcl} \sum_{i=1}^{n} A(S_i) & = & \sum_{i=1}^{n} \int_{a_i}^{b_i} (u_i(x) - l_i(x))\, dx \\ & = & \sum_{i=1}^{n} \int_{a}^{b} (u_i(x) - l_i(x))\, dx \\ & = & \int_{a}^{b} \sum_{i=1}^{n} (u_i(x) - l_i(x))\, dx \\ & = & \int_{a}^{b} (u(x) - l(x))\, dx \\ & = & A(S), \end{array} \quad (17)$$

which completes the proof.