3. The concept of integral and the fundamental theorem

T: As we did before, let me to start with a problem. It is more challenge than the one we discussed about slope and helpfully is more interesting.

Discussion Question 3.1. *Find the area bounded by x axis, $x = 4$ and $y = x^2$ as shown in Figure 10.*

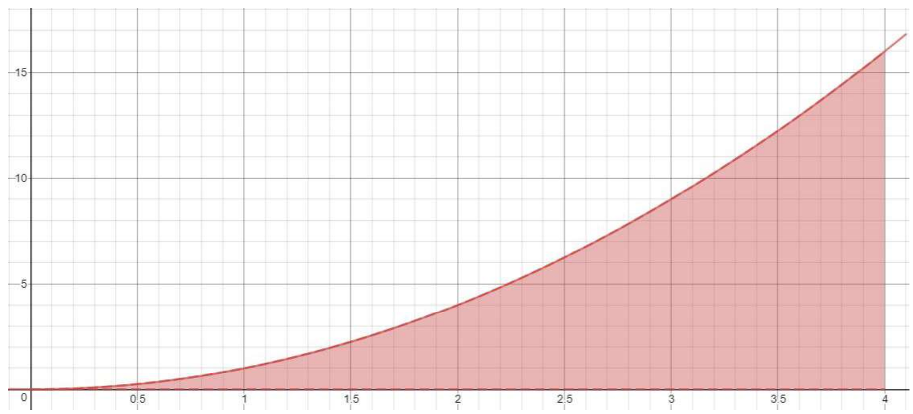

FIG 10. *The area bounded by $y = x^2, y = 0, x = 0, x = 4$*

S: I have to admit that I can only work on a shape with straight boundaries, basically polygons. Do we actually have a formula for that shape. Should we start with some sort of approximation like we did with the slope. But seems to me that there is no straightforward method even for approximation.

T: It is indeed less obvious on how to do approximation, which need to be done in a way that we are able to control the approximation error. One way is to slice the shape into n parts, as shown in Figure 12, whose areas are denoted by $s_1, s_2, \ldots, s_n$, by $n + 1$ vertical lines

$$x = \frac{4i}{n} := x_i, \qquad i = 0, 1, ..., n.$$

The area S of the shape is the summation of the areas of the small pieces of the partition:

$$S = s_1 + s_2 + \cdots + s_n$$

S: Each of piece looks more or less like trapezoid. Should we estimate s_i by the area of the associated trapezoid?

T: Quite natural idea, isn't it? As we refine the partition of the shape by increasing n, each small part will be more like trapezoid, and the estimation by the areas of those trapezoids indeed should work. I have some other reasons to use rectangles as shown in Figure 12. Let us estimate a_i by the area of

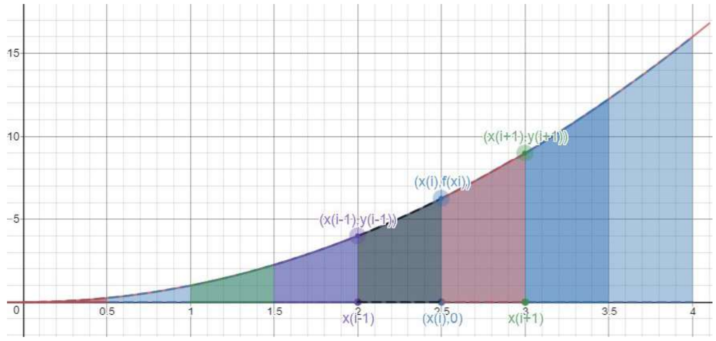

FIG 11. *Estimate the area by slicing it into small pieces*

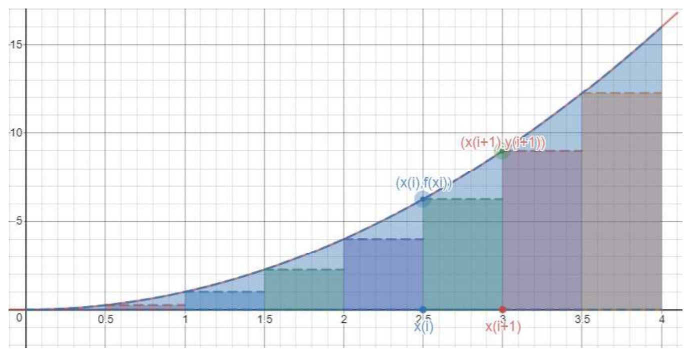

FIG 12. *Estimate the area of small pieces by rectangles-I*

the associated rectangle. Notice that the dimension for the $i-th$ rectangle $(i = 1, 2, ..n)$ is

$$base = x_i - x_{i-1} = \frac{4}{n}, \qquad height = f(x_{i-1}) = x_{i-1}^2 = (\frac{4(i-1)}{n})^2.$$

So the area of $i-th$ rectangle is

$$a_i = f(x_{i-1})(x_i - x_{i-1}) = \frac{4^3}{n^3} \times (i-1)^2.$$

Adding them to get the estimation of the area of the whole shape

$$\begin{aligned} S &\approx \sum_{1\le i\le n} f(x_{i-1})(x_i - x_{i-1}) = \frac{4^3}{n^3} \times \sum_{1\le i\le n}(i-1)^2 \\ &= \frac{4^3}{3} - \frac{4^3}{2n} + \frac{4^3}{6n^2} =: A_n \end{aligned} \tag{3.1}$$

For the last step, I use an algebraic identity for any integer m

$$\sum_{1\le i\le m} i^2 = \frac{m^3}{3} + \frac{m^2}{2} + \frac{m}{6} \tag{3.2}$$

S: The estimation is much more complicated than what we did on slope. We directly estimate the slope by a single quantity. Here we have to divide the area into small pieces and estimate each individual. The results can then be used to get an estimation of full area.

T: Exactly. It is part of reason that I prefer to estimate by using rectangles rather than trapezoid. The calculation is already hard for using rectangles and using trapezoid will add extra complexity. what can we say about A_n in (3.1) if n becomes very large?

S: It should approach to $\frac{4^3}{3}$ as the other two terms in the definition (3.1) of A decreases to 0.

T: Right! notice that A_n is increasing and therefore indeed takes its supreme $\frac{4^3}{3}$ as the limit:

$$\lim_{n\to\infty} A_n = \lim_{n\to\infty}(\frac{4^3}{3} - \frac{4^3}{2n} + \frac{4^3}{6n^2}) = \frac{4^3}{3}. \tag{3.3}$$

S: Are you saying that S is equal to $\frac{4^3}{3}$? It is still not clear to me why S has to be exactly equal to $\frac{4^3}{3}$. We add many terms and each term has an estimation error. How can we measure the overall impact of the estimation error?

T: Your concern is legitimate. Do you agree that at least $S > A_n$ for ANY n and therefore S is a upper bound for the sequence $(A_n)_{n\geq 1}$ and

$$S \geq \frac{4^3}{3}. \tag{3.4}$$

S: That is clear to me since each $s_i > a_i$.

T: OK. As we did on slope problem, we now estimate the area from the other direction. Instead using small rectangle, let's pick up larger rectangles as shown in Figure 13

S: So we can show $S \leq \frac{4^3}{3}$ and then conclude that S has to be **equal** to $\frac{4^3}{3}$?

T: Exactly. We play the same sandwich trick as we did with slope. The dimension for the $i-th$ rectangle $(i = 1, 2, ..n)$ changes to

$$base = x_i - x_{i-1} = \frac{4}{n}, \qquad height = f(x_i) = x_i^2 = (\frac{4i}{n})^2.$$

So the area of $i-th$ rectangle is

$$b_i = \frac{4}{n} \times \frac{4^2 i^2}{n^2} = \frac{4^3}{n^3} \times i^2.$$

Adding them to get the estimation of the area of the whole shape

$$S \approx \frac{4^3}{n^3} \times \sum_{1\leq i\leq n} i^2 = \frac{4^3}{3} + \frac{4^3}{2n} + \frac{4^3}{6n^2} =: B_n.$$

It is clear that $S \leq B_n$ since $s_i < b_i$ for all i and therefore S is a lower bound for the sequence $(B_n)_{n\geq 1}$. It is easy to see that B_n is decreasing and converges

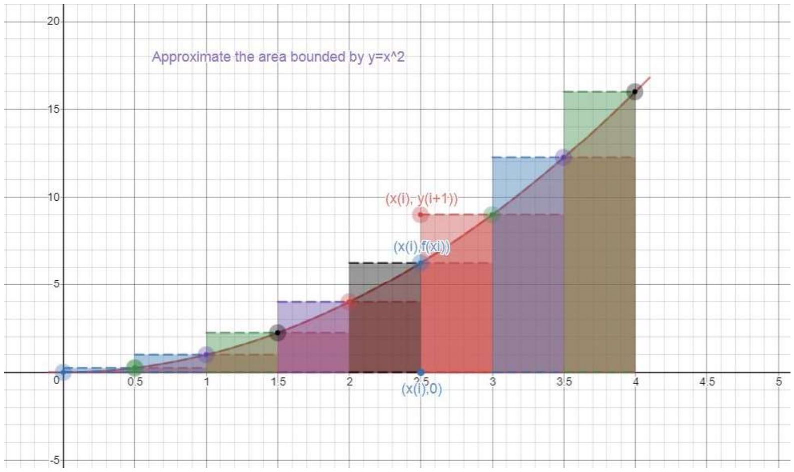

FIG 13. *Estimate the area of small pieces by rectangles-II*

to inf $B = \frac{4^3}{3}$:

$$\lim_{n\to\infty} B_n = \frac{4^3}{3}, \tag{3.5}$$

which implies

$$S \le \frac{4^3}{3}. \tag{3.6}$$

Combining with (3.4), we obtain $S = \frac{4^3}{3}$.

S: Cool! So both approximation methods work. I mean we can use small rectangles and large rectangles to approximate, they all approach to the final value $S = \frac{4^3}{3}$.

T: Right. The approximation method is not unique, but they end up with same limit. Let's rewrite two approximation summations as follows with $f(x) = x^2$, denote $\Delta x_i := x_i - x_{i-1}$,

$$A_n = \sum_{1\le i\le n} f(x_{i-1})\Delta x_i, \tag{3.7}$$

$$B_n = \sum_{1\le i\le n} f(x_i)\Delta x_i, \tag{3.8}$$

We have show that A_n and B_n both approach to the area S as n increases. Notice that x_{i-1} and x_i are two end point of the interval $[x_{i-1}, x_i]$. If we select another point $\hat{c}_i$ over the interval $[x_{i-1}, x_i]$, using the rectangle with height $f(c_i)$ to estimate the area s_i of the associated small piece as shown in Figure 14, then we get another approximation:

$$D_n := \sum_{1\le i\le n} f(c_i)\Delta x_i \tag{3.9}$$

FIG 14. *Estimate the area of small pieces by rectangles-III*

Does D_n approach to S as n increases too?

S: Why not?! It is clear that $f(c_i)$ is between $f(x_{i-1})$ and $f(x_i)$, so D_n is between A_n and B_n. If both A_n and B_n approach to S, D_n should also approach to S.

T: Excellent! Now you can see the reason why I choose rectangles, rather than trapezoids, to estimate S. By using rectangles, I have two special methods, one takes the lowest height and the other take highest height, they both converge to the desired value S. Since $f(x_{i-1})$ is the minimum value and $f(x_i)$ is the maximum value over $[x_{i-1}, x_i]$, A_n is called a **lower Darboux sum**. and B_n is called a **upper Darboux sum**.

S: So there is no difference at all? Intuitively, I do feel that trapezoid method is better.

T: There is no difference after we take limit, i.e. they all approach to S. But trapezoid method goes much faster to real value and a computer will certainly sense the difference [21]. It is quite surprising that the same argument works in a general setting where the upper boundary of the area is given by a general function $y = f(x)$ as shown in Figurehave 15.

To extend lower Darboux sum and upper Darboux sum to a general function $f(x)$ over $[a, b]$, we assume it meet the following requirement:

Assumption 3.1. *$f(x)$ has the maximum value and minimum value on any sub interval $[\alpha, \beta] \subset [a, b]$, i.e. there exist c and d in $[\alpha, \beta]$ such that*

$$f(c) \leq f(x) \leq f(d), \quad \forall x \in [\alpha, \beta]$$

Remark 7. *A general function $f(x)$ might not be able to reach its maximum or minimum as shown in Example 2.2 where it dose not have the maximum value*

[21] As an exercise, you can actually calculated the values for first 10 items of A_n, D_n, C_n and compare how close they are to S.

over $[-2, 4]$[22]*. We shall show that so called "continuous" function always meets above assumption in Theorem 4.5.*

Follow the idea in above example, $[a, b]$ is divided into n sub intervals by points

$$x_i = a + i\frac{b-a}{n}, \quad i = 0, 1, \ldots, n$$

and $\Delta x_i = x_i - x_{i-1}$ for $1 \le i \le n$. We can make the lower Darboux sum and the upper Darboux sum similarly as in (3.7) and (3.8) [23].

$$A_n = \sum_{1 \le i \le n} f(c_i)\Delta x_i, \tag{3.10}$$

$$B_n = \sum_{1 \le i \le n} f(d_i)\Delta x_i, \tag{3.11}$$

where $f(x)$ reaches the minimum value and the maximum value at c_i and d_i respectively over the interval $[x_{i-1}, x_i]$ for $i = 1, \ldots, n$.

Definition 3.1. *Let f satisfy the condition in Assumption 3.1. If both $\lim_{n\to\infty} A_n$ and $\lim_{n\to\infty} B_n$ exist and are equal, then the common limit is called the* **integral** *of $f(x)$ over $[a, b]$ and denoted by:*

$$\int_a^b f(x)dx := \lim_{n\to\infty} A_n = \lim_{n\to\infty} B_n \tag{3.12}$$

The $f(x)$ is called as the **integrand** *of the integral.*

The result of the area example can be expressed as

$$\int_0^4 x^2 dx = \frac{4^3}{3} \tag{3.13}$$

S: So the integral is just the area of the shape?

T: Yes if $f(x) \ge 0$ and it represents the upper boundary of the shape as we discussed before, see Figure 15. But integral can represent many other quantities. What $\sum_{1 \le i \le n} f(c_i)\Delta x_i$ represents if x represents time and $f(x)$ represent the velocity someone travels along a straight line at time x?

S: Each $f(c_i)\Delta x_i$ approximates the distance traveled during the time period $[x_{i-1}, x_i]$[24], the sum estimates the total distance traveled over the time period $[a, b]$. looks that the integral represents the exact distance traveled.

T: Exactly. Recall that we can get instantaneous velocity function from a distance function by calculating derivative. Using integral, we can derive the

[22] The supremum of f is 4, which is not reached by function at any point $x \in [-2, 4]$ and therefore can not be treated as the maximum value.

[23] The partition of the interval $[a, b]$ do not need to be evenly divided by above x_i, but can be defined in a more general setting.

[24] Although the speed is not constant over $[x_{i-1}, x_i]$, the instantaneous speed $f(c_i)$ is used as approximation for the speed over that time period.

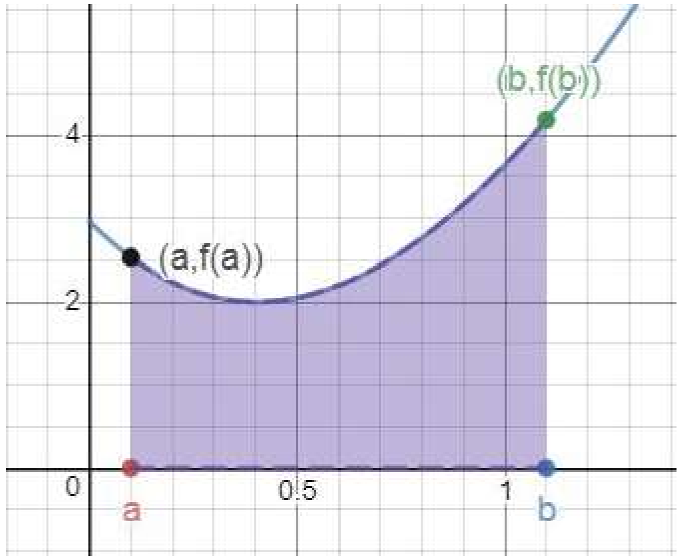

FIG 15. *Integral for a general function $f(x)$ over the interval $[a, b]$*

distance function by instantaneous velocity function. At this moment, you can simply consider the integral as the area under the curve for a intuitive understanding. Here are a few obvious properties when we interpret integral as areas.

1. If $f(x) \equiv C$ is a constant function, then

$$\int_a^b f(x)dx = C \times (b-a) \tag{3.14}$$

which is consistent to what we expect for the area of a rectangle.

2. If $f(x) \le g(x)$ over $[a, b]$, then

$$\int_a^b f(x)dx \le \int_a^b g(x)dx \tag{3.15}$$

In particular, if $m \le f(x) \le M$, then

$$m(b-a) \le \int_a^b f(x)dx \le M(b-a) \tag{3.16}$$

3. For $a < c < b$, we have

$$\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx \tag{3.17}$$

S: For a simple function $f(x) = x^2$, it takes us quite lot of effort to get the solution (3.13). I wonder how much we can do for the integral with more complicated integrand $f(x)$.

T: It would be very challenging if we have to go through the whole estimation process as we did for $f(x) = x^2$ over $[0, 4]$. Fortunately, we have so-called the fundamental theorem, that provides an effective way to calculate a integral if the integrand $f(x)$ is "continuous". We are going to investigate continuous function in Section 4 and show that

Theorem 3.2. *The integral of continuous function $f(x)$ exits. More specifically, if $f(x)$ is continuous over $[a, b]$, then f meets the condition in Assumption 3.1, and the lower Darboux sum and upper Darboux sum converges to the same limit, which is denoted by $\int_a^b f(x)dx$.*

We have

Theorem 3.3. The Fundamental Theorem. *For a continuous function $f(x)$ over $[a, b]$, if there exists another function, denoted by $F(x)$, such that $F'(x) = f(x)$, then*

$$\int_a^b f(x)dx = F(b) - F(a) \tag{3.18}$$

Remark 8. *Since $f(x)$ is the derivative function of $F(x)$, $F(x)$ is usually called an* **anti-derivative** *of $f(x)$. Notice that if $F(x)$ is an anti-derivative, so is $F(x) + C$. As such, Anti-derivative is unique up to a constant by theorem 2.3 (4).*

S: So the question is how to find anti-derivative function.

T: Like in dealing with derivative, we need

Task 3.1. *Develop tools to calculate anti-derivative functions for general functions.*

Let us now focus on the polynomial $f(x)$ defined in 2.3. First of all, any idea for the anti-derivative for $f(x) = x^2$?

S: So we need $F'(x) = x^2$. We know that $(x^3)' = 3x^2$ which is close to x^2 except an constant 3. If assuming $F(x) = cx^3$ with some constant c, then $x^2 = F'(x) = 3cx^2$, so $c = \frac{1}{3}$ and $F(x) = \frac{1}{3}x^2$.

T: If applying the fundamental theorem, we have

$$\int_0^4 x^2 dx = \frac{1}{3}4^3 - \frac{1}{3}0^3 = \frac{1}{3}4^3$$

and recover the value that take us a lot of efforts to get!

S: I start to sense how powerful calculus can be.

T: It is just a start. Your arguments apply for $f(x) = x^n$ to find $F(x) = \frac{1}{n+1}x^{n+1}$. Notice that if $F(x)$ and $G(x)$ are anti-derivatives for $f(x)$ and $g(x)$, then $aF(x) + bG(g)$ is the anti-derivative for $af(x) + bg(x)$. Therefore, we can find an anti-derivative for the polynomial $f(x)$.

$$F(x) = \frac{a_n}{n+1}x^{n+1} + \cdots + \frac{a_1 x^2}{2} + a_0 x \tag{3.19}$$

S: That is really amazing. it means that we can find the area bounded by any polynomial.

T: Calculus actually can calculate areas and volumes of geometric objects bounded by much more sophisticated boundaries.

S: It is like playing a magic with the fundamental theorem and I am curious why it is true.

T: It is not easy, but we are not far away from prove it. Let us stop here since we already cover enough in this section.