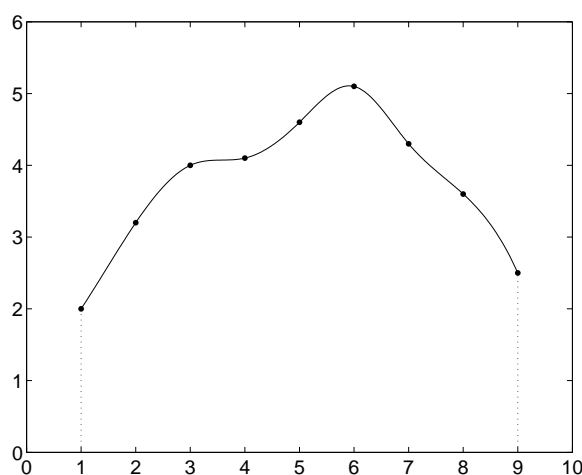


Numerical integration of a function known only through data points

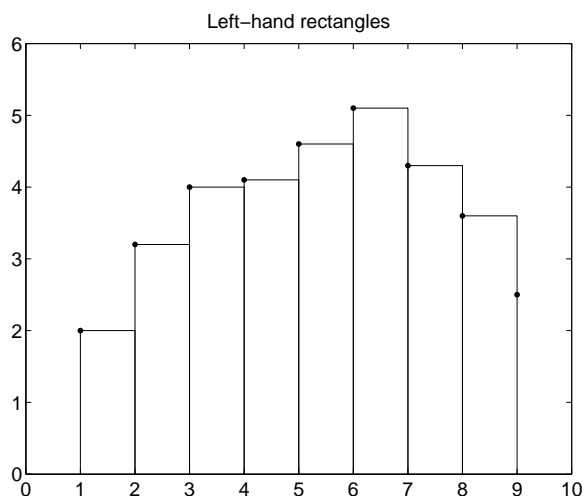
Suppose you are working on a project to determine the total amount of some quantity based on measurements of a rate. For example, you might measure the rate of flow of water at certain times and use these to determine the total amount of water that flowed. Or, you might record the speed of an airplane at certain times and try to figure out the total distance traveled. In cases such as this, the total quantity is given by the integral of the rate, and we can think of it as the area under a graph of the rate as a function of time.

Suppose you collect data points on the rate at times 1, 2, 3, ..., 9. These are displayed in the table at the right; the x values are the times, and the y values are the rates you measure. You believe that these data are points from a nice smooth function, but you only have the data points, not the function itself. Even so, you could imagine graphing the data points and then drawing smooth curves through them to define a function f , as in the figure below. The total quantity that you are trying to determine is equal to the area under the function f . The question is: how do you use the data points to approximate the area under f without knowing anything more about f than the data points you have?

x	y
1	2
2	3.2
3	4
4	4.1
5	4.6
6	5.1
7	4.3
8	3.6
9	2.5



One way to approximate the area under the graph is to approximate the region under the graph by rectangles and add up the areas of the rectangles. Using the data points as the upper left corners of the rectangles leads to these “left-hand” rectangles:

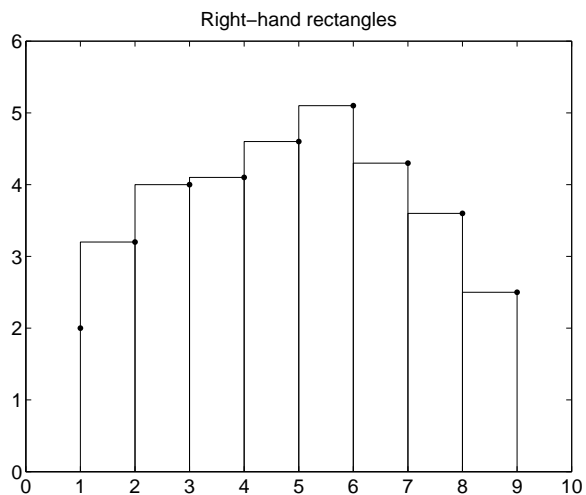


It's easy to add up the areas of the rectangles by hand, since they all have width 1 (notice that the last data point is not used):

$$L = 1 \cdot 2 + 1 \cdot 3.2 + 1 \cdot 4 + \cdots + 1 \cdot 3.6 = 30.9$$

On the other hand, the tops of the rectangles don't look like a smooth function through the data points, so the area under those horizontal lines won't be exactly the area under the function we have in mind.

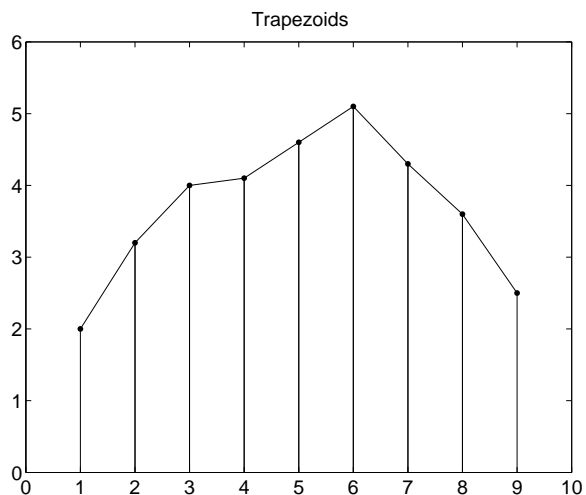
You can just as well use the data points to define the upper **right** corners of the rectangles. The graph looks slightly different:



The total area of all the rectangles is different too (here we don't use the first data point):

$$R = 1 \cdot 3.2 + 1 \cdot 4 + \cdots + 1 \cdot 3.6 + 1 \cdot 2.5 = 31.4$$

If we simply connected the data points with straight lines, the "function" that would give you looks a lot more like the function we have in mind:



The regions below the straight lines here are trapezoids, and it's not hard to calculate and add up the areas of the trapezoids, because the area of a trapezoid is the width times the average of the heights on the left and on the right, so the total area of the trapezoids is:

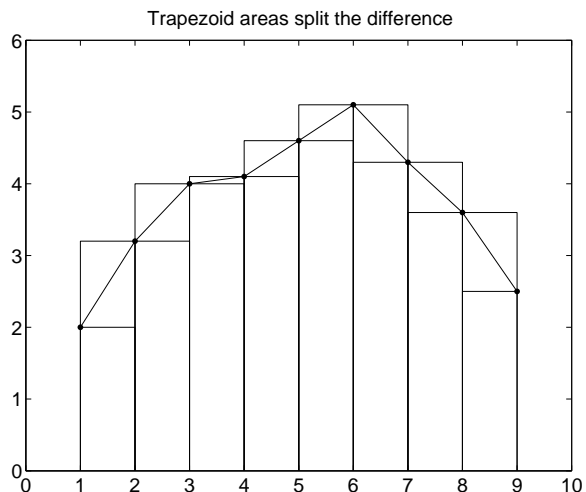
$$T = 1 \cdot \frac{2 + 3.2}{2} + 1 \cdot \frac{3.2 + 4}{2} + \cdots + 1 \cdot \frac{3.6 + 2.5}{2} = 31.15$$

It happens that this is the average of the left-rectangle value L and the right-rectangle value R :

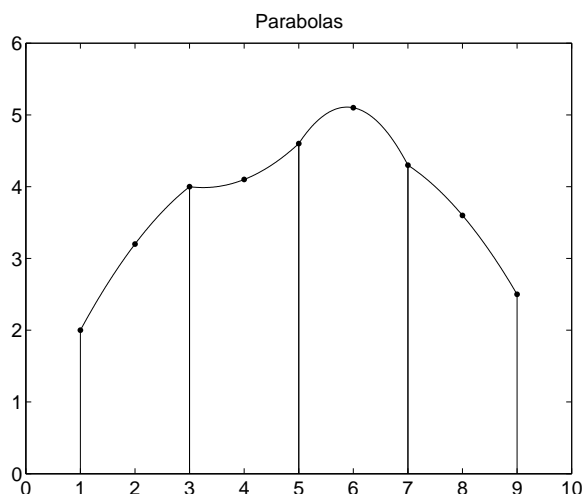
$$T = \frac{L + R}{2} = \frac{30.9 + 31.4}{2} = 31.15$$

You can see this by mentally adding together the expressions for L and R above and dividing by 2.

You can also see it by drawing left-hand rectangles and right-hand rectangles together with the trapezoids. The area of each trapezoid splits the difference (or averages) between the areas of the left-hand and right-hand rectangles.



In many calculus classes, one learns how to calculate the area below a parabola exactly. We can use this idea to get a better approximation of the area. First, imagine connecting data points 1, 2, 3 by a parabola through the points, and think about the area below that parabola. Then connect data points 3, 4, 5 with another parabola, and so on. This results in a smoother version of a function through the data points:



Fortunately, after some work with calculus, one can get a nice formula for the area under each parabola. The area under the first parabola is:

$$A = 1 \cdot \frac{2 + 4 \cdot 3.2 + 4}{3}$$

and the area under the four parabolas is:

$$P = 1 \cdot \frac{2 + 4 \cdot 3.2 + 4}{3} + 1 \cdot \frac{4 + 4 \cdot 4.1 + 4.6}{3} + \dots = 31.4333$$

General formulas

In general, we have a table of data like the one shown at the right. Suppose the x values are evenly spaced. We can calculate the areas of left-hand rectangles, right-hand rectangles, trapezoids, area under one parabola, and area under all parabolas using these formulas:

x	y
x_1	y_1
x_2	y_2
x_3	y_3
x_4	y_4
\vdots	\vdots
x_n	y_n

$$L = (x_2 - x_1)(y_1 + y_2 + \cdots + y_{n-1})$$

$$R = (x_2 - x_1)(y_2 + y_3 + \cdots + y_n)$$

$$T = (x_2 - x_1)(y_1 + 2y_2 + 2y_3 + \cdots + 2y_{n-1} + y_n)/2$$

$$A = (x_2 - x_1)(y_1 + 4y_2 + y_3)/3$$

$$P = (x_2 - x_1)(y_1 + 4y_2 + 2y_3 + 4y_4 + \cdots + 2y_{n-2} + 4y_{n-1} + y_n)/3$$

The coefficients 1, 2, 2, \dots , 2, 1 in the trapezoid calculation come from the fact that each y value besides the first and the last comes up twice in the calculation. The coefficients 1, 4, 2, 4, 2, \dots , 2, 4, 1 in the parabola calculation come up because of the coefficient of 4 on the middle y value in each parabola and the fact that y values that are common to two parabolas come up twice in the calculation.

An easy way to keep track of the coefficients is to organize them into a table like the one below. Notice that the left-hand rectangles use all the data points except the last one. The right-hand rectangles skip the first data point but use all the rest. The trapezoids use all the data points, but give larger weight to the middle data points, since they have coefficients of 2. The parabola method uses all the data points, but weights the middle data points differently, and so is slightly more complicated.

x	y	Left	Right	Trapezoid	Parabola
x_1	y_1	1	0	1	1
x_2	y_2	1	1	2	4
x_3	y_3	1	1	2	2
x_4	y_4	1	1	2	4
x_5	y_5	1	1	2	2
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
x_{n-2}	y_{n-2}	1	1	2	2
x_{n-1}	y_{n-1}	1	1	2	4
x_n	y_n	0	1	1	1

Calculating areas on the TI-83/TI-84

It is easy to put the data points into a graphing calculator and get the calculator to compute the areas of the rectangles, trapezoids, or parabolas. Use STAT, EDIT, and enter the data under lists L1 and L2, then the coefficients under L3, L4, L5, and L6. For our example, it looks like this:

L1	L2	L3	L4	L5	L6
1	2	1	0	1	1
2	3.2	1	1	2	4
3	4	1	1	2	2
4	4.1	1	1	2	4
5	4.6	1	1	2	2
6	5.1	1	1	2	4
7	4.3	1	1	2	2
8	3.6	1	1	2	4
9	2.5	0	1	1	1

You need to have an odd number of data points in order to use the areas below parabolas; this is the only way the 1, 4, 2, 4, . . . , 2, 4, 1 pattern will work out.

You can have the calculator add things up with the commands below. To type them in, do LIST, MATH, sum(, then LIST L2, *, etc.

Left: $\text{sum}(L2 * L3) * (x2 - x1)$

Right: $\text{sum}(L2 * L4) * (x2 - x1)$

Trapezoids: $\text{sum}(L2 * L5) * (x2 - x1) / 2$

Parabolas: $\text{sum}(L2 * L6) * (x2 - x1) / 3$

You need to calculate the difference $x2 - x1$ by hand, but it's not hard, because the spacing between the x values of the data points is always the same. Of course, you don't need to type in lists L3, L4, and L5 if you are just going to use the area under parabolas. You can also use Excel to do these calculations.

Accuracy of the calculations

Finally, a note about the accuracy of these calculations. When the function f is constant (a rare case!), the left-hand and right-hand rectangles give the exact area. When the function f is constant or linear, the trapezoids give the exact area. But when the function f is constant, linear, quadratic or even cubic, the parabola calculation gives the exact area. So the parabola calculation is exact for a wider collection of functions, and this is the main reason to recommend using it when the function you have in mind really is smooth, even though it is slightly more complicated. Using parabolas to find the area below a function is called *Simpson's method*.

Cubic splines

You might wonder where the smooth function in the very first graph came from. It is a cubic spline fit to the data points using the Matlab `spline` function. Using this function might give an even better approximation of the area under the actual function f , but you need to have access to a method for determining the best cubic spline. The area under the cubic spline is approximately 31.3878.