Approximate integration

(Section 11.1 of CLP textbook)

The goal is to find better approximations of $\int_a^b f(x)\,dx$ beyond the right & left Riemann sums.

We'll work with a specific example: $\int_1^7 \log(x)\,dx$.

First, let's compute explicitly

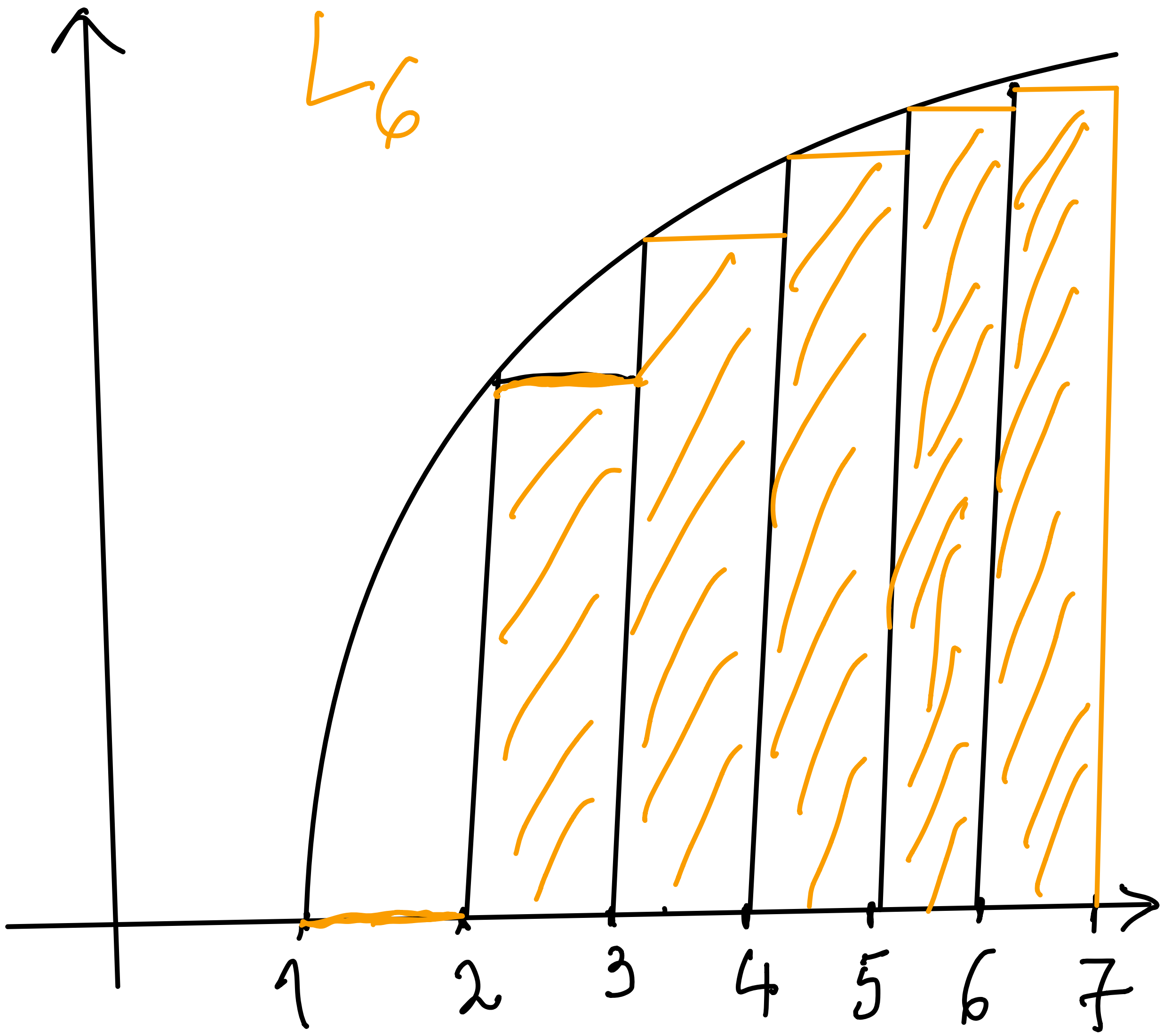
$$\int_1^7 \log(x)\,dx = \left(x\log(x) - x\right)\Big|_1^7$$

$$= \left(7\log 7 - 7\right) - \left(-1\right) = 7\log 7 - 6$$

With the use of a calculator,

$$\int_1^7 \log(x)\,dx = 7\log(7) - 6 = 7.62137\ldots$$

However, let's see first what would be the approximations provided by left, respectively right Riemann sums with 6 subintervals (i.e. $n=6$).



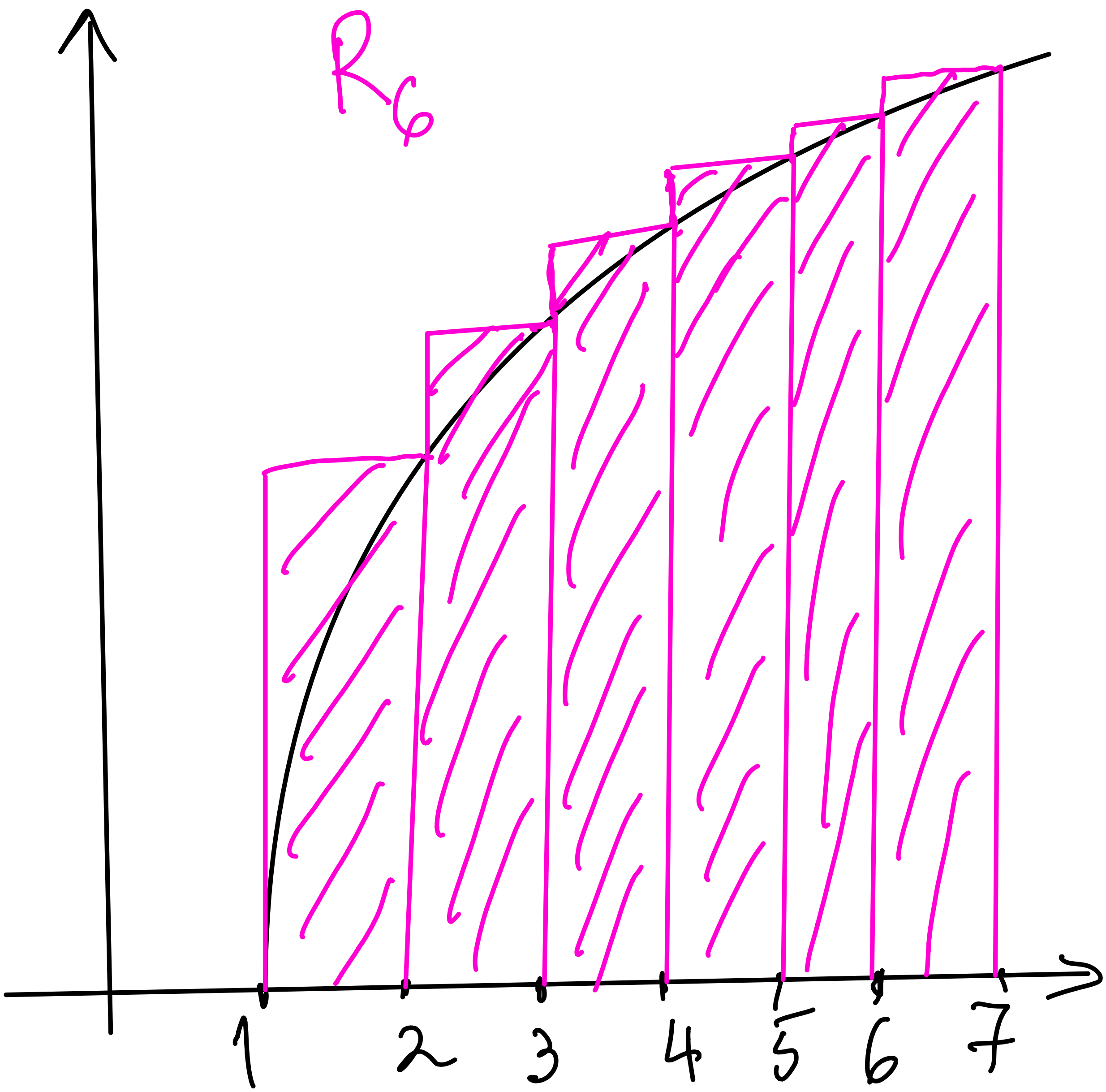
So, the left sum L_6 equals:

$\log(1)+\log(2)+\log(3)+\log(4)+\log(5)+\log(6)$

(note that $\Delta x = \frac{7-1}{6} = 1$)

So, $L_6 = \sum_{i=1}^{6} \log(i) = 6.57925\ldots$

Next, we use a right Riemann sum to estimate $\int_1^7 \log(x)\,dx$, again with $n=6$.



So, $R_6 = \sum_{i=2}^{7} \log(i) = \log(2) + \log(3) + \log(4) + \log(5) + \log(6) + \log(7)$

With the use of a calculator:

$R_6 = 8.52516\ldots$

Not surprising (since $\log(x)$ is increasing on $[1,7]$), we have that:

$$\begin{array}{ccccc} L_6 & < & \int_1^7 \log(x)\,dx & < & R_6 \\ \| & & & & \| \\ 6.57925\ldots & < & 7.62137\ldots & < & 8.52516\ldots \end{array}$$

Next we try better approximation

Clearly, in this case, the average of L_6 & R_6 is a much better approximation to the actual value of $\int_1^7 \log(x)\,dx$

Indeed: $\frac{L_6+R_6}{2} = 7.55220\ldots$ which is quite close to $7.62137\ldots$

We see that this average:

$$\frac{L_6+R_6}{2} = \frac{1}{2}\Big[\log(1)+2\log(2)+2\log(3)+2\log(4)+2\log(5)+2\log(6)+\log(7)\Big]$$

represents the area of 6 trapezoids which correspond to connecting the points on the graph of $\log(x)$:

$(1, \log(1)) \longrightarrow (2, \log(2))$

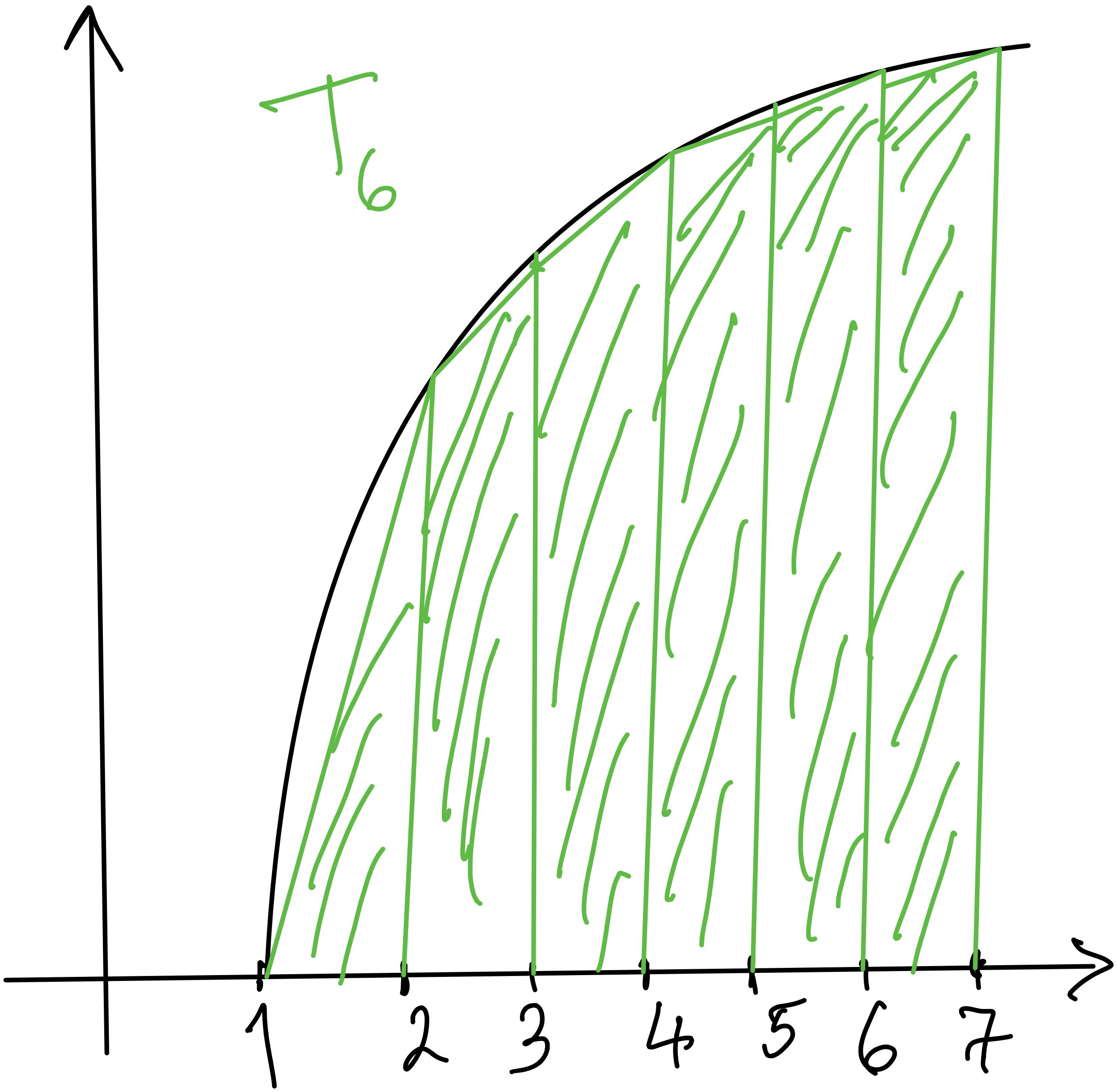
$(2, \log(2)) \longrightarrow (3, \log(3))$

$(3, \log(3)) \longrightarrow (4, \log(4))$

$(4, \log(4)) \longrightarrow (5, \log(5))$

$(5, \log(5)) \longrightarrow (6, \log(6))$

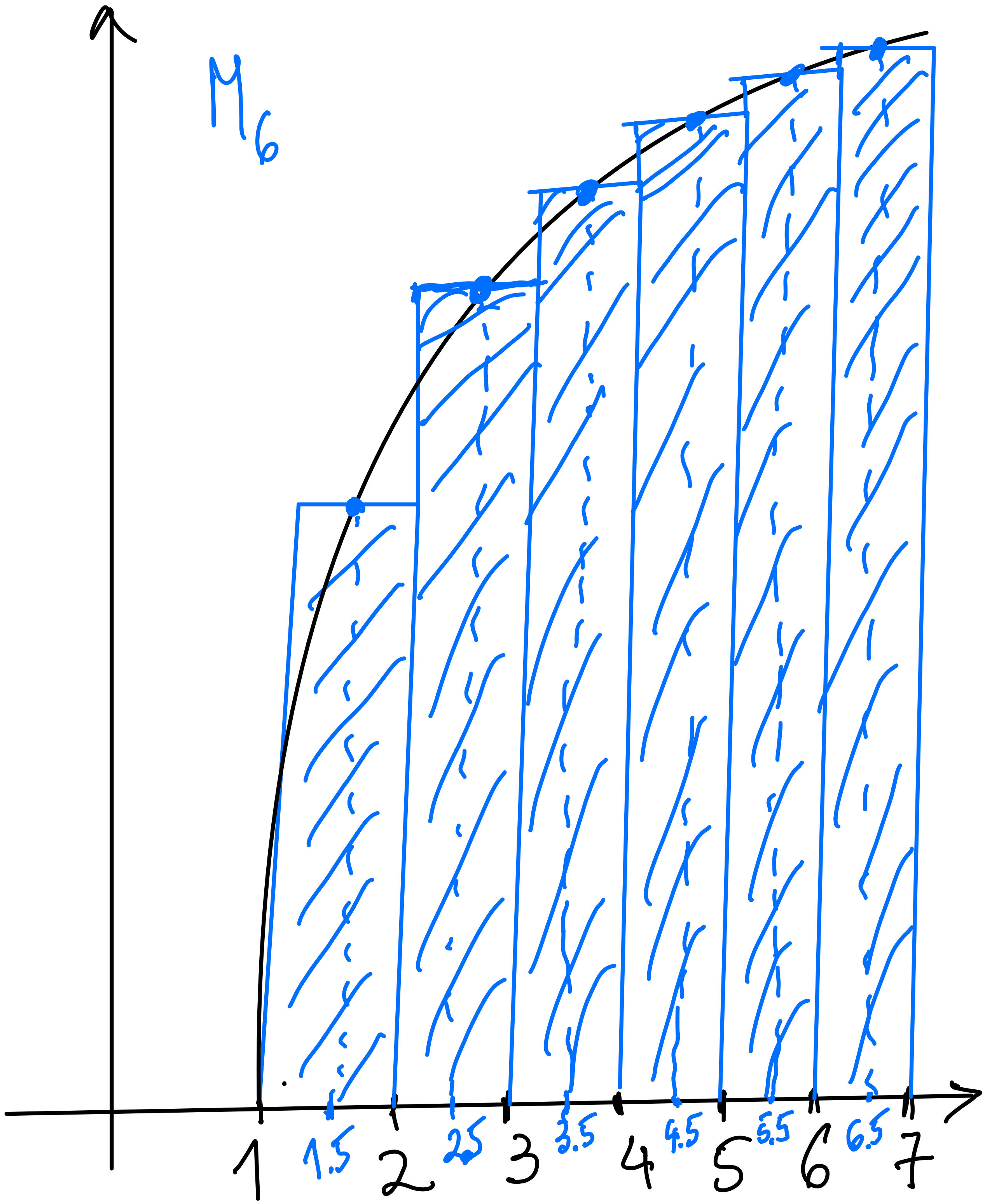
$(6, \log(6)) \longrightarrow (7, \log(7))$



$$T_6 = \sum_{i=1}^{6} \frac{\log(i+1) + \log(i)}{2}$$

is a very good approximation

Another option which also gives a good approximation is the Midpoint Rule. This amounts to constructing rectangles, but this time based on the height corresponding to the midpoint of each subinterval.



M6
1
1.5
2
2.5
3
3.5
4
4.5
5
5.5
6
6.5
7

So, $M_6 = \log(1.5) + \log(2.5) + \log(3.5) + \log(4.5) + \log(5.5) + \log(6.5) = 7.65514\ldots$

(again $\Delta x = 1$)

This approximation of:

$$\int_1^7 \log(x)\,dx = 7.62137\ldots$$

is even better than the previous ones.

In general, for a function $f(x)$ on an interval $[a, b]$, we have the Trapezoidal Rule with n subintervals is:

$$T_n = \Delta x \cdot \sum_{i=1}^{n} \frac{f(x_i) + f(x_{i-1})}{2},$$

where $\Delta x = \dfrac{b-a}{n}$, while

$x_i = a + i \cdot \Delta x$, for each $i = 0, 1, 2, \ldots, n$ $\left(\text{so, } x_0 = a \text{ \& } x_n = b\right)$

Similarly, the Midpoint Rule with n subintervals yields:

$$M_n = \Delta x \cdot \sum_{i=1}^{n} f\left(\frac{x_i + x_{i-1}}{2}\right),$$

again with $\Delta x = \frac{b-a}{n}$ &

$x_i = a + i \cdot \Delta x$, for $i = 0, 1, 2, \ldots, n$.

We define the errors in our approximations as follows:

$$E_T = \left| \int_a^b f(x)\,dx - T_n \right|$$

(the error in the Trapezoidal Rule with n subintervals)

and ~~also~~, we have:

$$E_M = \left| \int_a^b f(x)\,dx - M_n \right|$$

(the error in the Midpoint Rule with n subintervals)

Going back to our example of approximating $\int_1^7 \log(x)\,dx$

with 6 subintervals, then we have the 2 errors be:

$$E_T = \left| \int_1^7 \log(x)\,dx - T_6 \right|$$

$$= \left| (7\log(7) - 6) - \frac{1}{2}\begin{pmatrix} \log(1) + 2\log(2) \\ + 2\log(3) + 2\log(4) \\ + 2\log(5) + 2\log(6) \\ + \log(7) \end{pmatrix} \right|$$

$$= \left| 7.62137\ldots - 7.55220\ldots \right|$$

$$= 0.06917\ldots \longrightarrow \text{so, quite small!}$$

Similarly, we compute the error in the Midpoint Rule for approximating $\int_1^7 \log(x)\,dx$ with $n = 6$; we get:

$$E_M = \left| \int_1^7 \log(x)\,dx - \left(\log(1.5) + \log(2.5) + \log(3.5) + \log(4.5) + \log(5.5) + \log(6.5) \right) \right|$$

$$= \left| 7.62137\ldots - 7.65514\ldots \right|$$

$= 0.03377\ldots \longrightarrow$ so, even smaller!

Now, the fact that the Midpoint & the Trapezoidal Rules provide better approximations to $\int_a^b f(x)dx$ is provided by the following:

<u>Theorem</u> if K is a positive real number with the property that for

$a \leq x \leq b$, we have: $|f''(x)| \leq K$, then

$$E_T \leq \frac{K \cdot (b-a)^3}{12 \cdot n^2}$$

and

$$E_M \leq \frac{K \cdot (b-a)^3}{24 \cdot n^2}$$

We'll check next this theorem for the special case of $\int_1^7 \log(x)\,dx$, with $n=6$.

So, $f(x) = \log(x)$ and in the Theorem, we compute first $f''(x) = \left(\log(x)'\right)' =$

$$= \left(\frac{1}{x}\right)' = -\frac{1}{x^2}.$$

Then the Theorem says to look for a positive real number K for which:

$$\left| f''(x) \right| \leq K \text{ for all } a \leq x \leq b$$

<u>i.e</u>

$$\left| -\frac{1}{x^2} \right| \leq K \text{ for } 1 \leq x \leq 7.$$

Since $\left| -\frac{1}{x^2} \right| = \frac{1}{x^2}$

is the largest when $x = 1$, it means we can take

$K = \frac{1}{1^2} = 1$ and then

we'd be certain that:

$$|f''(x)| \leq K \text{ for all } a \leq x \leq b$$

<u>i.e.</u>

$$\left| -\frac{1}{x^2} \right| \leq 1 \text{ for } 1 \leq x \leq 7.$$

Then the Theorem says that the errors in the Trapezoidal respectively Midpoint Rules are guaranteed to satisfy:

$$E_T \leq \frac{K \cdot (b-a)^3}{12 n^2}$$

respectively

$$E_M \leq \frac{K \cdot (b-a)^3}{24 n^2}.$$

For our specific example: $n=6$, $a=1$, $b=7$ & we saw that $K=1$, which means that the 2 errors are guaranteed to be bounded by:

$$E_T \le \frac{1\cdot(7-1)^3}{12\cdot 6^2} = \frac{6^3}{12\cdot 6^2} = \frac{1}{2}$$

respectively

$$E_M \le \frac{1\cdot(7-1)^3}{24\cdot 6^2} = \frac{6^3}{24\cdot 6^2} = \frac{1}{4}$$

We've seen the actual errors are much smaller; in reality

$$E_T = 0.06917\ldots < 0.5$$

while

$E_M = 0.03377\ldots < 0.25$.

So, the Theorem only provides some information about the error, not the actual size of the error since the theorem applies to all possible functions & therefore, for some functions, the errors in the Midpoint or the Trapezoidal Rules are bigger.

However, the Theorem allows us to determine for which n, we know for sure that the error is below a certain bound.

For example, for which n, we know for sure that

$$E_T < 10^{-4} \; ?$$

So, we know that

$$E_T \le \frac{K \cdot (b-a)^3}{12 n^2}$$

and again we work with

$$\int_1^7 \log(x)\,dx$$ for which

we have: $a=1$, $b=7$, $K=1$.

So, $$E_T \le \frac{6^3}{12 \cdot n^2} = \frac{18}{n^2}$$

& so, we ask for which n, we'd know for

sure that $E_T \leq 10^{-4}$.
This means we want:

$$\frac{18}{n^2} \leq 10^{-4} \Rightarrow n^2 \geqslant \frac{18}{10^{-4}}$$

$$\Longleftrightarrow n^2 \geqslant 180{,}000 .$$

We see that $n = 500$ satisfies the inequality; actually, for $n = 425$, we already have: $n^2 = 425^2 = 180{,}625$

$$> 180{,}000$$

So, we know that for the Trapezoidal Rule with $n \geq 425$, we're guaranteed for $E_T < 10^{-4}$.

However, the actual value of n for which $E_T < 10^{-4}$ is most likely much smaller.

Once again, note that the Theorem gives us a bound which guarantees small error regardless of the function.

Re-doing the same analysis for the Midpoint Rule, for which we ask again for which n, we'd be certain that the error in the Midpoint Rule is less than 10^{-4}?

So, we have — for the

Same $\int_1^7 \log(x)\,dx$, the bound from the Theorem:

$$E_M \le \frac{K\cdot(b-a)^3}{24n^2},$$

where (as before): $a=1$, $b=7$, $K=1$

$$\Rightarrow E_M \le \frac{6^3}{24n^2} = \frac{9}{n^2}.$$

Since we want $E_M \le 10^{-4}$, it means we need:

$$\frac{9}{n^2} \le 10^{-4} \Rightarrow \frac{9}{10^{-4}} \le n^2$$

$$\Rightarrow 90{,}000 \le n^2 \Rightarrow \boxed{n \ge 300}$$

Just as in the case of the Trapezoidal Rule, also for the Midpoint Rule, the exact value of n for which the error E_M in evaluating $\int_1^7 \log(x)\,dx$ is less than 10^{-4} must definitely be smaller than 300 but finding the first n for which this happens is difficult. So, we settle for what the Theorem gives us: for $n \geq 300$, we know for sure that $E_M \leq 10^{-4}$.

We conclude with yet another method of approximating an integral.

Simpson's Rule

This time we approximate

$$\int_1^7 \log(x)\,dx \qquad \text{by}$$

$$S_6 = \frac{1}{3}\Big(\log(1) + 4\log(2) + 2\log(3) + 4\log(4) + 2\log(5) + 4\log(6) + \log(7)\Big)$$

The construction of S_6 comes from using 2 adjacent subintervals at a time in order to form the sums:

- $\frac{1}{3}\left(\log(1)+4\log(2)+\log(3)\right)$ from the subintervals $[1,2]$ & $[2,3]$
- $\frac{1}{3}\left(\log(3)+4\log(4)+\log(5)\right)$ from the subintervals $[3,4]$ & $[4,5]$
- $\frac{1}{3}\left(\log(5)+4\log(6)+\log(7)\right)$ from the subintervals $[5,6]$ & $[6,7]$

The actual mathematics beyond

this construction is more subtle & appears in the CLP textbook (section 11.1); the point is that now we consider parabolas passing through $\boxed{3}$ points at a time from the graph of our function.

In general, for an arbitrary function $f(x)$ on $[a,b]$, the integral $\int_a^b f(x)dx$ is approximated through Simpson's Rule with n subintervals (where n must be <u>even</u>) by the following sum:

$$S_n = \frac{\Delta x}{3} \cdot \left(f(x_0) + 4f(x_1) + 2f(x_2) + 4f(x_3) + 2f(x_4) + 4f(x_5) + 2f(x_6) + 4f(x_7) + \ldots + 2f(x_{n-2}) + 4f(x_{n-1}) + f(x_n) \right),$$

where $\Delta x = \frac{b-a}{n}$ &

$x_i = a + i \cdot \Delta x$ for $0 \leq i \leq n$.

Also, here there is a Theorem which provides an upper bound for the error in the Simpson's Rule: if L is a positive real number such that $\left| f^{(4)}(x) \right| \leq L$ for $a \leq x \leq b$ (i.e. the 4th derivative of $f(x)$)

is less than or equal to L in absolute value, for all $x \in [a,b]$)

then the error

$$E_S = \left| \int_a^b f(x)\,dx - S_n \right|$$

satisfies

$$E_S \leq \frac{L \cdot (b-a)^5}{180\, n^4}.$$

Next we compare S_6 with $\int_1^7 \log(x)\,dx$ and estimate E_S.

So, $S_6 = \frac{1}{3}\Big(\log(1) + 4\log(2) + 2\log(3) + 4\log(4) + 2\log(5) + 4\log(6) + \log(7)\Big) = 7.61560....$

The answer is very close to

$$\int_1^7 \log(x)\,dx = 7.62137....$$

So, the error in this Simpson's Rule application is:

$$E_S = \left|\int_1^7 \log(x)\,dx - S_6\right| =$$

$$= |7.62137.... - 7.61560....|$$

$$\boxed{E_S = 0.00577....}$$ VERY SMALL!

Now, our theorem provides a much bigger possible margin for the error, as we can see as follows.

$f(x) = \log(x)$; $f'(x) = \frac{1}{x}$;

$f''(x) = -\frac{1}{x^2}$; $f'''(x) = \frac{2}{x^3}$ & so,

$f^{(4)}(x) = -\frac{6}{x^4}$. Therefore the bound L which has to be larger (or equal) than $\left|f^{(4)}(x)\right|$ for $1 \leq x \leq 7$

is : $L = \max\limits_{1 \leq x \leq 7} \left|-\frac{6}{x^4}\right|$

$$L = \max_{1 \leq x \leq 7} \frac{6}{x^4}$$

$$L = \frac{6}{1^4} = 6 .$$

(since the largest value in $\frac{6}{x^4}$ corresponds to $x=1$)

So, our Theorem only guarantee that the error E_S when estimating $\int_1^7 \log(x)\,dx$ with the Simpson's Rule with 6 subintervals is bounded by:

$$E_S \leq \frac{L \cdot (b-a)^5}{180\, n^4} \quad \& \text{ so,}$$

because $L = 6$; $b = 7$; $a = 1$ & $n = 6$, we get:

$$E_S \leq \frac{6 \cdot 6^5}{180 \cdot 6^4} = \frac{36}{180} = \frac{1}{5} = 0.2$$

However, we already saw that "in reality", the error is much smaller:

$$E_S = 0.00577\ldots < 0.2$$

The advantage with our Theorem is that if we ask that $E_S \leq 10^{-6}$, we can find some n for which S_n is guaranteed to be within 10^{-6} of the actual value of the integral; here's how we do it.

So, we want n such that $E_S \leq 10^{-6}$; but we know that $E_S \leq \frac{L \cdot (b-a)^5}{180 \cdot n^4}$

& so, using again $L=6$, $b=7$, $a=1$, we get:

$$E_S \leq \frac{6 \cdot 6^5}{180 \cdot n^4} = \frac{1296}{5n^4}$$

So, we need n such that:

$$\frac{1296}{5n^4} \leq 10^{-6} \Rightarrow n^4 \geq \frac{1296}{5 \cdot 10^{-6}}$$

$$\Rightarrow n^4 \geq 259{,}200{,}000$$

We get $n \geq 127$ since

$$127^4 = 260{,}144{,}641 > 259{,}200{,}000$$

Finally, even though $n = 127$

is most likely way larger than what we'd need in reality for the error be less than 10^{-6}, we can already appreciate that it's pretty small number for subdivisions which are definitely guaranteeing an error less than 10^{-6}.

This is ~~way~~ better than either the Trapezoidal or the Midpoint Rules!