7 Techniques of integration

7.7 Approximate integration

- Exact integration is a very useful thing (especially for differential equations) — when we can do it.

- In order to do it we need to be able to find the anti-derivative of a function — we can't always do this.

$$\int e^{-x^2} \mathrm{d}x = ?$$

  We saw that we can write this down in terms of Taylor series, but we can also approximate it.

- There are several different ways of doing this — perhaps the most straight forward relies on interpretting the definite integral as "the area under the curve".

- By approximating the area we approximate the integral.

- We already know 1 way of approximating the integral — Riemann sums.

$$\int_a^b f(x)\mathrm{d}x \approx \sum_{i=1}^{n} f(x_i^*)\Delta x$$

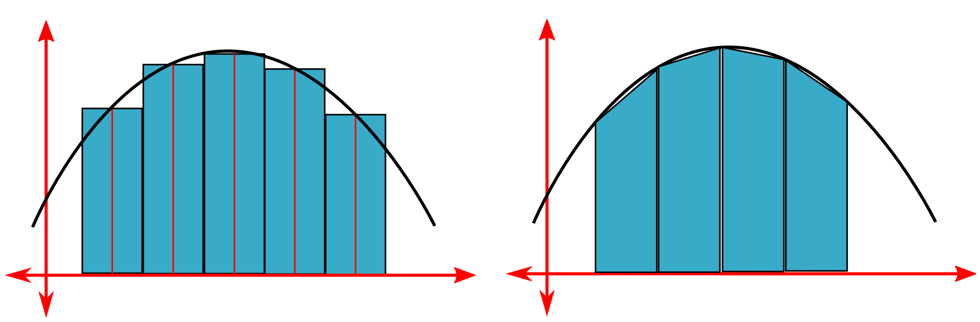
  with $x_i^* \in [x_{i-1}, x_i]$.

- It turns out that a good way to choose x_i^* is to take the mid-point of the interval $x_i^* = (x_{i-1} + x_i)/2$ — *ie* the error is smaller.

  Theorem (Midpoint rule)**.**

$$\int_a^b f(x)\mathrm{d}x \approx M_n = \sum f(\bar{x_i})\Delta x = \Delta x(f(\bar{x_1}) + f(\bar{x_1}) + f(\bar{x_2}) + \cdots + f(\bar{x_n}))$$

$$\Delta x = \frac{b-a}{n} \qquad x_i = a + i\Delta x \qquad \bar{x_i} = (x_{i-1} + x_i)/2$$



  This is perhaps the simplest way of doing things — inside our little subinterval $[x_i, x_{i+1}]$ we approximate the function as being a constant.

- In our Riemann sum we approximate the area by little rectangles. Another method is to approximate it by little trapezoids — ie approximate the function on $[x_i, x_{i+1}]$ by a line. So approximate the whole function as a sequence of lines.
- Left-hand side of trapezoid has height $f(x_{i-1})$. Right-hand side has height $f(x_i)$. The area is therefore
$$\Delta x \cdot (f(x_i) + f(x_{i-1}))/2$$

  Theorem (Trapezoid rule)**.**
$$\begin{aligned} \int_a^b f(x)\mathrm{d}x \approx T_n &= \sum (f(x_i) + f(x_{i-1}))/2 \cdot \Delta x \\ &= \frac{\Delta x}{2}\left(f(x_0) + 2f(x_1) + 2f(x_2) \cdots + 2f(x_{n-1}) + f(x_n)\right) \\ \Delta x &= \frac{b-a}{n} \qquad x_i = a + i\Delta x \end{aligned}$$

- So how good are these approximations — what is the error?
- The error is the difference between our approximation and the value of the integral — it is a function of n
$$Error(n) = \left| \int_a^b f(x)\mathrm{d}x - M_n \right|$$
- We know that as $n \to \infty$ the error goes to zero (since it is a riemann sum), but how quickly? The difference between a good method and a bad method.
- This depends on the function — in particular how "curvy" is the function — how much does its gradient change.
- If the gradient is constant — function is linear. Need to look at second derivative.
- Also — how wide are the slices — how big is Δx?

  Theorem (Error bounds)**.** *Suppose $|f''(x)| < K$ on $a \le x \le b$. If $E_M(n)$ and $E_T(n)$ are the errors of the midpoint and trapezoid rules respectively then*
$$|E_T(n)| \le \frac{K(b-a)^3}{12n^2} \qquad\qquad |E_M(n)| \le \frac{K(b-a)^3}{24n^2}$$

  Midpoint rule is slightly better.

Some examples. Approximate the integral $\int_1^2 1/x\mathrm{d}x$ using the trapezoid rule and $n = 5$.
$$\begin{aligned} &\int_a^b f(x)\mathrm{d}x \approx T_n = \frac{\Delta x}{2}\left(f(x_0) + 2f(x_1) + 2f(x_2) \cdots + 2f(x_{n-1}) + f(x_n)\right) \\ &a = 1, b = 2\Delta x = 0.2 \\ &\qquad T_n = 0.1(f(1) + 2f(1.2) + 2f(1.4) + 2f(1.6) + 2f(1.8) + f(2)) \\ &\qquad\quad = 0.1(1/1 + 2/1.2 + 2/1.4 + 2/1.6 + 2/1.8 + 1/2) = 0.6956\ldots \end{aligned}$$

Now let us bound the error.

$$f'(x) = -1/x^2 \qquad\qquad f''(x) = 2/x^3$$

$$|f''(x)| = 2|x^{-3}| \leq 2 \qquad\qquad \text{since } 1 \leq x \leq 2$$

$$|E_T(5)| \leq \frac{K(b-a)^3}{12n^2} = \frac{2 \cdot 1^3}{12 \cdot 25} = \frac{1}{150} \approx 0.00666$$

Let us turn the question around. How big do we need n to be in order for the error to be less than 10^{-6}?

$$10^{-6} \geq \frac{K(b-a)^3}{12n^2} = \frac{2}{12n^2}$$

$$10^6 \leq 6n^2$$

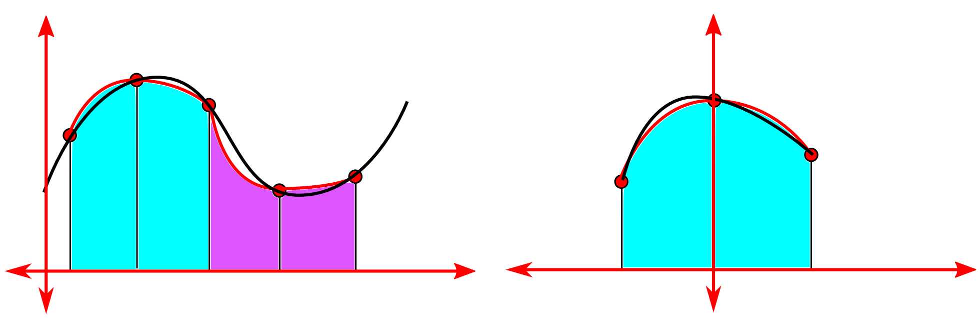
$$\frac{10^6}{6} \leq n^2$$

$$\sqrt{\frac{10^6}{6}} \approx 408.25 \leq n$$

So we need n to be ≥ 409.

To get the trapezoid rule, you approximate the function by a sequence of lines. A much better approximation is Simpson's rule. Approximate the curve by a sequence of parabolas — this is significantly more effective since it takes some change in slope into account.

- Divide $[a, b]$ into n segments with n even. Width is $\Delta x = (b-a)/n$.
- Each triple of points $(x_{i-1}, f(x_{i-1}), (x_i, f(x_i)), (x_{i+1}, f(x_{i+1})$ can be used to define a parabola. We then add up the area contributions from each parabola.



- Simplify things and look at a parabola going through 3 points $(-h, y_{-1}), (0, y_0), (h, y_1)$.
- The general equation for a parabola is $y = Ax^2 + Bx + C$, so the area under it is

$$\int_{-h}^{h} (Ax^2 + Bx + C)\mathrm{d}x = [Ax^3/3 + Bx^2/2 + Cx]_{-h}^{h}$$

$$= 2Ah^3/3 + 2Ch = \frac{h}{3}(2Ah^2 + 6C).$$

- Since we don't know A, B, C, we need to relate this to our 3 points.

$$y_{-1} = Ah^2 - Bh + Cy_0 \qquad\qquad = Cy_{-1} = Ah^2 + Bh + C$$

- Hence the area under the parabola is $\frac{h}{3}(y_{-1} + y_1 + 4y_0)$.
- Each triple of points gives us an area like this. Hence the total area is

$$\begin{aligned}\int_a^b f(x)\mathrm{d}x &\approx \frac{h}{3}(y_0 + 4y_1 + y_2) + \frac{h}{3}(y_2 + 4y_3 + y_4) + \cdots + \frac{h}{3}(y_{n-2} + 4y_{n-1} + y_n) \\ &= \frac{h}{3}(y_0 + 4y_1 + 2y_2 + 4y_3 + 2y_4 + \cdots 2y_{n-2} + 4y_{n-1} + y_n)\end{aligned}$$

Theorem (Simpson's rule)**.**

$$\begin{aligned}&\int_a^b f(x)\mathrm{d}x \approx S_n \\ &\quad = \frac{\Delta x}{3}\left[f(x_0) + 4f(x_1) + 2f(x_2) + 4f(x_3) + 2f(x_4) + \cdots 2f(x_{n-2}) + 4f(x_{n-1}) + f(x_n)\right]\end{aligned}$$

Where $\Delta x = \frac{b-a}{n}$ and n is an even number.
If $|f^{(4)}(x)| \leq K$ for all $a \leq x \leq b$ then the error $E_s(n)$ is bounded by

$$|E_s(n)| \leq \frac{K(b-a)^5}{180n^4}$$

Back to the previous example — we want to the error in estimating $\int_1^2 1/x\mathrm{d}x$ to be less than 10^{-6}. The 4th derivative is $24/x^5$ and so is bounded by 24.

$$\begin{aligned}10^{-6} &\geq \frac{K(b-a)^5}{180n^4} = \frac{24}{180n^4} \\ 10^6 &\leq \frac{15n^4}{2} \\ \frac{2}{15} \times 10^6 &\leq n^4 \\ 19.1 &\leq n\end{aligned}$$

Hence we need $n \geq 20$. Compare this to $n \geq 409$ for the trapezoid rule.

What do you need to know......

> Students will have to know the formulae for the Midpoint Rule (p. 496), the Trapezoidal Rule (p. 497) and Simpson's Rule (p. 502) for the final exam. But reassure them that they will be provided with the error bounds [3] (p. 499) and [4] (p. 503) on the final exam if they are needed.