Numerical integration

Dr.ir. Ivo Roghair, Prof.dr.ir. Martin van Sint Annaland

Chemical Process Intensification group
Eindhoven University of Technology

Numerical Methods (6BER03), 2024-2025

Today's outline

What is numerical integration?

To determine the integral $I(x)$ of an integrand $f(x)$, which can be used to compute the area underneath the integrand between $x = a$ and $x = b$.

$$I(x) = \int_a^b f(x)dx$$

Today we will outline different numerical integration methods.

- Riemann integrals
- Trapezoidal rule
- Simpson's rule

Why do chemical engineers need integration?

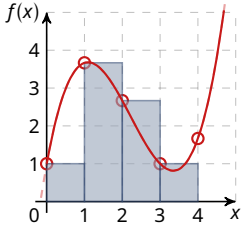
- Obtaining the cumulative particle size distribution from a particle size distribution
- The concentration outflow over time may be integrated to yield the residence time distribution
- Integration of a varying product outflow yields the total product outflow
- Quantitative analysis of mixture components via e.g. GC/MS
- Not all function have an explicit antiderivative, e.g. $\int e^{x^2} dx$ or $\int \frac{1}{\ln x} dx$

Today's outline

Riemann integrals

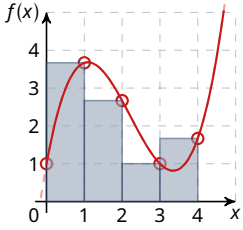
Basic idea: Subdivide the interval $[a, b]$ into n subintervals of equal length $\Delta x = \frac{b-a}{n}$ and use the sum of area to approximate the integral.

Left endpoint rule



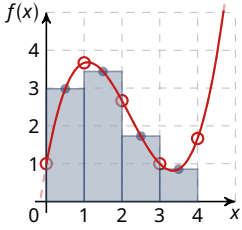
$$L_n = \sum_{i=0}^{n-1} f(x_i)\Delta x_i$$

Right endpoint rule



$$R_n = \sum_{i=0}^{n-1} f(x_{i+1})\Delta x_i$$

Midpoint rule



$$M_n = \sum_{i=0}^{n-1} f(\bar{x}_i)\Delta x_i$$

with $\bar{x}_i = \frac{x_i + x_{i+1}}{2}$

Errors in Riemann integrals

We define the exact integral as $I = \int_a^b f(x)dx$, and L_n, R_n and M_n represent the left, right and midpoint rule approximations of I based on n intervals.

Writing $f^{(k)}_{\max}$ for the maximum value of the k-th derivative, the upper-bounds of the errors by Riemann integrals are:

- $|I - L_n| \leq \frac{f^{(1)}_{\max}(b-a)^2}{2n}$
- $|I - R_n| \leq \frac{f^{(1)}_{\max}(b-a)^2}{2n}$
- $|I - M_n| \leq \frac{f^{(2)}_{\max}(b-a)^3}{24n^2}$

Note that while $|I - L_n|$ and $|I - R_n|$ give the same *upper-bounds* of the error, this does not mean the same error. Rather, the error is of opposite sign!

Today's outline

Trapezoid rule

Since the sign of the approximation error of the left and right endpoint rules is opposite, we can take the average of these approximations:

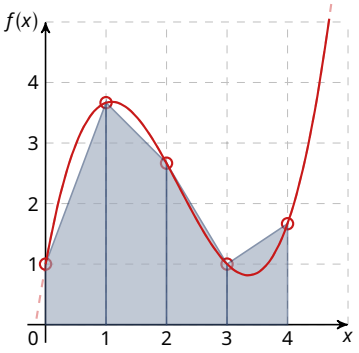
$$T_n = \frac{L_n + R_n}{2}$$

The total area is obtained by geometric reconstruction of trapezoids:

$$T_n = \sum_{i=0}^{n-1} \frac{f(x_{i+1}) + f(x_i)}{2} \Delta x_i$$

Note that this can be rewritten for equidistant intervals:

$$T_n = \frac{b-a}{2n} \left(f(x_0) + 2f(x_1) + \ldots + 2f(x_{n-1}) + f(x_n) \right)$$



Error in trapezoid integration

The trapezoid rule result over n intervals T_n approximates the exact integral $I = \int_a^b f(x)dx$. The upper-bounds of the error is given as:

$$|I - T_n| \leq \frac{f^{(2)}_{\max}(b-a)^3}{12n^2}$$

Recall that the midpoint rule approximates with an upper-bound error of

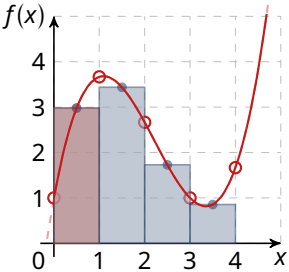
$$|I - M_n| \leq \frac{f^{(2)}_{\max}(b-a)^3}{24n^2}$$

The midpoint rule approximation has lower error bounds than the trapezoid rule. A linear function is, however, better approximated by the trapezoid rule.

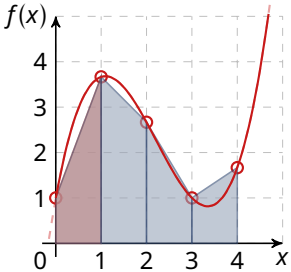
Today's outline

Towards higher-order integration

Compare how the midpoint and trapezoid functions behave on convex and concave parts of a graph.



Midpoint rule



Trapezoid rule

In convex parts (bending down), the midpoint rule tends to overestimate the integral (trapezoid underestimates). In concave parts (bending up), the midpoint rule tends to underestimate the integral (trapezoid overestimates).

Towards higher-order integration

The errors of the midpoint rule and trapezoid rule behave in a similar way, but have opposite signs.

- Midpoint: $|I - M_n| \leq \frac{f^{(2)}_{\max}(b-a)^3}{24n^2}$
- Trapezoid: $|I - T_n| \leq \frac{f^{(2)}_{\max}(b-a)^3}{12n^2}$

For a quadratic function, the errors relate as:

$$|I - M_n| = \frac{1}{2}|I - T_n|$$

Taking the weighted average of these two yields the Simpson's rule:

$$S_{2n} = \frac{2}{3}M_n + \frac{1}{3}T_n$$

The $2n$ means we have $2n$ subintervals: the n trapezoid intervals are subdivided by the midpoint rule.

Simpson's rule

Consider the interval $i \in [x_0, x_2]$, subdivided in three equidistant interpolation points: x_0, x_1, x_2.

- Midpoint: $M_i = f(\frac{x_0 + x_2}{2})2\Delta x = f(x_1)2\Delta x$
- Trapezoid: $T_i = \frac{f(x_0) + f(x_2)}{2}2\Delta x$
- Simpson: $S_i = \frac{2}{3}M_i + \frac{1}{3}T_i$

Note that M_i and T_i were computed on interval $x_2 - x_0 = 2\Delta x$.

Now we have:

$$S_i = \frac{2}{3}\left[f(x_1)2\Delta x\right] + \frac{1}{3}\left[\frac{f(x_0) + f(x_2)}{2}2\Delta x\right]$$
$$= \frac{4\Delta x}{3}f(x_1) + \frac{\Delta x}{3}f(x_0) + f(x_2) = \frac{\Delta x}{3}\left(f(x_0) + 4f(x_1) + f(x_2)\right)$$

Simpson's rule

We write $f(x_k) = f_k$. The integral of an interval $i \in [x_0, x_2]$ is approximated as:

$$S_i = \frac{\Delta x}{3}(f_0 + 4f_1 + f_2)$$

The next interval, S_j with $j \in [x_2, x_4]$ with midpoint $x_3 = \frac{x_2+x_4}{2}$ is approximated as:

$$S_j = \frac{\Delta x}{3}(f_2 + 4f_3 + f_4)$$

If we sum these two intervals we obtain:

$$\begin{aligned} I \approx S_i + S_j &= \left[\frac{\Delta x}{3}(f_0 + 4f_1 + f_2)\right] + \left[\frac{\Delta x}{3}(f_2 + 4f_3 + f_4)\right] \\ &= \frac{\Delta x}{3}(f_0 + 4f_1 + 2f_2 + 4f_3 + f4) \end{aligned}$$

Simpson's rule

In general, Simpson's rule can be written as:

$$\int_a^b f(x)dx \approx \sum_{\substack{k=2 \\ k \text{ even}}}^{n} \frac{\Delta x}{3}(f_{k-2} + 4f_{k-1} + f_k)$$

$$= \frac{\Delta x}{3}(f_0 + 4f_1 + 2f_2 + 4f_3 + 2f_4 + \ldots + 2f_{n-2} + 4f_{n-1} + f_n)$$

The error is given by:

$$|I - S_n| \leq \frac{f^{(4)}_{\max}(b-a)^5}{180n^4}$$

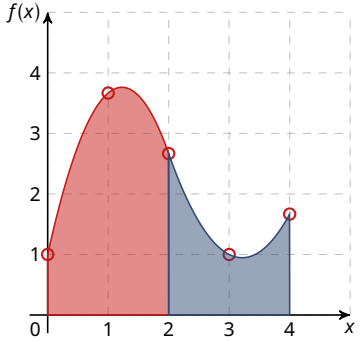
if integrand f is differentiable on $[a, b]$.

Simpson's rule: example

Recall our example data, described by $f(x) = \frac{x^3}{2} - \frac{10x^2}{3} + \frac{11x}{2} + 1$

$I = \int_0^4 \frac{x^3}{2} - \frac{10x^2}{3} + \frac{11x}{2} + 1 = \frac{80}{9} \approx 8.888\ldots$

- Interpolating x_0, x_1 and x_2: $p_{2a}(x) = -\frac{11}{6}x^2 + 4\frac{1}{2}x + 1$
  $\int_0^2 p_{2a} = \frac{55}{9} \approx 6.1111$
- Interpolating x_2, x_3 and x_4: $p_{2b}(x) = \frac{7x^2}{6} - 7\frac{1}{2}x + 13$
  $\int_2^4 p_{2b} = \frac{25}{9} \approx 2.777\ldots$
- Adding the separate integrals:
  $\int_0^2 p_{2a} + \int_2^4 p_{2b} = \frac{80}{9}$



Using Simpson's rule:

$I \approx \frac{\Delta x}{3}(f_0 + 4f_1 + 2f_2 + 4f_3 + f_4) = \frac{1}{3}(1 + 4 \cdot 3.6667 + 2 \cdot 2.6667 + 4 \cdot 1.0000 + 1.6667) = 8.88888 = \frac{80}{9}$

Simpson's method is of fourth order, and it gives exact approximations of third order polynomials!

Integration in Python

Integration can be done numerically in Python.

- `np.trapz(y, x)` uses the trapezoid rule to integrate the data. Make sure you use the `x` variable if your data is not spaced with $\Delta x = 1$. Can handle non-equidistant data.

```python
import numpy as np
x = np.linspace(-2, 2, 2001)
y = 1 / (x**2 + 1)
I = np.trapz(y, x) # Or: scipy.integrate.trapezoid
print(I)
```

```
2.214297328921525
```

- Integration of functions can be done using the `quad(func, a, b)` function:

```python
import numpy as np
from scipy.integrate import quad
f = lambda x: np.exp(-x**2)
I, err = quad(f, 0, 10)
print(I, err)
```

```
0.886226925452758 1.8483380528941764e-13
```

Today's outline

What hasn't been discussed?

This course is by no means complete, and further reading is possible.

- Gaussian quadrature: A third-order integration method that requires only two base points (in contrast to the third order Simpson's method, which requires three points)
- Adaptive techniques: Parts of a function that are relatively steady (no wild oscillations) and differentiable can be integrated with much larger step sizes than other parts of the function.
- Simpson's 3/8-rule: Yet another integration technique, requiring an additional data point

Summary

- Several techniques for numerical integration were discussed:
  - Riemann sums, trapezoid rule, Simpson's rule
  - Upper-bound errors were given for each technique
  - Built-in Python functions were illustrated
- Continue with characterization of convergence of the integration methods in the tutorials!

Integration tutorials

1. Implement a function to integrate a mathematical function for a specific number of integration intervals. Implement it as a function, which can be called with arguments:
   - Function (handle) to integrate
   - Integration boundaries (as separate arguments or as a 2×1 numpy array)
   - Number of integration intervals

   For instance: `def leftrule(func, x0, x1, N):.`
2. Set up a function to integrate:

```
def myfunction(x):
    return x**2 - 4*x + 6 + np.sin(5*x)
```

3. Integrate the function, e.g. `int_left = leftrule(myfunction, 0, 10, 25)`
4. Assess how the number of intervals affects the deviation from the true integral value.
5. Create a log-log plot of the deviation vs. number of intervals used.
6. Do this for all methods discussed[1] and compare their performance in a graph

[1] Riemann left, right, midpoint, trapezoid, and Simpson