Queens College, CUNY, Department of Computer Science

Numerical Methods

CSCI 361 / 761

Fall 2017

Instructor: Dr. Sateesh Mane

March 26, 2018

6 Lecture 6b: the Euler–Maclaurin formula

- **This lecture is classified as an 'advanced topic' and is not for examination.**
- The formulas in Lecture 6 for the error term of the trapezoid rule, leading to Romberg integration, are applicable ***only if $f(x)$ is infinitely differentiable for all $a \le x \le b$.***
- This lecture is devoted to a more careful analysis of the Euler–Maclaurin formula.

6.21 Euler–Maclaurin formula

- Our starting point is the integral we are given to evaluate:

$$I = \int_a^b f(x)\,dx\,. \tag{6.21.1}$$

- Let us denote the exact value of the integral by I_{ex}.
- We employ the trapezoid rule with n equal subintervals to obtain the approximate value

$$T_n = h\left[\frac{f(x_0)}{2} + f(x_1) + f(x_2) + \cdots + f(x_{n-1}) + \frac{f(x_n)}{2}\right]. \tag{6.21.2}$$

- Here $h = (b-a)/n$ is the length of each subinterval.
- The **Euler–Maclaurin formula** states that, if f is p times continuously differentiable for all $x \in [a,b]$, then

$$\begin{aligned} T_n - I_{\text{ex}} &= \frac{h^2 B_2}{2!}(f'(b) - f'(a)) + \frac{h^3 B_3}{3!}(f''(b) - f''(a)) \\ &\qquad + \cdots + \frac{h^p B_p}{p!}(f^{(p-1)}(b) - f^{(p-1)}(a)) + R_p \\ &= \left(\sum_{k=2}^{p} \frac{h^k B_k}{k!}(f^{(k-1)}(b) - f^{(k-1)}(a))\right) + R_p\,. \end{aligned} \tag{6.21.3}$$

- Here the B_k are the **Bernoulli numbers** and R_p is a remainder term.
- Actually $B_k = 0$ for all odd k except B_1, so we can omit all the odd values of k in eq. (6.21.3).
- It is convenient to assume that $f(x)$ is $2p$ times continuously differentiable for all $x \in [a,b]$.
- Then we obtain the following sum:

$$T_n - I_{\text{ex}} = \left(\sum_{k=1}^{p} \frac{h^{2k} B_{2k}}{(2k)!}(f^{(2k-1)}(b) - f^{(2k-1)}(a))\right) + R_{2p}\,. \tag{6.21.4}$$

- Many (most?) books state that if f is infinitely differentiable for all $x \in [a,b]$, then

$$T_n - I_{\text{ex}} = \sum_{k=1}^{\infty} \frac{h^{2k} B_{2k}}{(2k)!}(f^{(2k-1)}(b) - f^{(2k-1)}(a))\,. \tag{6.21.5}$$

- They extend the sum to ∞ and drop the remainder term R_{2p} for $p \to \infty$.
- It is, however, recognized in the literature that the right hand side of eq. (6.21.5) is in general **not a convergent series.**
- We say that the right hand side of eq. (6.21.5) is in general an **asymptotic series.**
- **We shall therefore analyze the remainder term R_{2p} as $p \to \infty$, for a few cases.**

6.22 Example: infinitely differentiable function

6.22.1 Euler–Maclaurin series

- Let us use $f(x) = \cos(4x)/(2\pi)$ and analyze the integral

$$I_{c4} = \frac{1}{2\pi}\int_0^{2\pi} \cos(4x)\,dx\,. \tag{6.22.1.1}$$

- We know of course that the exact value is $I_{c4} = 0$.
- The function $\cos(4x)/(2\pi)$ is infinitely differentiable for all x.
- The function $\cos(4x)/(2\pi)$ is also periodic with a period of 2π, although we shall not require this property in the analysis below.
- Let us analyze the Euler–Maclaurin series for this example.
- It is easily derived that

$$f^{2k-1}(x) = (-1)^k 4^{2k-1}\,\frac{\sin(4x)}{2\pi}\,. \tag{6.22.1.2}$$

- Hence evaluation at $x = 0$ and 2π yields $f^{2k-1}(0) = f^{2k-1}(2\pi) = 0$ for all $k = 1, 2, \dots$.
- **Hence all the terms in the Euler–Maclaurin series vanish.**
- **If we accept eq. (6.21.5) literally, it therefore tells us that for for this example**

$$T_n - I_{c4} = 0\,. \tag{6.22.1.3}$$

- *This is clearly false.* We can easily verify that T_n does not equal $I_{c4}(=0)$ for all values of n:

$$n=1: \qquad T_1 = \frac{\cos(0)+\cos(8\pi)}{2} \qquad =1\,, \tag{6.22.1.4a}$$

$$n=2: \qquad T_2 = \frac{1}{2}\left[\frac{\cos(0)+\cos(8\pi)}{2} + \cos(4\pi)\right] \qquad =1\,, \tag{6.22.1.4b}$$

$$n=3: \qquad T_3 = \frac{1}{2}\left[\frac{\cos(0)+\cos(8\pi)}{2} + \cos(\frac{8\pi}{3}) + \cos(\frac{16\pi}{3})\right] \qquad =0\,, \tag{6.22.1.4c}$$

$$n=4: \qquad T_4 = \frac{1}{4}\left[\frac{\cos(0)+\cos(8\pi)}{2} + \cos(2\pi) + \cos(4\pi) + \cos(6\pi)\right] \qquad =1\,. \tag{6.22.1.4d}$$

- For all $n > 4$ we obtain $T_n = 0$ exactly. (Also $T_3 = 0$.)
  1. Hence the value of T_n does not approach the exact value of zero "gradually" as n increases. Instead $T_n \neq I_{c4}$ for some values of n (specifically $n = 1$, 2 and 4) and then abruptly equals the exact result of zero for $n > 4$.
  2. Whatever the numerical error in the value of T_n is, it does not decrease to zero gradually as n increases.
- **Because all the terms in the Euler–Maclaurin series vanish, the numerical error in the value of T_n arises solely from the remainder term R_{2p}.**
- **We must therefore analyze the remainder term R_{2p}, in particular the limit $p \to \infty$.**

6.22.2 Remainder term

- An upper bound on the magnitude $|R_{2p}|$ is given by

$$|R_{2p}| \le \frac{2\zeta(2p)h^{2p}}{(2\pi)^{2p}} \int_0^{2\pi} |f^{2p}(x)|\,dx\,. \tag{6.22.2.1}$$

- Here $\zeta(s)$ is the **Riemann zeta function,** which for complex s with $\Re\{s\} > 1$ is given by

$$\zeta(s) = 1 + \frac{1}{2^s} + \frac{1}{3^s} + \cdots = \sum_{n=1}^{\infty} \frac{1}{n^s}\,. \tag{6.22.2.2}$$

- Hence for $2p \gg 1$ we can approximate $\zeta(2p) \simeq 1$ and bound it by $|\zeta(2p)| < 2$.
- Note also that $h = 2\pi/n$.
- Next note that $|f^{2p}(x)| = 4^{2p}|\cos(4x)|/(2\pi)$.
- Hence for $f(x) = \cos(4x)/(2\pi)$, the upper bound in eq. (6.22.2.1) is

$$|R_{2p}| < 4\,\frac{4^{2p}}{(2\pi)^{2p}}\frac{(2\pi)^{2p}}{n^{2p}} \int_0^{2\pi} \frac{|\cos(4x)|}{2\pi}\,dx < 4\,\frac{4^{2p}}{n^{2p}}\,. \tag{6.22.2.3}$$

- Hence the limit as $p \to \infty$ depends on the value of n.
  1. If $n = 1$ or 2, then the upper bound goes to ∞ as $p \to \infty$.
  2. If $n = 4$, then the limit of the upper bound is a nonzero constant $\lim_{p\to\infty} |R_{2p}| < 4$.
  3. *If $n > 4$, then the upper bound approaches zero:* $\lim_{p\to\infty} |R_{2p}| = 0$.
- Hence if $n > 4$ then we obtain $T_n = 0$, as observed.
- **However if $n \le 4$, then the remainder term is not necessarily zero.**
- All the terms in the Euler–Maclaurin series vanish, and the answer is given solely by the limiting value of remainder term

$$T_n = \lim_{p\to\infty} R_{2p}\,. \tag{6.22.2.4}$$

- Using the Kronecker delta, we obtained

$$T_n = \delta_{n,1} + \delta_{n,2} + \delta_{n,4}\,. \tag{6.22.2.5}$$

- Evidently the limiting value of the remainder term depends on n, and equals 1 for $n = 1$, 2 and 4, and is zero for $n > 4$. (It is also zero for $n = 3$, but this is not so easy to prove.)
- Hence, even for infinitely differentiable functions, we cannot assume $R_p \to 0$ as $p \to \infty$, as (tacitly?) assumed in eq. (6.21.5).

6.23 Revised formula for error term

- If $f(x)$ is infinitely differentiable for all $x \in [a, b]$, we need to modify eq. (6.21.5) to say instead

$$T_n - I_{\rm ex} = \lim_{p\to\infty} \left\{ \left(\sum_{k=1}^{p} \frac{h^{2k} B_{2k}}{(2k)!} (f^{(2k-1)}(b) - f^{(2k-1)}(a)) \right) + R_{2p}(h) \right\}. \tag{6.23.1}$$

- If $f(x)$ is moreover periodic with period $b - a$, then the terms in the Euler–Maclaurin series vanish and the error arises solely from the remainder term

$$T_n - I_{\rm ex} = \lim_{p\to\infty} R_{2p}(h)\,. \tag{6.23.2}$$

- Note the following (see eq. (6.21.5)):

  1. As already remarked above, in general the series in eq. (6.23.1) does not converge to a finite limit as $p \to \infty$.
  2. Therefore the value of the remainder term $R_{2p}(h)$ in eq. (6.23.1) also may not converge to a finite limit as $p \to \infty$.
  3. Hence, as $p \to \infty$, the right hand side of eq. (6.23.1) can consist of two diverging quantities which cancel to yield a finite result.

- In Romberg integration, we employ eq. (6.21.5), i.e. *without a remainder term*, and perform subtractions to systematically cancel the terms in h^2, h^4, etc.

- However, the Romberg integration procedure will in many cases *diverge* if we attempt to cancel too many powers of h.

6.24 Polynomials

- Let $f(x)$ be a polynomial of degree s.
- Then for all $r > s$, its derivatives are zero: $f^{(r)}(x) = 0$ for all $r > s$.
- Then the sum on the right hand side in eq. (6.23.1) terminates after a finite number of terms.
- We obtain the exact value for the integral by setting $p = \lfloor (s/2) \rfloor$:

$$T_n - I_{\rm ex} = \sum_{k=1}^{\lfloor (s/2) \rfloor} \frac{h^{2k} B_{2k}}{(2k)!} \left(f^{(2k-1)}(b) - f^{(2k-1)}(a) \right) . \qquad (6.24.1)$$

- Specifically, if $f(x)$ is linear in x, so $s = 1$, the right hand side of eq. (6.24.1) is zero.
- It is well known that the trapezoid method yields the exact result if $f(x)$ is a linear polynomial.

6.25 Infinitely differentiable periodic function and Discrete Fourier Transform

- Suppose $f(x)$ is infinitely differentiable for all x.
- Suppose also that $f(x)$ is periodic in x with period 2π, i.e. $f(x+2\pi) = f(x)$ for all x.
- Suppose also that $f(x)$ satisfies suitable integrability conditions which we shall not worry about here, e.g.

$$\int_0^{2\pi} |f(x)|\,dx < \infty\,. \tag{6.25.1}$$

- Then we can expand $f(x)$ in a Fourier series as follows:

$$\begin{aligned} f(x) = \tfrac{1}{2}a_0 &+ a_1\cos(x) + a_2\cos(2x) + \cdots \\ &+ b_1\sin(x) + b_2\sin(2x) + \cdots \end{aligned} \tag{6.25.2}$$

- Suppose we wish to calculate the integral

$$A_0 = \frac{1}{\pi}\int_0^{2\pi} f(x)\,dx\,. \tag{6.25.3}$$

- We know the exact answer is $A_0 = a_0$.
- Because $f(0) = f(2\pi)$, the trapezoid rule with n subintervals yields

$$T_n = \frac{1}{n}\left[f(0) + f\Big(\frac{2\pi}{n}\Big) + \cdots + f\Big(\frac{2\pi(n-1)}{n}\Big)\right]. \tag{6.25.4}$$

- The above expression is related to the **Discrete Fourier Transform.**
- In general, the value of T_n will not equal a_0 exactly.
- However, suppose the function $f(x)$ is **bandwidth limited.**
- This means that there exists an integer $m \ge 0$ such that $a_j = b_j = 0$ for all $j > m$.
- Then it is known from the theory of Fourier series that the value of T_n will equal a_0 exactly for all $n > m$.
- We can relate this example to eq. (6.23.1).
- First we recall that for an infinitely differentiable function, eq. (6.23.1) is applicable.
- Next we note that for a periodic function with period 2π, for all $k \ge 0$ we have

$$f^{(k)}(0) = f^{(k)}(2\pi)\,, \qquad (k = 0, 1, 2, \dots)\,. \tag{6.25.5}$$

- Therefore $f^{(k)}(2\pi) - f^{(k)}(0) = 0$ hence all the terms in the Euler–Maclaurin series vanish.

- Hence for an infinitely differentiable periodic function, eq. (6.23.1) simplifies to eq. (6.23.2):

$$T_n - I_{\rm ex} = \lim_{p\to\infty} R_{2p}(h)\,. \tag{6.25.6}$$

- Employing the same analysis as in Sec. 6.22.2, we obtain (see eq. (6.22.2.3))

$$|R_{2p}| < 4\,\frac{m^{2p}}{(2\pi)^{2p}}\frac{(2\pi)^{2p}}{n^{2p}}\int_0^{2\pi} |f(x)|\,\frac{dx}{\pi} < \text{const.} \times \frac{m^{2p}}{n^{2p}}\,. \tag{6.25.7}$$

- Hence if $n \le m$, then the limit of the remainder term is not zero: $\lim_{p\to\infty} |R_{2p}| \ne 0$, and we cannot expect the value of T_n to equal a_0 exactly.

- **However if $n > m$, then the limit of the remainder term is zero: $\lim_{p\to\infty} |R_{2p}| = 0$.**

- Therefore for all $n > m$ (the bandwidth limit), we obtain the exact answer $T_n = a_0$.

- A similar analysis show that for all $n > m$, the trapezoid rule yields the exact answer for the integrals

$$A_j = \frac{1}{\pi}\int_0^{2\pi} f(x)\cos(jx)\,dx\,, \qquad B_j = \frac{1}{\pi}\int_0^{2\pi} f(x)\sin(jx)\,dx\,. \tag{6.25.8}$$

- The results in eq. (6.25.8) are known from the theory of the Discrete Fourier Transform.

- However, the Discrete Fourier Transform is an application of the trapezoid rule, and the Euler–Maclaurin series must confirm that the trapezoid rule yields the exact answer for all $n > m$, ***and moreover that it is* not *guaranteed to do so for $n \le m$.***

- In this respect, eq. (6.21.5) is not good enough because all the terms on the right hand side of eq. (6.21.5) vanish, which implies that the trapezoid rule should yield the exact answer for *all values of n*, which is clearly not true.

- We require eq. (6.23.1) to justify the error analysis of the trapezoid rule, for infinitely differentiable periodic functions.

6.26 Romberg integration

- Romberg integration is a technique to employ **Richardson extrapolation** to systematically cancel higher order error terms for the numerical evaluation of integrals.
- The most common application (I think) employs the trapezoid rule.
- For example, the error term of $O(1/n^2)$, i.e. $O(h^2)$, can be cancelled by subtraction (see eq. (6.21.4))

$$R = \frac{4T_{2n} - T_n}{3}\,. \tag{6.26.1}$$

- The Romberg method can be systematized as follows.
  1. Define integers $j \geq 0$ and $k \geq 0$.
  2. Then set $k = 0$ and $n = 2^j$ and define the zeroth order Romberg terms for $j \geq 0$

$$R(j,0) = T(n = 2^j)\,. \tag{6.26.2}$$

  3. Then set $k = 1$. The first order Romberg terms are (here $j \geq 1$)

$$R(j,1) = \frac{4R(j,0) - R(j-1,0)}{3}\,. \tag{6.26.3}$$

  4. This cancels the term of $O(h^2) = O(1/n^2)$ in the Euler–Maclaurin series in eq. (6.21.4).
  5. Next set $k = 2$. The second order Romberg terms are (here $j \geq 2$)

$$R(j,2) = \frac{4^2 R(j,1) - R(j-1,1)}{4^2 - 1}\,. \tag{6.26.4}$$

  6. This cancels the term of $O(h^4) = O(1/n^4)$ in the Euler–Maclaurin series in eq. (6.21.4).
  7. For arbitrary $k > 0$, the k^{th} order Romberg terms are, with $j \geq k$,

$$R(j,k) = \frac{4^k R(j,k-1) - R(j-1,k-1)}{4^k - 1}\,. \tag{6.26.5}$$

  8. This cancels the term of $O(h^{2k}) = O(1/n^{2k})$ in the Euler–Maclaurin series in eq. (6.21.4).
- If $f(x)$ is a polynomial of degree s, then from Sec. 6.24, $R(k,k)$ yields the exact value of the integral for $k = \lfloor (s/2) \rfloor$. In particular, if $s = 3$ (cubic polynomials), then $k = \lfloor (s/2) \rfloor = 1$, which is Simpson's rule. This is a proof that Simpson's rule yields the exact result for all polynomials up to degree 3.
- It is known that Romberg integration runs into difficulties if the integrand $f(x)$ is only finitely differentiable $2p$ times for all $x \in [a,b]$.
- Then the Euler–Maclaurin series in eq. (6.21.4) contains only p terms.
- Hence if $k > p$ the Romberg method does not yield any improvement in numerical accuracy.

- However, we see that even for functions which are infinitely differentiable for all $x \in [a, b]$, the remainder term $R_{2p}(h)$ does not necessarily vanish as $p \to \infty$, and it can contribute significantly to the numerical error (perhaps even all of the error).
- The successive Romberg subtractions cause $R(j, k)$ to contain linear combinations of the remainder terms $R_{2p}(h)$, for different values of h.
- As an example, consider the following function

$$f_{p1}(x) = \begin{cases} \sin^2(2x)/(2\pi) & (0 \le x \le \pi)\,, \\ \\ \sin^4(2x)/(2\pi) & (\pi < x \le 2\pi)\,. \end{cases} \tag{6.26.6}$$

- The integral is

$$\begin{aligned} I_{p1} &= \int_0^{2\pi} f_{p1}(x)\,dx \\ &= \frac{1}{2\pi}\int_0^{\pi} \sin^2(2x)\,dx + \frac{1}{2\pi}\int_0^{2\pi} \sin^4(2x)\,dx \\ &= \frac{1}{2\pi}\int_0^{\pi} \frac{1-\cos(4x)}{2}\,dx + \frac{1}{2\pi}\int_{\pi}^{2\pi} \frac{3-4\cos(4x)+\cos(8x)}{8}\,dx \\ &= \frac{1}{4} + \frac{3}{16} = \frac{7}{16} = 0.4375\,. \end{aligned} \tag{6.26.7}$$

- The function $f_{p1}(x)$ is twice differentiable at $x = \pi$, and infinitely differentiable for all other values of x in the domain of integration.
- Hence $p = 1$ in the Euler–Maclaurin series in eq. (6.21.4).
- The function $f_{p1}(x)$ is periodic with period 2π, hence all the terms in the Euler–Maclaurin series vanish in eq. (6.21.4).
- The results are tabulated below for T_n (trapezoid rule) and Romberg first, second, third order.
- **Observe that the trapezoid rule T_n yields the exact answer 0.4375 for $n \ge 16$.**
- **However, the Romberg integration actually performs poorly and is worse than the trapezoid rule.**

j	n		T_n		$R(j,1)$		$R(j,2)$		$R(j,3)$
0	1		0						
1	2		0		0				
2	4		0		0		0		
3	8		0.5		0.666667		0.711111		0.722399
4	16		0.4375		**0.416667**		**0.4**		**0.395062**
5	32		0.4375		0.4375		**0.438889**		**0.439506**
6	64		0.4375		0.4375		0.4375		**0.437478**
7	128		0.4375		0.4375		0.4375		0.4375
8	256		0.4375		0.4375		0.4375		0.4375