Numerical Analysis

Daming Li

Department of Mathematics, Shanghai JiaoTong University,
Shanghai, 200240, China
Email: lidaming@sjtu.edu.cn

October 14, 2014

If the values of a function f are given at a few points, say $x_0, x_1, \cdots, x_n$, can that information be used to estimate a derivative $f'(c)$ or an integral $\int_a^b f(x)dx$? The answer is a qualified Yes.
The exact values of the derivative or the integral can be recovered only if f belongs to some relatively small family of function. For example, f is taken to be a polynomial of degree of n.
By Taylor expansion

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2}f''(\xi)$$

the first order derivative of f

$$f'(x) = \frac{f(x+h) - f(x)}{h} - \frac{h}{2}f''(\xi) \approx \frac{f(x+h) - f(x)}{h}$$

$-\frac{h}{2}f''(\xi)$ is called the truncation error, which arises at the some stage in the derivation. Thus the truncation error is $O(h)$.

Numerical Differentiation and Richardson Extrapolation

One problem is the subtractive cancellation in the numerical differentiation.

A better approximation is

$$f'(x) = \frac{f(x+h) - f(x-h)}{2h} - \frac{h^2}{6} f^{(3)}(\xi)$$

with the truncation error $-\frac{h^2}{6} f^{(3)}(\xi) = O(h^2)$. This can also be derived by Taylor expansion

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2} f''(x) + \frac{h^3}{6} f''(\xi_1)$$

$$f(x-h) = f(x) - hf'(x) + \frac{h^2}{2} f''(x) - \frac{h^3}{6} f'''(\xi_1)$$

$$f'(x) = \frac{f(x+h) - f(x-h)}{2h} - \frac{h^2}{12} [f'''(\xi_1) + f'''(\xi_2)]$$

An approximation of the second order derivative

$$f''(x) = \frac{f(x+h) - 2f(x) + f(x-h)}{h^2} - \frac{h^2}{12} f^{(4)}(\xi)$$

for some $\xi \in (x-h, x+h)$.

A general approach to numerical differentiation and integration can based on polynomial interpolation.

$$f(x) = \sum_{i=0}^{n} f(x_i) l_i(x) + \frac{1}{(n+1)!} f^{(n+1)}(\xi_x) w(x)$$

where $w(x) = \prod_{i=0}^{n} (x - x_i)$. Taking derivative in this formula,

$$f'(x) = \sum_{i=0}^{n} f(x_i) l_i'(x) + \frac{1}{(n+1)!} f^{(n+1)}(\xi_x) w'(x) + \frac{1}{(n+1)!} w(x) \frac{df^{(n+1)}(\xi_x)}{dx}$$

$$f'(x_\alpha) = \sum_{i=0}^{n} f(x_i) l_i'(x_\alpha) + \frac{1}{(n+1)!} f^{(n+1)}(\xi_x) \prod_{\alpha \neq j=0}^{n} (x_\alpha - x_j)$$

$$f'(x) = \frac{f(x+h) - f(x-h)}{2h} - \left[\frac{1}{3!}h^2 f^{(3)}(x) + \frac{1}{5!}h^4 f^{(5)}(x) + \frac{1}{7!}h^6 f^{(7)}(x) + \cdots\right]$$

I.e.,

$$L = \phi(h) + a_2 h^2 + a_4 h^4 + a_6 h^6 + \cdots$$

where L stands for $f'(x)$ and $\phi(h)$ stands for the numerical differentiation formula. Replacing h by $h/2$, we have

$$L = \phi(h/2) + a_2 h^2/4 + a_4 h^4/16 + a_6 h^6/64 + \cdots$$

which is called the Richardson extrapolation. Eliminating $a_2 h^2$ terms in the above two equality,

$$L = \frac{4}{3}\phi(h/2) - \frac{1}{3}\phi(h) - a_4 h^4/4 - 5a_6 h^6/16 - \cdots$$

which has accuracy $O(h^4)$.

Define

$$\psi(h) = \frac{4}{3}\phi(h/2) - \frac{1}{3}\phi(h)$$

Then we have

$$L = \psi(h) + b_4 h^4 + b_6 h^6 + \cdots$$

and one step of Richardson extrapolation

$$L = \psi(h/2) + b_4 h^4/16 + b_6 h^6/64 + \cdots$$

Eliminating the terms relating to b_4 in the above equalities,

$$L = \frac{16}{15}\psi(h/2) - \frac{1}{15}\psi(h) - b_6 h^6/20 - \cdots$$

with accuracy $O(h^6)$.

Select a convenient h (say $h = 1$) and compute

$$D(n, 0) = \phi(h/2^n), \quad n = 0, 1, \cdots, M$$

$$D(n, k) = \frac{4^k}{4^k - 1} D(n, k-1) - \frac{1}{4^k - 1} D(n-1, k-1)$$

for $k = 1, 2, \cdots, M$ and $n = k, k+1, \cdots, M$.

We can prove that

$$D(n, k-1) = L + \sum_{j=k}^{\infty} A_{jk} (h/2^n)^{2j} = L + O(h^{2k})$$

Proof. When $k = 1$,

$$D(n, 0) = \phi(h/2^n) = L - \sum_{j=1}^{\infty} a_{2j} (h/2^n)^{2j}$$

Thus we let $A_{j1} = -a_{2j}$. Now proceed by induction on k. We assume that the result is valid for some $k-1$, and on that basis we prove it for k.

$$\begin{aligned} D(n,k) &= \frac{4^k}{4^k-1}\Big[L+\sum_{j=k}^{\infty}A_{j,k}(\frac{h}{2^n})^{2j}\Big]-\frac{1}{4^k-1}\Big[L+\sum_{j=k}^{\infty}A_{j,k}(\frac{h}{2^{n-1}})^{2j}\Big] \\ &= L+\sum_{j=k}^{\infty}A_{j,k}\Big[\frac{4^k-4^j}{4^k-1}\Big](h/2^n)^{2j} \end{aligned}$$

Thus $A_{j,k+1}$ should be defined by

$$A_{j,k+1}=A_{j,k}\Big[\frac{4^k-4^j}{4^k-1}\Big]$$

$$\begin{pmatrix} D(0,0) & & & & \\ D(1,0) & D(1,1) & & & \\ D(2,0) & D(2,1) & D(2,2) & & \\ \vdots & \vdots & \vdots & \ddots & \\ D(M,0) & D(M,1) & D(M,2) & \cdots & D(M,M) \end{pmatrix}$$

Exercises: P.441

5,9,10,12,21

Integration Based on Interpolation

One powerful method for computing the integral

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

numerically by another function g that approximates f well and is easily integrated

$$\int_a^b f(x) \approx \int_a^b g(x)dx$$

Polynomials are good candidates for the function g, and indeed, g can be a polynomial that interpolates to f at a certain set of nodes. One method to obtain a polynomial is to truncate a Taylor series.

$$\int_0^1 e^{x^2} dx \approx \int_0^1 \sum_{k=0}^{n} \frac{x^{2k}}{k!} dx = \sum_{k=0}^{n} \frac{1}{(2k+1)k!}$$

However, it is desirable to have general procedures that require only evaluations of the integrand.

Integration Based on Interpolation

Methods based on interpolation fulfill this desideratum. Spline functions can also be used to interpolate f, and integrals of spline functions are easily computed.

Select nodes $x_0, x_1, \cdots, x_n$ in $[a, b]$ and define the Lagrange basis

$$l_i(x) = \prod_{i \neq j = 0}^{n} \frac{x - x_j}{x_i - x_j} \quad (0 \leq i \leq n)$$

The polynomial of degree $\leq n$ that interpolating f at the nodes is

$$p(x) = \sum_{i=0}^{n} f(x_i) l_i(x)$$

Then

$$\int_a^b f(x)dx \approx \int_a^b p(x)dx = \sum_{i=0}^{n} f(x_i) \int_a^b l_i(x)dx$$

Thus we have a formula of numerical integration

$$\int_a^b f(x)dx \approx \sum_{i=0}^{n} A_i f(x_i)$$

where

$$A_i = \int_a^b l_i(x)dx$$

The above numerical integration is called the Newton-Cotes Formula, if the nodes are equally spaced.

The trapezoid rule

$$\int_a^b f(x)dx \approx \frac{b-a}{2}[f(a)+f(b)]$$

It is exact for all $f \in \Pi_1$ (that is, polynomial of degree at most 1). Moreover, its error term is

$$-\frac{1}{12}(b-a)^3 f''(\xi)$$

This is determined by integrating the error term in the polynomial approximation $f(x) - P_1(x) = f''(\xi_x)(x-a)(x-b)/2$, and employing the Mean-Value Theorem for Integrals.

If the interval $[a, b]$ is partitioned like this

$$a = x_0 < x_1 < \cdots < x_n = b$$

the composite trapezoid rule

$$\int_a^b f(x)dx = \sum_{i=1}^{n} \int_{x_{i-1}}^{x_i} f(x)dx \approx \frac{1}{2} \sum_{i=1}^{n} (x_i - x_{i-1})[f(x_i) - f(x_{i-1})]$$

For the uniform spacing $h = (b-a)/n$ and $x_i = a + ih$, the composite trapezoid rule takes the form

$$\int_a^b f(x)dx \approx \frac{1}{2}\left[f(a) + 2\sum_{i=1}^{n-1} f(a+ih) + f(b)\right]$$

The error term for the composite trapezoid rule is

$$-\frac{1}{12}(b-a)h^2 f''(\xi)$$

The Simpson's rule

$$\int_a^b f(x)dx \approx \frac{b-a}{6}\left[f(a)+4f\left(\frac{a+b}{2}\right)+f(b)\right]$$

It is exact for all $f \in \Pi_3$. Its error term is

$$-\frac{1}{90}(b-a)^5 f^{(4)}(\xi)$$

Proof. If $h=(b-a)/2$, this numerical integration formula takes the form

$$\int_a^{a+2h} f(x)dx \approx \frac{h}{3}[f(a)+4f(a+h)+f(b)]$$

Using Taylor's theorem, the right-hand side can be written as

$$2hf(a)+2h^2f'(a)+\frac{4}{3}h^3f^{(2)}(a)+\frac{2}{3}h^4f^{(3)}(a)+\frac{100}{15}h^5f^{(4)}(a)+\cdots$$

The Taylor expansion of the left-hand side is

$$2hf(a)+2h^2f'(a)+\frac{4}{3}h^3f^{(2)}(a)+\frac{2}{3}h^4f^{(3)}(a)+\frac{32}{5!}h^5f^{(4)}(a)+\cdots$$

Combing these two expansions, we obtained the desired errors. The composite Simpson rule

$$\int_a^b f(x)dx = \sum_{i=1}^{n/2} \int_{x_{2i-1}}^{x_{2i}} f(x)dx \approx \frac{h}{3} \sum_{i=1}^{n/2} [f(x_{2i-2}) + 4f(x_{2i-1}) + f(x_{2i})]$$

Here we assume that n is even. The error term for the composite Simpson rule is

$$-\frac{1}{180}(b-a)h^4 f^{(4)}(\xi)$$

for some $\xi \in (a, b)$.

Integration Based on Interpolation

For any nodes x_i, the formula

$$\int_a^b f(x)dx \approx \sum_{i=0}^{n} A_i f(x_i)$$

where

$$A_i = \int_a^b l_i(x)dx$$

is exact for all polynomials of degree $\leq n$. On the other hand, If the above formula is exact for any polynomial of degree $\leq n$, then A_i must satisfies the above formula. This is because

$$\int_a^b l_i(x)dx = \sum_{i=0}^{n} A_i l_j(x_i) = A_j$$

This tell us how to determine the weight A_i. This method called the method of undermined coefficients.

For example, we want to find a formula

$$\int_0^1 f(x)dx \approx A_0 f(0) + A_1 f(\frac{1}{2}) + A_2 f(1)$$

which is exact for all polynomials of degree ≤ 2.

$$1 = \int_0^1 dx = A_0 + A_1 + A_2$$

$$1/2 = \int_0^1 x dx = \frac{1}{2} A_1 + A_2$$

$$1/3 = \int_0^1 x^2 dx = \frac{1}{4} A_1 + A_2$$

Thus $A_0 = 1/6$, $A_1 = 2/3$ and $A_2 = 1/6$.

Integration Based on Interpolation

The General integration formula

$$\int_a^b w(x)f(x)dx \approx \sum_{i=0}^{n} A_i f(x_i)$$

where w can be any fixed weight function. If this formula is exact for all polynomials of degree $\leq n$,

$$A_i = \int_a^b w(x) l_i(x) dx$$

From a formula for numerical integration on one interval, we can derive a formula for any other interval by making a linear change of variable. If the first formula is exact for polynomials of a certain degree, the same will be true of the second.

Suppose that a numerical integration formula is given:

$$\int_c^d f(t)dt \approx A_i f(t_i)$$

which is exact for all polynomials of degree $\leq m$. We want a numerical integration formula for some other interval, say $[a, b]$. Define

$$\lambda(t) = \frac{b-a}{d-c}t + \frac{ad-bc}{d-c}$$

and make the change of variable $x = \lambda(t)$,

$$\int_a^b f(x)dx = \frac{b-a}{d-c}\int_c^d dt f(\lambda(t)) \approx \frac{b-a}{d-c}\sum_{i=0}^{n} A_i f(\lambda(t_i))$$

Hence,

$$\int_a^b f(x)dx \approx \frac{b-a}{d-c}\sum_{i=0}^{n} A_i f\Big(\frac{b-a}{d-c}t_i + \frac{ad-bc}{d-c}\Big)$$

Integration Based on Interpolation

If p is the polynomial of degree $\leq n$ that interpolates f at $x_0, x_1, \cdots, x_n$,

$$f(x) - p(x) = \frac{1}{(n+1)!} f^{(n+1)}(\xi_x) \prod_{i=0}^{n} (x - x_i)$$

$$\int_a^b f(x)dx - \sum_{i=0}^{n} A_i f(x_i) = \int_a^b \frac{1}{(n+1)!} f^{(n+1)}(\xi_x) \prod_{i=0}^{n} (x - x_i) dx$$

If $|f^{(n+1)}(x)| \leq M$ on $[a, b]$, then

$$\left| \int_a^b f(x)dx - \sum_{i=0}^{n} A_i f(x_i) \right| \leq \frac{M}{(n+1)!} \int_a^b \left| \prod_{i=0}^{n} (x - x_i) \right| dx$$

The choice of nodes that makes the right-hand side of this inequality as small as possible is known to be

$$x_i = \frac{a+b}{2} + \frac{b-a}{2} \cos\left[\frac{(i+1)\pi}{n+2}\right] \quad (0 \leq i \leq n)$$

Integration Based on Interpolation

If the interval is $[-1, 1]$, these nodes have the simpler form

$$x_i = \cos\left[\frac{(i+1)\pi}{n+2}\right] \quad (0 \leq i \leq n)$$

These are the zeros of the function

$$U_{n+1}(x) = \frac{\sin[(n+2)\theta]}{\sin\theta} \quad (x = \cos\theta)$$

The function U_{n+1} is known as a Chebyshev polynomial of the second kind. For these nodes, a computation shows that

$$\int_{-1}^{1} |(x - x_0)(x - x_1)\cdots(x - x_n)| dx = 2^{-n}$$

Thus we have

$$\left| \int_a^b f(x)dx - \sum_{i=0}^{n} A_i f(x_i) \right| \leq \frac{M}{(n+1)!2^n}$$

Exercises: P.452

2,3,5,6,11,15,20

Gaussian Quadrature

We have give the quadrature formula

$$\int_a^b f(x)dx \approx \sum_{i=0}^{n} A_i f(x_i)$$

that are exact for polynomials of degree $\leq n$ where the nodes $x_0, \cdots, x_n$ are given. Is it possible to choose better nodes than these fixed nodes?

Consider a Gaussian quadrature rules

$$\int_a^b f(x)w(x)dx \approx \sum_{i=0}^{n} A_i f(x_i)$$

where w is a fixed positive weight function. Let this formula is exact for $f \in \Pi_n$, then

$$A_i = \int_a^b \prod_{i=0}^{n} \frac{x - x_j}{x_i - x_j} dx$$

Gaussian Quadrature

Let q be a nonzero polynomial of degree $n+1$ that is w-orthogonal to Π_n in the sense that for any $p \in \Pi_n$, we have

$$\int_a^b q(x)p(x)w(x)dx = 0$$

If $x_0, x_1, \cdots, x_n$ are zeros of q then the Gaussian quadrature formula will be exact for all $f \in \Pi_{2n+1}$.
Proof. Let $f \in \Pi_{2n+1}$. Divide f by q, obtaining a quotient p and remainder r

$$f = qp + r \quad (p, r \in \Pi_n)$$

So $f(x_i) = r(x_i)$

$$\int_a^b fwdx = \int_a^b rwdx = \sum_{i=0}^{n} A_i r(x_i) = \sum_{i=0}^{n} A_i f(x_i)$$

Let w be a positive weight function in $C[a, b]$. Let f a nonzero elements of $C[a, b]$ that is w-orthogonal to Π_n. Then f changes sign at least $n+1$ times on (a, b).

Proof. Since $1 \in \Pi_n$, we have $\int_a^b f(x)w(x)dx = 0$, and this shows that f changes sign at least once. Suppose that f changes sign only r times, with $r \le n$. Choose points t_i so that

$$a = t_0 < t_1 < t_2 < \cdots < t_r < t_{r+1} = b$$

and so that f is of one sign on each interval

$$(t_0, t_1), (t_1, t_2), \cdots, (t_r, t_{r+1})$$

The polynomial $p(x) = \prod_{i=1}^{r}(x - t_i)$ has the same sign property, and therefore $\int_a^b f(x)p(x)w(x)dx \neq 0$. Since $p \in \Pi_n$, this is a contradiction.

Gaussian Quadrature

In a Gaussian quadrature formula, the coefficients are positive, and moreover their sum is $\int_a^b w(x)dx$.

Proof. Fix n, and let q be a polynomial of degree $n+1$ that is w-orthogonal to Π_n. The zeros of q are denoted by $x_0, x_1, \cdots, x_n$, and these are the nodes in the Gaussian formula. Let p be the polynomial $q(x)/(x - x_j)$, for some fixed j. Since p^2 is of degree at most $2n$, the Gaussian formula will be exact for it. Consequently

$$0 < \int_a^b p^2(x)w(x)dx = \sum_{i=0}^{n} A_i p^2(x_i) = A_j p^2(x_j)$$

From this we conclude that $A_j > 0$. Since the Gaussian formula is exact for $f(x) \equiv 1$

$$\int_a^b w(x)dx = \sum_{i=0}^{n} A_i$$

Gaussian Quadrature

The Gaussian formula with error term:

$$\int_a^b f(x)w(x)dx = \sum_{i=0}^{n-1} A_i f(x_i) + \frac{f^{(2n)}(\xi)}{(2n)!}\int_a^b q^2(x)w(x)dx$$

where $a < \xi < b$ and $q(x) = \prod_{i=0}^{n}(x - x_i)$.

Proof. By Hermite interpolation, there is a polynomial p of degree at most $2n - 1$ such that

$$p(x_i) = f(x_i), \quad p'(x_i) = f'(x_i), \quad 0 \le i \le n-1$$

The error formula for this interpolation is

$$f(x) - p(x) = f^{(2n)}(\xi(x))q^2(x)/(2n)!$$

$$\int_a^b f(x)w(x)dx - \int_a^b p(x)w(x)dx = \frac{1}{(2n)!}\int_a^b f^{(2n)}(\xi(x))q^2(x)w(x)dx$$

$$\int_a^b p(x)w(x)dx = \sum_{i=0}^{n-1} A_i p(x_i) = \sum_{i=0}^{n-1} A_i f(x_i)$$

Exercises: P.462

1,8(a),10

Let trapezoid rule for I using n subintervals denoted by $T(n)$

$$T(n) = \frac{b-a}{n} \sum_{i=0}^{n}{}'' f\Big(a + i\frac{b-a}{n}\Big)$$

The double prime on the summation sign signifies that the first and last terms are to be halved. It is easy to prove that

$$T(2n) = \frac{1}{2}T(n) + h\sum_{i=1}^{n} f\Big(a + (2i-1)\frac{b-a}{2n}\Big)$$

$$T(2^n) = \frac{1}{2}T(2^{n-1}) + \frac{b-a}{2^n}\sum_{i=1}^{2^{n-1}} f\Big(a + (2i-1)\frac{b-a}{2^n}\Big)$$

Letting $R(n,0)$ denote the trapezoid estimate with 2^n subintervals,

$$\begin{cases} R(0,0) = \frac{1}{2}(b-a)[f(a)+f(b)] \\ R(n,0) = \frac{1}{2}R(n-1,0) + \frac{b-a}{2^n}\sum_{i=1}^{2^{n-1}} f\Big(a + (2i-1)\frac{b-a}{2^n}\Big) \\ R(n,m) = R(n,m) + \frac{1}{4^m-1}[R(n,m-1) - R(n-1,m-1)] \end{cases}$$

Using Richardson extrapolation, we obtain the Romberg algorithm.

$$\begin{pmatrix} R(0,0) & & & & \\ R(1,0) & R(1,1) & & & \\ R(2,0) & R(2,1) & R(2,2) & & \\ \vdots & \vdots & \vdots & \ddots & \\ R(M,0) & R(M,1) & R(M,2) & \cdots & R(M,M) \end{pmatrix}$$

To explain the Richardson extrapolation in the Romberg algorithm, we first note the Euler-Maclaurin formula

$$\int_0^1 f(t)dx = \frac{1}{2}[f(0)+f(1)]+\sum_{k=1}^{m-1} A_{2k}[f^{(2k-1)}(0)-f^{(2k-1)}(1)]-A_{2m}f^{(2m)}(\xi_0)$$

where $\xi_0 \in (0,1)$. A_k is called the Bernoulli numbers, which satisfies

$$\frac{x}{e^x - 1} = \sum_{k=0}^{\infty} A_k x^k$$

Using the Euler-Maclaurin formula for $\int_{x_i}^{x_{i+1}} f(x)dx$ and summing over k, we have

$$
\begin{aligned}
\int_a^b f(x)dx &= \frac{h}{2}\sum_{i=0}^{2^{n-1}}[f(x_i)+f(x_{i+1})] + \sum_{k=1}^{m-1} A_{2k}[f^{(2k-1)}(a) - f^{(2k-1)}(b)] \\
&\quad - A_{2m}(b-a)h^{2m}f^{(2m)}(\xi)
\end{aligned}
$$

for $\xi \in (a,b)$. Here we defined $h = (b-a)/2^n$. This formula tells us that

$$
I = R(n,0) + c_2h^2 + c_4h^4 + \cdots + c_{2m-2}h^{(2m-2)} + c_{cm}h^{2m}f^{(2m)}(\xi)
$$

It's Richardson extrapolation is just the Romberg algorithm.

If $f \in C[a,b]$ then each column in the Romberg array converges to the integral of f. Thus for each m,

$$\lim_{n\to\infty} R(n,m) = \int_a^b f(x)dx$$

Proof. We begin with column one, which contains trapezoidal estimates of the integral I. The trapezoid rule with k subintervals can be written as

$$R(n,0) = h\sum_{i=0}^{k} = \frac{h}{2}\sum_{i=0}^{k-1} f(a+ih) + \frac{h}{2}\sum_{i=1}^{k} f(a+ih)$$

The right side of this equation represents the average of two Riemann sums for I. Since $h = (b-a)/k$, $R(n,0)$ converges to I. As for the second column, we note that
$R(n,1) = \frac{4}{3}R(n,0) - \frac{1}{3}R(n-1,0)$ which also has a limit I. All subsequent columns can be analyzed in the same way.

Exercises: P.470

3,4,5