Trapezoidal rule

$$\int_a^b f(x)\,dx = \left[\frac{1}{2}f(x_0) + f(x_1) + f(x_2) + \ldots f(x_{N-2}) + f(x_{N-1}) + \frac{1}{2}f(x_N)\right]h + O(h^2)$$

The error is $O(h^2)$. "Big O notation". Simply means that for small enough h, the error goes as (is proportional to) h^2 (more strictly, $|\text{error}| < Ch^2$).

Keep in mind that we have obtained this from

$$\int_{x_{i-1}}^{x_i} f(x)\,dx = [f(x_{i-1}) + f(x_i)](x_i - x_{i-1})/2 + f''(\tilde{\xi}_i)\,O([x_i - x_{i-1}]^3).$$

That is, it assumes that the second derivative is finite everywhere.

A more detailed estimate of the error is $-\frac{(b-a)h^2}{12}f''(\xi)$.

Note that if the sign of f" is the same everywhere, the sign of the error is known. Generally, the integral of convex parts of the function is overestimated, and that of concave parts is underestimated.

Trapezoidal rule

$$\int_a^b f(x)\,dx = \left[\frac{1}{2}f(x_0) + f(x_1) + f(x_2) + \ldots f(x_{N-2}) + f(x_{N-1}) + \frac{1}{2}f(x_N)\right]h + O(h^2)$$

Interestingly, the error contains only even powers of h:

$$E = \int_a^b f(x)\,dx - \left[\frac{1}{2}f(x_0) + f(x_1) + f(x_2) + \ldots f(x_{N-2}) + f(x_{N-1}) + \frac{1}{2}f(x_N)\right]h = C_1 h^2 + C_2 h^4 + C_3 h^6 + \ldots$$

Of course, assumes that **all** derivatives exist and are finite.

Euler-Maclaurin formula gives the coefficients in terms of the function derivatives and Bernoulli numbers.

Another nice fact: if we used the trapezoidal rule with step h and now want to repeat with step h/2, we can reuse the result for step h and only do 1/2 as many summations as would be needed otherwise.

$$T(h) = \left[\frac{1}{2}f(x_0) + f(x_1) + f(x_2) + \ldots f(x_{N-2}) + f(x_{N-1}) + \frac{1}{2}f(x_N)\right]h$$

$$T(h/2) = \left[\frac{1}{2}f(x_0) + f(x_{1/2}) + f(x_1) + f(x_{3/2}) + \ldots + f(x_{N-3/2}) + f(x_{N-1}) + f(x_{N-1/2}) + \frac{1}{2}f(x_N)\right]h/2 =$$

$$T(h)/2 + [f(x_{1/2}) + f(x_{3/2}) + \ldots f(x_{N-3/2}) + f(x_{N-1/2})]h/2$$

$$I = T(h) + C_1 h^2 + C_2 h^4 + \ldots$$ *I* is the exact value of the integral

$$T(h) = I - C_1 h^2 - C_2 h^4 - \ldots$$

$$T(h/2) = I - C_1 (h/2)^2 - C_2 (h/2)^4 - \ldots$$

$$\frac{4}{3} T(h/2) - \frac{1}{3} T(h) = I + (C_4/4) h^4 + \ldots$$

So, while T(h) and T(h/2) are both 2nd order accurate, by combining them we get 4th order accuracy.

By combining T(h), T(h/2) and T(h/4) get 6th order, etc.

In general, this is known as Richardson extrapolation (also applicable to, e.g., numerical differentiation).

Specifically in this case, this is called Romberg integration. There is a nice algorithm doing this recursively (see, e.g, Numerical Recipes)

Trapezoidal rule is very accurate for periodic functions with the period equal to the integration range.

$$f(x) = a_0 + \sum_{n=1}^{\infty} \left[a_n \cos(2\pi n x/T) + b_n \sin(2\pi n x/T) \right]$$

Except for a_0, the integral is zero. So is the trapezoidal rule estimate, except when $n=kN$.

For a function whose mth derivative is discontinuous, the Fourier component amplitudes decay as a_n, $b_n \sim \nu^{-(m+1)}$.

$$\sum_{k=1}^{\infty} (kN)^{-(m+1)} \sim N^{-(m+1)} \sim h^{m+1}$$

For an infinitely differentiable function, the convergence is faster than any power law.

Midpoint rule

$$\int_a^b f(x)\,dx = [f(x_{1/2}) + f(x_{3/2}) + \ldots + f(x_{N-3/2}) + f(x_{N-1/2})]h + O(h^2)$$

Trapezoidal rule

$$\int_a^b f(x)\,dx = \left[\frac{1}{2}f(x_0) + f(x_1) + f(x_2) + \ldots f(x_{N-2}) + f(x_{N-1}) + \frac{1}{2}f(x_N)\right]h + O(h^2)$$

Midpoint rule has the same order of accuracy. Likewise, only even-order terms in the error: $E = A_1 h^2 + A_2 h^4 + A_3 h^6 + \ldots$

Cannot reuse the result when doubling the number of points, but can reuse when tripling.

Midpoint rule can be used directly when there is a singularity at the end(s), although the accuracy will be lower and it is better to transform the integrand to avoid the singularity, unless it is trivial, like sin x/x.

Not as accurate for periodic functions.

Rules of the midpoint type are called open, those of the trapezoidal type are called closed.

Often, names “midpoint” and “trapezoidal” are reserved to formulas for one subinterval only, and when subintervals are combined, they are called extended or composite midpoint and trapezoidal rules, respectively.

For the trapezoidal rule, we used piecewise-linear interpolation. Can produce higher-order rules by using higher-degree polynomials instead.

Use subintervals with 3 points in each – quadratic polynomials.

Subintervals $[x_{i-1},x_i]$, with points x_{i-1}, $x_{i-1/2}$, x_i. Consider the case when the points are equidistant, with the distance between adjacent points h/2.

For subinterval $[x_{i-1},x_i]$,

$$f(x)=2f(x_{i-1})\frac{(x-x_{i-1/2})(x-x_i)}{h^2}-4f(x_{i-1/2})\frac{(x-x_{i-1})(x-x_i)}{h^2}+2f(x_i)\frac{(x-x_{i-1})(x-x_{i-1/2})}{h^2}$$

$$+\frac{f'''(\xi_i)}{6}(x-x_{i-1})(x-x_{i-1/2})(x-x_i)$$

$$\int_{x_{i-1}}^{x_i}(x-x_{i-1/2})(x-x_i)dx=\int_{x_{i-1}}^{x_i}(x-x_{i-1})(x-x_{i-1/2})dx=h^3/12\,;\quad \int_{x_{i-1}}^{x_i}(x-x_{i-1})(x-x_i)dx=-h^3/6$$

$$\int_{x_{i-1}}^{x_i}f(x)dx=\frac{1}{6}[f(x_{i-1})+4f(x_{i-1/2})+f(x_i)]+O(h^4)$$

Combining all subintervals,

$$\int_a^b f(x)dx=\left[f(x_0)+4f(x_{1/2})+2f(x_1)+4f(x_{3/2})+2f(x_2)+\ldots+2f(x_{N-1})+4f(x_{N-1/2})+f(x_N)\right]\frac{h}{6}$$

$$+O(h^3)$$

Simpson rule

Simpson rule

$$\int_a^b f(x)\,dx = \left[f(x_0) + 4f(x_{1/2}) + 2f(x_1) + 4f(x_{3/2}) + 2f(x_2) + \ldots + 2f(x_{N-1}) + 4f(x_{N-1/2}) + f(x_N) \right] \frac{h}{6} + O(h^3)$$

Trapezoidal:

$$T(h) = \left[\frac{1}{2} f(x_0) + f(x_1) + f(x_2) + \ldots f(x_{N-2}) + f(x_{N-1}) + \frac{1}{2} f(x_N) \right] h$$

$$T(h/2) = \left[\frac{1}{2} f(x_0) + f(x_{1/2}) + f(x_1) + f(x_{3/2}) + \ldots + f(x_{N-3/2}) + f(x_{N-1}) + f(x_{N-1/2}) + \frac{1}{2} f(x_N) \right] h/2$$

Romberg procedure: $\frac{4}{3} T(h/2) - \frac{1}{3} T(h)$ – gives the Simpson rule

But the error was $O(h^4)$! So the $O(h^3)$ estimate above is too pessimistic!

Using Simpson directly is only very slightly faster, but doing Romberg also gives an error estimate and the process can stop or continue. Simpson cannot be reused in the same way as trapezoidal.

Can obtain even higher-order rules by using more points and higher-order polynomials. Some of them even have names.

Can obtain even higher-order rules by using more points and higher-order polynomials. Some of them even have names. The general name for rules with equidistant points is Newton-Cotes.

But these higher-order rules are rarely used. Remember that high-degree interpolation with equidistant points is often bad. Can be used with small intervals, but then the accuracy will probably be **too** high, and it is better to increase the accuracy gradually with Romberg or use other methods that I'll talk about later.

Dealing with infinite or semi-infinite intervals

$$\int_a^\infty f(x)\,dx\,; \quad \int_{-\infty}^\infty f(x)\,dx$$

3 basic approaches:

1) Replace the infinite limit with a finite one (cut off). Either throw away what's above the limit or approximate by something that can be integrated analytically. May work if the integrand decays very rapidly. If not, then to get a good approximation, the cutoff needs to be very high. Perhaps can divide into separate subranges and use different steps. There are 2 sources of error, so if you try to do extrapolation, you need to extrapolate in 2 variables.

2) Do a variable transformation. For instance, if x=a+(1-t)/t, then

$$\int_a^\infty f(x)\,dx = \int_0^1 f\left(a+\frac{1-t}{t}\right)\frac{1}{t^2}\,dt$$

This can introduce a singularity at 0 that needs to be dealt with.

3) Use some special methods I'll mention later.

Dealing with singularities

Functions with singularities at the ends of the integration range can be integrated using open rules, but convergence rate is poor. Can also be dealt with by a variable transformation.

For instance, if the integrand diverges at a as $(x-a)^{-\gamma}$, where $0<\gamma<1$, transform $x = a + t^{1/(1-\gamma)}$

$$\int_a^b f(x)\,dx = \int_0^{(b-a)^{1-\gamma}} t^{\gamma/(1-\gamma)} f\left(a + t^{1/(1-\gamma)}\right) dt$$

Similar transformations can also help if a higher derivative is singular. Also, can subtract and add a function that can be integrated analytically, but this approach is not as good.

If a singularity is in the middle, divide into subranges.

If the singularity position is unknown, perhaps some adaptive procedure.

Adaptive integration

Do integration with a particular h, then h/2. If the difference is too large, have to reduce further. However, it may be that the difference is concentrated in one part of the range. Calculate the difference for two halves separately, increase the number of points only where this is needed.

So far, considered methods based on interpolation using equidistant points. We know this is not optimal and higher-order methods may fail. We know that the optimal choice is Chebyshev nodes, which for [-1,1] are

$$x_j = \cos\left(\frac{\pi(2j-1)}{2N}\right), \quad j = 1,\ldots,N$$

This is known as Fejér quadrature. But a more popular choice is extrema (rather than roots) of Chebyshev polynomials:

$$x_j = \cos\left(\frac{\pi j}{N}\right), \quad j = 0,\ldots,N$$

The advantage is the possibility of reusing the values of the function. The method is called Clenshaw-Curtis quadrature.

As in all cases, $\int_{-1}^{1} f(x)\,dx \approx \sum_{j=0}^{N} w_j f(x_j)$. As it turns out, the weights can

be calculated by a Fourier transform.

http://www.sam.math.ethz.ch/~joergw/Papers/fejer.pdf

Gaussian quadratures

Chebyshev nodes are (almost) optimal for interpolation, but not necessarily for integration.

Recall that we have built our Newton-Cotes rules so that they are exact for linear (trapezoidal) or quadratic (Simpson) polynomials. In general, a method based on interpolation with N nodes should be exact for interpolating polynomials of degree N-1. If the integrand is a polynomial of degree ≤ N-1, then the rule should be exact for it.

So we can say that the rule using N points is built so that it is exact for polynomials of degree ≤ N-1, given the positions of the points, which are fixed. Essentially, we had N equations for N weights.

Let us now assume that the positions of the points are allowed to vary as well. For N points we now have 2N unknowns (N positions and N weights). So we can make the integral exact for polynomials of degree up to 2N-1.

A more general problem: choose N nodes and weights so that the integral $\int_a^b W(x) P_{2N-1}(x)\,dx$ is exact.

Consider a polynomial $Q_N(x)$ of order N, all roots of which are the nodes we are looking for. If we multiply it by an arbitrary polynomial $P_{N-1}(x)$ of order at most N-1, the resulting polynomial is of order at most 2N-1 and thus our scheme should be exact for it. But on the other hand, it still has zeros at the nodes, and thus our scheme gives zero. Thus,

$$\int_a^b W(x) Q_N(x) P_{N-1}(x)\,dx = 0$$

This is actually the definition of orthogonality.

For a given W(x), one can build a series of orthogonal polynomials $p_N(x)$, one of each order N, such that a given polynomial is orthogonal to all other polynomials of the series with weight W(x). Any polynomial of a lower order is a linear combination of polynomials of lower orders in the series, thus, if $Q_N(x) = p_N(x)$, it is guaranteed that the above condition is satisfied.

Weights $w_j = \int_a^b W(x) \prod_{i \neq j} \frac{x - x_i}{x_j - x_i}\,dx$

Example. Consider W(x) = 1 and the interval [-1,1].

$p_0(x)= 1$ $\quad p_1(x)=x+a. \quad \int_{-1}^{1} p_0(x)\,p_1(x)\,dx=\int_{-1}^{1} 1\times(x+a)\,dx=0\Rightarrow a=0.$

$$p_2(x)=x^2+bx+c. \quad \int_{-1}^{1} p_2(x)\,p_1(x)\,dx=\int_{-1}^{1}(x^2+bx+c)\,x\,dx=2\,b/3\Rightarrow b=0.$$

$$\int_{-1}^{1} p_2(x)\,p_0(x)\,dx=\int_{-1}^{1} x^2+c\,dx=2/3+2\,c=0\Rightarrow c=-1/3.$$

N = 1. Integration point at the root of p_1, which is obviously x = 0.

Weight - $\int_{-1}^{1} 1\,dx=2.$

$\int_{-1}^{1} f(x)\,dx\approx f(-1/\sqrt{3})+f(1/\sqrt{3}).$

$\int_{-1}^{1} f(x)\,dx\approx 2\,f(0)$ – midpoint rule

N = 2. Integration points at the roots of p_2: $\quad x^2-1/3=0\Rightarrow x_{1,2}=\pm 1/\sqrt{3}.$

Weights:

$$w_1=\int_{-1}^{1}\frac{x-x_2}{x_1-x_2}\,dx=-\frac{\sqrt{3}}{2}\int_{-1}^{1}(x-1/\sqrt{3})\,dx=1. \quad w_2=\int_{-1}^{1}\frac{x-x_1}{x_2-x_1}\,dx=\frac{\sqrt{3}}{2}\int_{-1}^{1}(x+1/\sqrt{3})\,dx=1.$$

$$\int_{-1}^{1} f(x)\,dx\approx f(-1/\sqrt{3})+f(1/\sqrt{3}).$$

For a given W(x), one can construct these series of orthogonal polynomials. There is always a recurrence relation

$$p_{j+1}(x)=(x-a_j)\,p_j(x)-b_j\,p_{j-1}(x),$$

where a_j and b_j are chosen so that p_j+1 is orthogonal to p_j and p_j-1.

In fact, for some frequently encountered W(x), the series of orthogonal polynomials are known and have names.

$W(x)=1,\ -1<x<1:$ Legendre polynomials $(j+1)P_{j+1}=(2j+1)xP_j-jP_{j-1}$

The corresponding integration method is called Gauss-Legendre quadrature

$W(x)=(1-x)^{\alpha}(1+x)^{\beta},-1<x<1:$ Jacobi polynomials

$W(x)=(1-x^2)^{-1/2},-1<x<1:$ familiar Chebyshev polynomials of the 1st kind

$$x_j=\cos\left(\frac{\pi(2j-1)}{2N}\right),\quad w_j=\frac{\pi}{N},\quad j=1,\ldots,N$$

$W(x)=(1-x^2)^{1/2},-1<x<1:$ Chebyshev polynomials of the 2nd kind

$$x_j=\cos\left(\frac{\pi j}{N+1}\right),\quad w_j=\frac{\pi}{N+1}\sin^2\left(\frac{\pi j}{N+1}\right),\quad j=1,\ldots,N$$

$W(x)=x^{\alpha}e^{-x},\ \ 0<x<\infty:$ Laguerre polynomials

$$(j+1)\,p_{j+1}=(-x+2\,j+\alpha+1)\,p_j-(j+\alpha)\,p_{j-1}$$

$W(x)=e^{-x^2},\ \ -\infty<x<\infty:$ Hermite polynomials

$$p_{j+1}=2\,xp_j-2\,jp_{j-1}$$

Can be used to do integrals with singularities and infinite limits.

In general, need to find roots numerically. Special methods for that – better than generic root-finding.

Obviously, the methods with W≠1 are applied to the whole integration range, not to subintervals, as Newton-Cotes. Not a problem, because they should always converge as the order increases.

In theory, should be superior to Clenshaw-Curtis. In practice, usually not the case.

Gauss-Kronrod rule

If one doubles the number of intervals in the Gauss-Legendre method (e.g., in order to do an error estimate), then there will be no points in common, so the values of the function cannot be reused. Add new points freely, but with the restriction that the old points remain in their places. If there were m points initially and then n new points are added, then there are m+2n free parameters, so the maximum degree of the polynomial for which the algorithm is exact is m+2n-1. This is lower than 2(m+n)-1, but for n≈m still higher than 2m-1.

If your function is represented as a set of data points, then interpolating with a cubic spline and integrating the resulting cubic polynomials analytically may be a sensible approach. But obviously, for integration it is not so important if the function is smooth.

Methods for oscillatory functions

If the integrand is a product of a slowly varying function and a rapidly oscillating function, there are ways to integrate with a step much larger than the period of the oscillation. E.g., Filon method.

Multidimensional integrals

If done straightforwardly (as repeated 1D integrals), computational cost grows exponentially with the number of dimensions. Monte Carlo methods avoid this, but accuracy is low. Sparse grids.

QUADPACK integration library http://www.netlib.org/quadpack/
Routines from QUADPACK are also included in GSL and NAG.

Mostly implement adaptive Gauss-Kronrod.