Numerical Quadrature

- *Quadrature* refers to any method for numerically approximating the value of a definite integral $\int_a^b f(x)dx$. The goal is to attain a given level of precision with the fewest possible function evaluations.

  The crucial factors that control the difficulty of a numerical integration problem are the dimension of the argument x and the smoothness of the integrand f.

- Any quadrature method relies on evaluating the integrand f on a finite set of points (called the *abscissas* or *quadrature points*), then processing these evaluations somehow to produce an approximation to the value of the integral. Usually this involves taking a weighted average.

  The goal is to determine which points to evaluate and what weights to use so as to maximize performance over a broad class of integrands.

- A natural strategy is to approximate f using a spline g with knots at a certain set of quadrature points. The integral $\int_a^b g(x)dx$ is easy to evaluate since it is a piecewise polynomial, and since g approximates f it makes sense to use $\int_a^b g(x)dx$ as an approximation to $\int_a^b f(x)dx$. When the quadrature points are evenly spaced the resulting quadrature rules are called *Newton-Cotes formulas*.

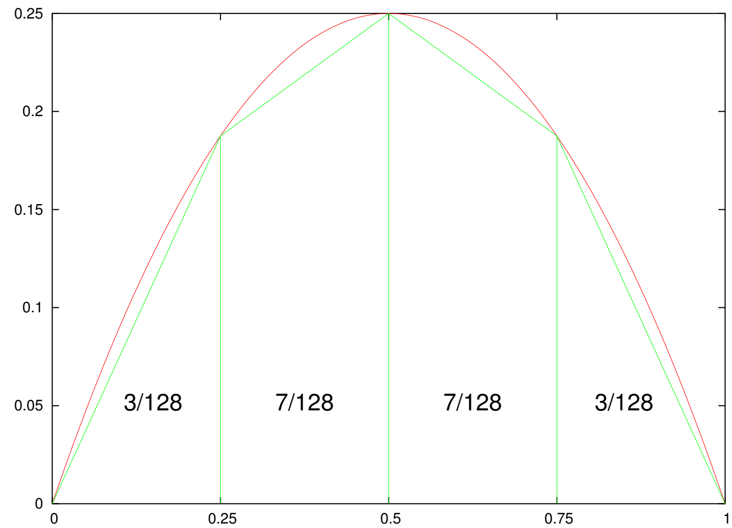
- Suppose we construct a grid on $[a, b]$, using mesh $h = (b - a)/m$, where m is an integer (the mesh is the distance between adjacent grid points). The grid points are $x_k = a + hk$. Let $f_k = f(x_k)$. Two straightforward rules are derived by approximating the integrand with linear or quadratic splines:

- *Trapezoidal rule:*

  The integral from x_k to x_{k+1} is $h(f_{k+1} + f_k)/2$. Therefore:

$$\int_a^b f(x)dx \approx h(\frac{1}{2}f_0 + f_1 + \ldots, +f_{m-1} + \frac{1}{2}f_m).$$

  The following shows the trapezoidal rule with $m = 4$ applied to $f(x) = x(1 - x)$ on $[0, 1]$. The approximate integral is $5/32$ while the exact value is $1/6$.



- *Simpson's rule:*

  Using Lagrange's formula, the quadratic interpolant through x_k, x_{k+1}, x_{k+2} can be shown to have integral $h(f_k + 4f_{k+1} + f_{k+2})/3$. Therefore:

$$\int_a^b f(x)dx \approx h(\frac{1}{3}f_0 + \frac{4}{3}f_1 + \frac{2}{3}f_2 + \ldots + \frac{4}{3}f_{m-1} + \frac{1}{3}f_m)$$

  Note that when applying Simpson's rule there must be an even number of intervals.

- Higher order (> 2) rules can be formulated, but are rarely used.

- In order to characterize the accuracy of these rules, we need to determine the accuracy of polynomial interpolation:

  Theorem: Let f be $n+1$ times continuously differentiable. Let $\tilde{f}$ be the degree n polynomial interpolation of f at $x_0, \ldots, x_n$. Then the error of interpolation at x can be written:

$$f(x) - \tilde{f}(x) = \frac{f^{(n+1)}(\eta)}{(n+1)!} \prod_j (x - x_j)$$

  for some $\min(x_j) \leq \eta \leq \max(x_j)$.

  Proof: Let $q(x) = \prod(x - x_j)$, and define g by:

$$g(y) \equiv f(y) - \tilde{f}(y) - q(y)\frac{f(x) - \tilde{f}(x)}{q(x)}.$$

Note that g is $n+1$ times continuously differentiable, and g vanishes at x and at $x_0, \ldots, x_n$. Therefore by Rolle's theorem g' has at least $n+1$ zeros. Repeated application of Rolle's theorem gives that $g^{(n+1)}$ has at least one zero, which we denote η. Evaluating $g^{(n+1)}$ at η gives the result.

Now we can derive the error estimate for the trapezoidal rule.

Theorem: Let f be twice continuously differentiable. Then for some $a \leq \eta \leq b$, the error for the trapezoidal rule is:

$$\int_a^b f(x)dx - \frac{b-a}{2}\left(f(a)+f(b)\right) = -\frac{(b-a)^3}{12} f''(\eta).$$

Proof: First observe that if $\tilde{f}$ denotes the linear interpolation of f on $[a,b]$, then the error can be represented as

$$\int_a^b f(x)dx - \frac{b-a}{2}\left(f(a)+f(b)\right) = \int_a^b \left(f(x) - \tilde{f}(x)\right) dx = \int_a^b (x-a)(x-b)\frac{f(x)-\tilde{f}(x)}{(x-a)(x-b)} dx.$$

The term $(x-a)(x-b)$ is non-positive, and the second factor is continuous. Therefore by the mean value theorem there exists η such that the above becomes equal to

$$\frac{f(\eta)-\tilde{f}(\eta)}{(\eta-a)(\eta-b)} \int_a^b (x-a)(x-b)dx.$$

The second term is equal to $-(b-a)^3/6$, and the first term is equal to $f''(\theta)/2$ for some $\theta \in [a,b]$, which gives the result.

- Applying this result, if we use mesh h then the error of the trapezoidal rule between successive abscissas decreases as h^3. Since there are L/h intervals, the total error decreases as h^2, or as $1/m^2$ if there are m abscissas. Thus in order to reduce the error by half, the number of quadrature points must be increased by approximately a factor of $\sqrt{2} \approx 1.4$.

- A direct extension of the theorem gives that the error of approximation for a degree n interpolant $\tilde{f}$ is bounded in magnitude by

$$\max_{a \leq u \leq b} \left| \frac{f^{(n+1)}(u)}{(n+1)!} \right| (b-a)^{n+2}.$$

Applying this to Simpson's rule ($n = 2$) gives that the error for a single quadratic interpolating polynomial on (a,b) decreases at order $(b-a)^4$. Thus the error for the interpolating spline with m abscissas decreases at order h^3, or as $1/m^3$ for m abscissas. Thus in order to reduce the error by half using Simpson's rule, the number of quadrature points must be increased by approximately a factor of $2^{1/3} \approx 1.26$.

- Surprisingly, Simpson's rule actually has error of magnitude $h^5 f^{(4)}(\eta)/90$ per interval, or $h^4 f^{(4)}(\eta)/90$ overall – an order better than calculated above.

  It is a general property that the order of the Newton-Cotes formula increases by two when moving from odd-order to even-order interpolations, and doesn't increase at all the other way around.

- Integration rules can be characterized in terms of the highest degree polynomial for which the error is zero. The trapezoidal rule is exact for degree 1 polynomials and Simpson's rule is exact for degree 2 polynomials – these statements are true because the interpolating polynomial is exactly equal to the integrand, $f \equiv \tilde{f}$.

  In fact, Simpson's rule is exact for cubics. To see this it is adequate to check that it holds for $f(x) = x^3$ (since Simpson's rule is additive and we already know that it is exact for quadratics). By direct evaluation $\int_0^1 x^3 dx = 1/4$, while Simpson's rule gives $(1 \cdot 0 + 4/8 + 1 \cdot 1)/6 = 1/4$ (in applying Simpson's rule the abscissas are $0, 1/2, 1$ and the mesh is $h = 1/2$).

  The reason for this is that the difference $f - \tilde{f}$, while not identically zero, has zero integral. The quadratic interpolant through $(0, 0)$, $(1/2, 1/8)$, $(1, 1)$ is

$$\tilde{f}(x) = 3x^2/2 - x/2$$

  and the pointwise error of interpolation is

$$\tilde{f}(x) - f(x) = 3x^2/2 - x/2 - x^3 = -(x - 1/2)^3 + (x - 1/2)/4 + 1/2$$

  which is an odd function with respect to $x = 1/2$, hence intergates to zero over any interval centered at $1/2$.

- For the trapezoidal rule, it is possible to descend through a sequence of meshes, each half the size of the previous mesh. This gives a sequence of approximations $A_n \rightarrow \int_a^b f(x)dx$ with mesh $(b - a)/2^n$.

  Note that this can be done efficiently without recomputing abscissas from the previous grid or even needing to save the individual function values, using

$$A_{n+1} = A_n/2 + (b - a)(\text{sum of new ordinates})/2^{n+1}.$$

  No such simple updating is available for Simpson's rule.

Unevenly spaced abscissas

- We can gain a lot of efficiency by using unevenly spaced abscissas. In particular, we can get exact results for polynomials of degree up to $2n-1$ with only n evaluations of f.

  Lemma 1: Let $w(x) > 0$ be integrable on $[a, b]$. There exists an orthogonal basis $\{q_0(x), \ldots, q_n(x)\}$ of monic polynomials with respect to $w(x)$, where $q_n(x)$ has degree n:

$$\int_a^b w(x) q_n(x) q_{n'}(x) dx \propto \mathcal{I}(n = n')$$

  Proof: Use the Gram-Schmidt procedure on the canonical basis $\{1, x, x^2, \ldots\}$.

  Lemma 2: Each $q_n(x)$ with $n > 0$ has n simple, real zeros.

  Proof: First observe that $q_0(x) \propto 1$, so for $n > 0$, $\int_a^b w(x) q_n(x) dx = 0$ by orthogonality. Therefore $q_n(x)$ has at least one real root. Let $x_1, \ldots, x_m$ denote the real roots of $q_n(x)$, so that $q_n(x) = g(x) \prod (x - x_j)^{d_j}$ where $g(x)$ has no real roots and the x_j are distinct. Let $r(x) = \prod_{\{\ell : d_\ell \text{ odd}\}} (x - x_\ell)$. If $r(x) = q_n(x)$ then we are done. Otherwise, $q_n(x) = g(x) r(x) h(x)$, where $h(x)$ has no real roots of odd multiplicity, and hence no sign changes. Since $r(x)$ has lower degree than $q_n(x)$, $\int_a^b w(x) r(x) q_n(x) = 0$ by orthogonality. But the integrand can be written $w(x) g(x) r(x)^2 h(x)$ which has no sign changes, which requires $g(x) r(x) h(x) = q_n(x)$ to be identically zero, which is a contradiction.

  Lemma 3: Let $x_0, \ldots, x_n$ be the roots of $q_{n+1}(x)$. Then there exists a unique set of numbers $a_0, \ldots, a_n$ such that:

$$\int_a^b w(x) q(x) dx = \sum_{j=0}^{n} a_j q(x_j)$$

  for all polynomials q of degree at most n.

  Proof: By linearity, and since $q_0, \ldots, q_n$ form a basis for the set of all polynomials of degree n, it is sufficient that the rule hold for this basis, that is,

$$\int_a^b w(x) q_k(x) dx = \sum_{j=0}^{n} a_j q_k(x_j) \quad k = 0, \ldots, n.$$

  Since $\int_a^b w(x) q_k(x) dx = 0$ for $k > 0$ by orthogonality, we can determine the a_j by solving the linear system

$$\begin{pmatrix} q_0(x_0) & \dots & q_0(x_n) \\ q_1(x_0) & \dots & q_1(x_n) \\ & \dots & \\ q_n(x_0) & \dots & q_n(x_n) \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \dots \\ a_n \end{pmatrix} = \begin{pmatrix} \int_a^b w(x)q_0(x)dx \\ 0 \\ \dots \\ 0 \end{pmatrix}.$$

It is a simple exercise to show that the coefficient matrix is non-singular, hence the solution exists.

Theorem: The values a_j and x_j defined in lemma 3 satisfy

$$\int_a^b w(x)p(x)dx = \sum_{j=0}^{n} a_j p(x_j),$$

for all polynomials $p(x)$ of degree at most $2n+1$.

Proof: Let $\tilde{p}(x)$ denote the degree n polynomial that agrees with $p(x)$ at $x_0, \dots, x_n$ (the roots of q_{n+1}). Write $p(x) = \tilde{p}(x) + \tilde{q}(x)$, and observe that since $\tilde{q}(x) = 0$ at $x_0, \dots, x_n$, we can write $\tilde{q}(x) = q_{n+1}(x)q(x)$ for $q(x)$ of degree n. Therefore

$$\begin{aligned} \int_a^b w(x)p(x)dx &= \int_a^b w(x)\tilde{p}(x)dx + \int_a^b w(x)q_{n+1}(x)q(x)dx \\ &= \int_a^b w(x)\tilde{p}(x)dx. \end{aligned}$$

By lemma 3,

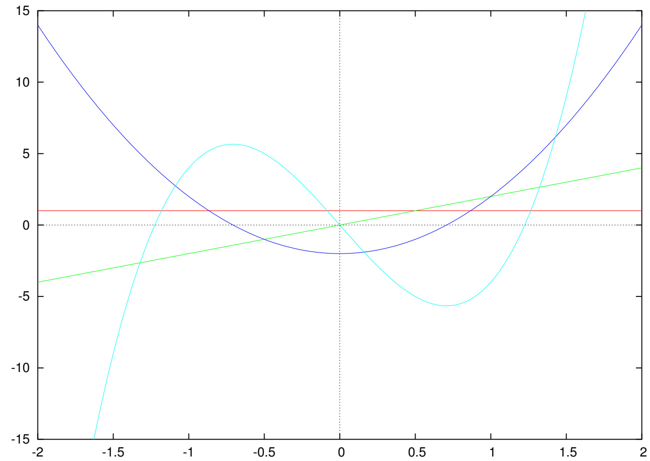
$$\int_a^b w(x)\tilde{p}(x)dx = \sum_{j=0}^{n} a_j \tilde{p}(x_j) = \sum_{j=0}^{n} a_j p(x_j),$$

which completes the proof.

Quadrature weight functions

- $w(x) = 1$ on $[a, b]$ (*Gauss-Legendre*)
- $w(x) = \sqrt{1-x^2}$ on $[-1, 1]$.
- $w(x) = 1/\sqrt{1-x^2}$ on $[-1, 1]$ (*Gauss-Chebyshev*)
- $w(x) = \exp(-x^2)$ on $(-\infty, \infty)$ (*Gauss-Hermite*)

- For example, the first four Guass-Hermite basis functions are shown below.



The procedure is used in two ways:

1. Integration of an arbitrary integrand:

$$\int_a^b g(x)dx = \int_a^b w(x)(g(x)/w(x))dx \approx \sum_j a_j g(x_j)/w(x_j)$$

2. Integration with respect to an explicit weight function, often a probability density function, e.g.

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} g(x) \exp(-x^2/2)dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} g(\sqrt{2}x) \exp(-x^2)dx \approx \frac{1}{\sqrt{\pi}} \sum_j a_j g(\sqrt{2}x_j).$$

- In the former case, the best results will be achieved if $g(x)/w(x)$ is approximately a low order polynomial. In the latter case, the best results will be achieved if $g(x)$ is approximately a low order polynomial.

- The error estimate for Gaussian quadrature is given by the following, for some $a \leq \eta \leq b$:

$$\int_a^b w(x)f(x)dx - \sum_{j=0}^{n} a_k f(x_k) = \frac{f^{(2n+2)}(\eta)}{(2n+2)!} \int_a^b w(x) q_{n+1}(x)^2 dx$$

- In general, the quadrature points x_j and weights w_j can be obtained by (i) searching for the roots, for instance using Newton's method, and (ii) solving a linear system for the weights. For the standard weights listed above, there are usually recurrence formulas that eliminate the second part. For instance, in the Gauss-Hermite case:

$$w_j = \frac{2^{n-1} n! \sqrt{\pi}}{n^2 H_{n-1}(x_j)^2},$$

where H_j is the j^{th} Gauss-Hermite polynomial (i.e. the j^{th} element of the basis described in Lemma 1 with $w(x) = \exp(-x^2)$). The following recurrences are useful when implementing the Newton search for the x_j:

$$H_{n+1}(x) = 2xH_n(x) - 2nH_{n-1}(x),$$

$$H_n'(x) = 2nH_{n-1}(x).$$

The squared norm of H_n (using the Gauss-Hermite weight function to determine the inner product) is $2^n n! \sqrt{\pi}$. Therefore the polynomials get large very quickly, so for numerical reasons it is essential to work with the normalized Gauss-Hermite polynomials. Letting $\tilde{H}_n(x)$ denote the normalized n^{th} Gauss-Hermite polynomial, we get the following recurrences:

$$\tilde{H}_{n+1}(x) = \sqrt{2/(n+1)} \cdot x\tilde{H}_n(x) - \sqrt{n/(n+1)}\tilde{H}_{n-1}(x).$$

$$\tilde{H}_n'(x) = \sqrt{2n}\tilde{H}_{n-1}(x).$$

- The one-dimensional quadrature rules discussed above can be directly generalized to higher dimensions. However as the dimension increases, the number of quadrature points grows geometrically. Consider the two-dimensional integral $\int f(x,y)dxdy$. We can integrate out one variable as follows:

$$\int f(x,y)dy \approx \sum_{j=0}^{n} w_j f(x, x_j).$$

Next we approximate the integral as

$$\int f(x,y)dxdy \approx \sum_{i=0}^{n} \sum_{j=0}^{n} w_i w_j f(x_i, x_j).$$

- There is a lot of work on *adaptive quadrature* that attempts to place the quadrature points efficiently in high dimensions. One strategy is to use more quadrature points in the coordinate directions where f is more variable. Another strategy is to reparametrize f to try to make it uniformly smooth in all directions.