Integrals in R^1 and their Approximation

One Dimensional Integration Let $f(x)$ be a continuous, or piecewise continuous, function defined on some universal interval $\mathcal{I}$ which includes the interval $[a, b]$. It will be familiar from elementary calculus that under these circumstances $\int_a^b f(x)\,dx$ is defined in terms of *Riemann sums*

$$\sum_{k=1}^{N} f(\xi_k)\,(x_k - x_{k-1})\,,$$

where, for a positive integer N, the x_k constitute a *partition* of the interval $[a,\, b]$: $P = \left\{ x_k \,\middle|\, k = 1, 2, ..., N \right\}$ with

$$a = x_0 < x_1 < \cdots < x_{k-1} < x_k < x_{k+1} < \cdots < x_{N-1} < x_N = b,$$

the *sample points* $\xi_k \in [x_{k-1},\, x_k]$ being arbitrary in each of these subintervals. The norm, $\|P\|$, of P is defined to be the largest of the numbers $x_k - x_{k-1}$, i.e., the length of the longest subinterval of $[a,\, b]$ defined by the partition. Using the calculus concept of *uniform continuity* one can show that as $\|P\| \to 0$, which entails $N \to \infty$ of course, these partitions approach a common limit, in terms of which the integral is defined:

$$\int_a^b f(x)\,dx \equiv \lim_{\|P\|\to 0} \sum_{k=1}^{N} f(\xi_k)\,(x_k - x_{k-1})\,.$$

So defined, the integral has some standard properties, including

Linearity: $\int_a^b (\alpha\, f(x) + \beta\, g(x))\,dx = \alpha \int_a^b f(x)\,dx + \beta \int_a^b g(x)\,dx,$
for two functions $f(x)$ and $g(x)$ as described and two scalars α and β. Further, we have the property of

Additivity: $\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx, \quad c \in \mathcal{I}.$

If $F(x)$ is an antiderivative for $f(x)$ the *Fundamental Theorem of Calculus* gives

$$\int_a^b f(x)\,dx \;=\; f(b) \;-\; F(a);$$

this is the standard way we calculate the integral if an antiderivative $F(x)$ can be found and expressed in terms of standard functions. However, it is often not possible to do this, necessitating the use of numerical procedures to approximate the integral. Perhaps the most convenient and accurate of these is the so-called *trapezoidal rule*, whereby, using the partition points x_k as described above, the integral is approximated by the sum

$$\int_a^b f(x)\,dx \approx \frac{1}{2}\sum_{k=1}^{N}\left(f(x_k)+f(x_{k-1})\right)(x_k - x_{k-1})\,,$$

reducing to $\frac{h}{2}\sum_{k=1}^{N}\left(f(x_k)+f(x_{k-1})\right)$ if the partition points are equally spaced with $h = (b-a)/N$.

It is natural to want to have some idea how accurate such an approximation procedure can be expected to be. In the case of equal spacing we have the following result.

Proposition 1 (Error Estimate in Use of the Trapezoidal Rule)
Let $f(x)$ be twice continuously differentiable on the interval $[a,\, b]$ of length $L = b-a$ and let $B = \max_{x\in[a,b]}\left\{|f''(x)|\right\}$. Let $N \geq L$ be a positive integer, $h = L/N \leq 1$. Then the error in the trapezoidal rule approximation, as shown above, is less than or equal to $\frac{2\,B\,L\,h^2}{3}$.

Proof Let $[c,\, d]$ be an arbitrary subinterval of $[a,\, b]$. Using integration by parts twice we have

$$\int_c^d f(x)\,dx \;=\; \int_c^d f(x)\cdot 1\,dx \;=\; f(x)(x-c)\Big|_c^d \;-\; \int_c^d f'(x)(x-c)\,dx$$

$$= f(d)(d-c) \; - \; f'(x)\frac{(x-c)^2}{2}\bigg|_c^d + \int_c^d f''(x)\frac{(x-c)^2}{2}\,dx$$

$$= f(d)(d-c) \; - \; f'(d)\frac{(d-c)^2}{2} \; + \int_c^d f''(x)\frac{(x-c)^2}{2}\,dx.$$

If we take $x-d$ as the antiderivative of 1 rather than $x-c$, much the same computation gives

$$\int_c^d f(x)\,dx \;=\; f(c)(d-c) \;+\; f'(c)\frac{(d-c)^2}{2} \;+ \int_c^d f''(x)\frac{(x-d)^2}{2}\,dx.$$

Adding the two results and dividing by 2 we obtain $\int_c^d f(x)\,dx =$

$$\frac{f(c)+f(d)}{2}(d-c) - \Big(f'(d) - f'(c)\Big)\,\frac{(d-c)^2}{2} + \int_c^d f''(x)\left(\frac{(x-c)^2+(x-d)^2}{2}\right)dx.$$

The first term here is the trapezoidal approximation for the single interval $[c,\, d]$; the sum of the last two terms forms the error in that approximation. We estimate these by

$$\left|\Big(f'(d) - f'(c)\Big)\,\frac{(d-c)^2}{2}\right| = \left|\int_c^d f''(x)\,dx\right|\,\frac{(d-c)^2}{2} \;\le\; B\,\frac{(d-c)^3}{2};$$

$$\left|\int_c^d f''(x)\left(\frac{(x-c)^2+(x-d)^2}{4}\right)dx\right| \;\le\; B\int_c^d \frac{(x-c)^2+(x-d)^2}{4}\,dx$$

$$= B\,\frac{(d-c)^3}{6}.$$

Thus

$$\left|\int_c^d f(x)\,dx - \frac{f(c)+f(d)}{2}(d-c)\right| \;\le\; B\left(\frac{(d-c)^3}{2} + \frac{(d-c)^3}{6}\right) = B\,\frac{2\,(d-c)^3}{3}.$$

Let the N equal subintervals of the proposition statement be $[x_{k-1},\, x_k]$, $k = 1, 2, ..., N$. Then, letting $[c,\, d]$ be replaced successively by these subintervals,

$$\left|\int_a^b f(x)\,dx \;-\; \sum_{k=1}^{N}\frac{f(x_{k-1})+f(x_k)}{2}(x_k - x_{k-1})\right|$$

$$\leq\ B \sum_{k=1}^{N} \frac{2(x_k - x_{k-1})^3}{3}\ \leq\ \frac{B\,L}{h}\left(\frac{2\,h^3}{3}\right)\ =\ \frac{2\,B\,L\,h^2}{3}$$

and the proposition is proved.

In general this estimate is very conservative, i.e., the trapezoidal estimates tend to be much better than the error estimate would lead one to think.

Example It is familiar that $\int_0^\pi sin(x)\,dx\ =\ 2$. Suppose we were to try to approximate this integral via the trapezoidal rule with $N\ =\ 30,\ \ h\ =\ \pi/30$. Since we have $B\ =\ 1,\ \ L\ =\ \pi$ here, the estimate of the Proposition 1 would lead one to expect an error $\leq\ \frac{2\,\pi}{3}\left(\frac{\pi}{30}\right)^2\ =\ .0238$. However, the trapezoidal rule estimates the integral as

$$\frac{\pi}{60}\sum_{k=1}^{30}\left(\sin\left(\frac{(k-1)\pi}{30}\right)+\sin\left(\frac{k\pi}{30}\right)\right)\ =\ 1.9982$$

corresponding to an error .0018, more than ten times smaller.

There are a large number of integral approximation methods; some more accurate than the above, including *Simpson's Rule* which, for even N and equal spacing, can be stated as

$$\int_a^b f(x)\,dx \approx \frac{h}{3}\sum_{j=1}^{N/2}\Big(f(x_{2j-2})\ +\ 4\,f(x_{2j-1})\ +\ f(x_{2j})\Big).$$

The error involves the fourth derivative of $f(x)$ and is proportional to h^4. Here we content ourselves with the trapezoidal rule; it is very convenient and readily adapted to other types of integrals.

Trapezoidal Rule for Arclength and Line Integrals Let $\mathcal{C}$ be a curve in a region $\mathcal{R} \subset R^n$, parametrized as

$$\mathcal{C}:\ \ X\ =\ X(t),\ \ a\ \leq\ t\ \leq\ b,$$

with endpoints $A = X(a)$, $B = X(b)$, and oriented from A to B. If $f(X)$ is a scalar-valued function defined in $\mathcal{R}$, the integral of $f(X)$, over $\mathcal{C}$, with respect to arclength s can be approximated in at least two ways by the trapezoidal rule. If we introduce a partition of $\mathcal{C}$ consisting of points $X_0 = X(a) = A,\ X_k = X(t_k),\ k = 1, 2, ..., N-1,\ X_N = X(b) = X(t_N)$, where $a = t_0 < t_1 < ... < t_{k-1} < t_k < t_{k+1} < ... < t_N = b$, we have, since $\frac{ds}{dt} = \|X'(t)\|$,

$$\begin{array}{ccc} \int_{\mathcal{C}} f(X)\, ds & \approx & \frac{1}{2}\sum_{k=1}^{N}\Big(f(X_k) + f(X_{k-1})\Big)\|X_k - X_{k-1}\| \\ \| & & \\ \int_a^b f(X(t))\|X'(t)\|\, dt & \approx & \frac{1}{2}\sum_{k=1}^{N} \tilde{f}_k\,(t_k - t_{k-1}), \end{array}$$

where $\tilde{f}_k = f(X(t_k))\|X'(t_k)\| + f(X(t_{k-1}))\|X'(t_{k-1})\|$.

If $F(X)$ is a continuous vector field defined on the region $\mathcal{R}$, the line integral of $F(X)$ over $\mathcal{C}$ also has two approximations based on the trapezoidal rule:

$$\begin{array}{ccc} \int_{\mathcal{C}} F(X)^* dX & \approx & \frac{1}{2}\sum_{k=1}^{N}\Big(F(X_k) + F(X_{k-1})\Big)^*(X_k - X_{k-1}) \\ \| & & \\ \int_a^b F(X(t))^* X'(t)\, dt & \approx & \frac{1}{2}\sum_{k=1}^{N} \tilde{f}_k\,(t_k - t_{k-1}), \end{array}$$

where now $\tilde{f}_k = F(X(t_k))^* X'(t_k) + F(X(t_{k-1}))^* X'(t_{k-1})$.

For both of these integrals the approximating expressions on the right hand sides will not, in general, be exactly equal.

Example We have seen in the section on Green's theorem that the area inside a simple closed curve $\mathcal{C}$ in the plane is given by $\frac{1}{2}\int_{\mathcal{C}} \begin{pmatrix} -y \\ x \end{pmatrix} \cdot \begin{pmatrix} dx \\ dy \end{pmatrix}$. Let us take $\mathcal{C}$ to be the unit circle and take $X_k = \begin{pmatrix} \cos(2\,k\,\pi/N) \\ \sin(2\,k\,\pi/N) \end{pmatrix}$, $k = 0, 1, 2, ..., N$. Using the first shown form of the trapezoidal rule for the line integral shown above we have the area approximated by

$$\frac{1}{4}\sum_{k=1}^{N}\begin{pmatrix}-\sin(2(k-1)\,\pi/N)-\sin(2\,k\,\pi/N)\\ \cos(2(k-1)\,\pi/N)+\cos(2\,k\,\pi/N)\end{pmatrix}\cdot\begin{pmatrix}\cos(2\,k\,\pi/N)-\cos(2(k-1)\,\pi/N)\\ \sin(2\,k\,\pi/N)-\sin(2(k-1)\,\pi/N)\end{pmatrix}$$

$$= \frac{1}{4}\sum_{k=1}^{N} 2\Big(\sin(2\,k\,\pi/N)\,\cos(2(k-1)\pi/N)-\sin(2(k-1)\pi/N)\,\cos(2\,k\,\pi/N)\Big)$$

$$= \tfrac{N}{2}\,\sin(2\,\pi/N) \;\approx\; \pi \text{ for large } N.$$