Math 107H

Topics for the first exam: addendum

Numerical Integration

Sometimes (most times?) the Fundamental Theorem of Calculus won't help us to compute a definite integral; we can't find an antiderivative. So we need to fall back on the definition:

$$\sum_{i=1}^{n} f(c_i)\Delta x_i \text{ approximates } \int_a^b f(x)\ dx,$$

where the interval $[a, b]$ is cut into n pieces of length $\Delta x_1, \ldots \Delta x_n$, and c_i lies in the i-th subinterval

Typically, for convenience, we choose the subintervals to have the same length $\Delta x_i = \Delta x = \frac{b-a}{n}$, and make "standard" choices of elements in the i-th subinterval $[x_{i-1}, x_i]$:

$L(f, n) = \sum_{i=1}^{n} f(x_{i-1})\Delta x$ (left endpoint estimate)

$R(f, n) = \sum_{i=1}^{n} f(x_i)\Delta x$ (right endpoint estimate)

$M(f, n) = \sum_{i=1}^{n} f(\frac{x_{i-1}+x_i}{2})\Delta x$ (midpoint estimate)

Of these, the midpoint estimate is probably best; $L(f, n)$ overestimates area when f is decreasing and underestimates it when f is increasing; $R(f, n)$ does the opposite. $M(f, n)$ tends to average these effects out. In fact, if we know that f'' doesn't get too large, say $|f''(x)| \leq K$ on $[a, b]$, then

$$|\int_a^b f(x)\ dx - M(f, n)| \leq K\frac{(b-a)^3}{24n^2}$$

In the end though, a midpoint estimate is throwing out a lot of information, since it approximates f on an interval by a constant. We can do better, taking into account more infmation about the function f, by approximating f by functions that better "fit" f on a subinterval, whose integrals we know how to compute.

The first is linear functions: we replace f on each subinterval by the linear function having the same values at the endpoints. This essentially replaces a rectangle in our sums with trapezoids. Since the area of a trapezoid is (length of base)(average of lengths of heights), we end up with the estimate

$T(f, n) = \sum_{i=1}^{n} \frac{f(x_{i-1})+f(x_i)}{2}\Delta x = \frac{1}{2}(\sum_{i=1}^{n} f(x_{i-1})\Delta x + \sum_{i=1}^{n} f(x_i)\Delta x)$
$= \frac{1}{2}(L(f, n) + R(f, n))$ (trapezoid estimate)

If f is close to being linear on each subinterval (i.e., f'' is not too big), this gives a better estimate of the integral than either of L or R alone. In fact, if $|f''(x)| \leq K$ on $[a, b]$, then

$$|\int_a^b f(x)\ dx - T(f, n)| \leq K\frac{(b-a)^3}{12n^2}$$

Not quite as good as we expect from midpoints, but it leads us to further impovements. Because: we expect we can do even better if we approximte f by "better" functions, e.g., quadratics!

To set this up better, we assume we cut $[a, b]$ into an even number $2n$ of subintervals, so $\Delta x = \frac{b-a}{2n}$. Then we deal with the subintervals in pairs, i.e., with endpoints three at a time:

$$x_{2i}, x_{2i+1} = x_{2i} + \Delta x, x_{2i+2} = x_{2i} + 2\Delta x.$$

There is exactly one quadratic function $g(x) = ax^2 + bx + c$ which takes the same value as f at these three points, and by plugging in those values at $x_{2i}, x_{2i} + \Delta x, x_{2i} + 2\Delta x$ we get three equations in three unknowns (a, b, and c), which we can solve to determine the quadratic g. This makes a good quadratic approximation to f on the interval $[x_{2i}, x_{2i+2}]$. But the real point is that we know how to integrate $\int_{x_{2i}}^{x_{2i+2}} g(x)\ dx$ exactly, since it is a quadratic, and a little arithmetic shows that this integral is equal to

$$\int_{x_{2i}}^{x_{2i+2}} g(x)\ dx = \frac{\Delta x}{3}(f(x_{2i}) + 4f(x_{2i+1}) + f(x_{2i+2}))$$

If we sum up these quantities, for each of the n pairs of intervals we have cut $[a, b]$ into, we get *Simpson's Rule*: for $\Delta x = (b - a)/2n$,

$$S(f, n) = \frac{\Delta x}{3}[f(x_0) + 4f(x_1) + 2f(x_2) + 4f(x_3) + 2f(x_4) + \cdots + 2f(x_{2n-2}) + 4f(x_{2n-1}) + f(x_{2n})]$$

It is an amazing fact that this estimate gives the precisely correct integral if f is quadratic <u>or</u> cubic. In fact, how good the estimate is depends on how much the third derivative of f is changing, i.e., on how big the fourth derivative is: if $|f''''(x)| \leq M$ on the interval $[a, b]$, then

$$|\int_a^b f(x)\ dx - S(f, n)| \leq K\frac{(b-a)^5}{180n^4}$$

So, typically, using twice as many intervals (i.e., doing twice the work) gives us an estimate about 16 times closer to the real value of the integral.

The importance of these estimates of the error is that they give us a means to decide beforehand how many subintervals to work with, in order to guarantee that our estimate is within some pre-determined error of the actual value of the integral. Note that, in some sense, every one of these estimates is computed as (length of subinterval)(sum of values of f, one for each subinterval), but some values are weighted more heavily than others. But <u>on average</u> the weight given to a value is one. The trapezoid rule chooses to take half of the values of both endpoints (instead of just one or the other, to avoid playing favorites), and Simpson's Rule gives the middle endpoint of a pair of subintervals twice as much weight as the endpoints.