Math 6500 Additional Material: The Lanczos Generalized Derivative

Numerical differentiation is the process of finding a formula which approximates the derivative of the function $f(x)$ from a collection of values $y_i = f(x_i)$, keeping in mind that our measurements of those values may be corrupted by errors. We have seen that one way to do this is to use Taylor series to come up with an approximation formula, and to use Richardson extrapolation to improve these approximations. Another is to proceed as follows: if we have a lot of x_i and y_i, we may find a least-squares fit of a linear function $g(x; a, b) = ax + b$ to our data points, and take a as our approximation of the derivative.

The least-squares procedure is to adjust the parameters a, b to minimize the sum of the squares

$$E(a,b) = \sum_i (g(x_i; a, b) - y_i)^2$$

In our case, we must minimize

$$E(a,b) = \sum_i (ax_i + b - y_i)^2$$

over all possible choices of a and b. At whatever values of a and b yield the minimizer, we must have

$$0 = \frac{\partial}{\partial a} E(a,b) = \sum_i 2(ax_i + b - y_i)x_i.$$

Now if we suppose that there are an odd number of x_i symmetrically spaced around 0, we can renumber them as $x_{-k}, \ldots, x_0 = 0, \ldots x_k$ with $x_{-i} = -x_i$ and observe that our right hand sum simplifies to

$$0 = \sum_{i=-k}^{k} 2x_i(ax_i - y_i)$$

as the sum $\sum_{i=-k}^{k} 2bx_i$ vanishes by symmetry. Notice that the sum $\sum_{i=-k}^{k} 2x_i y_i$ does *not* vanish by symmetry as the y_i are presumably not symmetric around 0. Solving for a, we get

$$a = \frac{\sum_{i=-k}^{k} x_i y_i}{\sum_{i=-k}^{k} x_i^2}.$$

Now suppose we have a lot of data points in a small interval $[-h, h]$. In the limit, if $e(x)$ is a Riemann-integrable function, we have

$$\lim_{k\to\infty} \sum_{i=-k}^{k} \frac{h}{k} e\left(\frac{h}{k} i\right) = \int_{-h}^{h} e(x)\, dx.$$

This means that if we assume that the unknown function $f(x)$ is Riemann integrable, let $x_i = \frac{hi}{k}$, recall $y_i = f(x_i)$, and take the limit as $k \to \infty$, we get the estimate

$$\begin{aligned} f'(0) &\sim \lim_{k\to\infty} \frac{\frac{h}{k}\sum_{i=-k}^{k} x_i y_i}{\frac{h}{k}\sum_{i=-k}^{k} x_i^2} \\ &= \frac{\int_{-h}^{h} x f(x)\,dx}{\int_{-h}^{h} x^2\,dx} \\ &= \frac{1}{\frac{h^3}{3} - \frac{(-h)^3}{3}} \int_{-h}^{h} x f(x)\,dx \\ &= \frac{3}{2h^3}\int_{-h}^{h} x f(x)\,dx. \end{aligned}$$

At an arbitrary point x, the formula becomes

$$f'(x) \sim \frac{3}{2h^3}\int_{-h}^{h} t f(x+t)\,dt.$$

and this formula defines the *Lanczos generalized derivative*

$$f_g'(x) = \lim_{h\to 0} \frac{3}{2h^3}\int_{-h}^{h} t f(x+t)\,dt.$$

The generalized derivative can be used both as a numerical and a theoretical tool. Here is the minihomework for this material:

1. Prove that if f is differentiable at x, $f_g'(x) = f'(x)$. Hint: L'Hospital's rule.

2. Find the error term in the numerical approximation formula:

$$f'(x) \sim \frac{3}{2h^3}\int_{-h}^{h} t f(x+t)\,dt.$$

   (Possibly bogus hint: Try expanding $f(x+t)$ as a Taylor series and integrating term-by-term.)

3. The function $f(x) = |x|$ is not differentiable at $x = 0$. Prove that $f_g'(0)$ exists by computing it.

4. (Extra credit) Prove that if the left derivative $f_l'(x) = \lim_{h\to 0^-} \frac{f(x+h)-f(x)}{h}$ and the right derivative $f_r'(x) = \lim_{h\to 0^+} \frac{f(x+h)-f(x)}{h}$ exist, then the generalized derivative $f_g'(x)$ exists and is equal to $\frac{1}{2}(f_l'(x) + f_r'(x))$.