Numerical Methods

Math 3338 – Spring 2022

Worksheet 9

Differentiation

Reading

CP	5.10
NMEP	5.1, 5.2

Table 1: Sections Covered

1 Derivatives

As you should know, the standard definition the derivative

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \tag{1}$$

is inadequate as there can be numeric error in the division. There actually isn't anything we can do about this.

Let's do some Taylor series. Recall,

$$\begin{aligned} f(x+h) &= f(x) + hf'(x) + \frac{h^2}{2} f''(x) + \frac{h^3}{6} f'''(x) + \cdots \\ f(x-h) &= f(x) - hf'(x) + \frac{h^2}{2} f''(x) - \frac{h^3}{6} f'''(x) + \cdots . \end{aligned}$$

Subtract these and solve for $f'(x)$ to see,

$$f'(x) = \frac{f(x+h) - f(x-h)}{2h} + \mathcal{O}(h^2). \tag{2}$$

This is called a *central difference* and it's order h^2 which is better than order h, like (1)[1]. Central differences are more accurate using larger values of h.

There are also *forward differences*, this is (1), and *backward differences*. You should be able to guess what a backward difference is, but just in case,

$$\frac{f(x) - f(x-h)}{h}$$

You'll primarily use forward and backward differences when dealing with sequential data. If your data looks like $x_0, x_1, x_2, \ldots$, you can't use central differences to evaluate the derivative at x_0, you must use a forward difference.

2 Second Derivatives and Higher Order

When posed with a problem, the only logical solution is Taylor series. The idea with be to expand $f(x+ih)$ for numbers i (preferably integers) and solve for $f^{(n)}(x)$ canceling lower order derivatives.

[1] Why is (1) order h?

For example, using the following expansions,

$$
\begin{aligned}
f(x+h) &= f(x) + hf'(x) + \frac{h^2}{2}f''(x) + \frac{h^3}{6}f'''(x) + \cdots \\
f(x-h) &= f(x) - hf'(x) + \frac{h^2}{2}f''(x) - \frac{h^3}{6}f'''(x) + \cdots \\
f(x+2h) &= f(x) + 2hf'(x) + \frac{4h^2}{2}f''(x) + \frac{8h^3}{6}f'''(x) + \cdots \\
f(x-2h) &= f(x) - 2hf'(x) + \frac{4h^2}{2}f''(x) - \frac{8h^3}{6}f'''(x) + \cdots .
\end{aligned}
$$

We can find that,

$$
\begin{aligned}
f''(x) &= \frac{f(x+h) - 2f(x) + f(x-h)}{h^2} + \mathcal{O}(h^2) \\
f^{(3)}(x) &= \frac{f(x+2h) - 2f(x+h) + 2f(x-h) - f(x-2h)}{2h^3} + \mathcal{O}(h^2) \\
f^{(4)}(x) &= \frac{f(x+2h) - 4f(x+h) + 6f(x) - 4f(x-h) + f(x-2h)}{h^4} + \mathcal{O}(h^2).
\end{aligned}
$$

These are all central differences, similar formulas exists for noncentral differences.

3 The Problem with Derivatives

We'll always be dividing by small numbers. This is especially true for $f^{(n)}(x)$. Combine this with,

1. Most problems being able to be phrased in terms of an integral
2. We can easily evaluate all explicit derivatives

and we tend to not use differentiation that frequently. There are situations where differentiation is essential. But that's a future problem.

4 Optimization

Finding the max or min of a function is quite important. This is relatively easy to do with calculus[2], but can we do it numerically? Yes, of course. We are going to focus on minimums, maximums are identical just different[3].

Consider $f(x) = x^2$. This has a minimum at $x = 0$. Let's find this numerically. Pick a starting point, let's say $x_0 = 2$. It turns out $x_1 = x_0 - \alpha \cdot f'(x_0)$ where α is a chosen number (sometimes called the "learning rate" for reasons), if $\alpha = .1$ then,

$$
x_1 = 2 - .1 \cdot (2 \cdot 2) = 2 - .4 = 1.6.
$$

This moved us closer to 0!

Why does this work? If $f'(x) > 0$ then the function will have a minimum to the left of x and if $f'(x) < 0$ the min will be to the right. That's about it. The parameter α controls how large a step we take, if it's too big we'll keep jumping over the minimum and if it's too small we won't go anywhere. Right now you should draw x^2 on your paper and verify what you read in this paragraph.

[2] For a single value function.
[3] So not identical?

Numerical Methods

Math 3338 – Spring 2022

Homework 9 (Due: Tuesday, February 15)

Problem 1 (1 pt) Write two functions, `diff` and `diff_2` for the first and second derivatives using central differences. The input of these should be `f,x,h=.01` where f is a function.

Problem 2 (1 pt) Given $f(x) = 3x^4 + 3x^3 - 2x + 1$ find $f'(x)$ and $f''(x)$ by hand. Use your programs to compare the exact answers with approximation for $x \in \{-5, -4, \ldots, 4, 5\}$. Make a table and put in a PDF.

Problem 3 (1 pt) Given $f(x) = \frac{1}{1+e^{-x}}$ find $f'(x)$ and $f''(x)$ by hand. Use your programs to compare the exact answers with approximation for $x \in \{-5, -4, \ldots, 4, 5\}$. Make a table and put in a PDF.

Also make a graph with f, f', f'' on the same axes. Label them accordingly. You should use your `diff` and `diff_2` programs for this.

Problem 4 (1 pt) Write a function called `gradient_descent` with inputs `f, a, alpha, tol=1e-9, max_iter=500` that approximates a minimum of $f(x)$ with starting point a and learning rate alpha. The tol value is a tolerance, if $|x_n - x_{n+1}| <$ tol then you can conclude x_n is a minimum. max_iter is the maximum number of iterations (to prevent an infinite loop). In `diff` use `h=1e-9`.

Return a tuple `(x_n, points)` where x_n is the x-value of the min and `points` is a list containing all the points generated by your process.

Test your function with x^2 and a few starting points.

Problem 5 (1 pt) Run `gradient_descent` with $f(x) = x^2$, $a = 2$. Answer each of these questions in a PDF (with your graphs) for $\alpha \in [.1, .5, 1]$.

1. How many iterations did the algorithm use?
2. How many of those iterations had values less than .1?
3. What do you think explains the large number of steps with small values?
4. How could we have fixed this issue?

Problem 6 (1 pt) Make a graph of $f(x) = x^2$ from -2 to 2 and the points you found in the previous problem for $\alpha = .1$. You should have a line (x^2) with a bunch of dots on it.