Numerical Optimization 02: Derivatives

Qiang Zhu

University of Nevada Las Vegas

July 14, 2020

Overview

Derivative

The goal of optimization is to find the point that minimizes an objective function. Knowing how the value of a function changes (derivative) is useful.

$$f(x + \Delta x) \approx f(x) + f'(x)\Delta x$$

$$f'(x) = \frac{\Delta f(x)}{\Delta x}$$

Derivatives in multiple dimensions

$$\text{Jacobian} \qquad \nabla f(x) = \left[\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \cdots \frac{\partial f(x)}{\partial x_n}\right]$$

$$\text{Hessian} \qquad \nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ & \vdots & \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix}$$

Numerical Differentiation

For practical applications, we rely on numerical methods to evaluate the derivatives.

- Finite Difference Methods

$$f`(x) \approx \begin{cases} \frac{f(x+h)-f(x)}{h} & \text{forward} \\ \frac{f(x+h/2)-f(x-h/2)}{h} & \text{central} \\ \frac{f(x)-f(x-h)}{h} & \text{backward} \end{cases}$$

- Complex Step Method

$$f`(x) = \mathrm{imag}(f(x+ih)/h)$$

Finite Difference - forward

$$f(x+h) = f(x) + \frac{f`(x)}{1!}h + \frac{f``(x)}{2!}h^2 + \frac{f```(x)}{3!}h^3 + \cdots$$

Finite Difference - forward

$$f(x+h) = f(x) + \frac{f'(x)}{1!}h + \frac{f''(x)}{2!}h^2 + \frac{f'''(x)}{3!}h^3 + \cdots$$

We can arrange it to

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

Therefore, forward difference has linear error.

Finite Difference - central

$$f(x+h/2) = f(x) + \frac{f`(x)}{1!}\frac{h}{2} + \frac{f``(x)}{2!}(\frac{h}{2})^2 + \frac{f```(x)}{3!}(\frac{h}{2})^3 + \cdots$$
$$f(x-h/2) = f(x) - \frac{f`(x)}{1!}\frac{h}{2} + \frac{f``(x)}{2!}(\frac{h}{2})^2 - \frac{f```(x)}{3!}(\frac{h}{2})^3 + \cdots$$
$$f`(x) = \frac{f(x+h/2) - f(x-h/2)}{h} + O(h^2)$$

Therefore, central difference has quadratic error.

Complex Step

According to Taylor expansion,

$$f(x+ih) = f(x) + ihf`(x) - h^2\frac{f``(x)}{2!} - ih^3\frac{f```(x)}{3!} + \cdots$$

Complex Step

According to Taylor expansion,

$$f(x+ih) = f(x) + ihf'(x) - h^2\frac{f''(x)}{2!} - ih^3\frac{f'''(x)}{3!} + \cdots$$

If we take only the imaginary part,

$$\mathrm{Im}(f(x+ih)) = hf'(x) - h^3\frac{f'''(x)}{3!} + \cdots$$

$$f'(x) = \frac{\mathrm{Im}(f(x+ih))}{h} + h^2\frac{f'''(x)}{3!} - \cdots = \frac{\mathrm{Im}(f(x+ih))}{h} + O(h^2)$$

Complex Step

According to Taylor expansion,

$$f(x+ih) = f(x) + ihf`(x) - h^2\frac{f``(x)}{2!} - ih^3\frac{f```(x)}{3!} + \cdots$$

If we take only the imaginary part,

$$\mathrm{Im}(f(x+ih)) = hf`(x) - h^3\frac{f```(x)}{3!} + \cdots$$

$$f`(x) = \frac{\mathrm{Im}(f(x+ih))}{h} + h^2\frac{f```(x)}{3!} - \cdots = \frac{\mathrm{Im}(f(x+ih))}{h} + O(h^2)$$

While the real part is

$$\mathrm{Re}(f(x+ih)) = f(x) - h^3\frac{f``(x)}{2!} + \cdots$$

$$f(x) = \mathrm{Re}(f(x+ih)) + O(h^2)$$

Complex Step

According to Taylor expansion,

$$f(x+ih) = f(x) + ihf'(x) - h^2\frac{f''(x)}{2!} - ih^3\frac{f'''(x)}{3!} + \cdots$$

If we take only the imaginary part,

$$\mathrm{Im}(f(x+ih)) = hf'(x) - h^3\frac{f'''(x)}{3!} + \cdots$$

$$f'(x) = \frac{\mathrm{Im}(f(x+ih))}{h} + h^2\frac{f'''(x)}{3!} - \cdots = \frac{\mathrm{Im}(f(x+ih))}{h} + O(h^2)$$

While the real part is

$$\mathrm{Re}(f(x+ih)) = f(x) - h^3\frac{f''(x)}{2!} + \cdots$$

$$f(x) = \mathrm{Re}(f(x+ih)) + O(h^2)$$

The complex step method is advantageous since

- Both $f(x)$ and $f'(x)$ can be evaluated in a single run
- $f'(x)$ has a quadratic error

Comparison

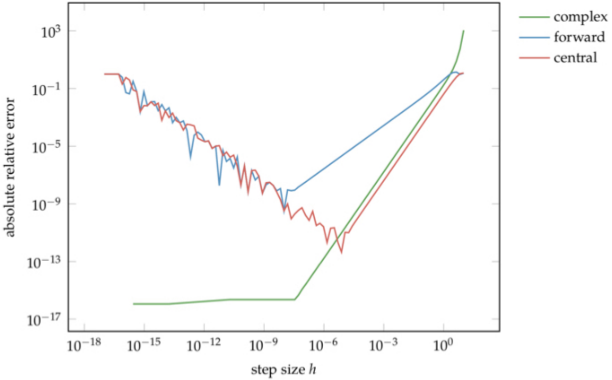


Figure 2.4. A comparison of the error in derivative estimate for the function $\sin(x)$ at $x = 1/2$ as the step size is varied. The linear error of the forward difference method and the quadratic error of the central difference and complex methods can be seen by the constant slopes on the right hand side. The complex step method avoids the subtractive cancellation error that occurs when differencing two function evaluations that are close together.

Homework: reproduce the above figure by yourself!

Why is complex step better than central difference?

$$f(x + ih) = u(x, h) + iv(x, h)$$
$$\mathrm{Im}(f(x + ih)) = h\frac{\partial v(x, y)}{\partial y}|_{y=0} + O(h^2)$$

If $v(x, 0) = 0$, $f(x) = u(x, 0)$. Dividing by h, we have f'(x)

$$\frac{\partial v(x, y)}{\partial y}|_{y=0} = \frac{\partial u(x, 0)}{\partial x}$$

The left side is what we used to compute the complex step. The right side is due to Cauchy-Riemann equation, which is used by the finite difference. Note that two method use two different functions (u and v).

Why is complex step better than central difference?

Consider a function $f(z) = z^2$,

$$f(z) = z^2 = x^2 - y^2 + i2xy$$

The finite difference does the function x^2, while the complex step gives $2x$, in the case for any $h = y > 0$.

Why is complex step better than central difference?

Consider a function $f(z) = z^2$,

$$f(z) = z^2 = x^2 - y^2 + i2xy$$

The finite difference does the function x^2, while the complex step gives $2x$, in the case for any $h = y > 0$.

Try another function:

$$\cos(x + iy) = \cos(x)\cosh(y) - i\sin(x)\sinh(y).$$

The imaginary part is $\sin(x)\sinh(y)$. For a small y, it gives -sin(x).

Dual numbers

Dual numbers can be expressed mathematically by including the abstract quantity ϵ, where ϵ^2 is 0. So that,

$$(a + b\epsilon) + (c + d\epsilon) = (a + c) + (b + d)\epsilon$$
$$(a + b\epsilon) * (c + d\epsilon) = (ac) + (ad + bc)\epsilon$$

The function's evaluation and derivative can be expressed simultaneously in an exact manner.

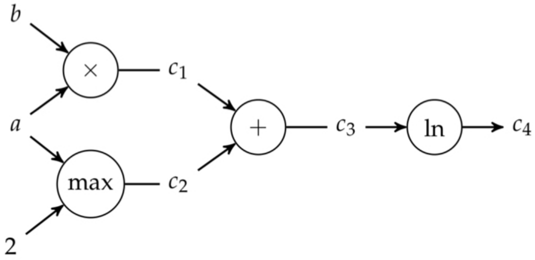
$$\begin{aligned} f(x) &= \sum_{k=0}^{\infty} \frac{f^k(a)}{k!}(x-a)^k \\ f(a+b\epsilon) &= \sum_{k=0}^{\infty} \frac{f^k(a)}{k!}(a+b\epsilon-a)^k = \sum_{k=0}^{\infty} \frac{f^k(a)b^k\epsilon^k}{k!} \\ &= f(a) + bf`(a)\epsilon + \epsilon^2 \sum_{k=2}^{\infty} \frac{f^k(a)b^k}{k!}\epsilon^{k-2} \\ &= f(a) + bf`(a)\epsilon \end{aligned}$$

Express a function as the computational graph

Suppose we have a target function

$$f(a,b) = \ln(ab + \max(a,2))$$

It can be expressed as



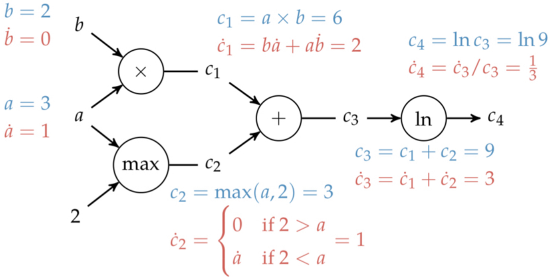
The derivative from the computational graph

Suppose we have a target function

$$f(a,b) = \ln(ab + \max(a,2))$$

The derivative is

$$\frac{df}{dx} = \frac{df}{dc_4}\frac{dc_4}{d_x} = \frac{df}{dc_4}\left(\frac{dc_4}{dc_3}\frac{dc_3}{dx}\right) = \frac{df}{dc_4}\left(\frac{dc_4}{dc_3}\left(\frac{dc_3}{dc_2}\frac{dc_2}{dx} + \frac{dc_3}{dc_1}\frac{dc_1}{dx}\right)\right)$$



Summary

- Derivatives are important for optimization.
- We rely on numerical derivatives in practical optimization
- Finite differences are the most easy ways to compute derivative
- The complex step method has better accuracy
- Dual numbers allow the exact evaluation of function and derivative simultaneously
- Analytic differentiation methods include forward and reverse accumulation on computational graphs