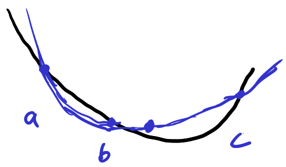
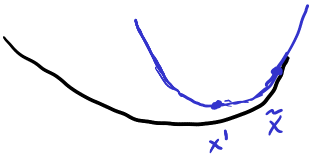
- Golden Section Search — linear convergence
- Schubert-Pyavskii
- Quadratic-Fit Search

Newton's Method





$f(\tilde{x}),\ f'(\tilde{x}),\ f''(\tilde{x})$

$$q(x) = f(\tilde{x}) + f'(\tilde{x})(x-\tilde{x}) + f''(\tilde{x})\frac{(x-\tilde{x})^2}{2}$$

$$\frac{\partial q}{\partial x} = 0 = f'(\tilde{x}) + (x'-\tilde{x})f''(\tilde{x})$$

$$\boxed{x' = \tilde{x} - \frac{f'(\tilde{x})}{f''(\tilde{x})}}$$

Informally, Newton's method <u>converges quadratically</u> near a smooth local minimum.

For an interval $I = [x^*-\delta,\ x^*+\delta]$ where

$f''(x) \neq 0 \quad \forall\, x \in I$

$f'''(x)$ continuous on I

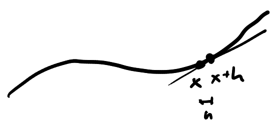
$\exists\, c$ s.t. $\frac{1}{2}\left|\frac{f'''(x^{(1)})}{f''(x^{(1)})}\right| < c \left|\frac{f'''(x^*)}{f''(x^*)}\right|$

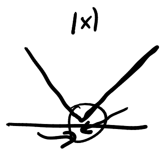
Then $\exists\, \gamma > 0$ s.t. $|x^* - x^{(k+1)}| \leq \gamma\, |x^* - x^{(k)}|^2$

Review of vector Calculus

Derivative

$$\frac{d}{dx}f(x) = f'(x) \equiv \lim_{h\to 0}\frac{f(x+h)-f(x)}{h}$$





Partial Derivative

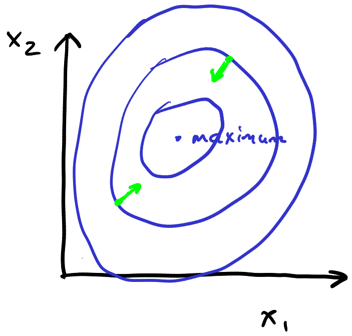
unit vector in direction i

$$\frac{\partial}{\partial x_i}f(\vec{x}) = \lim_{h\to 0}\frac{f(\vec{x}+h\hat{e}_i)-f(\vec{x})}{h}$$

Gradient

$f: \mathbb{R}^n \to \mathbb{R}$

$$\nabla f(\vec{x}) = g \equiv \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$$



Jacobian

$f: \mathbb{R}^n \to \mathbb{R}^m$

$$\begin{array}{cc} & n \\ m & \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} \end{array}$$

Hessian

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

symmetric if the second derivatives are all continuous in a neighborhood

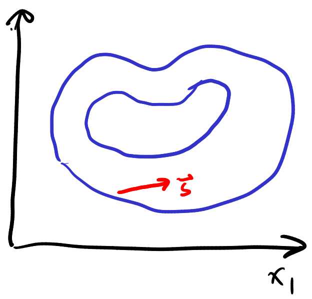
Directional Derivative

direction $\vec{s}$ (often a unit vector)

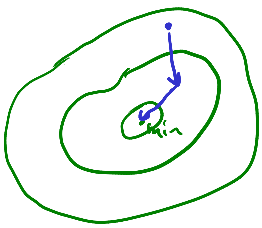
$$\nabla_{\vec{s}} f(\vec{x}) \equiv \lim_{h \to 0} \frac{f(\vec{x} + h\vec{s}) - f(\vec{x})}{h}$$

"instantaneous rate of change in f when moving at velocity $\vec{s}$.

N+W: $D_{\vec{s}}$



$$\nabla_{\vec{s}} f(\vec{x}) = \nabla f(\vec{x}) \cdot \vec{s}$$



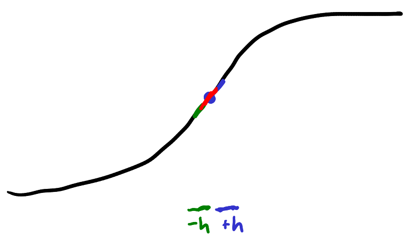
Calculating Derivatives

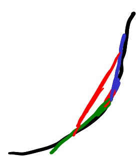
1. Symbolic
2. Numerical
3. Automatic

Numerical

Finite Difference

$$f'(x) \approx \underbrace{\frac{f(x+h) - f(x)}{h}}_{\text{forward}} \approx \underbrace{\frac{f(x) - f(x-h)}{h}}_{\text{backward}} \approx \underbrace{\frac{f(x+\frac{h}{2}) - f(x-\frac{h}{2})}{h}}_{\text{central}}$$





Errors

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

$$-f'(x)\,h = f(x) - f(x+h) - \frac{f''(x)}{2!}h^2 + \dots$$

$$\underbrace{f'(x) = \frac{f(x+h)-f(x)}{h}} + \underbrace{\frac{f''(x)}{2!}h + \dots}_{\text{error}}$$

$O(h)$ for forward and reverse

$$f\left(x+\tfrac{h}{2}\right) = \cancel{f(x)} + f'(x)\frac{h}{2} + \cancel{\frac{f''(x)}{2!}\left(\frac{h}{2}\right)^2} + \frac{f'''(x)}{3!}\left(\frac{h}{2}\right)^3 + \dots$$

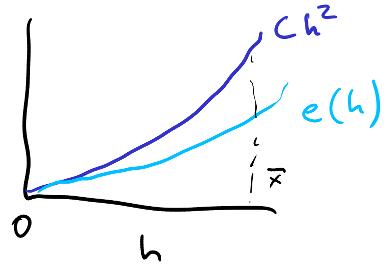
$$f\left(x-\tfrac{h}{2}\right) = \cancel{f(x)} - f'(x)\frac{h}{2} + \cancel{\frac{f''(x)}{2!}\left(\frac{h}{2}\right)^2} - \frac{f'''(x)}{3!}\left(\frac{h}{2}\right)^3 + \dots$$

$$f\left(x+\tfrac{h}{2}\right) - f\left(x-\tfrac{h}{2}\right) = 2f'(x)\frac{h}{2} + \frac{2}{3!}f'''(x)\left(\frac{h}{2}\right)^3 + \dots$$

$$\underbrace{f'(x) = \frac{f\left(x+\frac{h}{2}\right) - f\left(x-\frac{h}{2}\right)}{h}} - \underbrace{\frac{f'''(x)}{24}h^2 \dots}_{\text{error}}$$

$O(h^2)$ for central difference

$e(h) \in O(h^2)$ as $h \to 0$ $\qquad \exists\, C, \bar{x} > 0$ s.t. $\forall\, h < \bar{x}$ $\; e(h) < C h^2$



estimate $f'(x)$ by evaluating $f(x+hi)$

Complex-Step

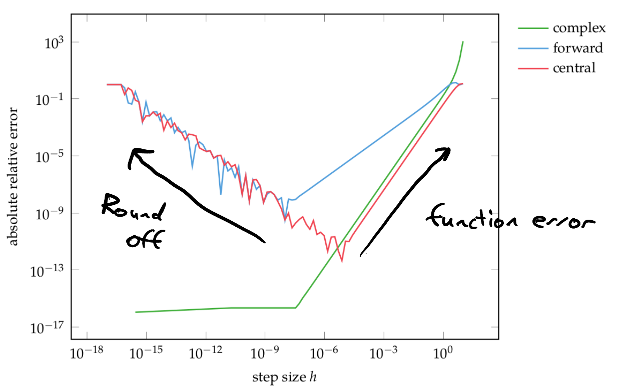
$$f(x+hi) = f(x) + ihf'(x) - h^2\frac{f''(x)}{2!} - i\frac{h^3 f'''(x)}{3!} \dots$$

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

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

$$f'(x) \simeq \frac{\mathrm{Im}(f(x+ih))}{h}$$

$$f(x) \simeq \mathrm{Re}(f(x+ih))$$



Automatic Differentiation

$i^2 = -1$

Dual Numbers

$$a + b\varepsilon \qquad \varepsilon^2 = 0$$

$$f(a+b\varepsilon) = \sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!}(a+b\varepsilon-a)^k$$

$$= \sum_{k=0}^{\infty} \frac{f^{(k)}(a)\, b^k \varepsilon^k}{k!}$$

$$= f(a) + bf'(a)\varepsilon + \cancelto{0}{\varepsilon^2 \sum_{k=2}^{\infty} \frac{f^{(k)}(a)\, b^k}{k!} \varepsilon^{(k-2)}}$$

$$f(a+b\varepsilon) = f(a) + bf'(a)\varepsilon$$

$$f'(a) = Du\left[f(a+\varepsilon)\right]$$

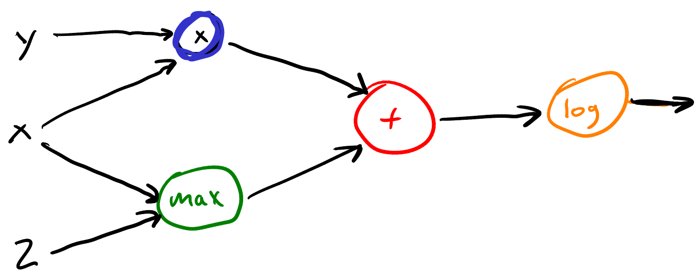
$Du \equiv$ dual part

everything multiplied by ε

$$f(a+\varepsilon) = f(a) + f'(a)\varepsilon$$

Computational Graphs + Chain Rule

$$f(x,y) = \log(xy + \max(x,2))$$



Chain Rule

$$\frac{\partial}{\partial x} f(g(x)) = \frac{\partial f}{\partial g}\frac{\partial g}{\partial x}$$

$$f(x,y) = \underbrace{\log}_{f}\underbrace{(xy + \max(x,2))}_{g}$$

$$\frac{\partial f}{\partial x} = \frac{\partial \log(g)}{\partial g}\frac{\partial g}{\partial x}$$

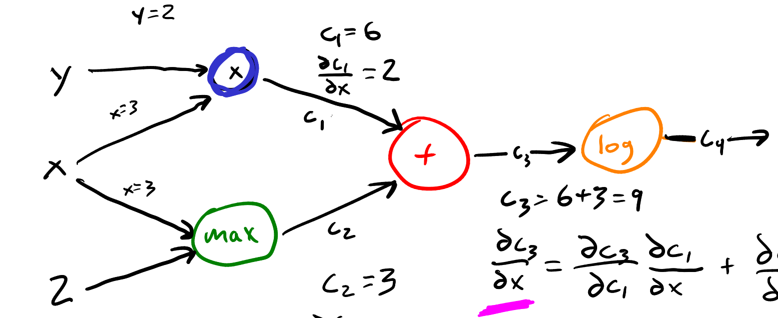
$$= \frac{1}{xy + \max(a,2)} \frac{\partial}{\partial x}(xy + \max(x,2))$$

$$= \quad '' \quad \left(1\frac{\partial(xy)}{\partial x} + 1\frac{\partial \max(x,2)}{\partial x}\right)$$

$$'' \quad (y + 1(x \geq 2) + 0(x<2))$$

$$= \frac{y + (x \geq 2)}{xy + \max(a,2)}$$

$\frac{\partial f}{\partial x}\Big|_{x=3,\ y=2}$



$c_1 = 6$, $\frac{\partial c_1}{\partial x} = 2$

$c_2 = 3$, $\frac{\partial c_2}{\partial x} = 1$

$c_3 = 6+3 = 9$

$$\frac{\partial c_3}{\partial x} = \frac{\partial c_3}{\partial c_1}\frac{\partial c_1}{\partial x} + \frac{\partial c_3}{\partial c_2}\frac{\partial c_2}{\partial x} = 1\cdot 2 + 1\cdot 1 = 3$$

n-dimensional

$c_4 = \log(9) \approx 2.197$

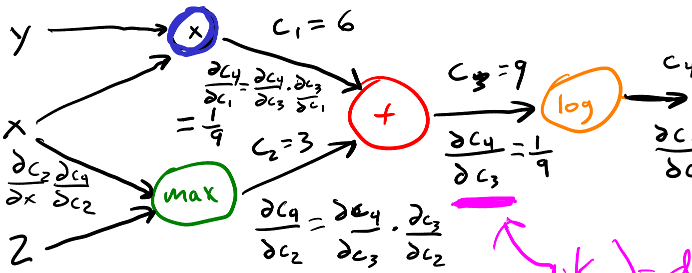
$$\boxed{\frac{\partial c_4}{\partial x}} = \frac{\partial c_4}{\partial c_3}\frac{\partial c_3}{\partial x} = \frac{1}{c_3}\, 3 = \frac{1}{9}\cdot 3 = \boxed{\frac{1}{3}}$$

Forward Mode

Reverse Mode

$y=2$

$x=3$



$c_1 = 6$

$\frac{\partial c_4}{\partial c_1} = \frac{\partial c_4}{\partial c_3}\cdot\frac{\partial c_3}{\partial c_1} = \frac{1}{9}$

$c_2 = 3$

$c_3 = 9$, $\frac{\partial c_4}{\partial c_3} = \frac{1}{9}$

dim(c_3)-dimensional

$c_4 = \log(9)$, $\frac{\partial c_4}{\partial c_4} = 1$

$\frac{\partial c_4}{\partial c_2} = \frac{\partial c_4}{\partial c_3}\cdot\frac{\partial c_3}{\partial c_2} = \frac{1}{9}\cdot 1 = \frac{1}{9}$

$$\boxed{\frac{\partial c_4}{\partial x}} = \frac{\partial c_1}{\partial x}\frac{\partial c_4}{\partial c_1} + \frac{\partial c_2}{\partial x}\frac{\partial c_4}{\partial c_2} = 2\cdot\frac{1}{9} + 1\cdot\frac{1}{9} = \boxed{\frac{1}{3}}$$

$f: \mathbb{R}^n \to \mathbb{R}^m$

$n \gg m$ Reverse Mode more efficient

For optimization, usually the case

$n \approx m$ roughly same efficiency