Lecture 3 - Differentiability

Functions of multiple variables

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

Partial derivatives $\frac{\partial f}{\partial x_i}(x) = \lim_{h \to 0} \frac{f(x+he_i) - f(x)}{h}$

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

Difference between functions of one and more variables:

One variable functions: differentiability $\Rightarrow$ continuity

Multi variable functions: can be differentiable in a point without being continous

Directional Derivative

When computing the partial derivative, we see what a slight nudge in either axis direction affects the output, but with the directional derivative we look at what the change to the output will be if we take that nudge in an arbitrary direction d.

Partial derivative:

function f of two variables a,b in direction x

$$\frac{\partial f}{\partial x}(a,b) = \lim_{h \to 0} \frac{f(a+h, b) - f(a,b)}{h}$$

can be written as:

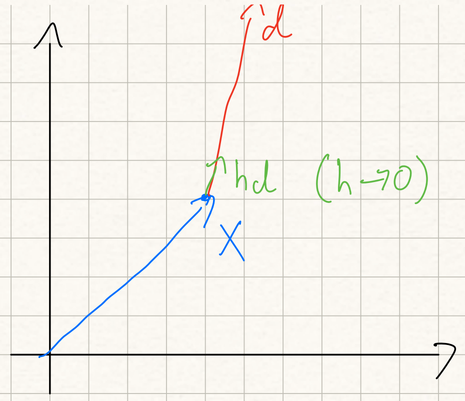
$$\frac{\partial f}{\partial x}(\vec{a}) = \lim_{h \to 0} \frac{f(\vec{a} + h\vec{e}_i) - f(\vec{a})}{h} \quad , \quad \vec{e}_i \text{ unit vector}$$

Can extend this to moving in different directions

directional derivative:

different notations: $f'(x, d)$, $\nabla_d f(x)$

$$\nabla_d f(x) = \lim_{h \to 0} \frac{f(x+hd) - f(x)}{h}$$



is a way to interpret the slope of a graph

if f is differentiable, the directional derivative exists in every direction and is

$$f'(x, d) = \nabla f(x)^T d$$

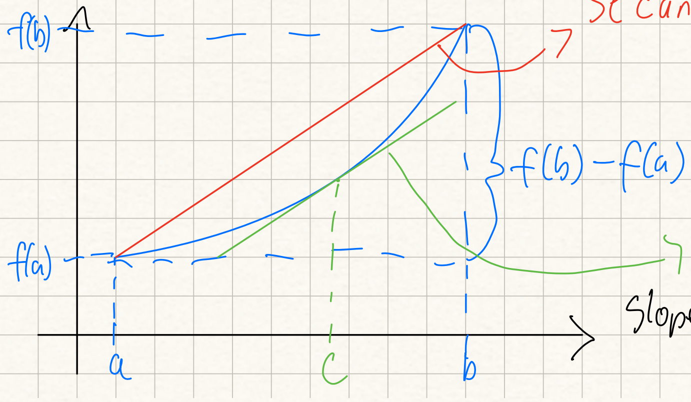
Mean value theorem (functions of a single variable)

Suppose $f: \mathbb{R} \to \mathbb{R}$ continuous on $[a, b]$ and differentiable on (a, b) $\Rightarrow$ there exists a c, where $a < c < b$ such that

$$f'(c) = \frac{f(b) - f(a)}{b - a}$$

Geometrically:

secant line: slope $\dfrac{f(b)-f(a)}{b-a}$



tangent line; slope $f'(c)$

Slope of secant and tangent is equal

Presentation from Claude

Let's start with single variable differentiability. A function $f(x)$ is differentiable at a point x_0 if the limit of the difference quotient exists:

$$f'(x_0) = \lim_{h \to 0} \frac{f(x_0+h) - f(x_0)}{h}$$

Geometrically, this means that we can draw a unique tangent line at the point $(x_0, f(x_0))$. However, in optimization, we often work with functions of multiple variables, which requires a more general definition.

For a function $f: \mathbb{R}^n \to \mathbb{R}$, differentiability at a point x_0 means we can approximate the function near x_0 with a linear function plus a small error term. Specifically, f is differentiable at x_0 if there exists a linear transformation L (represented by a vector) such that:

$$f(x_0+h) = f(x_0) + Lh + o(\|h\|)$$

where $o(\|h\|)$ represents a remainder term that goes to zero faster than $\|h\|$

This leads us to the concept of partial derivatives. For a function $f(x_1, \ldots, x_n)$, the partial derivative with respect to x_i is;

$$\frac{\partial f}{\partial x_i} = \lim_{h \to 0} \frac{f(x_1, \dots, x_i + h, \dots, x_n) - f(x_1, \dots, x_i, \dots, x_n)}{h}$$

$$= \lim_{h \to 0} \frac{f(x + he_i) - f(x)}{h}, \quad e_i \text{ unit vector}$$

These partial derivatives form the Jacobian matrix when f is vector-valued, or the gradient vector ∇f when f is scalar-valued. The gradient has a crucial geometric interpretation; it points in the direction of steepest increase of the function.

Let's consider a concrete example. Take $f(x,y) = x^2 + xy + y^2$
The gradient is:

$$\nabla f = \begin{bmatrix} 2x + y \\ x + 2y \end{bmatrix}$$

The directional derivative extends this concept. For a unit vector u, the directional derivative is:

$$\nabla_u f(x) = \lim_{t \to 0} \frac{f(x + tu) - f(x)}{t} = \nabla f(x)^T u$$

This gives us the rate of change of f in any direction u
An important theorem states that if f is differentiable, then all directional derivatives exist and can be computed using

the gradient.

The Hessian matrix $H(f)$, containing all second partial derivatives, helps us analyze the local behavior of f:

$$H(f) = \left[\frac{\partial^2 f}{\partial x_i \partial x_j} \right]$$

For our example $f(x,y) = x^2 + xy + y^2$, the Hessian is:

$$H(f) = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

A descent direction d at point x_0 satisfies $\nabla f(x_0)^T d < 0$. This concept is fundamental in optimization algorithms, as it tells us which directions will decrease our objective function.

Finally, quadratic functions are particulary important in optimization because they often appear as approximations to more complex functions. A quadratic function has the form:

$$f(x) = \frac{1}{2} x^T Q x + b^T x$$

where Q is symmetric and b is a vector.