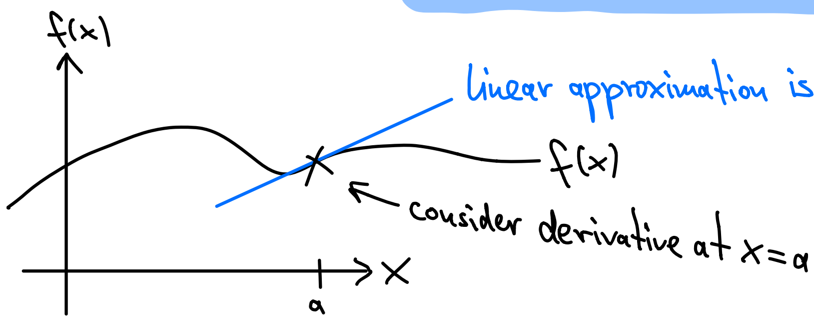
2.2 Total Derivative

Session 6
Feb. 19, 2020

Let us consider the usual derivative of $f: \mathbb{R} \to \mathbb{R}$ from a slightly different perspective: $\boxed{\text{derivative} = \text{best linear approximation}}$



the linear function $\ell(x) = mx + b$ must satisfy

- $\ell(a) = f(a)$

  $\Rightarrow ma + b = f(a) \Rightarrow b = f(a) - ma$

  $\Rightarrow \ell(x) = mx + f(a) - ma = m(x-a) + f(a)$

- good approximation near $x = a$: $f(x) - \ell(x)$ goes to zero faster than $x - a$ as $x \to a$

  $\Rightarrow \dfrac{f(x) - \ell(x)}{x - a} \longrightarrow 0$ as $x \to a$

Denote $x = a + h$, then $f(x) - \ell(x) = f(\underbrace{a+h}_{x}) - f(a) - m\underbrace{h}_{x-a} =: E(h)$

So a reasonable definition of differentiability is:

Definition:

$f: \mathbb{R} \to \mathbb{R}$ is differentiable at $a \in \mathbb{R}$ if there is some $m \in \mathbb{R}$ s.t.

$$f(a+h) = f(a) + mh + E(h) \quad \text{with} \quad \lim_{h \to 0} \frac{E(h)}{h} = 0.$$

Side remark: f continuous at $a \in \mathbb{R}$ if $f(a+h) = f(a) + R(h)$ with $\lim_{h \to 0} R(h) = 0$ (because this means $\lim_{h \to 0} f(a+h) = f(a)$).

What is m?

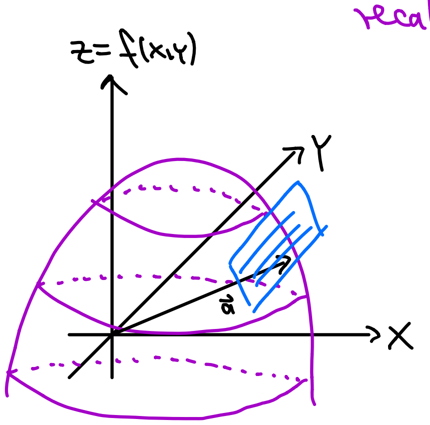
$$m = \frac{f(a+h)-f(a)-E(h)}{h} = \frac{f(a+h)-f(a)}{h} - \underbrace{\frac{E(h)}{h}}_{\to 0 \text{ as } h\to 0}$$

so taking the limit $h \to 0$ yields $m = \lim_{h\to 0} \frac{f(a+h)-f(a)}{h} = f'(a)$ (the usual derivative)

Conclusion: the definition above is equivalent to the one you know from Calculus and Linear Algebra I (limit of difference quotient)

Now: apply similar arguments for $f: \mathbb{R}^2 \to \mathbb{R}$, $\vec{x} = \begin{pmatrix} x \\ y \end{pmatrix} \mapsto f(\vec{x}) = f(x,y)$

How does $f(\vec{x})$ look like near $\vec{x} = \vec{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$?



recall: $z = f(x,y)$ defines a surface $=$ graph of f

$$= \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3 : z = f(x,y),\ x \in \mathbb{R}, y \in \mathbb{R} \right\}$$

"linear" (better: affine) function $=$ plane

$$\ell(x,y) = m_1 x + m_2 y + b = \underbrace{\vec{m}\cdot\vec{x}}_{} + b$$

scalar product $\vec{m}\cdot\vec{x} = \begin{pmatrix} m_1 \\ m_2 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} = m_1 x + m_2 y$

- from $\ell(\vec{a}) = f(\vec{a})$ we get $f(\vec{a}) = \vec{m}\cdot\vec{a} + b$

$$\Rightarrow \ell(\vec{x}) = \vec{m}\cdot\vec{x} + f(\vec{a}) - \vec{m}\cdot\vec{a} = \vec{m}\cdot(\vec{x}-\vec{a}) + f(\vec{a})$$

As before, we write $\vec{x} = \vec{a} + \vec{h}$, and we want to consider $\vec{h} \to 0$.

↳ note: $\vec{h} \to 0$ means: if $\vec{h} = \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}$, then both $h_1 \to 0$ and $h_2 \to 0$;

alternatively: $|\vec{h}| = \sqrt{h_1^2 + h_2^2} \to 0$

Definition:

$f: \mathbb{R}^n \to \mathbb{R}$ is called differentiable at $\vec{a} \in \mathbb{R}^n$ if there is a vector $\vec{m} \in \mathbb{R}^n$

such that
$$f(\vec{a}+\vec{h}) = f(\vec{a}) + \vec{m}\cdot\vec{h} + E(\vec{h}) \quad \text{with} \quad \frac{E(\vec{h})}{|\vec{h}|} \longrightarrow 0 \text{ as } \vec{h} \longrightarrow 0.$$

We call $\vec{m} =: (\vec{\nabla} f)(\vec{a}) = \vec{\nabla} f(\vec{a})$ the gradient of f at $\vec{a}$.

Remarks:

- sometimes $\vec{\nabla}$ is called "nabla" or "del"
- in other words: above we require $\displaystyle\lim_{\vec{h}\to 0} \frac{f(\vec{a}+\vec{h}) - f(\vec{a}) - \vec{m}\cdot\vec{h}}{|\vec{h}|} = 0$

but how do we solve for $\vec{m}$?
↳ see below

- geometrically in $\mathbb{R}^2$: $z = \ell(\vec{x}) = f(\vec{a}) + (\vec{\nabla} f)(\vec{a}) \cdot (\vec{x} - \vec{a})$ is the tangent plane

  to the surface $z = f(\vec{x})$ at $\vec{x} = \vec{a}$

  note: in $\mathbb{R}^n$ we call $z = \ell(\vec{x})$ a "tangent hyperplane" and $\vec{z} = f(\vec{x})$ a "hypersurface"

now: look at special case when $\vec{h} = \begin{pmatrix} h \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ and, say, $h > 0$

$$\Rightarrow 0 = \lim_{\vec{h}\to 0} \frac{f(\vec{a}+\vec{h}) - f(\vec{a}) - \vec{m}\cdot\vec{h}}{|\vec{h}|} = \lim_{h\to 0} \frac{f(a_1+h, a_2, \dots, a_n) - f(a_1, \dots, a_n) - m_1 h}{h}$$

$$= \lim_{h\to 0} \frac{f(a_1+h, a_2, \dots, a_n) - f(a_1, \dots, a_n)}{h} - m_1$$

$$\vec{m}\cdot\vec{h} = \begin{pmatrix} m_1 \\ \vdots \\ m_n \end{pmatrix} \begin{pmatrix} h \\ 0 \\ \vdots \\ 0 \end{pmatrix} = m_1 h$$

$$\vec{a}+\vec{h} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} + \begin{pmatrix} h \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} a_1+h \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$$

$$\Rightarrow m_1 = \lim_{h\to 0} \frac{f(a_1+h, a_2, \dots, a_n) - f(a_1, \dots, a_n)}{h}$$ which is the partial derivative $\frac{\partial f}{\partial x}(\vec{a})$!

Theorem:

If $f: \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\vec{a} \in \mathbb{R}^n$, then $\frac{\partial f}{\partial x_i}(\vec{a})$ exist $\underbrace{\forall}_{\text{for all}}\, i = 1, \dots, n$ and

$$(\vec{\nabla} f)(\vec{a}) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}.$$

Note: short-hand notation: $\frac{\partial f}{\partial x} = \partial_x f$, so $\vec{\nabla} f = \begin{pmatrix} \partial_{x_1} f \\ \vdots \\ \partial_{x_n} f \end{pmatrix}$

Example: $f(x, y) = x \sin(y)$

$$\Rightarrow (\vec{\nabla} f)(\vec{x}) = \begin{pmatrix} \frac{\partial f(\vec{x})}{\partial x} \\ \frac{\partial f(\vec{x})}{\partial y} \end{pmatrix} = \begin{pmatrix} \sin(y) \\ x\cos(y) \end{pmatrix}$$

$$\Rightarrow \text{with } \vec{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}: \quad (\vec{\nabla} f)(\vec{a}) = \begin{pmatrix} \sin(a_2) \\ a_1 \cos(a_2) \end{pmatrix}$$

the tangent plane at $\vec{a}$ is defined by

$$z = f(\vec{a}) + (\vec{\nabla} f)(\vec{a}) \cdot (\vec{x} - \vec{a}) = a_1 \sin(a_2) + \begin{pmatrix} \sin(a_2) \\ a_1 \cos(a_2) \end{pmatrix} \begin{pmatrix} x - a_1 \\ y - a_2 \end{pmatrix}$$

$$\Rightarrow z = a_1 \sin(a_2) + \sin(a_2)(x - a_1) + a_1 \cos(a_2)(y - a_2)$$

Note: the usual sum, product and quotient rules still hold, because they hold for the partial derivatives:

- $\vec{\nabla}(f+g) = \vec{\nabla} f + \vec{\nabla} g$
- $\vec{\nabla}(f \cdot g) = \begin{pmatrix} \partial_{x_1}(fg) \\ \vdots \\ \partial_{x_n}(fg) \end{pmatrix} = \begin{pmatrix} (\partial_{x_1} f) g + f(\partial_{x_1} g) \\ \vdots \\ (\partial_{x_n} f) g + f(\partial_{x_n} g) \end{pmatrix} = (\vec{\nabla} f) g + f(\vec{\nabla} g)$
- $\vec{\nabla}\left(\frac{f}{g}\right) = \frac{(\vec{\nabla} f) g - f(\vec{\nabla} g)}{g^2}$ for $g \neq 0$

Note: similar to one-dimension, we can define the $\boxed{\text{differential } df}$

↳ in finite difference notation: $\Delta f = f(\vec{a}+\vec{h}) - f(\vec{a}) = (\vec{\nabla} f)(\vec{a}) \cdot \vec{h} + \text{error}$

↳ in infinitesimal notation, writing $\vec{h} = d\vec{x} = \begin{pmatrix} dx_1 \\ \vdots \\ dx_n \end{pmatrix}$:

$$\boxed{df = (\vec{\nabla} f) \cdot d\vec{x} = \frac{\partial f}{\partial x_1} dx_1 + \dots + \frac{\partial f}{\partial x_n} dx_n}$$

<u>Ex.:</u> $f(\vec{x}) = x \sin(y)$

$$\Rightarrow df = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy = \sin(y) dx + x \cos(y) dy$$