CHAPTER TWO: DIFFERENTIATION

SECTION 2.1: GEOMETRY OF REAL VALUED FNS

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

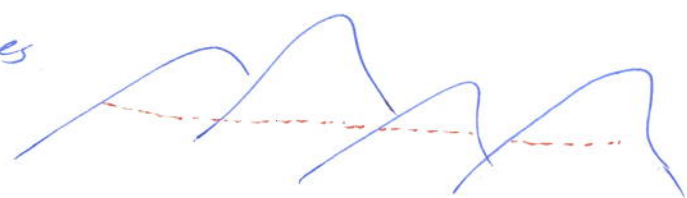
↳ vector valued if $m \geq 2$

scaler valued if $m = 1$

- $f: U \subset \mathbb{R}^n \longrightarrow \mathbb{R}$: graph$(f)$ $= \{(x_1, \dots, x_n, f(x_1, \dots, x_n)) : \vec{x} \in U\}$
- Level set is subset where f is constant

↳ level set of value c is $\{\vec{x} \in U : f(\vec{x}) = c\}$

↳ Think heights in mountain ranges



Lots of examples/additional notation in the book

↳ Many graphing programs will do this

↳ Do Mathematica Example

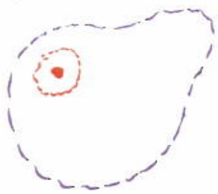
Homework: #1, #24

Suggested: #3, #30

Section 2.2: Limits and Continuity

Terminology

- **Open Disk/Ball**: $D_r(\vec{x}_0) = \{\vec{x} : \|\vec{x} - \vec{x}_0\| < r\}$
- **Open Set**: $U \subset \mathbb{R}^n$ open if for all $\vec{x}_0 \in U$ there is a r (which may depend on $\vec{x}_0$) st $D_r(\vec{x}_0) \subset U$



↳ Note: empty set $\emptyset$ considered open, $\mathbb{R}^n$ open

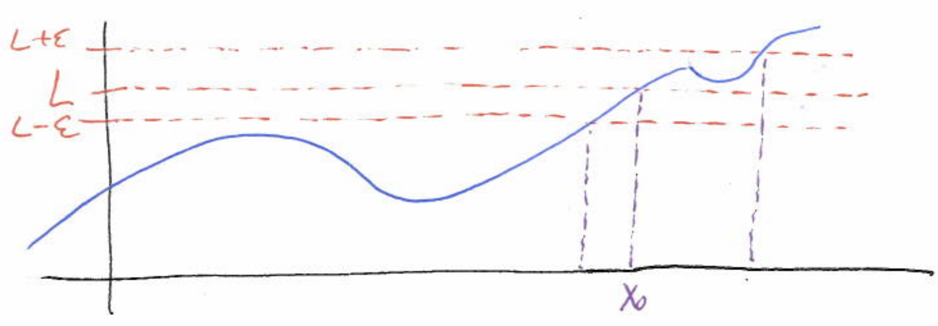
↳ use dotted lines to denote open

- **Neighborhood**: Mean any open set containing $\vec{x}_0$
- **Boundary Points**: Given $A \subset \mathbb{R}^n$, say $\vec{x}$ is a boundary point of A if every neighborhood of $\vec{x}$ contains at least one point in A and at least one point not in A.
- **Closed**: A set is closed if it contains all its ~~limit~~ boundary points.

- **Limit**: Say the limit of f as $\vec{x}$ approaches $\vec{x}_0$ is L if $f(\vec{x})$ gets closer and closer to L as $\vec{x}$ gets closer to $\vec{x}_0$. Denote $\lim_{x \to x_0} f(x) = L$.

↳ Can define in terms of neighborhoods

↳ Can do ϵ-δ: $\forall \epsilon > 0 \; \exists \delta$ st $|\vec{x} - \vec{x}_0| < \delta$ (and $\vec{x} \neq \vec{x}_0$) implies $|f(\vec{x}) - L| < \epsilon$



EXAMPLE: PROVE $f(x)$ is continuous at $x=3$ if $f(x)=x^2$

"Guess" $L=9$. Given ε find δ st $|x-3|<\delta \Rightarrow |f(x)-9|<L$

Well, $|f(x)-9| = |x^2-9| = |x-3|\cdot|x+3| < \varepsilon$

↳ wlog, assume $\delta<1$ so $|x+3|$ is between 2 and 4

Then $|x-3|\cdot 2 < \varepsilon$ or $|x-3| < \frac{\varepsilon}{2}$

So if we take $\delta < \frac{\varepsilon}{2}$ then $|x-3|<\delta \rightarrow |f(x)-9| < \varepsilon$ ▲

• Usually won't argue so rigorously, but good to know "how"

PROPERTIES OF LIMITS

• Uniqueness: $\lim_{x\to x_0} f(x)$ equals b_1 and b_2 then $b_1=b_2$

• Constant: $\lim_{x\to x_0} c\, f(x) = c \lim_{x\to x_0} f(x)$

• Sum/Diff: $\lim_{x\to x_0}\left(f(x)+g(x)\right) = \lim_{x\to x_0} f(x) + \lim_{x\to x_0} g(x)$ if at least one exists on RHS

• Product/Quotient: $\lim_{x\to x_0} \frac{f(x)}{g(x)} = \frac{\lim_{x\to x_0} f(x)}{\lim_{x\to x_0} g(x)}$ and $\lim_{x\to x_0} f(x)g(x) = \lim_{x\to x_0} f(x) \lim_{x\to x_0} g(x)$

so long as at least one of RHS exists, and for quotient $\lim_{x\to x_0} g(x)$ is non-zero

• Components $f(\vec{x}) = (f_1(\vec{x}),\dots, f_n(\vec{x}))$. Then $\lim_{x\to x_0} f(\vec{x}) = \vec{b}$ if and only if $\lim_{x\to x_0} f_i(\vec{x}) = b_i$ for $i\in\{1,2,\dots,n\}$

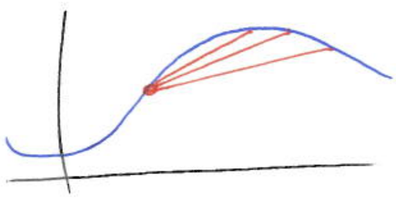
DANGERS: $\lim_{x\to\infty}(x^2-x)$, $\lim_{x\to\infty}(x^2-x^2)$, $\lim_{x\to\infty}(x^2-x^3)$

Do not define $\infty-\infty$, $\pm\infty\cdot 0$; do define $\infty\cdot\infty$, $\infty+\infty$

Section 2.3: Differentiation

Defn of the deriv (one variable)

$$f'(x) = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} \quad \text{or} \quad f'(a) = \lim_{x\to a} \frac{f(x) - f(a)}{x-a}$$



Interpretation: average speed

Partial Deriv

$$\frac{\partial f}{\partial x_j}(x_1,\dots,x_n) = \lim_{h\to 0} \frac{f(x_1, x_2, \dots, x_j + h, \dots, x_n) - f(x_1, \dots, x_n)}{h}$$

$$= \lim_{h\to 0} \frac{f(\vec{x} + h\vec{e}_j) - f(\vec{x})}{h}$$

↳ Just treat all other variables as constants

Example: Find partials of $f(x,y) = x\cos(xy)$

Outstanding Example: $f(x,y) = (xy)^{1/3}$

↳ $\frac{\partial f}{\partial x}(0,0) = \lim_{h\to 0} \frac{f(h,0) - f(0,0)}{h} = 0 = \frac{\partial f}{\partial y}(0,0)$

Let $g : \mathbb{R} \to \mathbb{R}^2$ be given by $g(x) = (x, x)$

Let $A(x) = f(g(x)) = (f \circ g)(x) = x^{2/3}$

↳ f and g differentiable

$A'(x) = \lim_{h\to 0} \frac{A(\cancel{x}+h) - A(0)}{h} = \lim_{h\to 0} \frac{h^{2/3}}{h}$ = undefined!

↳ Composition of diff is not necc diff

↳ Chain rule will be even harder

↳ Problem here: no good tangent plane at $(0,0)$

Section 2.3 (cont)

Linear Approximations

↳ Very important: locally complex fns well approx with simple fns

↳ tangent line: $y = f(a) + f'(a)(x-a)$

↳ Size of error?

↳ MVT: $f(x) = f(a) + f'(c)(x-a)$

$$\text{Thus } |f(x) - y| = |f'(c) - f'(a)| \cdot |x-a|$$
$$\leq \left| 2 \max_{w \in [a,c]} f'(w) \right| \cdot |x-a|$$

↳ better estimate if f'' exists

↳ Generalize to higher dimensions

Tangent Plane:

$$z = f(x_0, y_0) + \left[\frac{\partial f}{\partial x}(x_0, y_0)\right](x - x_0) + \left[\frac{\partial f}{\partial y}(x_0, y_0)\right](y - y_0)$$

Deriv in one-var: $\displaystyle \lim_{x \to x_0} \frac{f(x) - f(x_0) - f'(x_0)(x - x_0)}{x - x_0} = 0$

<u>Defn of the deriv</u>: Open $U \subset \mathbb{R}^n$, $f: \mathbb{R}^n \to \mathbb{R}^m$ is diff at $\vec{x}_0$ if partial derivs exist at $\vec{x}_0$, and if $T = Df(\vec{x}_0)$ is the $m \times n$ matrix with elements $\partial f_i / \partial x_j (\vec{x}_0)$ satisfies

$$\lim_{\vec{x} \to \vec{x}_0} \frac{\| f(\vec{x}) - f(\vec{x}_0) - T(\vec{x} - \vec{x}_0) \|}{\| \vec{x} - \vec{x}_0 \|} = 0,$$

with $\vec{x} - \vec{x}_0$ a column vector. Call T the deriv of f at x_0 or the matrix of partial derivs of f at $\vec{x}_0$

Section 2.3 (cont)

• **Example in 2-dim:** $f: \mathbb{R}^2 \to \mathbb{R}$ is diff at (x_0, y_0) if

$$\lim_{(x,y)\to(x_0,y_0)} \frac{f(x,y) - f(x_0,y_0) - \left[\frac{\partial f}{\partial x}(x_0,y_0)\right](x-x_0) - \left[\frac{\partial f}{\partial y}(x_0,y_0)\right](y-y_0)}{\|(x,y) - (x_0,y_0)\|}$$

tends to 0. In other words, locally tangent plane approxs well.

• **Example:** $f(x,y) = x^2 + y^4 + e^{xy}$ at $(1,0,2)$

Write it as $z = f(x,y)$, with $z_0 = f(x_0,y_0) = f(1,0) = 2$

$\frac{\partial f}{\partial x} = 2x + ye^{xy} \implies \frac{\partial f}{\partial x}(1,0) = 2$

$\frac{\partial f}{\partial y} = 4y^3 + xe^{xy} \implies \frac{\partial f}{\partial y}(1,0) = 1$

Tangent plane is $z = 2 + 2(x-1) + 1\cdot(y-0)$

Special Case: Gradient

$f: \mathbb{R}^n \to \mathbb{R}$ then $Df(\vec{x}) = \left[\frac{\partial f}{\partial x_1} \quad \cdots \quad \frac{\partial f}{\partial x_n}\right]$ is a $1\times n$ matrix, form corresponding vector $\nabla f = \text{grad}(f) = \left(\partial f/\partial x_1, \ldots, \partial f/\partial x_n\right)$, called the gradient of f.

Example: $f(x,y,z) = x^2+y^2+z^2$ then $\nabla f = (2x, 2y, 2z)$

Note $Df(\vec{x})(\vec{h}) = \nabla f(\vec{x}) \cdot \vec{h}$

SECTION 2.3 (CONT)

THM: $f: U \subset \mathbb{R}^n \to \mathbb{R}^m$ diff at $\vec{x}_0 \Rightarrow f$ is cont at $\vec{x}_0$

Proof: $f(\vec{x}) = f(\vec{x}_0) + Df(\vec{x}_0)(\vec{x}-\vec{x}_0) + E_{\vec{x}_0}(\vec{x})$

with $\lim_{x \to \vec{x}_0} \|E_{\vec{x}_0}(\vec{x})\| / \|\vec{x}-\vec{x}_0\| = 0$

THM: $f: U \subset \mathbb{R}^n \to \mathbb{R}^m$ st all partial derivs $\partial f_i / \partial x_j$ exist and continuous in a nbhood of $\vec{x} \in U$. Then f is differentiable at $\vec{x}$.

↳ Recall $f(x,y) = (xy)^{1/3}$

↳ partial derivs not continuous at $(0,0)$: $\frac{\partial f}{\partial x} = \frac{1}{3} x^{-\frac{2}{3}} y^{\frac{1}{3}}$

and if $y = x$ then $\frac{\partial f}{\partial x}(x,x) = \frac{1}{3} x^{-1/3}$

Following valid: Continuous Partial Derivatives $\Longrightarrow$ Function is Differentiable $\Longrightarrow$ Partial Derivatives Exist

Each converse statement, obtained by reversing arrows, can fail.

Say a function is C^1 if it is continuously differentiable, C^2 if twice continuously differentiable, and so on.

PROOF PARTIALS EXIST AND CONT $\Rightarrow$ FN IS DIFFERENTIABLE

- Will do $n=2$, generalization possible
- Highlights an important technique: adding zero.
- Natural guess for derivative: try it:

$$f(x,y) - f(0,0) - \left[\frac{\partial f}{\partial x}(0,0)\right]x - \left[\frac{\partial f}{\partial y}(0,0)\right]y$$

$$= \left\{\underbrace{f(x,y) - f(0,y)}_{MVT} - \left[\frac{\partial f}{\partial x}(0,0)\right]x\right\} + \left\{\underbrace{f(0,y) - f(0,0)}_{MVT} - \left[\frac{\partial f}{\partial y}(0,0)\right]y\right\}$$

$$= \left\{\left[\frac{\partial f}{\partial x}(c,y)\right]x - \left[\frac{\partial f}{\partial x}(0,0)\right]x\right\} + \left\{\left[\frac{\partial f}{\partial y}(0,\tilde{c})\right]y - \left[\frac{\partial f}{\partial y}(0,0)\right]y\right\}$$

$$= \underbrace{\left[\frac{\partial f}{\partial x}(c,y) - \frac{\partial f}{\partial x}(0,0)\right]}_{\to 0 \text{ (by continuity)}} x + \underbrace{\left[\frac{\partial f}{\partial y}(0,\tilde{c}) - \frac{\partial f}{\partial y}(0,0)\right]}_{\to 0 \text{ (by continuity)}} y$$

As $\frac{\|x\|}{\|(x,y)\|}$ and $\frac{\|y\|}{\|(x,y)\|}$ bounded, above tends to zero even after dividing by $\|(x,y)\| = \|(x,y)-(0,0)\|$. $\blacksquare$

Homework: #2ab, #4ab, #5, #7c, #12a (linear approx), #13a

Suggested: #3, #4cde, #10, #15, #18, #19

Section 2.5: Properties of the Derivative

THM: Properties of the Derivative:

Assume $f, g: U \subset \mathbb{R}^n \to \mathbb{R}^m$ diff at $\vec{x}_0$, let c be a constant.

- Const Rule: $h(\vec{x}) = c f(\vec{x})$ Then $Dh(\vec{x}_0) = c Df(\vec{x}_0)$
- Sum Rule: $h(\vec{x}) = f(\vec{x}) + g(\vec{x})$ Then $Dh(\vec{x}_0) = Df(\vec{x}_0) + Dg(\vec{x}_0)$
- Product Rule: $m=1$, $h(\vec{x}) = f(\vec{x})g(\vec{x})$ Then $Dh(\vec{x}_0) = \cancel{g(\vec{x}_0)}\, Df(\vec{x}_0)\, g(\vec{x}_0) + f(\vec{x}_0) Dg(\vec{x}_0)$
- Quotient Rule: $m=1$, $h(\vec{x}) = f(\vec{x})/g(\vec{x})$ Then $Dh(\vec{x}_0) = \dfrac{Df(\vec{x}_0)g(\vec{x}_0) - f(\vec{x}_0)Dg(\vec{x}_0)}{g(\vec{x}_0)^2}$

Example: $f(x,y,z) = x^2+y^2+z^2 \quad g(x) = x^3+y^3$

$$h(x,y,z) = f(x,y,z)\, g(x,y,z) = x^5 + x^3y^2 + x^3z^2 + x^2y^3 + y^5 + z^2y^3$$

$$Dh(x,y,z) = \left(\frac{\partial h}{\partial x}, \frac{\partial h}{\partial y}, \frac{\partial h}{\partial z}\right) = \big(5x^4 + 3x^2y^2 + 3x^2z^2 + 2xy^3,\; 2x^3y + 3x^2y^2 + 5y^4 + 3z^2y^2,\; 2x^3z + 2zy^3\big)$$

$$Df(x,y,z) = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\right) = (2x, 2y, 2z) \qquad f(x,y,z) = x^2+y^2+z^2$$

$$Dg(x,y,z) = \left(\frac{\partial g}{\partial x}, \frac{\partial g}{\partial y}, \frac{\partial g}{\partial z}\right) = (3x^2, 3y^2, 0)$$

$$Df(x,y,z)g(x,y,z) + f(x,y,z)Dg(x,y,z) = (2x, 2y, 2z) * (x^3+y^3) + (x^2+y^2+z^2) * (3x^2, 3y^2, 0)$$

$$\cancel{= 3x^4 + 2x^3y + 2x^3z}$$

$$= \big(5x^4 + 3x^2y^2 + 3x^2z^2 + 2xy^3,\; 5y^4 + 2x^3y + 3x^2y^2 + 3z^2y^2,\; 2x^3z + 2y^3z\big)$$

↳ Note agree. Faster is to note symmetry in $x \to y$ and $y \to x$

SECTION 2.5: (CONT)

Proof of product rule

↳ Use powerful technique of adding zero!

$$\lim_{\vec{x}\to\vec{x}_0} \frac{\| f(\vec{x})g(\vec{x}) - f(\vec{x}_0)g(\vec{x}_0) - [Df(\vec{x}_0)g(\vec{x}_0) + f(\vec{x}_0)Dg(\vec{x}_0)](\vec{x}-\vec{x}_0) \|}{\|\vec{x} - \vec{x}_0\|}$$

↳ add $-f(\vec{x}_0)g(\vec{x}) + f(\vec{x}_0)g(\vec{x})$

use triangle inequality: $\|\vec{P} + \vec{Q}\| \le \|\vec{P}\| + \|\vec{Q}\|$

$$\le \lim_{\vec{x}\to\vec{x}_0} \frac{\| [f(\vec{x}) - f(\vec{x}_0) - Df(\vec{x}_0)(\vec{x}-\vec{x}_0)] \| \, |g(\vec{x})|}{\|\vec{x} - \vec{x}_0\|}$$

$$+ \lim_{\vec{x}\to\vec{x}_0} \frac{\| Df(\vec{x}_0)(g(\vec{x}) - g(\vec{x}_0))(\vec{x}-\vec{x}_0) \|}{\|\vec{x} - \vec{x}_0\|}$$

$$+ \lim_{\vec{x}\to\vec{x}_0} \frac{\|f(\vec{x}_0)\| \, \|g(\vec{x}) - g(\vec{x}_0) - Dg(\vec{x}_0)(\vec{x}-\vec{x}_0)\|}{\|\vec{x} - \vec{x}_0\|}$$

and each piece tends to zero. ∎

SECTION 2.5 (CONT)

THM: THE CHAIN RULE: f: open $V \subset \mathbb{R}^m \to \mathbb{R}^p$ and g: open $U \subset \mathbb{R}^n \to \mathbb{R}^m$ with $g(U) \subset V$. Suppose g is diff at $\vec{x}_0$ and f is diff at $\vec{y}_0 = g(\vec{x}_0)$. Then $h(\vec{x}) = (f \circ g)(\vec{x}) = f(g(\vec{x}))$ is diff at $\vec{x}_0$ and $Dh(\vec{x}_0) = Df(\vec{x}_0)\, Dg(\vec{x}_0)$

Example: $g(x,y) = (\cos x, \sin x, y)$

$$f(u,v,w) = u^2 + v^2 + uvw$$

$$h(x,y) = f(g(x,y)) = f(\cos x, \sin x, y)$$

$$= \cos^2 x + \sin^2 x + \cos x \cdot \sin x \cdot y$$

$$(Dh)(x,y) = \left(\frac{\partial h}{\partial x}, \frac{\partial h}{\partial y}\right) = \left((-\sin^2 x + \cos^2 x)y,\ \cos x \cdot \sin x\right)$$

$$\cancel{Dg(x,y) = \left(\frac{\partial g}{\partial x}, \frac{\partial g}{\partial y}\right) = (-\sin x}$$

$$\cancel{Dg(x,y) = \begin{pmatrix} \frac{\partial g_1}{\partial x} & \frac{\partial g_2}{\partial x} & \frac{\partial g_3}{\partial x} \\ \frac{\partial g_1}{\partial y} & \frac{\partial g_2}{\partial y} & \frac{\partial g_3}{\partial y} \end{pmatrix}}$$

$$Dg(x,y) = \begin{pmatrix} \frac{\partial g_1}{\partial x} & \frac{\partial g_1}{\partial y} \\ \frac{\partial g_2}{\partial x} & \frac{\partial g_2}{\partial y} \\ \frac{\partial g_3}{\partial x} & \frac{\partial g_3}{\partial y} \end{pmatrix} = \begin{pmatrix} -\sin x, & 0 \\ \cos x, & 0 \\ 0, & 1 \end{pmatrix}$$

Consider $f(x(v,w), y(v,w), z(v,w)) - f(x(u,w), y(v,w), z(v,w))$

(MORE CHAIN RULE)

By the defn of the deriv

$$f(\vec{r}_1) - f(\vec{r}_0) - (Df)(\vec{r}_0)(\vec{r}_1 - \vec{r}_0) = \text{small}$$

$$f(\vec{r}_2) - f(\vec{r}_0) - (Df)(\vec{r}_0)(\vec{r}_2 - \vec{r}_0) = \text{small}$$

where $\vec{r}_0 = (x(u,w), y(u,w), z(u,w))$ all u's

$\vec{r}_1 = (x(u,w), y(v,w), z(v,w))$ one u, two v's

$\vec{r}_2 = (x(v,w), y(v,w), z(v,w))$ all v's

Thus $f(\vec{r}_2) - f(\vec{r}_1) = \left[f(\vec{r}_2) - f(\vec{r}_0)\right] - \left[f(\vec{r}_1) - f(\vec{r}_0)\right]$

$$= -(Df)(\vec{r}_0)(\vec{r}_2 - \vec{r}_0) + (Df)(\vec{r}_0)(\vec{r}_1 - \vec{r}_0) + \text{small}$$

$$= -(Df)(\vec{r}_0)\vec{r}_2 + (Df)(\vec{r}_0)\vec{r}_0 + (Df)(\vec{r}_0)\vec{r}_1 - (Df)(\vec{r}_0)\vec{r}_0 + \text{small}$$

cancel

$$= -(Df)(\vec{r}_0)(\vec{r}_2 - \vec{r}_1) + \text{small}$$

$$= -(Df)(\vec{r}_0)\cdot(x(v,w) - x(u,w), 0, 0)$$

This will lead to $\frac{\partial f}{\partial x}\frac{\partial x}{\partial u}$, as we divide

by $v-u$ and then $\frac{x(v,w) - x(u,w)}{v-u} \to \frac{\partial x}{\partial u}$

Math 350: The Chain Rule

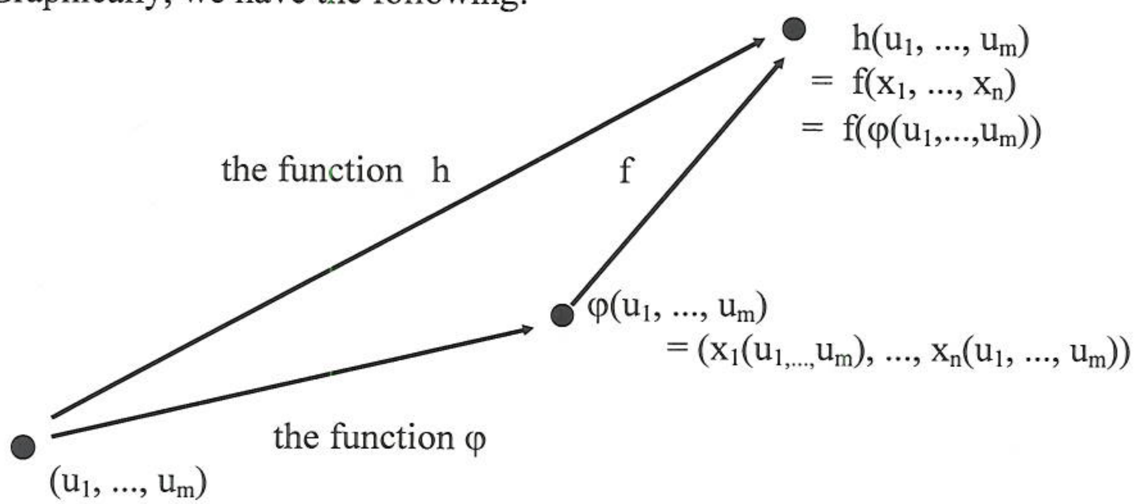
The Chain Rule is a very useful tool for analyzing the following: Say you have a function f of $(x_1, x_2, \ldots, x_n)$, and these variables are themselves functions of $(u_1, u_2, \ldots, u_m)$. How does our function f change as we vary u_1 thru u_m??? We'll state and explain the Chain Rule, and then give a DIFFERENT PROOF FROM THE BOOK, using *only* the definition of the derivative. This is a slight modification of notes I wrote years ago for a similar class at Princeton.

(I). Statement:

We'll state the Chain Rule. First, some notation:

Let $h: \mathbb{R}^m \to \mathbb{R}$	say h is a function of $(u_1, u_2, \ldots, u_m)$
$f: \mathbb{R}^n \to \mathbb{R}$	say f is a function of $(x_1, x_2, \ldots, x_n)$
$\varphi: \mathbb{R}^m \to \mathbb{R}^n$	say φ is a function of $(u_1, u_2, \ldots, u_m)$

Graphically, we have the following:



Our function h lives on $\mathbb{R}^m$. So, you give it an m-tuple, $(u_1, \ldots, u_m)$, and it will give you a real number back. The function f lives on $\mathbb{R}^n$. If you give it an n-tuple, $(x_1, \ldots, x_n)$, it will give you back a number. And what of the variables x_1 thru x_n? Well, they can be thought of as functions on $\mathbb{R}^m$: you give them an m-tuple, $(u_1, \ldots, u_m)$, and they'll return a number.

We cannot look at $f(x_1(u_1, \ldots, u_m))$, for f composed with x_1 doesn't make sense: x_1 gives us just ONE number; f needs n numbers.

What do we do? Remember, we're trying to understand the beast:

$$h(u_1, \ldots, u_m) = f(x_1(u_1, \ldots, u_m), \ldots, x_n(u_1, \ldots, u_m))$$

We define an auxiliary function, φ, to help us. What will $\varphi(u_1, \ldots, u_m)$ be? Whatever we want. We now look for something useful. Look at the Right Hand Side above—wouldn't it be nice if we could choose a φ that would give us this? We can! Just let:

$$\varphi(u_1, \ldots, u_m) = (x_1(u_1, \ldots, u_m), x_2(u_1, \ldots, u_m), \ldots, x_n(u_1, \ldots, u_m))$$

Now we can write $h = f \circ \varphi$, f composed with φ. The advantage of this is that we know that often compositions of nice functions are nice: if we compose two continuous functions, we get a continuous function. In one dimension, we have the 1-dimensional chain rule for compositions. We hope to be able to do something similar here. Anyway, here is the long awaited statement of:

The Chain Rule:

$$\begin{aligned}(Dh)(u_1, \ldots, u_m) &= (Df)(\varphi\,(u_1, \ldots, u_m))\,(D\varphi)(u_1, \ldots, u_m)\\ &= (Df)(x_1, \ldots, x_n)\,(D\varphi)(u_1, \ldots, u_m)\end{aligned}$$

Let's write out what this is: for the sake of space, I will not explicitly write WHERE the functions are being evaluated—we always evaluate h at $(u_1, \ldots, u_m)$, f at $\varphi(u_1, \ldots, u_m) = (x_1(u_1, \ldots, u_m), \ldots, x_n(u_1, \ldots, u_m))$, and φ at $(u_1, \ldots, u_m)$.

The Chain Rule:

$$\mathbf{Dh} = \left(\frac{\partial \mathbf{h}}{\partial \mathbf{u_1}}, \frac{\partial \mathbf{h}}{\partial \mathbf{u_2}}, \ldots, \frac{\partial \mathbf{h}}{\partial \mathbf{u_m}}\right) \quad \mathbf{Df} = \left(\frac{\partial \mathbf{f}}{\partial \mathbf{x_1}}, \frac{\partial \mathbf{f}}{\partial \mathbf{x_2}}, \ldots, \frac{\partial \mathbf{f}}{\partial \mathbf{x_n}}\right)$$

$D\varphi$ is more complicated: Unlike Df and Dh, which are vectors, $D\varphi$ is a matrix quantity. This is because φ is really a collection of m functions,

$$\begin{aligned}\varphi(u_1, \ldots, u_m) &= (\varphi_1(u_1, \ldots, u_m), \ldots, \varphi_n(u_1, \ldots, u_m))\\ &= (x_1(u_1, \ldots, u_m), \ldots, x_n(u_1, \ldots, u_m))\end{aligned}$$

We obtain:

We define an auxiliary function, φ, to help us. What will $\varphi(u_1, \ldots, u_m)$ be? Whatever we want. We now look for something useful. Look at the Right Hand Side above—wouldn't it be nice if we could choose a φ that would give us this? We can! Just let:

$$\varphi(u_1, \ldots, u_m) = (x_1(u_1, \ldots, u_m), x_2(u_1, \ldots, u_m), \ldots, x_n(u_1, \ldots, u_m))$$

Now we can write $h = f\circ\varphi$, f composed with φ. The advantage of this is that we know that often compositions of nice functions are nice: if we compose two continuous functions, we get a continuous function. In one dimension, we have the 1-dimensional chain rule for compositions. We hope to be able to do something similar here. Anyway, here is the long awaited statement of:

The Chain Rule:

$$\begin{aligned}(Dh)(u_1, \ldots, u_m) &= (Df)(\varphi\,(u_1, \ldots, u_m))\,(D\varphi)(u_1, \ldots, u_m)\\ &= (Df)(x_1, \ldots, x_n)\,(D\varphi)(u_1, \ldots, u_m)\end{aligned}$$

Let's write out what this is: for the sake of space, I will not explicitly write WHERE the functions are being evaluated—we always evaluate h at $(u_1, \ldots, u_m)$, f at $\varphi(u_1, \ldots, u_m) = (x_1(u_1, \ldots, u_m), \ldots, x_n(u_1, \ldots, u_m))$, and φ at $(u_1, \ldots, u_m)$.

The Chain Rule:

$$\mathbf{Dh} = \left(\frac{\partial \mathbf{h}}{\partial \mathbf{u}_1}, \frac{\partial \mathbf{h}}{\partial \mathbf{u}_2}, \ldots, \frac{\partial \mathbf{h}}{\partial \mathbf{u}_m}\right) \quad \mathbf{Df} = \left(\frac{\partial \mathbf{f}}{\partial \mathbf{x}_1}, \frac{\partial \mathbf{f}}{\partial \mathbf{x}_2}, \ldots, \frac{\partial \mathbf{f}}{\partial \mathbf{x}_n}\right)$$

$D\varphi$ is more complicated: Unlike Df and Dh, which are vectors, $D\varphi$ is a matrix quantity. This is because φ is really a collection of m functions,

$$\begin{aligned}\varphi(u_1, \ldots, u_m) &= (\varphi_1(u_1, \ldots, u_m), \ldots, \varphi_n(u_1, \ldots, u_m))\\ &= (x_1(u_1, \ldots, u_m), \ldots, x_n(u_1, \ldots, u_m))\end{aligned}$$

We obtain:

$$
(D\varphi) = \begin{pmatrix}
\frac{\partial x_1}{\partial u_1}, & \frac{\partial x_1}{\partial u_2}, & \ldots, & \frac{\partial x_1}{\partial u_m} \\
\frac{\partial x_2}{\partial u_1}, & \frac{\partial x_2}{\partial u_2}, & \ldots, & \frac{\partial x_2}{\partial u_m} \\
 & & & \\
\frac{\partial x_n}{\partial u_1}, & \frac{\partial x_n}{\partial u_2}, & \ldots, & \frac{\partial x_n}{\partial u_m}
\end{pmatrix}
$$

Combining the above expressions for Dh, Df, and $D\varphi$ yields:

Chain Rule:

$$\frac{\partial h}{\partial u_1} = \frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial u_1} + \frac{\partial f}{\partial x_2}\frac{\partial x_2}{\partial u_1} + \ldots + \frac{\partial f}{\partial x_n}\frac{\partial x_n}{\partial u_1}$$

$$\frac{\partial h}{\partial u_2} = \frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial u_2} + \frac{\partial f}{\partial x_2}\frac{\partial x_2}{\partial u_2} + \ldots + \frac{\partial f}{\partial x_n}\frac{\partial x_n}{\partial u_2}$$

and so on till

$$\frac{\partial h}{\partial u_m} = \frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial u_m} + \frac{\partial f}{\partial x_2}\frac{\partial x_2}{\partial u_m} + \ldots + \frac{\partial f}{\partial x_n}\frac{\partial x_n}{\partial u_m}$$

(II). One Dimensional Case:

OK. We now have the above formula, but WHERE DID IT COME FROM? Let's go back to one-dimension, and take a look at what is happening:

Translating from our language to what we spoke in High School:

$$h(u) = f(\varphi(u)) \rightarrow h'(u) = f'(\varphi(u))\,\varphi'(u)$$

How do we go about proving this? Always go back to what you know: here we're trying to find the derivative. Okay, so, let's recall the definition of the derivative. We know that. The derivative is defined by:

$$\begin{aligned} h'(u) &= \lim_{y \to u} \{h(y) - h(u)\} / \{y-u\} \\ &= \lim_{y \to u} \{f(\varphi(y)) - f(\varphi(u))\} / \{y-u\} \\ &= \lim_{y \to u} \frac{f(\varphi(y)) - f(\varphi(u))}{\varphi(y) - \varphi(u)} * \frac{\varphi(y) - \varphi(u)}{y - u} \end{aligned}$$

All we did was multiply by 1 in a very clever way. Why did we do this? Our function f is a function of one variable. The second term looks like $\varphi'(u)$ in the limit, and the first term looks like f' evaluated at $\varphi(u)$. As the two limits exist, the limit of the product is the product of the limits, so we can conclude:

$$h'(u) = f'(\varphi(u))\,\varphi'(u)$$

Why isn't this proof rigorous? The definition of $f'(z)$ is the following:

$$f'(z) = \lim_{w \to z} \{f(w) - f(z)\} / \{w - z\}$$

We cheated in the above: this limit has to hold *FOR ALL* paths where w heads to z. We didn't consider *all* paths, only a special path. But maybe this isn't too bad: if the limit exists, then it doesn't matter *WHICH* path we take. In better words: look, I know $f'(z)$ exists, and I know the value is *INDEPENDENT* of the path I take. So why don't I just make life easy on myself and take this nice path? What a great idea! We leave for the interested, rigorous reader what to do if $\varphi(y)$ equals $\varphi(u)$ infinitely often (this cannot happen if $\varphi'(u) \neq 0$). Hint: go back to the definition of $\partial h/\partial u$ and calculate it directly, going along points where $\varphi(y) = \varphi(u)$.

(III). Higher Dimensions:

We now argue as in above, but in higher dimensions. To make things easier to view, let's just look at $n = 3$, $m = 2$, so we have (x_1, x_2, x_3), which we denote by (x, y, z) for convenience, and (u_1, u_2), which we denote by (u, w).

$$h(u,w) = f(x(u,w), y(u,w), z(u,w))$$

We calculate $\partial h/\partial u$, at the point (u,w), and compare with $\partial h/\partial u_1$ from page 3.

$$\partial h/\partial u = \lim_{v \to u} \{ h(v, w) - h(u, w) \} / \{ v - u \}$$

$$= \lim_{v \to u} \frac{f(x(v,w), y(v,w), z(v,w)) - f(x(u,w), y(u,w), z(u,w))}{v - u}$$

So, we start at the point (x(u,w), y(u,w), z(u,w)) and we finish at the point (x(v,w), y(v,w), z(v,w)). We cannot directly mimic the 1-dimensional case, but what if our starting point were (x(u,w), y(v,w), z(v,w))? Then all we would've done is change the x-coordinate of the 3-tuple, and we could multiply and divide by x(v,w) - x(u,w). We would then have:

$$\partial f/\partial x \quad \partial x/\partial u$$

Sadly, life isn't quite that simple: we don't have that as our starting point. But, what if we added and subtracted f(x(u,w), y(v,w), z(v,w)) in the numerator? Then we would get:

$$\frac{\partial h}{\partial u} = \lim_{v \to u} \frac{f(x(v,w),y(v,w),z(v,w)) - f(x(u,w),y(v,w),z(v,w))}{v - u} +$$

$$\lim_{v \to u} \frac{f(x(u,w),y(v,w),z(v,w)) - f(x(u,w),y(u,w),z(u,w))}{v - u}$$

We now multiply the first term by 1:

$$\frac{\partial h}{\partial u} = \lim_{v \to u} \frac{f(x(v,w),y(v,w),z(v,w)) - f(x(u,w),y(v,w),z(v,w))}{x(v,w) - x(u,w)} * \frac{x(v,w) - x(u,w)}{v - u}$$

$$+ \lim_{v \to u} \frac{f(x(u,w),y(v,w),z(v,w)) - f(x(u,w),y(u,w),z(u,w))}{v - u}$$

$$\frac{\partial h}{\partial u} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial u} + \lim_{v \to u} \frac{f(x(u,w),y(v,w),z(v,w)) - f(x(u,w),y(u,w),z(u,w))}{v - u}$$

Now we just repeat what we did before! We've got two points, start at (x(u,w),y(u,w),z(u,w)), end at (x(u,w),y(v,w),z(v,w)). Again, what if our first point were (x(u,w),y(u,w),z(v,w))? Then all we would've done is change the y-coordinate of the 3-tuple, and we could multiply and divide by y(v,w) - y(u,w). We would then (in the limit) get $\partial f/\partial y\ \partial y/\partial u$, plus another term, the difference of the point we added and our *true* first point. Let's do it!

$$\frac{\partial h}{\partial u} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial u} + \lim_{v \to u} \frac{f(x(u,w),y(v,w),z(v,w)) - f(x(u,w),y(u,w),z(v,w))}{v - u}$$

$$+ \lim_{v \to u} \frac{f(x(u,w),y(u,w),z(v,w)) - f(x(u,w),y(u,w),z(u,w))}{v - u}$$

Multiplying the first limit by {y(v,w) - y(u,w)} / {y(v,w) - y(u,w)} we get:

$$\frac{\partial h}{\partial u} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial u} + \lim_{v \to u} \frac{f(x(u,w),y(u,w),z(v,w))-f(x(u,w),y(u,w),z(u,w))}{v - u}$$

Multiplying the last term by {z(v,w) - z(v,w)} / {z(v,w) - z(v,w)}, we get that this term, in the limit, is just $\partial f/\partial z\ \partial z/\partial u$.

Hence we get:

$$\boxed{\frac{\partial h}{\partial u} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial u} + \frac{\partial f}{\partial z}\frac{\partial z}{\partial u} \quad \textbf{which is The Chain Rule!}}$$

SECTION 2.6: GRADIENTS AND DIRECTIONAL DERIVATIVES

• **Gradient**: $\text{grad}(f) = \nabla f = \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}\right)$

↳ This is derivative Df written as a vector

• **Directional Deriv**: Directional derivative of f at $\vec{x}$ along vector $\vec{v}$ (usually unit length) is $\frac{d}{dt} f(\vec{x} + t\vec{v})$. If $\|\vec{v}\| = 1$ say the directional derivative in the direction of $\vec{v}$.

↳ if $\|\vec{v}\| \neq 1$, changing scale

↳ Equivalent to $\lim_{h \to 0} \frac{f(\vec{x} + h\vec{v}) - f(\vec{x})}{h}$

THM: $f: \mathbb{R}^n \to \mathbb{R}$ diff. Then all derivs exist and the dir deriv in the dir of $\vec{v}$ is $Df(\vec{x})\vec{v} = \nabla f(\vec{x}) \cdot \vec{v} = \frac{\partial f}{\partial x_1}(\vec{x}) v_1 + \cdots + \frac{\partial f}{\partial x_n}(\vec{x}) v_n$

Proof: Let $c(t) = \vec{x} + t\vec{v}$ and use chain rule

Note $c'(0) = \vec{v}$, $c(0) = \vec{x}$

Thus $\frac{d}{dt} f(c(t)) = Df(c(0))\, c'(t) = \nabla f(c(0)) \cdot \vec{v}$ ∎

THM: GEOMETRIC INTERPRETATION:

If $\nabla f(\vec{x}_0) \neq \vec{0}$ then $\nabla f(\vec{x}_0)$ points in dir of fastest increase of f

Proof: Rate of change of f in unit dir $\vec{n}$ is $\nabla f(\vec{x}_0) \cdot \vec{n}$

Have magnitude $\|\nabla f(\vec{x}_0)\| \cdot \|\vec{n}\| \cdot |\cos\theta|$, largest when $\theta = 0, \pi$ so parallel ($\theta = 0$ gives max, $\theta = \pi$ gives min)

SECTION 2.6 (CONT)

THM: Gradient is normal to level surfaces. Specifically, $f: \mathbb{R}^n \to \mathbb{R}$ a C^1 map, $\vec{x}_0$ in level surface S defined by $f(\vec{x}) = k$. If curve $c(t)$ in S with $c(0) = \vec{x}_0$ and $\vec{v} = c'(t)$ is the tangent vector at $t=0$, then $\nabla f(\vec{x}_0) \cdot \vec{v} = \vec{0}$

Proof: Chain rule again!

Apply to $h(t) = f(c(t)) = k$ ∎

Note: These results will be VERY useful for max/min problems

Defn: Tangent Plane: S be surface $f(\vec{x}) = k$. The tangent plane at $\vec{x}_0$ is defined by $\{\vec{x} : \nabla f(\vec{x}_0) \cdot (\vec{x} - \vec{x}_0) = 0\}$

Often call ∇f the gradient vector field; means at point $\vec{P}$ draw vector $\nabla f(\vec{P})$.

↳ Example: Gravity: $\vec{F}(x,y,z) = -G\frac{m_1 m_2}{r^2}\vec{n} = \nabla\left(\frac{G m_1 m_2}{r}\right)$

where $\vec{n} = \vec{r}/r$, $\vec{r} = (x,y,z)$

Homework: #2ab, #4a, #6a, #16 (Ralph), #18

Suggested: #5a, #12, #17, #21, #23

Review Problems

HW: Pg 176: #23 homog
#47 chemistry

Suggested: Pg 176: #26, #41
#42