MAT257 Lecture Notes

ARKY!! :3C

'24 Fall & '25 Winter Semester

Contents

§1 Day 1: Course Administrative Details and Equivalent Norms (Sep. 4, 2024)

Course administrative details!

- The prerequisites to this class are calculus and linear algebra; specifically, this class is about calculus on manifolds.
- The main three theorems this class will cover are the inclusive function theorem, change of variable in integral, and Stokes' theorem (generalization of FTC).
- The main techniques we will study this class are manifolds, theory of integration, and differential forms.

To give an example of how linear algebra can be used in defining differentiability, we start by recalling that for a function $f : \mathbb{R} \to \mathbb{R}$ to be differentiable in 1 variable, we need

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

to exist for any $a \in \mathbb{R}$. Specifically, this may be rewritten as

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

where we have $f(a) + f'(a)h$ as the best local linear approximation of $f(a+h)$ at $h = 0$, and $h \mapsto ch$ is a linear transformation $\mathbb{R} \to \mathbb{R}$. For the higher dimensional analogue of the above, let us take $f : \mathbb{R}^n \to \mathbb{R}^m$, and have

$$\begin{aligned} a &= (a_1, \ldots, a_n), \\ x &= (x_1, \ldots, x_n), \\ h &= (h_1, \ldots, h_n). \end{aligned}$$

If we let $y = f(x)$, then $y \in \mathbb{R}^m$, and we can write it as a column vector with components $y_1, \ldots, y_m$;

$$f(x) = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} f_1(x_1, \ldots, x_n) \\ \vdots \\ f_n(x_1, \ldots, x_n) \end{pmatrix}.$$

In this way, each y_i for $1 \leq i \leq n$ can be viewed as a function in n variables. We say f is differentiable at $a \in \mathbb{R}^n$ if there exists a linear transformation $\lambda : \mathbb{R}^n \to \mathbb{R}^m$ such that[1]

$$\lim_{h \to 0} \frac{|f(a+h) - (f(a) + \lambda h)|}{|h|} = 0$$

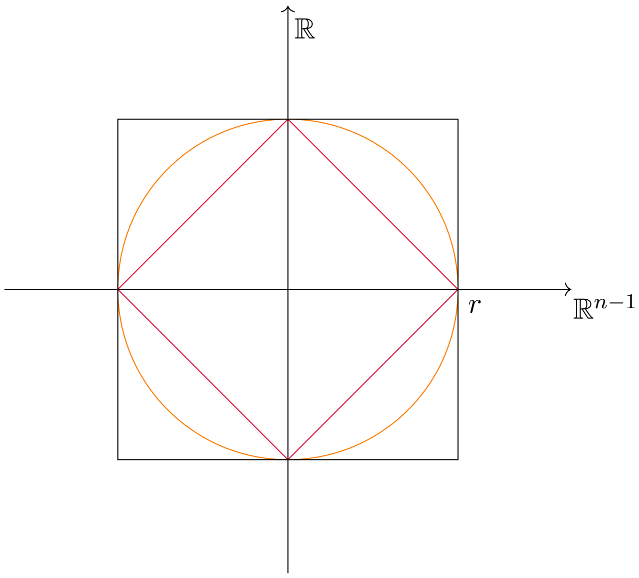
In this way, we have that $f(a) + \lambda h$ is the best linear approximation of $f(a+h)$ at $h = 0$. Notice that compared to the $\mathbb{R} \to \mathbb{R}$ differentiability condition, we impose a norm on $f(a+h) - (f(a) + \lambda h)$ and h, since we cannot divide by $\mathbb{R}^n, \mathbb{R}^m$ vectors.

[1] i think we need a norm in the numerator for the limit below; bierstone didn't do it in class though, so... idk. also lambda is jacobian?

This also brings the topic of what sort of norms we could be looking at; for example, we have[2]

$$
\begin{aligned}
|x| &= \sqrt{x_1^2 + \cdots + x_n^2}, \\
\|x\| &= \max\{|x_1|, \ldots, |x_n|\}, \\
|||x||| &= |x_1| + \cdots + |x_n|.
\end{aligned}
$$

We say that two norms p, q are equivalent to each other if we may find a constant C where $p(x) \le Cq(x)$ (and vice versa). In fact, the norms $|x|, \|x\|, |||x|||$ are all equivalent to each other; here is a visual proof,



where each figure is the level sets of their respective norms. In particular, $|x| \le \sqrt{n}\,\|x\|$, and $|||x||| \le \sqrt{n}\,|x|$. To prove the latter analytically, we may write the norm as an inner product, $|||x||| = \langle x, u(x)\rangle < |x|\,|u(x)| \le \sqrt{n}r$, where u_i is the sign of each component of x.[3]

[2] these are (in order) euclidean, maximum, and taxicab norms; or $\ell^2, \ell^\infty, \ell^1$

[3] i kinda get where this proof is going, but at the same time i dont really know how he defined u in the first place, so yeah. alternate proof would be to give Cauchy-Schwarz

§2 Day 2: Review of MAT247, Metric Spaces (Sep. 6, 2024)

Course administrative details!

- Fall Office Hours will be held on Mondays from 10:30 to 11:30AM (likely in Bahen 6114).

We start with a review on functions and continuity. Given $\mathbb{R}^n$ with elements $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$, we define the following addition and scalar multiplication,

$$\begin{aligned} x + y &= (x_1 + y_1, \ldots, x_n + y_n), \\ \lambda x &= (\lambda x_1, \ldots, \lambda x_n), \lambda \in \mathbb{R}. \end{aligned}$$

We may also equip it with the Euclidean inner product and norm,

$$\begin{aligned} \langle x, y \rangle &= x_1 y_1 + \cdots + x_n y_n, \\ |x| &= (x_1^2 + \cdots + x_n^2)^{\frac{1}{2}} = \sqrt{\langle x, x \rangle}, \end{aligned}$$

of which it has the properties

(a) Non-negativity; $|x| \geq 0$, and is equal to 0 if and only if $x = 0$.

(b) Absolute Homogeneity; $|\lambda x| = |\lambda| \, |x|$ for any scalar $\lambda \in \mathbb{R}$.

(c) Triangle Inequality; $|x + y| \leq |x| + |y|$ for any vectors $x, y \in \mathbb{R}^n$.

We may check that the properties above hold for the Euclidean norm;

(a) Trivial, since the square of each component is non-negative.

(b) Not discussed in class, but we may directly expand the norm to get the equality,

$$|\lambda x| = \left(\sum_{i=1}^{n} (\lambda x_i)^2 \right)^{\frac{1}{2}} = \left(\lambda^2 \sum_{i=1}^{n} x_i^2 \right)^{\frac{1}{2}} = |\lambda| \, |x| \, .$$

(c) Start by observing that from inner product properties, we have $|\langle x, y \rangle| \leq |x| \, |y|$, with equality if and only if x, y are linearly dependent. If x, y are LI, observe $\lambda y - x \neq 0$ for all scalars $\lambda \in \mathbb{R}$, and we may write

$$\begin{aligned} 0 < |\lambda y - x|^2 &= \langle \lambda y - x, \lambda y - x \rangle \\ &= \lambda^2 \langle y, y \rangle - 2\lambda \langle x, y \rangle + \langle x, x \rangle \\ &= |y|^2 \lambda^2 - 2 \langle x, y \rangle \lambda + |x|^2 \, . \end{aligned}$$

Reading the above as a polynomial in λ, we observe that it must not have any real roots, and so the discriminant is necessarily negative, i.e. $4 \langle x, y \rangle^2 - 4 |x|^2 |y|^2 < 0$, which implies $|\langle x, y \rangle| < |x| \, |y|$. Returning to the triangle inequality, let us start by squaring both sides to obtain

$$|x + y|^2 = \langle x, x \rangle + 2 \langle x, y \rangle + \langle y, y \rangle \leq |x|^2 + 2 |x| \, |y| + |y|^2 \, .$$

This leaves us with $2 \langle x, y \rangle \leq 2 |x| \, |y|$, which, from the above, we find equality if x, y are linearly dependent, and "$<$" if LI.

Consider the linear transformation, $T : \mathbb{R}^n \to \mathbb{R}^m$, equipped with standard bases $\{e_1, \ldots, e_n\}$ of $\mathbb{R}^n$, and $\{f_1, \ldots, f_m\}$ of $\mathbb{R}^m$; let the matrix representation of T be given by the below matrix A,

$$y = T(x) \implies \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = \underbrace{\begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \ldots & a_{mn} \end{pmatrix}}_{A} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

In terms of basis vectors, we may visualize T by

$$T(e_i) = \sum_{j=1}^{m} a_{ji} f_j.$$

Now, suppose $S : \mathbb{R}^m \to \mathbb{R}^\ell$ be a linear transformation, with corresponding matrix representation $B \in M_{\ell\times m}(\mathbb{R})$. Then the composition $S \circ T : \mathbb{R}^n \to \mathbb{R}^\ell$ has matrix $B \cdot A \in M_{\ell\times n}(\mathbb{R})$.

Here is an example from class; if $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear map, then $|Tx| \leq M\,|x|$ for some $M > 0$. To see this, let us start by writing

$$Tx = T\left(\sum_{i=1}^{n} x_i e_i\right) = \sum_{i=1}^{n} x_i T(e_i).$$

This gives us

$$\begin{aligned} |Tx| \leq \sum_{i=1}^{n} |x_i T(e_i)| &= \sum_{i=1}^{n} |x_i|\,|T(e_i)| \\ &= \left(\sum_{i=1}^{n} a_{ji}^2\right)^{\frac{1}{2}} \\ &\leq \left(\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2\right)^{\frac{1}{2}} =: |A|\,. \end{aligned}$$

Then we have

$$|Tx| \leq \left(\sum_{i=1}^{n} |x_i|\right)|A| \leq \sqrt{n}\,|A|\,|x|\,.$$

We simply pick $M \geq \sqrt{n}\,|A|$ and we are done. □

§2.1 Topology of $\mathbb{R}^n$

Let us equip $\mathbb{R}^n$ with the Euclidean norm; this is an example of a metric space (i.e., a set equipped with a distance function $d(x, y)$). The function $d : X \times X \to \mathbb{R}$ has the following properties,

(a) Symmetry, $d(x, y) = d(y, x)$.

(b) Non-negativity, $d(x, y) \geq 0$, with $d(x, y) = 0$ if $x = y$.

(c) Triangle inequality, $d(x, y) \leq d(x, z) + d(z, y)$.

§3 Day 3: Topology in $\mathbb{R}^n$ (Sep. 9, 2024)

What are some n-dimensional analogues of closed intervals $[a, b] \in \mathbb{R}$? We have

- The closed rectangle $[a_1, b_1] \times \cdots \times [a_n, b_n] \subset \mathbb{R}^n$,
- The closed ball $\{x \in \mathbb{R}^n \mid |x - a| \leq r\}$.

For n-dimensional analogues of open intervals, we have open rectangles and balls, i.e.

- $(a_1, b_1) \times \cdots \times (a_n, b_n)$,
- $\{x \in \mathbb{R}^n \mid |x - a| < r\}$.

We say a subset $U \subseteq \mathbb{R}^n$ is open (two definitions) if:

- For any $a \in U$, we may pick $\varepsilon > 0$ such that the ball $B(a, \varepsilon) \subseteq U$.
- For any $a \in U$, there exists an open rectangle R such that $a \in R \subseteq U$.

We say a subset $C \subseteq \mathbb{R}^n$ is closed if $\mathbb{R}^n \setminus C$ is open. Here are some examples of closed sets,

- $\emptyset, \mathbb{R}^n$;
- Closed rectangles and balls as per earlier;
- Finite sets.

Let us take the closed subset $A \subseteq \mathbb{R}^n$, and have A contain all rationals in $(0, 1)$. We claim that $[0, 1] \subseteq A$. To see this, consider $x \in \mathbb{R}^n \setminus A$. Since A is closed, $\mathbb{R}^n \setminus A$ is open, which means there exists $\varepsilon > 0$ such that $B(x, \varepsilon) \subseteq A$. By density of $\mathbb{Q}$ in $\mathbb{R}$, we may always find a rational in $B(x, \varepsilon)$, meaning $B(x, \varepsilon) \cap A = \emptyset$, and $x \notin [0, 1]$. $\square$

A few remarks;

- Any union of open sets is open.
- The finite union of closed sets is closed.
- Arbitrary unions of closed sets are not necessarily closed; observe

$$\bigcup_{n \in \mathbb{N}} \left[\frac{1}{n} - 1, 1 - \frac{1}{n}\right] = (-1, 1).$$

Now, consider $A \subseteq \mathbb{R}^n$, and $x \in \mathbb{R}^n$; there are 3 possibilities:

1. There exists an open ball B such that $x \in B \subset A$ (i.e. the interior of A, $\operatorname{int} A$).
2. There exists an open ball B such that $x \in B \subset \mathbb{R} \setminus A$ (i.e. the exterior of A, $\operatorname{ext} A$).
3. For all open balls B such that $x \in B$, there exists $y_1 \in A$ and $y_2 \in \mathbb{R}^n \setminus A$ such that $y_1, y_2 \in B$ (i.e. the boundary of A, $\operatorname{bdry} A$).

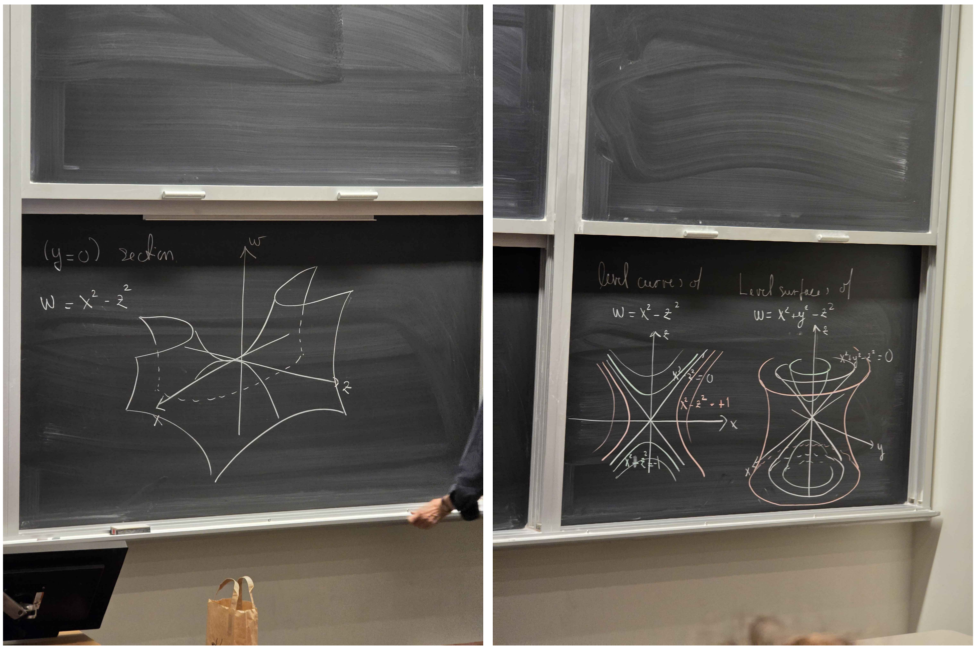
For example, we may consider $A \subset \mathbb{R}^n$ to be the rationals in $(0, 1)$; then the interior of A is $\emptyset$, the exterior of A is $\mathbb{R}^n \setminus [0, 1]$, and the boundary of A is $[0, 1]$.

Let $A \subset \mathbb{R}^n$; given the function $f : A \to \mathbb{R}^m$ then we define the graph of f,

$$\operatorname{graph} f = \{(x, f(x)) \mid x \in A\} \subseteq A \times \mathbb{R}^m.$$

§4 Day 4: Graphing Example; Operations on Functions, Definition of Continuity by Open Sets (Sep. 11, 2024)

We graphed $w = x^2 + y^2 - z^2$ in class today. Since I don't really know how to LaTeX these kinds of graphs, I'll just drop these pictures in;



The main idea was to start by considering $z = 0$ and observing that $w = x^2 + y^2$ is really a parabola ($w = x^2$, $w = y^2$) rotated about the w-axis; setting x or y to 0, we get the left picture. If we examine the level sets of $w = x^2 - z^2$, we may combine the two to obtain the level surfaces of $w = x^2 + y^2 - z^2$ (as per the rightmost diagram on the blackboard).

Let $A \subset \mathbb{R}^n$, and consider a function $f : A \to \mathbb{R}^m$ (i.e., a function $f : \mathbb{R}^n \to \mathbb{R}^m$ with domain restricted onto A). If we want to create another function $g : B \to \mathbb{R}^p$ to be composed with A, we implicitly ask that $f(A) \subset B$; with this, we may write

$$(g \circ f)(x) = g(f(x))$$

where $\mathrm{dom}(g \circ f) = f^{-1}(B)$. Now, let us consider the inner product $\langle , \rangle : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$. Let f, g be functions from $\mathbb{R}^n$ to $\mathbb{R}^m$. Then we may construct

$$(f, g)(x) : \mathbb{R}^n \to \mathbb{R}^m \times \mathbb{R}^m,$$
$$f \cdot g = \langle , \rangle \circ (f \cdot g),$$

which we may indeed check sends $\mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$ as expected of the definition of dot product.

Returning to earlier, let us consider $f : (A \subset \mathbb{R}^n) \to \mathbb{R}^m$, and let us consider $f(x)$ written in its components, $f(x) = (f_1(x), \dots, f_m(x))$. If we wish to be specific, observe that we may write each f_i as the following composition,

$$f_i = \pi_i \circ f \text{ where } \pi_i : \mathbb{R}^m \to \mathbb{R},$$

where π_i is the mapping $(x_1, \dots, x_m) \mapsto x_i$.

On the topic of limits, recall from MAT157 that $\lim_{x\to a} f(x) = b$ means that for every $\varepsilon > 0$, there exists $\delta > 0$ such that if x is in a δ-ball of a, then $f(x)$ is in an ε-ball of b. We may extend this idea to define continuity; we say f is continuous at a if

$$\lim_{x\to a} f(x) = f(a),$$

and that f is a continuous function if it is continuous for all $a \in A$. If we want to define continuity in its topological notion, though, we have that $f : \mathbb{R}^n \to \mathbb{R}^m$ is continuous if and only if $f^{-1}(U)$ is open in $\mathbb{R}^n$ for all $U \subset \mathbb{R}^m$.

Theorem 4.1 (Spivak 1.8)**.** We say $f : (A \subset \mathbb{R}^n) \to \mathbb{R}^m$ is continuous if and only if, for all $U \subset \mathbb{R}^m$, $f^{-1}(U) = A \cap V$ for some open set $V \subset \mathbb{R}^n$.[4]

($\Rightarrow$) Consider any open set $U \subset \mathbb{R}^m$. If $a \in f^{-1}(U)$, we have $b := f(a) \in U$. Since U is open, there exists some $\varepsilon > 0$ such that $B(a, \varepsilon) \subset U$; using the fact that f is continuous at a, we may construct a δ-ball $B(a, \delta(a))$ about a such that

$$A \cap B(a, \delta(a)) \overset{f}{\mapsto} B(f(a), \varepsilon).$$

With this, we may take the union of all such balls $B(a, \delta(a))$ and observe that

$$f^{-1}(U) = A \cap \underbrace{\left(\bigcup_{a\in f^{-1}(U)} B(a, \delta(a)) \right)}_{:=V},$$

where we may note V is open (since the arbitrary union of open sets is open). We may note that $A \cap V$ indeed covers $f^{-1}(U)$; if it did not, then we would be able to pick a new $a \in f^{-1}(U)$ and repeat the above process, contradicting the definition of V.

($\Leftarrow$) For any $a \in A$, let us have $U = B(f(a), \varepsilon)$ for any $\varepsilon > 0$. Then $f^{-1}(U) = A \cap V$ for some open set $V \subset \mathbb{R}^n$; by definition of open sets, we may find a ball about a contained in V; let it be $B(a, \delta)$. Then this fulfills the $\varepsilon - \delta$ definition of continuity, and we are done. $\square$

[4] do note that this proof is different from lecture, since i either mis-transcribed or bierstone made a mistake concerning balls for all $a \in A$ sending to b.

§5 Day 5: Composition of Continuous Functions; Uniform Continuity, Distance Metric, Compactness (Sep. 13, 2024)

We start with a few examples:

- Is $f(x,y) = \frac{x^2-y^2}{x^2+y^2}$ continuous? The answer is no, since the limit on $(x,y) \to 0$ on ℓ_x (read: X-axis) is equal to -1, while the limit on ℓ_y is equal to 1. This means there are two conflicting limits approaching $(x,y) \to 0$, meaning that f is not continuous at 0.
- Is $f(x,y) = e^{-\frac{|x-y|}{x^2-2xy+y^2}}$ continuous? Observe that

$$e^{-\frac{|x-y|}{x^2-2xy+y^2}} = e^{-\frac{1}{|x-y|}};$$

since $|x-y|$ and $e^{-\frac{1}{x}}$ are continuous, we see that the composition is continuous as well, and so $f(x,y)$ is continuous.

While the composition of continuous functions property has been proved in 327, we provide a 257 variant of the proof;

Theorem 5.1 (Composition of Continuous Functions is Continuous)**.** Let $f : (A \subset \mathbb{R}^n) \to \mathbb{R}^m$ and $g : (B \subset \mathbb{R}^m) \to \mathbb{R}^p$ be continuous, with $f(A) \subset B$. Then we have that $g \circ f$ is continuous.

Let U be open in $\mathbb{R}^p$; then $g^{-1}(U) = B \cap V$ where V is open in $\mathbb{R}^m$ (as per our definition of continuity). Furthermore, we also have $(g \circ f)^{-1}(U) = f^{-1}(B \cap V) = f^{-1}(B) \cap f^{-1}(V)$; let $f^{-1}(V) = A \cap W$ for some open $W \subset \mathbb{R}^n$, then we have

$$f^{-1}(B) \cap f^{-1}(V) = \underbrace{f^{-1}(B) \cap A}_{f^{-1}(B) \supset A} \cap W = A \cap W,$$

which is as desired. $\square$

We now provide examples of continuity.

1. Let us have a linear function $T : \mathbb{R}^n \to \mathbb{R}^m$ such that T is uniformly continuous. Then

$$|T(x) - T(y)| = |T(x-y)| \leq C\,|x-y|$$

for some scalar C. In particular, when constructing an $\varepsilon - \delta$ proof for continuity here, for any $\varepsilon > 0$ we may pick $\delta < \frac{\varepsilon}{C}$.

2. Let $f = (f_1, \ldots, f_m) : \mathbb{R}^n \to \mathbb{R}^m$; we say that f is continuous if and only if each of f_i for $1 \leq i \leq m$ are continuous as well.

   ($\Rightarrow$) By contrapositive; if any f_i is discontinuous at any point a, then f is discontinuous at a as well.

   ($\Leftarrow$) If every f_i is continuous, there exists δ_i for each f_i such that $|f_i(x) - f_i(a)| < \frac{\varepsilon}{\sqrt{n}}$ whenever $|x-a| < \delta_i$. Then we may set $\delta = \min\{\delta_1, \ldots, \delta_m\}$ to see

$$|f(x) - f(a)|^2 = \sum_{i=1}^{m} |f_i(x) - f_i(a)| < n\left(\frac{\varepsilon}{\sqrt{n}}\right)^2 = \varepsilon^2,$$

   yielding $|f(x) - f(a)| < \varepsilon$ whenever $|x-a| < \delta$. $\square$

In terms topological definitions though, we may write

($\Rightarrow$) If f is continuous, then we may use the composition of continuous functions to see that the projections $f_i = \pi_i \circ f$ are indeed continuous.

($\Leftarrow$) Take any open subset $U \subset \mathbb{R}^m$. Then let R be an open rectrangle in U, and let us consider the union of all possible rectangles,

$$f^{-1}(U) = \bigcup_{R \in U} f^{-1}(R).$$

Since each rectangle is defined as the cartesian product of open intervals on each respective f_i, we see that $f^{-1}(R)$ is given by

$$f^{-1}(R) = \bigcap_{i=1}^{m} f_i^{-1}(R_i),$$

where R_i are the said respective open intervals. Since each f_i is continuous and the intersection of open sets is open, we see $f^{-1}(R)$ is open for all R, and so $f^{-1}(U)$ is open as well. $\square$

3. Let $X \subset \mathbb{R}^n$, and define the metric $d(x, X) = \inf_{a \in X} |x - a|$, i.e. the smallest distance from x to some point $a \in X$. We want to show that $f(x) = d(x, X)$ is uniformly continuous on $\mathbb{R}^n$.

   To start, let us consider $|f(x) - f(y)| = |d(x, X) - d(y, X)| \leq |x - y|$; we wish to prove the inequality. To do this, start by taking $d(x, X) - |x - y|$, and consider

$$\begin{aligned} d(x, X) - |x - y| &\leq d(x, X) + \underbrace{|y - a| - |x - a|}_{\text{Triangle Ineq.}} \\ &\leq |y - a| && (d(x, X) - |x - a| \leq 0) \\ &\leq d(y, X). \end{aligned}$$

   In this way, we get $|x - a| \leq |x - y| + |y - a|$, which yields $d(x, X) - d(y, X) \leq |x - y|$. $\square$

We also briefly touched on compactness at the end of the class. We call a subset $X \subset \mathbb{R}^n$ *compact* if every open covering $\mathcal{O}$ of X has a finite subcovering (i.e., a subset of $\mathcal{O}$ that covers X). Here are some examples,

- $\mathbb{R}$ (equipped with the standard topology) is not compact. If we let $\mathcal{O}$ be an open covering given by
$$\mathcal{O} = \{(a, a+1) \mid a \in \mathbb{R}\},$$
  we have that $\mathcal{O}$ covers $\mathbb{R}$, but there does not exist a finite subcovering. Thus, $\mathbb{R}$ cannot be compact.

- The open interval $(0, 1)$ (once again, equipped with the standard topology) is not compact. This time, let
$$\mathcal{O} = \left\{ \left(\frac{1}{n}, 1 - \frac{1}{n} \right) \mid n \in \mathbb{N} \right\}.$$
  Clearly, $\mathcal{O}$ covers X, but it does not admit a finite subcovering.

- Any topology on a finite set X is compact. In particular, any covering is necessarily finite, since there are finitely many elements in $\mathcal{P}(X)$; this means any subcovering, even if it is the same as the covering, is finite as well.

§6 Day 6: Compactness (Sep. 16, 2024)

We start by giving some properties on compactness on $\mathbb{R}^n$.

Theorem 6.1 (Compactness $\iff$ Closed and Bounded)**.** A subset $X \subset \mathbb{R}^n$ is compact if and only if it is also closed and bounded.

We prove both directions now.

($\Rightarrow$) Suppose $X \subset \mathbb{R}^n$ is compact. Then consider an open cover of X by open balls of radius 1; then we may find a finite subcover (by definition of compactness). Since there are a finite number of open balls of radius 1 covering X, all points are finitely close to each other.

We now show that X is closed, i.e. $\mathbb{R}^n \setminus X$ is open. Take any $a \in \mathbb{R}^n \setminus x$. We wish to find $\delta > 0$ such that $B(a, \delta) \subset \mathbb{R}^n \setminus X$; then we may consider the closed balls of radius $\frac{1}{k}$ given by $\overline{B}(a, \frac{1}{k})$ for $k = 1, 2, \ldots,$ and denote

$$U_k = \mathbb{R}^n \setminus \overline{B}\left(a, \frac{1}{k}\right).$$

Clearly, $\bigcup_k U_k = \mathbb{R}^n \setminus \{a\}$, and so U_k forms an open covering of X, and so there exists a finite subcover of it. Thus, $X \subset U_k$ for some k, where $B(a, \frac{1}{k}) \subset \mathbb{R}^n \setminus X$. We conclude that X is closed.

($\Leftarrow$) Will be done next lecture (?)

Theorem 6.2 (Continuous Image of Compact is Compact)**.** Let $X \subset \mathbb{R}^n$ be compact, and consider a continuous function $f : X \to \mathbb{R}^n$. Then $f(x)$ is compact.

To start, let $\mathcal{O} = \{U\}$ (read: multiple open sets U) be an open cover of $f(X)$. For every $U \in \mathcal{O}$, consider by continuity we have

$$f^{-1}(U) = X \cap V_U,$$

where V_U is some open set in $\mathbb{R}^n$. Then $\{V_U\}_{U \in \mathcal{O}}$ is an open covering of X. Since X is compact, we may write

$$X \subset V_{U_1} \cup \cdots \cup V_{U_k}$$

for some open sets $U_1, \ldots, U_k$ in the covering $\mathcal{O}$. Therefore, $f(X) \subset U_1 \cup \cdots \cup U_k$, which is indeed a finite covering. $\square$

Theorem 6.3 (Extreme Value Theorem)**.** A continuous function $f : X \to \mathbb{R}$ on a compact $X \subset \mathbb{R}^n$ takes on a minimum and maximum value.

Since $f(X)$ is compact, we know it is closed and bounded; let $M = \sup\{f(x) \mid x \in X\} < \infty$. If $M \notin f(X)$, then there is an open interval around M outside $f(X)$ (since the complement of $f(X)$ is open), contradicting that M is the supremum; thus, f attains M at some point, and we may consider $-f(X)$ to obtain the infimum / minimum. $\square$

Theorem 6.4 (ε-neighborhood Theorem)**.** If we have a compact X in an open set $U \subset \mathbb{R}^n$, then there is $\varepsilon > 0$ such that the ε-neighborhood of X in $\mathbb{R}^n$ lies in U; specifically, the ε-neighborhood of X can be defined as

$$\{y \in \mathbb{R}^n \mid d(y, X) < \varepsilon\} = \bigcup_{x \in X} B(x, \varepsilon). \qquad \textit{(Left as exercise)}$$

To prove this, let $f(x) = d(x, \mathbb{R}^n \setminus U)$ be continuous (continuity has been proven previously I think). Then $f(x) > 0$ for all $x \in X$, because we may always pick a ball centered at x in U. Since X is compact, f has a minimum value at $\varepsilon > 0$; this means the ε-neighborhood of X lies in U. $\square$

§7 Day 7: Uniform Continuity, Hard Direction of Heine-Borel (Sep. 18, 2024)

Theorem 7.1 (Cont. Function on Compact Set is Uniformly Cont.)**.** A continuous function $f : (X \subset \mathbb{R}^n) \to \mathbb{R}^n$, where X is compact, is uniformly continuous. [5]

Recall that uniform continuity means that for all $\varepsilon > 0$, there exists some one-size-fits-all $\delta > 0$ such that $|f(x) - f(y)| < \varepsilon$ whenever $|x - y| < \delta$ for any $x, y \in X$.

Let us have $X \times X \subset \mathbb{R}^n \times \mathbb{R}^n$. To start, the diagonal $\Delta = \{(x, x) \mid x \in X\}$ is compact, because Δ is the image of X under the map $x \mapsto (x, x)$, which is a continuous function. Thus, we have that $g : X \times X \to \mathbb{R}$ where $g(x, y) = |f(x) - f(y)|$ is continuous, as per the composition of continuous functions.

Given $\varepsilon > 0$, consider $g^{-1}((-\varepsilon, \varepsilon)) = (X \times X) \cup U$ where U is an open set in $\mathbb{R}^n \times \mathbb{R}^n$; clearly, $\Delta \subset U$, since Δ is compact and U is open. By the ε-neighborhood theorem, there exists $\delta > 0$ such that the δ-neighborhood of Δ is in U. Then consider x, y such that $|x - y| < \delta$, and observe we have

$$|(x, x) - (x, y)| \leq |x - y| + |y - y| < \delta$$

by the triangle inequality. This means (x, y) is in a δ-neighborhood of Δ, and so is in U. By construction, we see that $(x, y) \in g^{-1}((-\varepsilon, \varepsilon))$, and we conclude $g((x, y)) = |f(x) - f(y)| < \varepsilon$. $\square$

Theorem 7.2 (Closed Interval is Compact)**.** A closed interval $X = [a, b] \subset \mathbb{R}$ is compact.

Consider an open cover $\mathcal{O}$ of $[a, b]$, and let A be the set of all $x \in [a, b]$ such that $[a, x]$ can be covered by finitely many sets in $\mathcal{O}$. Then we want to show that $a, b \in A$, and that A is bounded above by b.

Let $\alpha = \sup A$; we start by showing that $\alpha \in A$. First, observe that $\alpha \in [a, b]$, meaning $\alpha \in U$ for some $U \in \mathcal{O}$. Since U is open, we may find a δ-ball around α in U, i.e. $(\alpha - \delta, \alpha + \delta) \subset U$. Since α is the supremum, there must exist some x in the interval to the left of α such that $[a, x]$ is covered by $U_1, \ldots, U_k \in \mathcal{O}$. This means $[a, \alpha]$ is covered by $U \cup \bigcup_{i=1}^k U_i$.

Now, we show that $\alpha = b$. In the opposite direction to the above, suppose we pick $x' \in (\alpha, \alpha + \delta) \cap [a, b]$ (i.e., to the right of α). Since we know $(a - \delta, a + \delta)$ is covered by U and $[a, x]$ is covered by finitely many sets in $\mathcal{O}$, $[a, x']$ is covered by $U \cup \bigcup_{i=1}^k U_i$ as well, which would contradict that α is the supremum of A. The only situation in which there is no contradiction is if $\alpha = b$, since x' would be at most b in this case. $\square$

Lemma 7.3 (Closed Rectangles in $\mathbb{R}^n$ are Compact)**.** Closed rectangles $R = [a_1, b_1] \times \cdots \times [a_n, b_n] \subset \mathbb{R}^n$ are compact.

We may prove this by inducting on n. As per earlier, we have that a closed interval on R is compact. See Day 8 for the complete proof; class was interrupted by a fire alarm :c

[5] In general, this works on any $f : X \to Y$ if X is a compact metric space. for here, we let them both be subsets of $\mathbb{R}^n$

§8 Day 8: Finishing Heine-Borel; Differentiation (Sep. 20, 2024)

Theorem 8.1 (Finishing Heine-Borel; Closed and Bounded implies Compact)**.** We now extend our proof that $X \subset \mathbb{R}^n$ is compact if it is closed and bounded to $n > 1$.

As a preliminary case, start by considering the rectangle $R = [a_1, b_1] \times \cdots \times [a_n, b_n] = Q \times [a_n, b_n]$. Let Q be a closed set in $\mathbb{R}^{n-1}$; continuing on our inductive proof from last time, we have that Q is compact. For any $t \in [a_n, b_n]$, consider $Q_t = \{(y, t) \mid y \in Q\}$ (where we may let y be the rest of the $n-1$ components in Q).

Now, consider $\mathcal{O}$ to be an open cover of R. Then

$$Q_t \subset U_1 \cup \cdots \cup U_k =: U$$

where $U_1, \ldots, U_k \in \mathcal{O}$, then by the ε-neighborhood theorem, there is $\varepsilon > 0$ such that $Q_t \times (t - \varepsilon, t + \varepsilon) \subset U$. Let us consider the set $\{(t - \varepsilon_t, t + \varepsilon_t) \mid t \in [a_n, b_n]\}$. This is an open cover of the closed interval $[a_n, b_n]$; by compactness of $[a_n, b_n]$, there is a finite subcover by the open intervals $(t - \varepsilon_t, t + \varepsilon_t)$. This means $Q \times (t - \varepsilon_t, t + \varepsilon_t)$ for finitely many $t \in [a_n, b_n]$ covers R, and we are done.

Returning to Heine-Borel, now, consider any $X \subset \mathbb{R}^n$. Since X is bounded, we may enclose $X \subset R$ where R is a closed rectangle construction as per above. Then let $\mathcal{O}$ be an open cover of R. Since $\mathbb{R}^n \setminus X$ is open, we have that $\mathcal{O} \cup \{\mathbb{R}^n \setminus X\}$ is an open cover of R, meaning that as per above, $R \subset U_1 \cup \cdots \cup U_k \cup (\mathbb{R}^n \setminus X)$ where $U_i \in \mathcal{O}$. Since $R \supset X$, we conclude that $U_1, \ldots, U_k$ is a finite subcover of $\mathcal{O}$ for X. $\square$

We now cover differentiation. Suppose $f : (U \subset \mathbb{R}^n) \to \mathbb{R}^m$ with U being open. We say that f is differentiable at a if there is a linear transformation $\lambda : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{h \to 0} \frac{f(a+h) - f(a) - \lambda h}{|h|} = 0.$$

Specifically, $f(a+h) - f(a) - \lambda(h) = o(|h|)$; i.e., it is equal to a function $\varphi(h)$ where $\lim_{h \to 0} \frac{y(h)}{|h|} = 0$.

Lemma 8.2 (Differentiability implies Continuity)**.** If f is differentiable at a, then f is continuous at a.

To see this, take $h \to 0$ in $f(a+h) - f(a) - \lambda(h) = o(|h|)$.

Lemma 8.3. If f is differentiable at a, then there is a unique affine function $h \mapsto c - \lambda(h)$ such that

$$\lim_{h \to 0} \frac{f(a+h) - f(a) - \lambda(h)}{|h|} = 0.$$

Let $c = f(a)$ by continuity. So we have to show that if $\lambda, \mu : \mathbb{R}^n \to \mathbb{R}^m$, with both satisfying the above limit, then we have $\lambda = \mu$. Observe that if we write

$$\lim_{x \to 0} \frac{\lambda(x) - \mu(x)}{|x|} = 0,$$

then we may take $x = ty$ and take $t \to 0$ to get

$$\lim_{t \to 0} \frac{\lambda(t_y) - \mu(t_y)}{|t_y|} = 0,$$

so $\frac{\lambda(y) - \mu(y)}{|y|} = 0$. This means $\lambda = \mu$. $\square$

With this, we say that λ is the derivative of f at a. We then may write $Df(a)$ or $f'(a)$ or $\partial_a f$. If f is differentiable at every point of U, then we say it is differentiable on the open set U.

§9 Day 9: Partial Differentiation, Jacobians (Sep. 23, 2024)

Let us have a function[6] $f : (U \subset \mathbb{R}^n) \to \mathbb{R}^m$. Then the matrix $D_f(a)$ or $f'(a)$ w.r.t. the standard bases of $\mathbb{R}^n, \mathbb{R}^m$ is called the Jacobian matrix of f at a.

We now give some examples.

(a) For an example, consider $g : \mathbb{R} \to \mathbb{R}$ differentiable at all $a \in \mathbb{R}$. Let $f(x, y) = g(x)$, where $(x, y) \in \mathbb{R}^2$. Then f is differentiable at (a, b), for any $b \in \mathbb{R}$ and $D_f(a, b) : (h, k) \mapsto g'(a)h$. Writing the derivative out, we have

$$\frac{f(a+h, b+k) - f(a, b) - g'(a)h}{|(h, k)|} \to 0$$

as $(h, k) \to 0$, meaning the above is equal to

$$\frac{g(a+h) - g(a) - g'(a)h}{|h|} \cdot \frac{|h|}{|(h, k)|} = 0.$$

Thus, the Jacobian is $D_f(a, b) = (g'(a), 0)$.

(b) Let $f(x, y) = \sqrt{|xy|}$. Is it differentiable at 0? To check this, we want to either find or disprove the existence of λ, μ such that

$$\frac{\sqrt{|hk|} - 0 - (\lambda h + \mu k)}{|(h, k)|} \to 0$$

as $(h, k) \to 0$. Now, suppose $h = k$. Then we have

$$\frac{|h| - (\lambda + \mu)h}{\sqrt{2}\,|h|} = \frac{1}{\sqrt{2}} - \frac{\lambda + \mu}{\sqrt{2}} \cdot \frac{h}{|h|}.$$

If $\lambda + \mu = 0$, then $\frac{1}{\sqrt{2}} \not\to 0$. If $\lambda + \mu \neq 0$, then the limit approaches 2 instead as $h \to 0$. Thus, we conclude that f is not differentiable at 0.

Let the directional derivative of f at a along a vector v be given by

$$D_v f(a) = \lim_{t \to 0} \frac{f(a + tv) - f(a)}{t}.$$

Then we define the ith partial derivative of f at a, for $i = 1, \ldots, n$, to be

$$\frac{\partial f}{\partial x_i}(a) = D_{e_i} f(a) = \lim_{h \to 0} \frac{f(a_1, \ldots, a_i + h, \ldots, a_n) - f(a_1, \ldots, a_n)}{h}.$$

Lemma 9.1 (Differentiability at Point implies Directional Derivatives Exist)**.** If f is differentiable at a, then all directional derivatives (i.e., for all v) at a exist, and $D_v f(a) = Df(a)v$.

To start, we know that

$$\frac{f(a + tv) - f(a) - D_f(a)(tv)}{|tv|} \to 0$$

as $t \to 0$. If $t \to 0^+$, then $|tv| = t\,|v|$ and we may multiply the LHS by $|v|$ to get

$$\frac{f(a + tv) - f(a)}{t} - Df(a)(v) \xrightarrow[t \to 0^+]{} 0$$

In the other way, if $t \to 0^-$, then $|tv| = -t\,|v|$; multiply the LHS by $-|v|$, and we get the same thing. □

[6] wanted to clarify this isn't correct notation, it just looks correct to me so i do it

§10 Day 10: Differentiation (Sep. 25, 2024)

Today we will go over differentiation. We start with a few examples;

(a) Let $f(x, y) = \sin(x \cos y)$. Then

$$\frac{\partial f}{\partial y} = \cos(x \sin y) x \cos y.$$

Note that in the partial derivative above, we hold x as a constant.

(b) Let $f(x, y) = x^{x^{x^{x^y}}} = e^{(log x) \cdot x^{x^{x^y}}}$. Then

$$D_2(1, y) = x^{x^{x^{x^y}}}; \quad \log x \frac{\partial}{\partial y} \left(x^{x^{x^y}} \right) = 0.$$

We now discuss higher order derivatives; consider $f : (U \subset \mathbb{R}^n) \to \mathbb{R}^n$. Suppose $D_i f : U \to \mathbb{R}$ exists for all i, and we consider that $D_j(D_i f)$ may also be written as

$$D_{ij} f(x), \quad f_{x_i x_j}(x), \quad \frac{\partial^2 f}{\partial x_j \partial x_i}(x).$$

However, we must note that order of differentiation is important; observe the example below (with corresponding link),

(c) Let $f(x, y) = \begin{cases} xy \frac{x^2 - y^2}{x^2 + y^2} & (x, y) \neq (0, 0) \\ 0 & (x, y) = (0, 0) \end{cases}$.

Then we have $D_2 f(x, 0) = x$, $D_2 f(0, y) = -y$ by symmetry, and $D_{12} f(0, 0) = -1$ since $D_{21} f(0, 0) = 1$.

Note that second order mixed partial derivatives at a are equal if they are both defined, and they are continuous on an open set containing a. For clarity, we introduce multi-index notation, given as below,

$$\frac{\partial^{|\alpha|} f}{\partial x^\alpha} = \frac{\partial^{a_1 + \cdots + a_n} f}{\partial x_1^{\alpha_1} \ldots \partial x_n^{a_n}},$$

i.e. $\alpha = (\alpha_1, \ldots, \alpha_n)$ and $|\alpha| = \alpha_1 + \cdots + \alpha_n$. A relevant point worth bringing up is Schwarz's theorem, which discusses when the orders of partials commute (note that this was not covered in class).

Theorem 10.1 (Chain Rule)**.** Let $f : (U \subset \mathbb{R}^n) \to \mathbb{R}^m$, and $g : (V \subset \mathbb{R}^m) \to \mathbb{R}^n$ (where $f(U) \subset V$). If f is differentiable at $a \in U$, and g is differentiable at $f(a)$, then $g \circ f : U \to \mathbb{R}^n$ is differentiable at a, and $D(g \circ f)(a) = D_g(f(n)) \circ Df(a)$, i.e. $(g \circ f)'(a) = g'(f(a)) \cdot f'(a)$ (note that this denotes matrix multiplication).

To start, we're given that $f(x) - f(a) - f'(a)(x - a) = \varphi(x)$ from $\lim_{x \to a} \frac{\varphi(x)}{|x - a|} = 0$, and $g(y) - g(b) - g'(b)(y - a) = \psi(y)$ from $\lim_{y \to b} \frac{\psi(y)}{|y - b|} = 0$. We want to show that

$$\frac{(g \circ f)(x) - g(b) - g'(b) f'(a)(x - a)}{|x - a|} \xrightarrow[x \to a]{} 0.$$

To start, we reduce the numerator as folows,

$$\begin{aligned} &(g \circ f)(x) - g(b) - g'(b) f'(a)(x - a) \\ &= \underbrace{(g \circ f)(x) - g(b) - g'(b)(f(x) - f(a)}_{\psi(f(x))} - \varphi(x)) \\ &= \psi(f(x)) + g'(b) \varphi(x). \end{aligned}$$

Then observe that the latter half of the reduced expression yields,

$$\lim_{x\to a} \frac{g'(b)\varphi(x)}{|x-a|} = g'(b) \lim_{x\to a} \frac{\varphi(x)}{|x-a|} = 0.$$

We now claim that the former half satisfies

$$\lim_{x\to a} \frac{\psi(f(x))}{|x-a|} = 0.$$

For all $\varepsilon > 0$, we have $|\psi(f(x))| < \varepsilon\,|f(x)-b|$ if $|f(x)-b| < \delta'$ for some $\delta' = \delta'(\varepsilon)$, which happens when $|x-a| < \delta$ for some $\delta = \delta(\delta')$ by continuity. Therefore, if $|x-a| < \delta$, then

$$\begin{aligned} |\psi(f(x))| &< \varepsilon\,|f(x)-b| \\ &= \varepsilon\left|f'(a)(x-a)+\varphi(x)\right| \\ &\le \varepsilon M\,|x-a| + \varepsilon\,|\varphi(x)|\,, \end{aligned}$$

and so $\frac{|\psi(f(x))|}{|x-a|} \le \varepsilon M + 0$ as $x \to a$, which means we may conclude the fraction goes to 0 as $x \to 0$ as desired. $\square$

§11 Day 11: Computations of Derivatives (Sep. 27, 2024)

We start with a few useful facts: for the following examples, we let $f : (U \subset \mathbb{R}^n) \to \mathbb{R}^m$.

(a) If f is a constant function $f = b$, then $Df(a) = 0$.

$$\lim_{h\to 0} \frac{f(a+h) - f(a) - 0}{|h|} = \lim_{h\to 0} \frac{b-b}{|h|} = 0.$$

(b) If f is a linear transformation, then $Df(a) = f$.

$$\lim_{h\to 0} \frac{f(a+h) - f(a) - f(h)}{|h|} = \lim_{h\to 0} \frac{f(a+h) - f(a+h)}{|h|} = 0.$$

(c) If $f = (f_1, \dots, f_m)$, f is differentiable at a if and only if each f_i is differentiable at a, and $Df(a) = (Df_1(a), \dots, Df_m(a))$. We call $Df(a)$ to be the Jacobian (read: $m \times n$ matrix) where the ith row is given by $Df_i(a)$. [7]

($\Rightarrow$) Observe that if f is differentiable at a, then $f_i = \pi_i \circ f$, where π_i is the projection from $\mathbb{R}^m \to \mathbb{R}$, which is differentiable by the chain rule.

($\leftarrow$) Now suppose each f_i is differentiable at a; write $\lambda = (Df_1(a), \dots, Df_m(a))$, and consider

$$\begin{aligned}&\lim_{h\to 0} \frac{f(a+h) - f(a) - \lambda(h)}{|h|} && \text{(Expand component-wise)} \\ &= \lim_{h\to 0} \frac{(f_1(a+h) - f_1(a) - Df_1(a), \dots, f_m(a+h) - f_m(a) - Df_m(a))}{|h|} = 0\end{aligned}$$

by the triangle inequality. $\square$

(d) Suppose $S : \mathbb{R}^2 \to \mathbb{R}$, and $S(x, y) = x + y$. Then we may observe that S is linear, meaning $D_S(a, b) = S$ by (b).

(e) Suppose $P : \mathbb{R}^2 \to \mathbb{R}$, and $P(x, y) = xy$. Then[8]

$$D_P(a, b)(x, y) = bx + ay = \begin{pmatrix} b & a \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

Then we may write (using $|hk| \le \sqrt{h^2 + k^2}$)

$$\lim_{(h,k)\to 0} \frac{|P(a+h, b+k) - P(a, b) - (bh + ak)|}{|(h, k)|} = \lim_{(h,k)\to 0} \frac{|hk|}{\sqrt{h^2 + k^2}} \le \sqrt{h^2 + k^2}.$$

[7] check this... i'm not so sure..? nvm jacobian wikipedia link

[8] yes i used binom i'm a lazy fuck

Theorem 11.1 (Addition, Product, Quotient Rule)**.** If $f, g : (U \subset \mathbb{R}^n) \to \mathbb{R}^m$ are differentiable at $a \in U$, then

(a) $D(f+g)(a) = Df(a) + Dg(a)$.

(b) If $m = 1$, then $D(fg)(a) = g(a)Df(a) + f(a)Dg(a)$.

(c) If $m = 1$ and $g(x) \neq 0$, then $D\left(\frac{1}{g}\right)(a) = -\frac{1}{g(a)^2}Dg(a)$.

The proofs simply follows from our previous examples.

(a) $f + g = S \circ (f+g)$;

(b) $fg = P \circ (f, g)$;

(c) $\frac{1}{g} = \frac{1}{y} \circ g$.

Theorem 11.2 (Jacobian Construction)**.** If $f : (U \subset \mathbb{R}^n) \to \mathbb{R}^m$ is differentiable at a, then $\frac{\partial f_i}{\partial x_j}(a)$ exists for all $1 \leq i \leq n$ and $1 \leq j \leq m$. Then $Df(a)$ is given by the matrix (which we will call the *Jacobian*),

$$\begin{pmatrix} \frac{\partial f_1}{\partial x_1}(a) & \dots & \frac{\partial f_1}{\partial x_n}(a) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(a) & \dots & \frac{\partial f_m}{\partial x_n}(a) \end{pmatrix} = \left(\frac{\partial f_i}{\partial x_j}(a)\right).$$

To prove this, observe that when $m = 1$, $f : U \to \mathbb{R}$, and $\frac{\partial f}{\partial x_j}(a) = (f \circ h)'(a_j)$, where $h(x) = (a_1, \dots, a_{j-1}, x, a_{j+1}, \dots)$. Then

$$\frac{\partial f}{\partial x_j}(a) = (f \circ h)(a_j) = f'(a) \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} j,$$

and so $\frac{\partial f}{\partial x_j}(a)$ is the jth component of $Df(a)$.

For general choices of m, $f = (f_1, \dots, f_m)$; since each f_i is differentiable at a, and $f'(a) = (f_1'(a), \dots, f_m'(a))$, we have that $f_i'(a)$ is the ith row of the matrix $f'(a)$. $\square$

For example, let $f(x, y) = g(x+y)$. Then $Df(c, d)(x, y) = D_g(c+d)(x+y)$, of which the LHS may be expanded as

$$(D_g \circ s)(c, d) \circ \underbrace{D_s(c, d)}_{S}(x, y) = D_g(c+d) \circ s(x, y) = D_g(c+d)(x+y)$$

for more clarity.

§12 Day 12: Computations of Derivatives II (Sep. 30, 2024)

We start with a few examples.

(a) Let $f(x,y) = g(x+y)$ be differentiable. Then

$$D_f(c,d)(x,y) = D_g(c+d)(x+y).$$

(b) Let $f(x,y) = g(xy) = (g \circ p)(x,y)$ be differentiable.

$$D_f(c,d)(x,y) = D_g \circ p(c,d) \circ D_p(c,d)(x,y) = D_g(cd)(dx+cy).$$

(c) Let $f(x,y) = \int_a^{x+y} g$ with g continuous. Then $D_f(c,d)(x,y) = g(c+d)(x+y)$.

(d) Let $f(x,y) = \int_a^{xy} g$ with g continuous. Then $D_f(c,d)(x,y) = g(cd)(dx+cy)$.

(e) Let $F(x) = f(g_1(x), \dots, g_m(x))$. Let $g_1, \dots, g_m : \mathbb{R}^n \to \mathbb{R}$ be differentiable at a. Then $f : \mathbb{R}^m \to \mathbb{R}$ is differentiable at $(g_1(a), \dots, g_m(a))$.

$$\frac{\partial F}{\partial x_i}(a) = \sum_{j=1}^{m} \frac{\partial f}{\partial y_j}(g_1(a), \dots, g_m(a)) \frac{\partial g_i}{\partial x_i}(a)$$

Using $F = f \circ g$, where $g = (g_1, \dots, g_m)$ and $F'(a) = Df(g(a)) \cdot D_g(a)$, we get that

$$\left(\frac{\partial F}{\partial x_1}(a), \dots, \frac{\partial F}{\partial x_n}(a)\right) = \left(\frac{\partial f_1}{\partial y_1}(g(a)), \dots, \frac{\partial f_m}{\partial y_m}(g(a))\right) \begin{pmatrix} \frac{\partial g_1}{\partial x_1}(a) & \dots & \frac{\partial g_1}{\partial x_n}(a) \\ \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial x_1}(a) & \dots & \frac{\partial g_m}{\partial x_n}(a) \end{pmatrix}.$$

With $y = g(x)$ and $z = f(y)$, we get that

$$\frac{\partial z}{\partial x_i} = \sum_{j=1}^{m} \frac{\partial z}{\partial y_i} \frac{\partial y_i}{\partial x_i}; \qquad \frac{\partial z}{\partial x} = \frac{\partial z}{\partial y} \frac{\partial y}{\partial x}.$$

Recall that whenever f is differentiable at a, then $D_v f(a) = Df(a)(v)$. In particular, $D_v f(a)$ is linear in v. Also recall that f is not necessarily differentiable at a, even if all the directional derivatives at a exist. For example, let

$$f(x,y) = \begin{cases} \frac{x^2 y}{x^4 + y^2} & \text{if } (x,y) \neq (0,0), \\ 0 & \text{if } (x,y) = (0,0). \end{cases}$$

Then letting $v = (h,k)$, we get $D_v f(0,0) = \lim_{t \to 0} \frac{f(th,th) - f(0,0)}{t} = \lim_{t \to 0} \frac{h^2 k}{t^2 h^4 + k^2}$, which is equal to $\frac{h^2}{k}$ if $k \neq 0$ and 0 if $k = 0$. Clearly, the directional derivatives are not commensurate to a single value, and so the derivative does not exist. In fact, f is differentiable on every straight line through 0, but it isn't even continuous at 0.

For example, let $f(x, x^2) = \frac{1}{2}$. This function is not continuous at 0. To see this, consider $f(x, mx)$ with $y = mx$, $\frac{mx}{x^2+y^2}$.[9]

[9] what

Theorem 12.1 (Differentiability Condition)**.** Let $f : (U \subset \mathbb{R}^n) \to \mathbb{R}^n$. If all $\frac{\partial f_i}{\partial x_j}$ exist in an open neighborhood of a and are continuous at a, then $Df(a)$ exists.

Start by assuming $m = 1$. Then let us have $a = (a_1, \ldots, a_n), h = (h_1, \ldots, h_n)$.[10]

$$\begin{aligned} f(a+h) - f(a) &= f(a_1 + h_1, a_2, \ldots, a_n) - f(a_1, \ldots, a_n) \\ &\quad + f(a_1 + h_1, a_2 + h_2, a_3, \ldots, a_n) - f(a_1 + h_1, a_2, \ldots, a_n) \\ &\quad + f(a_1 + h_1, \ldots, a_n + h_n) - f(a_1 + h_1, \ldots, a_{n-1} + h_{n-1}, a_n) \\ &= h_1 \frac{\partial f}{\partial x_1}(b_1, \ldots, a_n) \\ &\quad + h_i \frac{\partial f}{\partial x_i}(a_1 + h_1, \ldots, b_i, a_n) \\ &\quad + \ldots . \end{aligned}$$

Let's write the above as $\sum_{i=1}^n h_i \frac{\partial f}{\partial x_i}(c_i)$, with $c_i \to a$ as $h \to 0$.

$$\begin{aligned} &\lim_{h \to 0} \frac{f(a+h) - f(a) - \sum_{i=1}^n \frac{\partial f}{\partial x_i}(a) \cdot h_i}{|h|} \\ &= \lim_{h \to 0} \frac{\sum_{i=1}^n \left(\frac{\partial f}{\partial x_i}(c_i) - \frac{\partial f}{\partial x_i}(a) \right) \cdot h_i}{|h|} \\ &= 0 \end{aligned}$$

by continuity of the partial at a, and $\frac{|h_i|}{|h|} = 1$. $\square$

[10]no idea what's going on here btw

§13 Day 13: Gradient (Oct. 2, 2024)

What is the different of fastest increase of a differentiable function? Let $f : (U \subset \mathbb{R}^n) \to \mathbb{R}$ be differentiable at $a \in U$. Then

$$D_x f(a) = Df(a)(x) = \underbrace{\left(\frac{\partial f}{\partial x_1}(a), \dots, \frac{\partial f}{\partial x_n}(a)\right)}_{\nabla f(a)} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \langle \nabla f(a), x \rangle,$$

where ∇ denotes the gradient of the function f. Invariably, $Df(a) \in (\mathbb{R}^n)^*$ (dual space), so there exists a unique $y \in \mathbb{R}^n$ such that $Df(a)(x) = \langle y, x \rangle$. We denote y by $\operatorname{grad} f(a)$ or $\nabla f(a)$. Let e be a unit vector; then $D_e f(a)$ is the slope of the curve at a, obtained by intersecting the graph of f with vertical planes through the line $x = a + te$ (for $t \in \mathbb{R}$). Then

$$\begin{aligned} D_e f(a) &= \langle \nabla f(a), e \rangle \\ &= |\nabla f(a)| \cos\theta, \end{aligned}$$

where θ is given by the angle between e and $\nabla f(a)$ (let's just assume it's nonzero); i.e., the directional derivative $D_e f(a)$ attains its largest value when $\theta = 0 \implies \cos\theta = 1$, where e is in the direction of $\nabla f(a)$. For example, if $f(x_1, x_2)$ denotes the temperature of a point (x_1, x_2) in the plane, then a heat-seeking bug will move in the direction of the gradient. If $\nabla f(a) = 0$, then there may still be directions of fastest increase, but they are not necessarily unique; for example, graph $z = x^2 - y^2$; we see that z increases along the x axis in both directions.

We now discuss differentiation under the integral sign. Given that $f(x, y)$ is continuous on $[a, b] \times [c, d]$, we have that $\frac{\partial f}{\partial y}$ is also continuous on the rectangle. Let

$$F(y) = \int_a^b f(x, y)\, dx.$$

Then $F(y)$ is continuously differentiable on $[c, d]$, and $F'(y) = \int_a^b \frac{\partial f}{\partial y}(x, y)\, dx$. More generally, f being C^n implies F being $C^n = \{f : \mathbb{R}^2 \to \mathbb{R} \mid f \text{ is } n \text{ times cont. diff.}\}$. To prove this, let us consider

$$\begin{aligned} &F(y + h) - F(y) - \int_a^b \frac{\partial f}{\partial y}(x, y)\, dx \cdot h \\ &= \int_a^b f(x, y + h) - f(x, y)\, dx - \frac{\partial f}{\partial y}(x, y) \cdot h\, dx. \end{aligned} \tag{13.1}$$

Then $\frac{\partial f}{\partial y}$ is uniformly continuous on $[a, b] \times [c, d]$, so given $\varepsilon > 0$, there exists $\delta > 0$ such that

$$|y_1 - y_2| < \delta \implies \left| \frac{\partial f}{\partial y}(x, y_1) - \frac{\partial f}{\partial y}(x, y_2) \right| < \varepsilon,$$

meaning $f(x, y + h) - f(x, y) = \frac{\partial f}{\partial y}(x, y_n) \cdot h$ for some $y_n \in (y, y + h)$ by MVT. Thus,

$$(13.1) = h \int_a^b \frac{\partial f}{\partial y}(x, y_n - y)\, dx,$$

then $\int_a^b \frac{\partial f}{\partial y}(x, y_n - y)\, dx \le \varepsilon(b - a)$, if $|h| < \delta$, by uniform continuity. Thus, we have that

$$\frac{F(y + h) - F(y) - \int_a^b \frac{\partial f}{\partial y}(x, y)\, dx}{h} \xrightarrow{h \to 0} 0. \quad \square$$

§14 Day 14: Inverse Function Theorem (Oct. 4, 2024)

Let A be an open set on $\mathbb{R}^n$, and let $f : A \to \mathbb{R}^n$ be a C^r map (where $1 \le r < \infty$). Let $a \in A$; if $Df_a : \mathbb{R}^n \to \mathbb{R}^n$ is invertible, then there exist an open neighborhood U of a where $U \subset A$ such that V of $f(a)$ in $\mathbb{R}^n$, $f|_u : U \to V$ has a C^r inverse $g = (f|_u)^{-1} : V \to U$. Consider the following,

(a) If g is merely differentiable on V, using $f \circ g = \mathrm{id}_V$, for all $y \in V$, we have $Df_{g(y)} \circ Dg_y = D(\mathrm{id}_V)_y = \mathrm{id}_{\mathbb{R}^n}$.

(b) (a) says that $D_g =$ "Inversion" $\circ D_f \circ f^{-1} \in C^\infty \circ C^{r-1} \circ g$. By induction, we see that g is C^r. Thus, in the inverse function theorem, it suffices to prove that g is differentiable (Neumann series).

(c) The proof is trivial in 1 dimension; it is non-trivial in higher dimensions, involving compactness and the Banach contraction fixed point theorem.

(d) Continuity of the derivative is essential; for example, if $f : \mathbb{R} \to \mathbb{R}$ is such that $f'(0) \ne 0$, but f is not invertible in any neighborhood of zero, then we may pick

$$f(x) = \begin{cases} cx + x^2 \sin\left(\frac{1}{x}\right) & x \ne 0, 0 < c < 1, \\ 0 & x = 0. \end{cases}$$

Then

$$f'(x) = \begin{cases} c + 2x \sin\left(\frac{1}{x}\right) - \cos\left(\frac{1}{x}\right) & x \ne 0, 0 < c < 1, \\ c & x = 0. \end{cases}$$

But $f'(\frac{1}{n}) = c + 0 - (-1)^n \in \{c-1, c+1\}$.

(e) The theorem gives sufficient (but not necessary) conditions. For example, take $f : \mathbb{R} \to \mathbb{R}$. $f(x) = x^3$ is C^∞, and $f'(0) = 0$. Yet it has a globally continuous inverse $f^{-1} : \mathbb{R} \to \mathbb{R}$, $y \mapsto y^{\frac{1}{3}}$. Of course, f^{-1} is not differentiable at $f(0) = 0$, otherwise it would contradict the chain rule.

(f) The theorem is purely local: consider $f : \mathbb{R}^2 \to \mathbb{R}^2$, and $f(x, y) = (e^x \cos y, e^x \sin y)$. Then

$$f'(x, y) = \begin{bmatrix} e^x \cos y & -e^x \sin y \\ e^x \sin y & e^x \cos y \end{bmatrix},$$

so $\det f'(x, y) = e^{2x} > 0$, but f is not invertible (it is periodic in y).

We now present some examples.

(a) Let $f : \mathbb{R}^2 \to \mathbb{R}^2$, $f(r, \theta) = (r \cos\theta, r \sin\theta)$. Then

$$f'(r, \theta) = \begin{bmatrix} \cos theta & -r \sin\theta \\ \sin\theta & r \cos\theta \end{bmatrix} \implies det f'(r, \theta) = r,$$

which is obviously nonzero for $r \ne 0$. A typical restriction of f is to $(0, \infty) \times (\theta_1, \theta_2)$ such that $0 < \theta_2 - \theta_1 \le 2\pi$; for example, $(0, \infty) \times (-\pi, \pi)$ or $(0, \infty) \times (0, 2\pi)$ are both suitable.

Let $g : \mathbb{R}^2 \to \mathbb{R}$, $g(x, y) = \sqrt{x^2 + y^2}$. Then $(g \circ f)(r, \theta) = g(r \cos\theta, r \sin\theta) = \sqrt{r^2} = r$; i.e., changing coordinates (composing f resp. f^{-1}) can simplify things greatly!

(b) Let $f(r, \theta, \varphi) = (r \sin\theta \cos\varphi, r \sin\theta \sin\varphi, r \cos\theta)$, where $r \in (0, \infty)$ specifies the radius, $\theta \in (0, \pi)$ specifies the inclination, and $\varphi \in (0, 2\pi)$ is the angle within the plane. This is the parametrization of a sphere (I'm not sure if we did anything else with this in class...).

§15 Day 15: Inverse Function Theorem, Pt. 2 (Oct. 7, 2024)

Given $f : (U \subset \mathbb{R}^n) \to \mathbb{R}^n$ with U open, if $a \in U$ and $\det f'(a) \neq 0$, then

(a) There exists open sets $V \ni a$, $W \ni f(a)$ such that $f : V \to W$ has a continuous inverse $f^{-1} : W \to V$,

(b) f^{-1} is differentiable, and $(f^{-1})'(y) = \big(f'(f^{-1}(y))\big)^{-1}$.

We now make a few remarks;

(a) If we know that f^{-1} is differentiable, then the formula follows directly from the chain rule,
$$f(f^{-1}(y)) = y \implies f'(f^{-1}(y)) \cdot (f^{-1})'(y) = \mathrm{id}.$$

(b) f^{-1} may exist even if $f'(a) = 0$. For example, let $y = x^3$; then $\frac{dy}{dx}\Big|_{x=0} = 0$. However, if $\det f'(a) = 0$, then f^{-1} is not differentiable at $f(a)$, because if it were, then
$$(f^{-1})'(f(a)) \cdot \underbrace{f'(a)}_{\det = 0} = \mathrm{id},$$
which is contradictory. Also, we can't eliminate the hypothesis that f is continuously differentiable.

(c) It follows from the IFT theorem that f is C^1; we want to show that the entries of the matrix $f'(f^{-1}(y))^{-1}$ are continuous. Let A be invertible, and consider $A^{-1}_{ji} = \pm\frac{\det A^{(ij)}}{\det A}$, where $A^{(ij)}$ is A with its ith row and jth column eliminated. Clearly, we see that the entries of A^{-1} are rational functions of entries in A.

(d) If f is C^r (i.e., all partial derivatives of order up to r exist and are continuous), then f^{-1} is C^r:
$$f \text{ is } C^r \iff f \text{ cts. and } D_f \text{ is } C^{r-1},$$
and so f is C^r implies that f is C^{r-1} and D_f is C^{r-1}, meaning f^{-1} is C^{r-1} by induction on r, i.e.
$$(f^{-1})'(y) = (f'(f^{-1}(y)))^{-1}$$
is C^{r-1} by the chain rule. □

We'll need the following consequences of the mean value theorem;

Lemma 15.1. Given us have a continuous function $f : B \to \mathbb{R}^m$, where B is a closed ball in $\mathbb{R}^n$, such that f is differentiable on $\mathrm{Int}(B)$, if $\left|\frac{\partial f_i}{\partial x_j}\right| \le M$ on $\mathrm{Int}(B)$ for all i, j, then $|f(x) - f(y)| \le \sqrt{mn}M\,|x - y|$ for $x, y \in B$.

If $f = (f_1, \dots, f_n)$ and $|f(x) - f(y)| \le \sqrt{m}\max_i |f_i(x) - f_i(y)|$, so it is enough to prove that the lemma for $m = 1$, i.e. $f : B \to \mathbb{R}$, $B \subset \mathbb{R}^n$. Let $g(t) = f((1-t)x + ty)$ for given x, y; then
$$\begin{aligned}
f(y) - f(x) &= g(1) - g(0) \\
&= g'(t_0) && \text{(for some } t_0 \in (0,1) \text{ by MVT)} \\
&= D_f(c_0)(y - x) && \text{(where } c_0 = (1-t_0)x + t_0 y) \\
&= \langle \nabla f(c_0), y - x \rangle,
\end{aligned}$$
and so we have that $|f(y) - f(x)| \le |\nabla f(c_0)|\,|y - x| \le \sqrt{n}\max\left|\frac{\partial f}{\partial x_j}\right|$. □

§16 Day 16: Inverse Function Theorem, Pt. 3 (Oct. 9, 2024)

(Note that this is my formulation) If f is a continuously differentiable function from an open subset $U \subset \mathbb{R}^n$ to $\mathbb{R}^n$, and the derivative $f'(a)$ is invertible at a (i.e., $\det f'(a) \neq 0$), then

(a) There exist open neighborhoods V of a, W of $f(a)$ such that $f(V) \subset W$, $f : V \to W$ is bijective, and $f^{-1}W \to V$ is continuously differentiable.

(b) f^{-1} is given by
$$(f^{-1})'(y) = \frac{1}{f'(f^{-1}(y))}.$$

To prove the theorem, we can assume that $Df(a) = \text{id}$; let $\lambda = Df(a)$, and $g = \lambda^{-1} \circ f$; and so we have that $D_g(a) = \text{id}$ as well. If the theorem is true for g, then it is true for $f = \lambda \circ g$, i.e.
$$f : V \xrightarrow{g} W \xrightarrow{\lambda} \lambda(W),$$
with inverse $g^{-1} \circ \lambda^{-1}$, which is continuous.

(a) To start, we can't have that $f(x) = f(a)$ if $x \neq a$ is sufficiently close to a:
$$\lim_{h \to a} \frac{f(a+h) - f(a) - \overbrace{Df(a)}^{\text{id}}\, h}{|h|} = 0.$$

If $f(a+h) = f(a)$, then the limit goes to $\frac{h}{|h|}$, which has norm 1. So $f(a+h) \neq f(a)$ for small enough h, i.e. there exists a closed ball B, centered at a, such that $f(x) \neq f(a)$ whenever $x \neq a$ is in B.

We can also assume that for $x \in B$, we have that $\det f'(x) \neq 0$ and
$$\left|\frac{\partial f_i}{\partial x_j}\right|(x) - \frac{\partial f_i}{\partial x_j}(a) < \frac{1}{2n}$$
for all i, j. With this, let $g(x) = f(x) - x$. Then
$$\frac{\partial f_i}{\partial x_j}(x) - \frac{\partial f_i}{\partial x_j}(a) = \frac{\partial g_i}{\partial x_j}(x),$$
where we may note $\frac{\partial f_i}{\partial x_j}(a)$ is literally just entries from the identity matrix. By the mean value lemma, we have that
$$\begin{aligned}
|g(x_1) - g(x_2)| &\leq n\frac{1}{2n}|x_1 - x_2|, && (x_1, x_2 \in B)\\
\implies |f(x_1) - x_1 - [f(x_2) - x_2]| &\leq \frac{1}{2}|x_1 - x_2| \\
\implies |x_1 - x_2| &\leq \text{LHS} + |f(x_1) - f(x_2)| \\
&\leq \frac{1}{2}|x_1 - x_2| + |f(x_1) - x_2| \\
&\leq 2\,|f(x_1) - f(x_2)|. && (*)
\end{aligned}$$

In particular, f is one to one on B. □

Claim 16.1. $f(\text{Int}(B))$ is open in $\mathbb{R}^n$.

Once we prove this, we get (a) by taking $V = \text{It}(B)$, $W = f(V)$; then f^{-1} is continuous by $(*)$, i.e.

$$\left|f^{-1}(y) - f^{-1}(y_2)\right| \leq 2\left|y_1 - y_2\right|.$$

To prove the claim, consider $x_0 \in \text{Int}(B)$, $y_0 \in f(x_0)$. We have to find an open ball containing y_0 in $f(\text{Int}(B))$. Let $d = d(y_0, f(\text{Bdry}(B))$. Let us show that $B(y_0, \frac{d}{2}) \subset f(\text{Int}(B))$; we will define $B_{\frac{d}{2}} := B(y_0, \frac{d}{2})$ from here on. Let $y \in B_{\frac{d}{2}}$; if $x \in \text{Bdry}(B)$, then $|y - y_0| < |y - f(x)|$. Then we just need to find $x \in \text{Int}(B)$ such that $y = f(x)$. Define $h : B \to \mathbb{R}$, where $h(x) = |y - f(x)|^2 = \sum_{i=1}^n (y_i - f_i(x))^2$. h is continuous, and so it attains minimum on B. By $|y - y_0| < |y - f(x)|$, it cannot occur on the boundary of B, and so it is strictly in the interior, at a critical point x, i.e. $\frac{\partial h}{\partial x_j} = 0$ for all j.

Thus, we have that

$$\sum_{i=1}^n 2(y_i - f_i(x))\frac{\partial f_i}{\partial x_j}(x) = 0,$$

where $\frac{\partial f_i}{\partial x_j}(x)$ are the entries of an invertible matrix. Therefore, $y_i - f_i(x) = 0$ for all i, which concludes the proof. $\square$

§17 Day 17: Inverse Function Theorem, Pt. 4 (Oct. 11, 2024)

If $f : (U \subset \mathbb{R}^n) \to \mathbb{R}^n$ is continuously differentiable, given $a \in U$ such that $\det f'(a) \neq 0$, we have that

(a) There are open neighborhoods V of a, W of $f(a)$, such that $f : V \to W$ has a continuous inverse $f^{-1} : W \to V$ with the property

$$\left|f^{-1}(y_1) - f^{-1}(y_2)\right| \leq 2\,|y_1 - y_2|$$

for any $y_1, y_2 \in W$.

(b) Today in class, we will show that f^{-1} is differentiable, and $(f^{-1})'(y) = f'(f^{-1}(y))^{-1}$.

Let $y_0 \in W$, $x_0 = f^{-1}(y_0)$, and $\mu = Df(x_0)$. We want to show that f^{-1} is differentiable at y_0, and $(f^{-1})'(y_0) = \mu^{-1}$. Then observe that $f(x) = f(x_0) + \mu(x - x_0) + \varphi(x - x_0)$, where

$$\lim_{x \to x_0} \frac{\varphi(x - x_0)}{|x - x_0|} = 0.$$

We have that $\mu^{-1}(f(x_0) - f(x)) = x - x_0 + \mu^{-1}\varphi(x - x_0)$, and each $y \in W$ can be written $y = f(x)$, $x \in V$. Write

$$\mu^{-1}(y - y_0) = f^{-1}(y) - f^{-1}(y_0) + \mu^{-1}\varphi(f^{-1}(y) - f^{-1}(y_0)).$$

We have to show that

$$\lim_{y \to y_0} \frac{\mu^{-1}\varphi(f^{-1}(y) - f^{-1}(y_0))}{y - y_0} = 0.$$

It is enough to show that the inside term goes to 0. Write

$$\lim_{y \to y_0} \frac{\varphi(f^{-1}(y) - f^{-1}(y_0))}{y - y_0} = \lim_{y \to y_0} \underbrace{\frac{\varphi(f^{-1}(y) - f^{-1}(y_0))}{|f^{-1}(y) - f^{-1}(y_0)|}}_{=0} \cdot \underbrace{\frac{\left|f^{-1}(y) - f^{-1}(y_0)\right|}{|y - y_0|}}_{\leq 2}.$$

Notice that the first term goes to 0 as $y \to y_0$ because $f^{-1}(y) \to f^{-1}(y_0)$ by continuity of f^{-1}. The second term is bounded above by 2, as per our lemma in (a). Thus, we see that the limit as a whole goes to 0 as $y \to y_0$. Since we already have $\frac{\varphi(t)}{|t|} \to 0$ as $t \to 0$ (we may just take $t = f^{-1}(y) - f^{-1}(y_0)$), we are done as per our earlier observation. $\square$

At the beginning of the proof, we said that we could assume that $f'(a) = \mathrm{id}$. Let $\lambda = f'(a)$, $g = \lambda^{-1} \circ f$. Then $Dg(a) = \lambda^{-1} \circ Df(a) = \mathrm{id}$.

Claim 17.1. If g satisfies the theorem, then so does f.

We have that $f = \lambda \circ g$, meaning $f : V \xrightarrow{g} W \xrightarrow{\lambda} \lambda(W)$; we also have $f^{-1} = g^{-1} \circ \lambda^{-1}$. If g^{-1} is continuous, we have that f^{-1} is continuous; the same goes with differentiability, where if g^{-1} is differentiable, then so is f^{-1} by the chain rule. We may also observe that W is an open set for g, and $\lambda(W)$ an open set for f. $\square$

Test administrative details! There will be 3 to 4 problems, similar to the ones on the problem set. Make sure you understand how to do all the problems on the past homeworks, even better if textbooks one are understood well too. Material is covered up to more or less a week before the test. The test is supposed to be straightforward

§18 Day 18: Implicit Function Theorem (Oct. 16, 2024)

We start with an example. Let $f(x,y) = x^2 + y^2 - 1$, and consider the region on $\mathbb{R}^2$ in which $f(x,y) = 0$. Can we solve $f(x,y) = 0$ for $y = g(x)$ near $x = a$, with $g(a) = b$? If $a \neq \pm 1$, then there are open intervals $I \ni a$ and $J \ni b$ such that, for all $x \in I$, there is a unique $y \in J$ such that $f(x,y) = 0$. Writing $y = g(x)$, we are done.

In general, we can solve for $y = g(x)$ whenever $\partial_y f(a,b) \neq 0$. We may proceed in the example with implicit differentiation; let $f(x, g(x)) = 0$, and consider

$$\frac{\partial f}{\partial x}(x, g(x)) + \frac{\partial f}{\partial y}(x, g(x))\frac{dg}{dx} = 0.$$

In our example, this is given by $2x + 2y\frac{dg}{dx}$. Then

$$g'(x) = -\frac{\partial_x f(x, g(x))}{\partial_y f(x, g(x))}$$

near $x = a$, which means $g'(x) = -\frac{x}{y}$ on the curve near (a,b) with $b \neq 0$. Of course, $g(x) = \sqrt{1-x^2}$; we may consider $(a, -b)$ and proceed to solve for $g_1(x) = -\sqrt{1-x^2}$.

In general, in a system of n equations, let

$$\begin{aligned} f_1(x_1, \dots, x_m, y_1, \dots, y_n) &= 0, \\ &\vdots \\ f_n(x_1, \dots, x_m, y_1, \dots, y_n) &= 0, \end{aligned}$$

where $f_1, \dots, f_n$ are functions in $m+n$ variables, i.e. $f(x,y) = 0$ where $x = (x_1, \dots, x_m) \in \mathbb{R}^m$, $y = (y_1, \dots, y_n) \in \mathbb{R}^n$, and $f = (f_1, \dots, f_n) : \mathbb{R}^{m+n} \to \mathbb{R}^n$, given (a,b), can we find, for all x close to $a \in \mathbb{R}^m$, a unique y near b such that $f(x,y) = 0$? (note that $y = g(x)$)

Consider the linear equation $Ax + By = 0$, where $A \in M_{n\times m}(\mathbb{R})$ and $B \in M_{n\times n}(\mathbb{R})$. If $\det B \neq 0$, then B is invertible, meaning we can solve for y as a function of x, i.e. $y = -B^{-1}Ax$. Then $Ax + By = B(B^{-1}Ax + y)$.

Theorem 18.1 (Implicit Function Theorem)**.** If U is an open set in $\mathbb{R}^m \times \mathbb{R}^n$, consider $f : U \to \mathbb{R}^n$ and let f be of class C^r, i.e. all partials exist up to order r and are continuous. Then for $(a,b) \in U, f(a,b) = 0$, then let

$$M = \left(\frac{\partial f_i}{\partial y_j}(a,b)\right) = \frac{\partial(f_1, \dots, f_n)}{\partial(y_1, \dots, y_n)}(a,b).$$

If $\det M \neq 0$, then there exist open sets $A \ni a$ in $\mathbb{R}^m$, $B \ni b$ in $\mathbb{R}^n$ such that, for all $x \in A$, there exists a unique $y \in B$ with $f(x,y) = 0$, i.e. $y = g(x)$. Moreover, g is of class C^r.

Let $F : U \to \mathbb{R}^m \times \mathbb{R}^n$, where $F(x,y) = (x, f(x,y))$ with $F(a,b) = (a, 0)$. Then

$$F'(a,b) = \left(\begin{array}{c|c} I_m & 0 \\ \hline * & M \end{array}\right) \implies \det F'(a,b) = \det M \neq 0.$$

By the inverse function theorem, there are open sets $U \ni (a,b)$ and $V \ni (a,0)$ such that $F : U \to V$ has C^r inverse $F^{-1} : V \to U$. We can assume U is of the form $A \times B$ by taking a smaller open set. Let $(u, v) = F(x, y)$, which implies that F^{-1} is of the form $(u, h(u, v))$. Then

$$\begin{aligned} F(F^{-1}(u,v)) &= (u,v) \\ \implies F(u,h(u,v)) &= (u,v) \\ \implies (u, f(u,h(u,v))) &= (u,v) \\ \implies f(u,h(u,v)) &= (u,v) \\ \implies f(u,h(u,0)) &= 0. \end{aligned}$$

Thus, $f(x, g(x)) = 0$ where $g(x) = h(x, 0)$ is C^r. $\square$

§19 Day 19: Implicit Function Theorem, Pt. 2 (Oct. 18, 2024)

Recall that given $f(x, y) = 0$ where

$$\begin{aligned} f &= (f_1, \ldots, f_n), \\ x &= (x_1, \ldots, x_m), \\ y &= (y_1, \ldots, y_n), \end{aligned}$$

with f being C^r in a neighborhood of (a, b) in $\mathbb{R}^{m+n}$, and $f(a, b) = 0$, if $\det \frac{\partial f_i}{\partial y_j}(a, b) \neq 0$, then we can solve $f(x, y) = 0$ for a C^r function $y = g(x)$ satisfying $g(a) = b$, i.e. $f(x, g(x)) = 0$, i.e. there are open sets $A \ni a$, $B \ni b$ such that, for all $x \in A$, there exists a unique $y \in B$ such that $f(x, y) = 0$. Write $y = g(x)$; moreover, g is C^r.

We may find $g'(x)$ by implicit differentiation, i.e. at $f_i(x, g(x)) = 0$, for all $i = 1, \ldots, n$, we have

$$\frac{\partial f_i}{\partial x_j}(x, g(x)) + \underbrace{\sum_{k=1}^{n} \frac{\partial f_i}{\partial y_k}(x, g(x))}_{\text{Matrix w/ invt. entries near } (a,b)} \frac{\partial g_k}{\partial x_j}(x)$$

$$\left(\frac{\partial g_k}{\partial x_j}(x)\right) = -\left(\frac{\partial f_i}{\partial y_k}(x, g(x))\right)^{-1} \left(\frac{\partial f_i}{\partial x_j}(x, g(x))\right).$$

The answer, of course, depends on $g(x)$. Moreover, since the implicit function theorem is more or less a generalization of the inverse function theorem, we prove that ImFT implies IFT. Let $U \subset \mathbb{R}^n$, and given $f : U \to \mathbb{R}^n$, consider that f is C^r in an open set U, and for $a \in U$, $\det f'(a) \neq 0$. Then let $F(x, y) = y - f(x)$, $b = f(a)$, $F(a, b) = 0$, where F is C^r in a neighborhood of (a, b). Then $\det \partial_x F(a, b) = \det(-f'(a))$. By the implicit function theorem, there are open neighborhoods A of G, B of A, such that for all $y \in A$, there exist unique $x \in B$ (using $x = g(y)$) such that $F(g(y), y) = 0$, i.e. $y = f(g(y))$. We may take $V = B \cap f^{-1}(A), W = A$ to see that $f : V \to W$ has C^r inverse g. $\square$

We go over a few examples now. Let $y^2 = x^2(x + 1)$ be a curve in $\mathbb{R}^2$, and let $f(x, y) = y^2 - x^2(x + 1)$; on $f(x, y) = 0$, $\frac{\partial f}{\partial y} = 2y \neq 0$ at every point of the curve where $y \neq 0$, so we can solve for y as a function of x.

$$\partial_x f = -3x^2 - 2x = -x(3x + 2) \neq 0$$

except when $x = 0$ or $-\frac{2}{3}$. IVT allows us to distinguish between smooth points and singularities; if f is smooth everywhere in 1 dimension, we call it a manifold. For example, Whitney's umbrella, $X : x^2 - xy^2 = 0$ is smooth according to the implicit function theorem except on the x axis.

§20 Day 20: Implicit Function Theorem, Pt. 3; Rank Theorem (Oct. 21, 2024)

We start with a few remarks on the implicit function theorem.

(a) Consider $f(x, y, z) = z^2 - xy^2$;

$$\frac{\partial f}{\partial z} = 2z, \qquad \frac{\partial f}{\partial x} = -y^2.$$

For the first partial, observe that we can solve for $z = g(x, y)$ at any point except where $z = 0$. For the second partial, we see that $-y^2 \neq 0$ as long as $y \neq 0$, meaning we may solve $x = h(y, z)$ near any point $(0, b, 0)$ with $b \neq 0$. Near $(a, 0, 0)$, with $a < 0$, we see that x is defined locally by $y = z = 0$, i.e. $f : (U \subset \mathbb{R}^3) \to \mathbb{R}^2$, we have $f(x, y, z) = (y, z)$.

(b) For coordinate changes, recall that if we take $f : \mathbb{R}^{m+n} \to \mathbb{R}^n$ and consider f as a function on (x, y), where $x \in \mathbb{R}^m$ and $y \in \mathbb{R}^n$, if $f(a, b) = 0$, and if

$$\det \frac{\partial f}{\partial y}(a, b) = \det \left(\frac{\partial f_i}{\partial y_j}(a, b) \right) \neq 0,$$

then there exists a function $F(x, y) = (x, f(x, y))$ with an inverse H near (a, b). We note that H is given by $H(u, v) = (u, h(u, v))$, with $F \circ H = \mathrm{id}$, and $(u, F \circ H(u, v)) = (u, v)$. That is, $(F \circ H)(u, v) = v$.

(c) Suppose $f : \mathbb{R}^p \to \mathbb{R}^n$ has rank n at a and $f(a) = 0$, i.e.

$$\left(\frac{\partial f_i}{\partial x_j}(a) \right)$$

has rank n for $1 \leq i \leq n$, $1 \leq j \leq p$. This means we may choose indices $j_1 < \cdots < j_n$ such that $\partial_{j_1, \ldots, y_n}(f_1, \ldots, f_n)(a)$ is invertible. For convenience, let us write $(x_{j_1}, \ldots, x_{j_n}) = (y_1, \ldots, y_n)$ and $(z_1, \ldots, z_{p-n})$ be the other x's. By the implicit function theorem, we can solve for $(y_1, \ldots, y_n)$ as a function of $(z_1, \ldots, z_{p-n})$. Let $x = P(z, y)$, where P is a linear transformation given by a change of coordinates. Then $(f \circ P)(z, y)$ satisfies $\partial_y (f \circ P)(P(n))$ being invertible, which means there necessarily exists a change of coordinates $(z, y) = H(u, v)$ such that $(f \circ P \circ H)(u, v) = v$. $\square$

We now introduce the rank theorem. Let $f : \mathbb{R}^p \to \mathbb{R}^n$ be C^n, and assume that f has rank r at every point m in a neighborhood of a given point a (so $r \leq p, r \leq n$). In (c) as above, $r = n$; in general, when $r = n$ is rank r at a, then it is rank r in some open neighborhood about a by continuity of determinant. Then there exists a coordinate change $x = H(u, v)$ with $v = (v_1, \ldots, v_r)$ and $u = (u_1, \ldots, u_{p-r})$, and some coordinate change K in the target space such that $(K \circ f \circ H)(u, v) = (v, 0)$. Then we may find indices $i_1 < \cdots < i_r$, $j_1 < \cdots < j_r$ such that $\partial_{j_1, \ldots, j_r}(f_{i_1}, \ldots, f_{i_r})(a)$ has rank r. We can assume that $(i_1, \ldots, i_r) = (1, \ldots, r)$ by permuting the coordinates in the target space; then there is a coordinate change $x = H(u, v)$, with $v = (v_1, \ldots, v_r)$ such that $(f_1, \ldots, f_r) \circ H(u, v) = v$, i.e.

$$\underbrace{(f \circ H)}_{=(f_1, \ldots, f_r)}(u, v) = (v, (f_{r+1}, \ldots, f_n) \circ H(u, v)).$$

§21 Day 21: Rank Theorem, Tangent Spaces (Oct. 23, 2024)

Let $f : \mathbb{R}^p \to \mathbb{R}^n$ be C^r in a neighborhood of a. Suppose f has constant rank r in a neighborhood of a; then there exists a C^r coordinate change $x = H(u, v)$ near a, and K near $f(a)$ such that $K \circ f \circ H : (u, v) \to (v, 0)$, where u, v are $p - 1, r$-tuples, i.e. $u = u_1, \dots, u_{p-1}$, and $v = v_1, \dots, v_r$. We also showed that after a permutation of coordinates in the target space, there exists a coordinate change $x = H(u, v)$ such that

$$(f \circ H)(u, v) = (v, \varphi(u, v)),$$
$$v = v_1, \dots, v_r,$$
$$\varphi = \varphi_1, \dots, \varphi_{n-r}.$$

(a) φ follows the index of u, i.e.

$$D(f \circ H) = \left(\begin{array}{c|c} 0 & I_r \\ \hline \partial_{u_i}\varphi_k & \partial_{v_j}\varphi_k \end{array}\right),$$

where $i = 1, \dots, p - r$, $j = 1, \dots, r$, and $k = 1, \dots, n - r$. Thus, rank r implies that $\partial_{u_i}\varphi_k = 0$ for all i, k, and so $\varphi(u, v) = \varphi(v)$; this means we have $(f \circ H)(u, v) = (v, \varphi(v))$ with (v, w) being the coordinates in the target space.

(b) We can make the change of coordinates K in the target space *after* the specific φs vanish, i.e. $(v, w) \mapsto (v, w - \varphi(v))$.

We now move onto tangent spaces.[11]

(a) Let us have a plane curve C, given by $f(x, y) = 0$ of which is C^r, with $\partial_y f(a, b) = 0$. By IVT, we can solve this as $y = g(x)$ being C^r near a, where $\varphi(a) = b$. The tangent space at (a, b) is given by $y - b = g'(a)(x - a)$, i.e. $f(x, g(x)) = 0$ and $\partial_x f(a, b) + \partial_y f(a, b)g'(a) = 0$. Expanding $g'(a)$, we get

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

is a tangent space to C at (a, b). By symmetry, we obtain an equivalent result if $\partial_x f(a, b) \neq 0$: if $\nabla f(a, b) \neq 0$, we can define a tangent line at (a, b) as $\partial_x f(a, b)(x - a) + \partial_y f(a, b)(y - b) = 0$. We say C is C^r smooth at (a, b), or a C^r manifold near (a, b).

(b) Let $M \subset \mathbb{R}^n$ be given by the set of points $f(x_1, \dots, x_n) = 0$, with f being C^r. Suppose $\nabla f(a) \neq 0$, i.e. $\partial_{x_i} f(a) \neq 0$ for some i. For example, if $i = n$, then $x_n = g(x_1, \dots, x_{n-1})$ near $(a_1, \dots, a_{n-1})$, where g is C^r, and $g(a_1, \dots, a_{n-1}) = a_n$. Then the tangent space at a is given by

$$x_n - a_n = g'(a_1, \dots, a_{n-1}) \begin{pmatrix} x_1 - a_1 \\ \vdots \\ x_{n-1} - a_{n-1} \end{pmatrix} = \sum_{i=1}^{n-1} \frac{\partial g}{\partial x_i}(a_1, \dots, a_{n-1})(x_i - a_i),$$

where $f(x_1, \dots, x_{n-1}, g(x_1, \dots, x_{n-1})) = 0$. Then for all $i = 1, \dots, n - 1$, we have

$$\frac{\partial f}{\partial x_i}(a) + \frac{\partial f}{\partial x_n}(a)\frac{\partial g}{\partial x_i}(\tilde{a}) = 0.$$

[11] mb if information isn't the most accurate, i am comprehending it myself atm x3

Then the $x(x_i - a_i)$ terms add up for $i = 1, \ldots, n-1$ to give

$$\frac{\partial f}{\partial x_i}(a)(x_1 - a_1) + \cdots + \frac{\partial f}{\partial x_i}(a)(x_n - a_n) = 0.$$

Specifically, if $\nabla f(a) \neq 0$, we can define a tangent space at a as the above formula, i.e. $\langle \nabla f(a), x - a \rangle = 0$. We say that M is C^r smooth at a, or a C^r manifold at a of dimension $n - 1$.

§22 Day 22: Tangent Spaces Pt. 2 (Oct. 25, 2024)

We start with an example;

(a) Let us find the tangent planes to the ellipsoid

$$X : \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$$

at a point $(x_0, y_0, z_0) \in X$. Since tangent planes are orthogonal to the derivative at (x_0, y_0, z_0), we have

$$\frac{2x_0}{a^2}(x - x_0) + \frac{2y_0}{b^2}(y - y_0) + \frac{2z_0}{c^2}(z - z_0) = 0,$$

or

$$\frac{x_0 x}{a^2} + \frac{y_0 y}{b^2} + \frac{z_0 z}{c^2} = 1.$$

In general, for $f : \mathbb{R}^p \to \mathbb{R}^n$, with $p \geq n$ and $M : \{x \in \mathbb{R}^p \mid f(x) = 0\}$, given a point $a \in M$ where f has rank n, we have

$$\det \frac{\partial(f_1, \dots, f_n)}{\partial(x_1, \dots, x_n)}(a) \neq 0.$$

For example, $f : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^n$, where $\det \partial_y f(a, b) \neq 0$, by the implicit function theorem, we can solve for $y = g(x)$ near (a, b), i.e. $g(a) = b$, $f(x, g(x)) = 0$ near $x = a$. The tangent space at (a, b) is given by $y - b = g'(a)(x - a)$, i.e.

$$\begin{pmatrix} y_1 - b_1 \\ \vdots \\ y_n - b_n \end{pmatrix} = \frac{\partial g_i}{\partial x_j}(a) \begin{pmatrix} x_1 - a_1 \\ \vdots \\ x_n - a_n \end{pmatrix}.$$

Given $f(x, g(x)) = 0$, we have $\partial_{x_i} f(a, b) + \sum_{j=1}^n \partial_{y_j} f(a, b) \partial_{x_i} g(a) = 0$ for $i = 1, \dots, m$, and $\partial_x f(a, b) + \partial_y f(a, b) g'(a) = 0$. Thus, we have

$$g'(a) = -\frac{\partial_y f(a, b)}{\partial_x f(a, b)},$$

i.e. the tangent space to M at a is $Df(a)(x - a) = 0$. When f has rank n at a, i.e., wherever tangent spaces may be defined, we say M is C^r smooth at a, or a manifold of dimension $p - n$ near a.

(a) Let $f(x, y) = 0$ and $g(x, y) = 0$ be two smooth curves in $\mathbb{R}^2$; we consider them at a common point (a, b) (meaning (a, b) lies on both the curves), i.e. $\nabla f(a) \neq 0$ and $\nabla g(a) \neq 0$. For example, the tangent lines to $f = 0$ at (a, b) are represented by

$$\frac{\partial f}{\partial x}(a, b)(x - a) + \frac{\partial f}{\partial y}(a, b)(y - b) = 0.$$

The curves are orthogonal at (a, b) if

$$\left(\frac{\partial f}{\partial x}\frac{\partial g}{\partial x} + \frac{\partial f}{\partial y}\frac{\partial g}{\partial y} \right)(a, b) = 0,$$

and tangent at (a, b) if

$$\left(\frac{\partial f}{\partial x}\frac{\partial g}{\partial x} - \frac{\partial f}{\partial y}\frac{\partial g}{\partial y} \right)(a, b) = 0.$$

(b) How to find a point of a surface $g(x, y, z) = 0$ which has least (or greatest) distance from the origin? i.e., find the extreme values[12] of $f(x, y, z) = x^2 + y^2 + z^2$ on a given surface $g(x, y, z) = 0$. Supposing we may solve $g(x, y, z)$ as $z = h(x, y)$, we now want to find the extreme values of $f(x, y, h(x, y))$; if the functions are C^r for $r \geq 1$, then extreme points occur at critical points, i.e.

$$\begin{aligned} f_x + f_z h_x &= 0, \\ f_y + f_z h_y &= 0, \\ g_x + g_z h_x &= 0, \\ g_y + g_z h_y &= 0. \end{aligned}$$

Then $f_x + \lambda g_x = 0$, $f_y + \lambda g_y = 0$, $f_z + \lambda g_z = 0$, where $\lambda = -\frac{f_z}{g_z}$, and $g(x, y, z) = 0$. There are 4 equations and 4 unknowns, so we may solve this. $\square$

[12] root is order preserving, so we may just look at f as the squares of components without rooting them all after

§23 Day 23: Lagrange's Method (Nov. 4, 2024)

We introduce Lagrange's method for obtaining extreme points on an objective function subject to a constraint.

Theorem 23.1 (Lagrange Method)**.** Let $g : U \to \mathbb{R}^p$ be $\mathcal{C}^r$ and $f : U \to \mathbb{R}$ be differentiable, with a local maximum or mnimum at some point $a \in g^{-1}(0)$. If g has rank p at a, i.e. $g^{-1}(0)$ is $\mathcal{C}^r$ smooth at a or $\nabla g_i(a)$ is linearly independent for p choices of i, then there exists $\lambda_1, \ldots, \lambda_p \in \mathbb{R}$ called *multipliers* such that we may construct the system of equations (where $i = 1, \ldots, p$)

$$\frac{\partial f}{\partial x_j}(a) + \lambda_i \frac{\partial g_i}{\partial x_j}(a) = 0,$$

as well as $g_i(a) = 0$. In particular, this may be written as $\nabla f(a) + \lambda \nabla g(a) = 0$, where $\lambda = (\lambda_1, \ldots, \lambda_p)$.

To prove this, let us have

$$\det \frac{\partial(g_1, \ldots, g_n)}{\partial(x_{j_1}, \ldots, x_{j_p})} \neq 0$$

for some $j_1 < j_2 < \cdots < j_p$. By permutation of variables, we can assume that $(x_{j_1}, \ldots, x_{j_p}) = (x_{n-p+1}, \ldots, x_n)$; with this, let us write $x = (u, v)$, where $(u, v) = (u_1, \ldots, u_{n-p}, v_1, \ldots, v_p)$. By the implicit function theorem, we solve $g(u, v) = 0$ for v as a $\mathcal{C}^1$ function $v = h(u)$, i.e. $g(u, h(u)) = 0$ over $u = (a_1, \ldots, a_{n-p})$); this means $f(u, h(u))$ has a local extremum at $u = (a_1, \ldots, a_{n-p})$, i.e. $(a_1, \ldots, a_{n-p})$ is a critical point of $\varphi(v) = f(u, h(u)) = f \circ H(u)$, where H is given by $H(u) = (u, h(u))$. Then we may write

$$\varphi'(u) = f'(u, h(u)) \begin{pmatrix} I \\ h'(u) \end{pmatrix},$$

and we have

$$0 = f'(a) \begin{pmatrix} I \\ h'(a_1, \ldots, a_{n-p}) \end{pmatrix} = \underbrace{\nabla f(a)}_{\in \ker} \begin{pmatrix} I \\ h'(a_1, \ldots, a_{n-p}) \end{pmatrix}.$$

Since $g_i(u, h(u)) = 0$ for $i = 1, \ldots, p$, we have that

$$0 = g_i'(a) \begin{pmatrix} I \\ h'(a_1, \ldots, a_{n-p}) \end{pmatrix} = \nabla g_i(a) \begin{pmatrix} I \\ h'(a_1, \ldots, a_{n-p}) \end{pmatrix};$$

since $\nabla g_i(a)$ is linearly independent and the latter matrix is of rank $n - p$, we conclude that $\dim \ker = p$, i.e. $\nabla f(a)$ is a linear combination of $\nabla g_i(a)$. $\square$

> **Remark 23.2.** To apply Lagrange's method, let us have a constraint function g and objective function f. If $\nabla g \neq 0$ over $g^{-1}(\{0\})$ and the set of all points where $g = 0$ is compact (it is useful to check the continuity of g, then use Heine-Borel here), then we may apply EVT to $f : g^{-1}(\{0\}) \to \mathbb{R}$ to see that it must admit an extreme point.

We now provide some examples.

(a) Suppose we want to find the point on a $\mathcal{C}^1$ hypersurface $g(x_1, \ldots, x_n) = 0$ at a local maximum or minimum distance from a fixed point x. We may use the objective function $f(x) = |x - c|^2$.[13] This method may fail if $g^{-1}(0)$ is not smooth at a.

[13] what?

(b) Find the triangle with given perimeter and largest possible area. We use the formula $f(x, y, z) = s(s-x)(s-y)(s-z)$, where the perimeter is given by $2s$, to find the area of the triangle.

Here, we want to find the maximum of f subject to the condition $g(x, y, z) = x + y + z - 2s = 0$, where $x, y, z \geq 0$, $x + y \geq z$, $y + z \geq x$, and $z + x \geq y$. On the boundary of the feasible set, one of the inequalities above becomes an equality, and so we have $f = 0$; thus, the maximum of f occurs at a critical point on the interior. Write

$$\frac{\partial f}{\partial x} + \lambda \frac{\partial g}{\partial x} = 0,$$

which yields the system

$$\begin{aligned} -s(s-y)(s-z) + \lambda &= 0, \\ -s(s-x)(s-z) + \lambda &= 0, \\ -s(s-x)(s-y) + \lambda &= 0. \end{aligned}$$

Substituting z, we have $x = y$, along with $x = z$, $y = z$ by symmetry. Thus, the area of the triangle is maximized when it is an equilateral.