

MAT257 TERM TEST 1 PREP/NOTES

JACK CERONI

CONTENTS

1. Introduction	2
2. Open Sets, Compactness, and the Basic Stuff	2
3. Differentiation	2
4. The Chain Rule	3
5. Rules for Taking Derivatives	3
6. The Inverse Function Theorem	3
6.1. Introduction	3
6.2. The Proof	3
7. The Implicit Function Theorem	7
8. Integration	8
9. Munkres Exercises	8
9.1. Munkres Problem 1.3	8
9.2. Munkres Problem 2.5.2	8
9.3. Munkres Problem 2.5.3	8
9.4. Munkres Problem 2.5.4	9
9.5. Munkres Problem 2.5.5	9
9.6. Munkres Problem 2.5.6	9
9.7. Munkres Problem 2.5.7	9
9.8. Munkres Problem 2.6.1	10
9.9. Munkres Problem 2.6.4	10
9.10. Munkres Problem 2.7.4	10
10. Spivak Exercises	10
11. Other Stuff	10
12. MAT257: 2020 Term Test 1 Rejects	11
13. MAT257: 2020 Term Test	11
13.1. Problem 1	11
13.2. Problem 2	11
13.3. Problem 3	11
13.4. Problem 4	11
13.5. Problem 5	11
14. MAT257: 2020 Final Exam Rejects	12
15. MAT257: 2016 Term Test	12
16. Axis Crawl Stuff	13
17. Proof of Inverse Function Theorem and Implicit Function Theorem (Again)	13

1. INTRODUCTION

2. OPEN SETS, COMPACTNESS, AND THE BASIC STUFF

The first thing that will be discussed in these notes is the notion of open and closed sets, compact sets, interiors, exteriors, and boundaries. All of these concepts are necessary to understand

3. DIFFERENTIATION

The first big idea developed in these notes is the notion of a derivative, which is simply a generalization of the definition of a derivative introduced in single-variable calculus.

Definition 1. A function $f : A \rightarrow \mathbb{R}^n$, for A open in \mathbb{R}^m is said to be differentiable at a if there exists some open neighbourhood U around a , such that for $a + h \in U$, we have:

$$f(a + h) = f(a) + Lh + o(h)$$

where L is linear, and o is a so-called “tiny” function, meaning that $\lim_{k \rightarrow 0} \frac{|o(k)|}{|k|} = 0$. The linear operator $L : A \rightarrow \mathbb{R}^m$ is called the differential of f at a , and is denoted by $[Df](a)$.

Notice that we called L the differential, instead of a differential, and for good reason!

Proposition 1. The differential at a , of a function that is differentiable at a is unique.

Proof. Suppose that there exist two linear operators L_1 and L_2 that satisfy the above criteria, so there exists U_1 such that:

$$f(a + h) = f(a) + L_1h + o_1(h)$$

and U_2 such that:

$$f(a + h) = f(a) + L_2h + o_2(h)$$

Note that $U = U_1 \cap U_2$ is open as well. We then have for $a + h \in U$:

$$(L_1 - L_2)h = o_2(h) - o_1(h)$$

for small enough h (so for $0 \leq |h| < \epsilon$, for some ϵ). Since o_1 and o_2 are tiny, it follows that choosing $h = tx$, where t is some vector, we will have:

$$\lim_{t \rightarrow 0} \frac{|o_2(tx) - o_1(tx)|}{|t|} = 0 \Rightarrow \lim_{t \rightarrow 0} \frac{|(L_1 - L_2)tx|}{|t|} = |(L_1 - L_2)x| = 0$$

so it follows that $(L_1 - L_2)x = 0$, so $L_1x = L_2x$. This holds for all x , so $L_1 = L_2$. Thus, the differential must be uniquely defined (it is always the same linear map from A to \mathbb{R}^m). \square

We can also prove this using the limit definition of the derivative in the same way:

Proof. Simply noting that:

$$\lim_{h \rightarrow 0} \frac{|f(a + h) - f(a) - L_1h|}{|h|} = 0 \quad \text{and} \quad \lim_{h \rightarrow 0} \frac{|f(a + h) - f(a) - L_2h|}{|h|} = 0$$

as well as:

$$0 \leq \frac{|(L_1 - L_2)h|}{|h|} \leq \frac{|f(a + h) - f(a) - L_1h|}{|h|} + \frac{|f(a + h) - f(a) - L_2h|}{|h|}$$

so by squeeze theorem, we have:

$$\lim_{h \rightarrow 0} \frac{|(L_1 - L_2)h|}{|h|} = 0$$

Thus, setting $h = tx$, for some x , we have $|(L_1 - L_2)x| = 0$, so $L_1 = L_2$. \square

4. THE CHAIN RULE

One of the most important results covered early on in these notes is the chain rule, which

5. RULES FOR TAKING DERIVATIVES

6. THE INVERSE FUNCTION THEOREM

6.1. Introduction. The goal of this first section will be to provide a detailed proof of the inverse function theorem, as it was done by Dror in class.

6.2. The Proof. A large fraction of the proof will rely on the principle that Dror referred to as “all-scale fidelity”, which means more-or-less that the distance between the vectors $x_1 - x_2$ and $f(x_1) - f(x_2)$ will be much less than the distance between x_1 and x_2 .

First, we must prove a small lemma which will allow us to make use of the “all-scale fidelity” principle in our proof in the inverse function theorem:

Lemma 1. *Given a map $f : A \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ such that f is continuously differentiable, if we have $|D_i f_j(a)| \leq M|a|$ for all $a \in \text{Int}(A)$, then:*

$$|f(x) - f(y)| \leq nmM|x - y|$$

for all $x, y \in A$.

Remark 1. *Intuitively, this lemma is making a fairly mundane claim: if the derivative of the function is bounded in all directions, then the “slope” of the line between two points is similarly bounded.*

In fact, an attentive reader may notice that for functions $f : \mathbb{R} \rightarrow \mathbb{R}$, this lemma is an immediate consequence of the mean-value theorem, which will turn out to be an important part of our proof. In fact, we can understand the proof of this lemma as crawling along a bunch of lines, each of which are parallel to a coordinate axis, applying mean-value theorem to each of these paths, and then summing up the result.

Without further delay:

Proof. Pick □

Now, have the machinery we need to prove the inverse function theorem. We begin by stating the theorem:

Theorem 1 (Inverse Function Theorem). *Suppose $f : A \rightarrow \mathbb{R}^n$ is continuously differentiable, with $A \subset \mathbb{R}^n$ open, with $\det[Df](x) \neq 0$ at some $x \in A$. Suppose in addition that $[Df](x) = \mathbb{I}$. Then there is an open set V containing x and an open set W containing $f(x)$, such that $f : V \rightarrow W$ has a continuous inverse $f^{-1} : W \rightarrow V$, which is differentiable for all $y \in W$ and satisfies:*

$$[Df^{-1}](y) = [Df](f^{-1}(y))^{-1}$$

There is a bunch of stuff we have to do to prove this theorem:

- (1) We have to find/prove existence of open sets V and W , and prove that f^{-1} is well-defined. In other words, f must be surjective and injective when treated as a map from V to W .
- (2) We have to prove that f^{-1} is continuous and differentiable.
- (3) We have to prove the formula for the differential of f^{-1} .

Part 1: Prep

First, we have to do a bit of “prep work” to make sure that we can prove all of the statement we outlined.

Lemma 2. *Given some $\epsilon > 0$ and $x \in A$ with f continuously differentiable at x and $[Df](x) = \mathbb{I}$, there exists some open neighbourhood of x inside which:*

$$|(f(y) - f(z)) - (y - z)| \leq \epsilon |y - z|$$

Proof. We define the “error function” $g(x) = f(x) - x$. This is simply the error between f and the identity. Since we know f behaves like the identity close to x , it follows that g will become small.

Firstly, since $[Df](x) = \mathbb{I}$, it is easy to see that $[Dg](x) = 0$. Thus, $D_i g_j(x) = 0$, for all i and j .

Since f and therefore g is continuously differentiable, it follows that each $D_i g_j$ is continuous. Thus, for each i and j , we can choose some open neighbourhood U_{ij} around x such that:

$$|D_i g_j(z) - D_i g_j(x)| = |D_i g_j(z)| < \frac{\epsilon}{n^2}$$

for all $z \in U_{ij}$. We can take the intersection of all such neighbourhoods to ensure that all partials are bounded above by $\frac{\epsilon}{n^2}$. Thus, applying the lemma we proved earlier, we have:

$$|g(y) - g(z)| - |(f(y) - f(z)) - (y - z)| \leq n^2 \cdot \frac{\epsilon}{n^2} |y - z| = \epsilon |y - z|$$

□

What have we just done? We have used the assumptions stated in the inverse function theorem to show that f satisfies “all-scale fidelity”, for whatever value of ϵ we choose.

We know that $\det[Df](x) \neq 0$. Note that the determinant of $[Df](x)$ just ends up being a sum of product of $D_j f_i$, which are all continuous. Thus, the determinant is also a continuous function of x . It follows that we can choose a neighbourhood V around x in which the determinant is non-zero. This will become useful in the next section.

Part 2: Existence

Now, we turn our attention to finding open sets V and W . In the previous section, we managed to find some open sets with some pretty nice properties.

We claim that $V = B_r$ and $W = B_r$ are the open sets we desire. This is a fairly big claim to make, and we will have to prove a bunch of stuff in order to justify this claim. To start, we can attempt to show that $f : V \rightarrow W$ is bijective, so at the very least, the inverse is well-defined.

Remark 2. *Proving injectivity is fairly easy! We make use of all-scale fidelity. Since the distance between vectors remain roughly the same when mapped under f in V , it follows that two points x and y can't be mapped to the same point under f , or else the lengths of the vectors would be “too different”.*

More formally, pick $x, y \in V$. Assume $f(x) = f(y)$. Note that by all-scale fidelity, we have:

$$|(f(x) - f(y)) - (x - y)| = |x - y| \leq \frac{1}{2} |x - y|$$

which can clearly only be the case when $x = y$. Thus, by definition, f is injective. Proving surjectivity is bit more difficult. We first make the following remark:

Remark 3. *The idea behind the proof of surjectivity is fairly intuitive: since all-scale fidelity essentially corresponds to f behaving locally like the identity, if we pick some y , then we can define a displacement vector from $f(a)$ of the form $y - f(a)$. Our hope is that if we “step” along this displacement vector from a in the domain, then the endpoint of this step will be close to y .*

Indeed, this is the case. If we pick some $y \in W$, and define $x = a + (y - f(a))$, then we have:

$$|f(x) - y| = |(f(x) - f(a)) - (x - a)| \leq \epsilon|x - a| = \epsilon|f(a) - y|$$

so the distance between $f(x)$ and y is smaller than the distance between y and $f(a)$.

Seeing this, our naive hope is that we can continue to apply this logic arbitrarily: we define a sequence of points x_n , and then apply f repeatedly, so that the distance between $f(x_n)$ and y is less than the distance between $f(x_{n-1})$, repeating this procedure until the distance converges to 0, and we have a point x such that $f(x) = y$. As it turns out, this works! This is exactly how we will go about proving surjectivity.

Lemma 3. *The function $f : V \rightarrow W$ is surjective.*

Proof. We will define a sequence of points as follows:

$$x_0 = a + (y - f(a)) \quad x_{n+1} = x_n + (y - f(x_n))$$

We assert that if $y \in B_{r/2}(f(a))$, then the sequence of points given here will be in $B_r(a)$. Indeed, the point $x_0 \in B_r$, as $|x_0 - a| = |y - f(a)| < \frac{r}{2}$. We proceed by induction. Assume $x_n \in B_r$. We have:

$$|x_{n+1} - a| = |x_n - a + y - f(x_n)| \leq |(f(a) - f(x_n)) - (a - x_n)| + |y - f(a)| < \frac{1}{2}|a - x_n| + \frac{r}{2} < \frac{r}{2} + \frac{r}{2} = r$$

so $x_{n+1} \in B_r(a)$. Thus, every x_n is in B_r . We then note that:

$$|x_{n+1} - x_n| = |y - f(x_n)|$$

In addition, we have $y = x_n - x_{n-1} + f(x_{n-1})$. Thus:

$$|y - f(x_n)| = |f(x_{n-1}) - f(x_n) - (x_{n-1} - x_n)| \leq \frac{1}{2}|x_n - x_{n-1}|$$

Thus, we have $|x_{n+1} - x_n| = \frac{1}{2^n}|y - f(x_0)|$. This sequence is clearly Cauchy, so it converges. We say $x_n \rightarrow x$. Intuitively, we know that x is the point such that $f(x) = y$, but we have to prove this. Since f is continuous, we have $f(x_n) \rightarrow f(x)$. Thus:

$$|f(x) - y| = \lim_{n \rightarrow \infty} |f(x_n) - y| \leq \lim_{n \rightarrow \infty} \frac{1}{2}|x_n - x_{n-1}| = 0$$

Thus, $f(x) = y$. Finally, note that:

$$|x - a| \leq |(f(x) - f(a)) - (x - a)| + |y - f(a)| < \frac{1}{2}|x - a| + \frac{r}{2} \Rightarrow |x - a| < r$$

so $x \in B_r(a)$. Thus, the map f is surjective. \square

Part 3: The Differential

Before jumping into the proof of differentiability, we can do something a bit simpler: prove the formula for the differential of the inverse function.

More specifically, let us assume that $f^{-1} : W \rightarrow V$ is differentiable, with an invertible differential. Note that:

$$f^1 \circ f = \mathbb{I}$$

Thus, by the chain rule:

$$[D(f^{-1} \circ f)](x) = [D\mathbb{I}](x) \Rightarrow [Df^{-1}](f(x)) \circ [Df](x) = \mathbb{I} \Rightarrow [Df^{-1}](x) = [Df](f^{-1}(x))^{-1}$$

so we have proved the formula for the differential. All that is left to do is to show that f^{-1} is in fact differentiable, which we will do in the next section.

Part 4: Differentiability

The above proof is the main part of the proof: proving continuity and differentiability of f^{-1} is a bit easier. More specifically, we are going to modify the statement of all-scale fidelity so that it works backwards (in a sense).

Remark 4. *All-scale fidelity is a very powerful tool. For example, if we know that all-scale fidelity holds, we automatically know that f is Lipschitz continuous, as we have:*

$$|f(x) - f(y)| - |x - y| \leq |(f(x) - f(y)) - (x - y)| < \epsilon|x - y| \Rightarrow |f(x) - f(y)| < (1 + \epsilon)|x - y|$$

for all x and y in the region where all-scale fidelity holds. In light of this power, using all-scale fidelity to prove differentiability of f^{-1} is likely a good plan.

Remember that we now have a well-defined inverse function $f^{-1} : W \rightarrow V$, so we can certainly reap the benefits of having a bijective function in some regions V and W . Before jumping into the new all-scale fidelity, let us stop for a moment and consider what we want to show. We need a neighbourhood of x in which:

$$f^{-1}(x + h) = f^{-1}(x) + Lh + o(h)$$

where L is linear. Now, from the previous section, we know that the only choice for L is the differential $[Df](f^{-1}(x))^{-1}$. Thus, we need to show the the following quantity is small:

$$f^{-1}(x + h) - f^{-1}(x) - [Df](f^{-1}(x))^{-1}h$$

Recall that in the region we are examining, $Df(x)$ is almost the identity, whose inverse is the identity. Thus, you may notice that the above expression is “almost”:

$$f^{-1}(x + h) - f^{-1}(x) - h = f^{-1}(x + h) - f^{-1}(x) - (x + h - x)$$

which looks very similar to the difference of vectors in the definition of all-scale fidelity. This suggests to us (a bit more explicitly) that coming up with a version of all-scale fidelity for f^{-1} might be a good idea.

Indeed, we will do just this. It’s not too tricky.

Lemma 4 (Inverse All-Scale Fidelity). *Pick y_1 and y_2 . Note that letting $x_1 = f^{-1}(y_1)$ and $x_2 = f^{-1}(y_2)$ gives us, by all-scale fidelity:*

$$|f^{-1}(y_1) - f^{-1}(y_2)| - |y_1 - y_2| \leq |(f^{-1}(y_1) - f^{-1}(y_2)) - (y_1 - y_2)| \leq \epsilon|f^{-1}(y_1) - f^{-1}(y_2)|$$

which implies that for $\epsilon < 1$:

$$|f^{-1}(y_1) - f^{-1}(y_2)| \leq \frac{1}{1 - \epsilon}|y_1 - y_2|$$

Thus, going back to the original equation for all-scale fidelity, we have:

$$|(f^{-1}(y_1) - f^{-1}(y_2)) - (y_1 - y_2)| \leq \epsilon|f^{-1}(y_1) - f^{-1}(y_2)| \leq \frac{\epsilon}{1 - \epsilon}|y_1 - y_2|$$

Note that for something like $\epsilon = 1/10$, we get $\epsilon/(1-\epsilon) = 1/9$, so our “new value” of ϵ is pretty close to the old value: an approximation which only improves as ϵ becomes smaller.

Proving continuity is basically trivial now (recall we showed that all-scale fidelity implies Lipschitz continuous, which obviously implies continuous). To prove differentiability, we return to what we were saying above. We need to show that:

$$f^{-1}(x+h) - f^{-1}(x) - (x+h-x)$$

is small. But from all-scale fidelity, we know that:

$$|f^{-1}(x+h) - f^{-1}(x) - (x+h-x)| \leq \frac{\epsilon}{1-\epsilon}|h|$$

for any $\epsilon < 1$. Thus, given some $\epsilon' > 0$, we choose $\frac{1}{n} < \epsilon'$, and note that if $\epsilon = \frac{1}{n+1}$, then:

$$|f^{-1}(x+h) - f^{-1}(x) - (x+h-x)| \leq \frac{\frac{1}{n+1}}{\frac{n}{n+1}}|h| = \frac{1}{n}|h| < \epsilon'|h|$$

so it follows by definition that the function given is small. Thus, f^{-1} is differentiable for $x \in W$. And we are done!

Part 5: Generalizing

Above, we managed to prove a special case of the inverse function theorem, namely, the version where $f'(a) = \mathbb{I}$. What if this is not the case? It turns out, it doesn't matter!

Let f be a function which satisfies the conditions of the inverse function theorem. Thus, $\lambda = [Df](a)$ is invertible. We let $g = \lambda^{-1} \circ f$. Clearly:

$$[Dg](x) = [D\lambda^{-1}](f(x)) \circ [Df](x) = \lambda^{-1} \circ \lambda = \mathbb{I}$$

In addition, $\lambda^{-1} \circ g$ clearly obeys the conditions of the inverse function theorem (it is continuously differentiable in A , and $[Dg](a)$ is invertible). In addition, $g'(a) = \mathbb{I}$. Thus, we can apply the lite-version of the inverse function theorem we proved above to conclude that there exist open V and W around a and $g(a)$, such that g has a differentiable inverse $h : W \rightarrow V$ for which:

$$[Dh](y) = [[Dg](h(y))]^{-1}$$

Consider the restriction $f : V \rightarrow \lambda(W)$, where $\lambda(W)$ is clearly open. Clearly, $f(x) = (\lambda \circ g)(x)$ for all $x \in V$. Since λ is a bijection from W to $\lambda(W)$, and g is a bijection from V to W , it follows that f has an inverse $k = g^{-1} \circ \lambda^{-1}$. Clearly, this function is differentiable, as g^{-1} and λ^{-1} are. Finally, note that:

Finally, we proved in generality that if f is a function, and its inverse is differentiable, then its differential is given by the formula. Thus, we are done: we have proved the inverse function theorem for f with arbitrary $f'(a)$.

7. THE IMPLICIT FUNCTION THEOREM

This next section of the notes will be dedicated to proving the implicit function theorem, which can be stated as follows:

Theorem 2 (Implicit Function Theorem).

Note that the proof of the implicit function theorem is quite simple, now that we have proved the inverse function theorem.

The implicit function theorem basically gives us some nice, sufficient conditions for solving a system of equations. Basically, if we have a function $f : \mathbb{R}^n \times \mathbb{R}^k \rightarrow \mathbb{R}^k$, such that f is continuously differentiable in some neighbourhood around a point a , with $f(a) = 0$, and it is also the case that the submatrix of partial derivatives with respect to each of the \mathbb{R}^k coordinates is invertible, then we can solve for the \mathbb{R}^k coordinates in terms of the \mathbb{R}^n ones. More specifically, there exists a function $g : A \rightarrow B$, where $(a_1, \dots, a_n) \in A$ and $(a_{n+1}, \dots, a_{n+k}) \in B$ and A and B are open, such that g “assigns” to the \mathbb{R}^n coordinates in A to some coordinates \mathbb{R}^k in B which “solve” f . In other words, $f(x, g(x)) = 0$, for all $x \in A$. We can also say that g is differentiable, so it is a pretty nice function!

8. INTEGRATION

Now, we can finally move onto integration. First, we need to do a bit of housekeeping involving the notion of oscillation:

Definition 2. *Given some*

9. MUNKRES EXERCISES

9.1. Munkres Problem 1.3. Clearly, X is bounded, as $\langle e_i - e_j, e_i - e_j \rangle = 0, 2$. Now, choose some $y \notin X$. Note that $y_j = 0$ for all $j \geq N$, for some N . It follows that $s = \min\{\|y_j - e_j\| \mid 1 \leq j < N\}$ is well-defined. Note that for $j \geq N$, we have $\|y_j - e_j\| = \sqrt{\langle y, y \rangle + 1} = r$. Taking $\epsilon = \min\{s, r\}$, and we can define an open ball around y that does not contain any e_j . Thus, X is closed.

Finally, we note that X is not compact. We simply take the $\epsilon = 1$ ball around each e_j , and note that since $\|e_j - e_i\| = \sqrt{2}$ for $j \neq i$, each of these sets are disjoint. Thus, this open cover has no finite subcover.

9.2. Munkres Problem 2.5.2. Note that:

$$\frac{f(t(h, k))}{t} = \frac{1}{t} \frac{hk}{h^2 + k^2}$$

for h, k and t non-zero. Thus, $f'(0, u)$, with $u = (h, k)$ exists only when $h = 0$ or $k = 0$, and in these cases it is 0. This implies immediately that $D_1 f(0)$ and $D_2 f(0)$ exist and equal 0. Note that f is not differentiable, as there are directional derivatives that do not exist. Finally, note that f is not continuous. For example:

$$\lim_{t \rightarrow 0} f(t, t) = \lim_{t \rightarrow 0} \frac{1}{2} = \frac{1}{2} \neq f(0)$$

9.3. Munkres Problem 2.5.3. Note that:

$$\left| \frac{f(t(h, k))}{t} \right| = \left| \frac{th^2k^2}{t^2h^2k^2 + (h - k)^2} \right|$$

for $h \neq k$, the above is less than $|t| \left| \frac{h^2k^2}{(h - k)^2} \right|$, which goes to 0 as $t \rightarrow 0$. Thus, in this case, the directional derivative exists. In the case that $h = k$, we have:

$$\frac{th^2k^2}{t^2h^2k^2 + (h - k)^2} = \frac{th^2k^2}{t^2h^2k^2} = \frac{1}{t}$$

so the limit does not exist. Thus, the partials exist, but the function is not differentiable at 0, as some directional derivatives are not defined. This function is not continuous: setting $x = y$, and letting $x \rightarrow 0$ yields 1, which is clearly not equal to $f(0)$.

9.4. **Munkres Problem 2.5.4.** We have:

$$\frac{f(t(h,k))}{t} = \frac{h^3}{k^2 + h^2}$$

so the directional derivative is always defined. It follows that the partials are defined and equal to 1 and 0. If the function is differentiable, it must be the case that $Df(0) = (1, 0)$. Thus, we must conclude that:

$$f(h, k) = \frac{h^3}{k^2 + h^2} - h = -\frac{hk^2}{k^2 + h^2}$$

is small. We need to have:

$$\frac{f(h, k)}{|(h, k)|} = -\frac{hk^2}{(k^2 + h^2)^{3/2}}$$

go to 0 as $(h, k) \rightarrow 0$. However this is not the case: taking $h = k$, the above expression becomes $-\frac{1}{\sqrt{8}}\text{sign}(h)$, the limit of which clearly does not exist as $h \rightarrow 0$. Note that this function is continuous. Note that $h^2 \leq k^2 + h^2$, so:

$$\left| \frac{h^3}{k^2 + h^2} \right| \leq \left| \frac{h^3}{h^2} \right| = |h|$$

which clearly goes to 0 as $(h, k) \rightarrow 0$. Hence, by squeeze theorem, $\lim_{(x,y) \rightarrow 0} f(x, y) = f(0, 0) = 0$, so the function is continuous.

9.5. **Munkres Problem 2.5.5.** Note that the limit as $t \rightarrow 0$ of:

$$\frac{f(t(h, k))}{t} = \frac{|t|(|h| + |k|)}{t} = \text{sign}(t)(|h| + |k|)$$

is clearly not defined, for any $u = (h, k)$. Thus, no directional derivatives exist, including the partials, and as a result, the function is not differentiable at 0. It is obvious that this function is continuous at 0.

9.6. **Munkres Problem 2.5.6.** Note that the limit as $t \rightarrow 0$:

$$\frac{f(t(h, k))}{t} = \text{sign}(t)\sqrt{|hk|}$$

is only defined when $h = 0$ or $k = 0$. In either case, the limit is 0. The function is not differentiable, as some directional derivatives are not defined. Clearly, it is continuous, as $\sqrt{|xy|} \leq \sqrt{x^2 + y^2} = |(x, y)|$.

9.7. **Munkres Problem 2.5.7.** Note that:

$$\frac{f(t(h, k))}{t} = \frac{t|t| |h| |k|}{t|t| \sqrt{h^2 + k^2}} = \frac{h|k|}{\sqrt{h^2 + k^2}}$$

Thus, the directional derivative is always defined, with the partials both being 0. Thus, if this function is differentiable, we must have $Df(0) = 0$. We then must have:

$$\frac{f(h, k)}{|(h, k)|} \rightarrow 0$$

as $(h, k) \rightarrow 0$. But we have:

$$\frac{f(h, k)}{|(h, k)|} = \frac{h|k|}{h^2 + k^2}$$

Choosing $h = k$, the above is equal to $\frac{1}{2}\text{sign}(h)$, the limit of which does not exist as $(h, k) \rightarrow 0$. Hence, the limit does not exist, so f is not differentiable at 0. Clearly, it is continuous:

$$|f(h, k)| = \left| \frac{h|k|}{\sqrt{h^2 + k^2}} \right| \leq \sqrt{h^2 + k^2} \rightarrow 0$$

as $(h, k) \rightarrow 0$.

9.8. Munkres Problem 2.6.1. We assert that this function has derivative 0. Indeed, note that the function:

$$f(0 + h) - f(0) = f(h) = |ab|$$

is small, as:

$$0 \leq \frac{|ab|}{\sqrt{a^2 + b^2}} \leq \frac{a^2 + b^2}{\sqrt{a^2 + b^2}} = \sqrt{a^2 + b^2}$$

which clearly goes to 0 as $(a, b) \rightarrow 0$. Thus, the function is differentiable. Now, if we choose a neighbourhood U around $(0, 0)$, note that U contains a point of the form $(0, a)$. Note that $D_1 f(0, a)$ isn't defined, as:

$$\frac{f(t, a) - f(0, a)}{t} = \frac{|t||a|}{t} = \text{sign}(t)|a|$$

where the limit of the above clearly does not exist as $t \rightarrow 0$. Thus the function is not C^1 .

9.9. Munkres Problem 2.6.4. We can perform an axis crawl. Note that:

$$f(a + h) - f(a) = \sum_j f(t_{j+1}) - f_j = \sum_j f(t_j + h_j e_j) - f_j(t_j)$$

By mean value theorem, there exists some $c_j \in (0, h_j)$ such that $D_j f(c_j) = [f(t_j + h_j e_j) - f_j(t_j)]h_j$. Thus:

$$\left| \sum_j f(t_j + h_j e_j) - f_j(t_j) \right| = \left| \sum_j D_j f(c_j) h_j \right| \leq M \sum_j |h_j|$$

which clearly goes to 0 as $h \rightarrow 0$. It follows that f is continuous at a .

9.10. Munkres Problem 2.7.4.

10. SPIVAK EXERCISES

11. OTHER STUFF

There are a few other important ideas that I feel are glossed over too much in Spivak and Munkres. For example, the following proof:

Proposition 2. *A limit with respect to a vector-valued function exists only when the limit exists when restricted to a line.*

12. MAT257: 2020 TERM TEST 1 REJECTS

13. MAT257: 2020 TERM TEST

13.1. **Problem 1.** Since the sequence is Cauchy, it converges to some point x . It may be the case that $x \in A$, or $x \notin A$. Either way, we can consider the set $A \cup \{x\}$, and note that any open cover contains an open set around x , which contains all but finitely many of the x_k . Taking elements of the cover around the remaining x_k gives a finite subcover, so A is compact.

Consider the sequence $x_k = k$, for $k \in \mathbb{N}$. Clearly, this sequence is not Cauchy, as $|x_j - x_i| = |j - i|$. Clearly, A is not compact, as it is not bounded. Adding a point will not make it bounded, so the above does not hold.

13.2. **Problem 2.** First, consider $\mu \circ \gamma$. Pick $\epsilon > 0$. We choose δ such that $|x| < \delta \Rightarrow |\mu(x)| < \frac{\epsilon}{C}|x|$. We choose δ' such that $|x| < \delta' \Rightarrow |\lambda(x)| \leq C|x|$. If we pick $r < \min\{\delta', \delta/C\}$, we will have $|x| < r$ imply that $|\lambda(x)| \leq C|x|$, and $C|x| < \delta$, so $|\lambda(x)| < \delta$. Thus:

$$|x| < r \Rightarrow |\lambda(x)| < \delta \Rightarrow |\mu(\lambda(x))| < \frac{\epsilon}{C}|\lambda(x)| \leq \epsilon|x|$$

which implies that $\mu \circ \gamma$ is tiny.

Now, consider $\lambda \circ \mu$. Pick $\epsilon > 0$. We choose δ such that $|x| < \delta \Rightarrow |\lambda(x)| \leq C|x|$. We choose δ' such that $|x| < \delta' \Rightarrow |\mu(x)| < \frac{\epsilon}{C}|x|$. We then note that picking $r = \min\{\delta', \delta C/\epsilon\}$ gives $|x| < r \Rightarrow |\mu(x)| < \frac{\epsilon}{C}|x|$, and $\frac{\epsilon}{C}|x| \leq \delta$, so $|\mu(x)| < \delta$ as well. Thus:

$$|x| < r \Rightarrow |\mu(x)| < \delta \Rightarrow |\lambda(\mu(x))| \leq C|\mu(x)| < \epsilon|x|$$

so $\lambda \circ \mu$ is tiny, by definition.

13.3. **Problem 3.** Note that:

$$|x_1 - x_2| - |f(x_1) - f(x_2)| \leq |(x_1 - x_2) - (f(x_1) - f(x_2))| \leq \epsilon|x_1 - x_2|$$

which implies that:

$$(1 - \epsilon)|x_1 - x_2| \leq |f(x_1) - f(x_2)|$$

Thus, if we consider $x_1 = f^{-1}(y_1)$ and $x_2 = f^{-1}(y_2)$, we will have:

$$(1 - \epsilon)|f^{-1}(y_1) - f^{-1}(y_2)| \leq |y_1 - y_2| \Rightarrow |f^{-1}(y_1) - f^{-1}(y_2)| \leq \frac{1}{1 - \epsilon}|y_1 - y_2|$$

as we have assumed $\epsilon < 1$. This allows us to conclude that:

$$|(f^{-1}(y_1) - f^{-1}(y_2)) - (y_1 - y_2)| \leq \epsilon|f^{-1}(y_1) - f^{-1}(y_2)| \leq \frac{\epsilon}{1 - \epsilon}|y_1 - y_2|$$

and the proof is complete. Note that for $\epsilon = 1/10$, we have $\frac{\epsilon}{1 - \epsilon} = \frac{1}{9}$.

13.4. **Problem 4.** See Munkres Problem 2.6.4, given above.

13.5. **Problem 5.** We apply the chain rule. Note that:

$$[Dh](x, y) = [D(f \circ k)](x, y) = [Df](k(x, y)) \circ [Dk](x, y)$$

where $k(x, y) = (x, y, g(x, y))$. Clearly:

$$[Dk](x, y) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ D_1g(x, y) & D_2g(x, y) \end{pmatrix}$$

$$\text{and } [Df](k(x, y)) = \begin{pmatrix} D_1f(x, y, g(x, y)) & D_2f(x, y, g(x, y)) & D_3f(x, y, g(x, y)) \end{pmatrix}$$

so it follows that:

$$\begin{aligned} & [Df](k(x, y)) \circ [Dk](x, y) \\ &= (D_1f(x, y, g(x, y)) + D_3f(x, y, g(x, y))D_1g(x, y) \quad D_2f(x, y, g(x, y)) + D_3f(x, y, g(x, y))D_2g(x, y)) \end{aligned}$$

In addition, if we know that $h(x, y) = 0$, then it follows that the entries of the matrix given above are both 0. Hence:

$$D_1f(x, y, g(x, y)) + D_3f(x, y, g(x, y))D_1g(x, y) = 0$$

and

$$D_2f(x, y, g(x, y)) + D_3f(x, y, g(x, y))D_2g(x, y) = 0$$

Thus, for (x, y) where $D_3f(x, y, g(x, y))$ is invertible (in this case, non-zero), we have:

$$D_1g(x, y) = -\frac{D_1f(x, y, g(x, y))}{D_3f(x, y, g(x, y))} \quad \text{and} \quad D_2g(x, y) = -\frac{D_2f(x, y, g(x, y))}{D_3f(x, y, g(x, y))}$$

for cases where $D_3f(x, y, g(x, y))$ is equal to 0, we cannot say anything about the partials of g .

14. MAT257: 2020 FINAL EXAM REJECTS

15. MAT257: 2016 TERM TEST

16. AXIS CRAWL STUFF

17. PROOF OF INVERSE FUNCTION THEOREM AND IMPLICIT FUNCTION THEOREM (AGAIN)

In this section, we will prove the inverse and implicit function theorems, once again. We begin by proving a small lemma, which will be useful to us:

Proposition 3. *Given some $f : \mathbb{R}^m \rightarrow \mathbb{R}^n$, such that f is C^1 in some neighbourhood A , and $|D_i f_j(x)| \leq M$ for all i and j , then $|f(x) - f(y)| \leq Mmn|x - y|$.*

Proof. Pick some component function f_j . We will have:

$$|f_j(x) - f_j(y)| = \left| \sum_k f_j(t_k) - f_j(t_{k-1}) \right|$$

where $t_0 = x$, and $t_k = x + (y_1 - x_1)e_1 + \cdots + (y_k - x_k)e_k$, so we are effectively “crawling along the axes”. By mean value theorem, we can find $c_k \in A$ with $c_k = t_{k-1} + h_k e_k$, with $h_k \in (0, y_k - x_k)$, such that $D_i f_j(c_k)(t_k - t_{k-1}) = f_j(t_k) - f_j(t_{k-1})$. Thus:

$$\left| \sum_k f_j(t_k) - f_j(t_{k-1}) \right| \leq \sum_k |f_j(t_k) - f_j(t_{k-1})| \leq M \sum_k |t_k - t_{k-1}| = M \sum_k |y_k - x_k| \leq Mm|x - y|$$

Finally, we have:

$$|f(x) - f(y)| \leq \sum_j |f_j(x) - f_j(y)| \leq Mmn|x - y|$$

□