Math 140C Lecture Notes (Professor: Luca Spolaor)

Isabelle Mills

May 22, 2024

Lecture 1: 4/2/2024

A set $X \subseteq \mathbb{R}^n$ where $X \neq \emptyset$ is a vector space if:

- $\vec{x}, \vec{y} \in X \implies \vec{x} + \vec{y} \in X$
- $\vec{x} \in X$ and $c \in \mathbb{R} \implies c\vec{x} \in X$.

If $\phi = \{\vec{x}_1, \ldots, \vec{x}_k\} \subset \mathbb{R}^n$, then we define:

$$\operatorname{span} \phi = \operatorname{span}\{\vec{x}_1, \ldots, \vec{x}_k\} = \{c_1\vec{x}_1 + \ldots + c_k\vec{x}_k \mid c_1, \ldots, c_k \in \mathbb{R}\}.$$

If $E \subseteq \mathbb{R}^n$ and $E = \operatorname{span} \phi$, then we say ϕ generates E.

Note that $\operatorname{span}\{\vec{x}_1, \ldots, \vec{x}_2\}$ forms a vector space (this is trivial to check).

$\{\vec{x}_1, \ldots, \vec{x}_k\} \subseteq \mathbb{R}^n$ is called linearly independent if:

$$\sum_{i=1}^{k} c_i\vec{x}_i = 0 \implies \forall i \in \{1, \ldots, k\}, \;\; c_i = 0.$$

If the above implication does not hold, then we call the set linearly dependent.

If $X \subseteq \mathbb{R}^n$ is a vector space, then we define the dimension of X as:
$\dim(X) = \sup\{k \in \mathbb{N} \cup \{0\} \mid \exists\{\vec{x}_1, \ldots, \vec{x}_k\} \subset X$ which is linearly independent$\}$.

Also, we define any set containing $\vec{0}$ to be automatically linearly dependent. This includes the singleton: $\{\vec{0}\}$.

$Q = \{\vec{x}_1, \ldots, \vec{x}_k\}$ is a basis for X if:

- Q is linearly independent.
- $\operatorname{span} Q = X$

As an example of a basis, for $\mathbb{R}^n$ we define the standard basis as the set $\{e_1, e_2, \ldots, e_n\}$ where e_i is the vector whose ith element is 1 and whose other elements are 0. It is pretty trivial to check that this set is in fact a basis of $\mathbb{R}^n$.

Proposition: If $B = \{\vec{x}_1, \ldots, \vec{x}_k\}$ is a basis of a vector space X, then:

1. $\forall \vec{v} \in X, \;\; \exists c_1, \ldots, c_k \in \mathbb{R} \;\; s.t. \;\; \vec{v} = \sum_{i=1}^{k} c_i\vec{x}_i$

   This is true because $X = \operatorname{span} B$. So by definition of a span, $\vec{v}$ can be expressed as a linear combination of the vectors of B.

2. The c_i such that $\vec{v} = \sum_{i=1}^{k} c_i \vec{x}_i$ are unique.

   Suppose that $\vec{v} = \sum c_i \vec{x}_i = \sum \alpha_i \vec{x}_i$. Then $\vec{0} = \sum (c_i - \alpha_i)\vec{x}_i$. Then since $\{\vec{x}_1, \ldots, \vec{x}_k\}$ are linearly independent, we know for all i that $c_i - \alpha_i = 0$. Hence, $c_i = \alpha_i$ for each i.

Theorem 9.2: Let $k \in \mathbb{N} \cup \{0\}$. If $X = \text{span}\{\vec{x}_1, \ldots, \vec{x}_k\}$, then $\dim(X) \leq k$.

Proof:
Suppose for the sake of contradiction that for any $m \in \mathbb{Z}_+$, there exists a linearly independent set $Q = \{\vec{y}_1, \ldots, \vec{y}_{k+m}\} \subset X$ which spans X. Then, define $S_0 = \{\vec{x}_1, \ldots, \vec{x}_k\}$ and note that S_0 spans X.

Now by induction, assume for $i \in \{0, 1, \ldots, k-1\}$, that S_i contains the first i vectors of Q in addition to $k - i$ vectors of S_0, and that $\text{span } S_i = X$. Then since S_i spans X, we know that $y_{i+1} \in X$ is in the span of S_i. So, letting $\vec{x}_{n_1}, \ldots, \vec{x}_{n_{k-i}}$ be the elements from S_0 in S_i, we know that there exists scalars $a_1, \ldots, a_{i+1}, b_1, \ldots, b_{k-i} \in \mathbb{R}$ where $a_{i+1} = 1$ such that:

$$\sum_{j=1}^{i+1} a_j \vec{y}_j + \sum_{j=1}^{k-i} b_j \vec{x}_{n_j} = \vec{0}$$

If all $b_j = 0$, then we have a contradiction. This is because $\{\vec{y}_1, \ldots, \vec{y}_{k+1}\}$ is assumed to be linearly independent. So, having all $b_j = 0$ implies that:

$$\sum_{j=1}^{i+1} a_j \vec{y}_j = \sum_{j=1}^{i+1} a_j \vec{y}_j + \sum_{j=i+2}^{k+1} 0 \cdot \vec{y}_j = \vec{0}$$

In turn this means that all $a_j = 0$, which contradicts that $a_{i+1} = 1$.

So, not all $b_j = 0$. This means that for some j we must have that $\vec{x}_{n_j}$ is in the span of $(S_i \setminus \{\vec{x}_{n_j}\}) \cup \{\vec{y}_{i+1}\}$. Call this set S_{i+1}. Clearly, S_{i+1} contains the first $i + 1$ vectors of Q. Also:

$$\text{span } S_{i+1} = \text{span } (S_i \cup \{\vec{y}_{i+1}\}) = \text{span } S_i = X.$$

So S_{i+1} satisfies the same conditions S_i did.

Now we get to the contradiction. Using the above reasoning, we will eventually construct $S_k = \{\vec{y}_1, \ldots, \vec{y}_k\}$ which still spans X. However, since $\vec{y}_{k+1} \in X$, that means that $\vec{y}_{k+1}$ equals some linear combination of the other $\vec{y}$ in Q. This contradicts that Q is linearly independent. ■

Corollary: If $B = \{\vec{x}_1, \ldots, \vec{x}_k\}$ is a basis for X, then $\dim(X) = k$.

Proof:
Since B is linearly independent, by definition $\dim(X) \geq k$. Meanwhile, since B spans X, we know by the above theorem that $\dim(X) \leq k$.
So $\dim(X) = k$.

Theorem 9.3: Suppose X is a vector space and $\dim(X) = n$. Then:

(A) For $E = \{\vec{x}_1, \ldots, \vec{x}_n\} \subset X$, we have that $X = \operatorname{span} E$ if and only if E is linearly independent.

Proof:
First, assume E is linearly independent. Then, note that for any $\vec{y} \in X$, we must have that $E \cup \{\vec{y}\}$ is linearly dependent because $|E \cup \{\vec{y}\}| > \dim(X)$. So, there exists $c_1, \ldots, c_n, c_{n+1} \in \mathbb{R}$ such that at least one c_i is nonzero and:

$$\sum_{i=1}^{n} c_i \vec{x}_i + c_{n+1} \vec{y} = \vec{0}$$

Now if $c_{n+1} = 0$, we have a contradiction because E is linearly independent. So, we conclude that $c_{n+1} \neq 0$. Then, by rearranging terms we can express y as a linear combination of the vectors of E. Therefore, $\operatorname{span} E = X$ since y can be any vector in X.

Secondly, assume E is not linearly independent. Then for some $\vec{x}_i \in E$, we have that $\operatorname{span} E = \operatorname{span}(E \setminus \{\vec{x}_i\})$. However, $|E \setminus \{\vec{x}_i\}| = n - 1$. So if $X = \operatorname{span} E$, then $\dim(X) \leq |E \setminus \{\vec{x}_i\}| = n - 1$, which contradicts our assumption that $\dim(X) = n$. Hence, $X \neq \operatorname{span} E$.

(B) X has a basis and every basis of X consists of n vectors.

Proof:
By the definition of $\dim(X)$, we know that there exists a linearly independent set of n vectors. By the previous part of this theorem, we also know that that set spans X. So, it is a basis of X. Meanwhile, by the corollary to theorem 9.2, we know that the number of vectors in a basis of X equals the dimension of X. Hence, all bases of X must have n vectors.

(C) If $1 \leq m \leq n$ and $\{\vec{y}_1, \ldots, \vec{y}_m\} \subset X$ is linearly independent, then X has a basis that contains $\vec{y}_1, \ldots, \vec{y}_m$.

Proof:
Let $S_0 = \{\vec{x}_1, \ldots, \vec{x}_n\}$ be a basis of X and $Q = \{\vec{y}_1, \ldots, \vec{y}_m\}$. Then by the same induction which we used to prove theorem 9.2, we can construct a basis: S_m, of X which contains $\vec{y}_1, \ldots, \vec{y}_m$.

Let X and Y be vector spaces. A map $\mathbf{A} : X \longrightarrow Y$ is linear if $\mathbf{A}(c_1 \vec{x}_1 + c_2 \vec{x}_2) = c_1 \mathbf{A}(\vec{x}_1) + c_2 \mathbf{A}(\vec{x}_2)$ for all $\vec{x}_1, \vec{x}_2 \in X$ and $c_1, c_2 \in \mathbb{R}$.

Observations:

1. A linear map sends $\vec{0}$ to $\vec{0}$. This is because:
$$\mathbf{A}(\vec{0}) = \mathbf{A}(\vec{v} - \vec{v}) = \mathbf{A}(\vec{v}) - \mathbf{A}(\vec{v}) = \vec{0}.$$
2. If $\mathbf{A} : X \longrightarrow Y$ is a linear map and $B = \{\vec{x}_1, \ldots, \vec{x}_k\}$ is a basis of X, then $\mathbf{A}\left(\sum_{i=1}^{k}(c_i \vec{x}_i)\right) = \sum_{i=1}^{k} c_i \mathbf{A}(\vec{x}_i)$ for all $c_1, \ldots, c_k \in \mathbb{R}$.

Given two vector spaces X and Y, we define $L(X, Y)$ to be the set of all linear mappings from X into Y. Also, we shall abbreviate $L(X, X)$ as $L(X)$.

$\mathscr{N}(\mathbf{A}) =$ "null space / kernel of $\mathbf{A}$" $= \{\vec{x} \in X \mid \mathbf{A}(\vec{x}) = \vec{0}\}$.

$\mathscr{R}(\mathbf{A}) =$ "range of $\mathbf{A}$" $= \{\vec{y} \in Y \mid \exists \vec{x} \in X \ \ s.t. \ \ \mathbf{A}\vec{x} = \vec{y}\}$.

Proposition: For any linear map $\mathbf{A} : X \longrightarrow Y$, $\mathscr{N}(\mathbf{A})$ and $\mathscr{R}(\mathbf{A})$ are vector spaces.

Proof:

- Assume $\vec{x}_1, \vec{x}_2 \in \mathscr{N}(\mathbf{A}) \subset X$ and $c \in \mathbb{R}$. Then:
  - $\mathbf{A}(\vec{x}_1 + \vec{x}_2) = \mathbf{A}(\vec{x}_1) + \mathbf{A}(\vec{x}_2) = \vec{0} + \vec{0} = \vec{0}$, which means that $\vec{x}_1 + \vec{x}_2 \in \mathscr{N}(\mathbf{A})$.
  - $\mathbf{A}(c\vec{x}_1) = c\mathbf{A}(\vec{x}_1) = c\vec{0} = \vec{0}$. So $c\vec{x}_1 \in \mathscr{N}(\mathbf{A})$.

  This shows that $\mathscr{N}(\mathbf{A})$ is a vector space.

- Assume $\vec{y}_1, \vec{y}_2 \in \mathscr{R}(\mathbf{A}) \subset Y$ and $c \in \mathbb{R}$. Then:
  - We know there exists $\vec{x}_1, \vec{x}_2 \in X$ such that $\mathbf{A}(\vec{x}_1) = \vec{y}_1$ and $\mathbf{A}(\vec{x}_2) = \vec{y}_2$. In turn, $\mathbf{A}(\vec{x}_1 + \vec{x}_2) = \mathbf{A}(\vec{x}_1) + \mathbf{A}(\vec{x}_2) = \vec{y}_1 + \vec{y}_2$. So $\vec{y}_1 + \vec{y}_2 \in \mathscr{R}(\mathbf{A})$.
  - Now continue letting $\vec{x}_1 \in X$ be a vector such that $\mathbf{A}(\vec{x}_1) = \vec{y}_1$. Then $\mathbf{A}(c\vec{x}_1) = c\mathbf{A}(\vec{x}_1) = c\vec{y}_1$. So $c\vec{y}_1 \in \mathscr{R}(\mathbf{A})$.

  This shows that $\mathscr{R}(\mathbf{A})$ is a vector space.

$\mathrm{rk}(\mathbf{A}) =$ "rank of $\mathbf{A}$" $= \dim(\mathscr{R}(\mathbf{A}))$.

$\mathrm{null}(\mathbf{A}) =$ "nullity of $\mathbf{A}$" $= \dim(\mathscr{N}(\mathbf{A}))$.

Rank-Nullity Theorem: Given any $\mathbf{A} \in L(X, Y)$, we have that
$$\dim(X) = \mathrm{rk}(\mathbf{A}) + \mathrm{null}(\mathbf{A}).$$

Proof:
Let $\dim(X) = n$.

$\mathscr{N}(\mathbf{A}) \subseteq X$ is a vector space. So pick a basis $\{\vec{v}_1, \ldots, \vec{v}_k\}$ for $\mathscr{N}(\mathbf{A})$ where $k = \mathrm{null}(\mathbf{A}) \leq \dim(X)$. Then by theorem 9.3, choose $\vec{w}_1, \ldots, \vec{w}_{m-k}$ such that $\{\vec{v}_1, \ldots, \vec{v}_k, \vec{w}_1, \ldots, \vec{w}_{n-k}\}$ is a basis of X. Note that $\dim(X) = n$.

Claim: $B = \{\mathbf{A}(\vec{w}_1), \ldots, \mathbf{A}(\vec{w}_{n-k})\}$ is a basis of $\mathscr{R}(\mathbf{A})$.

- $\mathbf{A}(\vec{v}_i) = \vec{0}$ for all $i \in \{1, \ldots, k\}$. So:

$$\begin{aligned}\mathscr{R}(\mathbf{A}) &= \operatorname{span}\{\mathbf{A}(\vec{v}_1), \ldots, \mathbf{A}(\vec{v}_k), \mathbf{A}(\vec{w}_1), \ldots, \mathbf{A}(\vec{w}_{n-k})\} \\ &= \operatorname{span}\{\mathbf{A}(\vec{w}_1), \ldots, \mathbf{A}(\vec{w}_{n-k})\} = \operatorname{span} B\end{aligned}$$

- B is linearly independent.
  To see this, note that: $\sum_{i=1}^{n-k} (c_i \mathbf{A}(\vec{w}_i)) = \vec{0} \implies \mathbf{A}\left(\sum_{i=1}^{n-k} c_i \vec{w}_i\right) = \vec{0}$

  Since we picked each $\vec{w}_1, \ldots, \vec{w}_{n-k} \in B$ so that they were not in $\mathscr{N}(A)$, we know that any vector in the span of B is not mapped to 0 by $\mathbf{A}$ unless it is the zero vector. So

$$\sum_{i=1}^{n-k} c_i \vec{w}_i = \vec{0}$$

  And since all the $\vec{w}_i$ are linearly independent, all constants c_i equal 0.

So $\operatorname{rk}(\mathbf{A}) = n - k = \dim(X) - \operatorname{null}(\mathbf{A})$.

Lecture 2: 4/4/2024

Proposition: Given $\mathbf{A} \in L(X, Y)$, then:

- $\mathbf{A}$ is injective if and only if $\operatorname{null}(\mathbf{A}) = 0$.

  Proof:
  ($\implies$) If $\mathbf{A}$ is injective, then since $\mathbf{A}(\vec{0}) = \vec{0}$, we have that any vector $\vec{v} \neq \vec{0}$ is not in $\mathscr{N}(\mathbf{A})$. So $\mathscr{N}(\mathbf{A}) = \{\vec{0}\}$, meaning $\operatorname{null}(\mathbf{A}) = 0$.

  ($\impliedby$) If $\operatorname{null}(\mathbf{A}) = 0$, then $\mathbf{A}(\vec{v}) = \vec{0} \implies \vec{v} = \vec{0}$. So now assume $\mathbf{A}(\vec{v}) = \mathbf{A}(\vec{u})$. Then $\mathbf{A}(\vec{v} - \vec{u}) = \vec{0}$, meaning $\vec{v} = \vec{u}$. Hence $\mathbf{A}$ is injective.

- $\mathbf{A}$ is surjective if and only if $\operatorname{rk}(\mathbf{A}) = \dim(Y)$.

  Proof:
  ($\implies$) If $\mathbf{A}$ is surjective then $\mathscr{R}(\mathbf{A}) = Y$. So we automatically have that $\operatorname{rk}(\mathbf{A}) = \dim(Y)$

  ($\impliedby$) If $\operatorname{rk}(\mathbf{A}) = \dim(Y)$, then there exists a linearly independent set of vectors $B \subset \mathscr{R}(\mathbf{A})$ containing $\dim(Y)$ many vectors and spanning $\mathscr{R}(\mathbf{A})$. Then by theorem 9.3, since $B \subset \mathscr{R}(\mathbf{A}) \subseteq Y$, we know $\operatorname{span} B = Y$. So, $\mathscr{R}(\mathbf{A}) = Y$, meaning $\mathbf{A}$ is surjective.

Corollary: Let $\mathbf{A} \in L(X)$. Then $\mathbf{A}$ is bijective if and only if $\text{null}(\mathbf{A}) = 0$.

Proof: (let $\mathbf{A} : X \longrightarrow X$ be a linear map)

($\Longrightarrow$) If $\mathbf{A}$ is bijective, then automatically $\mathbf{A}$ is injective. So $\text{null}(\mathbf{A}) = 0$ by the previous proposition.

($\Longleftarrow$) If $\text{null}(\mathbf{A}) = 0$, then by the rank-nullity theorem, we know that $\text{rk}(\mathbf{A}) = \dim(X)$. Thus $\mathbf{A}$ is both injective and surjective, meaning $\mathbf{A}$ is bijective.

For $\mathbf{A} \in L(X)$, when $\text{null}(\mathbf{A}) = 0$, we call $\mathbf{A}$ invertible and define $\mathbf{A}^{-1} : X \longrightarrow X$ such that $\mathbf{A}^{-1}(\mathbf{A}(\vec{x})) = \vec{x}$ for all $\vec{x} \in X$.

Because $\mathbf{A}$ must be a bijective set function, we know that $\mathbf{A}^{-1}$ must also be a right-inverse of $\mathbf{A}$, meaning $\mathbf{A}(\mathbf{A}^{-1}(\vec{x})) = \vec{x}$.

Additionally, consider any $\vec{x}_1, \vec{x}_2 \in X$ and let $\vec{x}_1' = \mathbf{A}^{-1}(\vec{x}_1)$ and $\vec{x}_2' = \mathbf{A}^{-1}(\vec{x}_2)$. Then since $\mathbf{A}$ is a linear mapping, we know that for any $c_1, c_2 \in \mathbb{R}$:

$$\mathbf{A}(c_1\vec{x}_1' + c_2\vec{x}_2') = c_1\mathbf{A}(\mathbf{A}^{-1}(\vec{x}_1)) + c_2\mathbf{A}(\mathbf{A}^{-1}(\vec{x}_2)) = c_1\vec{x}_1 + c_2\vec{x}_2$$

So: $\mathbf{A}^{-1}(c_1\vec{x}_1 + c_2\vec{x}_2) = c_1\vec{x}_1' + c_2\vec{x}_2' = c_1\mathbf{A}^{-1}(\vec{x}_1) + c_2\mathbf{A}^{-1}(\vec{x}_2)$. Hence, we've shown that $\mathbf{A}^{-1}$ is a linear mapping, meaning that $\mathbf{A}^{-1} \in L(X)$.

Let $\mathbf{A} \in L(X, Y)$ and $\mathbf{B} \in L(Y, Z)$. Then we define $\mathbf{BA} : X \longrightarrow Z$ by the rule that $\vec{x} \mapsto \mathbf{B}(\mathbf{A}(\vec{x}))$.

We can trivially show that $\mathbf{BA}$ is a linear mapping. Consider any $\vec{x}_1, \vec{x}_2 \in X$ and $c_1, c_2 \in \mathbb{R}$. Then:

$$\begin{aligned}\mathbf{BA}(c_1\vec{x}_1 + c_2\vec{x}_2) &= \mathbf{B}(c_1\mathbf{A}(\vec{x}_1) + c_2\mathbf{A}(\vec{x}_2)) \\ &= c_1\mathbf{B}(\mathbf{A}(\vec{x}_1)) + c_2\mathbf{B}(\mathbf{A}(\vec{x}_2)) \\ &= c_1\mathbf{BA}(\vec{x}_1) + c_2\mathbf{BA}(\vec{x}_2)\end{aligned}$$

This means that $\mathbf{BA} \in L(X, Z)$.

Let $\mathbf{A}, \mathbf{B} \in L(X, Y)$ and $c_1, c_2 \in \mathbb{R}$. Then we define $(c_1\mathbf{A} + c_2\mathbf{B}) : X \longrightarrow Y$ by the rule: $\vec{x} \mapsto c_1\mathbf{A}(\vec{x}) + c_2\mathbf{B}(\vec{x})$.

It is even more trivial to show that $(c_1\mathbf{A} + c_2\mathbf{B})$ is a linear map.

Let $\mathbf{A} \in L(\mathbb{R}^n, \mathbb{R}^m)$. We define the norm of $\mathbf{A}$ as:

$$\|\mathbf{A}\| = \sup\{\|\mathbf{A}(\vec{x})\| \mid \vec{x} \in \mathbb{R}^n \text{ and } \|\vec{x}\| \leq 1\}.$$

Throughout this section, we shall prove that $\|\cdot\| : L(\mathbb{R}^n, \mathbb{R}^m) \longrightarrow \mathbb{R}$ is well-defined and fulfills the properties of a general norm function.

Proposition: If $\mathbf{A} \in L(\mathbb{R}^n, \mathbb{R}^m)$, then $\|\mathbf{A}\|$ exists and is finite.

Proof:
Let $\{e_1, \ldots, e_n\}$ be the standard basis in $\mathbb{R}^n$. Then for any $\vec{x} \in \mathbb{R}^n$, there are unique $c_1, \ldots, c_n \in \mathbb{R}$ such that $\vec{x} = c_1e_1 + \ldots + c_ne_n$.

Since we are working with the standard basis, we know: $\|\vec{x}\| = \sqrt{\sum_{i=1}^{n} c_i^2}$.

Thus, for $\|\vec{x}\| \leq 1$, we must have that $|c_i| \leq 1$ for each c_i. This means:

$$\|\mathbf{A}(\vec{x})\| = \left\|\sum_{i=1}^{n} c_i\mathbf{A}(e_i)\right\| \leq \sum_{i=1}^{n} \|c_i\mathbf{A}(e_i)\| = \sum_{i=1}^{n} |c_i|\|\mathbf{A}(e_i)\| \leq \sum_{i=1}^{n} \|\mathbf{A}(e_i)\|$$

Importantly, we must have that $\sum_{i=1}^{n} \|\mathbf{A}(e_i)\|$ is finite. Additionally, it is an upper bound to the set: $\{\|\mathbf{A}(\vec{x})\| \mid \vec{x} \in \mathbb{R}^n \text{ and } \|\vec{x}\| \leq 1\} \subseteq \mathbb{R}$.

So, we showed that the above set is bounded above. Also, the above set is nonempty because it must contain $\|\vec{0}\| = 0$. Thus by the least upper bound property of $\mathbb{R}$, we know that the supremum of this set exists in $\mathbb{R}$.

Hence, $\|\mathbf{A}\|$ exists and is finite.

Side note, the above proof also shows that $\|\mathbf{A}\| \geq 0$.

Lemma: For $\mathbf{A} \in L(\mathbb{R}^n, \mathbb{R}^m)$ and $\vec{x} \in \mathbb{R}^n$, we have that $\|\mathbf{A}(\vec{x})\| \leq \|\mathbf{A}\|\|\vec{x}\|$.

Proof:
Case 1: $\vec{x} \neq \vec{0}$.
Then since $\|\vec{x}\| \neq 0$, we can say that:

$$\|\mathbf{A}(\vec{x})\| = \left\|\mathbf{A}\left(\|\vec{x}\|\frac{\vec{x}}{\|\vec{x}\|}\right)\right\| = \left\|\|\vec{x}\|\mathbf{A}\left(\frac{\vec{x}}{\|\vec{x}\|}\right)\right\| = \left\|\mathbf{A}\left(\frac{\vec{x}}{\|\vec{x}\|}\right)\right\|\|\vec{x}\|$$

Now $\frac{\vec{x}}{\|\vec{x}\|} \in \mathbb{R}^n$ and $\left\|\frac{\vec{x}}{\|\vec{x}\|}\right\| = 1$. So, $\left\|\mathbf{A}\left(\frac{\vec{x}}{\|\vec{x}\|}\right)\right\|\|\vec{x}\| \leq \|\mathbf{A}\|\|\vec{x}\|$

Case 2: $\vec{x} = \vec{0}$.
Then trivially $\|\mathbf{A}(\vec{x})\| = \|\mathbf{A}(\vec{0})\| = 0 = \|\mathbf{A}\|\|\vec{0}\| = \|\mathbf{A}\|\|\vec{x}\|$

Proposition: If $\mathbf{A} \in L(\mathbb{R}^n, \mathbb{R}^m)$, then $0 \leq \|\mathbf{A}\|$. Also $\|\mathbf{A}\| = 0$ if and only if $\mathbf{A}$ is the unique function mapping all of $\mathbb{R}^n$ to $\vec{0}$.

Proof:
We already showed previously that $\|\mathbf{A}\| \geq 0$. So, it now suffices to show that $\|\mathbf{A}\| = 0 \iff \mathscr{N}(\mathbf{A}) = \mathbb{R}^n$.

($\Longrightarrow$) Assume that $\mathscr{N}(\mathbf{A}) \neq \mathbb{R}^n$. Then there exists $\vec{x} \in \mathbb{R}^n$ such that $\mathbf{A}(\vec{x}) \neq \vec{0}$. Since $\vec{x}$ can't be $\vec{0}$, consider the vector $\hat{x} = \frac{\vec{x}}{\|\vec{x}\|}$. By the linearity of $\mathbf{A}$, we know $\mathbf{A}(\hat{x}) = \frac{1}{\|\vec{x}\|}\mathbf{A}(\vec{x}) \neq \vec{0}$. So, $\|\mathbf{A}(\hat{x})\| > 0$. But $\|\mathbf{A}(\hat{x})\|$ is in the set that $\|\mathbf{A}\|$ is a supremum of, which means that $\|\mathbf{A}\| \geq \|\mathbf{A}(\hat{x})\| > 0$. Or in other words, $\|\mathbf{A}\| \neq 0$.

($\Longleftarrow$) Assume that $\mathscr{N}(\mathbf{A}) = \mathbb{R}^n$. Then,
$$\sup\{\|\mathbf{A}(\vec{x})\| \mid \vec{x} \in \mathbb{R}^n \text{ and } \|\vec{x}\| \leq 1\} = \sup\{0\} = 0$$

Corollary: Given $\mathbf{A} \in L(\mathbb{R}^n, \mathbb{R}^m)$, we have that $\mathbf{A}$ is uniformly continuous.

Proof:
Case 1: $\|\mathbf{A}\| \neq 0$, meaning we can divide by $\|\mathbf{A}\|$.
By the previous proposition, $\|\mathbf{A}(\vec{x}) - \mathbf{A}(\vec{y})\| \leq \|\mathbf{A}\|\|\vec{x} - \vec{y}\|$ for all $\vec{x}, \vec{y} \in \mathbb{R}^n$. Hence, for any $\varepsilon > 0$, if we make $\|\vec{x} - \vec{y}\| < \frac{\varepsilon}{\|\mathbf{A}\|}$, then $\|\mathbf{A}(\vec{x}) - \mathbf{A}(\vec{y})\| < \varepsilon$.

Case 2: $\|\mathbf{A}\| = 0$.
Then $\mathbf{A}$ is a constant function, making it automatically uniformly continuous.

Subcorollary: Given $\mathbf{A} \in L(\mathbb{R}^n, \mathbb{R}^m)$, there exists $\vec{x} \in \mathbb{R}^n$ with $\|\vec{x}\| \leq 1$ such that $\|\mathbf{A}(\vec{x})\| = \|\mathbf{A}\|$.

Proof:
Let $S = \{\vec{x} \in \mathbb{R}^n \mid \|\vec{x}\| \leq 1\}$ and consider the restriction $\mathbf{A}|_S$.

Since S is a closed and bounded subset of $\mathbb{R}^n$, we know that S is compact by the Heine-Borel theorem (see proposition 28 in Math 140A notes). This combined with the fact that $\mathbf{A}|_S$ is still continuous means that by the extreme value theorem, there is $\vec{x} \in S$ with:
$$\mathbf{A}(\vec{x}) = \mathbf{A}|_S(\vec{x}) = \sup\{\|\mathbf{A}(\vec{x})\| \mid \vec{x} \in \mathbb{R}^n \text{ and } \|\vec{x}\| \leq 1\}.$$

Proposition: If $\mathbf{A}, \mathbf{B} \in L(\mathbb{R}^n, \mathbb{R}^m)$, then $\|\mathbf{A} + \mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|$.

Proof:
Let $\vec{x} \in \mathbb{R}^n$ be a vector such that $\|\vec{x}\| \leq 1$ and $\|\mathbf{A}(\vec{x})\| = \|\mathbf{A}\|$. Then:
$$\begin{aligned}\|\mathbf{A} + \mathbf{B}\| = \|(\mathbf{A} + \mathbf{B})(\vec{x})\| &= \|\mathbf{A}(\vec{x}) + \mathbf{B}(\vec{x})\| \\ &\leq \|\mathbf{A}(\vec{x})\| + \|\mathbf{B}(\vec{x})\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|\end{aligned}$$

Proposition: If $\mathbf{A} \in L(\mathbb{R}^n, \mathbb{R}^m)$ and $c \in \mathbb{R}$, then $\|c\mathbf{A}\| = |c|\|\mathbf{A}\|$.

Proof:
Pick $\vec{x} \in \mathbb{R}^n$ satisfying $\|\vec{x}\| \leq 1$ and $\|\mathbf{A}(\vec{x})\| = \|\mathbf{A}\|$. Then:
$$|c|\|\mathbf{A}\| = |c|\|\mathbf{A}(\vec{x})\| = \|c\mathbf{A}(\vec{x})\| = \|(c\mathbf{A})(\vec{x})\| \leq \|c\mathbf{A}\|.$$

Next, pick $\vec{y} \in \mathbb{R}^n$ satisfying $\|\vec{y}\| \leq 1$ and $\|(c\mathbf{A})(\vec{x})\| = \|c\mathbf{A}\|$. Then:
$$\|c\mathbf{A}\| = \|(c\mathbf{A})(\vec{y})\| = \|c\mathbf{A}(\vec{y})\| = |c|\|\mathbf{A}\,\vec{y}\| \leq |c|\|\mathbf{A}\|.$$

Specifically because of the four propositions above, we have shown that $\|\cdot\| : L(\mathbb{R}^n, \mathbb{R}^m) \longrightarrow \mathbb{R}$ is well-defined and a valid norm. Consequently, by defining $d(\mathbf{A}, \mathbf{B}) = \|\mathbf{A} - \mathbf{B}\|$ for all $\mathbf{A}, \mathbf{B} \in L(\mathbb{R}^n, \mathbb{R}^m)$, we naturally get that $L(\mathbb{R}^n, \mathbb{R}^m)$ is a metric space.

Given any $\mathbf{A}, \mathbf{B}, \mathbf{C} \in L(\mathbb{R}^n, \mathbb{R}^m)$, we have:

- $d(\mathbf{A}, \mathbf{B}) = \|\mathbf{A} - \mathbf{B}\| \geq 0$ with $d(\mathbf{A}, \mathbf{B}) = 0$ if and only if $\mathbf{A} = \mathbf{B}$.
- $d(\mathbf{A}, \mathbf{B}) = \|\mathbf{A} - \mathbf{B}\| = |-1|\|\mathbf{B} - \mathbf{A}\| = d(\mathbf{B}, \mathbf{A})$
- $d(\mathbf{A}, \mathbf{C}) = \|\mathbf{A} - \mathbf{C}\| \leq \|\mathbf{A} - \mathbf{B}\| + \|\mathbf{B} - \mathbf{C}\| = d(\mathbf{A}, \mathbf{B}) + d(\mathbf{B}, \mathbf{C})$

Before moving on, here is another corollary of the above statements.

Corollary: If $\mathbf{A} \in L(\mathbb{R}^n, \mathbb{R}^m)$ and $\mathbf{B} \in L(\mathbb{R}^m, \mathbb{R}^k)$, then $\|\mathbf{BA}\| \leq \|\mathbf{B}\|\|\mathbf{A}\|$.

Proof:
Pick $\vec{x} \in \mathbb{R}^n$ satisfying $\|\vec{x}\| \leq 1$ and $\|(\mathbf{BA})(\vec{x})\| = \|\mathbf{BA}\|$. Then:
$$\|\mathbf{BA}\| = \|(\mathbf{BA})(\vec{x})\| = \|\mathbf{B}(\mathbf{A}(\vec{x}))\| \leq \|\mathbf{B}\|\|\mathbf{A}(\vec{x})\| \leq \|\mathbf{B}\|\|\mathbf{A}\|.$$

Theorem 9.8: Let $\Omega \subset L(\mathbb{R}^n)$ be the set of all invertible linear mappings on $\mathbb{R}^n$.

(A) If $\mathbf{A} \in \Omega$, $\mathbf{B} \in L(\mathbb{R}^n)$, and $\|\mathbf{B} - \mathbf{A}\| < \dfrac{1}{\|\mathbf{A}^{-1}\|}$, then $\mathbf{B} \in \Omega$.

Proof:
Pick $\vec{x} \in \mathbb{R}^n$ such that $\|\vec{x}\| \leq 1$. Then:
$$\begin{aligned}\|\mathbf{A}(\vec{x})\| &= \|(\mathbf{A} - \mathbf{B} + \mathbf{B})(\vec{x})\| \\ &\leq \|(\mathbf{A} - \mathbf{B})(\vec{x})\| + \|\mathbf{B}(\vec{x})\| \\ &\leq \|\mathbf{A} - \mathbf{B}\|\|\vec{x}\| + \|\mathbf{B}(\vec{x})\| = \|\mathbf{B} - \mathbf{A}\|\|\vec{x}\| + \|\mathbf{B}(\vec{x})\|\end{aligned}$$

Meanwhile, note that $\|\mathbf{A}^{-1}\| \neq 0$. We know this because $\mathbf{A}^{-1}$ must be invertible (because $\mathscr{N}(\mathbf{A}^{-1}) = \{\vec{0}\}$) and the one linear mapping in $L(\mathbb{R}^n)$ with norm 0 is not invertible. So:
$$\frac{\|\vec{x}\|}{\|\mathbf{A}^{-1}\|} = \frac{\|\mathbf{A}^{-1}\mathbf{A}(\vec{x})\|}{\|\mathbf{A}^{-1}\|} \leq \frac{\|\mathbf{A}^{-1}\|\|\mathbf{A}(\vec{x})\|}{\|\mathbf{A}^{-1}\|} = \|\mathbf{A}(\vec{x})\|$$

Hence, $\frac{\|\vec{x}\|}{\|\mathbf{A}^{-1}\|} \leq \|\mathbf{B}-\mathbf{A}\|\|\vec{x}\| + \|\mathbf{B}(\vec{x})\|$. By rearranging terms, we get this expression: $\left(\frac{1}{\|\mathbf{A}^{-1}\|} - \|\mathbf{B}-\mathbf{A}\|\right)\|\vec{x}\| \leq \|\mathbf{B}(\vec{x})\|$.

Now, note that if $\|\mathbf{B}(\vec{x})\| = 0$ but $\vec{x} \neq \vec{0}$, then we must have that: $\frac{1}{\|\mathbf{A}^{-1}\|} - \|\mathbf{B}-\mathbf{A}\| \leq 0$. Or in other words, $\|\mathbf{B}-\mathbf{A}\| \geq \frac{1}{\|\mathbf{A}^{-1}\|}$. So, if $\|\mathbf{B}-\mathbf{A}\| < \frac{1}{\|\mathbf{A}^{-1}\|}$, then $\|\mathbf{B}(\vec{x})\| = 0$ only when $\vec{x} = \vec{0}$.
Hence, $\text{null}(\mathbf{B}) = 0$ and $\mathbf{B}$ is invertible.

(B) Ω is an open subset of $L(\mathbb{R}^n)$, and the mapping over Ω with the rule: $\mathbf{A} \mapsto \mathbf{A}^{-1}$, is continuous.

Proof:
Firstly, by part A we know that for any $\mathbf{A} \in \Omega$, if $r = \frac{1}{\|\mathbf{A}^{-1}\|}$, then $B_r(\mathbf{A}) \subseteq \Omega$. So, Ω is an open set in the metric space $L(\mathbb{R}^n)$.

Now let $\mathbf{A}, \mathbf{B} \in \Omega$ and recall from part A that:

$$\left(\frac{1}{\|\mathbf{A}^{-1}\|} - \|\mathbf{B}-\mathbf{A}\|\right)\|\vec{x}\| \leq \|\mathbf{B}(\vec{x})\|.$$

Since we know $\mathbf{B}^{-1}$ exists, set $\vec{x} = \mathbf{B}^{-1}(\vec{y})$. Then the above expression becomes: $\left(\frac{1}{\|\mathbf{A}^{-1}\|} - \|\mathbf{B}-\mathbf{A}\|\right)\|\mathbf{B}^{-1}(\vec{y})\| \leq \|\vec{y}\|$.
Because we are interested in $\mathbf{B}$ close to $\mathbf{A}$, we can assume that $\|\mathbf{B}-\mathbf{A}\| < \frac{1}{\|\mathbf{A}^{-1}\|}$. Thus it is safe to divide by $\frac{1}{\|\mathbf{A}^{-1}\|} - \|\mathbf{B}-\mathbf{A}\|$.
So, setting $\vec{y} \in \mathbb{R}^n$ to be the vector satisfying $\|\vec{y}\| \leq 1$ and $\|\mathbf{B}^{-1}(\vec{y})\| = \|\mathbf{B}^{-1}\|$, we have that:

$$\|\mathbf{B}^{-1}\| = \|\mathbf{B}^{-1}(\vec{y})\| \leq \frac{\|\vec{y}\|}{\frac{1}{\|\mathbf{A}^{-1}\|}-\|\mathbf{B}-\mathbf{A}\|} \leq \frac{1}{\frac{1}{\|\mathbf{A}^{-1}\|}-\|\mathbf{B}-\mathbf{A}\|} = \frac{\|\mathbf{A}^{-1}\|}{1-\|\mathbf{A}^{-1}\|\|\mathbf{B}-\mathbf{A}\|}$$

Lemma: Given $\mathbf{A} \in L(Z,W)$, $\mathbf{B}, \mathbf{C} \in L(Y,Z)$, and $\mathbf{D} \in L(X,Y)$, we have that $\mathbf{A}(\mathbf{B}+\mathbf{C}) = \mathbf{AB}+\mathbf{AC}$ and $(\mathbf{B}+\mathbf{C})\mathbf{D} = \mathbf{BD}+\mathbf{CD}$.

Proof:
- $\mathbf{A}((\mathbf{B}+\mathbf{C})(\vec{v})) = \mathbf{A}(\mathbf{B}(\vec{v})+\mathbf{C}(\vec{v})) = \mathbf{A}(\mathbf{B}(\vec{v})) + \mathbf{A}(\mathbf{C}(\vec{v}))$
- $(\mathbf{B}+\mathbf{C})(\mathbf{D}(\vec{v})) = \mathbf{B}(\mathbf{D}(\vec{v})) + \mathbf{C}(\mathbf{D}(\vec{v}))$

Based on the above lemma, we have that $\mathbf{B}^{-1}-\mathbf{A}^{-1} = \mathbf{B}^{-1}(\mathbf{A}-\mathbf{B})\mathbf{A}^{-1}$.
So:

$$0 \leq \|\mathbf{B}^{-1}-\mathbf{A}^{-1}\| = \|\mathbf{B}^{-1}(\mathbf{A}-\mathbf{B})\mathbf{A}^{-1}\|$$
$$\leq \|\mathbf{B}^{-1}\|\|(\mathbf{A}-\mathbf{B})\|\|\mathbf{A}^{-1}\| \leq \frac{\|\mathbf{A}^{-1}\|^2}{1-\|\mathbf{A}^{-1}\|\|\mathbf{B}-\mathbf{A}\|}\|\mathbf{B}-\mathbf{A}\|$$

Finally, assume $\mathbf{A} \in \Omega'$. This is fine because the mapping is automatically continuous at $\mathbf{A}$ if $\mathbf{A} \notin \Omega'$. Then we have that:

$$\lim_{\mathbf{B}\to\mathbf{A}} \left(\frac{\|\mathbf{A}^{-1}\|^2}{1-\|\mathbf{A}^{-1}\|\|\mathbf{B}-\mathbf{A}\|} \|\mathbf{B}-\mathbf{A}\| \right) = \|\mathbf{A}^{-1}\|^2 \cdot 0 = 0.$$

So, $0 \leq \lim_{\mathbf{B}\to\mathbf{A}} (\|\mathbf{B}^{-1} - \mathbf{A}^{-1}\|) \leq 0.$

This means that $d(\mathbf{B}^{-1},\ \mathbf{A}^{-1}) = \|\mathbf{B}^{-1} - \mathbf{A}^{-1}\| \to 0$ as $\mathbf{B} \to \mathbf{A}$. Or in other words:

$$\lim_{\mathbf{B}\to\mathbf{A}} (\mathbf{B}^{-1}) = \mathbf{A}^{-1}. \blacksquare$$

Lecture 3: 4/9/2024

Let X and Y be vector spaces and fix two bases $\{\vec{x}_1, \ldots, \vec{x}_n\}$ and $\{\vec{y}_1, \ldots, \vec{y}_m\}$ of X and Y respectively. Then given any $\mathbf{A} \in L(X, Y)$, since $\mathbf{A}(\vec{x}_j) \in Y$ for each $j \in \{1, \ldots, n\}$, we have that there are unique scalars $a_{i,j}$ such that:

$$\mathbf{A}(\vec{x}_j) = \sum_{i=1}^{m} a_{i,j}\, \vec{y}_i$$

For convenience, we can visualize these numbers in an $m \times n$ matrix:

$$[\mathbf{A}] = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix}$$

Note that for each $j \in \{1, \ldots, n\}$, we have that the jth column of $[\mathbf{A}]$ gives the coordinates of $\mathbf{A}(\vec{x}_j)$ with respect to the basis $\{\vec{y}_1, \ldots, \vec{y}_m\}$. Thus, we call the vectors $\mathbf{A}(\vec{x}_j)$ the column vectors of $[\mathbf{A}]$.

Fact 1: Given any $\vec{x} \in X$, there are unique scalars $c_1, \ldots, c_n$ such that $\vec{x} = \sum_{j=1}^{n} c_j\, \vec{x}_j$. Then, the coordinates of $\mathbf{A}(\vec{x})$ with respect to our basis of Y is given by the commonly defined matrix-vector product:

$$\begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}$$

This is true because $\mathbf{A}$ is linear. Therefore:

$$\mathbf{A}(\vec{x}) = \mathbf{A}\left(\sum_{j=1}^{n} c_j \vec{x}_j\right) = \sum_{j=1}^{n} c_j \mathbf{A}(\vec{x}_j)$$
$$= \sum_{j=1}^{n} c_j \left(\sum_{i=1}^{m} a_{i,j}\, \vec{y}_i\right) = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} c_j a_{i,j}\right) \vec{y}_i$$

Fact 2: When we said how to generate an $m \times n$ matrix $[\mathbf{A}]$ for any $A \in L(X, Y)$, we were implicitely creating a mapping $\phi : L(X, Y) \longrightarrow \mathcal{M}_{m\times n}(\mathbb{R})$ (the set of $m \times n$ real matrices). Importantly, this map is invertible.

Let us define a mapping $\varphi : \mathcal{M}_{m\times n}(\mathbb{R}) \longrightarrow L(X, Y)$ such that for any $[\mathbf{B}] \in \mathcal{M}_{m\times n}(\mathbb{R})$ where

$$[\mathbf{B}] = \begin{bmatrix} b_{1,1} & b_{1,2} & \cdots & b_{1,n} \\ b_{2,1} & b_{2,2} & \cdots & b_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m,1} & b_{m,2} & \cdots & b_{m,n} \end{bmatrix}$$

...we define $\varphi([\mathbf{B}]) \in L(X, Y)$ by $\varphi([\mathbf{B}])(\vec{x}) = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} c_j b_{i,j}\right) \vec{y}_i$ where $c_1, \ldots, c_n$ are the coefficients such that $\vec{x} = \sum_{j=1}^{n} c_j \vec{x}_j$.

Firstly, $\varphi([\mathbf{B}])$ is well defined because $c_1, \ldots, c_n$ are unique with respect to our basis of X. Also, $\varphi([\mathbf{B}])$ is linear because sums are linear.

Meanwhile, by fact 1 we know that for any $\mathbf{A} \in L(X, Y)$, $\mathbf{A} = \varphi(\phi(\mathbf{A}))$. At the same time, you can easily check that for any $[\mathbf{B}] \in \mathcal{M}_{m\times n}(\mathbb{R})$, $[\mathbf{B}] = \phi(\varphi([\mathbf{B}]))$. Hence, $\varphi = \phi^{-1}$.

Thus, from now on we shall say that the linear map $\mathbf{A}$ and matrix $[\mathbf{A}]$ are associated with each other if $\phi(\mathbf{A}) = [\mathbf{A}]$ and $\varphi([\mathbf{A}]) = \mathbf{A}$.

Fact 3: In addition to our bases for X and Y, fix $\{\vec{z}_1, \ldots, \vec{z}_p\}$ as our basis for Z. Then, given the linear maps $\mathbf{A} \in L(X, Y)$ and $\mathbf{B} \in L(Y, Z)$ and their associated matrices $[\mathbf{A}] \in \mathcal{M}_{m\times n}(\mathbb{R})$ and $[\mathbf{B}] \in \mathcal{M}_{p\times m}(\mathbb{R})$, we have that the map $\mathbf{BA}$ is associated with the matrix $[\mathbf{B}][\mathbf{A}]$.

Let us use $a_{i,j}$ and $b_{k,i}$ to refer to the entries of $[\mathbf{A}]$ and $[\mathbf{B}]$ respectively. Then note that:

$$\mathbf{BA}(\vec{x_j}) = \mathbf{B}(\mathbf{A}(\vec{x_j})) = \mathbf{B}\left(\sum_{i=1}^{m} a_{i,j}\, \vec{y}_i\right) = \sum_{i=1}^{m} a_{i,j} \mathbf{B}(\vec{y}_i)$$
$$= \sum_{i=1}^{m} a_{i,j} \left(\sum_{k=1}^{p} b_{k,i}\, \vec{z}_k\right) = \sum_{k=1}^{p} \left(\sum_{i=1}^{m} (a_{i,j} b_{k,i})\right) \vec{z}_k$$

So, the (k,j)th. entry of the matrix associated with $\mathbf{BA}$ is $\sum_{i=1}^{m} b_{k,i}a_{i,j}$.

Hence, the matrix associated with the map $\mathbf{BA}$ is precisely the matrix product $[\mathbf{B}][\mathbf{A}]$.

Fact 4: Suppose that $\mathbf{A}$ and $\mathbf{B}$ are linear maps in $L(X,Y)$ and that c_1 and c_2 are scalars. Then the matrix $c_1[\mathbf{A}] + c_2[\mathbf{B}]$ is associated with the linear map $c_1\mathbf{A} + c_2\mathbf{B}$.

This is rather trivial to prove compared to the other facts. So, since I'm really behind, I'm just not going to prove it here. Frick you <3.

Now from a rigor point of view, we'd rather work with linear maps than matrices. This is because the defintion of a matrix depends on what bases we fix, whereas linear maps are defined independently of any bases. That said, matrices are too convenient to not be discussed.

Going foward, here are three notational things from linear we shall adopt when talking about linear maps:

1. We shall abbreviate $\mathbf{A}(\vec{x})$ as $\mathbf{A}\vec{x}$.
2. We shall denote $\mathbf{0} \in L(X,Y)$ as the linear map with $\mathcal{N}(\mathbf{0}) = X$. After all, $[\mathbf{0}]$ is the zero matrix.
3. We shall denote $\mathbf{I} \in L(X)$ as the identity map on X. After all, $[\mathbf{I}]$ is the identity matrix.

Since an $m \times n$ matrix can be thought of as a list of $m \cdot n$ numbers, the "natural" norm to equip $\mathcal{M}_{m\times n}(\mathbb{R})$ with is:

$$\|[\mathbf{A}]\|_F = \left(\sum_{i=1}^{m}\sum_{j=1}^{n}(a_{i,j})^2\right)^{\frac{1}{2}}$$

Note on my notation:

Since I view $|\cdot|$ as having already been reserved for the absolute value function, I am not going to use the same notation as Rudin and my professor use for this matrix norm. Rather, because this norm is also called the Frobenius norm, I shall denote it by $\|\cdot\|_F$.

Also, this is a valid norm for the same reasons that the vector Euclidean norm is a valid norm.

If we define $d([\mathbf{B}],[\mathbf{A}]) = \|[\mathbf{B}]-[\mathbf{A}]\|_F$, then we can treat $\mathcal{M}_{m\times n}(\mathbb{R})$ as a metric space with the metric d.

Proposition: Using the standard bases for $\mathbb{R}^n$ and $\mathbb{R}^m$, we have that for any associated linear map $\mathbf{A} \in L(\mathbb{R}^n,\mathbb{R}^m)$ and matrix $[\mathbf{A}] \in \mathcal{M}_{m\times n}(\mathbb{R})$ with coefficients $a_{i,j}$ for $1 \le i \le m$ and $1 \le j \le n$:

$$\|\mathbf{A}\| \le \|[\mathbf{A}]\|_F$$

Proof:
Let $\vec{x} = (x_1,\ldots,x_n) \in \mathbb{R}^n$. Then by the Cauchy-Schwarz inequality:

$$\|\mathbf{A}\,\vec{x}\|^2 = \sum_{i=1}^{m}\left(\sum_{j=1}^{n} a_{i,j}x_j\right)^2 \le \sum_{i=1}^{m}\left(\sum_{j=1}^{n} a_{i,j}^2 \cdot \sum_{j=1}^{n} x_j^2\right) = \|\vec{x}\|^2 \cdot \sum_{i=1}^{m}\sum_{j=1}^{n} a_{i,j}^2$$

So $\|\mathbf{A}\,\vec{x}\|^2 \le \|\vec{x}\|^2 \cdot \|[\mathbf{A}]\|_F^2$. Or in other words, $\|\mathbf{A}\|^2 \le 1 \cdot \|[\mathbf{A}]\|_F^2$.

Corollary 1: Using the standard bases for $\mathbb{R}^n$ and $\mathbb{R}^m$, consider any matrix $[\mathbf{A}] \in \mathcal{M}_{m\times n}(\mathbb{R})$. Then the mapping $[\mathbf{A}] \mapsto \mathbf{A}$ is continuous.

Proof:
Pick any matrices $[\mathbf{A}],[\mathbf{B}] \in \mathcal{M}_{n\times n}(\mathbb{R})$ and let $\varepsilon > 0$. Then if $\|[\mathbf{B}]-[\mathbf{A}]\|_F < \varepsilon$, we have that $\|\mathbf{B}-\mathbf{A}\| \le \|[\mathbf{B}]-[\mathbf{A}]\|_F < \varepsilon$.

Corollary 2: Suppose that S is a matric space, that $a_{1,1},\ldots,a_{m,n}$ are real continuous functions on S, and that for each $p \in S$, $\mathbf{A}_p$ is the linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$ whose associated matrix is:

$$[\mathbf{A}_p] = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix}$$

Then the mapping $p \mapsto \mathbf{A}_p$ is a continuous mapping of S into $L(\mathbb{R}^n,\mathbb{R}^m)$.

Proof:
Firstly, the mapping $p \mapsto [\mathbf{A}_p]$ is continuous for the same reason that a vector valued function is continuous if and only if its component functions are all continuous. Additionally, the mapping $[\mathbf{B}] \mapsto \mathbf{B}$ is continuous. Thus, because the composition of two continuous functions is itself continuous, we have that the mapping $p \mapsto \mathbf{A}_p$ is continuous.

To finish this lecture, here is one last helpful fact:

Lemma: For any associated map $\mathbf{A} \in L(X)$ and matrix $[\mathbf{A}] \in \mathcal{M}_{n\times n}(\mathbb{R})$, we have that $\mathbf{A}$ is invertible if and only if $[\mathbf{A}]$ is invertible.

Proof:
If $\mathbf{B}$ exists such that $\mathbf{BA} = \mathbf{I}$ and $\mathbf{AB} = \mathbf{I}$, then we have that $[\mathbf{B}][\mathbf{A}] = [\mathbf{I}]$ and $[\mathbf{A}][\mathbf{B}] = [\mathbf{I}]$. So $[\mathbf{B}] = [\mathbf{A}]^{-1}$.

Similarly, if $[\mathbf{B}]$ exists such that $[\mathbf{B}][\mathbf{A}] = [\mathbf{I}]$ and $[\mathbf{A}][\mathbf{B}] = [\mathbf{I}]$, then we have that $\mathbf{BA} = \mathbf{I}$ and $\mathbf{AB} = \mathbf{I}$. Hence, $\mathbf{B} = \mathbf{A}^{-1}$.

So, we have that $\mathbf{A}^{-1}$ exists if and only if $[\mathbf{A}]^{-1}$ exists.
Also, $[\mathbf{A}^{-1}] = [\mathbf{A}]^{-1}$.

Lecture 4: 4/11/2024

Suppose that E is an open set in $\mathbb{R}^n$, and that f is a function from E to $\mathbb{R}^m$. Then consider any $\vec{x} \in E$. We say that f is differentiable at $\vec{x}$ if there exists $\mathbf{A} \in L(\mathbb{R}^n, \mathbb{R}^m)$ such that:

$$\lim_{\vec{h}\to\vec{0}} \frac{\|f(\vec{x}+\vec{h}) - f(\vec{x}) - \mathbf{A}\vec{h}\|}{\|\vec{h}\|} = 0$$

Theorem 9.12: Suppose both $\mathbf{A}_1$ and $\mathbf{A}_2$ satisfy the above limit. Then $\mathbf{A}_1 = \mathbf{A}_2$.

Proof:
Note that:

$$\begin{aligned}\|\mathbf{A}_2\vec{h} - \mathbf{A}_1\vec{h}\| &= \|\mathbf{A}_2\vec{h} + f(\vec{x}+\vec{h}) - f(\vec{x}+\vec{h}) + f(\vec{x}) - f(\vec{x}) - \mathbf{A}_1\vec{h}\| \\ &\le \|\mathbf{A}_2\vec{h} + f(\vec{x}) - f(\vec{x}+\vec{h})\| + \|f(\vec{x}+\vec{h}) - f(\vec{x}) - \mathbf{A}_1\vec{h}\| \\ &= \|f(\vec{x}+\vec{h}) - f(\vec{x}) - \mathbf{A}_2\vec{h}\| + \|f(\vec{x}+\vec{h}) - f(\vec{x}) - \mathbf{A}_1\vec{h}\|\end{aligned}$$

It then follows that $\lim_{\vec{h}\to\vec{0}} \frac{\|\mathbf{A}_2\vec{h} - \mathbf{A}_1\vec{h}\|}{\|\vec{h}\|} = 0$.

So, let's fix $\vec{h_0} \in \mathbb{R}^n \setminus \{\vec{0}\}$. Then we know that $\lim_{t\to 0} \frac{\|(\mathbf{A}_2-\mathbf{A}_1)t\vec{h_0}\|}{\|t\vec{h_0}\|} = 0$.

But note that $\frac{\|(\mathbf{A}_2-\mathbf{A}_1)t\vec{h_0}\|}{\|t\vec{h_0}\|} = \frac{|t|\|(\mathbf{A}_2-\mathbf{A}_1)\vec{h_0}\|}{|t|\|\vec{h_0}\|} = \frac{\|(\mathbf{A}_2-\mathbf{A}_1)\vec{h_0}\|}{\|\vec{h_0}\|}$.

Thus $\lim_{t\to 0} \frac{\|(\mathbf{A}_2-\mathbf{A}_1)t\vec{h_0}\|}{\|t\vec{h_0}\|} = \frac{\|(\mathbf{A}_2-\mathbf{A}_1)\vec{h_0}\|}{\|\vec{h_0}\|}$ for all $\vec{h_0} \in \mathbb{R}^n \setminus \{\vec{0}\}$.

Thus we know that $(\mathbf{A}_2 - \mathbf{A}_1)\vec{h_0} = \vec{0}$ for all $\vec{h_0} \in \mathbb{R}$. Or in other words, $\mathbf{A}_2 = \mathbf{A}_1$.

Since any $\mathbf{A} \in L(\mathbb{R}^n, \mathbb{R}^m)$ satisfying that $\lim_{\vec{h} \to \vec{0}} \frac{\|f(\vec{x}+\vec{h}) - f(\vec{x}) - \mathbf{A}\vec{h}\|}{\|\vec{h}\|} = 0$ is unique, we denote $f'(\vec{x}) = \mathbf{A}$ and call $f'(\vec{x})$ the differential of f at $\vec{x}$.

Notes:

- If f is differentiable on all of E, then we say f is "differentiable in" E. In that case, note that f' can be interpreted as a function from E to $L(\mathbb{R}^n, \mathbb{R}^m)$.
- If we define $r(\vec{h}) = f(\vec{x} + \vec{h}) - f(\vec{x}) - f'(\vec{x})\vec{h}$, then we can say that $f(\vec{x} + \vec{h}) - f(\vec{x}) = f'(x)\vec{h} + r(\vec{h})$ where $\lim_{\vec{h} \to \vec{0}} \frac{\|r(\vec{h})\|}{\|\vec{h}\|} = 0$.
    - Proposition: If f is differentiable at $\vec{x}$, then f is continuous at $\vec{x}$.

      Proof:

      $$\begin{aligned} \|f(\vec{x} + \vec{h}) - f(\vec{x})\| &= \|f'(x)\vec{h} + r(\vec{h})\| \\ &\leq \|f'(x)\vec{h}\| + \|r(\vec{h})\| \\ &= \|f'(x)\vec{h}\| + \|\vec{h}\|\frac{\|r(\vec{h})\|}{\|\vec{h}\|} \end{aligned}$$

      Now because any $\mathbf{A} \in L(\mathbb{R}^n, \mathbb{R}^m)$ is uniformly continuous, we know that $f'(\vec{x})\vec{h} \to \vec{0}$ as $\vec{h} \to \vec{0}$. Also, since both $\|\vec{h}\|$ and $\frac{\|r(\vec{h})\|}{\|\vec{h}\|}$ approach $\vec{0}$ as $\vec{h} \to \vec{0}$, we know their product does as well.

      So by comparison we know that $\lim_{\vec{h} \to \vec{0}} \|f(\vec{x} + \vec{h}) - f(\vec{x})\| = 0$.

      Hence, $f(\vec{y}) \to f(\vec{x})$ as $\vec{y} \to \vec{x}$, which means that f is continuous at $\vec{x}$.

Here are some simple facts whose proofs are trivial.

- Sum Rule:

  Suppose that both f and g are functions going into $\mathbb{R}^m$, and that both are differentiable at $\vec{x} \in \mathbb{R}^n$. Then $(f + g)'(\vec{x}) = f'(\vec{x}) + g'(\vec{x})$.

- Scalar Multiplication Rule:

  Suppose that f is a function going into $\mathbb{R}^m$ that is differentiable at $\vec{x} \in \mathbb{R}^n$, and that $c \in \mathbb{R}$. Then $(cf)'(\vec{x}) = cf'(\vec{x})$.

- If $\mathbf{A} \in L(\mathbb{R}^n, \mathbb{R}^m)$, then for all $\vec{x} \in \mathbb{R}^n$, $\mathbf{A}'(\vec{x}) = \mathbf{A}$.

Theorem 9.15 (The Chain Rule): Suppose that $E \subseteq \mathbb{R}^m$ is open and that $f : E \longrightarrow \mathbb{R}^m$ is differentiable at $\vec{x}_0 \in E$. Also suppose that $f(\vec{x}_0)$ is in an open subset of $f(E)$, and that $g : f(E) \longrightarrow \mathbb{R}^k$ is differentiable at $f(\vec{x}) = \vec{y}_0$. Then $F = g \circ f$ is differentiable at $\vec{x}_0$ and:

$$F'(\vec{x}_0) = g'(f(\vec{x}_0))f'(\vec{x}_0)$$

Proof:
Set $\mathbf{A} = f'(\vec{x}_0)$ and $\mathbf{B} = g'(\vec{y}_0)$. Then define the functions:

- $u(\vec{h}) = f(\vec{x}_0 + \vec{h}) - f(\vec{x}_0) - \mathbf{A}\vec{h}$
- $v(\vec{k}) = g(\vec{y}_0 + \vec{k}) - g(\vec{y}_0) - \mathbf{B}\vec{k}$

Next, we define the function $\eta(\vec{k}) = \begin{cases} \frac{\|v(\vec{k})\|}{\|\vec{k}\|} & \text{if } v(\vec{k}) \neq \vec{0} \\ 0 & \text{if } v(\vec{k}) = \vec{0} \end{cases}$

Then, we always have that $\|\vec{k}\|\eta(\vec{k}) = \|v(\vec{k})\|$. Also, η is continuous at $\vec{k} = 0$. After all $v(\vec{0}) = \vec{0}$ and $\frac{\|v(\vec{k})\|}{\|\vec{k}\|} \to 0$ as $\vec{k} \to \vec{0}$.

With all that setup out of the way, we now need to show that:

$$\lim_{\vec{h} \to \vec{0}} \frac{\|F(\vec{x}_0 + \vec{h}) - F(\vec{x}_0) - \mathbf{BA}\vec{h}\|}{\|\vec{h}\|} = 0.$$

So, put $\vec{k} = f(\vec{x} + \vec{h}) - f(\vec{x})$ and note that:

$$\begin{aligned} F(\vec{x}_0 + \vec{h}) - F(\vec{x}_0) - \mathbf{BA}\vec{h} &= g(\vec{y}_0 + \vec{k}) - g(\vec{y}_0) - \mathbf{B}(\mathbf{A}\vec{h}) \\ &= \mathbf{B}\vec{k} + v(\vec{k}) - \mathbf{B}(\mathbf{A}\vec{h}) \\ &= v(\vec{k}) + \mathbf{B}(\vec{k} - \mathbf{A}\vec{h}) \end{aligned}$$

Meanwhile, notice that $\vec{k} = \mathbf{A}\vec{h} + u(\vec{h})$. Therefore, we have that:

$$\begin{aligned} \frac{\|F(\vec{x}_0 + \vec{h}) - F(\vec{x}_0) - \mathbf{BA}\vec{h}\|}{\|\vec{h}\|} &= \frac{\|v(\vec{k}) + \mathbf{B}(\vec{k} - \mathbf{A}\vec{h})\|}{\|\vec{h}\|} \\ &\leq \frac{\|v(\mathbf{A}\vec{h} + u(\vec{h}))\|}{\|\vec{h}\|} + \frac{\|\mathbf{B}(u(\vec{h}))\|}{\|\vec{h}\|} \\ &\leq \frac{\|\mathbf{A}\vec{h} + u(\vec{h})\|}{\|\vec{h}\|}\eta(\mathbf{A}\vec{h} + u(\vec{h})) + \|\mathbf{B}\|\frac{\|u(\vec{h})\|}{\|\vec{h}\|} \\ &\leq \left(\frac{\|\mathbf{A}\|\|\vec{h}\|}{\|\vec{h}\|} + \frac{\|u(\vec{h})\|}{\|\vec{h}\|}\right)\eta(\vec{k}) + \|\mathbf{B}\|\frac{\|u(\vec{h})\|}{\|\vec{h}\|} \end{aligned}$$

Now we know f is continuous at $\vec{x}_0$ because f is also differentiable there. So, $\vec{k} = f(\vec{x}_0 + \vec{h}) - f(\vec{x}_0) \to 0$ as $\vec{h} \to 0$. That combined with the fact that η is continuous at $\vec{k} = \vec{0}$ means that $\eta(\vec{k}) \to 0$ as $\vec{h} \to 0$.

Combining that with the fact that $\frac{\|u(\vec{h})\|}{\|\vec{h}\|} \to 0$ as $\vec{h} \to \vec{0}$, we know that:

$$\left(\frac{\|\mathbf{A}\|\|\vec{h}\|}{\|\vec{h}\|} + \frac{\|u(\vec{h})\|}{\|\vec{h}\|}\right)\eta(\vec{k}) + \|\mathbf{B}\|\frac{\|u(\vec{h})\|}{\|\vec{h}\|} \to 0 \text{ as } \vec{h} \to 0.$$

Hence, we can conclude by comparison that:

$$\lim_{\vec{h}\to\vec{0}} \frac{\|F(\vec{x}_0+\vec{h})-F(\vec{x}_0)-\mathbf{BA}\,\vec{h}\|}{\|\vec{h}\|} = 0.$$

Let $E \subseteq \mathbb{R}^n$ be open and consider a function $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$. Also let $\{e_1, \ldots, e_n\}$ and $\{u_1, \ldots, u_m\}$ be the standard bases of $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively. Then, expressing f in terms of its component functions, we have that:

$$f(\vec{x}) = \sum_{i=1}^{m} f_i(\vec{x})u_i = (f_1(\vec{x}), \ldots, f_m(\vec{x})).$$

Equivalently, we can write that $f_i(\vec{x}) = f(\vec{x}) \cdot u_i$.

Now for each $1 \leq j \leq n$ and $1 \leq i \leq m$, we define:

$$(D_j f_i)(\vec{x}) = \lim_{t\to 0} \frac{f_i(\vec{x} + te_j) - f_i(\vec{x})}{t}$$

Each of these $D_j f_i$ are called partial derivatives.

Note that to calculate the partial derivative $(D_j f_i)(\vec{x})$, all you need to do is treat all the components of $\vec{x}$ as constant except the jth component. Then, the partial derivative is just a single variable limit with respect to that component.

Theorem 9.17: Suppose $E \subseteq \mathbb{R}^n$ is open and that $f : E \longrightarrow \mathbb{R}^m$ is differentiable at a point $\vec{x} \in E$. Then all partial derivatives $(D_j f_i)(\vec{x})$ exist and:

$$f'(\vec{x})e_j = \sum_{i=1}^{k} (D_j f_i)(x)u_i.$$

Proof:
Fix $j \in \{1, \ldots, n\}$ and note that since f is differentiable, we have that $f(\vec{x} + te_j) - f(\vec{x}) = f'(\vec{x})(te_j) + r(te_j)$ such that $\frac{\|r(te_j)\|}{\|te_j\|} \to 0$ as $t \to 0$.

Then as $\|te_j\| = |t|$, we have that $\left\|\frac{r(te_j)}{t}\right\| = \frac{\|r(te_j)\|}{|t|} = \frac{\|r(te_j)\|}{\|te_j\|} \to 0$ as $t \to 0$. Thus, $\frac{r(te_j)}{t} \to \vec{0}$ as $t \to 0$. And since $f'(\vec{x})(te_j) = tf'(\vec{x})e_j$, we have that: $\frac{f'(\vec{x})(te_j)}{t} + \frac{r(te_j)}{t} \to f'(\vec{x})e_j$ as $t \to 0$.

So, we now know that $\lim_{t\to 0} \frac{f(\vec{x}+te_j)-f(\vec{x})}{t} = f'(\vec{x})(e_j)$.

Next, consider that:

$$\begin{aligned}(f'(x)e_j)\cdot u_i &= u_i \cdot \lim_{t\to 0}\tfrac{f(\vec{x}+te_j)-f(\vec{x})}{t}\\ &= \lim_{t\to 0}\left(u_i\cdot \tfrac{f(\vec{x}+te_j)-f(\vec{x})}{t}\right)\\ &= \lim_{t\to 0}\tfrac{(f(\vec{x}+te_j)\cdot u_i)-(f(\vec{x})\cdot u_i)}{t}\\ &= \lim_{t\to 0}\tfrac{f_i(\vec{x}+te_j)-f_i(\vec{x})}{t} = (D_jf_i)(\vec{x})\end{aligned}$$

It immediately follows that $f'(\vec{x})e_j = \sum_{i=1}^{k}(D_jf_i)(x)u_i$.

As a result of the above theorem, we have that if f is differentiable at $\vec{x}$, then when using $\{e_1,\ldots,e_n\}$ and $\{u_1,\ldots,u_m\}$ as our bases for $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively, then:

$$[f'(\vec{x})] = \begin{bmatrix}(D_1f_1)(\vec{x}) & (D_2f_1)(\vec{x}) & \cdots & (D_nf_1)(\vec{x})\\ (D_1f_2)(\vec{x}) & (D_2f_2)(\vec{x}) & \cdots & (D_nf_2)(\vec{x})\\ \vdots & \vdots & \ddots & \vdots\\ (D_1f_m)(\vec{x}) & (D_2f_m)(\vec{x}) & \cdots & (D_nf_m)(\vec{x})\end{bmatrix}$$

However, as I'm about to demonstrate, the converse of theorem 9.17 is not true. Thus, we can't automatically rely on calculating the partial derivatives of f to find the differential of f at $\vec{x}$.

Exercise 9.6: For $(x,y)\in\mathbb{R}^2$, define $f(x,y) = \begin{cases}\frac{xy}{x^2+y^2} & \text{if } (x,y)\neq(0,0)\\ 0 & \text{if } (x,y)=(0,0)\end{cases}$

Then we can show that the partial derivatives of f exist at every point of $\mathbb{R}^2$.

Clearly, we have that $(D_1f)(x,y) = \frac{y(x^2+y^2)}{(x^2+y^2)^2}$ and $(D_2f)(x,y) = \frac{x(x^2+y^2)}{(x^2+y^2)^2}$ when $(x,y)\neq 0$. Meanwhile, at $(x,y)=0$ we have when $h\neq 0$ that:

$$\tfrac{f(h,0)-f(0,0)}{h} = \tfrac{\frac{0}{h^2+0}-0}{h} = 0 \quad \text{and} \quad \tfrac{f(0,h)-f(0,0)}{h} = \tfrac{\frac{0}{0+h^2}-0}{h} = 0$$

Therefore:

$$(D_1f)(0,0) = \lim_{h\to 0}\tfrac{f(h,0)-f(0,0)}{h} = 0 \quad \text{and} \quad (D_2f)(0,0) = \lim_{h\to 0}\tfrac{f(0,h)-f(0,0)}{h} = 0$$

That said, $f(x,y)$ isn't even continuous at $(0,0)$.

Whenever $y=x$, we have that $f(x,y)=\frac{1}{2}$. Hence, given the sequence $(x_n,y_n)=(\frac{1}{n},\frac{1}{n})$, we have that $(x_n,y_n)\to(0,0)$ and that $(x_n,y_n)\neq(0,0)$ for any $n\in\mathbb{Z}_+$. But, $f(x_n,y_n)\not\to 0$.

Thus, $\lim\limits_{\substack{x\to 0\\ y\to 0}} f(x,y)\neq f(0,0)$, which means f is not continuous at $(0,0)$.

Since continuity is necessary for differentiability, this also demonstrates that the existence of partial derivatives does not imply a function is differentiable.

Let γ be a differentiable mapping from $(a,b) \subset \mathbb{R}$ to $E \subseteq \mathbb{R}^n$ where E is open and a and b are finite. Also let $f : E \longrightarrow \mathbb{R}$ be a differentiable function. Finally, define $g(t) = f(\gamma(t))$ for $a < t < b$. Then we know g is differentiable and that:

$$g'(t) = f'(\gamma(t))\gamma'(t).$$

Since g' is a real function, letting $\gamma(t) = (\gamma_1(t), \ldots, \gamma_n(t))$ we can rewrite the above expression as:

$$g'(t) = \sum_{i=1}^{n} D_i f(\gamma(t))\gamma_i'(t)$$

Now, this situation is so common that we have special notation just for it. Letting $\{e_1, \ldots, e_n\}$ be the standard basis for $\mathbb{R}^n$, we define the gradient of f at $\vec{x}$ as:

$$\nabla f(\vec{x}) = \sum_{i=1}^{n} D_i f(\vec{x}) e_i$$

Thus, $g'(t) = \nabla f(\gamma(t)) \cdot \gamma'(t)$.

Now suppose that $\gamma : (a,b) \longrightarrow \mathbb{R}^n$ is defined such that $a < 0 < b$, $\vec{x} \in E$, and $\gamma(t) = \vec{x} + t\vec{u}$ where $\vec{u}$ is a unit vector. Then we define the directional derivative of f at $\vec{x}$ in the direction of $\vec{u}$ as:

$$(D_{\vec{u}} f)(\vec{x}) = (f \circ \gamma)'(0) = \nabla f(\vec{x}) \cdot \vec{u}$$

From this it is really trivial to see that $|D_u f(\vec{x})|$ is maximized when $\vec{u}$ is a scalar multiple of $\nabla f(\vec{x})$.

To finish off lecture, note that if $E \subseteq \mathbb{R}^n$ is open and $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is differentiable and written as $f(\vec{x}) = (f_1(\vec{x}), \ldots, f_m(\vec{x}))$, then using the standard bases for $\mathbb{R}^n$ and $\mathbb{R}^m$, we can abuse notation and say that:

$$[f'(\vec{x})] = \begin{bmatrix} \nabla f_1(\vec{x}) \\ \vdots \\ \nabla f_m(\vec{x}) \end{bmatrix}$$

Lecture 5: 4/16/2024

A set $E \subseteq \mathbb{R}^k$ is convex if $t\vec{x} + (1-t)\vec{y} \in E$ whenever $x \in E$, $y \in E$, and $0 < t < 1$.

Theorem 9.19: Suppose that $E \subseteq \mathbb{R}^n$ is convex and open, that $f : E \longrightarrow \mathbb{R}^m$ is differentiable, and there is a real number M such that $\|f'(\vec{x})\| \leq M$ for every $\vec{x} \in E$. Then $\|f(\vec{b}) - f(\vec{a})\| \leq M\|\vec{b} - \vec{a}\|$ for all $\vec{a}, \vec{b} \in E$.

Proof:
Fix $\vec{a}, \vec{b} \in E$ and define $\gamma(t) = (1-t)\vec{a} + t\vec{b}$ for $0 < t < 1$. Next, define $\vec{g}(t) = f(\gamma(t))$ for $0 < t < 1$. Since E is both convex and differentiable, we know that both $\vec{g}$ and $\vec{g}'$ are well defined on the interval $(0, 1)$. So, by the mean value theorem for vector-valued functions (proposition 93 in the Math 140B notes), we know that for some $x \in (0, 1)$:

$$\|\vec{g}(1) - \vec{g}(0)\| \leq |1 - 0|\|\vec{g}'(x)\|$$

Or in other words, $\|f(\vec{b}) - f(\vec{a})\| \leq \|f'(\gamma(x))\gamma'(x)\| \leq M\|\vec{b} - \vec{a}\|$.

Corollary: If additionally $f'(\vec{x}) = \mathbf{0}$ for all $\vec{x} \in E$, then f is constant.

Proof:
If $f'(\vec{x}) = \mathbf{0}$ for all $\vec{x}$, then $M = 0$. So $\|f(\vec{b}) - f(\vec{a})\| = 0$ for all $\vec{a}, \vec{b} \in E$.

Let $E \subseteq \mathbb{R}^n$ be an open set and $f : E \longrightarrow \mathbb{R}^m$ be differentiable. Then f is called continuously differentiable if $f' : E \longrightarrow L(\mathbb{R}^n, \mathbb{R}^m)$ is continuous. Or in other words, $\forall \varepsilon > 0, \ \exists \delta > 0 \ \ s.t. \ \ \|\vec{y} - \vec{x}\| < \delta \Longrightarrow \|f'(\vec{y}) - f'(\vec{x})\| < \varepsilon$

When this is the case, we say f is a $\mathscr{C}^1$-mapping and that $f \in \mathscr{C}^1(E, \mathbb{R}^m)$.

Theorem 9.21: Suppose f is a function from an open set $E \subseteq \mathbb{R}^n$ to $\mathbb{R}^m$. Then $f \in \mathscr{C}^1(E, \mathbb{R}^m)$ if and only if the partial derivatives $D_j f_i$ exist and are continuous on E for all $1 \leq i \leq m$ and $1 \leq j \leq n$.

Proof:
Let $\{e_1, \ldots, e_n\}$ and $\{u_1, \ldots, u_m\}$ be the standard bases for $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively.

($\Longrightarrow$) Say $f \in \mathscr{C}^1(E, \mathbb{R}^m)$. Then f is differentiable, which means that $(D_j f_i)(\vec{x}) = (f'(\vec{x})e_j) \cdot u_i$ exists for all $1 \leq i \leq m, \ 1 \leq j \leq n$, and $\vec{x} \in E$.

Thus for any $\vec{x}, \vec{y} \in E$, we have that:

$$\begin{aligned}|(D_j f_i)(\vec{y}) - (D_j f_i)(\vec{x})| &= |(f'(\vec{y})e_j) \cdot u_i - (f'(\vec{x})e_j) \cdot u_i| \\ &= |((f'(\vec{x}) - f'(\vec{x}))e_j) \cdot u_i| \\ &\leq \|((f'(\vec{x}) - f'(\vec{x}))e_j)\|\|u_i\| \\ &\leq \|f'(\vec{x}) - f'(\vec{x})\|\|e_j\|\|u_i\| \\ &= \|f'(\vec{x}) - f'(\vec{x})\|\end{aligned}$$

Now, since $f \in \mathscr{C}^1(E, \mathbb{R}^m)$, we know that for any $\varepsilon > 0$, there exists $\delta > 0$ such that $\|\vec{y} - \vec{x}\| < \delta \Longrightarrow \|f'(\vec{x}) - f'(\vec{x})\| < \varepsilon$. Thus, by the above inequality, we also have that:

$$\|\vec{y} - \vec{x}\| < \delta \Longrightarrow |(D_j f_i)(\vec{y}) - (D_j f_i)(\vec{x})| < \varepsilon.$$

($\Longleftarrow$) Firstly, we'll assume that $m = 1$. That way f is a real-valued function.

Fix $\vec{x} \in E$ and $\varepsilon > 0$. Then because E is open, there exists $r_0 > 0$ such that $B_{r_0}(\vec{x}) \subseteq E$. Additionally, because each partial derivative is continuous, there exists $r_j > 0$ such that:

$$\|\vec{y} - \vec{x}\| < r_j \Longrightarrow |(D_j f)(\vec{y}) - (D_j f)(\vec{x})| < \tfrac{\varepsilon}{n}.$$

Set $R = \min\{r_0, r_1, \ldots, r_n\}$. Then choose $\vec{h} = (h_1, \ldots, h_n) \in B_R(\vec{x})$, and define $\vec{v}_j$ such that $v_0 = \vec{0}$ and $v_j = \vec{v}_{j-1} + h_j e_j$ for each $1 \leq j \leq n$. Then note that:

$$f(\vec{x} + \vec{h}) - f(\vec{x}) = \sum_{j=1}^{n} \Big(f(\vec{x} + \vec{v}_j) - f(\vec{x} + \vec{v}_{j-1}) \Big)$$

Now for each $1 \leq j \leq n$, define $\widetilde{f}_j(t) = f(\vec{x} + \vec{v}_{j-1} + t h_j e_j)$. Then firstly, $\widetilde{f}_j$ is well defined on the interval $[0, 1]$ because:

- $\widetilde{f}_j(1) = f(\vec{x} + \vec{v}_j) \in B_R(\vec{x})$
- $\widetilde{f}_j(0) = f(\vec{x} + \vec{v}_{j-1}) \in B_R(\vec{x})$
- $B_R(\vec{x}) \subseteq E$ is convex.

Additionally, $\widetilde{f}_j$ is differentiable on $[0, 1]$ with:

$$\widetilde{f}_j'(t) = h_j (D_j f)(\vec{x} + \vec{v}_{j-1} + t h_j e_j)$$

Therefore, by the mean value theorem, there exists $\theta \in (0, 1)$ such that:

$$\widetilde{f}_j(1) - \widetilde{f}_j(0) = \widetilde{f}_j'(\theta) = h_j (D_j f)(\vec{x} + \vec{v}_{j-1} + \theta h_j e_j)$$

Importantly, we have that $(\vec{x} + \vec{v}_{j-1} + \theta h_j e_j) \in B_R(\vec{x})$. Therefore:

$$|(D_j f)(\vec{x} + \vec{v}_{j-1} + \theta h_j e_j) - (D_j f)(\vec{x})| < \tfrac{\varepsilon}{n}$$

In turn, this means that:

$$\begin{aligned}&|f(\vec{x} + \vec{v}_j) - f(\vec{x} + \vec{v}_{j-1}) - h_j (D_j f)(\vec{x})| \\ &\qquad = |h_j (D_j f)(\vec{x} + \vec{v}_{j-1} + \theta h_j e_j) - h_j (D_j f)(\vec{x})| < |h_j| \tfrac{\varepsilon}{n}\end{aligned}$$

So now, we can say that:

$$\begin{aligned}
&\left| f(\vec{x}+\vec{h}) - f(\vec{x}) - \sum_{j=1}^{n} h_j (D_j f)(\vec{x}) \right| \\
&\qquad = \left| \sum_{j=1}^{n} \Big(f(\vec{x}+\vec{v}_j) - f(\vec{x}+\vec{v}_{j-1}) - h_j (D_j f)(\vec{x}) \Big) \right| \\
&\qquad \leq \sum_{j=1}^{n} \left| f(\vec{x}+\vec{v}_j) - f(\vec{x}+\vec{v}_{j-1}) - h_j (D_j f)(\vec{x}) \right| \\
&\qquad < \sum_{j=1}^{n} |h_j| \tfrac{\varepsilon}{n} \leq \varepsilon \sum_{j=1}^{n} \tfrac{\|\vec{h}\|}{n} = \varepsilon \|\vec{h}\|
\end{aligned}$$

And thus we have that: $\lim\limits_{\vec{h} \to \vec{0}} \dfrac{\left| f(\vec{x}+\vec{h}) - f(\vec{x}) - \sum\limits_{j=1}^{n} h_j (D_j f)(\vec{x}) \right|}{\|\vec{h}\|} = 0.$

Note that $f'(\vec{x}) = \sum\limits_{j=1}^{n} h_j (D_j f)(\vec{x}) \in L(\mathbb{R}^n, \mathbb{R})$, and that $f'(\vec{x})$ is the sum of a bunch of continuous functions. Thus, $f'(\vec{x})$ is itself continuous at $\vec{x}$, meaning that f is a $\mathscr{C}^1$-mapping.

Now that we've shown this theorem holds when $m = 1$, lets consider when $m > 1$. Let $f(\vec{x}) = (f_1(\vec{x}), \ldots, f_m(\vec{x}))$ and then note that f is continuous at $\vec{x}$ if and only if f_i is continuous at $\vec{x}$ for each $1 \leq i \leq m$. Luckily, we already showed that all $D_j f_i$ being continuous for each $1 \leq j \leq n$ implies that f_i is continuously differentiable. Hence, having all $D_j f_i$ be continuous for each $1 \leq j \leq n$ and $1 \leq i \leq m$ implies that $f \in \mathscr{C}^1(E, \mathbb{R}^m)$.

Exercise 9.7: Suppose that f is a real-valued function defined in an open set $E \subseteq \mathbb{R}^n$, and that the partial derivatives $D_1 f, \ldots, D_n f$ are defined and bounded in E. Then we can prove that f is continuous in E.

Pick $M > |(D_j f)(\vec{x})|$ for all $1 \leq j \leq n$ and $\vec{x} \in E$. Then fix $\vec{x} \in E$ and $\varepsilon > 0$. Next, choose $r > 0$ such that $B_r(\vec{x}) \subseteq E$ and set $R = \min\{r, \frac{\varepsilon}{nM}\}$.

Like in the proof above, choose $\vec{h} = (h_1, \ldots, h_n) \in B_R(\vec{x})$ and define $\vec{v}_j$ such that $v_0 = \vec{0}$ and $v_j = \vec{v}_{j-1} + h_j e_j$ for each $1 \leq j \leq n$. That way:

$$f(\vec{x}+\vec{h}) - f(\vec{x}) = \sum_{j=1}^{n} \Big(f(\vec{x}+\vec{v}_j) - f(\vec{x}+\vec{v}_{j-1}) \Big)$$

Next, for each $1 \leq j \leq n$, define $\widetilde{f}_j(t) = f(\vec{x} + \vec{v}_{j-1} + th_j e_j)$. Then like before, $\widetilde{f}_j$ is well defined on the interval $[0,1]$ because:

- $\widetilde{f}_j(1) = f(\vec{x} + \vec{v}_j) \in B_R(\vec{x})$
- $\widetilde{f}_j(0) = f(\vec{x} + \vec{v}_{j-1}) \in B_R(\vec{x})$
- $B_R(\vec{x}) \subseteq E$ is convex.

Additionally, $\widetilde{f}_j$ is differentiable on $[0,1]$ with:

$$\widetilde{f}_j'(t) = h_j(D_j f)(\vec{x} + \vec{v}_{j-1} + th_j e_j)$$

Therefore, by the mean value theorem, there exists $\theta \in (0,1)$ such that:

$$\widetilde{f}_j(1) - \widetilde{f}_j(0) = \widetilde{f}_j'(\theta) = h_j(D_j f)(\vec{x} + \vec{v}_{j-1} + \theta h_j e_j)$$

And then because $|D_j f|$ is bounded by M:

$$\begin{aligned} |f(\vec{x} + \vec{v}_j) - f(\vec{x} + \vec{v}_{j-1})| &= |\widetilde{f}_j(1) - \widetilde{f}_j(0)| \\ &= |h_j(D_j f)(\vec{x} + \vec{v}_{j-1} + \theta h_j e_j)| \\ &= |h_j||(D_j f)(\vec{x} + \vec{v}_{j-1} + \theta h_j e_j)| < |h_j| M \end{aligned}$$

So now:

$$|f(\vec{x} + \vec{h}) - f(\vec{x})| \leq \sum_{j=1}^{n} \left| f(\vec{x} + \vec{v}_j) - f(\vec{x} + \vec{v}_{j-1}) \right| < \sum_{j=1}^{n} |h_j| M \leq \|\vec{h}\| \cdot nM$$

Hence, since $R = \frac{\varepsilon}{nM}$, we have that $|f(\vec{x} + \vec{h}) - f(\vec{x})| < \varepsilon$ for any $\vec{h} \in B_R(\vec{x})$. So, f is continuous.

Exercise 9.8: Suppose that f is a differentiable real function on an open set $E \subseteq \mathbb{R}^n$, and that f has a local maximum at a point $x \in E$. Then $f'(x) = \mathbf{0}$.

Proof:
Let $r > 0$ such that $B_r \subseteq E$ and $\|\vec{h}\| < r \implies f(\vec{x} + \vec{h}) \leq f(\vec{x})$. Then pick any such $\vec{h} \in B_r$ and define $\gamma : (-1,1) \longrightarrow \mathbb{R}^n$ by $\gamma(t) = \vec{x} + t\vec{h}$. Then note that $(f \circ \gamma)'(t) = f'(\vec{x} + t\vec{h})\,\vec{h}$.

Crucially, we know that $(f \circ \gamma)'$ reaches a local maximum at $t = 0$. And since $f \circ \gamma$ is a function mapping an interval of $\mathbb{R}$ to $\mathbb{R}$, we know from Math 140B that $(f \circ \gamma)'(0) = f'(\vec{x})\,\vec{h} = 0$. Then as $\vec{h}$ can be a scalar multiple of any vector, we must have that $f'(\vec{x}) = \mathbf{0}$.

Let X be a metric space with the metric d, and consider some $\varphi : X \longrightarrow X$. We call φ a contraction if there exists $c < 1$ such that for all $x, y \in X$:

$$d(\varphi(x), \varphi(y)) \leq c \cdot d(x, y).$$

Theorem 9.23 (Banach Fixed Point Theorem):
Let X be a complete metric space with metric d, and let $\varphi : X \longrightarrow X$ be a contraction. Then there exists a unique $x \in X$ such that $\varphi(x) = x$

Proof:
We can establish that x is unique if it exists very easily.

Assume $x, y \in X$ such that $\varphi(x) = x$ and $\varphi(y) = y$. Then: $d(x, y) = d(\varphi(x), \varphi(y)) \leq c \cdot d(x, y)$. Since $c < 1$, the only way that this is possible is if $d(x, y) = 0$. So, $x = y$.

Therefore, we now will focus on showing that x exists.

Pick any $x_0 \in X$ and then recursively define the sequence (x_n) by setting $x_{n+1} = \varphi(x_n)$ for all n. Then because f is a contraction, there exists $c < 1$ such that $d(\varphi(a), \varphi(b)) \leq c \cdot d(a, b)$ for all $a, b \in X$. Hence, for all $n \geq 1$, we have that:

$$d(x_{n+1}, x_n) = d(\varphi(x_n), \varphi(x_{n-1})) \leq c \cdot d(x_n, x_{n-1}).$$

By induction, we thus get that $d(x_{n+1}, x_n) \leq c^n d(x_1, x_0)$ for all n (including 0).

Now consider that for any $m, n \in \mathbb{Z}_+$ such that $m > n \geq 1$, we have that:

$$d(x_m, x_n) \leq \sum_{i=n}^{m-1} d(x_{i+1}, x_i) \leq d(x_1, x_0)(c^n + \ldots + c^{m-1})$$
$$\leq d(x_1, x_0)c^n \sum_{i=1}^{m-n-1} c^i \leq \tfrac{c^n}{1-c} d(x_1, x_0)$$

This shows that (x_n) is a Cauchy sequence. After all, for any $\varepsilon > 0$, we can set N such that $\frac{c^N}{1-c} d(x_1, x_0) < \varepsilon$. That way, for all $m > n > N$ we have that $d(x_m, x_n) < \varepsilon$.

So because X is complete, we know that (x_n) converges. Therefore let x equal the limit of (x_n) and consider any $\varepsilon > 0$.

Since $x_n \to x$, we know that there exists N such that $\forall m > N$, we have that $d(x, x_m) < \varepsilon/2$. And since m is a contraction, we also know that $d(\varphi(x), \varphi(x_m)) \leq c\frac{\varepsilon}{2} < \varepsilon/2$ for all $m > N$. Hence letting $m > N$:

$$d(\varphi(x), x) \leq d(\varphi(x), \varphi(x_m)) + d(x_{m+1}, x) < \varepsilon/2 + \varepsilon/2 = \varepsilon$$

And as ε is arbitrary, we thus know that $d(\varphi(x), x) = 0$. So, $\varphi(x) = x$.

Lecture 6: 4/18/2024

The Inverse Function Theorem:
Suppose that $f \in \mathscr{C}^1(E, \mathbb{R}^n)$ where $E \subseteq \mathbb{R}^n$ is open, that $f'(\vec{a})$ is invertible for some $\vec{a} \in E$, and that $\vec{b} = f(\vec{a})$. Then:

(A) There exist open subsets U and V of $\mathbb{R}^n$ such that $\vec{a} \in U$, $\vec{b} \in V$, $f(U) = V$, and f is one-to-one on U.

(B) If g is defined such that $(g \circ f)(\vec{x}) = \vec{x}$ for all $x \in U$, then $g \in \mathscr{C}^1(V, \mathbb{R}^n)$ with $g'(\vec{y}) = f'(g(\vec{y}))^{-1}$ for all $\vec{y} \in V$.

Proof of part A:

Set $\mathbf{A} = f'(\vec{a})$ and set $\lambda = \frac{1}{2\|\mathbf{A}\|^{-1}}$. Then since $f \in \mathscr{C}^1(E, \mathbb{R}^n)$, we know that there exists $\delta > 0$ such that $\vec{x} \in B_\delta(\vec{a}) \implies \|f'(\vec{x}) - \mathbf{A}\| < \lambda$.

So, for any r satisfying that $0 < r < \delta$ and $B_r \subseteq E$, set $U = B_r(\vec{a})$. Then let $\vec{y} \in \mathbb{R}^n$ and consider the function:

$$\varphi(\vec{x}) = \vec{x} + \mathbf{A}^{-1}(\vec{y} - f(\vec{x})) \quad \text{(defined for all } \vec{x} \in E\text{)}.$$

Now firstly, observe that $f(\vec{x}) = \vec{y}$ if and only if $\vec{x}$ is a fixed point of φ.

Proof:
($\implies$) If $f(\vec{x}) = \vec{y}$, then $\varphi(\vec{x}) = \vec{x} + \mathbf{A}^{-1}(\vec{0}) = \vec{x}$.

($\impliedby$) If $\varphi(\vec{x}) = \vec{x}$, then $\vec{0} = \mathbf{A}^{-1}(\vec{y} - f(\vec{x}))$. As $\operatorname{null}(\mathbf{A}^{-1}) = 0$, we therefore must have that $\vec{y} - f(\vec{x}) = \vec{0}$. So $f(\vec{x}) = \vec{y}$.

Secondly, note that $\varphi'(\vec{x}) = \mathbf{I} - \mathbf{A}^{-1} f'(\vec{x}) = \mathbf{A}^{-1}(\mathbf{A} - f'(\vec{x}))$. Therefore, assuming $\vec{x} \in U$, we know that:

$$\|\varphi'(\vec{x})\| \leq \|\mathbf{A}^{-1}\|\|\mathbf{A} - f'(\vec{x})\| < \|\mathbf{A}^{-1}\|\lambda = \tfrac{1}{2}$$

In turn, since $U = B_r(\vec{a})$ is convex, we can apply theorem 9.19 to say that $\|\varphi(\vec{x}_2) - \varphi(\vec{x}_1)\| \leq \frac{1}{2}\|\vec{x}_2 - \vec{x}_1\|$. Now, be aware that this does not imply that φ is a contraction because we still don't know where φ maps U to. However, this is enough information to do the proof on the previous page showing that φ has at most one fixed point on U.

Hence, f is one-to-one on U since for any $\vec{y} \in \mathbb{R}^n$, there can be at most one $\vec{x} \in U$ such that $\vec{y} = f(\vec{x})$

Now, we move on to showing that $V = f(U)$ is open.

Let $\vec{y}_0 \in V$ and note that there exists a unique $\vec{x}_0 \in U$ such that $f(\vec{x}_0) = \vec{y}_0$. Also, because U is open, we know there exists $R > 0$ such that $\overline{B_R(\vec{x}_0)} \subset U$. So let B equal that closed ball.

Next, fix $\vec{y}$ such that $\|\vec{y} - \vec{y_0}\| < \lambda R$ and define φ using $\vec{y}$ as was described on the last page. Then note that:

$$\|\varphi(\vec{x_0}) - \vec{x_0}\| = \|\mathbf{A}^{-1}(\vec{y} - \vec{y_0})\| < \|\mathbf{A}^{-1}\|\lambda R = \tfrac{R}{2}$$

Therefore, we have that for all $\vec{x} \in B$:

$$\begin{aligned}\|\varphi(\vec{x}) - \vec{x_0}\| &\leq \|\varphi(\vec{x}) - \varphi(\vec{x_0})\| + \|\varphi(\vec{x_0}) - \vec{x_0}\| \\ &< \tfrac{1}{2}\|\vec{x} - \vec{x_0}\| + \tfrac{R}{2} < \tfrac{R}{2} + \tfrac{R}{2} = R\end{aligned}$$

Or in other words, $\vec{x} \in B \implies \phi(\vec{x}) \in B$.

Hence, φ is a contraction over B (for real this time). Also, because B is a closed and bounded subset of $\mathbb{R}^n$, we know B is compact and thus complete. So by theorem 9.23, φ has a fixed point $\vec{x} \in B$. It follows then that $\vec{y} \in V$.

Since $\vec{y}$ was arbitary, we thus know that $B_{\lambda R}(\vec{y_0}) \subseteq V$. So, V is open.

Proof of part B:

Based on the previous part, we know that f is a bijective map from U to V. So, there does exist g such that $(g \circ f)(\vec{x}) = \vec{x}$ for all $\vec{x} \in U$.

Now pick $\vec{y} \in V$ and $\vec{y} + \vec{k} \in V$. Then there exists $\vec{x} \in U$ and $\vec{x} + \vec{h} \in U$ such that $f(\vec{x}) = \vec{y}$ and $f(\vec{x} + \vec{h}) = \vec{y} + \vec{k}$.

Next, having defined φ one more time using any $\vec{y_0} \in \mathbb{R}^n$ we have that:

$$\begin{aligned}\varphi(\vec{x} + \vec{h}) - \varphi(\vec{x}) &= \vec{x} + \vec{h} + \mathbf{A}^{-1}(\vec{y_0} - f(\vec{x} + \vec{h})) - \vec{x} - \mathbf{A}^{-1}(\vec{y_0} - f(\vec{x})) \\ &= \vec{h} + \mathbf{A}^{-1}(f(\vec{x}) - f(\vec{x} + \vec{h})) \\ &= \vec{h} - \mathbf{A}^{-1}\vec{k}\end{aligned}$$

At the same time, we know from part A that $\|\varphi(\vec{x} + \vec{h}) - \varphi(\vec{x})\| < \frac{1}{2}\|\vec{h}\|$. Therefore, we have that $\|\vec{h} - \mathbf{A}^{-1}\vec{k}\| \leq \frac{1}{2}\|\vec{h}\|$, and in turn this means that $\|\mathbf{A}^{-1}\vec{k}\| \geq \frac{1}{2}\|\vec{h}\|$.

This is because $\|\vec{h}\| \leq \|\vec{h} - \mathbf{A}^{-1}\vec{k}\| + \|\mathbf{A}^{-1}\vec{k}\| \leq \frac{1}{2}\|\vec{h}\| + \|\mathbf{A}^{-1}\vec{k}\|$.

So $\|\vec{h}\| \leq 2\|\mathbf{A}^{-1}\vec{k}\| \leq 2\|\mathbf{A}^{-1}\|\|\vec{k}\| = \frac{1}{\lambda}\|\vec{k}\|$.

Then, because $\vec{x} \in U$, we know that $\|f'(\vec{x}) - \mathbf{A}\| \leq \lambda < \frac{1}{\|\mathbf{A}^{-1}\|}$. Thus, $f'(\vec{x})$ has an inverse which we shall call $\mathbf{T}$. Now note that:

$$\begin{aligned}g(\vec{y} + \vec{k}) - g(\vec{y}) - \mathbf{T}\vec{k} &= \vec{h} - \mathbf{T}\vec{k} \\ &= \mathbf{T}f'(\vec{x})\vec{h} - \mathbf{T}(f(\vec{x} + \vec{h}) - f(\vec{x})) \\ &= -\mathbf{T}(f(\vec{x} + \vec{h}) - f(\vec{x}) - f'(\vec{x})\vec{h})\end{aligned}$$

Combining that with the fact that $\|\vec{h}\|\lambda \leq \|\vec{k}\|$, we know that:

$$\frac{\|g(\vec{y}+\vec{k})-g(\vec{y})-\mathbf{T}\vec{k}\|}{\|\vec{k}\|} \leq \frac{\|\mathbf{T}\|}{\lambda} \cdot \frac{\|f(\vec{x}+\vec{h})-f(\vec{x})-f'(\vec{x})\vec{h}\|}{\|\vec{h}\|}$$

So because $\vec{h} \to \vec{0}$ as $\vec{k} \to \vec{0}$, we know that $\lim_{\vec{k}\to\vec{0}} \frac{\|g(\vec{y}+\vec{k})-g(\vec{y})-\mathbf{T}\vec{k}\|}{\|\vec{k}\|} = 0$. And thus $g'(\vec{y}) = \mathbf{T} = f'(g(\vec{y}))^{-1}$.

Importantly, $\vec{y}$ was arbitrary. So we generally have that $g : V \longrightarrow \mathbb{R}^n$ is differentiable with $g'(\vec{y}) = f'(g(\vec{y}))^{-1}$. Additionally, note that:

- g being differentiable implies that g is continuous.
- f' is continuous by the assumption of this theorem.
- The map $\mathbf{A} \mapsto \mathbf{A}^{-1}$ is continuous by theorem 9.8.

Therefore, g' is the composition of three continuous functions, which means that g' is continuous. So $g \in \mathscr{C}^1(V, \mathbb{R}^n)$. ■

Corollary: If $E \subseteq \mathbb{R}^n$ is open and $f \in \mathscr{C}^1(E, \mathbb{R}^n)$ satisfies that $f'(\vec{x})$ is invertible for every $\vec{x} \in E$, then for every open $W \subseteq E$, $f(W)$ is open.

This is a direct result of part A of the previous theorem.

Lecture 7: 4/25/2024

For the next theorem we need to go over some notation.

Given any $\vec{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ and $\vec{y} = (y_1, \ldots, y_m) \in \mathbb{R}^m$, we shall write $(\vec{x}, \vec{y})$ for the vector $(x_1, \ldots, x_n, y_1, \ldots, y_m) \in \mathbb{R}^{n+m}$.

Also, given any $\mathbf{A} \in L(\mathbb{R}^{n+m}, \mathbb{R}^n)$, we shall define $\mathbf{A}_x(\vec{h}) = \mathbf{A}(\vec{h}, \vec{0})$ and $\mathbf{A}_y(\vec{k}) = \mathbf{A}(\vec{0}, \vec{k})$ for all $\vec{h} \in \mathbb{R}^n$ and $\vec{k} \in \mathbb{R}^m$. That way, $\mathbf{A}_x \in L(\mathbb{R}^n)$, $\mathbf{A}_y \in L(\mathbb{R}^m, \mathbb{R}^n)$, and $\mathbf{A}(\vec{h}, \vec{k}) = \mathbf{A}_x\vec{h} + \mathbf{A}_y\vec{k}$.

Proposition: If $\mathbf{A} \in L(\mathbb{R}^{n+m}, \mathbb{R}^n)$ and if $\mathbf{A}_x$ is invertible, then there corresponds to every $\vec{k} \in \mathbb{R}^m$ a unique $\vec{h} \in \mathbb{R}^n$ such that $\mathbf{A}(\vec{h}, \vec{k}) = \vec{0}$. Specifically, this $\vec{h}$ is equal to $-\mathbf{A}_x^{-1}\mathbf{A}_y\vec{k}$.

Proof:
$\mathbf{A}(\vec{h}, \vec{k}) = \vec{0} \implies \mathbf{A}_x\vec{h} + \mathbf{A}_y\vec{k} = \vec{0} \implies \vec{h} = -\mathbf{A}_x^{-1}\mathbf{A}_y\vec{k}$

Now what we want to do on the next page is extend this theorem somehow to a more general function $f : E \longrightarrow \mathbb{R}^n$ where $E \subseteq \mathbb{R}^{n+m}$ is open.

Implicit Function Theorem:

Suppose that $E \subseteq \mathbb{R}^{n+m}$ is open and that $f \in \mathscr{C}^1(E, \mathbb{R}^n)$ satisfies that $f(\vec{a}, \vec{b}) = \vec{0}$ for some $(\vec{a}, \vec{b}) \in E$. Also, set $\mathbf{A} = f'(\vec{a}, \vec{b})$ and suppose that $\mathbf{A}_x$ is invertible. Then:

(A) There exists open sets $U \subseteq \mathbb{R}^{n+m}$ and $W \subseteq \mathbb{R}^m$ such that $(\vec{a}, \vec{b}) \in U$, $\vec{b} \in W$, and $\forall \vec{y} \in W$ there exists a unique $\vec{x}$ such that $(\vec{x}, \vec{y}) \in U$ and $f(\vec{x}, \vec{y}) = \vec{0}$.

(B) If we define $g : W \longrightarrow \mathbb{R}^n$ such that each $\vec{y} \in W$ is mapped to the unique $\vec{x}$ described above, then $f(g(\vec{y}), \vec{y}) = \vec{0}$ for all $\vec{y} \in W$, $g(\vec{b}) = \vec{a}$, and $g \in \mathscr{C}^1(W, \mathbb{R}^m)$ with $g'(\vec{b}) = -\mathbf{A}_x^{-1}\mathbf{A}_y$.

Proof of part A:

Define $F : \mathbb{R}^{n+k} \longrightarrow \mathbb{R}^{n+k}$ by the rule $F(\vec{x}, \vec{y}) = (f(\vec{x}, \vec{y}), \vec{y})$. Then, $F \in \mathscr{C}^1(E, \mathbb{R}^{n+m})$.

(You can check that all the partial derivatives of F are continuous)

So, let us now consider $F'(\vec{a}, \vec{b})$. Importantly, you can check that $F'(\vec{a}, \vec{b})(\vec{h}, \vec{k}) = (f'(\vec{a}, \vec{b})(\vec{h}, \vec{k}), \ \vec{k})$. Immediately, this means if $F'(\vec{a}, \vec{b})(\vec{h}, \vec{k}) = \vec{0}$, we must have that $\vec{k} = \vec{0}$. Also note that $f'(\vec{a}, \vec{b})(\vec{h}, \vec{k}) = \mathbf{A}(\vec{h}, \vec{k}) = \mathbf{A}_x\vec{h} + \mathbf{A}_y\vec{k}$. Since $\mathbf{A}_x$ is invertible and $\vec{k} = 0$, we know that $\mathbf{A}_x\vec{h} + \mathbf{A}_y\vec{k} = \vec{0} \Longrightarrow \vec{h} = \vec{0}$.

Therefore, we have shown that:

$$F'(\vec{a}, \vec{b})(\vec{h}, \vec{k}) = \vec{0} \Longrightarrow (\vec{h}, \vec{k}) = (\vec{0}, \vec{0}).$$

Or in other words, $F'(\vec{a}, \vec{b})$ is invertible.

Now by the inverse function theorem let U and V be open subsets of $\mathbb{R}^{n+m}$ such that $(\vec{a}, \vec{b}) \in U$, $F(\vec{a}, \vec{b}) = (\vec{0}, \vec{b}) \in V$, and F is invertible going from U to V. Then set $W = \{\vec{y} \in \mathbb{R}^m \mid (\vec{0}, \vec{y}) \in V\}$.

Clearly, W is open since V is open. Also, $\vec{b} \in W$. Finally, let $\vec{y} \in W$ and consider that $(\vec{0}, \vec{y}) \in V$. Therefore there is a unique $(\vec{x}, \vec{y}) \in U$ such that $F(\vec{x}, \vec{y}) = f((\vec{x}, \vec{y}), \vec{y}) = (\vec{0}, \vec{y})$. Obviously, this $\vec{x}$ is also the unique vector satisfying that $(\vec{x}, \vec{y}) \in U$ and $f(\vec{x}, \vec{y}) = \vec{0}$.

Proof of part B:

Now let $g : W \longrightarrow \mathbb{R}^n$ such that $g(\vec{y})$ satisfies that $f(g(\vec{y}), \vec{y}) = \vec{0}$ and $(g(\vec{y}), \vec{y}) \in V$ for all $\vec{y} \in U$. Also define $G : V \longrightarrow U$ as the function inverting F on U.

By the inverse function theorem, $G \in \mathscr{C}^1(V, \mathbb{R}^{n+m})$. Also, for all $y \in \vec{W}$ we have that $(g(\vec{y}), \vec{y}) = G(F(g(\vec{y}), \vec{y})) = G(\vec{0}, \vec{y})$. Thus, by defining $\Phi : W \longrightarrow V$ such that $\Phi(\vec{y}) = (g(\vec{y}), \vec{y})$, we know that $\Phi \in \mathscr{C}^1(W, \mathbb{R}^{n+m})$. And in turn, this means that $g \in \mathscr{C}^1(W, \mathbb{R}^n)$.

Finally, consider that $g(\vec{b}) = \vec{a}$ trivially. At the same time, note that $\Phi'(\vec{y})\vec{k} = (g'(\vec{y})\vec{k}, \vec{k})$ for all $\vec{y} \in W$ and $\vec{k} \in \mathbb{R}^m$. Thus, because $f(\Phi(\vec{y})) = \vec{0}$ for all $\vec{y} \in W$, we know that $f'(\Phi(\vec{y}))\Phi'(\vec{y}) = \mathbf{0}$ by chain rule. And when $\vec{y} = \vec{b}$, this becomes $\mathbf{A}\Phi'(\vec{b}) = \mathbf{0}$. So for all $\vec{k} \in \mathbb{R}^m$, we have that:

$$\mathbf{A}_x g'(\vec{b})\vec{k} + \mathbf{A}_y \vec{k} = \mathbf{A}(g'(\vec{b})\vec{k}, \vec{k}) = \mathbf{A}\Phi'(\vec{b})\vec{k} = \vec{0}.$$

It follows that $\mathbf{A}_x g'(\vec{b}) + \mathbf{A}_y = \mathbf{0}$. Or in other words, $g'(\vec{b}) = -\mathbf{A}_x^{-1}\mathbf{A}_y$. ■

Note that it is really easy to check if $\mathbf{A}_x$ is invertible.

First, let $e_1, \ldots, e_n$ and $u_1, \ldots, u_m$ be the standard bases of $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively. Next, fix $\{(e_1, \vec{0}), \ldots, (e_n, \vec{0}), (\vec{0}, u_1), \ldots, (\vec{0}, u_m)\}$ as our basis for $\mathbb{R}^{n+m}$ and $\{e_1, \ldots, e_n\}$ as our basis for $\mathbb{R}^n$. Then we have that:

$$[\mathbf{A}] = \left[\begin{array}{c|c} [\mathbf{A}_x] & [\mathbf{A}_y] \end{array}\right]$$

So finally, check if $[\mathbf{A}_x]$ is invertible or not.

Let $E \subseteq \mathbb{R}^n$ be open and suppose that $f : E \longrightarrow \mathbb{R}$ has the partial derivatives $D_1 f, \ldots, D_n f$. Then note that each partial derivative is itself a function from $\mathbb{R}^n$ to $\mathbb{R}$. So, if each $D_j f$ is itself differentiable, then we define the second-order partial derivatives of f as $D_{i,j} = D_i(D_j f)$.

Additionally, if $D_{i,j} f$ is continuous in E for all $1 \leq i, j \leq n$, then we say that $f \in \mathscr{C}^2(E)$ (which is an abbreviation of the notation: $f \in \mathscr{C}^2(E, \mathbb{R})$).

Note that we're only limiting ourselves to non-vector-valued functions here for the sake of simplicity. Also, hopefully you can guess from this definition how an nth-order partial derivative is defined, and what it means for a function to be a $\mathscr{C}^n$-mapping.

Theorem 9.41: Let $f : E \longrightarrow \mathbb{R}$ where $E \subseteq \mathbb{R}^n$ is open. If $f \in \mathscr{C}^2(E)$, then $(D_{i,j}f)(\vec{x}) = (D_{j,i}f)(\vec{x})$ for all $1 \leq i, j \leq n$ and $\vec{x} \in E$.

Proof:
Let $e_1, \ldots, e_n$ be the standard basis of $\mathbb{R}^n$ and let $Q \subset E$ be a closed rectangle with opposite vertices at $\vec{x}$ and $\vec{x} + he_i + ke_j$ where $h > 0$ and $k > 0$. Then define:

$$\Delta(f, Q) = f(\vec{x} + he_i + ke_j) - f(\vec{x} + he_j) - f(\vec{x} + ke_j) + f(\vec{x})$$

Next, let $u(s) = f(\vec{x} + se_i + ke_j) - f(\vec{x} + se_i)$. Importantly, u is real function that is differentiable on $[0, h]$. So:

$$\begin{aligned}\Delta(f, Q) &= u(h) - u(0) \\ &= hu'(s_1) \quad \text{for some } s_1 \in (0, h) \quad \text{(by the mean value theorem)} \\ &= h((D_i f)(\vec{x} + s_1 e_i + ke_j) - (D_i f)(\vec{x} + s_1 e_i))\end{aligned}$$

Similarly, define $v(t) = (D_i f)(\vec{x} + s_1 e_i + te_j) - (D_i f)(\vec{x} + s_1 e_i)$. Then $(D_i f)(\vec{x} + s_1 e_i + ke_j) - (D_i f)(\vec{x} + s_1 e_i) = v(k) - v(0)$ and v is a real function that is differentiable on $[0, k]$. So:

$$\begin{aligned}\Delta(f, Q) &= h(v(k) - v(0)) \\ &= h(kv'(t_1)) \quad \text{for some } t_1 \in (0, k) \quad \text{(by the mean value theorem)} \\ &= hk((D_{j,i}f)(\vec{x} + s_1 e_i + t_1 e_j)).\end{aligned}$$

Now importantly, we can use the exact same reasoning to say that $\Delta(f, Q) = hk((D_{i,j}f)(\vec{x} + s_2 e_i + t_2 e_j))$ for some $s_2 \in (0, h)$ and $t_2 \in (0, k)$.

So finally, consider that because $D_{j,i}f$ and $D_{i,j}f$ are continuous, we know that for all $\varepsilon > 0$, there exists $\delta_1, \delta_2 > 0$ such that:

- $\sqrt{s_1^2 + t_1^2} < \delta_1 \implies |(D_{j,i}f)(\vec{x} + s_1 e_i + t_1 e_j) - (D_{j,i}f)(\vec{x})| < \varepsilon/2$
- $\sqrt{s_2^2 + t_2^2} < \delta_2 \implies |(D_{i,j}f)(\vec{x} + s_2 e_i + t_2 e_j) - (D_{i,j}f)(\vec{x})| < \varepsilon/2$

Thus, set $\delta = \min(\delta_1, \delta_2)$ and pick any $h, k > 0$ such that $h^2 < \frac{\delta^2}{2}$ and $k^2 < \frac{\delta^2}{2}$. Then as $0 < s_1, s_2 < h$ and $0 < t_1, t_2 < k$, we know that both $\sqrt{s_1^2 + t_1^2} < \delta_1$ and $\sqrt{s_2^2 + t_2^2} < \delta_2$. So:

$$\begin{aligned}&|(D_{j,i}f)(\vec{x}) - (D_{i,j}f)(\vec{x})| \\ &\quad \leq \left|\tfrac{\Delta(f,Q)}{hk} - (D_{j,i}f)(\vec{x})\right| + \left|\tfrac{\Delta(f,Q)}{hk} - (D_{i,j}f)(\vec{x})\right| \\ &\quad = |(D_{j,i}f)(\vec{x} + s_1 e_i + t_1 e_j) - (D_{j,i}f)(\vec{x})| \\ &\qquad\qquad + |(D_{i,j}f)(\vec{x} + s_2 e_i + t_2 e_j) - (D_{i,j}f)(\vec{x})| \\ &\quad < \varepsilon/2 + \varepsilon/2 = \varepsilon\end{aligned}$$

And since ε was arbitary, we conclude that $(D_{j,i}f)(\vec{x}) = (D_{i,j}f)(\vec{x})$.
(Note that Rudin proves something slightly stronger in his book.)

Exercise 9.9 If f is a differentiable mapping of a connected open set $E \subseteq \mathbb{R}^n$ into $\mathbb{R}^m$ such that $f'(\vec{x}) = \mathbf{0}$ for every $\vec{x} \in E$, then f is constant on E.

Pick any $\vec{x}_0 \in E$ and define the sets $A = \{\vec{x} \in E \mid f(\vec{x}) = f(\vec{x}_0)\}$, and $B = E \setminus A$. Clearly $A \cup B = E$. Also, $A \neq \emptyset$ as $\vec{x}_0 \in A$. Therefore, since E is connected we know that if $B \neq \emptyset$, then either $\overline{A} \cap B \neq \emptyset$ or $A \cap \overline{B} \neq \emptyset$.

First, assume $A \cap \overline{B} \neq \emptyset$ and let $\vec{x} \in A \cap \overline{B}$. Because E is open, there exists $r > 0$ such that $B_r(\vec{x}) \subseteq E$. Additionally, $B_r(\vec{x})$ is convex. Therefore, we know by the corollary to theorem 9.19 that $\forall \vec{y} \in B_r(\vec{x}),\ \ f(\vec{y}) = f(\vec{x}) = f(\vec{x}_0)$. But then we have a contradiction because $B_r(\vec{x}) \cap B = \emptyset$ and thus $\vec{x} \notin \overline{B}$. So, we conclude that $A \cap \overline{B} = \emptyset$.

Secondly, assume $\overline{A} \cap B \neq \emptyset$ and let $\vec{x} \in \overline{A} \cap B$. Once again, there exists $r > 0$ such that $B_r(\vec{x}) \subseteq E$. So, by a similar argument as above, we can show that $B_r(\vec{x}) \cap A = \emptyset$, thus contradicting that $\vec{x} \in \overline{A}$. As a result, we conclude that $\overline{A} \cap B = \emptyset$.

Hence, by the contrapositive of the definition of connectedness, we conclude that $B = \emptyset$, meaning that f is constant on E.

Lecture 8: 4/30/2024

Let X be a vector space. We call $\mathbf{P} \in L(X)$ a projection if $\mathbf{P}^2 = \mathbf{P}$
i.e: $\forall \vec{x} \in X,\ \ \mathbf{P}\mathbf{P}\vec{x} = \mathbf{P}\vec{x}$

Here are some properties of projections:

(A) If $\mathbf{P}$ is a projection, then $\forall \vec{x} \in X,\ \ \exists \vec{x}_1, \vec{x}_2 \in \mathbf{X}\ \ s.t.\ \ \vec{x} = \vec{x}_1 + \vec{x}_2$, $\vec{x}_1 \in \mathscr{R}(\mathbf{P})$, and $\vec{x}_2 \in \mathscr{N}(\mathbf{P})$.

Proof:
Obviously, we have that $\mathbf{P}\vec{x} \in \mathscr{R}(\mathbf{P})$. Meanwhile:
$\mathbf{P}(\vec{x} - \mathbf{P}\vec{x}) = \mathbf{P}\vec{x} - \mathbf{P}\mathbf{P}\vec{x} = \mathbf{P}\vec{x} - \mathbf{P}\vec{x} = \vec{0}$. So,
$(\vec{x} - \mathbf{P}\vec{x}) \in \mathscr{N}(\mathbf{P})$ and $\vec{x} = (\mathbf{P}\vec{x}) + (\vec{x} - \mathbf{P}\vec{x})$.

(C) If $\mathbf{P}$ is a projection and $\vec{x} \in \mathscr{R}(\mathbf{P})$, then $\vec{x} = \mathbf{P}\vec{x}$.

Proof:
Suppose $\vec{x} \in \mathscr{R}(\mathbf{P})$. Then there exists $\vec{y}$ such that $\mathbf{P}\vec{y} = \vec{x}$, and in turn this means that $\vec{x} = \mathbf{P}\vec{y} = \mathbf{P}\mathbf{P}\vec{y} = \mathbf{P}\vec{x}$.

(C) If X is finite dimensioned and X_1 is a vector subspace of X, then there exists a projection $\mathbf{P} : \mathbf{X} \longrightarrow \mathbf{X}$ such that $\mathscr{R}(\mathbf{P}) = X_1$.

Proof:
Let $\dim X = n$. If $X_1 = \{\vec{0}\}$, this statement is trivial. Just set $\mathbf{P} = \mathbf{0}$. So now assume that $k = \dim(X_1) \neq 0$. Then by theorem 9.3.C, we can find a basis $\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n\}$ of X such that $\{\vec{x}_1, \ldots, \vec{x}_k\}$ is a basis of X_1.

Having done that, for any $\vec{y} \in X$ there are unique scalars $c_1, \ldots, c_n$ such that $\vec{y} = c_1\vec{x}_1 + \ldots + c_k\vec{x}_k + \ldots + c_n\vec{x}_n$. So, assign $\mathbf{P}\vec{y} = c_1\vec{x}_1 + \ldots + c_k\vec{x}_k$.

Clearly, $\mathbf{P}$ is a linear operator with $\mathscr{R}(\mathbf{P}) = \text{span}\{x_1, \ldots, x_k\} = X_1$. Also, $\vec{y} \in X_1 \Longrightarrow \vec{y} = \mathbf{P}\vec{y}$. Thus, $\mathbf{P}$ is a projection since $\mathbf{P}\vec{x} = \mathbf{P}(\mathbf{P}\vec{x})$ for all $\vec{x} \in X$.

With this background, we now present the following theorem:

The Rank Theorem:
Suppose that m, n, r are nonnegative integers with $m, n \geq r$, that $E \subseteq \mathbb{R}^n$ is open, that $F \in \mathscr{C}^1(E, \mathbb{R}^m)$, and that $F'(\vec{x})$ has rank r for every $\vec{x} \in E$. Fix $\vec{a} \in E$, and set $\mathbf{A} = F'(\vec{a})$. Next, having set $Y_1 = \mathscr{R}(\mathbf{A})$, let $\mathbf{P} \in L(\mathbb{R}^m)$ be a projection with $\mathscr{R}(\mathbf{P}) = Y_1$ and let $Y_2 = \mathscr{N}(\mathbf{P})$. Then:

There exist open sets U and V in $\mathbb{R}^n$ satisfying that $\vec{a} \in U \subseteq E$, that $\mathbf{A}(V)$ is open relative to Y_1, and that there exist $H \in \mathscr{C}^1(V, U)$ with a $\mathscr{C}^1$ inverse and $\varphi \in \mathscr{C}^1(\mathbf{A}(V), Y_2)$ such that $F(H(\vec{x})) = \mathbf{A}\vec{x} + \varphi(\mathbf{A}\vec{x})$.

Proof:
If $r = 0$, then letting $R > 0$ be such that $B_r(\vec{a}) \subseteq E$, we know that f is constant on $B_R(\vec{a})$. Thus, set $V = U = B_R(\vec{a})$, $H(\vec{x}) = \vec{x}$, and $\varphi(\vec{0}) = F(\vec{a})$. Clearly, $H \in \mathscr{C}^1(V, U)$ and is its own inverse. Also, while it doesn't make sense to say that $\varphi \in \mathscr{C}^1(\mathbf{A}(V), Y_2)$ since $\mathbf{A}(V) = \{\vec{0}\}$ has no limit points, we do have that since $\mathbf{A}(V) = Y_1$, trivially $\mathbf{A}(V)$ is open relative to Y_1. And since $Y_2 = \mathbb{R}^m$, we know that $\phi(\vec{0}) \in Y_2$. Finally, we have for all $\vec{x} \in V$ that:

$$F(H(\vec{x})) = F(\vec{x}) = F(\vec{a}) = \vec{0} + \varphi(\mathbf{A}\vec{x}) = \mathbf{A}\vec{x} + \varphi(\mathbf{A}\vec{x})$$

So, this theorem is "true" trivially when $r = 0$.

Now we shall assume that $r > 0$ for the rest of this proof. Let $\{\vec{y}_1, \ldots, \vec{y}_r\}$ be a basis for Y_1 and choose a $\vec{z}_i$ for each $1 \leq i \leq r$ such that $\mathbf{A}\vec{z}_i = \vec{y}_i$. Then let $\mathbf{S} \in L(Y_1, \mathbb{R}^n)$ be the linear map $\mathbf{S}(c_1\vec{y}_1 + \ldots + c_r\vec{y}_r) = c_1\vec{z}_1 + \ldots + c_r\vec{z}_r$.

Note that $\mathbf{AS}\vec{y}_i = \mathbf{A}\vec{z}_i = \vec{y}_i$ for all $i \in \{1, \ldots, r\}$.

Next, define $G(\vec{x}) = \vec{x} + \mathbf{SP}(F(\vec{x}) - \mathbf{A}\vec{x})$ for all $\vec{x} \in E$. Then note that G is a $\mathscr{C}^1$-map with $G'(\vec{a}) = \mathbf{I} + \mathbf{SP}(\mathbf{A} - \mathbf{A}) = \mathbf{I}$. So by the inverse function theorem, there are open sets U_1 and V_1 in $\mathbb{R}^n$ with a $\vec{a} \in U$ such that G is a one-to-one mapping of U_1 onto V_1 whose inverse H is also a $\mathscr{C}^1$-map.

(Also note that $H'(\vec{x}) = G'(H(\vec{x}))^{-1}$ is invertible for all $\vec{x} \in V_1$.)

Let $\vec{b} = G(\vec{a})$ and let $V \subseteq V_1$ be a convex open ball around $\vec{b}$.
By theorem 4.8 (proposition 69 in 140A notes), we know that $U = G^{-1}(V) = H(V)$ is open. Plus, $\vec{a} \in U$. Thus, restricting G and H to U and V, we still have that U, V, G, and H satisfy the properties guarenteed by the inverse function theorem. Only now, V is also convex.

Also, we can show that $\mathbf{A}(V)$ is open relative to Y_1 as follows:

Let $\vec{y}_0 \in \mathbf{A}(V)$ and pick $\vec{x}_0$ such that $\mathbf{A}\vec{x}_0 = \vec{y}_0$. Next pick $\varepsilon > 0$ such that $B_\varepsilon(\vec{x}_0) \subseteq V$ and then set $\delta = \frac{\varepsilon}{2\|\mathbf{S}\|}$. Now for any $\vec{y} \in B_\delta(\vec{y}_0) \cap Y_1$, we have that:

- $\vec{x}_0 + \mathbf{S}(\vec{y} - \vec{y}_0) \in B_\varepsilon(\vec{x}_0) \subseteq V$
- $\vec{y} = \mathbf{A}(\vec{x}_0 + \mathbf{S}(\vec{y} - \vec{y}_0)) \subseteq \mathbf{A}(V)$

So, $B_\delta(\vec{y}_0) \cap Y_1 \subseteq \mathbf{A}(V)$.

Meanwhile, notice that $\mathbf{ASP}(\mathbf{A}\vec{x}) = \mathbf{AS}(\mathbf{A}\vec{x}) = \mathbf{A}\vec{x}$ for all $\vec{x} \in \mathbb{R}^n$. So, $\mathbf{A}(G(\vec{x})) = \mathbf{P}F(\vec{x})$ for all $\vec{x} \in E$

Proof:
$\mathbf{A}(G(\vec{x})) = \mathbf{A}\vec{x} + \mathbf{ASP}(F(\vec{x}) - \mathbf{A}\vec{x}) = \mathbf{AS}(\mathbf{P}F(\vec{x})) + \mathbf{A}\vec{x} - \mathbf{A}\vec{x} = \mathbf{P}F(\vec{x})$

In turn, we can then replace $\vec{x}$ with $H(\vec{x})$ and say that for all $\vec{x} \in V$, $\mathbf{P}F(H(\vec{x})) = \mathbf{A}(G(H(\vec{x}))) = \mathbf{A}\vec{x}$.

Then, having defined $\Psi(\vec{x}) = F(H(\vec{x})) - \mathbf{A}\vec{x}$, note that for all $\vec{x} \in V$, $\mathbf{P}\Psi(\vec{x}) = \mathbf{P}F(H(\vec{x})) - \mathbf{A}\vec{x} = \vec{0}$. Also, $\Psi'(\vec{x}) = F'(H(\vec{x}))H'(\vec{x}) - \mathbf{A}$ is continuous. So, $\Psi \in \mathscr{C}^1(V, P_2)$.

Finally, all that's left to do is show the existence of a $\mathscr{C}^1$-mapping φ from $\mathbf{A}(V)$ to Y_2 satisfying that $\varphi(\mathbf{A}\vec{x}) = \Psi(\vec{x})$ for all $\vec{x} \in V$. To start, we first show that if $\vec{x}_1 \in V$, $\vec{x}_2 \in V$, and $\mathbf{A}\vec{x}_1 = \mathbf{A}\vec{x}_2$, then $\Psi(\vec{x}_1) = \Psi(\vec{x}_2)$.

Put $\Phi(\vec{x}) = F(H(\vec{x}))$ for all $\vec{x} \in V$. Then, $\Phi'(\vec{x}) = F'(H(\vec{x}))H'(\vec{x})$.
Also, since $\mathscr{R}(H'(\vec{x})) = \mathbb{R}^n$ for all $\vec{x} \in V$, we know that:

$$\mathscr{R}(F'(H(\vec{x}))H'(\vec{x})) = \mathscr{R}(F'(H(\vec{x}))).$$

Thus, $\mathrm{rk}(\Phi'(\vec{x})) = \mathrm{rk}(F'(H(\vec{x}))) = r$ for all $\vec{x} \in V$.

Next, fix $\vec{x} \in V$ and set $M = \mathscr{R}(\Phi'(\vec{x}))$. Then because $\mathbf{P}\Phi'(\vec{x})\,\vec{v} = \mathbf{A}\,\vec{v}$ for all $\vec{v} \in \mathbb{R}^m$, we know that $\mathbf{P}$ maps M surjectively onto P_1. In turn, by the rank-nullity theorem and the fact that $\dim(M) = \dim(P_1)$, we know that $\mathbf{P}$ restricted to M is injective.

Now suppose $\mathbf{A}\,\vec{h} = \vec{0}$. Then we know that $\mathbf{P}\Phi'(\vec{x})\,\vec{h} = \vec{0}$, and since $\mathbf{P}$ is one-to-one on M, we have that $\Phi'(\vec{x})\,\vec{h} = \vec{0}$. This proves that $\mathbf{A}\,\vec{h} = \vec{0}$ implies that for all $\vec{x} \in V$:

$$\Psi'(\vec{x})\,\vec{h} = F'(H(\vec{x}))H'(\vec{x})\,\vec{h} - \mathbf{A}\,\vec{h} = \Phi'(\vec{x})\,\vec{h} - \mathbf{A}\,\vec{h} = \vec{0}.$$

Finally, assume $\vec{x}_1 \in V$, $\vec{x}_2 \in V$, and $\mathbf{A}\,\vec{x}_1 = \mathbf{A}\,\vec{x}_2$. Then set $\vec{h} = \vec{x}_2 - \vec{x}_1$ and define $g(t) = \Psi(\vec{x}_1 + t\,\vec{h})$ for all $h \in [0, 1]$ (note that g is well defined because V is convex).

Importantly, $\mathbf{A}\,\vec{h} = \vec{0}$. So, $g'(t) = \Psi'(\vec{x}_1 + t\,\vec{h})\,\vec{h} = \vec{0}$ for all t and that means that $g(0) = g(1)$. But $g(0) = \Psi(\vec{x}_1)$ and $g(1) = \Psi(\vec{x}_2)$. Hence, $\Psi(\vec{x}_1) = \Psi(\vec{x}_2)$.

Therefore, we are able to umambiguously define $\varphi : V(A) \longrightarrow Y_2$ by the expression $\varphi(\mathbf{A}\,\vec{x}) = \Psi(\vec{x})$. Furthermore, we can show that φ is $\mathscr{C}^1$.

Fix $\vec{y}_0 \in \mathbf{A}(V)$ and set $\vec{x}_0 \in V$ such that $\mathbf{A}\,\vec{x}_0 = \vec{y}_0$. Next, let $W \subseteq \mathbf{A}(V)$ be open relative to Y_1 such that for any $\vec{y} \in W$, $\vec{x} = \vec{x}_0 + \mathbf{S}(\vec{y} - \vec{y}_0) \in V$.

Then because $\vec{y} = \mathbf{A}(\vec{x}_0 + \mathbf{S}(\vec{y} - \vec{y}_0))$ for all $\vec{y} \in W$, we have that $\varphi(\vec{y}) = \Psi(\vec{x}_0 - \mathbf{S}(\vec{y}_0 - \vec{y}))$. Hence, φ is continuously differentiable with respect to $\vec{y} \in W$. And since W and $\vec{y}_0$ was arbitrary, we have that φ is continuously differentiable on $\mathbf{A}(V)$. ■

But what does this mean?

Letting $\vec{y} \in F(U)$, we know there exists $\vec{x} \in V$ such that $\vec{y} = F(H(\vec{x}))$. In turn, we can find that $\mathbf{P}\,\vec{y} = \mathbf{P}\mathbf{A}\,\vec{x} + \mathbf{P}\varphi(\mathbf{A}\,\vec{x}) = \mathbf{A}\,\vec{x} + \vec{0}$ and that $\vec{y} = (\mathbf{P}\,\vec{y}) + \varphi(\mathbf{P}\,\vec{y})$.

This essentially means we can view $F(U)$ as the graph φ over $\mathbf{P}(F(U))$.

I'm behind and so am going to avoid going further in depth on this. However, this will be useful if you study differential geometry and manifolds.

To finish off this chapter (I'm skipping the segment on determinants), here is a theorem about differentiation of integrals.

Theorem 9.42: Let $\varphi(x,t)$ be a real-valued function whose partial derivative with respect to t is defined for all $a \leq x \leq b$ and $c \leq t \leq d$. Also, let us define $\varphi^t(x) = \varphi(x,t)$ for each $t \in [c,d]$. Then suppose:

(A) $\alpha : [a,b] \longrightarrow \mathbb{R}$ is monotone increasing.

(B) $\varphi^t \in \mathscr{R}_a^b(\alpha)$ for every $t \in [c,d]$

(C) $c < s < d$ and to every $\varepsilon > 0$, there corresponds $\delta > 0$ such that $|(D_2\varphi)(x,t) - (D_2\varphi)(x,s)| < \varepsilon$ for all $x \in [a,b]$ and $t \in (s-\delta, s+\delta)$.

Define $f(t) = \int_a^b \varphi(x,t)d\alpha(x)$. Then $(D_2\varphi)(x,s) = (D_2\varphi)^s(x) \in \mathscr{R}_a^b(\alpha)$, $f'(s)$ exists, and $f'(s) = \int_a^b (D_2\varphi)(x,s)d\alpha(x)$.

Proof:
Let $\varepsilon > 0$ and pick a corresponding $\delta > 0$ by assumption (C).

Define $\psi(x,t) = \frac{\varphi(x,t)-\varphi(x,s)}{t-s}$ for all $x \in [a,b]$ and $t \in B_\delta(s) \setminus \{s\}$. Then note that by the mean value theorem, for all $t \neq s$ and $x \in [a,b]$ there exists u between s and t such that $\psi(x,t) = (D_2\varphi)(x,u)$. Hence, we have that:

$$|\psi(x,t) - (D_2\varphi)(x,s)| < \varepsilon \text{ for all } x \in [a,b] \text{ and } t \in B_\delta(s) \setminus \{s\}.$$

This tells us that $\lim_{t\to s} \psi(x,t) = (D_2\varphi)(x,s)$ for all $x \in [a,b]$. In fact, $\psi(x,t)$ converges uniformly since the same δ works for all x.

So by theorem 7.16 (proposition 120 in the 140B notes), we know that:

$$\lim_{t\to s} \int_a^b \psi(x,t)d\alpha(x) = \int_a^b \lim_{t\to s} \psi(x,t)d\alpha(x) = \int_a^b (D_2\varphi)(x,s)d\alpha(x).$$

Also, note that $\frac{f(t)-f(s)}{t-s} = \int_a^b \psi(x,t)d\alpha(x)$. Thus:

$$f'(s) = \int_a^b (D_2\varphi)(x,s)d\alpha(x).$$

Exercise 9.26: The existence and continuity of $D_{1,2}f$ does not imply the existence of $D_1 f$.

Let $g : \mathbb{R} \longrightarrow \mathbb{R}$ be everywhere continuous but nowhere differentiable on $\mathbb{R}$. Then define $f(x,y) = g(x)$ for all $(x,y) \in \mathbb{R}^2$.

Clearly, $D_1 f$ doesn't exist because g' doesn't exist. But, $(D_2 f)(x,y) = 0$ for all $(x,y) \in \mathbb{R}^2$, and in turn that means that $(D_{1,2}f)(x,y) = 0$ for all $\mathbb{R}^2$.

Exercise 9.27: Put $f(0,0)=0$ and $f(x,y)=\frac{xy(x^2-y^2)}{x^2+y^2}$ when $(x,y)\neq(0,0)$. Then:

(A) f, D_1f, and D_2f are all continuous in $\mathbb{R}^2$:

When $(x,y)\neq(0,0)$, we have that:

$$(D_1f)(x,y)=\frac{yx^4-y^5+4x^2y^3}{(x^2+y^2)^2} \qquad (D_2f)(x,y)=\frac{-xy^4+x^5-4x^3y^2}{(x^2+y^2)^2}$$

Meanwhile, when $(x,y)=(0,0)$, we have that:

$$(D_1f)(0,0)=\lim_{x\to0}\frac{0}{x}=0=\lim_{y\to0}\frac{0}{y}=(D_2f)(0,0)$$

Also note that:

$$\begin{aligned}0\leq\left|\frac{yx^4-y^5+4x^2y^3}{(x^2+y^2)^2}\right|&=\left|\frac{y(x^4+2x^2y^2-y^4)+2x^2y^3}{x^4+2x^2y^2+y^4}\right|\\&\leq|y|\left|\frac{x^4+2x^2y^2-y^4}{x^4+2x^2y^2+y^4}+\frac{2x^2y^2}{x^4+2x^2y^2+y^4}\right|\\&\leq|y|\left(\left|\frac{x^4+2x^2y^2-y^4}{x^4+2x^2y^2+y^4}\right|+\left|\frac{2x^2y^2}{x^4+2x^2y^2+y^4}\right|\right)\leq|y|(1+1)=2|y|\end{aligned}$$

$$\begin{aligned}0\leq\left|\frac{-xy^4+x^5-4x^3y^2}{(x^2+y^2)^2}\right|&=\left|\frac{x(x^4-2x^2y^2-y^4-2x^2y^2)}{x^4+2x^2y^2+y^4}\right|\\&\leq|x|\left|\frac{x^4-2x^2y^2-y^4}{x^4+2x^2y^2+y^4}-\frac{2x^2y^2}{x^4+2x^2y^2+y^4}\right|\\&\leq|x|\left(\left|\frac{x^4-2x^2y^2-y^4}{x^4+2x^2y^2+y^4}\right|+\left|\frac{2x^2y^2}{x^4+2x^2y^2+y^4}\right|\right)\leq|x|(1+1)=2|x|\end{aligned}$$

Thus, $\lim_{(x,y)\to(0,0)}(D_1f)(x,y)=0=\lim_{(x,y)\to(0,0)}(D_2f)(x,y)$.

And this means that D_1f and D_2f are both continuous in $\mathbb{R}^2$. Also, since both partial derivatives are continuous, we know that $f\in\mathscr{C}^1(\mathbb{R}^2)$. So, f is continuous.

(C) $(D_{1,2})f(0,0)=1$ but $(D_{2,1})f(0,0)=-1$.

This is because $(D_1f)(0,y)=-y$ and $D_2f(x,0)=x$.

(B) $D_{1,2}f$ and $D_{2,1}f$ are continuous at every point of $\mathbb{R}^2$ except $(0,0)$.

When $(x,y)\neq(0,0)$, we have that:

$$(D_{1,2}f)(x,y)=\frac{x^6+9x^4y^2-9x^2y^4-y^6}{(x^2+y^2)^3}=(D_{2,1}f)(x,y)$$

So clearly, $D_{1,2}f$ and $D_{2,1}f$ are continuous at $(x,y)\neq(0,0)$. But consider approaching $(0,0)$ along the path $\gamma(t)=(t,t)$. Then:

$$\lim_{t\to0}(D_{1,2}f)(\gamma(t))=\frac{0}{8t^6}=0=\lim_{t\to0}(D_{2,1}f)(\gamma(t)).$$

This shows that $D_{1,2}f$ and $D_{2,1}f$ have a discontinuity at $(0,0)$.

Exercise 9.28: For $t \geq 0$, define $\varphi(x,t) = \begin{cases} x & \text{if } 0 \leq x \leq \sqrt{t} \\ -x + 2\sqrt{t} & \text{if } \sqrt{t} \leq x \leq 2\sqrt{t} \\ 0 & \text{otherwise} \end{cases}$
Also let $\varphi(x,t) = -\varphi(x,-t)$ if $t < 0$. Then $(D_2\varphi)(x,0) = 0$ for all x.

Fix any $x \neq 0$. Then, there exists $t_0 > 0$ such that $t_0 < \frac{1}{4}x^2$. So, for all $t \in (0,t_0)$, $\varphi(x,t) = 0$. Similarly, for all $t \in (-t_0,0)$, $\varphi(x,t) = 0$. Hence, $(D_2\varphi)(x,0) = 0$.

Meanwhile, if $x = 0$, then $\varphi(x,t) = x = 0$ for all $t \in \mathbb{R}$. So, $(D_2\varphi)(0,0) = 0$.

Now define $f(t) = \int_{-1}^{1} \varphi(x,t)dx$. Then for $|t| < \frac{1}{4}$, we have that $f(t) = t$.

First, consider any $0 < t < \frac{1}{4}$. Then:

$$f(t) = \int_0^{\sqrt{t}} x dx + \int_{\sqrt{t}}^{2\sqrt{t}} (-x + 2\sqrt{t})dx$$
$$= [\tfrac{1}{2}x^2]_0^{\sqrt{t}} + [-\tfrac{1}{2}x^2 + 2\sqrt{t}x]_{\sqrt{t}}^{2\sqrt{t}} = \tfrac{1}{2}t + -2t + 4t + \tfrac{1}{2}t - 2t = t$$

Secondly, consider any $-\frac{1}{4} < t < 0$. Then:

$$f(t) = \int_{-\sqrt{t}}^{0} -x dx + \int_{-2\sqrt{t}}^{-\sqrt{t}} (x - 2\sqrt{t})dx$$
$$= [-\tfrac{1}{2}x^2]_{\sqrt{t}}^{0} + [\tfrac{1}{2}x^2 - 2\sqrt{t}x]_{-2\sqrt{t}}^{-\sqrt{t}} = \tfrac{1}{2}t + \tfrac{1}{2}t + 2t + 2t - 4t = t$$

Finally, $f(0) = \int_{-1}^{1} 0dx = 0$. So clearly, $f'(0) = 1 \neq 0 = \int_{-1}^{1} (D_2\varphi)(x,0)dx$.

Exercise 9.29: Let E be an open set in $\mathbb{R}^n$. Inductively, we can say that $f \in \mathscr{C}^k(E)$ if all the partial derivatives: $D_1 f, \ldots, D_n f$ of f belong to $\mathscr{C}^{k-1}(E)$. Now assume $f \in \mathscr{C}^k(E)$. Then the kth-order partial derivative $D_{i_1,i_2,\ldots,i_k} f = D_{i_1} D_{i_2} \ldots D_{i_k} f$ is unchanged if the subscripts $i_1, \ldots, i_k$ are permuted.

By theorem 9.41, we know this is true for $k = 2$. So, we now proceed by induction on k.

Assume $f \in \mathscr{C}^k(E)$ and consider any kth-order partial derivative $D_{i_1,i_2,\ldots,i_k} f$. Then, consider any permutation $\sigma \in S_k$ and assume the claim of the problem holds for smaller k.

Observation 1: Equality is maintained if the permutation doesn't effect i_1.

Suppose $\pi \in S_k$ such that $\pi(1) = 1$. Then:

$$D_{i_1,i_2,\ldots,i_k} f = D_{i_1}\left(D_{i_2,\ldots,i_k} f\right) = D_{i_{\pi(1)}}\left(D_{i_{\pi(2)},\ldots,i_{\pi(k)}} f\right) = D_{i_{\pi(1)} i_{\pi(2)},\ldots,i_{\pi(k)}} f$$

Observation 2: We can definitely swap i_1 with i_j so long as $j \neq k$.

If $j \neq 1$ and $j \neq k$, then by induction:

$$D_{i_1,i_2,\ldots,i_k} f = D_{i_1,\ldots,i_j}\left(D_{i_{j+1},\ldots,i_k} f\right) = D_{i_j i_2\ldots,i_{j-1} i_1}\left(D_{i_{j+1},\ldots,i_k} f\right) = D_{i_j i_2,\ldots,i_{j-1} i_1 i_{j+1}\ldots i_k} f$$

Observation 3: We can definitely swap i_{k-1} and i_k.

By theorem 9.41:

$$D_{i_1,\ldots,i_{k-2} i_{k-1} i_k} f = D_{i_1,\ldots,i_{k-2}}(D_{i_{k-1} i_k} f) = D_{i_1,\ldots,i_{k-2}}(D_{i_k i_{k-1}} f) = D_{i_1,\ldots,i_{k-2} i_k i_{k-1}} f$$

Let $\tau_{i,j}$ be the permutation swapping i and j and leaving everything else the same.

If $\sigma(1) \neq k$, then we can say that $\sigma = \pi \circ \tau_{1,j}$ where π is a permutation such that $\pi(1) = 1$.

If $\sigma(1) = k$, then we can say that $\sigma = \pi \circ \tau_{1,k-1} \circ \tau_{k-1,k}$ where π is a permutation such that $\pi(1) = 1$.

In either case, by combining the three observations above we can say for any permutation $\sigma \in S_k$ that $D_{i_1,i_2,\ldots,i_k} f = D_{i_{\sigma(1)} i_{\sigma(2)},\ldots,i_{\sigma(k)}} f$.

Lecture 9: 5/2/2024

A note on notation:

- If A and B are any two sets, we shall write $A - B$ to mean the set $\{x \in A \mid x \notin B\}$ (which we've previously referred to as $A \setminus B$).
- We call a family of sets $(A_\alpha)_{\alpha \in A}$ pairwise disjoint if $A_\alpha \cap A_\beta = \emptyset$ for all $\alpha, \beta \in A$ such that $\alpha \neq \beta$.

A family of sets $\mathcal{R}$ is called a ring if $\forall A, B \in \mathcal{R}$, $A \cup B \in \mathcal{R}$ and $A - B \in \mathcal{R}$.

Since $A \cap B = A - (A - B)$, we also have that $A \cap B \in \mathcal{R}$.

A ring $\mathcal{R}$ is called a σ-ring if whenever $A_n \in \mathcal{R}$ for each $n \in \{1, 2, \ldots\}$, then:

$$\bigcup_{n=1}^{\infty} A_n \in \mathcal{R}.$$

And since $\bigcap_{n=1}^{\infty} A_n = A_1 - \bigcap_{n=1}^{\infty} (A_1 - A_n)$, we know $\bigcap_{n=1}^{\infty} A_n \in \mathcal{R}$ if $\mathcal{R}$ is a σ-ring.

Let $\mathcal{R}$ be a ring and $\phi : \mathcal{R} \longrightarrow \mathbb{R} \cup \{-\infty, \infty\}$. We say ϕ is additive if $A \cap B = \emptyset$ implies that $\phi(A \cup B) = \phi(A) + \phi(B)$. Also, if $\mathcal{R}$ is a σ-ring, we say ϕ is countably additive if $A_i \cap A_j = \emptyset$ for all $i \neq j$ implies that:

$$\phi\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \phi(A_n).$$

Some notes:

1. While our definition allows for the range of ϕ to include $+\infty$ or $-\infty$, we shall always assume that that is not the case.

2. In the formula $\phi\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \phi(A_n)$, we can arrange the A_n in any order on the left-side. This tells us that $\sum \phi(A_n)$ must converge absolutely (if at all) since rearranging terms does not change the series' value.

If ϕ is additive, then it satisfies the following properties:

- $\phi(\emptyset) = 0$.
  Proof:
  $\phi(A) = \phi(\emptyset \cup A) = \phi(\emptyset) + \phi(A)$ since $\emptyset \cap A = \emptyset$. Thus $\phi(\emptyset) = 0$.

- $\phi(A_1 \cup \ldots \cup A_n) = \phi(A_1) + \ldots + \phi(A_n)$ if $A_i \cap A_j = \emptyset$ whenever $i \neq j$.

- $\phi(A_1 \cup A_2) + \phi(A_1 \cap A_2) = \phi(A_1) + \phi(A_2)$.
  This is true because $\phi(A_1 \cup A_2) = \phi(A_1) + \phi(A_2 - A_1)$ and $\phi(A_2) = \phi(A_2 - A_1) + \phi(A_2 \cap A_1)$.

- If $\phi(A) \geq 0$ for all $A \in \mathcal{R}$ and $A_1 \subseteq A_2$, then $\phi(A_1) \leq \phi(A_2)$.
  This is true because $\phi(A_2) = \phi(A_1) + \phi(A_2 - A_1) \geq \phi(A_1)$.
  For this reason we call a nonnegative additive ϕ monotonic.

- $\phi(A - B) = \phi(A) - \phi(B)$ if $B \subseteq A$ and $|\phi(B)| < +\infty$.

Theorem 11.3: Suppose that ϕ is countably additive on a ring $\mathcal{R}$, that $A_n \in \mathcal{R}$ for all $n \in \mathbb{N}$, that $A_1 \subseteq A_2 \subseteq \ldots$, and that $A = \bigcup_{n=1}^{\infty} A_n$. Then $\phi(A_n) \to \phi(A)$ as $n \to \infty$.

Proof:
Put $B_1 = A_1$ and $B_n = A_n - A_{n-1}$ for all $n \in \{2, 3, \ldots\}$. Then $B_i \cap B_j = \emptyset$ for all $i \neq j$, $A_n = \bigcup_{k=1}^{n} B_k$, and $A = \bigcup_{n=1}^{\infty} B_n$.

As a result, we know that $\phi(A_n) = \sum_{k=1}^{n} \phi(B_k)$ and $\phi(A) = \sum_{k=1}^{\infty} \phi(B_k)$.

And so, by the definition of an infinite series, $\phi(A_n) \to \phi(A)$.

We define an interval I in $\mathbb{R}^p$ as the set:

$$I = \{(x_1, \ldots, x_p) \mid a_i \leq x_i \leq b_i, \ \ \forall i \in \{1, \ldots, p\}\}$$

Note: we can also makes any of the inequalities in this definition strict. Also, a_i is allowed to be equal to b_i.

Also, if $A \subseteq \mathbb{R}^p$ is the union of a finite number of intervals, we say A is an elementary set.

Lemma: Suppose that I and J are intervals. Then $I - J$ is the union of a finite number of disjoint intervals.

Proof:
To start, let's write $I = |a_1, b_1| \times |a_2, b_2| \times \ldots \times |a_n, b_n|$ and $J = |c_1, d_1| \times |c_2, d_2| \times \ldots \times |c_n, d_n|$ where "|" is a placeholder for either "(" or "[". Also, let's write "||" to flip the strictness of an inequality of an interval.

Now let $\mathcal{I}$ be a list of intervals. Then, for each $i \in \{1, \ldots, j\}$, add the intervals below to $\mathcal{I}$:

- $|\min(c_1, b_1),\ \max(a_1, d_1)| \times \ldots \times |\min(c_{i-1},\ b_{i-1}), \max(a_{i-1},\ d_{i-1})| \times |a_i,\ \min(c_i, b_i)\| \times |a_{i+1}, b_{i+1}| \times \ldots \times |a_j, b_j|$
- $|\min(c_1, b_1),\ \max(a_1, d_1)| \times \ldots \times |\min(c_{i-1},\ b_{i-1}), \max(a_{i-1},\ d_{i-1})| \times \|\max(a_i, d_i),\ b_i| \times |a_{i+1}, b_{i+1}| \times \ldots \times |a_j, b_j|$

You're hopefully smart enough to figure out what the intervals added to $\mathcal{I}$ in step 1 look like. Also, it's ok if any intervals added to $\mathcal{I}$ are empty.

Then $I - J = \bigcup_{l \in \mathcal{I}} l$ and all the sets in $\mathcal{I}$ are disjoint to each other.

Now, let $\mathcal{E}$ denote the family of all elementary subsets of $\mathbb{R}^p$.

Claim 1 (homework): $\mathcal{E}$ is a ring but not a σ-ring.

First, let's show that $\mathcal{E}$ is a ring.

Let $A, B \in \mathcal{E}$. Then we know that $A = \bigcup_{i=1}^{m} I_i$ and $B = \bigcup_{i=m+1}^{m+n} I_i$ where each I_i is an interval.

So immediately, $A \cup B = \bigcup_{i=1}^{m+n} I_i \in \mathcal{E}$.

Meanwhile, because $(X \cup Y) - Z = (X - Z) \cup (Y - Z)$ and $X - (Y \cup Z) = (X - Y) - Z$ for all sets X, Y, Z, we can apply the prior lemma a bajillion times to get that $A - B$ is the union of m (not necessarily disjoint) unions of bajillions of disjoint intervals. So, $A - B \in \mathcal{E}$.

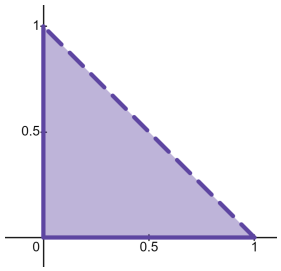
Second, let's show that $\mathcal{E}$ is not a σ-ring.

For a basic example in $\mathbb{R}$, consider the set $\bigcup_{n=1}^{\infty}(2n-2, 2n]$.

Clearly, that cannot be rewritten as a finite union of intervals.

For a more interesting original example, consider the set $T = \bigcup_{q \in \mathbb{Q} \cap [0,1]} [0, q) \times [0, 1-q)$

Then T is the union of countably many intervals. But T also consists of all the points in the triangle to the right minus its longest edge.



Now let $\mathcal{I}$ be any set of intervals contained in T. By expanding all the intervals when possible, we can assume without loss of generality that all intervals in $\mathcal{I}$ have the form $[0, t) \times [0, 1-t)$.

Then, to show that the sets in $\mathcal{I}$ do not cover all of T, note that the union of $\mathcal{I}$ will "look" like a staircase. So, you can find points "between" the steps not covered by any sets in $\mathcal{I}$ but which are in T.

I'm stressed on time so I'm not going to go into more details.

Claim 2 (homework): If $A \in \mathcal{E}$, then A is the union of a finite collection of pairwise disjoint intervals.

Proof:
Since $A \in \mathcal{E}$, we know $A = \bigcup_{i=1}^{n} I_i$ where each I_i is an interval.

Then note that $A = \bigcup_{i=1}^{n} \left(I_i - \bigcup_{j=i+1}^{n} I_j \right)$ where each of the $I_i - \bigcup_{j=i+1}^{n} I_j$ are pairwise disjoint.

Also, by repeatedly applying our prior lemma, $I_i - \bigcup_{j=i+1}^{n} I_j$ can be expressed as the union of finitely many disjoint intervals.

Next, given any interval $I = [a_1, b_1] \times \ldots [a_p, b_p]$, let $m(I) = \prod_{i=1}^{p} (b_i - a_i)$.

Also, if any of the inequalities defining I are made strict, we still define m the same.

Then given any $A \in \mathcal{E}$, if $\mathcal{I}$ is a collection of intervals that are pairwise disjoint such that the union of the intervals in $\mathcal{I}$ equals A, then we define:

$$m(A) = \sum_{I \in \mathcal{I}} m(I)$$

Claim 3 (homework): Let $\mathcal{I}$ and $\mathcal{J}$ both be collections of pairwise disjoint intervals such that $\bigcup_{I \in \mathcal{I}} I = A = \bigcup_{J \in \mathcal{J}} J$. Then $\sum_{I \in \mathcal{I}} m(I) = \sum_{J \in \mathcal{J}} m(J)$.

Proof:
Let $\mathcal{I} = \{I_1, \ldots, I_n\}$ and $\mathcal{J} = \{J_1, \ldots, J_m\}$ Then since $I_i \cap J_j \in \mathcal{E}$ for all $1 \leq i \leq n$ and $1 \leq j \leq m$ and because $I_{i_1} \cap J_{j_1}$ and $I_{i_2} \cap J_{j_2}$ are disjoint whenever $i_1 \neq i_2$ or $j_1 \neq j_2$, we can define a collection $\mathcal{K}$ satisfying that:

- $\mathcal{K}$ is a collection of pairwise disjoint intervals.
- Each I_i and J_j is the union of a subcollection of $\mathcal{K}$.
- $\bigcup_{l \in \mathcal{K}} l = A$.

Now if $l_1 \cup \ldots \cup l_k = I_i$, then clearly $m(l_1) + \ldots + m(l_k) = m(I_i)$.
To prove this, you can just use the distributive property.

So, by rearranging terms in the sum, we can say that:

$$\sum_{I \in \mathcal{I}} m(I) = \sum_{l \in \mathcal{K}} m(l) = \sum_{J \in \mathcal{J}} m(J).$$

This tells us that m is a well defined function on $\mathcal{E}$.

Claim 4 (homework): m is additive.

Proof:
Suppose A and B are disjoint and $\mathcal{I}$ and $\mathcal{J}$ are collections of pairwise disjoint intervals such that $\bigcup_{I \in \mathcal{I}} I = A$ and $\bigcup_{J \in \mathcal{J}} J = B$.

Then $\mathcal{I} \cap \mathcal{J} = \emptyset$ and $\mathcal{I} \cup \mathcal{J}$ is pairwise disjoint with $\bigcup_{l \in \mathcal{I} \cup \mathcal{J}} l = A \cup B$.
So, we can say that:

$$m(A \cup B) = \sum_{l \in \mathcal{I} \cup \mathcal{J}} m(l) = \sum_{I \in \mathcal{I}} m(I) + \sum_{J \in \mathcal{J}} m(J) = m(A) + m(B)$$

A nonnegative additive set function ϕ defined on $\mathcal{E}$ is said to be regular if:

For all $A \in \mathcal{E}$ and $\varepsilon > 0$, there exists $F, G \in \mathcal{E}$ such that F is closed, G is open, $F \subseteq A \subseteq G$, and $\phi(G) - \varepsilon \leq \phi(A) \leq \phi(F) + \varepsilon$.

In other words, $\phi(G) \leq \phi(A) + \varepsilon$ and $\phi(F) \geq \phi(A) - \varepsilon$.

Examples:

- Hopefully it is obvious to you that m is regular.
- Let $\mathbb{R}^p = \mathbb{R}$ and $\alpha : \mathbb{R} \longrightarrow \mathbb{R}$ be a monotonically increasing function, then put for all intervals in $\mathbb{R}$:

$$\mu([a,b)) = \alpha(b-) - \alpha(a-)$$
$$\mu([a,b]) = \alpha(b+) - \alpha(a-)$$
$$\mu((a,b]) = \alpha(b+) - \alpha(a+)$$
$$\mu((a,b)) = \alpha(b-) - \alpha(a+)$$

Also, extend this μ to all of $\mathcal{E}$ the same way we extended m. Then μ is regular (although we won't prove that).

Let μ be additive, regular, and finite on $\mathcal{E}$. Then for any $E \subset \mathbb{R}^p$ define:

$$\mu^*(E) = \inf \left\{ \sum_{n=1}^{\infty} \mu(A_n) \;\middle|\; E \subset \bigcup_{n=1}^{\infty} A_n, \text{ where each } A_n \in \mathcal{E} \text{ and is open} \right\}$$

Then, $\mu^* : \mathcal{P}(\mathbb{R}^p) \longrightarrow [0,\infty]$ is called the outer measure of E corresponding to μ (where $\mathcal{P}(\mathbb{R}^p)$ is the power set of $\mathbb{R}^p$).

Remark 1: You can set all but finitely many $A_n = \emptyset$ in order to emulate E being a subset of a finite open cover of A_n.

Remark 2: If $E_1 \subset E_2$, then clearly $\mu^*(E_1) \leq \mu^*(E_2)$.

Theorem 11.8:

(A) For every $A \in \mathcal{E}$, $\mu^*(A) = \mu(A)$.

Proof:
Let $A \in \mathcal{E}$ and $\varepsilon > 0$. Then because μ is regular, there exists an open set $G \supseteq A$ such that $\mu(G) \leq \mu(A) + \varepsilon$. Also, $\mu^*(A) \leq \mu(G)$ as $A \subseteq G \cup \emptyset \cup \emptyset \cup \ldots$. Thus, we know that $\mu^*(A) \leq \mu(A) + \varepsilon$.

Meanwhile, since $\mu^*(A)$ is by definition an infimum of a set, there exists a sequence (A_n) of open sets in $\mathcal{E}$ whose union contains A such that:

$$\sum_{n=1}^{\infty} \mu(A_n) \leq \mu^*(A) + \varepsilon.$$

Also, by the regularity of μ there exists a closed set F such that $\mu(F) \geq \mu(A) - \varepsilon$. Importantly, F is compact and (A_n) is an open cover of F. So $F \subset A_1 \cup \ldots \cup A_N$ for some N and hence:

$$\begin{aligned} \mu(A) \leq \mu(F) + \varepsilon &\leq \mu(A_1 \cup \ldots \cup A_N) + \varepsilon \\ &\leq \sum_{n=1}^{N} \mu(A_n) + \varepsilon \leq \sum_{n=1}^{\infty} \mu(A_n) + \varepsilon \leq \mu^*(A) + 2\varepsilon \end{aligned}$$

Since ε was arbitrary, we thus have that $\mu^*(A) = \mu(A)$.

(B) If $E = \bigcup_{n=1}^{\infty} E_n$, then $\mu^*(E) \leq \sum_{n=1}^{\infty} \mu^*(E_n)$.

This is called σ-subadditivity.

Proof:
Now suppose $E = \bigcup_{n=1}^{\infty} E_n$.

If $\mu^*(E_n) = +\infty$ for any n, then this statement is trivially true. So, assume $\mu^*(E) < +\infty$ for all n. Then for any $\varepsilon > 0$, there are coverings $(A_{n,k})$ of E_n by open elementary sets such that:

$$\sum_{k=1}^{\infty} \mu(A_{n,k}) \leq \mu^*(E_n) + 2^{-n}\varepsilon$$

Then because all the $(A_{n,k})$'s form an open cover for E, we have that:

$$\mu^*(E) \leq \sum_{n=1}^{\infty}\sum_{k=1}^{\infty} \mu(A_{n,k}) \leq \sum_{n=1}^{\infty} \mu^*(E_n) + \tfrac{1}{2}\varepsilon \cdot \tfrac{1}{1-\frac{1}{2}} = \sum_{n=1}^{\infty} \mu^*(E_n) + \varepsilon$$

Lecture 10: 5/7/2024

For any $A, B \subseteq \mathbb{R}^p$ and μ^* defined as in last lecture, we define:

- $S(A,B) = (A - B) \cup (B - A)$ (This is the symmetric difference of A and B.)
- $d(A,B) = \mu^*(S(A,B))$.

Also, given any sequence (A_n) of sets, we write $A_n \to A$ if $\lim_{n\to\infty} d(A_n, A) = 0$.

If there is a sequence (A_n) of elementary sets such that $A_n \to A$, then we say A is finitely μ-measurable and write $A \in \mathfrak{M}_F(\mu)$.

If A is the union of countably many finitely μ-measurable sets, we say that A is μ-measurable and write $A \in \mathfrak{M}_F(\mu)$.

Some properties of S:

(A) $S(A,B) = S(B,A)$ and $S(A,A) = \emptyset$.
Hopefully this is just obvious to you.

(B) $S(A,B) \subseteq S(A,C) \cup S(C,B)$.
This is because $(A - B) \subseteq (A - C) \cup (C - B)$ and $(B - A) \subseteq (C - A) \cup (B - C)$.

(C) $S(A^C, B^C) = S(A, B)$.

This is because:

$$S(A^C, B^C) = (A^C \cap B) \cup (B^C \cap A)$$
$$= (A^C \cup B^C) \cap (A \cup B) = (A \cup B) - (A \cap B)$$
$$= (A - B) \cup (B - A) = S(A, B)$$

(D) $S(A_1 \cup A_2,\ \ B_1 \cup B_2) \subseteq S(A_1, B_1) \cup S(A_2, B_2)$.

This is because:

$$(A_1 \cup A_2) - (B_1 \cup B_2) = (A_1 - (B_1 \cup B_2)) \cup (A_2 - (B_1 \cup B_2))$$
$$\subseteq (A_1 - B_1) \cup (A_2 - B_2)$$

(E) $S(A_1 \cap A_2,\ \ B_1 \cap B_2) \subseteq S(A_1, B_1) \cup S(A_2, B_2)$.

This is because:

$$S(A_1 \cap A_2,\ \ B_1 \cap B_2) = S(A_1^C \cup A_2^C,\ \ B_1^C \cup B_2^C)$$
$$\subseteq S(A_1^C, B_1^C) \cup S(A_2^C, B_2^C) = S(A_1, B_1) \cup S(A_2, B_2)$$

(F) $S(A_1 - A_2,\ \ B_1 - B_2) \subseteq S(A_1, B_1) \cup S(A_2, B_2)$.

This is because:

$$S(A_1 - A_2,\ \ B_1 - B_2) = S(A_1 \cap A_2^C,\ \ B_1 \cap B_2^C)$$
$$\subseteq S(A_1, B_1) \cup S(A_2^C, B_2^C) = S(A_1, B_1) \cup S(A_2, B_2)$$

Some properties of d: (Homework problem 5:2)

(1) $d(A, B) = d(B, A)$ and $d(A, A) = 0$.

The first identity is true because $S(A, B) = S(B, A)$. So $\mu^*(S(A, B)) = \mu^*(S(B, A))$.

The second identity is true because $S(A, A) = \emptyset$ and $\emptyset \in \mathcal{E}$. So, $d(A, A) = \mu^*(S(A, A)) = \mu^*(\emptyset) = \mu(\emptyset) = 0$.

(2) $d(A, B) \leq d(A, C) + d(C, B)$

Proof:

Because $S(A, B) \subseteq S(A, C) \cup S(C, B)$ and μ^* is subadditive, we have that:

$$d(A, B) = \mu^*(S(A, B)) \leq \mu^*(S(A, C) \cup S(C, B))$$
$$\leq \mu^*(S(A, C)) + \mu^*(S(C, B)) = d(A, C) + d(C, B)$$

(3) $d(A_1 \cup A_2, B_1 \cup B_2) \leq d(A_1, B_1) + d(A_2, B_2)$
$d(A_1 \cap A_2, B_1 \cap B_2) \leq d(A_1, B_1) + d(A_2, B_2)$
$d(A_1 - A_2, B_1 - B_2) \leq d(A_1, B_1) + d(A_2, B_2)$

The proof of these inequalities are identical to the proof of the previous property. Just use the subadditivity of μ^* and the fact that:

- $S(A_1 \cup A_2,\ \ B_1 \cup B_2) \subseteq S(A_1, B_1) \cup S(A_2, B_2)$
- $S(A_1 \cap A_2,\ \ B_1 \cap B_2) \subseteq S(A_1, B_1) \cup S(A_2, B_2)$
- $S(A_1 - A_2,\ \ B_1 - B_2) \subseteq S(A_1, B_1) \cup S(A_2, B_2)$

(Not part of homework problem 5:2)

(4) If either $\mu^*(A)$ or $\mu^*(B)$ is finite, then $|\mu^*(A) - \mu^*(B)| \leq d(A,B)$.

Proof:
Suppose $0 \leq \mu^*(B) \leq \mu^*(A)$ where $\mu^*(B)$ is finite. Then $\mu^*(A) = d(A,\emptyset) \leq d(A,B) + d(B,\emptyset) = d(A,B) + \mu^*(B)$. So, because $\mu^*(B)$ is finite, we know that:

$$\mu^*(A) - \mu^*(B) = |\mu^*(A) - \mu^*(B)| \leq d(A,B)$$

Meanwhile, if $0 \leq \mu^*(A) \leq \mu^*(B)$ and $\mu^*(B)$ is still finite, then note that $d(B,\emptyset) \leq d(B,A) + d(A,\emptyset)$. So, because $d(B,A) = d(A,B)$, we have that:

$$\mu^*(B) - \mu^*(A) = |\mu^*(A) - \mu^*(B)| \leq d(A,B)$$

So the identity is true if $\mu^*(B)$ is finite. Needless to say, the same reasoning works for when $\mu^*(A)$ is finite.

Now it is worth noting that d isn't quite a metric on $\mathcal{P}(\mathbb{R}^p)$ since $d(A,B) = 0$ does not imply that $A = B$.

To prove this, set $\mu = m$ and suppose that $A \in \mathcal{E}$ and $B = A \cup \{\vec{p}\,\}$ where $\vec{p} \in \mathbb{R}^p \setminus A$. Then $d(A,B) = m^*(S(A,B)) = m^*(\{\vec{p}\,\}) = 0$.

That said, if we define the equivalence relation: $A \sim B \iff d(A,B) = 0$, then the set $\mathcal{P}(\mathbb{R}^p)/_\sim$ of equivalence classes of $\sim$ is a metric space when equipped with d. Then in that metric space, we can view $\mathfrak{M}_F(\mu)$ as the closure of $\mathcal{E}$.

Theorem 11.10: $\mathfrak{M}(\mu)$ is a σ-ring and μ^* is countably additive on $\mathfrak{M}(\mu)$.

Proof:

Part 1: Showing $\mathfrak{M}_F(\mu)$ is a ring and μ^* is additive on $\mathfrak{M}_F(\mu)$

Suppose $A, B \in \mathfrak{M}_F(\mu)$. Then choose sequences (A_n) and (B_n) such that each $A_n, B_n \in \mathcal{E}$, $A_n \to A$, and $B_n \to B$.

Observe that:

- $A_n \cup B_n \to A \cup B$
- $A_n \cap B_n \to A \cap B$
- $A_n - B_n \to A - B$

To prove the first statement here, note that:
$0 \leq d(A_n \cup B_n,\ A \cup B)$
$\leq d(A_n, A) + d(B_n, B) \to 0$

Likewise, the other statements are proven similarly.

This proves that $A, B \in \mathfrak{M}_F(\mu) \implies A \cup B,\ A - B \in \mathfrak{M}_F(\mu)$. So, $\mathfrak{M}_F(\mu)$ is a ring.

Also, since $d(A_n, A) \to 0$, we must have that $\mu^*(A)$ is finite. So, $\mu^*(A_n) \to \mu^*(A)$ since $|\mu^*(A_n) - \mu^*(A)| \leq d(A_n, A)$. Similarly, we know that $\mu^*(B_n) \to \mu^*(B)$, $\mu^*(A_n \cup B_n) \to \mu^*(A \cup B)$ and etc.

Thus, because $\mu(A_n) + \mu(B_n) = \mu(A_n \cup B_n) + \mu(A_n \cap B_n)$ for all n (see page 41), we can conclude by letting $n \to \infty$ that:

$$\mu^*(A) + \mu^*(B) = \mu^*(A \cup B) + \mu^*(A \cap B)$$

In turn, because $\mu^*(\emptyset) = \mu(\emptyset) = 0$, we can conclude that if $A \cap B = \emptyset$, then $\mu^*(A) + \mu^*(B) = \mu^*(A \cup B)$. Hence, μ^* is additive on $\mathfrak{M}_F(\mu)$.

Part 2: Showing that μ^* is countably additive on $\mathfrak{M}(\mu)$

Lemma 1: If $A = \bigcup_{n=1}^{\infty} A_n \in \mathfrak{M}(\mu)$ with each $A_n \in \mathfrak{M}_F(\mu)$ and pairwise disjoint, then $\mu^*(A) = \sum_{n=1}^{\infty} \mu^*(A_n)$

Suppose $A \in \mathfrak{M}(\mu)$. Then, A equals the union of a countable collection (A_n) of sets in $\mathfrak{M}_F(\mu)$. Without loss of generality, we can also assume that (A_n) is pairwise disjoint.

To see why this is, assume $A'_n \in \mathfrak{M}_F(\mu)$ for all n and $A = \bigcup_{n=1}^{\infty} A'_n$.
Then set $A_1 = A'_1$ and $A_n = \bigcup_{k=1}^{n} A'_k - \bigcup_{k=1}^{n-1} A'_k$.

Then, (A_n) is pairwise disjoint, each $A_n \in \mathfrak{M}_F(\mu)$, and $A = \bigcup_{n=1}^{\infty} A_n$.

By theorem 11.8.B, we know that $\mu^*(A) \leq \sum_{n=1}^{\infty} \mu^*(A_n)$.

On the other hand, because $A \supset A_1 \cup \ldots \cup A_n$ for all $n \in \mathbb{N}$ and μ^* is additive on $\mathfrak{M}_F(\mu)$, we have that for all n:

$$\mu^*(A) \geq \mu^*(A_1 \cup \ldots \cup A_n) = \mu^*(A_1) + \ldots + \mu^*(A_n).$$

This means $\mu^*(A) \geq \sum_{n=1}^{\infty} \mu^*(A_n)$. And so, the lemma is proved.

Lemma 2: $A \in \mathfrak{M}(\mu)$ and $\mu^*(A) < \infty \implies A \in \mathfrak{M}_F(\mu)$.

Continue letting (A_n) be a countable collection of disjoint sets in $\mathfrak{M}_F(\mu)$ such that $A = \bigcup A_n$ and suppose $\mu^*(A)$ is finite. Then define $B_n = A_1 \cup \ldots \cup B_n$ and note that:

$$0 \leq d(A, B_n) = \mu^*(\bigcup_{i=n+1}^{\infty} A_i) \leq \sum_{i=n+1}^{\infty} \mu^*(A_i) \to 0 \text{ as } n \to 0.$$

Because $B_n \to A$ and each $B_n \in \mathfrak{M}_F(\mu)$, we know that $A \in \mathfrak{M}_F(\mu)$.

And now, it is clear that μ^* is countably additive on $\mathfrak{M}(\mu)$.

Suppose $A = \bigcup A_n$ where (A_n) is a pairwise disjoint sequence of sets in $\mathfrak{M}(\mu)$.

If all $\mu^*(A_n)$ are finite, then each $A_n \in \mathfrak{M}_F(\mu)$ by lemma 2. So, we know by lemma 1 that $\mu^*(A) = \sum_{n=1}^{\infty} \mu^*(A_n)$.

Meanwhile, if any $\mu^*(A_n)$ is infinite, then we trivially have that $\sum_{n=1}^{\infty} \mu^*(A_n) = \infty$ and that $\mu^*(A) = \infty$ because $A_n \subseteq A$.

Part 3: Showing that $\mathfrak{M}(\mu)$ is a σ-ring

Suppose $A_n \in \mathfrak{M}(\mu)$ for all $n \in \mathbb{N}$. Then because the union of countably many sets is still countable, we know that:

$$\bigcup_{n=1}^{\infty} A_n \in \mathfrak{M}(\mu)$$

Meanwhile, suppose that $A, B \in \mathfrak{M}(\mu)$ and that (A_n) and (B_n) are collections of sets in $\mathfrak{M}_F(\mu)$ such that $A = \bigcup_{n=1}^{\infty} A_n$ and $B = \bigcup_{n=1}^{\infty} B_n$.

Since $A_n \cap B = \bigcup_{i=1}^{\infty} (A_n \cap B_i)$, we know that $A_n \cap B \in \mathfrak{M}(\mu)$ for all n.

Furthermore, since $A_n \cap B \subseteq A_n$ and $\mu^*(A_n)$ is finite because $A_n \in \mathfrak{M}_F(\mu)$ (we showed this in part 1 of the proof), we know that $\mu^*(A_n \cap B)$ is finite. Hence, $A_n \cap B \in \mathfrak{M}_F(\mu)$.

It follows that $A - B \in \mathfrak{M}_\mu$ because $A - B = \bigcup_{n=1}^{\infty} (A_n - B)$. ■

The significance of this theorem is that given any additive, regular, and finite set function μ defined on $\mathcal{E}$, we know that it's outer measure is a countably additive set function on the σ-ring: $\mathfrak{M}(\mu)$.

More succinctly, we can (and will) just refer to μ^* as μ since μ^* essentially extends the definition of μ from $\mathcal{E}$ to all of $\mathfrak{M}(\mu)$. This extended set function is called a measure. And when $\mu = m$, then this extended function is called the Lebesgue measure on $\mathbb{R}^p$.

Homework 5.3: A σ-ring $\mathcal{R}$ of subsets of a set X is either finite or has uncountably many elements.

If X is finite, than the largest possible set of subsets of X is also finite. So, we can assume that X is infinite.

Now, assume that $\mathcal{R}$ is countably infinite, meaning that $\mathcal{R} = \{A_1, A_2, \ldots\}$. Then for all $x \in X$, we can say that $B_x = \bigcap_{x \in A_i} A_i \in \mathcal{R}$. In turn, this means that $\forall x, y \in X, \quad B_x \cap B_y \in \mathcal{R}$.

Suppose $z \in B_x \cap B_y$. Then $B_z \subseteq B_x \cap B_y$. Also, we can prove by contradiction that $x, y \in B_Z$.

If $x \notin B_z$, then $x \in B_x - B_z$ and $z \in B_x \subseteq (B_x - B_z) \not\ni z$.
The same thing happens if $y \notin B_z$.

Therefore, we would have that $B_x, B_y \subseteq B_z$. And so, $B_x = B_z = B_y$.

Finally, let $\mathcal{F} = \{B_x \mid x \in X\}$. Since $\mathcal{F}$ is a subset of $\mathcal{R}$, we know that $\mathcal{F}$ can't be uncountable. Meanwhile, we also can't have that $\mathcal{F}$ is finite. This is because for any $A \in \mathcal{R}$, we know that $B_x \subseteq A$ for all $x \in A$ and $A = \bigcup_{x \in A} B_x$. So if $\mathcal{F}$ is finite, then there can only be finitely many A in $\mathcal{R}$.

Thirdly, $\mathcal{F}$ can't be countable. To see why, assume $\mathcal{F} = \{B_{x_1}, B_{x_2}, \ldots\}$. Then for any $S \subseteq \mathbb{N}$, we have that $\bigcup_{i \in S} B_{x_i} \in \mathcal{R}$.

But also, because $\mathcal{F}$ is pairwise disjoint, if $S \neq T \subseteq \mathbb{N}$, then $\bigcup_{i \in S} B_{x_i} \neq \bigcup_{i \in T} B_{x_i}$. Hence, there is an injective map from $\mathcal{P}(\mathbb{N})$ to $\mathcal{R}$.

So, we've shown a contradiction if $\mathcal{R}$ is countable.

Let $\mathcal{E}$ be the family of elementary subsets of $\mathbb{R}$ and denote $m : \mathcal{E} \to [0, \infty)$ the Lebesgue set function and by $\mu^*(A)$ the outer measure of a set $A \subseteq \mathbb{R}$. Also, for a scalar $t \in \mathbb{R}$, let the set $A + t = \{x + t \mid x \in A\}$.

1. If $A \in \mathcal{E}$, then $m(A + t) = m(A)$ for all $t \in \mathbb{R}$.

   If $A \in \mathcal{E}$, then there exists disjoint intervals $I_1, I_2, \ldots, I_n$ such that $A = I_1 \cup I_2 \cup \ldots \cup I_n$. Then, $m(A) = m(I_1) + m(I_2) + \ldots + m(I_n)$.

   Now note that $A + t = (I_1 + t) \cup (I_2 + t) \cup \ldots \cup (I_n + t)$. Also, $(I_i + t) \cap (I_j + t) = \emptyset$ for all $i \neq j$, meaning that:

   $$m(A + t) = m(I_1 + t) + m(I_2 + t) + \ldots + m(I_n + t).$$

   But clearly, $m(I_i + t) = m(I_i)$ for all $1 \leq i \leq n$. So, we conclude that $m(A) = m(A + t)$.

2. If $A \in \mathbb{R}$, then $\mu^*(A + t) = \mu^*(A)$ for all $t \in \mathbb{R}$.

   Let $\varepsilon > 0$. Then by the definition of $\mu^*(A)$, there exists a sequence (A_n) of open sets from $\mathcal{E}$ such that:

   $$A \subseteq \bigcup_{n=1}^{\infty} A_n \text{ and } \mu^*(A) \leq \sum_{n=1}^{\infty} m(A_n) \leq \mu^*(A) + \varepsilon.$$

Importantly, $(A_n + t)$ is also an open cover of $A + t$. So, we know that:

$$\mu^*(A+t) \leq \sum_{n=1}^{\infty} m(A_n + t) = \sum_{n=1}^{\infty} m(A_n) \leq \mu^*(A) + \varepsilon$$

Meanwhile, by the definition of $\mu^*(A+t)$ there exists a sequence (B_n) of open sets from $\mathcal{E}$ such that:

$$(A+t) \subseteq \bigcup_{n=1}^{\infty} B_n \text{ and } \mu^*(A+t) \leq \sum_{n=1}^{\infty} m(B_n) \leq \mu^*(A+t) + \varepsilon.$$

Then as $(B_n + (-t))$ is also an open cover of A, we have that:

$$\mu^*(A) \leq \sum_{n=1}^{\infty} m(B_n + (-t)) = \sum_{n=1}^{\infty} m(B_n) \leq \mu^*(A+t) + \varepsilon$$

And as ε was arbitrary, we're now done.

Some notes:

(a) Every open set in $\mathbb{R}^p$ is the union of a countable collection of open intervals (look at exercise 3.23 which we did for homework in math 140A). Thus if A is open, then $A \in \mathfrak{M}(\mu)$.

By taking complements, we also get that every closed set is in $\mathfrak{M}(\mu)$. This is because if A is closed, then $A = \mathbb{R}^k - A^{\mathsf{C}} \in \mathfrak{M}(\mu)$.

(b) If $A \in \mathfrak{M}(\mu)$ and $\varepsilon > 0$, then there exists a closed set F and open set G such that $F \subseteq A \subseteq G$, $\mu(G - A) < \varepsilon$, and $\mu(A - F) < \varepsilon$.

The existence of G satisfying this claim is immediate from our definition of $\mu = \mu^*$