Math 140C Lecture Notes (Professor: Luca Spolaor)

Isabelle Mills

April 21, 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. $\forall \vec{v} \in X,\ \ c_1, \ldots, c_k \in \mathbb{R}\ \ s.t.\ \ \vec{v} = \sum_{i=1}^{k} c_i\vec{x}_i$

   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 $c_{n+1} \neq 0$. Thus, 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 transformations 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:
  ($\Longrightarrow$) 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 $\operatorname{null}(\mathbf{A}) = 0$.

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

  ($\Longleftarrow$) 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 $\vec{x} = c_1 e_1 + \ldots + c_n e_n$.

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

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 $\sum\limits_{i=1}^{n} \|\mathbf{A}(e_i)\|$ 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 transformation 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, $\mathrm{null}(\mathbf{B}) = 0$ 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:
$$\|\mathbf{B}^{-1}\| = \|\mathbf{B}^{-1}(\vec{y})\| \leq \frac{\|\vec{y}\|}{\frac{1}{\|\mathbf{A}^{-1}\|} - \|\mathbf{B}-\mathbf{A}\|} \leq \frac{1}{\frac{1}{\|\mathbf{A}^{-1}\|} - \|\mathbf{B}-\mathbf{A}\|} = \frac{\|\mathbf{A}^{-1}\|}{1-\|\mathbf{A}^{-1}\|\|\mathbf{B}-\mathbf{A}\|}$$

<u>Lemma</u>: 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:
$$\begin{aligned} 0 \leq \|\mathbf{B}^{-1} - \mathbf{A}^{-1}\| &= \|\mathbf{B}^{-1}(\mathbf{A}-\mathbf{B})\mathbf{A}^{-1}\| \\ &\leq \|\mathbf{B}^{-1}\|\|(\mathbf{A}-\mathbf{B})\|\|\mathbf{A}^{-1}\| \leq \frac{\|\mathbf{A}^{-1}\|^2}{1-\|\mathbf{A}^{-1}\|\|\mathbf{B}-\mathbf{A}\|}\|\mathbf{B}-\mathbf{A}\| \end{aligned}$$

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

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 $\vec{x} = \sum\limits_{j=1}^{n} c_j\, \vec{x}_j$. 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:

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

<u>Fact 2</u>: 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 $\varphi([\mathbf{A}]) = \mathbf{A}$.

<u>Fact 3</u>: 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:

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

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 $\mathscr{N}(\mathbf{0}) = X$. 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:

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

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

So, let's fix $\vec{h}_0 \in \mathbb{R}^n \setminus \{\vec{0}\}$. Then we know that $\lim\limits_{t\to 0} \frac{\|(\mathbf{A}_2 - \mathbf{A}_1)t\vec{h}_0\|}{\|t\vec{h}_0\|} = 0$.

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 $\lim\limits_{t\to 0} \frac{\|(\mathbf{A}_2 - \mathbf{A}_1)t\vec{h}_0\|}{\|t\vec{h}_0\|} = \frac{\|(\mathbf{A}_2 - \mathbf{A}_1)\vec{h}_0\|}{\|\vec{h}_0\|}$ 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 $\lim\limits_{t\to 0} \frac{f(\vec{x}+te_j)-f(\vec{x})}{t} = f'(\vec{x})(e_j)$.

Next, consider that:

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

It immediately follows that $f'(\vec{x})e_j = \sum\limits_{i=1}^{k}(D_j f_i)(x)u_i$.

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 $\gamma(t) = (\gamma_1(t), \dots, \gamma_n(t))$ 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 $\{e_1, \dots, e_n\}$ 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:

$$D_{\vec{u}} f(\vec{x}) = \lim_{t \to 0} f(\gamma(t)) = \nabla f(\vec{x}) \cdot \vec{u}$$

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 $f(\vec{x}) = (f_1(\vec{x}), \dots, f_m(\vec{x}))$, then using the standard bases for $\mathbb{R}^n$ and $\mathbb{R}^m$, we can abuse notation and say that:

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

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 $\|f'(x)\| \leq M$ for every $x \in E$. 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 $f \in \mathscr{C}^1(E)$.