Analysis, Topology, Optimization, Machine Learning, and Computational Analysis

Daniel Drake

June 29, 2020

1 Notation, Set Theory, and Logic

Definition 1.1. ***Sets, Set Builder Notation and Hats***
A set is a collection of objects.
Example: A = The set of all hats.
We call the objects in the set elements.
Example: A Cowboy hat is a type of hat and thus belongs in the set of all hats: A.

Set builder notation is a way of describing a set using mathematical, logical symbols, or words. Look at the following example:

$$E = \{x \in \mathbb{N} : x = 2n \ where \ n \in \mathbb{N}\}$$

This reads: E is the set of all x in $\mathbb{N}$ such that $x = 2n$ where n is in $\mathbb{N}$
This set is also known as the even numbers.
When talking about functions, another common way of describing a set is:

$$C_X^Y = \{f : X \to Y | f \ is \ continuous\}$$

This reads: C_X^Y is the set of all functions f mapping from X to Y such that f is continuous.

Definition 1.2. ***Common Sets of Numbers***

$$\mathbb{N} = \{1, 2, ...\}$$
$$\mathbb{N}_0 = \{0, 1, 2, ...\}$$
$$\mathbb{N}_m = \{1, 2, ..., m\} \ where \ m \in \mathbb{N}$$
$$\mathbb{R} \ is \ the \ set \ of \ Real \ Numbers$$
$$\mathbb{R}^+ = \{x \in \mathbb{R} : x > 0\}$$
$$\mathbb{R}_0^+ = \{x \in \mathbb{R} : x \geq 0\}$$
$$\mathbb{R}^- = \{x \in \mathbb{R} : x < 0\}$$
$$\mathbb{R}_0^- = \{x \in \mathbb{R} : x \leq 0\}$$

Definition 1.3. _**Primitives and Symbols**_
The logical or and not are both primitives and are written:

$$logical\ or:\ \vee$$

and

$$not:\ \neg$$

Respectively.
Statements are written: $L, M, N, O, P, Q, ...etc$
A statement is a sequence of words or symbols which is either true or false.
So then $L \vee M$ is a new statement composed of $L, \vee,$ and M.
We can then use this as the definition of a new statement:

$$N := L \vee M$$

Which is read: N is defined as L or M.
So N is true if:

- _L is true._

- _M is true._

- _L and M are both true._

Further, N is false if: L and M are both false.

Similarly, we can define a new statement: $N := \neg M$
In this case, if M is true then N is false.
In the same vain, if M is false then N is true.

Definition 1.4. _**And**_
Let A, B be statements.

$$A \wedge B := \neg((\neg A) \vee (\neg B))$$

Definition 1.5. _**Intersection and Union**_
Let A, B be sets.

$$A \cap B := \{x : x \in A \wedge x \in B\}$$

$$A \cup B := \{x : x \in A \vee x \in B\}$$

Definition 1.6. _**Compliment**_
Let A be a set.
Then: $x \notin A := \neg(x \in A)$ and $A^c := \{x : x \notin A\}$

Definition 1.7. _**Logical Implication and Equivalence**_
Let A, B be statements:
$$A \Rightarrow B := (\neg A) \vee B$$

Further:
$$A \Leftrightarrow B := (A \Rightarrow B) \wedge (B \Rightarrow A)$$

Definition 1.8. _**Theory Crafting**_
A **mathematical theory** is a collection of well formed statements.
We write $\vdash P$ if and only if a statement P is said to be a part of our theory M

Axiom 1.1. _**Propositional Logic**_
Let A, B and C be a statements.
Then we have the following axioms:

- $\vdash A \vee A \Rightarrow A$

- $\vdash A \Rightarrow A \vee B$

- $\vdash A \vee B \Rightarrow B \vee A$

- $\vdash (A \Rightarrow C) \Rightarrow (A \vee B \Rightarrow B \vee C)$

These statements I am taking as the initial part of our theory M.

Rule Of Inference 1.1. _**Modus Ponens**_
Let A, B be statements.
If $\vdash A \Rightarrow B$ and $\vdash A$ then $\vdash B$
Note that you can't use the Principle of Substitution after you use Modus Ponens. Written as a single theorem:
$\vdash [(A \Rightarrow B) \wedge A] \Rightarrow B$

Rule Of Inference 1.2. _**Logical Deduction**_
Let P and Q be statements.
First, we suppose that $\boxed{\vdash} P$.
If we can then show that $\boxed{\vdash} Q$,
then it is proven that $\vdash P \Rightarrow Q$
Note, this does not mean that either $\vdash P$ or $\vdash Q$
To show this, instead of writing: $\vdash$ we will write $\boxed{\vdash}$ meaning that there is a dependence on an assumption.

Theorem 1.1. *__First Theorem__*

Let A be a statement.

Then $A \Rightarrow A \vee A$

Proof:

We know that:

$\vdash A \Rightarrow A \vee B$ *[Axiom 2]*

By replacing B with A we then have:

$\vdash A \Rightarrow A \vee A$ *[Principle of Substitution.]*

Theorem 1.2. *__Second Theorem__*

Let A be a statement.

Then $A \Leftrightarrow A \vee A$

Proof:

Since $\vdash A \Rightarrow A \vee A$ and $\vdash A \vee A \Rightarrow A$ *[Axiom 1 and the first theorem]*

And so we know:

$\vdash A \Leftrightarrow A \vee A$ *[Definition of Equivalence]*

Theorem 1.3. *__The Commutative Property of OR__*

Let A, B be statements.

Then: $A \vee B \Leftrightarrow B \vee A$

Proof:

By Axiom 3 we know:

$\vdash A \vee B \Rightarrow B \vee A$ *[Axiom 3]*

We can then Substitute B for C:

$\vdash A \vee C \Rightarrow C \vee A$ *[Principle of Substitution]*

Next we substitute A for B:

$\vdash B \vee C \Rightarrow C \vee B$ *[Principle of Substitution]*

Finally we substitute C for A:

$\vdash B \vee A \Rightarrow A \vee B$ *[Principle of Substitution]*

And so by definition of equivalence we have that:

$\vdash A \vee B \Leftrightarrow B \vee A$ *[Definition of Equivalence]*

Theorem 1.4. *__Shakespeare's theorem__*

Let B be a statement.

Then $B \vee \neg B$

Proof:

We have the Axiom:

$\vdash A \Rightarrow A \vee B$ *[Axiom 2]*

We can then replace A with B and we have:

$\vdash B \Rightarrow B \vee B$ *[Principle of Substitution.]*

$\vdash \neg B \lor (B \lor B)$ *[Def. of Implication]*

$\vdash \neg B \lor B$ *[Second Theorem]*

By the commutative property of OR we can then have that:

$\vdash B \lor \neg B$ *[Commutative Property of Or]*

Theorem 1.5. *The Tautology Theorem*

Let A be a statement.

Then: $A \Rightarrow A$ and further $A \Leftrightarrow A$

Proof: ───────────────────

Let B be a statement.

From the proof Shakespeare's theorem we know that:

$\vdash \neg B \lor B$ *[Proof of Shakespeare's theorem.]*

Replace B with A.

$\vdash \neg A \lor A$ *[Principle of Substitution.]*

By the definition of Implication we have that:

$\vdash A \Rightarrow A$ *[Def. of Logical Implication]*

Finally by the definition of equivalence we have that:

$\vdash A \Leftrightarrow A$ *[Def. of Equivalence]*

Theorem 1.6. *The Double Negative Theorem*

Let A be a statement.

Then $\neg(\neg A) \Leftrightarrow A$

Proof: ───────────────────

From Shakespeare's theorem we have:

$\vdash B \lor \neg B$ *[Shakespeare's Theorem]*

Replace B with $\neg A$

$\vdash \neg A \lor \neg(\neg A)$ *[Principle of Substitution]*

Then by the definition of implication we have:

$\vdash A \Rightarrow \neg(\neg A)$ *[Def. of Logical Implication]*

For the next proof we will need the following, substitute A with $\neg A$

$\vdash \neg A \Rightarrow \neg(\neg(\neg A))$ *[Principle of Substitution] *1*

Now for the other direction, start with axiom 4:

$\vdash (A \Rightarrow C) \Rightarrow (A \lor B \Rightarrow B \lor C)$ *[Axiom 4]*

Substitute C with $\neg(\neg A)$

$\vdash (A \Rightarrow \neg(\neg A)) \Rightarrow (A \lor B \Rightarrow B \lor \neg(\neg A))$ *[Principle of Substitution]*

Substitute A with $\neg A$

$\vdash (\neg A \Rightarrow \neg(\neg(\neg A))) \Rightarrow (\neg A \lor B \Rightarrow B \lor \neg(\neg(\neg A)))$ *[Principle of Substitution]*

Substitute B with A

$\vdash (\neg A \Rightarrow \neg(\neg(\neg A))) \Rightarrow (\neg A \lor A \Rightarrow A \lor \neg(\neg(\neg A)))$ *[Principle of Substitution]*

We have that $\vdash (\neg A \Rightarrow \neg(\neg(\neg A)))$ from the last part:

$\vdash \neg A \lor A \Rightarrow A \lor \neg(\neg(\neg A))$ *[Modus Ponens knowing: *1]*

We have Shakespeare's theorem and so:

$\vdash A \lor \neg(\neg(\neg A))$ *[Modus Ponens knowing: $\vdash \neg A \lor A$]*

We know that OR is commutative and so:

$\vdash \neg(\neg(\neg A)) \lor A$ *[The Commutative Property of OR]*

By the definition of implication we know:

$\vdash \neg(\neg A) \Rightarrow A$ *[Def. of Logical Implication.]*

And finally by the definition of equivalence we know:

$\vdash \neg(\neg A) \Leftrightarrow A$ *[Def. of Logical Equivalence.]*

Theorem 1.7. _Contrapositive_

Let A, B be statements.

$(A \Rightarrow B) \Leftrightarrow (\neg B \Rightarrow \neg A)$ *and* $(A \Leftrightarrow B) \Leftrightarrow (\neg A \Leftrightarrow \neg B)$

Theorem 1.8. _Captain Morgan's Laws for Logic_

Let A, B be statements.

Then: $\neg(A \land B) \Leftrightarrow \neg A \lor \neg B$ *and* $\neg(A \lor B) \Leftrightarrow \neg A \land \neg B$

Proof:

Using the Deductive method we start with the assumption that:

$\boxed{\vdash} \ \neg(A \land B)$ *[Deductive Method]*

By the definition of and we know the following are equivalent.

$\boxed{\vdash} \ \neg(\neg(\neg A \lor \neg B))$ *[Def. And]*

By the previous theorem we know that:

$\boxed{\vdash} \ \neg A \lor \neg B$ *[Double Negative Theorem]*

Then the Deductive Method tells us that:

$\vdash \neg(A \land B) \Leftrightarrow \neg A \lor \neg B$ *[Deductive Method Conclusion]*

And so we are done with the first part.

Using the Deductive Method again, we start with the assumption that:

$\boxed{\vdash} \ \neg A \land \neg B$ *[Deductive Method]*

By definition of and, we know that:

$\boxed{\vdash} \ \neg(\neg(\neg A) \lor \neg(\neg B))$ *[Def. of And]*

By the double negative theorem we know that:

$\boxed{\vdash} \ \neg(A \lor B)$ *[Double Negative Theorem]*

And so by the Deductive Method we know:

$\vdash \neg A \land \neg B \Leftrightarrow \neg(A \lor B)$ *[Conclusion of the Deductive Method]*

Which gives us both statements.

Theorem 1.9. _Shakespeare's Contradiction_

Let B be a statement and define: $\bot := B \land \neg B$

Then $\neg\bot$

Start with Shakespeare's theorem.

$\vdash B \lor \neg B$

$\vdash \neg\neg(B \lor \neg B)$

$\vdash \neg(\neg B \land \neg\neg B)$

$\vdash \neg(\neg B \land B)$

$\vdash \neg(B \land \neg B)$

$\vdash \neg\bot$

And so it's not the case that contradictions are part of our theory.

Theorem 1.10. ***Logical Selection***

Let A, B be statements.

$\vdash A \Rightarrow A \vee B$

$\vdash \neg A \Rightarrow \neg A \vee B$

$\vdash \neg A \Rightarrow \neg A \vee \neg B$

$\vdash \neg A \Rightarrow \neg(A \wedge B)$

$\vdash \neg\neg A \vee \neg(A \wedge B)$

$\vdash A \vee \neg(A \wedge B)$

$\vdash \neg(A \wedge B) \vee A$

$\vdash (A \wedge B) \Rightarrow A$

Theorem 1.11. ***Modus Ponens***

Let A, B be statements.

Then: $[(A \Rightarrow B) \wedge A] \Rightarrow B$

$\boxed{\vdash} (A \Rightarrow B) \wedge A$

$\boxed{\vdash} (\neg A \vee B) \wedge A$

$\boxed{\vdash} \neg(\neg(\neg A \vee B) \vee \neg A)$

Theorem 1.12. ***Logical Reduction***

Let A, B be statements.

$[(A \vee B) \wedge \neg B] \Rightarrow A$

From Modus Ponens we have:

$\vdash [(A \Rightarrow B) \wedge A] \Rightarrow B$

$\vdash [(\neg A \vee B) \wedge A] \Rightarrow B$

$\vdash [(\neg C \vee B) \wedge C] \Rightarrow B$

$\vdash [(\neg C \vee A) \wedge C] \Rightarrow A$

$\vdash [(\neg B \vee A) \wedge B] \Rightarrow A$

$\vdash [(A \vee \neg B) \wedge B] \Rightarrow A$

$\vdash [(A \vee \neg\neg B) \wedge \neg B] \Rightarrow A$

$\vdash [(A \vee B) \wedge \neg B] \Rightarrow A$

Rule Of Inference 1.3. ***Contradiction***

Let P be a statement.

If $P \Rightarrow \bot$ then $\neg P$

Proof: ⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯

Suppose $\boxed{\vdash}$ $P \Rightarrow \bot$ *[Deductive Method]*

$\boxed{\vdash}$ $\neg P \vee \bot$ *[Def. Implication]*

$\boxed{\vdash}$ $\neg P$ *[We know $\vdash \neg\bot$]*

$\therefore$ $\vdash [(P \Rightarrow \bot) \Rightarrow (\neg P)]$

By this theorem we have an extension of the deductive method and our new rule of Inference:

Method of Contradiction:

If we assume P

And then we can deduce $\bot$

Then we know $\neg P$

Theorem 1.13. _Principle of Explosion_

Let A be a statement.

Definition 1.9. _Subsets and Set Equality_

Let A, B be sets.

$(A \subset B) := (x \in A \Rightarrow x \in B)$

$(A = B) := (A \subset B) \wedge (B \subset A)$

Theorem 1.14. _De Morgan's laws_

Let A, B be sets.

Then: $(A \cap B)^c = A^c \cup B^c$ and $(A \cup B)^c = A^c \cap B^c$

Proof:

Let $x \in (A \cap B)^c$

$\Leftrightarrow x \notin A \cap B$ *[Def. Compliment]*

$\Leftrightarrow \neg(x \in A \cap B)$ *[Def. $\notin$]*

$\Leftrightarrow \neg(x \in A \wedge x \in B)$ *[Def. Intersection]*

$\Leftrightarrow \neg(x \in A) \vee \neg(x \in B)$ *[Captain Morgan's laws for logic]*

$\Leftrightarrow x \notin A \vee x \notin B$ *[Def. $\notin$]*

$\Leftrightarrow x \in A^c \vee x \in B^c$ *[Def. Compliment]*

$\Leftrightarrow x \in A^c \cup B^c$ *[Def. Union]*

$\therefore (A \cap B)^c = A^c \cup B^c$ *[Def. Subset and Set Equality]*

Let $x \in (A \cup B)^c$

$\Leftrightarrow x \notin A \cup B$

$\Leftrightarrow \neg(x \in A \cup B)$

$\Leftrightarrow \neg(x \in A \wedge x \in B)$

$\Leftrightarrow \neg(x \in A) \vee \neg(x \in B)$

$\Leftrightarrow x \notin A \vee x \notin B$

$\Leftrightarrow x \in A^c \vee x \in B^c$

$\Leftrightarrow x \in A^c \cup B^c$

$\therefore (A \cup B)^c = A^c \cap B^c$

Theorem 1.15. ***The union only makes things larger***

Let A, B be sets.

Then: $A \subset A \cup B$

Proof: ───────────────────────

Let $x \in A$ and $x \in B$ be statements.

Then by Axiom 2 we know:

$x \in A \Rightarrow x \in A \vee x \in B$ *[Axiom 2 where the statements are $x \in A$ and $x \in B$]*

Then by definition of union we know:

$x \in A \Rightarrow x \in A \cup B$ *[Def. Union]*

And so by definition of subsets we know:

$A \subset A \cup B$ *[Def. of Subsets]*

Theorem 1.16. ***Union and Intersection Distributive Properties***

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

Proof: ───────────────────────

Let $x \in A \cap (B \cup C) \Rightarrow x \in A \wedge x \in B \cup C$

Suppose that $x \notin C$

$\Rightarrow x \in B$

$\Rightarrow x \in A \wedge x \in B$

$\Rightarrow x \in A \cap B$

$\Rightarrow x \in (A \cap B) \cup (A \cap C)$ *[$(A \cap B) \subset (A \cap B) \cup (A \cap C)$]*

Suppose that $x \notin B$

$\Rightarrow x \in C$

$\Rightarrow x \in A$ and $x \in C$

$\Rightarrow x \in A \cap C$

$\Rightarrow x \in (A \cap B) \cup (A \cap C)$ *[$(A \cap C) \subset (A \cap B) \cup (A \cap C)$]*

Therefore:

$$A \cap (B \cup C) \subset (A \cap B) \cup (A \cap C)$$

Definition 1.10. ***Power Set***

Let $X \neq \phi$

$$2^X := \{V : V \subseteq X\}$$

Definition 1.11. _**Injection, Surjection, and Bijection**_
Let A, B be sets and function $f : A \to B$. f is said to be an Injection if:

$$(\forall x, y \in A)(f(x) = f(y) \Rightarrow x = y)$$

f is said to be a Surjection if:

$$(\forall y \in B)(\exists x \in A)(f(x) = y)$$

f is said to be a Bijection if it is both an Injection and a Surjection.

Theorem 1.17. _**It's pretty big.**_
Let $X \neq \phi$ and $M := card(X) < card(\mathbb{N})$
Then: $card(2^X) = 2^M$

Definition 1.12. _**K - Combinations**_
Let S be a non empty finite set where $n = card(S)$
A K-combination of S is a subset: $K \subset S$ where $card(K) = k$
We then have the collection: $C(S, k) = \{K \subset S : card(K) = k\}$
The number of K-combinations $= card(C(S, k)) = \binom{n}{k}$

Definition 1.13. _**Equivalence Relations**_
Let $S \neq \phi$
Then $\cong$ is called an Equivalence Relation if:

- $(\forall a \in S)(a \cong a)$ _[Reflexive]_

- $(\forall a, b \in S)(a \cong b \Leftrightarrow b \cong a)$ _[Symmetric]_

- $(\forall a, b, c \in S)(a \cong b \land b \cong c \Rightarrow a \cong c)$ _[Transitive]_

Reference

Definition 1.14. _**Equivalence Class**_
Let $S \neq \phi$ and $a \in S$ and $\cong$ be an equivalence relation on S.
Then the equivalence class $[a]$ is defined as follows:

$$[a] := \{x \in S : x \cong a\}$$

2 Topology

Definition 2.1. *Topology*
Let $X \neq \phi$
Further let $\tau \subseteq 2^X$ such that:

$$\phi, X \in \tau$$

$$(\forall A \neq \phi) \left(\{U_\alpha\}_{\alpha \in A} \subseteq \tau \Rightarrow \bigcup_{\alpha \in A} U_\alpha \in \tau \right)$$

$$(\forall m \in \mathbb{N}) \left(\{U_j\}_{j \in \mathbb{N}_m} \Rightarrow \bigcap_{j=1^m} U_j \in \tau \right)$$

Definition 2.2. *Relative Topology*
Let $X \neq \phi$ and $Z \subset X$
Then the relative topology on Z is written as follows:

$$\tau_Z = \{Z \cap U : U \in \tau_X\}$$

Theorem 2.1. *The Relative Topology is a Topology on Z*
Let $E \in \tau_Z$

$$\Rightarrow E = Z \cap U \subset Z$$

$$\Rightarrow \tau_Z \subseteq 2^Z$$

And so we have met the first criteria.
Next:

$$\phi \in \tau \Rightarrow Z \cap \phi \in \tau_Z \Rightarrow Z \cap \phi = \phi \in \tau_Z$$

Next:

$$X \in \tau \Rightarrow Z \cap X \in \tau_Z \Rightarrow Z \cap X = Z \in \tau_Z$$

Next: Let $A \neq \phi$ and $\{U_\alpha\}_{\alpha \in A} \in \tau_Z$

$$\Rightarrow \exists \{V_\alpha\}_{\alpha \in A} \subset \tau \text{ such that: } U_\alpha = Z \cap V_\alpha$$

$$\Rightarrow \bigcup_{\alpha \in A} U_\alpha = \bigcup_{\alpha \in A} Z \cap V_\alpha$$

3 Change

Definition 3.1. <u>***Metric***</u>
Let X be a non-empty set..
Let $d : X \times X \to \mathbb{R}_0^+$ such that:

- *$(\forall x, y \in X)d(x, y) = 0 \Leftrightarrow x = y$*

- *$(\forall x, y \in X)d(x, y) = d(y, x)$*

- *$(\forall x, y, z \in X)d(x, z) \leq d(x, y) + d(y, z)$*

Then d is called a metric and (X, d) is called a metric space.
Reference

Definition 3.2. <u>***Limit of a function***</u>
Let $T : X \to Y$ where (X, d_X) and (Y, d_Y) are metric spaces.
Then fix $x_0 \in X$.
If:
$$(\exists L \in Y)(\forall \epsilon > 0)(\exists \delta > 0)(\forall x \in X)(d(x, x_0) < \delta \Rightarrow d(f(x), L) < \epsilon)$$
Then:
$$\lim_{x \to x_0} f(x) = L$$

Reference

Definition 3.3. <u>***Derivative***</u>
Let $\hat{f} : \mathbb{R} \to \mathbb{R}$
Further let $f = \hat{f}|_U$ where $U \in \tau_{\mathbb{R}}$
Then f is said to be differentiable at $x \in U$ if there exists an L_x such that:

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

If L_x exists for all $x \in U$ then we write:

$$\frac{d}{dx} f(x) = \lim_{h \to 0} \frac{f(x + h) - f(x)}{h}$$

Reference

Theorem 3.1. <u>***Fundamental increment lemma***</u>
Let f be described as above and be differentiable at x.
Then there exists a function $\phi : \mathbb{R} \to \mathbb{R}$ such that:

$$f(x + h) = f(x) + \frac{d}{dx} f(x)h + \phi(x)h$$

and

$$\lim_{h \to 0} \phi(h) = 0$$

Proof:

Define: $\phi(h) = \frac{f(x+h)-f(x)}{h} - \frac{d}{dx}f(x)$

Then: $\phi(h)h = f(x+h) - f(x) - \frac{d}{dx}f(x)h$

Then: $\phi(h)h + f(x) - \frac{d}{dx}f(x)h = f(x+h)$

And so we have property 1.

Next:

$$\lim_{h \to 0} \phi(h) = \lim_{h \to 0} \left[\frac{f(x+h) - f(x) - \frac{d}{dx}f(x)h}{h} \right] = \lim_{h \to 0} \left[\frac{f(x+h) - f(x)}{h} - \frac{d}{dx}f(x) \right]$$

$$= \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} - \lim_{h \to 0} \frac{d}{dx}f(x) = \frac{d}{dx}f(x) - \frac{d}{dx}f(x) = 0$$

Definition 3.4. *Partial Derivative*

Let $\hat{f} : \mathbb{R}^n \to \mathbb{R}$

Further let $f = \hat{f}|_U$ where $U \in \tau_{\mathbb{R}^n}$

Then f is said to be differentiable at $x \in U$ with respect to the i'th component of x if there exists an L_{x_i} such that:

$$L_{x_i} = \lim_{h \to 0} \frac{f(x_1, ..., x_i + h, ..., x_n) - f(x_1, ..., x_i, ..., x_n)}{h}$$

If L_{x_i} exists for all $x \in U$ then we write:

$$\frac{\partial}{\partial x_i}f(x) = \lim_{h \to 0} \frac{f(x_1, ..., x_i + h, ..., x_n) - f(x_1, ..., x_i, ..., x_n)}{h}$$

Reference

Theorem 3.2. *Equivalent characterization*

Let $\hat{f} : \mathbb{R}^n \to \mathbb{R}$

Further let $f = \hat{f}|_U$ where $U \in \tau_{\mathbb{R}^n}$

And let f be differentiable at $x \in U$ with respect to the i'th component of x, then:

$$L_{x_i} = \lim_{h \to 0} \frac{f(x_1, ..., x_i + h, ..., x_n) - f(x_1, ..., x_i, ..., x_n)}{h}$$

$$\Leftrightarrow 0 = \lim_{h \to 0} \left[\frac{f(x_1, ..., x_i + h, ..., x_n) - f(x_1, ..., x_i, ..., x_n)}{h} - L_{x_i} \right]$$

$$\Leftrightarrow 0 = \lim_{h \to 0} \left[\frac{f(x_1, ..., x_i + h, ..., x_n) - f(x_1, ..., x_i, ..., x_n)}{h} - \frac{L_{x_i} \cdot h}{h} \right]$$

$$\Leftrightarrow 0 = \lim_{h \to 0} \left[\frac{f(x_1, ..., x_i + h, ..., x_n) - f(x_1, ..., x_i, ..., x_n) - \langle L_{x_i}, h \rangle}{h} \right]$$

Definition 3.5. _Gradient_
Let $\hat{f} : \mathbb{R}^n \to \mathbb{R}$ and let $f : U \to \mathbb{R}$ such that $f = \hat{f}|_U$ where $U \in \tau_{\mathbb{R}^n}$
f is said to be differentiable at $x \in U$ if $\exists \nabla f : \mathbb{R}^n \to \mathbb{R}^n$ such that:

$$\lim_{h \to 0} \frac{|f(x+h) - f(x) - \langle \nabla f(x), h \rangle|}{||h||} = 0$$

Theorem 3.3. _Form of the Gradient_
Let f be defined as above.
Then $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$ where:

$$\nabla f(x) = \begin{bmatrix} \frac{\partial}{\partial x_1} f(x) \\ \vdots \\ \frac{\partial}{\partial x_n} f(x) \end{bmatrix} \forall x \in \mathbb{R}^n$$

is the form of ∇f which satisfies the above statement if f is differentiable.
Reference

Proof:
Suppose ∇f is defined as above and all the partial derivatives exist.
Then:

$$\frac{1}{||h||} |f(x+h) - f(x) - \langle \nabla f(x), h \rangle| = \frac{1}{||h||} \left| f(x+h) - f(x) - \sum_{j=1}^{n} \frac{\partial}{\partial x_j} f(x) \cdot h_j \right|$$

Definition 3.6. _Matrix Functional Differentiability_
Let $\hat{T} : \mathbb{R}^{n \times m} \to \mathbb{R}$ and let $T : U \to \mathbb{R}$ such that $T = \hat{T}|_U$ where $U \in \tau_{\mathbb{R}^{n \times m}}$
T is said to be differentiable at $x \in U$ if $\exists D : \mathbb{R}^{n \times m} \to \mathbb{R}^{n \times m}$ such that:

$$\lim_{h \to 0} \frac{|T(x+h) - T(x) - \langle DT(x), h \rangle|}{||h||} = 0$$

where $\langle \cdot, \cdot \rangle$ is an inner product defined on $\mathbb{R}^{n \times m}$

Definition 3.7. _Frobenius inner product_
The Frobenius inner product is defined as:

$$\langle \cdot, \cdot \rangle_{FB} : \mathbb{R}^{n \times m} \times \mathbb{R}^{n \times m} \to \mathbb{R} \text{ such that: } \langle A, B \rangle_{FB} = \sum_{i=1}^{n} \sum_{j=1}^{m} a_{i,j} b_{i,j} \text{ for all } A, B \in \mathbb{R}^{n \times m}$$

Theorem 3.4. _Form of Matrix Functional Derivative_

$$DT(x) = \begin{bmatrix} \frac{\partial}{\partial x_{1,1}} & \cdots & \frac{\partial}{\partial x_{1,m}} \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_{n,1}} & \cdots & \frac{\partial}{\partial x_{n,m}} \end{bmatrix}$$

14

Definition 3.8. **_Differentiability of a multi-variable function._**
Let $\hat{f} : \mathbb{R}^m \to \mathbb{R}^n$ such that:

$$\hat{f}(x) = \begin{bmatrix} f_1(x) \\ \vdots \\ f_n(x) \end{bmatrix} \ and \ (\forall j \in \mathbb{N}_n)(f_j : \mathbb{R}^m \to \mathbb{R})$$

Further let $f = \hat{f}|_U$ where $U \in \tau_{\mathbb{R}^m}$
Then f is said to be differentiable at $x \in U$ if there exists a linear operator $J_{f(x)} : \mathbb{R}^m \to \mathbb{R}^n$ such that:

$$\lim_{h \to \vec{0}} \frac{||f(x+h) - f(x) + J_{f(x)}(h)||_{\mathbb{R}^m}}{||h||_{\mathbb{R}^n}} = 0$$

Reference

Theorem 3.5. **_The Jacobian matrix linear operator for the above system._**
So our guess is that:

$$J_{f(x)} = \begin{bmatrix} \frac{\partial}{\partial x_1} f_1(x) & \cdots & \frac{\partial}{\partial x_n} f_1(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_1} f_m(x) & \cdots & \frac{\partial}{\partial x_n} f_m(x) \end{bmatrix}$$

since this form is a linear operator mapping from the appropriate space to the appropriate space. It should be noted that the transpose of this matrix can not satisfy the definition of differentiability of a multi-variable function and so it is not the correct linear operator.

Definition 3.9. **_Matrix operator differentiability_**
Let $T : \mathbb{R}^{n \times m} \to \mathbb{R}^n$ such that:

$$T(A) = \begin{bmatrix} T_1(A) \\ \vdots \\ T_n(A) \end{bmatrix} \forall A \in \mathbb{R}^{n \times m} \ and \ (\forall j \in \mathbb{N}_n)(T_j : \mathbb{R}^{n \times m} \to \mathbb{R})$$

Then T is said to be differentiable at $A \in \mathbb{R}^{n \times m}$ if there exists a linear operator $D : \mathbb{R}^{n \times m} \to \mathbb{R}^n$ where:

$$\lim_{h \to 0} \frac{||T(A+h) - T(A) + D(h)||_{\mathbb{R}^n}}{||h||_{\mathbb{R}^{n \times m}}} = 0$$

If D exists then it is called the Matrix operator derivative and is written: $D_{\mathbb{R}^{n \times m}} T(A)$

Theorem 3.6. _The form of the Matrix operator derivative._
Let T be described as above and differentiable at $A \in \mathbb{R}^{n \times m}$

$$\frac{T(A+h) - T(A)}{||h||} = \begin{bmatrix} \frac{T_1(A+h) - T_1(A)}{||h||} \\ \vdots \\ \frac{T_n(A+h) - T_n(A)}{||h||} \end{bmatrix}$$

and so:

$$\lim_{h \to 0} \frac{T(A+h) - T(A)}{||h||} = \begin{bmatrix} \lim_{h \to 0} \frac{T_1(A+h) - T_1(A)}{||h||} \\ \vdots \\ \lim_{h \to 0} \frac{T_n(A+h) - T_n(A)}{||h||} \end{bmatrix}$$

Definition 3.10. _Subspace Differentiability_
Let $X = \{X_j\}_{j=1}^n$ be a sequence of finite dimensional vector spaces where $dim(X_j) = k_j = m_j \times n_j$
Let $T : \prod_{j=1}^n X_j \to Y$ where Y is a finite dimensional vector space with $dim(Y) = k_y$
Let $x_j \in X_j$ for some $j \in \mathbb{N}_n$
Where

$$x_j = \begin{bmatrix} x_{1,1} & \cdots & x_{1,n_j} \\ \vdots & \ddots & \vdots \\ x_{m_j,1} & \cdots & x_{m_j,n_j} \end{bmatrix}$$

T is said to be differentiable at $x \in X$ where $x = (x_0, ..., x_j, ..., x_{n-1})$ with respect to X_j if there exists a linear operator $D : X_j \to Y$:
Given $h \in X_j \setminus \{\vec{0}\}$ define $\hat{h} = (0, ..., h, ..., 0) \in X$ where h is in the j'th place of $\hat{h}$:

$$\lim_{h \to 0} \frac{||T(x + \hat{h}) - T(x) + D(h)||_X}{||h||_{X_j}} = 0$$

Then D is called the subspace derivative of T at x with respect to X_j and is written: $D_{x_j}T(x)$

Definition 3.11. _Product space Derivative_
Let $X = \{X_j\}_{j=0}^{n-1}$ be a sequence of finite dimensional vector spaces where $dim(X_j) = k_j$
Let $T : \prod_{j=0}^{n-1} X_j \to Y$ where Y is a finite dimensional vector space with $dim(Y) = k_y$
Let $\{x_j\}_{j=0}^{n-1}$ be a sequence of vectors such that: $(\forall j \in \{0, ..., n-1\})(x_j \in X_j)$
The product space derivative at the point $z \in X$ is:

$$D_X T(z) = \begin{bmatrix} D_{x_0}T(z) \\ \vdots \\ D_{x_{n-1}}T(z) \end{bmatrix}$$

Definition 3.12. _**Fréchet derivative**_

Let V, W be normed vector spaces and $U \subset V$ be an open set.

An operator $f : U \to W$ is said to be Fréchet differentiable if there exists a bounded linear operator $A : V \to W$ such that:

$$\lim_{||h|| \to 0} \frac{||f(x + h) - f(x) + Ah||_W}{||h||_V} = 0$$

Reference

Theorem 3.7. _**Fréchet derivative of a bounded linear operator**_

Let V, W be normed vector spaces and $U \subset V$ be an open set.

Let $\hat{f} : V \to W$ be a bounded linear operator.

Then lets look at $f = \hat{f}|_U$

My guess is that $A = \hat{f}$

Let $x \in U$ and $h \in U \pitchfork ||h|| \neq 0$ and $x + h \in U$, Then:

$$\frac{||f(x + h) - f(x) + Ah||_W}{||h||_V} = \frac{||f(x) + f(h) - f(x) + \hat{f}(h)||_W}{||h||_V} = \frac{||f(x) + f(h) - f(x) + f(h)||_W}{||h||_V} = 0$$

Thus let $\epsilon > 0$ and $\delta > 0$

Then if $0 < ||h|| < \delta$ we know that $\frac{||f(x+h) - f(x) + Ah||_W}{||h||_V} = 0 < \epsilon$

Therefore:

$$\lim_{||h|| \to 0} \frac{||f(x + h) - f(x) + Ah||_W}{||h||_V} = 0$$

Thus $A = \hat{f}$ is the Fréchet derivative of f.

3.1   Finite Composition Operator

Definition 3.13. _**Finite Composition Operator**_

Let the collection $X = \{X_j\}_{j=0}^{n}$ be a finite sequence of sets.

Further let $\{T_j\}_{j=0}^{n-1}$ be a finite sequence of operators such that $(\forall j \in \mathbb{N}_{n-1})(T_j : X_j \to X_{j+1})$

Then $T^n : X_0 \to X_n$ defined by:

$$T^n := \bigcirc_{j=0}^{n-1} T_j$$

is called the **Finite Composition Operator defined on X**.

Definition 3.14. _**Multi-variable Finite Composition Iteration**_

Let the collection $X = \{X_j\}_{j=0}^{n}$ and $Y = \{Y_j\}_{j=0}^{n-1}$ be finite sequences of sets.

Further let $\{T_j\}_{j=0}^{n-1}$ be a finite sequence of operators such that: $(\forall j \in \mathbb{N}_{n-1})(T_j : X_j \times Y_j \to X_{j+1})$

Let $T^n : X_0 \times \prod_{j=0}^{n-1} Y_j \to X_n$ where:

$$T^n(x, y) = z_n \text{ where } z_{j+1} = T_j(z_j, \pi_j(y)) \text{ or } z_{j+1} = T_j(z_j) \text{ and } z_0 = x \in X_0$$

Definition 3.15. ***Gradient Descent***
Let $E : \mathbb{R}^n \to \mathbb{R}$ be a differentiable operator.
The method of Gradient Descent says that a local minimum of E can be found using the following iteration:

$$a_{n+1} = a_n - \gamma \nabla E(a_n)$$

Where $\gamma > 0$

Example 3.1. ***Objective Operator for Data Set Defined Operator Approximation***
Let $X \subset \mathbb{R}^n, Y \subset \mathbb{R}^m$ such that $X \times Y$ defines an operator T.

$$E(a) = \sum_{x \in X} ||T(x) - T^n(x, a)||$$

4 Surjective Continuous Non-decreasing Bounded Functionals

Let $B = \{f : \mathbb{R} \to [0, 1] | f \text{ is surjective, continuous, and non-decreasing.}\}$

Theorem 4.1. ***B is convex.***
Let $f, g \in B$ and $h(x) := \lambda f(x) + (1 - \lambda)g(x)$ where $\lambda \in [0, 1]$
Then h is still continuous since the linear combination of continuous functions is continuous.

Proof:
Since both f and g are surjective and non-decreasing, then there exists x_0, y_0, x_1, y_1 in $\mathbb{R}$ such that:
$f(x_0) = 0 = g(y_0)$ and $f(x_1) = 1 = g(y_1)$
Suppose WLOG that $x_0 \leq y_0$ and $x_1 \leq y_1$
Then we know that:
$$h(x_0) = \lambda f(x_0) + (1 - \lambda)g(x_0) = \lambda 0 + (1 - \lambda)0 = 0$$

and

$$h(y_1) = \lambda f(y_1) + (1 - \lambda)g(y_1) = \lambda 1 + (1 - \lambda)1 = 1$$

Now if we pick $\alpha \in [0, 1]$ by the intermediate value theorem, we know that there exists an $x_\alpha \in [x_0, y_1]$ such that:

$$h(x_\alpha) = \alpha$$

Since α was arbitrary element, I have shown that h is surjective.

Finally, let $x_0 < x_1$ be elements in $\mathbb{R}$
Then we know that $f(x_0) \leq f(x_1)$ and $g(x_0) \leq g(x_1)$
$\Rightarrow \lambda f(x_0) \leq \lambda f(x_1)$ and $(1 - \lambda)g(x_0) \leq (1 - \lambda)g(x_1)$
$\Rightarrow \lambda f(x_0) + (1 - \lambda)g(x_0) \leq \lambda f(x_1) + (1 - \lambda)g(x_1)$
$\Rightarrow h(x_0) \leq h(x_1)$

18

Thus h is non-decreasing.
Since h is surjective, continuous, and non-decreasing, then $h \in B$
Thus B is convex.

Theorem 4.2. <u>B is translation invariant.</u>
Let $f \in B$ and $g(x) := f(x + c)$ where $c \in \mathbb{R}$
f is continuous and so is the addition operator so g is continuous.
Proof:
Let $\alpha \in [0,1]$ since f is surjective then $\exists x \in \mathbb{R} \pitchfork f(x) = \alpha$
Then $g(x - c) = f(x + c - c) = f(x) = \alpha$ and so g is surjective.
Let $x < y$ be elements in $\mathbb{R}$
Then $f(x) \leq f(y) \Rightarrow f(x + c) \leq f(y + c)$
$\Rightarrow g(x) \leq f(y)$ and so g is non-decreasing.
Thus $g \in B$ and B is therefore translation invariant.

Theorem 4.3. <u>B is not complete.</u>

Theorem 4.4. <u>Every element in B can be decomposed as a finite non-trivial convex combination</u>

5 A topological description of finite metric spaces.

Finite spaces are of interest because they describe the world of computers. This being the case, we would still like to do analysis on these spaces and analysis starts with topological descriptions.

Theorem 5.1. <u>The First Rule of Induced Topologies on Finite Metric Subspaces is:</u>
Let $V \subset X$ where $0 < card(V) =: N < card(\mathbb{N})$ and X is a metric space.
Then the subspace topology on V is the discrete topology.
Proof:
The associated topological space on V is:

$$\tau_V = \{V \cap U : U \in \tau_X\}$$

Since V is of finite carnality, we can uniquely number each element.
Thus:

$$V = \bigcup_{i=1}^{N} \{v_i\}$$

Further:
$$V_{\min} := \min\{d(x, y) : x, y \in V\}$$

Let $v \in V$ and $\epsilon_V = \frac{V_{\min}}{2}$
Then $B(v, \epsilon_V) \in \tau_X \Rightarrow V \cap B(v, \epsilon_V) \in \tau_V$
However $V \cap B(v, \epsilon_V) = \{v\}$ and since v was arbitrary, we thus know that: $(\forall v \in V)(\{v\} \in \tau_V)$
We can now prove that $\tau_V = 2^V$
By definition we know that $\tau_V \subset 2^V$
Let $E \in 2^V$ then $E = \bigcup_{j=1}^{M}\{v_j\}$ where $M \leq N$
Since we know ever $\{v_j\}$ is open we know that E is open and thus: $E \in \tau_V$
Thus the induced topology on a finite subset of a Metric space is the discrete topology.

Lemma 5.1. _Everything is Continuous when your domain is finite._
Let $V \subset X$ and $Y \neq \phi$ where X, Y are metric spaces and $0 < card(V) = N < card(\mathbb{N})$
Then ever $f : V \to Y$ is continuous.
Proof:
Let $U \in \tau_W$
Then: $f^{-1}(U) \subset V$ where $f : V \to Y$ is arbitrary.
Thus: $f^{-1}(U) \in \tau_V$ by the previous theorem.
And so f is continuous.
And so every $f : V \to Y$ is continuous.

Theorem 5.2. _Everything is Lipschitz continuous when your domain is finite._
So if we want to have some form of meaningful topological description of continuity for these spaces, we are going to need a "stronger" form of continuity.
Let $V \subset X$ and $Y \neq \phi$ where X, Y are metric spaces and $0 < card(V) = N < card(\mathbb{N})$
*Further let $f : V \to Y$, then f is **Lipschitz continuous**.*
Proof:
We can simply look at:

$$K := \max\left(\left\{\frac{d_Y(f(x), f(y))}{d_X(x, y)} : x, y \in V\right\}\right)$$

Then we know that:

$$(\forall x, y \in V)\left(\frac{d_Y(f(x), f(y))}{d_X(x, y)} \leq K\right)$$

And therefore:

$$(\forall x, y \in V)\left(d_Y(f(x), f(y)) \leq K d_X(x, y)\right)$$

And thus f is Lipschitz continuous.

Definition 5.1. <u>***K - Families***</u>

Let $V \subset X$ and $Y \neq \phi$ where X, Y are Normed spaces and $0 < card(V) = N < card(\mathbb{N})$

First we have the set of all operators.

$$\Lambda := \{f | f : V \to Y\}$$

Next we have the $K - Families$

$$\Lambda(K) := \{f \in \Lambda | K \text{ is the smallest Lipschitz constant for } f.\}$$

The set: $\Lambda(K)$ is called a K - Family. Next we have the $K - Nests$

$$B(K) := \{f : V \to Y | (\forall x, y \in V)(||f(x) - f(y)|| \leq K||x - y||)\}$$

Theorem 5.3. <u>***K - Properties***</u>

- $B(K) = \bigcup_{L \in [0,K]} \Lambda(L)$

  Proof:

  Let $f \in B(K) \Rightarrow f$ is lipschitz.

  Which means that there exists a unique smallest constant K' where: $0 \leq K' \leq K$

  Thus $f \in \Lambda(K') \subset \bigcup_{L \in [0,K]} \Lambda(L)$

  Therefore: $B(K) \subseteq \bigcup_{L \in [0,K]} \Lambda(L)$

  Let $f \in \bigcup_{L \in [0,K]} \Lambda(L) \Rightarrow \exists L \in [0, K] \pitchfork (\forall x, y \in V)(||f(x) - f(y)|| \leq L||x - y|| \leq K||x - y||)$

  $\Rightarrow f \in B(K)$

  $\Rightarrow \bigcup_{L \in [0,K]} \Lambda(L) \subseteq B(K)$

  $\therefore B(K) = \bigcup_{L \in [0,K]} \Lambda(L)$

- $L \leq K \Rightarrow B(L) \subseteq B(K)$

  Proof:

  Let $f \in B(L)$ and $x, y \in V$

  $\Rightarrow ||f(x) - f(y)|| \leq L||x - y|| \leq K||x - y||)$

  $\Rightarrow f \in B(K)$

- $B(K)$ is convex.

  Proof:

  Let $f, g \in B(K)$ and $h_\lambda(x) = \lambda f(x) + (1 - \lambda)g(x)$ for some $\lambda \in [0, 1]$

  Further let $x, y \in V$:

  $$||h_\lambda(x) - h_\lambda(y)|| = ||\lambda f(x) + (1-\lambda)g(x) - (\lambda f(y) + (1-\lambda)g(y))|| = ||\lambda f(x) + (1-\lambda)g(x) - \lambda f(y) - (1-\lambda)g(y)||$$

  $$= ||\lambda(f(x) - f(y)) + (1 - \lambda)(g(x) - g(y))|| \leq \lambda||(f(x) - f(y))|| + (1 - \lambda)||(g(x) - g(y))||$$

  $$\leq \lambda K||x - y|| + (1 - \lambda)K||x - y|| = K||x - y||$$

  $\Rightarrow h_\lambda \in B(K)$

- $\Lambda(K)$ is not convex.

  Proof:

  Let $K > 0$

  If $f \in \Lambda(K) \Rightarrow -f \in \Lambda(K)$

  However $\frac{f + -f}{2} = 0 \in \Lambda(0) \neq \Lambda(K)$

Lemma 5.2. *__Paths through Function spaces__*
Let $f \in \Lambda(K)$ and $g \in \Lambda(L)$ then $h_\lambda(x) := \lambda f(x) + (1-\lambda)g(x)$ at most belongs to the family: $\Lambda(\lambda K + (1-\lambda)L)$ Further: $H : [0,1] \to \Gamma$ where $H(\lambda) = h_\lambda$ is a path through the function space:

$$\Gamma = \bigcup_{\nu \in [0,1]} \Lambda(\nu \cdot \max(L, K))$$

Proof: ▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬

$$||h(x)-h(y)|| = ||\lambda f(x)+(1-\lambda)g(x)-(\lambda f(y)+(1-\lambda)g(y))|| = ||\lambda f(x)+(1-\lambda)g(x)-\lambda f(y)-(1-\lambda)g(y)||$$

$$= ||\lambda(f(x)-f(y))+(1-\lambda)(g(x)-g(y))|| \le \lambda ||(f(x)-f(y))|| + (1-\lambda)||(g(x)-g(y))||$$

$$\le \lambda K ||x-y|| + (1-\lambda)L||x-y|| = (\lambda K + (1-\lambda)L)||x-y||$$

Theorem 5.4. *__Ordering__*
For this theorem, we will need both the domain and co-domain to be finite.
First, Let $V \subset X$ and $W \subset Y$ where X, Y are normed spaces.
Next let $\Lambda := \{f | f : V \to W\}$ where $0 < card(V) =: N < card(\mathbb{N})$ and $0 < card(W) =: M < card(\mathbb{N})$
And: $\Lambda(K) := \{f \in \Lambda | K$ is the Lipschitz constant for f.$\}$ as before.
Then:

$$K < L \Rightarrow card(\Lambda(K)) < card(\Lambda(L))$$

6 Convex Operators

Theorem 6.1. *__Invariance under Convex Monotonic Operator Composition__*
Reference

Theorem 6.2. *__Invariance under Affine Composition__*
Let $f : \mathbb{R}^m \to \mathbb{R}^n$ be a convex operator, then:

$$g(x) = f(Ax + b)$$

Is also a convex operator.