# Baby Rudin Summary Notes

Bryan Johnston

October 20, 2020

## Contents



## 0 Introduction

These are my notes on Walter Rudin’s *Principles of Mathematical Analysis (Third Edition)* chapters 1-9. Chapters 9 and 10 are left out, since it’s not uncommon to feel that differential forms and Lebesgue theory are covered better in other places.

I tried to include items, from either the main text or the exercises, which are frequently used later and so are important in that way.

I hope these notes are useful to someone.

## 0.1 Preparation

Officially, *Principles of Mathematical Analysis* requires very little background material. Rudin explicitly mentions the axioms governing the arithmetic of the integers $\mathbb{Z}$ and familiarity of the rational number field $\mathbb{Q}$. So this means he is saying that he expects you to know about:

- Associative, commutative, and distributive properties for the integers
- How fractions (rational numbers) behave
- Induction
- The well-ordering principle for the natural numbers $\mathbb{N}$
- Inductive/recursive definitions of sets, functions, etc...

But in fact, the text actually requires that the reader knows about things like:

- The greatest common divisiors and algorithm
- The unique factorization of integers into primes
- The binomial theorem/expansion
- The sum of a finite geometric series

You also need to be familiar with the way mathematicians reason about and prove things. Velleman's book *How to Prove It: A Structured Approach* is good. Hopefully one of these two links to it will work:

1. `http://users.metu.edu.tr/serge/courses/111-2011/textbook-math111.pdf`
2. `https://www.amazon.com/How-Prove-Structured-Daniel-Velleman/dp/1108439535`

In addition, there are these notes on background material, which come at things from a different perspective.

Here is a good example of the kind of thinking that you should be comfortable with:

> Let $f(l, m, n)$ be a function of three integer variables, where $f(l, m, n)$ is also an integer.
> Then the truth of *Statement #1* implies the truth of *Statement #2*.
>
> - *Statement #1*
>
>   Shorthand:
>   $$\forall W \exists V \forall m \forall n \big[V < |m - n| \implies W < f(l, m, n)\big]$$
>   Verbose:

For every $W$, there exists a $V$ such that for every $m$, and for every $n$,

$$V < |m - n| \implies W < f(l, m, n)$$

Super verbose:

For every $W$ it is true that: $\Big\{$

There exists a $V$ such that: $\Big[$

For every $m$ it is true that: $\Big($

For every $n$ it is true that: $\Big\langle$

$$V < |m - n| \implies W < f(l, m, n)$$

$\Big\rangle \Big) \Big] \Big\}$

- *Statement #2*

  Shorthand:

  $\forall W \forall m \exists V \forall n \big[V < |m - n| \implies W < f(l, m, n)\big]$

  Verbose:

  For every $W$, and for every $m$, there exists a $V$ such that for every $n$,

  $$V < |m - n| \implies W < f(l, m, n)$$

  Super verbose:

  For every $W$ it is true that: $\Big\{$

  For every $m$ it is true that: $\Big[$

  There exists a $V$ such that: $\Big($

  For every $n$ it is true that: $\Big\langle$

  $$V < |m - n| \implies W < f(l, m, n)$$

  $\Big\rangle \Big) \Big] \Big\}$

You should be able to *understand* and *prove* the above without too much trouble.

## 0.2 The Axiom of Choice

Rudin freely uses the *Axiom of Choice*, without mentioning that he is doing so.

One example is when he shows that the $\epsilon - \delta$ definitions of a limit is equivalent to the convergent sequence definition of a limit.

His proof of the *Stone-Weierstrass Theorem* is another example.

A lot of people (including me) tend to worry about whether or not certain results do or do not require *some version of Choice* in order to be true.

It's a really interesting thing - but don't spend so much time worrying about it that you lose track of the bigger picture.

# 1 Chapter 1: The Real and Complex Number Systems

1. Proof that $\sqrt{2}$ is irrational
2. Def of *total order* = *linear order* = *strict order*
3. $\leq$ vs $<$
4. *supremum/l.u.b., infimum/g.l.b.*
5. *maximum/largest, minimum/smallest* of a set are unique, if they exist
6. "least upper bound property" for an ordered set $S$:
   Every non-empty subset $E \subset S$ which is bounded above has a supremum in $S$
7. Def of *field*, and properties which follow from definition
8. Def of *ordered* field: "+" preserves order, $xy > 0$ when $x$ and $y$ are greater than 0
9. Properties of ordered fields (including $1 > 0$ as a consequence)
10. Existence of $\mathbb{R} \supset \mathbb{Q}$, where $\mathbb{R}$ has l.u.b. property
11. Archimedean property of $\mathbb{R}$ follows from lub property.
12. $\mathbb{Q}$ *dense* in $\mathbb{R}$ follows from Archimedean property.
    Means that $\mathbb{R}$ is "basically unique" (left vague for the time being).
13. Existence and uniqueness of *n-th roots* of positive reals $x$, for $n \in \mathbb{N}$.
    Written $x^{1/n}$ or $\sqrt[n]{x}$
    Tricky proof if you haven't seen it.
    Rudin doesn't provide a lot of motivation or intuition here.
    It's on the reader to figure out where the intuition comes from.
    Main tricks in proof are:
    - Finding a candidate as a supremum
    - "telescoping " $b^n - a^n$
    - $b^n - a^n < (b-a)nb^{n-1}$ when $0 < a < b$

- Proof by contradiction, constructing clever counterexamples.

14. Baby Rudin version of the *extended reals*
    - Written $\overline{\mathbb{R}}$ or $[-\infty, +\infty]$
    - Slightly different properties than the version in Papa Rudin
    - Is not a field!
    - Some operations left undefined.
    - Is totally ordered, and has l.u.b. property
15. The *complex numbers* $\mathbb{C}$
    - Defined as set of ordered pairs $(x, y)$ of real numbers
    - $i$ is defined as $(0, 1)$ without mentioning "the mysterious squre root of $-1$"
    - Introduce alternate notation/shorthand:
      "$x + iy$" means $(x, y)$
16. You should think "$\mathbb{R} \subset \mathbb{C}$" by "identifying" $a \in \mathbb{R}$ with $(a, 0) \in \mathbb{C}$.
    Yes, what exactly this means may be a little fuzzy for now...Don't worry about it.
17. *Triangle Inequality* for $\mathbb{R}^2$ and $\mathbb{C}$.
    Proof is standard and fundamental.
18. *Schwarz Inequality* for $\mathbb{C}^n$.
    Proof is standard and fundamental.
19. *Schwarz Inequality.* for $\mathbb{R}^n$ as a consequence of the one for $\mathbb{C}^n$ (because $\mathbb{R} \subset \mathbb{C}$)
20. *Triangle Inequality* on $\mathbb{R}^n$ as a consequence of *Schwarz Inequality* for $\mathbb{R}^n$
21. (Looking forward to p.88) *Triangle Inequality for Subtraction* on $\mathbb{R}^n$ as an easy consequence.

    $$\Big| |x| - |y| \Big| \leq \Big| x - y \Big|$$

# 2 Chapter 2: Basic Topology

1. Rudin freely uses the *Axiom of Choice*.
2. Definition of *function.*
   The *codomain* is part of the function in his definition.
   Image, pre-image/inverse-image
3. Onto/surjective, one-to-one/injective
4. Bijection = one-to-one "correspondence"

5. "Same cardinal number" = "equivalent sets" = exists a bijection

6. finite, infinite, countable, uncountable, at-most-countable

7. Definition: An *(infinite) sequence* is a *function* with domain $\mathbb{N}$.

8. Infinite subset of countable set is countable.

9. *Indexed collection of sets*, unions, intersections, (pairwise) disjoint.

10. A countable union of countable sets in countable.
    Same proof is one that shows $\mathbb{Q}$ is countable.

11. $\mathbb{Q}$ is countable, $\mathbb{Z}^n$ is countable

12. Set of infinite binary-valued sequences $\{0, 1\}^\infty$ is uncountable.

13. $\mathbb{R}$ is uncountable.

14. Definition of *metric space*.
    There are three requirements for the *metric* function $d(x, y)$.

15. Important! The empty set $\emptyset = \{\}$ is allowed to be a metric space in Rudin's framework!

16. Subspaces

17. Segment = $(a, b)$ vs interval = $[a, b]$ in $\mathbb{R}$

18. Convex subsets of $\mathbb{R}^n$

19. *Basic topology in metric spaces:*

    - neighborhood
    - limit point
    - open
    - closed
    - perfect
    - bounded
    - dense

    Some important examples of these.

20. De Morgan's rules for unions and intersections of indexed collections of sets.

21. Unions and intersections of open and closed sets.

22. Fact (#6, p.43): Set of limit points is closed.

23. Definition of *closure* $\overline{E}$ of a subset $E$ of metric space.
    "Smallest" closed set containing it.
    Intersection of all closed sets containing it.
    Union of the set and its limit points.

24. Fact (preview chapter 3):
    $x \in \overline{E}$ i.f.f. there exists a sequence of points in $E$ "converging" to $x$.
    Convergence is defined in chapter 3.

25. Supremum of a non-empty bunded set in $\mathbb{R}$ is in the set's closure.
    Consequence: If the set is closed, then the set has a maximum.

26. *Subspace* topology, *relatively open* subset.

27. Open covers, compact subsets, compact spaces.

28. *Hausdorff* spaces (metric spaces as an example).

29. Compact subspaces of Hausdorff spaces are closed.

30. Closed subsets of compact spaces are compact.

31. (p.38) If you have a non-empty collection of compact sets in a Hausdorff space such that for every finite subcollection the intersection of that subcollection is non-empty, then the intersection of all of them is non-empty.

32. Compact guarantees *"limit point compact"*

33. (p.39) Closed intervals and *k-cells* are compact.

34. (p.40) *Heine-Borel Theorem* for $\mathbb{R}^n$:
    Equivalent:
    (a) Compact
    (b) Closed and blounded
    (c) Limit point compact

35. *Bolzano-Weierstrass Theorem*:
    Every bounded and infinite subset of $\mathbb{R}^n$ has a limit point in $\mathbb{R}^n$.

36. A non-empty perfect set is uncountable.
    Important proof.
    Key point: It's a variation on Cantor's "diagonal construction".

37. *The Cantor Set*

38. *Connected sets*, *separations.*
    Two equivalent definitions.

39. *Interval property* of connected sets in $\mathbb{R}$.

40. Subset of $\mathbb{R}$ is connected if and only if it is a bounded or unbounded interval.

# 3 Chapter 3: Numerical Sequences and Series

1. Definition of *sequence* with values in a metric space.

2. $(\epsilon, \delta)$ definition of *convergence* of a sequences.

3. *Limit* of a sequence.

4. Note! These are different but related concepts:
   - Limit of a convergent sequence
   - Limit point in a metric space

5. *Divergence*

6. "Convergence to $\infty$" for sequences in $\mathbb{R}$.
   (Technically *not* convergence to a limit in $\mathbb{R}$...)

7. Convergence/divergence depends on the ambient space!

8. *Bounded* sequence: if the range is bounded.

9. (p.48) Characteristics of convergent sequences:
   - The limit of a convergent sequence is unique.
   - Convergent sequence is bounded.
   - Given a limit point (chapter 2) $p \in X$ of a set $E$ in a metric space $X$, there is a sequence of points in $E$ which converges to that limit point $p$.
   - If $E$ is a subset of a metric space $X$, then the closure (chapter 2) of $E$ in $X$ consists exactly of all points in $X$ which are limits of convergent sequences of points from $E$.

10. *Theorem: Preserving an Inequality in a Limit*
    This is not officially presented in the book, but is used all over the place.

    Let $\{a_n\}$ be a convergent sequence of real numbers such that $a_n \to a$ as $n \to \infty$.

    If $a_n \leq L$ for every $n$, then $a \leq L$.

    Proof: Trivial proof by contradiction (or contrapositive).

11. Finite termwise sums and products of convergent sequences in $\mathbb{C}$.

12. Convergent sequences in $\mathbb{R}^n$.

13. Finite termwise sums and dot-products of convergent sequences in $\mathbb{R}^n$.

14. *Subsequences* and *subsequential limits.*

15. An obvious but useful fact:

    If $\{c_n\}$ is a subsequence of $\{b_n\}$, and $\{b_n\}$ is a subsequence of $\{a_n\}$, then $\{c_n\}$ is a subsequence of $\{a_n\}$.

16. Useful fact:

    If $\{a_n\}$ is a sequence in $\mathbb{N}$ such that $a_n < a_{n+1}$ for each $n$
    then $n \leq a_n$ for each $n$.

    Prove by induction. Use the fact that there are no integers between 0 and 1.

17. (p.51) Sequences in compact spaces have convergent subsequences.

18. (p.52) Set of subsequential limits of a sequence is closed.
    A limit point of a set of subsequential limits is itself a subsequential limit.

19. *Cauchy* sequences.

20. *Diameter* of a subset of a metric space.

21. *Tails* of a sequences.

22. Sequence is Cauchy i.f.f. the sequence of diameters of its tails tends to zero.

23. The diameter of a closure is the same as the diameter of the original.

24. (p.53) Cauchy sequences vs convergent sequences:

    (a) In general metric space

    (b) In compact metric space

    (c) In $\mathbb{R}^n$

25. Definition of *complete* metric space.

26. Fact: A compact metric space is complete.

27. Fact: A closed subset of a complete metric space is complete.

28. Definition of *monotone increasing/decreasing* sequences in $\mathbb{R}$.

29. A monotone sequence converges i.f.f. it's bounded.

30. Definitions of $\limsup$ and $\liminf$ for sequences in $\mathbb{R}$.

31. Equivalent definitions of $\limsup$ and $\liminf$.

32. (p.57) Properties of $\limsup$ and $\liminf$:

    (a) $\{s_n\}$ converges i.f.f. $\limsup = \liminf$

(b) $s_n \le t_n$ in some tail guarantees
$\liminf\{s_n\} \le \liminf\{t_n\}$
$\limsup\{s_n\} \le \limsup\{t_n\}$

(c) Respects termwise multiplcation by convergent sequence of positive real numbers.

33. Three special convergent sequences:

(a) $p > 0$ guarantees $p^{1/n} \to 1$ as $n \to \infty$.

(b) $n^{1/n} \to 1$ as $n \to \infty$.

(c) If $\alpha \in \mathbb{R}$, then $p > 0$ guarantees $\dfrac{n^\alpha}{(1+p)^n} \to 0$ as $n \to \infty$

Prove these using the Binomial Theorem!

34. Definition of *series* of values in $\mathbb{C}$ or $\mathbb{R}$.

- Crucial: A *series* is a *sequence* whose terms are *partial sums* derived from another sequence.
- Notation/shorthand:
  $\sum_n a_n$ is the series (sequence of partial sums) derived
  from the underlying sequence $\{a_n\}$.
- Convergence of series is *defined as* convergence of the sequence of partial sums.
- $\sum_n a_n = l$ is shorthand for
  "the sequence of partial sums for $\sum_n a_n$ converges to $l$".

35. (Not explicitly stated in book?) *Fact: Finite Sum Of Convergent Series (FSOCS)*

If all $K$ of the series

$$\sum_n a_{1,n} \quad , \quad \sum_n a_{2,n} \quad , \quad \dots \quad , \quad \sum_n a_{K,n}$$

converge, then $\sum_n (a_{1,n} + a_{2,n} + \cdots + a_{K,n})$ converges, and

$$\sum_n \left( \sum_{k=1}^{K} a_{k,n} \right) = \sum_{k=1}^{K} \left( \sum_n a_{k,n} \right)$$

Proof: We already know how finite sums of limits of convergent *sequences* behave. Apply that to the sequences of partial sums for the $K$ series we're looking at. There's a small detail where you have to look at a finite sum of finite sums in two ways:

$$\sum_{k=1}^{K} \sum_{n=1}^{N} a_{k,n} = \sum_{n=1}^{N} \sum_{k=1}^{K} a_{k,n}$$

36. Cauchy criterion for series convergence is the Cauchy criterion for its sequence of partial sums.

37. *Boundedness* criterion for series convergence.

38. *Comparison test* for series convergence (actually gives "absolute" convergence... coming later...)

39. *Geometric series*:

    (a) If $0 \leq x < 1$, then $\sum_{n\geq 0} x^n = \frac{1}{1-x}$.

    (b) Otherwise, $\sum_{n\geq 0}$ diverges.

40. *Cauchy condensation/subsequence criterion* for series $\sum_n a_n$

    whose underlying sequence $\{a_n\}$ is monotone decreasing and non-negative.

41. Convergence/divergence of $\sum_n \frac{1}{n^p}$.

    (a) Case $p > 1$: converges.

    (b) Case $p \leq 1$: diverges.

    Prove by comparing the *Cauchy condensations* with geometric series.

42. Definition of the number $e = \sum_{n\geq 0} \frac{1}{n!}$

43. Proof that $e = \lim_{n\to\infty} \left(1 + \frac{1}{n}\right)^n$

44. (p.65) Fast convergence of partial sums $s_n$ for $e$:

$$0 < e - s_n < \frac{1}{n!n}$$

    Straightforward proof by looking at the tails.

45. Beautiful proof that $e$ is irrational, using previous item.

46. *Root Test* for series convergence using $\alpha = \limsup_n |a_n|^{1/n}$

    (a) Case $\alpha < 1$: *Absolute* convergence.

    (b) Case $\alpha > 1$: Divergence.

    (c) Case $\alpha = 1$: Is inconclusive.

47. *Ratio Test* for series convergence. Works differently than the Root Test!

    (a) *Absolute* convergence, if $\limsup_n \left|\frac{a_{n+1}}{a_n}\right| < 1$.

    (b) Divergence, if $\left|\frac{a_{n+1}}{a_n}\right| \geq 1$ for *every* $n$ sufficiently large (past some cutoff).

    (c) Otherwise is inconclusive.

    Important! Even knowing that $\lim_n \left|\frac{a_{n+1}}{a_n}\right| = 1$ gives you no information.
    Compare $a_n = \frac{1}{n}$ and $a_n = \frac{1}{n^2}$ to see this.
    Also notice how "shuffling terms" can "confuse" the Ratio Test.

48. (p.68) Theorem: For any sequence $\{a_n\}$ in $\mathbb{R}$,

    - $\liminf_n \left|\frac{a_{n+1}}{a_n}\right| \leq \liminf_n |a_n|^{1/n}$
    - $\limsup_n |a_n|^{1/n} \leq \limsup_n \left|\frac{a_{n+1}}{a_n}\right|$

49. (p.68) The Root Test versus the Ratio Test.
    From previous item, have:

    - If Ratio Test indicates convergence (case (a)), then so does the Root Test.
    - If the Root Test is inconclusive (case (c)), then the Ratio Test either indicates divergence or is inconclusive (cases (b) and (c)).

50. (p.69) *Power series* for a sequence $\{c_n\}$ in $\mathbb{C}$ and some *fixed* $z \in \mathbb{C}$:

    $$\sum_n c_n z^n$$

    This is just a series!
    Don't try to think of it as a function quite yet.
    ***!!! Wait until Chapter 8 to think of power series as "functions of $z$" !!!***

51. (p.69) *Radius of convergence* of a sequence $\{c_n\}$ of coefficients used in a power series.

    $$\alpha = \limsup_n |c_n|^{1/n}$$

    $$R = \frac{1}{\alpha}$$

    Can verify using Root Test:

    (a) Case $|z| < R$: The power series $\sum_n c_n z^n$ converges *absolutely*.

(b) Case $|z| > R$: The power series $\sum_n c_n z^n$ diverges.

(c) Case $|z| = R$: The Root Test is inconclusive.
But! See the result following the Abel/Dirichlet Convergence Test.

52. (p.70) *Summation By Parts* (this is so great)
The proof of this is a straightforward manipulation of terms.

Given sequences $\{a_n\}$ and $\{b_n\}$,
put $A_{-1} = 0$, and $A_n = a_0 + a_1 + \cdots + a_n$ for $n \geq 0$.

Then if $0 \leq p \leq q$, you get

$$\sum_{n=p}^{q} a_n b_n = A_q b_q - A_{p-1} b_p - \sum_{n=p}^{q-1} A_n (b_{n+1} - b_n)$$

This formula is entirely analogous to "Integration By Parts", which is covered later in this book in Chapter 6 on Integration.

53. (p.70) *Abel/Dirichlet Convergence Test*
    - Proof using Summation By Parts!
    - (p.71) Application #1: *Alternating Series Test*
    - (p.71) Application #2: A power series with radius of convergence = 1 and a decreasing sequence of positive coefficients converges at all $z$ with $|z| = 1$, except possibly $z = 1$.

54. Fact: (Preview of Chapter 7) The Abel/Dirichlet Convergence Test works for sequences of functions where the hypotheses hold "uniformly".

55. Definition of *absolute convergence.*

56. The *Cauchy product* of two series.

57. (p.74) Fact: The Cauchy product of two convergent series converges to the "right value" if *at least one* of them converges absolutely.

58. (p.75, p.175) Fact: If the Cauchy product converges, then it converges to the "right value".
Proof: Delayed until p.175.

59. (p.76) *Rearrangements* of non-absolutely convergent series:
Can rearrange to converge to arbitrary values, including $\pm\infty$, or even diverge with arbitrary $\limsup$ and $\liminf$.

60. (p.78) *Rearrangements* of absolutely convergent series:
All rearrangements converge, and in fact converge to the sum of the original series.

61. At this point (the end of Chapter 3), Rudin could've stated and proved the

    (p.175, Theorem 8.3) *Cauchy Double Series Theorem*

    He could've given an elementary proof which uses only material presented so far, basically Knopp's proof on p.143 of Knopp's *Theory and Application of Infinite Series*. Instead Rudin waits until he he needs the theorem in Chapter 8, then gives a slick proof which uses uniform convergence and continuity.

# 4 Chapter 4: Continuity

1. The $(\epsilon, \delta)$ definition of the *limit* of a function at a *limit point* of its domain.
    - This limit point doesn't need to actually be in the domain!
    - The usual $(\epsilon, \delta)$ definition which is used in metric spaces.
    - The value of the function at the lmit point (if it's in the domain) is not used in the definition of *limit*.
      Using only values "close to" but not equal to the limit point.
2. Alternate definition of the *limit* of a function, in terms of convergent sequences.
    - Only need to worry about sequences which *converge to but never hit* the limit point.
    - Is equivalent to $(\epsilon, \delta)$ definition if you believe in *Axiom of Countable Choice*.
3. Fact: If the limit of $f$ at a domain limit point $a$ exists, then it is unique.
4. Fact: These limits behave like you'd hope:
   Limits of pointwise sum, product, difference, quotient, and dot product.
   Limits of the *real-valued* components of functions in $\mathbb{R}^n$ (or $\mathbb{C}^n$) vs. limits of the original function.
5. The $(\epsilon, \delta)$ definition of *continuity* of a function at a point in its domain.
6. Definition of *continuity* of a function on a subset of its domain.
7. Definition of *(globally) continuous* function on a domain.
8. Trivial but important fact: If $f$ is continuous on $X$ and $A \subset X$, then the restriction $f|_A$ is continuous on $A$.
9. (p.86) Continuity of a function at
    (a) *isolated points* of the domain:
    Nothing to check...is always continuous at these points.
    (b) *limit points* of the domain:
    Continuous there i.f.f. the limit exists and equals the function value there.
    Easy to verify/prove that this is true.

10. (p.86) Composition of continuous is continuous.

11. Here's a theorem that's very useful but is not formally presented in the book. I've never actually seen it stated and authors will frequently apply it without making any special mention, so I conclude that the theorem must be super obvious to most mathematicians. I guess, then, the fact that I seem to pay such special attention to it probably says...something... about me. :) This is also useful in some of the exercises here (see for example Chapter 5 #2 p.114).

    *Theorem: Composition of Limits in Metric Spaces*

    If

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

    (b) $\lim_{y \to b} g(y) = c$

    (c) $f(x) \neq b$ in some neighborhood of $a$

    Then

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

    (Existence of this limit is part of the conclusion.)

    *Proof:* Trivial and straight from definitions.
    Assumption (c) is essential and that there are easy counterexamples without it.

12. Here's another theorem that's super useful and used all over the place but not formally presented in the book.

    *Theorem: Continuous Functions Respect Limits in Metric Spaces*

    If

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

    (b) $g$ is defined near $b$ and is continuous at $b$

    Then

    $$\lim_{x \to a} \Big( g(f(x)) \Big) = g(b) = g\Big( \lim_{x \to a} f(x) \Big)$$

    (Existence of this limit is part of the conclusion.)

    *Proof:* Trivial and straight from definitions.

13. (p.86) *"Pre-images of open sets" characterization* of continuous functions.
    Straightforward proof of equivalence with $(\epsilon, \delta)$ continuity at every point in the domain.

14. Corollary: *"Pre-images of closed sets" characterization* of continuous functions.
    Because pre-image respects set complement.

15. Fact: These behave like you'd hope:
Continuity of pointwise sum, product, difference, quotient, dot product.
Continuity of the *real-valued* components of functions in $\mathbb{R}^n$ (or $\mathbb{C}^n$) vs. continuity of the original function.

16. (p.88) Fact: The function

$$w \mapsto |w|$$

is continuous on $\mathbb{R}^n$ (or $\mathbb{C}^n$).
Easy proof using *Triangle Inequality for Subtraction*.

17. Definition of *bounded* function $f : E \longrightarrow \mathbb{R}^n$
When the image $f(E)$ is a bounded subset of $\mathbb{R}^n$.

18. (p.89) Theorem:
The image of a compact set under a continuous function is compact.
In other words: Continuous functions preserve compactness.

19. (p.89) Corollary #1:
If $X$ is compact and $f : X \longrightarrow \mathbb{R}^n$ is continuous, then $f(X)$ is closed and bounded.

20. (p.89) Corollary #2:
Real-valued continuous functions on compact sets "attain their minimum and maximum values".
In other words: $\sup f(X) \in f(X)$ and $\inf f(X) \in f(X)$.

21. (p.90) Corollary #3:

*Continuity of the inverse of a continuous bijection on a compact domain:*

If $f : X \longrightarrow Y$ is a continuous *injective* function from a *compact* metric space $X$ *onto* a metric space $Y$,
then the inverse function $f^{-1} : Y \longrightarrow X$ is also continuous.

In other words, $f$ is *homeomorphism*.

The trick in the proof is to work with closed wsets instead of open sets.

22. (p.90) Definition of *uniformly continuous* function.
Continuity on $X$ vs uniform continuity on $X$:

- Continuity on $X$:

$$\forall x \in X \ \forall \epsilon > 0 \ \exists \delta \ \left( \forall y \in X \ldots \text{etc} \ldots \right)$$

Here $\delta$ depends on $x$ and $\epsilon$.

- Uniform continuity on $X$:

$$\forall \epsilon > 0 \ \exists \delta \ \left( \forall x \in X \ \forall y \in X \ldots \text{etc} \ldots \right)$$

  This time $\delta$ depends on $\epsilon$ but *not* $x$.

23. (p.91) Theorem:
    A continuous function on a *compact* domain is actually *uniformly continuous* on that domain.

    The proof is important, since it is a common technique.

    (a) Pick arbitrary $\epsilon > 0$.
    (b) Use $\epsilon/2$ for continuity at each point $x$.
    (c) Get $\gamma_x$ neighborhood at each point $x$.
    (d) Apply compactness to the open cover provided by $\gamma_x/2$ neighborhoods.
    (e) Let $\delta$ be the minimum of the resulting finite collection of $\gamma_x/2$ values.
    (f) Apply the triangle inequality a couple of times.

24. (p.91) Counterexamples of previous theorem on non-compact subsets of $\mathbb{R}^1$.

25. (p.93) Theorem:
    Images under continuous functions of connected sets are connected.
    In other words: Continuous functions preserve connectedness.

26. (p.93) Corollary:
    *Intermediate Value Theorem* for continuous functions $f : [a, b] \longrightarrow \mathbb{R}$.

27. Discontinuities of functions.

28. (p.94) Right-hand and left-hand limits of functions defined on $E \subset \mathbb{R}$.
    $f(x+)$ and $f(x-)$ are *defined only at* limit points $x$ of $E$.

29. Types of discontinuities at $x \in (a, b)$ for $f$ definied on $(a, b)$:

    (a) Simple/first-kind: When $f(x+)$ and $f(x-)$ both exist.
    (b) Second-kind: When at least one of them doesn't exist.

30. Definition of *monotone increasing (or decreasing)* function defined on $(a, b) \subset [+\infty, -\infty]$.

31. (p.95) Easy fact:
    Left-hand and right-hand limits exist at all points for a *monotone* function defined on $(a, b)$.

32. (p.95) Theorem:

    If $f$ is a monotone increasing function defined on $(a, b)$:

(a) If $a < x < b$, then

$$\sup_{t<x} f(t) = f(x-) \le f(x) \le f(x+) = \inf_{x<t} f(t)$$

(b) If $a < x < y < b$, then

$$f(x+) \le f(y-)$$

(Similarly for monotone decreasing functions.)

Proof: Straightforward.

33. Corollaries:

    (a) Monotone functions don't have discontinuities of the second kind.

    (b) The set of discontinuities of a monotone function is at most countable (but don't need to all be isolated points).

34. (p.97) Fact:
    Given an aribrary countable set $\{x_n\} \subset (a, b)$, it's easy to construct (using a series) a monotone function on $(a, b)$ whose set of discontinuities is precisely the set $\{x_n\}$.

    Here's how:

    Let $\{c_n\}$ be a sequence of positive reals such that $\sum_n c_n$ converges.

    Define

    $$f(x) = \sum_{\substack{n \text{ such that} \\ x_n < x}} c_n$$

    Verification that this works is straightforward.

35. (p.97) *Left/right continuous* at a point.

36. (p.98) "Infinite limits" and "limits at infinity" in the extended reals $[-\infty, +\infty]$.

37. (p.98) "Neighborhoods of $\pm\infty$"

38. (p.98) Reformulation of limit points in $[-\infty, +\infty]$:

    If $E \subset [-\infty, +\infty]$, then $x \in [-\infty, +\infty]$ is a *limit point* of $E$
    if every *neighborhood* of $x$ contains a point of $E$ other than $x$ itself.

39. (p.98) Reformulation and unified treatment of limits and convergence in $[-\infty, +\infty]$:

    Suppose that

    (a) $E \subset [-\infty, +\infty]$
    (b) $f : E \longrightarrow [-\infty, +\infty]$
    (c) $x \in [-\infty, +\infty]$ is a limit point of $E$

(d) $L \in [-\infty, +\infty]$

(e) For every neighborhood $L \in V \subset [-\infty, +\infty]$,
there exists a neighborhood $x \in U \subset [-\infty, +\infty]$,
such that $f(t) \in V$ whenever $t \in U$ and $t \neq x$.

In this case you say "$L$ is the limit of $f$ as $t \to x$", and write

$$L = \lim_{t \to x} f(t)$$

or

$$f(t) \to L \text{ as } t \to x$$

# 5 Chapter 5: Differentiation

1. (p.103) Defining the *derivative of a function at a point* $x \in [a, b]$ for real-valued functions defined on closed intervals $[a, b] \subset \mathbb{R}$.

2. Definition applies even at the *endpoints* of the interval.

$$f'(x) = \lim_{t \to x} \frac{f(t) - f(x)}{t - x}$$

   Obviously, use one-sided limits when $x$ is an endpoint of $[a, b]$.

3. $f'$ is then a function whose domain is points of $[a, b]$ where the above limit exists.

4. (p.104) Definition: If $f'$ is defined at each point of $E \subset [a, b]$, then $f$ is *differentiable* on $E$.

5. (p.104) Easy fact: $f$ differentiable at $x \implies f$ continuous at $x$.
   Proof: Recall that limits respect multiplcation, and notice that (when $x \neq t$)

$$f(t) - f(x) = \frac{f(t) - f(x)}{t - x} \cdot (t - x)$$

6. (p.104) Derivatives of sums, products, and quotients.

7. (p.105) Theorem: *The Chain Rule*
   - Rudin's statement and proof are a little goofed up.
     - Only need $f$ *defined* on $[a, b]$. Don't actually need *continuity* on $[a, b]$.
     - The required continuity of $f$ at $x$ actually follows from differentiability at $x$.
   - The Chain Rule is a pointwise thing.
   - The proof is an application of linear approximations provided by the derivatives.

8. (p.114, #2) *Baby Inverse Function Theorem*

   If $f' > 0$ on $(a, b)$, then

(a) $f$ is *strictly* increasing on $(a, b)$.
(b) The image $f((a, b)) = (c, d)$ is a (possibly infinite) open interval.
(c) The inverse function $g = f^{-1}$ exists and is continuous on $(c, d)$.
(d) $g$ is differentiable on $(c, d)$ and $g'(f(x)) = \dfrac{1}{f'(x)}$

Proof: First show $g$ exists and is continuous. Then use the definition of derivative and "compose limits".

9. (p.107) Definition of *local minima/maxima.*

10. (p.107) Main Fact: If $f$ differentiable at local maximum or minimum, then $f'$ equals zero there.
Proof: Look at difference quotients on *both sides* near the point.

11. (p.107) *The Cauchy/Generalized Mean Value Theorem*

- Continuity of $f$ and $g$ required only on the closed interval $[a, b]$.
- Differentiability of $f$ and $g$ required only on the open interval $(a, b)$.
- The point $x$ that is guaranteed by the theorem is actually in the open interval $(a, b)$:
In other words: $a < x < b$.
- Proof: Uses the fact that $f(a, b])$ is compact.
Construct a clever function $h$ which is zero at both $a$ and $b$.
If $h$ is identically zero on the interval, then basically done.
Otherwise, look at what happens if $h$ is not identically zero on the interval.
Proceed...

12. (p.108) Corollary: *The usual/simple Mean Value Theorem ("MVT")*

13. (p.108) Obvious facts you get from the Mean Value Theorem:

(a) $f' \geq 0$ on $(a, b)$ guarantees $f$ is monotone increasing.
(b) $f' > 0$ on $(a, b)$ guarantees $f$ is monotone *strictly* increasing.
(c) $f' = 0$ on $(a, b)$ guarantees $f$ is constant.

14. (p.108) Interesting Fact: *Intermediate Value Theorem for Derivatives*

If $f'$ exists on the closed interval $[a, b]$ and $f'(a) < \lambda < f'(b)$,
then there exists a $x$ in the open interval$(a, b)$
such that $\lambda = f'(x)$.

Proof: Look at the helper function $g(x) = f(x) - \lambda x$ on $[a, b]$. Look at $g'$ and notice that the minimum of $g$ can't occur at the endpoints, so it must occur in the open interval $(a, b)$.

Warning! This theorem does *not* imply that derivatives are always continuous! They are not.

15. (p.109) Corollary:
If $f'$ exists on $[a, b]$, then $f'$ cannot have simple discontinuities on $[a, b]$.
That is, if $f'$ is not continuous at $x$, then at least one of $f'(x-)$ or $f'(x+)$ does not exist.

16. (p.109) Theorem: *l'Hostiptal's Rule*

    - Applies to various combinations of finite open intervals, infinite open intervals, and finite/infinite limiting values.
    - Note the individual limits of $f'$ and $g'$ are not assumed to exist!
    - Proof is an application of Cauchy/Generalized Mean Value Theorem.

17. (p.110) Higher-order derivatives $f^{(n)}$ of functions $f$ defined on $[a, b]$.
Things to pay attention to:

    - Use one-sided limits and neighborhoods where appropriate/necessary.
    Context matters!
    - If you know $f^{(n)}(x)$ exists, then that guarantees:
        (a) $f^{(n-1)}(x)$ must exist and that $f^{(n-1)}$ must exist at all points in some sufficiently small neighborhood of $x$. Otherwise, you can't make sense of the limiting process necessary to calculate $f^{(n)}(x)$!
        (b) And so since $f^{(n-1)}$ exists at all points of some neighborhood of $x$, it follows similarly that $f^{(n-2)}$ exists and is differentiable (and so obviously continuous) at all points of some neighborhood of $x$.

18. (p.110) *Taylor's Theorem*
Note: Is a generalization of the *MVT*.

    (a) Requirements: $f^{(n-1)}$ continuous on $[a, b]$ and $f^{(n)}$ exists on $(a, b)$.
    (Analogous to the *MVT*.)
    (b) (p.111) Rudin's beautiful proof is important to understand and involves clever repeated applications of the *MVT*.

19. (p.111) Differentiating vector-valued functions of one real variable

$$[a, b] \to \mathbb{R}^n$$

and complex-valued functions of one real variable

$$[a, b] \to \mathbb{C}$$

    - Same "limit of the difference quotient " definition as for derivatives of real-valued functions.
    - Everything winds up being done componentwise, because limits in $\mathbb{R}^n$ and $\mathbb{C}$ work componentwise.

20. (p.112) "Product Rule" for differentiating inner-product of vetor-valued functions works like you want it to.
Proof: Just a verification, using the fact that the components are differentiable.

21. (p.112) Crucial Fact!

    The *MVT* and corollaries (like *l'Hospital's Rule*) *fail to hold* in the complex/vector-valued scenarios!

    This is because they are all consequences of compactness and the theorem that if $f$ differentiable at *local maximum or minimum*, then $f'$ equals zero there.

    Good counterexample:
    The function $f(t) = e^{it}$ on $[0, 2\pi]$ is an example that shows the *MVT* does not hold for complex-valued functions.

22. (p.113) But not all hope is lost...
...because there is the following extremely useful theorem.

23. Theorem: *Vector-valued Mean Value Bound*

    If $f[a,b] \to \mathbb{R}^k$ is continuous on $[a,b]$ and differentiable on $(a,b)$, then

    $$|f(b) - f(a)| \leq |b - a||f'(x)|$$

    for *some* $x \in (a,b)$ (open interval!).

    Proof: Look at this function which is a real-valued inner product:

    $$\phi(t) = \langle f(b) - f(a), f(t) \rangle >$$

    Apply the *MVT* and *Schwarz Inequality*.

# 6 Chapter 6: The Riemann ("Riemann-Stieltjes") Integral

There are a couple things to mention here before we get going on the chapter.

Note #1: Many people think it's good to have exposure to the Riemann integral before looking (later) at the Lebesgue integral, which is a generalization of the Riemann one, and is presented for example in Rudin's followup book *Real & Complex Analysis*. In addition, Rudin covers a lot of integration-related material (uniform convergence, Fudamental Theorem of Calculus, power series, Gamma function, Stone-Weierstrass theorem, etc.) in *Principles of Mathematical Analysis*, and so we we need some kind of integral at this point if we want to cover that material.

Note #2: But, you can probably safely ignore the material in this chapter related to the *Stieltjes* generialization on the Riemann integral. This is the material which allows "$dx$" to be generalized to "$d\alpha(x)$", where $\alpha$ is a monotone increasing function. However this *Riemann-Stieltjes* integral is also just a Lebesgue integral using a "measure" (which you learn about in Lebesge integration) obtained from the function $\alpha$. For more details see this page about Lebesgue–Stieltjes integration.

Anyway, here are the chapter 6 notes.

1. (p.120) The *Riemann-Stieltjes Integral* as Rudin presents it is the *Darboux Integral*.
2. Note: Defined on $[a, b]$ only for *bounded* functions.
3. (p.120) Partitions and upper/lower sums for a partition.
4. (p.121) Upper and lower Riemann integrals for bounded functions on $[a, b]$. Written:

$$\overline{\int_a^b} f(x)\ dx \text{ and } \underline{\int_a^b} f(x)\ dx$$

   or

$$\overline{\int_a^b} f \text{ and } \underline{\int_a^b} f$$

5. Definition: $f$ is *Riemann integrable* if the upper and lower integrals are the equal. That is, if

$$\overline{\int_a^b} f = \underline{\int_a^b} f$$

   In which case the integral $\int_a^b f$ is defined as that common value.

6. Important! $dx$ or $dt$ or $ds$ or whatever is sometimes a meaningless waste of space. Try to use the better notation when its possible to avoid writing the bound "variable of integration".
   Sometimes it is, and sometimes it isn't.
   - It's possible, for example, when a function has already been given a name:
     Let $h$ be the function defined by $s \mapsto s^2$.
     Then you can write the integral as

$$\int_a^b h$$

- It's not possible, for example, when you have an "anonymous" function of, say, two variables $x$ and $y$.
  For example the function $(x, y) \mapsto xy$ defined on a recangle.
  In this case you need to specify which variable is "bound" by the integration process:

$$v(x) = \int_a^b xy \; dy$$

  versus

$$v(y) = \int_a^b xy \; dx$$

7. (p.123) *Refinement* of a partition of $[a, b]$.
   *Common refinement* of two partitions.

8. (p.123) Fact: Refining a partition decreases the upper sum closer to the upper integral, and increases the lower sum closer to the lower integral.

   This is due to basic behavior of sups and infs:

   If $X \subset Y \subset \mathbb{R}$, then

   - $\inf Y \leq \inf X$, because $\inf Y$ is one of the lower bounds of $X$.
   - $\sup X \leq \sup Y$, because $\sup Y$ is one of the upper bounds of $X$.

9. (p.124) *Theorem:*

$$\underline{\int_a^b} f \leq \overline{\int_a^b} f$$

   Proof: Prove that every lower sum is at most equal to any (possibly different) upper sum, by looking at common refinements. Take sups and infs.

10. (p.124) *Theorem: Criterion for Riemann Integrability*

    A bounded function on $[a, b]$ is integrable
    if and only if
    given any $\epsilon > 0$ there exists a partition whose upper and lower sums differ by at most $\epsilon$.

    Proof: Straightforward application of $\underline{\int_a^b} f \leq \overline{\int_a^b} f$

11. (p.125) *Theorem: Continuous Functions are Integrable*
    Proof: Compactness of $[a, b]$ implies uniform continuity.
    Use it, and the previous theorem.

12. (p.126) *Theorem: Monotone Functions are Integrable*
    (Even when they're not continuous!)

    Proof: The main trick here is that the difference between an upper sum and lower sum for the same partition "telescopes". The maximum value on one subinterval equals the minimum value on the next sub interval, because the function is monotone. It's really quite lovely. :)

13. Definition: The *mesh (or norm)* of a partition is the length of the longest subinterval. Rudin doesn't actually define this in the book (unless I missed it somehow), but the word is commonly used elsewhere.
    See `https://en.wikipedia.org/wiki/Partition_of_an_interval`.

14. (p.127) *Theorem:*
    The composite of an integrable function followed by a continuous function is integrable.

    Sketch Proof:
    Use uniform continuity of the second function.
    Find a partition that makes the upper and lower sums of the first function very close to each other.
    Main Trick: Divide the points of the partition (for the first function) into two buckets: Where $M_i - m_i$ is "small", and 2) where $M_i - m_i$ is "big".
    Observe things.

15. (p.128) *Theorem: Basic facts about the Riemann integral*

    (a) It's a "linear functional":
    $$\int_a^b (\alpha f + \beta g) = \alpha \int_a^b f + \beta \int_a^b g$$
    The integral of a (pointwise) linear combination is the linear combination of the integrals.

    (b) Respects pointwise "$\leq$":
    If $f \leq g$, then $\int_a^b f \leq \int_a^b g$.

    (c) If $f \leq M$, then $\int_a^b f \leq M|b-a|$.

    (d) $\int_a^b + \int_b^c = \int_a^c$

16. (p.129) *Theorem:*

    (a) The (pointwise) product $fg$ of two integrable functions $f$ and $g$ is integrable.

(b) The absolute value $|g|$ of an integrable function $g$ is integrable.

(c) $\left| \int_a^b f \right| \leq \int_a^b |f|$ (This is a *Triangle Inequality for Integrals.*)

Proof: One way to prove (a) and (b) is to do it from scratch, straight from the definitions.
If you want to do it that way for (a), then you'll have to first prove it for the case where $f$ and $g$ are both non-negative, and then proceed from there.

It's a little more clever to prove them by using the theorem which says an integrable function followed by a continuous one is integrable.

In particular, it's definitely worth checking out Rudin's proof of (a), where he applies the Polarization Identity.

For (c), just notice that both $f \leq |f|$ and $-f \leq |f|$, that the absolute value of the integral is either the integral of $f$ or $-f$, and that (as you already saw) integration respects pointwise "$\leq$".

17. (p.129) The *unit step function*:

$$I(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 & \text{if } x > 0 \end{cases}$$

Notice is left continuous at 0.

He presents it so that he can look at a certain kind Stieltjes integral (p.130 Theorems 6.15 and 6.16).
I don't think we *necessarily* need this for the rest of the material in the text, even though it is a very important function.

18. (p.133-134) *Relationship Between Riemann Integration and Differentiation* The hypotheses of each theorem are important.

(a) *Theorem 6.20:* If $f$ is integrable on $[a, b]$, then the function

$$F(x) = \int_a^x f$$

is *uniformly* continuous on $[a, b]$.

Note that Rudin's version only says that the function is continuous, but uniform continuity is apparent.

Proof: Because $f$ bounded, by definition of integrable,
so for some $M$ and all $x$ and $y$,

$$|F(y) - F(x)| = \left| \int\limits_x^y f \right| \leq \int\limits_x^y |f| \leq M\,|y - x|$$

(b) *Theorem Theorem 6.20: Fundamental Theorem of Calculus #1 (FTC #1)*
If $f$ and $F$ are as above, and $f$ is continuous at $x_0 \in [a, b]$ (could be an endpoint), then $F$ is differentiable at $x_0$ and $F'(x_0) = f(x_0)$.

Proof: Straightforward from the definition of derivative.

(c) *Theorem Theorem 6.21: Fundamental Theorem of Calculus #2 (FTC #2)*
If $f$ is integrable on $[a, b]$ and if $F$ is a differentiable function on $[a.b]$ such that $F' = f$ on $[a, b]$, then

$$\int\limits_a^b f = F(b) - F(a)$$

Proof: Pick a partition of $[a, b]$. Use the points of the partition to "telescope" $F(b) - F(a)$:

$$F(b) - F(a) = [F(b) - F(x_{n-1})] + [F(x_{n-1}) - F(x_{n-2})] + \cdots + [(F(x_1) - F(a)]$$

Apply *MVT* to each term on the right side and notice that this traps $F(b) - F(a)$ between the lower and upper sums for the partition.

(d) *Theorem 6.22: Integration by Parts*
Suppose $F$ and $G$ are differentiable on $[a, b]$, $f = F'$, $g = G'$ on $[a, b]$, and $f$ and $g$ are integrable.
Then $Fg$ and $fG$ are integrable, and

$$\int\limits_a^b Fg = F(b)G(b) - F(a)G(a) - \int\limits_a^b fG$$

Proof: There's really there's not much to do - only verifying that it's legal to apply *FTC #2* to a function $Fg + fG$ obtained from the product rule for derivatives.
$F$ and $G$ are continuous, so they're integrable.
The product of integrable functions is integrable, so $Fg$ and $fG$ are integrable.
The derivative of $FG$ is $fG + Fg$, which is integrable since the sum of integrable functions is integrable.

By part *FTC #2* we know that

$$\int_a^b Fg + \int_a^b fG = \int_a^b (Fg + fG) = \int_a^b (FG)' = F(b)G(b) - F(a)G(a)$$

That's pretty much it.

(e) *Theorem: Change of Variables (not in book)*
Let $g$ be differentiable on $[a, b]$, $f = F'$ be integrable on $[g(a), g(b)]$ (swap end-points if necessary), and suppose that $(f \circ g)g'$ is integrable on $[a, b]$.
Then

$$\int_{g(a)}^{g(b)} f = \int_a^b (f \circ g)g'$$

Proof: This one is also straightforward.
Apply *FTC #2* to the function $F$, whose derivative is $f$, and is integrable by assumption:

$$\int_{g(a)}^{g(b)} f = F(g(b)) - F(g(a))$$

Apply *FTC #2* to the composite function $F \circ g$, whose derivative is $(f \circ g)g'$ by the *Chain Rule*, and is integrable by assumption:

$$\int_a^b (f \circ g)g' = (F \circ g)(b) - (F \circ g)(a) = F(g(b)) - F(g(a))$$

So, they're the same.

19. (p.135) Integration of vector-valued and complex-valued functions.
Upshot: Everything is definied and done componentwise.
The integral is definied when each componenent is integrable.

20. (p.135) *Theorem: Triangle Inequality for Integrals of Vector-valued Functions*

If $f : [a, b] \to \mathbb{R}^k$ (or $\mathbb{C}$) is integrable, then

$$\left| \int_a^b f \right| \le \int_a^b |f|$$

Proof: This proof involves an important technique for manipulating sums and integrals. It's worth understanding, since you'll probably see the similar maneuver appear in other places.

Write $f = (f_1, f_2, \ldots, f_k)$.

Write $\displaystyle\int_a^b f = \int_a^b (f_1, f_2, \ldots, f_k) = \left( \int_a^b f_1, \int_a^b f_2, \ldots, \int_a^b f_k \right)$.

Notice that

$$
\begin{aligned}
\left| \int_a^b f \right|^2 &= \left( \int_a^b f_1 \right)^2 + \left( \int_a^b f_2 \right)^2 + \cdots + \left( \int_a^b f_k \right)^2 \\
&= A_1 \int_a^b f_1 + A_2 \int_a^b f_2 + \cdots + A_k \int_a^b f_k \qquad \left(\text{where } A_j = \int_a^b f_j \right) \\
&= \int_a^b A_1 f_1 + \int_a^b A_2 f_2 + \cdots + \int_a^b A_k f_k \\
&= \int_a^b (A_1 f_1 + A_2 f_2 + \cdots + A_k f_k) \\
&\le \int_a^b \Big( \left| (A_1, A_2, \ldots, A_k) \right| \left| (f_1, f_2, \ldots, f_k) \right| \Big) \qquad (\textit{Schwarz Inequality}, \text{ integral respects ``}\le\text{''}) \\
&= \left| (A_1, A_2, \ldots, A_k) \right| \left( \int_a^b \left| (f_1, f_2, \ldots, f_k) \right| \right) \\
&= \left| (A_1, A_2, \ldots, A_k) \right| \left( \int_a^b |f| \right) \\
&= \left| \int_a^b f \right| \left( \int_a^b |f| \right)
\end{aligned}
$$

We just showed that

$$\left| \int_a^b f \right|^2 \le \left| \int_a^b f \right| \left( \int_a^b |f| \right)$$

That's the important part. The rest is simple.

21. (p.136) Definition of *curves* $\gamma : [a, b] \to \mathbb{R}^k$.

22. Crucial: A curve is a *function*, and *not* the *image* of that function.

23. (p.136) A curve $\gamma$ is an *arc* if it's one-to-one (injective).

24. (p.136) A curve $\gamma$ is an *closed* if $\gamma(a) = \gamma(b)$.

25. *Continuously differentiable* on a domain means the derivative exists on and is continuous on the domain.

26. The *length* $\Lambda(\gamma)$ of a curve $\gamma$ which is *continuously differentiable* on $[a, b]$.

$$\Lambda(\gamma) = \int_a^b |\gamma'|$$

# 7 Chapter 7: Sequences and Series of Functions

Note #1: This chapter should be called "When Can You Interchange Limits?".

Note #2: The functions in this chapter are complex-valued (which includes real-valued, since $\mathbb{R} \subset \mathbb{C}$), unless indicated otherwise.

Note #3: Most of the results actually generalize to functions with more general codomains.

1. (p.143) *Pointwise limit* of a sequence $\{f_n\}$ of functions.

2. *Pointwise limit* of a series $\sum_n f_n$ of functions defined accordingly.

3. The domain of the limiting function is set of points where have convergence.

4. (p.144) The main questions:

   - When are pointwise limits continuous?
   - When can you interchange pointwise limit and derivative operations?
   - When can you interchange pointwise limit and integral operations?
   - In general, when are limiting operations preserved under pointwise limits?

   Example:

   When is it true that

   $$\lim_{t \to x} \left( \lim_{n \to \infty} f_n(t) \right) ?{=}? \lim_{n \to \infty} \left( \lim_{t \to x} f_n(t) \right)$$

   Counterexample (with $m \to \infty$ and $n \to \infty$):

   $$s_{m,n} = \frac{m}{m+n}$$

5. (p.145-146) Useful counterexamples.

6. (p.147) Definition of *uniform convergence* of sequence or series of functions.
   Note! It is *with respect to* a given domain.
   Might have uniform convergence on a smaller domain but not a larger one.

7. (p.147) It is easy to see (look at the order of the four involved quantifiers $\forall$, $\exists$, $\forall$, and $\forall$) that uniform convergence implies pointwise convergence.

8. (p.147) Definition: *Uniformly Cauchy* sequence of functions $\{f_n\}$ on a common domain:

   For every $\epsilon > 0$ there is an $N$ such that $|f_n(x) - f_m(x)| < \epsilon$
   whenever $m > N$, $n > N$, and $x$ is in the domain.

   Note #1: Prove to yourself that a *uniformly Cauchy* sequence of functions is *pointwise Cauchy*, meaning that $\{f_n(x)\}$ is a Cauchy sequence at each point $x$ in the domain.

   Note #2: Prove to yourself that for a *pointwise Cauchy* sequence $\{f_n\}$, there is a limit function $f$ such that, for every $x$ in the domain, $f_n(x) \to f(x)$ as $n \to \infty$.

9. (p.147) *Theorem: Cauchy criterion for uniform convergence*

   A sequence $\{f_n\}$ of functions on a common domain converges uniformly (to some $f$) on that domain if and only if the sequence $\{f_n\}$ is *uniformly Cauchy*.

   Proof: The "only if" direction is an easy $\epsilon/2$ argument.

   The "if" direction is more interesting, and worth looking at since it's' a standard technique. Suppose that $\{f_n\}$ is *uniformly Cauchy*, pick $\epsilon$, and find $N$ accordingly. Let $x$ be a point in the domain, and let $n > N$. Then

   $$|f_n(x) - f_m(x)| < \epsilon$$

   if $m > N$. Let $f$ be the limit function, which exists since $\{f_n\}$ is *pointwise Cauchy*. Keeping $n$ and $x$ fixed, and letting $m \to \infty$, it follows that that

   $$|f_n(x) - f(x)| \le \epsilon$$

   The last step follows from *Preserving an Inequality in a Limit*.

10. (p.148) Very useful and trivial-to-prove fact:

    $f_n \to f$ uniformly on $E$ if and only if $\sup_{x \in E} |f_n(x) - f(x)| \to 0$.

11. (p.148) *Theorem: Weierstrass M-Test*

Let $\{M_n\}$ be some sequence of positive numbers.

Suppose $\{f_n\}$ is a sequence of functions on $E$ such that $|f_n(x)| \leq M_n$ for all $n$ and $x \in E$.

If the series $\sum_n M_n$ converges, then the series $\sum_n f_n$ converges uniformly and absolutely.

(Rudin does not mention that convergence is absolute, but it should come out in the proof.)

Proof: This is an easy application of the stuff about *uniformly Cauchy* sequences.

12. (p.149) *Main Uniform Convergence Limit Interchange Theorem ("MUCLIT")*

If

(a) $f_n \to f$ uniformly on a set $E$ in a metric space,
(b) $x$ is a limit point of $E$, but not necessarily in $E$, and
(c) $\lim_{t \to x} f_n(t) = A_n$ exists for every $n$,

then

(a) $\lim_{n \to \infty} A_n$ exists,
(b) $\lim_{t \to x} f(t)$ exists, and
(c) $\lim_{n \to \infty} A_n = \lim_{t \to x} f(t)$

Conclusion (c) says that you can interchange limits:

$$\lim_{n \to \infty} \left( \lim_{t \to x} f_n(t) \right) = \lim_{t \to x} \left( \lim_{n \to \infty} f_n(t) \right)$$

This theorem is extremely important. It is the basis used to prove, for example, that the uniform limit of continuous functions is continuous, and that power series are termwise differentiable inside the radius of convergence.

Proof: It's an important one, sometimes called an "$\epsilon/3$ argument". Here's how it goes.

- Pick $\epsilon > 0$.
- Find some $N$ such that $|f_m - f| < \epsilon$ on $E$ (uniformly) for every $m > N$.
- If $t \in E$, $m > N$, and $n > N$, then the *Triangle Inequality* gives

$$|f_m(t) - f_n(t)| \leq |f_m(t) - f(t)| + |f(t) - f_n(t)| \leq 2\epsilon$$

- The absolute value function $z \mapsto |z|$ is continuous, so by *Continuous Functions Respect Limits in Metric Spaces* and *Preserving an Inequality in a Limit*, we get

$$|A_m - A_n| = \lim_{t \to x} |f_m(t) - f_n(t)| \leq 2\epsilon$$

- This shows $\{A_n\}$ is Cauchy and has some limit

$$A = \lim_{n \to \infty} A_n$$

.

- The following and final step is the "$\epsilon/3$ argument" part.
  By the *Triangle Inequality*,

$$|f(t) - A| \leq |f(t) - f_n(t)| + |f_n(t) - A_n| + |A_n - A|$$

  Given $\epsilon > 0$ there is an $N$ large enough so that both
  · $|f - f_N| < \epsilon$ (uniformly), and
  · $|A_N - A| < \epsilon$
  For this $N$, find $\delta > 0$ such that $|f_N(t) - A_N| < \epsilon$ when $|t - x| < \delta$. Then whenever $|t - x| < \delta$, we have

$$\begin{aligned} |f(t) - A| &\leq |f(t) - f_N(t)| + |f_N(t) - A_N| + |A_N - A| \\ &\leq \epsilon + \epsilon + \epsilon \\ &= 3\epsilon \end{aligned}$$

- The previous step completes the proof, if you replace $\epsilon$ with $\epsilon/3$.
  Also, notice that in this final step we only needed *one* value $N$ to get control over $f$. The full power of uniform convergence was used earlier to show that $\{A_n\}$ is Cauchy.

13. (p.150) *Corollary: The Uniform Limit of Continuous Functions is Continuous*

    If $\{f_n\}$ is a sequence of continuous functions on $E$, and $f_n \to f$ uniformly on $E$, then $f$ is continuous.

    Proof: There's nothing to show at isolated points. At limit points, just apply the previous theorem.

14. (p.150) Interesting fact:

    A (pointwise) monotone decreasing, (pointwise) convergent sequence of continuous functions on a compact metric space which happens to converge (pointwise) to a continuous function is necessarily *uniformly* convergent.

    Proof: A good example of how compactness lets you control things!
    Let $\{f_n\}$ be the sequence of functions on compact domain $K$.

Show that WLOG each $f_n \geq 0$ on $K$.
Pick $\epsilon > 0$, and look at $\bigcap_n K_n$, where $K_n = \{x \in K : f_n(x) \geq \epsilon\}$.
Apply result about intesections of compact sets in a Hausdorff space.

15. (p.150) Definition of

$$\mathscr{C}(X) = \{\text{complex-valued, continuous, bounded functions on } X\}$$

Boundedness is obviously redundant if $X$ is compact.

16. (p.150) Definition of of *supremum norm* (or *infinity norm*) for *bounded* functions $f$ on $X$:

    - $\|f\| = \|f\|_\infty = \sup_{x \in X} |f(x)|$.
    - Note: It's weird that Rudin officially defines $\|f\|_\infty$ only for $f \in \mathscr{C}(X)$ which means he's requiring $f$ to be both *continuous and bounded* on $X$. But *boundedness* is the only requirement neededed for the definition to make sense.
    - Note: This definition of *supermum norm* might be slightly different from one you might see in a measure theory course.
    - The supremum norm has a *Triangle Inequality*:

    $$\|f + g\|_\infty \leq \|f\|_\infty + \|g\|_\infty$$

    since for all $x \in X$,

    $$\big|f(x) + g(x)\big| \leq \big|f(x)\big| + \big|g(x)\big| \leq \|f\|_\infty + \|g\|_\infty$$

    - Turns $\mathscr{C}(X)$ into a metric space with metric

    $$d(f, g) = \|f - g\|_\infty$$

    - Finally, notice that the supremum norm is "submultiplicative":

    $$\|fg\|_\infty \leq \|f\|_\infty \|g\|_\infty$$

    since for all $x \in X$,

    $$\big|f(x)g(x)\big| = \big|f(x)\big|\big|g(x)\big| \leq \|f\|_\infty \|g\|_\infty$$

17. (p.151) *Theorem: Some Facts About The Metric Space* $\mathscr{C}(X)$

    With the *supermum norm* metric on $\mathscr{C}(X)$, we have:

    (a) Let $f \in \mathscr{C}(X)$ and $f_n \in \mathscr{C}(X)$ for each $n = 1, 2, 3, \ldots$
    Then $f_n \to f$ *uniformly* on $X$ if and only if $f_n \to f$ in the metric space $\mathscr{C}(X)$.

    (b) $\mathscr{C}(X)$ is a *complete* metric space.

Proof:

Part (a) follows immediately from a fact on p.148.

Part (b) is also easy. If $\{f_n\}$ is sequence in $\mathscr{C}(X)$ which is Cauchy w.r.t. the $\mathscr{C}(X)$ metric, then (verify this!) it is *uniformly Cauchy* on $X$. By the (p.147) *Cauchy criterion for uniform convergence*, $\{f_n\}$ converges *uniformly* on $X$ to some $f$.

Notice that $f$ must be in $\mathscr{C}(X)$, because it's bounded (compare with $f_N$ for some large $N$) and continuous (it's the uniform limit of continuous functions). Finally, apply Part (a).

18. (p.151) *Theorem: Termwise Integration of Uniformly Convergent Sequences (TIUCS)* (Integration is a process involving limits, so this is a theorem about interchanging limits.)

    If

    (a) $f_n \to f$ uniformly on $[a, b]$ as $n \to \infty$
    (b) each $f_n$ is integrable on $[a, b]$

    then

    (a) $f$ is integrable on $[a, b]$
    (b) $\int_a^b f_n \to \int_a^b f$ as $n \to \infty$

    Proof: Not bad at all.

    First, use uniform convergence to find $N$ with $|f_n - f| < \epsilon$ uniformly when $n > N$. Verify that this traps the upper and lower integrals of $f$ within $\epsilon(b-a)$ of the integral of $f_n$, and so within $2\epsilon(b-a)$ of each other. Since $\epsilon$ is arbitrary, the upper and lower integrals of $f$ are equal and so $f$ is integrable.

    Next, just notice that if $n > N$ (as above), then $|f_n - f| < \epsilon$ guarantees that

$$\left| \int_a^b f_n - \int_a^b f \right| \leq \int_a^b |f_n - f| \leq \epsilon$$

    by the *Triangle Inequality for Integrals.*

19. (p.152) Corollary:

    If the series $\sum_n f_n \to f$ uniformly on $[a, b]$ as $n \to \infty$, and each $f_n$ is integrable, then

$$\sum_n \left( \int_a^b f_n \right) = \int_a^b f = \int_a^b \left( \sum_n f_n \right)$$

That the limiting function $f = \sum_n f_n$ is integrable and the series $\sum_n \int_a^b f_n$ on the left side converges are parts of the conclusion.

Proof: Use the previous *(TIUCS)* theorem, along with the property that the integral of a *finite* sum is the sum of the integrals.

20. (p.152 & p.146) Fact: Uniform convergence by itself doesn't give you control over derivatives of *real-valued* functions.
    So differentiation is not as nice as integration in this way.

    Consider this counterexample on $\mathbb{R}$: $f_n(x) = \dfrac{\sin nx}{\sqrt{n}}$

    Note: If you study *complex analysis* later, you will find that the *complex derivative* of a *complex-valued* function of one *complex* variable behaves more nicely. It's almost like a completely different thing.

21. (p.152) *Theorem: Criteria for Termwise Differentiation of Sequences (CTDS)*

    This can be used (see p.173) to give an easy proof that power series are termwise differentiable.
    The proof itself is also important.

    Let $\{f_n\}$ be a sequence of differentiable real-valued functions on $[a, b]$, and suppose

    (a) there is some $x_0 \in [a, b]$ such that the sequence $\{f_n(x_0)\}$ converges, and
    (b) the sequence $\{f_n'\}$ converges *uniformly* on $[a, b]$.

    Then

    (a) the sequence $\{f_n\}$ converges *uniformly* on $[a, b]$,
    (b) the limit function $f = \lim_{n\to\infty} f_n$ is differentiable on $[a, b]$, and
    (c) $f_n' \to f'$ on $[a, b]$ as $n \to \infty$ (with convergence uniform by assumption (b)).

    Proof: Here's the gist of it. Notice that

    $$|f_n(x) - f_m(x)| \le |f_n(x) - f_m(x) - f_n(x_0) + f_m(x_0)| + |f_n(x_0) - f_m(x_0)|$$

and then use the hypotheses along with the *MVT* to show $\{f_n\}$ is uniformly Cauchy, so there is a limit function $f$ on $[a, b]$. To show that $f'$ exists and equals the limit of $f_n'$ on $[a, b]$, apply *MUCLIT* to the sequence of functions

$$\phi_n(t) = \frac{f_n(t) - f_n(x)}{t - x}$$

where $x \in [a, b]$, and $t \neq x$. Use the *MVT* as before to show $\{\phi_n\}$ is uniformly Cauchy, to justify applying *MUCLIT*.

22. (p.154) *Theorem: Existence of a nowhere differentiable function*

There is a continous function on $\mathbb{R}$ which is "nowhere differentiable", i.e. not differentiable at any point.

The function is constructed in a "*fractal*-like" way as the uniform limit of a sequence of continuous functions obtained from $x \mapsto |x|$ ($x \in [-1, 1]$).

The example function Rudin gives is essentially the *Weierstrass Function*.
(See `https://en.wikipedia.org/wiki/Weierstrass_function`.)

There is a step in Rudin's proof of which you might get stuck on if you haven't seen something similar before, or if you're not especially fluent in manipulating inequalties (as I am not). The difficulty here is that there are actually several small steps rolled into one. Rudin is implicitly challenging you to work it out for yourself, and accepting such challenges is part of reading the book correctly.

Here it is (p.154):

*(Note: Rudin has already shown that $|\gamma_m| = 4^m$, and $|\gamma_n| \leq 4^n$ for $0 \leq n \leq m$.)*

Since $|\gamma_m| = 4^m$, we conclude that

$$\begin{aligned}\Big[\text{The Important Quantity}\Big] &= \left| \sum_{n=0}^{m} \left(\frac{3}{4}\right)^n \gamma_n \right| \\ &\geq 3^m - \sum_{n=0}^{m-1} 3^n \\ &= \frac{1}{2}\left(3^m + 1\right)\end{aligned}$$

There are two things going on here:

- The *Triangle Inequality*:

$$3^m = \left| \left(\frac{3}{4}\right)^m \gamma_m \right| = \left| \sum_{n=0}^{m} \left(\frac{3}{4}\right)^n \gamma_n + \left[(-1) \sum_{n=0}^{m-1} \left(\frac{3}{4}\right)^n \gamma_n \right] \right|$$

$$
\begin{aligned}
&\le \left|\sum_{n=0}^{m}\left(\frac{3}{4}\right)^n\gamma_n\right| + \left|\sum_{n=0}^{m-1}\left(\frac{3}{4}\right)^n\gamma_n\right| && \textit{(Triangle Inequality)}\\
&= \Big[\text{The Important Quantity}\Big] + \left|\sum_{n=0}^{m-1}\left(\frac{3}{4}\right)^n\gamma_n\right|\\
&\le \Big[\text{The Important Quantity}\Big] + \sum_{n=0}^{m-1}\left(\frac{3}{4}\right)^n|\gamma_n| && \textit{(Triangle Inequality)}\\
&\le \Big[\text{The Important Quantity}\Big] + \sum_{n=0}^{m-1}3^n
\end{aligned}
$$

so

$$3^m - \sum_{n=0}^{m-1}3^n \le \Big[\text{The Important Quantity}\Big]$$

- The sum of a *finite geometric series* for $r \neq 1$:

$$\sum_{n=0}^{m-1} r^n = \frac{r^m - 1}{r - 1}$$

so when $r = 3$

$$\sum_{n=0}^{m-1} 3^n = \frac{1}{2}\left(3^m - 1\right)$$

The thing that remains is to notice that

$$3^m - \sum_{n=0}^{m-1}3^n = 3^m - \frac{1}{2}\left(3^m - 1\right) = \frac{1}{2}\left(3^m + 1\right)$$

23. (p.154) Question motivating looking at *equicontinuous families* of functions:

    Given a sequence of functions, when is there a pointwise or uniformly convergent susequence?

    We will see that, on compact domains, *equicontinuity* is closely related to uniform converge.

24. (p.155) Definition: *pointwise bounded* sequence $\{f_n\}$ of functions on a common domain $E$:

    $\{f_n(x)\}$ is a bounded sequence for every point $x \in E$ in the domain.

25. (p.155) Definition: *uniformly bounded* sequence $\{f_n\}$ of functions on a common domain $E$:

    There is a (common) bound $M$ such that $|f_n(x)| < M$ for all $x \in E$ and all $n = 1, 2, 3, \ldots$

26. (p.155) As usual (like with uniform continuity and uniform convergence), being *uniformly bounded* is a relatively "strong" condition:

    If the sequence $\{f_n\}$ is *uniformly bounded* then it is *pointwise bounded*, and each function $f_n$ is a *bounded function.*

    But *the converse is not true in general (meaning without uniform convergence)*, even with continuous complex-valued functions on a compact domain.
    See the counterexample below.

27. (Not in text) Counterexample:

    Define a sequence $\{f_n\}$ of functions on $[0, 1]$ by

    $$f_n(x) = \begin{cases} n^2 x & \text{if } 0 \le x \le \frac{1}{n} \\ \frac{1}{x} & \text{if } \frac{1}{n} < x \end{cases}$$

    This is a sequence of continuous complex-valued functions on a compact domain which such that:

    - $\{f_n\}$ is *pointwise convergent*:
      $$f_n(0) = 0 \text{ for all } n, \text{ and } f_n(x) \to \frac{1}{x} \text{ when } x > 0.$$
    - $\{f_n\}$ is *pointwise bounded*, since it's pointwise convergent.
    - Each function $f_n$ is *a bounded function*, since $f_n(x) \le n$ for every $x$.

    But $\{f_n\}$ is *not uniformly bounded*, since $f_n(\frac{1}{n}) = n$.

    Note that the sequence $\{f_n\}$ in this counterexample is *not uniformly convergent.*

28. (Not in text) *Fact: Uniformly Convergent Sequences of Bounded Functions (UCSBF)*

    If $\{f_n\}$ *converges uniformly* to $f$, and each function $f_n$ is a *bounded function*, then $\{f_n\}$ is *uniformly bounded*, and $f$ is a *bounded function.*

    Proof:

    The sequence is uniformly Cauchy, which means that for some (fixed) large $N$ and $N < n$, $f_n$ is "uniformly within" $\epsilon$ of $f_N$. So if $|f_N| < M$ then $|f_n| \le M + \epsilon$. There are only finitely many remaining functions $f_1, f_2, \ldots, f_{N-1}$ to take into conseration. Look at the the maximum of $M + \epsilon$ and the bounds for $f_1, f_2, \ldots, f_{N-1}$.

    Since $\{f_n\}$ is *uniformly bounded*, there is a constamt $K$ such that

    $$|f_n(x)| \le K$$

    for all $x$ in the (common) domain and all $n$. "Passing to the limit" (make sure you understand what's happening here!), we see that

    $$|f(x)| \le K$$

for all $x$, and so $f$ is bounded.

29. (p.155-156) Important Counterexamples:

    - (p.155) $f_n(x) = \sin nx$, with $x \in [0, 2\pi]$

      This is a *uniformly bounded* sequence of continuous functions on a compact domain.
      But there does not exist a pointwise convergent subsequence.

      Note: It is tricky to prove this at this stage. Rudin invokes a result, called *Lebesgue Dominated Convergence*, from Chapter 11. You should skip Chapter 11 in this book, and instead learn the related material from Rudins other book *Real & Complex Analysis*.

    - (p.156) $f_n(x) = \dfrac{x^2}{x^2 + (1 - nx)^2}$, with $x \in [0, 1]$

      This is a *pointwise* convergent and *uniformly bounded* sequence of continuous functions on a compact domain.
      But there does not exist a *uniformly* convergent subsequence.

      Proof: Notice that $f_n \to 0$ pointwise, but $f_n(1/n) = 1$.

    So basically, we need to find some new condition or concept to get better guarantees...
    The needed condition is *equicontinuity*.

30. (p.156) *Equicontinuity* relates three things:

    - uniform continuity,
    - uniform boundedness, and
    - uniform convergence.

31. (p.156) Definition: *equicontinuous*

    A family of complex functions $\mathscr{F}$ defined on a metric space $E$ is *equicontinuous* if:

    For every $\epsilon$ there exists a $\delta$ such that

    $$|f(x) - f(y)| < \epsilon$$

    whenever $f \in \mathscr{F}$, $x \in E$, and $y \in E$, and $d(x, y) < \delta$.

    - *Equicontinuity* is a pretty "strong" condition.
    - Given $\epsilon$, the *same* $\delta$ works for *all* functions and pairs of points.
    - Obviously each function $f \in \mathscr{F}$ is uniformly continuous (and so continuous).

32. (p.156) *Theorem: Diagonal Construction Theorem*

A *pointwise bounded* sequence $\{f_n\}$ of complex-valued functions on a *countable* domain has a *pointwise convergent* subsequence.

This theorem provides a kind of "selection process", which is useful not only here in this chapter but also in other more general situations.

The technique in the proof is also a very important - maybe more important than the theorem itself. It will help a lot to draw a picture or look at Rudin's on p.157.

Proof: This amounts basically to repeated applications of compactness.

Let $x_1, x_2, x_3, \dots$ be an enumeration of the (common) domain of the functions $f_n$.

The sequence $\{f_n(x_1)\}$ is bounded, so (compactness applied here) has a convergent subsequence, say $f_{n(k)}(x_1)$, $k = 1, 2, 3, \dots$. We now have a pointwise bounded subsequences of functions $\{f_{n(k)}\}$ which converges on the point $x_1$. Let $\{f_n^1\}$ be this subsequence $\{f_{n(k)}\}$.

The sequence $\{f_n^1(x_2)\}$ is bounded, so (compactness applied here) has a convergent subsequence, say $f_{n(k)}^1(x_2)$, $k = 1, 2, 3, \dots$. We now have a pointwise bounded subsequences of functions $\{f_{n(k)}^1\}$ which converges on the point $x_2$. Let $\{f_n^2\}$ be this subsequence $\{f_{n(k)}^1\}$.

"Continue in this way", meaning inductively define $\{f_n^{k+1}\}$ as a subsequnce of $\{f_n^k\}$.

We now have a "sequence of sequences" $\{f_n^k\}$ such that:

- $\{f_n^k(x_k)\}$ converges
- $\{f_n^{k+1}\}$ is a subsequnce of $\{f_n^k\}$
- Each $\{f_n^k\}$ is a subsequnce of the original sequence $\{f_n\}$

The final step in the proof is the "diagonal step". Define the sequence of functions $\{g_n\} = \{f_n^n\}$, whose $n$-th term is the $n$-th function of the $n$-th sequence.

Verify that $\{g_n\}$ does what we want it to do. The only tricky thing might be convincing yourself that $\{g_n\}$ *actually is a subsequence* of $\{f_n\}$ and every $\{f_n^k\}$. Recall (or prove if you haven't noticed it before) the obvious fact that the composite of strictly increasing functions is a strictly increasing function, and apply to the sequence indices.

33. (p.157) *Theorem:*

A uniformly convergent sequence $\{f_n\}$ of continuous, complex-valued functions on a compact metric space is equicontinuous.

Proof: This is another important proof.

(Don't be confused by the fact that Rudin introduced the *Diagonal Construction Theorem* immediately before this. It's not used in this proof.)

This is basically an "$\epsilon/3$-argument". The main idea is that "most" of the functions $f_n$ are very close to some particular $f_N$ for a "large-enough" $N$, and you you only have to manage finitely many "exceptions".

Note that

- each function $f_n$ is uniformly continuous on the domain since it's compact, and
- the sequence $\{f_n\}$ is uniformly Cauchy.

Picka n arbitrary $\epsilon > 0$.
First, find $N$ large so that $\|f_N - f_n\|_\infty < \dfrac{\epsilon}{3}$ when $n > N$.
Next, for $k = 1, 2, \ldots, N$ find $\delta_k$ by uniform continuity so that $|f_k(x) - f_k(y)| < \dfrac{\epsilon}{3}$ when $d(x, y) < \delta_k$.
Finally, let $\delta = \min\{\delta_1, \delta_2, \ldots, \delta_N\}$.

Verify that $\delta$ does the job by applying the *Triangle Inequality* a couple of times. Handle cases $n \leq N$ and $n > N$ separately.

34. (p.158) *Theorem: Arzela's Theorem*

If $\{f_n\}$ is a pointwise bounded and equicontinuous sequence of complex functions on a compact domain $K$, then

(a) $\{f_n\}$ is uniformly bounded on $K$, and
(b) $\{f_n\}$ contains a uniformly convergent subsequence.

Proof of conclusion (a):

Pick any $\epsilon > 0$, say $\epsilon = 1$ for example. The actual value doesn't matter.
By equicontinuity find $\delta$ and then by compactness cover $K$ with *finitely* many open balls $B_\delta(x_i)$ of radius $\delta$ centered at $x_i$, $i = 1, 2, \ldots, M$, such that

$$|f_n(y) - f_n(x_i)| < \epsilon$$

for all $n$, all $x_i$, and all $y$ such that $y \in B_\delta(x_i)$.

Since $\{f_n\}$ is pointwise bounded, for each of the points $x_i$, $i = 1, 2, \ldots, M$, there is bound $L_i$ with

$$|f_n(x_i)| \leq L_i, \text{ for every } n$$

Let $L = \max\{L_1, L_2, \ldots, L_M\}$
Given an arbitrary $n = 1, 2, \ldots$, and given an arbitrary $y \in K$, we see that

$$\left|f_n(y)\right| = \left|f_n(x_i) + \big(f_n(y) - f_n(x_i)\big)\right| \le \left|f_n(x_i)\right| + \epsilon \le L + \epsilon$$

for whichever $x_i$ is within $\delta$ of $y$.

This proves conclusion (a) of the theorem.

Note: The proof of conclusion (b) relies on exercise Chapter 2 #25 (p.45), illustrating why completing the exercises is integral to understanding the text.

Proof of conclusion (b):

Crucial step #1: Let $E \subset K$ be a countable dense subset provided by exercise Chapter 2 #25 (p.45). By the *Diagonal Construction Theorem*, there is a subsequence $\{g_n\}$ of $\{f_n\}$ which converges at each point of $E$.

We'll now make a straightforward $\epsilon/5$ argument, which is application of and illustrates the strength of equicontinuity.

Crucial step #2: Pick an arbitrary $\epsilon > 0$, and by equicontinuity find $\delta$ and then by compactness cover $K$ with *finitely* many open balls $B_\delta(x_i)$ of radius $\delta$ centered at $x_i$, $i = 1, 2, \ldots, M$, such that

$$|g_n(y) - g_n(x_i)| < \epsilon$$

for all $n$, all $x_i$, and all $y$ such that $y \in B_\delta(x_i)$.

Since $E$ is dense, each ball $B_\delta(x_i)$, $i = 1, 2, \ldots, M$, contains some point of $E$, call it $p_i$.

Note that $\{g_n\}$ converges uniformly on $\{p_1, p_2, \ldots, p_M\}$.

(*Convince yourself, if you have not already, that pointwise and uniform convergence are the same thing on a finite set such as* $\{p_1, p_2, \ldots, p_M\}$.)

So there exists an $N$ such that whenever $n > N$ and $m > N$ we get

$$|g_m(p_i) - g_n(p_i)| < \epsilon, \text{ for each } p_i \in \{p_1, p_2, \ldots, p_M\}$$

An arbitrary $y \in K$ is within $\delta$ of some $x_i$, and $x_i$ is within $\delta$ of $p_i$.

If $n > N$ and $m > N$ we get

$$\begin{aligned}
|g_m(y) - g_n(y)| &= \Big|(g_m(y) - g_m(x_i)\Big| \\
&\qquad + \Big|g_m(x_i) - g_m(p_i)\Big| + \Big|g_m(p_i) - g_n(p_i)\Big| \\
&\qquad + \Big|g_n(p_i) - g_n(x_i)\Big| + \Big|g_n(x_i) - g_n(y))\Big| \\
&\leq \epsilon \\
&\qquad + \epsilon + \epsilon \\
&\qquad + \epsilon + \epsilon \\
&= 5\epsilon
\end{aligned}$$

Since $\epsilon$ was arbitrary we can repeat the whole argument with $\epsilon$ replaced by $\epsilon/5$, and this shows uniform convergence of $\{g_n\}$ on all of $K$.

We have proven conclusion (b).

This completes the proof of the theorem.

35. (p.159) *Theorem: Weierstrass Polynomial Uniform Approximation Theorem (WPUAT)*

Given a continuous complex function $f$ on $[a, b]$ and $\epsilon > 0$, there is a polynomial $P$ with

$$\Big|f(x) - P(x)\Big| < \epsilon \text{ for all } x \in [a, b]$$

In other words, *any continuous on* $[a, b]$ *can be "uniformly approximated" on* $[a, b]$ *within* $\epsilon$ *by a polynomial.*

In terms of the metric space of continuous (and by compactness necessarily bounded) functions on $[a, b]$, with the metric obtained from $\|\cdot\|_\infty$, we have

$$\|f - P\|_\infty \leq \epsilon$$

Said yet another way, the set of polynomials on $[a, b]$ is *dense* in the metric space of continuous functions on $[a, b]$.

*Note: I'm going to dig much more deeply into this result than previous ones, because there is a lot in the proof, and if you're like me then things might seem pretty mysterious at first.*

Rudin has given us a super-important-to-digest-and-internalize proof (p.159-160), which is in some ways kind of difficult, but in other ways also quite simple.

- The proof is difficult because you might not have any idea what the heck he's thinking, where he's coming from, or where he's going.

  Rudin is leaving it up to you to figure out the motivation for what's going on. In the long run, you will likely thank him for this.

- The proof is simple because once you figure out what the general idea is, you can pretty much just plow forward using brute force with a little cleverness sprinkled in.

Here's a sketch of the proof.

(a) Motivational Observation #1:
If $f$ happens to be uniformly continuous on some open interval $(c, d)$ containing $[a, b]$, and $g_n(x)$ is the average value of $f(x)$ on $[x - \frac{1}{2n}, x + \frac{1}{2n}]$, then the sequence $\{g_n\}$ converges uniformly to $f$ on $[a, b]$.

More precisely, suppose $f$ is uniformly continuous on some "sufficiently large" open interval $(c, d)$ containing $[a, b]$, and define $g_n(x)$, for $x \in [a, b]$, to be the average value of $f$ on $[x - \frac{1}{2n}, x + \frac{1}{2n}]$, i.e.

$$g_n(x) = \int_{x-\frac{1}{2n}}^{x+\frac{1}{2n}} nf(t)\ dt$$

Notice, recall, or prove to yourself that $g_n \to f$ uniformly on $[a, b]$ as $n \to \infty$.

(b) Motivational Observation #2:
What's about to happen here is an idiomatic thing in analysis.
It might at first just look like a trick, but...

> *"If you use a trick enough times then it becomes a technique."*
> -David Barrett
>
> (This was paraphrased; I heard it in his single variable complex analysis class around 1996.)

You can rewrite the integrals from the previous observation by "integrating against" auxiliary functions $h_n$:

$$\begin{aligned} g_n(x) &= \int_{x-\frac{1}{2n}}^{x+\frac{1}{2n}} h_n(t-x)f(t)\ dt \\ &= \int_a^b h_n(t-x)f(t)\ dt \end{aligned}$$

where

$$h_n(s) = \begin{cases} n & \text{if } -\frac{1}{2n} \leq s \leq \frac{1}{2n} \\ 0 & \text{otherwise} \end{cases}$$

(c) Motivational Observation #3:

If $Q(x)$ is a polynomial in $x$, and

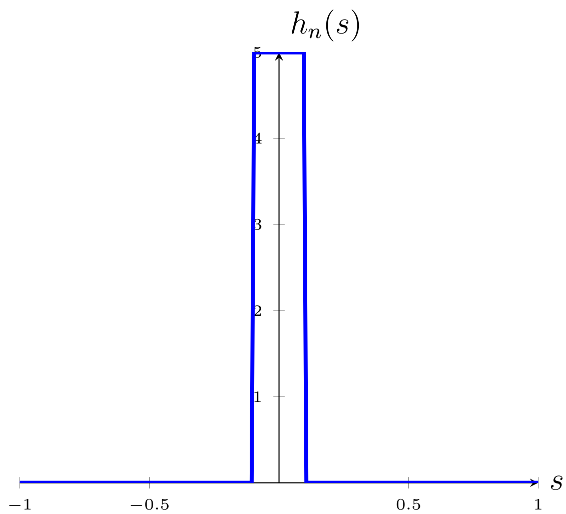
$$P(x) = \int_a^b Q(t-x)f(t)\ dt$$

then $P(x)$ is a polynomial in $x$.

This becomes clear when you think about how the integrand looks after fully expanding all powers $(t-x)^k$ in $P(t-x)$. What's left is just a sum of terms $c_{(k,j)}f(t)t^{k-j}x^j$, and the $x^j$ is a constant with respect to the variable of integration.

(d) Putting together the previous three observations, we see that there is definitely hope. We are looking for some collection of polynomials $Q_n$ which resemble the previously defined functions $h_n$ in terms of how they behave when they are used to integrate against $f$.

Here's a graph of $h_n$ for $n = 5$:




and also notice that for every $n$

$$\int_{-1}^{1} h_n = 1$$

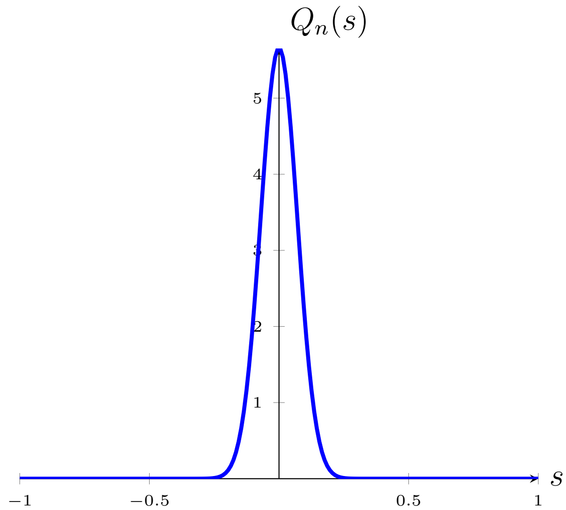
OK here we go. Let's take a look at polynomials

$$Q_n(s) = c_n(1-s^2)^n$$

where the constants $c_n$ are *chosen* so that

$$\int_{-1}^{1} Q_n = 1$$

Here's the graph of $Q_n$ for $n = 100$:

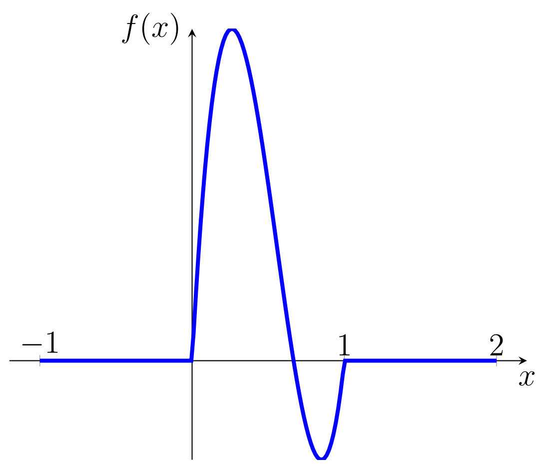



(e) That's the intuition.
The remainder of the proof is showing that the polynomials $P_n$ defined by

$$P_n(x) = \int_0^1 Q(t-x)f(t)$$

do what we want them to do.

(f) Prove to yourself that WLOG you can assume that $[a, b] = [0, 1]$, and $f(0) = 0 = f(1)$.
For convenience later on (change of variable stuff), extend $f$ to all of $[-1, 2]$ by setting $f(x) = 0$ when $x \notin [0, 1]$.
Now $f$ now looks something like this:




(g) Let's try to get control of $|P_n(x) \ - f(x)|$ when $0 \le x \le 1$.
This is brute force with some cleverness.

$$\begin{aligned} \Big|P_n(x) \ - f(x)\Big| &= \left|P_n(x) - f(x) \int_{-1}^1 Q_n(t)\ dt\right| \\ &= \left|\int_0^1 Q_n(t-x)f(t)\ dt - \int_{-1}^1 Q_n(t)f(x)\ dt\right| \end{aligned}$$

(Cleverness: now change/substitute variables in the first integral.)

$$= \left| \int_{-x}^{1-x} Q_n(t) f(t+x)\ dt - \int_{-1}^{1} Q_n(t) f(x)\ dt \right|$$

(Recall $f = 0$ outside of $[0,1]$.)

$$= \left| \int_{-1}^{1} Q_n(t) f(t+x)\ dt - \int_{-1}^{1} Q_n(t) f(x)\ dt \right|$$

$$= \left| \int_{-1}^{1} Q_n(t) \big( f(t+x) - f(x) \big)\ dt \right|$$

$$\leq \int_{-1}^{1} Q_n(t) \Big| f(t+x) - f(x) \Big|\ dt \qquad (***)$$

(h) $f$ is uniformly continuous on $[-1,2]$, so given $\epsilon > 0$ there is a $\delta$ such that

$$|f(x) - f(y)| < \epsilon$$

whenever $|x - y| < \delta$.

(i) Brute force using the uniform continuity $\epsilon$ and $\delta$ pair.
Let $M$ be the maximum value of $|f|$. From the line (***) above, we can get

$$\begin{aligned}
\Big| P_n(x) \ - f(x) \Big| &\leq \int_{-1}^{1} Q_n(t) \Big| f(t+x) - f(x) \Big|\ dt \\
&\leq \int_{-1}^{-\delta} Q_n(t) 2M\ dt + \int_{-\delta}^{\delta} Q_n(t)\epsilon\ dt + \int_{\delta}^{1} Q_n(t) 2M\ dt \\
&= \int_{-\delta}^{\delta} Q_n(t)\epsilon\ dt + 4M \int_{\delta}^{1} Q_n(t)\ dt \\
&\leq \int_{-1}^{1} Q_n(t)\epsilon\ dt + 4M \int_{\delta}^{1} Q_n(t)\ dt \\
&= \epsilon + 4M \int_{\delta}^{1} Q_n(t)\ dt \\
&= \epsilon + 4M \int_{\delta}^{1} c_n (1-t^2)^n\ dt \\
&\leq \epsilon + 4M \int_{\delta}^{1} c_n (1-\delta^2)^n\ dt \\
&\leq \epsilon + 4M c_n (1-\delta^2)^n \qquad (+++)
\end{aligned}$$

What we've done here is another standard technique in analysis. We have broken the integral up into different pieces, and the way we did it depends on $\epsilon$ and ther function $f$, but does not depend on the index $n$.

(j) Line (+++) above applies with the same $\epsilon$ and $\delta$ to all values of the index $n$. This means that the proof will (except for some details like replacing $\epsilon$ with $\epsilon/2$) be complete if we can show

$$0 = \lim_{n \to \infty} c_n (1-\delta^2)^n$$

We will be in good shape as long as the $c_n$ don't grow "too quickly" with $n$, because $(1-\delta^2)^n$ decays exponentially with $n$.

(k) Get a handle on the growth of the $c_n$.
Recalling that

$$\frac{1}{c_n} = \int_{-1}^{1} (1-x^2)^n \, dx$$

our plan will be to show that the integral $\int_{-1}^{1} (1-x^2)^n \, dx$ on the right can't *decay* eponentially fast with $n$.

One way to do this is to "get $n$ out of the exponent" of the integrand. The key observation is that when $-1 \le x \le 1$, we have

$$(1-x^2)^n \ge 1 - nx^2$$

(You can observe/prove this in either of two ways:

i. Apply the *MVT* to the difference $(1-x^2)^n - (1-nx^2)$.
ii. Recall that (when $r \neq 1$)

$$\frac{1-r^n}{1-r} = 1 + r + r^2 + \cdots + r^{n-1}$$

Apply with $r = 1 - x^2$.)

Now that we have this, notice that

$$\begin{aligned} 1 = c_n \int_{-1}^{1} (1-x^2)^n \, dx &= 2c_n \int_{0}^{1} (1-x^2)^n \, dx \\ &\ge 2c_n \int_{0}^{1/\sqrt{n}} (1-x^2)^n \, dx \\ &\ge 2c_n \int_{0}^{1/\sqrt{n}} (1-nx^2) \, dx \\ &= 2c_n \left( \frac{2}{3\sqrt{n}} \right) \end{aligned}$$

by direct integration at the end.

An important thing to notice about the last chain of inequalities is that the domain of integration was restricted from $[0,1]$ to $[0, 1/\sqrt{n}]$. Without doing this we wouldn't get the result that we want (verify this!).

How did we know to "do surgery" like this?

We showed before that $(1-x^2)^n \ge 1 - nx^2$ on $[0,1]$. But the function $1 - nx^2$ is *negative* when $x \in (1/\sqrt{n}, 1]$.

We can get a more refined bound for $(1-x^2)^n$ by comparing with the function $x \mapsto 1-nx^2$ when $x \in [0, 1/\sqrt{n}]$ and the function $x \mapsto 0$ when $x \in (1/\sqrt{n}, 1]$. This is basically the same as just restricing the range of the integral of $1-nx^2$.

It is a common experience in analysis that otherwise solid proofs will break because of greedy or sloppy estimates or bounds.

(l) We have shown that

$$\frac{3\sqrt{n}}{4} \geq c_n$$

which shows

$$0 = \lim_{n\to\infty} c_n(1-\delta^2)^n$$

and completes the proof (except any details you want to flesh out).

36. (p.161) Corollary of *WPUAT*:

On $[-a, a]$, the function $t \mapsto |t|$ can be *uniformly approximated* by polynomials.
That is, there is a sequence $\{P_n\}$ of polynomials, which can be chosen so that $P_n(0) = 0$, such that

$$P_n(t) \to |t| \text{ as } n \to \infty$$

*uniformly* on $t \in [-a, a]$.

Proof: Apply *WPUAT* to the function $t \mapsto |t|$ to get a sequence of polynomials $P_n^*$ which converge uniformly to $|t|$ on $[-a, a]$. Verify that the sequence $\{P_n\}$ has the desired properties, where $P_n(t)$ is defined as $P_n^*(t) - P_n^*(0)$.

37. (p.161) Definition of an *algebra* complex-valued (or real-valued) functions on some common domain $E$:

"Closed" under pointwise addition, pointwise multiplication, and multiplication by scalars in $\mathbb{C}$ (or $\mathbb{R}$).

38. (p.161) Definition of a *uniformly closed* set of functions on a common domain.

A set of functions $\mathscr{A}$ on a common domain $E$ is *uniformly closed* if $f \in \mathscr{A}$ whenever $f_n \in \mathscr{A}$ and $f_n \to f$ uniformly on $E$.

Intuitively, this just means that whenever a function $f$ on $E$ can be uniformly approximated by members of $\mathscr{A}$, then $f$ is also in $\mathscr{A}$.

*Important!*

- Rudin is a little sloppy here, but don't let it confuse you. There is no requirement at this point that the functions are bounded, so it *does not actually make sense* to think of uniform convergence in terms of convergence in the supremum norm ($\|\cdot\|_\infty$) metric.
- Rudin actually defines this concept only for *algebras* of functions on $E$, but it obviously appleis to any collection of functions on a common domain $E$.

39. (p.161) Definition of the *uniform closure* of a set of functions on a common domain.

    If $\mathscr{A}$ is a set of functions on a common domain $E$, then the *uniform closure* of $\mathscr{A}$ is the set of all functions $f$ on $E$ which can be "uniformly approximated" by functions in $\mathscr{A}$.

    In other words, $f$ is in the *uniform closure* of $\mathscr{A}$ if and only if there exists a sequence $\{f_n\}$ with $f_n \in \mathscr{A}$ and $f_n \to f$ uniformly on $E$.

    Obviously $\mathscr{A}$ is contained in its *uniform closure*.
    Make sure this is clear to you!

    *Important!*

    - Again, Rudin is a little sloppy here, and it is confusing. He references Definition 7.14 of the supremum norm ($\|\cdot\|_\infty$). But there is no requirement at this point that the functions are bounded, so it *does not actually make sense* to think of uniform convergence in terms of convergence in the supremum norm ($\|\cdot\|_\infty$) metric.
    - Rudin actually defines this concept only for *algebras* of functions on $E$, but it obviously appleis to any collection of functions on a common domain $E$.

40. (Not in book) *Observations:*
    The following observations are very natural given the definitions that have been set up, however Rudin is relying on you to notice and convince yourself that:

    If $\mathscr{A}$ is any set of (not necessarily bounded) functions on domain $E$, then

    - The *uniform closure* of $\mathscr{A}$ is *uniformly closed*.
      Why?

      This is might be obvious, but if not then here are the details.
      Let $\mathscr{B}$ be the uniform closure of $\mathscr{A}$, and let $\{f_n\}$ be a sequence of functions from $\mathscr{B}$ which converges uniformly on $E$ to a function $f$. We must show that $f \in \mathscr{B}$.

      To do this, we will find a sequence $\{g_n\}$ of functions in $\mathscr{A}$ which converges uniformly to $f$.

      The key idea here is another standard technique in analysis:

*Approximate each term (in this case $f_n$) in an approximating sequence by something else which is more useful (in this case $g_n$).*

Each function $f_n$ is in $\mathscr{B}$, so there exists (verify!) a function $g_n$ in $\mathscr{A}$ such that, for all $x \in E$,

$$|f_n(x) - g_n(x)| < \frac{1}{n}$$

Then, for all $x \in E$,

$$|f(x) - g_n(x)| < |f(x) - f_n(x)| + |f_n(x) - g_n(x)| < |f(x) - f_n(x)| + \frac{1}{n}$$

Show that this means $g_n \to f$ uniformly on $E$.

If $\mathscr{A}$ is a set of *bounded* functions on domain $E$, then the meaning of the words (uniformly) "closed" and (uniformly) "closure" agrees with how we use them in metric spaces:

- The supremum norm $\|\cdot\|_\infty$ is defined on $\mathscr{A}$.
- $\mathscr{A}$ is thus a subspace of the metric space of all bounded functions on $E$.
- $\mathscr{A}$ is *uniformly closed* if and only if $\mathscr{A}$ is closed in the $\|\cdot\|_\infty$ metric space of all bounded functions on $E$.
  Why?

  Recall and combine these relevant facts:

  (a) $\mathscr{A}$ is closed in all bounded functions on $E$ exactly when it equals its closure in all bounded functions on $E$.

  (b) The closure of $\mathscr{A}$ in all bounded functions on $E$ consists exactly of all the bounded functions on $E$ which are limits (in the $\|\cdot\|_\infty$ metric!) of convergent sequences of points from $\mathscr{A}$.

  (c) A sequence $\{f_n\}$ of points from $\mathscr{A}$ converges in the $\|\cdot\|_\infty$ metric to a bounded function $f$ if and only if $f_n \to f$ uniformly on $E$.
- The *uniform closure* of $\mathscr{A}$ is the same as its closure in the $\|\cdot\|_\infty$ metric space of all bounded functions on $E$.
  Why? The pieces are the same as above:

  If $f$ is in the *uniform closure* of $\mathscr{A}$, then there is some sequence $\{f_n\}$ of functions from $\mathscr{A}$ which converges uniformly on $E$ to $f$. So $f$ is bounded and $\{f_n\}$ converges to $f$ in the $\|\cdot\|_\infty$ metric. This means $f$ is in the $\|\cdot\|_\infty$ metric closure of $\mathscr{A}$.

  The other direction is similar.
- The *uniform closure* of $\mathscr{A}$ is is *uniformly closed*.
  Why?

  Combine the previous two items.

41. (p.161) *Theorem:*

> Let $\mathscr{A}$ be an algebra of bounded functions, and let $\mathscr{B}$ be the uniform closure of $\mathscr{A}$. Then $\mathscr{B}$ is a uniformly closed algebra of bounded functions.

Proof:
*Note: The word "closed" is used in two different ways here. Don't get confused.*

We've already shown that $\mathscr{B}$ is *uniformly closed.* We only need to show that

i) the functions in $\mathscr{B}$ are bounded, and that
ii) $\mathscr{B}$ is closed under pointwise addition, pointwise multiplication, and multiplication by scalars, so that it's actually an *algebra* of functions.

Let $f$ and $g$ be in $\mathscr{B}$. There are then sequences $\{f_n\}$ and $\{g_n\}$ of (bounded) functions from $\mathscr{A}$ which converge uniformly to $f$ and $g$.

We observed previously (see *UCSBF*) that $f$ and $g$ must then be *bounded* functions.

Let $\lambda$ be a scalar from $\mathbb{C}$ (or $\mathbb{R}$).
Notice/show that $\lambda f_n \to \lambda f$ uniformly, so that $\lambda f$ is in $\mathscr{B}$.

Similarly, notice/show $f_n + g_n \to f + g$ uniformly, so that $f + g$ is in $\mathscr{B}$.

The only tricky part is showing that $\mathscr{B}$ is closed under pointwise multiplication. We must show that $f_n g_n \to fg$ uniformly as $n \to \infty$. This is where we need the functions in $\mathscr{A}$ to be bounded.
(See exercise #3 p.165 in the next item for a counterexample.)

We showed previously (see *UCSBF*) that $\{f_n\}$ and $\{g_n\}$ must be uniformly bounded. Find a constant $M$ such that

(a) $|f_n(x)| < M$ and $|g_n(x)| < M$ for all $x$ and $n$, and
(b) $|f(x)| < M$ and $|g(x)| < M$ for all $x$.

Now the proof is just like showing that the termwise product of convergent sequences is convergent.

$$\begin{aligned}
\|f_n g_n - fg\|_\infty &= \|f_n g_n - f_n g + f_n g - fg\|_\infty \\
&\le \|f_n(g_n - g)\|_\infty + \|g(f_n - f)\|_\infty \\
&\le \|f_n\|_\infty \|g_n - g\|_\infty + \|g\|_\infty \|f_n - f\|_\infty \\
&\qquad\qquad \text{(Recall that } \|\cdot\|_\infty \text{ is "submultiplicative"...)} \\
&\le M\|g_n - g\|_\infty + M\|f_n - f\|_\infty
\end{aligned}$$

This shows $\|f_n g_n - fg\|_\infty$ goes to zero as $n \to \infty$.

42. (Exercise #3 p.165)
Counterexample:

Without requiring the functions $f_n$ and $g_n$ to be bounded, the termwise product $f_n g_n$ of uniformly convergent sequences $\{f_n\}$ and $\{g_n\}$ *might not* converge uniformly to the product of the limits.

Define sequences of functions $f_n$ and $g_n$ on $(0, 1)$ as

$$f_n(x) = \frac{1}{n}$$
$$g_n(x) = \frac{1}{x}$$

Then

- $f_n(x) \to 0$ uniformly on $(0, 1)$,
- $g_n(x) \to \dfrac{1}{x}$ uniformly on $(0, 1)$,
- but $f_n(x)g_n(x) = \dfrac{1}{nx}$ does *not* converge uniformly to $0 = 0 \cdot \dfrac{1}{x}$ on $(0, 1)$.

43. (p.162) Definition: *separate points*

    A set $\mathscr{A}$ of complex (or real) functions on common domain $E$ *separate points*, if whenever $x_1$ and $x_2$ are distinct points of $E$, then there is a function $f \in \mathscr{A}$ with $f_(x_1) \neq f(x_2)$.

    - Example: Polynomials on $[-1, 1]$
    - Counterexample: *Even* polynomials on $[-1, 1]$

44. (p.162) Definition: *vanishes at not point*

    A set $\mathscr{A}$ of complex (or real) functions on common domain $E$ *vanishes at not point*, if whenever $x \in E$ there is a function $f \in \mathscr{A}$ with $f(x) \neq 0$.

    - Example: Polynomials on $[-1, 1]$
    - Counterexample: Polynomials on $[-1, 1]$ which have zero constant term.

45. (p.162) *Theorem: Stone-Weierstrass Helper Lemma*

    Suppose:

    - $\mathscr{A}$ is an algebra of complex (or real) functions on $E$, which *separate points* and *vanishes at not point*,
    - $x_1$ and $x_2$ are points in $E$, and
    - $c_1$ and $c_2$ are complex (or real) constants.

    Then there is a function $f \in \mathscr{A}$ such that $f(x_1) = c_1$ and $f(x_2) = c_2$.

Proof: This is a fun exercise, so try to figure it out.
Build functions $v$ and $u$ in $\mathscr{A}$ which let you express

$$f = \frac{c_1 v}{v(x_1)} + \frac{c_2 u}{u(x_2)}$$

46. (p.162) *Theorem: The Stone-Weierstrass Theorem*

If

- $\mathscr{A}$ is an algebra of *real-valued and continuous* functions on a compact metric space domain $K$, and
- $\mathscr{A}$ separates points of $K$ and vanishes at no point of $K$,

then

the uniform closure $\mathscr{B}$ of $\mathscr{A}$ is the (uniformly closed) algebra whose members are exactly the real-valued and continuous functions on $K$

Note #1: This says that if $\mathscr{A}$ then every real-valued continuous function can be uniformly approximated arbitrarily well by members of $\mathscr{A}$.

Note #2: In terms of metric spaces, using the $\|\cdot\|_\infty$ supremum metric, this says that $\mathscr{A}$ is *dense* in the space of all continuous real-valued functions on $K$.

Note #3: This theorem *does not hold as is* for complex algebras! I will point out the steps in the proof which fail in the complex case.
But, if you add an additional condition to the complex algebra, then the conclusion of the theorem *does hold*.
See the corollary after the theorem.

Proof:

Convince yourself (if it's not obvious to you), that to prove the theorem it's good enough to prove the following:

Given a continuous function $f$ on $K$, and given $\epsilon > 0$, there is a function $g \in \mathscr{B}$ such that $\|f - g\|_\infty \le \epsilon$.

That's what we'll prove.

(a) Use the corollary of *WPUAT* which says that $|y|$ can be uniformly approximated by polynomials in $y$, and the fact that $\mathscr{B}$ is uniformly closed, to show that $|g| \in \mathscr{B}$ whenever $g \in \mathscr{B}$.

This step relies on $y = g(x)$ taking values in some compact interval $[-a, a] \subset \mathbb{R}$.

**(!) This step doesn't make sense when $f$ is allowed to take values in $\mathbb{C}$.**

(b) Show that if $h$ and $g$ are in $\mathscr{B}$ then so are $\max(h, g)$ and $\min(h, g)$.
Express these using the absolute value function :)

**(!) This step doesn't make sense when $f$ is allowed to take values in $\mathbb{C}$.**

(c) Let $f$ be a real-valued function which is continuous on $K$, and fix this function for the rest of the proof.

Let $x \neq y$ be distinct points of $K$. Use the previous lemma to show that there is a (continuous) function $h_{x,y} \in \mathscr{A}$ such that

$$h_{x,y}(x) = f(x) \text{ and } h_{x,y}(y) = f(y)$$

(d) Let $f$ be as in step (c).
Fix $x \in K$ and $\epsilon > 0$. For each $y \in K$, the function $h_{x,y}$ is continuous and agrees with $f$ at $y$. Show that each $y \in K$ has an open neighborhood $U_y$ such that

$$\left| h_{x,y}(t) - f(t) \right| < \epsilon$$

when $t \in U_y$. This gives an open cover $\{U_y : y \in K\}$ of $K$, and by compactness there's a finite subcover $\{U_{y_1}, U_{y_2}, U_{y_3}, \ldots U_{y_N}\}$.

Define $g_x = \max\{h_{x,y_1}, h_{x,y_2}, h_{x,y_3}, \ldots, h_{x,y_N}\}$.

If $t \in K$, then $t$ is in some $U_{y_n}$, and so

$$g_x(t) \geq h_{x,y_n}(t) > f(t) - \epsilon$$

(e) Fix $\epsilon > 0$.
The previous step showed that for each $x \in K$ we can find a function $g_x \in \mathscr{B}$ such that

$$\begin{aligned} g_x(x) &= f(x) \\ g_x(t) &> f(t) - \epsilon \quad \text{for every } t \in K \end{aligned}$$

(f) Let $f$ be as in step (c).
This step is just like step (d), but with the goal of "trapping" $f$ from the other side.

Show that each $x \in K$ has an open neighborhood $V_x$ such that

$$\left| g_x(t) - f(t) \right| < \epsilon$$

when $t \in V_x$. This gives an open cover $\{V_x : x \in K\}$ of $K$, and by compactness there's a finite subcover $\{V_{x_1}, V_{x_2}, V_{x_3}, \ldots V_{x_M}\}$.

Define $g = \min\{g_{x_1}, g_{x_2}, g_{x_3}, \ldots, g_{x_M}\}$.

If $t \in K$, then $t$ is in some $V_{x_n}$, and so

$$g(t) \leq g_{x_n}(t) < f(t) + \epsilon$$

(g) Notice from step (e) that

$$f(t) - \epsilon < g(t) = \min\{g_{x_1}(t), g_{x_2}(t), g_{x_3}(t), \ldots, g_{x_M}(t)\}$$

for every $t \in K$.

(h) We have found a function $g$ has these properties:
   - $g \in \mathscr{B}$, by step (b).
   - $\|g - f\|_\infty \leq \epsilon$, by step (f) and step (g).

This completes the proof.

47. (p.165) Definition of *self-adjoint* algebra of complex-valued functions.
Basically just means "closed under" pointwise complex conjugation.

48. (p.165) *Corollary: Stone-Weierstrass for Self-Adjoint Complex Algebras*

   If $\mathscr{A}$ is a *self-adjoint* algebra of continuous complex-valued functions on a compact metric space domain $K$, which separates points of $K$ and vanishes at no point of $K$, then its uniform closure is the space of *all* complex-valued and continuous functnios on $K$.

The proof is beautiful and is actually really easy once you figure out what the main idea should be.

(a) Show that the real and imaginary parts of the functions in $\mathscr{A}$ are also functions in $\mathscr{A}$.

(b) Show that these real and imaginary parts actually form an algebra which satisfies the conditions of *The Stone-Weierstrass Theorem.*

(c) Proceed.

# 8 Chapter 8: Some Special Functions

1. Recall:

   - p.69 Definition of *power series* for a sequence $\{c_n\}$ in $\mathbb{C}$ and some *fixed* $z \in \mathbb{C}$:

   $$\sum_n c_n z^n \text{ or } \sum_n c_n |z - a|^n$$

   - p.69 Definition of the *radius of convergence* $R$ for the *coefficients* $\{c_n\}$ of a power series:

   $$R = \frac{1}{\alpha} \text{ with } \alpha = \limsup_n |c_n|^{1/n}$$

   - If $|z| < R$, then the power series $\displaystyle\sum_n c_n z^n$ converges absolutely to a complex number.

2. This means that if a sequence $\{c_n\}$ of complex coefficients has radius of convergence $R$, then can define *functions*

$$z \mapsto \sum_n c_n z^n \qquad \text{with domain } \{z \in \mathbb{C} : |z| < R\}$$

$$w \mapsto \sum_n c_n (w-a)^n \qquad \text{with domain } \{w \in \mathbb{C} : |w-a| < R\}$$

3. (p.172) At this point Rudin introduces and uses some standard jargon.
   If you're like me, then you might get confused by what exactly this jargon means.
   Here are the relevant definitions:

   - A function $f(z)$ is *"represented by"* a power series $\sum_n c_n(z-a)^n$ in $(a-r, a+r)$.

     This means:

     The radius of convergence $R$ of $\{c_n\}$ is at least as big as $r$, and that $\sum_n c_n(z-a)^n$ converges to $f(z)$ when $z \in (a-r, a+r)$.

     In other words,

     $$f(z) = \lim_{k\to\infty} \sum_{n=0}^{k} c_n (z-a)^n$$

     when $z \in (a-r, a+r)$.

   - A function $f(z)$ is *"of the form"* $\sum_n c_n(z-a)^n$

     Means the same thing as above.

   - A function $f(z)$ can be *"expanded"* in a power series *"near the point"* $z = a$
     This means that there *exists* some $r > 0$ and a sequence $\{c_n\}$ of coefficients such that $f(z)$ is *represented by* the power series $\sum_n c_n(z-a)^n$ in $(a-r, a+r)$.

4. (p.173) *Theorem: Main Power Series Theorem (MPST)*

   Let the power series $f(z) = \sum_{n\geq 0} c_n z^n$ have radius of convergence $R > 0$.

   Then

   (a) The sequence $\{f_k\}$ of partial sums

   $$f_k(z) = \sum_{n=0}^{k} c_n z^n \qquad k = 1, 2, 3, \ldots$$

   for $f$ converges absolutely on the open disk $|z| < R$.

   (b) The sequence $\{f_k\}$ of partial sums for $f$ converges uniformly on every compact disk $|z| \leq R - \delta$.

   (c) $f$ is continuous on the open disk $|z| < R$.

(d) The power series

$$\sum_{n\geq 0} c_{n+1}(n+1)z^n$$

has the same radius of convergence $R$ as the original power series.

(e) *Restricting the domain of $f$ real values $x \in (-R, R)$ so that we can talk about differntiation:*
$f$ is "termwise differentiable" on the real interval $(-R, R)$, meaning that $f$ is differentiable on $(-R, R)$ and that

$$f'(x) = \lim_{n\to\infty} f_k'(x) = \sum_{n\geq 0} c_n n x^{(n-1)} = \sum_{n\geq 0} c_{n+1}(n+1)x^n$$

when $x \in (-R, R)$.

Proof:

Proof of (a):

We already saw this.

Proof of (b):

If $|z| \leq R - \delta$, then $|c_n z^n| \leq |c_n(R-\delta)^n|$.
Apply the *Weierstrass M-Test*.

Proof of (c):

You can prove this directly using uniform convergence from part (b), but it also follows from part (e) about differentiability.

Proof of (d):

You need to show that

$$\limsup_n |(n+1)c_{n+1}|^{1/n} = \limsup_n |nc_n|^{1/n} = \limsup_n |c_n|^{1/n}$$

Recall from chapter 3 p.57 that lim sup respects termwise multiplcation by convergent sequence of positive real numbers.
Recall also from chapter 3 that $\lim_{n\to\infty} n^{1/n} = 1$.

Comments before proving part (e):

Rudin provides a proof of part (e) which is little more than a quick application of the *CTDS Theorem* for termwise differentiability of sequences. While his proof leverages and avoids repeating work already done, it's not particularly enlightening in terms of understanding power series *per se*. There are, however, two other common proofs which are more "hands on" and probably helpful. So I'll give three proofs here.

Note: By parts (a), (b), and (c) of the theorem applied to the series from part (d), we already know that the series

$$\sum_{n\geq 0} c_{n+1}(n+1)z^n$$

converges absolutely on the open disk $|z| < R$, converges uniformly on every compact disk $|z| \leq R - \delta$, and that the resulting function of $z$ is continuous on the open disk $|z| < R$.

Proof of (e) #1 (Rudin's proof)

This proof is basically just an application of the *CTDS Theorem*.

Pick an arbitrary $x \in (-R, R)$, and choose any $\delta$ with

$$x \in (a, b) \subset [a, b] \subset (-R, R)$$

Let

$$f_k(t) = \sum_{n=0}^{k} c_n t^n \qquad t \in [a, b],\ k = 1, 2, 3, \ldots$$

The hypotheses of *CTDS Theorem* are satisfied on $[a, b]$ for the sequence $\{f_k\}$. So $f$ is differentiable on $[a, b]$, and in particular at the point $x$, and

$$f'(x) = \lim_{k\to\infty} f_k'(x) = \lim_{k\to\infty} \sum_{n=0}^{k} c_{n+1}(n+1)x^n = \sum_{n\geq 0} c_{n+1}(n+1)x^n$$

Poof we're done. It's like magic - all the real work was packaged up in the *CTDS Theorem*.

Proof of (e) #2 (From Spivak's *Calculus*)

This proof basically just involves combining *FTC #1* with the *TIUCS* theorem.

Let

$$g(x) = \sum_{n\geq 0} c_{n+1}(n+1)x^n$$

for $x \in (-R, R)$.

Define the function

$$G(x) = \int_0^x g$$

for $x \in (-R, R)$. By *FTC #1*, the function $G$ is differentiable at $x$ and $G'(x) = g(x)$.

Now let $g_k$ be a partial sum for $g$. That is, let

$$g_k(t) = \sum_{n=0}^{k} c_{n+1}(n+1)t^n$$

for $t \in (-R, R)$.

We know from part (d) that $g_k \to g$ uniformly on the closed interval $[0, x]$, if $x \geq 0$, and on the closed interval $[x, 0]$, if $x < 0$.

Applying the *TIUCS* theorem from p.151 to $g$ and $g_k$, we get

$$\begin{aligned}
G(x) &= \int_0^x g(t)\ dt \\
&= \lim_{k\to\infty} \int_0^x g_k(t)\ dt \\
&= \lim_{k\to\infty} \int_0^x \left( \sum_{n=0}^{k} c_{n+1}(n+1)t^n \right)\ dt \\
&= \lim_{k\to\infty} \left( \sum_{n=0}^{k} c_{n+1}x^{n+1} - \sum_{n=0}^{k} c_{n+1}0^{n+1} \right) \\
&= \lim_{k\to\infty} \sum_{n=1}^{k} c_n x^n \\
&= \lim_{k\to\infty} \left( -c_0 + \sum_{n=0}^{k} c_n x^n \right) \\
&= -c_0 + \lim_{k\to\infty} \sum_{n=0}^{k} c_n x^n \\
&= -c_0 + f(x)
\end{aligned}$$

This shows that $f(x) = G(x) + c_0$, so $f$ is differentiable on $(-R, R)$ since $G$ is, and

$$f'(x) = G'(x) = g(x) = \sum_{n\geq 0} c_{n+1}(n+1)x^n$$

as desired.

Proof of (e) #3 (From Ahlfors' *Complex Analysis*)

Unlike the previous two proofs, this third proof also works for *complex derivatives of complex functions of one complex variable.* (That material is not covered in the particular Rudin book we're looking at here in these notes, but to find out more, check out for example

Lars Ahlfors' book *Complex Analysis* book or Rudin's other book *Real & Complex Analysis*.)

This proof is definitely the most "hands-on" one of the three, and is absolutely worth knowing. It is an "$\epsilon/3$ argument", and also involves another extremely common technique where you "control the tail" of a series. You might want to take a quick look back at *MUCLIT*, so get the big picture of what's going on.

First, for $|z| < R$, let

$$s_n(z) = \sum_{k=0}^{n-1} c_n z^k$$
$$R_n(z) = \sum_{k=n}^{\infty} c_n z^k$$
$$g(z) = \sum_{n=0}^{\infty} c_{n+1}(n+1)z^n$$

so that you get (verify that this makes sense!)

$$f(z) = s_n(z) + R_n(z)$$

Pick an arbitrary point $z$ with $|z| < R$.
We will show differentiability of $f$ at the point $z$.

We need to prove that

$$\lim_{w \to z} \frac{f(w) - f(z)}{w - z} - g(z) = 0$$

The motivation behind the proof consists of two things:

- First, notice that $s_n(z)$ is just a polynomial, so we know it's differentiable, and that $s_n'(z)$ is a "head" of the series $g(z)$.
- Second, $f(z)$ and $g(z)$ converge *uniformly* on compact sets inside $\{z : |z| < R\}$.
  This means that we will be able to make the "tails" of our series *"uniformly small"*, which gives hope that the series might behave a lot like polynomials in terms of differentiation.

For our chosen $z$, there exists some $\rho$ ("rho") with $|z| < \rho < R$.
Fix this value $\rho$.

Let $w$ be a point such that $|w| < \rho < R$.

The first key trick, which is a standard analysis manipulation, is to "telescope" the expression

$$\frac{f(w)-f(z)}{w-z} - g(z)$$

in such a way that the resulting pieces are manageable. This results in an $\epsilon/3$ proof.

Write

$$\begin{aligned}\frac{f(w)-f(z)}{w-z} - g(z) = &\left(\frac{s_n(w)-s_n(z)}{w-z} - s_n'(z)\right) && \text{(A)}\\ &+\left(s_n'(z)-g(z)\right) && \text{(B)}\\ &+\frac{R_n(w)-R_n(z)}{w-z} && \text{(C)}\end{aligned}$$

This looks pretty good so far, since term (A) goes to zero as $w \to z$ as $n$ is held fixed, and term (B) goes to zero as $n \to \infty$. If we only had to worry about these then we'd be nearly done, since we could first choose $n$ large enough to make term (B) small, then make $|w-z|$ small enough to make term (A) small.

But we need to get control of term (C). It's not immediately clear what happens to term (C) for a fixed $n$ as $w \to z$. This leads us to the second key trick, which involves another standard analysis manipulation, which is to "telescope"

$$w^k - z^k = (w-z)(w^{k-1}z^0 + w^{k-2}z^1 + \cdots + w^0 z^{k-1}) = (w-z)\sum_{j=1}^{k} w^{k-j}z^{j-1}$$

Write

$$\begin{aligned}\frac{R_n(w)-R_n(z)}{w-z} &= \frac{1}{w-z}\left[\sum_{k=n}^{\infty} c_k w^k - \sum_{k=n}^{\infty} c_k z^k\right]\\ &= \frac{1}{w-z}\sum_{k=n}^{\infty} c_k(w^k - z^k) \qquad \text{(Verify this!)}\\ &= \frac{1}{w-z}\sum_{k=n}^{\infty}\left\{c_k(w-z)\sum_{j=1}^{k} w^{k-j}z^{j-1}\right\}\\ &= \sum_{k=n}^{\infty}\left\{c_k\sum_{j=1}^{k} w^{k-j}z^{j-1}\right\} \qquad \text{(Verify this!)}\end{aligned}$$

We are in great shape now! By the *Triangle Inequality*, we have

$$\left|c_k\sum_{j=1}^{k} w^{k-j}z^{j-1}\right| \le |c_k|\sum_{j=1}^{k}|w|^{k-j}|z|^{j-1} \le |c_k|\sum_{j=1}^{k}\rho^{k-1} = |c_k|k\rho^{k-1}$$

But we already know that the series

$$\sum_{k=n}^{\infty} |c_k| k \rho^{k-1}$$

converges, since $0 < \rho < R$, and so the power series $g(\rho)$ converges *absolutely*. So the series

$$\frac{R_n(w) - R_n(z)}{w - z} = \sum_{k=n}^{\infty} \left\{ c_k \sum_{j=1}^{k} w^{k-j} z^{j-1} \right\}$$

coverges absolutely (by the comparison test). But in fact since absolute value is continuous, we can get even more information:

$$\begin{aligned}
\left| \frac{R_n(w) - R_n(z)}{w - z} \right| &= \left| \sum_{k=n}^{\infty} \left\{ c_k \sum_{j=1}^{k} w^{k-j} z^{j-1} \right\} \right| \\
&= \left| \lim_{N \to \infty} \sum_{k=n}^{N} \left\{ c_k \sum_{j=1}^{k} w^{k-j} z^{j-1} \right\} \right| \\
&= \lim_{N \to \infty} \left| \sum_{k=n}^{N} \left\{ c_k \sum_{j=1}^{k} w^{k-j} z^{j-1} \right\} \right| \qquad \text{(continuity used here)} \\
&\leq \lim_{N \to \infty} \sum_{k=n}^{N} \left| c_k \sum_{j=1}^{k} w^{k-j} z^{j-1} \right| \\
&\leq \lim_{N \to \infty} \sum_{k=n}^{N} |c_k| k \rho^{k-1} \\
&\leq \sum_{k=n}^{\infty} |c_k| k \rho^{k-1}
\end{aligned}$$

This is exactly what we needed to complete the proof, since

$$\frac{R_n(w) - R_n(z)}{w - z} \leq \sum_{k=n}^{\infty} |c_k| k \rho^{k-1}$$

*does not depend* on $w$.

*Remark:*
We have actually shown that

$$\frac{R_n(w) - R_n(z)}{w - z} \to 0$$

*uniformly* as $n \to \infty$, as a function of $w$ on the set $\{w : |w| < \rho < R \text{ and } w \neq z\}$. This is exactly what we needed to frame things in terms of *MUCLIT*.

Putting things back together, we obtain

$$\left|\frac{f(w)-f(z)}{w-z} - g(z)\right| \le \left|\frac{s_n(w)-s_n(z)}{w-z} - s_n'(z)\right| \qquad \text{(A*)}$$

$$+ \left|s_n'(z) - g(z)\right| \qquad \text{(B*)}$$

$$+ \sum_{k=n}^{\infty} |c_k| k \rho^{k-1} \qquad \text{(C*)}$$

We may then show that

$$\lim_{w\to z} \frac{f(w)-f(z)}{w-z} - g(z) = 0$$

by first fixing $n$ sufficient large, making terms (B*) and (C*) each smaller than $\epsilon$, then finally making $|w-z|$ small enough so that term (A*) is smaller than $\epsilon$.

5. (p.173) *Corollary: MPST Corollary*

If $f(x)$ is "represented by" the power series $\sum_{n\ge 0} c_n x^n$ on $(-r, r)$, then

(a) $f$ is infinitely differentiable on $(-r, r)$.
(b) Each successive derivative $f^{(k)}(x)$ is represented on $(-r, r)$ by the termwise derivative of the previous series which represents $f^{(k-1)}(x)$.
(c) $f^{(k)}(0) = k! c_k \quad$ for $k = 0, 1, 2, 3, \ldots$

In addition (from (c)), if $f$ is represented by a power series on $(-r, r)$, then it is unique in the following sense:

If $\sum_{n\ge 0} b_n x^n$ and $\sum_{n\ge 0} c_n x^n$ both represent $f$ on $(-r, r)$,
then $b_n = c_n$ for all $n = 0, 1, 2, 3, \ldots$

Proof: The *MPST* shows that $f$ is termwise differentiable, and also that the termwise derivative series has the same radius of convergence as the original. This means that the *MPST* can be applied repeatedly.

6. *Counterexample* (Chapter 8, p.196 exercise #1)

The converse of part (a) of *MPST Corollary* does not hold.
Being infinitely differentiable *does not imply* having a power series representation.

$$f(x) = \begin{cases} e^{-1/x^2} & \text{when } x \ne 0 \\ 0 & \text{when } x = 0 \end{cases}$$

7. (Not in book) *Converse to Abel's Theorem*

   *Note: I'm putting this before* Abel's Theorem *in hopes that it might help some people get better intuition about the proof of* Abel's Theorem.

   > Suppose that the sequence of partial sums for
   >
   > $$\sum_{n\geq 0} c_n x^n$$
   >
   > converges uniformly on some interval $(t, 1)$.
   > Then the series
   >
   > $$\sum_{n\geq 0} c_n$$
   >
   > converges.

   Proof: Straightforward application of *MUCLIT* at the limit point $x = 1$.
   As an aside, also notice that the hypotheses imply the radius of convergence is at least 1.

8. (Not in book) *Lemma for Abel's Theorem*

   > Let $-1 < \lambda < 1$.
   > If $\{a_n\}$ is a sequence of complex numbers such that $a_n \to 0$ as $n \to \infty$,
   > then the sequence of functions
   >
   > $$g_k(x) = (1-x)\sum_{n=0}^{k} a_n x^n$$
   >
   > converges uniformly on $[\lambda, 1)$ as $k \to \infty$ to the function
   >
   > $$g(x) = (1-x)\sum_{n\geq 0} a_n x^n$$

   Proof: Straightforward. Do it as an exercise - if you've successfully made it this far into the book then it should be pretty easy for you.

9. (p.174) *Theorem: Abel's Theorem For Behavior On Circle Of Convergence*
   This simple-looking theorem is actually often quite useful when dealing with "edge cases" in limiting processes.

   > Suppose that $\sum_{n\geq 0} c_n$ converges.
   > Put $f(x) = \sum_{n\geq 0} c_n x^n$ for $-1 < x \leq 1$.
   > Then $f(x)$ is continuous at $x = 1$.

   *Notes:*

- *The statement of the theorem makes sense. You should verify that the hypothesis that $\sum_{n\geq 0} c_n$ converges guarantees the radius of convergence of $\sum_{n\geq 0} c_n x^n$ must be at least equal to one.*
- *We've already seen that the* MPST *covers this result in the situation where the radius of convergence is strictly greater than one. But the proof we're about to give covers both that case and also the case where the radius of convergence is precisely one.*
- *Once again, Rudin provides a standard but slick proof without providing motivation. It's up to us to figure out how you might come up with such a proof. So I'm spending a little extra time on this theorem since I had trouble getting any inutition for it.*
- *The obvious way to approach the proof is using* MUCLIT *on the series in question to show that*
$$\lim_{x\to 1}\left(\lim_{k\to\infty}\sum_{n=0}^{k} c_n x^n\right) = \lim_{k\to\infty}\left(\lim_{x\to 1}\sum_{n=0}^{k} c_n x^n\right)$$
*But to apply* MUCLIT *we'd first have to show that* $\sum_{n=0}^{k} c_n x^n$ *converges uniformly on some interval* $(1-\delta, 1)$. *While uniform convergence does in fact hold, it isn't necessarily super obvious how to directly prove it.*

  *In fact, the* Converse to Abel's Theorem *tolds us that the uniform convergence must be equivalent to the convergence of the series* $\sum_{n\geq 0} c_n$.

  *What this means for us is that just using the weaker condition that* $c_n \to 0$ *cannot provide a proof of uniform convergence. So we might as well plan on trying to find a proof that involves summing the coefficients somehow.*

Proof: Let
$$f_k(x) = \sum_{n=0}^{k} c_n x^n$$
To prove the theorem, it is sufficient to show that $f_k \to f$ uniformly on $(0,1)$, because then an easy application of *MUCLIT* shows

$$\lim_{x\to 1} f(x) = \lim_{x\to 1}\left(\lim_{k\to\infty}\sum_{n=0}^{k} c_n x^n\right) = \lim_{k\to\infty}\left(\lim_{x\to 1}\sum_{n=0}^{k} c_n x^n\right) = \sum_{n\geq 0} c_n$$

But we'll have to find a way to cleverly rewrite $f_k(x)$ to use the the hypothesis that $\sum_{n\geq 0} c_n$ converges. We'd also like to leverage *Lemma for Abel's Theorem*, which means we want to find a factor of $(1-x)$ somehow.
Let
$$s_n = c_0 + c_1 + \cdots + c_n$$

and

$$s = \sum_{k=0}^{\infty} c_k$$

- *Approach #1: Use Summation By Parts*

$$\begin{aligned} f_k(x) &= \sum_{n=0}^{k} c_n x^n \\ &= s_k x^{k+1} + \sum_{n=0}^{k} s_n \left(x^n - x^{n+1}\right) \\ &= s_k x^{k+1} + (1-x) \sum_{n=0}^{k} s_n x^n \end{aligned}$$

- *Approach #2: Multiply by* $1 = (1-x)(1+x+x^2+\cdots)$

$$\begin{aligned} f_k(x) &= \sum_{n=0}^{k} c_n x^n \\ &= (1-x)(1+x+x^2+\cdots) \sum_{n=0}^{k} c_n x^n \\ &= (\text{...reader verify this manipulation...}) \\ &= (1-x) \left( \sum_{n=0}^{k} s_n x^n + \sum_{n \geq k+1} s_k x^n \right) \\ &= (1-x) \sum_{n=0}^{k} s_n x^n + s_k (1-x) \sum_{n \geq k+1} x^n \\ &= (1-x) \sum_{n=0}^{k} s_n x^n + s_k x^{k+1} (1-x)(1+x+x^2+\cdots) \\ &= s_k x^{k+1} + (1-x) \sum_{n=0}^{k} s_n x^n \end{aligned}$$

That was the main trick, and either way we got

$$f_k(x) = s_k x^{k+1} + (1-x) \sum_{n=0}^{k} s_n x^n$$

Let us coerce things into the form required by *Lemma for Abel's Theorem*. The issue is that in general it's not the case that $s_n \to 0$ as $n \to \infty$.

But it *is* the case that $(s_n - s) \to 0$, so:

$$\begin{aligned}
f_k(x) &= s_k x^{k+1} + (1-x)\sum_{n=0}^{k} s_n x^n \\
&= s_k x^{k+1} + (1-x)\sum_{n=0}^{k} s x^n + (1-x)\sum_{n=0}^{k}(s_n - s)x^n \\
&= s_k x^{k+1} + (1-x)s\sum_{n=0}^{k} x^n + (1-x)\sum_{n=0}^{k}(s_n - s)x^n \\
&= s_k x^{k+1} + s(1-x)\sum_{n=0}^{k} x^n + (1-x)\sum_{n=0}^{k}(s_n - s)x^n \\
&= s_k x^{k+1} + s(1-x^{k+1}) + (1-x)\sum_{n=0}^{k}(s_n - s)x^n \\
&= \left\{ s + (s_k - s)x^{k+1} \right\} + \left\{ (1-x)\sum_{n=0}^{k}(s_n - s)x^n \right\} \\
&= \left\{ \text{(A)} \right\} + \left\{ \text{(B)} \right\}
\end{aligned}$$

We are in very good shape now, since (A) is obviously uniformly convergent, and (B) is uniformly convergent by *Lemma for Abel's Theorem.*

This completes the proof.

10. (p.174) *Corollary: Corollary #1 of Abel's Theorem*

    Let $\sum_{n\geq 0} c_n x^n$ have radius of convergence $R$,
    and let $x_0$ be an endpoint $\pm R$ of the interval $-R < x < R$.

    If $\sum_{n\geq 0} c_n {x_0}^n$ converges, then

    $$\lim_{x\to x_0} \sum_{n\geq 0} c_n x^n = \sum_{n\geq 0} c_n {x_0}^n$$

    In other words, if a power series converges at and endpoint of the interval of convergence, then it is continuous there.

Proof:

Put $b_n = c_n(x_0)^n$. Notice that the radius of convergence $R'$ for the series $\sum_{n\geq 0} b_n x^n$

is $R' = 1$, and that $\sum_{n\geq 0} b_n = \sum_{n\geq 0} c_n x_0^n$.

We can then define $f(x) = \sum_{n\geq 0} b_n x^n$ for $-1 < x \leq 1$, which is continuous at $x = 1$ by *Abel's Theorem.*

Then the composite

$$g(x) = f(x/x_0) = \sum_{n\geq 0} b_n (x/x_0)^n = \sum_{n\geq 0} c_n x^n$$

is continuous at $x = x_0$. This leads immediately to the desired conclusion.

11. (p.75, p.175) *Corollary: Corollary #2 of Abel's Theorem*

    Let $A = \sum_{n\geq 0} a_n$ and $B = \sum_{n\geq 0} b_n$ be convergent series.
    If the Cauchy product of $\sum_{n\geq 0} a_n$ and $\sum_{n\geq 0} b_n$ happens to converge, then it converges to $AB$.

    *Note that neither of $\sum_{n\geq 0} a_n$ or $\sum_{n\geq 0} b_n$ is required to be absolutely convergent.*

    In other words, if the Cauchy product of two series happens to converge, then it converges to the "right value".

    The proof is delightfully easy and beautiful.

    Proof:

    Let

$$f(x) = \sum_{n\geq 0} a_n x^n \qquad g(x) = \sum_{n\geq 0} b_n x^n$$
$$c_n = \sum_{k=0}^{n} a_k b_{n-k} \qquad h(x) = \sum_{n\geq 0} c_n x^n$$

    We already know that the series for $f(x)$, $g(x)$, and $h(x)$ each converge *absolutely* on $x \in (-1, 1)$. This means we know that $h(x) = f(x)g(x)$ when $x \in (-1, 1)$, since the series for $h(x)$ is the Cauchy product of those for $f(x)$ and $g(x)$. Now pass to the limit using *Abel's Theorem.*

12. (Not in book) *Theorem: Single To Double Series Conversion Theorem (STDSCT)*
    *Note: This theorem can be found, for example, in Knopp's Theory And Application Of Infinite Series. It's a very useful result, but will use it only in Knopp's proof of the Cauchy's Double Series Theorem (CDST) – which is the next theorem after this one.*

    Suppose the series $\sum_{n\geq 0} c_n$ is absolutely convergent.
    Let $\tau : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ be any bijection.

Then

$$\sum_{n\geq 0} c_n = \sum_{i\geq 0}\left(\sum_{j\geq 0} c_{\tau(i,j)}\right)$$

(The convergence of $\sum_{j\geq 0} c_{\tau(i,j)}$ for each $i$ and of $\sum_{i\geq 0}\left(\sum_{j\geq 0} c_{\tau(i,j)}\right)$ is part of the conclusion.)

Proof:

Fix any bijection $\tau : \mathbb{N}\times\mathbb{N} \to \mathbb{N}$.

Notice that for each fixed $i$, the series $\sum_{j\geq 1} c_{\tau(i,j)}$ is absolutely convergent, since each partial sum of $\sum_{j\geq 1}\left|c_{\tau(i,j)}\right|$ is bounded above by $\sum_{n\geq 0}|c_n|$.

Let us show that

$$\sum_{i=1}^{k}\left(\sum_{j\geq 1} c_{\tau(i,j)}\right) \text{ is “close to” } \sum_{n=1}^{k} c_n$$

when $k$ is “large enough”.

Let $\epsilon > 0$ be given.
First, find $K_1$ large enough so that

$$\sum_{n>K_1} |c_n| < \epsilon$$

Second, you can – since $\tau$ is a bijection – find $K_2$ and $J_0$ large enough so that the image of $\mathbb{N}\times\{1,2,3,\ldots,K_2\}$ under $\tau$ satisfies

$$\{1,2,3,\ldots,K_1\} \subset \tau\left(\{1,2,3,\ldots,K_2\}\times\{1,2,3,\ldots,J_0\}\right)$$

(So, we're just making the finite rectangle

$$\{1,2,3,\ldots,K_2\}\times\{1,2,3,\ldots,J_0\}$$

big enough so its image contains every value $1,2,3,\ldots K_1$.)

Let $k > \max(K_1, K_2)$.

The crucial observation is that we can then get a lot of “cancellation” in

$$\left|\sum_{i=1}^{k}\left(\sum_{j\geq 1} c_{\tau(i,j)}\right) - \sum_{n=1}^{k} c_n\right|$$

so that what remains is "small". We have:

$$\begin{aligned}
&\sum_{i=1}^{k}\left(\sum_{j\geq 1} c_{\tau(i,j)}\right) - \sum_{n=1}^{k} c_n \\
&\quad = \sum_{i=1}^{k}\left(\sum_{j=1}^{J_0} c_{\tau(i,j)} + \sum_{j>J_0} c_{\tau(i,j)}\right) - \sum_{n=1}^{k} c_n \\
&\quad = \left(\sum_{i=1}^{k}\sum_{j=1}^{J_0} c_{\tau(i,j)} - \sum_{n=1}^{k} c_n\right) + \left(\sum_{i=1}^{k}\sum_{j>J_0} c_{\tau(i,j)}\right) \\
&\qquad\qquad \text{(next line by "}\mathit{FSOCS}\text{" from chapter 3)} \\
&\quad = \left(\sum_{i=1}^{k}\sum_{j=1}^{J_0} c_{\tau(i,j)} - \sum_{n=1}^{k} c_n\right) + \left(\sum_{j>J_0}\sum_{i=1}^{k} c_{\tau(i,j)}\right) \\
&\quad = \left(\sum_{i=1}^{K_2}\sum_{j=1}^{J_0} c_{\tau(i,j)} - \sum_{n=1}^{K_1} c_n\right) + \left(\sum_{i=K_2+1}^{k}\sum_{j=1}^{J_0} c_{\tau(i,j)} - \sum_{n=K_1+1}^{k} c_n\right) + \left(\sum_{j>J_0}\sum_{i=1}^{k} c_{\tau(i,j)}\right) \\
&\quad = \qquad \left(\mathbf{A}\right) \qquad + \qquad \left(\mathbf{B}\right) \qquad + \qquad \left(\mathbf{C}\right)
\end{aligned}$$

Observe that each partial sum for series "**C**" is a sum of a finite subsequence of the tail sequence $\{c_n\}$, $n = K_1+1, K_1+2, K_1+3, \ldots$. By choice of $K_1$, then, we see that $|\mathbf{C}| < \epsilon$. Similarly, by choice of $K_2$, injectivity of $\tau$, and making cancellations, we see that $|\mathbf{A}| < \epsilon$ and $|\mathbf{B}| < \epsilon$.
This shows

$$\left|\sum_{i=1}^{k}\left(\sum_{j\geq 1} c_{\tau(i,j)}\right) - \sum_{n=1}^{k} c_n\right| < 3\epsilon$$

This completes the proof.

*Note about the proof:*
*Actually, by looking more closely and not using the Triangle Inequality so soon, you can get the improved inequality*

$$\left|\sum_{i=1}^{k}\left(\sum_{j\geq 1} c_{\tau(i,j)}\right) - \sum_{n=1}^{k} c_n\right| < \epsilon$$

*Here is how. After making cancellations terms "$\boldsymbol{A}$" and "$\boldsymbol{B}$" are each a sum of a finite subsequence of the tail sequence $\{c_n\}$, $n = K_1+1, K_1+2, K_1+3, \ldots$. We observed already that each partial sum for series "$\boldsymbol{C}$" is also such a sum. Now the trick is to see that the three sets of indices of $\{c_n\}$ which make contributions to "$\boldsymbol{A}$", "$\boldsymbol{B}$" or "$\boldsymbol{C}$" are pairwise disjoint. So $\boldsymbol{A}+\boldsymbol{B}+\boldsymbol{C}$ is actually the sum of a subsequence of $\{c_n\}$, $n = K_1+1, K_1+2, K_1+3, \ldots$ So we get*

$$\left|\boldsymbol{A}+\boldsymbol{B}+\boldsymbol{C}\right| < \epsilon$$

13. (p.175) *Theorem: Cauchy's Double Series Theorem (CDST)*
*Note: Don't underestimate the importance of this theorem.*

Suppose "$a_{ij}$" $= a_{(i,j)} \in \mathbb{C}$ for $i, j = 1, 2, 3, \ldots$ are such that

(i) $\sum_{j\geq 1} a_{ij}$ converges absolutely for every $i$, and

(ii) $\sum_{i\geq 1}\left(\sum_{j\geq 1} |a_{ij}|\right)$ also converges.

Then

(a) Given any bijection $\sigma : \mathbb{N} \to \mathbb{N} \times \mathbb{N}$,
the series $\sum_{n\geq 1} a_{\sigma(n)}$ converges absolutely,
and the value does not depend on $\sigma$;

(b) $\sum_{i\geq 1}\left|\sum_{j\geq 1} a_{ij}\right|$ converges;

(c) $\sum_{i\geq 1} a_{ij}$ coverges absolutely for each $j$;

(d) $\sum_{j\geq 1}\left(\sum_{i\geq 1} |a_{ij}|\right)$ converges;

(e) $\sum_{j\geq 1}\left|\sum_{i\geq 1} a_{ij}\right|$ converges; and

(f) $\sum_{i\geq 1}\left(\sum_{j\geq 1} a_{ij}\right) = \sum_{j\geq 1}\left(\sum_{i\geq 1} a_{ij}\right) = \sum_{n\geq 1} a_{\sigma(n)}$ (for $\sigma$ like in part (a)).

Rudin's proof of conclusion (f) is beautiful, and clever enough to quite possibly leave you wondering how anybody could've thought of it. There is also another proof of conclusion (f) which is more standard and more hands-on, and can be found – for example – in Knopp's *Theory And Application Of Infinite Series*. We'll look at both proofs.

Proofs of (a) - (e) (needed by both Rudin's and Knopp's proof of (f))

Proof of (a):

The key here is to just look at the situation the right way.

Let $\sigma : \mathbb{N} \to \mathbb{N} \times \mathbb{N}$ be any bijection

$$\sigma(n) = (\sigma_1(n), \sigma_2(n)) = \text{“}(i(n), j(n))\text{”}$$

(So, $\sigma$ is a one-to-one correspondence between $\mathbb{N}$ and an "infinite quadrant" $\mathbb{N} \times \mathbb{N} = \{(i, j) : i \geq 1 \text{ and } j \geq 1\}$. One famous example of such a map is the "snake-like and diagonal" one typically used to show that the rationals $\mathbb{Q}$ are countable.)

We then have

$$a_{\sigma(n)} = a_{(\sigma_1(n),\sigma_2(n))} = a_{(i(n),j(n))}$$

Consider a partial sum $s_k$ for the series $\sum_{n\geq 1} \left|a_{\sigma(n)}\right|$:

$$s_k = \sum_{n=1}^{k} \left|a_{\sigma(n)}\right|$$

Here is the crucial observation:

> The image $\sigma(\{1, 2, \ldots, k\})$ of the set $\{1, 2, \ldots, k\}$ under $\sigma$ is finite, and so it's a bounded subset of $\mathbb{N} \times \mathbb{N}$.

This means that there are $I, J \in \mathbb{N}$ such that

$$\begin{aligned} i(n) = \sigma_1(n) &\leq I \qquad \text{for } n \in \{1, 2, \ldots, k\} \\ j(n) = \sigma_2(n) &\leq J \qquad \text{for } n \in \{1, 2, \ldots, k\} \end{aligned}$$

Recalling that $\sigma$ is one-to-one, we see (you should verify this!) that this means

$$s_k = \sum_{n=1}^{k} \left|a_{\sigma(n)}\right| \leq \sum_{i=1}^{I} \left( \sum_{j=1}^{J} |a_{ij}| \right) \leq \sum_{i=1}^{I} \left( \sum_{j\geq 1} |a_{ij}| \right) \leq \sum_{i\geq 1} \left( \sum_{j\geq 1} |a_{ij}| \right)$$

The sequence $\{s_k\}$ of partial sums is thus monotone and bounded, and so converges.

We now know that $\sum_{n\geq 1} \left|a_{\sigma(n)}\right|$ converges, so $\sum_{n\geq 1} a_{\sigma(n)}$ converges absolutely. The fact that the sum of the series does not depend on $\sigma$ follows from the theorem saying that rearranging an absolutely convergent series does not change it's convergence behavior or its limiting sum.

This completes the proof of (a).

Proof of (b):

Once you recall that $\left|\sum_{j\geq 1} a_{ij}\right| \leq \sum_{j\geq 1} |a_{ij}|$ (for any given $i$), the proof of (b) is trivial, since

$$\sum_{i=1}^{k} \left| \sum_{j\geq 1} a_{ij} \right| \leq \sum_{i=1}^{k} \sum_{j\geq 1} |a_{ij}| \leq \sum_{i\geq 1} \sum_{j\geq 1} |a_{ij}|$$

The sequence of partial sums is thus monotone and bounded, and so converges.

This completes the proof of (b).

Proof of (c):

The "from-scratch" proof of (c) is essentially the same as the proof of (a). But in fact, (c) actually also follows immediately from part (a). For a fixed value of $j$, it's easy to find a bijection $\sigma$ such that the sequence $\{a_{ij}\}$ , $i = 1, 2, 3 \ldots,$ is a subsequence of $\{\sigma(i)\}$. For example, you can easily convince yourself by drawing a picture that the "snake-like and diagonal" bijection mentioned above will work.

This completes the proof of (c).

Proof of (d):

Consider a partial sum $\tilde{s}_k$ for (d), and notice that

$$\tilde{s}_k = \sum_{j=1}^{k} \left( \sum_{i \geq 1} |a_{ij}| \right) = \sum_{i \geq 1} \left( \sum_{j=1}^{k} |a_{ij}| \right) \leq \sum_{i \geq 1} \left( \sum_{j \geq 1} |a_{ij}| \right)$$

Convergence follows since the partial sums are a monotone increasing sequence.

This completes the proof of (d).

Proof of (e):

This is like the proof of (b).

<u>Proof of (f) #1 (from Knopp's *Theory And Application Of Infinite Series*)</u>

TODO

<u>Proof of (f)#2 (Rudin's proof)</u>

This is yet another example showing the power of *MUCLIT*.

TODO

14. (Not in book) *Corollary: Corollary #1 of CDST*
TODO: Double series vs single series. From Knopp's proof of part(f).

15. (Not in book) *Corollary: Corollary #2 of CDST*
TODO: Cauchy product vs arbitrary orderings of products of terms.

# 9 Chapter 9: Functions of Several Variables

To be completed.