These lecture notes are based on Rudin’s
*Principles of Mathematical Analysis, 2nd Edition* (McGraw Hill)
Can two different power series define the same function?

(Starting on page 177 of Rudin...)

Suppose $\sum a_n x^n$ and $\sum b_n x^n$ both converge to the same function $f(x)$ on the interval $(-R, R)$.

Question: Is it possible that $a_n \neq b_n$ for some $n$?

Answer: No!

Why? Because by Theorem 8.1, $(\forall n \geq 0)$

$$a_n = \frac{f^{(n)}(0)}{n!} = b_n,$$

So, if two power series are equal on some interval, they must have identical coefficients!
The following theorem shows that we can conclude that two power series must have identical coefficients under a much weaker hypothesis than agreeing on a whole interval!

**Theorem (8.5)**

Suppose the series \( \sum a_n x^n \) and \( \sum b_n x^n \) converge for all \( x \in (-R, R) \). Let

\[
E = \left\{ x \in (-R, R) : \sum a_n x^n = \sum b_n x^n \right\}.
\]

If \( E \) has a limit point in \((-R, R)\), then \((\forall n \geq 0), a_n = b_n\); i.e. the two power series are identical.
Proof of Theorem 8.5

For all \( n \geq 0 \), let \( c_n = a_n - b_n \), and set

\[
f(x) = \sum c_n x^n = \sum a_n x^n - \sum b_n x^n.
\]

Note that \( E = f^{-1}(0) \cap (-R, R) \). Since \( f \) is continuous, \( E \) is a relatively closed subset of \((-R, R)\). Let

\[
A = \{ x \in (-R, R) : x \text{ is a limit point of } E \}
\]

\[
B = (-R, R) - A
\]

Note that \( A \subseteq E \) because \( E \) is closed; also, \( A \) is a closed subset of \((-R, R)\), since the set of limit points of a set is always closed by Exercise 6 in Ch. 2. Therefore, \( B \) is open.

We will now show that \( A \) is also open!
Claim: $A$ is open

Proof of Claim:
Let $x_0 \in A$. By Theorem 8.4, we can expand $f$ as a power series about $x_0$:

$$f(x) = \sum_{n=0}^{\infty} d_n(x - x_0)^n.$$ 

We will now prove that $d_n = 0$ for all $n = 0, 1, 2, 3, \ldots$: Suppose this is not true; let $k$ be the smallest non-negative integer such that $d_k \neq 0$. Then factoring out $(x - x_0)^k$ gives

$$f(x) = (x - x_0)^k g(x)$$

where

$$g(x) = \sum_{m=0}^{\infty} d_{k+m}(x - x_0)^m.$$
Since \( g(x_0) = d_k \neq 0 \) and \( g \) is continuous, it follows that \( \exists \delta > 0 \) such that \( g(x) \neq 0 \) for all \( x \in N_\delta(x_0) \).

Hence, \( f(x) = (x - x_0)^k g(x) \neq 0 \) for all \( x \in N_\delta(x_0) - \{x_0\} \). Hence \( x_0 \) is not a limit point of \( E \). But this contradicts the fact that \( x_0 \in A \)! This contradiction proves that \( d_n = 0 \) for all \( n = 0, 1, 2, 3, \ldots \).

It follows that \( f(x) = 0 \) for all \( x \) such that \( |x - x_0| < R - |x_0| \), proving that \( A \) contains a neighborhood of \( x_0 \), which implies that \( A \) is open since \( x_0 \in A \) was arbitrary.

Thus

\[
(-R, R) = A \cup B
\]

expresses \(( -R, R ) \) as the disjoint union of two open sets!
Since \((-R, R)\) is connected, we conclude that \(A\) or \(B\) must be empty. But \(A \neq \emptyset\), so we conclude that \(B = \emptyset\), implying that \((-R, R) = A\). We showed above that

\[
A \subseteq E \subseteq (-R, R),
\]

so we now know that

\[
(-R, R) \subseteq E \subseteq (-R, R).
\]

Of course it follows that \(E = (-R, R)\). But we observed above that if two power series are equal on a whole interval, their coefficients must all be equal, proving Theorem 8.5.
Remark (illustrated on next slide)

For Theorem 8.5 to apply, the limit point of $E$ must be in $(-R, R)$. If the limit point is $R$ or $-R$, the theorem is not necessarily true. For example, let

$$f(x) = \sin(x/(1 - x^2))$$
$$g(x) = 0$$

These functions can both be expressed as convergent power series on $(-1, 1)$. The set of points at which they are equal has both 1 and -1 as limit points. However, the two functions are obviously not equal to each other on $(-1, 1)$.
$f(x) = \sin\left(\frac{x}{1 - x^2}\right)$ and $g(x) = 0$ for $x \in (-1, 1)$
Define

\[ E(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!} \quad (0! = 1) \]

Ratio Test \(\implies\) \(E(z)\) converges for all \(z \in \mathbb{C}\). Hence we have a well defined function

\[ E : \mathbb{C} \to \mathbb{C}. \]

**Theorem**

For all \(z, w \in \mathbb{C}\), \(E(z)E(w) = E(z + w)\).
Proof

\[ E(z)E(w) = \sum_{n=0}^{\infty} \frac{z^n}{n!} \sum_{m=0}^{\infty} \frac{w^m}{m!} = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \frac{z^n w^m}{n! m!} \]

\[ = \sum_{n=0}^{\infty} \sum_{k=0}^{n} \frac{z^k w^{n-k}}{k!(n-k)!} = \sum_{n=0}^{\infty} \sum_{k=0}^{n} \frac{1}{n!} \frac{z^k w^{n-k}}{k!(n-k)!} \]

\[ = \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{k=0}^{n} \binom{n}{k} z^k w^{n-k} = \sum_{n=0}^{\infty} \frac{1}{n!} (z + w)^n = E(z + w). \]

The third equal sign is obtained by grouping together sets of monomials with constant total degree. The change in order of summation is justified by Theorem 3.50, since the series are absolutely convergent. \qed
Since, for all $z \in \mathbb{C}$,
\[ E(z)E(-z) = E(0) = 1, \]
we conclude that for all $z \in \mathbb{C}$,
\[ E(z) \neq 0 \text{ and } E(-z) = 1/E(z). \]

$x \geq 0 \implies E(x) = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \cdots > 0$.

Since $E(-x) = 1/E(x)$, it follows that
\[ (\forall x \in \mathbb{R}), E(x) > 0. \]

Since for all $x > 0$, $E(x) > 1 + x$, we conclude that
\[ \lim_{x \to \infty} E(x) = \infty. \]

Also,
\[ \lim_{x \to -\infty} E(x) = \lim_{x \to \infty} E(-x) = \lim_{x \to \infty} \frac{1}{E(x)} = 0. \]
Monotonicity of $E(x)$ on $\mathbb{R}$

By definition of $E(z)$, $0 \leq x < y \implies E(x) < E(y)$. Using $E(x) = 1/E(-x)$, one easily deduces:

$E$ is strictly increasing on the entire real axis.
Computing the derivative $E'(x)$

**Lemma**

\[
\lim_{h \to 0} \frac{E(h) - 1}{h} = 1.
\]

**Proof.**

\[
\lim_{h \to 0} \frac{E(h) - 1}{h} = \lim_{h \to 0} \frac{(1 + \frac{h}{1!} + \frac{h^2}{2!} + \ldots) - 1}{h}
\]

\[
= \lim_{h \to 0} \left( \frac{1}{1!} + \frac{h}{2!} + \frac{h^2}{3!} + \ldots \right) = 1.
\]
Theorem

For all $x \in \mathbb{R}$, $E'(x) = E(x)$

Proof.

$$E'(x) = \lim_{h \to 0} \frac{E(x + h) - E(x)}{h} = \lim_{h \to 0} \frac{E(x)E(h) - E(x)}{h}$$

$$= \lim_{h \to 0} E(x) \left( \frac{E(h) - 1}{h} \right) = E(x).$$
Verifying that $E(n) = e^n$ for $n \in J$

From $E(z + w) = E(z)E(w)$, it is easy to deduce using mathematical induction that for all $z_1, \ldots, z_n \in \mathbb{C}$,

$$E(z_1 + z_2 + \cdots + z_n) = E(z_1)E(z_2) \cdots E(z_n).$$

In Chapter 3, we defined $e = 1 + \frac{1}{1!} + \frac{2}{2!} + \ldots$ so obviously $E(1) = e$. Hence, for all $n \in J$,

$$E(n) = E(1 + \cdots + 1) = E(1) \cdots E(1) = e^n.$$
Verifying the $E(p) = e^p$ for rational numbers $p$

Suppose $p = n/m$, $(n, m \in \mathbb{Z})$. Then

$$E(p)^m = E(p) \ldots E(p) = E(mp) = E(n) = e^n.$$  

Since $E(p)$ is a positive real number whose $m$-th power $= e^n$, it follows that

$$E(p) = e^{n/m} = e^p$$  

by the definition of rational exponents given in Ex. 6 of Chapter 1.

If $p \in \mathbb{Q}$ is negative, then

$$E(p) = \frac{1}{E(-p)} = \frac{1}{e^{-p}} = e^p.$$  

So we have shown that for all $p \in \mathbb{Q}$,

$$E(p) = e^p.$$
Since \( \mathbb{Q} \) is dense in \( \mathbb{R} \), \( E(x) \) is the unique continuous function \( \mathbb{R} \to \mathbb{R} \) that agrees with \( e^x \) on \( \mathbb{Q} \). So for now on, we use the notation \( e^z \) or \( \exp(z) \) for the function \( E(z) \).

The following theorem gathers together several important properties of the function \( e^x : \mathbb{R} \to \mathbb{R} \).
Theorem (8.6)

1. \( e^x \) is continuous and differentiable on \( \mathbb{R} \)
2. \((e^x)’ = e^x\)
3. \( e^x \) is strictly increasing and \( e^x > 0 \) for all \( x \in \mathbb{R} \)
4. \( e^{x+y} = e^x e^y \) for all \( x, y \in \mathbb{R} \)
5. \( \lim_{x \to \infty} e^x = \infty; \lim_{x \to -\infty} e^x = 0 \)
6. \( \lim_{x \to \infty} x^n e^{-x} = 0 \) for all \( n \in J \).
We have already proved (1)-(5). For a quick proof of (6), note that the definition of $e^x$ implies that for $x > 0$,

$$e^x > \frac{x^{n+1}}{(n + 1)!}$$

Inverting both sides, reversing the inequality, and multiplying by $x^n$ yields

$$0 < x^n e^{-x} < \frac{(n + 1)!}{x}.$$ 

Letting $x \to \infty$, (6) then follows by the squeeze theorem.
Let $\mathbb{R}_+ = \{x \in \mathbb{R} : x > 0\}$.

Because $e^x$ is continuous and $\lim_{x \to \infty} e^x = \infty$ and $\lim_{x \to -\infty} e^x = 0$, it follows that $\exp : \mathbb{R} \to \mathbb{R}_+$ is surjective.

Because $e^x$ is strictly increasing, $\exp : \mathbb{R} \to \mathbb{R}_+$ is injective; hence $\exp : \mathbb{R} \to \mathbb{R}_+$ is bijective. Therefore $\exp$ has an inverse

$$L : \mathbb{R}_+ \to \mathbb{R}.$$
Some facts about Logarithms

Lemma

$L$ is strictly increasing.

Proof.

We need to prove that for all $x, y \in \mathbb{R}^+$, $x < y \implies L(x) < L(y)$. Proceeding by contrapositive, we will instead prove that $L(x) \geq L(y) \implies x \geq y$. So, assume that $L(x) \geq L(y)$. There are two cases.

Case 1. Suppose $L(x) > L(y)$. Because $\exp$ is increasing, we know $\exp(L(x)) > \exp(L(y))$, so $x > y$.

Case 2. Suppose $L(x) = L(y)$. Then $\exp(L(x)) = \exp(L(y))$, so $x = y$, establishing the contrapositive.
Lemma

$L$ is continuous.

Proof.

Fix $x \in \mathbb{R}$. Note that

$$\exp : [x - 1, x + 1] \to [\exp(x - 1), \exp(x + 1)]$$

is continuous and bijective.

Since $[x - 1, x + 1]$ is compact, Theorem 4.17 implies that

$$L = \exp^{-1} : [\exp(x - 1), \exp(x + 1)] \to [x - 1, x + 1]$$

is also continuous. Since $x$ was arbitrary, $L$ is continuous on $\mathbb{R}_+$. □
Derivative of $L$

**Lemma**

$L : \mathbb{R}_+ \to \mathbb{R}$ is differentiable and $L'(x) = 1/x$. 
Proof.

Let \( x \in \mathbb{R} \). Compute

\[
L'(x) = \lim_{y \to x} \frac{L(y) - L(x)}{y - x} = \frac{1}{\lim_{y \to x} \left( \frac{y-x}{L(y) - L(x)} \right)}
\]

\[
= \frac{1}{\lim_{y \to x} \left( \frac{E(L(y)) - E(L(x))}{L(y) - L(x)} \right)} = \frac{1}{\lim_{z \to L(x)} \left( \frac{E(z) - E(L(x))}{z - L(x)} \right)}
\]

\[
= \frac{1}{E'(L(x))} = \frac{1}{E(L(x))} = \frac{1}{x}
\]
Properties of $L$

Since $E(0) = 1$, $L(1) = 0$. It follows that for all $x \in \mathbb{R}_+$,

$$L(x) = \int_1^x \frac{1}{t} \, dt.$$ 

In some treatments of this subject, this formula is taken as the definition of log, and exp is defined to be $\log^{-1}$!

**Claim:**

For all $u, v \in \mathbb{R}_+$, $L(uv) = L(u) + L(v)$. 
Proof.

\[ L(uv) = L(E(L(u))E(L(v))) = L(E(L(u) + L(v))) = L(u) + L(v). \]

Properties we’ve already proved of \( E \) easily imply that

\[ \lim_{x \to \infty} L(x) = \infty \quad \text{and} \quad \lim_{x \to 0^+} L(x) = -\infty. \]
More properties of $L$

It’s easy to show that for all $p \in \mathbb{Q}$, and for all $x \in \mathbb{R}_+$,

$$pL(x) = L(x^p)$$

and, hence

$$E(pL(x)) = x^p.$$

Motivated by this, the expression “$x^p$” is often defined for all $x \in \mathbb{R}_+$ and for all $p \in \mathbb{R}$ by the formula

$$x^p = E(pL(x))(= e^{p \log x})$$

Recall that a different definition was given in Ex. 6 of Chapter 1, but this one is actually easier to use.
Properties of real exponents

Using this definition of $x^p$, it is easy to derive laws of exponents for real exponents, for example:

\[ x^{\alpha + \beta} = x^\alpha x^\beta \quad \text{and} \quad (x^\alpha)^\beta = x^{\alpha \beta} \]

\[ (xy)^\alpha = x^\alpha y^\alpha \]

\[ \frac{d}{dx} (x^\alpha) = \alpha x^{\alpha - 1} \quad \text{for all} \ x \in \mathbb{R}_+, \alpha \in \mathbb{R}. \]

Henceforth, we write

\[ e^x = E(x) \quad \text{and} \quad \log(x) = L(x) \]
Claim:

\[ \lim_{x \to \infty} x^{-\alpha} \log x = 0 \text{ for all } \alpha > 0. \]

Proof.

L’Hospital’s rule:

\[
\lim_{x \to \infty} x^{-\alpha} \log x = \lim_{x \to \infty} \frac{\log x}{x^\alpha} = \lim_{x \to \infty} \frac{1/x}{\alpha x^{\alpha-1}}
\]

\[= \lim_{x \to \infty} \frac{1}{\alpha x^\alpha} = 0\]
For all $z \in \mathbb{C}$, define

$$C(z) = \frac{1}{2}(e^{iz} + e^{-iz})$$

$$S(z) = \frac{1}{2i}(e^{iz} - e^{-iz})$$
By definition of $e^z$, $\overline{e^z} = e^{\overline{z}}$. Hence if $x \in \mathbb{R}$,

$$\overline{C(x)} = \frac{1}{2}(\overline{e^{ix}} + \overline{e^{-ix}})$$

$$= \frac{1}{2}(e^{-ix} + e^{ix}) = C(x).$$

Similarly for $S(x)$. This shows

$$x \in \mathbb{R} \implies C(x) \in \mathbb{R} \text{ and } S(x) \in \mathbb{R}.$$ 

From the definitions, for all $z \in \mathbb{C}$,

$$e^{iz} = C(z) + iS(z).$$
It follows that if \( x \in \mathbb{R} \), then

\[
C(x) = \text{Re}(e^{ix}) \quad \text{and} \quad S(x) = \text{Im}(e^{ix}).
\]

Also, if \( x \in \mathbb{R} \),

\[
|e^{ix}|^2 = e^{ix} \overline{e^{ix}} = e^{ix} e^{-ix} = 1.
\]

Therefore, for all \( x \in \mathbb{R} \),

\[
|e^{ix}| = 1
\]

Hence for all \( x \in \mathbb{R} \),

\[
C(x)^2 + S(x)^2 = 1.
\]
The definitions of these functions imply that

\[ C(0) = 1 \text{ and } S(0) = 0. \]

Also, differentiating both sides of \( e^{ix} = C(x) + iS(x) \) shows easily that for all \( x \in \mathbb{R} \),

\[ C'(x) = -S(x) \text{ and } S'(x) = C(x). \]

Because \( C(x)^2 + S(x)^2 = 1 \), it follows that the parametrized curve

\[ \gamma(t) = (C(t), S(t)) \]

lies entirely on the unit circle. Note that

\[ |\gamma'(t)| = |(-S(t), C(t))| = \sqrt{C(x)^2 + S(x)^2} = 1. \]
Hence the arclength of $\gamma$ from $\gamma(0)$ to $\gamma(t)$ equals

$$\int_0^t |\gamma'(s)| \, ds = \int_0^t 1 \, ds = t.$$ 

Thus if we start at $(1, 0)$ and move counterclockwise a distance of $t$ along the unit circle,

$$x\text{-coordinate} = C(t)$$

$$y\text{-coordinate} = S(t)$$

This proves that according to the usual definitions of the trig functions,

$$C(t) = \cos t \quad \text{and} \quad S(t) = \sin t.$$
Rudin proves that there exists $x \in \mathbb{R}_+$ such that $C(x) = 0$. Let $x_0$ be the smallest such $x$.

**Definition**

\[
\pi = 2x_0.
\]

$C(\pi/2) = 0 \implies S(\pi/2) = \pm 1$. Since, on $[0, \pi/2]$, $S'(x) = C(x) \geq 0$, $S(x)$ is increasing on $[0, \pi/2]$, so $S(\pi/2) = 1$. It follows that

\[
e^{\pi i/2} = C(\pi/2) + iS(\pi/2) = i.
\]

Hence

\[
e^{\pi i} = i^2 = -1
\]

and

\[
e^{2\pi i} = (-1)^2 = 1
\]

.
Periodicity of exponentials and trig functions

For all $z \in \mathbb{C}$, $n \in \mathbb{Z}$, $e^{z+2\pi in} = e^{z}(e^{2\pi i})^{n} = e^{z}$.

**Theorem (8.7)**

(a) For all $z \in \mathbb{C}$, $e^{z+2\pi i} = e^{z}$

It then follows directly from the definitions that

(b) For all $x \in \mathbb{R}$,

\[ S(x + 2\pi) = S(x) \]
\[ C(x + 2\pi) = C(x) \]

(c) If $0 < t < 2\pi$, then $e^{it} \neq 1$

(d) If $z$ is a complex number with $|z| = 1$, there is a unique $t \in [0, 2\pi)$ such that $e^{it} = z$. 
Therefore, on the domain $0 \leq t \leq 2\pi$,

$$\gamma(t) = e^{it}$$

is a simple closed curve whose range is precisely the unit circle in $\mathbb{C}$ centered at 0.
Suppose $P(z)$ is a polynomial with complex coefficients, and suppose $P(r) = 0$ for some $r \in \mathbb{C}$.

If we divide $(z - r)$ into $P(z)$ by polynomial long division, we get a quotient polynomial $Q(z)$ and a remainder $c \in \mathbb{C}$. Hence

$$P(z) = (z - r)Q(z) + c.$$ 

Substituting $z = r$ in both sides $\Rightarrow c = 0$. Hence, if $P(r) = 0$, then

$$P(z) = (z - r)Q(z)$$

for some polynomial $Q(z)$. 

Eric Klassen, Florida State University
Thus if we can prove that every polynomial $P(z)$ of degree $n$ has a root, we get

$$P(z) = (z - r_1)Q_1(z) = (z - r_1)(z - r_2)Q_2(z) = \ldots$$

$$P(z) = c(z - r_1)(z - r_2)\ldots(z - r_n)$$

Thus it would follow that every polynomial $P(z)$ over $\mathbb{C}$ factors completely into linear factors, and hence has at most $n$ roots.
Theorem (8.8; Fundamental Theorem of Algebra)

Suppose $a_0, \ldots, a_n \in \mathbb{C}$, $n \geq 1$, and $a_n \neq 0$. Let

$$P(z) = \sum_{k=0}^{n} a_k z^k$$

Then there exists $z \in \mathbb{C}$ such that $P(z) = 0$. 
Proof of Theorem 8.8

Without loss of generality, assume $a_n = 1$. Let $\mu = \inf\{|P(z)| : z \in \mathbb{C}\}$. The strategy of our proof is as follows:

1. We prove this inf is actually attained, i.e., that there exists $z_0 \in \mathbb{C}$ such that $|P(z_0)| = \mu$.
2. We prove that $\mu = 0$.

It then follows that $P(z_0) = 0$, proving the theorem!
Proof that $\inf |P(z)|$ is attained

Suppose $z \in \mathbb{C}$, $|z| = R$, where $R > 0$.

Then
\[ z^n = P(z) - a_{n-1}z^{n-1} - \cdots - a_0 \]

\[ \therefore |z^n| \leq |P(z)| + |a_{n-1}| |z|^{n-1} + \cdots + |a_0| \]

\[ R^n \leq |P(z)| + |a_{n-1}| |R|^{n-1} + \cdots + |a_0| \]

\[ \therefore |P(z)| \geq R^n - |a_{n-1}| |R|^{n-1} - \cdots - |a_0| \]

\[ |P(z)| \geq R^n(1 - |a_{n-1}| |R|^{-1} - \cdots - |a_0| |R|^{-n}) \]

Clearly the $RHS \to \infty$ as $R \to \infty$

\[ \therefore \exists R_0 \in \mathbb{R}_+ \text{ such that } \forall R > R_0, \text{ RHS } > \mu + 1 \]
Therefore, $|z| > R_0 \implies |P(z)| > \mu + 1$

$\forall n \in J$, choose $z_n \in \mathbb{C}$ such that $|P(z_n)| < \mu + \frac{1}{n}$.

Clearly, $\{z_n\}$ is a sequence in $NR_0(0)$.

Since $NR_0(0)$ is compact, $\{z_n\}$ has a subsequence $\{y_n\}$ that converges to a point $z_0 \in NR_0(0)$. 
Since $|P(z)|$ is continuous on $N_{R_0}(0)$, it follows by Theorem 4.16 that

$$|P(z_0)| = |P \left( \lim_{n \to \infty} y_n \right)| = \lim_{n \to \infty} |P(y_n)| = \mu;$$

the last equality follows by the squeeze theorem since for all $n \in J$

$$\mu \leq |P(y_n)| \leq \mu + \frac{1}{n}.$$ 

Thus we have found $z_0 \in \mathbb{C}$ such that $|P(z_0)| = \mu$. 

Eric Klassen, Florida State University

Online Notes for Advanced Calculus II (MAA4227/MAA5307)
Proof that $\mu = 0$

Suppose $\mu = |P(z_0)| \neq 0$. Let $Q(z) = \frac{P(z+z_0)}{P(z_0)}$.

$Q(z)$ is a non-constant polynomial, $Q(0) = 1$, and $|Q(z)| \geq 1$ ($\forall z \in \mathbb{C}$). Write

$$Q(z) = 1 + b_k z^k + \cdots + b_n z^n,$$

where $b_k \neq 0$.

By Theorem 8.7 (4), $\exists \theta \in \mathbb{R}$ such that

$$e^{ik\theta} = -\frac{|b_k|}{b_k},$$

since $-\frac{|b_k|}{b_k}$ has magnitude $= 1$, i.e.,

$$e^{ik\theta} b^k = -|b_k|.$$
If $r > 0$, it follows that

$$|1 + b_k r^k e^{ik\theta}| = |1 - r^k|b_k||$$

If we also assume $r^k < \frac{1}{|b_k|}$, then $|1 + b_k r^k e^{ik\theta}| = 1 - r^k|b_k|$

By definition of $Q$,

$$|Q(re^{i\theta})| \leq |1 + b_k r^k e^{ik\theta}| + |b_{k+1}|r^{k+1} + \cdots + |b_n|r^n$$

$$\therefore |Q(re^{i\theta})| \leq 1 - r^k|b_k| + |b_{k+1}|r^{k+1} + \cdots + |b_n|r^n$$

$$\therefore |Q(re^{i\theta})| \leq 1 - r^k(|b_k| - r|b_{k+1}| - \cdots - r^{n-k}|b_n|)$$

For sufficiently small $r > 0$, the last parenthesis is $> 0$.

Hence, for sufficient small $r$, $|Q(re^{i\theta})| < 1$. Contradiction!

Therefore $\mu = 0$, so $P(z_0) = 0$.
"$V$ is a vector space over $\mathbb{C}$"

Completely analogous to vector spaces over $\mathbb{R}$, except scalars $\in \mathbb{C}$.

E.g. if $\alpha, \beta \in \mathbb{C}$, $u, v, w \in V$

\[
\begin{align*}
\nu + \omega &= \omega + \nu \\
(u + \nu) + \omega &= u + (\nu + \omega) \\
\alpha(\nu + \omega) &= \alpha\nu + \alpha\omega \\
(\alpha + \beta)v &= \alpha v + \beta v \\
1 \cdot \nu &= \nu \\
0 \cdot \nu &= 0 \in V
\end{align*}
\]

Definitions of span, basis, linear independence, dimension all the same as over $\mathbb{R}$. 
The most familiar example: \( \mathbb{C}^n \) is a vector space over \( \mathbb{C} \) of dimension \( n \).

**Definition**

A *Hermitian inner product (HIP)* on a vector space \( V \) over \( \mathbb{C} \) is a function \( \langle \ , \ \rangle : V \times V \to \mathbb{C} \) such that, assuming \( u, v \in V, \lambda \in \mathbb{C} \)

\[
\begin{align*}
1 & : \quad \langle u, v \rangle = \overline{\langle v, u \rangle} \quad \text{(conjugate – symmetric)} \\
2 & : \quad \langle \lambda u, v \rangle = \lambda \langle u, v \rangle \\
3 & : \quad \langle u_1 + u_2, v \rangle = \langle u_1, v \rangle + \langle u_2, v \rangle 
\end{align*}
\]

**Notes:**

\( 1 + 3 \implies \langle u, v_1 + v_2 \rangle = \langle u, v_1 \rangle + \langle u, v_2 \rangle \)

\( 1 + 2 \implies \langle u, \lambda v \rangle = \overline{\lambda} \langle u, v \rangle \)
① $\Rightarrow (\forall u \in V), \langle u, u \rangle \in \mathbb{R}$

since $\langle u, u \rangle = \overline{\langle u, u \rangle}$

To be a HIP, we also require

④ $\langle u, u \rangle \geq 0, \forall u \in V,$ and
$\langle u, u \rangle = 0 \iff u = 0$

The norm (or magnitude) is determined by

$$|u| = \sqrt{\langle u, u \rangle}.$$
**Standard example:** \( V = \mathbb{C}^n; \) if

\[
\mathbf{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \quad \text{and} \quad \mathbf{w} = \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix}
\]

Define

\[
\langle \mathbf{v}, \mathbf{w} \rangle = \sum_{i=1}^{n} v_i \overline{w_i}
\]

Just like ordinary dot-product in \( \mathbb{R}^n \), except for taking complex conjugates of second factors!
Definitions

If \( v, w \in V \), \( v \) is orthogonal to \( w \) if \( \langle v, w \rangle = 0 \).
(Note: \( \langle v, w \rangle = 0 \iff \langle w, v \rangle = 0 \))

A set \( \{v_1, \ldots, v_k\} \subseteq V \) is orthonormal \iff \( \langle v_i, v_j \rangle = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases} \)

Suppose \( \{v_1, \ldots, v_n\} \) is an orthonormal set and
\( u = \alpha_1 v_1 + \cdots + \alpha_n v_n \). Then

\[
\langle u, u \rangle = \alpha_1 \overline{\alpha_1} + \cdots + \alpha_n \overline{\alpha_n} = |\alpha_1|^2 + \cdots + |\alpha_n|^2
\]

So \( |u| = \sqrt{\sum |\alpha_i|^2} \)
Still assuming that $\{v_1, \ldots, v_n\}$ is orthonormal and 
$u = \alpha_1 v_1 + \cdots + \alpha_n v_n$,

$$\langle u, v_i \rangle = \alpha_i \quad (\forall i)$$

Hence, if we know $u$ is a linear combination of the $\{v_i\}$, we can recover the coefficients by this formula.

**Important example**

Let $V = \{\text{continuous functions } [a, b] \to \mathbb{C}\}$

This is a complex vector space.

Define a Hermitian inner product on $V$ by

$$\langle f, g \rangle = \int_a^b f(x)\overline{g(x)} \, dx$$
Lemma

If $V$ is a Hermitian $\mathbb{C}$-vector space, $v, w \in V$, and $v \perp w$, then

$$\|\vec{v} - \vec{w}\|^2 = \|\vec{v}\|^2 + \|\vec{w}\|^2,$$

also

$$\|\vec{v} + \vec{w}\|^2 = \|\vec{v}\|^2 + \|\vec{w}\|^2$$

Proof.

$$\|\vec{v} - \vec{w}\|^2 = \langle v - w, v - w \rangle$$

$$= \langle v, v \rangle - \langle w, v \rangle - \langle v, w \rangle + \langle w, w \rangle$$

$$= \|v\|^2 + \|w\|^2$$

[Note: if $\langle v, w \rangle = 0$, then $\langle w, v \rangle = 0$]
Lemma

Suppose \( \{e_1, \ldots, e_n\} \subseteq V \) are orthonormal and \( v \in V \); let
\[
c_i = \langle v, e_i \rangle \quad \forall i = 1, \ldots, n.
\]
Then
\[
\left( v - \sum_{i=1}^{n} c_i e_i \right) \perp e_j \quad \forall j = 1, \ldots, n
\]

Proof.

\[
\langle v - \sum_{i=1}^{n} c_i e_i, e_j \rangle = \langle v, e_j \rangle - c_j \langle e_j, e_j \rangle = 0
\]
Lemma

Let \( \{e_1, \ldots e_n\} \subseteq V \) be orthonormal and \( v \in V \). Let \( c_i = \langle v, e_i \rangle \) for all \( i = 1, \ldots, n \).

The element of \( \text{span}\{e_1, \ldots, e_n\} \) that is closest to \( v \) is \( \sum c_i e_i \), and no other element of \( \text{span}\{e_1, \ldots, e_n\} \) is equally close to \( v \).

Proof.

Let \( \sum d_i e_i \in \text{span} \{e_1, \ldots, e_n\} \).

\[
|v - \sum d_i e_i|^2 = |(v - \sum c_i e_i) + \sum (c_i - d_i) e_i|^2 \\
= |v - \sum c_i e_i|^2 + |\sum (c_i - d_i) e_i|^2 \\
= |v - \sum c_i e_i|^2 + \sum |c_i - d_i|^2
\]

Clearly, this is minimized when \( d_i = c_i \) (\( \forall i \)).
Definition
Suppose $V$ is an HIP, $\{e_1, \ldots, e_n\}$ is an orthonormal set, and $v \in V$. For each $i = 1, \ldots, n$, let $c_i = \langle v, e_i \rangle$. Then we call
$$\sum_{i=1}^{n} c_i e_i$$
the projection of $v$ into $\text{span}\{e_1, \ldots, e_n\}$. It has two important properties:

1. Of all vectors in $\text{span}\{e_1, \ldots, e_n\}$, it is closest to $v$.
2. The difference $v - \sum_{i=1}^{n} c_i e_i$ is perpendicular to $e_i$ for all $i = 1, \ldots, n$. 

Fourier Series

**Definition**

A **trigonometric polynomial** is a finite sum of the form

\[ f(x) = a_0 + \sum_{n=1}^{N} (a_n \cos(nx) + b_n \sin(nx)) \] (⋆)

where \( x \in \mathbb{R} \) and \( a_i, b_i \in \mathbb{C} \)

**Note:** \( f : \mathbb{R} \to \mathbb{C} \)

By the definitions of \( \cos x \) and \( \sin x \) in terms of \( e^{ix} \), (⋆) can also be written

\[ f(x) = \sum_{n=-N}^{N} c_n e^{inx} \] (⋆′)
where

\[ c_n = \frac{1}{2}(a_n - ib_n) \quad \text{and} \quad c_{-n} = \frac{1}{2}(a_n + ib_n) \]

for \( n = 1, \ldots, N \) and \( c_0 = a_0 \). Clearly every trigonometric polynomial is periodic with period \( 2\pi \).

By the Fundamental Theorem of Calculus,

\[ \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{inx} \, dx = \begin{cases} 1, & n = 0 \\ 0, & n \neq 0, \ n \in \mathbb{Z} \end{cases} \]
Lemma
\[ \left\{ \frac{1}{\sqrt{2\pi}} e^{inx} \right\}_{n=-\infty}^{\infty} \] is an orthonormal set in the Hermitian vector space of continuous functions \([-\pi, \pi] \rightarrow \mathbb{C} \)

Proof.
\[
\left\langle \frac{1}{\sqrt{2\pi}} e^{inx}, \frac{1}{\sqrt{2\pi}} e^{imx} \right\rangle = \int_{-\pi}^{\pi} \frac{1}{2\pi} e^{inx} e^{-imx} \, dx \\
= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(n-m)x} \, dx = \begin{cases} 1, & n = m \\ 0, & n \neq m \end{cases}
\]
Lemma

If \( f(x) = \sum_{n=-N}^{N} c_n e^{inx} \) then for all \( n = -N, \ldots, N \),

\[
c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-inx} \, dx.
\]

Proof.

\[
\int_{-\pi}^{\pi} f(x) e^{-imx} \, dx = \sum_{n=-N}^{N} c_n \int_{-\pi}^{\pi} e^{i(n-m)x} \, dx = 2\pi c_m
\]

\[
\therefore \quad \forall m = -N, \ldots, N \quad c_m = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-imx} \, dx
\]
Lemma

If \( f(x) = \sum_{n=-N}^{N} c_n e^{inx} \), then

\[
f(x) \in \mathbb{R} \quad (\forall x) \iff c_{-n} = \overline{c_n} \quad (\forall n = 0, \ldots, N)
\]

Proof.

( \implies \) If \( f(x) \) is real, then

\[
c_{-n} = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x)e^{inx} \, dx = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x)e^{-inx} \, dx = \overline{c_n}
\]
If each $c_n = \overline{c_n}$, then

$$f(x) = c_0 + \sum_{n=1}^{N} c_n e^{inx} + \overline{c_n} e^{-inx}$$

It follows immediately that

$$f(x) = \overline{f(x)} \quad (\forall x), \quad \text{so} \quad f(x) \in \mathbb{R}$$

Continuing to assume that $f : \mathbb{R} \to \mathbb{C}$ is defined by

$$f(x) = \sum_{n=-N}^{N} c_n e^{inx}, \quad (\star)$$

clearly $f$ is periodic on $\mathbb{R}$, with period $2\pi$. 

Eric Klassen, Florida State University  
Online Notes for Advanced Calculus II (MAA4227/MAA5307)
A trigonometric series is a series of the form

\[ \sum_{n=-\infty}^{\infty} c_n e^{inx} \quad x \in \mathbb{R}, \ c_n \in \mathbb{C} \]

Its \( N \)-th partial sum is defined to be

\[ \sum_{n=-N}^{N} c_n e^{inx} \]
If \( f \in \mathcal{R} \) (i.e., if \( f \) is Riemann integrable) on \([-\pi, \pi]\), then for all \( n \in \mathbb{Z} \), we define

\[
c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-inx} \, dx
\]

and then define

\[
\sum_{n=-\infty}^{\infty} c_n e^{inx}
\]

to be the Fourier series of \( f(x) \).

(The numbers \( \{c_n\} \) are called the Fourier coefficients of \( f \).)
Main Question:

Under what circumstances (and in what sense) does the Fourier series of $f$ converge to $f$?

We have already shown that if $f$ is a finite trigonometric polynomial, then it is equal to its Fourier series (since its Fourier series will be the same trigonometric polynomial).

Definition

An orthogonal system of functions on $[a, b]$ is a sequence $\{\phi_n\}$ of functions $[a, b] \rightarrow \mathbb{C}$ such that,

$$\forall n \neq m, \quad \int_a^b \phi_n(x)\overline{\phi_m(x)}dx = 0$$
Definition

If, in addition,

\[(\forall n) \int_a^b |\phi_n(x)|^2 \, dx = 1\]

then \(\{\phi_n\}\) is said to be orthonormal.

We’ve verified that the functions \(\left\{\frac{e^{inx}}{\sqrt{2\pi}}\right\}_{n \in \mathbb{Z}}\) form an orthonormal system on \([-\pi, \pi]\).

It’s easy to verify that another orthonormal system on \([-\pi, \pi]\) is

\(\left\{\frac{1}{\sqrt{2\pi}}, \frac{\cos x}{\sqrt{\pi}}, \frac{\sin x}{\sqrt{\pi}}, \frac{\cos 2x}{\sqrt{\pi}}, \frac{\sin 2x}{\sqrt{\pi}}, \ldots\right\}\)
We now have two different ways of making $\{\text{functions } [a, b] \to \mathbb{C}\}$ into a metric space.

1. $\text{dist}(f, g) = \sup\{|f(x) - g(x)| : x \in [a, b]\}$.

2. $\text{dist}(f, g) = \left(\int_{a}^{b} (f(x) - g(x))(f(x) - g(x)) \, dx\right)^{1/2}$.

The first is often referred to as the “sup-metric” while the second is the “$L^2$-metric”. Note that the $L^2$-metric is simply the magnitude $|f - g|$ with respect to the HIP we have defined on the complex vector space of functions.
Given any \( f \in \mathcal{R} \) on \([a, b]\), and any orthonormal system \( \{\phi_n\} \) on \([a, b]\), we define the \( n \)-th Fourier coefficient of \( f \) relative to \( \{\phi_n\} \) by

\[
c_n = \int_a^b f(t)\phi_n(t)\,dt
\]

We write

\[
f(x) \sim \sum_n c_n\phi_n(x) \quad (\triangle)
\]

and call this the Fourier series of \( f \) relative to \( \{\phi_n\} \).

\textit{Note:} We are saying nothing about the convergence of the series \((\triangle)\); just naming it the Fourier series of \( f \).
Theorem (8.11)

Let \( \{\phi_n\}_{n \in \mathbb{J}} \) be an orthonormal system on \([a, b]\).

Let

\[
s_n(x) = \sum_{m=1}^{n} c_m \phi_m(x)
\]

be the \( n \)-th partial sum of the Fourier series of \( f \), and let

\[
t_n(x) = \sum_{m=1}^{n} \gamma_m \phi_m(x)
\]

be an arbitrary function in \( \text{span}\{\phi_1, \ldots, \phi_n\} \), where each \( \gamma_m \in \mathbb{C} \).
Theorem (8.11)

Then
\[ \int_a^b |f - s_n|^2 \, dx \leq \int_a^b |f - t_n|^2 \, dx \]
and equality holds \( \iff \gamma_m = c_m, \quad \forall m = 1, \ldots, n. \)

Proof

Mimics earlier proof for general HIP spaces.

First, note that
\[ \int_a^b |t_n|^2 \, dx = \sum_{m=1}^n |\gamma_m|^2. \]
\[ \int |f - t_n|^2 = \int (f - t_n)(\overline{f} - \overline{t_n}) \]
\[ = \int f\overline{f} - \int f\overline{t_n} - \int \overline{f}t_n + \int t_n\overline{t_n} \]
\[ = \int |f|^2 - \sum c_m\gamma_m - \sum \overline{c_m}\gamma_m + \sum \gamma_m\overline{\gamma_m} \]
\[ = \int |f|^2 - \sum c_m\overline{c_m} + \sum c_m\overline{c_m} \]
\[ - \sum c_m\overline{\gamma_m} - \sum \overline{c_m}\gamma_m + \sum \gamma_m\overline{\gamma_m} \]
\[ = \int |f|^2 - \sum |c_m|^2 + \sum (c_m - \gamma_m)(\overline{c_m} - \overline{\gamma_m}) \]
\[ = \int |f|^2 - \sum |c_m|^2 + \sum |c_m - \gamma_m|^2 \]

Clearly, this achieves its minimum \( \iff \gamma_m = c_m \) for all \( m \). \( \square \)
Setting $\gamma_m = c_m$, $t_n = s_n$ in last equation, shows

\[
\int |f - s_n|^2 = \int |f|^2 - \sum_{m=1}^{n} |c_m|^2
\]

\[
\int |f - s_n|^2 = \int |f|^2 - \int |s_n|^2
\]

\[
\therefore \int |s_n|^2 \leq \int |f|^2 \quad (\star)
\]

(Projection onto subspace spanned by $\phi_1, \ldots, \phi_n$ is smaller than original vector).
Theorem (8.12)

If \( \{ \phi_n \} \) is an orthonormal system on \([a, b]\), and if

\[
f(x) \sim \sum_{n=1}^{\infty} c_n \phi_n(x)
\]

then

\[
\sum_{n=1}^{\infty} |c_n|^2 \leq \int_a^b |f(x)|^2 \, dx
\]

In particular,

\[
\lim_{n \to \infty} c_n = 0
\]
Proof.

Equation (⋆) in the last proof shows

\[ \sum_{m=1}^{n} |c_m|^2 \leq \int |f|^2 \] for all \( n \).

Since

\[ \left\{ \sum_{m=1}^{n} |c_m|^2 \right\}_{n=1}^{\infty} \]

is a monotone increasing sequence that is bounded above by \( \int |f|^2 \), it converges to a number \( \leq \int |f|^2 \).

The fact that \( \sum |c_n|^2 \) converges \( \implies \lim_{n \to \infty} c_n = 0 \).
We now assume that $f : \mathbb{R} \rightarrow \mathbb{C}$ has period $2\pi$ and $f \in \mathcal{R}$ on $[-\pi, \pi]$ (hence on every bounded interval). Recall

**Definition (Fourier series of $f$)**

$$f(x) \sim \sum_{n=-\infty}^{\infty} c_n e^{inx}$$

where each

$$c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x)e^{-inx} \, dx$$

Let

$$S_N(x) = s_N(f; x) = \sum_{n=-N}^{N} c_n e^{inx}$$

be the $N$-th partial sum of the Fourier series.
We have proved that (8.12)

\[
\frac{1}{2\pi} \int_{-\pi}^{\pi} |s_N(x)|^2 \, dx = \sum_{n=-N}^{N} |c_n|^2 \leq \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(x)|^2 \, dx
\]

Define the Dirichlet kernel by

\[
D_N(x) = \sum_{n=-N}^{N} e^{inx}
\]

Note:

\[
(e^{ix} - 1)D_N(x) = e^{i(N+1)x} - e^{-iNx}
\]

Multiplying both sides by \(e^{-\frac{1}{2}x}\)

\[
(e^{\frac{1}{2}ix} - e^{-\frac{1}{2}ix})D_N(x) = e^{i(N+\frac{1}{2})x} - e^{-i(N+\frac{1}{2})x}
\]
\[ D_N(x) = \frac{\sin \((N + \frac{1}{2})x\)}{\sin \(\frac{1}{2}x\)} \]

Now, using \( t = x - u \), \( u = x - t \), \( du = -dt \),

\[
s_N(x) = \sum_{n=-N}^{N} \frac{1}{2\pi} \left( \int_{-\pi}^{\pi} f(t)e^{-int} dt \right) e^{inx}
\]

\[
= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) \left( \sum_{n=-N}^{N} e^{i(n(x-t))} \right) dt
\]

\[
= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) D_N(x - t) dt
\]

Now change variables from \( t \) to \( u \):
\[ s_N(x) = -\frac{1}{2\pi} \int_{x+\pi}^{x-\pi} f(x-u)D_N(u) \, du \]

\[ = \frac{1}{2\pi} \int_{x-\pi}^{x+\pi} f(x-t)D_N(t) \, dt \]

\[
\therefore s_N(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x-t)D_N(t) \, dt
\]

(Periodicity \(\Rightarrow\) the interval of integration doesn’t matter, as long as its length is \(2\pi\).)
Finally, a theorem about pointwise convergence!

Theorem (8.14)

Assume \( f \) has period \( 2\pi \) and \( f \in \mathcal{R} \) on \( [-\pi, \pi] \), and suppose \( x \in \mathbb{R} \). If \( \exists \delta > 0 \) and \( M < \infty \) such that

\[
|f(x + t) - f(x)| \leq M|t| \quad (\forall t \in (-\delta, \delta))
\]

then

\[
\lim_{N \to \infty} s_N(x) = f(x)
\]

In other words, the Fourier series of \( f \) converges to \( f \) at the point \( x \).
Clearly, by Mean Value Theorem, the hypothesis holds if $f$ is differentiable on $(x - \delta, x + \delta)$ and $f'$ is bounded on $(x - \delta, x + \delta)$. Can just take

$$M = \sup_{t \in (-\delta, \delta)} |f'(x + t)|.$$

Thus Theorem 8.14 implies that if $f$ is differentiable and $f'$ is bounded on $\mathbb{R}$,

$$s_N(x) \to f(x) \quad \forall x \in \mathbb{R}$$

**Proof of 8.14**

Fixing $x \in \mathbb{R}$, define:

$$g(t) = \begin{cases} 
\frac{f(x-t)-f(x)}{\sin(t/2)} & t \in [-\pi, \pi] - \{0\} \\
0 & t = 0
\end{cases}$$
An easy computation shows that \( \frac{1}{2\pi} \int_{-\pi}^{\pi} D_N(t) \, dt = 1 \).

Therefore

\[
\begin{align*}
  s_N(x) - f(x) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x - t)D_N(t) \, dt - \frac{1}{2\pi}f(x) \int_{-\pi}^{\pi} D_N(t) \, dt \\
  &= \frac{1}{2\pi} \int_{-\pi}^{\pi} (f(x - t) - f(x)) \, D_N(t) \, dt \\
  &= \frac{1}{2\pi} \int_{-\pi}^{\pi} (f(x - t) - f(x)) \frac{\sin ((N + \frac{1}{2})t)}{\sin(\frac{1}{2}t)} \, dt \\
  &= \frac{1}{2\pi} \int_{-\pi}^{\pi} g(t) \sin \left( (N + \frac{1}{2})t \right) \, dt \\
  &= \frac{1}{2\pi} \int_{-\pi}^{\pi} g(t) \cos \left( \frac{t}{2} \right) \sin Nt \, dt + \frac{1}{2\pi} \int_{-\pi}^{\pi} g(t) \sin \left( \frac{t}{2} \right) \cos Nt \, dt
\end{align*}
\]
It is easy to show \( g(t) \cos \left( \frac{t}{2} \right) \) and \( g(t) \sin \left( \frac{t}{2} \right) \) are bounded and Riemann integrable (bounded follows from the hypothesis on \( f \) near \( x \); integrable follows from 6.10, which says that if \( f \) is bounded and has finitely many discontinuities, then it is integrable).

By Theorem 8.12, both these terms \( \to 0 \) as \( N \to \infty \) because they represent Fourier coefficients with respect to a certain orthonormal basis (sines and cosines).

This proves \( s_N(x) \to f(x) \) as \( N \to \infty \).

\[ \square \]

**Corollary**: If \( f(x) = 0 \) for all \( x \in (a, b) \), then

\[ s_N(x) \to 0 \quad \forall x \in (a, b). \]
Example:

\[ f(x) = 0 \quad \text{and} \quad g(x) \]

\( f \) and \( g \) are both pointwise limits of Fourier series; they agree on \((0, \pi/2)\), but not everywhere!
Contrast:
If two functions are analytic (i.e. sums of power series) and they agree on an interval, they agree everywhere by Theorem 8.5.

If two functions are sums of Fourier series, they can agree on an interval, but disagree elsewhere!

“By changing the Fourier coefficients, can change behavior of $f$ in one interval, while not changing it in another interval.”
Theorem (8.15)

If \( f \) is continuous with period \( 2\pi \) and \( \varepsilon > 0 \), then \( \exists \) a trigonometric polynomial \( P \) such that

\[
|P(x) - f(x)| < \varepsilon \quad \forall x \in \mathbb{R}
\]

Proof.

Think of periodic functions as functions \( S^1 \to \mathbb{C} \)
\( (S^1 = \text{unit circle in } \mathbb{C}) \) by

\[
\tilde{f}(e^{ix}) = f(x)
\]

Then use Theorem 7.33.
Theorem (8.16 - Parseval’s Theorem)

Suppose \( f, g \in \mathcal{R} \) on \([-\pi, \pi]\), both have period \(2\pi\), and have Fourier series

\[
f(x) \sim \sum_{-\infty}^{\infty} c_n e^{inx} \quad g(x) \sim \sum_{-\infty}^{\infty} \gamma_n e^{inx}
\]

Then:

1. \[
\lim_{N \to \infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(x) - s_N(f; x)|^2 \, dx = 0
\]

2. \[
\frac{1}{2\pi} \int_{-\pi}^{\pi} f(x)g(x) \, dx = \sum_{-\infty}^{\infty} c_n \gamma_n
\]

3. \[
\frac{1}{2\pi} \int_{-\pi}^{\pi} |f(x)|^2 \, dx = \sum_{-\infty}^{\infty} |c_n|^2
\]
In Theorem 8.14, we showed that under certain hypotheses, the Fourier series of $f$ converges to $f$ pointwise. Part \( \textcircled{1} \) of Parseval’s Theorem shows that, under weaker hypotheses, the Fourier series of $f$ converges to $f$ with respect to the $L^2$ metric. (In fact these hypotheses can further be weakened using Lebesgue integration.)

Parts \( \textcircled{2} \) and \( \textcircled{3} \) of Parseval’s Theorem show that we can calculate Hermitian inner products and $L^2$ norms of functions in terms of their Fourier coefficients, just as we calculate inner products and norms of vectors in $\mathbb{C}^n$ using their coefficients w.r.t. the standard orthonormal basis.
Proof of Theorem 8.16

For any $h \in \mathcal{R}$ on $[-\pi, \pi]$, define

$$\|h\|_2 = \sqrt{\frac{1}{2\pi} \int_{-\pi}^{\pi} |h(x)|^2 \, dx}$$

Let $\varepsilon > 0$. Though $f$ might not be continuous, it is shown in Exercise 12 of Chapter 6 that by linearly interpolating values of $f$ from a sufficiently fine partition of $[-\pi, \pi]$, we obtain a continuous function $h$ with period $2\pi$ such that

$$\|f - h\|_2 < \varepsilon$$
Then Theorem 8.15 $\implies \exists$ a trigonometric polynomial $P$ such that $|h(x) - P(x)| < \varepsilon$ for all $x$. Hence $\|h - P\|_2 < \varepsilon$.

Assuming $P$ has degree $N_0$, Theorem 8.11 $\implies$

$$\|h - s_N(h)\|_2 \leq \|h - P\|_2 < \varepsilon \quad \forall N \geq N_0$$

Using the inequality right before Theorem 8.12, we showed

$$\|s_N(h) - s_N(f)\|_2 = \|s_N(h - f)\|_2 \leq \|h - f\|_2 < \varepsilon$$

[In fact, we showed $\int |s_N|^2 \leq \int |f|^2$, $\forall f \in \mathcal{R}$]
Now, the triangle inequality for $\| \cdot \|_2$ (from Ex. 11, Chap. 6) \[ \Rightarrow \]
\[ \| f - s_N(f) \|_2 \leq \| f - h \|_2 + \| h - s_N(h) \|_2 + \| s_N(h) - s_N(f) \| \]
\[ < 3\varepsilon \]
for all $N \geq N_0$. This proves $\textcircled{1}$. 

Next

\[ \frac{1}{2\pi} \int_{-\pi}^{\pi} s_N(f) \overline{g} \, dx = \sum_{n=-N}^{N} c_n \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{inx} \overline{g(x)} \, dx \]
\[ = \sum_{n=-N}^{N} c_n \overline{\gamma_n} \quad \text{(*)} \]
\[
\left| \int f \overline{g} - \int s_N(f) \overline{g} \right| \leq \int |f - s_N(f)||g|
\]

(Schwarz inequality) \leq \left( \int |f - s_N|^2 \int |g|^2 \right)^{\frac{1}{2}}

By ①, the RHS → 0 as \( N \to \infty \).

Substituting (⋆) in the LHS, implies ②.

③ follows from ② by substituting \( f \) for \( g \)!
In Exercise 10 on page 139 of Rudin, we worked out a proof of Hölder’s Inequality. It states that if $p$ and $q$ are positive real numbers that satisfy

$$\frac{1}{p} + \frac{1}{q} = 1$$

and $f$ and $g$ are Riemann-integrable functions $[a, b] \to \mathbb{C}$, then

$$\left| \int_a^b fg \, dx \right| \leq \left( \int_a^b |f|^p \, dx \right)^{1/p} \left( \int_a^b |g|^q \, dx \right)^{1/q}.$$

The special case of $p = q = 2$ is just the Schwarz inequality.
Gamma function

Definition

For $0 < x < \infty$, define

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} \, dt$$

*Note*: Need to look separately at $\int_0^1$ and $\int_1^\infty$ - good Calculus II problem to show it converges!

(For $\int_1^\infty$, first do for integers $x = n$ by math induction. Then a non-integer is dominated by an integer.)
Theorem (8.18)

1. $\forall \, 0 < x < \infty$, $\Gamma(x + 1) = x\Gamma(x)$ “functional equation”
2. $\Gamma(n + 1) = n! \quad \forall n = 0, 1, 2, 3, \ldots$
3. $\log \circ \Gamma$ is convex (concave up) on $(0, \infty)$

$g$ is “convex” (Exercise 23, page 101) if and only if

$$g(\lambda x + (1 - \lambda)y) \leq \lambda g(x) + (1 - \lambda)g(y)$$
Proof.

1. Integration by parts

2. First note $\Gamma(1) = 1$; then use mathematical induction and part ①.

3. Assume $\frac{1}{p} + \frac{1}{q} = 1$.

Hölder’s inequality $\Rightarrow \Gamma \left( \frac{x}{p} + \frac{y}{q} \right) \leq \Gamma(x)^{\frac{1}{p}} \Gamma(y)^{\frac{1}{q}}$

③ follows.

Recall that Hölder’s inequality was proved in Exercise 10, page 139 of Rudin.
Theorem (8.19)

If \( f : (0, \infty) \to (0, \infty) \) and, for all \( x > 0 \), it satisfies:

1. \( f(x + 1) = xf(x) \quad \forall x \in (0, \infty) \)
2. \( f(1) = 1 \)
3. \( \log \circ f \) is convex

then

\[ f(x) = \Gamma(x) \]

Proof

Since \( \Gamma \) satisfies these conditions, it is enough to show that \( f \) is uniquely determined by the conditions. Because of (1), it is enough to do this for \( 0 < x \leq 1 \).
Let $\varphi(x) = \log(f(x))$.

Then

$$\varphi(x + 1) = \log(f(x + 1)) = \log(xf(x)) = \log(x) + \log(f(x)) = \log(x) + \varphi(x) \quad (0 < x < \infty)$$

Also, $\varphi(1) = 0$ and $\varphi$ is convex.

By ① and ②, $f(n + 1) = n! \quad (\forall n \in J)$. So,

$$\varphi(n + 1) = \log(n!)$$

$$\therefore \varphi(n + 1) - \varphi(n) = \log(n!) - \log((n - 1)!) = \log(n)$$

$$\varphi(n + 1) - \varphi(n) = \log(n) \quad (\text{for } n \in J)$$
By convexity,

\[
\varphi \left( \frac{x}{1+x}n + \frac{1}{1+x}(n+1+x) \right) \leq \frac{x}{1+x} \varphi(n) + \frac{1}{1+x} \varphi(n+1+x)
\]

\[
\varphi(n+1) \leq \frac{x \varphi(n) + \varphi(n+1+x)}{1+x}
\]

\[
x \varphi(n+1) + \varphi(n+1) \leq x \varphi(n) + \varphi(n+1+x)
\]

\[
\varphi(n+1) - \varphi(n) \leq \frac{\varphi(n+1+x) - \varphi(n+1)}{x}
\]

\[
\log(n) \leq \frac{\varphi(n+1+x) - \varphi(n+1)}{x}
\]
Similarly,

\[ \varphi((1 - x)(n + 1) + x(n + 2)) \leq (1 - x)\varphi(n + 1) + x\varphi(n + 2) \]
\[ \varphi(n + 1 + x) - \varphi(n + 1) \leq x(\varphi(n + 2) - \varphi(n + 1)) \]

\[ \frac{\varphi(n + 1 + x) - \varphi(n + 1)}{x} \leq \log(n + 1) \]

So,

\[ 2 \leq \log(n) \leq \frac{\varphi(n + 1 + x) - \varphi(n + 1)}{x} \leq \log(n + 1) \]
\[ \varphi(x + 1) = \varphi(x) + \log(x) \]
\[ \varphi(x + 2) = \varphi(x + 1) + \log(x + 1) \]
\[ = \varphi(x) + \log(x) + \log(x + 1) \]
\[ \varphi(x + 3) = \varphi(x + 2) + \log(x + 2) \]
\[ = \varphi(x) + \log(x) + \log(x + 1) + \log(x + 2) \]
\[ \vdots \]
\[ \varphi(x + n + 1) = \varphi(x) + \log(x) + \log(x + 1) + \cdots + \log(x + n) \]

\[ \text{(1) } \varphi(x + n + 1) = \varphi(x) + \log[x(x + 1) \ldots (x + n)] \]
Substituting $\varphi(x)$ into $\varphi(x)$, 

$$\log(n) \leq \frac{\varphi(x) + \log[x(x + 1) \ldots (x + n)] - \varphi(n + 1)}{x}$$ 

$$\leq \log(n + 1)$$

$$x \log(n) \leq \varphi(x) + \log[x(x + 1) \ldots (x + n)] - \varphi(n + 1)$$ 

$$\leq x \log(n + 1)$$

$$0 \leq \varphi(x) + \log[x(x + 1) \ldots (x + n)] - \log(n!)$$ 

$$-x \log(n) \leq x \log \left( \frac{n + 1}{n} \right)$$

$$0 \leq \varphi(x) - \log \left[ \frac{n^x n!}{x(x + 1) \ldots (x + n)} \right] \leq x \log \left( 1 + \frac{1}{n} \right)$$
By squeeze theorem,

\[ \varphi(x) = \lim_{n \to \infty} \log \left[ \frac{n^x n!}{x(x + 1) \ldots (x + n)} \right] \]

\[ \therefore f(x) = \lim_{n \to \infty} \frac{n^x n!}{x(x + 1) \ldots (x + n)} \]

Thus \( f(x) \) is determined by \( 1 \), \( 2 \) and \( 3 \),

\[ \therefore f(x) = \Gamma(x) \]

**Note:** We have established:

\[ \Gamma(x) = \lim_{n \to \infty} \frac{n^x n!}{x(x + 1) \ldots (x + n)} \]
Theorem (8.20)

If $x, y > 0$, then

$$
\int_0^1 t^{x-1}(1-t)^{y-1} \, dt = \frac{\Gamma(x) \Gamma(y)}{\Gamma(x+y)}
$$

Define the beta function by

$$
B(x, y) = \int_0^1 t^{x-1}(1-t)^{y-1} \, dt
$$

Proof

$\forall y > 0$

$$
B(1, y) = \int_0^1 (1-t)^{y-1} \, dt = -\frac{(1-t)^y}{y} \bigg|_0^1 = \frac{1}{y}
$$
Next, prove that as a function of $x$, $\log(B(x, y))$ is convex for each fixed $y$.

Given $p, q$ such that $\frac{1}{p} + \frac{1}{q} = 1$:

$$B\left(\frac{x}{p} + \frac{z}{q}, y\right) = \int_0^1 t^{\frac{x}{p} + \frac{z}{q} - \frac{1}{p} - \frac{1}{q}} \left(1 - t\right)^{\frac{y}{p} + \frac{y}{q} - \frac{1}{p} - \frac{1}{q}} dt$$

$$= \int_0^1 t^{\frac{x-1}{p}} \left(1 - t\right)^{\frac{y-1}{p}} \ t^{\frac{z-1}{q}} \left(1 - t\right)^{\frac{y-1}{q}} dt$$

by Hölder’s inequality

$$\leq \left(\int_0^1 t^{x-1} (1 - t)^{y-1} dt\right)^{\frac{1}{p}} \left(\int_0^1 t^{z-1} (1 - t)^{y-1} dt\right)^{\frac{1}{q}}$$

$$\leq B(x, y)^{\frac{1}{p}} B(z, y)^{\frac{1}{q}}$$
\[
\therefore \log \left( B \left( \frac{x}{p} + \frac{z}{q}, y \right) \right) \leq \frac{1}{p} \log(B(x, y)) + \frac{1}{q} \log(B(z, y))
\]
proving \( \log(B(x, y)) \) is convex as a function of \( x \).

Next, prove

\[
B(x + 1, y) = \frac{x}{x + y} B(x, y) \quad (\forall x, y > 0)
\]

\[
B(x + 1, y) = \int_{0}^{1} t^x(1 - t)^{y-1} \, dt = \int_{0}^{1} \left( \frac{t}{1 - t} \right)^x (1 - t)^{x+y-1} \, dt
\]
\[ \frac{t}{1-t} = -1 + \frac{t}{1-t} \]

Integrating by parts,

\[ u = \left( \frac{t}{1-t} \right)^x \]

\[ dv = (1 - t)^{x+y-1} \, dt \]

\[ du = x \left( \frac{t}{1-t} \right)^{x-1} \left[ \frac{1}{(1-t)^2} \right] \, dt \]

\[ v = -\frac{(1-t)^{x+y}}{x+y} \]
\[ B(x + 1, y) \]
\[
= -\frac{t^x(1 - t)^y}{x + y} \bigg|_0^1 + \int_0^1 \frac{x}{x + y} t^{x-1}(1 - t)^{y+1} \left[ \frac{1}{(1 - t)^2} \right] dt 
\]
\[
= \frac{x}{x + y} B(x, y)
\]

Now, define \( f(x) = \frac{\Gamma(x + y)}{\Gamma(y)} B(x, y) \)
Check:

(b) \[ f(1) = \frac{\Gamma(1 + y)}{\Gamma(y)} B(1, y) \]

\[ = y \cdot \frac{1}{y} = 1 \quad \checkmark \]

(a) \[ f(x + 1) = \frac{\Gamma(x + 1 + y)}{\Gamma(y)} B(x + 1, y) \]

\[ = (x + y) \frac{\Gamma(x + y)}{\Gamma(y)} \frac{x}{x + y} B(x, y) \]

\[ = xf(x) \quad \checkmark \]
(c) Since

\[
\log(f(x)) = \log(\Gamma(x + y)) + \log(B(x, y)) - \log(\Gamma(y))
\]

\log(f(x)) \text{ is convex.}

Therefore, by Theorem 8.19, \( f(x) = \Gamma(x) \).

\[
\Gamma(x) = \frac{\Gamma(x + y)}{\Gamma(y)} B(x, y)
\]

\[
\therefore B(x, y) = \frac{\Gamma(x) \Gamma(y)}{\Gamma(x + y)} \quad \checkmark
\]
Consequences

Making the substitution $t = \sin^2 \theta$, $dt = 2 \sin \theta \cos \theta \, d\theta$ gives

$$B(x, y) = 2 \int_0^{\pi/2} (\sin \theta)^{2x-1} (\cos \theta)^{2y-1} \, d\theta$$

So, this

$$= \frac{\Gamma(x) \Gamma(y)}{\Gamma(x + y)}$$

Plugging in $x = y = \frac{1}{2}$ gives

$$2 \int_0^{\pi/2} d\theta = \frac{\Gamma(\frac{1}{2})^2}{\Gamma(1)}$$
or

\[ \pi = \Gamma \left( \frac{1}{2} \right)^2 \implies \Gamma \left( \frac{1}{2} \right) = \sqrt{\pi} \]

Substituting \( t = s^2 \) in the definition of \( \Gamma \), gives

\[ \Gamma(x) = 2 \int_0^\infty s^{2x-1} e^{-s^2} \, ds \]

Plug \( x = \frac{1}{2} \) \implies

\[ \Gamma \left( \frac{1}{2} \right) = 2 \int_0^\infty e^{-s^2} \, ds = \int_{-\infty}^\infty e^{-s^2} \, ds \]

\[ \therefore \int_{-\infty}^\infty e^{-s^2} \, ds = \sqrt{\pi} \]
Claim \( \forall x \in (0, \infty) \)

\[
\Gamma(x) = \frac{2^{x-1}}{\sqrt{\pi}} \Gamma \left( \frac{x}{2} \right) \Gamma \left( \frac{x + 1}{2} \right)
\]

Proof Define

\[
f(x) = \frac{2^{x-1}}{\sqrt{\pi}} \Gamma \left( \frac{x}{2} \right) \Gamma \left( \frac{x + 1}{2} \right)
\]

Using Theorem 8.19, check:

(a) \( f(1) = \frac{1}{\sqrt{\pi}} \Gamma \left( \frac{1}{2} \right) \Gamma(1) = 1 \) \( \checkmark \)

(b) \( \log(f(x)) = (x - 1) \log 2 - \log(\sqrt{\pi}) + \log \left( \Gamma \left( \frac{x}{2} \right) \right) + \log \left( \Gamma \left( \frac{x + 1}{2} \right) \right) \) \( \text{convex} \) \( \checkmark \)
(c) $f(x + 1) = \frac{2^x}{\sqrt{\pi}} \Gamma \left( \frac{x + 1}{2} \right) \Gamma \left( \frac{x}{2} + 1 \right)$

$$= \frac{2^x}{\sqrt{\pi}} \Gamma \left( \frac{x + 1}{2} \right) \frac{x}{2} \Gamma \left( \frac{x}{2} \right)$$

$$= x \frac{2^{x-1}}{\sqrt{\pi}} \Gamma \left( \frac{x}{2} \right) \Gamma \left( \frac{x + 1}{2} \right)$$

$$= xf(x) \quad \checkmark$$

\[ \therefore \text{ By Theorem 8.19, } f(x) = \Gamma(x) \]
**Stirling’s formula**

\[
\lim_{x \to \infty} \frac{\Gamma(x + 1)}{(x/e)^x \sqrt{2\pi x}} = 1
\]

Sometimes written

\[\Gamma(x + 1) \sim \left(\frac{x}{e}\right)^x \sqrt{2\pi x}\]

I won’t go through the proof of this formula here... you can read it in Rudin!
Please read the first two pages of Chapter 9 on your own, up to 9.4 Definitions. It should be familiar to you from Linear Algebra. Note: in this chapter all vector spaces are assumed to be subspaces of $\mathbb{R}^n$.

**Definition (9.4)**

Let $X, Y$ be vector spaces. A function $A : X \to Y$ is called a linear transformation if

\[ \forall x, y \in X, c \in \mathbb{R} \]

1. $A(x + y) = Ax + Ay$
2. $A(cx) = cA(x)$
Note: Often write $Ax$ instead $A(x)$ if $A$ is linear.

Note that if $A$ is linear, the action of $A$ on $X$ is determined by its action on a basis $\{x_1, \ldots, x_n\}$ of $X$, because if $x \in X$ is arbitrary, then we can write $x = c_1x_1 + \cdots + c_nx_n$, and hence

$$A(x) = c_1A(x_1) + \cdots + c_nA(x_n)$$

A linear transformation $A : X \to X$ is called a linear operator on $X$.

If $A : X \to X$ is bijective (i.e. 1 − 1 and onto), we say $A$ is invertible.

In that case, its inverse $A^{-1} : X \to X$ is defined, is another linear operator on $X$ and $A \circ A^{-1} = A^{-1} \circ A = \text{Id}_X$
Theorem (9.5)

Assume \( X \) is a finite dimensional vector space. Then a linear operator \( A : X \to X \) is 1−1 if and only if it is onto.

Proof

See book. You learned this in Linear Algebra!

Definitions (9.6)

If \( X, Y \) are vector spaces, let \( L(X, Y) = \{ \text{linear transformations } X \to Y \} \).

Write \( L(X) \) for \( L(X, X) \).

Note:
If \( c_1, c_2 \in \mathbb{R} \) and \( A_1, A_2 \in L(X, Y) \), then \( c_1 A_1 + c_2 A_2 \in L(X, Y) \).
If $X, Y$ and $Z$ are vector spaces, $A \in L(X, Y)$ and $B \in L(Y, Z)$, then $B \circ A \in L(X, Z)$.

We often write $BA$ for $B \circ A$.

**Note:**
Even if $X = Y = Z$, in general $AB \neq BA$.

If $A \in L(\mathbb{R}^n, \mathbb{R}^m)$, define the norm of $A$, $\|A\|$, by

$$\|A\| = \sup \{|Ax| : x \in \mathbb{R}^n \text{ and } |x| \leq 1\}$$

It follows easily that $(\forall x \in \mathbb{R}^n)$

$$|Ax| \leq \|A\| |x|.$$ 

Also, if $\lambda \in \mathbb{R}$ and $|Ax| \leq \lambda |x|$ for all $x \in \mathbb{R}^n$, then $\|A\| \leq \lambda$. 
Theorem (9.7)

1. **If** \( A \in L(\mathbb{R}^n, \mathbb{R}^m) \), **then** \( \|A\| \leq \infty \) **and** \( A : \mathbb{R}^n \to \mathbb{R}^m \) **is uniformly continuous.**

2. **If** \( A, B \in L(\mathbb{R}^n, \mathbb{R}^m) \) **and** \( c \in \mathbb{R} \), **then** \( \|A + B\| \leq \|A\| + \|B\| \) **and** \( \|cA\| = |c| \|A\| \).
   *If you define* \( d(A, B) = \|A - B\| \), **this makes** \( L(\mathbb{R}^n, \mathbb{R}^m) \) **into a metric space.**

3. **If** \( A \in L(\mathbb{R}^n, \mathbb{R}^m) \) **and** \( B \in L(\mathbb{R}^m, \mathbb{R}^k) \), **then** \( \|BA\| \leq \|B\| \|A\| \).
Proof

1: Let \( \{e_1, \ldots, e_n\} \) be the standard basis for \( \mathbb{R}^n \) and suppose \( x = c_1e_1 + \cdots + c_ne_n \) and \( |x| \leq 1 \).

It follows that \( |c_i| \leq 1 \) (\( \forall i = 1, \ldots, n \)).

Then

\[
|Ax| = \left| \sum c_i Ae_i \right| \leq \sum |c_i| |Ae_i| \leq \sum |Ae_i|
\]

Hence

\[
\|A\| \leq \sum_{i=1}^{n} |Ae_i| < \infty
\]

To prove \( A \) uniformly continuous, let \( \varepsilon > 0 \) and let \( \delta = \frac{\varepsilon}{\|A\|} \).

Then if \( |x - y| < \delta \), \( |Ax - Ay| \leq \|A\| |x - y| < \varepsilon \).

(Case where \( \|A\| = 0 \) is trivial, since in that case \( A = 0 \).)
②: The inequality in ② follows from:

\[(A + B)x = |Ax + Bx| \leq |Ax| + |Bx| \leq (\|A\| + \|B\|) |x|\]

Proof of \(\|cA\| = |c| \|A\|\) is similar.

Skip proof that \(L(\mathbb{R}^n, \mathbb{R}^m)\) is a metric space - it’s easy!

③ follows from:

\[(BA)x = |B(Ax)| \leq \|B\| |Ax| \leq \|B\| \|A\| |x|\]

We use the metric space structure on \(L(\mathbb{R}^n, \mathbb{R}^m)\) in the following.
Theorem (9.8)

Let $\Omega = \{ A \in L(X) \text{ such that } A \text{ is invertible} \}.$

\begin{enumerate}
\item If $A \in \Omega$, $B \in L(\mathbb{R}^n)$ and 
\[ \| B - A \| \| A^{-1} \| < 1 \]
then $B \in \Omega$
\item $\Omega$ is an open subset of $L(\mathbb{R}^n)$ and the function $\Omega \rightarrow \Omega$ given by $A \rightarrow A^{-1}$ is continuous on $\Omega$.
\end{enumerate}
Proof

1: Let $\alpha = \frac{1}{\|A^{-1}\|}$, so $\|A^{-1}\| = \frac{1}{\alpha}$,

and let $\beta = \|B - A\|$. Since $\beta \cdot \frac{1}{\alpha} < 1$, $\beta < \alpha$.

For every $x \in \mathbb{R}^n$,

$$\alpha|\mathbf{x}| = \alpha|A^{-1}A\mathbf{x}| \leq \alpha\|A^{-1}\| |A\mathbf{x}|$$

$$= |A\mathbf{x}| \leq |(A - B)\mathbf{x}| + |B\mathbf{x}| \leq \beta|\mathbf{x}| + |B\mathbf{x}|$$

Hence,

$$(\alpha - \beta)|\mathbf{x}| \leq |B\mathbf{x}| \quad (\forall \mathbf{x} \in \mathbb{R}^n) \quad (\star)$$
Since $\alpha - \beta > 0$, this $\implies B\mathbf{x} \neq 0$ if $\mathbf{x} \neq 0$.

Therefore $B$ is injective and, by Theorem 9.5, bijective.

Hence

$$B \in \Omega$$

(2) This also proves

$$N_{\frac{1}{\|A^{-1}\|}}(A) \subseteq \Omega$$

so $\Omega$ is open in $L(\mathbb{R}^n)$.

Just need to prove the function $A \to A^{-1}$ is continuous.

In $(\star)$ replace $\mathbf{x}$ by $B^{-1}\mathbf{y}$, giving

$$(\forall \mathbf{y} \in \mathbb{R}^n) \quad (\alpha - \beta)|B^{-1}\mathbf{y}| \leq |BB^{-1}\mathbf{y}| = |\mathbf{y}|$$
This proves

$$
\|B^{-1}\| \leq \frac{1}{\alpha - \beta}
$$

Now, write $B^{-1} - A^{-1} = B^{-1}(A - B)A^{-1}$ and apply 9.7(3):

$$
\|B^{-1} - A^{-1}\| \leq \|B^{-1}\| \|A - B\| \|A^{-1}\| \leq \frac{\beta}{(\alpha - \beta)\alpha}
$$

Since $\beta = \|B - A\|$, we can make $B^{-1}$ as closed as desired to $A^{-1}$ by making $B$ sufficiently close to $A$ (i.e. $\beta$ close to 0).

This establishes continuity.
Matrices

Suppose $X$ and $Y$ are vector spaces and \( \{x_1, \ldots, x_n\} \) is a basis of $X$ and \( \{y_1, \ldots, y_m\} \) is a basis of $Y$. Let \( A \in L(X, Y) \). For each \( j = 1, \ldots, n \) we can write

\[
A x_j = \sum_{i=1}^{m} a_{ij} y_i
\]

where \( \forall \ 1 \leq j \leq n, \forall \ 1 \leq i \leq m, \ a_{ij} \in \mathbb{R} \).

These \( \{a_{ij}\} \) form an \( m \times n \) matrix

\[
[A] = \begin{bmatrix}
a_{11} & a_{12} & \ldots & a_{1n} \\
a_{21} & a_{22} & \ldots & a_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
a_{m1} & a_{m2} & \ldots & a_{mn}
\end{bmatrix}
\]
$[A]$ is called the **matrix of $A$** with respect to the given bases $\{x_j\}$ and $\{y_i\}$.

*Note:* The $j$-th column of $[A]$ gives the coefficients of $A x_j$ with respect to the basis $\{y_i\}$.

Suppose $x \in X$ is arbitrary; write

$$x = \sum_{j=1}^{n} c_j x_j$$
Then

\[ A\mathbf{x} = \sum_{j=1}^{n} c_{j}A\mathbf{x}_{j} = \sum_{j=1}^{n} c_{j} \sum_{i=1}^{m} a_{ij} \mathbf{y}_{i} = \sum_{i=1}^{m} \left( \sum_{j=1}^{n} a_{ij} c_{j} \right) \mathbf{y}_{i} \]

Thus the coordinates of \( A\mathbf{x} \) (with respect to \( \{ \mathbf{y}_{i} \} \)) are

\[ \sum_{j=1}^{n} a_{ij} c_{j} \]
To sum up: If we express $\mathbf{x} \in X$ as a column vector
\[
\begin{bmatrix}
c_1 \\
\vdots \\
c_n
\end{bmatrix}
\]
with respect to the basis $\{\mathbf{x}_j\}$, then matrix multiplication
\[
[A]
\begin{bmatrix}
c_1 \\
\vdots \\
c_n
\end{bmatrix}
\]
gives us $A\mathbf{x}$, expressed as a column vector with respect to the basis $\{\mathbf{y}_i\}$ of $Y$.

This correspondence $A \leftrightarrow [A]$ gives a 1–1 correspondence between
\[
L(X, Y) \leftrightarrow \{m \times n \text{ matrices over } \mathbb{R}\}\]
Don’t forget: the matrix \([A]\) depends not only on \(A \in L(X, Y)\), but also on our choice of bases for \(X\) and \(Y\).

Now, suppose we have a third vector space \(Z\) with basis \(\{z_1, \ldots, z_p\}\) and linear transformations

\[
A : X \rightarrow Y \quad B : Y \rightarrow Z
\]

Then \(B \circ A\) (written \(BA\)) has matrix

\[
[B \circ A] = [B][A]
\]

(\(\Leftarrow\) matrix multiplication)

with respect to the given bases.
Suppose \( \{x_1, \ldots, x_n\} \) and \( \{y_1, \ldots, y_m\} \) are the standard bases of \( \mathbb{R}^n \) and \( \mathbb{R}^m \).

Assume \( A \in L(\mathbb{R}^n, \mathbb{R}^m) \) has matrix \([A] = (a_{ij})\) with respect to these bases.

Suppose \( x = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} \in \mathbb{R}^n \).
Then

$$|Ax|^2 = \sum_i \left( \sum_j a_{ij}c_j \right)^2$$

(Schwarz inequality) \[ \leq \sum_i \left[ \left( \sum_j a_{ij}^2 \right) \cdot \left( \sum_j c_j^2 \right) \right] \]

$$= \left( \sum_{i,j} a_{ij}^2 \right) \|x\|^2$$

Hence

$$\|A\| \leq \left( \sum_{i,j} a_{ij}^2 \right)^{\frac{1}{2}}$$ \((\star)\)
Theorem

Suppose $S$ is a metric space and $\forall \ i = 1, \ldots, m, \ \forall \ j = 1, \ldots, n, \ a_{ij} : S \to \mathbb{R}$ is continuous. For each $p \in S$, define $A_p \in L(\mathbb{R}^n, \mathbb{R}^m)$ to be the linear transformation whose matrix is $(a_{ij}(p))$. Then the function

$$S \to L(\mathbb{R}^n, \mathbb{R}^m)$$

defined by

$$p \mapsto A_p$$

is continuous. (Using the metric space structure on $L(\mathbb{R}^n, \mathbb{R}^m)$, recently introduced.)

Proof.

This follows easily by applying $(\star)$ to the difference $B - A$. Omit details.
Goal: Define the derivative of a function

\[ f : \mathbb{R}^n \to \mathbb{R}^m. \]

What kind of object should it be?

Recall case \( f : \mathbb{R} \to \mathbb{R} \). We defined

\[ f'(x) = \lim_{h \to 0} \frac{f(x + h) - f(x)}{h} \]

Rewrite as

\[ \lim_{h \to 0} \frac{f(x + h) - f(x) - f'(x)h}{h} = 0 \]
Thus, if we define \( r(h) = f(x + h) - f(x) - f'(x)h \), we see

\[
\lim_{h \to 0} \frac{r(h)}{h} = 0
\]

So \( r(0) = 0 \) and \( r(h) \to 0 \) as \( h \to 0 \) “very quickly”.

**Sum up:** The linear transformation

\[ \mathbb{R} \to \mathbb{R} \]

given by

\[ h \mapsto f'(x)h \]

is a very good approximation of the non-linear map \( h \mapsto f(x + h) - f(x) \) in the sense that the error term \( r(h) = f(x + h) - f(x) - f'(x)h \) has the property that

\[
\lim_{h \to 0} \frac{r(h)}{h} = 0.
\]
**Note:** This is equivalent to

\[
\lim_{h \to 0} \frac{|r(h)|}{|h|} = 0
\]

**Moral of the Story:** Instead of thinking of \( f'(x) \) simply as a number, we think of it as a linear transformation

\[
\mathbb{R} \to \mathbb{R}
\]

given by

\[
h \mapsto f'(x)h
\]

**Generalize:**
Definition

Let $E \subseteq \mathbb{R}^n$ be open, $x \in E$, and assume $f : E \to \mathbb{R}^m$ is a function. If there exists a linear transformation $A \in L(\mathbb{R}^n, \mathbb{R}^m)$ such that

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

(\star 1)

then we say $f$ is differentiable at $x$ and

$$f'(x) = A$$

(Note: in the above limit, $h \in \mathbb{R}^n$)

If $f$ is differentiable at each $x \in E$, we say $f$ is differentiable on $E$.

Repeat: For each given $x \in E$, $f'(x)$ is an element of $L(\mathbb{R}^n, \mathbb{R}^m)$. This element changes as $x$ changes.
Theorem

\[ f \text{ has at most one derivative at } \mathbf{x} \]

I.e., if \( A_1, A_2 \in L(\mathbb{R}^n, \mathbb{R}^m) \) and both \( A_1 \) and \( A_2 \) satisfy \((\star 1)\), then \( A_1 = A_2 \).

Proof

Let \( B = A_1 - A_2 \). Then, \( \forall \mathbf{h} \in \mathbb{R}^n \),

\[
|B\mathbf{h}| = |(f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) - A_2\mathbf{h}) - (f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) - A_1\mathbf{h})| \\
\leq |f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) - A_2\mathbf{h}| + |f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) - A_1\mathbf{h}| \\
\therefore \frac{|B\mathbf{h}|}{|\mathbf{h}|} \to 0 \text{ as } \mathbf{h} \to 0
\]
For fixed $h \in \mathbb{R}^n$, this \[
\lim_{t \to 0} \frac{|B(th)|}{|th|} = 0
\]

But \[
\frac{|B(th)|}{|th|} = \frac{|tB(h)|}{|th|} = \frac{|t| |B(h)|}{|t||h|} = \frac{|B(h)|}{|h|}
\]

which doesn’t depend on $t$!

\[
\therefore B(h) = 0 \quad \forall h \in \mathbb{R}^n
\]
Recall definition of derivative

Let \( E \subseteq \mathbb{R}^n \) be open, and \( f : E \to \mathbb{R}^m \). Fix \( x \in E \). If there is a linear transformation \( A \in L(\mathbb{R}^n, \mathbb{R}^m) \) such that

\[
\lim_{h \to 0} \frac{|f(x + h) - f(x) - Ah|}{|h|} = 0 \quad (\star)
\]

we write \( f'(x) = A \), and say “\( f \) is differentiable at \( x \)”.

We proved last time that, given \( f \) and \( x \), there is at most one such \( A \). Hence, \( f'(x) \) is well-defined.
Remarks

1. (⋆) can be written as
   \[ f(x + h) - f(x) = f'(x)h + r(h) \]
   where
   \[
   \lim_{h \to 0} \frac{|r(h)|}{|h|} = 0
   \]
   i.e. “\( f(x + h) - f(x) \) is very well approximated by \( f'(x)h \)”.

2. If \( f \) is differentiable on \( E \), then \( f': E \to L(\mathbb{R}^n, \mathbb{R}^m) \) is a function.

3. The formula in 1⃝ shows that if \( f \) is differentiable at \( x \), then \( f \) is continuous at \( x \).

4. “\( f'(x) \)” is sometimes called the differential of \( f \) at \( x \) or the total derivative of \( f \) at \( x \).
Ex.
Suppose $A : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation.

Then

$$
\lim_{h \to 0} \frac{|A(x + h) - A(x) - A(h)|}{|h|} = \lim_{h \to 0} \frac{|Ax + Ah - Ax - Ah|}{|h|}
$$

$$
= 0
$$

\[\therefore\] For all $x \in \mathbb{R}^n$, $A'(x) = A$
Theorem (9.15 - Chain Rule)

Suppose $E \subseteq \mathbb{R}^n$ is open, $f : E \rightarrow \mathbb{R}^m$, and $g$ maps a neighborhood of $f(E)$ (in $\mathbb{R}^m$) to $\mathbb{R}^k$.

Also, assume $f$ is differentiable at $x_0 \in E$ and $g$ is differentiable at $f(x_0)$.

Define $F : E \rightarrow \mathbb{R}^k$ by $F = g \circ f$.

Then $F$ is differentiable at $x_0$ and

$$F'(x_0) = g'(f(x_0))f'(x_0)$$

Note $g'(f(x_0)) \in L(\mathbb{R}^m, \mathbb{R}^k)$, and $f'(x_0) \in L(\mathbb{R}^n, \mathbb{R}^m)$.

Proof.

Omit. Very similar to proof of ordinary Chain Rule.
Partial Derivatives

Recall from Calculus III:
Suppose $f : \mathbb{R}^n \rightarrow \mathbb{R}$. The partial derivatives of $f$ are defined as

$$\frac{\partial f}{\partial x_j}(a_1, \ldots, a_n) = \lim_{h \to 0} \frac{f(a_1, \ldots, a_j + h, \ldots, a_n) - f(a_1, \ldots, a_n)}{h}$$

“The derivative of $f(x_1, \ldots, x_n)$ with respect to $x_j$, holding the other variables constant”.

If $a = (a_1, \ldots, a_n)$ and $e_j = (0, \ldots, 0, 1, 0, 0)$

We can also write

$$\frac{\partial f}{\partial x_j}(a) = \lim_{h \to 0} \frac{f(a + he_j) - f(a)}{h}$$
Suppose \( E \subseteq \mathbb{R}^n \) is open, and \( f : E \to \mathbb{R}^m \). Let \( \{e_1, \ldots, e_n\} \) and \( \{u_1, \ldots, u_m\} \) be standard bases of \( \mathbb{R}^n \) and \( \mathbb{R}^m \).

The components of \( f \) are functions \( f_1, \ldots, f_m \) (each \( f_i : \mathbb{R}^n \to \mathbb{R} \)) defined by

\[
f(x) = (f_1(x), f_2(x), \ldots, f_m(x))
\]

Equivalently, \( f_i(x) = f(x) \cdot u_i \)

If \( x \in E, 1 \leq i \leq m, 1 \leq j \leq n \), we can take partial derivative

\[
\frac{\partial f_i}{\partial x_j}(x) = \lim_{t \to 0} \frac{f_i(x + te_j) - f_i(x)}{t}
\]

Other notation:

\[
\frac{\partial f_i}{\partial x_j}(x) = (D_{j}f_i)(x)
\]
Theorem (9.17)

Suppose $E \subseteq \mathbb{R}^n$ is open, and $f : E \to \mathbb{R}^m$ is differentiable at $x \in E$. Then the partial derivatives $(D_j f_i)$ all exist at $x$ and

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

$$1 \leq j \leq n$$

In other words, with respect to the standard bases

$$[f'(x)] = \begin{bmatrix} (D_1 f_1)(x) & \ldots & (D_n f_1)(x) \\ \vdots & \ddots & \vdots \\ (D_1 f_m)(x) & \ldots & (D_n f_m)(x) \end{bmatrix}$$
Thus, the partial derivatives of the components of $f$ are the entries of the matrix for $f'(x)$ with respect to the standard bases.

**Proof**

By definition of $f'(x)$,

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

$$\therefore f(x + te_j) - f(x) = f'(x)(te_j) + r(te_j)$$

where $\frac{|r(te_j)|}{t} \to 0$ as $t \to 0$ (since $|te_j| = |t|$).
\[ \frac{f(x + te_j) - f(x)}{t} - \frac{f'(x)(te_j)}{t} \rightarrow 0 \text{ as } t \rightarrow 0 \]

\[ f(x + te_j) - f(x) \rightarrow f'(x)(e_j) \text{ as } t \rightarrow 0 \]

\[ \lim_{t \to 0} \frac{f(x + te_j) - f(x)}{t} = f'(x)(e_j) \quad (\star) \]

The \( i \)-th entry of the vector on the RHS of \((\star)\) is the \( ij \)-th entry of the matrix \([f'(x)]\).

The \( i \)-th entry of the vector on the LHS of \((\star)\) is

\[ \lim_{t \to 0} \frac{f_i(x + te_j) - f_i(x)}{t} = (D_jf_i)(x) \]
\[ x_0 \in \mathbb{R}^n; \{e_1, \ldots, e_n\} \text{ standard basis for } \mathbb{R}^n \text{ (column vectors)}; \]
\[ f : \mathbb{R}^n \to \mathbb{R} \text{ differentiable}; \ u \in \mathbb{R}^n, \text{ with } |u| = 1. \]

Define the directional derivative of \( f \) at \( x_0 \) in the direction \( u \) by

\[
D_u f(x_0) = \left. \frac{d}{dt} \right|_{t=0} f(x_0 + tu)
\]

Interpretation: Define \( \gamma : \mathbb{R} \to \mathbb{R}^n \) by \( \gamma(t) = x_0 + tu \).
\( \gamma \) is a parametrized straight line in \( \mathbb{R}^n \) with a constant velocity \( \gamma'(t) = u \) passing through \( x_0 \) at \( t = 0 \).

\[
\mathbb{R} \xrightarrow{\gamma} \mathbb{R}^n \xrightarrow{f} \mathbb{R}
\]
\[ D_u f(x_0) = (f \circ \gamma)'(0) = (\text{by Chain Rule}) \quad f'(\gamma(0))\gamma'(0) = f'(x_0)u \]

Note: \( u = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} \) and \( f'(x_0) = [D_1 f(x_0), \ldots, D_n f(x_0)] \)

\[ \therefore \quad D_u f(x_0) = D_1 f(x_0)u_1 + \cdots + D_n f(x_0)u_n \]
If we define the gradient of $f$ at $x_0$ \( \nabla f(x_0) \in \mathbb{R}^n \) by

\[
\nabla f(x_0) = D_1 f(x_0) e_1 + \cdots + D_n f(x_0) e_n
\]

then we have

\[
D_u f(x_0) = \nabla f(x_0) \cdot u
\]

By Schwarz inequality, the value of $u$ giving largest $D_u f(x_0)$ is

\[
u = \frac{\nabla f(x_0)}{|\nabla f(x_0)|}
\]

So we say “\( \nabla f(x_0) \) points in the direction of most rapid increase of $f$.”
For this value of $u$,

$$D_u f(x_0) = \nabla f(x_0) \cdot \frac{\nabla f(x_0)}{|\nabla f(x_0)|} = |\nabla f(x_0)|$$

“$\nabla f(x_0)$ points in the direction of most rapid increase of $f$ starting at $x_0$, and $|\nabla f(x_0)|$ gives the magnitude of this most rapid increase.”
The Inverse Function Theorem

Suppose \( E \subseteq \mathbb{R}^n \) is open, \( f : E \rightarrow \mathbb{R}^m \), and \( f \) is differentiable on \( E \). Then

\[
f' : E \rightarrow L(\mathbb{R}^n, \mathbb{R}^m).
\]

If \( f' \) is continuous, we say \( f \) is continuously differentiable or write

\[
f \in C'(E).
\]

(More modern notation: \( f \in C^1(E) \))

**Theorem (9.21)**

\[
f \in C'(E) \iff \text{all the partial derivatives } D_j f_i \text{ are defined and continuous on } E.
\]

**Proof.**

Omit
Theorem (9.24 - Inverse Function Theorem)

Suppose $E \subseteq \mathbb{R}^n$ is open, $f : E \to \mathbb{R}^n$, and $f \in C'(E)$.
Suppose $a \in E$, $f'(a) : \mathbb{R}^n \to \mathbb{R}^n$ is invertible, and $b = f(a)$.
Then

1. ∃ open sets $U, V \in \mathbb{R}^n$ such that $a \in U$, $b \in V$, $f$ is injective on $U$, and $f(U) = V$

2. If $g : V \to U$ is the inverse of $f$ (which exists by (1)), then $g \in C'(V)$ and $g'(b) = f'(a)^{-1}$

“For a $C'$ function, invertibility of $f'(a)$ at a single point $a \implies$ invertibility of $f$ on a neighborhood of $a$.”
Thank you!

Questions?