Physics 171 - Mathematical Methods in Physics

0.1 Lecture 1

In this lecture I discuss definitions and notation that will be used in the rest of the course.

A set is a collection of objects, called members of the set. Sets are denoted by capital letters $A, B, C, \cdots$

Some special sets appear often and have their own special notation.

$\mathbb{N}$:= the set of natural numbers, $\{0, 1, 2, 3 \cdots\}$.

$\mathbb{Z}$:= the set of integers, $\{\cdots - 2, -1, 0, 1, 2, 3 \cdots\}$.

$\mathbb{Q}$:= the set of rational numbers, $x = a/b$ where $a, b \in \mathbb{Z}$ with $b \neq 0$.

$\mathbb{R}$:= the set of real numbers.

$\mathbb{C}$:= the set of complex numbers.

$\{\emptyset\}$:= the empty set (set with no elements).

I will sometimes use the following mathematical abbreviations:

$\in$: $a \in A \equiv a$ is an member of the set A.

$\forall$: $\equiv$ "for all".

$\exists$: $\equiv$ "there exists"

$\Rightarrow$: $\equiv$ "implies"

$\iff$: $\equiv$ "if and only if"

$\ni$: $\equiv$ "such that"

Two sets A and B are equal if they have exactly the same elements. In the above notation this can be expressed as:

$$A = B \quad \iff \quad a \in A \Rightarrow a \in B \quad \textit{and} \quad b \in B \Rightarrow b \in A.$$

Set A is a subset of set B, written $A \subset B$ or $B \supset A$ if $a \in A \Rightarrow a \in B$. This means that every element of A is also an element of B.
Example: Let $E \equiv$ even natural numbers. Then

$$E \subset \mathbb{N}$$

$$\{0, 2, 4, 6 \cdots\} \subset \{0, 1, 2, 3, 4, \cdots\}$$

The sum or union of two sets A and B

$$A + B \qquad \text{or} \qquad A \cup B$$

is the set whose members are members of A or B:

$$A = \{1, 2, 5\} \qquad B = \{2, 3, 4\}$$

$$A + B = A \cup B = \{1, 2, 3, 4, 5\}.$$

The intersection of two sets A and B is the set

$$A \cap B$$

whose members are members of both A and B:

$$A = \{1, 2, 5\} \qquad B = \{2, 3, 4\}$$

$$A \cap B = \{2\}$$

The difference $A - B$ of two sets A and B is the set whose members are members of A and not members of B

$$A = \{1, 2, 5\} \qquad B = \{2, 3, 4\}$$

$$A - B = \{1, 5\}$$

There are standard rules for manipulating sets. These are usually expressed in terms of axioms. It turns out that some naive axioms of set theory are inconsistent. The Zermelo-Frankel axioms are an acceptable set of axioms. They are not intuitive, but they are designed to allow standard manipulations of sets without allowing the existence of certain "bad" sets.

Example: Show

$$D = A \cup (B \cap C) = (A \cup B)) \cap (A \cup C) = E \tag{1}$$

To show this first note that $x \in D$ means $x \in A$ or x is a member of both B and C. If $x \in A$ then $x \in (A \cup B)$ and $x \in (A \cup C)$. If $x \in (B \cap C)$ then $x \in (A \cup B)$ and $x \in (A \cup C)$. In either case $x \in (A \cup B)) \cap (A \cup C)$. This shows that $D \subset E$. To show $E \subset D$ note that $x \in E$ means that x is a member of both $(A \cup B)$ and $(A \cup C)$. If $x \in A$ then $x \in A \cup (B \cap C)$. If $x \notin A$, then it must be in both B and C or $B \cap C$. Combining these cases we see $x \in E \Rightarrow x \in D$. Thus $E = D$.

In this course we will deal with sets of three different sizes.

1. Finite sets.

2. Countably infinite sets.

3. Uncountably infinite sets.

Two sets, A and B, are the same size if there is a 1 to 1 correspondence between the members of A and the members of B.

A set has N members if the members can be put into a 1 to 1 correspondence with the first N natural numbers, $\{0, 1, 2 \cdots N-1\}$. Sets with N members are finite sets. Countably infinite sets are sets whose members can be put into a $1-1$ correspondence with the natural numbers $\mathbb{N}$. Infinite sets that are not countable are uncountable. Uncountable sets also come in different sizes, however we will not need to concern ourselves with the ordering of different uncountable sets.

The real numbers $\mathbb{R}$ are uncountable while the rational numbers $\mathbb{Q}$ are countable.

It might seem odd that the set of irrational numbers is much larger than the set of rational numbers since between every pair of irrational numbers there is a rational number.

Later we will show that the volume occupied by the irrational numbers between zero and one is one, while the volume occupied by the rational numbers between zero and one is zero. For this lecture I give a standard proof that the rational numbers are countable.

To show this it is necessary to construct a $1-1$ correspondence between the elements of $\mathbb{Q}$ and $\mathbb{N}$. To construct the desired correspondence note that any rational number can be expressed in the form

$$x \in \mathbb{Q} \Rightarrow x = (-1)^l \frac{n}{m} \qquad n, m \in \mathbb{N}, \quad l \in \{0, 1\} \tag{2}$$

where $m \neq 0$. This correspondence is unique if m and n have no common divisors except 1 and if $n = 0$ we take $m = 1$ and $l = 0$.

Define the mapping f from $\mathbb{Q}$ to a subset S of $\mathbb{N}$ by

$$f((-)^l \frac{n}{m}) := 2^l \cdot 3^n \cdot 5^m \tag{3}$$

The members of the set S, $2^l \cdot 3^n \cdot 5^m$ are all members of $\mathbb{N}$ and can be ordered according to their size. A given rational number x is identified with $s_k = f(x)$ where s_k is the k^{th} element of S. The desired $1-1$ correspondence is $k \leftrightarrow x$. This shows that the rational numbers are countable.

0.2 Lecture 2

In working with real, complex numbers, and vectors we will have to work with sets of points one a line, in a plane, or in a finite dimensional space. In each of these cases there is a natural definition of the distance between the points

$$\rho(a,b) := \text{distance between } a \text{ and } b$$

Different distance functions will be introduced later, however all of the distance functions that we introduce are real valued functions with the following three properties:

$$\rho(a,b) \geq 0 \tag{4}$$

$$\rho(a,b) = 0 \iff a = b \tag{5}$$

and for any c

$$\rho(a,b) \leq \rho(a,c) + \rho(c,b) \tag{6}$$

Distance functions with these three properties are called metrics. Sets with metric functions defined on the pairs of members are called metric spaces.

A standard metric for the real numbers is the absolute value of the difference of the numbers

$$\rho(x_1,x_2) = |x_1 - x_2|. \tag{7}$$

We can use the metric function to define a number of special sets. The abstract characterization of the metric function will allow us to extend familiar concepts, like continuity and convergence, involving real numbers to complex numbers and vector spaces.

A neighborhood of a point $p \in S$ is the set of points of the form

$$N_{p,R} := \{p' \in S | \rho(p,p') < R\} \tag{8}$$

where R is a positive real number.

A point p of a set $S' \subset S$ of is isolated if there is a neighborhood of p containing no other element of S'. For example the integer 4 is an isolated member of the integers.

A point p accumulation point of the subset $S' \subset S$ if *every* neighborhood of p in S contains a member of S'.

A point p interior point of the subset $S' \subset S$ there is a neighborhood of p in S such that all elements of this neighborhood are elements of S'.

A set is open if all of its points are interior points.

A set is closed if it contains all of its accumulation points.
Example:

$$(a, b) := \{x \in \mathbb{R} | a < x < b\} \quad \text{is open}$$

$$[a, b] := \{x \in \mathbb{R} | a < x < b\} \quad \text{is closed}$$

A region is an open set with the property that any two points can be connected by a continuous line in the set.
Example:

$$(a, b) \quad \text{is an open region}$$

$$(a, b) \cup (c, d), \qquad c > b, \qquad \text{is open, it is not a region.}$$

We say $F(a) = b$ from metric space A to metric space B is continuous at a if for every $\epsilon > 0$ there exists a $R > 0$ such that for every $a' \in N_{a,R}$, $\rho_B(b, F(a')) < \epsilon$.

We say $F(a) = b$ from metric space A to metric space B is uniformly continuous on $C \subset A$ if for every $\epsilon > 0$ and every $a \in C$ exists a $R >$ (independent of $a \in C$) such that for every $a' \in N_{a,R}$, $\rho_B(F(a), F(a')) < \epsilon$.

Note that in formulating this definition we only used the metric functions on A and B (the metric function on A was used to define the region $N_{a,R}$. You should convince yourselves that this reduces to the standard definition that you already know when it is applied to real valued functions of a real variable.

In mathematics there are a number of ways of proving results. Understanding how this is done helps to understand when certain results can be extended or not. In this class we will use three methods to prove results:

1. Proof by construction: In this method the existence of an object satisfying certain mathematical properties is established by an explicit construction of the desired object.

   Example: The statement that there exists a prime number greater than 11 can be established by showing that 13 is prime and that it is greater than 11.

   Example: In physics quantum field theory can be characterized by axioms that define the expected physically sensible properties of quantum fields:

   A1: The theory is a quantum theory.

A2: The theory consistent with special relativity.

A3: The energy of the system is bounded from below.

A4: The theory can describe particles.

A5: Microscopic causality

It turns out that the non-trivial examples of theories satisfying these axioms are difficult to construct. One can however prove the *consistency* of these axioms by showing that free field theories are an example of a (trivial) theory that is consistent with these axioms. This is an example of proof by construction.

2. Proof by contradiction: To prove something by contradiction assume that the desired result is false. Use that assumption to establish a contradiction. This then implies that the starting assumption is incorrect.

   Example: (Aristotle) To prove $\sqrt{2}$ is irrational assume by contradiction that it is rational. Then we can write

$$2 = \frac{m}{n}$$

   where m and n are assumed to have no common divisors. Squaring this equation gives

$$2n^2 = m^2.$$

   Since m^2 is even (by above) and the square of any odd number is odd, it follows that m is even. This means that m^2 is divisible by 4. If this is true then n^2 is divisible by 2. This means that n cannot be odd, so it must be even. This contradicts the assumption that m and n have no common divisors. Thus the assumption that $\sqrt{2}$ must be false. This establishes that $\sqrt{2}$ is irrational.

3. Proof by induction: When proving results that apply to an infinite or large number of results, the next best thing to proof by construction is proof by induction. In this case we label the desired results $R_1 \cdots R_n \cdots$. We begin by establishing that R_1 is true by construction or contradiction. Then we show by construction or contradiction that if R_n is true for $n < m$ then R_m is true. The principle of mathematical induction then implies that R_k is true for any $k \geq 1$.

Example: For real numbers x it is easy to show $|x_1 x_2| = |x_1||x_2|$ by going through all of four possible sign combinations for the two numbers. If we assume (out induction assumption) $|x_1 \cdots x_{n-1}| = |x_1| \cdots |x_{n-1}|$ then

$$|x_1 \cdots x_n| = |x_1 \cdots x_{n-1}||x_n| = |x_1| \cdots |x_n| \tag{9}$$

- which shows that if the result holds for $N - 1$ factors it also holds for N factors.

Example: Cluster properties in relativistic quantum mechanics is the statement that if a relativistically invariant system is divided into isolated subsystems, then each isolated subsystem is relativistically invariant.

It turns out it is easy to establish the result for a system of two particles. One can then show that if the result holds for $M < N$ particles it also holds for N particles. This is a problem that is central to my own research program.

Complex numbers

Consider the polynomial equation

$$P(z) = z^2 + 1 = 0. \tag{10}$$

This equation has no solutions when z is restricted to be a real number. Complex numbers are an extension of the real numbers so that solutions of polynomial equations have roots.

In order to construct the desired extension it is enough to introduce the new number

$$i = \sqrt{-1}. \tag{11}$$

This notation was introduced by Euler in 1779. With the introduction of i equation (**??**) can be factored

$$P(z) = (z + i)(z - i) \tag{12}$$

with roots

$$z = \pm i. \tag{13}$$

A general complex number z has the form

$$x = x + iy \tag{14}$$

where $x, y \in \mathbb{R}$. The number x is called the real part of z, denoted by

$$x = \Re(z). \tag{15}$$

The number y is called the imaginary part of z, denoted by

$$y = \Im(z). \tag{16}$$

A complex number $z = x + iy$ is real if $y = 0$. It is imaginary if $x = 0$. It is zero if $x = y = 0$.

Complex numbers can be added, subtracted, and multiplied:

$$z_1 + z_2 = (x_1 + iy_1) + (x_2 + iy_2) = (x_1 + x_2) + i(y_1 + y_2) \tag{17}$$

$$z_1 - z_2 = (x_1 + iy_1) - (x_2 + iy_2) = (x_1 - x_2)i(y_1 - y_2) \tag{18}$$

$$z_1 z_2 = (x_1 + iy_1)(x_2 + iy_2) = x_1 x_2 + i(x_1 y_2 + y_2 x_1) + i^2 y_1 y_2 =$$
$$(x_1 x_2 - y_1 y_2) + i(x_1 y_2 + y_2 x_1). \tag{19}$$

0.3 Lecture 3

If a complex number $z \neq 0$ then it has an multiplicative inverse

$$\frac{1}{z} = \frac{1}{x+iy}\frac{x-iy}{x-iy} = \frac{x}{x^2+y^2} - i\frac{y}{x^2+y^2} \tag{20}$$

The complex conjugate z^* of a complex number $z = x + iy$ is

$$z^* = (x+iy)^* = x - iy \tag{21}$$

It follows that

$$i^* = -i \tag{22}$$

A direct calculation shows that for any complex number $z = x + iy$ that

$$zz^* = (x+iy)(x-iy) = x^2 + y^2 \geq 0 \tag{23}$$

The real number

$$|z| := \sqrt{zz^*} = \sqrt{x^2+y^2} \tag{24}$$

is called the modulus of the complex number z.

Let $z = x + iy$ and write

$$z = |z|(\frac{x}{|z|} + i\frac{y}{|z|}) \tag{25}$$

This can be written

$$z = |z|(\cos(\phi) + i\sin(\phi)) \tag{26}$$

where

$$\sin(\phi) = \frac{x}{|z|} \qquad \cos(\phi) = \frac{y}{|z|}. \tag{27}$$

The angle ϕ is the argument of z The argument of a complex number is only defined up to an integer multiple of 2π.

Any complex number can be expressed in terms of its real and imaginary part or in terms of its modulus and argument. It can be represented by a vector in the complex plane with x coordinate x and y coordinate y. The modulus and phase are like a representation of the same vector in terms of polar coordinates. The modulus is the length of the vector and the argument is the angle that the vector makes with the x axis.

Note that

$$z_1 z_2 = |z_1|(\cos(\phi_1) + i\sin(\phi_1))|z_2|(\cos(\phi_2) + i\sin(\phi_2)) =$$

$$|z_1||z_2|[(\cos(\phi_1)\cos(\phi_2) - \sin(\phi_1)\sin(\phi_2)) + i(\cos(\phi_1)\sin(\phi_2) + \sin(\phi_1)\cos(\phi_2))] \tag{28}$$

Using the trigonometric identities

$$\cos(\phi_1 + \phi_2) = \cos(\phi_1)\cos(\phi_2) - \sin(\phi_1)\sin(\phi_2) \tag{29}$$

$$\sin(\phi_1 + \phi_2) = \cos(\phi_1)\sin(\phi_2) + \sin(\phi_1)\cos(\phi_2) \tag{30}$$

gives

$$z_1 z_2 = |z_1||z_2|(\cos(\phi_1 + \phi_2) + i\sin(\phi_1 + \phi_2)) \tag{31}$$

This means that the modulus of the product of two complex numbers is the product of the modulus of each complex number,

$$|z_1 z_2| = |z_1||z_2| \tag{32}$$

and the argument of the product of two complex numbers is the sum of the arguments of each individual complex number,

$$\arg(z_1 z_2) = \arg(z_1) + \arg(z_2) \tag{33}$$

The addition of complex numbers can be represented graphically in the complex plane.

From the geometric observation that the length of any side of a triangle is less that the sum of the lengths of the two other sides gives

$$|z_1 + z_2| \leq |z_1| + |z_2| \tag{34}$$

$$|z_1| \leq |z_1 + z_2| + |z_2| \tag{35}$$

$$|z_2| \leq |z_1 + z_2| + |z_1| \tag{36}$$

which for obvious reasons are called triangle inequalities.

If we define

$$\rho(z_1, z_2) := |z_1 - z_2| \tag{37}$$

the using the definition of modulus (23) and the triangle inequality (34) it is easy to show

$$\rho(z_1, z_2) \geq 0 \qquad \rho(z_1, z_2) = 0 \qquad \Rightarrow \qquad z_1 = z_2 \tag{38}$$

$$\rho(z_1, z_2) \leq \rho(z_1, z_3) + \rho(z_3, z_2) \tag{39}$$

which means that $\rho(z_1, z_2)$ is a metric function for the complex numbers.

We started our discussion of complex numbers by looking for roots of the equation

$$P(z) = z^2 + 1 \tag{40}$$

Initially this was considered as a real valued function of a real variable, however showed that it could also be considered as a complex valued function of a complex variable.

This new interpretation has the advantage that this equation has two roots when considered as a complex valued function of a complex variable.

For this reason it is useful to introduce a new class of complex valued functions of a complex variable.

In general a function f is a mapping from a set A to a set B. The set A is called the domain of f and the set B is called the range of f. When A and B are both subsets of the complex plane $\mathbb{C}$ f is called a complex function of a complex variable.

It is also possible to have real valued functions of a complex variable such as

$$f(z); = zz^* + 4(zz^*)^2 \tag{41}$$

$$f(z) := \Im(z) = y \tag{42}$$

and complex valued functions of a real variable such as

$$f(\phi) = \cos(\phi) + i\sin(\phi) \tag{43}$$

Polynomials in z are a class of complex valued functions of complex arguments. A degree N polynomial is a function of the form

$$P(z) = \sum_{n=0}^{N} a_n z^n \tag{44}$$

where z is a complex variable and a_n are complex coefficients.

One of the great features of complex numbers is that while we introduced the complex number i to find roots of $0 = z^2 + 1$, we will show later that *any* degree N polynomial has N complex roots. No further extensions of the complex numbers are needed to completely factorize any polynomial. This property of the complex is called algebraic completeness. The theorem that the complex numbers are algebraically complete is called the fundamental theorem of algebra.

One of the simplest non-trivial complex functions of a complex argument is the exponential function. It is defined by

$$e^z := 1 + \sum_{n=1}^{\infty} \frac{z^n}{n!}. \tag{45}$$

Note that

$$|e^z - 1 + \sum_{n=1}^{N} \frac{z^n}{n!}| =$$

$$| \sum_{n=N+1}^{\infty} \frac{z^n}{n!}| \tag{46}$$

Mathematical induction can be used to show $|z^n| = |z|^n$ follows from (32), $|z^2| = |z|^2$. Combining this result with repeated use of the triangle inequality gives

$$\leq \sum_{n=N+1}^{\infty} \frac{|z|^n}{n!}. \tag{47}$$

This is a real sum which is known to converges to zero as $N \to \infty$ because this is a property of the real valued exponential function. (I will show this shortly)

Thus this series representation converges to the exponential function as N goes to infinity for every z. This shows that the definition (45) makes sense.

One special case of the exponential function is

$$e^{i\phi} = 1 + \sum_{n=1}^{\infty} \frac{i^2 \phi^n}{n!} =$$

$$\sum_{n=0}^{\infty} \frac{(i)^{2n}}{(2n)!} \phi^{2n} + i \sum_{n=0}^{\infty} \frac{(i)^{2n}}{(2n+1)!} \phi^{2n+1} =$$

$$\sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} \phi^{2n} + i \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} \phi^{2n+1} = \cos(\phi) + i \sin(\phi) \tag{48}$$

The series on the last line of (48) are recognized as the series representations for $\cos(\phi)$ and $\sin(\phi)$.

This allows us to write

$$z = |z|(\cos(\phi) + i\sin(\phi)) = |z|e^{i\phi} \tag{49}$$

and leads to

$$z_1 z_2 = |z_1||z_2|(\cos(\phi_1 + \phi_2) + i\sin(\phi_1 + \phi_2)) = |z_1||z_2|e^{i(\phi_1+\phi_2)} \tag{50}$$

One way to construct new functions is to approximate them by known functions. This can be done formally by using the metric function on $\mathbb{C}$.

A sequence of functions $f_n(x) : A \to B$ converges to a function $f(x)$ at $x = p$ if for every $\epsilon > 0$ there is an N such that for every $n > N$

$$\rho(f(p), f_n(p)) < \epsilon \tag{51}$$

where $\rho(\cdot, \cdot)$ is a metric function on B.

In general for a given ϵ the number N may depend on the point p. In the special case when the N can be chosen independent of $p \in A$ the convergence is uniform on A:

A sequence of functions $f_n(x) : A \to B$ converges uniformly to a function $f(x)$ if for every $\epsilon > 0$ there is an N such that for every $n > N$

$$\rho(f(x), f_n(x)) < \epsilon \tag{52}$$

for all $x \in A$.

Since convergence can be expressed in terms of a metric function, these definitions can be applied equally to real or complex valued functions.

In many cases one does not know the limiting function. It is possible to test for convergence of a sequence of approximate functions by considering only the terms in the sequence.

A Cauchy sequence f_n is a sequence of elements of a metric space B with the property that for every $\epsilon > 0$ there is an N such that whenever $m, n > N$

$$\rho(f_n - f_m) < \epsilon \tag{53}$$

To prove this assume that f_n converges to f. Pick $\epsilon > 0$ so there is an N with

$$\rho(f, f_n) < \frac{\epsilon}{2} \tag{54}$$

Let $m, n > N$. Then

$$\rho(f_n, f_m) \leq \rho(f_n, f) + \rho(f_m, f) \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \tag{55}$$

29:171 - Homework Assignment #1

1. Prove
$$e^{z_1}e^{z_2} = e^{z_1+z_2}$$
for any pair of complex numbers z_1 and z_2. Use the definition
$$e^z = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \cdots \qquad = \sum_{n=0}^{\infty} \frac{z^n}{n!}.$$

2. Find the real and imaginary part of $\sin(z)$ and $\cos(z)$. Express your results in terms of real valued functions.

3. Show that zz^* is always real and non-negative.

4. Use the quadratic formula to factorize the polynomial
$$P(z) = z^2 + 3z + 12.$$
Use complex arithmetic to verify that the polynomial is recovered by multiplying the factors.

5. Calculate the real and imaginary parts of
$$\frac{10 + i5}{7 - i6}.$$

6 Find the modulus and argument of $\cos(ix)$.

0.4 Lecture 4

Conversely assume that $\{f_n\}$ is a Cauchy sequence. We can choose a subsequence $\{f_{m_1}, f_{m_2}, \cdots, f_{m_k}, \cdots\}$ with m_l sufficiently large so

$$\rho(f_{m_k}, f_{m_{k+n}}) < \frac{1}{2^k}. \tag{56}$$

To do this first choose m_1 so $n > m_1$ implies $\rho(f_{m_1}, n) < \frac{1}{2}$. Next choose $m_2 > m_1$ so $n > m_2$ implies $\rho(f_{m_2}, n) < \frac{1}{2^2}$. This process can be continued to satisfy the inequalities in (56). Define

$$f := f_{m_1} + \sum_{k=1}^{\infty} (f_{m_{k+1}} - f_{m_k}) \tag{57}$$

Since

$$\rho(f, f_{m_l}) \leq \sum_{k=l+1}^{\infty} \rho(f_{m_k+1}, f_{m_k}) \leq \frac{1}{2^l} (\sum_{m=0}^{\infty} \frac{1}{2^m}) = \frac{1}{2^{l-1}}. \tag{58}$$

The right hand side can be made as small as desired by choosing a large enough l. Specifically, for any $\epsilon > 0$ it is possible to find an such that $\epsilon > \frac{1}{2^{l-1}}$. This shows that the sequence f_{n_k} converges to f as $k \to \infty$.

This shows that a necessary and sufficient for condition for a series to converge is that it is a Cauchy sequence. We also showed how to express the limit as a convergent series.

In the above the Cauchy sequence was a sequence of numbers. The above proof also extends to the case that $f \to f(z)$ and $f_n \to f_n(z)$ are functions.

Cauchy sequences can be used to show either convergence or uniform convergence of sequences of functions:

A sequence $f_n(z)$ converges for $z = p$ if for every $\epsilon > 0$ there is an N such that whenever $m, n > N$

$$\rho(f_n(p), f_m(p)) < \epsilon \tag{59}$$

A sequence $f_n(z)$ is uniformly convergent on a set S if for every $\epsilon > 0$ there is an N such that whenever $m, n > N$

$$\rho(f_n(z), f_m(z)) < \epsilon \tag{60}$$

independent of $z \in S$.

Example: We show that the partial sums of the exponential series are a Cauchy sequence for each z, they are not however uniformly Cauchy. First define the partial sums:

$$z_m := \sum_{n=0}^{m} \frac{z^n}{n!} \tag{61}$$

To show $\{z_m\}$ is a Cauchy sequence consider (for $k > m$)

$$z_m - z_k = \sum_{n=m+1}^{k} \frac{z^n}{n!} \tag{62}$$

We can write the right hand side of (**??**) as

$$\frac{z^{m+1}}{(m+1)!}(1 + \frac{x}{m+2} + \frac{x^2}{(m+2)(m+3)} + \cdots \frac{x^{k-m-1}}{(m+2)\cdots k})$$

Repeated use of the triangle inequality means that the amplitude of this complex number is bounded by

$$\begin{aligned} &\leq \frac{|z|^{m+1}}{(m+1)!}(1 + \frac{|z|}{m+2} + \frac{|z|^2}{(m+2)^2} + \cdots \frac{|z|^{k-m-1}}{(m+2)^{k-m-1}}) \leq \\ &\leq \frac{|z|^{m+1}}{(m+1)!}(1 + \frac{|z|}{m+2} + \frac{|z|^2}{(m+2)^2} + \cdots = \\ &\frac{|z|^{m+1}}{(m+1)!}\frac{1}{1-|z|/(m+2)} \end{aligned} \tag{63}$$

where we have used the geometric series

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x} \tag{64}$$

for $|x| < 1$. For any fixed $|z|$ it is possible to choose m sufficiently large to make (63) as small as desired.

This same proof can be used to show that the series for the the following functions

$$\sin(z) := \sum_{n=0}^{\infty} \frac{(-)^n}{(2n+1)!} z^{2n+1} \tag{65}$$

$$\cos(z) := \sum_{n=0}^{\infty} \frac{(-)^n}{(2n)!} z^{2n} \tag{66}$$

$$\sinh(z) := \sum_{n=0}^{\infty} \frac{1}{(2n+1)!} z^{2n+1} \tag{67}$$

$$\cosh(z) := \sum_{n=0}^{\infty} \frac{1}{(2n)!} z^{2n} \tag{68}$$

converge for all z in the complex plane.

Because these functions are represented by absolutely convergent series, many identities that one derives for the corresponding functions of a real variable have generalizations to a complex variable.

An immediate consequence of equations (65), (66),(67),(68) are the identities

$$\cosh(iz) = \cos(z) \qquad \cos(iz) = \cosh(z) \tag{69}$$

$$\sinh(iz) = i\sin(z) \qquad \sin(iz) = i\sinh(z) \tag{70}$$

$$\cosh(z) = \frac{1}{2}(e^z + e^{-z}) \qquad \sinh(z) = \frac{1}{2}(e^z - e^{-z}) \tag{71}$$

$$e^{\pm z} = \cosh(z) \pm \sinh(z) \tag{72}$$

$$e^{\pm iz} = \cos(z) \pm i\sin(z) \tag{73}$$

$$\cos(z) = \frac{1}{2}(e^{iz} + e^{-iz}) \qquad \sin(z) = \frac{1}{2i}(e^{iz} - e^{-iz}) \tag{74}$$

Four homework I will ask you to prove

$$\cos(z_1 + z_2) = \cos(z_1)\cos(z_2) - \sin(z_1)\sin(z_2) \tag{75}$$

$$\sin(z_1 + z_2) = \sin(z_1)\cos(z_2) + \cos(z_1)\sin(z_2) \tag{76}$$

From (69)- (76) we get

$$\cos(x + iy) = \cos(x)\cosh(y) - i\sin(x)\sinh(y) \tag{77}$$

$$\sin(x+iy) = \sin(x)\cosh(y) + i\cos(x)\sinh(y) \tag{78}$$

$$\cosh(z_1+z_2) = \cosh(z_1)\cosh(z_2) + \sinh(z_1)\sinh(z_2) \tag{79}$$

$$\sinh(z_1+z_2) = \sinh(z_1)\cosh(z_2) + \cosh(z_1)\sinh(z_2) \tag{80}$$

$$\cosh(x+iy) = \cosh(x)\cos(y) + i\sinh(x)\sin(y) \tag{81}$$

$$\sinh(x+iy) = \sinh(x)\cos(y) + i\cosh(x)\sin(y) \tag{82}$$

We can also define the logarithm of a complex variable. It is defined to satisfy the relations

$$e^{\ln(z)} = z \tag{83}$$

If we set

$$\ln(z) = u(z) + iv(z) \tag{84}$$

then using the result of the homework problem 1 gives

$$e^{\ln(z)} = e^{u(z)+iv(z)} = e^{u(z)}e^{iv(z)} = |z|e^{i\phi} \tag{85}$$

or

$$u(z) = \ln|z| \qquad v(z) = \phi(z) + 2n\pi \tag{86}$$

Putting all of this together gives

$$\ln(z) = \ln|z| + i(\phi(Z) + 2n\pi). \tag{87}$$

Here n is any integer. The complex logarithm is an example of a multiple valued complex function.

0.5 Lecture 5

Complex derivatives

A general complex-valued function of a complex variable $z = x + iy$ has the form

$$f(z) = u(x, y) + iv(x, y). \tag{88}$$

This function has a complex derivative if

$$\frac{df}{dz}(z) = \lim_{\Delta z \to 0} \frac{f(z + \Delta z) - f(z)}{\Delta z} \tag{89}$$

exists and is unique. The existence of a complex derivative is a restrictive condition. The problem is that while this looks similar to the ordinary definition of the derivative of a real valued function of a real variable, the uniqueness requirement means that the result is independent of the direction of Δz approaches zero in the complex plane.

If we write

$$\Delta z = |\Delta z| e^{i\Delta\phi} \tag{90}$$

it follows that $|\Delta z| \to 0$ implies $\Delta z \to 0$, however in general the limit may have a residual dependence on ϕ. When the function has a complex derivative there can be no residual ϕ dependence.

Example 1: Let $z = x + iy$ and $f(z) = x + 2iy$. Then

$$\frac{df}{dz}(0) = \lim_{\Delta z \to 0} \frac{x + 2iy}{x + iy} = \lim_{\Delta z \to 0} \frac{x^2 + 2y^2 + ixy}{x^2 + y^2} \tag{91}$$

If we set $y = 0$ and let $x \to 0$ then $\frac{df}{dz}(0) = 1$; on the other hand if we set $x = 0$ and let $y \to 0$ we get $\frac{df}{dz}(0) = 2$. Clearly the derivative of this function at zero depends on how Δz approaches 0. This means that $f(z) = x + 2iy$ does not have a complex derivative.

Example 2: Consider $f(z) = z^2$. Then

$$\frac{df}{dz} = \lim_{\Delta z \to 0} \frac{z^2 + 2z\Delta z + (\Delta z)^2}{\Delta z} = \lim_{\Delta z \to 0} (2z + \Delta z) = 2z \tag{92}$$

which is clearly independent of how $\Delta z \to 0$.

When $f(z)$ does have a complex derivative, since the definition of the derivative is essentially the same as it is for real valued functions of a real argument, it follows that the complex derivative of $f(z)$ obeys the standard rules of differential calculus:

$$\frac{d}{dz}(f_1(z) + f_2(z)) = \frac{df_1}{dz}(z) + \frac{df_2}{dz}(z) \tag{93}$$

$$\frac{d}{dz}(f_1(z)f_2(z)) = \frac{df_1}{dz}(z)f_2(z) + f_1(z)\frac{df_2}{dz}(z) \tag{94}$$

$$\frac{d}{dz}(f_1(f_2(z))) = \frac{df_1}{dz}(f_2(z))\frac{df_2}{dz}(z) \tag{95}$$

$$\frac{d}{dz}\frac{f_1(z)}{f_2(z)} = \frac{df_1}{dz}(z)\frac{1}{f_2(z)} - \frac{f_1(z)}{f_2(z)^2}\frac{df_2}{dz}(z) = \tag{96}$$

$$\frac{1}{f_2(z)^2}\left(f_2(z)\frac{df_1}{dz}(z) - f_1(z)\frac{df_2}{dz}(z)\right). \tag{97}$$

The proof of these results are identical to the corresponding proofs for real valued functions.

Since the existence of a complex derivative is not always immediately apparent, it is useful to have a simple test to see if a function has a complex derivative. Let

$$f(z) = u(x, y) + iv(x, y). \tag{98}$$

If $f(z)$ has a complex derivative then we can compute it by letting $\Delta z = \Delta x$ or $\Delta z = i\Delta y$. Both must give the same result:

$$\frac{df}{dz}(z) = \frac{\partial f}{\partial x}(z) = -i\frac{\partial f}{\partial y}(z) \tag{99}$$

Express this in terms of the real and imaginary components to get

$$\frac{df}{dz}(z) = \frac{\partial u}{\partial x}(z) + i\frac{\partial v}{\partial x}(z) = \frac{\partial v}{\partial y}(z) - i\frac{\partial u}{\partial y}(z). \tag{100}$$

Equating the real and imaginary parts of these expressions gives

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x} \tag{101}$$

Equation (101) are called the Cauchy-Riemann equations. They are a necessary condition for $f(z) = u(x, y) + iv(x, y)$ to have a complex derivative.

One immediate consequence of the existence of a complex derivative is that it can be computed by letting $\Delta z \to 0$ in any direction in the complex plane.

The Cauchy-Riemann equations have some interesting consequences. If

$$f(z) = u(x, y) + iv(x, y) \tag{102}$$

and the second partial derivatives of u and v with respect to x and y exist, then we have

$$\frac{\partial^2 u}{\partial x^2} = \frac{\partial^2 v}{\partial x \partial y} \tag{103}$$

$$\frac{\partial^2 u}{\partial y^2} = -\frac{\partial^2 v}{\partial y \partial dx} \tag{104}$$

Adding equations (103) and (104) gives

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0 \tag{105}$$

which means that if u is the real part of a function with a complex derivative that it is necessarily a solution of Laplace's equation in two dimensions.

Using the same method it also follows that the imaginary part of a function with a complex derivative is also a solution of Laplace's equation

$$\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} = 0 \tag{106}$$

These two solutions are not independent, they are necessarily related by the Cauchy-Riemann equations.

This has an interesting geometric interpretation. Consider the gradients of the functions $u(x, y)$ and $v(x, y)$:

$$\vec{\nabla} u(x, y) = (\frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}) \qquad \vec{\nabla} v(x, y) = (\frac{\partial v}{\partial x}, \frac{\partial y}{\partial y}) \tag{107}$$

The dot product of these two vectors is

$$\vec{\nabla} u(x, y) \cdot \vec{\nabla} v(x, y) = \frac{\partial u}{\partial x}\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}\frac{\partial v}{\partial y} \tag{108}$$

Use the Cauchy Riemann equations in the last two terms to get

$$\vec{\nabla} u(x,y) \cdot \vec{\nabla} v(x,y) = \frac{\partial u}{\partial x}\frac{\partial v}{\partial x} + (-\frac{\partial v}{\partial x})\frac{\partial u}{\partial x} = 0. \tag{109}$$

To understand the meaning of this relation note that the equations

$$u(x,y) = c_1 \tag{110}$$

and

$$v(x,y) = c_2, \tag{111}$$

where c_1 and c_2 are constants, define curves in the complex plane. The gradient of u is perpendicular to the curve $u(x,y) = c_1$, while the gradient of v is perpendicular to the curve $v(x,y) = c_2$. Since we have shown that the gradients are perpendicular to each other, it follows that if the curves $u(x,y) = c_1$ and $v(x,y) = c_2$ intersect, they necessarily intersect at right angles!

Thus the real and imaginary parts of complex function with a complex derivative are solutions of Laplace's equations whose level surfaces are everywhere perpendicular. Functions satisfying Laplace's equation (in any number of variables) are called harmonic functions.

The requirement that the real and imaginary parts of a function with a complex derivative are harmonic indicates the restrictive nature of functions that have Complex derivatives.

We have shown that the Cauchy-Riemann equations are a necessary condition for complex function of a complex variable to have a complex derivative.

I will show that if a complex valued function of a complex variable satisfies the Cauchy Riemann equations and the first partial derivatives of that function are continuous, then that function has a complex derivative. This provides a simple means to test for the existence of a complex derivative.

Theorem 5.1 Let $f(z) = u(x,y) + iv(x,y)$ and assume

$$\frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}, \frac{\partial v}{\partial x}, \frac{\partial v}{\partial y} \tag{112}$$

exist, are continuous in a region R containing z, and satisfy the Cauchy Riemann identities. Then $f(z)$ has a complex derivative.

To prove this we first pick a point $z = x + iy$. The existence of the partial derivatives means that

$$u(x+\Delta x, y) - u(x,y) = \Delta x[\frac{\partial u}{\partial x}(x,y) + \delta_{ux}(x,y,\Delta x)] \qquad (113)$$

where

$$\delta_{ux}(x,y;\Delta x) \to 0 \qquad \text{as} \qquad \Delta x \to 0 \qquad (114)$$

Similarly

$$u(x, y+\Delta y) - u(x,y) = \Delta y[\frac{\partial u}{\partial y}(x,y) + \delta_{uy}(x,y;\Delta y)] \qquad (115)$$

where

$$\delta_{uy}(x,y;\Delta y) \to 0 \qquad \text{as} \qquad \Delta y \to 0 \qquad (116)$$

and

$$v(x+\Delta x, y) - v(x,y) = \Delta x[\frac{\partial v}{\partial x}(x,y) + \delta_{vx}(x,y;\Delta x)] \qquad (117)$$

where

$$\delta_{vx}(x,y;\Delta x) \to 0 \qquad \text{as} \qquad \Delta x \to 0 \qquad (118)$$

and

$$v(x, y+\Delta y) - v(x,y) = \Delta y[\frac{\partial v}{\partial y}(x,y) + \delta_{vy}(x,y;\Delta y)] \qquad (119)$$

where

$$\delta_{vy}(x,y;\Delta y) \to 0 \qquad \text{as} \qquad \Delta y \to 0. \qquad (120)$$

Continuity of these partial derivatives means

$$\frac{\partial u}{\partial x}(x+\Delta x, y) = \frac{\partial u}{\partial x}(x,y) + \xi_{uxx}(x,y;\Delta x) \qquad (121)$$

$$\frac{\partial u}{\partial x}(x, y+\Delta y) = \frac{\partial u}{\partial x}(x,y) + \xi_{uxy}(x,y;\Delta y) \qquad (122)$$

$$\vdots \qquad (123)$$

where continuity requires

$$\xi_{uxx}(x,y;\Delta x) \to 0 \qquad (124)$$

$$\xi_{uxy}(x,y;\Delta y) \to 0 \qquad (125)$$

etc. as $\Delta x \to 0$ and $\Delta y \to 0$

I use these relations along with the Cauchy Riemann equations to compute the complex derivative of $f(z)$:

$$\frac{f(z+\Delta z)-f(z)}{\Delta z} =$$

$$\frac{u(x+\Delta x, y+\Delta y)+iv(x+\Delta x, y+\Delta y)-u(x,y)-iv(x,y)}{\Delta x+i\Delta y} \tag{126}$$

I expand all terms using the above relations and use the Cauchy -Riemann equations to factor out a Δz. I first do the expansion for u and v individually.

$$u(x+\Delta x, y+\Delta y) =$$

$$u(x, y+\Delta y)+\Delta x[\frac{\partial u}{\partial x}(x, y+\Delta y)+\delta_{ux}(x, y+\Delta y; \Delta x)] =$$

$$u(x, y+\Delta y)+\Delta x[\frac{\partial u}{\partial x}(x,y)+\xi_{uxy}(x, y; \Delta y)+\delta_{ux}(x, y+\Delta y; \Delta x)] =$$

$$u(x,y)+\Delta y[\frac{\partial u}{\partial y}(x,y)+\delta_{uy}(x,y;\Delta y)]+$$

$$\Delta x[\frac{\partial u}{\partial x}(x,y)+\xi_{uxy}(x,y;\Delta y)+\delta_{ux}(x, y+\Delta y; \Delta x)] \tag{127}$$

It follows that

$$u(x+\Delta x, y+\Delta y)-u(x,y) =$$

$$\Delta y[\frac{\partial u}{\partial y}(x,y)+\delta_{uy}(x,y;\Delta y)]+$$

$$\Delta x[\frac{\partial u}{\partial x}(x,y)+\xi_{uxy}(x,y;\Delta y)+\delta_{ux}(x, y+\Delta y, \Delta x)] \tag{128}$$

An identical calculation replacing u with v gives

$$v(x+\Delta x, y+\Delta y)-v(x,y) =$$

$$\Delta y[\frac{\partial v}{\partial y}(x,y)+\delta_{vy}(x,y;\Delta y)]+$$

$$\Delta x[\frac{\partial v}{\partial x}(x,y)+\xi_{vxy}(x,y,\Delta y)+\delta_{vx}(x, y+\Delta y, \Delta x)] \tag{129}$$

Using the Cauchy-Riemann equations and adding i times the second term to the first gives

$$u(x+\Delta x, y+\Delta y)-u(x,y)+iv(x+\Delta x, y+\Delta y)-iv(x,y) =$$

$$
\begin{gathered}
\Delta y[-\frac{\partial v}{\partial x}(x,y) + \delta_{uy}(x,y,\Delta y)]+ \\
\Delta x[\frac{\partial u}{\partial x}(x,y) + \xi_{uxy}(x,y,\Delta y) + \delta_{ux}(x,y+\Delta y,\Delta x)]+ \\
+i\Delta y[\frac{\partial u}{\partial x}(x,y) + \delta_{vy}(x,y,\Delta y)]+ \\
+i\Delta x[\frac{\partial v}{\partial x}(x,y) + \xi_{vxy}(x,y,\Delta y) + \delta_{vx}(x,y+\Delta y,\Delta x)) = \\
(\Delta x + i\Delta y)[\frac{\partial u}{\partial x}(x,y) + i\frac{\partial v}{\partial x}(x,y)]+ \\
\Delta y[\delta_{uy}(x,y,\Delta y) + i\delta_{vy}(x,y,\Delta y)]+ \\
\Delta x[\xi_{uxy}(x,y,\Delta y)+\delta_{ux}(x,y+\Delta y,\Delta x))+i\xi_{vxy}(x,y,\Delta y)+i\delta_{vx}(x,y+\Delta y,\Delta x)]
\end{gathered} \tag{130}
$$

Dividing by $\Delta z = \Delta x + i\Delta y$ gives

$$
\begin{gathered}
\frac{f(z+\Delta z) - f(z)}{\Delta z} = \\
[\frac{\partial u}{\partial x}(x,y) + i\frac{\partial v}{\partial x}(x,y)]+ \\
\frac{\Delta y}{\Delta z}[\delta_{uy}(x,y,\Delta y) + i\delta_{vy}(x,y,\Delta y)]+ \\
\frac{\Delta x}{\Delta z}[\xi_{uxy}(x,y,\Delta y)+\delta_{ux}(x,y+\Delta y,\Delta x))+i\xi_{vxy}(x,y,\Delta y)+i\delta_{vx}(x,y+\Delta y,\Delta x)]
\end{gathered} \tag{131}
$$

If I take the limit as $\Delta z \to 0$ the last two lines of equation (131) vanish. The surviving term is the complex derivative:

$$
\begin{gathered}
\frac{df}{dz}(z) = \lim_{\Delta z \to 0} \frac{f(z+\Delta z) - f(z)}{\Delta z} = \\
\frac{\partial u}{\partial x}(x,y) + i\frac{\partial v}{\partial x}(x,y)
\end{gathered} \tag{132}
$$

This result has no residual dependence on the argument of Δz.

This completes the proof of theorem 5.1. This shows that the existence of a complex derivative of $f(z)$ can be established by (1) checking that the real and imaginary parts of $f(z)$ satisfy the Cauchy-Riemann equations and that the partial derivatives are continuous functions.

0.6 Lecture 6

One more simple calculation will help clarify the meaning of a function with a complex derivative. Let

$$f(x,y) = u(x,y) + iv(x,y) \tag{133}$$

and assume that the double Taylor series in x and y about the origin converges. Then

$$\text{series} = \sum_{m,n=0}^{\infty} \frac{x^n y^m}{n!m!} \frac{\partial^n}{\partial x^n} \frac{\partial^m}{\partial y^m} (u+iv)_{|0,0}. \tag{134}$$

The Cauchy-Riemann equations imply

$$\frac{\partial}{\partial y}(u+iv)_{|0,0} = \frac{\partial}{\partial x}(-v+iu)_{|0,0} = i\frac{\partial}{\partial x}(u+iv)_{|0,0} \tag{135}$$

I use this to convert all of the y derivatives in the Taylor expansion to x derivatives to get

$$\text{series} = \sum_{m,n=0}^{\infty} \frac{x^n (iy)^m}{n!m!} \frac{\partial^n}{\partial x^{n+m}} (u+iv)_{|0,0} \tag{136}$$

Next I let $k = m+n$ replace the sum over m and n by a sum over k and n:

$$\text{series} = \sum_{k=0}^{\infty} \frac{1}{k!} \sum_{n=0}^{k} \frac{k!}{n!(k-n)!} x^n (iy)^{k-n} \frac{\partial^n}{\partial x^k} (u+iv)_{|0,0} \tag{137}$$

The binomial theorem gives

$$\text{series} = \sum_{k=0}^{\infty} \frac{(x+iy)^k}{k!} \frac{\partial^n}{\partial x^k} (u+iv)_{|0,0} \tag{138}$$

This shows that if the function has a convergent Taylor series and the real and imaginary parts satisfy the Cauchy-Riemann equations, then the function is depends only on the combination $(x+iy)$, rather than independently on x and y.

Since we can alway write a general functions of x and y as a function of z and z^*, what we have shown that if the function satisfies the Cauchy Riemann equations that

$$\frac{\partial f(z,x^*)}{\partial z^*} = 0 \tag{139}$$

For Homework I will ask you to show that equation (139) is equivalent to the Cauchy Riemann equations.

Integrals of functions of a complex variable

Complex functions can also be integrated. I will define the integral of a complex valued functions of a complex variable along a path in the complex plane.

By a path in the complex plane we mean a function of the form

$$z(t) = u(t) + iv(t) \tag{140}$$

where t is a real parameter that varies between t_a and t_b with

$$z(t_a) = a \qquad z(t_b) = b \tag{141}$$

We are interested in integrals where the path is continuous. In addition, while we can tolerate kinks in the path, we do not want to consider paths that have an infinite number of kinks when t ranges over a finite interval.

A regular curve C is a curve in the complex plane where the functions $u(t)$ and $v(t)$ are piecewise differentiable. For any finite subinterval $[t_a, t_b]$, this means that it is possible to divide $[t_a, t_b]$ into a finite number of subintervals

$$[t_a, t_b] = [t_a, t_{a_1}] \cup [t_{a_1}, t_{a_2}] \cup \cdots \cup \cdots \cup [t_{a_{N-1}}, t_{a_{N-1}}][t_{a_{N-1}}, t_b] \tag{142}$$

where $u(t)$ and $v(t)$ continuous on $[t_a, t_b]$ and have continuous derivatives on each of the subintervals $[t_{a_k}, t_{a_{k+1}}]$.

To define the integral of a complex function $f(z)$ along a regular path C first subdivide the curve into n segments

$$a = z_0, z_1, z_2 \cdots z_{n-1}, z_n = b \tag{143}$$

On each of the sub intervals I choose a point ζ_i and define the sum:

$$I_n := \sum_{k=1}^{N} f(\zeta_k)(z_k - z_{k-1}) \tag{144}$$

This quantity is a sum of products of complex numbers and is thus a complex number.

This can be repeated for any number of points n. I choose the points so $|z_k - z_{k+1}| \to 0$ and $n \to \infty$.

The limit

$$I := \lim_{n\to\infty} I_n, \tag{145}$$

provided it exists and is independent of how the points z_k and ζ_k are chosen is called the contour integral of $f(z)$ along C and is written as

$$I := \int_C f(z)dz \tag{146}$$

In general the value of a contour integral will depend on both the endpoints and the choice of path, C.

While a contour integral can be evaluated directly from the definition, it can be reduced to an ordinary integral by writing

$$I_n := \sum_{k=1}^{N} f(\zeta_k)(z_k - z_{k-1}) =$$

$$\sum_{k=1}^{N}(u(\zeta_k) + iv(\zeta_k))(x_k - x_{k-1} + iy_k - iy_{k-1}) =$$

$$\sum_{k=1}^{N}[u(\zeta_k)(x_k - x_{k-1}) - v(\zeta_k))(y_k - iy_{k-1})]+$$

$$i\sum_{k=1}^{N}[v(\zeta_k)(x_k - x_{k-1}) + u(\zeta_k)(y_k - y_{k-1})] \tag{147}$$

Since

$$|z_k - z_{k+1}| \to 0 \tag{148}$$

implies

$$|x_k - x_{k+1}| \to 0 \qquad |y_k - y_{k+1}| \to 0 \tag{149}$$

this become the definition of the integral of a function of two variables along a path in a plane:

$$I = \int [u(x,y)dx - v(x,y)dy]+$$

$$i\int [v(x,y)dx + u(x,y)dy] \tag{150}$$

and if the path $C = (x(t), y(t))$ this becomes

$$I = \int_a^b [u(x,y)\frac{dx}{dt} - v(x,y)\frac{dy}{dt}] +$$

$$i\int_a^b [v(x,y)\frac{dx}{dt} + u(x,y)\frac{dy}{dt}]dt \tag{151}$$

which is an ordinary integral. If C has continuous partial derivatives, then $\frac{dx}{dt}$ and $\frac{dy}{dt}$ are continuous functions of t. If C has only piecewise continuous partial derivatives, then the integral is replaced by a sum of terms of the above form, corresponding to integrals between the points where the derivatives have discontinuities.

Equation (151) is the standard way to calculate the Contour integral. In general there will be many ways to parameterize the same curve. Different parameterizations will give the same value of the integral.

29:171 - Homework Assignment #2

1. Show
$$\cos(z_1 + z_2) = \cos(z_1)\cos(z_2) - \sin(z_1)\sin(z_2)$$
   You may use the results from problem 1 of the first set, but be careful to work from definitions.

2. Show that if $f_1(z) = u_1(x, y) + iv_1(x, y)$ and $f_2(z) = u_2(x, y) + iv_2(x, y)$ have continuous partial derivatives and satisfy the Cauchy-Riemann equations then $g(z) = f_1(f_2(z))$ has continuous partial derivatives and satisfies the Cauchy-Riemann equations.

3. Show that $f(z) = e^{z^2}$ has a complex derivative.

4. Find the real and imaginary parts of $f(z) = \sin(z)$. Show explicitly that they both are solutions to Laplace's equation and show that their gradients are orthogonal at any point.

5. For $z = re^{i\phi} = x + iy$, let $f(z) = u(r, \phi) + iv(r, \phi)$. Find the form of the Cauchy-Riemann equations in terms of the (r, ϕ) variables.

6. Let $u(x, y) = ax^3 + bx^2y + cxy^2 + dy^3$. Find the values of a, b, c, d for which this function satisfies Laplace's equation. For these values of a, b, c, d, find a corresponding $v(x, y)$ that satisfies Laplace's equation and the Cauchy-Riemann equations with $u(x, y)$?

7. Let a and b be complex numbers. Show that
$$f(z) = \frac{a - z}{b - z}$$
   has a complex derivative for $z \neq b$.

0.7 Lecture 7

One property of ordinary real integrals that is shared by contour integrals is that the integral of a derivative depends on the value of the function at the endpoints.

Equation (151) can be written as

$$I = \int_a^b f(z(t)) \frac{dz}{dt} dt \tag{152}$$

Assume that

$$f(z) = \frac{dF(z)}{dz} \qquad F(z) = U(z) + iV(z) \tag{153}$$

where $F(z)$ has a complex derivative at all points on the path C. Then using the definition of the complex derivative it is easy to show, because the derivative is independent of direction, that

$$\frac{dF(z(t))}{dz} \frac{dz}{dt} = \frac{dF(z(t))}{dt} \tag{154}$$

which implies

$$\begin{gathered} I = \int_a^b f(z(t)) \frac{dz}{dt} dt = \\ \int_a^b \frac{dF(z(t))}{dt} dt = \\ \int_a^b [\frac{dU(z(t))}{dt} + i\frac{dU(z(t))}{dt}] dt = \\ (U(b) + iU(b)) - (U(a) + iV(a)) = F(b) - F(a) \end{gathered} \tag{155}$$

This result is the complex version of the fundamental theorem of calculus.

In applying this theorem, some extra care is required to treat integrals of derivatives of multiple valued functions, like the complex logarithm. This will be discussed later.

Contour integrals satisfy standard properties that are shared by ordinary Riemann integrals, since both are expressed as limits of sums of products:

$$\int_C af(z)dz = a \int_C f(z)dz \tag{156}$$

$$\int_C (f(z)+g(z))dz = \int_C f(z)dz + \int_C g(z)dz \tag{157}$$

If c is a point on C that divides C into two paths C_1 and C_2 then

$$\int_C f(z)dz = \int_{C_1} f(z)dz + \int_{C_2} f(z)dz. \tag{158}$$

The integration by parts formula follows from

$$\int_a^b \frac{d}{dz}(f(z)g(z))dz =$$

$$f(b)g(b) - f(a)g(a) = \int_a^b f(z)\frac{dg}{dz}(z)dz + \int_a^b g(z)\frac{df}{dz}(z)dz \tag{159}$$

Example: 1 Let $f(z) = x + 2iy$ and let C be the path composed by joining the two curves

$$C_1: \quad z_1(t) = 2t \qquad 0 \le t \le 1 \tag{160}$$

$$C_2: \quad z_2(t) = 2 + it \qquad 0 \le t \le 1 \tag{161}$$

$$I = \int_{C_1+C_2} f(z)dz = \int_{C_1} f(z)dz + \int_{C_2} f(z)dz =$$

$$\int_0^1 f(z_1(t))\frac{dz_1(t)}{dt}dt + \int_0^1 f(z_2(t))\frac{dz_2(t)}{dt}dt =$$

$$\int_0^1 f(z_1(t))\frac{dz_1(t)}{dt}dt + \int_0^1 f(z_2(t))\frac{dz_2(t)}{dt}dt =$$

$$\int_0^1 2t \cdot 2dt + \int_0^1 (2+2it)idt = 2 + 2i - 11 - 2i \tag{162}$$

In this example the value of the integral will dependent on the path in the complex plane.

Example: 2 Let $f(z) = z^2$ and let C be the curve

$$C: \quad z_(t) = t + it \qquad 0 \le t \le 1 \tag{163}$$

$$I = \int_C f(z)dz = \int_0^1 f(z(t))\frac{dz(t)}{dt}dt +$$

$$\int_0^1 (t+it)^2(1+i)dt = \frac{1}{3}(t+it)^3 \tag{164}$$

In this case we note that since

$$z^2 = \frac{1}{3}\frac{dz^3}{dz} \tag{165}$$

we can also use the fundamental theorem of calculus to show that

$$I = \int_C f(z)dz = \int_C \frac{1}{3}\frac{d}{dz}z^3 dz = \frac{1}{3}z^3|_0^{1+i} = \frac{1}{3}(1+i)^3 \tag{166}$$

This gives the same result at the explicit integral over C, but since this is the integral of a derivative, all that matters in the value of the function being differentiated at the endpoints of C. In this case the integral is independent of the choice of path.

Example: 3 Let $f(z) = 1/z$ and let C be the a circle of radius 1 about the origin

$$C: \quad z_1(t) = e^{i\phi} \qquad 0 \leq \phi \leq 2\pi \tag{167}$$

$$I = \int_C f(z)dz = \int_0^{2\pi} f(z(\phi))\frac{dz(\phi)}{d\phi}d\phi + =$$

$$\int_0^{2\pi} \frac{1}{e^{i\phi}} ie^{i\phi} d\phi = i\int_0^{2\pi} d\phi = 2\pi i \tag{168}$$

The interesting thing about this integral is that

$$\frac{1}{z} = \frac{d}{dz}\ln(z), \tag{169}$$

however the integral around a closed curve does not give zero. The problem in this case is that the natural logarithm is a multiple valued function. While the value of the integral does not depend on the starting value, going around the loop once always increases the imaginary part by 2π above its previous value. So while the fundamental theorem of calculus still works, we have to pay attention for which part of the multiple valued function is evaluated at each point. This example will be important later.

In applications that follow it will be necessary to compute bounds on the modulus of certain contour integrals. The simplest such bound follows immediately from the definition:

$$I_n = \sum_{k=1}^{n} f(\zeta_k)(z_k - z_{k-1}) \tag{170}$$

Using our standard inequalities

$$|I_n| \leq \sum_{k=1}^{n} |f(\zeta_k)||z_k - z_{k-1}| \tag{171}$$

Since

$$|f(\zeta_k)| < |f| \tag{172}$$

where $|f|$ is the maximum value of the modulus of $f(z)$ on C the above can be replaced by

$$|I_n| \leq |f| \sum_{k=1}^{n} |z_k - z_{k-1}| \tag{173}$$

Put the sum is the length of a polygon inscribed inside of the curve, which bounded by the length of the curve. If L is the length of the curve we get

$$|I_n| \leq |f|L \tag{174}$$

Since the right hand side is independent of n it follow that if the integral exists then

$$|\int_C f(z)dz| \leq |f|L \tag{175}$$

This inequality is called Darboux's inequality.

Analytic functions

A complex function $f(z)$ is analytic at a point $z = z_0$ if there is a neighborhood on z_0 where $f(z)$ is single valued and has a complex derivative.

The set of points in the complex plane where $f(z)$ is analytic is called the domain of analyticity of $f(z)$.

A complex function $f(z)$ whose domain of analyticity is the entire complex plane is called an entire function.

A point z_0 where $f(z)$ is analytic is called a regular point of $f(z)$.

A point z_0 where $f(z)$ is not analytic is called a singular point of $f(z)$.

A singular point z_0 of $f(z)$ is isolated if there is a neighborhood of z_0 where $f(z)$ is analytic for $z \neq z_0$

Conformal Mapping

For a function $f(x)$ of a single real variable, the condition that the equation $y = f(x)$ can be solved for $x = g(y)$ in a neighborhood of $y_0 = f(x_0)$ is that

$$\frac{df}{dx}(x_0) \neq 0 \tag{176}$$

and is continuous in a neighborhood of x_0 This theorem is called the inverse function theorem. It can be proved using concepts from calculus.

The basic idea behind the proof is to note that the existence of a derivative at $x = x_0$ means that it is possible to write

$$y = f(x) = f(x_0) + \frac{df}{dx}(x_0)(x - x_0) + R(x_0, x) \tag{177}$$

where the remainder

$$R(x, x_0) = f(x) - f(x_0) - \frac{df}{dx}(x_0)(x - x_0) \tag{178}$$

is a known function of x, and x_0 that vanishes as $x \to x_0$.

Using $y_0 = f(x_0)$ and the assumption that

$$\frac{df}{dx}(x_0) \neq 0 \tag{179}$$

gives

$$x = x_0 + \frac{1}{\frac{df}{dx}(x_0)}[y - y_0 - R(x_0, x)]. \tag{180}$$

Equation (181) is not a solution of the inverse problem because it has an x dependence on both sides, however when y is close to y_0 one expects that the remainder $R(x_0, x)$ will be small. This suggests trying to solve this equation by iteration

$$x_0(y) = x_0$$

$$x_1(y) = x_0 + \frac{1}{\frac{df}{dx}(x_0)}[y - y_0 - R(x_0, x_0)]$$

$$\vdots$$

$$x_n(y) = x_0 + \frac{1}{\frac{df}{dx}(x_0)}[y - y_0 - R(x_0, x_{n-1}(y))] \tag{181}$$

The sequence $\{x_n(y)\}$ are functions of y that converges to the desired inverse function provided that the sequence is Cauchy. It can be shown that this sequence will be a Cauchy for small enough $|y - y_0|$ if $\frac{df}{dx}(x)$ is non-zero and continuous in a neighborhood of x_0.

Clearly this construction can be generalized to complex functions. This implies that an analytic function with a non-zero complex derivative at $z = z_0$ can be interpreted as a locally invertible mapping for a neighborhood U of z_0 in the complex plane to a neighborhood V of $z'_0 = f(z_0)$.

The mapping can be expressed as a mapping from (x, y) to (x', y'):

$$z' = x' + iy' = f(z) = u(x, y) + iv(x, y) \tag{182}$$

Equating the real and imaginary parts gives

$$x' = u(x, y) \qquad y' = v(x, y) \tag{183}$$

A mapping of this form, where $f(z)$ is analytic in a neighborhood of $z = z_0$ with a non-zero derivative at z_0, is called a conformal transformation.

0.8 Lecture 8

In general conformal mappings are transformations of N dimensional spaces into themselves that preserve angles between intersecting curves. I demonstrate that complex conformal transformations have this property.

Let

$$f(z) = u(x, y) + iv(x, y) \tag{184}$$

be a conformal transformation from a neighborhood U of z_0 to a neighborhood V of $z'_0 = f(z_0)$. Let $c_1(t) = x_1(t) + iy_1(t)$ and $c_2(t) = x_2(t) + iy_2(t)$ be two curves in U that intersect at z_0 when $t = 0$.

The image of these two curves are two curves in V that intersect at z'_0 when $t = 0$:

$$c'_1(t) = u(x_1(t), y_1(t)) + iv(x_1(t), y_1(t)) \tag{185}$$

$$c'_2(t) = u(x_2(t), y_2(t)) + iv(x_2(t), y_2(t)) \tag{186}$$

To show that both of these curves intersect at the same angle let

$$c_1(t) - z_0 = r_1(t)e^{i\phi_1(t)} \tag{187}$$

$$c'_1(t) - z'_0 = r'_1(t)e^{i\phi'_1(t)} \tag{188}$$

Computing the complex derivative at $z = z_0$ along the curve $z_1(t)$ gives

$$\frac{dz'}{dz}(z_0) = \lim_{t\to 0} \frac{c'_1(t) - z'_0}{c_1(t) - z_0} = \lim_{t\to 0} \frac{r'_1(t)}{r_2(t)} e^{i(\phi'_1(t) - \phi_1(t)} =$$

$$\frac{r'_1(0)}{r_1(0)} e^{i(\phi'_1(0) - \phi_1(0)} \tag{189}$$

I can also compute this same derivative along the second curve. Since the value of the complex derivative is independent of how z_0 is approached it follows that

$$\frac{r'_1(0)}{r_1(0)} e^{i(\phi'_1(0) - \phi_1(0)} = \frac{r'_2(0)}{r_2(0)} e^{i(\phi'_2(0) - \phi_2(0)} \tag{190}$$

Comparing the arguments of these two complex numbers gives

$$\phi'_1(0) - \phi_1(0) = \phi'_2(0) - \phi_2(0) \tag{191}$$

While it is possible to satisfy these equations by adding multiples of 2π, that does not change the position of the tangent curve in the plane.

Equation (191) can be written as

$$\phi_1'(0) - \phi_2'(0) = \phi_1(0) - \phi_2(0) \tag{192}$$

which show that the two curves in U and their images in V intersect with the same angle, $\phi_1(0) - \phi_2(0)$.

Application : Heat equation

Consider thermal energy of a two dimensional solid of volume V. If we ignore thermal expansion and assume that the heat capacity c per unit volume of the solid is constant, then the thermal enegy of a small volume ΔV is $\Delta E = cT\Delta V$. Integrating this over a finite volume V gives the thermal energy in the volume:

$$E = \int_V cTdV. \tag{193}$$

If there are no sources of heat in the volume, the change in thermal energy in the volume is due to the heat transported across the boundary of the volume V. This can be expressed in terms of the heat current $\vec{J}$ by the equation

$$\frac{d}{dt}\int cTdV = -\int \vec{J}\cdot d\vec{S}. \tag{194}$$

The heat current is proportional to the temperature gradient, and points in the opposite direction of the gradient. This gives

$$\vec{J} = -\kappa\vec{\nabla}T \tag{195}$$

where the constant κ depends on the conductivity of the plate. Using the divergence theorem gives

$$\frac{d}{dt}\int cTdV = \int \kappa\vec{\nabla}T\cdot d\vec{S} =$$

$$\frac{d}{dt}\int cTdV = \int \kappa\nabla^2 TdV. \tag{196}$$

Since the difference

$$\int [\frac{\partial T}{\partial t} - (\kappa/c)\nabla^2 T]dV = 0 \tag{197}$$

must vanish for all volumes, it follows that the temperature is a solution of the heat equation

$$\frac{\partial T}{\partial t} = (\kappa/c)\nabla^2 T. \tag{198}$$

Once the system has reached a steady state condition the left side of this equation becomes 0 and then the temperature is given by a solution of Laplace's equation subject to the appropriate boundary conditions.

Conformal mapping techniques can be used to transform complex boundary conditions into simpler ones.

Let us assume that we have a plate in the quarter of the complex plane $x > 0$ and $y > 0$. Assume that the boundary along the y axis is maintained at temperature $T = 0$. On the x axis, for $0 \leq x \leq 1$ the plate is insulated so $(dT/dy) = 0$, and for $x > 1$ it is maintained at $T = 1$. When the system reaches a steady state the temperature of the plate will satisfy Laplace's equation

$$\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} = 0 \qquad x, y > 0 \tag{199}$$

subject to the boundary conditions

$$T(0, y) = 0, \qquad \frac{\partial T}{\partial y}(x, 0) = 0 \qquad 0 \leq x \leq 1, \qquad T(x, 0) = 1 \qquad x \geq 1 \tag{200}$$

These boundary conditions are fairly complex. I show how this problem can be solved using conformal mapping techniques.

Let $z = x + iy$ and $w = u + iv$ and let

$$z = \sin(w) \tag{201}$$

$$x + iy = \sin(u + iv) = \sin(u)\cosh(v) + i\cos(u)\sinh(v) \tag{202}$$

$$x = \sin(u)\cosh(v) \qquad y = \cos(u)\sinh(v) \tag{203}$$

The boundary $x = 0$ corresponds to $u = 0$. The boundary component $y = 0$, $0 \leq x \leq 1$ corresponds to $v = 0$ and $0 \leq u \leq \pi/2$, and the region $y = 0$, $x \geq 1$ corresponds $u = \pi/2$ and $v > 0$. Thus the quarter plane is mapped into a half infinite strip of width $\pi/2$.

The boundary conditions on the strip are simpler

$$T = 0 \qquad \text{for} \qquad u = 0, \quad v > 0 \tag{204}$$

$$T = 1 \qquad \text{for} \qquad u = \frac{\pi}{2}, \quad v > 0 \tag{205}$$

$$\frac{dT}{dv} = 0 \qquad \text{for} \qquad 0 \leq u \leq \frac{\pi}{2}, \quad v = 0 \tag{206}$$

A simple solution of Laplace's equation in this region subject to these boundary conditions is

$$T(u, v) = U(u, v) = \frac{2}{\pi} u \tag{207}$$

The desired solution to the original problem is

$$T(x, y) = U(u(x, y)), v(x, y)) = \frac{2}{\pi} u(x, y) \tag{208}$$

This means the we have to solve for u as a function of x and y.

$$x^2 = \sin^2 u \cosh^2 v \qquad y^2 = \cos^2 u \sinh^2 v \tag{209}$$

$$\frac{x^2}{\sin^2 u} - \frac{y^2}{\cos^2 u} = \cosh^2 v - \sinh^2 v = 1 \tag{210}$$

We can solve this for $u(x, y)$:

$$1 = \frac{x^2}{\sin^2 u} - \frac{y^2}{\cos^2 u} = \frac{x^2}{\sin^2 u} - \frac{y^2}{1 - \sin^2 u} \tag{211}$$

This can be solved by multiplying through by $(1 - \sin^2 u) \sin^2 u$

$$(1 - \sin^2 u) \sin^2 u = x^2 (1 - \sin^2 u) - y^2 \sin^2 u \tag{212}$$

This can be expressed as a quadratic equation for $\sin^2(u)$:

$$(\sin^2 u)^2 - (1 + x^2 + y^2)(\sin^2 u) + x^2 = 0 \tag{213}$$

$$\sin^2(u) = \frac{1 + x^2 + y^2}{2} \pm \frac{1}{2}\sqrt{(1 + x^2 + y^2)^2 - 4x^2} \tag{214}$$

The root with the $-$ sign vanishes when $x = y = 0$, which is consistent with $x + iy = \sin(u + iv)$. Using some more algebra it is possible write the right side as

$$\frac{1}{4}(\sqrt{(1 + x)^2 + y^2} - \sqrt{(1 - x)^2 + y^2})^2 \tag{215}$$

This leads to

$$T(x, y) = \frac{2}{\pi} \text{Arcsin}[\frac{1}{2}(\sqrt{(1 + x)^2 + y^2} - \sqrt{(1 - x)^2 + y^2})] \tag{216}$$

In this problem the solution to the transformed problem was the real part of
$$f(z') = \frac{2}{\pi} z' = \frac{2}{\pi}(u + iv) \tag{217}$$

Homographic transformations

The class of conformal transformations that arise from the linear relations
$$czz' = dz' - az - b \tag{218}$$
will be used next semester to generate new solutions of the hypergeometric differential equation from existing solutions. These transformations are called homographic transformations.

It is customary to write this relation in solved form
$$z' = \frac{az + b}{cz + d} \tag{219}$$
Direct calculation shows that
$$\frac{dz'}{dz} = \frac{a}{cz + d} - \frac{(az + d)c}{(cz + d)^2} = \frac{ad - bc}{(cz + d)^2} \tag{220}$$
which will be non-zero if $ad - bc \neq 0$. It is conformal in the complex plane for $z \neq -d/c$. One can add a point at infinity to the complex plane, then one can interpret $z = -d/c$ as the point that gets mapped to complex infinity.

Homographic transformations can be generated from three elementary Homographic transformations corresponding to

Translations
$$z \to z' = z + e \tag{221}$$

Changes of scale
$$z \to z' = ez \tag{222}$$

Inversions
$$z \to z' = \frac{1}{z} \tag{223}$$

To use these to generate a general homographic transformation consider the following sequence of transformations

$$z \rightarrow z_1 = z + \frac{d}{c} \tag{224}$$

$$z_1 \rightarrow z_2 = c^2 z_1 \tag{225}$$

$$z_2 \rightarrow z_3 = \frac{1}{z_2} \tag{226}$$

$$z_3 \rightarrow z_4 = (bc - ad) z_3 \tag{227}$$

$$z_4 \rightarrow z' = \frac{a}{c} + z_4 \tag{228}$$

Making successive substitutions gives

$$z' = \frac{a}{c} + z_4 = \frac{a}{c} + (bc - ad) z_3 =$$

$$\frac{a}{c} + (bc - ad)\frac{1}{z_2} = \frac{a}{c} + (bc - ad)\frac{1}{c^2 z_1} =$$

$$\frac{a}{c} + (bc - ad)\frac{1}{c^2(z + d/c)} = \frac{a}{c} + (bc - ad)\frac{1}{c(cz + d)} =$$

$$\frac{cza + ad + bc - ad}{c(cz + d)} = \frac{az + b}{(cz + d)} \tag{229}$$

Translations, changes of scale, and inversions have the property that they map circle in the complex plane to circles (or lines in the in the case of inversions). It is customary to consider a line as a circle through the point at infinity.

These transformations all involve functions that are analytic in the complex plane (except for the inversion at $z = 0$). Analyticity can be established using the Cauchy Riemann equations or by directly using the definitions.

Since each of the elementary transformations map circles to circles, and every homographic transformations can be expressed as a product of a finite number of elementary transformations, it follows that every homographic transformations maps circles to circles.

It remains to show that each of the elementary transformations map circles to circles.

The equation for a circle of radius R and center c in the complex plane is

$$|z - c| = R \tag{230}$$

If $z \rightarrow z' = z + e$ then

$$|z' - (e + c)| = R \tag{231}$$

is a circle of radius R centered at $e + c$. If $z \rightarrow z' = ez$ then

$$|z' - ec| = R|e| \tag{232}$$

which is a circle with center ec and radius $|e|R$.

I will leave it as a homework exercise to show that inversions also map circles to circles (circles through the origin get mapped to lines = circles through the point at infinity)

0.9 Lecture 9

These is one more useful property of of homographic transformations. Consider the composition of two successive homographic transformations

$$z' = \frac{az+b}{cz+d}$$

$$z'' = \frac{a'z'+b'}{c'z'+d'} = \frac{a'\frac{az+b}{cz+d}+b'}{c'\frac{az+b}{cz+d}+d'} =$$

$$\frac{a'(az+b)+b'(cz+d)}{c'(az+b)+d'(cz+d)} = \frac{(a'a+b'c)z+(a'b+b'd)}{(c'a+d'c)z+(c'b+d'd)} \tag{233}$$

This expression is another homographic transformation.

To get more insight into this transformation consider the matrix product

$$\begin{pmatrix} a'' & b'' \\ c'' & d'' \end{pmatrix} = \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a'a+b'c & a'b+b'd \\ c'a+d'c & c'b+d'd \end{pmatrix}$$

Inspection of these equations show that the transformation of the coefficients of successive homographic transformations transform exactly like multiplication of complex 2×2 matrices. The condition $ab - cd \neq 0$ means that the matrices have inverses.

We have a $1-1$ correspondence between complex invertible 2×2 matrices and homographic transformations. Under this correspondence matrix multiplication gets mapped into composition of homographic transformations.

A group is a set S with a product $\cdot$ such that

1 The product of two elements of S is an element of S.

2 The product is associative: $(a \cdot b) \cdot c = a \cdot (b \cdot c)$.

3 There is an identity element e satisfying $a \cdot e = e \cdot a = a$ for every a in S.

4 Every element a of S has an inverse a^{-1} in S satisfying $aa^{-1} = a^{-1}a = e$.

Both the complex 2×2 matrices with non-zero determinant under matrix multiplication and the set of homographic transformations under composition of functions are groups. The set of complex 2×2 matrices with non-zero determinant under matrix multiplication is a group called $GL(2, \mathbb{C})$ (complex general linear group in 2 dimensions).

An invertible mapping ϕ from a group G_1 to G_2 that satisfies

$$\phi(e_1) = e_2 \tag{234}$$

$$\phi(a \cdot b) = \phi(a) \cdot \phi(b) \tag{235}$$

is called a group isomorphism. What we have shown it that the group of homographic transformation in the complex plane is isomorphic to the group $GL(2, \mathbb{C})$.

The nice feature about this relation is that any property of $GL(2, \mathbb{C})$ translates to a property of homographic transformations.

The Cauchy Goursat Theorem:

The Cauchy Goursat theorem is the most important theorem in complex analysis. The precise statement of the theorem is

Cauchy-Goursat Theorem: Let C denote a piecewise regular closed curve in the complex plane. Let $f(z)$ be analytic on the curve C and within the region enclosed by C. Then

$$\int_C f(z) dz = 0. \tag{236}$$

Remarks:

1. The integral around a closed curve C is sometimes denoted by

$$\oint_C f(z) dz. \tag{237}$$

2. Analyticity on C and in the interior are essential requirements.

3. Cauchy proved this assuming that $\frac{df}{dz}$ was continuous. Goursat's proved the theorem without this assumption.

To prove the Cauchy Goursat theorem first note that for any integer $n \geq 0$

$$z^n = \frac{1}{n+1} \frac{d}{dz} z^{n+1}. \tag{238}$$

It follows that

$$\oint z^n dz = \frac{1}{n+1}\oint \frac{d}{dz} z^{n+1} dz = \frac{z_0^{n+1}}{n+1} - \frac{z_0^{n+1}}{n+1} = 0 \tag{239}$$

where z_0 is the starting and ending point of the curve C. Since complex integration is a linear operator this result extends to any polynomial $P(z)$ in z:

$$\oint_C P(z) dz = 0. \tag{240}$$

Next consider the general case. Imagine putting a grid of small squares over the region enclosed by C. Because of the linearity of the integral, the integral around the closed curve can be expressed as a sum of integrals around the interior squares and geometric figures bound by parts of C and interior squares

$$\oint_C f(z) dz = \sum_{n=1}^{N} \oint_{C_n} f(z) dz. \tag{241}$$

The orientation of the integrals are all taken in the same sense (counter-clockwise or clockwise). This works because the integrals on common boundaries are in opposite directions and cancel. What remains are a number of integrals over segments that add up to give C.

The triangle inequality gives

$$|\oint_C f(z) dz| \leq \sum_{n=1}^{N} |\oint_{C_n} f(z) dz| \tag{242}$$

The next step is to show that this sum can be made as small as desired. To do this choose an arbitrary $\epsilon > 0$. Let A be the set consisting of C and the part of the complex plane enclosed by C.

The existence of a derivative of $f(z)$ at $z_0 \in A$ means that

$$\frac{f(z) - f(z_0)}{z - z_0} = \frac{df}{dz}(z_0) + r(z, z_0) \tag{243}$$

where the remainder $r(z, z_0) \to 0$ as $z \to z_0$.

Specifically for every $\epsilon > 0$ there is a δ such that $|z - z_0| < \delta$ means that $|r(z, z_0)| \leq \epsilon$.

Subdivide A into a grid with sides of length l plus boundary terms. Ask if in the intersection of A with a given square of side l there is a z_0 such that

$|r(z, z_0)| < \epsilon$ for every z in this set. If a square fails to have this property subdivide the edges by a factor of two and ask the same question again.

Either this process terminates, and there is a finite square size where this inequality holds on all squares that intersect A, or there are an infinite sequence of nested smaller squares where in each square there is no z_0 where this holds for every z in the square. The intersection of this nested sequence of squares is a single point z_∞ that is in A and in all of the squares in this sequence. This point has the property that for any $l > 0$ there is point z satisfying $|z - z_\infty| \leq \sqrt{2}l$ where $r(z, z_\infty) > \epsilon$. The point z_∞ is in the set A, since it is in every one of the nested sets, each of which is in A.

It follows that no matter how small we choose δ at z_∞ there is at least one z with $|z - z_\infty| < \delta$ such that

$$|\frac{f(z) - f(z_\infty)}{z - z_\infty} - \frac{df}{dz}(z_0)| > \epsilon \tag{244}$$

which contradicts the assumption that $f(z)$ is differentiable at $z_\infty \in A$.

What this argument shows is that for every positive ϵ there is a finite grid size l such that $|z - z_0| < \sqrt{2}l$ implies $|r(z, z_0)| < \epsilon$

29:171 - Homework Assignment #3

1. Let $f(z) = e^z$. Let C be the curve in the complex plane that starts at the origin, $z = 0$, goes along the positive real axis to the point $z = 2$, and then proceeds in a straight line in positive imaginary direction from $z = 2$ to $z = 2 + 3i$.

   Calculate the contour integral

   $$\int_C f(z)dz$$

   directly. Check your answer by noting that $\frac{df}{dz}(z) = f(z)$

2. Show that the conformal mapping

   $$z \rightarrow z' = \frac{1}{z}$$

   maps the circle

   $$|z - a| = r$$

   into another circle. Find the origin and radius of the transformed circle. Determine the condition for the transformed circle to become a line (i.e. circle of infinite radius).

3. Show that any real valued analytic function is a constant.

4. If C is the circle $|z| = 1$, calculate the line integral (in the counter-clockwise direction) of

   $$\int_C \frac{dz}{z}$$

5. Consider the homographic transformation

   $$z' = \frac{az + b}{cz + d} \qquad bc - ad \neq 0$$

   Calculate the inverse transformation. Is it homographic?

6. Use Darboux's theorem to put a bound on the integral

   $$|\int_C \sin(z)dz|$$

   where C is the circle $|z| = 5$.

0.10 Lecture 10

We are now in a position to complete the proof of the Cauchy Goursat theorem.

Let z_{n0} be a point in each square where the remainder term is bounded by a fixed ϵ and note

$$f(z) = f(z_{0n}) + \frac{df}{dz}(z_{0n})(z - z_{0n}) + r(z, z_{0n})(z - z_{0n}) \tag{245}$$

Using (242) in (245) along with the triangle inequality again gives

$$|\oint_C f(z)dz| \leq$$

$$\sum_{n=1}^{N} |\oint_{C_n} (f_{z_{0n}} + \frac{df}{dz}(z_{0n})(z - z_{0n}))dz|$$
$$+ \sum_{n=1}^{N} |\oint_{C_n} (r(z, z_{0n})(z - z_{0n}))dz| \tag{246}$$

The integrands in the first term are polynomials, which have been shown to have vanishing integral.

Darboux's theorem can be used to put upper bounds on the integrals over the small squares. The path length of C_n is $4l$ for squares, and less than $4l + s_n$ where s_n is the length of the C in the square, and l is the edge length of a square.

The integrand is bounded by

$$|(r(z, z_{0n})(z - z_{0n}))| \leq \sqrt{2} l\epsilon \tag{247}$$

This means that the n^{th} integral is bounded by

$$|\oint_{C_n} (r(z, z_{0n})(z - z_{0n}))dz| \leq \epsilon\sqrt{2} l(4l + s_n) \tag{248}$$

Summing over all squares

$$\sum \epsilon\sqrt{2} 4l^2 \leq \epsilon 4\sqrt{2} A \tag{249}$$

where A is the area of a rectangle bounding the curve C and

$$\sum \epsilon\sqrt{2} l s_n \leq \epsilon\sqrt{2} l S \tag{250}$$

where S is the length of C. This gives

$$|\oint_C f(z)dz| \leq \epsilon \times (4\sqrt{2}A + \sqrt{2} l S) \tag{251}$$

Since ϵ was arbitrary, the right hand side can be taken smaller than any given positive quantity. This requires

$$\oint_C f(z)dz = 0 \tag{252}$$

which completes the proof of the Cauchy Goursat theorem.

I will now use Cauchy's theorem to prove some useful results.

For the first result assume that the closed curve C in Cauchy's theorem can be expressed as a sum two curves, C_1 and C_2. By this I mean that the end of C_1 is the beginning of C_2, and the end of C_2 is the beginning of C_1. It follows

$$0 = \oint_C f(z)dz = \int_{C_1+C_2} f(z)dz = \int_{C_1} f(z)dz + \int_{C_2} f(z)dz =$$

$$\int_{C_1} f(z)dz - \int_{-C_2} f(z)dz \tag{253}$$

where $-C_2$ indicates the curve that follows the same path as C_2 but goes in the opposite direction. It has the same start and end point as C_1. Therefore

$$\int_{C_1} f(z)dz = \int_{-C_2} f(z)dz \tag{254}$$

This shows that the integral of an analytic function along any two paths in the complex plane with the same starting and end points are independent of path provided $f(z)$ is analytic in the region bounded by both curves.

In developing further consequences of Cauchy's theorem it is useful to introduce a definition.

Definition: A region R in the complex plane is called simply connected if the interior of every closed curve in R contains only points of R.

Definition: A region R in the complex plane is called multiply connected if it is not simply connected.

Theorem: Let $f(z)$ be analytic throughout a simply connected region R and let $z \cup C = \{\emptyset\}$. Then

$$\frac{1}{2\pi i}\oint_C \frac{f(z')}{z'-z}dz' = \begin{cases} f(z) & : \quad z \quad \text{interior to } C \\ 0 & : \quad z \quad \text{exterior to } C \end{cases} \tag{255}$$

Proof:$\frac{1}{z'-z}$ is analytic as a function of z' except at the point $z' = z$. The product $\frac{f(z')}{z'-z}$ is the product of analytic functions in z', except when $z' = z$.

If $z \notin R$ then $\frac{f(z')}{z'-z}$ is analytic on C and at every point interior to C so by Cauchy's theorem

$$\oint_C \frac{f(z')}{z'-z}dz' = 0 \tag{256}$$

The same result follows if $z \in R$ but strictly outside of the curve C.

If z is in the interior of C, using Cauchy's theorem it is possible to replace C by small circle of radius r about z such that the circle in bounded by C. This can be done by drawing a line from a point on the circle to a point on C that does not intersect z. The path that starts at the line goes around C, extends along the line, goes around the small circle in the opposite direction, and returns to the starting point along the line does not bound z so the value of this integral is zero. The two integrals along the line are in the opposite direction and cancel. The integral over C and the circle in the opposite direction cancel. If the direction of the integral around the circle is reversed the sign changes and it must be identical to the integral around C in the same direction.

Let $\odot$ denote the circle. Then

$$\frac{1}{2\pi i}\oint_C \frac{f(z')}{z'-z}dz' =$$

$$\frac{1}{2\pi i}\oint_\odot \frac{f(z')}{z'-z}dz' =$$

$$\frac{f(z)}{2\pi i}\oint_\odot \frac{1}{z'-z}dz' + \frac{1}{2\pi i}\oint_\odot \frac{f(z')-f(z)}{z'-z}dz' \tag{257}$$

Since $f(z')$ is analytic at z it is continuous at z. This means that for every $\epsilon >$ there is an $r >$ such that $|z'-z| < r \Rightarrow |f(z) - f(z')| < \epsilon$. Choose the radius of $\odot$ to be r. Then by Darboux's theorem

$$|\frac{1}{2\pi i} \oint_{\odot} \frac{f(z') - f(z)}{z' - z} dz'| \leq \frac{1}{2\pi} \frac{\epsilon}{r} 2\pi r = \epsilon \tag{258}$$

On the other hand if we let $z' = z + re^{i\phi}$

$$\frac{f(z)}{2\pi i} \oint_{\odot} \frac{1}{z' - z} dz' =$$

$$\frac{f(z)}{2\pi i} \int_0^{2\pi} \frac{1}{e^{i\phi}} \frac{dz'}{d\phi} d\phi =$$

$$\frac{f(z)}{2\pi i} \int_0^{2\pi} \frac{1}{re^{i\phi}} ire^{i\phi} d\phi = f(z) \tag{259}$$

What we have shown is that for any $\epsilon > 0$ we have

$$|f(z) - \frac{1}{2\pi i} \oint_C \frac{f(z')}{z' - z} dz'| < \epsilon \tag{260}$$

where $\epsilon > 0$ is arbitrary. Since the left hand side is independent of ϵ it follows that

$$f(z) = \frac{1}{2\pi i} \oint_C \frac{f(z')}{z' - z} dz' \tag{261}$$

as desired.

Note that implicit in the proof of this result is the assumption that C is a counter clockwise curve. It C was in the opposite direction the formula would have a $-$ sign.

0.11 Lecture 11

The implication of this theorem is that if $f(z)$ is known to be analytic in a simply connected region then the knowledge of $f(z)$ on a curve uniquely fixes the value of $f(z)$ at all points interior to the curve.

The value of $f(z)$ on the curve can be thought of as boundary conditions that determine the solution of Laplace's equation in the interior.

The Cauchy integral formula can be used to prove that analytic functions are infinitely differentiable. To show this let C be a regular curve in a simply connected region. Let $g(z)$ be continuous on C and on the interior of C. Define the function

$$f(z) = \frac{1}{2\pi i}\oint_C \frac{g(z')}{z'-z}dz' \tag{262}$$

Consider the difference

$$\Delta := \frac{f(z+\Delta z)-f(z)}{\Delta z} - \frac{1}{2\pi i}\oint_C \frac{g(z')}{(z'-z)^2}dz' \tag{263}$$

The assumption the $z \notin C$ implies

$$|\Delta| = |\frac{1}{2\pi i}\oint_C g(z')[\frac{1}{z'-z-\Delta z} - \frac{1}{z'-z} - \frac{1}{(z'-z)^2}]dz'| \tag{264}$$

Elementary algebra implies

$$[\frac{1}{z'-z-\Delta z} - \frac{1}{z'-z} - \frac{1}{(z'-z)^2}] = \frac{(\Delta z)^2}{(z'-z)^2(z'-z-\Delta z)}. \tag{265}$$

Using (265) in (264) gives

$$|\Delta| = \frac{|\Delta z|}{2\pi}|\int_C \frac{g(z')}{(z'-z)^2(z'-z-\Delta z)}|. \tag{266}$$

The integral is bounded by Darboux's theorem. This shows that $\frac{df}{dz}(z')$ exists for all z in the interior of the region bounded by C. More important, it gives an integral expression for the derivative:

$$\frac{df}{dz}(z) = \frac{1}{2\pi i}\oint_C \frac{g(z')}{(z'-z)^2}dz' \tag{267}$$

Now we use mathematical induction to show that all derivatives of $f(z)$ exist in this region. The induction assumption is that the $n-th$ derivative exists and is given by

$$f^{(n)}(z) := \frac{d^n f}{dz^n}(z) = \frac{n!}{2\pi i} \oint_C \frac{g(z')}{(z'-z)^{n+1}} dz' \tag{268}$$

Assuming (268) I show that the same equation holds for $n \to n+1$. Consider

$$\Delta = \frac{f^{(n)}(z+\Delta z) - f^{(n)}(z)}{\Delta z} - \frac{(n+1)!}{2\pi i} \oint_C \frac{g(z')}{(z'-z)^{n+2}} dz' =$$

$$\frac{n!}{2\pi i} \oint_C g(z')[\frac{1}{(z'-z-\Delta z)^{n+1}} - \frac{1}{(z'-z)^{n+1}} - \frac{n+1}{(z'-z)^{n+2}}]dz' \tag{269}$$

To estimate this note

$$\frac{1}{(z'-z-\Delta z)^{n+1}} = \left(\frac{1}{z'-z} + \frac{1}{z'-z} \Delta z \frac{1}{z'-z-\Delta z} \right)^{n+1} =$$

$$\sum_{k=0}^{n+1} \frac{(n+1)!}{k!(n+1-k)!} \frac{1}{(z'-z)^k} \left(\frac{1}{z'-z} \Delta z \frac{1}{z'-z-\Delta z} \right)^{n+1-k} =$$

$$\frac{1}{(z'-z)^{n+1}} + \frac{(n+1)}{(z'-z)^{n+1}} \Delta z \frac{1}{z'-z-\Delta z} +$$

$$\sum_{k=2}^{n+1} \frac{(n+1)!}{k!(n+1-k)!} \frac{1}{(z'-z)^k} \left(\frac{1}{z'-z} \Delta z \frac{1}{z'-z-\Delta z} \right)^{n+1-k} \tag{270}$$

Using (270) in (269) gives

$$\Delta =$$

$$\frac{n!}{2\pi i} \oint_C g(z')[\frac{(n+1)}{(z'-z)^{n+1}} \frac{1}{z'-z-\Delta z} - \frac{n+1}{(z'-z)^{n+2}} +$$

$$\sum_{k=2}^{n+1} \frac{(n+1)!}{k!(n+1-k)!} \frac{1}{(z'-z)^k} \frac{1}{\Delta z} \left(\frac{1}{z'-z} \Delta z \frac{1}{z'-z-\Delta z} \right)^{n+1-k}]dz' \tag{271}$$

The last term is of the form Δz times a quantity that is bounded by Darboux's theorem. The other two terms can be estimated by noting

$$\frac{(n+1)}{(z'-z)^{n+1}} \frac{1}{z'-z-\Delta z} - \frac{n+1}{(z'-z)^{n+2}} =$$

$$\frac{(n+1)}{(z'-z)^{n+1}}[\frac{1}{z'-z-\Delta z} - \frac{1}{z'-z}] =$$
$$\frac{\Delta z(n+1)}{(z'-z)^{n+2}}\frac{1}{z'-z-\Delta z} \tag{272}$$

This leads to a another term of the form Δz multiplied by a quantity bounded by Darboux's theorem. Letting $|\Delta z| \to 0$ shows the $n+1$ derivative exists and reproduces the formula assumed in the induction assumption.

Thus unlike calculus of real functions, if a function is analytic in a region then all complex derivatives of the function exist in that region.

Equation (**??**) is an example of an integral representation of an analytic function. This result can be generalized as follows:

Theorem: Given the integral representation

$$f(z) := \int_C K(z,z')g(z')dz; \qquad z \in R \tag{273}$$

then the complex derivative of $f(z)$ exists and is given by

$$\frac{df}{dz}(z) := \int_c \frac{\partial K(z,z')}{\partial z} g(z')dz; \qquad z \in R \tag{274}$$

provided the following condition are satisfied

1. For $z \in R$, $K(z,z')$ an analytic of z for any $z' \in C$
2. For $z \in R$, $K(z,z')g(z')$ an a continuous function of z'

To prove this note the analyticity in z allows us to use the Cauchy integral formula

$$K(z,z') = \frac{1}{2\pi i}\int_{C'} \frac{K(t,z')}{t-z}dt \qquad C' \subset R \tag{275}$$

Next we use the definition along with the above to write

$$f(z) := \frac{1}{2\pi i}\int_C [\int_{C'} \frac{K(t,z')}{t-z}dt]g(z')dz'; \qquad z \in R \tag{276}$$

Since $\frac{K(t,z')}{t-z}g(z')$ is continuous for $z \neq t$ the order of the integrals can be interchanged (recall these two complex integrals can be converted to ordinary integrals where continuity of the integrand is sufficient to interchange the order of the integrals:

$$f(z) := \frac{1}{2\pi i}\int_{C'}[\int_C \frac{K(t,z')g(z')}{t-z}dz']dt; \qquad z \in R \tag{277}$$

Using (267) gives

$$\frac{df}{dz}(z) := \frac{1}{2\pi i}\int_{C'}[\int_C \frac{K(t,z')g(z')}{(t-z)^2}dz']dt; \qquad z \in R \tag{278}$$

Note that (267) gives

$$\frac{\partial K(z,z')}{\partial z} = \frac{1}{2\pi i}\int_{C'} \frac{K(t,z')}{(t-z)^2}dt \qquad C' \subset R \tag{279}$$

If we interchange the order of integration in (279) and use (278) we get

$$\frac{df}{dz}(z) := \frac{1}{2\pi i}\int_C \frac{\partial K(z,z')}{\partial z}g(z')dz' \tag{280}$$

0.12 Lecture 12

Next I derive some properties of analytic functions

Theorem 12.1: The modulus of an analytic function cannot have a local maximum within the region of analyticity.

proof: Let z_0 be a regular point z_0 of $f(z)$. Then for small enough r, $f(z)$ is analytic in the region $|z - z_0| \leq r$. Denote the circle $|z - z_0| \leq r$ by $\odot$. The Cauchy integral formula implies

$$f(z_0) = \frac{1}{2\pi i} \int_{\odot} \frac{f(z)}{z - z_0} dz. \tag{281}$$

Applying Darboux's inequality to this integral gives

$$|f(z_0)| \leq \max|f(z)|_{z \in \odot}. \tag{282}$$

It follows that there must be at least one point z on $\odot$ where $|f(z)| \geq |f(z_0)|$.

This must also true on every circle of radius $r' < r$, since r was chosen arbitrarily. This means that there is no neighborhood of z_0 where $f(z_0)$ is a local maximum. This proves the theorem. ∎

If we let $g(z) = 1/f(z)$ and note that $g(z)$ is analytic in a neighborhood of $z = z_0$ provided $f(z_0) \neq 0$, then $g(z_0)$ cannot be a local maximum of $g(z)$. This is equivalent to saying that $f(z_0)$ can not be a local minimum.

The same applies separately to the real and imaginary parts of an analytic function since

$$|e^{f(z)}| = e^{\Re(z)} \qquad |e^{if(z)}| = e^{\Im(z)} \tag{283}$$

and e^x is increasing.

Theorem 12.2: A bounded entire function must be constant.

proof: Since $f(z)$ is entire it is possible to write

$$\frac{df}{dz}(z) = \frac{1}{2\pi i} \int_{\odot} \frac{f(z')}{(z' - z)^2} dz'. \tag{284}$$

Choosing $\odot$ to have a radius R and using Darboux's theorem gives

$$|\frac{df}{dz}(z)| = \frac{\max|f(z')|_{z' \in \odot}}{R} \leq \frac{\max|f(z')|}{R}. \tag{285}$$

Since the right hand side can be made as small as desired by increasing R and the left hand side is independent of R, it follows that

$$\frac{df}{dz}(z) = 0 \tag{286}$$

which means that $f(z) = constant$.■

The implication of this result is that any analytic function that are bounded at ∞ necessarily have a singularity on the complex plane.

By looking are higher derivatives this result can be generalized to show that entire functions bounded by polynomials have to be polynomials. This will be the subject of homework.

Cauchy's theorem also has a converse, called Morera's theorem.

Theorem 12.3(Morera's Theorem): If the integral

$$\int_C f(z)dz \tag{287}$$

of a continuous function $f(z)$ vanishes for any closed contour C of a region R, then $f(z)$ is analytic in R.

proof: The proof of Morerra's theorem is by construction. Define the function

$$F(z) = \int_a^z f(z')dz'. \tag{288}$$

Since the integral is independent of path, this function is well defined and only depends on the choice of z and a in R. Compute the derivative of $F(z)$ directly from the definition

$$\frac{F(z+\Delta z) - F(z)}{\Delta z} = \frac{1}{\Delta z}[\int_a^{z+\Delta z} f(z')dz' - \int_a^z f(z')dz'] =$$

$$\frac{f(z)}{\Delta z}\int_z^{z+\Delta z} dz' + \frac{1}{\Delta z}\int_z^{z+\Delta z}[f(z') - f(z)]dz'. \tag{289}$$

The first term is $f(z)$ while Darboux's theorem implies that the second is bounded by

$$\max(|f(z') - f(z)|)|z' - z| < \Delta \tag{290}$$

which vanishes in the limit that $|\Delta_z| \to 0$ by the assumed continuity of $f(z)$. This shows that

$$f(z) = \frac{dF}{dz}(z). \tag{291}$$

Since $F(z)$ has all derivatives, it follows that $f(z)$ is analytic.■

Next I discuss manipulations of sequences of analytic functions. Consider a sequence of analytic functions $f_1(z), f_2(z), \cdots$ defined in some region such that

$$f(z) = \sum_{n=1}^{\infty} f_n(z) \tag{292}$$

converges uniformly for all z in some curve C. Then assuming all integrals exist

$$\int_C f(z) = \int_C \sum_{n=1}^{\infty} f_n(z) = \sum_{n=1}^{\infty} \int_C f_n(z) \tag{293}$$

This means that whenever we have uniform convergence the order of the sum and integral can interchanged. To prove this let

$$s_n(z) = \sum_{k=1}^{n} f_k(z) \tag{294}$$

be the partial sum consisting of the first n terms.

From Darboux's theorem

$$|\int_C (f(z) - s_n(z))dx| \leq \max_{z\in C}|f(z) - s_n(z)|L \tag{295}$$

where L is the length of the curve. Uniform comvergence of the series means that for a given ϵ it is possibe to choose n large enough so

$$|f(z) - s_n(z)| < \epsilon \tag{296}$$

independent of z. On the other hand

$$\int s_n(z)dz = \sum_{k=1}^{n} \int f_k(z)dz \tag{297}$$

which shows that

$$|\int_C f(z)dz - \sum_{k=1}^{n} \int_C f_k(z)dz| \to 0 \tag{298}$$

as $n \to \infty$. This means

$$\int_C f(z)dz = \sum_{k=1}^{\infty} \int_C f_k(z)dz. \tag{299}$$

29:171 - Homework Assignment #4

1. Let $f(z)$ and $g(z)$ be entire functions. Assume that they agree on a line segment of the real axis. Show that they agree for all z.

2. If $f(z)$ is analytic and non-vanishing in a region R, and continuous in R and its boundary, show that $|f|$ assumes its minimum and maximum values on the boundary of R.

3. Show that if $f(z)$ is entire and $|f(z)| < C|z^n|$ for sufficiently large values of $|z|$, where C is a constant, then $f(z)$ must be a polynomial of degree $\leq n$.

4. Show that any non-constant polynomial in z has at least one complex root.

5. Let $f(t)$ be continuous for $t \in [a, b]$. Show

$$F(z) = \int_a^b e^{izt} f(t) dt$$

   is an entire function. Calculate the derivative.

6. Prove that for charge free two-dimensional space the value of the electrostatic potential at any point is equal to the average of the potential over the circumference of any circle centered on the point. You may assume that the potential is the real part of an analytic function (the electrostatic potential in a charge free region is a solution of Laplace's equation).

0.13 Lecture 13

We can use the previous result to show that $f(z)$ is analytic. To show this apply Morera's theorem to the previous result:

$$\oint_C f(z)dz = \sum_{n=1}^{\infty} \oint_C f_n(z)dz = 0 \tag{300}$$

for any closed curve in R. This implies that $f(z)$ is analytic.

Since the function $f(z)$ is analytic it follows that

$$\frac{df}{dz}(z) = \frac{1}{2\pi i} \frac{f(z)}{(z-z_0)^2} \tag{301}$$

By the uniform convergence of the partial sums of functions f_n we have uniform convergence of the series on an interior point z_0 of C

$$\frac{1}{2\pi i} \frac{f(z)}{(z-z_0)^2} = \sum_{n=1}^{\infty} \frac{1}{2\pi i} \frac{f_n(z)}{(z-z_0)^2} \tag{302}$$

If we integrate this it follows that

$$\frac{1}{2\pi i} \int \frac{f(z)}{(z-z_0)^2} = \sum_{n=1}^{\infty} \frac{1}{2\pi i} \int \frac{f_n(z)}{(z-z_0)^2} \tag{303}$$

which is equivalent to the uniform convergence of

$$\frac{df}{dz}(z) = \sum_{n=1}^{\infty} \frac{df_n}{dz}(z) \tag{304}$$

This generalizes to all higher derivatives by induction.

The two important lessons from these exercises are that it is always possible to change the order of sums and integrals provided the sums converge uniformly on the path of integration (and the path has a finite length). The second important observation is that is a series of analytic functions converges uniformly, the resulting function is analytic and the sum of the derivatives of each of the terms in the series is uniformly convergent and converges to the derivative of the function.

Having an infinite number of derivatives does not guarantee that a function has a convergent Taylor series. A standard example is

$$f(x) = e^{-\frac{1}{x^2}} \tag{305}$$

All derivatives of this function exist at $x = 0$ and are zero. This function can be computed for any $x \neq 0$ and the result is not zero. It follows that the radius of convergence of the Taylor series is zero.

When the function is analytic we can use the Cauchy integral representation to show that the Taylor series actually converges.

Theorem 13.1: Taylor's Theorem. Let $f(z)$ be analytic in a region R and let $\odot$ be a circle of radius r centered at z_0 in the region R. Then for any point z in the interior of $\odot$ $f(z)$ is equal to the uniformly convergent power series

$$f(z) = \sum_{n=1}^{\infty} a_n (z - z_0)^n \tag{306}$$

where

$$a_n = \frac{1}{n!}\frac{d^n}{dz^n} f(z)_{|z=z+0} = \frac{1}{2\pi i}\int_{\odot} \frac{f(z')}{(z'-z_0)^{n+1}} dz' \tag{307}$$

proof:

The Cauchy integral formula gives for the interior points of $\odot$

$$\begin{aligned} f(z) = \frac{1}{2\pi i}\int_{\odot} \frac{f(z')}{z'-z} dz' = \\ \frac{1}{2\pi i}\int_{\odot} \frac{f(z')}{z'-z_0+z_0-z} dz' \\ \frac{1}{2\pi i}\int_{\odot} \frac{f(z')}{z'-z_0}\frac{1}{1-\frac{z-z_0}{z'-z_0}} dz'. \end{aligned} \tag{308}$$

Since $z' \in \odot$ and z is an interior point of the circle of radius r it follows that

$$|\frac{z-z_0}{z'-z_0}| < \frac{|z-z_0|}{r} < 1 \tag{309}$$

$$\frac{1}{1-\frac{z-z_0}{z'-z_)}} dz' = \sum_{n=0}^{\infty} \frac{(z-z_0)^n}{(z'-z_0)^n} \tag{310}$$

converges uniformly for all points z' on $\odot$. This gives

$$f(z) = \frac{1}{2\pi i}\int_{\odot} f(z')\sum_{n=0}^{\infty}\frac{(z-z_0)^n}{(z'-z_0)^{n+1}} \tag{311}$$

Because of the uniform convergence we can interchange the order of the sum and integral to get

$$f(z) = \sum_{n=0}^{\infty}\frac{1}{n!}\frac{n!}{2\pi i}\int_{\odot} f(z')\frac{(z-z_0)^n}{(z'-z_0)^{n+1}} \tag{312}$$

where we have previously shown that

$$\frac{d^n f}{dz^n}(z_0) = \frac{n!}{2\pi i}\int_{\odot}\frac{f(z')}{(z'-z_0)^{n+1}}dz' \tag{313}$$

which gives

$$f(z) = \sum_{n=0}^{\infty}\frac{1}{n!}\frac{d^n f}{dz^n}(z_0)(z-z_0)^n \tag{314}$$

Which is exactly the Taylor series about z_0.

Note that we have shown that this series converges for z in the interior of $\odot$. The circle $\odot$ was only required to be in the region of analyticity. It can be replaced by the largest circle centered at z_0 that is still in the region of analyticity and by the above proof the Taylor series will converge provided z is in this larger circle.

Taylor's theorem has a converse. Assume that

$$f(z) = \sum_{n=0}^{\infty} a_n (z-z_0)^n \tag{315}$$

is absolutely convergent in a neighborhood of z_0. By the ratio test convergence fails if

$$|\frac{a_{n+1}r^{n+1}}{a_n r^n}| > 1 \tag{316}$$

as $n \to \infty$. Convergence requires that $|a_n r^n| < A$ where A is a finite constant (if this is not true the ratio test fails). This means that there is a constant A with the property

$$|a_n| \leq \frac{A}{r^n}. \tag{317}$$

Choose z so $|z-z_0|<r$. Then

$$|\sum_{n=0}^{\infty} a_n(z-z_0)^n| \leq \sum_{n=0}^{\infty} |a_n(z-z_0)^n| \leq \tag{318}$$

$$\sum_{n=0}^{\infty} A|\frac{(z-z_0)}{r}|^n \leq \frac{A}{1-|\frac{(z-z_0)}{r}|}. \tag{319}$$

The means that the convergence is uniform for $|z-z_0|<r$.

It follows that the integral of $f(z)$ around any close curve in the region bounded by a circle of radius r about z_0 can be computed by interchanging the order of the sum and integral

$$\oint_C f(z)dz = \sum_{n=0}^{\infty} a_n \oint_C (z-z_0)^n dz = 0 \tag{320}$$

We can not apply Morera's theorem to conclude that $f(z)$ is analytic for z inside the circle of radius r.

What we have show is that at least locally any analytic function behaves like a convergent series in powers of z.

Next I consider a generalization of Taylor's theorem that can be applied to a function that is analytic in a region R bounded by two concentric circles. The function does not have to be analytic inside the inner circle.

Pick a point $z \in R$ that lies between the two circles. Consider a curve consisting of a large circle, a small circle, and a small circle between these to circles that surrounds z. Include a pairs of lines the connect the outer circle to the small circle and the inner circle to the small circle. THe combined path lies in R and the integral

$$\oint \frac{f(z')}{z'-z}dz' = 0 \tag{321}$$

by Cauchy's theorem.

Since the integrals over the lines cancel this means that the integral over the circle around z can be expressed as the difference of the integral over the large circle minus the integral over small circle

$$\oint_{\odot_z} \frac{f(z')}{z'-z}dz' = \oint_{\odot_L} \frac{f(z')}{z'-z}dz' - \oint_{\odot_S} \frac{f(z')}{z'-z}dz' \tag{322}$$

The integral around the circle around z if $2\pi i f(z)$ so we have

$$f(z) = \frac{1}{2\pi i}\oint_{\odot_L}\frac{f(z')}{z'-z}dz' - \frac{1}{2\pi i}\oint_{\odot_S}\frac{f(z')}{z'-z}dz' \tag{323}$$

Let z_0 be the center of the two concentric circles and write the above as

$$f(z) = \frac{1}{2\pi i}\oint_{\odot_L}\frac{f(z')}{z'-z_0+z_0-z}dz' - \frac{1}{2\pi i}\oint_{\odot_S}\frac{f(z')}{z'-z_0+z_0-z}dz' \tag{324}$$

On the large circle $|z_0-z'| > z_0-z$ while on the small circle $|z_0-z| > z_0-z'$. Use these inequalities to write the above as

$$f(z) = \frac{1}{2\pi i}\oint_{\odot_L}\frac{f(z')}{z'-z_0}\frac{1}{1-\frac{z-z_0}{z'-z_0}}dz' - \frac{1}{2\pi i}\oint_{\odot_S}\frac{f(z')}{z_0-z}\frac{1}{1-+\frac{z_0-z'}{z_0-z}}dz' \tag{325}$$

Using the uniform convergence of the geometric series this becomes

$$f(z) = \frac{1}{2\pi i}\oint_{\odot_L}\frac{f(z')}{z'-z_0}\sum_{n=0}^{\infty}\left(\frac{z-z_0}{z'-z_0}\right)^n dz' - \frac{1}{2\pi i}\oint_{\odot_S}\frac{f(z')}{z_0-z}\sum_{n=0}^{\infty}\left(\frac{z_0-z'}{z_0-z}\right)^n dz' \tag{326}$$

Since the sums are uniformly convergent of the respective circles we can interchange the order of the sum and the integral and the result still converges.

$$\begin{aligned} f(z) &= \sum_{n=0}^{\infty} a_n (z-z_0)^n + \sum_{n=0}^{\infty} b_{n+1}(z-z_0)^{-n-1} \\ &= \sum_{n=0}^{\infty} a_n (z-z_0)^n + \sum_{n=1}^{\infty}\frac{b_n}{(z-z_0)^n} \end{aligned} \tag{327}$$

where

$$a_n = \frac{1}{2\pi i}\oint_{\odot_L}\frac{f(z')}{(z'-z_0)^{n+1}}dz' \tag{328}$$

$$b_{n+1} = -\frac{1}{2\pi i}\oint_{\odot_S} f(z')(z'-z_0)^n dz' \tag{329}$$

The series is is called the Laurent series. It is clear from the proof that it converges when z lies between any pair of concentric circles in R.

What is different is that the convergence of the Laurent series is not restricted to simply connected sets. As in the case of Taylor's theorem, the proof of the theorem gives integral formulas to compute the coefficients of the series in terms of the function $f(z)$.

As an example, the function $f(z) = 1/z$ is analytic except at $z = 0$. If we consider a point $z_0 \neq$ this function has a convergent Taylor series for all z satisfying $|z - z_0| < |z_0|$. On the other hand it has a Laurent series that converges everywhere in the complex plane except at $z = 0$. Bu using the Laurent series the region where the series converges is much bigger.

We have primarily been considering functions that are analytic. The points where the function is not analytic are called singularities. It is useful to discuss both the structure of $0's$ and the singularities of analytic functions.
Definition: An analytic function $f(z)$ has a zero of order n at z_0 if

$$0 = f(z_0) = \frac{df}{dz}(z_0) = \frac{d^2 f}{dz^2}(z_0) = \cdots \frac{d^{n-1} f}{dz^{n-1}}(z_0) \tag{330}$$

and

$$\frac{d^n f}{dz^n}(z_0) \neq 0 \tag{331}$$

If $f(z)$ has a zero of order n then Taylor's theorem implies

$$f(z) = (z - z_0)^n h(z) \tag{332}$$

where $h(z)$ is analytic and does not vanish at $z = z_0$. By continuity it follows that $f(z)$ must be non-vanishing in a neighborhood of z_0.

Note also that if the function has a zero of infinite order then the function must be identically zero by Taylor's theorem. This means that the zeros of all analytic functions, except the zero function, are *isolated*.

If $f(z)$ has an isolated singularity at $z = z_0$ then we can always compute the Laurent series of $f(z)$ about this point. The function $f(z)$ is singular at z_0 if at least one of the Laurent coefficients $b_k \neq 0$.

The function $f(z)$ has a pole of order n if $b_n \neq 0$ and $b^k = 0$ for $k > n$.

The sum

$$\sum_{k=1}^{n} \frac{b_k}{(z - z_0)^k} \tag{333}$$

is called the principal part of $f(z)$ at the singular point $z = z_0$.

A pole of order 1 is called a simple pole.

A function $f(z)$ that is analytic, except at a set of isolated singularities, where the function has poles, is called meromorphic.

If an infinite number of the $b_n \neq 0$ in the Laurent series, then $f(z)$ is said to have an isolated essential singularity at $z = z_0$.

The function $f(z)$ behaves wildly in the neighborhood of an isolated essential singularity. This is illustrated by the next theorem.

0.14 Lecture 14

Theorem 14.1:(Weierstrass) If $f(z)$ has an isolated essential singularity at $z = z_0$ then for arbitrary positive numbers ϵ and δ and any complex number a one has

$$|f(z) - a| < \epsilon \tag{334}$$

for some z satisfying $|z - z_0| < \delta$

This theorem means that $f(z)$ oscillates so violently near z_0 it gets arbitrarily close to every complex number!

Since the singularity is isolated the Laurent series converges, with

$$b_n = \frac{1}{2\pi i}\int_C (z' - z_0)^{n-1} f(z')dz' \tag{335}$$

for a circle of radius r about z_0. Darboux's theorem gives

$$|b_n| \leq \frac{2\pi r}{2\pi} r^{n-1} |f| \tag{336}$$

where $|f|$ is the largest value of $|f(z)|$ on the circle.

If $f(z)$ is bounded in a neighborhood of z_0, then the above implies $b_n \to 0$ so $f(z)$ cannot be bounded if any of the $b_n, \quad n > 0$ are non-zero.

Pick an arbitrary complex number a. Either z_0 is an accumulation point of $f(z) - a$, in which case we done, or it is not. If z_0 is not an accumulation point of $f(z) - a$ then there is an $\eta > 0$ such that $|z - z_0| < \eta$ implies $|f(z) - a| > 0$.

This means that

$$g(z) = \frac{1}{f(z) - a} \tag{337}$$

is well defined for $|z - z_0| < \eta$. Solving for $f(z)$ allows us to express f in terms of g:

$$f(z) = a + \frac{1}{g(z)}. \tag{338}$$

If $g(z_0)$ is finite then $f(z)$ is analytic. If $g(z)$ has a zero of finite order at z_0, then $f(z)$ has a pole of the same order. The only other possibility is that $g(z)$ has an essential singularity.

By our previous argument, $g(z)$ cannot be bounded in a neighborhood of z_0, which means that

$$|f(z) - a| = |\frac{1}{g(z)}| \tag{339}$$

can be made as small as desired in any neighborhood of z_0. This completes the proof of the theorem ■.

While the residue theorem is straightforward consequence of Cauchy's theorem, it is one of the most useful theorems of complex analysis.

Recall that if $f(z)$ is analytic in a neighborhood of z_0 then Cauchy's theorem implies

$$0 = \frac{1}{2\pi i} \oint_c f(z')dz' \tag{340}$$

for any closed regular curve in the neighborhood about z_0

When $f(z)$ is an isolated singularity of $f(z)$ we define the residue of $f(z)$ at z_0

$$\text{Res}(f(z_0) = \frac{1}{2\pi i} \oint_c f(z')dz' \tag{341}$$

for any closed regular curve in the neighborhood of z_0 containing no other singularities.

If a curve encloses N isolated singularities then is it possible, using Cauchy's theorem, to replace the single curve by N curves, where the n^{th} curve C_n only contains the n^{th} singular point. Application of Cauchy's theorem and the definition of the residue gives

Theorem 14.1:(residue theorem) If $f(z)$ is meromorphic in a region R and C is a regular curve in R then

$$\frac{1}{2\pi i} \oint_c f(z')dz' = \sum_{n=1}^{N} \frac{1}{2\pi i} \oint_{c_n} f(z')dz' = \sum_{n=1}^{N} \text{Res}(f(z_0)). \tag{342}$$

where the sum is over the residues of singularities enclosed by R.

The theorem is only useful if we can compute the residue of different kinds of curves around different kinds of isolated singularities.

Given an isolated singularity at $z = z_0$ we can alway use Cauchy's theorem to replace a given curve by a circle of small radius about z_0. We still must specify whether the curve is counterclockwise (positive) or clockwise (negative) and how many times the curve winds around z_0 in the counterclockwise n_+ and clockwise n_- direction.

We first consider the case that $f(z)$ has a simple pole at $z = z_0$. In that case we can write

$$f(z) = \frac{g(z)}{z - z_0}. \tag{343}$$

where $g(z_0) \neq 0$ and $g(z)$ is analytic in a neighborhood of z_0.

$$\text{Res}(f(z)) = \frac{1}{2\pi i}\oint \frac{g(z_0)}{re^{i\phi}} ire^{i\phi}d\phi + \frac{1}{2\pi i}\oint \frac{g(z)-g(z_0)}{z-z_0}dz. \tag{344}$$

The integrand of the second term is analytic so this integral gives zero. The integral of the first term is

$$\text{Res}(f(z)) = \frac{g(z_0)}{2\pi i}\oint id\phi \tag{345}$$

If the curve c is counterclockwise residue is

$$g(z_0), \tag{346}$$

if the curve c is clockwise residue is

$$-g(z_0), \tag{347}$$

and if the curve c goes around z_0 n_+ times in the counterclockwise and n_- time in the clockwise direction the residue is

$$(n_+ - n_-)g(z_0) \tag{348}$$

Next we consider the case of poles of order $n > 1$. First let

$$f(z) = \frac{b_n}{(z-z_0)^n} \tag{349}$$

Let c be a single counterclockwise circle. The

$$\text{Res}f(z_0) = \frac{1}{2\pi i}\int_0^{2\pi} \frac{ire^{i\phi}d\phi}{r^n e^{in\phi}} =$$

$$\frac{1}{2\pi i}\int_0^{2\pi r^{n-1}} ie^{-i(n-1)\phi}d\phi = 0 \tag{350}$$

In general, if $f(z)$ has a pole of order n at z_0 then

$$f(z) = \frac{g(z)}{(z-z_0)^n} \tag{351}$$

where $g(z)$ is analytic at $z = z_0$. If we expand $g(z)$ in a power series we get

$$f(z) = \frac{g(z_0)}{(z-z_0)^n} + \frac{dg}{dz}(z_0)\frac{1}{(z-z_0)^{n-1}} + \cdots +$$

$$\frac{1}{(n-1)!}\frac{d^{n-1}g}{dz^{n-1}}(z_0)\frac{1}{(z-z_0)^1} + h(z) \tag{352}$$

where $h(z)$ is analytic. Integrating this and computing the residue gives

$$\text{Res}(f(z_0)) = \frac{1}{(n-1)!}\frac{d^{n-1}g}{dz^{n-1}}(z_0) \tag{353}$$

where

$$g(z) = f(z)(z-z_0)^n \tag{354}$$

This must be non-zero if $f(z)$ has a pole of order n.

Note that the residue of a function with $f(z)$ with an order n pole at $z = z_0$ is always the Laurent coefficient b_1.

When a function has multiple poles at different points consider

$$f(z) = \prod_{n=1}^{N} \frac{1}{z-z_n} \qquad z_i \neq z_j,\ i \neq j \tag{355}$$

Then

$$f(z) = \sum_{n=0}^{N} \frac{c_n}{z-z_n} \tag{356}$$

The coefficients c_n can be computed by taking the difference of these functions and evaluating the integral over a small circle that only contains the $n-th$ singularity. In that case we get the identity

$$c_n = \frac{2\pi i}{2\pi i}\prod_{k\neq n}^{N}\frac{1}{z_n - z_k} = \prod_{k\neq n}^{N}\frac{1}{z_n - z_k} \tag{357}$$

This leads to the useful representation

$$f(z) = \prod_{n=1}^{N}\frac{1}{z-z_n} = \sum_{n=0}^{N}\prod_{k\neq n}^{N}\frac{1}{z_n - z_k}\frac{1}{z-z_n} \tag{358}$$

This has the immediate generalization to the case that the product is multiplied by an arbitrary polynomial $P(z)$:

$$f(z) = \prod_{n=1}^{N} \frac{P(z)}{z - z_n} = \sum_{n=0}^{N} \prod_{k \neq n}^{N} \frac{1}{z_n - z_k} \frac{P(z_n)}{z - z_n} \tag{359}$$

Note that this formula still holds if $P(z)$ vanishes at some of the roots.

0.15 Lecture 15

Theorem 14.2:(Jordan's Lemma) Let Γ_R be a semicircle in the upper half of the complex plane of radius R with center at the origin. Let $f(z) \to 0$ uniformly in $\arg(z)$ as $|z| \to \infty$ when $0 \leq \phi \leq \pi$. Let

$$I_R := \int_{\Gamma_R} e^{i\alpha z} f(z) dz \tag{360}$$

Then for $\alpha > 0$

$$\lim_{R\to\infty} I_R = 0 \tag{361}$$

proof: The proof uses Darboux's theorem

$$I_R := \int_{\Gamma_R} e^{i\alpha z} f(z) dz =$$

$$\int_0^\pi e^{i\alpha R\cos(\phi) - \alpha R\sin(\phi)} f(Re^{i\phi}) iRe^{i\phi} d\phi \tag{362}$$

Since $f(z)$ vanishes uniformly in the argument of z as $|z|\infty$ for $0 \leq \phi \leq \pi$ then for any $\epsilon >$ we can find a large enough R so

$$|I_R| :=$$

$$\epsilon R \int_0^\pi e^{-\alpha R\sin(\phi)} d\phi = 2\epsilon R \int_0^{\frac{\pi}{2}} e^{-\alpha R\sin(\phi)} d\phi \tag{363}$$

Since

$$\sin(\theta) \geq \frac{2\theta}{\pi} \qquad 0 \leq \theta \leq \frac{\pi}{2} \tag{364}$$

this is bound by

$$|I_R| :=$$

$$2\epsilon R \int_0^{\frac{\pi}{2}} e^{\frac{-\alpha R 2\theta}{\pi}} d\phi = 2R\epsilon \frac{\pi}{2\alpha R}(1 - e^{-\alpha R}) = \frac{\epsilon\pi}{\alpha}(1 - e^{-\alpha R}) \to 0 \tag{365}$$

as $R \to \infty$

This completes the proof of Jordan's lemma.

Next I consider some applications of the residue theorem

Example 1: Let

$$f(z) = \frac{g(z)}{h(z)} \tag{366}$$

where $g(z)$ and $h(z)$ are analytic and $h(z)$ has a simple zero at z_0 and $g(z_0) \neq 0$. Then we have

$$h(z) = (z - z_0)[\frac{dh}{dz}(z_0) + (z - z_0)w(z)] \tag{367}$$

where $w(z)$ is analytic and can be computed from Taylor's theorem. If C is a curve around the point $z = z_0$ in the counter clockwise direction then

$$\oint_C f(z)dz = 2\pi i \frac{g(z_0)}{\frac{dh}{dz}(z_0)} \tag{368}$$

Example 2: Let

$$I = \int_0^\infty \frac{x^2 dx}{(x^2+1)(x^2+4)} \tag{369}$$

Since this is an even function it can be written as

$$I = \frac{1}{2}\int_{-\infty}^\infty \frac{x^2 dx}{(x^2+1)(x^2+4)} \tag{370}$$

Next we convert this to an integral of a complex function around a closed contour. To do this replace all of the x's by a complex variable z. Then extend the integral to include a large semicircle and take the limit as the radius of the semicircle goes to infinity. If we can shoe that the integral over the semicircle is zero, then the integral over the line is equal to the contour integral, which is $2\pi i$ times the sum of the residues of the poles bounded by real axis and the semicircle.

To show the integral over the semicircle gives no contribution let $z = Re^{i\phi}$ and $dz = iRe^{i\phi}d\phi$ giving that

$$\frac{i}{2}\int_0^\pi \frac{R^3 e^{3i\phi} d\phi}{(R^2 e^{2i\phi}+1)(R^2 e^{2i\phi}+1)} = \tag{371}$$

$$\frac{i}{2R}\int_0^\pi \frac{e^{-i\phi} d\phi}{(1+e^{-2i\phi}/R^2)(1+4e^{-2i\phi}/R^2)} = \tag{372}$$

using Darboux's theorem this the modulus of this is bounded by

$$\leq \frac{1}{2R}\frac{2\pi}{(1-1/R^2)(1-4/R^2)} \tag{373}$$

for $R > 1$. This clearly vanishes as $R \to \infty$

Next factor the denominator to get

$$I = \frac{1}{2}\oint_C \frac{z^2 dz}{(z+i)(z-i)(z+2i)(z-2i)} \tag{374}$$

This has simple poles inside C at i and $2i$ giving

$$I = \frac{2\pi}{2} i [\frac{(i^2}{(i+i)(i+2i)(i-2i)} + \frac{(2i)^2}{(2i+i)(2i-i)(2i+2i)} = \pi[-\frac{1}{6} + \frac{1}{3}] = \frac{\pi}{6} \tag{375}$$

Example 3: Let

$$I = \int_0^{2\pi} \frac{d\phi}{1 + a\sin(\theta)} \qquad 0 \leq a^2 < 1 \tag{376}$$

Let $z = e^{i\phi}$, $dz = izd\phi$ and note

$$\sin(\phi) = \frac{1}{2i}(z - z^{-1}) \tag{377}$$

It follows that the integral can be written as a contour integral

$$\oint_C \frac{dz}{iz} \frac{1}{1 + a(\frac{1}{2i})(z - 1/z)} = \tag{378}$$

$$\oint_C dz \frac{2}{2iz + a(z^2 - 1)} = \tag{379}$$

$$\oint_C dz \frac{2}{a} \frac{1}{(z - i[\frac{1}{a} + \sqrt{\frac{1}{a^2} - 1})(z - i[\frac{1}{a} - \sqrt{\frac{1}{a^2} - 1})} \tag{380}$$

The pole at $i[\frac{1}{a} - \sqrt{\frac{1}{a^2} - 1}$ is in the unit circle, giving

$$I = \frac{2}{a} \frac{2\pi i}{(i[\frac{1}{a} - \sqrt{\frac{1}{a^2} - 1}] - i[\frac{1}{a} + \sqrt{\frac{1}{a^2} - 1}])} = \frac{2\pi}{\sqrt{1 - a^2}} \tag{381}$$

29:171 - Homework Assignment #5

1. Find the Laurent series for $e^{1/z}$ about the origin. What kind of isolated singularity does this function have at $z = 0$?

2. Find the Laurent series for $\cosh(z + 1/z)$ about the origin. What kind of isolated singularity does this function have at $z = 0$?

3. Let $f_1(z)$ be analytic in a region R_1. Assume that $f_2(z)$ is analytic in a region R_2. Assume that R_1 and R_2 have a not empty, simply connected intersection that contains an open set, and that f_1 and f_2 agree on the intersection. Use Morerra's theorem to show that the function $g(z) = f_1(z)$ on R_1 and $f_2(z)$ on R_2 is analytic.

4. What is the radius of convergence of the Taylor series of the analytic function
$$f(z) = \frac{1}{(z-4)(z^2+5)}$$
about the point $z_0 = 10i$?

5. Use Cauchy's theorem to evaluate the integral
$$\int_0^\infty e^{iby^2}$$
where b is real.

6. What can you say about an entire function that is bounded by $|z^{3/2}|$ for large z?

0.16 Lecture 16

Example 4: Let

$$I = \int_{-\infty}^{\infty} e^{-iax-bx^2} dx. \tag{382}$$

First complete the square in the exponent.

$$I = \int_{-\infty}^{\infty} e^{-b(x+\frac{ia}{2b})^2 - \frac{ba^2}{4b^2}} dx \tag{383}$$

Next let $x' = x + i\frac{a}{2b}$ and $dx = dx'$ to get

$$I = e^{-\frac{a^2}{4b}} \int_{-\infty+i\frac{a}{2b}}^{\infty+i\frac{a}{2b}} e^{-b(x')^2} dx. \tag{384}$$

In order to evaluate this make a rectangle bounded by $[-R, R]$ on the real axis, the line $[-R+i\frac{a}{2b}, R+i\frac{a}{2b}]$, and the edges $[R, R+i\frac{a}{2b}]$, and $[-R, -R+i\frac{a}{2b}]$. The integrand is analytic inside of this rectangle. Applying Cauchy's theorem gives

$$0 = e^{-\frac{a^2}{4b}} \int_{-R}^{R} e^{-bx^2} dx + e^{-\frac{a^2}{4b}} \int_{0}^{\frac{a}{2b}} e^{-b(R+iy)^2} dy$$
$$-e^{-\frac{a^2}{4b}} \int_{-R+i\frac{a}{2b}}^{R+i\frac{a}{2b}} e^{-bx^2} dx + e^{-\frac{a^2}{4b}} \int_{\frac{a}{2b}}^{0} e^{-b(R-iy)^2} dy. \tag{385}$$

In the limit that $R \to \infty$ the first integral becomes a standard Gaussian integral, while the third integral becomes the integral that we are trying to compute. Thus this reduces to a standard Gaussian integral if we can show that the boundary terms give no contribution in the limit that $R \to \infty$. To do this consider

$$|e^{-\frac{a^2}{4b}} \int_{0}^{\frac{a}{2b}} e^{-b(R+iy)^2} dy| = e^{-\frac{a^2}{4b}} \int_{0}^{\frac{a}{2b}} e^{-bR^2} e^{-2iyb} e^{by^2} dy \leq \tag{386}$$

$$e^{-\frac{a^2}{4b}} e^{-bR^2} e^{b(\frac{1}{2b})^2} \to 0 \tag{387}$$

as $R \to \infty$. It follows that

$$I = \lim_{R\to\infty} e^{-\frac{a^2}{4b}} \int_{-R}^{R} e^{-bx^2} dx = \sqrt{\frac{\pi}{b}} e^{\frac{-a^2}{4b}} \tag{388}$$

Example 5: Let

$$I = \int_0^\infty \frac{\sin(x)}{x} dx. \tag{389}$$

To calculate this note

$$I = \frac{1}{2i} \int_0^\infty (\frac{e^{ix}}{x} - \frac{e^{-ix}}{x}). \tag{390}$$

Consider the curve C in the upper half plane bounded by a small half circle of radius r and a large half circle of radius R about the origin and the line segments on the real line connecting these circles.

Compute

$$0 = \int_C \frac{e^{iz}}{z} dz = \int_{\odot_R} \frac{e^{iz}}{z} dz + \int_{-R}^{-r} \frac{e^{ix}}{x} dx +$$

$$i \int_\pi^0 e^{(ir\cos(\phi) - r\sin(\phi))} d\phi + \int_r^R \frac{e^{ix}}{x} dx. \tag{391}$$

The integral over the large semicircle vanishes in the limit $R \to \infty$ by Jordan's lemma. Changing $x \to -x$ in the second integral gives

$$- \int_r^R \frac{e^{-ix}}{x} dx +$$

$$i \int_\pi^0 e^{ir\cos(\phi) - r\sin(\phi)} d\phi + \int_r^R \frac{e^{ix}}{x} dx. \tag{392}$$

Since the integrand in the integral over the small circle can be expanded in a uniforly convergent series in r. Changing the order of the sum and integral gives i plus a function that vanishes as, $r \to 0$, giving

$$i\pi = \lim_{R\to\infty, r\to o} \int_r^R (\frac{e^{ix}}{x} - \frac{e^{-ix}}{x}) dx \tag{393}$$

which leads to

$$I = \frac{2}{\pi}. \tag{394}$$

One type of integral that often appears in problems involving scattering is

$$I = \int_{-\infty}^\infty \frac{f(x)}{x - x_0} dx \tag{395}$$

where $|f(x)| < c|x|^{-\alpha}$ with $\alpha > 0$ and $f(z)$ is analytic in a neighborhood of x_0.

It is useful to define two semicircular paths around x_0:

$$\Gamma_r^+ = x_0 + re^{i\phi} \qquad \phi : \pi \to 0 \tag{396}$$

$$\Gamma_r^- = x_0 + re^{i\phi} \qquad \phi : \pi \to 2\pi \tag{397}$$

The integral is not well-defined due to the singularity, however it can be made into a well-defined integral deforming the path to avoid the singularity:

$$I_\pm :=$$

$$\int_{-\infty}^{x_0-r} \frac{f(x)}{x-x_0}dx + \int_{\Gamma_r^\pm} \frac{f(x)}{x-x_0}dx + \int_{x_0+r}^{\infty} \frac{f(x)}{x-x_0}dx. \tag{398}$$

0.17 Lecture 17

For very small r the middle integral becomes

$$\lim_{r\to 0}\int_{\Gamma_r^\pm}\frac{f(x)}{x-x_0}dx =$$

$$i\int \frac{f(x_0+re^{i\phi})}{re^{i\phi}}ire^{i\phi}d\phi =$$

$$i\int f(x_0+re^{i\phi})d\phi =$$

$$\mp i\pi f(x_0) + \int [f(x_0+re^{i\phi})-f(x_0)] \tag{399}$$

The last terms vanishes by Darboux's theorem and the continuity of $f(x)$ as $r\to 0$. This gives

$$\lim_{r\to 0}\int_{\Gamma_r^\pm}\frac{f(x)}{x-x_0}dx = \mp i\pi f(x_0) \tag{400}$$

The value of the middle integral depends on the path taken around the singularity.

The remaining integral is called the principal value of the integral.

I show that it is finite:

$$P\int_{-\infty}^{\infty}\frac{f(x)}{x-x_0}dx := \lim_{r\to 0}[\int_{-\infty}^{x_0-r}\frac{f(x)}{x-x_0}dx + \int_{x^0+r}^{\infty}\frac{f(x)}{x-x_0}dx]$$

The term in $[\cdots]$ can be written as a sum of integrals

$$[\cdots] =$$

$$\int_{-\infty}^{x_0-a}\frac{f(x)}{x-x_0}dx + \int_{x_0-a}^{x_0-r}\frac{f(x)-f(x_0)}{x-x_0}dx +$$

$$f(x_0)\underbrace{[\ln(\frac{-r}{-a})+\ln(\frac{b}{r})]}_{\ln(\frac{b}{a})} +$$

$$\int_{x_0+r}^{x_0+b}\frac{f(x)-f(x_0)}{x-x_0}dx + \int_{x_0+b}^{\to\infty}\frac{f(x)}{x-x_0}dx \tag{401}$$

The singular terms ($r = 0$) in the middle integral cancel before we take the limit. Every other term in the is integral is well defined. In the limit that $r \to 0$ this is well defined and defines the principal value of the integral.

Note

$$\frac{f(x) - f(x_0)}{x - x_0} = \sum_{n=1}^{\infty} \frac{1}{n!} \frac{d^n f}{x^n}(x_0)(x - x_0)^{n-1} \tag{402}$$

converges uniformly in a neighborhood of x_0. The points a and b can be chosen small enough so these expressions converge on $x \in [x_0 - a, x_0 + b]$.

If this integral is combined with our two integrals over the semicircle, the integral over the lower semicircle can be replaced by an integral over the line with the singularity raised infinitesimally above the line

$$I_\pm = \int_{-\infty}^{\infty} \frac{f(x)}{x - x_0 \mp i0^+}$$
$$P \int_{-\infty}^{\infty} \frac{f(x)}{x - x_0} \pm i\pi f(x_0) \tag{403}$$

This formula is often written as

$$\frac{1}{x - x_0 \mp i0^+} = P \frac{1}{x - x_0} \pm i\pi\delta(x - x_0) \tag{404}$$

We will introduce the "δ-function" later.

Note that in this integral a slight change in the position of the singularity can change the value of the integral. When these integrals arise in problems, the physics determines the position of the singularity.

Example: Compute the principal value of

$$\int_0^\infty \frac{dx}{x^2 - x_0^2} \tag{405}$$

We write this as the

$$P \int_0^\infty \frac{dx}{x^2 - x_0^2} =$$
$$\lim_{\epsilon \to 0} [\int_0^{x_0 - \epsilon} \frac{dx}{x^2 - x_0^2} + \int_{x_0 + \epsilon}^\infty \frac{dx}{x^2 - x_0^2}] \tag{406}$$

Let $u = x/x_0$, $du = dx/x_0$ in the first integral and $v = x_0/x$, $dv = x_0 dx/x^2$ in the second to get

$$\lim_{\epsilon \to 0} [\int_0^{1 - \epsilon/x_0} \frac{du}{x_0^2(u^2 - 1)} + \int_0^{\frac{1}{1 - \epsilon/x_0}} \frac{dv}{x_0^2(1 - v^2)}] \tag{407}$$

Relabeling $v \to u$ in the second integral gives

$$\lim_{\epsilon \to 0} \int_{\frac{1}{1-\epsilon/x_0}}^{1-\epsilon/x_0} \frac{du}{x_0^2(u^2-1)} \tag{408}$$

The combined integral can be bounded using Darboux's theorem

$$\int_{\frac{1}{1-\epsilon/x_0}}^{1-\epsilon/x_0} \frac{du}{x_0^2(u^2-1)} \leq |\frac{1}{x_0^2}\frac{x_0}{\epsilon}|\frac{1}{1-\epsilon/x_0} - 1 - \epsilon/x_0|. \tag{409}$$

The last term vanishes like ϵ^2, and when combined with the $1/\epsilon$, the integral vanishes like a constant $\times\epsilon$ as $\epsilon \to 0$. This shows that

$$P\int_0^\infty \frac{dx}{x^2 - x_0^2} = 0 \tag{410}$$

Multi-valued functions;

Recall that

$$e^{\ln z} = z \tag{411}$$

can solved to get

$$\ln(z) = \ln(|z|) + i\phi \tag{412}$$

where

$$\phi = \arg(z) + 2\pi n. \tag{413}$$

If we start by considering the branch of the complex logarithm corresponding to a $0 \leq \phi < 2\pi$ then let ϕ increase by 2π this function does not return to its original value. It returns to its original value plus $2\pi i$.

Definition: A point z_0 in the complex plane such that $f(z)$ does not return to its initial value after going around *any* closed curve is called a branch point of $f(z)$.

Definition: A line connecting two branch points is called a branch cut of $f(z)$.

By these definitions we see that 0 is a branch point of $\ln(z)$. Since $\ln(1/z) = -\ln(z)$ we see that ∞ is also a branch point. Any line from 0 to ∞ can be taken as a branch cut for the logarithm.

While the branch points of a function are fixed by the function, the branch cuts can be chosen as desired.

The $\ln(z)$, while multiple valued, is analytic in a small region about any z, provided z is not a branch point. If we do not choose to identify the points in the complex plane that differ by $2\pi n$, we get a more complex geometrical surface called a Riemann surface. The Riemann surface for $\ln(z)$ can be thought of as an infinite spiral of complex planes that wind around the origin. Each plane in the Riemann surface is called a Riemann sheet. Branch cuts are places where one passes from one sheet to the next. For example, if we choose the branch cut of $\ln(z)$ to be the real axis, $0 < \phi < 2\pi$ is the $n = 0$ sheet of $\ln(z)$, negative $-2\pi < \phi < 0$ corresponds to the $n = -1$ sheet, while, $2\pi < \phi < 4\pi$ corresponds to the $n = 1$ sheet.

We can write the complex logarithm as an infinite collection of functions

$$\ln_n(z) = \ln(|z|) + i(\arg(\phi) + 2n\pi) \qquad 0 \leq \phi < 2\pi \tag{414}$$

This function has complex derivatives at each point on the branch cut, but there are different values at the points where different adjacent sheets meet. This is easy to see by noting that adding $2\pi ni$ to a function with a complex derivative does not impact its differentiability.

The structure of the Riemann surface depends on the function. Multi-valued functions often arise when complex variables are raised to non-integer powers.

0.18 Lecture 18

For the next example of a multi-valued function consider the square root of z. Let

$$z = re^{i\phi}. \tag{415}$$

If ϕ is increased by 2π in

$$z^{1/2} = r^{1/2}e^{i\phi/2} \tag{416}$$

the function does *not* return to its original value. Instead it becomes

$$z^{1/2} \to r^{1/2}e^{i(\phi/2+\pi)} = -r^{1/2}e^{i\phi/2} \tag{417}$$

which is the well-known second value of the square root. If the phase ϕ is increased by an additional factor of 2π this returns to the original function.

This shows that $\sqrt{z}$ has a branch point at $z = 0$. Since we can make the same argument for $\frac{1}{\sqrt{z}}$ it follows that $\sqrt{z}$ also has a branch point at infinity.

A branch cut is any line that starts at 0 and extends to ∞. This function is double valued. If we start at the branch cut and increase ϕ by 2π we change the sign of the square root; if we continue increasing by a second factor of 2π we return to the original branch of the function.

A straight forward computation shows that the two branches of $\sqrt{z}$ are

$$z^{1/2} = r^{1/2}e^{i\phi/2} \tag{418}$$

$$z^{1/2} = -r^{1/2}e^{i\phi/2} \tag{419}$$

Another example of a multiple valued function is

$$f(z) := z^{\alpha} \tag{420}$$

for real α. Like the square root this has branch points at zero and infinity, since each curve around $z = 0$ increases the phase by $e^{i2\pi\alpha}$.

This function will have an infinite number of branches if α is irrational, while it will have a finite number when α is rational.

Treating functions with multiple branch points requires some care. Consider the function

$$f(z) = (z^2 - 1)^{1/2} = (z-1)^{1/2}(z+1)^{1/2}. \tag{421}$$

This function is a product of square roots. The first term has branch points at 1 and ∞ while the second term has branch points at -1 and ∞.

If we let $z \rightarrow \frac{1}{z}$ in this function the resulting function becomes

$$g(z) = f(\frac{1}{z}) = \frac{(1-z)^{1/2}(z+1)^{1/2}}{z} \tag{422}$$

which has a simple pole at $z = 0$. Thus this product has no branch point at infinity.

The result is that if we consider a curve that goes around one of the branch points the phase increases by π, on the other hand if the curve goes around both branch points the function returns to its original value.

In this case the branch cut is any line between -1 and 1. Sometimes it is convenient to use the line of length 2 that connects these points. It is also possible to deform this curve so it goes through infinity, i.e. $[-\infty, -1]U[1, \infty]$. How it is chosen is a matter of convenience.

Integrals where the integrands have branch cuts

The value of a multivalued function changes discontinuously across a branch cut. It is still possible to use Cauchy's theorem and the residue theorem with functions that have branch cuts; we just have to make sure that our curves do not cross branch cuts and we must realize that the integrand is not continuous across a branch cut.

As an example consider the integral

$$I := \int_0^\infty \frac{x^{p-1}}{x^2+1} \qquad 0 < p < 2 \tag{423}$$

The integrand of this function has simple poles at $z = \pm i$ and branch points at 0 and ∞ (for $p \neq 1$).

Take the branch cut along the positive real axis.

To do this integral we first choose a branch of the integrand. The value of this function on the n^{th} branch for $z = re^{i\phi}$ can be labeled by

$$f_n(z) = \frac{r^{p-1}e^{i(p-1)(\phi+2\pi n)}}{r^2 e^{2i\phi}+1} \tag{424}$$

To do this integral choose the branch $n = 0$ with this choice value of $f(z)$ just above the real line is

$$f_0(x) = r^{p-1}e^{i(p-1)} \tag{425}$$

while the value just below the real line is

$$f_0(x) = r^{p-1}e^{i(p-1)2\pi} \tag{426}$$

To evaluate this integral chose a contour that starts near the origin just above the real axis and follows the real axis out an amount R. To this add a circle of radius R around the entire plane. It returns to the positive x axis on a different Riemann sheet, corresponding to $n = 1$. Next integrate back to the origin, this time below the branch cut. Finally connect the two line at the origin by a small circle.

The final curve does not cross the branch cut. The integrand has two simple poles at $z = \pm i$ in the interior of the curve so the integral around the closed curve can be computed using the residue theorem. In order to use the theorem it has to be applied to the chosen branch. In out case $n = 0$ branch was convenient. In this case the residue theorem gives

$$I_t = 2\pi i[\frac{1^{p-1}e^{i(p-1)(\pi/2)}}{2i} - \frac{1^{p-1}e^{i(p-1)(3\pi/2)}}{2i}] = 2\pi i e^{i\pi(p-1)} \sin(\frac{\pi}{2}(p-1)) \quad (427)$$

Here we used $3\pi/2$ rather than $-\pi/2$ because $3\pi/2$ corresponds to the chosen $n = 0$ branch.

The contributions from the large and small circle can be estimated using Darboux's. theorem. The conditions on p are chosen so the contribution to the integral from the large and small circles vanish in the limit that the radii become infinity and zero respectively.

The surviving integrals are

$$I_t = \lim_{r\to\infty} \lim_{r\to 0}[\int_r^R dx \frac{x^{p-1}}{x^2-1} - \int_r^R dx \frac{x^{p-1}}{x^2-1} e^{i(p-1)2\pi} =$$

$$\int_0^\infty dx \frac{x^{p-1}}{x^2-1} \times [1 - e^{i2\pi(p-1)}] = \int_0^\infty dx \frac{x^{p-1}}{x^2-1} \times -2ie^{ip\pi} \sin(p\pi) \quad (428)$$

which is the desired integral multiplied by a complex factor. Using this with (**??**) gives an expression for this integral gives

$$\int_0^\infty \frac{x^{p-1}}{x^2+1} dx = \pi \frac{\cos(\frac{\pi p}{2})}{\sin(\pi p)} = \frac{\pi}{2\sin(\frac{\pi p}{2})} \quad (429)$$

29:171 - Homework Assignment #6

Compute the following integrals

1.
$$\int_0^{2\pi} \frac{d\theta}{a + b\sin(\theta)} \qquad a > b > 0$$

2.
$$\int_0^{2\pi} \frac{\sin^2(\theta)d\theta}{a + b\cos(\theta)} \qquad a > b > 0$$

3.
$$\int_0^{2\pi} \frac{d\theta}{(a + b\cos(\theta))^2} \qquad a > b > 0$$

4.
$$\int_0^{\infty} \frac{dx}{1 + x^4}$$

5.
$$\int_0^{\infty} \frac{x^2 dx}{(x^2 + a^2)^3}$$

6.
$$\int_0^{\infty} \frac{\sin(x)dx}{x(x^2 + a^2)^2}$$

7.
$$\int_0^{\infty} \frac{\sin^2(x)dx}{x^2}$$

0.19 Lecture 19

The residue theorem can also be employed to sum infinite series. The following theorem useful.

Theorem 19.1: Let $f(z)$ be meromorphic and let C be a regular curve that encloses the zeros of $\sin(\pi z)$ located at $z = m, m+1, \cdots m+n$. If the poles of $f(z)$ and zeros of $\sin(\pi z)$ are distinct then

$$\sum_{k=m}^{m+n} f(m) = \frac{1}{2\pi i}\oint_C \pi\cot(\pi z) f(z)dz - \sum_{\text{poles}_{f(z)}} \text{Res}[\pi\cot(\pi z)f(z)] \quad (430)$$

The proof of this theorem is a direct application of the residue theorem to the meromorphic function $\pi\cot(\pi z)f(z)$ which has poles for $z\in\mathbb{Z}$ in addition to the poles of $f(z)$.

The residue of the poles of $\pi\cot(\pi z)$ are all 1 which can be seen by expanding the Taylor series of the $\sin(\pi z)$ about $z = n\pi$:

$$\pi\cot(\pi z) \approx \pi\frac{\cos(\pi n)}{\pi\cos(\pi n)(z-n)} + \cdots \quad (431)$$

To use this note

$$|\cot(\pi z)| = |\frac{\cos(\pi(x+iy))}{\sin(\pi(x+iy))}$$
$$= |\frac{\cos(\pi x)\cosh(\pi y)) - i\sin(\pi x)\sinh(\pi y))}{\sin(\pi x)\cosh(\pi y)) + i\cos(\pi x)\sinh(\pi y))}| =$$
$$|\frac{\cos^2(\pi x) + \sinh^2(\pi y)}{\sin^2(\pi x) + \sinh^2(\pi y))}|^{1/2} \quad (432)$$

This is bounded for large y. When $y = 0$ it is bounded for $x = n + \frac{1}{2}$.

If this is multiplied by a function that vanishes for large $|z|$ and the contour is chosen to be a large rectangle that intersects the x-axis at half integers, then in the limit of infinite rectangle size

$$\frac{1}{2\pi i}\oint_C \pi\cot(\pi z)f(z)dz \to 0. \quad (433)$$

This means that

$$\sum_{n=-\infty}^{\infty} f(n) = -\sum_{\text{poles}_{f(z)}} \text{Res}[\pi\cot(\pi z)f(z)]. \quad (434)$$

This is the most useful form of this theorem.

As an example let

$$f(z) = \frac{1}{a^2 + z^2}. \tag{435}$$

This has poles at $z = \pm ia$. The above theorem implies

$$\sum_{n=-\infty}^{\infty} \frac{1}{a^2 + n^2} = -\pi(\cot(\pi ia)\frac{1}{2ia} + \cot(-\pi ia)\frac{1}{-2ia}). \tag{436}$$

note

$$\cot(\pi ia) = \frac{\cos(i\pi a)}{\sin(i\pi a)} = -i\coth(\pi a). \tag{437}$$

$$\cot(-\pi ia) = i\coth(\pi a). \tag{438}$$

The right hand side of this expression gives

$$\frac{\coth(\pi a)}{a}. \tag{439}$$

The sequence can be rewritten as

$$\sum_{n=1}^{\infty} \frac{1}{a^2 + n^2} = \frac{\coth(\pi a)}{2a} - \frac{1}{a^2}. \tag{440}$$

While the $\pi\cot(\pi z)$ is useful for summing series with positive entries, $\pi\csc(\pi z)$ can be used to sum alternating series. It can also be shown to be bounded in the same sense as $\pi\cot(\pi z)$, however the residue of the poles at $z = n$ are $\cos(n\pi) = (-)^n$.

Next I consider the problem of extending analytic functions. First I state a uniqueness theorem for analytic functions.

Theorem 17.1: Let $f_1(z)$ and $f_2(z)$ be analytic in a region R. Let S be a set in R with an accumulation point z_0 in R. If $f_1(z) = f_2(z)$ for $z \in S$ the $f_1(z) = f_2(z)$ for all $z \in R$.

To prove this note that $g(z) = f_1(z) - f_2(z)$ is analytic in R. By assumption $g(z_n) = 0$ for all $z \in S$. Since z_0 is an accumulation point of S we can find a subsequence z'_n in S such that $z'_n \to z_0$. Since $g(z)$ is analytic, it is also continuous. It follows that

$$g(z_0) = \lim_{n\to\infty} g(z'_n) = 0 \tag{441}$$

Thus z_0 is a zero of an analytic function that is not isolated. The only analytic function that does not have isolated zeros is the zero function. Therefore

$$f_1(z) = f_2(z) \tag{442}$$

in all of R.

Since curves and small neighborhoods contain points with accumulation points, this theorem shows that analytic functions that agree on a curve or a small neighborhood are also identical.

0.20 Lecture 20

Next assume that $f(z)$ is analytic in a region $D \subset R$ and known explicitly in a neighborhood of z_0. Consider a curve C between z_0 and any other z in R.

At each point $z \in C \cap D$ the function has a Taylor expansion with a a radius of convergence that is determined by the largest circle that can be drawn around z_0 in D. Pick a point on C in this circle near the boundary of the circle. It is possible to generate a new Taylor series about this point that has a radius of convergence that extends the boundary of D to the nearest singularity. In general it will extend outside of the original circle. Since the two series agree in the intersection of the circles, they define a single analytic function in the combined region. This process can be repeated until we either hit a singularity on C or reach the final point z. This shows that the function can be extended to a single analytic function to all points that can be connected to the original point by a curve in the new domain of analyticity. If there is more than one path to z and the two path go around a branch point the resulting function could be multiple valued.

The process where an analytic function can be extended to a single analytic function in an extended region is called analytic continuation.

Theorem 20.1: Let R_1 and R_2 be non overlapping regions with a common boundary B. Assume that $f_1(z)$ is analytic in R_1 and $f_2(z)$ is analytic in R_2 and $f_1(z)$ and $f_2(z)$ are continuous and equal on B. Then

$$h(z) := \left\{ \begin{array}{ll} f_1(z) & z \in R_1 \cup B \\ f_2(z) & z \in R_2 \cup B \end{array} \right. \tag{443}$$

is analytic in $R_1 \cup R_2 \cup B$.

This theorem can be proved using Morerra's theorem. Any curve that is not in R_1 or R_2 can be broken up into parts in B and parts that cross B. By continuity the parts in B can be moved infinitesimally in R_1 or R_2 without changing the value of the integral. For a closed curve the number of crossing must be even. The curve can be replaced by a sum of curves that are entirely in R_1 or R_2 by adding pairs of curves in opposite direction one either side of B. Using Cauchy's theorem in either region shows that the integral around any close curve in $R_1 \cup R_2 \cup B$ vanishes. Since $h(z)$ is also continuous in this region it is analytic by Morerra's theorem.

Theorem 20.2: (Schwartz reflection principle) Let $f(z)$ be analytic in a region R that has part of the real line as the boundary. Assume that $f(z)$ is real for z real, and continuous on the boundary. Define

$$\bar{R} = \{a|z^* \in R\} \tag{444}$$

Then

$$g(z) := (f(z^*))^* \tag{445}$$

is an analytic continuation of $f(z)$ into

$$\bar{R} \cup R \tag{446}$$

To prove this note that if $z \in \bar{R}$ then $z' = z^* \in R$. For z_1, $z_2 \in \bar{R}$

$$[\frac{f(z_1') - f(z_2')}{z_1' - z_2'}]^* = [\frac{g(z_1) - g(z_2)}{z_1 - z_2}] \tag{447}$$

since the term on the left has a limit as $z_1' \to z_2'$ independent of direction, the same is true for the term on the right. This means that $g(z)$ is analytic in $\bar{R}$. The reality for real z means that $f(z) = g(z)$ for z real. By theorem 17.1 $f(z)$ and $g(z)$ define a single analytic function on $\bar{R} \cup R$. This function has the property that

$$f^*(z) = f(z^*) \tag{448}$$

The nature of this theorem is easy to understand. The reality means if x is not the real line

$$f(x) = g(x) \sum_{n=0}^{\infty} c_n (x - x_0)^n \tag{449}$$

where $c_n = c_n^*$. Taking the conjugate of f conjugates $c_n \to c_n^* = c_n$ and $(z - x_0)^n \to (z^* - x_0)^n$. The conjugation of z then gives that the coefficients of the Taylor series for expanding $(z^* - x_0)^n \to (z - x_0)^n$ leading to the original series, which is now valid for z in either region.

Dispersion Relations

Many functions that appear in physics are analytic except for a branch cut along the real axis from $x_0 \to \infty$, are real for $x < x_0$ and fall off faster than $\frac{c}{|z|}$ for large $|z|$ and vanish at the branch point.

The source of these functions normally comes from expressions involving resolvents of self-adjoint operators (defined later). If O is a linear operator, and z is a complex number we can ask if operator $(z-O)$ has an inverse. The operator $(O-z)^{-1}$ considered as an operator valued function of the complex variable z is called the resolvent of O. While we will study these operators next semester, this operator has poles when O has isolated eigenvalues, and branch cuts where O has continuous eigenvalues. The operator is analytic in z when the inverse exists and is continuous. In this class of problems branch cuts on a half line are typical associated with operators that have eigenvalues bounded from below, like energy operators.

Normally the information on the branch cut is experimentally accessible. Dispersion relations provide a simple relation between the value of such a function anywhere and the imaginary part of the function along the branch cut. They follow as an application of the Schwartz reflection principle.

Consider

$$f(z) = \frac{1}{2\pi i} \oint_c \frac{f(z')}{z' - z} dz' \tag{450}$$

where curve is the sum of a circle at infinity and the lines in either side of the branch cut. The assumption that $|f(z)| < c\frac{1}{|z|}$ means that there is no contribution from the large circle at ∞. Vanishing at the origin means that there is no contribution from a small circle around the branch point. What remains are the integrals on either side of the branch cut:

$$f(z) = \frac{1}{2\pi i} \int_{x_0}^{\infty} \frac{f(x + i0^+)}{x + i0^+ - z} dx - \frac{1}{2\pi i} \int_{x_0}^{\infty} \frac{f(x - i0^+)}{x - i0^+ - z} dx \tag{451}$$

Taking the limit as $\epsilon \to 0$ for z not on the branch cut, using the Schwartz reflection principle ($f * (z) = f(z^*)$), gives

$$f(z) = \frac{1}{2\pi i} \int_{x_0}^{\infty} 2\Im \frac{f(x)}{x - z} dx \tag{452}$$

This is called a dispersion relation. It expresses the value of $f(z)$ in the cut plane in terms of the imaginary part of the function along the branch cut.

0.21 Lecture 21:

The fundamental Theorem of Algebra:

We initially introduced complex numbers to find roots to the equation $x^2 + 1 = 0$. We then mentioned that introducing the new number i was enough to factor any Polynomial. We are now in a position to prove this result.

Theorem 19.3: Let $f(z)$ be meromorphic in a region R and $g(z)$ be analytic in R. Let C be a closed regular curve in R on which $f(z)$ is analytic and nowhere zero. If $f(z)$ has M zeros $\{z_j\}_{j=1}^M$ of order $\{m_j\}_{j=1}^M$ and N poles $\{p_j\}_{j=1}^N$ or order $\{n_j\}_{j=1}^N$ in C then

$$\frac{1}{2\pi i}\oint_C g(z)\frac{\frac{df}{dz}}{f(z)}dz = \sum_{j=1}^{M} m_j g(z_j) - \sum_{n=1}^{N} m_j g(p_j) \tag{453}$$

To prove this note that in general near a pole or zero $f(z)$ has the form

$$f(z) = a_1(z-z_0)^n + a_2(z-z_0)^{n+1} + \cdots \tag{454}$$

and

$$\frac{df(z)}{dz} = na_1(z-z_0)^{n-1} + (n+1)a_2(z-z_0)^n + \cdots \tag{455}$$

where n may be positive (zero) or negative (pole). Near a pole or zero the ratio has the form

$$\frac{\frac{df}{dz}}{f(z)} = \frac{na_1(z-z_0)^{n-1} + (n+1)a_2(z-z_0)^n + \cdots}{a_1(z-z_0)^n + a_2(z-z_0)^{n+1} + \cdots} = \frac{n}{z-z_0} + \text{analytic function} \tag{456}$$

Multiplying by $g(z)$ gives in the neighborhood of a zero or pole

$$g(z)\frac{\frac{df}{dz}}{f(z)} = \frac{ng(z)}{z-z_0} + \text{analytic function} \tag{457}$$

If we apply the residue theorem to this expression we get

$$\oint g(z)\frac{\frac{df}{dz}}{f(z)} = \sum 2\pi i\left(\sum_{j=1}^{M} m_j g(z_j) - \sum_{j=1}^{N} n_j g(p_j)\right) \tag{458}$$

which is the desired result.

We use this result to count the zeros of an n-th degree polynomial. It is obvious that for large z and n-th degree polynomial grows like $|z|^n$. This means that all of the zeros must lie inside of a circle of sufficiently large radius R. From the above theorem

$$\frac{1}{2\pi 1}\oint \frac{\frac{dP_n}{dz}}{P_n(z)} = \sum_{j=1}^{M} m_j \tag{459}$$

which adds up to the total number of zeros if we count order m zeros as m zeros. On the other hand as the circle gets very large

$$\frac{1}{2\pi 1}\oint \frac{\frac{dP_n}{dz}}{P_n(z)} \to \tag{460}$$

$$\frac{1}{2\pi 1}\oint \frac{n}{z}dz = n \tag{461}$$

where n is the degree of the polynomial.

This proves the fundamental theorem of algebra - i.e. that any polynomial of degree n has n complex roots.

A powerful technique for approximating integrals involving integrals of the form

$$I(w) = \int_C e^{wf(z)} g(z) dz \tag{462}$$

for very large $|w|$ is called the method of saddle point integration. The intuitive observation is that when $|w|$ is very large the major contribution to the above integral is due to the points on the curve where $Re(f(z)w)$ is largest. Near these points we also have to deal with high frequency oscillations.

These integrals were first studied by Debye. He deformed the path C so one a part of the deformed path C_0

a. $\Im(f(z))$ is locally constant.

b. There is a point $z_0 \in C_0$ where

$$\frac{df}{dz}(z_0) = 0 \tag{463}$$

c. At $z = z_0$ along the path $\Re(z)$ goes through a relative maximum.

To understand the precise meaning of these conditions consider the behavior of an analytic function in the neighborhood of a point z_0 where its first complex derivative is zero:

$$f(z) = f(z_0) + \frac{1}{2}\frac{d^2 f}{dz^2}(z_0)(z - z_0)^2 + \cdots \tag{464}$$

We first assume that $\frac{d^2 f}{dz^2}(z_0) \neq 0$. Let

$$z - z_0 = re^{i\phi} \tag{465}$$

and

$$\frac{1}{2}\frac{d^2 f}{dz^2} = Re^{i\psi} \tag{466}$$

Then we have for sufficiently small r the approximation:

$$f(z) = f(z_0) + Rr^2 e^{i\psi + 2i\phi} + \cdots =$$

$$f(z_0) + Rr^2(\cos(\psi + 2i\phi) + i\sin(\psi + 2i\phi) + \cdots \tag{467}$$

Express this in terms of the real and imaginary parts of $f(z) = u(z) + iv(z)$:

$$u(z) = u(z_0) + Rr^2(\cos(\psi + 2i\phi) + \cdots \tag{468}$$

$$v(z) = v(z_0) + Rr^2(\sin(\psi + 2i\phi) + \cdots \tag{469}$$

The condition that $f(z)$ has a relative maximum at $f(z_0)$ means that the curve is designed to goes through z_0 in the direction given by

$$\cos(\psi + 2\phi) = -1 \qquad \sin(\psi + 2\phi) = 0 \tag{470}$$

which means that

$$\psi + 2\phi = (2n + 1)\pi \tag{471}$$

$$\psi = \frac{1}{2}((2n + 1)\pi - \phi) = n\pi + \frac{\pi}{2} - \frac{\phi}{2} \tag{472}$$

Along this path the imaginary part of the function is constant to third order, which minimizes oscillations, while the real part reaches a local maximum *along the deformed curve* at $z = z_0$.

In this case the integral is replaced by

$$I(w) = \int_C e^{wf(z)} g(z) dz = \int_{C_0} e^{wf(z)} g(z) dz \approx \tag{473}$$

On C_0 near z_0 we have

$$\frac{w}{2}\frac{d^2f}{dz^2}(z_0)(z-z_0)^2 = -w\tau^2 \tag{474}$$

which gives

$$I(w) \approx e^{wf(z_0)}\int_{C_0} e^{-w\tau^2} g(z(\tau))\frac{dz}{d\tau} \tag{475}$$

29:171 - Homework Assignment #7

Compute the following integrals

1. $$\int_0^\infty \frac{\ln(x)}{b^2 + x^2} dx$$

2. $$P \int_{-\infty}^{\infty} \frac{1}{(x - x_0)^2 + a^2} \frac{1}{(x - x_1)} dx$$

3. $$\int_0^\infty \frac{dx}{a^3 + x^3}$$

4. $$\int_0^\infty \frac{dx}{(a^3 + x^3)^2}$$

5. $$\int_{-\infty}^{\infty} \frac{e^{ikx} dx}{(x^2 + a^2)}$$

6. $$\int_C e^{1/z} dz \qquad C = \text{unit circle}$$

7. Calculate

$$\sum_{n=1}^{\infty} \frac{1}{n^2}$$

and

$$\sum_{n=1}^{\infty} (-1)^{n+1} \frac{1}{n^2}$$

0.22 Lecture 22

The last step is to extend the integral from the small region of C_0 near z_0 to $(-\infty, \infty)$. We also expand

$$g(z(\tau))\frac{dz}{d\tau} = \sum_n c_n \tau^n \tag{476}$$

giving the approximation

$$I(w) \approx e^{wf(z_0)} \int_{-\infty}^{\infty} e^{-w\tau^2} \sum_n c_n \tau^n d\tau \tag{477}$$

The integrals are zero for odd n and for even n they are determined by

$$I_{2n} := \int e^{-w\tau^2} \tau^{2n} d\tau = \left(-\frac{d}{dw}\right)^n \int e^{-w\tau^2} d\tau \left(-\frac{d}{dw}\right)^n \sqrt{\frac{\pi}{w}} \tag{478}$$

$$\sqrt{\pi}\frac{1}{2} \cdot \frac{5}{2} \cdots \frac{2n-1}{2} w^{-(2n+1)/2} \tag{479}$$

Thus

$$I(w) \approx e^{wf(z_0)} \sum_{n=0}^{\infty} I_{2n} \tag{480}$$

In general this series may not converge. It is often the case that the series generated using the method of steepest decent is an asymptotic series. We write the series

$$\sum_{n=0}^{\infty} f_n z^{-n} \tag{481}$$

is asymptotic to $f(z)$ if for any n:

$$\lim_{|z| \to \infty} \{z^n (f(z) - \sum_{k=0}^{n} f_k z^{-k})\} = 0 \tag{482}$$

What this equation means is that for any fixed n

$$|f(z) - \sum_{k=0}^{n} f_k z^{-k}| < \frac{\epsilon}{|z|^n} \tag{483}$$

which means the the error by using the finite sum can be made as small as desired by choosing z large enough.

Unfortunately this does *not* mean that the full series converges for any fixed z. While asymptotic series are not always convergent, a given function can have at most one asymptotic series. In addition, different functions can have the same asymptotic series.

Example: The method of steepest descent is best illustrated by example. Consider the integral:

$$I(w) := \int_0^\infty e^{-z} z^w \tag{484}$$

where w is large. First transform this integral to an integral of the desired form by letting $z = w\nu$:

$$I(w) := \int_0^\infty e^{-w\nu} w^w \nu^w w d\nu = w^{w+1} \int_0^\infty e^{w(\ln(\nu)-\nu)} d\nu \tag{485}$$

In this example $f(\nu) = \ln(\nu) - \nu$. Note that

$$\frac{df}{d\nu}(\nu) = \frac{1}{\nu} - 1 \tag{486}$$

which vanishes when $\nu = 1$. Expanding about $\nu = 1$

$$f(\nu) = -1 - \frac{1}{2\nu^2_{\nu=1}}(\nu-1)^2 + \cdots = 1 - \frac{1}{2}(\nu-1)^2 + \cdots \tag{487}$$

If we let $\nu = -1 + re^{i\phi}$ we get

$$f(\nu) = -1 = 1 - \frac{1}{2} r^2 e^{2i\phi} + \cdots \tag{488}$$

Choose $\phi = 0, \pi, \cdots$ which gives

$$I(w) = w^{w+1} \int_0^\infty e^{-w - \frac{w}{2} r^2} dr = e^{-w} e^{w+\frac{1}{2}} \sqrt{2\pi} \tag{489}$$

In general this is only an approximation that is expected to improve as w gets large.

The Gamma and Beta functions

Special functions are an important element of mathematical physics. These functions can be expressed in many different forms. Integral representations are often useful, especially when considering analytic properties of these functions.

The Gamma function is defined for all values of z by the following integral representation

$$\frac{1}{\Gamma(z)} := \frac{1}{2\pi i}\int_C \frac{e^t}{t^z}dt. \tag{490}$$

The function

$$t^z = e^{z\ln t} \tag{491}$$

has branch points in t at zero and ∞ unless z is an integer. It is analytic in z when $t \neq 0$. When t is not an integer $\ln(t)$ has branch points at 0 and ∞. The branch cut is along the negative real axis, from $x = -\infty$ to $x = 0$. The contour C in the above expression goes from $-\infty$ to 0 below the cut and from 0 to $-\infty$ above the branch point. The term e^t falls off exponentially for large negative t. The curve also includes a circle around the origin that avoids the point $t = 0$ When $z = n$ is an integer this curve is equivalent to integrating around a closed circle about the origin. When $z = -n$ we have

$$\frac{1}{\Gamma(-n)} := \frac{1}{2\pi i}\int_C e^t t^n dt = 0 \tag{492}$$

by Cauchy's theorem; when n is a positive integer

$$\frac{1}{\Gamma(n)} := \frac{1}{2\pi i}\int_C \frac{e^t}{t^n}dt = \frac{2\pi i}{2\pi i}\frac{1}{(n-1)!} \tag{493}$$

Taking inverses gives

$$\Gamma(n) = \infty \qquad n < 0 \tag{494}$$

$$\Gamma(n+1) = n! \qquad n \geq 0 \tag{495}$$

The properties of this integral representation imply that $\frac{1}{\Gamma(z)}$ is defined and differentiable for all z. $\Gamma(z)$ is analytic, except to poles alone the negative real axis.

An important property of the Gamma function is

$$\Gamma(z+1) = z\Gamma(z) \tag{496}$$

To prove this note that

$$\frac{1}{\Gamma(z+1)} = \int_c e^t t^{-(z+1)} dt = -\frac{1}{z}\int_c e^t \frac{d}{dt} t^{-(z)} \tag{497}$$

Integrate by parts noting that the endpoint contributions on the curve C have the real part of $t = -\infty$ where the integrand vanishes for all z.

$$\frac{1}{\Gamma(z+1)} = \int_c e^t t^{-(z+1)} dt = \frac{1}{z} \int_c e^t t^{-(z)} = \frac{1}{z\Gamma(z)} \tag{498}$$

Taking inverses gives the desired result:

$$\Gamma(z+1) = z\Gamma(z) \tag{499}$$

0.23 lecture 23

Next we show that there is an alternative integral representation for the Γ function given by

$$\Gamma_1(z) = \int_0^\infty e^{-t} t^{z-1} dt \qquad \Re(z) > 0 \tag{500}$$

that is valid for z in the right half plane.

Again using integration by parts

$$\Gamma_1(z+1) = \int_0^\infty e^{-t} t^z dt = \int_0^\infty (-\frac{d}{dt}) e^{-t} t^z dt$$

$$\int_0^\infty e^{-t} \frac{d}{dt} t^z dt = z\Gamma_1(z) \qquad (z \neq 0) \tag{501}$$

where we have used that the endpoint contributions vanish for z in the right half plane.

Next I argue that the two representations, $\Gamma(z)$ and $\Gamma_1(z)$, of the Γ function are equal. To do this we first define the β functions for $\Re(a), \Re(b) > 0$:

$$\beta(a,b) := \int_0^1 t^{a-1}(1-t)^{b-1} dt \tag{502}$$

Let $t \to \frac{1}{\tau}$ and $t \to \sin^2(\theta)$ in the above expression to get two equivalent expressions for the β function:

$$\beta(a,b) = \int_1^\infty t^{-a-b}(t-1)^{b-1} dt = 2\int_0^{\frac{\pi}{2}} \sin^{2a-1}(\theta)\cos^{2b-1}(\theta) d\theta \tag{503}$$

Next note that

$$\Gamma(a)\Gamma(b) = \int_0^\infty dt dw e^{-t-w} t^{a-1} w^{b-1} \tag{504}$$

Next let $t = y^2$ $w = x^2$ to get

$$\Gamma(a)\Gamma(b) = 4\int_0^\infty dx dx e^{-x^- y^2} x^{2a-1} y^{2b-1} \tag{505}$$

Next change to polar coordinates

$$\Gamma(a)\Gamma(b) = 4\int_0^\infty r^{2a+2b-1} e^{-r^2} dr \int_0^{\frac{\pi}{2}} d\theta \sin^{2a-1}(\theta)\cos^{2b-1}(\theta) \tag{506}$$

Finally let $u = r^2$ to get

$$\Gamma(a)\Gamma(b) = \int_0^\infty u^{a+b-1} e^{-u} du 2 \int_0^{\frac{\pi}{2}} d\theta \sin^{2a-1}(\theta) \cos^{2b-1}(\theta) \tag{507}$$

Comparing with (503) we get

$$\Gamma(a)\Gamma(b) = \Gamma(a+b)\beta(a,b) \tag{508}$$

or

$$\beta(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \tag{509}$$

which is all in terms of our second expression for the Γ function.

Let $a = z$ $b = 1 - z$ so $a + b = 1$. Then the above becomes

$$\Gamma(z)\Gamma(1-z) = \int_0^\infty e^{-u} du 2 \int_0^{\frac{\pi}{2}} d\theta \sin^{2z-1}(\theta) \cos^{1-2z}(\theta) = \tag{510}$$

$$2 \int_0^{\frac{\pi}{2}} d\theta \tan^{2z-1}(\theta)(\theta) \tag{511}$$

Let $x = \tan(\theta)$, $dx = (1 + x^2)d\theta$ to get

$$2 \int_0^\infty d\theta \frac{x^{2z-1}}{x^2+1} dx = \pi \csc(\pi z) \tag{512}$$

which is the integral done is section 18 for real z between 0 and 1 . This extends to $x \to z$ at points of analyticity.

One convenient by-product of this formula is for $z = 1/2$ we get

$$\Gamma(\frac{1}{2})^2 = \pi \tag{513}$$

or

$$\Gamma(\frac{1}{2}) = \sqrt{\pi}. \tag{514}$$

To show the desired equivalence change variables $t \to -t' = e^{-i\pi} t'$ in the expression of $\Gamma(z)$:

$$\frac{1}{\Gamma(1-z)} = \frac{1}{2\pi i} \int_C e^t t^{-(1-z)} dt = -\frac{1}{2\pi i} \int_C \frac{e^{-t'}}{t'^{(1-z)}} e^{i\pi(1-z)} dt' \tag{515}$$

In this transformation the phase of bottom half of the original curve goes from $e^{-i\pi} \rightarrow e^{i\pi}$ to 0 to $e^{2\pi i}$ and the curve remains in the counter clockwise direction. This can be broken up into integrals alone the two semi infinite lines and the circle:

$$-\frac{1}{2\pi i}\int_{\infty}^{0} \frac{e^{-t'}}{t'^{(1-z)}} e^{i\pi(1-z)} dt$$
$$-\frac{1}{2\pi i}\int_{0}^{\infty} \frac{e^{-t'}}{t'^{(1-z)}} e^{i\pi(1-z)} e^{i2\pi(z-1)} dt$$
$$-\frac{1}{2\pi i}\int_{0}^{2\pi} \frac{e^{-re^{i\phi}}}{r^{(1-z)}} e^{i\phi(z-1)} e^{i\pi(1-z)} ir d\phi \qquad (516)$$

The integral over the circle vanishes when $\Re(z) > 0$ as the radius $\rightarrow 0$. The first two integrals can be combined to get

$$\frac{1}{2\pi i} e^{i\pi(1-z)} \int_{0}^{\infty} [\frac{e^{-t}}{t^{(1-z)}} (1 - e^{i2\pi(z-1)}) dt =$$
$$\frac{1}{2\pi i}\int_{0}^{\infty} \frac{e^{-t}}{t^{(1-z)}} [e^{i\pi(1-z)} - e^{-i\pi(1-z)}] dt =$$
$$-\frac{1}{2\pi i}\int_{0}^{\infty} e^{-t} t^{z-1} [e^{-i\pi z} - e^{i\pi z}] dt =$$
$$\frac{\sin(\pi z)}{\pi}\int_{0}^{\infty} e^{-t} t^{z-1} dt =$$
$$\frac{\sin(\pi z)}{\pi}\Gamma_1(z) = \frac{1}{\Gamma_1(z)} \qquad (517)$$

This shows that

$$\Gamma_1(z) = \Gamma(z) \qquad (518)$$

for $\Re(z) > 0$. Since both Γ functions satisfy $\Gamma(z+1) = z\Gamma(z)$, we can alway choose n large enough to $Re(n+z) > 0$, which then implies the equality in the plane provided z in not a negative integer.

0.24 Lecture 24:

This section begins the study of vectors and vector spaces. Following the text, I will use a notation that was introduced by Dirac for quantum mechanics. Also, since I will be using both finite and infinite dimensional vector spaces, what I discuss applies to all types of vector spaces, unless specifically stated.

Vectors, $|a\rangle$, $|b\rangle$, $|c\rangle$, $\cdots$ are elements of a set S. There are two operations defined on vectors.

Vector addition: Adding vectors $|a\rangle$ and $|b\rangle$ gives a new vector $|c\rangle$. Vector addition is expressed as

$$|a\rangle + |b\rangle = |c\rangle. \tag{519}$$

Scalar multiplication: If $|a\rangle$ is a vector and α is a complex number then a new vector $|c\rangle$ is defined by

$$|c\rangle = \alpha|a\rangle. \tag{520}$$

I also assume the existence of some special vectors.

Zero vector: The zero vector, $|0\rangle$, satisfies

$$|a\rangle + |0\rangle = |a\rangle \tag{521}$$

for any vector $|a\rangle \in S$.

Inverse vector: Given a vector $|a\rangle \in S$, the inverse vector $|-a\rangle \in S$ satisfies

$$|a\rangle + |-a\rangle = |0\rangle. \tag{522}$$

A complex vector space is a set S with the operations of vector addition and scalar multiplication, where S has a zero vector, every vector in S has an inverse in S, and the rules of vector addition and scalar multiplications for vectors $|a\rangle$, $|b\rangle$, $|c\rangle \in S$ and complex numbers $\alpha, \beta \in \mathbb{C}$ are:

1.)

$$(|a\rangle + |b\rangle) + |c\rangle = |a\rangle + (|b\rangle + |c\rangle) \tag{523}$$

$$|a\rangle + |b\rangle = |b\rangle + |a\rangle \tag{524}$$

2.)

$$1 \cdot |a\rangle = |a\rangle \tag{525}$$

3.)

$$\beta(\alpha|a\rangle) = (\alpha\beta)|a\rangle \tag{526}$$

$$(\alpha+\beta)|a\rangle) = (\alpha|a\rangle) + (\beta|a\rangle) \tag{527}$$

$$\alpha(|a\rangle + |b\rangle) = (\alpha|a\rangle) + (\alpha|b\rangle) \tag{528}$$

Examples of vector spaces are below:
Example 1: Complex numbers with addition and multiplication of complex numbers:

$$|a\rangle = x + iy \tag{529}$$

Example 2: Complex 2×2 matrices

$$|a\rangle = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \tag{530}$$

where $\alpha, \beta, \gamma, \delta \in \mathbb{C}$ and addition is addition of complex matrices and scalar multiplication is multiplication of the matrix by a complex constant.
Example 3: Degree 2 polynomials:

$$|a\rangle = \alpha + \beta z + \gamma z^2 \tag{531}$$

where $\alpha, \beta, \gamma \in \mathbb{C}$ and addition and scalar multiplication are addition and scalar multiplication of polynomials.

There are some elementary consequences that follow directly from the definition of a vector space.
The vector $|0\rangle$ is unique. Assume that there are two zero vectors, $|0_1\rangle$ and $|0_2\rangle$, and let $|a\rangle$ and $|-a\rangle$ be inverses. Then

$$|0_1\rangle = |a\rangle + |-a\rangle = |0_2\rangle \tag{532}$$

Therefore

$$|0_1\rangle = |0_2\rangle. \tag{533}$$

Given a vector $|a\rangle$ the inverse $|-a\rangle$ is unique. Assume that $|a\rangle$ has two inverses, $|-a_1\rangle$ and $|-a_2\rangle$. Then

$$|a\rangle + |-a_1\rangle = |0\rangle = |a\rangle + |-a_2\rangle. \tag{534}$$

Add $|-a_1\rangle$ to both sides of this equation to get

$$|0\rangle + |-a_1\rangle = |0\rangle + |-a_2\rangle \tag{535}$$

or

$$| - a_1\rangle = | - a_2\rangle. \tag{536}$$

$0|a\rangle = |0\rangle$: To show this note

$$|a\rangle = 1|a\rangle = (1 + 0)|a\rangle = 1|a\rangle + 0|a\rangle = |a\rangle + 0|a\rangle \tag{537}$$

Add the inverse $| - a\rangle$ of $|a\rangle$ to both sides to get

$$| - a\rangle + |a\rangle = | - a\rangle + |a\rangle + 0|a\rangle \tag{538}$$

which becomes

$$|0\rangle = |0\rangle + 0|a\rangle = 0|a\rangle. \tag{539}$$

The inverse $| - a\rangle$ of $|a\rangle$ is

$$| - a\rangle = (-1)|a\rangle. \tag{540}$$

To prove this note that

$$|a\rangle + (-1)|a\rangle = 1|a\rangle + (-1)|a\rangle = (1 - 1)|a\rangle = 0|a\rangle = |0\rangle. \tag{541}$$

Given the identification above it is customary to define

$$-|a\rangle = (-1)|a\rangle. \tag{542}$$

These properties apply to any vector space.

Just because a set of vectors with a definition of addition and scalar multiplication form a vector space, it does not automatically follow that the vectors have a length or scalar product.

A vector space is a metric space if there is real valued function, $\rho(\cdot, \cdot)$, defined on pairs of vectors satisfying

1. $\rho(|a\rangle, |b\rangle) \geq 0$
2. $\rho(|a\rangle, |b\rangle) = 0 \iff |a\rangle = |b\rangle$
3. $\rho(|a\rangle, |b\rangle) = \rho(|b\rangle, |a\rangle)$
4. $\rho(|a\rangle, |b\rangle) \leq \rho(|a\rangle, |c\rangle) + \rho(|c\rangle, |b\rangle)$

Not that metric spaces do not have to be vector spaces.
Example 1: Let S be the set of points on the surface of a unit three dimensional sphere. Define a metric on this space by

$$\rho(|a\rangle, |b\rangle) = \text{minimum arc length of curve between } |a\rangle \text{ and } |b\rangle \text{ on sphere} \tag{543}$$

It follows from this definition that

$$\rho(|b\rangle, |a\rangle) \geq 0 \tag{544}$$

$$\rho(|a\rangle, |b\rangle) = 0 \iff |a\rangle = |b\rangle \tag{545}$$

$$\rho(|a\rangle, |c\rangle) + \rho(|c\rangle, |b\rangle) \geq \rho(|a\rangle, |b\rangle) \tag{546}$$

This means the $\rho(\cdot, \cdot)$ is a metric on S. This metric space is not a vector space.

A set of vectors is a normed linear space if there is real valued function, $\|\cdot\|$, defined on vectors satisfying

1. $\||a\rangle\| \geq 0$
2. $\||a\rangle\| = 0 \iff |a\rangle = 0$
3. $\|\alpha|a\rangle\| = |\alpha|\||a\rangle\|$
4. $\|(|a\rangle + |b\rangle)\| \leq \||a\rangle\| + \||b\rangle\|$

A set of vectors is a inner product space if there is complex valued function, $\langle\cdot|\cdot\rangle$, defined on pairs vectors satisfying

1. $\langle a|b\rangle^* = \langle b|a\rangle$
2. For $|c\rangle = |a\rangle + |b\rangle$
$$\langle d|c\rangle = \langle d|a\rangle + \langle d|b\rangle \tag{547}$$
3. For $|c\rangle = \alpha|a\rangle$, $\alpha \in \mathbb{C}$
$$\langle b|c\rangle = \alpha\langle b|a\rangle \tag{548}$$
4.
$$\langle a|a\rangle \geq 0 \tag{549}$$
$$\langle a|a\rangle = 0 \equiv |a\rangle = |0\rangle \tag{550}$$

All inner product spaces are normed linear spaces with the norm

$$\||a\rangle\| = \langle a|a\rangle^{1/2} \tag{551}$$

and all normed linear spaces are metric spaces with the metric

$$\rho(|a\rangle, |b\rangle) = \|(|a\rangle - |b\rangle)\|. \tag{552}$$

The distinctions are important because there are important metric spaces that are not normed spaces, and important normed linear spaces that are not inner product spaces. Most examples where these distinctions are relevant involve infinite dimensional vector spaces, which will discussed later.

To show that all normed linear spaces are metric spaces we use

$$\rho(|a\rangle, |b\rangle) := \|(|a\rangle - |b\rangle)\| \geq 0 \tag{553}$$

$$\rho(|a\rangle, |b\rangle) := \|(|a\rangle - |b\rangle)\| = 0 \qquad \Longleftrightarrow \qquad |a\rangle = |b\rangle \tag{554}$$

$$\rho(|a\rangle, |b\rangle) := \|(|a\rangle - |b\rangle)\| = \|(|b\rangle - |a\rangle)\| = \rho(|b\rangle, |a\rangle) \tag{555}$$

$$\rho(|a\rangle, |b\rangle) + \rho(|b\rangle, |c\rangle) := \|(|a\rangle - |b\rangle)\| + \|(|b\rangle - |c\rangle)\| \geq \tag{556}$$

$$\|(|a\rangle - |c\rangle)\| = \rho(|a\rangle, |c\rangle) \tag{557}$$

To show that all inner product spaces are normed linear spaces we first prove the Cauchy Schwartz inequality.

Consider the vector

$$|c\rangle = |a\rangle - \lambda\langle b|a\rangle|b\rangle \tag{558}$$

It follows that

$$\langle c|c\rangle = (\langle a| - \lambda^*\langle b|a\rangle^*\langle b|)(|a\rangle - \lambda\langle b|a\rangle|b\rangle) = \tag{559}$$

$$\langle a|a\rangle - \lambda\langle b|a\rangle\langle a|b\rangle - \lambda^*\langle b|a\rangle^*\langle b|a\rangle + |\lambda|^2\langle b|a\rangle\langle a|b\rangle\langle b|b\rangle \geq 0 \tag{560}$$

This inequality holds for any λ. Restrict to the case that λ is real. The inequality also holds for all real λ. This means that this polynomial in real λ can have not real roots in λ The roots of this polynomial in λ

$$r = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \tag{561}$$

This polynomial will have no real roots if

$$b^2 - 4ac < 0 \tag{562}$$

where

$$b = -2|\langle a|b\rangle|^2 \tag{563}$$

$$a = \langle a|a\rangle \tag{564}$$

$$c = \langle b|b\rangle|\langle a|b\rangle|^2 \tag{565}$$

This condition is equivalent to

$$\sqrt{\langle a|a\rangle}\sqrt{\langle b|b\rangle} \geq |\langle a|b\rangle| \tag{566}$$

This is the Cauchy-Schwartz inequality. It is a property of any inner product space.

Using the Cauchy Schwartz inequality it is possible to show that every inner product space is a normed linear space. First note

$$\||a\rangle\| := \langle a|a\rangle^{1/2} \geq 0 \tag{567}$$

$$0 = \||a\rangle\| = \langle a|a\rangle^{1/2} \iff |a\rangle = |0\rangle \tag{568}$$

$$\|\alpha|a\rangle\| := \langle \alpha a|\alpha a\rangle^{1/2} = (\alpha^*\alpha)^{1/2}\langle a|a\rangle^{1/2} = |\alpha|\||a\rangle\| \tag{569}$$

$$\||a+b\rangle\|^2 = \langle a+b|a+b\rangle = \tag{570}$$

$$\langle a|a\rangle + \langle b|b\rangle + \langle a|b\rangle + \langle b|a\rangle \leq \tag{571}$$

$$\langle a|a\rangle + \langle b|b\rangle + 2|\langle a|b\rangle| \leq \tag{572}$$

$$\langle a|a\rangle + \langle b|b\rangle + 2\sqrt{\langle a|a\rangle\langle b|b\rangle}| = (\sqrt{\langle a|a\rangle} + \sqrt{\langle b|b\rangle})^2 \tag{573}$$

Taking square root of both sides gives the triangle inequality

$$\|(|a\rangle + |b\rangle)\| \leq \||a\rangle\| + \||b\rangle\| \tag{574}$$

29:171 - Homework Assignment #8

1. By integrating
$$\int \frac{zdz}{a - e^{-iz}}$$
over a rectangular curve with corners at $-\pi$, π, $\pi + in$ $-\pi + in$ and letting $n \to \infty$ show
$$\int_0^\pi \frac{x\sin(x)dx}{1 + a^2 - 2a\cos(x)} = \frac{\pi}{a}\log(1+a) \qquad (0 < a < 1)$$

2. Evaluate
$$\int_0^\infty \frac{\ln^2(z)}{z^2+1}dz$$

3. Express the integral
$$\int_0^\infty e^{-\alpha x^2} x^\beta dx$$
where α and β are real and positive in terms of the Gamma function.

4. Prove that if $a > 0, -\frac{1}{2}\pi < a\lambda < \frac{1}{2}\pi$
$$\int_0^\infty e^{-r^a\cos(a\lambda)}\cos(r^a sin(a\lambda))dr = \cos(\lambda)\frac{1}{a}\Gamma(\frac{1}{a})$$

5. Calculate
$$\int_0^{\pi/2} \sin^\alpha(\theta)\cos^\beta(\theta)d\theta$$
for $\alpha, \beta > 0$.

6. Evaluate $\beta(m,n)$ and relate it to the binomial coefficients.

0.25 Lecture 25

Example 2: Let S be the set of continuous complex valued functions of a real variable in the interval $[0, 1]$ is a vector space. The function,

$$\|f\| = \sup_{x \in [0,1]} |f(z)| \tag{575}$$

where sup means least upper bound, is a norm on this space. This norm cannot be constructed from an inner product in the manner discussed above.
Example 3: Let $S = \mathbb{C}$. Then

$$\langle a|b\rangle = a^* b \tag{576}$$

is an inner product on the vector space of complex numbers.
Example 4: Let S the vector space of complex 2×2 matrices. If

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \qquad \text{and} \qquad A^\dagger = \begin{pmatrix} a_{11}^* & a_{21}^* \\ a_{12}^* & a_{22}^* \end{pmatrix} \tag{577}$$

Then

$$\langle A|B\rangle = \mathrm{Tr}(A^\dagger B) \tag{578}$$

is a scalar product. Here AB is the matrix product and $\mathrm{Tr}(A)$ is the sum of diagonal elements.
Example 5: Let S the vector space of second degree complex polynomials. Then

$$\langle P|Q\rangle = \int_0^1 dx P^*(x) Q(x) \tag{579}$$

is a scalar product.

From these examples it is clear the there are many different types of vector spaces. Our goal is to exploit the common features of these different looking spaces.
Definition: A vector space with a metric $\rho(|a\rangle, |b\rangle)$ is complete if every Cauchy sequence of vectors converges to a vector in the space.

Since all vectors spaces with norms or scalar products are also metric spaces, we have the following definitions.
Definition: A complete normed linear space is called a Banach space.
Definition: A complete inner product space is called a Hilbert space.

Spaces that are not complete can be made complete including new vectors defined as Cauchy sequences of vectors in the space. This is exactly how the real numbers are constructed by completing the rational numbers.

Consider a vector space with an inner product. The inner product of the vectors

$$|a\rangle := \beta|b\rangle + \gamma|c\rangle \tag{580}$$

$$|a'\rangle := \delta|d\rangle + \epsilon|e\rangle \tag{581}$$

is

$$\langle a'|a\rangle = \beta\delta^*\langle d|b\rangle + \beta\epsilon^*\langle e|b\rangle + \gamma\delta^*\langle d|c\rangle + \gamma\epsilon^*\langle e|c\rangle. \tag{582}$$

This can be factored as

$$\langle a'|a\rangle = \delta^*\langle d|a\rangle + \epsilon^*\langle e|a\rangle = \tag{583}$$

$$\beta\langle a'|b\rangle + \gamma\langle a'|c\rangle \tag{584}$$

This shows that the inner product is linear in the arguments on the right while it is conjugate linear in the arguments on the left.

It is possible to treat this more symmetrically by introducing a space S^* of dual vectors denoted by $\langle a|$.

Dual vectors form a vector space by definition. In particular linear combinations of dual vectors with complex coefficients are dual vectors. There is no complex conjugation involved.

Each dual vector $\langle a|$ is associated with a linear functional on the original vector space. A linear functional is an operator L that gives a complex number when it is applied to a vector. It also satisfies the linearity condition

$$L(|b\rangle + \gamma|c\rangle) = L(|b\rangle) + \gamma L(|c\rangle) \tag{585}$$

The linear functional associated with the dual vector $\langle a|$ on an inner product space is

$$L_{\langle a|}(|b\rangle) = \langle a|b\rangle \tag{586}$$

This dual vector acts linearly on S:

$$L_{\langle a|}(|b\rangle + \gamma|c\rangle) =$$

$$L_{\langle a|}|b\rangle + \gamma L_{\langle a|}|c\rangle =$$

$$\langle a|b\rangle + \gamma\langle a|c\rangle \tag{587}$$

Consistency with the previous definitions requires

$$L_{\langle(a+\beta b|} = L_{\langle a|} + \beta^* L_{\langle b|} \tag{588}$$

which means that the relation that assigns a dual vector to a vector is not linear.

Note that while the above discussion applies to dual vectors in an inner product space, the concept of a continuous linear functional makes sense on all vector spaces. These functionals define the dual space of any vector space, but in the general case there is not necessarily a $1-1$ correspondence between vector and dual vectors.

Operators:

Given two vectors spaces, S_1 and S_2, an operator, f, is a function that assigns vectors in a subset D of S_1 to vectors in a subset R of S_2. This is written

$$f : D \to R \tag{589}$$

The set $D \subset S_1$ is called the domain of f The set $R \subset S_2$ is called the range of f.

The function f is onto S_2 if $R = S_2$. It is one to one if $|a\rangle \neq |b\rangle$ implies $f(|a\rangle) \neq f(|b\rangle)$.

This section primarily concerns the class of linear operators. An operator $f : D \subset S_1 \to R \subset S_2$ is linear if for and $|a\rangle, \quad |b\rangle \in D$ and $\alpha \in \mathbb{C}$

$$f(\alpha|a\rangle + |b\rangle) = \alpha f(|a\rangle) + f(|b\rangle) \tag{590}$$

The following vector space operations are defined on linear operators

addition of linear operators

$$(f_1 + f_2)(|a\rangle) = f_1(|a\rangle) + f_2(|a\rangle) \tag{591}$$

multiplication of linear operators by complex numbers

$$(\alpha f_1(|a\rangle)) = \alpha f_1(|a\rangle) \tag{592}$$

composition of linear operators: Let $f_1 : D_1 \subset S_1 \to R_2 \subset S_2$ and let $f_2 : D_2 = R_2 \subset R_3 \subset S_3$. Then

$$f_2 \cdot f_1(|a\rangle) := f_2(f_1(|a\rangle)) \tag{593}$$

In general if $f_1 \cdot f_2$ is defined it does not follow that $f_2 \cdot f_1$ is defined. If they are both defined,in general $f_1 \cdot f_2 \neq f_2 \cdot f_1$. This means that multiplication of linear operators is not generally commutative.

Two linear operators f_1 and f_1 are equal if $D_1 = D_2$, $R_1 = R_2$ and for every $|a\rangle \in D_1$

$$f_1(|a\rangle) = f_2(|a\rangle) \tag{594}$$

The zero operator, O maps every vector in S_1 to the zero vector $|0\rangle_2$ in S_2:

$$O|a\rangle = |0\rangle_2 \tag{595}$$

In what follows I assume that $D = S_1 = S_2$. I will also use upper case Latin letters to represent linear functions:

$$f \to A \tag{596}$$

$$f_1 \to A_1 \qquad f_2 \to A_2 \Rightarrow f_2 \cdot f_1 \to A_2 A_1 \tag{597}$$

If $D = R = S$ then the identity operator, I, is the the linear operator that satisfies

$$I|a\rangle = |a\rangle \qquad \forall |a\rangle \in S \tag{598}$$

The following identities are consequences of the definitions

$$A0 = 0A = 0 \qquad IA = AI = A \tag{599}$$

To prove these note

$$0A|a\rangle = 0|Aa\rangle = |0\rangle \tag{600}$$

$$A0|v\rangle = A|0\rangle = A(|0\rangle - |0\rangle) = |0 \tag{601}$$

$$IA|a\rangle = A|a\rangle = AI|a\rangle \tag{602}$$

The commutator of linear operators A and B is defined by

$$[A, B] := AB - BA \tag{603}$$

The anticommutator of linear operators A and B is defined by

$$\{A, B\} := AB + BA \tag{604}$$

It is obvious from the definitions that

$$[A, B] = -[B, A] \tag{605}$$

$$\{A, B\} = \{B, A\} \tag{606}$$

$$AB = \frac{1}{2}\{B, A\} + \frac{1}{2}[A, B] \tag{607}$$

Example 1: Products of anti-commuting operators commute. Assume

$$\{A_i, A_j\} = 0 \qquad i \neq j$$

Then

$$[A_1A_2, A_3A_4] = A_1A_2A_3A_4 - A_3A_4A_1A_2 =$$

$$A_1A_2A_3A_4 + A_1A_3A_2A_4 - A_1A_3A_2A_4 - A_3A_1A_2A_4 + A_3A_1A_2A_4 + A_3A_1A_4A_2$$

$$-A_3A_1A_4A_2 - A_3A_4A_1A_2 + A_3A_4A_1A_2 - A_3A_4A_1A_2 =$$

$$A_1\{A_2, A_3\}A_4 - \{A_1, A_3\}A_2A_4A_3A_1\{A_2, A_4\}A_3\{A_1, A_4\}A_2 = 0 \tag{608}$$

This identity is very important in quantum field theory. It explains why fermions are observable.

0.26 Lecture 26

Since linear operators are closed under addition, operator multiplication, and multiplication by complex scalars, it is possible to use these operations to create new linear operators.

For example we can construct linear operators that are polynomials in a given linear operator

$$P(A) = p_0 I + p_1 A + p_2 AA + \cdots + p_n \underbrace{AA \cdots AA}_{\text{n-times}} \tag{609}$$

It is customary to define

$$A^n := \underbrace{AA \cdots AA}_{\text{n-times}} \tag{610}$$

so the above polynomial has the traditional form

$$P(A) = p_0 I + p_1 A + p_2 A^2 + \cdots + p_n A^n \tag{611}$$

It is possible to consider polynomials of infinite degree. As in the case of ordinary functions, they can be defined in terms of Cauchy sequences of finite degree polynomials. In order to define a Cauchy sequence it is necessary to find a notion of convergence for sequences of linear operators.

While there are many ways to do this I consider the special case of linear operators on normed linear spaces. On these spaces it is possible to define the norm of a linear operator as follows:

$$\||A|\| := \underbrace{\sup}_{\||v\rangle=1} \|A|v\rangle\| = \underbrace{\sup}_{\||v\rangle\|>1} \frac{\|A|v\rangle\|}{\||v\rangle\|} \tag{612}$$

where sup is the supremum or least upper bound.

It is an immediate consequence of this definition that

$$\|A|a\rangle\| \leq \||A|\| \cdot \||v\rangle\| \tag{613}$$

for any $|a\rangle$

A linear operator A on a normed linear space is bounded if

$$\||A|\| < \infty \tag{614}$$

Bounded operators A on a normed linear space are continuous. To see this note if

$$\|(|a_m\rangle - |a_n\rangle)\| < \epsilon \tag{615}$$

then

$$\|(A|a_m\rangle - A|a_n\rangle)\| = \|A(|a_m\rangle - |a_n\rangle)\| \leq \||A|\|\epsilon \tag{616}$$

This means that

$$\|(|a_m\rangle - |a_n\rangle)\| \to 0 \Rightarrow \|(A|a_m\rangle - A|a_n\rangle)\| \to 0 \tag{617}$$

or that it is permissible to pull limits through bounded operators.

I now show that $\||A|\|$ is a norm on the vector space of bounded operators. By definition

$$\||A|\| \geq 0 \tag{618}$$

if

$$0 = \||A|\| = \|A|v\rangle\| \Rightarrow A|v\rangle = |0\rangle \Rightarrow A = 0 \tag{619}$$

where the zero on the left is complex zero while the zero on the right is the zero operator.

$$\||\alpha A|\| := \underbrace{\sup}_{\||v\rangle=1} \|\alpha A|v\rangle\| = |\alpha| \underbrace{\sup}_{\||v\rangle=1} \|A|v\rangle\| = |\alpha| \cdot \||\alpha A|\| \tag{620}$$

$$\||A+B|\| := \underbrace{\sup}_{\||v\rangle=1} \|A+B|v\rangle\| \leq \underbrace{\sup}_{\||v\rangle=1} \|A|v\rangle\| + \underbrace{\sup}_{\||v\rangle=1} \|B|v\rangle\| = \||A|\| + \||B|\| \tag{621}$$

In addition to addition and multiplication by scalars, operators can be multiplied. It is useful to know how to calculate the operator norm of products of operators

$$\||AB|\| = \underbrace{\sup}_{\||v\rangle=1} \|AB|v\rangle\| \leq \tag{622}$$

$$\||A|\| \underbrace{\sup}_{\||v\rangle=1} \|B|v\rangle\| \leq \tag{623}$$

$$\||A|\|\||B|\| \tag{624}$$

Now that I have a norm on the space of bounded linear operators I can define the exponential function of an operator in terms Cauchy sequences of finite degree polynomials:

$$(e^A)_N = I + \sum_{n=1}^{N} \frac{1}{n!} A^n \tag{625}$$

The limit is the infinite series

$$e^A = I + \sum_{n=1}^{\infty} \frac{1}{n!} A^n \tag{626}$$

It is a simple consequence of the definitions that

$$|||e^A||| = |||I + \sum_{n=1}^{\infty} \frac{1}{n!} A^n||| \tag{627}$$

Since $|||\cdot|||$ is a norm, repeated use of the triangle inequality gives

$$|||e^A||| = \leq 1 + \sum_{n=1} \frac{1}{n!} |||A^n||| \tag{628}$$

Equation (624) implies

$$|||A^n||| \leq |||A|||^n \tag{629}$$

which when used in (628) gives

$$|||e^A||| \leq 1 + \sum_{n=1} \frac{1}{n!} |||A|||^n = e^{|||e^A|||} < \infty \tag{630}$$

This shows that the series converges in norm when applied to any vector in our normed linear space. In addition, this convergence is uniform in the sense that the bound above holds for all vectors and does not depend on the vector.

For the same reason it is possible to define $f(A)$ for any entire function $f(z)$ of a bounded operator A since the functions can always be expressed as a convergent power series.

Certain classes of linear operators appear often in applications.

Inverse operators: If A is a linear operator and there exists another linear operator A_r^{-1} with the property

$$AA_r^{-1} = I \tag{631}$$

then A_r^{-1} is called a right inverse of A.

If A is a linear operator and there exists another linear operator A_l^{-1} with the property

$$A_r^{-1}A = I \tag{632}$$

then A_l^{-1} is called a left inverse of A.

Theorem: If both A_l^{-1} and A_r^{-1} exist then they are unique and $A_l^{-1} = A_r^{-1}$

$$A_{l_1}^{-1}A - A_{l_2}^{-1}A = I - I = 0 \tag{633}$$

$$0 = A_{l_1}^{-1}AA_r^{-1} - A_{l_2}^{-1}AA_r^{-1} == A_{l_1}^{-1}I - A_{l_2}^{-1}I \tag{634}$$

or

$$A_{l_1}^{-1} - A_{l_2}^{-1} \tag{635}$$

An analogous argument shows that A_r^{-1} is unique. The simple calculation

$$A_r^{-1} = A_r^{-1}AA_l^{-1} = A_l^{-1} \tag{636}$$

shows that both operators must be the same.

A linear operator A has an inverse if it has both a left and right inverse. The inverse operator of A is denoted by A^{-1}.

Theorem: If both A^{-1} and B^{-1} exist then $(AB)^{-1}$ exists and

$$(AB)^{-1} = B^{-1}A^{-1} \tag{637}$$

$$B^{-1}A^{-1}AB = B^{-1}IB = I \tag{638}$$

$$ABB^{-1}A^{-1} = AIA^{-1} = I \tag{639}$$

which shows that $B^{-1}A^{-1}$ is a left and right inverse of AB.

If A is a linear operator and z is a complex number then the resolvent operator, $R(z, A)$ of A is defined by

$$R(z, A) = (zI - A)^{-1} \tag{640}$$

when it exists.

It satisfies two important identities called the first resolvent identity:

$$R(z_1, A) - R(z_2, A) = R(z_1, A)(z_2 - z_1)R(z_2, A) = R(z_2)(z_2 - z_1)R(z_1) \tag{641}$$

and the second resolvent identity:

$$R(z,A)-R(z,B) = R(z,B)(A-B)R(z,A) = R(z,B)(A-B)R(z,A). \quad (642)$$

A nice property of the resolvent of a bounded linear operator is that if $R(z,A)$ is bounded at $z = z_0$ then it is analytic as an operator valued function of z for z near z_0. This follows by iterating the first resolvent equation to get the series expansion

$$R(z,A) = R(z_0,A)[I + \sum_{n=1}^{\infty}(z_0 - z)^n R(z_0,A)^n] \quad (643)$$

which converges uniformly for

$$|z_0 - z| \||R(z_0,A)\|| < 1 \quad (644)$$

The points z in the complex plane where $R(z,A)$ is bounded is called the resolvent set of A. I have just demonstrated that $R(z,A)$ is analytic on the resolvent set of A.

When linear operators act on an inner product space there are additional ways to classify linear operators that are important in applications.

Let A be a linear operator on an inner product space. Define the adjoint operator $A^\dagger$ by

$$\langle b|(A|a\rangle) = \big(\langle a|(A^\dagger|b\rangle)\big)^* \quad (645)$$

On inner product spaces we use the notation

$$\langle b|A|a\rangle := \langle b|(A|a\rangle) \quad (646)$$

In this notation equation (645) becomes

$$\langle b|A|a\rangle = \langle a|A^\dagger|b\rangle^* \quad (647)$$

Note that if

$$|c\rangle = A|b\rangle \quad (648)$$

then

$$\langle c|d\rangle = \langle d|c\rangle^* = \langle d|A|b\rangle^* = \langle b|A^\dagger|d\rangle \quad (649)$$

This means that $\langle c| := \langle b|A^\dagger$ is the dual vector to $|c\rangle = A|b\rangle$.

The adjoint operation has a number of elementary properties that follow from the definition.

$(A^\dagger)^\dagger = A$:

$$\langle a|A|b\rangle = \langle b|A^\dagger|a\rangle^* = \langle a|(A^\dagger)^\dagger|b\rangle^{**} = \tag{650}$$

Since this holds for all $|a\rangle$ and $|b\rangle$ is follows that

$$(A^\dagger)^\dagger = A \tag{651}$$

$(AB)^\dagger = B^\dagger A^\dagger$:

$$\langle a|AB|b\rangle = \langle b|(AB)^\dagger|a\rangle^* \tag{652}$$

Let $|c\rangle = B|b\rangle$ and $|d\rangle = A^\dagger|a\rangle$. Then

$$\langle a|AB|b\rangle = \langle a|A|c\rangle = \langle c|A^\dagger|a\rangle^* = \langle d|c\rangle = \tag{653}$$

$$\langle d|B|b\rangle = \langle b|B^\dagger|d\rangle^* = \langle b|B^\dagger A^\dagger|a\rangle^* \tag{654}$$

Comparing (652) and (654) gives $(AB)^\dagger = B^\dagger A^\dagger$
$(A + \beta B)^\dagger = A^\dagger + \beta^* B^\dagger$:

$$\langle b|(A + \beta B)^\dagger|a\rangle^* = \langle a|(A + \beta B)|b\rangle = \langle a|A|b\rangle + \beta\langle a|B|b\rangle = \tag{655}$$

$$\langle b|A^\dagger|a\rangle^* + \beta\langle b|B^\dagger|a\rangle^* \tag{656}$$

Taking complex conjugates of both sides of this equation gives the desired result.

0.27 Lecture 27

The adjoint is an essential part of the definition of some kinds of operators:
Definition: A linear operator A on an inner product space is hermitian if and only if $A = A^\dagger$.
Definition: A linear operator A on an inner product space is unitary if and only if $A^{-1} = A^\dagger$.
Definition: A linear operator A on an inner product space is normal if and only if $[A, A^\dagger] = 0$.

By these definitions it follows that all hermitian and unitary operators are normal.
Definition: A linear operator A on an inner product space is a projection operator if and only if

$$A = A^\dagger \qquad \text{and} \qquad A^2 = A \tag{657}$$

Definition: A linear operator A on an inner product space is a positive if and only if

$$A = A^\dagger \qquad \text{and} \qquad \langle a|A|a\rangle \geq 0 \tag{658}$$

for all $|a\rangle$ in the inner product space.

The Cauchy Schwartz inequality implies

$$|\langle a|A|a\rangle|^2 \leq \||a\rangle\|^2 \|A|a\rangle\|^2 \leq \||a\rangle\|^4 \||A|\|^2 \tag{659}$$

Taking square roots gives

$$|\langle a|A|a\rangle| \leq \||a\rangle\|^2 \||A|\| \tag{660}$$

One property of a positive operator A is that it satisfies a generalized Cauchy Schwartz inequality:

$$|\langle a|A|b\rangle|^2 \leq \langle a|A|a\rangle\langle b|A|b\rangle \tag{661}$$

The proof of this result is left as a homework exercise.

It has the following useful consequence. If A is positive

$$\|A|a\rangle\|^4 = \langle a|A^2|a\rangle^2 \leq \langle a|A|a\rangle\langle Aa|A|Aa\rangle \tag{662}$$

which is equivalent to

$$\|A|a\rangle\|^2 \leq \langle a|A|a\rangle \frac{\langle Aa|}{\|A|a\rangle\|} A \frac{|Aa\rangle}{\|A|a\rangle\|} \tag{663}$$

The means

$$\sup_{\||a\rangle\|=1} \|A|a\rangle\|^2 \leq \sup_{\||a\rangle\|=1} \langle a|A|a\rangle \frac{\langle Aa|}{\|A|a\rangle\|} A \frac{|Aa\rangle}{\|A|a\rangle\|} \leq \sup_{\||a\rangle\|=1} \langle a|A|a\rangle^2 \tag{664}$$

Comparing (660) and (664) gives

$$\||A|\| = \sup_{\||a\rangle\|=1} \langle a|A|a\rangle \tag{665}$$

when A is a positive operator. The right hand side of this equation is easier to compute than the left hand side.

Definition: Let A be a linear operator. An eigenvector $|v\rangle$ of A is a vector satisfying

$$A|v\rangle = \lambda|v\rangle \tag{666}$$

where λ is a complex constant called the eigenvalue of A associated with the eigenvector $vertv\rangle$.

Example 1: Every vector is an eigenvector of the identity operator I with eigenvalue 1:

$$I|a\rangle = 1|a\rangle \tag{667}$$

Example 2: If P is a projection operator and $P|a\rangle \neq 0$ then $P|a\rangle$ is an eigenvector of P with eigenvalue 1:

$$P^2|a\rangle = 1P|a\rangle \tag{668}$$

Theorem 26.1: The eigenvalues of a Hermitian operator are real

Proof:

$$A = A^\dagger \qquad A\langle v\rangle = \lambda|v\rangle \tag{669}$$

$$\langle v|A|v\rangle = \langle v|v\rangle\lambda = (\langle v|A^\dagger|v\rangle)^* = (\langle v|A|v\rangle)^* = \langle^* v|v\rangle\lambda \tag{670}$$

It follows that

$$(\lambda - \lambda^*)\||v\rangle|^2 = 0 \tag{671}$$

which means that if $|v\rangle$ is not the zero vector that $\lambda = \lambda^*$

Theorem 26.2: The eigenvector of a Hermitian operator corresponding to different eigenvalues are orthogonal.

Proof:

$$\langle v_1|A|v_2\rangle = \lambda_2\langle v_1|v_2\rangle = (\langle v_2|A^\dagger|v_1\rangle)^* =$$
$$(\langle v_2|A|v_1\rangle)^* = \lambda_1^*(\langle v_2|v_1\rangle)^* = \lambda_1\langle v_1|v_2\rangle \tag{672}$$

which gives

$$(\lambda_1 - \lambda_2)\langle v_1|v_2\rangle = 0 \tag{673}$$

Thus if $\lambda_1 - \lambda_2 \neq 0$ then

$$\langle v_1|v_2\rangle = 0 \tag{674}$$

Theorem 26.3: If λ is an eigenvalue of a Unitary operator then $\lambda\lambda^* = 1$

Proof:

$$\langle v|v\rangle = \langle v|U^\dagger U|v\rangle = \lambda\langle v|U^\dagger|v\rangle =$$
$$\lambda(\langle v|U|v\rangle)^* = \lambda\lambda^*(\langle v|v\rangle)^* = \lambda\lambda^*\langle v|v\rangle \tag{675}$$

Which gives

$$(1 - \lambda\lambda^*)\||v\rangle\|^2 = 0 \tag{676}$$

or $\lambda\lambda^* = 1$ if $|v\rangle \neq |0\rangle$.

Theorem 26.4: The eigenvectors of a unitary operator corresponding to different eigenvalues are orthogonal

Proof:

$$\langle v_1|v_2\rangle = \langle v_1|U^\dagger U|v_2\rangle = \lambda_2\langle v_1|U^\dagger|v_2\rangle =$$
$$\lambda_2(\langle v_2|U|v_1\rangle)^* = \lambda_2\lambda_1^*(\langle v_2|v_1\rangle)^* = \lambda_2\lambda_1^*\langle v_2|v_2\rangle \tag{677}$$

Which gives

$$(\lambda_2\lambda_1^* - 1)\langle v_1|v_2\rangle = 0 \tag{678}$$

or $\langle v_1|v_2\rangle = 0$ for $\lambda_1 \neq \lambda_2$.

In your homework you will show that eigenvectors of normal operators with different eigenvalues are also orthogonal/

Let $|a\rangle$ and $|b\rangle$ be elements of an inner product space and let $\langle b|$ be the dual vector to $|b\rangle$. Define the linear operator

$$|a\rangle\langle b| \tag{679}$$

by

$$|a\rangle\langle b|(|c\rangle + \alpha|d\rangle) := |a\rangle(\langle b|c\rangle + \alpha\langle b|d\rangle) \tag{680}$$

Note that

$$(\langle c|a\rangle\langle b|d\rangle)^* = \langle d|(|a\rangle\langle b|)^\dagger|c\rangle = \langle d|b\rangle\langle a|d\rangle \tag{681}$$

which immediately gives

$$(|a\rangle\langle b|)^\dagger = (|b\rangle\langle a|) \tag{682}$$

Consider the special case when $|a\rangle = |b\rangle$ with $\langle a|a\rangle = 1$:

$$P_a := |a\rangle\langle a| = P_a^\dagger \tag{683}$$

It follows immediately that

$$P_a P_a = |a\rangle\langle a|a\rangle\langle a| = |a\rangle 1\langle a| = |a\rangle\langle a| = P_a \tag{684}$$

These are the equation that define a projection operator. This projection operator maps every vector to a complex multiple of the vector $|a\rangle$.

This is not the most general projection operator. To see this let $|a\rangle$ and $|b\rangle$ be orthogonal unit vector

$$\langle a|b\rangle = 0 \tag{685}$$

Let

$$P = P_a + P_b \tag{686}$$

It follows from (685) that

$$P_a P_b = |a\rangle\langle a|b\rangle\langle b| = |a\rangle 0\langle b| = 0 \tag{687}$$

Taking adjoints gives

$$0 = (0)^\dagger = (P_a P_b)^\dagger = P_b^\dagger P_a^\dagger = P_b P_a \tag{688}$$

It follows that

$$P^\dagger = (P_a + P_b)^\dagger = (P_a)^\dagger + (P_b)^\dagger = P_a + P_b = P \tag{689}$$

$$P^2 = P_a P_a + P_a P_b + P_b P_a + P_b P_b = P_a + 0 + 0 + P_b = P \tag{690}$$

Thus P is a projection operator.

The identity operator is another operator that is trivially a projection operator.

If P is a projection operator and P has an inverse then P is the identity. To prove this note

$$I = PP^{-1} = PPP^{-1} = PI = P \tag{691}$$

If P_1 and P_2 are projection operators then $P = P_1 + P_2$ is a projection operator if and only if $P_1 P_2 = 0$

If $P_1 P_2 = 0$ then $P_2 P_1 = (P_1 P_2)^\dagger = 0$ and

$$(P_1+P_2)^2 = P_1P_1+P_1P_2+P_2P_1+P_2P_2 == P_1+0+0+P_2 = (P_1+P_2) \tag{692}$$

Conversely if

$$(P_1 + P_2)^2 = (P_1 + P_2) \tag{693}$$

then

$$P_1P_2 + P_2P_1 = 0 \tag{694}$$

Multiplying by P_1 on the left gives

$$0 = P_1(P_1P_2 + P_2P_1) = P_1P_2 + P_1P_2P_1 \tag{695}$$

while multiplying by P_1 on the right gives

$$0 = (P_1P_2 + P_2P_1)P_1 = P_1P_2P_1 + P_2P_1 \tag{696}$$

Comparing (696) and (696) gives

$$P_2P_1 = P_2P_1 \tag{697}$$

which when combined with (694) gives

$$P_1P_2 = P_2P_1 = 0 \tag{698}$$

Projection operators $\{P_i\}$ satisfying

$$P_iP_j = \delta_{ij}P_j \tag{699}$$

are called orthogonal projectors. The Kronecker delta function, δ_{ij}, is

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \tag{700}$$

29:171 - Homework Assignment #9

1. The commutator and anti-commutator of two linear operators are defined by
$$[A, B] := AB - BA \qquad \{A, B\} := AB - BA$$
Prove the following identities
$$[A[B, C]] + [B[C, A]] + [C[A, B]] = 0$$
$$[A, BC] = [A, B]C + B[A, C]$$
$$[A, BC] = \{A, B\}C - B\{A, C\}$$

2. Let K be a linear Hermitian operator. Define
$$W := (I + iK)(I - iK)^{-1}$$
Show that W is a unitary operator.

   Express K in terms of W.(K is called the Cayley transform of W)

3. Let P be an orthogonal projection operator. Let $Q := I - P$.

   Show that Q is an orthogonal projection operator.

   Evaluate QP.

4. A linear operator N is Nilpotent if for some finite n, $N^n = 0$. Show that e^N is a finite degree polynomial in N if N is nilpotent. Show that $e^{\alpha N} e^{\beta N} = e^{(\alpha+\beta)N}$ still holds when N is nilpotent.

5. Let A be a bounded linear operator on a normed linear space. Define the partial sums
$$F_n(A) = I + \sum_{m=1}^{n} \frac{1}{m!} A^m$$
Show that this is a Cauchy sequence of operators.

6. Show that if $[A, B] = 0$ that

$$exp(A + B) = exp(A)exp(B) = exp(B)exp(A)$$

   What happens to these relations if $[A, B] = \alpha I$ where $\alpha \in \mathbb{C}$ and I is the identity operator?

7. Let P be a positive operator. Prove the generalized Cauchy Schwartz inequality:

$$|\langle a|P|b\rangle|^2 \leq \langle a|P|a\rangle\langle b|P|b\rangle$$

0.28 Lecture 28

A nice application of constructing new operators in terms of polynomials given by the square root theorem.
Theorem: Every bounded positive operator has a unique positive square root.
Proof: Without loss of generality I assume that $\||A\|| < 1$, Otherwise I can define $A' = \frac{1}{2\||A\||}A$ and

$$\sqrt{A} = \sqrt{2\||A\||}\sqrt{A'} \tag{701}$$

where $\||A'\|| = \frac{1}{2}$

I construct $\sqrt{A}$ as a limit of polynomials in with real positive coefficients A. Define

$$C := 1 - A \tag{702}$$

and replace the unknown $\sqrt{A}$ by X:

$$X := 1 - \sqrt{A} \tag{703}$$

Then

$$X^2 = I - 2\sqrt{A} + A = I - 2(1 - X) + (1 - C) \tag{704}$$

or

$$X = \frac{1}{2}(C + X^2) \tag{705}$$

Next I note that C is a positive operator. Using (665) with $\||a\rangle\| = 1$

$$\langle a|C|a\rangle = 1 - \langle a|A|a\rangle \geq 1 - \||A\|| > 0 \tag{706}$$

On the other hand since $\langle a|A|a\rangle < 1$

$$\langle a|C|a\rangle \leq 1 \tag{707}$$

Using (660) again gives

$$\||C\|| = \sup_{\||a\rangle\|=1} \langle a|C|a\rangle \leq 1 \tag{708}$$

Given this bound consider the iteration of (705)

$$X_0 = \frac{1}{2}C \tag{709}$$

$$X_1 = \frac{1}{2}(C + X_0^2) = \frac{1}{2}(C + \frac{1}{4}C^2) \tag{710}$$

$$\vdots \tag{711}$$

$$X_{n+1} = \frac{1}{2}(C + X_n^2) \tag{712}$$

Note that X_0 is a polynomial in C with positive coefficients.

By induction assume that X_n is a polynomial in C with constant coefficients, then by (712) X_{n+1} is a polynomial in C with positive coefficients. This implies that X_n is positive.

Assume that $|||X_n||| \leq 1$. Then

$$|||X_{n+1}||| = \frac{1}{2}|||C + X_n^2||| \leq \frac{1}{2}|||C||| + \frac{1}{2}|||X_n||||||X_n||| \leq \frac{1}{2} + \frac{1}{2} = 1 \tag{713}$$

This shows that the X_n are positive and bounded in norm by 1.

Next I show that the sequence of operators X_n is a Cauchy sequence

$$X_n - X_{n-1} = \frac{1}{2}(C + X_{n-1}^2) - \frac{1}{2}(C + X_{n-2}^2) = \frac{1}{2}(X_{n-1}^2 - X_{n-2}^2) \tag{714}$$

Since all of the X_n are polynomials in C they commute so

$$X_n - X_{n-1} = \frac{1}{2}(X_{n-1} + X_{n-2})(X_{n-1} - X_{n-2}) \tag{715}$$

Repeating this $n-1$ times gives

$$X_n - X_{n-1} = \frac{1}{2}(X_{n-1} + X_{n-2})\frac{1}{2}(X_{n-2} + X_{n-3}) \cdots \frac{1}{2}(X_1 + X_0)\frac{1}{2}X_0 = \tag{716}$$

$$X_n - X_{n-1} = \frac{1}{2^n}X_0 \prod_{k=2}^{n}(X_{k-1} + X_{k-2}) \tag{717}$$

This immediately implies

$$|||X_n - X_{n-1}||| \leq \frac{1}{2^n}|||X_0||| \prod_{k=2}^{n}(|||X_{k-1}||| + |||X_{k-2}|||) < 1 \tag{718}$$

On the other hand (717) also implies that

$$X_n - X_{n-1} \tag{719}$$

is a polynomial in C with positive coefficients.

Consider the sequence of vectors

$$X_n|a\rangle \tag{720}$$

Note that

$$\|(X_n - X_m)|a\rangle\|^4 = \langle a|(X_n - X_m)(X_n - X_m)|a\rangle^2 \tag{721}$$

Since $X_n - X_m$ is positive the generalized Cauchy Schwartz inequality (homework) gives

$$\langle a|(X_n - X_m)(X_n - X_m)|a\rangle^2 \leq \langle a|(X_n - X_m)|a\rangle\langle a|(X_n - X_m)^3|a\rangle^2 \leq$$

$$\langle a|(X_n - X_m)|a\rangle 1 \tag{722}$$

Note $\{\langle a|X_n|a\rangle\}$ is an increasing sequence of positive numbers, bounded by 1. This sequence must be Cauchy, otherwise we can find infinite subsequences with $\langle a|(X_m - X_n)|a\rangle > 1/K$. Using $K+1$ elements leads to a violation of the bound.

The left side of the inequality implies

$$\|(X_m - X_n)|a\rangle\| \to 0 \tag{723}$$

which means that the sequence of vectors $X_n|a\rangle$ is Cauchy. If the inner product space is complete this converges to a vector in the space. Define

$$X|a\rangle = \lim_{n\to\infty} X_n|a\rangle = |a\rangle - \sqrt{A}|a\rangle \tag{724}$$

Solving for $\sqrt{A}$ gives

$$\sqrt{A}|a\rangle = |a\rangle - \lim_{n\to\infty} X_n|a\rangle \tag{725}$$

Since $\||X_n|\| \leq 1$ for all n, and X_n are all Hermitian, it follows that $I - X_n$ is positive.

Note that while there may be many operators that satisfy $B^2 = A$, there is only one positive square root.

Since $\sqrt{A}$ is a limit of polynomials in A it follows that

$$[\sqrt{A}, A] = 0 \tag{726}$$

If A and B are both positive operators and $[A, B] = 0$ then AB is positive. To prove this note

$$(AB)^\dagger = B^\dagger A^\dagger = BA = AB \tag{727}$$

and

$$AB = \sqrt{A}\sqrt{A}B = \sqrt{A}B\sqrt{A} \tag{728}$$

Let

$$|c\rangle = \sqrt{A}|a\rangle \tag{729}$$

so

$$\langle a|AB|a\rangle = \langle c|B|c\rangle \geq 0 \tag{730}$$

Assume that A is positive and A^{-1} exists, then

$$\sqrt{A}^{-1} = A^{-1}\sqrt{A} \tag{731}$$

Assume that A^{-1} exists. Then

$$A = (AA^\dagger)^{1/2}(AA^\dagger)^{-1/2}A \tag{732}$$

Note that

$$P := (AA^\dagger)^{1/2} \geq 0 \tag{733}$$

is positive (square root of a positive operator) and

$$(AA^\dagger)^{-1/2}A[(AA^\dagger)^{-1/2}A]^\dagger = (AA^\dagger)^{-1/2}AA^\dagger(AA^\dagger)^{-1/2} = I \tag{734}$$

This means that $U := (AA^\dagger)^{-1/2}A$ is unitary. It follows that every invertible operator can be written as the product of a positive and a unitary operator:

$$A = PU \tag{735}$$

Note that we also have

$$A = A(AA^\dagger)^{-1/2}(AA^\dagger)^{1/2} \tag{736}$$

with a (different) unitary operator on the left and a positive operator on the right. This result is called the polar decomposition theorem.

0.29 Lecture 29

Linear combinations of vectors can be used to generate new vectors:

$$|b\rangle = \sum_{i=1}^{N} \alpha_i |a_i\rangle \tag{737}$$

Definition: A set of vectors $\{|a_i\rangle\}_{i\in I}$ are linearly independent if the only solution to

$$\sum_{i\in I} \alpha_i |a_i\rangle = |0\rangle \tag{738}$$

is $\alpha_i = 0$ for all $i \in I$ where I is an index set.
Definition: A set of vectors $\{|a_i\rangle\}_{i\in I}$ span a vector space $\mathcal{V}$ if any vector $|b\rangle \in \mathcal{V}$ can be expressed as

$$|b\rangle = \sum_{i\in I} \alpha_i |a_i\rangle \tag{739}$$

Theorem 29.1: Let $n < \infty$ vectors span a vector space containing r linearly independent vectors. Then $n \geq r$.

To prove this let $\{|b_i\rangle\}_{1\leq i\leq r}$ be the linear independent vectors and let $\{|a_i\rangle\}_{1\leq i\leq n}$ be the vectors that span the vector space. I assume that $|b_i\rangle \neq |0\rangle$.

First note

$$|b_1\rangle = \sum_{i=1}^{n} \alpha_{1j} |a_j\rangle \tag{740}$$

At least one of the $\alpha_{1j} \neq 0$. By relabeling the index set on the vectors $|a_i\rangle$ we can assume without loss of generality that $\alpha_{11} \neq 0$. It follows that

$$|a_1\rangle = \frac{1}{\alpha_{11}} \left(|b_1\rangle - (\sum_{i=2}^{n} \alpha_{1j} |a_j\rangle) \right) \tag{741}$$

It follows that $\{|b_1\rangle, |a_2\rangle, \cdots |a_n\rangle\}$ span the same vector space.

Next express $|b_2\rangle$ in terms of this new spanning set.

$$|b_2\rangle = \beta_{21}|b_1\rangle + \sum_{i=2}^{n} \alpha_{2j} |a_j\rangle \tag{742}$$

Since $|b_1\rangle$ and $|b_2\rangle$ are independent, at least one of the $\alpha_{2j} \neq 0$. Again I can relabel indices so $\alpha_{22} \neq 0$. This allows me to replace $|a_2\rangle$ by $|b_2\rangle$ in the spanning set, giving a new spanning set

$$\{|b_1\rangle, |b_2\rangle, |a_3\rangle \cdots |a_n\rangle\} \tag{743}$$

This process can be continued until all n of the $|a_i\rangle$ are replaced by the first n $|b_i\rangle$s. In this case the first n $|b_i\rangle$ span the space. If $r > n$ then $|b_{n+1}\rangle \cdots |b_r\rangle$ can all be expressed in terms of the $|b_i\rangle$ for $1 \leq i \leq n$. This means that the vectors $|b_{n+1}\rangle \cdots |b_r\rangle$ are not linearly independent of $|b_1\rangle \cdots |b_n\rangle$. This completes the proof of the theorem.

Definition: The dimension of a vector space is the maximal number of linearly independent vectors.

Definition: A basis of a vector space is a linearly independent set of vectors that span the space.

Theorem 29.2: The number of basis vectors in an n dimensional space is n.

Proof: Let $\{|b_l\rangle\}_{l=1}^{m}$ be a basis for a n-dimensional vector space. Since the $|b_l\rangle$ are linearly independent, by definition of dimension, $m \leq n$. On the other hand since $|b_l\rangle$ span the vector space, Theorem 29.1 implies $n \geq m$. To satisfy both inequalities requires $m = n$.

Corollary: Any two bases of the same vector space have the same number of basis vectors.

Let $\{|n\rangle\}_{n=1}^{N}$ be a basis for a vector space $\mathcal{V}$. Then any vector can be written as

$$|a\rangle = \sum_{n=1}^{N} |n\rangle a_n \tag{744}$$

This decomposition is unique because if

$$|a\rangle = \sum_{n=1}^{N} |n\rangle a'_n \tag{745}$$

then

$$|0\rangle = |a\rangle - |a\rangle = \sum_{n=1}^{N} |n\rangle (a_n - a'_n) \tag{746}$$

Since the basis vectors are linearly independent

$$a_n - a'_n = 0 \tag{747}$$

for every n.

This shows that there is a 1-1 correspondence between vectors in a N-dimensional vector space and ordered sets of N complex numbers. This means the study of abstract vector spaces can always be reduced to the study of ordered sets of complex numbers.

Definition: The numbers a_n are coordinates of the vector $|a\rangle$ in the basis $\{|n\rangle\}_{n=1}^N$.

Note that a vector $|a\rangle$ will have different coordinates in different bases:

$$|a\rangle = \sum_{n=1}^N |n\rangle a_n = \sum_{n=1}^N |\bar{n}\rangle b_n \tag{748}$$

Vector operations become operations on components. The components of the sum of two vectors is represented by the sum of the components of the individual vectors: The components of the scalar multiple of a vector by a complex constant α is the product of α with the components of the vector:

$$\sum_{n=1}^N |n\rangle a_n + \sum_{n=1}^N |n\rangle b_n = \sum_{n=1}^N |n\rangle (a_n + b_n) \tag{749}$$

$$\alpha(\sum_{n=1}^N |n\rangle a_n) = \sum_{n=1}^N |n\rangle (\alpha a_n) \tag{750}$$

Ii is sometimes useful to write the components of vectors in a given basis as

$$|a\rangle = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_{n-1} \\ a_n \end{pmatrix} \tag{751}$$

In this notation

$$|a\rangle + \beta|b\rangle \tag{752}$$

can be written as

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_{n-1} \\ a_n \end{pmatrix} + \beta \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_{n-1} \\ b_n \end{pmatrix} = \begin{pmatrix} a_1 + \beta b_1 \\ a_2 + \beta b_2 \\ \vdots \\ a_{n-1} + \beta b_{n-1} \\ a_n + \beta b_n \end{pmatrix} \tag{753}$$

Implicit in these expressions is a set of basis vectors. Using the same coordinates with different basis vectors results in a different vector.

Let A be a linear operator and let $\{|n\rangle\}_{n=1}^{N}$ be a basis. Each of the vectors $A|n\rangle$ can be expressed as a linear combination of the basis vectors:

$$A|n\rangle = \sum_{m=1}^{M} |m\rangle A_{mn} \tag{754}$$

It follows that

$$A \sum_{n=1}^{N} |n\rangle a_n = \sum_{n=1}^{N} A|n\rangle a_n = \sum_{n=1}^{N} \sum_{m=1}^{N} |m\rangle A_{mn} a_n \tag{755}$$

The means that if a_n are the components of $|a\rangle$ in the basis $\{|n\rangle\}_{n=1}^{N}$ then the components of $A|a\rangle$ in this basis are

$$b_m = \sum_{n=1}^{N} A_{mn} a_n \tag{756}$$

Definition: An $N \times M$ array of numbers that represent a linear operator A in a basis is called a matrix. Normally the first index is called the "row" index and thesecond index is called the "column" index. The matrix A_{mn} is soemtimes written as

$$A_{mn} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n-1} & q_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n-1} & q_{2n} \\ \vdots & \vdots & \ldots & \vdots & \vdots \\ a_{n-11} & a_{n-12} & \cdots & a_{n-1n-1} & a_{n-1n} \\ a_{n1} & a_{n2} & \cdots & a_{nn-1} & a_{nn} \end{pmatrix} \tag{757}$$

Definition: The operation

$$a'_n = \sum_{m=1}^{M} A_{nm} a_m \tag{758}$$

is called multiplication of a vector by a matrix.

The product of linear operators $C = AB$ can be represented by matrices as

$$C_{mn} = \sum_{k=1}^{M} A_{mk} B_{kn} \tag{759}$$

where M is the dimension of the first index of B_{kn} and the dimension of the last index of A_{mk} This operation is called multiplication of a matrix by a matrix.

To compute the elements of a matrix A_{mn} note that

$$\langle m|A|n\rangle = \sum_{k,l=1}^{N} \langle m|k\rangle A_{kn} \tag{760}$$

In order to extract the numbers A_{kl} is is necessary to invert $\langle m|k\rangle$. While this can always be done, it can be avoided by using orthonormal bases.

If

$$|a\rangle = \sum_{n=1}^{N} |n\rangle a_n \tag{761}$$

then the dual vector has the form

$$\langle a| = \sum_{n=1}^{N} a_n^* \langle n| \tag{762}$$

where $\langle n|$ is the dual of $|n\rangle$.

The action of the linear functional $\langle b|$ on the vector $|a\rangle$ is

$$\langle b|a\rangle = \sum_{m=1}^{N} \sum_{n=1}^{N} b_m^* \langle m|n\rangle a_n \tag{763}$$

Here the same matrix the appears in (760) appears in (763).

The relationship between the matrix elements of A and $A^\dagger$ can be computed from the definition:

$$\langle b|A|a\rangle = \langle a|A^\dagger|b\rangle^* \tag{764}$$

The left side of this expression is

$$\sum_{m=1}^{N} \sum_{n=1}^{N} \sum_{k=1}^{N} b_m^* \langle m|k\rangle A_{kn} a_n = \tag{765}$$

while the right side is

$$\sum_{m=1}^{N} \sum_{n=1}^{N} \sum_{k=1}^{N} a_m \langle k|m\rangle A_{kn}^{\dagger *} b_n^* \tag{766}$$

Comparing coefficients of the same vectors

$$\sum_{k=1}^{N} \langle k|n\rangle A^*_{kn} = \sum_{k=1}^{N} \langle n|k\rangle A^\dagger_{km} \tag{767}$$

The matrix the appears in (760) appears again here.
Definition: The transpose of the matrix A_{mn} is the matrix $A^T_{mn} = A_{nm}$.

This definition allows us to write

$$\sum_{k=1}^{N} A^{T*}_{mk} \langle k|n\rangle = \sum_{k=1}^{N} \langle m|k\rangle A^\dagger_{kn} \tag{768}$$

The matrices $M_{mk} := \langle m|k\rangle$ are characteristic of inner product spaces. While they are always invertible, it the basis vectors are almost parallel, the inverse matrix can get very large and lead to computational instabilities.

The matrix $\langle m|n\rangle$ becomes the identity when the basis is orthonormal.
Definition: A basis $\{|n\rangle\}_{n=1}^N$ is orthonormal if

$$\langle m|n\rangle = \delta_{mn} \tag{769}$$

In an orthonormal basis (768) becomes

$$A^{T*}_{mn} = A^\dagger_{mn} \tag{770}$$

and

$$A_{mn} = \langle m|A|n\rangle \tag{771}$$

If a basis $\{|n\rangle\}_{n=1}^N$ is not orthonormal it is possible to use it to construct a new orthonormal basis. This method is important in quantum mechanics and is called the Gram Schmidt method. The new basis is denoted by $\{|\bar{n}\rangle\}_{n=1}^N$

The construction is

$$|\bar{1}\rangle := c_1|1\rangle \tag{772}$$

where

$$c_1 = \frac{1}{\||1\rangle\|} \tag{773}$$

$$|\bar{2}\rangle = c_2[|2\rangle - |\bar{1}\rangle\langle\bar{1}|2\rangle] \tag{774}$$

$$c_2 = \frac{1}{\|[|2\rangle - |\bar{1}\rangle\langle\bar{1}|2\rangle]\|} \tag{775}$$

$$|\bar{3}\rangle = c_3[|3\rangle - |\bar{1}\rangle\langle\bar{1}|3\rangle - |\bar{2}\rangle\langle\bar{2}|3\rangle] \tag{776}$$

$$c_3 = \frac{1}{\| [|3\rangle - |\bar{1}\rangle\langle\bar{1}|3\rangle - |\bar{2}\rangle\langle\bar{2}|3\rangle] \|} \tag{777}$$

$$\vdots \tag{778}$$

$$|\bar{n}\rangle = c_n[|n\rangle - \sum_{k=1}^{n-1} |\bar{k}\rangle]] \tag{779}$$

$$c_n = \frac{1}{\| [|n\rangle - \sum_{k=1}^{n-1} |\bar{k}\rangle] \|} \tag{780}$$

This completes the construction of the new orthonormal basis.

0.30 Lecture 30

In an orthonormal basis a matrix is Hermitian if

$$A_{mn} = A^{T*}_{mn} = A^{\dagger}_{mn} \tag{781}$$

In an orthonormal basis a matrix is unitary if

$$A^{-1}_{mn} = A^{T*}_{mn} = A^{\dagger}_{mn} \tag{782}$$

Next I discuss inverses of linear operators. I what follows I assume that the dimension of the vector space, N, is finite.

Define

$$\epsilon^{1,2,\cdots,N-1,N}_{n_1,n_2,\cdots n_{N-1},n_N} \tag{783}$$

by the conditions

$$\epsilon^{1,2,3\cdots,n}_{1,2,3,\cdots,n} = 1 \tag{784}$$

and $\epsilon^{1,2,\cdots,N-1,N}_{n_1,n_2,\cdots n_{N-1},n_N}$ is antisymmetric on interchanging any n_i with n_j of $j \neq i$.

Given this definition the Determinant of an $N \times N$ matrix A_{mn} is defined by

$$\det(A) = \sum_{n_1,\cdots,n_N} \epsilon^{1,2,\cdots,N-1,N}_{n_1,n_2,\cdots n_{N-1},n_N} A_{1,n_1} A_{2,n_2} \cdots A_{N,n_N} \tag{785}$$

Theorem 30.1 If $\det(A) \neq 0$ then $A_{1,n}, \cdots, A_{N,n}$ are components of N linearly independent vectors, $|v_k\rangle$

$$|v_k\rangle = \sum_{n=1}^{N} A_{k,n}|n\rangle \tag{786}$$

To prove this assume by contradiction that $|v_k\rangle$ are linearly independent. Then for some m

$$|v_m\rangle = \sum_{l\neq m} c_l |v_l\rangle \tag{787}$$

It follows that

$$\det(A) = \sum \epsilon^{1,2,\cdots,N-1,N}_{n_1,n_2,\cdots n_{N-1},n_N} A_{1,n_1}A_{2,n_2}\cdots A_{N,n_N} =$$

$$\sum \epsilon^{1,2,\cdots,N-1,N}_{n_1,n_2,\cdots n_{N-1},n_N} A_{1,n_1}A_{2,n_2}\cdots (\sum_l c_L A_{l,n_m})\cdots A_{N,n_N} =$$

$$\sum_{l\neq m}\sum c_l \epsilon^{1,2,\cdots,N-1,N}_{n_1,n_2,\cdots n_{N-1},n_N} A_{1,n_1}A_{2,n_2}\cdots A_{l,n_m}\cdots A_{N,n_N} \tag{788}$$

Each l in the sum is repeated has terms of the form

$$\epsilon^{\cdots}_{\cdots n_m \cdots n_l \cdots} A_{l,n_m}\cdots A_{l,n_l} =$$

$$-\epsilon^{\cdots}_{\cdots n_l \cdots n_m \cdots} A_{l,n_m}\cdots A_{l,n_l} =$$

$$-\epsilon^{\cdots}_{\cdots n_m \cdots n_l \cdots} A_{l,n_l}\cdots A_{l,n_m} \tag{789}$$

where I have used the antisymmetry of $\epsilon^{1,2,\cdots,N-1,N}_{n_1,n_2,\cdots n_{N-1},n_N}$ and relabeled dummy indices. This shows that this term, and thus each term in the sum, is equal to zero.

This contradicts the assumption that $\det(A)$ is non-zero.

A similar proof shows that of $\det(A) \neq 0$ that the "columns" of the matrix A_{mn} must also be coordinates of linearly independent vectors.

There is an alternate way to write the determinant of a matrix. A permutation σ on the integers $\{1, \cdots, N\}$ is an invertible function on $\{1, \cdots, N\}$. There are $N!$ distinct permutations of N objects. They can be generated by taking products of pairwise interchanges. We define $|\sigma| = 0$ if σ can be obtained from the identity by an even number of pairwise interchanges, and $|\sigma| = 1$ if σ can be obtained from the identity by an odd number of pairwise interchanges. $\mathcal{P}(N)$ denotes the set of all $N!$ permutations on the integers $\{1, \cdots, N\}$.

With this definition the determinant can be written as

$$\det(A) = \sum_{\sigma \in \mathcal{P}(N)} (-)^{|\sigma|} A_{1,\sigma(1)} \cdots A_{N,\sigma(N)} \tag{790}$$

This follows because the sum over $n_1 \cdots n_N$ is non-zero only when all of the n_i are different, which corresponds to a permutation. In that case $\epsilon_{1\cdots N}^{\sigma(1)\cdots\sigma(N)} = (-)^{|\sigma|}$

Assume that $D = \det(A) \neq 0$. If I treat the components of the matrix A_{mn} as a set of N^2 independent variables then for any l:

$$D = \sum_n A_{ln} \frac{\partial D}{\partial A_{ln}} \tag{791}$$

This is an immediate consequence of the fact for fixed l each contribution to D has one term of the form A_{ln}.

On the other hand if $l \neq l'$ then

$$D = \sum_n A_{l'n} \frac{\partial D}{\partial A_{ln}} = 0 \tag{792}$$

since it is equivalent to replacing the $l-th$ row of the matrix with the l' row of the matrix. In this case the resulting expression is the determinant of a matrix with two identical rows, which must vanish.

Combining these two results gives

$$D\delta_{ll'} = \sum_n A_{ln} \frac{\partial D}{\partial A_{l'n}} \tag{793}$$

or if $D \neq 0$

$$\delta_{ll'} = \sum_n A_{ln} \frac{1}{D} \frac{\partial D}{\partial A_{l'n}} = \tag{794}$$

$$\delta_{ll'} = \sum_n A_{ln} \frac{\partial \ln(D)}{\partial A_{l'n}} \tag{795}$$

This shows that whenever $\det(A) \neq 0$ the matrix A_{mn} has an inverse given by

$$A^{-1}_{mn} = \frac{\partial \ln(D)}{\partial A_{nm}} \tag{796}$$

The matrix

$$C_{nl} := \frac{\partial D}{\partial A_{ln}} \tag{797}$$

is called the cofactor matrix to A_{nl}. I have just shown that

$$A^{-1}_{mn} = \frac{1}{\det(A)} C_{mn} \tag{798}$$

29:171 - Homework Assignment #11

1. Prove that the eigenvectors of a normal operator corresponding to distinct eigenvalues are orthogonal.

2. Consider the vector space of polynomials with inner product

$$\langle P_1|P_2\rangle = \int_0^\infty e^{-x} P_1^*(x) P_2(x) dx$$

   The polynomials $1, x, x^2, x^3, x^4$ are linearly independent vectors in this space. Use the Gram-Schmidt method to make a set of four orthonormal basis functions with this scalar product. Compare your results to expressions for the first four Laguerre polynomials.

3. Show that

$$A^{-1}_{ij} := \frac{\partial \ln(\det(A))}{\partial A_{ji}}$$

   is both a left and right inverse of the matrix A_{ij}.

4. Find the characteristic Polynomial $\Phi(\lambda)$ of

$$A = \begin{pmatrix} a & b \\ b & a \end{pmatrix}$$

a. Find the roots of the characteristic polynomial

b. Show $\phi(A) = 0$

c. Find $\phi_n(A)$ for each eigenvalue λ_n of A.

d. Show $\phi_1(A) + \phi_2(A) = I$

5. Let A be a Hermitian operator in a d-dimensional vector space with d distinct eigenvalues. Show that

$$P_m = \prod_{n \neq m, 1}^{d} \frac{A - \lambda_n}{\lambda_m - \lambda_n}$$

a. is an orthogonal projection operator

b. $P_m P_n = \delta_{mn} P_m$.

6. Let J be

$$J := \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

a. Find the eigenvalues of J.

b. Find the eigenvectors of J.

c. Show that J^3 can be expressed as a linear combination of J^2 ,J and I.

d. Calculate

$$R := \exp(iJ\theta)$$

where θ is a parameter.

0.31 Lecture 31:

Example 1:

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \tag{799}$$

$$\det(A) = a_{11}a_{22} - a_{12}a_{21} \tag{800}$$

$$\ln(\det(A)) = \ln(a_{11}a_{22} - a_{12}a_{21}) \tag{801}$$

$$a_{11}^{-1} = \frac{\partial \ln(\det(A))}{\partial a_{11}} = \frac{a_{22}}{a_{11}a_{22} - a_{12}a_{21}} \tag{802}$$

$$a_{12}^{-1} = \frac{\partial \ln(\det(A))}{\partial a_{21}} = \frac{-a_{12}}{a_{11}a_{22} - a_{12}a_{21}} \tag{803}$$

$$a_{21}^{-1} = \frac{\partial \ln(\det(A))}{\partial a_{12}} = \frac{-a_{21}}{a_{11}a_{22} - a_{12}a_{21}} \tag{804}$$

$$a_{22}^{-1} = \frac{\partial \ln(\det(A))}{\partial a_{22}} = \frac{a_{11}}{a_{11}a_{22} - a_{12}a_{21}} \tag{805}$$

Multiplication shows that this is the inverse of the matrix A.

Since the determinant of a matrix plays such an important role it is useful to establish some other properties of determinants.

The most useful of these is

$$\det(AB) = \det(A)\det(B) \tag{806}$$

To prove this I first note that

$$\det(A) = \sum_{n_1 \cdots n_N} \epsilon^{1,2,\cdots,N-1,N}_{n_1,n_2,\cdots n_{N-1},n_N} A_{1,n_1} \cdots A_{N,n_N} = \tag{807}$$

$$\sum_{\sigma \in P(N)} (-)^{|\sigma|} A_{1,\sigma(1)} \cdots A_{N,\sigma(N)} \tag{808}$$

where the sum is over all permutations σ on N objects where $|\sigma|$ is 1 if σ can be constructed out of an odd number of pairwise interchanges of $(1, 2, \cdots, n)$ and 0 if σ can be constructed out of an even number of pairwise interchanges of $(1, 2, \cdots, n)$.

It is easy to see that this definition is completely equivalent to the original definition. With this definition

$$\det(AB) = \sum_{n_1 \cdots n_N} \sum_{\sigma \in P(N)} (-)^{|\sigma|} A_{1,n_1} B_{n_1,\sigma(1)} \cdots A_{N,n_N} B_{N_n,\sigma(N)} = \tag{809}$$

$$\sum_{n_1\cdots n_N} A_{1,n_1}\cdots A_{N,n_N} \sum_{\sigma\in P(N)} (-)^{|\sigma|} B_{n_1,\sigma(1)}\cdots B_{N_n,\sigma(N)} \tag{810}$$

The second sum vanishes unless all of the n_i are different. This means that I can replace the n_i sums by a sum over permutations

$$\sum_{\sigma'\in P(N)} A_{1,\sigma'(1)}\cdots A_{N,\sigma'(N)} \sum_{\sigma\in P(N)} (-)^{|\sigma|} B_{\sigma'(1),\sigma(1)}\cdots B_{\sigma'(N),\sigma(N)} \tag{811}$$

which can be rewritten as

$$\sum_{\sigma'\in P(N)} (-)^{|\sigma'|} A_{1,\sigma'(1)}\cdots A_{N,\sigma'(N)} \sum_{\sigma\in P(N)} (-)^{|\sigma|+(-)^{|\sigma'|}} B_{\sigma'(1),\sigma(1)}\cdots B_{\sigma'(N),\sigma(N)} \tag{812}$$

The second term becomes

$$B_{\sigma'(1),\sigma(1)}\cdots B_{\sigma'(N),\sigma(N)} = B_{1,\sigma^{-1\prime}\sigma(1)}\cdots B_{N,\sigma^{-1\prime}\sigma(N)}$$

I can redefine the index

$$\sigma'' = \sigma'^{-1}\sigma, \tag{813}$$

where the sum still runs over all permutations, to get

$$|\sigma''| = |\sigma^{-1\prime}| + |\sigma| = -|\sigma^{-1\prime}| + |\sigma| = -|\sigma'| + |\sigma| \tag{814}$$

It follows that

$$\det(AB) =$$
$$\sum_{\sigma'\in P(N)} (-)^{|\sigma'|} A_{1,\sigma'(1)}\cdots A_{N,\sigma'(N)} \sum_{\sigma''\in P(N)} (-)^{|\sigma''|} B_{1,\sigma''(1)}\cdots B_{N,\sigma''(N)} =$$
$$\det(A)\det(B) \tag{815}$$

In a similar fashion it is possible to show

$$\det(A) = \det(A^T) \tag{816}$$

This will be left as a homework exercise.

Note also that

$$\det(I) = 1 \qquad 1 = \det(I) = \det(AA^{-1}) = \det(A)\det(A^{-1}) \tag{817}$$

which implies

$$\det(A^{-1}) = \frac{1}{\det(A)} \tag{818}$$

If $\{|n\rangle\}_{n=1}^N$ and $\{|\bar{n}\rangle\}_{n=1}^N$ are both orthonormal bases then any vector can be written as

$$|a\rangle = \sum_{n=1}^N |n\rangle a_n = \sum_{n=1}^N |\bar{n}\rangle \bar{a}_n \qquad (819)$$

The orthonormality of the basis vectors gives

$$a_n = \langle n|a\rangle \qquad \bar{a}_n = \langle \bar{n}|a\rangle \qquad (820)$$

which can be used to express **??** as

$$|a\rangle = \sum_{n=1}^N |n\rangle\langle n|a\rangle = \sum_{n=1}^N |\bar{n}\rangle\langle\bar{n}|a\rangle \qquad (821)$$

Comparing the left and right hand side of these equations, if I remove $|a\rangle$ from both sides I get two expressions for the identity operator

$$I = \sum_{n=1}^N |n\rangle\langle n| = \sum_{n=1}^N |\bar{n}\rangle\langle\bar{n}| \qquad (822)$$

Now consider

$$|a\rangle = I^2|a\rangle = \sum_{m=1}^N |\bar{m}\rangle\langle\bar{m}| \sum_{n=1}^N |n\rangle\langle n|a\rangle =$$

$$\sum_{m=1}^N |\bar{m}\rangle\langle\bar{m}| \sum_{n=1}^N |n\rangle a_n = \sum_{m=1}^N |\bar{m}\rangle \bar{a}_m \qquad (823)$$

Comparing these expressions

$$\bar{a}_m = \sum_{n=1}^N \langle\bar{m}|n\rangle a_n \qquad (824)$$

The quantities

$$\langle\bar{m}|n\rangle \qquad (825)$$

are matrix elements of the operator U

$$\langle\bar{m}|n\rangle = \langle\bar{m}|U|\bar{n}\rangle = \langle m|U|n\rangle \qquad (826)$$

where

$$U := \sum_n |n\rangle\langle\bar{n}| \tag{827}$$

Note that

$$UU^\dagger := \sum_{nm} |n\rangle\langle\bar{n}||\bar{m}\rangle\langle m| = \sum_m |m\rangle\langle m| = I \tag{828}$$

which show that U is unitary.

This shows that any operator that changes orthonormal bases is unitary.

It is also easy to show that any unitary operator can be written in the form

$$U := \sum_n |n\rangle\langle\bar{n}| \tag{829}$$

where $\{|n\rangle\}_{n=1}^N$ and $\{|\bar{n}\rangle\}_{n=1}^N$ are both orthonormal bases

To do this start with the basis $\{|\bar{n}\rangle\}_{n=1}^N$. Define

$$|n\rangle := U|\hat{n}\rangle \tag{830}$$

for $1 \leq n \leq N$.

The unitarity of U implies that these transformed vectors are also orthonormal and that

$$U = \sum_{n=1}^N |n\rangle\langle\hat{n}| \tag{831}$$

This shows that unitary operators can always be associated with changes of orthonormal bases. It is useful to contrast the transformation properties of basis vectors with the transformation properties of the components of vectors:

$$\bar{a}_m = \sum_{n=1}^N \langle\bar{m}|U|\bar{n}\rangle a_n \tag{832}$$

$$|\bar{m}\rangle = \sum_{n=1}^N |\bar{n}\rangle\langle\bar{m}|U|\bar{n}\rangle^* = \sum_{n=1}^N |\bar{n}\rangle\langle\bar{n}|U^\dagger|\bar{m}\rangle \tag{833}$$

which show that the basis vectors and components have transformation properties.

0.32 Lecture 32

In the discussion the on matrices I showed that the condition for a linear operator have an inverse was that the determinant of a matrix representation of the linear operator be non zero.

My definition of determinant appears to be basis dependent. Note however that for a linear operator A

$$A_{mn} = \langle m|A|n\rangle = \sum \langle m|\bar{k}\rangle\langle\bar{k}|A|\bar{l}\rangle\langle\bar{l}|n\rangle$$

$$=$$

$$\sum \langle m|U|k\rangle \bar{A}_{kl}\langle l|U^\dagger|n\rangle$$

$$= \sum_{kl} U_{mn}\bar{A}_{kl}U^\dagger_{ln} \tag{834}$$

It follows that

$$\det(A) = \det(U\bar{A}U^\dagger) = \det(U)\det(\bar{A})\det(U^\dagger) = \det(UU^\dagger)\det(\bar{A}) = \det(\bar{A}) \tag{835}$$

This shows that the determinant is invariant with respect to a change of basis.

I showed this assuming a unitary change of basis. If one does not care if the bases are orthonormal then U_{mn} can be replaced by any invertible matrix W

$$A_{mn} = \sum_{kl} W_{mn}\bar{A}_{kl}W^{-1}_{ln} \tag{836}$$

This will also have the same determinant. The proof is identical.
Tensors: Consider an inner product space.

Vectors and dual vectors can be written as

$$|a\rangle = \sum_n |n\rangle a_n \tag{837}$$

$$\langle a| = \sum_n a^*_N\langle n| \tag{838}$$

The scalar product can be expressed in terms of the components as

$$\langle b|a\rangle = \sum_{n=1}^N b^*_n a_n \tag{839}$$

In a different orthonormal basis this equation becomes

$$\langle b|a\rangle = \sum_{n=1}^{N} \bar{b}_n^* \bar{a}_n \tag{840}$$

which is equivalent to

$$\langle b|a\rangle = \sum_{n,m,k=1}^{N} \bar{b}_m^* U_{mk}^\dagger U_{kn} a_n \tag{841}$$

The relevant observation is that vector components transform with U_{kn} under change of basis while components of dual vectors transform with $U_{kn}^\dagger$ under change of basis.

Equation (**??**) can be written in the form

$$\langle b|a\rangle = \sum_{n,m,k=1}^{N} \bar{b}_m^* U_{mk}^\dagger U_{kn} a_n = \sum_{n,m,k=1}^{N} U_{km}^* U_{kn} \bar{b}_m^* a_n \tag{842}$$

It is possible to consider objects that transform like products of n vectors and m dual vectors. An object

$$T_{b_1 \cdots b_m}^{a_1 \cdots a_n} \tag{843}$$

is called a rank (m,n) tensor if it has the following transformation properties under the transformation U

$$\bar{T}_{b_1 \cdots b_m}^{a_1 \cdots a_n} = \sum{}' U_{a_1 a_1'} \cdots U_{a_n a_n'} U_{b_1 b_1'}^* \cdots U_{b_m b_m'}^* T_{b_1 \cdots b_m}^{a_1 \cdots a_n} \tag{844}$$

The way that I have constructed tensors, when a vector index and dual vector index are set to be equal and summed, the resulting sum is invariant. The remaining indices describe a tensor of rank $(n-1, m-1)$. This operation is called a contraction.

Tensors normally appear in a more abstract setting. Typically one has a vector space, a quadratic form, and a dual vectors space.

The scalar product is generalized to have the form

$$Q(a,b) = \sum_{m,n} a_m^* Q_{mn} b_n \tag{845}$$

The unitary transforms are replaced by linear transformations that leave the quadratic form Q invariant

$$\sum_{c,d} W^{\dagger}_{ac} Q_{cd} W_{db} = Q_{ab} \tag{846}$$

These transformations form a group under matrix multiplication

$$W'_{ab} = \sum_c W^1_{ac} W^2_{cb} \tag{847}$$

The vectors may be real or complex. A rank (mn) tensor associated with this quadratic form transform like

$$\bar{T}^{a_1 \cdots a_n}_{b_1 \cdots b_m} = \sum{}' W_{a_1 a'_1} \cdots W_{a_n a'_n} W^*_{b_1 b'_1} \cdots W^*_{b_m b'_m} T^{a'_1 \cdots a'_n}_{b'_1 \cdots b'_m} \tag{848}$$

The important feature of tensors is that the zero tensor is invariant with respect to the transformation W_{ab}. This means that if an equation implies that two tensor quantities are equal, then the transformed tensors are also equal.

Tensors of the same rank form a vector space. They can be added and multiplied by scalars.

Example: The vector space is a real four-dimensional vector space. The quadratic form is the diagonal matrix

$$\eta_{ab} := \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{849}$$

and the transformation are real transformations that preserve is quadratic form

$$\Lambda_{ac} \Lambda_{bd} \eta_{cd} = \eta_{ab} \tag{850}$$

These transformations are called Lorentz transformations.

The Maxwell field strength tensor

$$F_{ab} = \frac{\partial A_a}{\partial x_b} - \frac{\partial A_b}{\partial x_a} = \begin{pmatrix} 0 & -E_1 & -E_2 & -E_3 \\ E_1 & 0 & -B_3 & B_2 \\ E_2 & B_3 & 0 & -B_1 \\ E_3 & -B_2 & B_1 & 0 \end{pmatrix} \tag{851}$$

is rank two antisymmetric tensor. The structure of this tensor arises because it involves derivatives of a vector. Tensors associated with stress and strain and dielectric strength are also related to Taylor expansions.

Another well-known tensor is the moment of inertia tensor of a system of point masses:

$$I_{ij} = \sum_k m_k (\vec{x}_k \cdot \vec{x}_k \delta_{ij} - x_k^i x_k^j) \tag{852}$$

In this case the quadratic form is δ_{ij}, the coordinates are real, and the transformations W_{ij} are three dimensional orthogonal matrices.

Tensors appear in many situations. In general they are tied to a quadratic form and a group of transformations that leave the form invariant. This setting naturally leads to tensors with indices that transform like vectors (contravariant indices) and indices that transform like dual vectors (covariant indices.)

Normally the transformations W_{ab} and W^*_{ab} acting on a vector space of dimension d are the fundamental and conjugate representations of the group that preserve the d-dimensional quadratic form Q_{ab}. Tensors can be thought of as vectors in a higher dimensional vector space that transform under a higher dimensional representations of the same group.

0.33 Lecture 33

The Cayley Hamiltonian Theorem

Let A be a linear operator represented by a $N \times N$ matrix with components A_{mn} in an orthonormal basis.
Definition: The characteristic polynomial of the linear operator A is the function

$$\phi(\lambda) := \det(\lambda I - A). \tag{853}$$

1. Since I have shown that the determinant is independent of the choice of basis, it follows that this function depends on the operator A, independent of its matrix representation.

2. The definition of the determinant implies that $\phi(\lambda)$ is a polynomial of degree N.

Theorem 31.1 Cayley Hamiltonian Theorem:

$$\phi(A) = 0. \tag{854}$$

The Cayley-Hamilton is one of the most important theorems in linear algebra.

To prove this recall

$$\det(M)\delta_{ij} = \sum_{k=1}^{N} \frac{\partial \det(M)}{\partial M_{ik}} M_{jk}. \tag{855}$$

Applying this result to the characteristic polynomial

$$\phi(\lambda)\delta_{ij} = \sum_{k=1}^{N} \frac{\partial \det(\phi(\lambda))}{\partial M_{ik}} (\lambda I - A)_{jk}. \tag{856}$$

Replacing λ by A in this expression gives

$$\phi(A)\delta_{ij} = \sum_{k=1}^{N} \frac{\partial \det(\phi(\lambda = A))}{\partial M_{ik}} (A - A)_{jk} = 0. \tag{857}$$

While the proof of this result is elementary, it has strong consequences.

Theorem 31.2 Let $r > N$ then A^r can be expressed as an polynomial of degree $N - 1$.

To see this note that the Cayley Hamilton theorem implies

$$\phi(A) = 0 = A^N + \sum_{k=0}^{N-1} c_k A^k \tag{858}$$

which can be written as

$$A^N = -\sum_{k=0}^{N-1} c_k A^k. \tag{859}$$

By induction assume that

$$A^s = \sum_{k=0}^{N-1} d_k A^k \tag{860}$$

Then

$$A^{s+1} = \sum_{k=0}^{N-2} d_k A^{k+1} + d_{N-1} A^N =$$

$$\sum_{k=1}^{N-1} d_{k-1} A^k - \sum_{k=0}^{N-1} d_{N-1} c_k A^k = \sum_{k=1}^{N-1} (d_{k-1} - d_{N-1} c_k) A^k \tag{861}$$

This implies

$$A^{s+1} = \sum_{k=0}^{N-1} d'_k A^k \tag{862}$$

This completes the proof of the theorem.

It shows that any function of the operator A that can be approximated by a polynomial can in fact be represented by a finite degree polynomial. The order of the polynomial depends on the dimension of the space.

Since $\phi(\lambda)$ is a polynomial of degree N, the fundamental theorem of algebra implies that

$$\phi(\lambda) = \prod_{l=1}^{L} (\lambda - \lambda_l)^{r_l} \tag{863}$$

where $\sum_{l=1}^{L} r_l = N$, and the λ_l are isolated zeros of $\phi(\lambda)$.

The function

$$\frac{1}{\phi(\lambda)} \tag{864}$$

has isolated poles at $\lambda = \lambda_l$.

$$\frac{1}{\phi(\lambda)} = \frac{1}{\prod_{l=1}^{L}(\lambda - \lambda_l)^{r_l}} \tag{865}$$

I define functions of the form

$$f(\lambda) = \sum_{l=1}^{L}(\sum_{k=1}^{r_l} \frac{c_{kl}}{(\lambda - \lambda_l)^k}) \tag{866}$$

Note that if γ_l is a positively oriented curve abound λ_l that does not contain any of the other isolated singularities then

$$c_{kl} = \frac{1}{2\pi i}\int_{\gamma_i} f(\lambda)(\lambda - \lambda_l)^{k-1} d\lambda \tag{867}$$

if I define these constants by

$$c_{kl} = \frac{1}{2\pi i}\int_{\gamma_i} \frac{1}{\phi(\lambda)}(\lambda - \lambda_l)^{k-1} d\lambda \tag{868}$$

then

$$g(\lambda) = \frac{1}{\phi(\lambda)} - f(\lambda) \tag{869}$$

is entire by Morerra's theorem (integrals around the λ_l vanish and it is analytic everywhere else). It is constant because it is bounded. The constant must be zero because the constant function vanishes at $\lambda = \lambda_l$.

Thus I can write

$$\frac{1}{\phi(\lambda)} = \sum_{l=1}^{L}(\sum_{k=1}^{r_l} \frac{c_{kl}}{(\lambda - \lambda_l)^k}) \tag{870}$$

$$c_{kl} = \frac{1}{2\pi i}\int_{\gamma_i} \frac{1}{\phi(\lambda)}(\lambda - \lambda_l)^{k-1} d\lambda \tag{871}$$

I define the polynomials

$$f_l(\lambda) = \sum_{k=1}^{r_l} c_{kl}(\lambda - \lambda_l)^{r^l - k} \tag{872}$$

then

$$\frac{1}{\phi(\lambda)} = \sum_{l=1}^{L} f_l(\lambda) \frac{1}{(\lambda - \lambda_l)^{r_l}} \tag{873}$$

Now that I have this representation I have

$$1 = \phi(\lambda)\frac{1}{\phi(\lambda)} = \sum_{l=1}^{L} f_l(\lambda)\frac{1}{(\lambda-\lambda_l)^{r_l}}\prod_{k=1}^{L}(\lambda-\lambda_k)^{r_k} =$$

$$\sum_{l=1}^{L} f_l(\lambda)\prod_{k=1,k\neq l}^{L}(\lambda-\lambda_k)^{r_k} = \sum_{l=1}^{L}\phi_l(\lambda) \tag{874}$$

where

$$\phi_l(\lambda) = f_l(\lambda)\prod_{k=1,k\neq l}^{L}(\lambda-\lambda_k)^{r_k} \tag{875}$$

is a polynomial in */lambda*.

Inserting A for λ in this polynomial gives

$$I = \sum_{l=1}^{L}\phi_l(A) \tag{876}$$

Next I investigate properties of the polynomials $\phi_l(A)$:

Theorem 31.3: $\phi_m(A)\phi_l(A) = 0 \qquad l \neq m$

To prove this observe

$$\phi_l(A)\phi_m(A) = f_l(A)f_m(A)\prod_{k\neq l,m}^{L}(A-\lambda_k)\prod_{t=1}^{L}(A-\lambda_t) \tag{877}$$

The term $\prod_{t=1}^{L}(A-\lambda_t) = \phi(A) = 0$ by the Cayley Hamilton theorem.

Theorem 31.4: $\phi_m(A)\phi_m(A) = \phi_m(A)$

To prove this use theorem 31.3 and (876) to write

$$\phi_l(A)\phi_l(A) = \phi_l(A)\sum_{k=1}^{L}\phi_k(A) = \phi_l(A)I = \phi_l(A) \tag{878}$$

Pick any vector $|\xi\rangle$ and consider

$$\phi_m(A)|\xi\rangle := |\lambda_m\rangle, \tag{879}$$

then

$$\phi_m(A)|\lambda_m\rangle := |\lambda_m\rangle \tag{880}$$

$$\phi_n(A)|\lambda_m\rangle = 0 \qquad n \neq m \tag{881}$$

Define S_k as the linear subspace of vectors spanned by all vectors of the form $\phi_k(A)|\xi\rangle$. By construction all vectors in this subspace satisfy

$$\phi_k(A)|\chi\rangle := |\chi\rangle \tag{882}$$

$$\phi_n(A)|\chi\rangle := 0 \qquad n \neq k \tag{883}$$

On the other hand any vector can be expressed as

$$|v\rangle = I|v\rangle = \sum_{l=1}^{L} \phi_l(A)|v\rangle \tag{884}$$

as a sum of vectors in each of the subspaces S_k

Definition: A generalized eigenvector $|v\rangle$ of a linear operator A of order r with eigenvalue λ is a vector satisfying

$$(\lambda I - A)^r |v\rangle = 0 \tag{885}$$

and

$$(\lambda I - A)^{r-1} |v\rangle \neq 0 \tag{886}$$

Theorem 31.5: Every vector $|\xi\rangle \in S_k$ is a generalized eigenvector of order $r \leq r_k$ of A with eigenvalue λ_k

Proof:

$$|\xi\rangle \in S_k \qquad \Rightarrow \qquad |\xi\rangle = \phi_k(A)|\chi\rangle \tag{887}$$

$$(\lambda_k I - A)^{r_k} \phi_k(A)|\chi\rangle = f_k(A)\phi(A)|\chi\rangle = 0 \tag{888}$$

by the Cayley Hamilton theorem.

Theorem 31.6: If $r_k = 1$ then

$$|\xi\rangle = \phi_k(A)|\chi\rangle \tag{889}$$

is an eigenvector of A with eigenvalue λ_k.

This follows from Theorem 30.5 by setting $r_k = 1$.

Theorem 31.7: If $|\chi\rangle \in S_k$ then

$$|\xi\rangle, \quad (\lambda_k I - A)|\xi\rangle, \quad \cdots \quad (\lambda_k I - A)^{r_k - 1}|\xi\rangle \tag{890}$$

are either 0 or linearly independent.

To prove this let l be the smallest integer such that $(\lambda_k I - A)^l |\xi\rangle \neq 0$ and consider

$$\sum_{m=1}^{l} \alpha_m (\lambda_k I - A)^m |\xi\rangle = 0 \tag{891}$$

Multiplying by $(\lambda_k I - A)^l$ implies $\alpha_0 = 0$. Next multiply by $(\lambda_k I - A)^{l-1}$ to show $\alpha_1 = 0$. This can be repeated $l+1$ times to show $\alpha_0 = \alpha_1 = \cdots \alpha_l = 0$.

Putting everything together gives the following result:

Theorem 31.7: Any vector $|v\rangle$ can be expanded as a linear combination of the generalized eigenvectors of the linear operator A.

The expansion is

$$|v\rangle = \sum_{l=1}^{L} \phi_l(A) |v\rangle \tag{892}$$

Theorem 31.8: Let A be a normal operator. Then all generalized eigenvectors of A have rank 1.

Proof: Assume

$$(\lambda I - A)^m |\xi\rangle \neq 0 \qquad m \geq 1 \tag{893}$$

Then

$$\langle \xi | (\lambda I - A)^{\dagger m} (\lambda I - A)^m |\xi\rangle > 0 \tag{894}$$

This means that

$$(\lambda I - A)^{\dagger m} (\lambda I - A)^m |\xi\rangle \neq 0 \tag{895}$$

which gives

$$\langle \xi | (\lambda I - A)^{\dagger m} (\lambda I - A)^m (\lambda I - A)^{\dagger m} (\lambda I - A)^m |\xi\rangle = \tag{896}$$

$$\langle \xi | (\lambda I - A)^{\dagger 2m} (\lambda I - A)^{2m} |\xi\rangle > 0 \tag{897}$$

This can be repeated until $2^k m > r$ which gives a contradiction.

If A is unitary, Hermitian, or normal it follows that

$$|\xi\rangle = \sum_{l=1}^{L} \phi_l(A) |\xi\rangle \tag{898}$$

with

$$A \phi_l(A) |\xi\rangle = \lambda_l \phi_l(A) |\xi\rangle \tag{899}$$

This shows that every vector in the vector space is a linear combination of eigenvectors of A. Since $\sum_{k=1}^{L} r_k = N$ is means that each of the subspaces S_k

is r_k dimensional, or has r_k linearly independent eigenvectors with eigenvalue λ_k. Without loss of generality these can be chosen to be orthonormal using the Gram Schmid method.

This proves the theorem:

Theorem 31.9 Every normal operator has a complete set of eigenstates. These can always be chosen to be orthonormal.

If A is normal then A can be expressed as the following polynomial in A:

$$A = AI = \sum_{l=1}^{L} A\phi_l(A) = \sum_{l=1}^{L} \lambda_l \phi_l(A) \tag{900}$$

Since

$$A^m \phi_l(A) = \lambda_l A^{m-1} \phi_l(A) = \cdots = \lambda_l^m \phi_l(A) \tag{901}$$

it follows that it

$$f(A) = \sum_{m=0}^{\infty} c_m A^m \tag{902}$$

is a norm convergent series then

$$f(A) = f(A)I = f(A) \sum_{l=1}^{L} \phi_l(A) =$$

$$\sum_{l=1}^{L} \sum_{m=0}^{\infty} c_m \lambda_l^m \phi_l(A) \sum_{l=1}^{L} f(\lambda_l) \phi_l(A) \tag{903}$$

This expresses $f(A)$ as a polynomial in A.

Example: Let A be normal linear operator on an 3 dimensional space with distinct eigenvalues λ_1, λ_2, λ_3. It is an elementary exercise to show that in this simple case

$$\phi_1(A) = \frac{(A-\lambda_2)(A-\lambda_3)}{(\lambda_1-\lambda_2)(\lambda_1-\lambda_3)} \tag{904}$$

$$\phi_2(A) = \frac{(A-\lambda_1)(A-\lambda_3)}{(\lambda_2-\lambda_1)(\lambda_2-\lambda_3)} \tag{905}$$

$$\phi_3(A) = \frac{(A-\lambda_1)(A-\lambda_2)}{(\lambda_3-\lambda_1)(\lambda_3-\lambda_2)} \tag{906}$$

Using this in the above formula gives

$$e^A = e^{\lambda_1}\phi_1(A) + e^{\lambda_2}\phi_2(A) + e^{\lambda_3}\phi_3(A) \tag{907}$$

or more explicitly

$$e^A = e^{\lambda_1}\frac{(A-\lambda_2)(A-\lambda_3)}{(\lambda_1-\lambda_2)(\lambda_1-\lambda_3)}+$$
$$e^{\lambda_2}\frac{(A-\lambda_1)(A-\lambda_3)}{(\lambda_2-\lambda_1)(\lambda_2-\lambda_3)}$$
$$+e^{\lambda_3}\frac{(A-\lambda_1)(A-\lambda_2)}{(\lambda_3-\lambda_1)(\lambda_3-\lambda_2)} \tag{908}$$

Previously this was defined in terms of an infinite series of matrix products of all orders. In this representation the only products that appear are I, A, A^2.

This shows that whenever a normal operator has N distinct eigenvalues that

$$\phi_l(A) = \frac{\prod_{k\neq l}^N (A-\lambda_k)}{\prod_{n\neq l}^N (\lambda_l-\lambda_n)} \tag{909}$$

and

$$f(A) = \sum_{l=1}^N f(\lambda_l)\frac{\prod_{k\neq l}^N (A-\lambda_k)}{\prod_{n\neq l}^N (\lambda_l-\lambda_n)} \tag{910}$$

The origin of these results is the Cayley Hamilton theorem, which allows one to reduce any function of A to a polynomial of degree A.

Next I consider the case that A is any $N \times N$ matrix. As discussed above, the characteristic polynomial still has N complex roots. For each distinct eigenvalue it is possible to construct the operator $\phi_l(A)$. The range of each of these operators is a subspace S_m of the N dimensional vector space.

Since

$$\sum_{l=1}^L \phi_l(A) = I \tag{911}$$

any vector cah be witten as

$$|v\rangle = \sum_{l=1}^L \phi_l(A)|v\rangle \tag{912}$$

The vectors $\phi_l(A)|v\rangle$ are linearly independent becasue if

$$0 = \sum_l a_l \phi_l(A)|v\rangle \tag{913}$$

and I multiply by $\phi_m(A)$ it follows that

$$0 = \sum_l a_l \phi_m(A)\phi_l(A)|v\rangle = a_m \phi_m(A)|v\rangle \tag{914}$$

which requires that $a_m = 0$ if $\phi_m(A)|v\rangle \neq 0$

I construct a basis by first constructing basis on each of the subspaces S_m. Fix m and let r be the largest integer such that

$$(\lambda_m - A)^r |\xi\rangle \neq 0 \tag{915}$$

for some $|\xi\rangle \in S_m$. If follows that

$$|\xi_{m10}\rangle = |\xi\rangle$$

$$|\xi_{m11}\rangle = (A - \lambda_m)|\xi\rangle$$

$$\vdots$$

$$|\xi_{m1r_1}\rangle = (A - \lambda_m)^r |\xi\rangle$$

are all linearly independent and span a subspace of S_m. Next find r_2 to be the largest integer such that

$$(A - \lambda_m)^{r_2} |\xi\rangle \neq 0$$

for $|\xi\rangle$ in S_m but not in the subspaces constructed above. This leads to $r_2 + 1$ new independent vectors which are labeled

$$|\xi_{m2n}\rangle \qquad 0 \leq r \leq r_2$$

This process can be repeated until we exhaust all of the independent vectors in S_m. This can be repeated for all distinct eigenvalues.

This leads to a sert of basis vectors with labels $|mnl\rangle$ By construction

$$A|mnl\rangle = \lambda_m |mnl\rangle + |mnl + 1\rangle \tag{916}$$

for $l + 1 \leq r_{n_m}$. Recall that in a given basis

$$A|n\rangle = \sum_m |m\rangle A_{mn}$$

so it follows that in this new basis

$$A_{nm} = \delta_{mn}\lambda_n + 1\eta_{n+1,n} \tag{917}$$

where $\eta_{n+1,n}$ is zero or 1.

In this basis A has eigenvalues along the diagonal. When the same eigenvalue is repeated the may be small blocks where there may be 1's directly below the diagonal.

This is called the Jordan canonical form of the matrix. If $\{|\bar{n}\rangle\}$ is any other basis then

$$|\bar{n}\rangle = |m\rangle S_{mn} \qquad |n\rangle = |\bar{m}\rangle S^{-1}_{mn} \tag{918}$$

for some matrix S_{mn}. It follows that

$$A|\bar{m}\rangle = \sum_n |\bar{n}\rangle \bar{A}_{nm} =$$

$$A|k\rangle S_{km} = \sum_l \sum_n |l\rangle S_{ln}\bar{A}_{nm} \tag{919}$$

or

$$A|k\rangle = \sum_l \sum_n |l\rangle S_{ln}\bar{A}_{nm}S^{-1}_{mk} =$$

$$\sum_l |l\rangle A_{lk} \tag{920}$$

Identifying coefficients of independent basis vectors gives

$$A_{lk} = \sum_{nm} S_{ln}\bar{A}_{nm}S^{-1}_{mk} \tag{921}$$

This shows that any matrix can be brought into Jordan canonical form by a suitable change of basis.

Note that in the general case the matrix is not unitary, it is just an invertible matrix.

When the matrices are normal then the Jordan matrix is diagonal and the basis vectors can be chosen to be orthonormal.

In this case if $|\bar{n}\rangle$ is also an orthonormal basis then

$$|\bar{n}\rangle = \sum_m |m\rangle U_{mn} \qquad \langle m|\bar{n}\rangle = U_{mn} \qquad U^{-1}_{mn} = \langle \bar{m}|n\rangle \tag{922}$$

and the operator can be brought into diagonal form by a unitary tranfsformation. The entries on the diagonal are just eigenvalues of the operator.

When two matrices commute:

$$[A, B] = 0 \tag{923}$$

then polynomials in these operators also commute. If follows that

$$\phi_l(A)\phi_m(B) = \phi_m(B)\phi_l(A) \tag{924}$$

Applying this operator to an arbitrary vectors $|xi\rangle$ gives

$$\phi_l(A)\phi_m(B)|\xi\rangle \tag{925}$$

If this vector is not zero, then it follows that

$$(\lambda_{al} - A)^{r_{al}}\phi_l(A)\phi_m(B)|\xi\rangle = 0 \tag{926}$$

$$(\lambda_{bm} - B)^{r_{bm}}\phi_l(A)\phi_m(B)|\xi\rangle(\lambda_{bm} - B)^{r_{bm}}\phi_m(B)\phi_l(A)|\xi\rangle = 0 \tag{927}$$

which means that this vector is a generalized eigenvector of both A and B. We also have

$$I = \sum_{l=1}^{L}\sum_{m=1}^{M}\phi_l(A)\phi_m(B) \tag{928}$$

which means that commuting operators have complete sets of simultaneous generalized eigenstates.

When A and B are normal the generalized eigenvectors become eigenvectors. The above result means there is a basis of simultaneous eigenvectors of both A and B.

If $|a\rangle$ is the only eigenvector of a normal operator A with eigenvalue a and B is a noraml operator satisfying $[B, A] = 0$ then $|a\rangle$ must be an eigenvector of B for some eigenvalue. This follows becasue

$$\phi_l(A) = \sum_m \phi_m(B)\phi_l(A) \neq 0 \tag{929}$$

which means that there is at least one value of m with $\phi_m(B)\phi_l(A) \neq 0$. SInce $\phi_l(A)$ is one dimensional, there can be no more than one value of m with this property.

Weyl Pairs and irreducibility

Note that if A is an $N \times N$ matrix the Cayley Hamilton theorem implies that any function of A can be expressed as a polynomial of degree N in A. If I consider the space of N×N matrices as a complex vector space, the Cayley Hamilton theorem implies that the powers of A span a subspace of this space of at most $N+1$ dimensions. On the other hand this vector space has N^2 dimensions. This means every operator cannot be written as a function of A. What we want to show next is that polynomials in carefully chosen pairs of pairs of matrices can be used to represent any matrix.

Begin by assuming that A is normal and has N distinct eigenvalues with orthonormal eigenvectors

$$A|a_n\rangle = a_n|a_n\rangle \tag{930}$$

where $1 \leq n \leq N$.

Next define the shift operator

$$U|a_n\rangle = |a_{n+1}\rangle \qquad n < N \tag{931}$$

$$U|a_N\rangle = |a_1\rangle \tag{932}$$

It follows that

$$U = \sum_{n=0}^{N-1} |a_{n+1}\rangle\langle a_n| + |a_1\rangle\langle a_N| \tag{933}$$

has the form a unitary change of basis.

By construction

$$U^N|a_n\rangle = |a_n\rangle \tag{934}$$

for all basis vectors so

$$U^N - 1 = 0. \tag{935}$$

Thus the characteristic polynomial is $\lambda^n - 1$ which has roots

$$\eta_n = e^{2\pi in/N} \qquad n = 1, 2, \cdots, N \tag{936}$$

or equivalently

$$\eta_n = e^{2\pi in/N} \qquad n = 0, 2, \cdots, N-1 \tag{937}$$

This means that

$$U^N - 1 = \prod_{n=1}^{N} (U - \eta_n) \tag{938}$$

If I pick a particular fixed value k of n it follows that

$$(U/\eta_k)^N - 1 = 0 = (U/\eta_k - 1)\sum_{l=0}^{N-1}(U/\eta_k)^l \tag{939}$$

where I have used the factorization

$$(x^n - 1) = (x - 1)(1 + x + x^2 \cdots + x^{n-1})$$

Next I show that the operator

$$\Pi_k := \frac{1}{N}\sum_{l=0}^{N-1}(U/\eta_k)^l \tag{940}$$

is an orthogonal projection operator. To see this note

$$\Pi_k^\dagger := \frac{1}{N}\sum_{l=0}^{N-1}(U^\dagger\eta_k)^l = \frac{1}{N}\sum_{l=0}^{N-1}(U/\eta_k)^{-l}$$

Multiply this by $(U/\eta_k)^N = 1$ to get

$$\frac{1}{N}\sum_{l=0}^{N-1}(U/\eta_k)^{N-l}$$

$$\frac{1}{N}\sum_{m=1}^{N}(U/\eta_k)^m = \Pi_k \qquad m = N - l$$

I also have

$$\Pi_k^2 = \frac{1}{N^2}\sum_{l,m=0}^{N-1}(U/\eta_k)^{l+m}$$

Clearly for each fixed value of l, $l+m$ goes from l to $l+N-1$ which means each power of U/η_k appears once. Since there are N values of l this sum is repeated N times giving

$$\Pi_k^2 = \frac{N}{N^2}\sum_{l=0}^{N-1}(U/\eta_k)^l = \Pi_k$$

where

$$(U - \eta_k)\Pi_k = 0$$

This means each Π_k projects on a different invariant subspace of U. Since there are N such operators with distinct eigenvalues, each of these subspaces must be one dimensional. This demonstrates that

$$\Pi_k = |u_k\rangle\langle u_k| \tag{941}$$

where

$$U|u_k\rangle = \eta_k|u_k\rangle$$

These equations determine the basis vectors $|u_k\rangle$ up to phase. To choose the phase of $|u_k\rangle$ note that

$$\langle a_N|u_k\rangle\langle u_k|a_N\rangle = \frac{1}{N}\sum_{l=0}^{N-1}\langle a_N|(U/\eta_k)^l|a_N\rangle = \frac{1}{N}$$

I choose the phase so

$$\langle a_N|u_k\rangle = \sqrt{\frac{1}{N}}$$

Using this choice gives

$$\langle a_N|u_k\rangle\langle u_k|a_m\rangle = \frac{1}{N}\sum_{l=1}^{N}\langle a_N|(U/\eta_k)^l|a_m\rangle$$

The only surviving term has $N = l + m$ or $l = N - m$ which gives

$$\frac{1}{\sqrt{N}}\langle u_k|a_m\rangle = \frac{1}{N}\eta_{km}$$

or

$$\langle u_k|a_m\rangle = \frac{1}{\sqrt{N}}\eta_{km}$$

What is distinctive about these two operators is the the magnitude of the inner product of any basis function in one set with any basis function in the other set is entirely independent of the choice of basis states. In all cases the relevant magnitude is $\frac{1}{\sqrt{N}}$.

It follows that

$$|a_m\rangle = \sum |u_k\rangle\langle u_k|a_m\rangle = \sum |u_k\rangle\frac{1}{\sqrt{N}}\eta_{km}$$

The next step is to define an adjoint shift operator on the eigenstates $|u_k\rangle$ of U. We define a new operator V by

$$\langle v_k|V = \langle v_{k+1}|$$

As in the case with U we have

$$V^N - 1 = 0,$$

with characteristic roots η_k. I also have

$$0 = (V - \eta_k)\sum_{l=1}^{N}(V/\eta_k)^l$$

with

$$|v_k\rangle\langle v_k| = \sum_{l=1}^{N}(V/\eta_k)^l$$

where I choose the phase on $\langle v_k|$ so

$$\langle u_N|v_k\rangle = \frac{1}{\sqrt{N}}$$

It follows that

$$\langle u_m|v_k\rangle\langle v_k|u_N\rangle = \langle u_m|\frac{1}{N}\sum_{l=1}^{N}(V/\eta_k)^l|u_N\rangle =$$

$$\langle u_m|\frac{1}{N}\sum_{l=1}^{N}(V/\eta_k)^l|u_N\rangle = \frac{1}{N}(1/\eta_k)^{N-m} = \frac{1}{N}\eta_{km}$$

This gives

$$\langle u_m|v_k\rangle = \frac{1}{\sqrt{N}}\eta_{km}$$

and

$$|v_m\rangle = \sum_k |u_k\rangle\langle u_k|v_m\rangle = \sum_k |u_k\rangle\frac{1}{\sqrt{N}}\eta_{km}$$

Comparing this to the corresponding expression for a_m

$$|a_m\rangle = \sum |u_k\rangle\frac{1}{\sqrt{N}}\eta_{km} = |v_m\rangle$$

From this it follows that $|v_k\rangle = |a_k\rangle$, or that we return to the original set of starting vectors.

Note that

$$UV|v_k\rangle = \eta_k|k+1\rangle$$

$$VU|v_k\rangle = V|v_{k+1}\rangle = \eta_{k+1}|v_{k+1}\rangle$$

Comparing these expressions gives the Weyl relations

$$VU = UV\eta = UVe^{\frac{2\pi i}{N}}$$

Note that

$$|u_m\rangle\langle u_n|$$

$$= |u_m\rangle\langle u_m|V^{n-m} = \frac{1}{N}\sum_{k=1}^{N} e^{\frac{-2\pi imk}{N}} U^k V^{n-m}$$

A general operator O can be written as

$$O = \sum_{mn} |u_m\rangle O_{mn} \langle u_n| =$$

$$\frac{1}{N}\sum_{k,m,n=1}^{N} O_{mn} e^{\frac{-2\pi imk}{N}} U^k V^{n-m}$$

This show that any operator has the general form

$$O = \sum_{nm=1}^{N} o(m,n) U^m V^n$$

Unlike the case of A^m the $N \times N$ operators $U^n V^m$ can be used to expand any $N \times N$ matrix.

The operators U and V are called Weyl pairs. In quantum mechanics the pairs U and V are called complementary operators. They are associated with a maximal mixing of eigenstates.

In many cases the Weyl pairs can be decomposed into smaller parts. Assume that the dimension N can factored as a product $N = KM$. Define

$$U_1 = U^M \qquad U_2 = U^K \tag{942}$$

Then $U_1^K = 1$ and $U_2^M = 1$. Since they are both powers of the same operator they commute. Define $\eta_{1m} = e^{2\pi im/M}$ and $\eta_{2k} = e^{2\pi ik/K}$

We label eigenstates of U using pairs of indices

$$|u_n\rangle \rightarrow |u_1 u_2\rangle \tag{943}$$

The eigenstates of U can be labeled by ordered pairs of integers $n \rightarrow (k, m)$ where $1 \leq k \leq K$ $1 \leq m \leq M$. In this notation

$$U_1|u_{1k_1}u_{2k_2}\rangle = \eta_{1k_1}|u_{1k_1}u_{2k_2}\rangle \tag{944}$$

$$U_2|u_{1k1}u_{2k_2}\rangle = \eta_{2k_2}|u_{1k1} + 1u_{2k_2}\rangle \tag{945}$$

Next I define V operators

$$\langle u_{1k_1}u_{2k_2}|V_1 = \langle u_{1k_1} + 1u_{2k_2}| \tag{946}$$

$$\langle u_{1k_1}u_{2k_2}|V_2 = \langle u_{1k_1} + 1u_{2k_2} + 1| \tag{947}$$

where the last basis vector is also mapped to the first in both of these expressions. These equations lead to

$$V_1^K - I = V_2^M - I = 0 \tag{948}$$

These characeristic polynomials imply that U_i and V_i have the same eigenvalues. We also have

$$V_2 V_2 = V_2 V_1 \tag{949}$$

so these operators have simultaneous eigenstates

$$|v_{1k1}v_{2k_2}\rangle \tag{950}$$

Appealing to the previous construction it also follows that

$$U_1|v_{1k1}v_{2k_2}\rangle = |v_{1k1} + 1v_{2k_2}\rangle \tag{951}$$

$$U_2|v_{1k1}v_{2k_2}\rangle = |v_{1k1} + 1v_{2k_2} + 1\rangle \tag{952}$$

The V operators are defined by

$$\langle u_{1k_1}u_{2k_2}|V_1 = \langle u_{1k_1} + 1u_{2k_2} \tag{953}$$

$$\langle u_{1k_1}u_{2k_2}|V_2 = \langle u_{1k_1} + 1u_{2k_2} + 1 \tag{954}$$

The standard construction gives

$$U_1^K = V_1^K = U_2^M = V_2^M = I \tag{955}$$

$$V_1U_1 = e^{\frac{2\pi i}{K}} U_1V_1 \qquad V_2U_2 = e^{\frac{2\pi i}{M}} U_2V_2 \tag{956}$$

Simple computations on the states gives

$$U_1U_2 = U_2U_1 \qquad V_1V_2 = V_2V_1 \qquad U_1V_2 = V_2U_1 \qquad V_1U_2 = U_2V_1 \tag{957}$$

For example

$$\langle u_{1k_1} u_{2k_2} | U_2V_2 = \eta_{1k_1} \langle u_{1k_1} u_{2k_2} + 1 = \langle u_{1k_1} u_{2k_2} | V_2U_2 \tag{958}$$

We see that the U operators commute with each other. They have common eigenvectors.

The decomposition above can be repeated if any of the smaller dimensions K or M admit a factorization into pairs of smaller numbers. This process can be continued until all U and V operators correspond to prime factors.

In the general case all of the U and V operators are needed as a basis for the most general operator. Specifically the dimension of the vector space can be factored into a product of prime numbers. Each of these primes appears twice in caclulating the dimension of the linear space of matrices on this space. One prime corresponds to a U and the other idenical prime corresponds to a corresponding V.

Thus the most general operator has the form.

$$F = \sum f(a_1, \cdots, a_k, b_1, \cdots, b_k) U_1^{a_1} \cdots U_k^{a_k} V_1^{b_1} \cdots V_k^{b_k} \tag{959}$$

Thus we see that the basic building blocks of any operatros are Weyl pairs of operators associated with prime demensions.

Moore Penrose Generalized Inverse

In our discussion of matrices the condition that a matrix has an inverse it the requirement that the determinant is non-vanishing. Sometimes computational method lead to instabilities when the determinant of a large matrix gets too small. It is often useful to have an algorithm for computing soemthing that becomes always exists and becomes the inverse when it exists.

A quantity with these properties is called the —colorblueMoore Penrose Generalized inverse. It A is a square matrix X is a matrix satisfying

$$AXA = A \tag{960}$$

$$XAX = X \tag{961}$$

$$AX = (AX)^\dagger \tag{962}$$

$$(XA) = (XA)^\dagger \tag{963}$$

First note this there is at most one X satisfying these equations. Let X_1 and X_2 satsify these equations. first note

$$AX_2 = (AX_2)^\dagger[(AX_1A)X_2)]^\dagger = [(AX_1)(AX_2)]^\dagger(AX_2)(AX_1) = (AX_2A)X_1 = AX_1 \tag{964}$$

Similarly

$$X_2A = (X_2A)^\dagger = (X_2AX_1A)^\dagger = (X_1A)(X_2A) = X_1(AX_2A) = X_1A \tag{965}$$

It now follows that

$$X_1 = X_1AX_1 = X_2AX_1 = X_2AX_2 = X_2 \tag{966}$$

which shows the uniqueness.

29:171 - Homework Assignment #10

1. In classical mechanics we can consider the space of functions $f(x,p)$ of coordinates and linear momentum as vectors in a vector space. The energy $E = E(x,p)$ of a system can be expressed the sum of a kinetic and potential energy, which is a function of coordinates and linear momentum.

   Define the operator A by

   $$(Af)(x,p) = \frac{\partial E}{\partial x}\frac{\partial f}{\partial p} - \frac{\partial E}{\partial p}\frac{\partial f}{\partial x}$$

   Show that A is a linear operator. We sometimes write

   $$Af = \{E, f\}_{p.b}$$

   which is called the Poisson Bracket of E and f

   Let $E = \frac{1}{2}(p^2+x^2)$ be the energy for a one dimensional simple harmonic oscillator with mass 1 and spring constant 1. Calculate

   $$Ax = \{E, x\}_{p.b}$$

   $$Ap = \{E, p\}_{p.b}$$

   Use these results to calculate

   $$x(t) = e^{-tA}x$$

   Show that this solution describes simple harmonic motion corresponding to an initial coordinate and momentum given by x and p

2. Let

   $$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \sigma_x = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \qquad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

   These matrices are called the Pauli spin matrices.

a. Let $\hat{n}$ be a real unit vector. Show that

$$\hat{n} \cdot \vec{\sigma} = \hat{n}_x \sigma_x + \hat{n}_y \sigma_y + \hat{n}_z \sigma_z$$

is a Hermitian matrix.

b. Show that $e^{\lambda \hat{n} \cdot \vec{\sigma}}$ is a positive operator for every real λ

3. Let N be a nilpotent operator satisfying $N^3 = 0$. Let $|v\rangle$ be an initial vector and define the polynomial in t

$$|v(t)\rangle = |v\rangle + tN|v\rangle + \frac{t^2}{2!} N^2 |v\rangle$$

Show that $|v(t)\rangle$ satisfies the differential equation

$$\frac{d|v(t)\rangle}{dt} := \lim_{\epsilon \to 0} \frac{|v(t+\epsilon)\rangle - |v(t)\rangle}{\epsilon} = N|v(t)\rangle$$

4. Show that three vectors, $\vec{a}, \vec{b}, \vec{c}$, in a three dimensional space are linearly dependent if and only if the component vectors satisfy

$$\vec{a} \cdot (\vec{b} \times \vec{c}) = 0$$

5. Show that $\det(A) = \det(A^T)$ where $A^T_{mn} = A_{nm}$

6. Complex numbers are vectors in a one-dimensional vector space. Positive operators on this space are given by multiplying by a real positive number. Let $A = 1/3$. Approximate the positive square root of this operator using the method used in class:

$$C = 1 - A = 2/3 \qquad X = 1 - \sqrt{A}$$

$$X_n = \frac{1}{2}(C + X_{n-1}^2) \qquad X_0 = C/2 \qquad X = \lim_{n \to \infty} X_n$$

Compare your approximation to what you get using a calculator.

29:171 - Homework Assignment #12

Consider the matrix

$$M := \begin{pmatrix} 1 & -i & 0 \\ i & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

1. Find the characteristic polynomial of M
2. Find the eigenvalues of M
3. Find the polynomials $\phi_i(\lambda)$
4. Calculate $\phi_i(M)$.
5. Let $|v\rangle$ be any vector such that $\phi_i(M)|v\rangle \neq 0$. Show that

   $$M\phi_i(M)|v\rangle = \lambda_i \phi_i(M)|v\rangle$$

   where λ_i is the i-th eigenvalue.
6. Calculate $\sin(M)$
7. Show that $\sum_i \phi_i(M) = I$
8. Find a similarity transform that diagonalizes M
9. Find M^{-1}